

**SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS BASADA EN MINERÍA DE DATOS  
PARA APOYAR LA TOMA DE DECISIONES EN EL PROCESO DE CIRUGIA DEL  
HOSPITAL DEPARTAMENTAL UNIVERSITARIO SANTA SOFIA DE CALDAS, EN LA  
CIUDAD DE MANIZALES**



Wilson Alejandro Rojas Calvo

Universidad Autónoma de Manizales

Maestría en Gestión y Desarrollo de Proyectos de Software

Manizales, junio de 2016

**SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS BASADA EN MINERÍA DE DATOS  
PARA APOYAR LA TOMA DE DECISIONES EN EL PROCESO DE CIRUGIA DEL  
HOSPITAL DEPARTAMENTAL UNIVERSITARIO SANTA SOFIA DE CALDAS, EN LA  
CIUDAD DE MANIZALES**

INFORME FINAL

Proyecto Tesis de Maestría

Wilson Alejandro Rojas Calvo

Director. Msc. Javier Hernández Cáceres

Universidad Autónoma de Manizales

Maestría en Gestión y Desarrollo de Proyectos de Software

Manizales, junio de 2016

## Dedicatoria

A mis padres Mario y Lourdes  
A mi esposa Claudia,  
A mis hermosas hijas Sofi y Mafe

## Agradecimientos

A Dios y mi familia, por todo el tiempo que me han obsequiado para la culminación de este proyecto.

A la ESE Hospital Departamental Universitario Santa Sofía de Caldas, por permitirme realizar el presente trabajo de investigación y facilitarme los recursos técnicos necesarios para alcanzar los objetivos.

A mi colega Fabio López, experto en Sistemas de Información Hospitalario, por su aporte y apoyo constante durante la realización del proyecto.

A los líderes y coordinadores de procesos que aportaron con su actitud y conocimiento para lograr las metas trazadas

Al asesor de proyecto Msc. Javier Hernández Cáceres por su aporte durante la ejecución del proyecto.

## TABLA DE CONTENIDO

INTRODUCCIÓN.....	14
1. REFERENTE CONTEXTUAL .....	17
1.1. Descripción del área problemática .....	17
1.2. Antecedentes.....	18
1.2.1 Conclusiones sección de antecedentes .....	23
1.3. Justificación .....	25
1.4. Formulación del problema.....	27
1.5. Objetivos.....	28
1.5.1. Objetivo General.....	28
1.5.2. Objetivos Específicos .....	28
1.6. Alcances y limitantes .....	29
1.6.1. Unidad de Análisis y Muestra a utilizar .....	29
1.6.2. Confidencialidad de los datos.....	29
1.6.3. Herramientas de software Candidatas .....	31
1.7. Resultados esperados.....	32
2. ESTRATEGIA METODOLÓGICA .....	34
2.1. Tipo de Estudio.....	34
2.2. Metodologías para proyectos de minería de datos .....	35
2.2.1. CRISP-DM (Cross Industry Standard Process for Data Mining).....	35
2.2.2. SEMMA (Sample – Explore – Modify – Model – Assess) .....	36
2.2.3. KDD ( <i>Knowledge Discovery and Data Mining</i> ) .....	37
2.2.4. Metodología Catalyst (P3TQ).....	38
2.2.5. Metodología Berry y Linoff .....	40
2.3. Comparación de metodologías de minería de datos .....	40
2.4. Selección de la metodología a implementar .....	42
2.5. Descripción metodología CRISP-DM.....	42
2.5.1. Fases de la metodología CRISP-DM .....	43
2.5.1.1. Comprensión del negocio.....	43
2.5.1.2. Comprensión de los datos.....	44
2.5.1.3. Preparación de los datos.....	46
2.5.1.4. Modelado .....	47
2.5.1.5. Evaluación .....	49
2.5.1.6. Despliegue de la Solución .....	50
2.6. Pruebas .....	52
2.7. Presupuesto .....	54
3. REFERENTE TEÓRICO .....	55
3.1 Información en el Sector Salud. Marco Legal .....	55
3.1.1. Tipos de Sistemas de Información .....	56
3.1.2. Sistemas de Información en las Instituciones Prestadoras de Servicios de Salud ..58	
3.2. Conceptos sobre almacenes de datos .....	60
3.2.1. Data warehouse .....	60
3.2.2. Data mart.....	61
3.2.3. OLTP .....	62
3.2.4. OLAP (On line Analytical Processing).....	62
3.2.5. OLAP vs OLTP .....	65
3.3. Modelos Dimensionales .....	65
3.3.1. Esquema en Estrella .....	68
3.3.2. Esquema copo de nieve .....	69

3.3.3.	Esquema de Constelación de hechos.....	70
3.4.	Inteligencia de Negocios .....	70
3.4.1.	Arquitectura de una solución de Inteligencia de Negocios .....	72
3.4.1.1.	Arquitectura Howard Dresner.....	72
3.4.1.2.	Arquitectura propuesta por William Inmon .....	73
3.4.1.3.	Arquitectura de Data Marts independientes.....	74
3.4.1.4.	Arquitectura propuesta por Ralph Kimball.....	74
3.4.2.	Software utilizado para soluciones de Inteligencia de negocios.....	77
3.4.2.1.	DB2 WebQuery for IBMi .....	78
3.4.2.2.	Herramientas para Mobile Analytics.....	78
3.5.	Data Mining .....	79
3.5.1.	Tipos de técnicas de minería de datos.....	82
3.5.1.1.	No supervisadas o descriptivas.....	82
3.5.1.2.	Supervisadas o predictivas.....	83
3.5.2.	Técnica de Agrupamiento Clustering. Descripción. ....	84
3.5.2.1.	Medidas de Similitud .....	85
3.5.2.2.	Tipos de algoritmos para <i>Clustering</i> .....	87
3.5.2.3.	Algoritmo K-Medias (K-means) .....	88
3.5.2.4.	Pasos para realizar un análisis de Clúster .....	90
3.5.3.	Reducción de la Dimensionalidad .....	92
3.5.4.	<i>Data Mining</i> y los <i>Data Warehouse</i> .....	93
3.5.5.	Software para Data Mining.....	94
3.6.	Desarrollo de la metodología .....	95
3.6.1.	Fase I. Comprensión del Negocio .....	95
3.6.1.1.	Caracterización Proceso Atención Integral en el Servicio de Quirófanos.....	95
3.6.1.2.	Procedimientos que integran el proceso .....	96
3.6.1.3.	Objetivo Proceso de Quirófanos .....	97
3.6.1.4.	Objetivos de la minería de datos .....	97
3.6.1.5.	Requerimientos Funcionales.....	99
3.6.1.6.	Plan del proyecto.....	102
3.6.2.	Fase II. Comprensión de los Datos .....	102
3.6.2.1.	Informe de exploración de datos .....	103
3.6.2.2.	Informe de verificación de calidad de los datos .....	104
3.6.3.	Fase III Preparación de los datos.....	104
3.6.3.1.	Objetivo de la Fase.....	104
3.6.3.2.	SCRIPT SQL Modelo de datos .....	108
3.6.3.3.	DISEÑO DEL MODELO DE DATOS.....	109
3.6.4.	Fase IV. Modelado .....	120
3.6.4.1.	Objetivo.....	120
3.6.4.2.	RF-1 OLAP Cirugía .....	124
3.6.4.3.	RF-2. OLAP Financiero .....	130
3.6.4.4.	RF-5. Tablero Facturación.....	134
3.6.4.5.	RF-8. Reportes para dispositivos móviles.....	136
3.6.4.6.	RF-9. Personalización y exportación de informes.....	139
3.6.4.7.	RF-11. Clustering para obtener grupos de pacientes. ....	141
3.6.5.	Fase V. Evaluación.....	160
3.6.5.1.	Objetivo.....	160
3.6.5.2.	Evaluación de los resultados.....	163
3.6.5.3.	Proceso de revisión .....	164
3.6.5.4.	Determinación de los pasos siguientes.....	166
3.6.6.	Fase VI. Despliegue .....	167
3.6.6.1.	Objetivo.....	167
3.7.	Pruebas .....	168
3.7.1.	Plan de Pruebas .....	168

3.8.	Análisis de resultados.....	173
4.	CONCLUSIONES.....	174
5.	RECOMENDACIONES .....	176
6.	TRABAJO FUTURO.....	177
7.	DIVULGACIÓN DE RESULTADOS .....	178
	BIBLIOGRAFIA.....	179

## LISTA DE FIGURAS

Figura 1 Logotipo Hospital Santa Sofía. Empresa objetivo del Estudio de Caso .....	29
Figura 2 Fases SEMMA. Fuente SAS .....	36
Figura 3. Fases de Metodología SEMMA. (Cattaneo et al, 2010) .....	37
Figura 4. Proceso de Descubrimiento de Conocimiento. Tomada de (Fayyad, 1996) .....	38
Figura 5. Fases de Metodología P3TQ .....	39
Figura 6. Interacción entre las cinco variables en metodología P3TQ .....	39
Figura 7. Fases de Metodología Berry y Linoff. ....	40
Figura 8. Metodologías utilizadas en proyectos de minería de datos 2014.....	41
Figura 9 Comparativa interrelaciones metodología SEMMA vs CRISP-DM.....	42
Figura 10 Modelo CRISP-DM.....	43
Figura 11 Metodología CRISP-DM. Fase I Comprensión del Negocio.....	43
Figura 12. Metodología CRISP-DM. Fase II. Comprensión de los datos .....	44
Figura 13. Metodología CRISP-DM. Fase III. Preparación de los datos .....	46
Figura 14. Metodología CRISP-DM. Fase IV Modelado. ....	47
Figura 15. Metodología CRISP-DM. Fase V. Evaluación. ....	49
Figura 16. Metodología CRISP-DM. Fase VI. Despliegue de la Solución. ....	50
Figura 17. Valoración cualitativa Soluciones de Soluciones de Inteligencia de Negocios. ....	53
Figura 18. Clasificación de los Sistemas de Información.....	56
Figura 19. Representación OLAP producción de cirugías.....	63
Figura 20. Representación gráfica de Modelo dimensional .....	66
Figura 21. Star Schema vs OLAP cube .....	67
Figura 22. Diagrama de Flujo, proceso de modelamiento dimensional.....	68
Figura 23. Esquema en Estrella (Modelo de datos Cirugía) .....	69
Figura 24. Esquema Copo de Nieve. (Modelo de datos Cirugía) .....	69
Figura 25. Esquema Constelación de Hechos. (Modelo de datos Cirugía).....	70
Figura 26. Arquitectura de Data Warehouse Corporativo (William Inmon).....	73
Figura 27. Arquitectura Data Mart Independientes .....	74
Figura 28. Elementos principales de la arquitectura DW/BI (Ralph Kimball).....	75
Figura 29. Análisis del cuadrante mágico de Gartner para 2014 “Business Intelligence” .....	77
Figura 30. Procedimiento de análisis de clustering.....	85
Figura 31. Ejemplo de dendograma para clustering jerárquico. ....	88
Figura 32. Ejemplo de distancias intra y entre clústeres .....	91
Figura 33. Ejemplo de conglomerados finales .....	91
Figura 34. Data mart para data mining extraído de data warehouse.....	94
Figura 35. Formatos de Descripción de datos .....	103
Figura 36a. Contenido tabla dimensional Hechos de cirugía. ....	106
Figura 37b. Contenido tabla dimensional Paciente.....	107
Figura 38. Creación de Esquema de Base de datos BI_SSofía .....	110
Figura 39. Data warehouse (tablas y vistas) para el desarrollo de la solución.....	111
Figura 40. Creación de la nueva solución de inteligencia de negocios (BI_SSofía).....	112
Figura 41. Proceso de creación de metadatas para el desarrollo de solución de BI.....	113

Figura 42. Listado de metadatos (Sinónimos Web Query) .....	114
Figura 43. Edición de sinónimos y atributos .....	114
Figura 44. Creación de relaciones entre tablas de hechos y dimensiones .....	115
Figura 45. Interfaz general vista de dimensiones .....	116
Figura 46. Modelo Conceptual de los datos.....	117
Figura 47. Diseño lógico de Data Warehouse. ....	118
Figura 48. Diseño físico de los Datos. ....	119
Figura 49. Proceso para la creación de un nuevo reporte en Web Query .....	121
Figura 50. Opciones de edición de reportes en Web Query.....	122
Figura 51. Framework DB2 Web Query Developer Workbench .....	123
Figura 52. Semaforización de atributos .....	123
Figura 53. Menú principal Solución de Inteligencia de Negocios .....	124
Figura 54. OLAP Cirugía. Ejecución del Informe para todos los años.....	125
Figura 55. OLAP Cirugía. Ejecución del Informe aplicando filtro para los años 2014-2015. ....	126
Figura 56. Dimensiones OLAP Cirugía. ....	127
Figura 57. OLAP Cirugía. Granularidad. Funcionalidad drill down. ....	128
Figura 58. Personalización atributos OLAP .....	129
Figura 59. Ejecución informe OLAP Financiera .....	130
Figura 60. Drill Down OLAP Financiera .....	131
Figura 61. Drill Down y agregación de atributos OLAP Financiera.....	132
Figura 62. Nivel de detalle Regimen – EPS. OLAP Financiera .....	133
Figura 63. Nivel de detalle seguimiento a glosas. OLAP Financiera .....	133
Figura 64. Tablero Monitoreo Facturación Ejemplo 1(2013-2015 todos los trimestres).....	134
Figura 65. Tablero Monitoreo Facturación Ejemplo 2(2013-2015 último trimestre) .....	135
Figura 66. Tablero Monitoreo Facturación Ejemplo 3 (Función PIVOT).....	135
Figura 67. Tablero Monitoreo Facturación Ejemplo 4 (Función PIVOT).....	136
Figura 68. Framework de trabajo ROAMBI ANALYTICS.....	137
Figura 69. Reporte Indicadores Históricos de producción 2193. Dispositivo Móvil. ....	137
Figura 70. Reporte Facturación de Cirugías. Dispositivo Móvil. ....	138
Figura 71. Reporte Indicadores de Cirugía. Dispositivo Móvil. ....	138
Figura 72. Generación de informe OLAP para exportar. ....	139
Figura 73. Funcionalidad de exportación de Resultados en DB2 Web Query .....	140
Figura 74. Vista Mitable para aplicación de Clustering.....	144
Figura 75. Archivo Arff para minería en WEKA.....	144
Figura 76. Exploración atenciones por grupos de cirugía.....	148
Figura 77. Exploración registros por tipo de atención.....	149
Figura 78. Exploración atenciones por grupos etareos. ....	150
Figura 79. Exploración atenciones por Zona de Residencia. ....	151
Figura 80. Exploración atenciones por Régimen. ....	152
Figura 81. Exploración atenciones por Género.....	152
Figura 82. Exploración atenciones por Estado Civil.....	153
Figura 83. Exploración atenciones por Etnias.....	154
Figura 84. Exploración atenciones por Estratos. ....	154
Figura 85. Centroides y Clústeres generados por K-Means en WEKA. ....	155
Figura 86. Resultados Clustering, set de datos 2 .....	157
Figura 87. Resultados Clustering, set de datos 3 .....	158
Figura 88. Evaluación de Satisfacción de Requerimientos – Calificado.....	161

Figura 89. Evaluación de Resultados obtenidos – Calificado .....	162
Figura 90. Caso de Prueba 2. Requerimiento Guardar Informes Personalizados en Web Query.....	169
Figura 91. Script SQL utilizado y su Visual Explain - Caso de Prueba 2.....	170
Figura 92. Caso de prueba 4. Funcionamiento OLAP Financiero. ....	171
Figura 93. Script SQL utilizado y su Visual Explain - Caso de Prueba 4.....	172

## LISTA DE TABLAS

Tabla 1. Resultados esperados del proyecto de investigación.....	32
Tabla 2. Presupuesto. Costos de Recurso Humano.....	54
Tabla 3. Presupuesto. Costos Insumos y Materiales.....	54
Tabla 4. Presupuesto. Costos Totales del Proyecto.....	54
Tabla 5. OLTP vs OLAP.....	65
Tabla 6. Componentes del proceso ETL.....	76
Tabla 7. Resumen de Algoritmos para minería.....	84
Tabla 8. Caracterización Proceso de Cirugía.....	96
Tabla 9. Resumen Procedimientos proceso Cirugía.....	96
Tabla 10. Definición Objetivos de Minería de Datos.....	98
Tabla 11. Resumen de Requerimientos Funcionales.....	99
Tabla 12. Detalle requerimiento funcional RF-1.....	100
Tabla 13. Caso de Uso. CU-001. Generar Informes mediante herramienta OLAP.....	101
Tabla 14. Plan del Proyecto.....	102
Tabla 15. Resumen de tablas dimensionales del modelo de datos.....	104
Tabla 16. Descripción tablas del Modelo Dimensional. Tabla de hechos Cirugía.....	105
Tabla 17. Descripción tablas del Modelo Dimensional. Tabla de dimensión Paciente.....	107
Tabla 18. Descripción Set de Datos para técnica de Minería (Clustering).....	141
Tabla 19. Ficha técnica de minería de datos Clustering.....	143
Tabla 20. Exploración de datos atributo Edad.....	145
Tabla 21. Exploración de datos atributo Grupo_CIE10.....	146
Tabla 22. Exploración de datos atributo Grupo_Cirugía.....	148
Tabla 23. Exploración de datos atributo Tipo_Atención.....	148
Tabla 24. Exploración de datos atributo Grupo Etareo.....	149
Tabla 25. Exploración de datos atributo Zona_residencia.....	150
Tabla 26. Exploración de datos atributo Régimen.....	151
Tabla 27. Exploración de datos atributo Género.....	152
Tabla 28. Exploración de datos atributo Estado_Civil.....	153
Tabla 29. Exploración de datos atributo Etnia.....	153
Tabla 30. Exploración de datos atributo Estrato.....	154
Tabla 31. Resumen Exploración de datos. Atributos relevantes.....	156
Tabla 32. Resumen Grupos de diagnóstico relevantes identificados.....	159
Tabla 33. Planificación del despliegue de la Solución de BI.....	167
Tabla 34. Planeación y Control de Casos de Prueba.....	168
Tabla 35. Resultados esperados vs Resultados obtenidos en el proyecto de investigación.....	173

## RESUMEN

La toma de decisiones en la gestión hospitalaria permite avanzar hacia la atención oportuna y humanizada de los usuarios. La administración de recursos relacionados con el proceso de cirugía, permite alcanzar mejores resultados tanto clínicos como financieros. El aumento exponencial de datos clínicos dificulta cada vez más el acceso a la información mediante métodos tradicionales. La Minería de datos y la Inteligencia de Negocios, ofrecen herramientas que permiten lograr altos niveles de autonomía, disponibilidad y facilidad de acceso a la información. La presente investigación busca implementar una solución de inteligencia de negocios que, basada en Minería de Datos, permita apoyar la Gestión Estratégica del Proceso de Cirugía en el Hospital Departamental Universitario Santa Sofía de Caldas, ubicado en la ciudad de Manizales, Colombia. Se aplica la metodología CRISP-DM y mediante el uso de modelos de datos y técnicas de visualización dimensionales, se generan informes dinámicos personalizables por el usuario, reportes activos y despliegue móvil de datos. Así mismo se aplica *clustering* como técnica de agrupamiento y usando el algoritmo *K-Means* se exploran datos buscando identificación de grupos de pacientes que han sido sometidos a procedimientos quirúrgicos y cuyos diagnósticos estaban asociados a cáncer. Se busca ofrecer a los líderes del servicio una solución orientada a procesos analíticos, que integre tanto datos administrativos como asistenciales.

**Metodología:** Se documentan las 6 fases de la metodología CRISP-DM; se utiliza DB2 Web Query como motor de reportes, DB2 for IBMi para el modelo de Data Warehouse, ROAMBI Analytics para la visualización en dispositivos móviles y WEKA como herramienta de minería de datos para la técnica Clustering y específicamente el algoritmo K-Means.

**Resultados:** Motor de reportes de Inteligencia de Negocios, OLAP, *Active Reports* y *dashboards* implementados y validados por los interesados, visualización en dispositivos móviles, técnica de minería aplicada, arrojando dos clústeres como grupos de pacientes quirúrgicos y documentación de cada fase abordada en la metodología.

**Conclusiones:** La evaluación de la solución demuestra altos niveles de aceptación por parte de los interesados.

Palabra clave: Data Warehouse, Data Mining, Business Intelligence, CRISP-DM, Minería de datos en Salud.

## ABSTRACT

Decision making in hospital management allows progress towards a timely and humane care of users. Administrative procedures related to surgery processes lead to achieve better results for both clinical and financial processes. Exponential growth of clinical data makes increasingly difficult the access to information through traditional methods. Data Mining and Business Intelligence offer tools for achieving high levels of autonomy, availability and ease of access to information. This research study aims at implementing a Business Intelligence Solution based on Data Mining to support Strategic Management of Surgical Procedures at the *Hospital Departamental Universitario Santa Sofia* in Manizales, Caldas, Colombia. The CRISP-DM methodology was applied and users' dynamic customizable reports, active reports and deployment of mobile data were generated by using data models and dimensional visualization techniques. Similarly, the clustering technique and the K-Means algorithm were used to explore data on identifying groups of patients who underwent surgical procedures and whose diagnoses were associated with cancer. This study seeks to provide leaders with a solution directed to analytical processes that integrate both data and administrative assistance.

**Methodology:** The 6 phases of CRISP-DM methodology were documented. It was used DB2 Web Query as a reporting engine, DB2 as IBMi for Data Warehouse model, ROAMBI Analytics for viewing on mobile devices and WEKA as a data mining tool for the clustering technique, particularly the K-Means algorithm.

**Results:** Results show Business Intelligence reports, OLAP, Active Reports, implemented and validated dashboards by the interested parties, display on mobile devices, as well as an applied mining technique that generates two clusters as groups of surgical patients and reports of each of the methodological phases.

**Conclusions:** The solution evaluation exhibits high levels of acceptance by the interested parties.

**Keywords:** Data Warehouse, Data Mining, Business Intelligence, CRISP-DM methodology, Data Mining in Health Care.

## INTRODUCCIÓN

Durante las últimas dos décadas, los hospitales públicos han sufrido grandes transformaciones. La aparición de las Empresas Sociales del Estado (ESE), trae consigo grandes desafíos gerenciales. La autonomía presupuestal, financiera y administrativa, exige de igual manera que los controles sobre la productividad y el impacto social sean gestionados con mayor rigurosidad (Mantilla, 2000). Los resultados no son confrontados únicamente desde la eficiencia administrativa, la sostenibilidad y productividad, sino también desde el nivel de calidad en la atención al usuario y su familia<sup>1</sup>.

Los frecuentes cambios normativos obligan al reporte específico de las actividades desarrolladas en el marco del Sistema Obligatorio de Garantía de la Calidad, con el fin de medir los avances y el impacto en la comunidad<sup>2</sup>. El objetivo principal de un Hospital es proteger la vida y procurar acciones de promoción de la salud y prevención de la enfermedad, los resultados trascienden de lo financiero a lo social. (Neves, 2011)

Gracias a los avances de las plataformas tecnológicas, se han incrementado exponencialmente la generación y almacenamiento de datos y por tal motivo, es necesario aplicar técnicas que mejoren la posibilidad de su aprovechamiento. (Howe, 2014). Tanto los datos estructurados como los no estructurados pueden aportar invaluable posibilidades de generar conocimiento.

Según (Valcárcel, 2004), dentro de estas enormes masas de datos existe una gran cantidad de información "oculta" de interés estratégico, a la cual no se puede acceder por las técnicas clásicas de recuperación de la información. El descubrimiento de esta información es posible gracias a la Minería de Datos (DM) por su sigla en inglés de *Data Mining*.

Al igual que muchas organizaciones, los Hospitales han venido aplicando diferentes modelos e instrumentos orientados a la gestión del conocimiento, recursos como intranets, extranets, bases de datos expertas, herramientas colaborativas, portales corporativos, blogs, directorios expertos, sistemas de recomendación, sistemas de gestión de clientes, gestores de contenido y aprendizaje virtual (Chaffey, 2005), sin embargo es necesario incorporar técnicas de análisis e inteligencia de negocios (BI) por su sigla en inglés de *Business Intelligence*, que maximicen el impacto de los sistemas de información y ofrezcan medios de reporte dinámicos, eficientes y al alcance de todos los niveles gerenciales. (Azma, 2012)

---

<sup>1</sup> Ministerio de la Protección Social. Resolución 123 de 2012. Por medio de la cual se modifica la Resolución 1445 de 2006. Por la cual se definen las funciones de la Entidad Acreditadora

<sup>2</sup> Ministerio de la Protección Social. Resolución 1446 de 2006. Por la cual se define el Sistema de Información para la Calidad y se adoptan los indicadores de monitoria del Sistema Obligatorio de Garantía de Calidad de la Atención en Salud. Diario Oficial 46271 de mayo 17 de 2006

(Marcano y Talavera, 2007), afirman que las empresas privadas tanto como las públicas deben tener la capacidad de ser adaptativas, aprender cómo resolver problemas y generar conocimiento, para establecer nuevos métodos en pro de la resolución de los mismos.

Actualmente el Hospital Departamental Universitario Santa Sofía de Caldas es la única Institución Pública de alto nivel de complejidad en el departamento de Caldas y su área de cobertura incluye además el Magdalena Caldense y Norte del Valle. Cuenta con infraestructura tecnológica IBM de alto nivel, importantes estándares de licenciamiento de software y sistemas transaccionales sólidos para todos sus registros. Sin embargo, no cuenta con procedimientos que le permitan aprovechar los datos recopilados y transformarlos en conocimiento a partir de técnicas modernas para el tratamiento de datos, como las ofrecidas por las técnicas de Minería de Datos (DM) e Inteligencia de negocios (BI).

Su carácter de Hospital Universitario le atribuye responsabilidades adicionales de promover la investigación, aportar nuevos conceptos y favorecer la formación profesional en el sector Salud. No es un reto sencillo y aunque la informática médica se ocupa de la aplicación de las computadoras en las ciencias médicas y biológicas y se ha considerado un campo de investigación por derecho propio desde hace más de veinte años, es imprescindible proponer investigaciones que puedan orientar y alinear esfuerzos para tal fin. (Abdul-Kareem, 2000)

Es por ello, que una solución de inteligencia de negocios (BI) basada en minería de datos (DM), representa un avance para que el Hospital logre obtener beneficios al comprender su información y así mismo fortalezca el sector salud con nuevo conocimiento. Como expresan (Albarrán y Salgado, 2013) la minería de datos es un soporte en la toma de decisiones, en los negocios permite elevar los niveles de competencia, con base en los rápidos procesamientos y extracción de información relevante del mismo, descubriendo conocimiento y patrones en las bases de datos

La dinámica de las organizaciones demanda mejores tiempos de respuesta y se hace necesaria la formulación de modelos simples y eficientes que permitan transformar el contenido de los sistemas transaccionales en conocimiento aplicable a la misión y objetivos del negocio. (Marcano y Talavera 2007)

De acuerdo a los datos estadísticos consolidados para Colombia, el número de procedimientos quirúrgicos, ha venido creciendo durante los últimos años pasando de 6.863.256 de cirugías en 2009, a 15.005.221 en 2012. Para el departamento de Caldas se registra un crecimiento de 140.182 procedimientos en el año 2009, a 217.626 en el año 2013. (MINSALUD, 2014). Lo que ha promovido que la Organización Mundial de la Salud (OMS), la Organización Panamericana de la Salud (OPS) y el Estado colombiano, se esmeren en fortalecer el sistema de información y la seguridad del paciente en los momentos quirúrgicos. (MINSALUD, 2008)

Estas cifras demuestran el impacto de los procedimientos quirúrgicos en la recuperación de la salud. El Hospital Departamental Universitario Santa Sofía de Caldas realizó para la

vigencia 2013 7.665 cirugías<sup>3</sup>, para la vigencia 2014 se incrementó a 8545<sup>4</sup> y en general, durante los últimos 10 años se superó la cifra de 10.000 cateterismos vesicales y 2000 cirugías de corazón abierto<sup>5</sup>, lo cual le permite ubicarse como una Institución con gran experiencia en el tratamiento de enfermedades cardiovasculares y otras disciplinas quirúrgicas.

El presente proyecto de investigación aplicado para el Hospital Santa Sofía, busca desarrollar una solución que apoye la gestión y toma de decisiones en el proceso de Cirugía y ofrezca mayor dinamismo durante la extracción de datos, mejore los tiempos de respuesta en la personalización y generación de reportes, eleve el grado de integración de los datos y brinde mayores niveles de calidad en la información obtenida.

Dado que el estudio se enfocará en el proceso de Cirugía, se espera lograr mayor granularidad y profundización en la investigación y se contemplarán tanto los datos clínicos como administrativos. Al final del proyecto, se espera contar con una metodología replicable, de tal manera que en caso de aportar valor agregado, pueda ser aplicada en otros procesos, tanto misionales, como estratégicos y de apoyo.

---

<sup>3</sup> E.S.E Hospital Departamental universitario Santa Sofía de Caldas. Evaluación Anual del desempeño de acuerdo al Convenio de desempeño 0188 de 2004. Informe vigencia 2013. P 53.

<sup>4</sup> E.S.E Hospital Departamental universitario Santa Sofía de Caldas. Informe de Gestión 2014. P 50

<sup>5</sup> E.S.E Hospital Departamental universitario Santa Sofía de Caldas, informe semestral de Rendición de Cuentas. Octubre de 2014.

## 1. REFERENTE CONTEXTUAL

### 1.1. Descripción del área problemática

Actualmente el Hospital Santa Sofía, siendo la única Institución Pública de alto nivel de complejidad para el Departamento de Caldas, cuenta con infraestructura tecnológica robusta y sistemas transaccionales sólidos para sus registros asistenciales y financieros. Sin embargo, y a pesar de su amplia gama de aplicaciones y recursos, no cuenta con una estructura definida y estandarizada para la generación de informes gerenciales por proceso.

Los análisis de gestión y autocontrol se basan en centenares de consultas en hojas de cálculo, que mediante conexiones ODBC (*Open DataBase Connectivity* por sus siglas en inglés) ofrecen consultas asíncronas con actualizaciones manuales y que ocasionan problemas de actualización de los orígenes de datos; reportes estáticos que dificultan el nivel de especificidad y debido a la multiplicidad de archivos se presenta ambigüedad al momento de elegir la fuente correcta, generando pérdida de reportes históricos, duplicidad en los datos y posibles inconsistencias.

Así mismo, se ha generado una dependencia notoria de los procesos asistenciales con el proceso de Administración de Tecnologías de la Información, que obliga a solicitar frecuentemente la generación de consultas, pues los usuarios no cuentan con los instrumentos necesarios para conformar sus propias estructuras de reporte. Esta dependencia conlleva a la pérdida de interés por parte de los usuarios en la generación de nuevos escenarios de verificación de sus datos, obligándose a trabajar con datos históricos desactualizados para definir soluciones a situaciones actuales y generando así tendencias equivocadas y peor aún conclusiones equivocadas acerca de la problemática real.

Es así, como el seguimiento a los indicadores de los procedimientos se hace a través de registros manuales, dificultando la exactitud de los valores consignados y mostrando tendencias que carecen de un soporte inalterable. De la misma manera, en análisis de casos en la toma de decisiones se torna anecdótico, pues a pesar de todos los formatos y controles físicos llevados de forma disciplinada, es difícil garantizar que se contemplen todos los datos generados en un quirófano de alta complejidad.

Los líderes del proceso de cirugía terminan invirtiendo tiempo valioso en la gestión de los datos y en el sector salud, cada segundo cuenta, cuando se trata de salvar vidas. Sin embargo es necesario que como E.S.E (Empresa Social del Estado) se logre mantener el equilibrio financiero y la optimización de recursos para la sostenibilidad de los procesos.

A pesar de contar con más de 10 años de registros clínicos estructurados, el hardware y software con el potencial suficiente para efectuar análisis de datos en tiempo real y el conocimiento en la gestión de bases de datos, no se ha logrado articular este conjunto de

elementos para proveer un servicio de análisis de datos, obtención y gestión del conocimiento.

Se continúa recurriendo a herramientas tradicionales para la explotación de los datos y esto evita que el tiempo, los recursos y la experiencia se optimicen. Desacelera la toma de decisiones y somete a la Institución a un desgaste administrativo que luego se refleja en el ámbito asistencial.

Es prioritario implementar soluciones modernas de tratamiento de datos, que mejore la oportunidad entre el registro clínico y su uso posterior en la planeación estratégica.

## 1.2. Antecedentes

El Ministerio de Salud y Protección Social monitorea la información correspondiente a los procedimientos quirúrgicos y desde diversos instrumentos ha fortalecido las actividades de seguimiento y acompañamiento.

La reglamentación del Sistema de Información para la Calidad (Resolución 0256 de 2016) del Sistema Obligatorio de Garantía de Calidad (Decreto 1011 de 2006), establece una serie de indicadores dirigidos a evaluar relacionados con calidad técnica, gerencia del riesgo, accesibilidad, oportunidad y lealtad del usuario en las Instituciones Prestadoras de Servicios de Salud (IPS). El decreto 2193 de 2004 exige de manera trimestral y anual acumulado, el reporte de indicadores de productividad asociados a las atenciones y específicamente los servicios de cirugía deben incorporar en sus procesos y procedimientos, las acciones tendientes al registro y reporte oportuno según lo establecido.

El Sistema único de Acreditación, mediante su anexo 123 de 2012, reglamenta los criterios obligatorios de acreditación y en uno de sus estándares, el de Gerencia de la Información, dispone que los hospitales deberán implementar procedimientos de **Minería de datos** para la toma de decisiones orientada a mejorar la calidad en la atención al usuario y su familia.

Lograr el aprovechamiento máximo de los datos almacenados durante estos años, representa un propósito a gran escala y los proveedores de salud exploran diversas técnicas visuales e interactivas en la generación y análisis de grandes y complejos conjuntos de datos, así como identificar sus conexiones y tendencias. (Bellucci, 2012)

La inteligencia de negocios y la minería de datos, se definen como herramientas analíticas especializadas para permitir la toma de decisiones basadas en una variedad de contextos organizacionales (Rohloff, 2011). Han sido utilizadas de manera satisfactoria en el sector de la salud, proporcionando beneficios a las Instituciones Prestadoras de Servicios (IPS) e impactando positivamente la mejora en la atención a los pacientes (Tremblay et al. 2012),

de la misma forma, facilitando la optimización, aprovechamiento del recurso humano, mejoramiento de procesos y procedimientos (Flower, 2006).

(Madsen, 2011) en su artículo “**Optimize your BI: five reasons healthcare business intelligence differs from other industries**”, publicado en la revista *US National Library of Medicine National Institutes of Health*, argumenta cinco aspectos que diferencian los proyectos de Inteligencia de Negocios en el sector salud, entre los cuales resalta el nivel de normatividad y regulación, los riesgos atribuibles a la misión crítica desempeñada, los datos no estandarizados que pueden dificultar las labores de ETL (extracción, transformación y carga) del modelo de datos y la perspectiva de los líderes al momento de analizar los datos.

Realiza un aporte relevante que permite tener en cuenta aspectos claves para el éxito del presente proyecto de investigación, entre los cuales se pueden contemplar:

- a. Convocar equipos multidisciplinarios que permitan integrar los datos clínicos, técnicos y de negocio, aprovechando personas con amplios conocimientos y a la vez otras con conocimientos especializados.
- b. Enfoque flexible y explorar diferentes métodos de acceso a los datos
- c. Centrarse en generar valor a la toma de decisiones
- d. Planificar y documentar la visión del proyecto.
- e. Usar métodos ágiles.
- f. Invertir el tiempo necesario en labores de ETL (*Extract, Transform and Load*)
- g. Desarrollar solución flexible con entregables funcionales
- h. Involucrar los usuarios, como primera razón de existencia de BI

Argumenta la autora, que en el desarrollo de proyectos de BI debemos ir más allá de los informes estáticos, que dominan el ámbito normativo, y tomar ventaja de la última visualización de datos y capacidades móviles. Estos dos aspectos son eje fundamental del presente proyecto de investigación y los planteamientos listados con anterioridad permiten orientar la implantación de la solución.

(Agnoletti et Al, 2013) en su artículo “**Operating room data management: improving efficiency and safety in a surgical block**” publicado en *BMC Surgery*, propone una herramienta de procesamiento de datos para dar a los gerentes de hospitales, anesthesiologists y cirujanos, la base de información para aumentar la eficiencia de los quirófanos y la seguridad del paciente. Un estudio aplicado en un hospital público de Italia que ofrece un conjunto de reportes de acuerdo al nivel de acceso del usuario, que permite obtener información consolidada sobre 14.675 procedimientos quirúrgicos realizadas entre los años 2009 hasta 2011. Entre otros, el estudio ofreció los siguientes tipos de informes por usuario:

**Jefe del Servicio:** datos generales de cirugía, total de procedimientos por sala, producción bruta por sala, número de cirugías, uso de horas de cirugía, cuadros comparativos entre salas, semaforización de indicadores.

**Anestesiólogo:** actividad quirúrgica total y tiempos promedio de anestesia, rutas de atención de los pacientes, promedios y desviación estándar de tiempos de anestesia por tipos de cirugía y otras estadísticas básicas sobre los procedimientos (mínimos, máximos, mediana, promedios)

**Cirujano:** top cinco (5) de cirugías realizadas, uso de quirófanos, indicadores de eficiencia, comparaciones por semanas, meses, trimestres, semestres y años, análisis sobre agendamiento, así como cifras por grupos quirúrgicos y específicos.

Como resultado se obtuvo mejora en indicadores sobre agendamiento de cirugías, el número de procedimientos no programados se redujo (del 25% en 2009 al 14% en 2011) y el uso de horas adicionales pasó del 28% en 2009 al 21% en 2011, incremento en cirugías de alta complejidad y disminución de mortalidad para algunos tipos de cirugía. Concluye el estudio que “Representa un experimento con éxito de la introducción de la innovación empresarial en un hospital público de Italia” y que “aunque el proyecto fue desarrollado por profesionales de la salud, su objetivo era alinear los objetivos gerenciales y profesionales”.

El estudio no menciona por ejemplo, la integración de otros datos relacionados con el servicio de cirugía, como puede ser, diagnósticos de las patologías integración de reportes históricos a Entes de control, dimensión asociadas a los pacientes (datos demográficos y geográficos), información sobre clientes (Empresas promotoras de salud), lo que se convierte en una oportunidad para la presente investigación.

(Wang, 2013) en el artículo: “**Predefined Three Tier Business Intelligence Architecture in Healthcare Enterprise**” publicado en *Journal of Medical Systems*, define inteligencia de negocios como un conjunto de métodos, procesos y tecnologías mediante las cuales se puede compartir, analizar y organizar la información de los hospitales y que permite impactar la toma de decisiones estratégica, táctica y operativa.

Propone una arquitectura *Top-Down* para su implementación en un Hospital en Massachusetts, a través de tres capas: *data Warehouse*, repositorios y procesos de análisis. Además, indica como esta arquitectura permite enfocarse en las necesidades del negocio, integrar diferentes plataformas y una implementación rápida.

Hace un recuento sobre la experiencia de implementar diversos proyectos utilizando estas tres capas, obteniendo escalabilidad y alto nivel de integración de fuentes de datos. Resalta también la importancia de aplicar los procesos de negocio para alinear los objetivos institucionales con los del proyecto.

(Foshay & Kuziemy, 2014) en su estudio “**Towards an implementation framework for business intelligence in healthcare**” publicado en *International Journal of Information Management*, mediante un estudio de caso en Nueva Escocia, Canadá, identifica los factores que pueden afectar la implementación de modelos de BI en los Hospitales, ayudando a confirmar la importancia de alinear los objetivos del negocio, la participación del equipo de salud en el diseño y logrando apoyo en los diversos niveles de la organización. Expone los factores críticos al momento de iniciar la implementación de

proyectos de inteligencia de negocios en los hospitales y plantea tres preguntas en su investigación:

- a. ¿Cuáles son los efectos adversos más significativos en los procesos y resultados atribuibles a la falta de capacidades de apoyo a las decisiones?
- b. ¿Cuáles son las causas fundamentales de estos efectos adversos, y cuáles son las posibles soluciones?
- c. En el contexto de las preguntas a y b, ¿cuáles son las consideraciones más importantes para la implementación de sistemas de BI en las Instituciones de Salud?

Expresan los autores “*Este trabajo se centra estrictamente en los factores de éxito fundamentales como punto de partida para el éxito final de un sistema de BI*” y concluyen entre otros aspectos:

- a. Proporcionar los tomadores de decisiones con un cómodo acceso a información de alta calidad es fundamental para el éxito del sistema de BI.
- b. Abordar problemas de calidad que se encuentran en los sistemas transaccionales y operacionales es un factor de éxito en el proyecto de BI.
- c. Desarrollar habilidades en el usuario final para el uso de los recursos ofrecidos por el proyecto. Fortalecer el despliegue y difusión de la información.

Proponen un *framework* de trabajo para identificar las necesidades de información en los tres niveles de procesos hospitalarios (misional, estratégico y de apoyo) y priorizar con claridad los objetivos estratégicos, la gestión de recursos, Gestión de operaciones, calidad y gestión de riesgos, útil para condensar las necesidades el proceso de Cirugía del Hospital Departamental Universitario Santa Sofía de Caldas.

En Cuba (Hernández & Denia, 2011) en su investigación “**Procedimiento para el desarrollo de un sistema de inteligencia de negocios en la gestión de ensayos clínicos en el Centro de Inmunología Molecular**”, para el Centro Nacional de Información de Ciencias Médicas, diseñaron un procedimiento que contribuyera al análisis y almacenamiento de los ensayos clínicos y que permitiera la aplicación de Inteligencia de negocios en el Centro Molecular.

El procedimiento propuesto, busca ofrecer a los médicos y especialistas un mecanismo mediante el cual puedan gestionar de manera íntegra, y viabilizar los estudios estadísticos que requieren. Como argumentan los investigadores, el procedimiento fue evaluado a partir del método de experto Delphi y se obtuvo el resultado de 'Muy adecuado' e indica de manera detallada los pasos llevados a cabo durante el proyecto.

En este caso, los investigadores exploraron varias metodologías para proyectos de minería de datos y a sin embargo, diseñaron su propia metodología y al final, ofrecieron al Centro de Inmunología Molecular un proyecto documentado y replicable para optimizar el análisis de las evoluciones clínicas por parte de los especialistas.

(Martínez, 2010) en su investigación “**La Inteligencia de Negocios como Herramienta para la Toma de Decisiones Estratégicas en las Empresas. Análisis de su aplicabilidad en el contexto corporativo colombiano**”, como tesis para optar al título de Magister en Administración para la Universidad Nacional de Colombia, define la Inteligencia de Negocios es un conjunto de técnicas y herramientas tanto de gestión empresarial como de aplicación tecnológica, que permiten la extracción e integración de los datos, su posterior procesamiento y distribución en forma de información.

En su estudio hace un recorrido sobre aspectos relacionados con el desarrollo de estrategia y tácticas en las organizaciones, el rol de la información en la sociedad moderna y conceptualiza sobre el uso de herramientas tecnológicas en la formulación estratégica.

Describe capas de infraestructura y funcionalidad de inteligencia de negocios, tales como la integración, el procesamiento de datos y la plataforma de distribución y entrega, lo cual permite conocer un enfoque sobre la arquitectura que se puede abordar para la solución de inteligencia de negocios propuesta en la presente investigación y establece un marco de referencia colombiano sobre el nivel de madurez de las empresas en cuanto a la aplicación de sistemas para el análisis de datos en la toma de decisiones.

(Van Oostrum et Al, 2011) en su artículo “**A Method for Clustering Surgical Cases to Allow Master Surgical Scheduling**”, publicado en *Information Systems and Operational Research*, aplicaron técnicas de **clustering** para apoyar la gestión de agendamiento en quirófanos aplicando el estudio de caso en el Hospital holandés Beatrix. Utilizaron tiempos de cirugía, tiempos de estancia, códigos de procedimiento y tipos de cirugía, que permitan optimizar el uso de los tiempos en las salas. Incluyeron solo las cirugías electivas, es decir, se excluyen aquellos procedimientos clasificados como urgentes.

A través de la técnica de minería **clustering**, agruparon las cirugías de acuerdo a su homogeneidad y de esta manera optimizaron los tiempos de agenda de las salas de operaciones, evitando que se filtraran cirugías que pudieran afectar el desempeño de las jornadas quirúrgicas.

Analizaron un solo año de datos (2006) y por ello uno de los temas de discusión es si las cifras son representativas para el año siguiente.

En la presente investigación se utilizarán diferentes tipos de informes que permitan a los líderes del servicio de cirugía optimizar su gestión de recursos. Así mismo se integrarán diez años de datos históricos. Este antecedente ejemplifica la aplicación de técnicas de agrupamiento en la toma de decisiones. Tema importante para el proyecto propuesto.

(Rivas et al, 2007) en su investigación “**Una metodología para sectorizar pacientes en el consumo de medicamentos aplicando data mart y data mining en un hospital**” para la Universidad Nacional San Marcos de Perú, presentan la aplicación de una técnica de “clúster” para agrupamiento de pacientes y que permite conocer los grupos a los cuales se les suministran clases específicas de medicamentos.

Este estudio es de gran utilidad para el presente proyecto, pues incluye la aplicación de **Clustering**, formando grupos de pacientes con unas características similares a través del algoritmo **Simple K-Means**.

Se aprecia que son consideradas dimensiones como diagnóstico, medicamentos, médicos, personas, servicios, tipo de seguro, entre otros y se analizan atributos como sexo, estado civil, código de diagnóstico, tipo de medicamento (controlado y no controlado) y presentación de medicamentos.

Expone las actividades realizadas para lograr los resultados y conceptualiza acerca de los elementos que conforman el *data warehouse*, tales como base de datos dimensional, cubos, dimensiones y OLAP.

Otro aspecto relevante en la investigación, es el uso de la aplicación **WEKA** (*Waikato Environment for Knowledge Analysis*) entorno para el análisis de datos de la Universidad de Waikato, y la descripción de las actividades llevadas a cabo, así como gráficos e imágenes de los resultados obtenidos.

Recomienda complementar el estudio, incluyendo aspectos como diagnósticos, o aplicando otras técnicas predictivas como redes neuronales.

Por su parte (Taié, 2008) en su Tesis “**Desarrollo de una metodología de extracción de conocimientos a partir de datos de micromatrices de DNA basada en ontologías genéticas**” para optar al Título de Magister en Explotación de Datos y Descubrimiento del Conocimiento, de la Universidad de Buenos Aires, desarrolla una metodología de análisis de datos que permite descubrir conocimientos biológicamente relevantes, buscando agrupaciones de genes, identificando aquellos que se expresan de manera similar en distintas condiciones experimentales.

Para ello, describe y aplica métodos de **agrupamiento jerárquico**, **K-Means**, perteneciente a técnicas de Clustering, aplicables al presente proyecto.

Realizar pruebas con los 5 métodos de agrupamiento y concluye que la aplicación de **K-Means** con 500 iteraciones fue óptima para su investigación.

Dado que el proyecto propuesto incluye además de inteligencia de negocios, la aplicación de técnica de minería de datos *clustering*, específicamente con el algoritmo *K-Means*, este antecedente ofrece herramientas útiles como marco de referencia.

### 1.2.1 Conclusiones sección de antecedentes

Es claro que a la luz de lo expuesto, se cuenta con una riqueza literaria de grandes proporciones en cuanto a la aplicación de técnicas de minería de datos e inteligencia de negocios en los hospitales. Las técnicas han sido diversas de acuerdo al objetivo y resultados buscados. Las prácticas modernas de administración, hacen uso de la minería de datos para la conformación de equipos de salud de alto desempeño, que basados en

el análisis de datos en áreas críticas como unidad de cuidados intensivos, sala de operaciones, sala de traumas, consulta externa. (Parker & Rosen, 2012) y cualquier otro tipo de atención dependiendo del nivel de complejidad del hospital, pueden alinear sus decisiones con los objetivos estratégicos.

A pesar de que durante la indagación de antecedentes en diversos recursos virtuales como *Science Direct*, *Elsevier*, *Research Gate*, *Scopus*, *IEEE* y otros, no arrojó artículos directamente relacionados con la gestión estratégica de servicios de cirugía, si es evidente que se han documentado experiencias con la aplicación de componentes de BI y DM en Hospitales y en otros tipos de procesos, sobre todo, relacionados directamente con patologías y no con la gestión gerencial.

### 1.3. Justificación

En Colombia, el recurso humano especializado para el sector salud es escaso. Desde hace ya varias décadas se ha presentado déficit en diversas especialidades y subespecialidades que dificultan la prestación de servicios y se convierten en factores que afectan la accesibilidad y oportunidad en la respuesta a los usuarios. Durante el 2011 solamente egresaron 1098 especialistas en todo el país, de los cuales 295 corresponden a disciplinas quirúrgicas. (MINSALUD, 2013)

Dada la creciente incidencia mundial en traumas (63 millones de personas) por año y tratamientos relacionados con cáncer y enfermedad cerebro vascular (31 millones de personas), el peso de la cirugía en los sistemas de salud públicos irá en aumento. Se calcula que en todo el mundo se realizan cada año 234 millones de operaciones de cirugía mayor, lo que equivale a una operación por cada 25 personas y a pesar de la relación costo eficacia que puede tener la cirugía en cuanto a vidas salvadas, la falta de acceso a una atención quirúrgica sigue siendo un grave problema. (MINSALUD, 2008)

Frente a estos datos alarmantes, la optimización de los recursos asociados a los quirófanos se convierte en una estrategia de gran impacto, pero compleja. Garantizar las coberturas requeridas por los pacientes, disminuir el porcentaje de cirugía cancelada, el aprovechamiento del recurso humano y la sincronización con los otros servicios hospitalarios, devenga gran esfuerzo y dedicación.

Se requiere sistemas capaces de garantizar la disponibilidad, integridad, confiabilidad y no repudio de la información, cubriendo requerimientos relacionados con el análisis de datos, visualización de resultados y soporte a decisiones, características propias de un modelo de BI. (Azma, 2009)

El Hospital Departamental Universitario Santa Sofía de Caldas, cuenta con 5 (cinco) quirófanos, más de 30 especialidades médicas y un recurso humano cualificado que le permite gozar de excelente prestigio en la región.

El proceso de Cirugía, representa uno de los servicios de mayor impacto en la atención de alta complejidad y los servicios adicionales derivados de los procedimientos quirúrgicos, incluyen apoyo diagnóstico, terapéutico, medicamentos, controles y estancias prolongadas de acuerdo al avance mostrado por los pacientes.

Para el registro de la historia clínica se ha dispuesto motores estables y eficientes de bases de datos (DB2 de IBM), servidores de alta gama (*Power System Blade IBM*), dispositivos de almacenamiento robustos (*Storwize V7000 IBM*) y licencias de uso para diferentes categorías de software (Incluido *WebQuery* para BI).

Es necesario que para el aprovechamiento de dicha arquitectura, se implemente un modelo que permita gestionar de forma dinámica sus datos y generar de manera autónoma los informes. Aplicación de técnicas de BI y DM que acerquen los datos a los niveles directivos y promuevan la optimización en análisis y toma de decisiones.

Una solución que ofrezca a los líderes la posibilidad de consolidar y visualizar de manera práctica los informes, garantizar fuentes de datos seguras y confiables, integridad de reportes históricos y promover interés de los usuarios por los datos de su servicio.

Se pretende entonces, mejorar los niveles de oportunidad, disponibilidad y confiabilidad de la información, así como facilitar la presentación de reportes ante los entes de control y el cumplimiento del marco normativo que rige la administración pública.

Se hace entonces necesario, que el Hospital Departamental Universitario Santa Sofía, construya una solución para unificar los criterios de búsqueda, consolidación y generación de informes y que mediante una estructura consistente, dinámica e integral, pueda apoyar el nivel ejecutivo para la toma de decisiones.

La investigación propuesta plantea el desarrollo de una solución para el proceso de Cirugía del Hospital, basada en BI y DM, accesible, centralizada y consistente que permita seleccionar, explorar, y modelar los datos en tiempo real. (Sheiderman, 2002)

Técnicas de BI y DM, que de acuerdo a (Mettler y Vimarlund, 2009) integren datos de una amplia variedad de fuentes internas y externas, ofrezcan una plataforma eficaz de información para los tomadores de decisiones de atención médica y apoyen la labor gerencial en todos los niveles de la Institución.

Recursos que permitan optimizar el uso de tiempo por parte del personal de salud y líderes de proceso, pues estudios realizados indican que la implementación de minería de datos e inteligencia de negocios en la organizaciones, tienen un retorno de inversión medio del 430% en 1,6 años. Esto debido a que en promedio un trabajador puede invertir hasta el 30% de su tiempo laboral para generar reportes repetitivos. (Dan Vesset, 2012)

Se requiere optimizar el uso de la infraestructura informática y potencializar el recurso humano con que cuenta el Hospital, a través de la implementación de soluciones que acerquen la información a los líderes, de tal manera que la gestión del conocimiento se encuentre a un par de clics y puedan dedicarse con mayor libertad al hermoso propósito de proteger la vida.

#### **1.4. Formulación del problema**

La pregunta de investigación que busca resolver la investigación es:

**¿Cómo desarrollar una solución de inteligencia de negocios basada en minería de datos, que permita apoyar la toma de decisiones en el proceso de Cirugía del Hospital Departamental Universitario Santa Sofía de Caldas?**

## 1.5. Objetivos

### 1.5.1. Objetivo General

Implementar una solución de inteligencia de negocios basada en minería de datos, para apoyar la gestión estratégica del proceso de Cirugía del Hospital Departamental Universitario Santa Sofía de Caldas, en la ciudad de Manizales

### 1.5.2. Objetivos Específicos

- a. Definir un modelo de datos dimensional que soporte procesos analíticos y de minería de datos.
- b. Implantar y documentar una solución de Inteligencia de Negocios usando la herramienta *DB2 WebQuery*, Software de *Mobile Analytics*, y la agrupación de pacientes quirúrgicos a través de técnicas de *Clustering*.
- c. Validar los resultados obtenidos por medio de líderes asistenciales y administrativos del Hospital Santa Sofía de la ciudad de Manizales.

## 1.6. Alcances y limitantes

El proyecto se llevará a cabo en el proceso de Cirugía del Hospital Departamental Universitario Santa Sofía de Caldas, de alto nivel de complejidad ubicado en la Ciudad de Manizales, departamento de Caldas, Colombia.



Figura 1 Logotipo Hospital Santa Sofía. Empresa objetivo del Estudio de Caso

### 1.6.1. Unidad de Análisis y Muestra a utilizar

La unidad de análisis para la solución de Inteligencia de negocios, estará conformada por los registros clínicos correspondientes a los procesos quirúrgicos que se han almacenado en el sistema de información ERP HOSVITAL entre los años 2007 y 2015, así como sus dimensiones de pacientes, profesionales, geográficas, temporal, clientes y aspectos financieros.

Se incluirán los datos sobre productividad quirúrgica publicados en el Sistema de Información Hospitalario (SIHO) del Ministerio de Salud entre los años 2010 y 2015, correspondientes a los requerimientos del Decreto 2193 de 2004.

La propuesta incluye la implementación de la solución de Inteligencia de Negocios y su respectivo despliegue con transferencia de conocimiento a los usuarios correspondientes, así como la documentación de las diferentes fases llevadas a cabo en la metodología.

### 1.6.2. Confidencialidad de los datos

La confidencialidad de los datos forma parte integral del proyecto de investigación, conforme al Artículo 15 de la Constitución Política que reza "*Todas las personas tienen derecho a su intimidad personal y familiar y a su buen nombre, y el Estado debe respetarlos y hacerlos respetar. De igual modo, tienen derecho a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en bancos de datos y en archivos de entidades públicas y privadas.*"

En esta misma línea, la Ley 23 de 1981, en su artículo 34 establece "*La historia clínica es el registro obligatorio de las condiciones de salud del paciente. Es un documento privado, sometido a reserva, que únicamente puede ser conocido por terceros previa autorización del paciente o en los casos previstos por la ley*"

Buscando la aplicación normativa, el Comité de Ética del Hospital deberá aprobar la ejecución del proyecto, debido al carácter privado de la Historia Clínica, la confidencialidad y control de acceso a los registros asistenciales, estipulado por la Resolución 1995 de 1999<sup>6</sup>.

Cualquier tipo de sustentación de los resultados del proyecto, deberá realizarse de manera privada, para evitar que los datos sean presentados al público en general y se incumpla con los requisitos de confidencialidad. Toda ilustración que incluya datos personales, deberá ser adaptada para ocultar dichos atributos.

Para el despliegue a nivel interno de la Institución se mostrarán siempre los datos reales de los pacientes.

De acuerdo con la Resolución 8430 de 1993,<sup>7</sup> la investigación puede ser clasificada sin riesgo, toda vez que su Artículo 11 establece en el literal a “*Investigación sin riesgo: Son estudios que emplean técnicas y métodos de investigación documental retrospectivos y aquellos en los que no se realiza ninguna intervención o modificación intencionada de las variables biológicas, fisiológicas, psicológicas o sociales de los individuos que participan en el estudio, entre los que se consideran: revisión de historias clínicas, entrevistas, cuestionarios y otros en los que no se le identifique ni se traten aspectos sensitivos de su conducta*”

En la construcción de los datos para la solución de inteligencia de negocios, no se efectuarán adaptaciones o transformaciones de campos asociados con la historia clínica, pues como establece la resolución 1995 de 1999 en su Artículo 16 “*El prestador de servicios de salud, debe archivar la historia clínica en un área restringida, con acceso limitado al personal de salud autorizado, conservando las historias clínicas en condiciones que garanticen la integridad física y técnica, sin adulteración o alteración de la información*” y en su artículo 18 “*Los programas automatizados que se diseñen y utilicen para el manejo de las Historias Clínicas, así como sus equipos y soportes documentales, deben estar provistos de mecanismos de seguridad, que imposibiliten la incorporación de modificaciones a la Historia Clínica una vez se registren y guarden los datos*” subrayado y cursiva fuera del texto original.

### **Ver Anexo 8 Aprobación de investigación por Comité de Ética del Hospital**

---

<sup>6</sup> Resolución 1995 de 1999 del Ministerio de Salud. Por la cual se establecen normas para el Manejo de la Historia Clínica.

<sup>7</sup> Por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud.

### 1.6.3. Herramientas de software Candidatas

En los últimos años, el mercado *Business Intelligence* se ha enriquecido con soluciones *open source* que cubren todo el espectro de necesidades de una organización para la explotación de la información. Algunas de estas herramientas tienen ya a sus espaldas varios años de recorrido y actualmente se hallan respaldadas por organizaciones que tienen un claro modelo de negocio orientado a los servicios de valor añadido. (Davenport et al, 2010)

Las tecnologías de la visualización son buenas para ubicar patrones en un conjunto de datos y pueden ser usadas en un proceso de Minería de Datos, para analizar la calidad del conjunto de datos. Los modelos de visualización pueden ser bidimensionales, tridimensionales o incluso multidimensionales. Se han desarrollado varias herramientas de visualización para integrarse con las bases de datos ofreciendo una visualización de forma interactiva a la Minería de Datos. (Marcano y Talavera, 2007)

Se utilizarán herramientas de software con las que cuenta el Hospital bajo el esquema de licencia propietaria y cuyo proveedor es IBM:

- a. Sistema Operativo *IBMi*, el cual incluye el motor de bases de datos DB2.
- b. Software *IBM WebQuery* para la implementación y despliegue de la solución de inteligencia de negocios.

Para complementar la visualización y análisis de variables, se proponen las siguientes herramientas:

- a. *ROAMBI Analytics, QLIK*: software para visualización y despliegue de Inteligencia de Negocios en dispositivos móviles.
- b. *WEKA*: plataforma para el aprendizaje automático y minería de datos.

## 1.7. Resultados esperados

Tabla 1.  
Resultados esperados del proyecto de investigación

Objetivo específico relacionado	Entregable	Descripción y Beneficios	Indicador
Definir un modelo de datos dimensional que soporte procesos analíticos y de minería de datos.	Modelo de datos especificado e implementado	<p>Estructura coherente de datos, después de su respectivo análisis, especificación, transformación y almacenamiento.</p> <p>Elemento fundamental para el despliegue seguro, eficiente y confiable de una solución de inteligencia de negocios, así como la fuente de datos de calidad para aplicación de técnicas de minería de datos <b>clustering</b></p>	<p>Porcentaje de modelo de datos especificado e implementado</p> <p>(Tablas de hechos y dimensiones implementadas/Tablas de hechos y dimensiones especificadas en el modelo) x 100</p> <p><b>Meta: 100%</b></p>
Implantar y documentar una solución de Inteligencia de Negocios usando la herramienta DB2 WebQuery, Software de Mobile Analytics, y la agrupación de pacientes quirúrgicos a través de técnicas de <i>Clustering</i> .	Solución de Inteligencia de negocios implantada en el proceso de cirugía del Hospital Departamental Universitario Santa Sofía de Caldas	<p>Solución de inteligencia que incluye las siguientes funcionalidades:</p> <ul style="list-style-type: none"> <li>• Módulo de reportes estáticos y activos</li> <li>• Módulo de Procesamiento analítico en línea (OLAP) como interfaz de usuario final que emplea una técnica llamada análisis multidimensional para presentar informes gráficos y permitir el análisis flexible e interactivo de datos. (Wang et al, 2012).</li> <li>• Dashboards para la visualización y monitoreo de los indicadores definidos por el proceso</li> <li>• Despliegue de informes a través de dispositivos móviles.</li> <li>• Resultados de aplicación de <b>clustering</b> como técnica de agrupamiento en minería de datos.</li> </ul>	<p>Porcentaje de implementación de funcionalidades de inteligencia de negocios propuestos</p> <p>(N° funcionalidades implementadas en la solución de inteligencia de negocios / N° funcionalidades propuestas (5)) x 100</p> <p><b>Meta: 100%</b></p>

Objetivo específico relacionado	Entregable	Descripción y Beneficios	Indicador
Validar los resultados obtenidos por medio de líderes asistenciales y administrativos del Hospital Santa Sofía de la ciudad de Manizales.	Resultado de la validación realizada por los pares académicos	Evaluar el porcentaje de aceptación sobre los requerimientos propuestos por el proyecto.	<p>Porcentaje obtenido de la evaluación por pares académicos, medido desde dos aspectos:</p> <p>a. Porcentaje de Satisfacción con requerimientos.</p> <p>Meta &gt;=80%</p> <p>b. Porcentaje de Evaluación de Resultados</p> <p>Meta &gt;=80%</p>

Elaboración propia. Resumen de las metas propuestas por el proyecto y los entregables para cada uno de los objetivos.

## 2. ESTRATEGIA METODOLÓGICA

### 2.1. Tipo de Estudio

Luego de efectuar la revisión de la literatura y el estado de arte, se aprecia un conjunto amplio de investigaciones y resultados acerca de la aplicación de la minería de datos y la inteligencia de negocios en diversos escenarios del sector salud.

La Minería de datos ha sido asociada a análisis con medicamentos, insumos, procedimientos quirúrgicos, morbilidad y mortalidad, y aplicación de metodologías para proyectos relacionados con Inteligencia de Negocios

Siendo así, el presente estudio se considera **descriptivo**, y debe apoyarse no sólo en metodologías documentadas según las técnicas a usar, sino que podrá verse sometido a factores propios de cada institución.

Se considera un **estudio de caso**, pues será aplicado en el Hospital Departamental Universitario Santa Sofía de Caldas, ubicado en la ciudad de Manizales, Departamento de Caldas, Colombia y particularmente, en el Proceso de Cirugía.

El proyecto deberá ser implementado de manera incluyente y dinámica. Requiere de comunicación directa con el usuario y tácticas de despliegue especiales para lograr altos niveles de adherencia.

Comprende una convergencia de recursos y de actividades, tales como tecnología, arquitecturas, requerimientos técnicos y del negocio, fuentes diversas de datos, procesos de *ETL* (Extracción, transformación y carga de datos), cubos *OLAP*, técnicas de minería de datos, bases de datos y herramientas de visualización, navegación y despliegue.

Dado que se han identificado varias metodologías para la aplicación en proyectos de minería de datos, se plantea una comparación previa en la cual se abordan sus respectivas características. Luego de la comparación se elegirá y se describirá una de ellas.

## 2.2. Metodologías para proyectos de minería de datos

Las metodologías permiten llevar a cabo el proceso de minería de datos en forma sistemática y no trivial. Ayudan a las organizaciones a entender el proceso de descubrimiento de conocimiento y proveen una guía para la planificación y ejecución de los proyectos (Moine et Al, 2010)

Existen diferentes metodologías para la implementación de proyectos de minería de datos. Algunas organizaciones implementan el proceso KDD, mientras que otras aplican un estándar más específico como CRISP-DM (*Cross Industry Standard Process for Data Mining*). Si la organización ha adquirido productos de la empresa SAS, tiene a su disposición una metodología especialmente desarrollada para los mismos, la metodología SEMMA (*Sample – Explore – Modify – Model – Assess*).

(Vanrell, 2010) en su Trabajo Final de Especialista En Ingeniería en Sistemas de Información, para la Universidad de Buenos Aires, titulado "**Elementos para un Modelo de Procesos de Explotación de Información**", muestra una comparación entre las metodologías y a la vez, las implicaciones de los procesos de explotación de información en todos los niveles de una empresa.

Es necesario realizar una comparación de las principales características de las metodologías mencionadas, con el objetivo de elegir aquella que se adapte a las necesidades del proyecto. Esto no significa que la presente investigación busque evaluar en profundidad sus ventajas y desventajas, sino más bien, determinar cual servirá como pauta para el resto del proyecto.

A continuación, se exponen brevemente las principales características de las siguientes metodologías:

- a. CRISP-DM (*Cross Industry Standard Process for Data Mining*)
- b. SEMMA (*Sample – Explore – Modify – Model – Assess*)
- c. KDD (*Knowledge Discovery and Data Mining*)
- d. Metodología Catalyst (P3TQ)
- e. Metodología Berry y Linoff

### 2.2.1. CRISP-DM (Cross Industry Standard Process for Data Mining)

CRISP-DM, creada por el grupo de empresas SPSS, NCR y Daimler Chrysler en el año 2000, es actualmente la guía de referencia más utilizada en el desarrollo de proyectos de minería de datos. Estructura el proceso en seis fases: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación e Implantación (Chapman et al, 2000). Además de metodología, es también un modelo gracias a las fases propuestas y su interacción.

Proporciona un marco para la realización de proyectos de minería de datos que es independiente tanto del sector de la industria y la tecnología utilizada. El modelo de proceso CRISP-DM tiene como objetivo realizar proyectos de minería de datos, menos costosos, fiables, repetibles, manejables, y ágiles. (Wirth & Hipp, 2000)

Las seis fases pueden repetirse bidireccionalmente hasta que el modelo de análisis creado se considera que ha alcanzado los resultados esperados. (Chapman, 2000)

- a. Comprensión del negocio (Objetivos y requerimientos)
- b. Comprensión de los datos (Recopilación, Descripción Exploración)
- c. Preparación de los datos (Selección, limpieza, integración)
- d. Modelado (aplicación de técnicas)
- e. Evaluación (evaluar modelos)
- f. Despliegue

### 2.2.2. SEMMA (Sample – Explore – Modify – Model – Assess)

La metodología SEMMA (*Sample – Explore – Modify – Model – Assess*) fue creada por SAS (Compañía desarrolladora de software orientado a procesos estadísticos, de analítica y minería de datos) y presenta cinco fases: muestra, exploración, modificación, modelización y asesoramiento (Rohanizadeh y Moghadama, 2009)

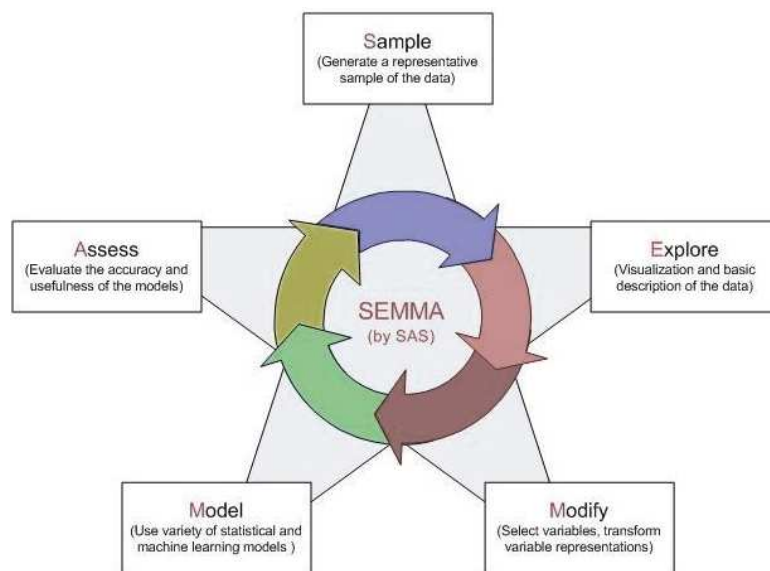


Figura 2 Fases SEMMA. Fuente SAS

Esta metodología está focalizada en las tareas de modelización. Consiste de cinco fases:

- **Sample** (Muestra): creación de una muestra significativa para el modelo.
- **Explore** (Exploración): comprensión de los datos buscando relaciones entre variables y anomalías.
- **Modify** (Modificación): transformación de las variables para las necesidades del modelo.

- **Model** (Modelización): aplicación de uno o varios modelos / técnicas sobre el conjunto de datos en la búsqueda de resultados.
- **Assessment** (Asesoramiento): evaluación de los resultados del modelo.

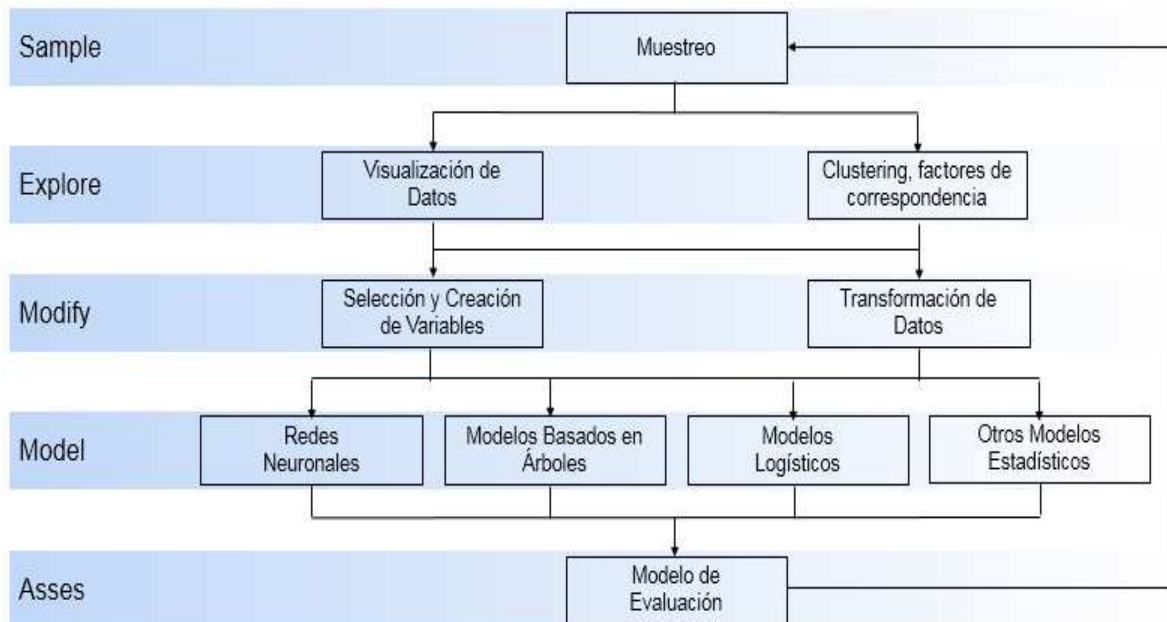


Figura 3. Fases de Metodología SEMMA. (Cattaneo et al, 2010)

### 2.2.3. KDD (*Knowledge Discovery and Data Mining*)

El objetivo principal de esta metodología es automatizar el procesamiento de los datos, permitiendo a los usuarios dedicar más tiempo a las tareas de análisis y al descubrimiento de relaciones entre los datos (Martínez, 2012)

KDD incluye la minería de datos como una de las etapas para obtener conocimiento en la organización, en total, plantea 6 fases (más una fase previa de generación de conocimiento) (Fayyad, 1996)

(Taié, 2008) utiliza el proceso de descubrimiento de conocimiento (KDD) para desarrollar su investigación “Desarrollo de una metodología de extracción de conocimientos a partir de datos de micromatrices de DNA basada en ontologías genéticas” y detalla las tareas de preparación, explotación y visualización de datos. En este caso, en la etapa de *Data Mining*, aplica técnicas como agrupamiento jerárquico y K-medias para análisis bioinformáticos.

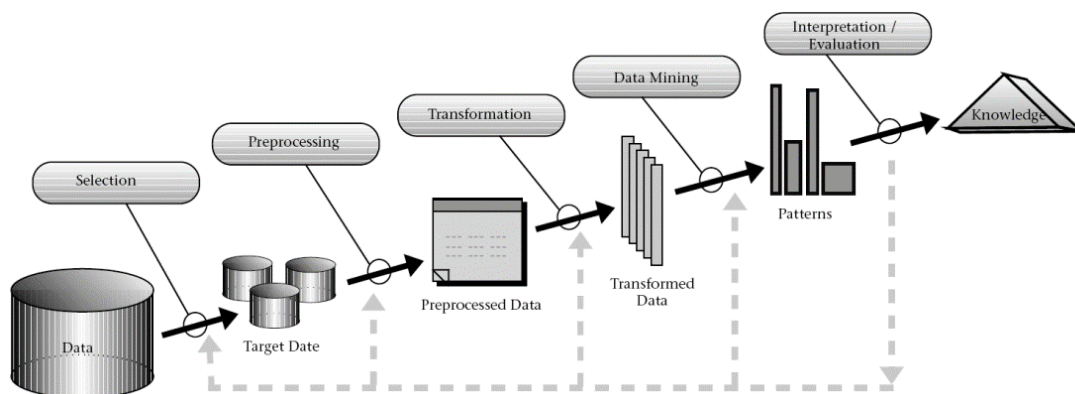


Figura 4. Proceso de Descubrimiento de Conocimiento. Tomada de (Fayyad, 1996)

Hace referencia al proceso de encontrar conocimiento en el dato y enfatiza el proceso de creación de aplicaciones de minería de datos.

- a. Pre KDD (Comprensión del negocio, requerimientos)
- b. Selección (Conjunto de datos)
- c. Pre-procesamiento y limpieza de datos (eliminación de ruido)
- d. Transformación (reducción, integración)
- e. Minería de datos (aplicación de técnicas de minería)
- f. Interpretación y evaluación
- g. Post-KDD (consolidar conocimiento adquirido)

Las etapas iniciales del proceso KDD son muy importantes porque son la base sobre la cual se ejecuta la minería de datos. Si la preparación de los datos no está bien hecha, los resultados obtenidos en los análisis no serán confiables (Martínez, 2012)

#### 2.2.4. Metodología Catalyst (P3TQ)

La metodología Catalyst, conocida como P3TQ (Product, Place, Price, Time, Quantity) está ganando cada vez mayor popularidad debido a su completitud y flexibilidad para adaptarse en distintos escenarios (Moine et al, 2010) y fue abordada por (Méndez & Rodríguez, 2009) en su Trabajo Profesional “**Herramienta de Estudio de Viabilidad para Proyectos que Utilizan la Metodología P3TQ**” en la Universidad de Buenos Aires, mediante el cual describe sus bondades y hace planteamientos a tener en cuenta en proyectos relacionados con minería. En este caso, se detallan las actividades realizadas en cada una de las etapas propuestas en la metodología. Plantea la formulación de dos modelos: el Modelo de Negocio y el Modelo de Explotación de Información. (Pyle, 2003)

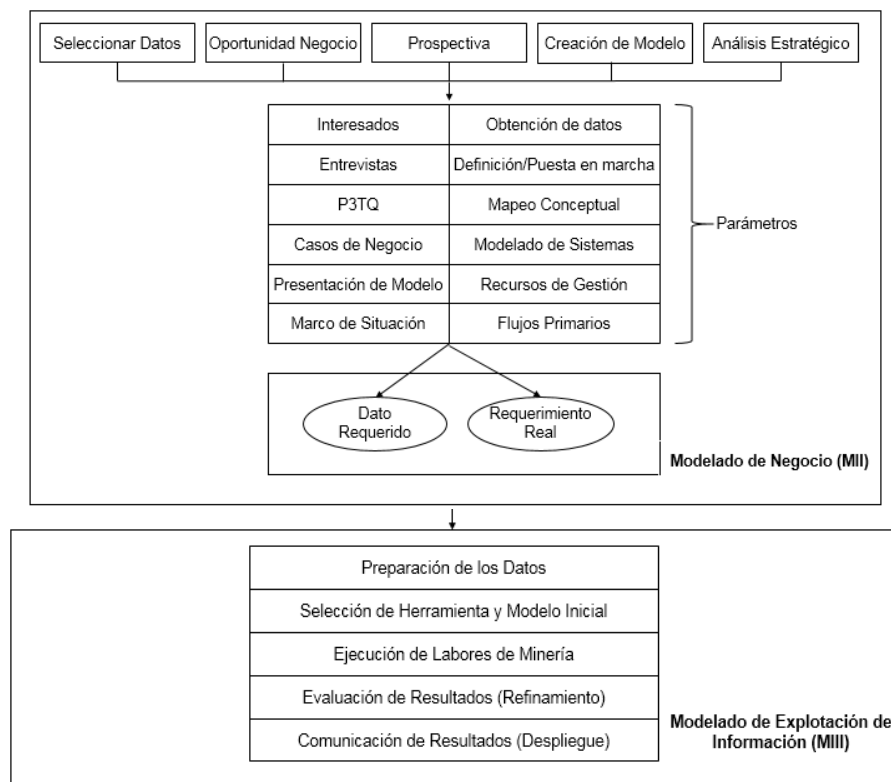


Figura 5. Fases de Metodología P3TQ  
Adaptado de (Cattaneo et al, 2010) y (Méndez y Rodríguez, 2009)

El primero de estos modelos “proporciona una guía de pasos para el desarrollo y la construcción de un modelo que permita identificar un problema de negocio o la oportunidad del mismo”, mientras que el segundo “proporciona una guía de pasos para la ejecución de modelos de Explotación de Información” (Vanrell, 2010)

En la descripción de esta metodología, se aprecia como las cinco variables de negocio que dan nombre a esta metodología, interactúan mutuamente (Méndez y Rodríguez, 2009) y (Pyle, 2003) afirma que una de las diferencias importantes entre P3TQ y las demás metodologías, es que considera quiénes son los interesados en el proyecto en la organización, considerando hasta la causa de su interés

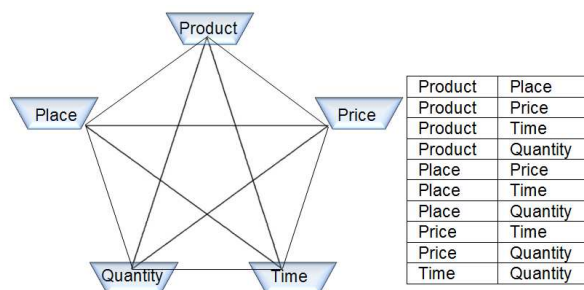


Figura 6. Interacción entre las cinco variables en metodología P3TQ  
Tomado de (Méndez y Rodríguez, 2009)

### 2.2.5. Metodología Berry y Linoff

A su vez, (Berry y Linoff ,1997) plantean como su metodología Berry y Linoff se vale de 10 fases como ciclo de implementación de minería de datos, incorporando actividades como: traducción de problema de negocio a problemas de (DM), selección apropiada de datos, conocimiento de datos, datos para el modelo, corrección de datos, transformaciones, creación del modelo, evaluación del modelo, despliegue y evaluación de resultados.

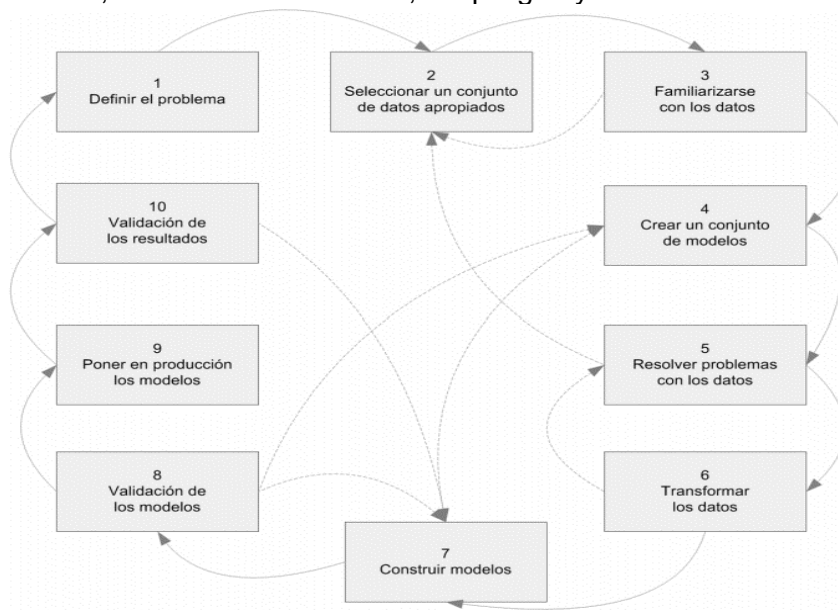


Figura 7. Fases de Metodología Berry y Linoff.  
Tomado de: (Gamarra, 2006)

### 2.3. Comparación de metodologías de minería de datos

Las diversas metodologías identifican técnicas de explotación de información utilizables. CRISP-DM identifica problemas de inteligencia de negocio. SEMMA y P3TQ no identifican problemas de inteligencia de negocio, ni relaciones entre técnicas de explotación de información. (Cattaneo et al, 2010)

El portal [www.Kdnuggets.com](http://www.Kdnuggets.com) en su sección *polls*, publica cifras de acuerdo a la metodología utilizada en proyectos de minería, y los resultados muestran una fuerte inclinación por la metodología CRISP-DM de acuerdo la siguiente figura, la cual además, representa la tendencia desde el año 2007:

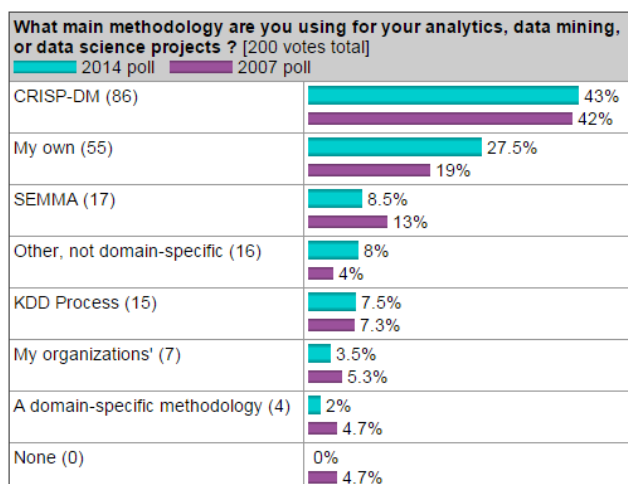


Figura 8. Metodologías utilizadas en proyectos de minería de datos 2014.  
<http://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html>

Al observar las estadísticas, se aprecia claramente que CRISP-DM, SEMMA y KDD, aparecen entre las cinco metodologías más usadas. Es por ello, que la comparación se centrará en estas tres opciones. Esto no significa que se desestime o desvirtúe el uso y alcances de las otras alternativas mencionadas, sino que se reduce el número de comparaciones a realizar en el presente proyecto.

SEMMA y CRISP-DM han crecido como estándares y definen un conjunto de pasos secuenciales que pretenden ser guía en la implementación de aplicaciones de minería de datos (Albarrán & Salgado, 2013)

La metodología CRISP-DM se describe en términos de un modelo de proceso jerárquico, que comprende cuatro niveles de abstracción (de lo general a lo específico): fases, tareas genéricas, tareas especializadas, e instancias de proceso y es descrita por (Cobos et al, 2010) como un proceso para el desarrollo de proyectos de minería de datos iterativo, abierto, personalizable y de gran reconocimiento por la industria y la academia.

La metodología CRISP-DM presenta el modelo de referencia y describe lo que se debe hacer en un proyecto de minería de datos, la Guía del usuario incluye sugerencias para cada fase (Wirth & Hipp, 2000), igualmente, de acuerdo a (Tomasello et al, 2011) CRISP-DM define los procesos y tareas que se deben realizar para desarrollar en forma exitosa un proyecto de explotación de información.

(Azevedo & Santos, 2008) concluyen en su estudio comparativo de metodologías, que SEMMA y CRISP-DM se pueden ver como una implementación del proceso de KDD y que primera vista, que CRISP-DM es más completo que el SEMMA. Recomiendan por ejemplo, integrar a la fase de muestra de SEMMA, la comprensión del dominio de aplicación, el conocimiento previo relevante y los objetivos del usuario final.

En la misma línea, (Rodríguez et al, 2003) argumenta que SEMMA se centra en técnicas de desarrollo del proceso, mientras CRISP-DM mantiene una perspectiva con respecto a los objetivos empresariales. Desde ese punto de vista más global se puede considerar que la metodología CRISP-DM está más cercana al concepto real de proyecto. Continúa

el mismo autor, resaltando que SEMMA está ligada a los productos de SAS y por su parte, CRISP-DM ha sido diseñada como una metodología neutra, libre y gratuita.

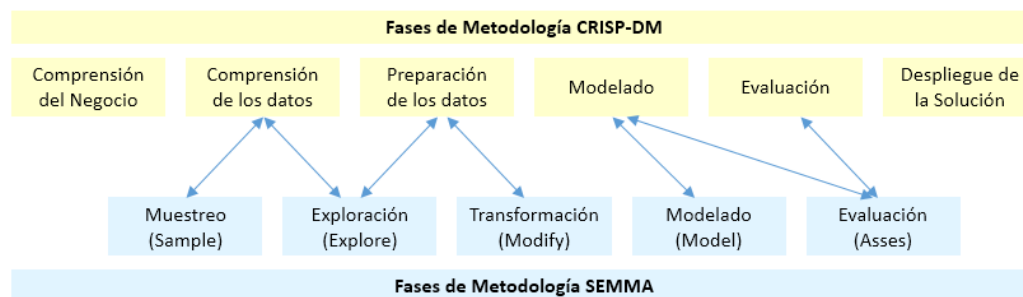


Figura 9 Comparativa interrelaciones metodología SEMMA vs CRISP-DM.  
Tomado de (Rodríguez et al, 2003)

Lo destacable de CRISP-DM es el detalle de las fases y la especificación de los subproductos, muchas veces informes, lo cual facilita las labores de seguimiento. (Hernández et al, 2005) quien resalta también la importancia de la primera fase “Comprensión del Negocio”, para el establecimiento de objetivos del negocio, evaluación de la situación, definición de objetivos de la minería de datos y la planificación del proyecto.

En los términos de (Salcedo et al, 2010) La utilización de la metodología CRISP en la construcción de un *Data Warehouse* proporciona confiabilidad, robustez y estandarización de los procesos y facilita la creación de futuros proyectos a través de planeación y seguimiento.

## 2.4. Selección de la metodología a implementar

Luego de realizar las respectivas comparaciones y tomar como referencia las cifras de uso presentadas por Knuggets, se determina CRISP-DM, como la metodología a utilizar durante la implementación del presente proyecto.

## 2.5. Descripción metodología CRISP-DM

**CRISP-DM (Cross Industry Standard Process for Data Mining) Proceso Estándar Industrial Híbrido para la Minería de Datos**

*El proyecto se realizará mediante la metodología CRISP-DM*, la cual se ha convertido en la más utilizada, según un estudio publicado en el año 2007 por la comunidad KDnuggets (*Data Mining Community's Top Resource*).

Esta metodología se enfoca en entender de la manera más completa posible el negocio y el problema que se desea resolver. Lo anterior permite hacer una correcta recolección de datos e interpretar bien los resultados de los análisis, alcanzando los objetivos que se hayan propuesto (Martínez, 2012)

### 2.5.1. Fases de la metodología CRISP-DM

Se describe a continuación y de manera breve las seis (6) fases planteadas. Puede apreciarse como se articulan entre sí para el logro de los resultados. En cada fase se definen sus entregables.

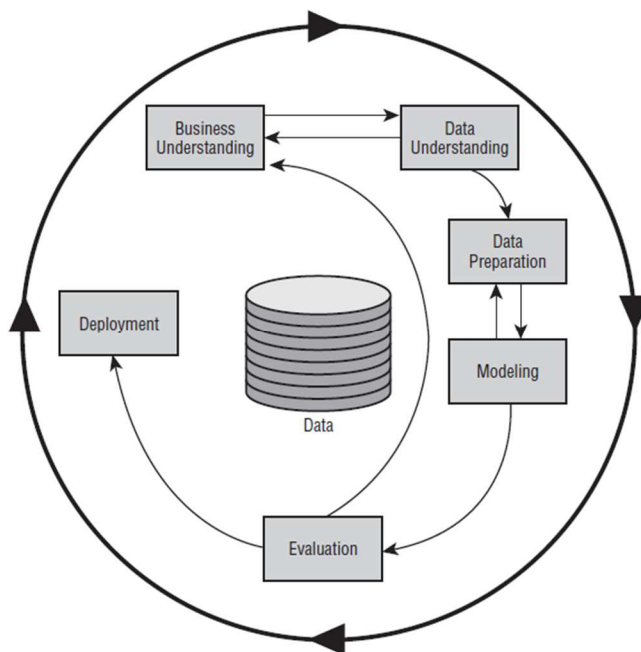


Figura 10 Modelo CRISP-DM  
[Chapman et al., 2000] Guía paso a paso de Minería de Datos.  
IBM Manual CRISP-DM de IBM SPSS Modeler

#### 2.5.1.1. Comprensión del negocio

Consiste en determinar los objetivos y requerimientos desde una perspectiva no técnica

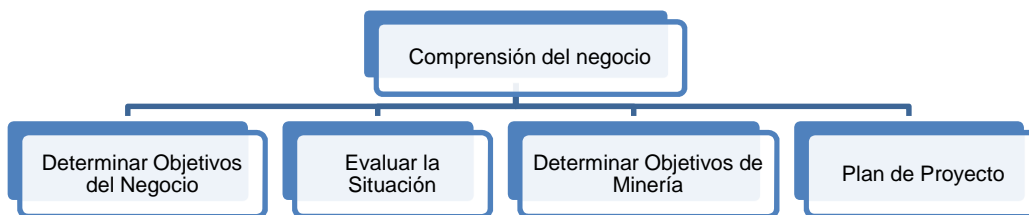


Figura 11 Metodología CRISP-DM. Fase I Comprensión del Negocio.  
Adaptado de [Chapman et al., 2000] Guía paso a paso de Minería de Datos y [Fernández, 2006]

- a) **Determinar los objetivos del negocio:** Se registran los objetivos del dominio de aplicación, los aspectos importantes del proceso que servirán como base para la ejecución del proyecto. Incluye los criterios de éxito.
- b) **Evaluación de la situación:** Se amplían los detalles del numeral anterior, se incluye el análisis de recursos como personas, datos y riesgos, así mismo se contemplan un diccionario de términos (del negocio y de minería).
- c) **Determinar los objetivos de la Minería de datos:** Se describe el tipo de problema de minería datos, como clúster, predicción, clasificación o visualización.
- d) **Plan de Proyecto:** se plantean las tareas del proyecto y se estiman los tiempos para cada una de las fases. Incluye los esfuerzos y recursos necesarios, resaltando puntos claves para lograr los resultados.

### Entregables de esta fase

Documentación en la cual se identifiquen los siguientes aspectos:

- Conceptos básicos sobre comprensión de la institución y el proceso de cirugía
- Objetivos del negocio
- Criterios de éxito para el cumplimiento de los objetivos
- Inventario de recursos (personal, expertos, recursos de TI, plataformas de hardware y software, herramientas para minería de datos)
- Riesgos y contingencias
- Terminología del negocio y del proyecto.
- Objetivos de la minería de datos
- Plan del proyecto

*Ver sección 3.6.1 para evidenciar el desarrollo de esta fase*

### 2.5.1.2. Comprensión de los datos

Consiste en familiarizarse con los datos teniendo presente los objetivos del negocio, acceder y explorar los datos para determinar su calidad.

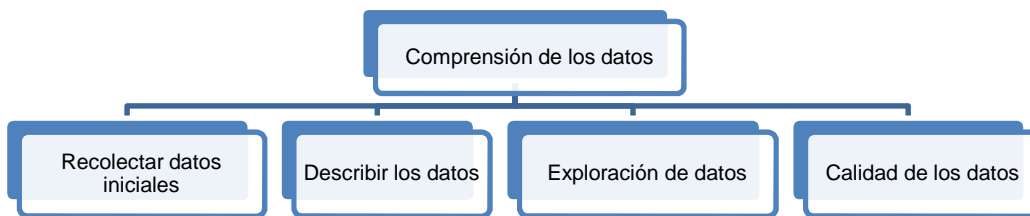


Figura 12. Metodología CRISP-DM. Fase II. Comprensión de los datos

Adaptado de [Chapman et al., 2000] Guía paso a paso de Minería de Datos y [Fernández, 2006]

### **a) Recopilación inicial de datos**

Establecer los tipos de datos con que se cuenta: *datos existentes* (transaccionales, registros...), *datos adquiridos* (datos externos que puede usar la institución) y *datos adicionales* (si es necesaria la creación de otros datos para complementar)

### **b) Descripción de los datos**

Describir cantidad de registros, cantidad de atributos, estadísticas para los conjuntos de datos. Tipos de valores que permitan establecer si los atributos son numéricos, categóricos, booleanos, lo cual será de gran utilidad en el modelado.

### **c) Exploración de datos**

Revisar los datos de tal manera que se puedan encontrar subconjuntos de datos interesantes, buscar errores en los datos, los tipos de EPS; usuarios, distribución geográfica, diagnósticos comunes de cirugía y en general, cualquier indicio que pueda brindar aplicabilidad a los datos en el proyecto de minería.

### **d) Verificación de calidad de los datos**

Verificar datos perdidos (null, 999, vacíos), errores en la introducción de datos clínicos, unidades de medida equivocadas, fallas en codificación de atributos, errores o ambigüedades. Listar los resultados de la verificación de calidad de datos; si existen problemas de calidad, listar las posibles soluciones. Las soluciones a los problemas de calidad de datos generalmente dependen tanto del conocimiento de los datos, como del negocio

### **Entregable de esta fase:**

Documentación que incluye los siguientes aspectos:

- Informe de recopilación de datos
- Informe de descripción de datos
- Informe de exploración de datos
- Informe de verificación de calidad de los datos

### **Formatos a Utilizar para la documentación de la Fase II**

*Ver sección 3.6.2 para evidenciar el desarrollo de esta fase.*

### 2.5.1.3. Preparación de los datos

Consiste en preparar el conjunto de datos adecuado

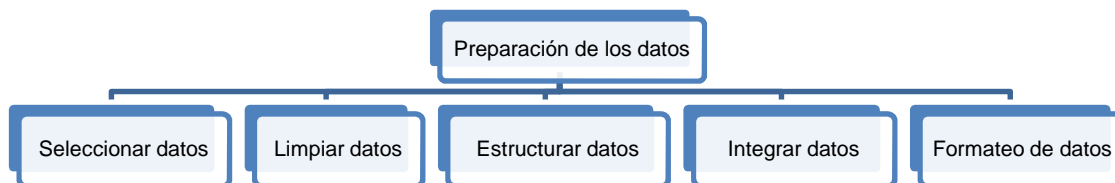


Figura 13. Metodología CRISP-DM. Fase III. Preparación de los datos  
Adaptación [Chapman et al., 2000] Guía paso a paso de Minería de Datos y [Fernández, 2006]

#### a) Selección de los datos

Decidir qué datos serán usados para el análisis. Los criterios incluyen la importancia a los objetivos de la minería de datos, la calidad, y las restricciones técnicas como límites sobre el volumen de datos los tipos de datos. Se deben seleccionar tanto los atributos (columnas) como los registros (filas) de las tablas. En caso de incluir o excluir o datos, se deben registrar las explicaciones respectivas.

#### b) Limpieza de datos

Elevar la calidad de los datos al nivel requerido por las técnicas de análisis seleccionadas. Esto puede implicar la selección de los subconjuntos de datos limpios, la inserción de datos por defectos. Se deben describir que decisiones y acciones fueron tomadas para los problemas de calidad de datos informados durante la tarea de Verificación de calidad de datos. Se deben considerar las transformaciones de los datos para una apropiada limpieza y el posible impacto en el análisis de resultados.

#### c) Construcción de datos

Realizar tareas que incluyen la construcción de operaciones como la producción de atributos derivados, el ingreso de nuevos registros, o la transformación de valores para atributos existentes. Los atributos derivados pueden ser por ejemplo campos calculados.

#### d) Integración de datos

Consiste en combinar o fusionar datos desde diferentes orígenes, tablas o registros. Se pueden realizar combinación de campos, tablas, archivos. Ejemplos: Creación de nueva tabla con total cirugías por cada EPS, glosas, valores facturados, pagados, pendientes de pago...

#### e) Formateo de datos

Se refiere a modificaciones principalmente *sintácticas* hechas a los datos que no cambian su significado, pero podría ser requerido por la herramienta de modelado. Describir si por

ejemplo, se requiere que la clave principal este al principio, o que los registros o campos estén en desorden, quitar comas en las evoluciones médicas, truncar la longitud de datos...

### Entregables de esta fase

- Conjunto de datos con sus respectivas transformaciones, fusiones, integraciones, de acuerdo a las exigencias de las técnicas de minería de datos
- *Data Warehouse – Data marts* (tablas de hechos y dimensiones) sobre el cual se estructurará la solución de inteligencia de negocios para el proceso de cirugía. (modelo de datos)
- Documento en el cual se registran las inclusiones, exclusiones, transformaciones, integración y formateo de los datos.

Se utilizará el **modelo dimensional constelación hechos** para la gestión de los datos y la construcción de *Data Warehouse*, de acuerdo a lo documentado en el apartado 3.3.3.

Ver sección 3.6.3 para evidenciar el desarrollo de esta fase.

#### 2.5.1.4. Modelado

Consiste en aplicar las técnicas de minería de datos a los conjuntos de datos

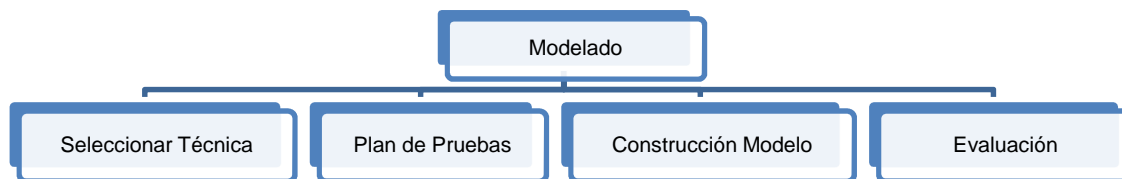


Figura 14. Metodología CRISP-DM. Fase IV Modelado.  
Adaptado de [Chapman et al., 2000] Guía paso a paso de Minería de Datos y [Fernández, 2006]

##### a) Selección de la técnica de modelado

Para el presente proyecto y de acuerdo a lo documentado en los capítulos previos, se aplicará la técnica de minería **clustering**, usando el algoritmo **K-Means** y la distancia de similitud **distancia euclidea**.

De igual manera se usarán herramientas de modelado para **inteligencia de negocios**, mediante la herramienta **WebQuery** y ofreciendo a los usuarios elementos de visualización.

**b) Generación de la prueba de diseño**

Establecer los mecanismos que se usarán para realizar las pruebas al diseño propuesto. Es decir, si se requieren datos de entrenamiento, prueba o validación.

**c) Construcción del Modelo**

Se documenta la aplicación de la técnica **clustering**, sus observaciones, pasos, parámetros. Como salida de esta tarea se debe contar *clusters* interpretados.

Se deben definir niveles de granularidad, uso de dimensiones, filtros, características de los cubos OLAP, tipos de visualizaciones. Se ejecutan los modelos propuestos y de tal ejercicio se obtienen resultados visibles.

**d) Evaluación del Modelo**

Luego de conocer los modelos iniciales, deben ser revisados con el objetivo de elegir los que serán considerarlos como finales. Estos modelos serán desplegados en el Hospital pues son los que arrojen los valores más relevantes para el proceso de Cirugía.

Se deberán revisar los datos arrojados en función de las actividades propias de la Gestión Estratégica del proceso de Cirugía y se utilizará el concepto de los expertos del proceso para calificar la pertinencia de dichos modelos

Igualmente se consideraran factores de usabilidad, de rendimiento, de tipo de despliegue (*active report, report bróker, OLAP, Dashboards, tecnología móvil*)

En esta tarea se debe analizar el impacto de los resultados obtenidos de acuerdo a los criterios establecidos y si cumplen o no con los objetivos planteados en la comprensión del negocio.

Es en ésta fase donde se determinar el apoyo que la solución le brinda realmente a la Gestión Estratégica del Proceso de Cirugía.

**Entregables de esta fase**

Documentación que describe los siguientes aspectos

- Descripción de técnicas de modelado elegidas
- Resultados de las pruebas de diseño
- Modelo construido
- Resultados de la evaluación del modelo
- Documentación técnica sobre el modelo (descripción del ambiente para solución de (BI), arquitectura, diagramas del modelo)

*Ver sección 3.6.4 para evidenciar el desarrollo de esta fase.*

### 2.5.1.5. Evaluación

Consiste en evaluar los modelos de la fase anteriores para determinar si son útiles a las necesidades del negocio. Se redactan los descubrimientos obtenidos a partir de los modelos y sus respectivas conclusiones, ya desde la óptica del hospital y su proceso de cirugía.

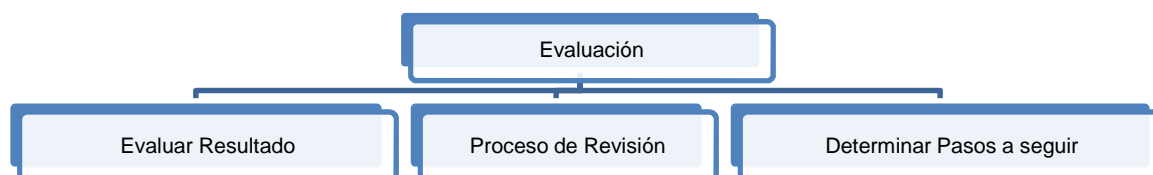


Figura 15. Metodología CRISP-DM. Fase V. Evaluación.  
Adaptado de [Chapman et al., 2000] Guía paso a paso de Minería de Datos y [Fernández, 2006]

#### a) Evaluación de los resultados

En esta evaluación se busca definir si los resultados realmente están alineados a los objetivos del proceso y se da respuesta a interrogantes como:

- ¿Los resultados se expresan con claridad y de forma que se puedan presentar con facilidad?
- ¿Se han realizado descubrimientos especiales o particularmente relevantes que deba resaltar?
- ¿En qué medida estos resultados se adaptan a los objetivos del Hospital en su proceso de Cirugía?
- ¿Qué cuestiones adicionales generan los resultados?
- ¿Cómo puede formular estas cuestiones en términos comerciales?

#### b) Proceso de revisión

Permite que la metodología cumpla con ciclos de mejoramiento. Cada fase debe ser revisada para resaltar lecciones aprendidas, errores, aspectos por mejorar. Pueden hacerse preguntas a las diferentes fases, tales como:

- ¿Esta fase ha contribuido al valor de sus resultados finales?
- ¿Existen formas de simplificar o mejorar esta fase u operación particular?
- ¿Cuáles fueron los fallos o errores cometidos en esa fase?
- ¿Cómo se pueden evitar la próxima vez?
- ¿Se han producido sorpresas (buenas y malas) en esta fase? A posteriori, ¿existe alguna forma de predecir esas instancias?
- ¿Existen decisiones alternativas o estrategias que se puedan utilizar en una fase concreta?
- ¿Construimos correctamente el modelo?
- ¿Existían otros atributos disponibles que no usamos?

### c) Determinación de los pasos siguientes

Según los resultados de la evaluación y la revisión de proceso, el equipo de proyecto decide cómo proceder. Si es necesario iniciar otras iteraciones, desplegar los resultados sobre el equipo de trabajo, y se tendrán en cuenta los recursos restantes y el presupuesto.

#### Entregables de esta fase:

- Evaluación del modelo por pares académicos
- Conclusiones luego de la revisión
- Plan de acciones a seguir en el proyecto

*Ver sección 3.6.5 para evidenciar el desarrollo de esta fase.*

### 2.5.1.6. Despliegue de la Solución

Explotar la utilidad de los modelos, integrándolos en las tareas de toma de decisiones del proceso de Cirugía del Hospital Departamental Universitario Santa Sofía de Caldas.

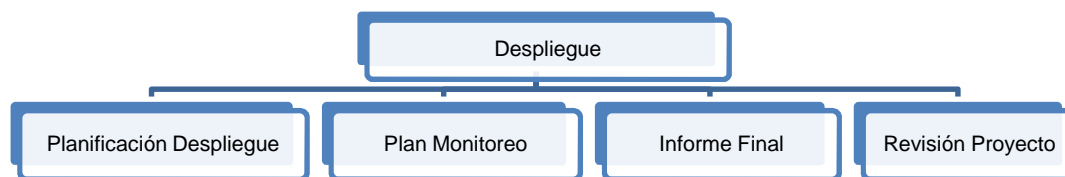


Figura 16. Metodología CRISP-DM. Fase VI. Despliegue de la Solución.  
Adaptado de [Chapman et al., 2000] Guía paso a paso de Minería de Datos y [Fernández, 2006]

#### a) Planificación del despliegue

Establecer cuáles serán las actividades a realizar para desplegar la solución en el proceso de Cirugía.

#### b) Planificación de la supervisión y el mantenimiento

Necesario para establecer los factores que se deben tener en cuenta para mantener vigente el modelo (esquemas de actualización de datos).

#### c) Creación del informe final

Documentar los pormenores del proyecto. Permite comunicar de manera eficiente los resultados y puede obedecer a temas como: descripción detallada, costo, desviaciones del plan inicial, resumen de resultados, recomendaciones futuras.

**d) Revisión del Proyecto**

Evaluar lo que fue correcto y lo que se equivocó, lo que fue bien hecho y lo que necesita para ser mejorado. Elaborar un resumen de las experiencias ganadas que puedan servir como base para otros proyectos similares de minería de datos.

**Entregables de esta fase:**

- **Modelo desplegado**
- **Manual de Usuario**
- **Evidencias transferencia de conocimiento**
- **Documento con el plan de despliegue, actualización, informe final y de revisión**

*Ver sección 3.6.6 para evidenciar el desarrollo de esta fase.*

## 2.6. Pruebas

Como argumenta (Boyer et al, 2010) en la implementación de proyectos de (BI), es necesario asegurar que una nueva funcionalidad, un nuevo informe, o una nueva actualización sea estable y entregue el enfoque adecuado para los usuarios. Para lograr la adopción y el nivel de confianza del proyecto, es necesario considerar que un informe con bajo rendimiento, información incorrecta o un resultado con errores, pondrá en riesgo el proyecto completo.

Las pruebas son el proceso de ejercitar al software con la finalidad de encontrar errores. En las aplicaciones Web se pueden considerar siete (7) aspectos: contenido, interfaz, navegación, componentes, configuración, desempeño y seguridad. (Presman, 2010)

De esta manera y luego de considerar el tamaño del proyecto, en enfoque que se tiene para el proceso de cirugía y el inventario de recursos tecnológicos con que cuenta la Institución, se estiman los siguientes tipos de prueba en la implementación del proyecto, en cuanto a lo relacionado con la aplicación Web (IBM WebQuery):

- a. **Pruebas de contenido:** permiten verificar los ámbitos sintácticos (ortografía, gramática y puntuación) y semántico (la exactitud de la información presentada, la consistencia de los datos publicados y la falta de ambigüedad).
- b. **Pruebas de interfaz:** ayudan a descubrir errores relacionados con los mecanismos específicos de la interfaz del usuario. Es decir, asegurar que las reglas de diseño, estética y contenido funcionen correctamente; garantizar que los navegadores web sean compatibles y que funcionan las opciones puestas a disposición del usuario. (Presman, 2010). Se deben probar los vínculos, los formatos utilizados en la solución (BI), ancho y tipos de datos adecuados, funcionamiento de menús desplegables, clasificación por niveles de datos, granularidad en los informes activos, funciones de ordenamiento y exportación.
- c. **Pruebas de facilidad de uso:** permiten calificar la manera en que el usuario percibe la complejidad en el manejo de la solución de (BI). El acceso a las opciones, la navegación a través de las opciones, la facilidad para utilizar por ejemplo, las dimensiones en la construcción de los informes dinámicos a través de OLAP. Así mismo, la complejidad que le representa al usuario utilizar las otras características como *report Broker* (envío automático de informes a correos), *Active Report*, uso de informes y reportes activos y el despliegue de informes a dispositivos móviles. Se miden factores como: Interactividad, plantillas, Legibilidad, Estética, características de despliegue, Sensibilidad en el tiempo, personalización y accesibilidad. (Presman, 2010)

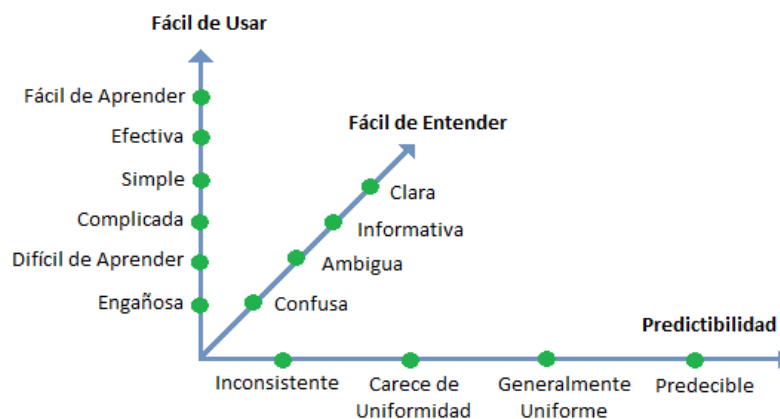


Figura 17. Valoración cualitativa Soluciones de Soluciones de Inteligencia de Negocios. Adaptado de (Pressman, 2010)

Así mismo se plantea la necesidad de realizar **pruebas de validación**, que a partir de los requerimientos iniciales del negocio y las características solicitadas identificadas durante la comprensión del negocio (fase I) permitan evidenciar cuando la solución de (BI), satisface las expectativas razonables de los usuarios. (Presman, 2010).

Se realizarán pruebas **alfa y beta**, con el objetivo de validar los resultados tanto en un ambiente controlado como de producción. (Presman, 2010)

Para ello, se realizarán pruebas que cubran cada uno de los requerimientos planteados. Basándose en los casos de uso descritos en le primera fase, se harán pruebas y se documentarán usando el formato de caso de prueba, en el cual se consignarán los pasos necesarios para producir los resultados, y se registrarán las observaciones correspondientes y la calificación de criterio de éxito

*Ver sección 3.7. Para evidenciar el desarrollo de esta fase.*

## 2.7. Presupuesto

Tabla 2.  
Presupuesto. Costos de Recurso Humano

COSTOS RECURSO HUMANO						
Nombre	Profesión	Rol	H/S	N/S	Vr. hora	Subtotal
Javier Hernández Cáceres	Ingeniero. Msc.	Director	4	28	36458	3.499.968
Fernando Uribe	Anestesiología	Cliente	2	10	93750	1.875.000
Coord. Cirugía	Jefe Enfermería Cirugía	Tester	2	12	20900	501.600
Gonzalo Calle	Director Científico	Cliente	2	8	33800	540.800
William Arias Betancurt	Gerente	Cliente	2	6	62500	750.000
Fabio López	Ingeniero Soporte Esp	Tester	2	16	36458	1.166.656
Luz Adriana Sánchez	Ingeniero Soporte	Tester	2	10	13541	270.820
Wilson Alejandro Rojas	Ingeniero	Autor	10	24	26800	6.432.000
TOTAL						15.036.844

H/S Horas por semana  
N/S Número de semanas en el proyecto

Tabla 3.  
Presupuesto. Costos Insumos y Materiales

COSTOS INSUMOS Y MATERIALES			
Descripción	Uso	Valor	Subtotal
Insumos de oficina y papelería	Documentación del proyecto	100.000	100.000
Tóner Impresora	Impresión de documentación fases del proyecto, pruebas, informes, formatos.	300.000	300.000
TOTAL			400.000

Tabla 4.  
Presupuesto. Costos Totales del Proyecto.

COSTOS TOTALES DEL PROYECTO	
Categoría de Costos	Valor
Costos de recurso humano	15.036.844
Costos Insumos y Materiales	400.000
Total Costos	15.436.844

### 3. REFERENTE TEÓRICO

#### 3.1 Información en el Sector Salud. Marco Legal

El Estado colombiano ha definido a través de diversas leyes, la obligatoriedad del reporte de información como instrumento de mejora continua del Sistema General de Seguridad Social en Salud. La Ley 100 de 1993, Ley 1122 de 2007, la Ley 1438 de 2011 y sus respectivos decretos reglamentarios y resoluciones, han definido un marco jurídico y legal que define la información como sustento para la toma de decisiones a nivel nacional y regional.

La Ley 1438 de 2011 en particular, establece la responsabilidad del Ministerio de Salud y Protección Social de establecer los mecanismos de medición, monitoreo y evaluación que permitan mejorar el desempeño de cada uno de los actores y sus resultados obtenidos. Igualmente, en su Artículo 12 “Articulación del Sistema de Información” define el plazo para la unificación de la Historia Clínica electrónica.

Así mismo, el Plan Nacional de Desarrollo 2014-2018 “Todos por un Nuevo País” continúa fortaleciendo el Sector Salud con el Sistema Integrado de la Protección Social (SISPRO) y el reporte de información como pilar fundamental para el desarrollo integral.

El Decreto 1011 de 2006 establece el Sistema Obligatorio de Garantía de Calidad de la Atención de Salud del Sistema General de Seguridad Social en Salud y a través de sus cuatro componentes: Sistema Único de habilitación (Resolución 2003 de 2014), Sistema Único de Acreditación (Resolución 2427 de 2014, Resolución 2082 de 2014 y Decreto 903 de 2014), Sistema de Información para la Calidad (Resolución 256 de 2016) y Auditoría para el mejoramiento de la Calidad, definen entre otros aspectos, el conjunto de datos e indicadores de reporte obligatorio que contribuye a la toma de decisiones a nivel nacional y a la estructuración de reportes periódicos. Particularmente, el anexo de la Resolución 123 de 2012 del Sistema Único de Acreditación, propone los estándares y criterios para lograr la excelencia en la prestación de servicios, y en su capítulo de Gerencia de la Información, incluye que los hospitales deberán implementar procedimientos de **Minería de datos** para la toma de decisiones orientada a mejorar la calidad en la atención al usuario y su familia.

La Superintendencia Nacional de Salud – SUPERSALUD- ha concebido la Circular Única con modificaciones de las Circulares Externas 049, 050, 051 y 052 de 2008, 057, 058 de 2009, 059, 060 061 y 062 de 2010, con el fin de propender por el control y vigilancia respectivo. De manera semestral so pena de las sanciones correspondientes, radica en su sistema electrónico los archivos planos con las variables predefinidas publicadas por los Hospitales.

Estos datos periódicos se suman a los exigidos por el Decreto 2193 de 2004 que de manera trimestral y anual, acumulan la cifras de productividad de los hospitales y junto con otros cuadros financieros, de contratación, talento humano y capacidad instalada,

permiten conocer los resultados obtenidos en la prestación de servicios de salud en todo el país.

### 3.1.1. Tipos de Sistemas de Información

Según (O'Brien, 2006) cuando los sistemas de información se enfocan en proporcionar apoyo para una toma eficaz de decisiones por parte de los directivos, se denominan **sistemas de apoyo a la administración** y propone una clasificación que incluye los **sistemas de información gerencial** (*MIS Management Information Systems*) los cuales ofrecen sistemas de reporte y pantallas para los directivos; los **sistemas de apoyo a la toma de decisiones** (*DSS Decisión Support Systems*) como apoyo directo en la toma de decisiones y con la posibilidad de simular escenarios y los **sistemas de información ejecutiva** (*EIS Executive Information Systems*) que proporcionan información desde una amplia variedad de fuentes internas y externas en pantallas fáciles de usar por gerentes y directivos.

Menciona el mismo autor en su libro "Sistemas de Información Gerencial", otras categorías como **sistemas expertos** y **sistemas de información estratégica** que ayudan a lograr ventajas comparativas y competitivas en el mercado. Todos estos sistemas basados en una diversidad de formatos fáciles de usar apoyan a los usuarios finales en un proceso particular de negocio y entregan información útil para basar sus decisiones. Gracias al pre-procesamiento de los datos pueden ofrecer un mejor rendimiento a la hora de visualizar los resultados.

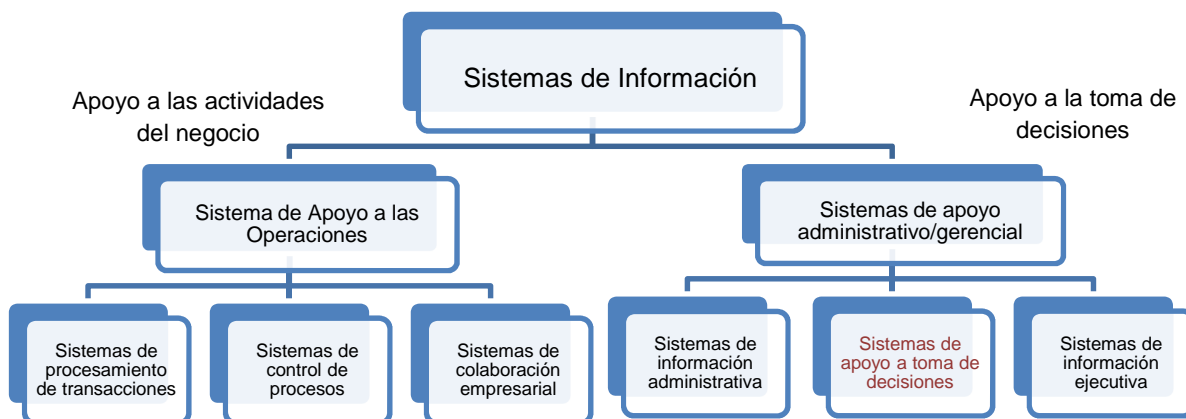


Figura 18. Clasificación de los Sistemas de Información  
Tomado de (O'Brien, 2006)

Los sistemas de información transaccionales son útiles en el registro de tareas, actividades y procesos diarios en las empresas. Permiten almacenar cada una de las

transacciones llevadas a cabo y conforman el repositorio de datos concerniente a la Historia Clínica de los pacientes, junto a sus registros administrativos.

Los Sistema de información transaccionales, se encargan de automatizar tareas y procesos que se realizan a diario en la empresa, manejando datos del funcionamiento de la organización. Se mantienen grandes cantidades de datos y a un nivel detallado (Martínez, 2012)

Son utilizados por los hospitales para apoyar las labores diarias, pero no son utilizados en la toma de decisiones de alto nivel, esto debido a que están diseñados para el registro de eventos, pero su rendimiento puede disminuir al gestionar grandes bases de datos para la generación de informes. A pesar de que en muchas Instituciones se cuentan con altos grados de sistematización, es común encontrar que los consolidados estadísticos son generados manualmente y aunque los registros electrónicos ofrecen la posibilidad de mejorar la oportunidad, calidad y continuidad en la atención clínica, es necesario aprovechar los datos estructurados y no estructurados. (West et. al. 2009)

Uno de los aplicativos más importantes del Hospital, es el software de Historias Clínicas y gestión financiera – Hosvital - , sistema transaccional con motor de base de datos DB2 de IBM que almacena un conjunto de datos invaluable. Aun así, (Dorda, 2002) criticó las limitaciones de la mayoría de los sistemas de información hospitalarios disponibles en el mercado, ya que según su experiencia, sólo ofrecen un conjunto predefinido de los procedimientos de recuperación de los datos de los pacientes, pero no presentan mecanismos flexibles para la recuperación de datos clínicos y su posterior análisis estadístico. Se han ido desarrollando aplicativos para la recuperación de datos de las historias clínicas con el fin de apoyar de alguna manera la información para la gestión estratégica.

Por su parte, (Rodríguez, 2011) en su trabajo de grado **“Diseño de un Sistema de Información Gerencial Alineado con la Orientación Estratégica de la Empresa para el Soporte en la Toma de Decisiones a Nivel Estratégico”** para optar al título de Magister en Ingeniería Industrial, para la Universidad Nacional de Colombia, Bogotá, hace énfasis en los Sistemas de Información Gerencial como herramienta para la toma de decisiones empresariales y acoge un ciclo de vida para su desarrollo el cual incluye las etapas de : planificación, análisis, diseño, implementación, uso y mantenimiento.

Plantea el diseño de elementos como entrevistas, definición de fuentes primarias y secundarias de datos y relaciona algunos de los formatos y matrices utilizadas en la documentación de la experiencia.

Además, incluye aspectos relacionados con herramientas aplicables a proyectos de inteligencia de negocios, como lo son *MicroStrategy*, *Cognos*, *Cisco Systems*, *Crystal Reports*, *Business Reports*, *SAP* y *SAS*. Utiliza criterios tecnológicos, financieros y de calidad, para la selección de una herramienta adecuada. Recomienda que las salidas del Sistema de Información Gerencial deben satisfacer los criterios de calidad, precisión, integridad, consistencia, completitud, validez, oportunidad y accesibilidad.

También resalta que es necesario generar un cambio total en la organización, capacitando, instruyendo y realizando los ajustes necesarios para que los usuarios vean y potencialicen las bondades de la solución.

### 3.1.2. Sistemas de Información en las Instituciones Prestadoras de Servicios de Salud

El Ministerio de Salud y La Protección Social ha dispuesto el Sistema Integral de Información de la Protección Social –SISPRO- que permite obtener, procesar y consolidar la información necesaria para la toma de decisiones que apoyen la elaboración de políticas, el monitoreo regulatorio y la gestión de servicios en cada uno de los niveles y en los procesos esenciales del sector: aseguramiento, financiamiento, oferta, demanda y uso de servicios<sup>8</sup> y aunque existen mecanismos para el reporte de datos por parte de los actores del Sistema General de Seguridad Social en Salud –SGSSS-<sup>9</sup> la labor de consolidar los datos de todo el país presenta dificultades que van mucho más allá de lo técnico.

Los procedimientos de captura, almacenamiento, transformación y envío de los datos, son difíciles de estandarizar en cada institución y por tal motivo, pueden generar brechas en la calidad de los reportes. A pesar de esto, el 97% de los registros reportados cumplen con criterios de validación y la base de datos que reposa en el Ministerio de Salud y Protección Social dispone de 2007 millones de registros validados y cargados entre los años 2001 a 2013. (MINSALUD, 2014) esto supone un gran esfuerzo de los Hospitales y EPS (Entidades Promotoras de Salud) para garantizar el envío de datos correctos.

Un sistema de información en salud es aquel que permite recopilar, integrar, procesar y ofrecer los instrumentos de reporte necesarios para gestionar todos los niveles del sistema y el mejoramiento continuo de los servicios de salud (OMS, 2004) es un conjunto de personas, datos, procesos, funciones, interfaces, redes y tecnologías que interactúan entre sí para apoyar y mejorar las operaciones diarias de la empresa, así como también la toma de decisiones (Álvarez, 2012)

En estudio “*Information systems in health sector in Colombia*” realizado por (Bernal & Forero, 2011) para la Universidad de los Andes y publicado en la Revista Gerencia y Políticas de Salud, se aprecia hace un recuento sobre la importancia de los sistemas de

---

<sup>8</sup> Bodega de Datos en la cual se concentra la información necesaria para la construcción de indicadores y reportes. Los datos provienen de fuentes de información internas y externas al Ministerio de Salud. [www.sispro.gov.co](http://www.sispro.gov.co)

<sup>9</sup> Ministerio de Protección Social. **Decreto 2193 de 2004**, Establece las condiciones y procedimientos para disponer de información periódica y sistemática que permita realizar el seguimiento y evaluación de la gestión de las instituciones públicas prestadoras de servicios de salud. Diario Oficial 45604 de julio 9 de 2004. **Resolución 3374 de 2000**. Por la cual se reglamentan los datos básicos que deben reportar los prestadores de servicios de salud y las entidades administradoras de planes de beneficios sobre los servicios de salud prestados.

información en salud, el Marco Normativo aplicable y los Ejes integradores establecidos por el Estado colombiano para que la información sea un pilar de desarrollo.

Los autores, mediante análisis de fuentes, componentes, flujos, necesidades de información e indicadores, identifican las debilidades, oportunidades, fortalezas y amenazas que enfrenta el desarrollo de los sistemas de información en salud en Colombia y manifiestan algunas recomendaciones, entre las cuales se pueden resaltar:

- a. El sistema de información debe soportarse en un marco jurídico y de principios de gobierno que lleven a su desarrollo. Hasta ahora, a pesar del contexto legal que define el funcionamiento, las instituciones no han tenido la fortaleza para hacer entrar en vigor lo contenido en la Ley.
- b. Un actuar organizacional orientado hacia la generación e intercambio de información es vital para el desarrollo de un sistema de información integral. Para lograr los objetivos y cumplir con los aspectos dispuestos por la Ley se deben desarrollar proyectos conjuntos que integren los datos de las diferentes organizaciones.
- c. El sistema de información en salud en Colombia, a pesar de ser una necesidad para la toma de decisiones informadas y el diseño de políticas, ha sido una de las grandes deficiencias del sector.

Igual que en otros sectores, el objetivo de un sistema de información en la salud, busca que los datos almacenados puedan ser transformados y aprovechados en la solución de problema de las Instituciones.

(Bonney, 2013), afirma que con el uso de la Historia Clínica Sistematizada, los sistemas de información hospitalarios (HIS), han ayudado a recopilar millones de registros basados en el proceso de atención, conformando repositorios de gran valor no sólo jurídico y legal, sino también de gran relevancia para nuevas investigaciones y avances en el sector de la Salud.

En ese sentido, (Prokosch & Ganslandt, 2009) resaltan como se están enfocando los sistemas de registro médico (EMR's), bodegas de datos y sistemas electrónicos de captura de datos (EDC) hacia el descubrimiento de conocimiento embebido en las historias clínicas y como retroalimentar en la enseñanza médica. (Kahn, Kaplan, Sokol & Dilaura, 2007) invitan a usar mecanismos que agilicen la búsqueda de información en las Historias Clínicas, para promover las labores investigativas y finalmente, contribuir a la gestión del conocimiento que requiere el sector de la salud para el desarrollo de guías y protocolos actualizados.

Para el caso del Hospital Santa Sofía, es fundamental el papel desempeñado por los (HIS) pues los datos registrados a partir de los eventos quirúrgicos pueden ser aprovechados para la definición de estrategias, el monitoreo y control de los resultados en el proceso de Cirugía. Más aún cuando la toma de decisiones está presente en todas las fases del proceso administrativo (planeación, organización, dirección y control) y por ende es una de las funciones que cumplen los directivos (Martínez, 2010)

## 3.2. Conceptos sobre almacenes de datos

Los almacenes o bodegas de datos, representan repositorios que contienen las transacciones u operaciones de un negocio. Pueden albergar datos históricos o permitir actualización en línea de acuerdo a la aplicación que se desee dar. Pueden ser soportados por diversas tecnologías en particular bases de datos para realizar operaciones de actualización o consulta. A continuación una breve descripción sobre algunos de los ejemplos como el Data Warehouse propuesto por (Inmon, 1997), Data mart propuesto por (Kimball & Ross, 2013) y detalles básicos sobre sistemas OLTP y OLAP.

### 3.2.1. Data warehouse

De acuerdo con (Inmon, 1997), un Data Warehouse es una colección de datos, orientados a hechos relevantes del negocio, integrados, que incluyen el tiempo como característica importante de referencia y no volátiles para el proceso de toma de decisiones. En este caso, datos relacionados con toda la empresa y agrupados respecto a los hechos. A continuación la descripción de algunas de las características que (Inmon, 1997) establece en un *Data Warehouse*:

- a. **Integración:** eliminando cualquier inconsistencia operacional existente entre los diversos sistemas de información de la empresa y garantizando una estructura consistente para los datos.
- b. **Temática:** organización de los datos por temas, agrupándolos para facilitar la comprensión al momento de los análisis y buscando coleccionar sólo aquellos datos que serán de utilidad en la gestión del conocimiento.
- c. **Histórico:** conformado por los diferentes valores que han tomado las variables en el tiempo, con el fin de ayuda en los análisis de tendencias y de comparación.
- d. **No Volátil:** los datos almacenados son de sólo consulta. En caso de agregar valores diferentes para las variables almacenadas, deberá hacerse sin modificar los datos existentes. Esto dará coherencia y confiabilidad a los análisis efectuados.

Para facilitar la gestión de un Data Warehouse es necesario documentar aspectos como: orígenes o fuentes de datos, periodicidad de actualización de datos, cálculos utilizados, niveles de confiabilidad, descripciones que permitan a los usuarios finales conocer qué tipo de información posee, su calidad y utilidad, ofreciéndole facilidades de acceso a partir de un lenguaje natural. De la misma manera, estas descripciones también conocidas como **metadatos** apoyan los procesos de administración, auditoría, especificación de interfaces y construcción de tareas de **ETL** (Extracción, transformación y carga) de su contenido. Los metadatos resuelven preguntas sobre la dimensión física de los datos (qué son, donde están ubicados, cuáles son las unidades de medida...)

En otra línea podemos citar a (Kimball, 2013) quien plantea la definición de Data Warehouse como una colección de datos en forma de una base de datos, que guarda y

ordena información extraída directamente desde los sistemas operacionales y datos externos. Expone la bodega de datos bajo el esquema de *Data marts*, los cuales representan conjuntos de datos específicos para las unidades de negocio y que pueden ser dependientes o independientes.

Tanto William Inmon, como Ralph Kimball, han contado con gran aceptación, siendo la primera utilizada para bodegas de datos corporativas, yendo de lo general a lo particular, mientras la segunda adoptada en procesos sistemáticos yendo de lo particular a lo general.

Ambos autores presentan los *Data Warehouses* (DW) como una fuente de información confiable y consolidada que está orientada al negocio, que proviene de distintos sistemas, se encuentra almacenado en un repositorio o base de datos y permite optimizar los tiempos de consulta por parte de los usuarios.

Para implementar un *Data Warehouse*, podemos hablar de los siguientes pasos:

- a. Extracción de los datos desde los diversos orígenes, su respectiva transformación de acuerdo a las necesidades u objetivos planteados.
- b. Carga de los datos a la estructura consistente y su actualización conforme a los periodos definidos.

Aprovechamiento o explotación de los datos, mediante recursos como *queries*, *reporting*, *OLAP*, *visualización*, *analytics*, minería de datos e interfaces a sistemas como *Executive Information Systems (EIS)* y *Decision Support Systems (DSS)*.

### 3.2.2. *Data mart*

Un *Data Mart* se puede considerar como una bodega de datos específica, estructurada para responder a las necesidades de un proceso de la empresa, o un conjunto de usuarios definido. Esto lo diferencia de los *Data Warehouse* cuya implementación se realiza para responder con un alcance de nivel corporativo, atendiendo a varias áreas del negocio a la vez (Kimball & Ross, 2013)

La ventaja de los *Data Mart* frente a los *Data Warehouses* tiene que ver con su implementación, ya que al ser de menor escala que los anteriores extraen información desde menos fuentes de datos y son más rápidos, fáciles y baratos de construir (Martínez, 2012)

Con la aparición de los *Data Marts*, surgen dos técnicas para la construcción de repositorios de datos (*data warehouse*)

- **Top-Down:** propuesto por William Inmon quien plantea constituir un gran *data warehouse* con todos los datos de la empresa y luego conformar varios *data marts* para uso de cada proceso.

- **Bottom-UP:** propuesto por Ralph Kimball, quien afirma que se puede ir construyendo *data marts* enfocados por procesos de la empresa y al final se tendrá un *data warehouse* corporativo.

La ventaja que tiene el enfoque *Bottom-Up* en relación al enfoque *Top-Down* es su rapidez de implementación, sin embargo se deberá prestar mucha atención y garantizar la integridad entre los *data marts* y la correcta relación entre ellos (Martínez, 2012)

### 3.2.3. OLTP

Los sistemas OLTP (*OnLine Transaction Processing*) surgen a mediados de 1970 y con el propósito de almacenar grandes volúmenes de datos y registrar las operaciones y transacciones empresariales en tiempo real. (Inmon, 2005) aunque no todos los datos que se guardan allí son útiles o relevantes para la generación de reportes gerenciales.

Los sistemas OLTP están diseñados para almacenar y modificar continuamente los datos, sin embargo no todos los datos son relevantes para la toma de decisiones. Están diseñados más para rendimiento en las aplicaciones que en los requerimientos de los usuarios para la consulta eficiente de información (Martínez, 2012)

De acuerdo a (Singhal, 2007) una de las características de los *Data Warehouse*, es precisamente integrar los datos provenientes de los sistemas manejadores de bases de datos relacionales (RDBMS) y de los sistemas OLTP, de tal manera que los ejecutivos de negocios puedan organizar, comprender y utilizar los datos para la gestión estratégica.

### 3.2.4. OLAP (On line Analytical Processing)

OLAP es un tipo de procesamiento de datos que se caracteriza, entre otras cosas, por permitir el análisis multidimensional e interactivo de la información de negocios a escala empresarial (Kimball & Ross, 2013). Consiste en combinar distintas áreas de la organización, y así ubicar ciertos tipos de información que revelen el comportamiento del negocio. (Molina & García, 2004) y Permiten que los *Data Warehouse* o *Datamarts* sean aprovechados en el soporte a las decisiones.

Los Sistemas OLAP trabajan con datos resumidos, la información se obtiene desde múltiples fuentes y la dispone de variados formatos, ya sea tablas, gráficos o reportes para que la utilicen áreas del negocio (Martínez, 2012)

Los OLAP soportan necesidades complejas de análisis, ayudan a los usuarios a realizar análisis de los datos desde diversas perspectivas o dimensiones y ofrecen herramientas para analizar grandes volúmenes de datos. Trabajan en línea a partir de datos resumidos y se basan en una estructura multidimensional, gracias a la cual presentan un tiempo más ágil de respuesta (Velásquez & Palade, 2008)

Una de sus mayores ventajas es la incorporación de dimensiones como (cliente, tiempo, espacio, producto, proveedor...) para lograr niveles de jerarquía a la visualización de informes. Básicamente, permiten observar unos datos concretos con sus respectivos atributos, frente a las dimensiones. (O'Brien, 2006)

Los usuarios de herramientas OLAP pueden pasar de perspectiva de negocio a otra, realizar simulaciones y relaciones de causa y efecto, por ejemplo, pueden estar observando los procedimientos quirúrgicos realizados en el Hospital, por día, semana o trimestre, año y pasar a detallar estos procedimientos por EPS (Empresa Promotora de Salud), regímenes de salud, regiones, especialidades o pacientes. Esta capacidad interactiva marca una gran diferencia con las herramientas simples de reporte.

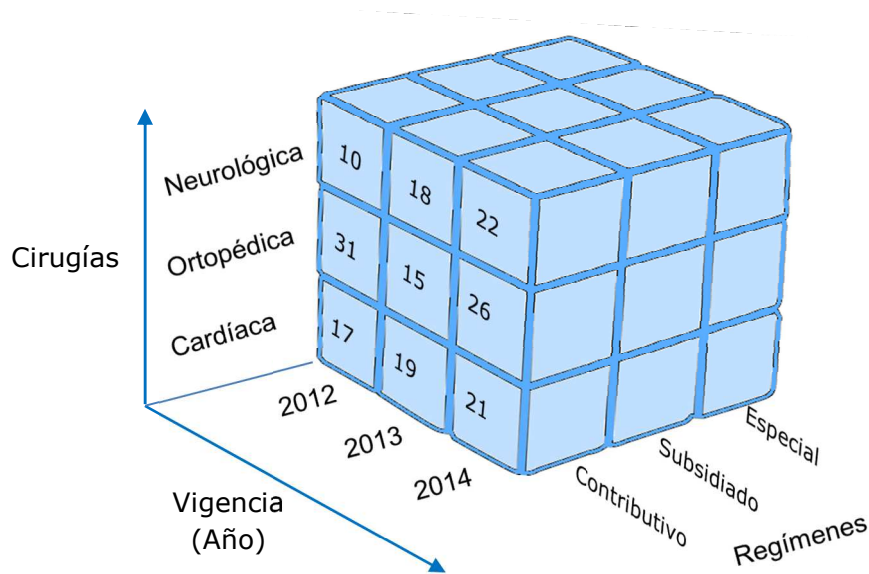


Figura 19. Representación OLAP producción de cirugías  
Elaboración propia

Se pueden realizar algunas operaciones para navegar entre las dimensiones, de tal manera que el usuario personaliza la vista de acuerdo a los datos que desee explorar. Entre ellas tenemos:

- Slice-and-dice:** corresponde a visualizar la información de un elemento en particular de una dimensión, Se trabaja con un subconjunto de los datos. Un valor determinado en un nivel de dimensión. Ejemplo: analizar toda la producción para el régimen contributivo de salud.
- Drill-Down:** permite ver la información del nivel inferior de la dimensión actual en una jerarquía definida, o sea, se muestran los datos detallados que en conjunto conforman el valor actual. Ejemplo, detallar uno de los años en trimestres.

- c) **Roll-up**: se encarga de pasar al nivel superior de la dimensión actual en una jerarquía definida, es decir, se consolidan los datos del nivel actual y se muestra el valor consolidado, correspondiente al nivel inmediatamente superior de la dimensión.
- d) **Pivot**: cambiar las dimensiones de posición. Con el objetivo de “rotar” las visualización, pero sin cambiar los datos correspondientes a las dimensiones.

Los cubos OLAP se incluyen como técnicas básicas en soluciones de inteligencia de negocios, pues son el último paso en el despliegue dimensional de los datos, incluso sobre la base de un esquema de estrella relacionado (Kimball & Ross, 2013)

Existen diferentes tipos de almacenamiento OLAP de acuerdo a su modelo físico de datos. Podemos definir:

#### a. MOLAP

Esta implementación almacena información pre-calculada en una base de datos multidimensional, optimizando los tiempos de respuesta. Utiliza una arquitectura de dos niveles: la base de datos multidimensional (BDMD) y el motor analítico. La BDMD es la encargada del manejo, acceso y obtención de los datos. El nivel de aplicación es el responsable de la ejecución de las consultas OLAP. (Tamayo & Moreno, 2010).

#### b. ROLAP

En este caso la información se guarda en una base de datos relacional, pero a diferencia del modelo entidad-relación, las tablas se conservan des-normalizadas. La des-normalización permite lograr un mejor nivel de consulta al disminuir *join*'s o conexiones entre las tablas de hechos y las tablas de dimensiones. (Kimball & Ross, 2010) Algunas representaciones de esto son el esquema en estrella (Star Schema) y el esquema en copo de nieve (Snowflake Schema). La principal ventaja de esta arquitectura es que permite el análisis de una enorme cantidad de datos.

En ROLAP se utiliza una arquitectura de tres niveles. La BD relacional maneja el almacenamiento de datos, el motor OLAP proporciona la funcionalidad analítica (capacidad de consulta), y alguna herramienta especializada es empleada para el nivel de presentación (Tamayo & Moreno, 2010)

#### c. HOLAP

Obedece a una arquitectura híbrida, en la cual algunos datos se conservan físicamente en modelos dimensionales y otros en modelos relacionales.

### 3.2.5. OLAP vs OLTP

Con el fin de contextualizar al lector, se presenta a continuación una comparación entre OLTP (*OnLine Transaction Processing*) y OLAP (*On line Analytical Processing*) propuesta entre por (Martínez, 2012)

Tabla 5.  
OLTP vs OLAP

OLTP	OLAP
Orientado a la operación diaria y al funcionamiento de las aplicaciones transaccionales.	Orientado al usuario que toma las decisiones del negocio.
La información se almacena en bases de datos relacionales, con datos normalizados, y se accede principalmente para insertar, modificar o eliminar datos.	La información se almacena en estructuras multidimensionales, y se accede para hacer consultas.
Muchos usuarios acceden a modificar los datos constantemente.	Los datos permanecen estáticos hasta su próxima actualización, los usuarios sólo acceden para lectura de los datos.
El tamaño de la base de datos incrementa rápidamente, por lo cual se le da preferencia a los datos más actuales. Se busca tener la mínima redundancia posible y consistencia en los datos.	El tamaño de la base de datos puede ser muy grande debido a la redundancia de datos. Los datos históricos y actuales son igual de importantes.

Tomado de (Martínez, 2012)

### 3.3. Modelos Dimensionales

Según (Kimball & Ross, 2013) el modelo dimensional es una técnica apropiada para el análisis de datos puesto que da respuesta a dos requerimientos fundamentales: a) Entrega información que eran incomprensibles a los usuarios del negocio y b) entrega consultas rápidas y de óptimo rendimiento.

El análisis multidimensional, permite examinar los datos de una organización, realizar comparativos, lograr niveles de detalle (granularidad) y gestionarlos desde diversas perspectivas. Busca entregar a los usuarios finales, una estructura común de información para apoyar la toma de decisiones estratégicas y tácticas (Martínez, 2012)

Consiste en modelar la información a través de **medidas**, **hechos** y **dimensiones**. Las medidas son los valores que puede tomar un dato en particular, las dimensiones son las

descripciones de las características que definen dicho dato y los hechos corresponden a la existencia de registros específicos para una combinación particular de dimensiones. (Abril & Pérez, 2007)

De acuerdo a (Tamayo & Moreno, 2010) encontramos las siguientes definiciones para los conceptos de hechos y dimensiones:

**Hechos:** el objeto a analizar, posee atributos llamados de hechos o de síntesis, y son de tipo cuantitativo. Sus valores (medidas) se obtienen generalmente por la aplicación de una función estadística que resume un conjunto de valores en un único valor. Por ejemplo: ventas en dólares, cantidad de unidades en inventario, cantidad de unidades de producto vendidas, horas trabajadas, promedio de piezas producidas, consumo de combustible de un vehículo, etcétera.

**Dimensiones:** representan cada uno de los ejes en un espacio multidimensional. Suministran el contexto en el que se obtienen las medidas de un hecho. Algunos ejemplos son: tiempo, producto, cliente, departamento, entre otras. Las dimensiones se utilizan para seleccionar y agrupar los datos en un nivel de detalle deseado. Los componentes de una dimensión se denominan niveles y se organizan en jerarquías, verbigracia, la dimensión tiempo puede tener niveles día, mes y año.

Las dimensiones suelen responder a preguntas de “quién, qué, cómo, cuándo, dónde y por qué” sucedieron los hechos y contienen la descripción de los atributos que luego permiten agrupar y filtrar los hechos en la solución de inteligencia de negocios.

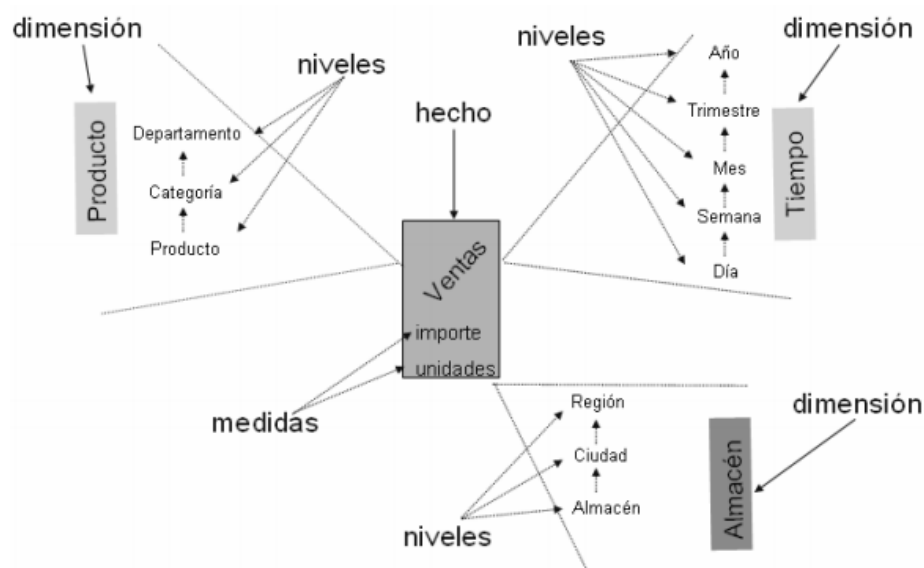


Figura 20. Representación gráfica de Modelo dimensional Tomado de (Tamayo & Moreno, 2010)

(Kimball & Ross, 2002) exponen en su libro “The Data Warehouse Toolkit” plantean una serie de buenas prácticas para lograr una implementación exitosa y detallan cuatro pasos para el diseño de modelos dimensionales:

- a. Seleccionar el proceso al cual se le va a diseñar el modelo dimensional
- b. Declarar el nivel de granularidad (detalle) que se contemplará en los datos
- c. Identificar las dimensiones que harán parte del modelo
- d. Identificar los hechos que se incluirán en el modelo.

Igualmente expresan que los **modelos dimensionales** pueden ser representados a través de sistemas manejadores de bases de datos relacionales o mediante bases de datos implementadas en entornos multidimensionales conocidos como *Online Analytical Processing* OLAP. Ambas implementaciones tiene un mismo concepto lógico, pero en el concepto físico se almacenan de manera diferente.

La representación dimensional con un modelo relacional puede realizarse bajo los siguientes esquemas:

- a) **Esquema en estrella** (*Star Schema*): está formado por una tabla de hechos y una tabla por cada una de las dimensiones requeridas
- b) **Esquema copo de nieve** (*Snowflake Schema*): es una variante del esquema en estrella que presenta las tablas de dimensión normalizadas.
- c) **Constelación de hechos**: son varios esquemas en estrella o copo de nieve que comparten dimensiones

En la siguiente imagen se observa la comparación realizada entre el **esquema en estrella** y los  **cubos OLAP**. Para el ejemplo, vemos como mediante el esquema en estrella, las dimensiones de mercado, fecha y producto se relacionan a los hechos (Ventas). Por su parte, mediante el cubo planteado podríamos hallar en sus intersecciones, los datos de mercado y fechas para un producto en particular.

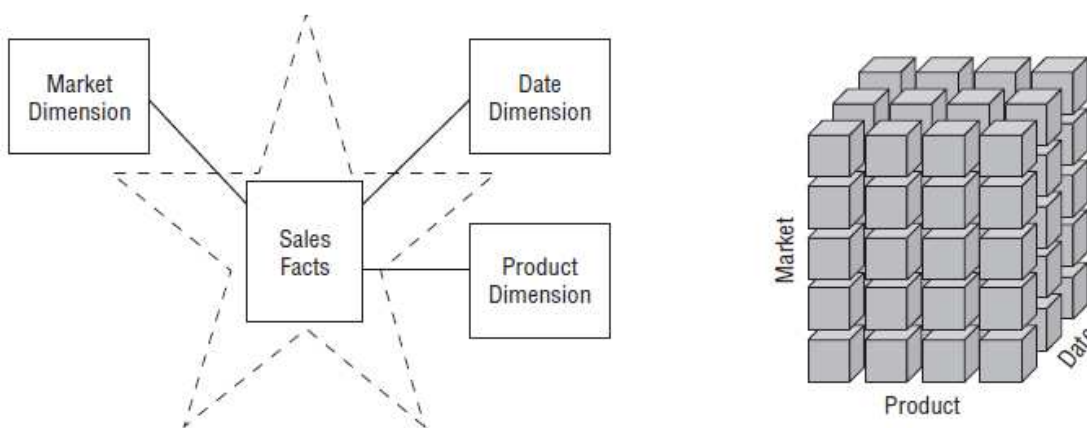


Figura 21. Star Schema vs OLAP cube  
Tomado de (Kimball & Ross, 2013)

La construcción de un modelo dimensional es una actividad dinámica e iterativa, en la cual se verifica constantemente el diseño y se compara con las necesidades identificadas del negocio. Se puede considerar iniciar con un modelo general y luego descender a uno detallado.

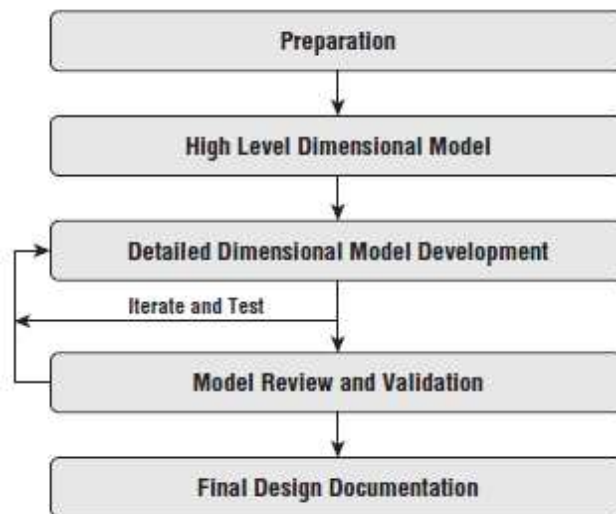


Figura 22. Diagrama de Flujo, proceso de modelamiento dimensional  
Tomado de (Kimball & Ross, 2013)

La **preparación** implica conocer los requerimientos del negocio, considerando el valor agregado del modelo y no sólo verlo como conjuntos de datos que darán solución a unos reportes. El **modelo dimensional de alto nivel**, establece los elementos generales que deben ser modelados a partir de los requerimientos definidos, el alcance y la granularidad. En esta instancia se pueden tener las dimensiones y hechos relevantes pero aun sin especificación completa. En la tercera etapa **desarrollo del modelo de datos detallado**, se definen las tablas de dimensiones y de hechos, sus atributos, transformaciones, fusiones, integraciones. Se realiza la carga de los datos al modelo. Esta etapa es iterativa y a partir de la **revisión y validación**, se recurre a detalles adicionales en el modelo hasta lograr el mejor resultado. Finalmente se **documenta el diseño final**

### 3.3.1. Esquema en Estrella

Bajo este esquema, cada proceso de negocio está representado por un modelo que consiste en una tabla de hechos (principal) conocida también como *fact-table* unida a varias que la rodean (*dimensional tables*). La tabla de hechos contiene generalmente, valores numéricos que representan los eventos ocurridos, mientras que las tablas de dimensiones contienen el contexto en el cual sucedieron. (Kimball & Ross, 2013).

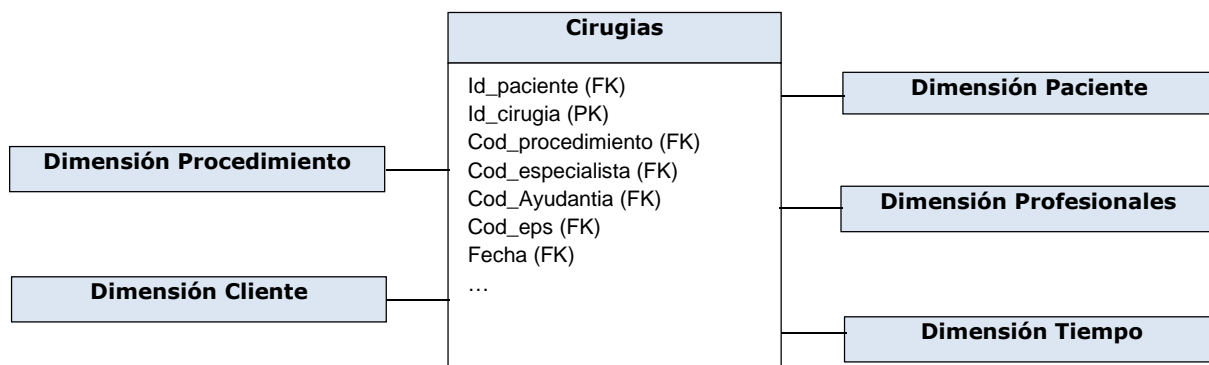


Figura 23. Esquema en Estrella (Modelo de datos Cirugía)  
Elaboración propia

Se observa en la Figura anterior, como la tabla de hechos es el conjunto de eventos quirúrgicos que se han realizado en el hospital, y alrededor se identifican algunas de las dimensiones que pueden explicar el contexto de cada una de las cirugías realizadas. Ampliando de esta manera los datos correspondientes a los procedimientos, pacientes, clientes y tiempo.

### 3.3.2. Esquema copo de nieve

Cuando se normaliza una jerarquía en una dimensión y aparecen tablas secundarias conectadas mediante atributos, se crea una estructura multinivel conocida como *Snowflake* (Copo de nieve) y a pesar de que representa los datos con precisión, puede dificultar el entendimiento de los datos para el usuario. De igual manera puede influir de manera negativa el rendimiento de las consultas. (Kimball & Ross, 2013).

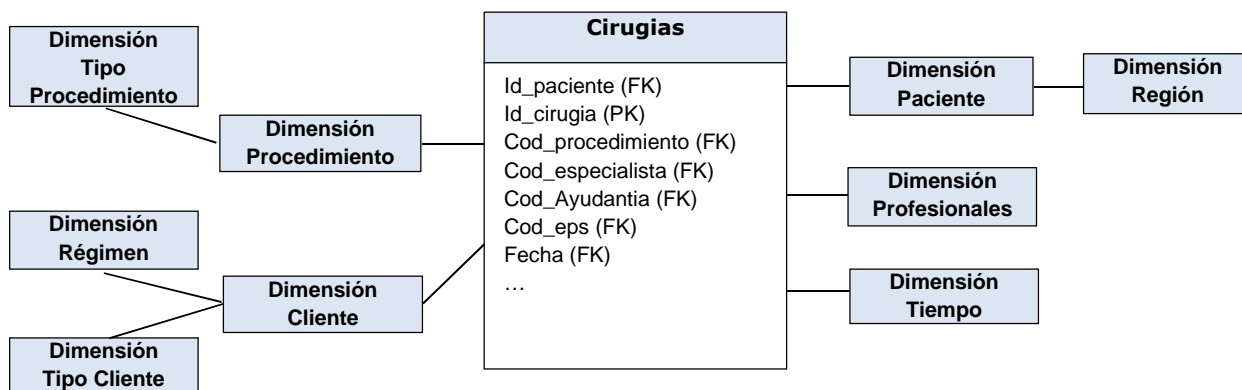


Figura 24. Esquema Copo de Nieve. (Modelo de datos Cirugía)  
Elaboración propia

### 3.3.3. Esquema de Constelación de hechos

Consiste en incluir en el modelo de datos, varias tablas de hechos y que a la vez, comparten las tablas de dimensiones especificadas, este reutilización de dimensiones es lo que permiten que los hechos se integren a nivel de toda la organización y que el modelo sea escalable. (Kimball & Ross, 2013)

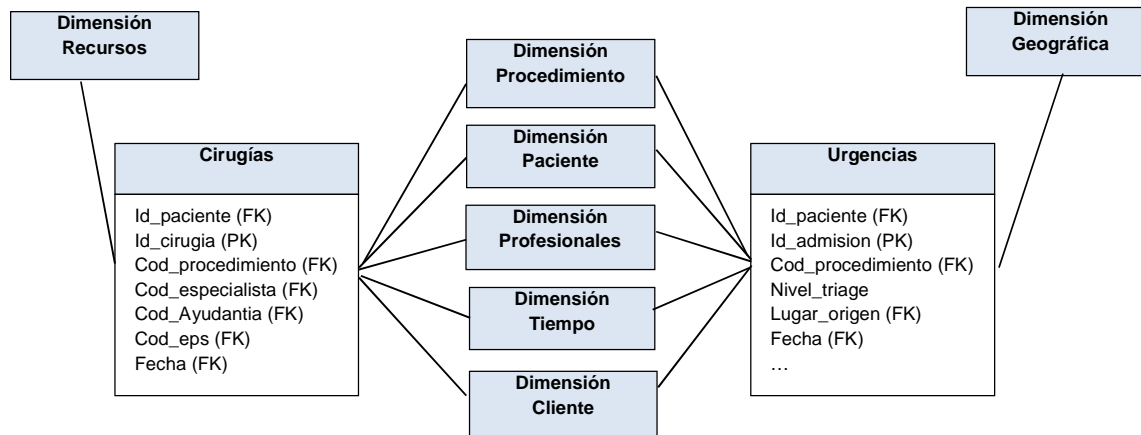


Figura 25. Esquema Constelación de Hechos. (Modelo de datos Cirugía)  
Elaboración propia

## 3.4. Inteligencia de Negocios

En un artículo publicado por International Business Machines Corp. (IBM), (Luhn, 1958) utilizó el término "*Business Intelligence System*" para referirse a un sistema automático que acepta información en su formato original, disemina los datos adecuada y rápidamente a los lugares correctos. A partir de allí, se conocen diversas definiciones para las soluciones de Inteligencia de Negocios, todas ellas resaltando su capacidad para la gestión de grandes volúmenes de información.

Según su aporte, el objetivo de un *Business Intelligence System* es "proporcionar la información adecuada para que las personas, equipos de trabajo, departamentos, divisiones y empresas puedan realizar sus actividades de manera eficiente."

En la actualidad la generación de informes claros, concisos y ante todo veraces, con base en los sistemas de información de las empresas, es fundamental en la toma de decisiones. (Salcedo et al, 2010). El volumen de datos que genera la sociedad actual continúa creciendo. El uso masivo de las redes sociales, dispositivos móviles, el internet de las cosas y las ciudades inteligentes, fortalecen aún más el concepto de Big data. Se hace necesario contar con sistemas que permitan analizar los datos y convertirlos en información útil. (Curto & Conesa, 2011)

A través de BI, las organizaciones cuentan con una nueva forma de ver sus datos, fortaleciendo gran variedad de aspectos como: planeación estratégica, inteligencia empresarial, finanzas, análisis de mercados, análisis de perfiles de usuarios y modelos de madurez de sus procesos que puedan alinear sus acciones hacia los objetivos del negocio. (Cleven et al., 2014)

Un sistema de BI puede procesar y analizar un gran volumen de datos, de diversos orígenes, la información consolidada sobre pacientes, procedimientos, flujo financiero, insumos, producción y resultados. Sin importar el número de datos, pueden ser representados a través de múltiples esquemas de visualización, identificando problemas y tendencias gracias a su granularidad. Gracias a las herramientas de (BI), el análisis de los datos no se reduce sólo a algunos gráficos estáticos, sino que además, pueden ser estudiados en profundidad, pasando de un escenario general a sus detalles específicos. (Luminița & Magdalena, 2012).

En algunos casos las herramientas analíticas son planteadas a través de software, fortaleciendo las capacidades de almacenamiento y procesamiento y así mismo se habilita la utilización de técnicas complementarias como análisis estadísticos, software de minería de datos y sistemas expertos (Martínez, 2010)

Una aplicación de Inteligencia de negocios, puede ser tan simple como una herramienta para consulta ad hoc, o tan compleja como un conjunto de técnicas de minería de datos. Se puede acceder a los datos mediante plantillas parametrizables que no requieren que los usuarios deban construir directamente las consultas. Siempre se busca que una mayor cantidad de usuarios pueda acceder a los datos.

Para (Porter, 1980) el diseño de sistemas inteligentes que permitieran analizar las enormes cantidades de datos provenientes de diversas fuentes, era clave en el desarrollo de estrategias competitivas. Incluyó además, estos sistemas como elemento fundamental para orientar la administración hacia la toma de decisiones en cuanto al mercado, los productos, los proveedores y los competidores.

Estos elementos de análisis constituyen el modelo cinco fuerzas de Porter, el cual visto desde la óptica de los sistemas de inteligencia de negocios permitió dividir los elementos sujetos de análisis bajo un modelo enfocado en la estrategia. En últimas permite identificar claramente qué tipo de información es la que se requiere y a partir de ello es posible identificar y dar manejo a las fuentes desde las cuáles esta información debe ser provista (Martínez, 2010)

Para (Negash & Gray, 2008) Los sistemas de inteligencia de negocios son herramientas que combinan la obtención y almacenamiento de datos, así como la gestión del conocimiento con herramientas analíticas que presentan información compleja y competitiva a los planificadores y decisores, permitiendo la estructuración de informes, funciones *drill down*, *drill Up*, *pivot*, análisis en tiempo real y predictivo.

En el ámbito de la Inteligencia de Negocios se puede encontrar implementación de sistemas relacionados con sistemas de almacenamiento y gestión de datos (*data warehouse*), sistemas de extracción, transformación y carga de datos (ETL), cubos para

análisis de información (*OLAP*), informes y reportes, cuadros de mando integral basados en indicadores (*balanced scorecard*), *analytics* y *big data*. (Curto & Conesa, 2011)

En particular el *Balanced Scorecard* propuesto por (Kaplan & Norton, 1992) incluye la aplicación de sistemas de información para la medición de resultados basados en indicadores. Enfoque aún utilizado en las empresas para gestionar por resultados y que fue más allá de la perspectiva financiera, incluyendo procesos internos, aprendizaje y el cliente.

(Popovič, Hackney, Coelho & Jaklič 2012) identifican la madurez de BI en términos de eficacia de integración de datos ETL (extracción, transformación y carga de los datos), y las capacidades analíticas como OLAP (*Online Analytical Processing*) y visualización de informes. Por supuesto, se hace énfasis en la calidad de los datos sobre los cuales se apliquen las técnicas, como factor de éxito para el aprovechamiento de BI y DM. (Isik, Jones & Sidorova 2011)

Para (Davenport, 2007), inteligencia de negocios significa “Utilizar los activos de datos para tomar mejores decisiones de negocios, con un sistema de información que provea un conjunto de tecnologías y productos para abastecer a los usuarios con la información que ellos necesitan para responder las preguntas de negocio y tomar decisiones tácticas y estratégicas.” En este sentido, también expresan (Salcedo et al, 2010) en su artículo “Metodología CRISP para la implementación Data Warehouse”:

*“La tendencia hacia la que apunta la ‘inteligencia de negocios’ es la divulgación de la información, tanto a nivel gerencial como a todo aquel que la necesite desde diferentes dimensiones y niveles asociados, para lograr obtener informes consolidados o detallados que faciliten la síntesis de determinado proceso empresarial y que repercutan directamente en la toma de decisiones, objeto que en últimas constituye el objetivo mismo de los Data Warehouses”*

### **3.4.1. Arquitectura de una solución de Inteligencia de Negocios**

La arquitectura de una solución de inteligencia de negocios, muestra la interacción y función de cada uno de los componentes que la conforman. Las organizaciones han implementado sistemas a partir de diversos planteamientos, buscando con ello alcanzar los mejores resultados.

A continuación expresaremos cuatro enfoques de arquitectura, sin profundizar en ellas, pero buscando ilustrar al lector.

#### **3.4.1.1. Arquitectura Howard Dresner**

En su libro “*The Business Intelligence Competency Center: An Essential Business Strategy*”, (Dresner, 2002) especifica cuatro capas para una solución de inteligencia de negocios.

- a. **Capa de Infraestructura (*Infrastructure Layer*):** la cual incluye los datos integrados que luego son aprovechados por las soluciones analíticas. Incorporar un concepto de una bodega de datos operacional (*Operational Data Store*) con el objetivo de mantener un nivel alto de actualización de los datos y por lo tanto información en “tiempo real”.
- b. **Capa de Funcionalidad (*Functionality Layer*):** Conjunto de aplicaciones y herramientas de análisis, consulta y generación de reportes.
- c. **Capa de organización (*Organization Layer*):** Se fomenta la cultura y las habilidades de análisis. Se establecen metodologías de BI, el orden o prioridad en el cual se implementarán las herramientas de acuerdo a los objetivos iniciales. Se elevan los niveles de madurez y del enfoque hacia el negocio.
- d. **Capa de Tendencia del Negocio (*Business Trend Layer*):** enfoque hacia la competitividad, nuevos mercados (globalización)

### 3.4.1.2. Arquitectura propuesta por William Inmon

William Inmon propone que desde los recursos transaccionales se efectúe una labor de adquisición de datos (ETL), llevándolos a una gran estructura denominada Data Warehouse Corporativo (EDW *Enterprise Data Warehouse*) dicha estructura se encuentra normalizada (3NF), puede ser consultada directamente por las aplicaciones de inteligencia de negocios, y a la vez entrega datos para la construcción de data marts dimensionales por departamento.

Podemos observar en la siguiente figura, como el EDW interactúa tanto con la capa de recursos y orígenes de datos, como con la capa de presentación y aplicaciones de BI.

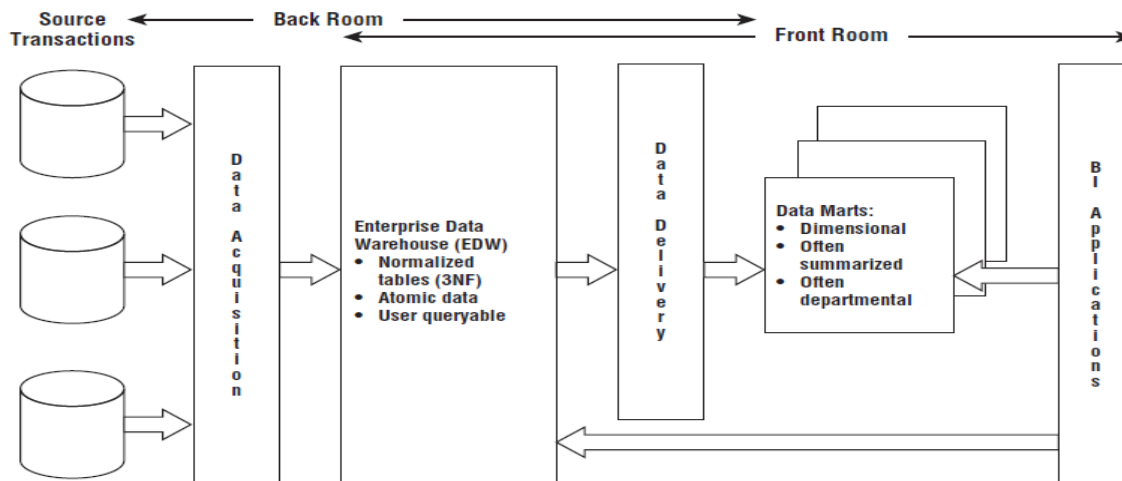


Figura 26. Arquitectura de Data Warehouse Corporativo (William Inmon)  
Tomado de (Kimball & Ross, 2013)

Esta arquitectura parte de conformar una “gran” estructura de datos y a partir de ella, obtener conjuntos de datos por departamento, es decir, de un todo hacia sus partes. Para esta arquitectura, la normalización (3NF) es obligatoria.

### 3.4.1.3. Arquitectura de Data Marts independientes

Como podemos ver en la figura, esta arquitectura parte de la idea de que cada *data mart* es independiente de los demás. Cada uno de los departamentos de la empresa crea su propio modelo de datos sin contemplar los demás. Esto conlleva a mayor inversión de tiempo y sobre todo, a que los datos no queden estandarizados en toda la empresa. Pueden aplicar reglas de ETL (Extracción, Transformación y Carga) distintas a un mismo tipo de dato.

Esta arquitectura puede dar solución a un departamento o área de la empresa en particular, pero no logra ofrecer herramientas de análisis a nivel directivo integral.

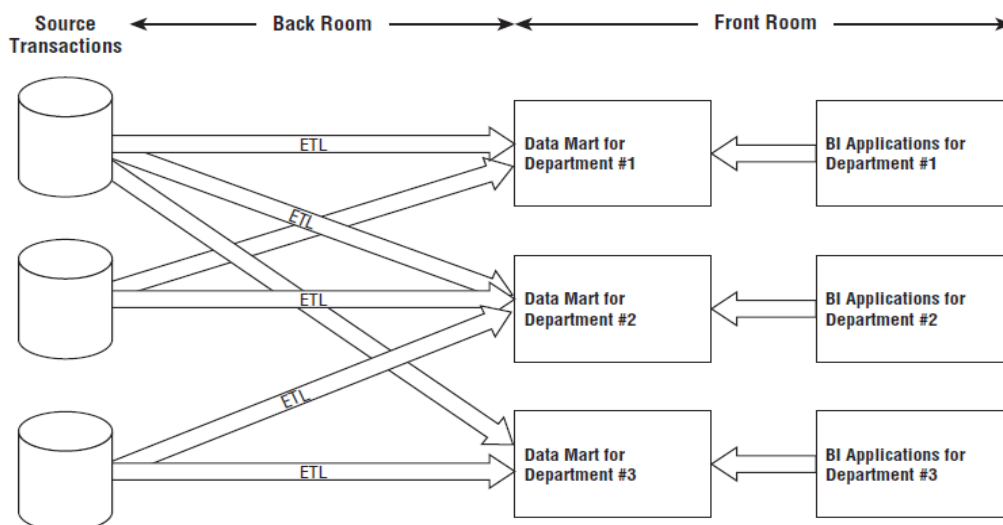


Figura 27. Arquitectura Data Mart Independientes  
Tomado de (Kimball & Ross, 2013)

### 3.4.1.4. Arquitectura propuesta por Ralph Kimball

De acuerdo a (Kimball & Ross, 2013) es posible sintetizar los elementos claves de la arquitectura en cuatro elementos:

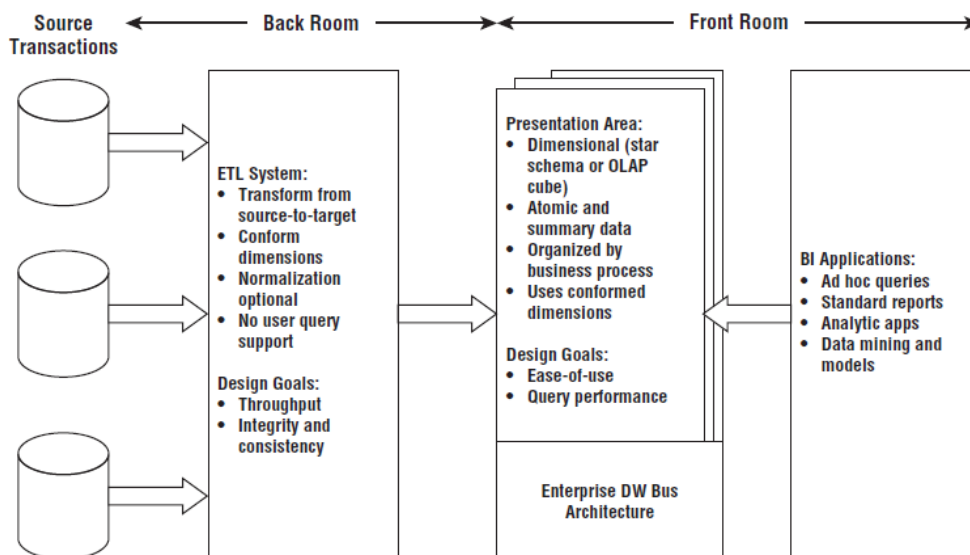


Figura 28. Elementos principales de la arquitectura DW/BI (Ralph Kimball)  
Tomado de (Kimball & Ross, 2013)

## Sistemas fuente

Representan los orígenes de datos. Pueden ser sistemas transaccionales que capturan todos los datos de la empresa en sus actividades cotidianas. Enfocados al registro y con pocos niveles de consulta. Aunque pueden almacenar datos históricos, no es su objetivo pues buscan mejor rendimiento en las operaciones de escritura. Pueden contemplarse otros repositorios, tales como hojas de cálculo, archivos de texto y otros formatos.

## Extracción, Transformación y Carga (ETL)

La **extracción** es el primer paso para insertar datos en un *data warehouse*. Consiste en leer los datos desde los recursos fuentes. Después de la extracción, cada dato puede ser sometido a diferentes **transformaciones** (limpieza, unificación, división, estandarización de formatos, combinación...) que les agreguen valor para el análisis posterior. Por último los datos transformados son **cargados** en el modelo dimensional, conformando las tablas de hechos y eventualmente actualizando el contenido de las tablas de dimensiones.

De su implementación adecuada dependen la integridad, uniformidad, consistencia y disponibilidad de los datos utilizados en el componente de análisis de una solución de BI (Bustamante et al, 2013)

Consiste por ejemplo, en importar los datos de los sistemas operacionales y bases de datos transaccionales, (por ejemplo, el sistema de información hospitalario HOSVITAL, el cual cuenta con un repositorio en el motor de bases de datos DB2) y los carga en el data Warehouse de la solución de inteligencia de negocios, convirtiéndolos en datos “que pueden tener un significado para los líderes del proceso de cirugía.”

Tabla 6.  
Componentes del proceso ETL

Componente	Entrada	Proceso	Salida
<b>Extracción</b>	Fuentes de datos, sistemas transaccionales, hojas de cálculo, archivos de texto.	Selección	Datos crudos(cargados en memoria o <i>staging</i> )
<b>Transformación</b>	Datos crudos(cargados en memoria o <i>staging</i> )	Limpieza, transformación, personalización, realización de cálculos y aplicación de funciones de agregación.	Datos formateados, estructurados y resumidos de acuerdo a las necesidades(aún en memoria o <i>staging</i> )
<b>Carga</b>	Datos formateados, estructurados y resumidos de acuerdo a las necesidades(aún en memoria o <i>staging</i> )	Inserción	Datos formateados, estructurados y resumidos con persistencia en el DW

Tomado de (Bustamante et al, 2013)

Las transformaciones más comunes que se realizan a los datos son:

- a. **Transformaciones simples:** son las más sencillas y todas las herramientas ETL poseen funcionalidades para reemplazar datos en volumen. Por ejemplo, cambiar los valores nulos por un valor por defecto, ya que los nulos pueden conllevar a consultas erradas. Estandarizar las unidades de medida, redefinir nombres de productos, clientes, recursos, son otros ejemplos sencillos.
- b. **Combinación – separación de datos:** unión de varios campos para formar uno, separación de un valor en varios campos. Unión de valores de diferentes tipos de fuentes para consolidar tablas de dimensión o de hechos.

### Área de presentación para soportar inteligencia de negocios

En esta área los datos son organizados y almacenados, de igual forma son dispuestos para que sean sometidos a las consultas de los usuarios, sistemas de reporte y aplicaciones de inteligencia de negocios. En esta área se accede a los modelos dimensionales, relacionales u OLAP. Afortunadamente, la industria ha madurado y se ha concluido que los modelos dimensionales técnicas muy utilizadas para entregar los datos a usuarios de inteligencia de negocios (Kimball & Ross, 2013)

Los datos incluidos en el área de presentación, deben ser detallados y atómicos, de tal manera que puedan soportar los niveles de granularidad exigidos por el usuario y soportar los tipos de consulta que se deseen hacer.

La arquitectura de bus pretende diseñar modelos ágiles, descentralizados e interactivos y resalta la importancia de compartir las dimensiones evitando así aplicaciones independientes. El área de presentación en una empresa grande, es al final un conjunto

de modelos con tablas dimensionales compartidas y asociadas a través de diversas tablas de hechos. Una constelación, de acuerdo a las denominaciones dadas.

### 3.4.2. Software utilizado para soluciones de Inteligencia de negocios

De acuerdo al grupo de consultoría Gartner, se identifica una gran diversidad de aplicaciones y proveedores de software para procesos de Inteligencia de Negocios. En este sentido, se encuentran proyectores de software libre con participación importante en el mercado, como es el caso de la Suite Pentaho. Se observan en el cuadrante de líderes, proveedores como **SAP, SAS, Qlikview, IBM, MicroStrategy y Microsoft** se mantienen en el mismo cuadrante de líderes, aunque bajaron su calificación con respecto al año anterior. **Tableau** Sigue en el cuadrante de líderes, con más valoración en “integridad de la visión”.



Figura 29. Análisis del cuadrante mágico de Gartner para 2014 “Business Intelligence”  
Fuente: Gartner 2014. <http://rollupconsulting.com/?p=1254>

A continuación una breve descripción de la herramienta IBM DB2 Webquery, la cual perteneció inicialmente al proveedor Information Builder y luego fue adquirido por IBM.

### 3.4.2.1. DB2 WebQuery for IBMi

IBM DB2 WebQuery es un producto de consulta y reporte basado en arquitectura web que permite acceder a sistemas de información soportados en tecnología ISeries de IBM, optimizado para el uso con manejador de bases de datos DB2 y que ofrece herramientas para la creación de aplicaciones de inteligencia de negocios, minería de datos, procesamiento analítico en línea (OLAP) e implementaciones de Data warehouse. (Bedoya et Al, 2010).

A través del IBM redbook “Getting Started with DB2 Web Query for i”, el colombiano Hernando Bedoya y un grupo de especialistas, describe cada una de las funcionalidades del IBM DB2 Webquery y detallan paso a paso su uso en la implementación de soluciones de inteligencia de negocios.

A través de sus 596 páginas, explican al lector cada una de las funcionalidades y a través de asistentes y descripción de tareas, ayudan a implementar el producto. Para el Hospital Departamental Universitario Santa Sofía de Caldas, IBM DB2 Webquery representa una oportunidad, pues se cuenta con la licencia de uso respectiva bajo el código de licencia para los sistemas operativos IBM i usados en los servidores Power 6 IBM Iseries tanto de producción como de contingencia.

El redbook mencionado incluye tutoriales para funcionalidades como:

- a. Report Assistant
- b. Graph Assistant
- c. Power Painter
- d. Active reports
- e. Developer Workbench
- f. Online Analytical Processing
- g. HTML Composer
- h. Report Broker

Durante la ejecución del presente proyecto se hará uso de estas características, con el objetivo de aprovechar la licencia con que cuenta el Hospital e integrar la tecnología de IBM instalada.

### 3.4.2.2. Herramientas para Mobile Analytics

Dado el auge de dispositivos móviles, se encuentran en el mercado diferentes soluciones que permiten visualizar resultados analíticos y reportes e informes para la toma de decisiones a través de teléfonos inteligentes y tablets. Roambi<sup>10</sup> es una herramienta que permite personalizar aplicaciones móviles para la visualización de datos. Con la posibilidad de conectarse a diferentes fuentes y configurar componentes como visualizaciones dinámicas, con capacidad interactiva y facilidad de uso.

---

<sup>10</sup> <http://roambi.com/analytics> Sitio oficial del fabricante del software.

Permite tomar datos de archivos .CSV, visualizarlas, explorarlas, filtrarlas y aplicar funciones de pivot (invertir filas y columnas).

También podemos mencionar suites como *QlikView*<sup>11</sup>, *SAS Mobile BI 7.31*<sup>12</sup>, *Pentaho Mobile BI*<sup>13</sup> y *WebFocus Mobile Faves de Informatio Builders*<sup>14</sup> que permiten dibujar tableros o dashboards, conectarse con fuentes de datos como drives, bases de datos, archivos planos, hojas de cálculo, servicios web, entre otras.

### 3.5. Data Mining

La minería de datos es un proceso que utiliza una variedad de herramientas de análisis de datos para descubrir patrones y relaciones en datos que pueden ser utilizados para hacer predicciones válidas (Two Crows, 2005).

Es la extracción de información oculta y predecible de grandes bases de datos, una tecnología para ayudar a las compañías a descubrir información relevante en sus bases de información, clasifican y predicen futuras tendencias y comportamientos (Rivas et al, 2007).

Ofrece la oportunidad de descubrir patrones en los datos que pueden ayudar a predecir el comportamiento de los clientes, productos y procesos. Sin embargo, las herramientas de minería de datos necesitan ser guiadas por expertos en el negocio y los datos. El "mejor" modelo suele obtenerse a partir de las pruebas con diferentes clases de técnicas y algoritmos (Two Crows, 2005).

Surge como una tecnología emergente que sirve de soporte para el descubrimiento de conocimiento, que se revela a partir de patrones observables en datos estructurados o asociaciones que usualmente eran desconocidas. (Taié, 2008)

Desde la década de los noventa, se han utilizado técnicas de minería de datos (DM). Inicialmente desde la estadística básica, procurando disminuir los costos y mejorar los indicadores productivos, luego incorporando las técnicas en la determinación de patrones relacionados con el diagnóstico, plan de tratamiento y recuperación y finalmente se han estado soportando las estrategias directivas para el desarrollo de acciones gerenciales. (Dwivedi & Naguib 2008)

En cuanto a gestión estratégica, encontramos por ejemplo a (Combes et al, 2008) cuyo estudio aplicado a través de la metodología KDD (*knowledge discovery in database*), permite obtener comparaciones predictivas de uso de quirófanos a partir de perfiles de pacientes, y en este caso, ofrece un acercamiento de la toma de decisiones a nivel gerencial para procesos de cirugía.

---

<sup>11</sup> <http://www.qlik.com/>

<sup>12</sup> [http://support.sas.com/documentation/onlinedoc/mobile\\_bi/](http://support.sas.com/documentation/onlinedoc/mobile_bi/)

<sup>13</sup> <http://www.pentaho.com/mobile-bi>

<sup>14</sup> <http://www.informationbuilders.com/products/webfocus/mobilefaves>

Se aprecian los esfuerzos que buscan plantear soluciones de minería de datos clínica y a través de experiencias vividas en países como Australia, China, Israel, Nueva Zelanda y Estados Unidos, enriquecen el avance de otras indagaciones (Lalayants, et al., 2013) y es que las condiciones específicas del sector salud exigen que se garantice la participación de todos los niveles de la organización. Que los altos directivos se comprometan con la implementación del modelo que se brinden espacios de inclusión con todos los interesados en el proyecto. (Nofal & Yusof, 2013)

Identificamos prácticas tan sencillas como tablas dinámicas para visualización en hojas de cálculo, hasta estudios que integran sistemas para soporte a toma de decisiones (DSS), técnicas de (DM) y métodos para la toma de decisiones con múltiples criterios (*Multiple Criteria Decision Making* – MCDM -). (Khademolqorani, 2013) que permitan mejorar el desempeño y evaluación de procesos. Estudios que han sido también ampliados por otros investigadores como (Aghdaie, 2014) y (Glover, 2010) buscando resaltar la importancia de técnicas de análisis predictivo y prescriptivo.

Igualmente se aprecia el uso de dichas técnicas para la detección de enfermedades crónicas y de salud pública como el consumo de tabaco en adolescentes (Montaño et al. 2014), eficiencia en determinación de factores de riesgo en diabetes tipo II (Rezaei, 2013; Bellazzi et al. 2009) y enfermedades con alto grado de mortalidad como cáncer de mama (Mclaren, 2009) a través de técnicas de regresión logística y redes neuronales.

La industria farmacéutica y de biotecnología, es tal vez, la que ha dado mayor uso a la minería de datos, promoviendo investigaciones no sólo orientadas a la producción de nuevos medicamentos (Olaleye, 2001), sino también a realizar análisis sobre identificación de reacciones adversas medicamentosas durante los procesos de atención y la aplicación de medicamentos en diversas patologías, tales como psiquiatría, cardiología y nefrología. Resultados que permiten evitar riesgos futuros a los pacientes (Hochberg et al, 2007).

Se han documentado algoritmos computacionales de la minería de datos como self-organizing map (SOM) and decision tree analysis (DTA), que ofrecen a los investigadores cualitativos un conjunto único de herramientas para el análisis de los datos informáticos de salud. (Castellani, 2003)

Se observa evidencia sobre estudios enfocados al apoyo diagnóstico en el tratamiento inicial de enfermedades y que han favorecido áreas particulares como la electroencefalografía (Flexer, 2000), correlaciones y clustering entre datos clínicos y las imágenes cerebrales (Megalooikonomou, 2000) para el descubrimiento de patrones que ayuden a la recuperación de los pacientes. Uso de inteligencia de negocios en los registros electrónicos de la historia clínica (Bonney, 2013), así como apoyo a la gestión estratégica de los procesos misionales mediante análisis de datos clínicos en el ciclo diagnóstico – tratamiento (Caron, 2013).

Igualmente se aprecia el uso de dichas técnicas para la detección de enfermedades crónicas y de salud pública como el consumo de tabaco en adolescentes (Montaño et al. 2014), eficiencia en determinación de factores de riesgo en diabetes tipo II (Rezaei, 2013; Bellazzi et al. 2009) y enfermedades con alto grado de mortalidad como cáncer de mama (Mclaren, 2009) a través de técnicas de regresión logística y redes neuronales.

Son notables los enfoques orientados a respaldar la gestión de procesos de tal manera que por medio de recopilación y análisis posterior de los datos se puedan estudiar características y factores de los pacientes durante la admisión y atención de urgencias (Clark & Normille, 2012), hospitalización y atenciones ambulatorias (West et al. 2009) así como descubrir patrones que permitan asegurar mejores niveles de calidad en los procesos de cirugía. (Jones, 2009).

A nivel del proceso de atención, los investigadores indagan por métodos que apoyen labores de asignación de camas en procesos de atención con calidad (Oliveira et al. 2013), eficiencia en cuidados de la salud en Unidades de Cuidados Intensivos (UCI) (Sarmiento et al, 2013), identificación de pacientes en riesgo de eventos adversos mediante minería y visualización de datos (Van Manen, 2014), gestión de ingresos y reingresos de pacientes a través de pre procesamiento, clasificación, *clustering* y visualización (Ag et al. 2014), el control y trazabilidad de los días de estancia hospitalaria, que tiene gran influencia en seguridad del paciente y la mitigación de riesgos de infecciones asociadas a la atención (Kudyba & Gregorio, 2010). Casos de estudio de gran importancia, teniendo en cuenta que todos estos servicios terminan derivando atención quirúrgicas.

El uso de minería de datos se debe entender como un apoyo para los analistas, y no reemplaza al conocimiento que tienen los expertos del negocio. No funciona por sí sólo, ya que los patrones que se encuentren en los datos deben ser interpretados y validados para verificar si son aplicables (Martínez, 2012)

Hay dos aspectos claves para el éxito en procesos de minería de datos. El primero es formular correctamente el problema que se desea resolver y el segundo garantizar los datos correctos (Two Crows, 2005)

Se encuentran casos documentados sobre la aplicación de minería de datos en los cuidados de la salud, en Holanda por ejemplo (Mans et al, 2008), presentan su investigación “**Process Mining Techniques: an Application to Stroke Care**”. Realizada para Dept. *Computer and System science, University of Pavia*, Italia, en la cual analizan datos relacionados con atención a enfermedades cerebro vasculares, para cuatro hospitales italianos, en los que la minería de datos ayudo a establecer factores desde inicio de ictus hasta llegada al servicio de emergencias. Definen el flujo de eventos que permite analizar la relación entre el conjunto de datos de los pacientes al momento de la atención y los datos de comportamiento previo registrados en la historia Clínica. Gracias a estos análisis pudieron encontrar como otros hospitales lograban inducir al paciente a las terapias y controles de hipertensión de manera más oportuna y con ello, aportan al análisis, diseño y medición de guías de atención médica.

En esta misma línea relacionada con la formación profesional e investigación clínica, (Epstein, 2014) demuestra el potencial de los estudios realizados por profesionales y estudiantes de doctorado, encargados de desplegar tácticas de gestión del conocimiento al interior de los Hospitales, y que a su vez, terminan favoreciendo el desarrollo integral del sector. El Hospital Departamental Universitario Santa Sofía, plasma en su Plan de Desarrollo 2012-2015 “Gestión por Resultados”, la necesidad de afianzar la gerencia de la información como factor estratégico y de la misma manera, construir procesos que le orienten al cumplimiento de los retos planteados por la Honorable Asamblea

Departamental de Caldas, que le atribuyó el calificativo de “Universitario” a través de la ordenanza 693 de 2012.

Podemos afirmar que en la actualidad la minería e inteligencia de negocios han encontrado un lugar importante en el enfoque investigativo el sector salud, incluso en la definición de políticas relacionadas con la confidencialidad de los datos de los pacientes (Cheung, 2002), estándares de calidad, perfiles epidemiológicos, gestión del riesgo, detección de fraudes en el sector y en la prestación de servicios (Pogue et al. 2013) y la búsqueda de nuevo conocimiento para la prevención de las enfermedades (Migut & Worrying, 2012).

La relación entre el proceso KDD y los *Data Warehouses* se da de forma natural, pues el primero busca contar con datos procesados, limpios y consolidados, mientras que los segundos ofrecen una estructura bien definida en donde almacenar la información con esas características (Martínez, 2012)

Es importante notar que no existe un “mejor” modelo o algoritmo de minería de datos, depende del problema en estudio y de los datos disponibles para decir cuál entrega resultados más confiables (Martínez, 2012)

Existen diferentes técnicas para llevar a cabo minería de datos, su elección dependerá del objetivo del negocio. Las técnicas descriptivas dan a conocer y como su nombre lo indica describir los datos utilizando resúmenes estadísticos, medias, desviaciones estándar, visualización y gráficos, buscando relaciones y vínculos potenciales entre variables (Two Crows, 2005)

Luego de la descripción y conocimiento de los datos, se pueden utilizar técnicas predictivas para construir un modelo basado en patrones determinados a partir de los resultados conocidos, que luego de ser sometido a pruebas de sus puede ser una guía útil para la comprensión de su negocio (Two Crows, 2005)

### **3.5.1. Tipos de técnicas de minería de datos**

Las técnicas de minería de datos se pueden clasificar en dos grandes categorías; Supervisadas o predictivas y no supervisadas o descriptivas. Una técnica constituye el enfoque conceptual para extraer la información de los datos, y, en general es implementada por varios algoritmos (Luzon, 2004)

#### **3.5.1.1. No supervisadas o descriptivas**

Estos modelos no cuentan con un resultado conocido y por ello se conocen como modelos de aprendizaje no supervisado y se va ajustando de acuerdo a las observaciones o datos entregados. Se utiliza una muestra de datos independiente de aquella utilizada para la fase de construcción y entrenamiento del modelo, con la intención de evaluar la

capacidad de predicción de éste (Martínez, 2012). Algunos algoritmos que se utilizan en estos modelos son los de *clustering* y las reglas de asociación.

- b. Clustering
- c. Asociación

### 3.5.1.2. Supervisadas o predictivas

Se cuenta con una variable con valor desconocido, y la finalidad es determinarlo. Esta variable se conoce como variable respuesta o variable dependiente, mientras que aquellas utilizadas para hacer la predicción se conocen como variables independientes (Two Crows, 2005)

Los modelos predictivos requieren ser “entrenados”, utilizando un conjunto de datos de cuyo valor de variable objetivo es conocido. La idea es que el modelo entregue resultados en base a un aprendizaje, en otras palabras, que se vaya ajustando a la realidad conocida (Martínez, 2012). Los resultados obtenidos se comparan con los valores conocidos en el entrenamiento, de allí que se conocen como *supervisados*. Dentro de los métodos predictivos se pueden señalar los relacionados con la predicción y la clasificación:

- a. **Predicción:** usados para establecer el comportamiento futuro, a partir de la asignación de un valor a un elemento en particular. se pueden utilizar por ejemplo para predecir comportamiento de la demanda futura, utilizando las ventas o el consumo pasado (Martínez, 2012). Encontramos técnicas como:
  - Árboles de predicción
  - Estimador de Núcleos
  
- a. **Clasificación** consiste en “mapear” un elemento dentro de un grupo de datos, de acuerdo a una clase predefinida, en otras palabras indicar a qué conjunto pertenece. Identifica las características o atributos que hacen que un elemento se vincule a un grupo. (Martínez, 2012). Encontramos técnicas como:
  - Tablas de decisión
  - Árboles de decisión
  - Inducción de reglas
  - Redes Bayesianas
  - Basado en Ejemplares
  - Redes Neuronales
  - Lógica Borrosa (difusa)
  - Algoritmos (Técnicas) Genéticas

José Hernández Orallo, coautor del libro “Introducción a la Minería de Datos” resume diversos algoritmos que pueden ser utilizados para tareas de clasificación, regresión, asociación y correlaciones.

Tabla 7.  
Resumen de Algoritmos para minería

Nombre del algoritmo	Predictivo		Descriptivo		
	Clasificación	Regresión	Agrupamiento	Asociación	Correlaciones
Redes Neuronales	✓	✓	✓		
Árboles de Decisión ID3, C4.5, C5.0	✓				
Árboles de Decisión CART	✓	✓			
Redes de Kohonen			✓		
Regresión Lineal y Logarítmica		✓			✓
Regresión Logística	✓			✓	
<b>K-Means</b>			✓		
A priori				✓	
Naive Bayes	✓				
Vecinos más próximos	✓	✓	✓		
Análisis Factorial y Componentes principales					✓
TwoStep, Cobweb			✓		
Algoritmos genéticos y evolutivos	✓	✓	✓	✓	✓
Máquinas de vectores soporte	✓	✓	✓		
Análisis discriminante Multivariante	✓				

Tomado de (Hernandez et Al, 2005)

### 3.5.2. Técnica de Agrupamiento Clustering. Descripción.

El Hospital Departamental Universitario Santa Sofía de Caldas, pretende incursionar en el análisis de datos a partir de técnicas de Minería de Datos. Por ello, teniendo en cuenta que la técnica *Clustering* permite realizar exploración (Xu, 2005) sobre los datos con que se cuenta y ofrecer un mayor entendimiento sobre el negocio, este proyecto de investigación complementará la solución de inteligencia de negocios con la aplicación de dicha técnica.

Esta técnica también se conoce como de agrupamiento y permite identificar grupos que poseen similitud entre sus elementos. Es utilizada para segmentar grupos de clientes (EPS), pacientes, grupos de procedimientos quirúrgicos y no quirúrgicos. Utiliza medidas de similitud entre los objetos.

El análisis de conglomerados o *Clustering*, es una técnica que permite analizar y examinar datos que no se encuentran etiquetados, formando conjuntos de grupos a partir de su similitud. (Xu, 2005) El análisis de datos juega un papel fundamental en la comprensión de los diversos fenómenos.

El principal objetivo del análisis clúster es dividir un conjunto de objetos en dos o más grupos, basándose en la similitud de un conjunto de variables que los caracterizan (Aldás, 2002).

Se divide el conjunto de datos en grupos que son muy diferentes unos de otros (máxima varianza), pero cuyos elementos sean muy similares entre sí (mínima varianza). Es un método descriptivo que identifica un grupo de categorías o “clusters” para describir los datos. (Martínez, 2012) (Two Crows, 2005) y la similitud puede medirse a través de funciones de distancia, las cuales juegan un papel crucial, ya que individuos cercanos deberían ir para el mismo grupo (Hernández et al, 2005) (Rivas et al, 2007)

En *clustering* no se sabe cuáles serán los grupos que se formarán según los atributos escogidos, por lo tanto, es necesario que los expertos del negocio interpreten las categorías que se formen y vean si hacen sentido o no (Martínez, 2012), de la misma manera, se agrupan los objetos de acuerdo a todas las variables y por ellos, una variable irrelevante puede generar ruido en los resultados obtenidos (Aldás, 2002). Es por ello que, como veremos más adelante se recomienda aplicar técnicas de reducción de dimensionalidad para obtener sólo aquellos atributos que son más “relevantes” y que describen mejor los datos.

(Xu & Wunsch, 2009) en su libro “*Clustering*” exponen de forma gráfica y descriptiva el procedimiento para llevar a cabo análisis de clúster. Mostrando como se itera entre el diseño del modelo y su validación, para obtener al final dos salidas. Los clústeres y el conocimiento generado después de la interpretación de los resultados.

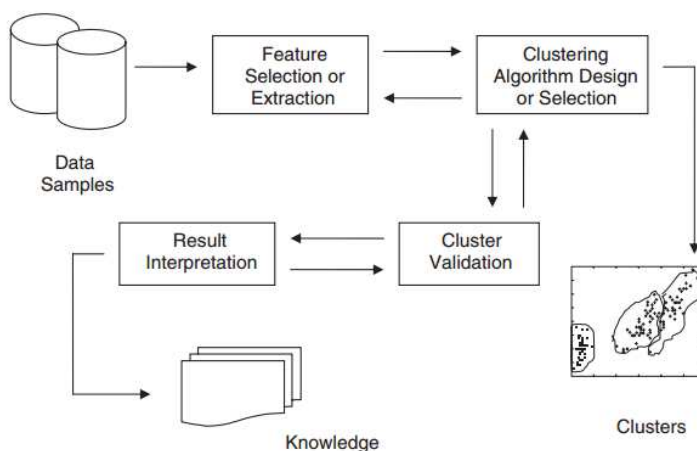


Figura 30. Procedimiento de análisis de clustering  
Tomado de (Xu & Wunsch, 2009) p 6

### 3.5.2.1. Medidas de Similitud

Las medidas de similitud establecen la forma en que se determina la proximidad que hay entre los datos. Miden la distancia entre dos objetos (Martínez, 2012). Los objetos están conformados por atributos o características y siendo (N) el número de objetos y (d) el

número de características, el conjunto obtenido será  $(N \times d)$ . Cada registro representa un objeto, mientras que las columnas sus atributos (Xu & Wunsch, 2009).

Una variable puede ser clasificada continua, discreta o binaria, además, de acuerdo a su nivel de medida puede ser nominal, ordinal o cuantitativa (Xu & Wunsch, 2009).

**Variable nominal:** pueden ser representados numéricamente, pero no tienen un orden específico. Por ejemplo los regímenes de un tipo de afiliación en salud, podría tomar valores como: contributivo=1, subsidiado=2, vinculado=3. Otro ejemplo podría ser el género de los pacientes que puede tomar valores como mujer=1 u hombre =2, sin que estos valores representen un orden en particular.

**Variable ordinal:** Como su nombre lo indica existe una jerarquía en sus definiciones. Bajo=1, Medio=2, Alto=3.

**Variable cuantitativa:** los valores adoptados por este tipo de variables son numéricos y tiene algún significado matemático y las operaciones entre ellos arrojan resultados interpretables. Costo de un procedimiento quirúrgico, promedio de ventas de servicios, costos de un insumo.

Dependiendo del tipo de variable (Xu & Wunsch, 2009) relaciona algunas de las medidas de similitud. (Hernández et al, 2005) en su libro "Introducción a la Minería de Datos" explica como las medidas de distancia formalizan a través de métricas la similitud entre los objetos.

Indica que las medidas de distancia tradicionales, se aplican sobre dos instancias o ejemplos numéricos  $x$  e  $y$  de dimensión  $n$ .

**Distancia Euclídea:** distancia clásica como longitud de la recta que une dos puntos en el espacio euclídeo.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

**Distancia de Manhattan:** hace referencia a recorrer un camino zigueando y no en línea recta.

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

**Distancia de Mahalanobis:** utiliza la matriz de covarianza  $S$ , a diferencia de las otras que asumen de cierto modo, que los atributos son independientes.

$$d(x, y) = \sqrt{(x - y)^T S^{-1} (x - y)}$$

Según (Hernández et al, 2005) en la mayoría de las métricas es muy conveniente normalizar todos los atributos, pues en caso de que algún atributo tenga una magnitud muchísimo mayor que otras, pesará mucho a la hora de calcular las distancias.

El mismo autor ilustra sobre el uso de conceptos de distancia para atributos nominales, como por ejemplo la función delta, es decir,  $\delta(a, b) = 0$  si y sólo si  $a=b$  y  $\delta(a, b) = 1$  en caso contrario. Se utiliza la siguiente función para medir la distancia, siendo  $\omega$  un factor de reducción.

$$d(x, y) = \omega \sum_{i=1}^n \delta(x_i, y_i)$$

Describe también distancias para tipos de datos complejos, por ejemplo si los datos son conjuntos:

$$d(x, y) = \frac{|x \cup y| - |x \cap y|}{|x \cup y|}$$

### 3.5.2.2. Tipos de algoritmos para *Clustering*

Teniendo en cuenta que la finalidad de *clustering* es hallar grupos con características comunes, encontramos algoritmos que requieren que se les establezca el número de *clusters* a buscar (algoritmos particionales) otros que parte de un número desconocido y los encuentran a partir de similitud (algoritmos jerárquicos) y algoritmos probabilísticos que busca a través de varias iteraciones la máxima verosimilitud, es decir, la función de distribución que se ajuste mejor a la muestra de datos.

#### Algoritmos particionales

En este caso primero se determina el número de clúster que se desea. Necesitan que se les diga cuál es el centroide inicial de cada conglomerado y el programa incluye todos los individuos que estén a una distancia preestablecida de ese centro. (Aldás, 2002)

Los algoritmos particionales dividen los objetos en un número de clústeres pre especificado, sin atender a una estructura jerárquica. (Xu & Wunsch, 2009)

KMeans por ejemplo, es un algoritmo de *clustering* particional, al cual se le debe dar el número de *clusters* a buscar. Es un método de agrupamiento por vecindad en el que se parte de un número determinado de prototipos. Este algoritmo utiliza normalmente la distancia euclidea y una vez que los prototipos han sido correctamente situados, los ejemplos nuevos son comparados con estos.

#### Algoritmos jerárquicos

Uno de los problemas del agrupamiento, es discernir cuantos grupos puede haber en los datos. Los métodos jerárquicos construyen un árbol en el que las hojas son los elementos

del conjunto de ejemplos y el resto de nodos son subconjuntos de ejemplos. (Hernández et al, 2005)

Los algoritmos jerárquicos usan dos métodos (Xu & Wunsch, 2009):

- a. Método aglomerativo, que define cada objeto como un clúster, y luego en cada iteración del algoritmo se van mezclando o agrupando aquellos objetos que se asemejan. El análisis comienza con tantos conglomerados como individuos, es decir, cada individuo es un conglomerado inicial. Se van formando nuevos aglomerados de manera ascendente hasta llegar a uno. (Aldás, 2002)
- b. Método divisivo, que parte por el conjunto total de objetos como un único *clúster*, y luego en cada iteración del algoritmo efectúa divisiones, separando los datos hasta llegar a un criterio de corte.

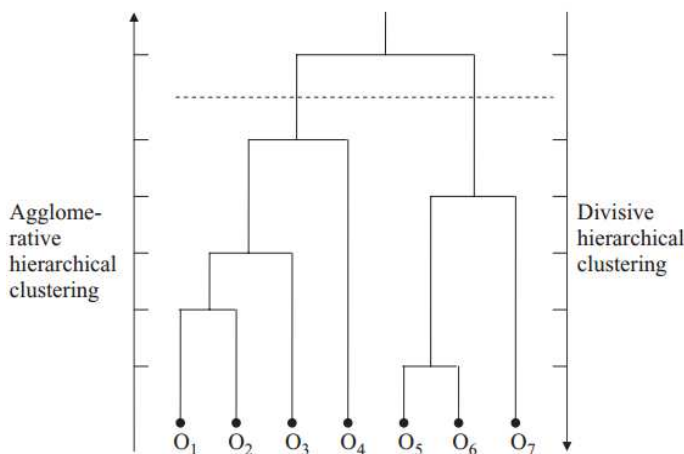


Figura 31. Ejemplo de dendrograma para clustering jerárquico. Tomado de (Xu & Wunsch, 2009)

Se recomienda utilizar primero un método jerárquico para establecer el número de clústeres, una aproximación de los centroides y detectar outliers. Con estos datos luego se puede proceder a un método no jerárquico. (Aldás, 2002)

### 3.5.2.3. Algoritmo K-Medias (K-means)

El algoritmo *K-medias* puede aplicarse para problemas de "agrupación por similitud" y puede ayudar al investigador a una comprensión cualitativa y cuantitativa de grandes cantidades de datos N-dimensionales. (MacQueen, 1967)

De acuerdo con la investigación de (Rivas et Al, 2007) se consideró al algoritmo *K-Means* como oportuno y se establece que permite modelar sistemas de minería de datos, con algoritmos simples pero de mucha robustez para cualquier proyecto de *clustering*.

Es necesario establecer al inicio cuantos clústeres se van a crear, esto se conoce como el parámetro  $K$ . se eligen  $K$  elementos de manera aleatorias, los cuales servirán como centro o media de cada clúster. Luego de esto, cada una de las instancias es asignada a uno de los clústeres (aquel más cercano) a partir de la distancia Euclídea, por ejemplo. Luego de que se tienen conformados dichos clústeres, se calculan su centroides y se repite la asignación. Esto hasta que los puntos centrales de los clústeres se han estabilizado y permanecen invariables en las iteraciones. (XU, 2009)

Tiene la ventaja de ser simple, no le afecta el orden de la muestra y se basa en el análisis de varianzas entre los datos. Sin embargo, es sensible a la elección inicial de los centroides y puede verse afectado por valores atípicos (*outliers*), por lo que es conveniente procesar los datos previamente

(Rivas et Al, 2007) aplica el algoritmo simple *K-Means* con el objetivo de formar grupos o clústeres, donde cada clúster equivale a un grupo de pacientes con características similares. Si bien es cierto que su investigación se enfoca en el agrupamiento de pacientes para analizar el consumo de medicamentos, es importante referenciar a (Xu & Wunsch, 2009) quienes afirman que en las ciencias médicas se vienen aplicando técnicas de *clustering* en entornos importantes que incluyen taxonomía, identificación de genes y proteínas, así como el diagnóstico y apoyo a los tratamientos.

Funciona de forma iterativa, dividiendo óptimamente el conjunto inicial de datos en un número ( $K$ ) de clústeres, el cual se indica como parámetro. Está basado en la minimización de la distancia interna.

(Hernández et Al, 2005) incluye una especificación matemática para describir el procedimiento llevado a cabo por el algoritmo *K-Means*, de la siguiente manera:

- a. Se calcula para cada ejemplo de  $x_k$ , el prototipo más próximo  $A_g$  y se incluye en la lista de dicho prototipo:

$$A_g = \arg \min_{A_i} \{d(x_k, A_i)\} \quad \forall_i = 1..n$$

- b. Luego de introducir todos los ejemplos, cada prototipo  $A_k$ , tendrá un conjunto de elementos

$$I(A_k) = \{x_{k1}, x_{k2} \dots x_{km}\}$$

- c. El prototipo es desplazado hacia el centro de su conjunto de ejemplos:

$$A_k = \frac{\sum_{i=1}^m x_{ki}}{m}$$

- d. Se repite el procedimiento hasta que ya no se desplazan los prototipos. Los ejemplos de entrada  $k$ , se dividen en regiones y el prototipo de cada región estará en el centro de cada una para reducir las distancias cuadráticas euclídeas entre los patrones de entrada y su centro más cercano. Minimizando así el valor  $J$ .

$$J = \sum_{i=1}^k \sum_{n=1}^m M_{i,n} d_{EUCL}(x_n - A_i)^2$$

Siendo (m) el conjunto de patrones,  $d_{EUCL}$  la distancia euclídea,  $x_n$  el ejemplo de entrada,  $A_i$  el prototipo de la clase  $i$  y  $M_{i,n}$  la función que indica la pertenencia del ejemplo  $n$  a la región  $i$ . Valiendo 1 si el prototipo  $A_i$  es el más cercano al ejemplo  $x_n$  y 0 en caso contrario.

$$M_{i,n} = \begin{cases} 1 & \text{si } d_{EUCL}(x_n - A_i) < d_{EUCL}(x_n - A_s) \quad \forall s \neq i, s = 1, 2, \dots, k \\ 0 & \text{en caso contrario} \end{cases}$$

#### 3.5.2.4. Pasos para realizar un análisis de Clúster

Joaquín Aldás Manzano en su documento “El Análisis de *Clúster*” de la Universidad de Valencia, explica 6 pasos para aplicar técnicas de *clustering*:

##### **Paso 1: Definir los objetivos del *clustering***

Concretando el objetivo por el cual se desea aplicar la técnica, entre los cuales se pueden contemplar el carácter exploratorio o confirmatorio. Establecer segmentos de mercado, grupos de pacientes, grupos de EPS de acuerdo a los registros por atenciones. Este paso incluye la selección de variables a usar.

##### **Paso 2: establecer un plan de análisis que incluye**

- a. Verificar si existen *outliers* y especificar su tratamiento o eliminación (Aldás, 2002)
- b. Definir la medida de similitud que se usará, pues los objetos se pondrán en el mismo grupo si están “cerca”. La distancia más común es la **distancia euclídea**, o sus variaciones como **distancia euclídea al cuadrado**. Estas distancias requieren estandarización de datos, lo cual no es necesario con la **distancia de Mahalanobis** que incorpora un procedimiento de estandarización y es muy recomendada cuando no hay mayor significancia entre las variables. (Aldás, 2002)
- c. Definir si los datos deben ser estandarizados para evitar la dispersión, de acuerdo a la distancia que se desea aplicar. Se recomienda ensayar diferentes distancias para comparar sus resultados.

##### **Paso 3: condiciones de aplicabilidad del análisis clúster**

El análisis clúster, no pretende inferir resultados de una muestra hacia una población, sino agrupar objetivamente por similitud los casos que estemos investigando. (Aldás, 2002)

##### **Paso 4: Estimación del modelo y ajuste global**

Elegir el procedimiento disponible para el agrupamiento que busque lo que se ha planteado: mayor distancia entre clústeres y menor distancia entre sus elementos.

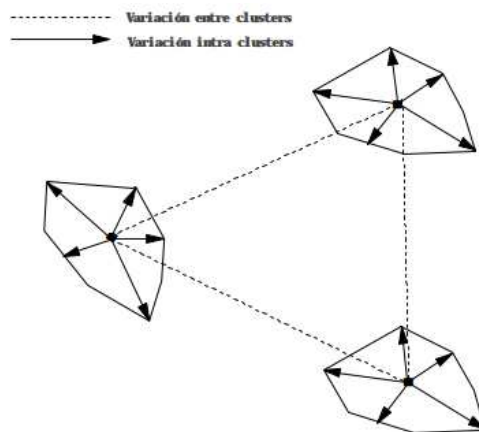


Figura 32. Ejemplo de distancias intra y entre clústeres  
Tomado de (Aldás, 2002) Análisis de clúster

### Paso 5: Interpretación de los conglomerados

Debe hacerse atendiendo a qué valores medios toman en cada uno de ellos las variables que se han utilizado para caracterizarlos. En el siguiente ejemplo se observa como para el conglomerado 1, se agruparon los clientes para quienes las variables **rapidez del servicio, flexibilidad de precios**; mientras el conglomerado de clientes 2, se preocupa más por **nivel de precios, imagen del fabricante y calidad del producto** (Aldás, 2002)

Centros de los conglomerados finales

	Conglomerado	
	1	2
Rapidez de servicio	4,38	2,57
nivel de precios	1,58	3,21
flexibilidad de precios	8,90	6,80
Imagen del fabricante	4,93	5,60
Servicio	2,96	2,87
Imagen de los vendedores	2,53	2,82
Calidad del producto	5,90	8,13

Figura 33. Ejemplo de conglomerados finales  
Tomado de (Aldás, 2002) Análisis de clúster

### Paso 6: Validación del Modelo

Asegurar que la solución es representativa de la población y que se puede aplicar a otros objetos por fuera de muestra. Para ellos, se puede usar un análisis no jerárquico, sin indicar el centroide inicial. Si el análisis es robusto, la solución final no debería diferir de la que se ha obtenido con anterioridad (Aldás, 2002)

### 3.5.2.5. Evaluación de técnicas de agrupamiento

De acuerdo a (Hernández et Al, 2005) los modelos descriptivos son complicados de evaluar, debido a que no se tiene una clase determinada para medir el grado de acierto. Sin embargo, se proponen algunas alternativas:

- a. Utilizar la distancia entre grupos para medir la calidad del agrupamiento. A mayor distancia y mayor separación de los grupos, mejor.
- b. Aprender varios modelos para el mismo conjunto de datos y comparar sus resultados.
- c. La mejor forma de evaluarlo es comparar la utilidad del resultado en el área de aplicación

### 3.5.3. Reducción de la Dimensionalidad

En el capítulo 4 “Limpieza y Transformación de datos” del libro Introducción a la Minería de Datos, (Hernández et Al, 2004) los autores argumentan la importancia de garantizar una buena calidad de los datos para obtener resultados confiables y precisos. Además de mostrar algunos ejemplos y recomendaciones acerca de la integración y limpieza de los datos a utilizar, plantean mecanismos para reconocer (explorar) los datos iniciales, tratar con valores faltantes (perdidos o ausentes) y erróneos (valores anómalos). En particular hacen énfasis en la *reducción de la dimensionalidad*, puesto que según afirman los autores, cuando se cuenta con muchos atributos respecto a la cantidad de instancias o ejemplos, los patrones extraídos pueden resultar caprichosos y poco robustos. Describen técnicas que permiten definir entonces, disminuir el número de atributos que conformarán la vista minable, para que, de acuerdo al algoritmo o técnica de minería a utilizar, se puedan alcanzar mejores resultados. A continuación se describen de manera muy breve la técnica de Análisis de componentes principales y otras prácticas que recomiendan los autores en la sección 5.4.2 “*Selección de características relevantes. Reducción de dimensionalidad*”.

#### Análisis de Componentes principales

La técnica más tradicional, conocida y eficiente para reducir la dimensionalidad por transformación, se denomina *Análisis de Componentes Principales* donde se busca transformar un conjunto de atributos o variables originales  $x_1, x_2, x_3 \dots x_m$  en otro conjunto de atributos  $f_1, f_2 \dots f_p$  donde  $p \leq n$ . (Hernández et Al, 2004).

Se busca que los atributos obtenidos sean lo más independientes posibles entre sí y al final, que los primeros atributos sean los más representativos del conjunto de datos.

En el libro mencionado, se explica de manera algebraica la representación de la técnica. (Hernández et Al, 2004) sostienen que *existen muchas herramientas que permiten realizar todo el proceso. Lo importante es que el resultado de todo el proceso serán los atributos más importantes.*

### **Eliminación de claves candidatas**

Propone también (Hernández et Al, 2004) que la regla general para reducir dimensionalidad es eliminar cualquier atributo que se pueda usar como clave primaria, ya sea como clave total o parcial. Una manera que recomiendan para determinar si es posible eliminar el atributo, es analizar si el número de apariciones es similar al número de registros. Id del paciente, número de historia clínica, códigos internos, nombres y apellidos, son ejemplos aplicables.

### **Eliminación de datos dependientes**

En la normalización de las bases de datos es común almacenar en una tabla diferente, los datos que se encuentran relacionados o que dependen de otros. Así por ejemplo, se usa en la tabla de personas el código de la ciudad, sabiendo que con este código se podrán ubicar otros datos en una tabla adicional, tales como nombre de la ciudad, departamento y país. Como se aprecia en el documento, la minería se puede realizar a partir de una “vista minable” desnormalizada (todo en una tabla) y en ella pueden existir datos dependientes unos de otros. (Hernández et Al, 2004) sugiere que sólo dejemos un campo y eliminemos los otros pues en caso de ser redundantes afectará los procesos de agrupamiento.

### **Análisis Correlacional y Análisis de varianza (Anova)**

Mediante la técnica de *análisis correlacional*, se pueden “cruzar” las variables del conjunto de datos y establecer cuales datos presentan mayor nivel de relación y que atributos parecen ser independientes. Argumenta (Hernández et Al, 2004) además, que mediante esta técnica se pueden también explorar los datos para comprender su interrelaciones.

El análisis correlacional forma parte de múltiples técnicas del análisis multivariante, al igual que el *Análisis de Varianza (ANOVA)* que ayudan a determinar si una o más variables influyen en otra. (Aldás, 2000) en su texto “El Análisis multivariable: conceptos básicos” para la Universidad de Valencia, relaciona ejemplos aplicados para el análisis multivariable, ilustra sobre la clasificación de dichas técnicas y ejemplifica seis pasos para llevar a cabo la actividad.

El objetivo del proyecto actual, no busca profundizar en dichas técnicas. Pero aun así, al momento de analizar los datos podremos recurrir a su ayuda para ajustar los atributos más representativos.

#### **3.5.4. Data Mining y los Data Warehouse**

Las vista minable puede ser el resultado de un subconjunto de datos proporcionado por un *data warehouse*, esto debido al beneficio obtenido por la transformación de los datos. Los problemas de limpieza y consolidación de datos son semejantes en ambos casos y es probable que si los datos provienen de un *dataware house*, no se requieran labores de limpieza más profundos. (Two Crows, 2005)

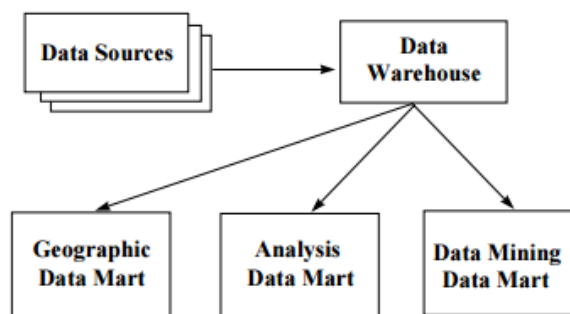


Figura 34. Data mart para data mining extraído de data warehouse  
Tomado de (Two Crows, 2005)

Aunque un *data warehouse* no es requisito fundamental para los proyectos de minería y los datos se podrían extraer directamente desde las bases de datos, es conveniente utilizarlo como origen, pues se realizan labores previas de integridad y calidad en los datos, criterios claves de éxito.

### 3.5.5. Software para Data Mining

#### WEKA

*Weka* (*Waikato Environment for Knowledge Analysis*) es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Los algoritmos bien se pueden aplicar directamente a un conjunto de datos o llamados desde su propio código Java. Weka contiene herramientas para pre-procesamiento, clasificación, regresión, *clustering*, reglas de asociación, y la visualización.<sup>15</sup> Desarrollado por la universidad de Waikato, ofrece la aplicación de diferentes técnicas y herramientas de minería. Es software libre distribuido bajo la licencia GNU-GPL (*General Public License*).

#### IBM SPSS

*IBM SPSS Statistics* es una familia de productos que permite realizar procesos de recopilación, análisis y presentación de datos. Incluye módulos para la aplicación de diversas técnicas estadísticas y de minería de datos. Está programado en Java, multiplataforma y con licencia propietaria. Fue creado por Norman H. Nie, C. Hadlai (Tex) Hull y Dale H. Bent en el año 1968.<sup>16</sup>

<sup>15</sup> <http://www.cs.waikato.ac.nz/ml/weka/book.html> Acceso web a información oficial sobre WEKA.

<sup>16</sup> <http://www-01.ibm.com/software/analytics/spss/> Sitio oficial de IBM SPSS

## 3.6. Desarrollo de la metodología

### 3.6.1. Fase I. Comprensión del Negocio

Objetivo de la Fase: Determinar los objetivos y requerimientos desde una perspectiva no técnica.

Siguiendo recomendaciones de (Chapman et Al, 1999) e (IBM, 2010), se llevaron a cabo actividades de revisión documental y registro de formatos para establecer los objetivos del negocio, los objetivos del proceso de cirugía, la estructura organizacional, la interdependencia de procesos y la caracterización del proceso de cirugía. Esto permitió definir los objetivos de minería de datos y documentar los requerimientos funcionales que la solución de inteligencia de negocios debía suplir.

Igualmente se identificaron los indicadores del proceso, los interesados en la solución, las características y riesgos a los cuales se podría enfrentar el desarrollo del proyecto.

En el **Anexo 1. Documentación adicional Fase I. Comprensión del negocio**, se podrán consultar datos y formatos adicionales que fueron utilizados durante esta labor.

Para facilitar el despliegue y de acuerdo a la metodología, el anexo incluye un glosario de definiciones tanto del negocio, como de la solución desarrollada.

A continuación algunas tablas, formatos e ilustraciones sobre las acciones mencionadas:

#### 3.6.1.1. Caracterización Proceso Atención Integral en el Servicio de Quirófanos

El Esquema definido para la implementación del Modelo de Operación por Procesos, implica una documentación básica. A continuación una breve descripción de los aspectos relevante sobre el proceso de quirófanos (Cirugía)<sup>17</sup>, con el fin de contextualizar al lector en esta Fase de Comprensión del Negocio.

---

<sup>17</sup> Información obtenida del Software para la documentación de Sistema de Gestión de Calidad.

Tabla 8.  
Caracterización Proceso de Cirugía.

Proceso	Atención integral al paciente en el Servicio de Quirófanos
<b>Objetivo</b>	Prestar la atención integral al paciente en el servicio de quirófanos; garantizando la calidad en cuanto a oportunidad, racionalidad técnico científica y seguridad; con personal idóneo; acorde con nuestro nivel de atención y dotación tecnológica, brindando información adecuada al paciente y su grupo familiar.
<b>Líder del Proceso</b>	Nathaly Zapata. Jefe de Enfermería
<b>Coordinador del Proceso</b>	Dr Fernando Uribe. Anestesiólogo.
<b>Recursos</b>	<b>Humanos:</b> Médicos generales y especialistas, Auxiliares de enfermería, Personal de apoyo diagnóstico y terapéutico, Personal en entrenamiento, Enfermero(a), Instrumentadora Quirúrgica, Personal de apoyo <b>Físicos:</b> Dotación de oficina, Insumos y equipos requeridos, Sistema de Información
<b>Inicio</b>	Solicitud de cirugía por parte del Cirujano.
<b>Clientes Internos</b>	Todos los procesos
<b>Clientes Externos</b>	Aseguradores, Usuarios, Ministerio de la Protección Social Secretaria de salud Municipal y Departamental, Entes de Control y Vigilancia
<b>Entradas</b>	Solicitud de procedimiento quirúrgico u observación, guías y protocolos de atención.
<b>Proveedores</b>	Todos los procesos del hospital, aseguradores, Ministerio de la Protección Social, Secretaria de Salud Municipal y Departamental, Sector Judicial, IPS públicas y privadas
<b>Termina</b>	Alta del paciente o retiro voluntario, traslado a otro servicio, remisión o contraremisión a otra institución, fallecimiento, fuga
<b>Resultados (Formatos/Documentos)</b>	Libro radicador ingreso de pacientes a recuperación, Libro entrega muestras patologías laboratorio, Programación cuadros de turno disponibilidad y compensatorios, Libro control de esterilización, Plan de Acción, Plan de Mejoramiento, actualización de guías y protocolos. (Ver tabla de retención documental del proceso). Sistema de Información Hosvital.
<b>Marco Legal</b>	Normatividad vigente de acuerdo a los requerimientos internos y externos (Ver normograma
<b>Riesgos</b>	Identificados a través del mapa de riesgos de cada proceso. Ver mapa de riesgos por proceso

### 3.6.1.2. Procedimientos que integran el proceso

Tabla 9.  
Resumen Procedimientos proceso Cirugía.

Código	Procedimiento	Ponderado dentro del Proceso
QU120-R1	Programación de Cirugías	0.10
QU120-R2	Atención en Quirófanos	0.40
QU120-R3	Atención en Recuperación	0.15
QU120-R4	Central de Esterilización	0.30
QU120-R5	Educación en Salud	0.05
Total Ponderado		1.00

La E.S.E Hospital Departamental Universitario Santa Sofía, incluye dentro de su plataforma estratégica, quince (15) objetivos institucionales, cuyo despliegue se da a través de diferentes estrategias. Una de ellas en particular “Afianzar la gerencia de la información como una ventaja competitiva”, promueve la gerencia de la información para la toma de decisiones y en el Plan de Gerencia de la Información ha incorporado acciones tendientes a desarrollar proyectos de **Inteligencia de Negocios y Minería de Datos**. De allí que los resultados del presente proyecto se encuentran enfocados hacia las metas propuestas por la institución.

Este interés por el aprovechamiento de los recursos tecnológicos, se encuentra enmarcado en el enfoque de la misión y visión institucional, además, el marco normativo colombiano ha venido incluyendo de manera gradual, la gestión de la información como eje clave para el desarrollo de las Empresas Sociales del Estado. El componente de Acreditación en Salud del Sistema Obligatorio de Garantía de la Calidad, establece en sus estándares la **minería de datos como herramienta de gestión y la identificación de grupos de interés**<sup>18</sup> como acciones importantes para garantizar el enfoque de los servicios hacia la población correcta. Por otro lado, la Estrategia de Gobierno en Línea promovida por el Estado colombiano, exige que se implementen medios tecnológicos para alcanzar la efectividad de las entidades públicas, se acerque el gobierno a la ciudadanía y se apoye la innovación e investigación a través de **datos abiertos** como conjuntos de datos tratados que sirvan como insumo para el desarrollo de aplicaciones y la explotación de información.

### 3.6.1.3. Objetivo Proceso de Quirófanos

Prestar la atención integral al paciente en el servicio de quirófanos; garantizando la calidad en cuanto a oportunidad, racionalidad técnico científica y seguridad; con personal idóneo; acorde con el nivel de atención y dotación tecnológica, brindando información adecuada al paciente y su grupo familiar.

Para verificar sus logros, mide el resultado de sus procesos con los siguientes indicadores, cuyos resultados serán visualizados a través de los reportes de la solución de inteligencia de negocios:

- Oportunidad en la realización de cirugía programada
- Proporción de cancelación de cirugía programada
- Tasa de complicaciones quirúrgicas en el transoperatorio
- Tasa de complicaciones anestésicas en el transoperatorio
- Porcentaje de complicaciones quirúrgicas presentadas en recuperación
- Porcentaje de complicaciones anestésicas presentadas en recuperación.

### 3.6.1.4. Objetivos de la minería de datos

---

<sup>18</sup> Clustering como técnica de minería de datos permite agrupar la población atendida en el hospital para la identificación de grupos de interés institucional.

Se definieron cinco (5) objetivos de minería de datos, hacia los cuales se orientan las funcionalidades de la solución de inteligencia de negocios y minería de datos.

Tabla 10.  
Definición Objetivos de Minería de Datos.

<b>FASE I: COMPRENSIÓN DEL NEGOCIO</b>			
<b>FORMATO OBJETIVOS DEL NEGOCIO</b>			
<b>ID</b>	<b>Objetivo</b>	<b>Fuente</b>	<b>Criterios de Éxito</b>
OB1	Apoyar la toma de decisiones, ofreciendo acceso a los datos asistenciales y financieros del proceso de cirugía	Gerencia Subdirección Científica	Solución de Inteligencia de negocios que permita analizar información relacionada con el proceso de cirugía.  Presentar informes generales y detallados sobre los procedimientos quirúrgicos realizados, incorporando datos por especialidades. Permite elegir el periodo de tiempo para el cual se desea el informe, así como las especialidades y especialistas a filtrar
OB2	Utilizar herramientas tecnológicas para definir grupos de pacientes que fueron sometidos a procedimientos quirúrgicos cuyos diagnósticos estaban asociados a cáncer.	Líder Garantía de Calidad	Se obtiene una caracterización por grupos de pacientes a quienes históricamente se les han realizado procedimientos quirúrgicos y cuyos diagnósticos están asociados a cáncer.
OB3	Realizar monitoreo al uso de recursos del proceso de quirófanos.	Líder proceso de Atención Integral en Quirófanos	Es posible filtrar por sala (quirófano, especialidad y profesional) los tiempos usados para los procedimientos realizados.
OB4	Facilitar el seguimiento a las re-intervenciones quirúrgicas.	Líder proceso de Atención Integral en Quirófanos	Reporte para identificar las posibles re intervenciones quirúrgicas.
OB5	Permitir la comparación de producción de procedimientos quirúrgicos de acuerdo a lo publicado en el informe del Decreto 2193 de 2004	Coordinadora del Talento Humano	Consolidar y visualizar las cifras reportadas en el Decreto 2193, por año, trimestre y tipo de indicador.
OB6	Visualizar los indicadores del proceso de quirófanos, de tal manera que se logre identificar su escala de valoración.	Líder proceso de Atención Integral en Quirófanos	La solución permite visualizar el estado de los indicadores, graficando de acuerdo a la escala de valoración de las fichas.
OB7	Ofrecer reportes para el monitoreo financiero del proceso de cirugía, incluyendo valores facturados, glosados, aceptados y soportados.	Jefe Facturación	Los informes permiten detallar los valores facturados, glosados, soportados y aceptados, según especialidad, procedimiento y asegurador.

### 3.6.1.5. Requerimientos Funcionales

Se plantearon entonces, once (11) requerimientos funcionales para apoyar la toma de decisiones en el proceso y para cada uno de ellos se diligenció el formato de detalle del requerimiento, explicado su alcance y significado. En la tabla siguiente, se aprecia el resumen con su respectivo peso porcentual sobre el proyecto.

Tabla 11.  
Resumen de Requerimientos Funcionales.

FASE I: COMPRENSIÓN DEL NEGOCIO			
FORMATO RESUMEN DE REQUERIMIENTOS FUNCIONALES Y NO FUNCIONALES			
ID	Descripción	Tipo Requerimiento	Ponderación
RF-1	La solución de Inteligencia de Negocios permitirá generar informes dinámicos mediante herramienta OLAP, para analizar datos asistenciales.	Funcional	10%
RF-2	La solución de Inteligencia de Negocios permitirá generar informes dinámicos mediante herramienta OLAP, para analizar datos financieros.	Funcional	10%
RF-3	La solución de Inteligencia de negocios incluirá un tablero para el monitoreo de tiempos quirúrgicos, por período, especialidad, especialista y procedimiento quirúrgico.	Funcional	10%
RF-4	La solución de inteligencia de negocios incluirá tablero para el seguimiento a indicadores de proceso de cirugía.	Funcional	5%
RF-5	La solución de inteligencia de negocios incluirá tablero para el seguimiento a valores facturados, glosados y soportados, por períodos de tiempo.	Funcional	10%
RF-6	La solución de inteligencia de negocios incluirá reporte para efectuar seguimiento a re- intervenciones quirúrgicas.	Funcional	5%
RF-7	La solución de inteligencia de negocios permitirá visualizar los indicadores de producción reportados históricamente en el Decreto 2193 de 2004.	Funcional	10%
RF-8	La solución permitirá el uso de aplicativo móvil para la visualización de informes	Funcional	10%
RF-9	La solución permitirá que los usuarios puedan personalizar los informes y exportarlos en diferentes formatos (xls y pdf), así como generar reportes activos que puedan ser usados en modo off line.	Funcional	10%
RF-10	Se podrán personalizar y guardar informes que permitan un acceso rápido de acuerdo a las necesidades de los líderes.	Funcional	10%
RF-11	Aplicar técnica <i>Clustering</i> de minería de datos, para agrupar pacientes que han sido sometidos a procedimientos quirúrgicos durante los años 2007-2015 y cuyos diagnósticos estaban asociados a cáncer.	Funcional	10%
RNF-1	El sistema deberá exigir autenticación de los usuarios para poder acceder a la información	No Funcional	N/A

La siguiente tabla muestra cómo se detallaron los requerimientos funcionales.

Tabla 12.  
Detalle requerimiento funcional RF-1.

FASE I: COMPRENSIÓN DEL NEGOCIO FORMATO DETALLE DE REQUERIMIENTOS FUNCIONALES Y NO FUNCIONALES				
<b>Identificador</b>	RF-1			
<b>Descripción</b>	La solución de Inteligencia de Negocios permitirá generar informes dinámicos mediante herramienta OLAP, <i>para analizar datos asistenciales</i> .			
<b>Entradas</b>	El usuario selecciona los parámetros o filtros para el informe a generar. <ul style="list-style-type: none"> <li>a. Vigencia – año (s)</li> <li>b. Estado de cirugía (facturada, realizada, confirmada)</li> </ul>			
	El usuario accede al cubo OLAP_CIRUGIA, navegando a través de la funcionalidad Drill-Down e incluye campos de diversas dimensiones hasta construir el informe deseado.			
	Ej. Informe sobre descripción de cirugía por parte del médico, que incluye los siguientes campos, para una vigencia determinada (Año) y estado de cirugía (Facturada, realizada):			
	No. De Cirugía	N ( 9.0 )	Tipo Herida	C ( 20 )
	Procedimiento	C ( 50 )	Cantidad De Sangrado	N ( 7.0 )
	Cedula Paciente	C ( 15 )	Vía	C ( 20 )
	Paciente	C ( 40 )	Tipo Anestesia	C ( 20 )
	Hora Inicial Cirugía	C ( 8 )	Descripción Cirugía	L(6000)
	Hora Final Cirugía	C ( 8 )	Descripción Complicación	L(4000)
	Diagnostico Entrada	C ( 5 )	Especialidad	C ( 20 )
Diagnostico Salida	C ( 5 )	Profesional	C ( 40 )	
<b>Salidas</b>	Informe personalizado y visualizado por el usuario.  El usuario elige en la solución de inteligencia de negocios, la opción de exportar informe en el formato deseado.			

La especificación de los otros requerimientos funcionales puede verse en el *Anexo 1 Documentación Adicional Fase I. Comprensión del Negocio*.

Para facilitar las labores posteriores de validación y pruebas, se documentaron los casos de uso principales, especificando una secuencia de pasos a ejecutar al momento de querer visualizar un reporte.

Tabla 13.  
Caso de Uso. CU-001. Generar Informes mediante herramienta OLAP

<b>FASE I: COMPRENSIÓN DEL NEGOCIO FORMATO CASOS DE USO</b>			
<b>ID</b>	CU -001		
<b>Nombre</b>	Generar Informes mediante herramienta OLAP		
<b>Descripción</b>	El usuario ingresa a la solución de (BI) para generar informes dinámicos mediante tecnología OLAP, dentro de un rango de tiempo determinado y según criterios de búsqueda por dimensiones		
<b>Autor</b>	Wilson Alejandro Rojas		
<b>Fecha creación</b>	17/03/2015	<b>Fecha última modificación</b>	
<b>Fuentes de Información</b>	Datos almacenados para la solución de inteligencia de negocios, proveniente del sistema de información hospital y que corresponde a procedimientos quirúrgicos realizados entre los años 2007-2015		
<b>Actores</b>	Líder de Cirugía, Jefe de Cirugía		
<b>Precondiciones</b>	<ul style="list-style-type: none"> <li>- El usuario debe haberse creado como usuario de la solución de inteligencia de negocios</li> <li>- El usuario debe haberse autenticado exitosamente en el sistema</li> </ul>		
<b>Post condiciones</b>	Caso exitoso: <ul style="list-style-type: none"> <li>- El usuario puede generar sus informes satisfactoriamente.</li> </ul>		
<b>Flujo normal de eventos</b>			
<ol style="list-style-type: none"> <li>1. El usuario se autentica en la Solución de Inteligencia de Negocios y ejecuta el informe tipo OLAP que desea abrir.</li> <li>2. El usuario establece los criterios básicos de búsqueda, entre los cuales tiene opciones:               <ol style="list-style-type: none"> <li>a. Periodos (año o vigencia) entre el 2007-2015. Selección Múltiple.</li> <li>b. Estado de los procedimientos que desea consultar: ("Cancelada", "Confirmada", "Facturada" y "Realizada")</li> </ol> </li> <li>3. El sistema presenta la estructura de datos opcionales para la construcción del informe tipo OLAP, provenientes de las dimensiones y hechos.</li> <li>4. El usuario construye el informe que requiere mediante los datos provenientes de las dimensiones, utiliza la técnica de arrastrar y soltar, o a través de la herramienta OLAP de Web Query for IBMi.               <ul style="list-style-type: none"> <li>• Dimensión Tiempo</li> <li>• Dimensión Empresa</li> <li>• Dimensión Especialidad</li> <li>• Dimensión Paciente.</li> <li>• Dimensión Profesional</li> <li>• Dimensión Diagnóstico.</li> <li>• Dimensión Procedimiento.</li> <li>• Dimensión Admisión.</li> </ul> </li> <li>5. El usuario obtiene el informe elaborado en WebQuery, que además ofrece opciones de ordenamiento, navegación (PIVOT, Drill-Down, Roll-Up), filtrado.</li> </ol>			
<b>Flujos alternos</b>			
<b>Excepciones</b>			
<ul style="list-style-type: none"> <li>o No existen datos para el período y estado de procedimientos solicitado. En el paso 2 del flujo normal, en caso de no existir datos, el informe aparecerá vacío. Sin embargo, permite que el usuario pueda elegir otros datos para la consulta.</li> </ul>			
<b>Criterios de aceptación</b>	El usuario puede acceder a los tableros definidos para el monitoreo de recursos y seguimientos al proceso.		
<b>Objetivos y Características relacionados</b>	OB1 OB7 C2 - C3 - C6.		
<b>Reglas del Negocio</b>	No especificadas		
<b>Requerimientos No funcionales</b>	RNF-1		
<b>Anotaciones</b>	No especificadas		

Por último en esta fase, se definió el **Plan de Proyecto**, en el cual se estimaron los tiempos y recursos para la ejecución de cada una de las fases que contempla la metodología CRISP-DM para el desarrollo de proyectos de minería de datos.

### 3.6.1.6. Plan del proyecto

Tabla 14.  
Plan del Proyecto.

FASE I: COMPRESIÓN DEL NEGOCIO FORMATO PLAN DE PROYECTO		
Fase	Tiempo	Recursos
Comprensión del Negocio	3 Semanas	Plataforma estratégica institucional, Plan de desarrollo Institucional, Manual de Procesos y procedimientos, Recurso humano del proceso de cirugía y procesos administrativos y de alta dirección. Autor tesis.
Comprensión de los Datos	4 Semanas	Sistema de información Hosvital (ERP Clínico y Asistencial). Motor de base de datos DB2. System iNavigator para acceder a las definiciones de esquemas de bases de datos. Autor de tesis
Preparación de Datos	3 Semanas	Motor de base de datos DB2. System iNavigator para la creación del modelo de datos físico. DB2 Web query Developer WorkBench para la creación del modelo de datos dimensional. Autor de tesis. Experto en Hosvital.
Modelado	4 semanas	DB2 Web query Developer WorkBench para la estructuración de cubos OLAP, reportes y dashboards, Software WEKA para la aplicación de técnica de minería de datos., ROAMBI Analytics para reportes móviles. Equipo de soporte básico y experto de Hosvital, recurso humano de diversos procesos, Director tesis, autor tesis.
Evaluación	3 Semanas	Recurso humano de diversos procesos, Equipo de soporte básico y experto de Hosvital. Autor tesis.
Despliegue	3 semanas	Equipo de soporte básico y experto de Hosvital. Autor tesis

### 3.6.2. Fase II. Comprensión de los Datos

#### Objetivo de la fase

Conocer, acceder y explorar los datos para determinar su calidad, teniendo presente los objetivos del negocio.

El sistema de información Hosvital (ERP Institucional), cuenta con siete esquemas de bases de datos para el almacenamiento de registros transaccionales clínicos y financieros. Cerca de 2000 tablas conforman la estructura relacional de dichas bases de datos<sup>19</sup>. Se procedió a identificar los objetos que albergaran los datos asociados a las atenciones del proceso de Atención Integral en el Servicio de Quirófanos, encontrando cincuenta (50) tablas y se procedió a realizar una especificación básica que permitiera conocer la estructura de sus datos.

La especificación tuvo en cuenta el origen de los datos, la calidad y su relevancia para el proyecto de investigación.

<sup>19</sup> Datos sobre objetos incluidos en los esquemas de bases de datos. System iNavigator de IBM.

En el **ANEXO 2. Documentación Adicional Fase II. Comprensión de los Datos**, se podrán evidenciar los formatos en los cuales se consignaron los detalles.

Igualmente se especificaron los datos de las tablas relacionadas, contemplando los campos, descripción, tipo de datos y estadísticas básicas, además de su relevancia para la integración del modelo de datos, de acuerdo a los objetivos y requerimientos establecidos en la Fase I. Comprensión del Negocio.

FASE II. COMPRESIÓN DE LOS DATOS FORMATO DE DESCRIPCIÓN DE LOS DATOS CLINICA.PROCIR									
FASE II. COMPRESIÓN DE LOS DATOS FORMATO DE DESCRIPCIÓN DE LOS DATOS CLINICA.MAEATE									
FASE II. COMPRESIÓN DE LOS DATOS FORMATO DE DESCRIPCIÓN DE LOS DATOS CLINICA.CAPBAS									
FASE II. COMPRESIÓN DE LOS DATOS FORMATO DE DESCRIPCIÓN DE LOS DATOS CLINICA.ADGLOSAS1									
Campo	Campo	Campo	Campo	Descripción	Tipo Dato	Valores Permitidos	Valor Mínimo	Valor Máximo	Relevancia
ProEmpCod									
ProMCDpto									
ProCirCod									
ProCitNum	MPNFac								
ProCitEst	MPCedu								
ProIndOpu	MPTDoc								
ProFlisSol									
ProFlisCx	MPCIPr								
ProCirAut	MPMeNi								
ProObsT	MAFchl	MPNFac	GloCtvo	Número de Factura	N ( 15 0 )	{(0-9999999)}	76002	1039499	Muy relevante
ProCtvin	MAHorl	MPNnom1	GloCtvo	Consecutivo	N ( 15 0 )	{(0-9999999)}	0	91652	Relevante
ProDxSall	MANoRe	MPNnom2	GloCod	Código Glosa	N ( 3 0 )	{(0-999)}	0	863	Relevante
ProCompli	MANoRe	MPApe1	GloItem	Item de la Glosa	C ( 15 )	{(0-9999999)}	0	997310	Relevante
PosTpHerd	MATeRe	MPApe2	GloHnoTip	Tipo de Honorario	C ( 2 )	{(0-99)}	0	9	Relevante
PosCntSng	MAVial	MPFchN	GloCtvl	Consecutivo de Item	N ( 4 0 )	{(0-9999999)}	0	9694	Relevante
DQxTmpPrf	MADI1l	MPSexo	GloPrcCiv	Consecutivo de Procedimiento en Maeate	N ( 5 0 )	{(0-9999999)}	0	99299	Baja
DQxTmpCim	MADI2l	MPDire	GloTpolT	Tipo Item ( S / P )	C ( 1 )	{(P,S)}	P	S	Relevante
DQxIndVia	MADI3l	MPTele	GloCnt	Cantidad de Items Glosados	N ( 12 2 )	{(0-999999999)}	0	896400	Relevante
	MAMeln	MDCodD	GloVir	Valor Glosado Motivo	N ( 18 2 )	{(0-999999999)}	0	185555126	Relevante
		MDCodM	GloVirAcp	Valor Aceptado Glosa	N ( 18 2 )	{(0-999999999)}	0	145902447	Relevante
		MDCodB	GloVirSop	Valor Soportado Glosa	N ( 18 2 )	{(0-999999999)}	0	185555126	Relevante
		MPEstC	GloVirConc	Valor Soportado notificación	N ( 18 2 )	{(0-999999999)}	0	127593960	Relevante
		MPGrEs	GloVirAcon	Valor Notificado	N ( 18 2 )	{(0-999999999)}	0	127593960	Relevante
			GloVirAcCo	Valor Aceptado en Notificación	N ( 18 2 )	{(0-999999999)}	0	127593960	Relevante
			GloFchRta	Fecha de Respuesta	D				
			GloUsrRta	Usuario que Responde	N				
			GloFchrtan	Fecha Rta. Notificación	D				
			GloUsrRtan	Usuario que Responde	N				

Figura 35. Formatos de Descripción de datos  
Elaboración Propia

### 3.6.2.1. Informe de exploración de datos

Los datos revisados permiten establecer que existen atributos con valores null, lo cual dificulta la integración y aprovechamiento de datos, más aún, cuando no pueden ser transformado por formar parte de una historia clínica, la cual está protegida normativamente para evitar modificaciones en su contenido. Sin embargo, son numerosos los campos de las tablas origen que se pueden utilizar para la estructuración del modelo de datos.

Los objetivos de minería no sufrieron cambios luego de la revisión del contenido de las tablas origen.

### 3.6.2.2. Informe de verificación de calidad de los datos

Al revisar el contenido de los campos en las tablas, se aprecia codificación propia del sector salud y la inclusión de códigos (llaves, claves primarias, claves foráneas) propios del software Hosvital.

Los campos que aparecen vacíos no pueden ser modificados por la prohibición normativa de modificar el contenido de la historia clínica.

Se observan contenidos cuya calidad depende de la digitación de los profesionales del servicio, pero que no afectan las funcionalidades de la solución a desarrollar.

No se excluyen datos por supuestos problemas de calidad, puesto que forman parte de la historia clínica y quedarán sujetos a los análisis que a futuro realice el hospital.

Todas las tablas mantienen coherencia en su estructura y no se encuentran problemas que pongan en riesgo la ejecución de proyecto.

### 3.6.3. Fase III Preparación de los datos

#### 3.6.3.1. Objetivo de la Fase

Preparar el conjunto de datos para el desarrollo de la solución de inteligencia de negocios y la aplicación de la técnica de minería *clustering*.

A partir de los orígenes de datos descritos en la fase anterior de la metodología CRISP-DM Fase II “Comprensión de los Datos”, se detalla la estructura de las tablas de hechos y de dimensiones que permitirán construir la solución de inteligencia de negocios.

A continuación algunas tablas e imágenes correspondientes a la descripción de las tablas del modelo dimensional. En el **Anexo 3 Documentación Adicional de la Fase III, Preparación de los Datos.**, se podrán consultar todas las descripciones efectuadas.

Tabla 15.  
Resumen de tablas dimensionales del modelo de datos

FASE III. PREPARACIÓN DE LOS DATOS FORMATO RESUMEN TABLAS DE DIMENSIONES		
Dimensión	Clave Primaria	Atributos
Empresa	MECNTR, MENNIT	MECNTR, EMPDSC, MENNIT, MENOMB, MTUDES, CASE, CASE1, CNTCOD, CNTDSC, RUBRO, NOMRUBRO

Paciente	MPCEDU	MPCEDU, MPTDOC, MPNOM1, MPNOM2, MPAPE1, MPAPE2, MPFCHN, MPESTC, MPGRES, MODESC, MPNIVEDU, MPGRPO, MPCODETN, MPSEXO, MPDIRE, MPTELE, MPTELE1, MPTELE2, MPMAIL, MDNOMD, MDNOMR, MDNOMM, MDNOMB, MPNOMC, HISTSANG, ANTECE_PER1, ANTECE_PER2, HCALERGIA, HISCGINECO, DC_AGE, DC_GEATREO
Procedimiento	CUPS	CUPS, NOM_PROC, NIVEL_ATE, PLAN_BENEF, TIPO_PROC, SOAT, NOM_SOAT, NOMC_SOAT, GRUPO_PROC, LAPAROSCOPIA, TIP_IMAG, SUBT_IMAG, C_CONTAB, CEN_COST, FINALIDAD, FCH_ACT, FCH_CRE, NOMC_CUPS
Profesionales	COD_MEDICO	COD_MEDICO, NOM_MEDICO, ESPEC_1, ESPEC_2, ESPEC_3, P_NOMBRE, S_NOMBRE, P_APELLIDO, S_APELLIDO, RED_MEDICO, TIP_DOC, CED_MEDICO, TEL_MEDICO, TIP_CONTRATO, ESTADO, FCH_ACT, FCH_CRE
Diagnostico	CIE10	CIE10, NOM_DIAG, NOM_CORTO, ALTO_COSTO, REP_OBLIG, SEXO, ESTADO, COD_AGRP, NOM_DXAGRP, NOM_C_AGRP, FCH_ACT, FCH_CRE
Tiempo	FECHA	FECHA, DIA_JULIAN, MM_DD_AAAA, ANIO, DIA_SEM, DIA_ANIO, SEM_ANIO, NUM_TRIM, SIGLO, NUM_ANIO, MES, DIA, DIA_NOMBRE, NOM_TRIM, FIN_SEMAN, ESTACION, MES_NOMBRE, MES_ABREV, FECH_EXCEL, FEC_IN_SEM, FEC_FN_SEM, SEM_EPI, PER_EPI, AN_PER_EPI, FCH_ACT
Admision	HISCKEY, HCTVIN1	HISCKEY, HISTIPDOC, HCTVIN1, MMNOMM, DIADESCRI, DIADSCGRU, DESCRIPCION, MPNOMP, INGFECADM, INGFECEGR, HISDESDET, INGFMHM
Geografica	PAIS, MDCODD, MDCODM	PAIS, MDCODD, MDNOMD, MDCODM, MDNOMM, MDURRU, MDCODB, MDNOMB, MDNOMR

Tabla 16.  
Descripción tablas del Modelo Dimensional. Tabla de hechos Cirugía.

FASE III. PREPARACIÓN DE LOS DATOS			
FORMATO DESCRIPCION TABLAS DEL MODELO DIMENSIONAL			
Tabla	Descripción de Cirugías		
Tipo de tabla	Hechos	X	Dimensión
Número Registros	206.162	Número Atributos	42
Claves foráneas	COD_PROC, DX_ENTRADA, DX_SALIDA, DX_COMPLICACION, CONTRATO, IDENTIFICACION, COD_PROC, COD_MED, PARTICIPANTE, FECHA_CX, CONS_INGRESO		
Métricas	Número de cirugías realizadas Número de pacientes. Duración cirugía Tiempo Estancia paciente		
Atributos	Descripción del Atributo	Titulo	
COD_CIRUGIA	Código Acto Quirúrgico	COD_CIRUGIA	
COD_CITA_ANESTESIA	Código Cita de Anestesia	COD_CITA_ANESTESIA	
DX_ENTRADA	Diagnóstico de Ingreso	DX_ENTRADA	
DX_SALIDA	Diagnóstico de Salida	DX_SALIDA	
DX_COMPLICACION	Diagnóstico de Complicación	DX_COMPLICACION	
DESC_COMPLICACION	Descripción de la Complicación	DESC_COMPLICACION	
CNT_SANGRADO	Cantidad de Sangrado	CNT_SANGRADO	
T_PERFUSION	Tiempo de Perfusión	T_PERFUSION	
TIEMPO_CLAMP	Tiempo de Clamp	TIEMPO_CLAMP	
FOL_SOLICIT	Folio que solicita la Cirugía	FOL_SOLICIT	
FOL_CIRUGIA	Folio en que registra la Cirugía	FOL_CIRUGIA	
CONS_INGRESO	Consecutivo de Ingreso a la Institución	CONS_INGRESO	
MOTIVO_CANCELACION	Motivo de Cancelación de la Cirugía	MOTIVO_CANCELACION	
FCH_CANCELA	Fecha de Cancelación de la Cirugía	FCH_CANCELA	
USU_CANCELA	Usuario que Cancela la Cirugía	USU_CANCELA	

CONTRATO	Código del Contrato al cual se carga la Cirugía	CONTRATO
NOM_CONTRATO	Nombre o Descripción del Contrato	NOM_CONTRATO
ESTADO_PAC	Estado del Paciente	ESTADO_PAC
TIPO_AGENDA	Tipo de Agenda	TIPO_AGENDA
FCH_PRG_SALA	Fecha de Programación de Sala	FCH_PRG_SALA
HORA_PRG_SALA	Hora de Programación de Sala	HORA_PRG_SALA
SALA	Sala – Quirófano	SALA
T_DOC_PAC	Tipo de document del paciente	T_DOC_PAC
IDENTIFICACION	Identificación del Paciente	IDENTIFICACION
ESTADO_CX	Estado de la Cirugía	ESTADO_CX
TIP_ANESTESIA	Tipo de Anestesia	TIP_ANESTESIA
FECHA_CX	Fecha de la Cirugía	FECHA_CX
TIPO_HERIDA	Tipo de Herida	TIPO_HERIDA
VIAS	Vías de acceso	VIAS
COD_PROC	Código de Procedimiento Quirúrgico	COD_PROC
CX_PROC	Código de Acto quirúrgico y Procedimiento Quirúrgico	CX_PROC
ESTADO_ITEM	Estado de cirugía dentro del Acto quirurgico	ESTADO_ITEM
COD_MED	Código del Médico	COD_MED
ESP_MEDICO	Especialidad del Médico	ESP_MEDICO
PARTICIPANTE	Código del Participante en Cirugía	PARTICIPANTE
HONORARIO	Profesional Principal de la Cirugía	HONORARIO
ESP_PART	Especialidad del Participante	ESP_PART
HORA_INI	Hora de Inicio de la Cirugía	HORA_INI
HORA_FIN	Hora finalización de la Cirugía	HORA_FIN
DESCRIPCION_QUIRURGICA	Descripción Quirúrgica	DESCRIPCION_QUIRURGICA
TIEMPO_MINUTOS	Duración en minutos de la cirugía	TIEMPO_MINUTOS

	CONTRATO	NOM_CONTRATO	ESTADO_PAC	TIPO_AGENDA	FCH_PRG_SALA	HORA_PRG_SALA	SALA	T_DOC_PAC
1	HSS2005055	PACIENTE INSTITUCIONAL SOAT + 10%	Vivo	Urgencia	2007-01-01	08:34:58	SALA CX No. 5	TI
2	HSS2005055	PACIENTE INSTITUCIONAL SOAT + 10%	Vivo	Urgencia	2007-01-01	08:34:58	SALA CX No. 5	TI
3	HSS2005046	LA PREVISORA SOAT	Vivo	Urgencia	2007-01-01	10:30:54	SALA CX No. 5	CC
4	HSS2005046	LA PREVISORA SOAT	Vivo	Urgencia	2007-01-01	10:30:54	SALA CX No. 5	CC
5	HSS2005046	LA PREVISORA SOAT	Vivo	Urgencia	2007-01-01	10:30:54	SALA CX No. 5	CC
6	HSS2005046	LA PREVISORA SOAT	Vivo	Urgencia	2007-01-01	10:30:54	SALA CX No. 5	CC
7	HSS2005046	LA PREVISORA SOAT	Vivo	Urgencia	2007-01-01	13:32:25	SALA CX No. 2	CC
8	HSS2005046	LA PREVISORA SOAT	Vivo	Urgencia	2007-01-01	13:32:25	SALA CX No. 2	CC
9	HSS2005046	LA PREVISORA SOAT	Vivo	Urgencia	2007-01-01	13:32:25	SALA CX No. 2	CC
10	HSS2005046	LA PREVISORA SOAT	Vivo	Urgencia	2007-01-01	13:32:25	SALA CX No. 2	CC
11	HSS2005046	LA PREVISORA SOAT	Vivo	Urgencia	2007-01-01	13:32:25	SALA CX No. 2	CC
12	HSS2005046	LA PREVISORA SOAT	Vivo	Urgencia	2007-01-01	13:32:25	SALA CX No. 2	CC
13	HSS2005055	PACIENTE INSTITUCIONAL SOAT + 10%	Vivo	Urgencia	2007-01-01	20:00:26	SALA CX No. 5	CC
14	HSS2005055	PACIENTE INSTITUCIONAL SOAT + 10%	Vivo	Urgencia	2007-01-01	20:00:26	SALA CX No. 5	CC
15	HSS2005055	PACIENTE INSTITUCIONAL SOAT + 10%	Vivo	Urgencia	2007-01-01	20:00:26	SALA CX No. 5	CC
16	HSS2005069	SEGUROS DEL ESTADO SOAT	Vivo	Urgencia	2007-01-01	17:47:45	SALA CX No. 2	CC
17	HSS2005069	SEGUROS DEL ESTADO SOAT	Vivo	Urgencia	2007-01-01	17:47:45	SALA CX No. 2	CC
18	HSS2005055	PACIENTE INSTITUCIONAL SOAT + 10%	Vivo	Urgencia	2007-01-01	23:25:33	SALA CX No. 2	CC
19	HSS2005055	PACIENTE INSTITUCIONAL SOAT + 10%	Vivo	Urgencia	2007-01-01	23:25:33	SALA CX No. 2	CC
20	HSS2005055	PACIENTE INSTITUCIONAL SOAT + 10%	Vivo	Urgencia	2007-01-01	23:25:33	SALA CX No. 2	CC
21	HSS2005055	PACIENTE INSTITUCIONAL SOAT + 10%	Vivo	Urgencia	2007-01-01	23:25:33	SALA CX No. 2	CC
22	HSS2005055	PACIENTE INSTITUCIONAL SOAT + 10%	Vivo	Urgencia	2007-01-01	23:25:33	SALA CX No. 2	CC
23	HSS2005055	PACIENTE INSTITUCIONAL SOAT + 10%	Vivo	Urgencia	2007-01-01	23:25:33	SALA CX No. 2	CC

Figura 36. Contenido tabla dimensional Hechos de cirugía.  
Elaboración propia

Tabla 17.  
Descripción tablas del Modelo Dimensional. Tabla de dimensión Paciente.

FASE III. PREPARACIÓN DE LOS DATOS				
FORMATO DESCRIPCION TABLAS DEL MODELO DIMENSIONAL				
Tabla	BI_SSOPIA.DIM_PACIENTE			
Tipo de tabla	Hechos		Dimensión	X
Número Registros	194.489	Número Atributos	31	
Claves foráneas				
Métricas	Número de pacientes.			
Atributos	Descripción del Atributo	Titulo		
MPCEDU	Documento de Identidad	Identificación		
MPTDOC	Tipo de Documento	Tipo_Doc		
MPNOM1	Primer Nombre	P Nombre		
MPNOM2	Segundo Nombre	S Nombre		
MPAPE1	Primer Apellido	P Apellido		
MPAPE2	Segundo Apellido	S Apellido		
MPFCHN	Fecha de Nacimiento	Fecha_Nac		
MPESTC	Estado Civil	Estado_Civil		
MPGRES	Grupo Etario	Grupo_Aten_Especial		
MODESC	Descripción de la Ocupación	Ocupación		
MPNIVEDU	Nivel Educativo	Nivel_Edu		
MPGRPO	Grupo Poblacional	Grupo Poblacional		
MPCODETN	Etnia	Etnia		
MPSEXO	Género	Sexo		
MPDIRE	Dirección	Dirección		
MPTLELE	Teléfono	Teléfono		
MPTLELE1	Teléfono 1	Teléfono 2		
MPTLELE2	Teléfono 2	Teléfono 3		
MPMAIL	Correo Electrónico	Email		
MDNOMD	Departamento de Residencia	Departamento		
MDNOMR	Región de Residencia	Región		
MDNOMM	Municipio de Residencia	Ciudad		
MDNOMB	Barrio de Residencia	Barrio		
MPNOMC	Nombre Completo	Nombre Completo Paciente		
HISTSANG	Tipo de Sangre del Paciente	Tipo de Sangre		
ANTECE_PER1	Antecedentes personales 1	Antecedentes Personales 1		
ANTECE_PER2	Antecedentes personales 2	Antecedentes Personales 2		
HICALERGIA	Alergias	Alergias		
HISCGINECO	Historia Clínica Ginecológica	HC_Ginecologica		
DC_AGE	Edad en años	Edad		
DC_GEATREO	Grupo Etario (quinquenos)	Grupo Etario		

Contenido de BI\_SSOPIA.DIM\_PACIENTE - 192.168.40.96(5102779f)

	MPCEDU	MPTDOC	MPNOM1	MPNOM2	MPAPE1	MPAPE2	MPFCHN	MPESTC	MPSEXO	MPDIRE	MPTLELE	MDNOMD	DC_AGE	DC_GEATREO
1												CALDAS	20	Entre 20 y 24
2												CALDAS	76	Entre 75 y 79
3												CALDAS	94	Entre 90 y 94
4												CALDAS	78	Entre 75 y 79
5												CALDAS	52	Entre 50 y 54
6												CALDAS	65	Entre 65 y 69
7												CALDAS	32	Entre 30 y 34
8												CALDAS	36	Entre 35 y 39
9												CALDAS	51	Entre 50 y 54
10												CALDAS	44	Entre 40 y 44
11												CALDAS	51	Entre 50 y 54
12												CALDAS	33	Entre 30 y 34
13												CALDAS	60	Entre 60 y 64
14												CALDAS	26	Entre 25 y 29
15												CALDAS	40	Entre 40 y 44
16												CALDAS	61	Entre 60 y 64
17												CALDAS	73	Entre 70 y 74
18												CALDAS	61	Entre 60 y 64
19														
20														
21														
22														
23														

Figura 37. Contenido tabla dimensional Paciente.

La documentación de la Fase III Preparación de los datos, incluye además los scripts utilizados para la construcción del modelo de datos que conforma el Data Mart sobre el cual se desarrollará el proyecto.

### 3.6.3.2. SCRIPT SQL Modelo de datos

#### SCRIPT SQL CREACION TABLA DE HECHOS HEC\_CIRUGIAS

Código hechos cirugía

```
-- Generar SQL
-- Versión: V7R1M0 100423
-- Generado en: 07/05/16 12:40:20
-- Base de datos relacional: S102F79F
-- Opción de estándares: DB2 for i
```

```
CREATE VIEW BI_SOFIA.HEC_CIRUGIAS (
  COD_CIRUGIA FOR COLUMN COD_C00001,
  COD_CITA_ANESTESIA FOR COLUMN COD_C00002,
  DX_ENTRADA,
  DX_SALIDA,
  DX_COMPLICACION FOR COLUMN DX_CO00001,
  DESC_COMPLICACION FOR COLUMN DESC_00001,
  CNT_SANGRADO FOR COLUMN CNT_S00001,
  T_PERFUSION FOR COLUMN T_PER00001,
  TIEMPO_CLAMP FOR COLUMN T_CLAMP ,
  FOL_SOLICIT FOR COLUMN FOL_S00001,
  FOL_CIRUGIA FOR COLUMN FOL_C00001,
  CONS_INGRESO FOR COLUMN CONS_00001,
  MOTIVO_CANCELACION FOR COLUMN MOTIV00001,
  FCH_CANCELA FOR COLUMN FCH_C00001,
  USU_CANCELA FOR COLUMN USU_C00001,
  CONTRATO,
  NOM_CONTRATO FOR COLUMN NOM_C00001,
  ESTADO_PAC,
  TIPO_AGENDA FOR COLUMN TIPO_00001,
  FCH_PRG_SALA FOR COLUMN FCH_P00001,
  HORA_PRG_SALA FOR COLUMN HORA_00001,
  SALA,
  T_DOC_PAC,
  IDENTIFICACION FOR COLUMN IDENT00001,
  ESTADO_CX,
  TIP_ANESTESIA FOR COLUMN TIP_A00001,
  FECHA_CX,
  TIPO_HERIDA FOR COLUMN TIPO_00002,
  VIAS,
  COD_PROC,
  CX_PROC,
  ESTADO_ITEM FOR COLUMN EST_ITEM ,
  COD_MED,
  ESP_MEDICO,
  PARTICIPANTE FOR COLUMN PARTI00001,
  HONORARIO,
  ESP_PART,
  HORA_INI,
  HORA_FIN,
  DESCRIPCION QUIRURGICA FOR COLUMN DESCR00001)
AS
(SELECT P.PROCIRCOD COD_CIRUGIA, P.PROCITNUM COD_CITA_AN, D.DIAENT DX_ENTRADA, P.PRODXSALI DX_SALIDA, D1.DXCOMPL
```

```
INNER JOIN INTRANET_EST_URGE EU
ON F_PROST=EU.PROST
INNER JOIN CLINICA_CONSUL SA
ON F_PROCONS=SA.CONSCOD
INNER JOIN INTRANET_ESTAGUM EST
ON F_PROSTA=EST.PROSTA
INNER JOIN CLINICA_MAEFRANS ANS
ON F_PROFRANS=ANS.FRANSCOD
INNER JOIN INTRANET_TPO_38882 TM
ON D_TIPHER=TM.TIPHER
INNER JOIN CLINICA_VIAS V
ON F1_VIACOD=V.VIACOD
INNER JOIN CLINICA_HOMROS H
ON F2_PERSTPH=H.PERSCOD
LEFT JOIN CLINICA_MAEFR MSP
ON F2_PERESP=MSP.PERSCOD
WHERE F_PROFEC=2007-01-01;
```

LABEL ON COLUMN BI\_SSOPIA.HEC\_CIRUGIAS  
( USU\_CANCELA TEXT IS 'AUSRDSCD' );

GRANT ALTER , REFERENCES , SELECT  
ON BI\_SSOPIA.HEC\_CIRUGIAS TO BI\_SSOPIA WITH GRANT OPTION ;

### SCRIPT SQLS CREACION TABLA DIMENSIONAL DIM\_PACIENTE

```
-- Generar SQL
-- Versión: V7R1M0 100423
-- Generado en: 07/05/16 12:40:54
-- Base de datos relacional: S102F79F
-- Opción de estándares: DB2 for i
```

```
CREATE VIEW BI_SSOPIA.DIM_PACIENTE (
  MPCEDU, MPTDOC, MPNOM1, MPNOM2,
  MPAPE1, MPAPE2, MPFCHN, MPESTC,
  MPGRES, MODESC, MPNIVEDU, MPGRPO,
  MPCODETN, MPSEXO, MPDIRE, MPTELE,
  MPTELE1, MPTELE2, MPMAIL, MDNOMD,
  MDNOMR, MDNOMM, MDNOMB, MPNOMC,
  HISTSANG, ANTECE_PER1 FOR COLUMN ANTEC00001,
  ANTECE_PER2 FOR COLUMN ANTEC00002,
  HICALERGIA, HISCGINECO, HISCAB, DC_AGE, DC_GEATREO)
AS
(SELECT MPCEDU, MPTDOC, MPNOM1, MPNOM2, MPAPE1, MPAPE2, MPFCHN, MPESTC, MPGRES, MODESC, MPNIVEDU, MPGRPO,
  MPCODETN, MPSEXO, MPDIRE, MPTELE, MPTELE1, MPTELE2, MPMAIL, MDNOMD, MDNOMR, MDNOMM, MDNOMB, MPNOMC,
```

GRANT ALTER, REFERENCES, SELECT  
ON BI\_SSOPIA.DIM\_PACIENTE TO BI\_SSOPIA WITH GRANT OPTION;

### 3.6.3.3. DISEÑO DEL MODELO DE DATOS

IBM DB2 Web Query for i (DB2 Web Query), es una herramienta de inteligencia de negocios basada en Java que proporciona herramientas para graficar, obtener consultas web y generar reportes. De acuerdo al RedBook “IBM DB2 Web Query for i Implementation Guide”, permite entre otras cosas: ofrecer capacidad a los usuarios finales

de obtener los datos cuando los necesitan, y en las formas en las que necesitan de manera personalizada; supervisión de cuadros de mando de indicadores de desempeño; detección de tendencias y excepciones en los datos con procesamiento analítico en tiempo real y la creación y entrega de informes de diversos formatos.

### 3.6.3.3.1. Metadatos

Una de las características de DB2 Web Query, es centralizar los datos para su gestión, descomposición automática de fechas, relaciones multidimensionales (Join's) y generación de campos calculados. Con la inclusión de las metadatos, es posible unificar los datos provenientes de los diferentes recursos y estandarizar sus definiciones. (Bedoya et Al 2014 p7,21). Las metadatos son simplemente datos sobre los datos, longitudes de registro, formatos de registro, nombres de campo, tipos de datos, atributos de campo, y longitudes de campo, que son utilizados por DB2 Webquery para la estructura de los reportes.

Antes de poder crear un reporte o un informe en DB2 Web Query, es necesario crear la metadata que le “presente” el origen de los datos y su estructura. Como lo describe el RedBook de IBM, el objetivo principal de la capa de metadatos es: “Mejorar la productividad de los desarrolladores de informes DB2 Web Query, proporcionando una capa de abstracción y evitando la complejidad de base de datos.”

Los datos transaccionales utilizados para la realización de este proyecto, provienen del ERP Hosvital con que cuenta el Hospital Departamental universitario Santa Sofia. En este caso el ERP cuenta con varios esquemas de bases de datos administradas por DB2 bajo el sistema operativo i, de IBM. Se procedió a la creación de un nuevo esquema de base de datos llamada **BI\_SSofia**, en la cual se crearon las tablas y las vistas que posteriormente son utilizadas como bodega de datos para la solución de Inteligencia de negocios.

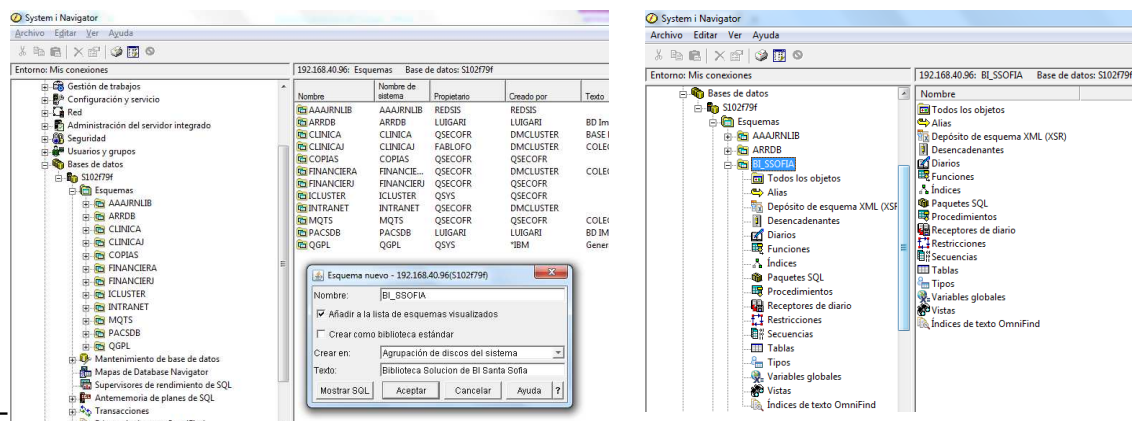


Figura 38. Creación de Esquema de Base de datos BI\_SSofia  
Elaboración propia

En la siguiente imagen se puede apreciar el conjunto de tablas y vistas creadas en el esquema mencionado.

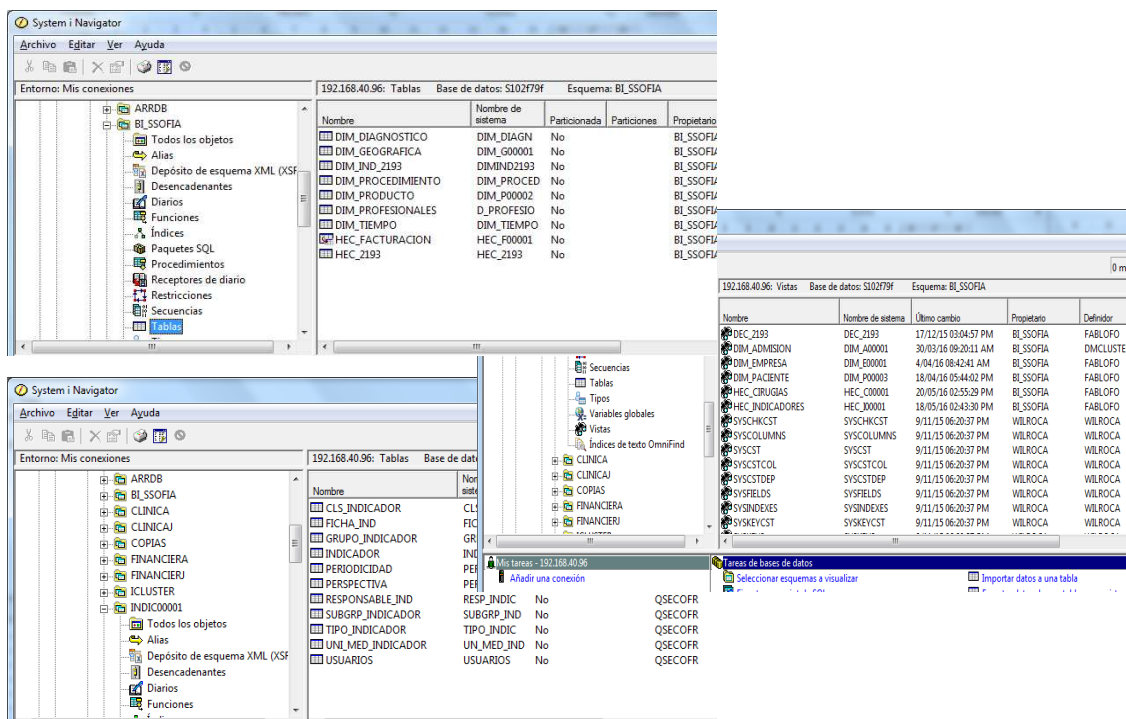


Figura 39. Data warehouse (tablas y vistas) para el desarrollo de la solución. Elaboración propia. Instantáneas de System i Navigator. Esquema BI\_SSofia

Estas tablas y vistas, conforman el esquema general de origen de datos, que luego es utilizado la construcción de las tablas de hechos y dimensiones, de acuerdo a lo referido en la teoría de data warehouse. Fueron generadas a partir de scripts de SQL en los cuales se integraron cerca de 50 tablas de las bases de datos Clínica y Financiera<sup>20</sup>, relacionadas con las transacciones del servicio de cirugía.

Antes de proceder a la creación de las metadatos en DB2 Web query for i, fue necesario instalar el aplicativo DB2 Web Query Developer WorkBench (Developer Studio) el cual es una herramienta de desarrollo cliente - servidor para administrar las aplicaciones de DB2 Web Query, las metadatos, las relaciones entre los sinónimos creados (modelo de datos) y la estructuración de reportes, dashboards y OLAP de la solución. A continuación algunas imágenes que evidencian la gestión básica del modelo de datos.

<sup>20</sup> Clínica y Financiera representan dos esquemas base del ERP Hosvital del Hospital. Ambos conforman un conjunto de 1950 tablas en las cuales se almacenan los datos de historia clínica y financieros.

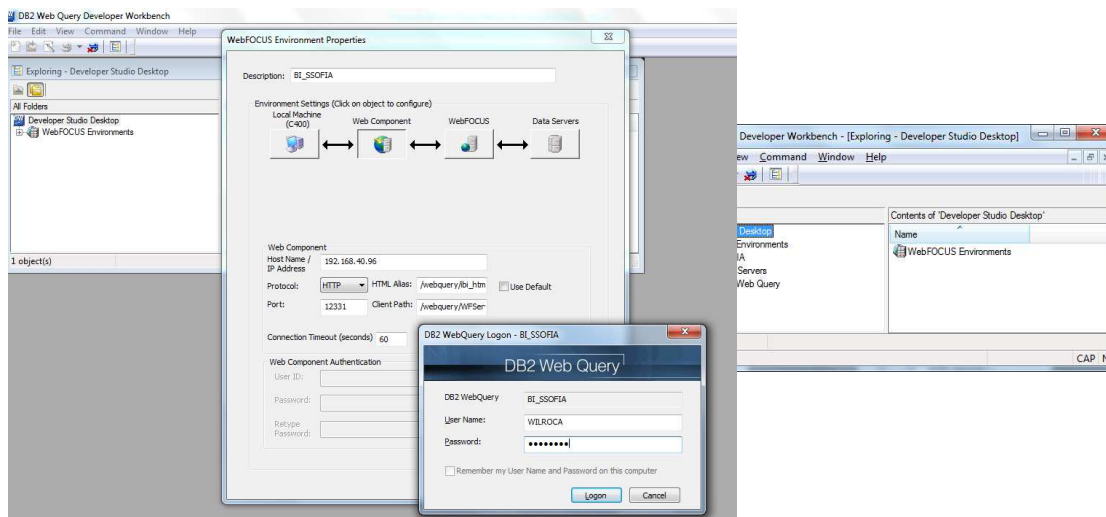


Figura 40. Creación de la nueva solución de inteligencia de negocios (BI\_SSofia)  
Elaboración propia. Instantáneas DB2 Web Query Developer Workbench

Luego de crear la aplicación, fue necesario proceder a enlazarla con los orígenes de datos. De esta manera se creó la *METADATA* necesaria para la funcionalidad del proyecto. A continuación, ilustraciones que muestran el proceso realizado, agregando los sinónimos necesarios.

Se puede apreciar la aplicación BI\_SSofia creada, así como las metadatos correspondientes al modelo de datos utilizado para la solución de inteligencia de negocios.

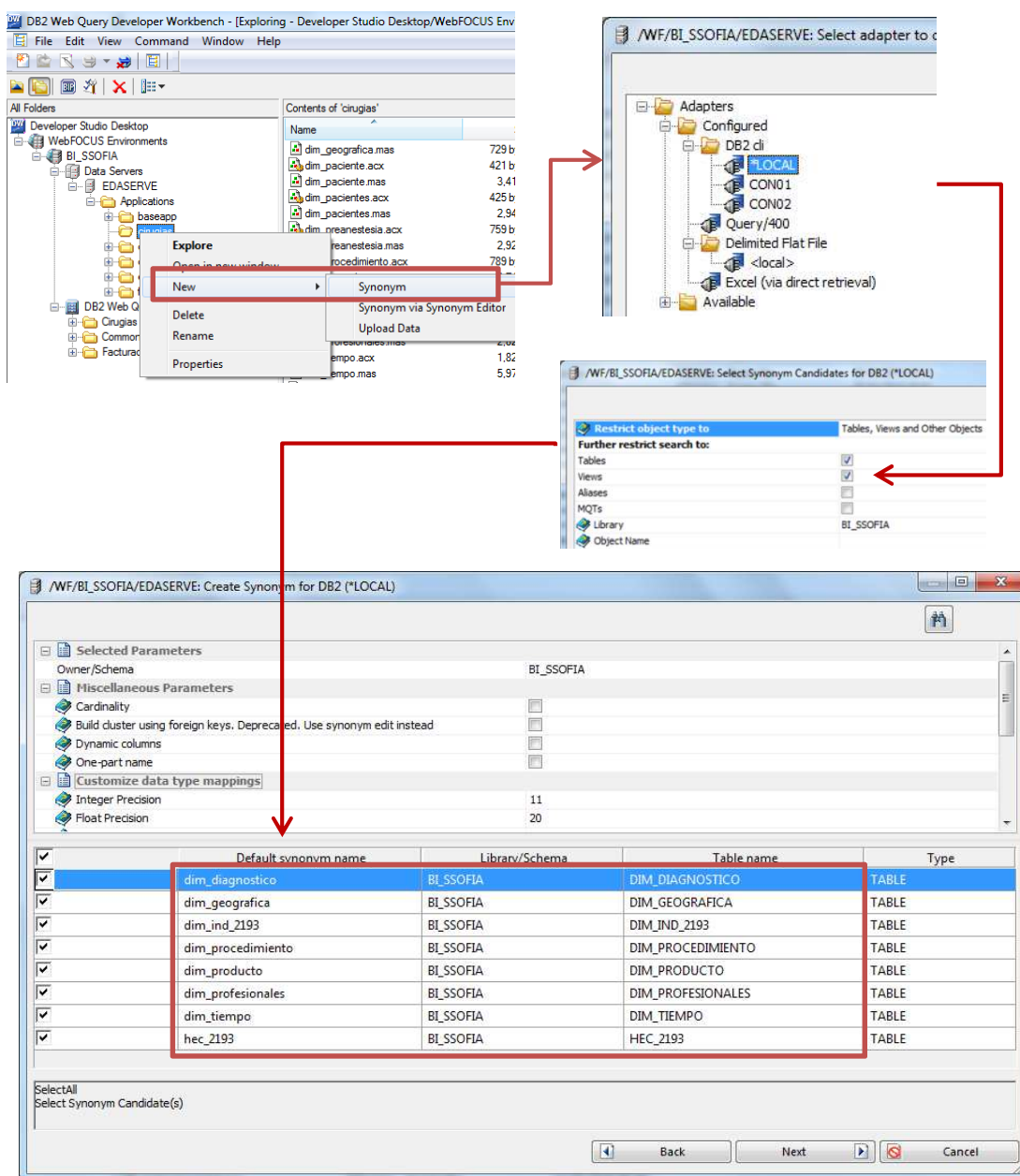


Figura 41. Proceso de creación de metadatos para el desarrollo de solución de BI Elaboración propia. Instantáneas DB2 Web Query Developer Workbench. Elaboración propia.

Al final del proceso, se obtienen las metadatos respectivas y de esta forma, la conexión entre Web Query como motor de reportes y los orígenes de datos proporcionados en este caso por IBM DB2.

Cada origen de datos genera dos tipos de archivo, Access file y maste file (.acx, .mas) el primero contiene metadatos utilizados por webquery para su modelo de datos y el segundo contiene los atributos que integran finalmente las dimensiones y hechos. Cada

una de estas metadatos (sinónimos para web query) pueden ser editadas, de tal manera que los campos sean personalizados (etiquetas, tipos de datos, tamaño de campos)

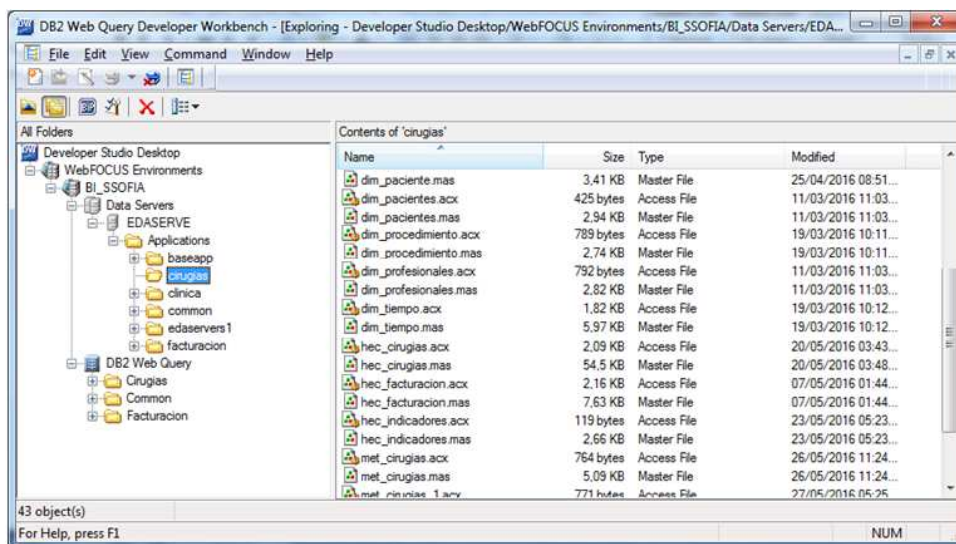


Figura 42. Listado de metadatos (Sinónimos Web Query)  
Elaboración propia. Instantáneas DB2 Web Query Developer Workbench.

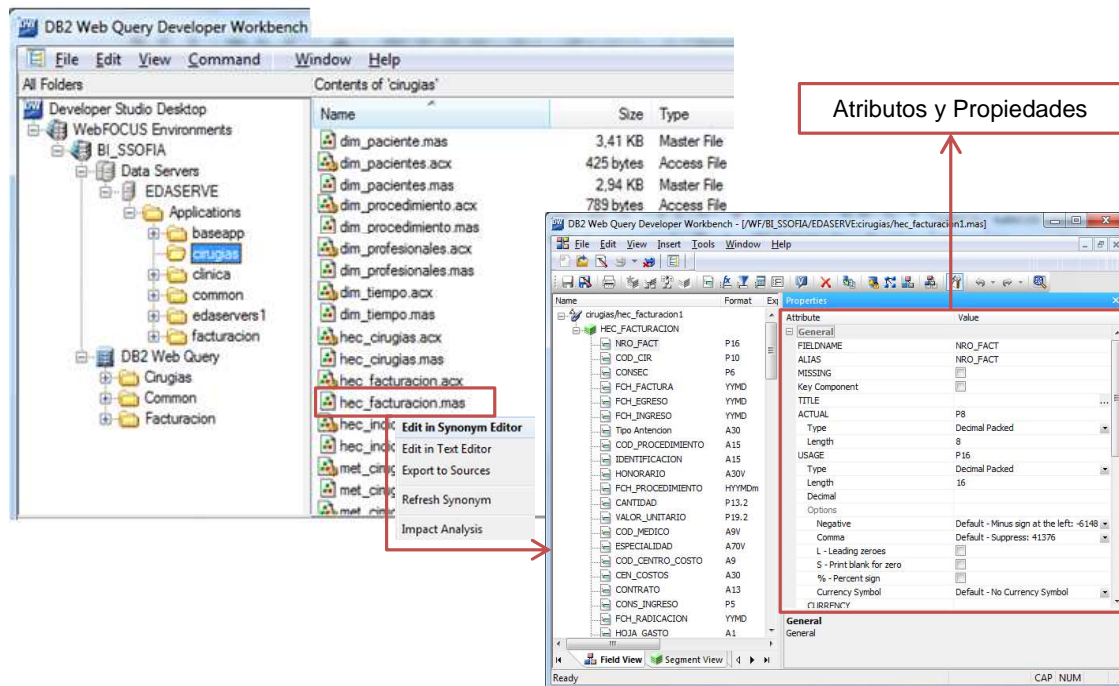


Figura 43. Edición de sinónimos y atributos  
Elaboración propia. Instantáneas DB2 Web Query Developer Workbench. Elaboración propia.

De igual forma, al editar el sinónimo de hechos de cirugías, se procedió a la creación de relaciones entre dicha tabla y las dimensiones. Los relaciones (join's) permiten enlazar los datos y garantizar la integridad y granularidad de los reportes a generar. (Bedoya et Al, 2014)

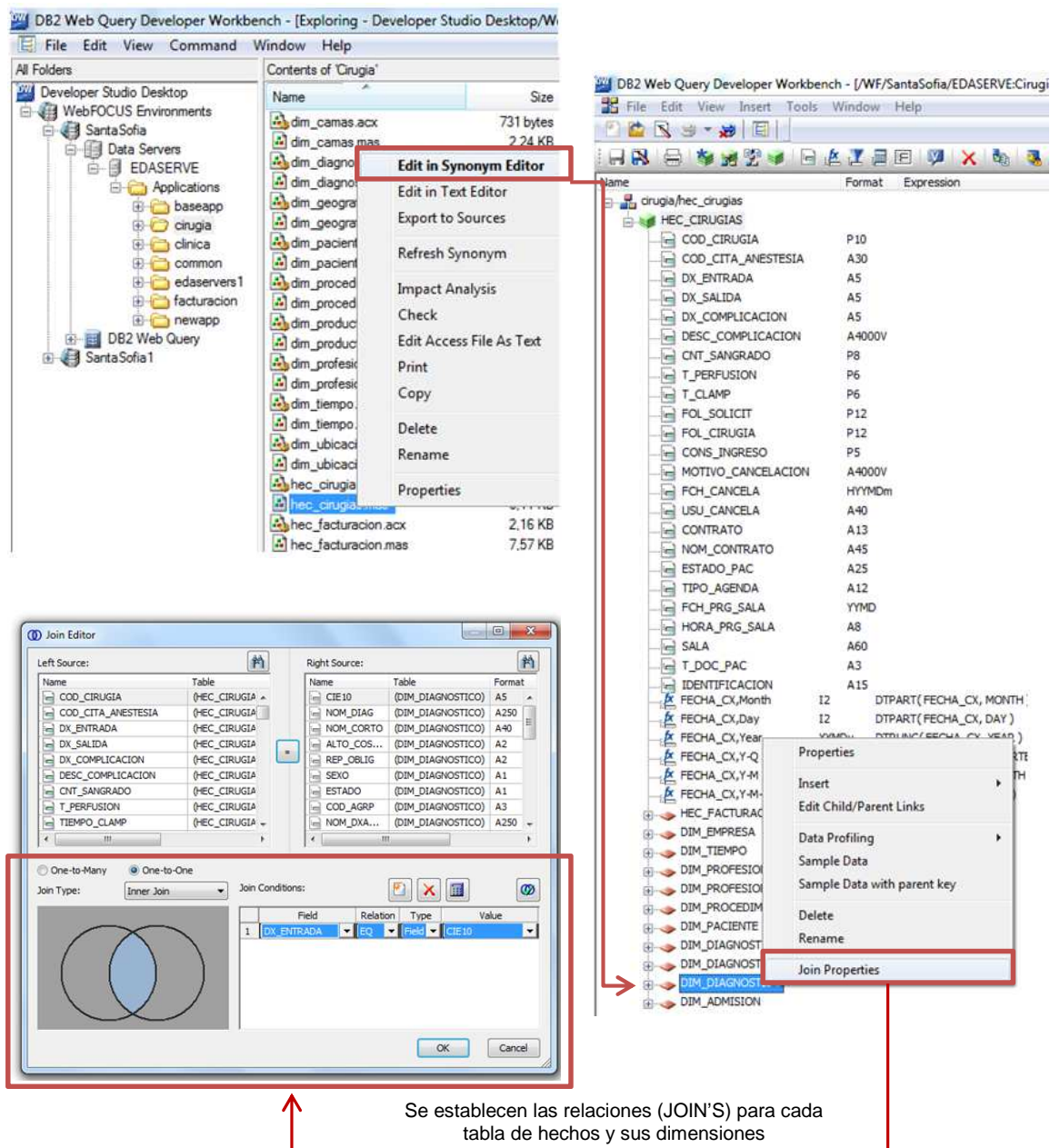


Figura 44. Creación de relaciones entre tablas de hechos y dimensiones  
Elaboración propia. Instantáneas DB2 Web Query Developer Workbench.

Se crearon los join's entre las tablas de hechos y cada una de las tablas dimensionales, buscando con ello alcanzar el nivel de granularidad necesario en la funcionalidad **drill down**, así como la integridad de los datos que se exploren a través de los reportes y cubos OLAP.

Las imágenes adicionales sobre el proceso llevado a cabo pueden ser consultados en el *Anexo 3 Documentación Adicional de la Fase III*, en la cuales se pueden apreciar conexiones del tipo JOIN y LEFT OUTER JOIN.

Luego de definir las conexiones entre la tabla de hechos y de dimensiones, se procedió a definir los campos de cada dimensión que serán visualizados a través del cubo OLAP. El usuario administrador de la solución puede agregar o quitar elementos de acuerdo a las necesidades el proceso de cirugía. Con la función “arrastrar y soltar” se pueden pasar del área izquierda (metadatos) los atributos que integrarán cada dimensión.

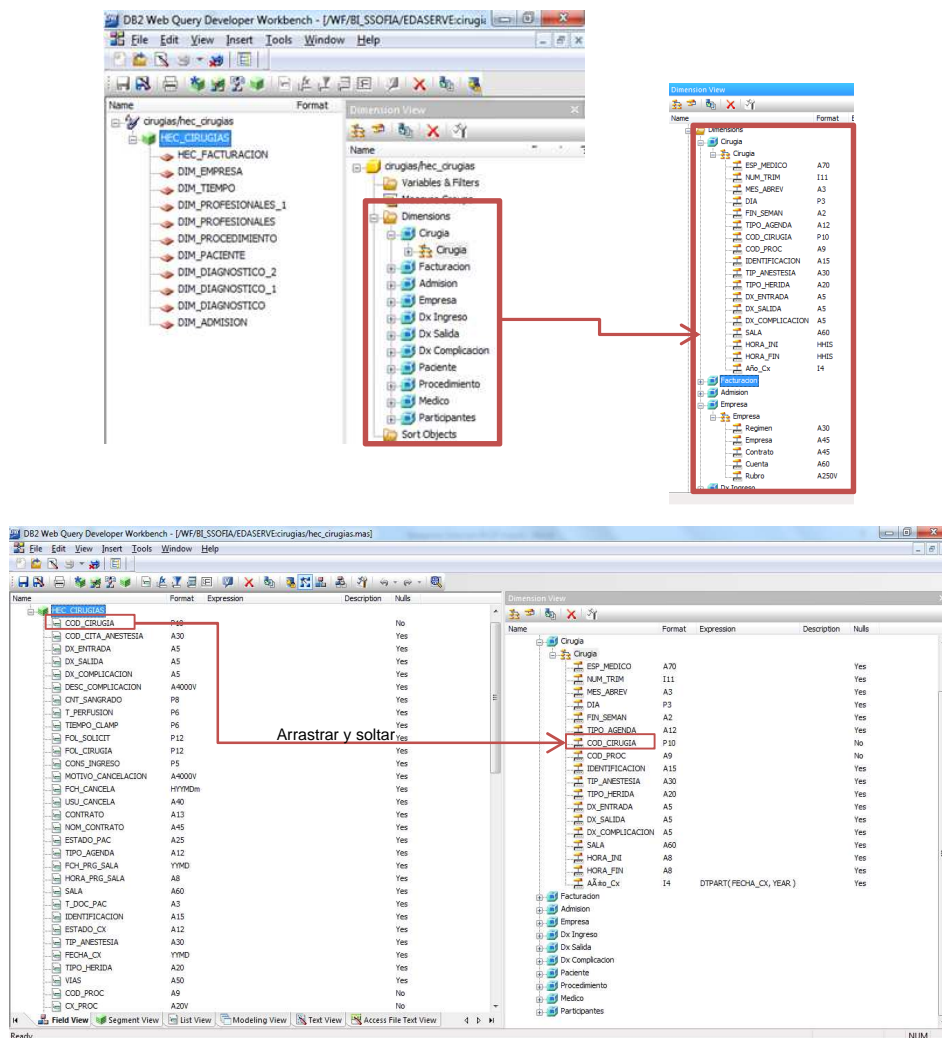


Figura 45. Interfaz general vista de dimensiones  
Elaboración propia. Instantáneas DB2 Web Query Developer Work Bench

### 3.6.3.3.2. Modelo conceptual de los datos

Creadas las metadatos y las respectivas relaciones entre las tablas de hechos y las tablas de dimensiones, se estableció un modelo de datos bajo el **esquema de constelación de hechos**, incluyendo, varias tablas de hechos y que a la vez, comparten las tablas de dimensiones especificadas, ofreciendo un modelo integral y escalable. (Kimball & Ross, 2013)

El modelo conceptual se basa en identificar qué tipo de procesos y vistas de negocio proporcionan la respuesta a las preguntas que tienen los usuarios finales (Curto y Conesa, 2011).

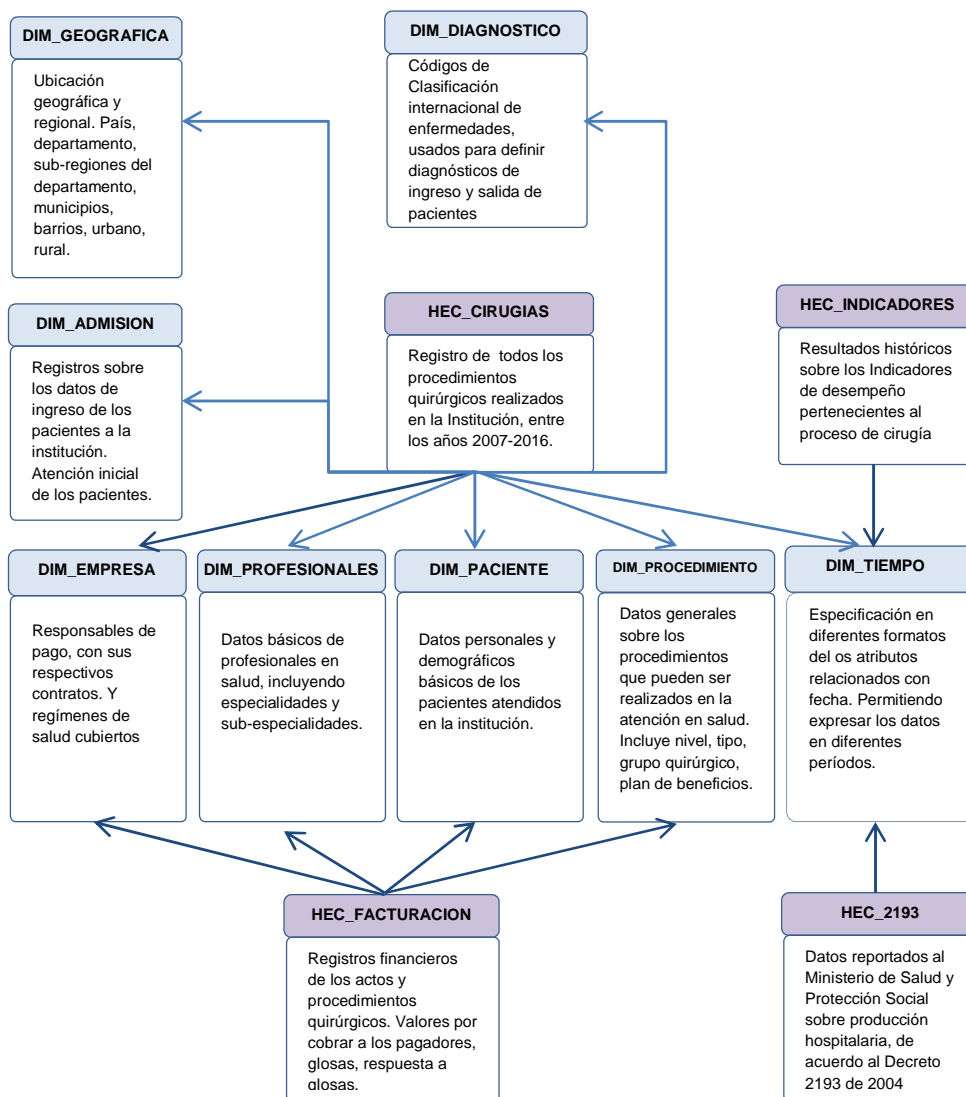


Figura 46. Modelo Conceptual de los datos  
Elaboración Propia.

### 3.6.3.3. Modelo Lógico de los datos

Luego de establecer las vistas del negocio que contienen las respuestas a los requerimientos de información, se procede a describir los atributos de las tablas de hechos y dimensiones. Indicando las claves subrogadas que identifican cada registro y las claves foráneas (Curto y Conesa, 2010)

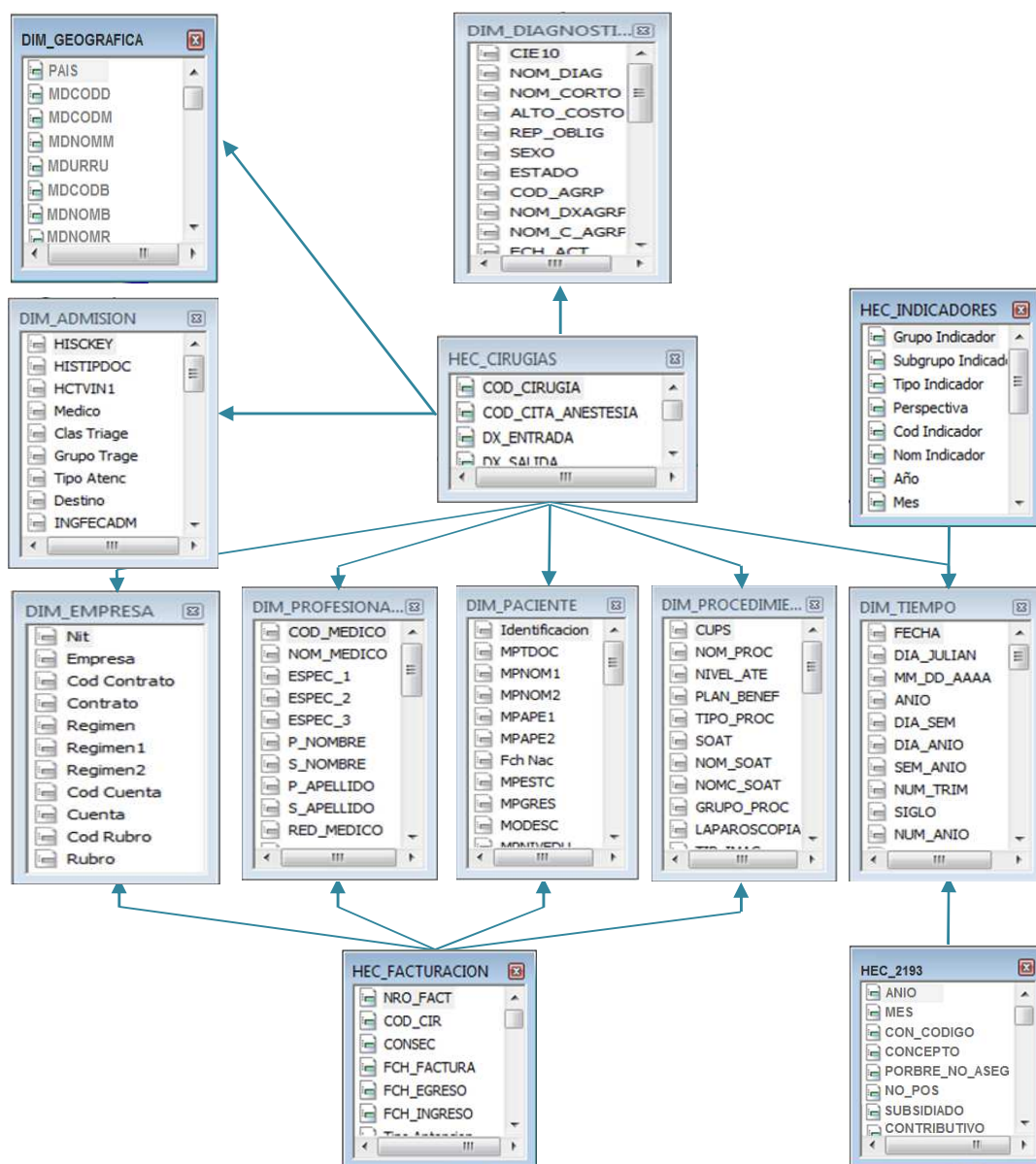


Figura 47. Diseño lógico de Data Warehouse.  
Adaptado de (Curto y Conesa, 2011)

### 3.6.3.3.4. Diseño físico de datos

Se presenta el diseño físico de los datos de tal manera que se puede observar el almacenamiento físico de los datos (Curto y Conesa, 2011), se utilizó el motor de bases de datos DB2 como Data Warehouse y la herramienta Data Studio para la modelización.

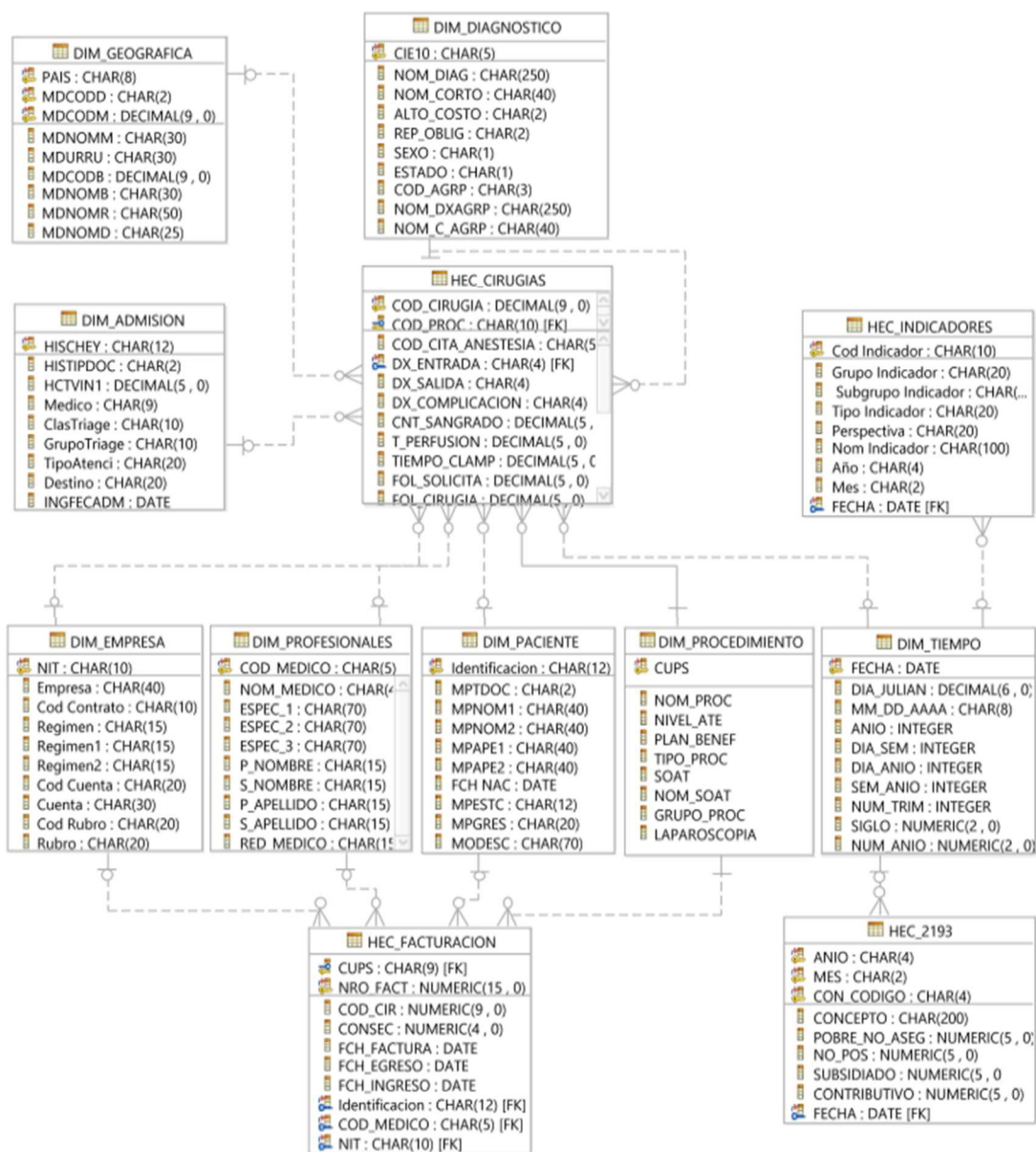


Figura 48. Diseño físico de los Datos.  
Exportad de IBM Data Studio Versión 4.1.2

### 3.6.4. Fase IV. Modelado

#### 3.6.4.1. Objetivo

Modelar la solución de inteligencia de negocios para el proceso de cirugía del Hospital Departamental Universitario Santa Sofía de Caldas.

En la documentación de la Fase III “Preparación de los Datos”, se pudo apreciar el modelo de datos conceptual, lógico y físico. Así mismo, la descripción detallada de cada una de las tablas que dan origen a los hechos y dimensiones de la solución de Inteligencia de negocios. Así mismo, se apreciaron imágenes del uso de IBM Web Query para la construcción del modelo dimensional.

En la fase IV “Modelado”, se presentan las ilustraciones y tablas resultantes luego de la construcción de la solución de inteligencia de negocios, se integran al documento imágenes de las herramientas IBM DB2 Web Query – IBM Developer Work Bench, para evidenciar la construcción de reportes, tableros de control (dashboard) y cubos OLAP, así como los resultados obtenidos de la aplicación de CLUSTERING como técnica de minería de datos y específicamente el algoritmo K-Means, para obtener grupos de pacientes que fueron sometidos a procedimientos quirúrgicos entre los años 2007 - 2015 y cuyos diagnósticos correspondieron a tumores malignos.

En el presente capítulo se muestra una secuencia de imágenes que permiten evidenciar el desarrollo de la solución para cinco (5) de los once (11) requerimientos funcionales establecidos en la Fase I “Comprensión del Negocio”:

- RF-1 OLAP Cirugía
- RF-2 OLAP Financiero
- RF-5. Tablero Facturación
- RF-8. Reportes para dispositivos móviles
- RF-9. Personalización y exportación de informes.
- RF-11. Clustering para obtener grupos de pacientes.

**En el Anexo 4 Documentación Adicional Fase IV Modelado, se podrán consultar las evidencias correspondientes a formatos de evaluación y los requerimientos funcionales restantes:**

- RF-3. Monitoreo de tiempos quirúrgicos
- RF-4. Tablero Indicadores de Cirugía
- RF-6. Re intervenciones quirúrgicas.
- RF-7. Indicadores de producción 2193
- RF-10. Funcionalidad Active Report.

La aplicación – IBM Developer Work Bench presenta una interfaz general dividida en dos secciones, la primera para administrar los sinónimos de Web Query o metadatos y la

segunda para administrar y construir cada uno de los tipos de reporte ofrecidos por el software. (Bedoya et Al, 2014. p25)

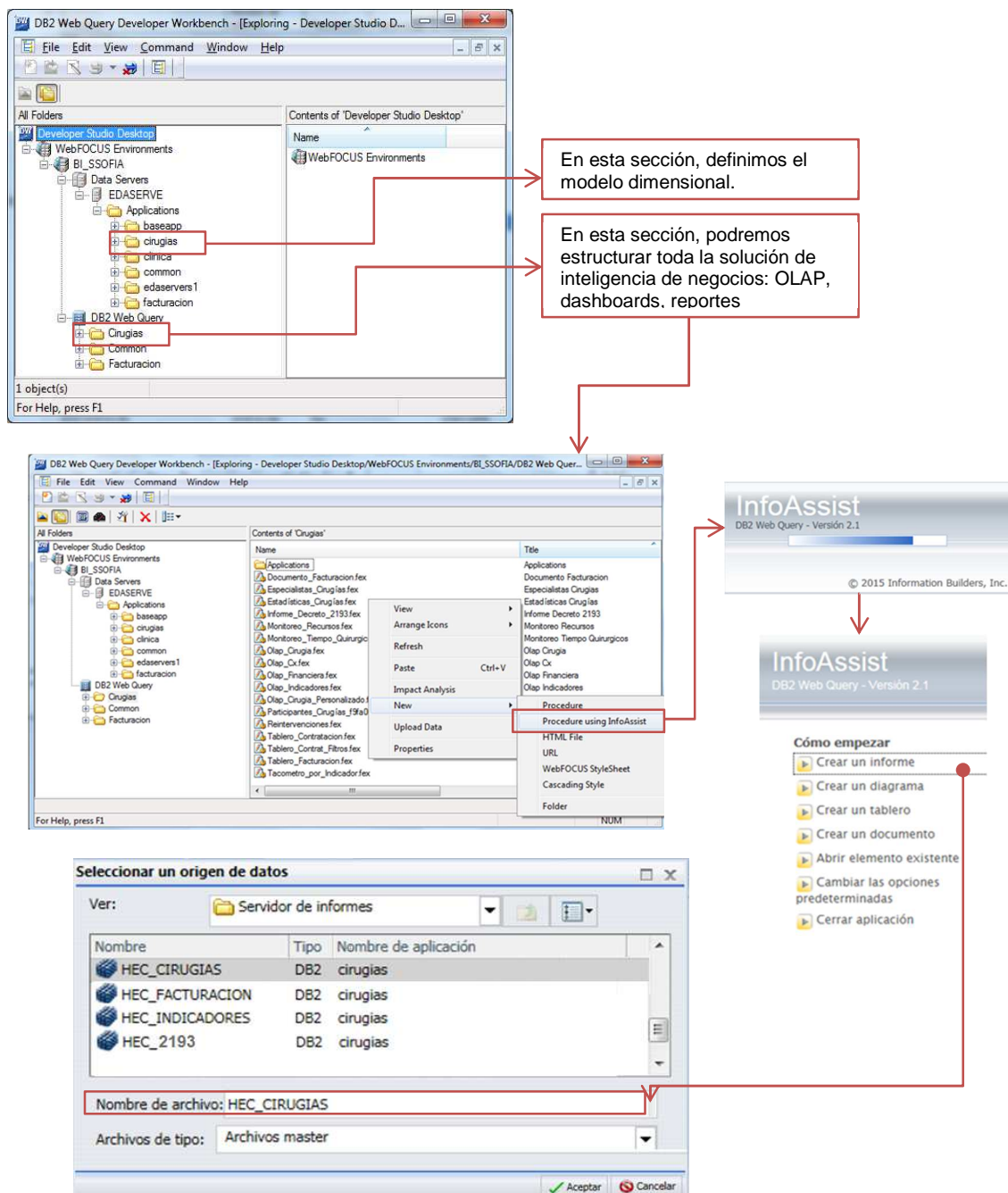


Figura 49. Proceso para la creación de un nuevo reporte en Web Query Instantánea tomada de DB2 Web Query Developer Workbench

Luego de creado el reporte, aparece en el listado el archivo de extensión .fex, el cual puede ser editado para proceder a la creación de los reportes que se deseen incorporar a la solución de inteligencia de negocios. (Bedoya et Al, 2014. p857 )

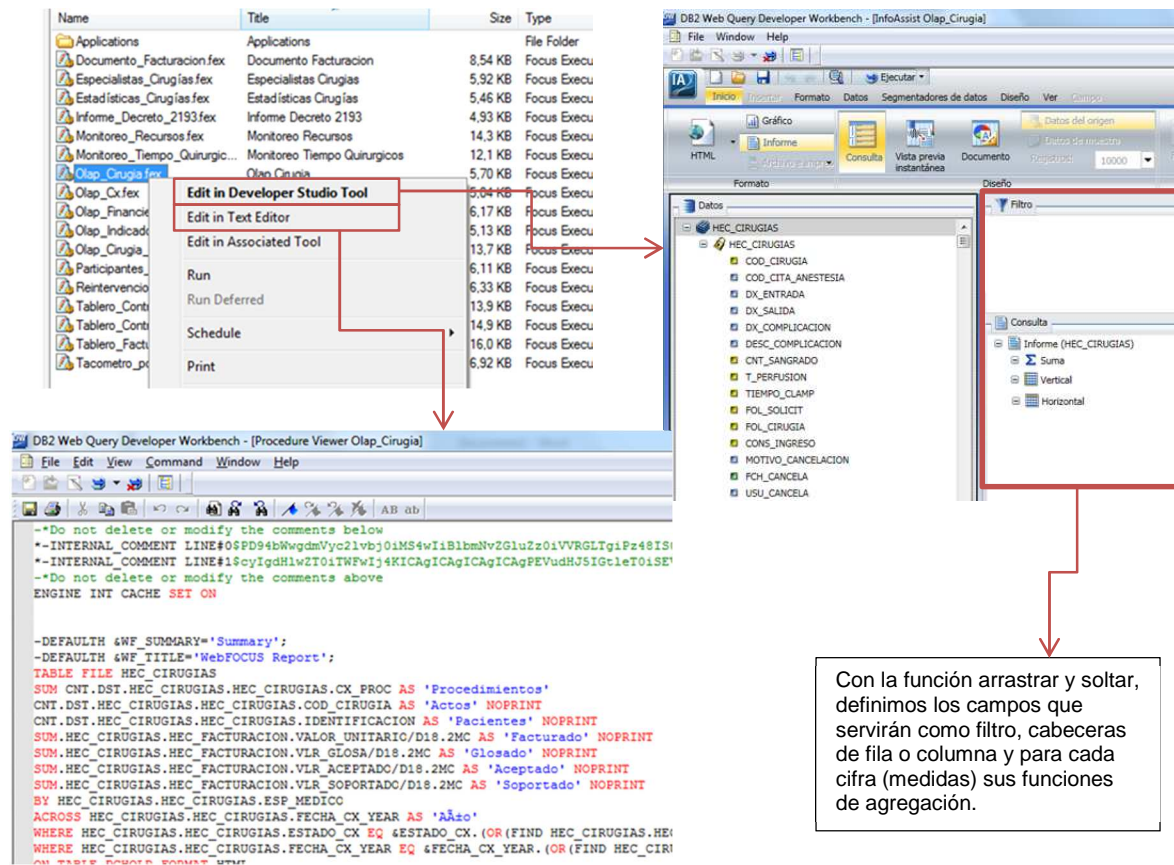


Figura 50. Opciones de edición de reportes en Web Query Instantánea tomada de DB2 Web Query Developer Workbench

En la siguiente ilustración se aprecia el framework sobre el cual se desarrollaron los diferentes tipos de reporte. A la izquierda, el árbol que incluye las tabla de hechos y sus dimensiones. En la parte superior derecha, la sección para insertar campos que servirán como parámetros de entrada a los informes y en la parte inferior derecha, con arrastrar y soltar se agregan los atributos que aparecerán en el reporte.

Es posible establecer funciones de agregación (suma, promedio, cuenta, porcentaje, máximo, mínimo, total, mediana, entre otras) a los atributos numéricos.

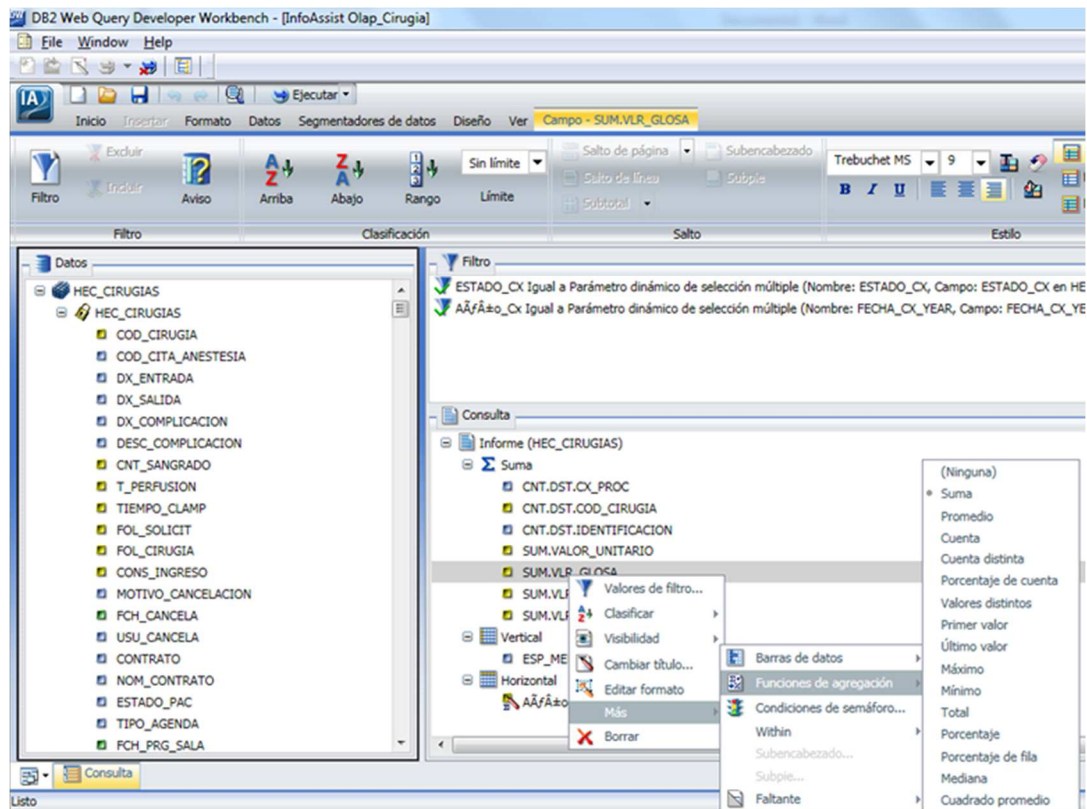


Figura 51. Framework DB2 Web Query Developer Workbench

Se implementaron también las funciones de semaforización que ofrece Web Query para personalizar los reportes y establecer escalas de valoración de acuerdo a las columnas calculadas.

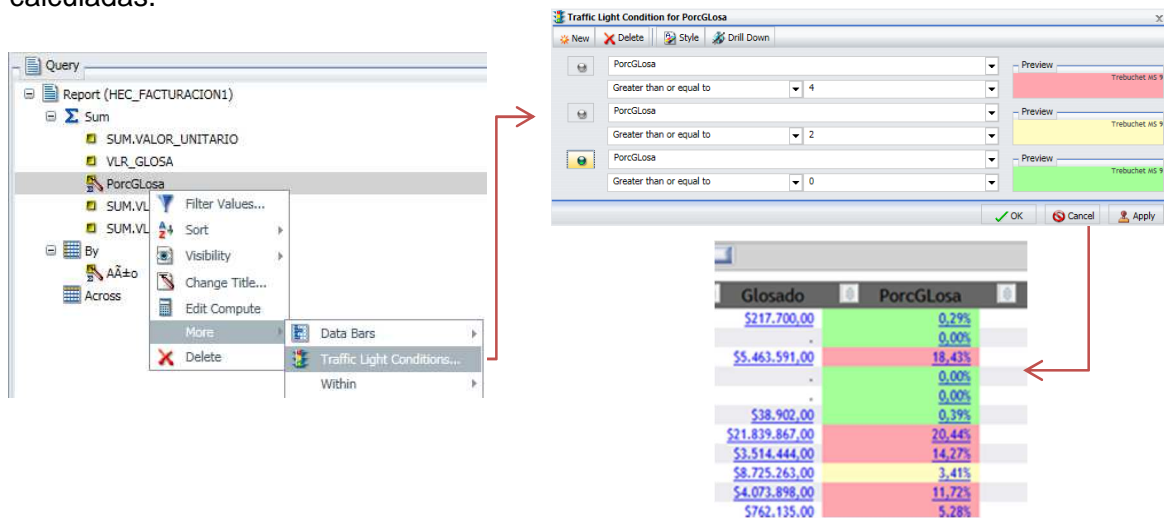


Figura 52. Semaforización de atributos  
Instantánea tomada de DB2 Web Query Developer Workbench

### 3.6.4.2. RF-1 OLAP Cirugía

La solución de Inteligencia de Negocios permitirá generar informes dinámicos mediante herramienta OLAP, para analizar datos asistenciales.

Se construyó el reporte “**OLAP Cirugías**” en la solución de Inteligencia de negocios, el cual permite estructurar reportes utilizando datos de diferentes dimensiones y agregar datos del proceso de atención de acuerdo a las necesidades.

En la siguiente ilustración se observa el menú de reportes creados en la solución de inteligencia de negocios.

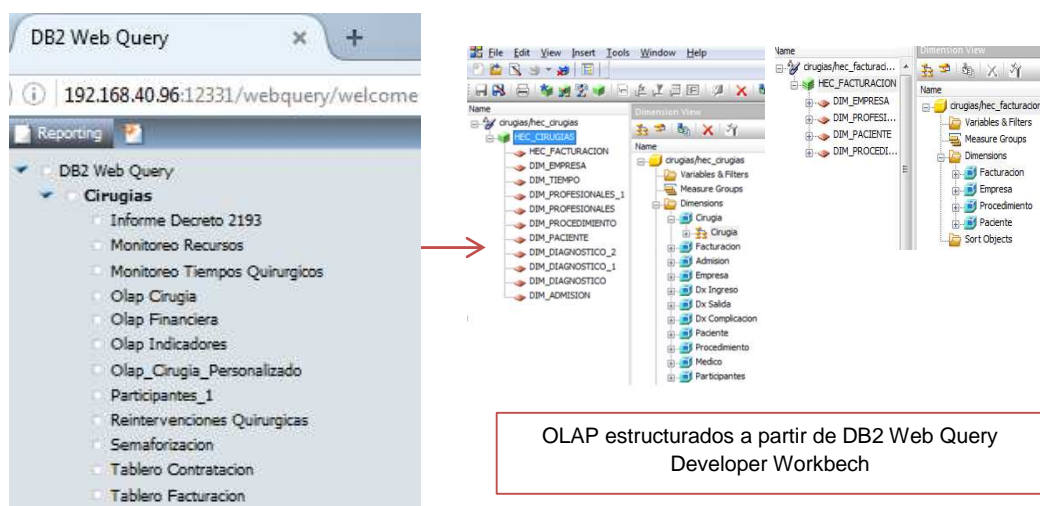


Figura 53. Menú principal Solución de Inteligencia de Negocios Instantáneas de DB2 Web Query Developer Workbench (derecha)

El OLAP construido permite obtener el consolidado de todos los procedimientos quirúrgicos realizados entre los años 2007-2016. Integra dimensiones como: Cirugía, facturación, Admisión, Empresa, Dx Ingreso (Diagnóstico de ingreso), Dx Egreso (Diagnóstico de egreso), Dx Complicación (Diagnóstico de complicación), Paciente, Procedimiento, Médico y Participantes.

Inicia con una cifra acumulada del número de cirugías por cada año, para cada especialidad y mediante la función **drill down**, el usuario podrá acceder a diversos niveles de granularidad. Con “arrastrar o soltar” o mediante el botón OLAP, se podrán agregar o quitar campos del reporte.

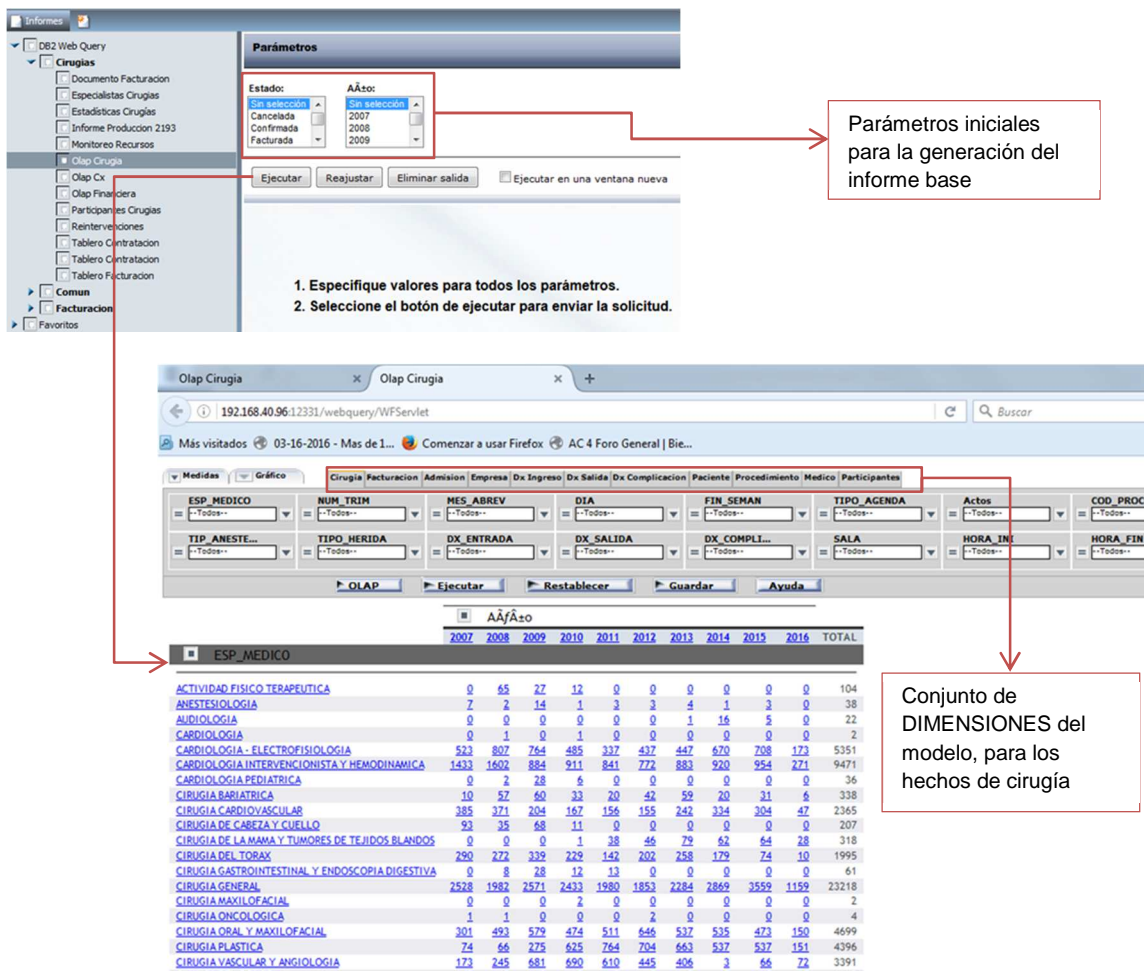


Figura 54. OLAP Cirugía. Ejecución del Informe para todos los años. Instantánea tomada de Solución de BI

Con el fin de permitir al usuario definir los datos iniciales para el reporte, se ofrece la opción de filtrado. En esta oportunidad se incluyeron dos variables: Estado de la cirugía (Cancelada, Confirmada, Facturada, Realizada) y una lista de selección múltiple para elegir los años que se desea incluir en la consulta.

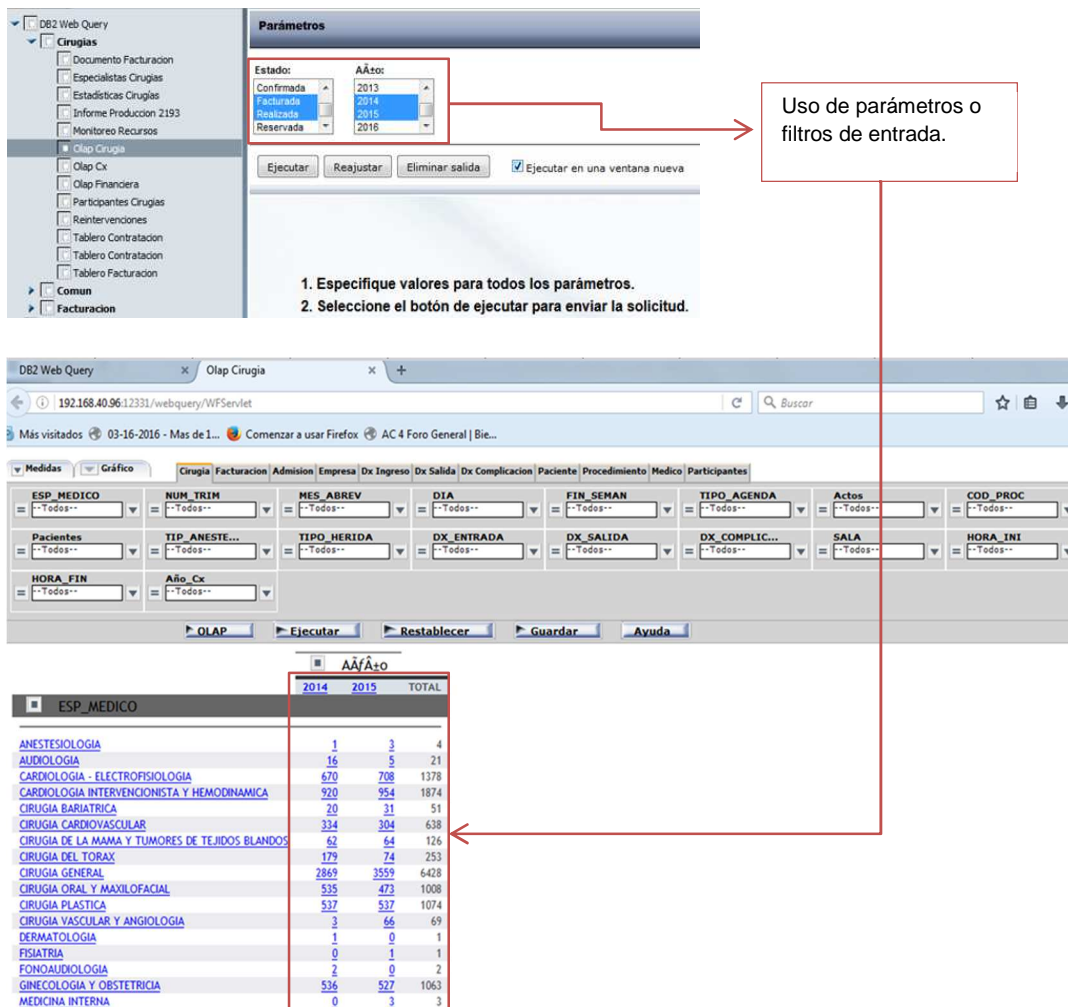


Figura 55. OLAP Cirugía. Ejecución del Informe aplicando filtro para los años 2014-2015. Instantánea tomada de Solución de BI

El informe puede ser ejecutado en la misma ventana o elegir una ventana nueva para brindar más espacio de visualización. Al dar clic en el botón ejecutar, Web Query muestra un resumen de total de cirugías, clasificadas por especialidad y año de realización.

### 3.6.4.2.1. Dimensiones del Cubo

En la documentación de la fase III “Preparación de los datos” se especificaron y construyeron las tablas dimensionales. Cada atributo definido puede ser agregado al informe, sin embargo para evitar saturación de la interfaz del usuario, se han elegido algunos de ellos. El usuario administrador podrá agregar o quitar atributos de acuerdo a la necesidad del proceso.

A continuación se presentan ilustraciones sobre algunas de las dimensiones que incluye el OLAP cirugía. Mediante “Arrastrar y soltar” el usuario puede agregar un atributo a la presentación del reporte, de igual forma utilizando el botón OLAP que aparece en la parte inferior de la sección de dimensiones.

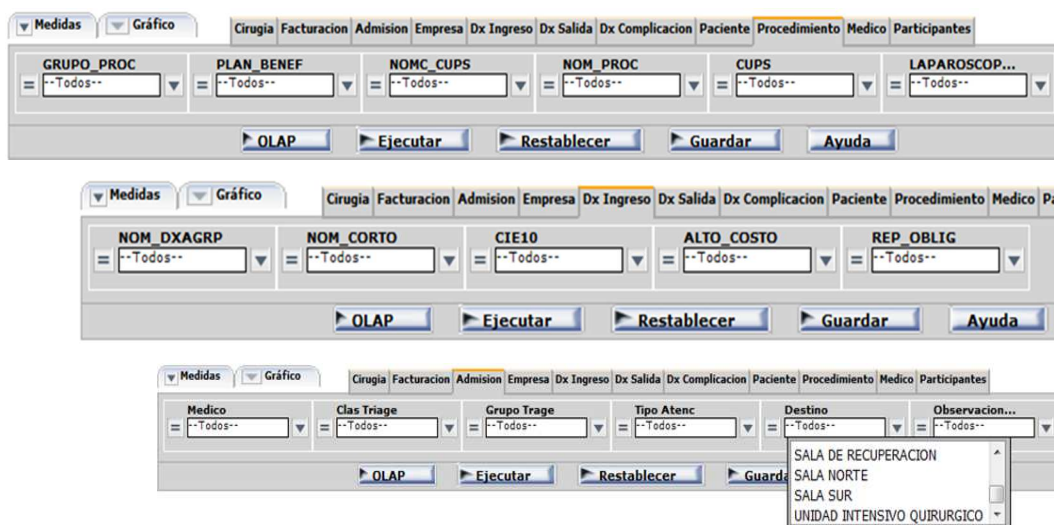


Figura 56. Dimensiones OLAP Cirugía.  
Instantánea tomada de Solución de BI

### 3.6.4.2.2. Granularidad del cubo OLAP Cirugía

La granularidad consiste en ofrecer al usuario la capacidad de adentrarse en los datos, dando clic sobre las medidas (valores) o sobre las etiquetas de fila o columna. (Bedoya et Al, 2014. p 421) de esta manera, se puede iniciar con un informe básico e ir logrando mayor nivel de especificidad.

En la siguiente ilustración se observa los resultados obtenidos al ir dando clic sobre los valores (drill down):



Figura 57. OLAP Cirugía. Granularidad. Funcionalidad drill down. Instantánea tomada de Solución de BI

Al dar clic sobre las 9 cirugías cardiovasculares que se realizaron el día 24 de julio de 2014, se observan los datos generales de los nueve procedimientos que se realizaron en 3 actos. En esta instancia, el usuario puede agregar otros datos de las dimensiones (paciente, admisión, procedimiento, diagnóstico, etc.)

Para ello se pueden arrastrar los campos que se quieran desde las dimensiones o personalizar el contenido a través de la opción **OLAP**

The screenshot displays a web browser window with a data table and an overlaid OLAP control panel. The table has columns: TIPO\_AGENDA, COD\_CIRUGIA, COD\_PROC, NOMC\_CUPS, and a numerical column. The OLAP panel, titled 'Panel de control de WebFOCUS OLAP', contains sections for 'Dimensiones-Pulse cuadros que añadir', 'Drill horizontal', 'Drill Down', and 'Medidas'. A red box highlights the 'OLAP' button in the table's toolbar, with a red arrow pointing to the 'Drill Down' section in the OLAP panel. The 'Medidas' section includes checkboxes for 'Procedimientos', 'Autos', 'Pacientes', 'Facturado', 'Closado', 'Aceptado', and 'Soportado'.

TIPO_AGENDA	COD_CIRUGIA	COD_PROC	NOMC_CUPS	
Programada	63034	361501	ANASTOMOSIS SIMPLE O SECUENCIAL DE ARTE	4446
		361701	ANASTOMOSIS CORONARIA PARA REVASCULARIZA	4446
		377401	INSERCIÓN O SUSTITUCIÓN DE ELECTRODO EPI	4446
	63049	361501	ANASTOMOSIS SIMPLE O SECUENCIAL DE ARTE	4602
		361701	ANASTOMOSIS CORONARIA PARA REVASCULARIZA	4602
		377401	INSERCIÓN O SUSTITUCIÓN DE ELECTRODO EPI	4602
	63050	361501	ANASTOMOSIS SIMPLE O SECUENCIAL DE ARTE	1419
		361701	ANASTOMOSIS CORONARIA PARA REVASCULARIZA	1419
		377401	INSERCIÓN O SUSTITUCIÓN DE ELECTRODO EPI	1419

Figura 58. Personalización atributos OLAP Instantánea tomada de Solución de BI

### 3.6.4.3. RF-2. OLAP Financiero

La solución de Inteligencia de Negocios permitirá generar informes dinámicos mediante herramienta OLAP, para analizar datos financieros.

El segundo OLAP construido para la solución de Inteligencia de Negocios fue denominado “**OLAP Financiera**” y permite integrar datos correspondientes a valores facturados, glosados, aceptados y soportados, así como integrar datos de profesionales, empresas y pacientes. Aunque no son obligatorios, se cuenta con parámetros iniciales; es posible elegir el tipo de registro (factura, prefactura), el año y el concepto quirúrgico (Honorarios, anestesiólogo, ayudante quirúrgico, derechos de sala y materiales)

En las siguientes imágenes se pueden apreciar resultados del cubo “OLAP Financiera”, de igual manera se satisface una de las funcionalidades exigidas como resultado de la investigación y es “La aplicación de semaforización de valores” (Bedoya et Al, 2014. p 417)

The image shows the OLAP Financiera application interface. The top part displays the 'Parámetros' (Parameters) section with three dropdown menus: 'Tipo' (Type) set to 'Factura', 'Año' (Year) set to '2013', and 'Concepto' (Concept) set to 'ANESTESIOLOGO'. Below the dropdowns are buttons for 'Ejecutar' (Execute), 'Reajustar' (Reset), and 'Eliminar salida' (Clear output), along with a checkbox for 'Ejecutar en una ventana nueva' (Execute in a new window). A red box highlights these parameters and buttons. Below the interface, a red arrow points to the 'Ejecutar' button.

The bottom part of the image shows the resulting data table in a browser window. The table has the following columns: 'Año' (Year), 'Facturado' (Factured), 'Glosado' (Billed), 'PorcGlosa' (Billed Percentage), 'Aceptado' (Accepted), and 'Soportado' (Supported). The data is color-coded: green for 2013, yellow for 2014, and red for 2015. A red box highlights the table, and a red arrow points to the 'Soportado' column for the year 2015.

Año	Facturado	Glosado	PorcGlosa	Aceptado	Soportado
2013	\$9.739.227.086,00	\$131.378.897,00	1,35%	\$30.625.276,00	\$90.273.153,00
2014	\$10.083.621.262,50	\$301.901.419,00	2,99%	\$43.241.612,00	\$250.985.286,00
2015	\$11.783.340.111,01	\$640.050.321,00	5,43%	\$82.223.413,00	\$498.904.577,00
TOTAL	\$31.606.188.459,51	\$1.073.330.637,00	3,40%	\$156.090.301,00	\$840.163.016,00

Figura 59. Ejecución informe OLAP Financiera Instantánea tomada de Solución de BI

Se observa en la siguiente imagen, la granularidad que permite aplicar el OLAP Financiero, descendiendo desde las vigencias anuales, a los períodos trimestrales y mensuales. Así mismo, la posibilidad de agregar atributos desde las dimensiones.



Figura 60. Drill Down OLAP Financiera Instantánea tomada de Solución de BI

También es posible especificar los datos a nivel de EPS y desde allí acceder a los valores financieros de los contratos.

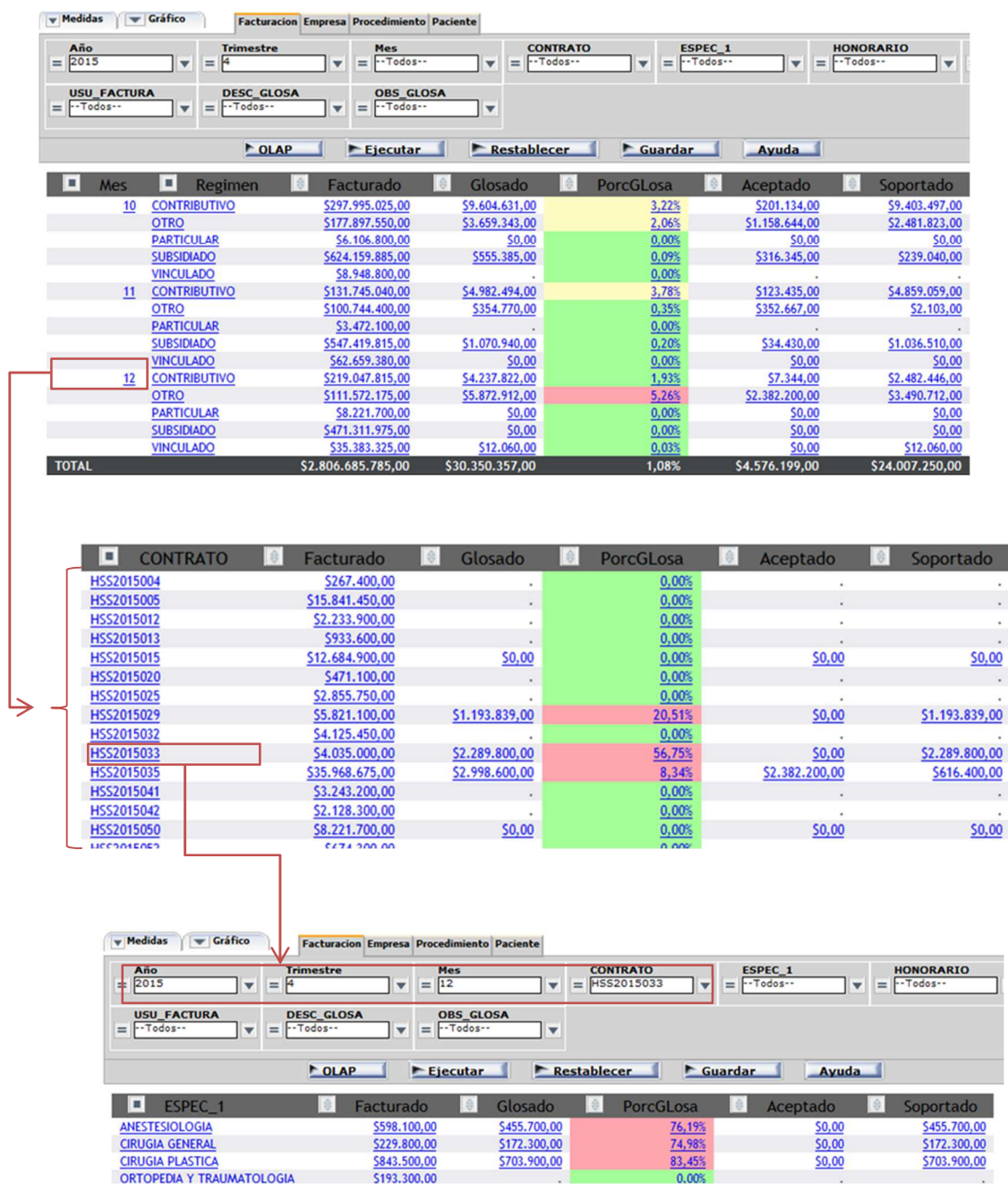


Figura 61. Drill Down y agregación de atributos OLAP Financiera. Instantánea tomada de Solución de BI

La siguiente imagen presenta otros ejemplos de personalización del OLAP Financiero, en este caso, las cifras facturadas, glosadas, soportadas y aceptadas para diferentes regímenes de salud y EPS. De igual forma, se detallan los tipos de cirugía que se han facturado por pagador.

Trimestre	Regimen	Empresa	Facturado	Glosado	PorcGlosa	Aceptado	Soportado
1	CONTRIBUTIVO	CAFESALUD	\$74.639.200,00	\$217.700,00	0,29%	\$217.700,00	\$0,00
		COMPENSAR	\$2.513.800,00	-	0,00%	-	-
		COOMEVA EPS	\$29.646.550,00	\$5.463.591,00	18,43%	\$1.494.200,00	\$3.969.391,00
		COSMITET LTDA	\$63.224.100,00	-	0,00%	-	-
		EPS FAMISANAR	\$3.416.900,00	-	0,00%	-	-
		EPS Y MEDICINA PREPAGADA SURAMERICANA SA	\$10.085.875,00	\$38.900,00	0,39%	\$38.900,00	\$0,00
		OTRO	\$106.836.635,00	\$21.839.867,00	20,44%	\$187.100,00	\$21.652.767,00
		NUEVA EMPRESA PROMOTORA DE SALUD	\$24.636.000,00	\$3.514.444,00	14,27%	\$279.557,00	\$1.782.457,00
		SALUD TOTAL EPS-S S.A	\$255.881.575,00	\$8.725.263,00	3,41%	\$1.696.163,00	\$7.029.100,00
		SALUDCOOP EPS	\$34.750.200,00	\$4.073.898,00	11,72%	\$0,00	\$3.401.336,00
		SANITAS EPS SA	\$14.447.600,00	\$762.135,00	5,28%	\$311.635,00	\$442.000,00
		SOS EPS	\$34.857.200,00	-	0,00%	-	-
		ASOCIACION INDIGENA DEL CAUCA AIC EPS-INDIGEN	\$6.595.400,00	-	0,00%	-	-
		AXA COLPATRIA SEGUROS DE VIDA S.A.	\$42.836.400,00	\$7.226.740,00	16,87%	\$6.828.940,00	\$385.700,00
		AXA COLPATRIA SEGUROS S.A.	\$4.517.400,00	\$35,00	0,00%	\$0,00	\$35,00
		COLSANITAS SA	\$447.600,00	-	0,00%	-	-
		COMPAÑIA DE SEGUROS PAN AMERICAN LIFE	\$19.813.100,00	-	0,00%	-	-
		CONSORCIO SAYP	\$19.813.100,00	-	0,00%	-	-
		TOTAL	\$29.646.550,00	\$5.463.591,00	18,43%	\$1.494.200,00	\$3.969.391,00

Contrato	NOMC_CUUPS	Facturado	Glosado	PorcGlosa	Aceptado	Soportado
COOMEVA E.P.S CONTRIBUTIVO	APLICACION DE Tutores EXTERNOS EN PUJA'O	\$415.500,00	\$415.500,00	100,00%	\$415.500,00	\$0,00
	BIOPSIA DE PA'LMON POR TORACOTOMIA	\$892.600,00	-	0,00%	-	-
	CIRUGIA MINIMAMENTE INVASIVA DE COLUMNA	\$9.718.400,00	\$427.200,00	4,40%	\$0,00	\$427.200,00
	COLGAJO LIBRE COMPUSTO CON TECNICA MICR	\$1.542.100,00	-	0,00%	-	-
	COLGAJO LIBRE CUTANEO CON TECNICA MICROV	\$918.600,00	-	0,00%	-	-
	COLGAJO LOCAL DE PIEL COMPUSTO DE VECIN	\$179.000,00	-	0,00%	-	-
	DESBRIDAMIENTO ESCISIONAL POR LESION SU	\$715.800,00	\$628.474,00	87,80%	\$41.000,00	\$587.474,00
	DESBRIDAMIENTO ESCISIONAL POR LESION DE	\$933.600,00	\$933.600,00	100,00%	\$933.600,00	\$0,00
	DESCOMPRESION DE NERVI EN TUNEL DEL CAR	\$179.000,00	\$23.392,00	13,07%	\$0,00	\$23.392,00
	ESOFAGOGASTRODUODENOSCOPIA [EGD] CON BIO	\$297.500,00	-	0,00%	-	-
	EXTRACCION DE DISPOSITIVO IMPLANTADO EN	\$104.100,00	\$104.100,00	100,00%	\$104.100,00	\$0,00
	FASCIOTOMIA DE MANO SOD	\$235.900,00	\$41.608,00	17,64%	\$0,00	\$41.608,00
	FASCIOTOMIA EN ANTEBRAZO CON LIBERACION	\$153.050,00	\$27.816,00	18,17%	\$0,00	\$27.816,00
	IMPLANTACI'N O SUSTITUCI'N DE PROTESIS C	\$3.000.000,00	-	0,00%	-	-
	INCISION Y DRENAJE DE ESCROTO Y TUNICA	\$131.450,00	-	0,00%	-	-
	OCCLUSION PINZAMIENTO O LIGADURA DE VASO	\$2.027.800,00	-	0,00%	-	-
	ORQUIDORRAFIA O SUTURA DE TESTICULO SOD	\$1.528.100,00	\$869.682,00	56,91%	\$0,00	\$869.682,00
	REDUCCION ABIERTA DE FRACTURA EN PLATILL	\$1.795.900,00	\$1.217.633,00	67,80%	\$0,00	\$1.217.633,00
	REDUCCION ABIERTA DE FRACTURA EN SEGMENT	\$1.425.150,00	\$774.586,00	54,35%	\$0,00	\$774.586,00
	RESECCION DE TUMOR SUPRATENTORIAL HEMISF	\$484.900,00	-	0,00%	-	-
	RETIRO DE CATETER PERITONEAL NCOC	\$1.172.200,00	-	0,00%	-	-
	TRAQUEOSTOMIA SOD	\$1.795.900,00	-	0,00%	-	-
	TOTAL	\$29.646.550,00	\$5.463.591,00	18,43%	\$1.494.200,00	\$3.969.391,00

Figura 62. Nivel de detalle Regimen – EPS. OLAP Financiera Instantánea tomada de Solución de BI

En las pruebas efectuadas al modelo, se puede llegar a niveles de especificidad que permiten hacer seguimiento a cada glosa, con sus respectivos conceptos y aclaraciones.

Cuenta	NOMC_CUUPS	OBS_GLOSA	Facturado	Glosado	PorcGlosa	Aceptado	Soportado
PLAN OBLIGATORIO DE SALUD POS	APLICACION DE Tutores EXTERNOS EN PUJA'O		\$0,00		0,00%		
		Se acepta glosa por cobro no pertinente. Luz Emilia Restrepo Marin (md auditor).	\$415.500,00	\$415.500,00	100,00%	\$415.500,00	\$0,00
	BIOPSIA DE PA'LMON POR TORACOTOMIA		\$892.600,00		0,00%		
	CIRUGIA MINIMAMENTE INVASIVA DE COLUMNA	No se acepta glosa se factura tarifa propia de hospital , para paquetes, se debe tener en cuenta que a la fecha no se llen contrato entre las partes , por lo que aplica su cobro.	\$9.718.400,00	\$427.200,00	4,40%	\$0,00	\$427.200,00
	COLGAJO LIBRE COMPUSTO CON TECNICA MICR		\$1.542.100,00		0,00%		
	COLGAJO LIBRE CUTANEO CON TECNICA MICROV		\$918.600,00		0,00%		
	COLGAJO LOCAL DE PIEL COMPUSTO DE VECIN		\$179.000,00		0,00%		
	DESBRIDAMIENTO ESCISIONAL POR LESION SU	No se acepta glosa, se hace la claridad que dicho procedimiento fue realizado el 04/12/2014 fecha diferente a la de los colgajos que fueron realizados el 24/11/2014 y 26/11/2014, por lo que	\$501.200,00	\$501.200,00	100,00%	\$0,00	\$501.200,00

Figura 63. Nivel de detalle seguimiento a glosas. OLAP Financiera Instantánea tomada de Solución de BI

### 3.6.4.4. RF-5. Tablero Facturación.

La solución de inteligencia de negocios incluirá tablero para el seguimiento a valores facturados, glosados y soportados, por períodos de tiempo.

Para complementar el OLAP Financiero, la Jefatura de facturación solicitó la elaboración de un tablero que resumiera los datos correspondientes a los valores facturados, glosados, soportados y aceptados. Permite al igual que los tableros presentados hasta el momento, el filtrado por año y trimestre.

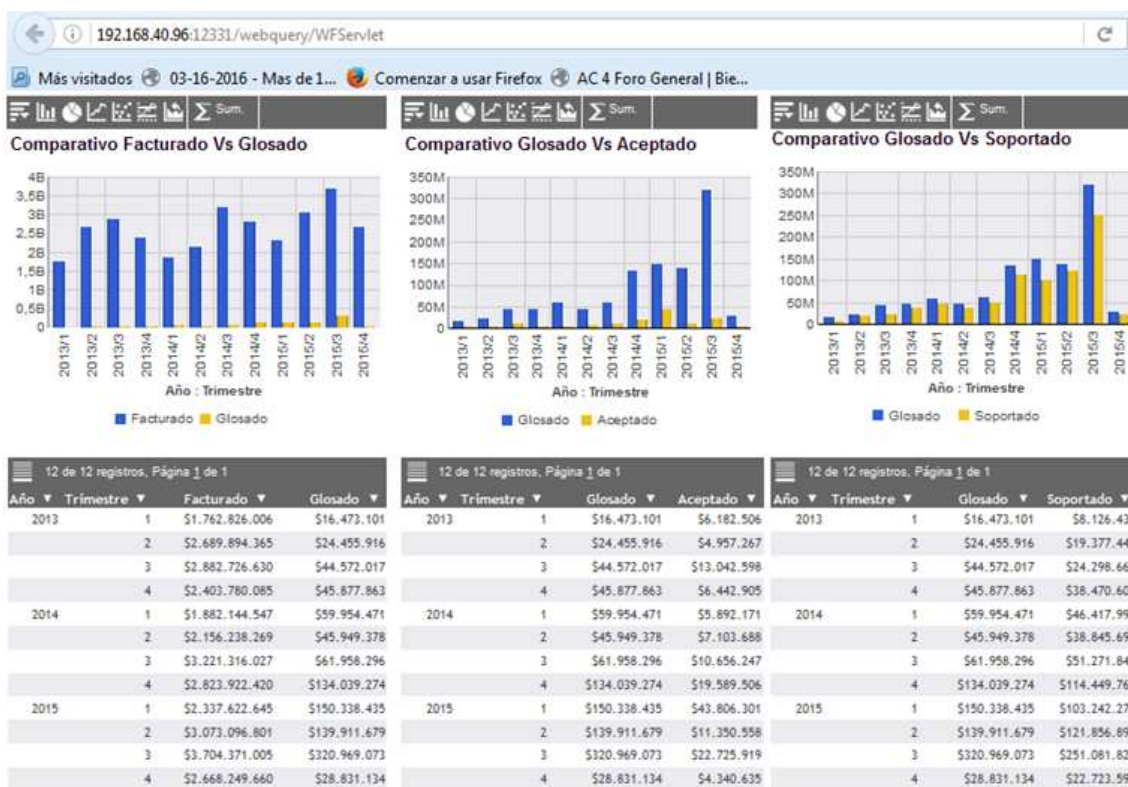


Figura 64. Tablero Monitoreo Facturación Ejemplo 1(2013-2015 todos los trimestres)  
Instantánea de Solución de BI



Figura 65. Tablero Monitoreo Facturación Ejemplo 2(2013-2015 último trimestre) Instantánea de Solución de BI

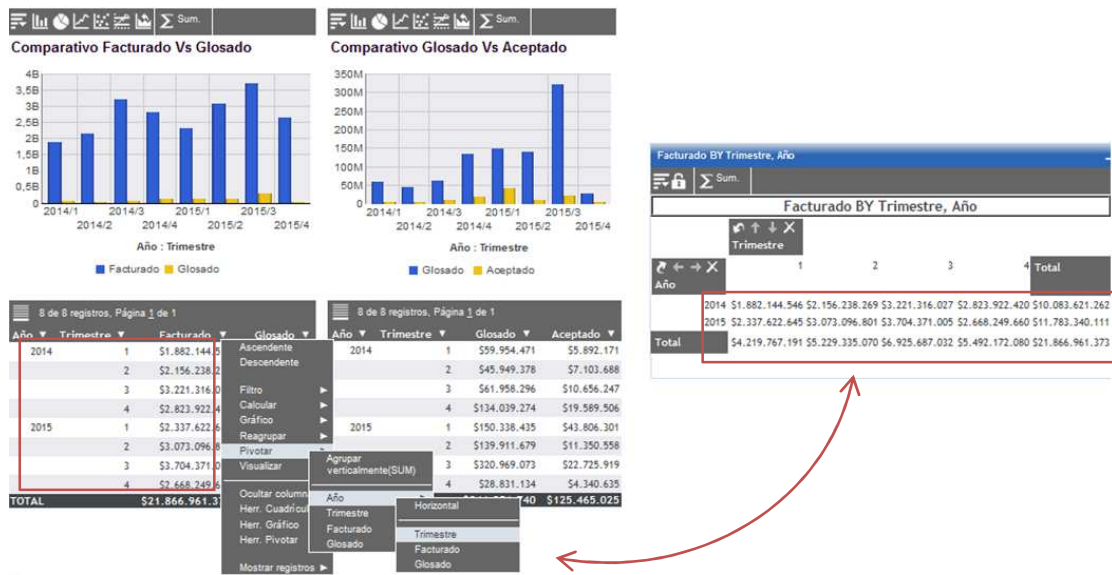


Figura 66. Tablero Monitoreo Facturación Ejemplo 3 (Función PIVOT) Instantánea de Solución de BI

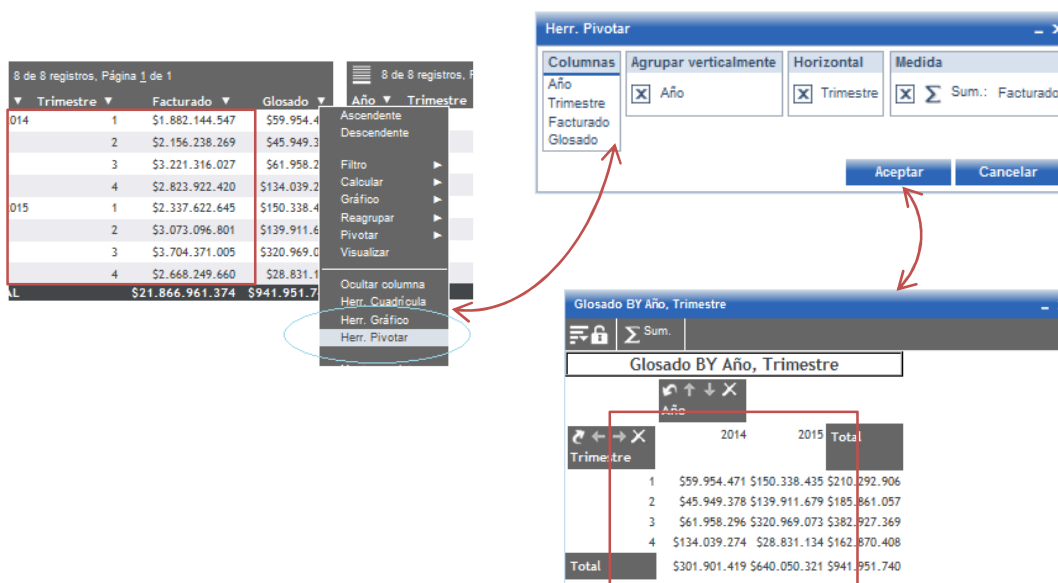


Figura 67. Tablero Monitoreo Facturación Ejemplo 4 (Función PIVOT)  
Instantánea de Solución de BI

### 3.6.4.5. RF-8. Reportes para dispositivos móviles.

#### La solución permitirá el uso de aplicativo móvil para la visualización de informes

Se ha complementado la solución de inteligencia de negocios, mediante la publicación de tres reportes a través de la herramienta ROAMBI ANALYTICS. A partir de los datos estructurados en la solución de inteligencia de negocios, se generaron tres (3) archivos en excel y se publicaron en un *drive* del correo corporativo. A partir de estos tres archivos se publicaron tres reportes que pueden ser visualizados en dispositivos móviles.

ROAMBI propone diversas plantillas de informe, así como distintos orígenes de datos (CSV, Excel, HTML, Google Drive, Box), incluir gráficos en las visualizaciones y definir varios niveles de agregación de la información.

Las siguientes corresponden a imágenes básicas de la interfaz del portal web de ROAMBI y de los reportes publicados y visualizados desde el dispositivo móvil.

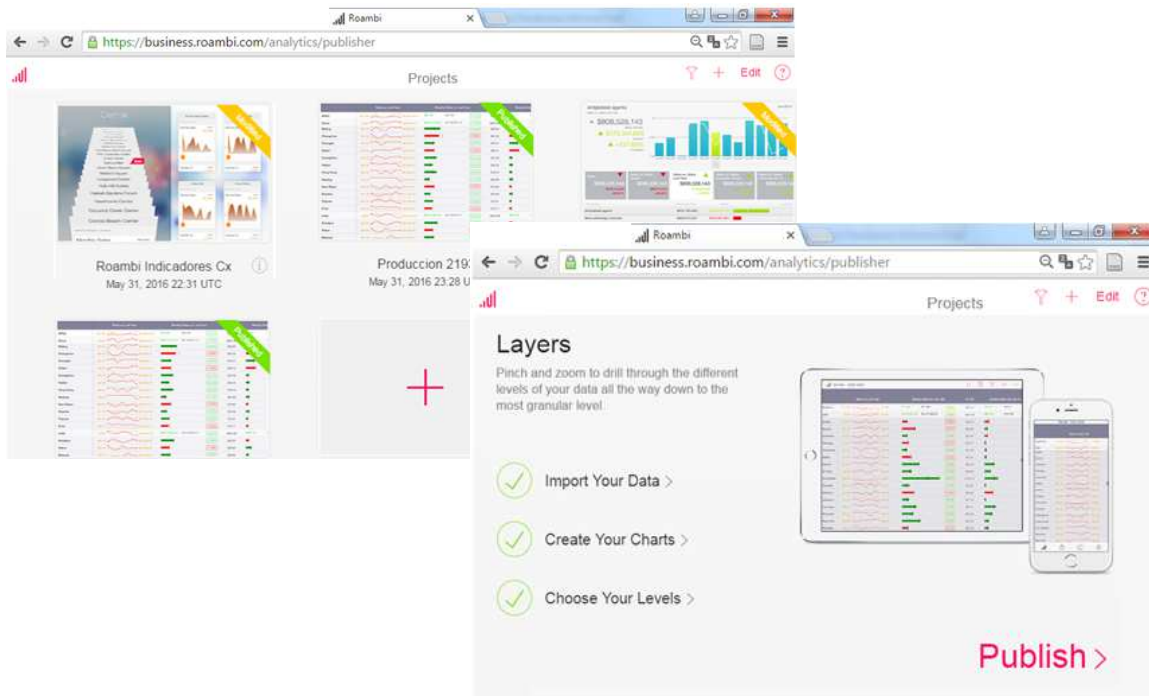


Figura 68. Framework de trabajo ROAMBI ANALYTICS. Instantáneas ROAMBI Analytics.



Figura 69. Reporte Indicadores Históricos de producción 2193. Dispositivo Móvil. Instantáneas ROAMBI Analytics.



Figura 70. Reporte Facturación Instantáneas ROAMBI Analytics.



Figura 71. Reporte Indicadores de Cirugía. Dispositivo Móvil. Instantáneas ROAMBI Analytics. Elaboración Propia

### 3.6.4.6. RF-9. Personalización y exportación de informes.

La solución permitirá que los usuarios puedan personalizar los informes y exportarlos en diferentes formatos (xls y pdf), así como generar reportes activos que puedan ser usados en modo off line.

Con el objetivo de presentar la funcionalidad para exportar los resultados en diferentes tipos de formato, se presentan a continuación dos ilustraciones. En la primera se ejecuta el OLAP Financiero y mediante *drill down* se accede a las cifras trimestrales del año 2015. En la segunda se agregan atributos desde las dimensiones y se muestra los tipos de exportación que ofrece DB2 Webquery.

The screenshot shows the Reporting tool interface. The top section is titled 'Parameters' and contains three dropdown menus: 'Tipo' (set to 'Factura'), 'Año' (set to '2015'), and 'Concepto' (set to 'ANESTESIOLOGO'). Below these are buttons for 'Run', 'Reset', and 'Clear Output', along with a checked option 'Run in a new window'. The bottom section shows the OLAP data tables. The first table is a summary for the year 2015, and the second table is a drill-down view for the year 2015, showing quarterly data (Trimestre 1, 2, 3, 4) and a total row. The columns include Facturado, Glosado, PorcGlosa, Aceptado, and Soportado. A red arrow points from the 'Concepto' dropdown to the 'DESC\_GLOSA' dropdown in the OLAP interface.

Año	Facturado	Glosado	PorcGlosa	Aceptado	Soportado
2015	\$11.731.046.061,01	\$640.050.321,00	5,46%	\$82.223.413,00	\$498.904.577,00
TOTAL	\$11.731.046.061,01	\$640.050.321,00	5,46%	\$82.223.413,00	\$498.904.577,00

Trimestre	Facturado	Glosado	PorcGlosa	Aceptado	Soportado
1	\$2.337.622.645,00	\$150.338.435,00	6,43%	\$43.806.301,00	\$103.242.270,00
2	\$3.070.515.801,00	\$139.911.679,00	4,56%	\$11.350.558,00	\$121.856.895,00
3	\$3.698.832.755,01	\$320.969.073,00	8,68%	\$22.725.919,00	\$251.081.821,00
4	\$2.624.074.860,00	\$28.831.134,00	1,10%	\$4.340.635,00	\$22.723.591,00
TOTAL	\$11.731.046.061,01	\$640.050.321,00	5,46%	\$82.223.413,00	\$498.904.577,00

Figura 72. Generación de informe OLAP para exportar. Instantáneas de la solución de BI

Se agregan atributos como Régimen, Empresa, Contrato, Grupo de Procedimientos quirúrgicos, código CUPS y nombre del procedimiento, obteniendo un informe más detallado y sobre el cual se muestra la funcionalidad de exportar los resultados. Se Utiliza el botón **Save** para guardar en los diversos formatos: archivo de Excel, Excel con fórmulas, pdf y Active Report (informe offline)

En la siguiente imagen se observa la ruta de navegación para exportar el archivo. Los informes tipo **Active Report** permiten visualización offline a través de navegadores e

incluyen funciones adicionales para agrupar, diagramar, ordenar, filtrar, etc. (Bedoya et Al, 2014)

The screenshot displays the DB2 Web Query interface with several key components:

- Top Navigation:** Includes browser tabs for 'DB2 Web Query' and 'Olap Financiera', and a search bar.
- Filtering and Action Bar:** Shows filters for 'GRUPO\_PROC' (GRUPO 20 - 23), 'NOMC\_CUPS', and 'CUPS'. Action buttons include 'OLAP', 'Run', 'Reset', 'Save', and 'Help'.
- Main Data Table:** A table with columns for 'Trimestre', 'Regimen', 'Empresa', and 'GRUPO\_PROC'. It lists various medical procedures and their associated costs.
- Export Options:** A menu is open, showing options: 'Save the data in an Excel file', 'Save the data in an Excel 2000 file', 'Display as a PDF Report', and 'Display as Active Report (Offline Analysis)'. A red box highlights the 'Save' button and the 'Display as Active Report' option.
- Active Report View:** A detailed view of a report for 'CONTRATO' and 'CAPRECOM HOSPITALARIO'. It includes a table with columns for 'Trimestre', 'Regimen', 'Empresa', 'GRUPO\_PROC', 'CUPS', 'NOMC\_CUPS', 'Facturado', 'Glosado', and 'PorcGlosa'. It lists specific medical procedures like 'DRENAJE DE COLECCION EPIDURAL SUPRATENTORIO' and 'DRENAJE DE ESPACIO SUBDURAL CON REPARO D'. A pie chart is visible at the bottom right of this view.
- Export Menu:** A separate 'Export' menu is shown with options: 'HTML', 'CSV (comma delim)', and 'XML (Excel)'. A red arrow points from the 'Display as Active Report' option to this menu.
- Other Elements:** Includes a PDF icon, a Microsoft Excel icon, and a 'PDF' button.

Figura 73. Funcionalidad de exportación de Resultados en DB2 Web Query Instantáneas solución BI

### 3.6.4.7. RF-11. Clustering para obtener grupos de pacientes.

**Aplicar técnica *Clustering* de minería de datos, para agrupar pacientes que han sido sometidos a procedimientos quirúrgicos durante los años 2007-2015 y cuyos diagnósticos estaban asociados a cáncer.**

Este capítulo presenta el análisis y resultados de la información obtenida en el desarrollo de la investigación “solución de inteligencia de negocios basada en minería de datos para apoyar la toma de decisiones en el proceso de cirugía del hospital departamental universitario santa Sofía de caldas, en la ciudad de Manizales.

Atendiendo a requerimientos normativos y de interés gerencia por iniciar labores de identificación de grupos de pacientes en la Institución, se propone la aplicación de Clustering y específicamente el algoritmo K-Means para identificar las características de los pacientes a quienes se les practicaron procedimientos quirúrgicos y cuyos diagnósticos estaban asociados a tumores malignos.

Se parte de un juego de datos que incluyó 8 atributos demográficos de los pacientes atendidos en cirugía y 6 datos asociados al proceso quirúrgico.

Tabla 18.  
Descripción Set de Datos para técnica de Minería (Clustering)

FASE II. COMPRESIÓN DE LOS DATOS FORMATO DE DESCRIPCIÓN DE LOS DATOS PROCEDIMIENTOS QUIRÚRGICOS ASOCIADOS A TUMORES MALIGNOS						
Campo	Descripción	Tipo Dato	Valores Permitidos	Valor Mínimo	Valor Máximo	Relevancia Investigación
Grupo_CIE10	Grupo de clasificación de enfermedades de acuerdo al Código Internacional. Para la técnica de minería se usan grupos asociados a tumores malignos	C ( 3 )	{[C00-C99], [D00-D09]}	C00	D09	Muy relevante
DiagnosticoAgrupado	Descripción del Grupo de clasificación de enfermedades.	C ( 99 )	Texto. {[a-z], [ [, ] ]}			Poco relevante
Cod_Cirugia	Código Único de Procedimiento en Salud (CUPS). Código de Cirugía.	C ( 6 )	Texto. {[0-9], [M]}	011304	M16312	Relevante
Cirugia	Nombre o descripción del procedimiento quirúrgico	N ( 9.0 )	Texto. {[a-z][-,/,%, [.,]]}			Poco relevante

Grupo_Cirugia	Grupo quirúrgico al que pertenece la cirugía. Util para clasificar de acuerdo a la complejidad del procedimiento. Util para medir productividad por nivel de complejidad	C ( 13 )	[GRUPO 2 - 6,GRUPO 7 - 10,GRUPO 11 - 13,GRUPO 20 - 23]	Grupo 2-6	Grupo 20-23	Muy Relevante
TipoAtencion	Determina si la cirugía corresponde a una urgencia o una cirugía programada	C ( 10 )	[Programada, Urgente]	Programada	Urgente	Relevante
Edad	Edad del paciente al momento de la cirugía (En años)	N	Entero. [0-9]	13	89	Relevante
Etareo	Grupo etáreo de acuerdo a la edad del paciente (por quinquenios)	C ( 10 )	[De 10 a 14,De 15 - 19,De 20 - 24,De 25 - 29,De 30 - 34,De 35 - 39,De 40 - 44,De 45 - 49,De 50 - 54,De 55 - 59,60 y mas]	De 10-14	60 y Mas	Muy relevante
ZonaResidencia	Zona del departamento de Caldas, de donde proviene el paciente	C ( 20 )	[Magdalena Caldense, Alto Occidente, Alto Oriente, Bajo Occidente, Centro Sur, Norte]			Muy relevante
Regimen	Regimen de salud al que pertenecía el paciente al momento de la cirugía	C ( 12 )	[SUBSIDIADO,C ONTRIBUTIVO,V INCULADO,PAR TICULAR,OTRO]			Muy relevante
Genero	Género del paciente	C ( 1 )	[F, M]			Muy relevante
Estado_Civil	Estado civil del paciente	C ( 11 )	[Viudo,Soltero,Casado,Union Libre]			Poco relevante
Etnia	Raza o comunidad cultural a la que pertenece el paciente	C ( 8 )	[Indigena, Ninguno]			Muy relevante
Estrato	Nivel socio económico del paciente al momento de la atención.	C ( 9 )	[Estrato 0,Estrato 1,Estrato 2,Estrato 3,Estrato 4]	Estrato 0	Estrato 4	Muy relevante

Se define la ficha técnica para técnica de Clustering elegida, describiendo el objetivo especificando el alcance de su aplicación

Tabla 19.  
Ficha técnica de minería de datos Clustering.

FASE IV. MODELADO			
FORMATO DESCRIPCION DE DATOS PARA TECNICA DE MINERIA			
<b>Técnica de Minería</b>	<b>Clustering (Algoritmo K-Means)</b>		
<b>Número Registros</b>	<b>2268</b>	<b>Número Atributos</b>	<b>14</b>
<b>Fuentes de datos</b>	Sistema de Información Hosvital. Modelo de datos para solución de inteligencia de negocios para el servicio de cirugía del Hospital Departamental Universitario Santa Sofía de Caldas. Período de las cirugías entre los años 2007 y 2016.		
<b>Descripción breve del objetivo</b>	Identificar grupos de interés, de acuerdo a los tipos de usuarios a los cuales se les realizó procedimientos quirúrgicos cuyos diagnósticos CIE-10 se encuentran asociados a tumores malignos.		
<b>Atributos</b>	<b>Descripción del Atributo</b>	<b>Título</b>	
Grupo_CIE10	Grupo de clasificación de enfermedades de acuerdo al Código Internacional	Grupo_CIE10	
DiagnosticoAgrupado	Descripción del Grupo de clasificación de enfermedades	DiagnosticoAgrupado	
Cod_Cirugia	Código Único de Procedimiento en Salud (CUPS). Código de Cirugía	Cod_Cirugia	
Cirugia	Nombre o descripción del procedimiento quirúrgico	Cirugia	
Grupo_Cirugia	Grupo quirúrgico al que pertenece la cirugía. Util para clasificar de acuerdo a la complejidad del procedimiento.	Grupo_Cirugia	
TipoAtencion	Determina si la cirugía corresponde a una urgencia o una cirugía programada	TipoAtencion	
Edad	Edad del paciente al momento de la cirugía (En años)	Edad	
Etareo	Grupo etéreo de acuerdo a la edad del paciente (por quinquenios)	Etareo	
ZonaResidencia	Zona del departamento de Caldas, de donde proviene el paciente	ZonaResidencia	
Regimen	Régimen de salud al que pertenecía el paciente al momento de la cirugía	Regimen	
Genero	Género del paciente	Genero	
Estado_Civil	Estado civil del paciente	Estado_Civil	
Etnia	Raza o comunidad cultural a la que pertenece el paciente	Etnia	
Estrato	Nivel socio económico del paciente al momento de la atención.	Estrato	
<b>Observaciones / Criterios según técnica de Minería</b>			
<p>Algoritmo de agrupación K-Means.</p> <p>Con la técnica de Clustering, se busca encontrar patrones entre los grupos de usuarios que fueron sometidos a procedimientos quirúrgicos durante los años 2007 a 2015 y cuyos diagnósticos estaban asociados a tumores malignos (cáncer). Esto, debido a que en el Hospital han venido apareciendo diagnósticos relacionados con cáncer entre las primeras 10 (diez) causas de atención y el hospital no cuenta con servicio de Oncología. Explorar los grupos de pacientes que han sido atendidos en los últimos 9 años, permitiría contar con información al momento de definir si es conveniente ampliar el portafolio de servicios y además, responder a requerimientos normativos de acreditación (identificar grupos de interés) y de la Política de Atención integral en Salud (Gestión de Riesgo Individual)</p>			

Se procede a reducir la dimensionalidad (Hernández et Al, 2004) puesto que se detectan atributos asociados a datos dependientes como los fueron (Cod\_cirugía, Cirugía y DiagnosticoAgrupado), conservando 11 atributos para aplicar la técnica de minería Clustering.

Grupo_CIE10	Grupo_Cirugia	TipoAtencion	Etareo	ZonaResidencia	Regimen	Genero	Estado_Civil	Etnia	Estrato
C44	GRUPO 2 - 6	Programada	60 y mas	Magdalena Caldense	SUBSIDIADO	F	Viudo	Ninguna	Estrato 0
C44	GRUPO 2 - 6	Programada	60 y mas	Magdalena Caldense	SUBSIDIADO	F	Viudo	Ninguna	Estrato 0
C23	GRUPO 7 - 10	Urgencia	60 y mas	Magdalena Caldense	CONTRIBUTIVO	F	Soltero	Ninguna	Estrato 1
C23	GRUPO 7 - 10	Urgencia	60 y mas	Magdalena Caldense	CONTRIBUTIVO	F	Soltero	Ninguna	Estrato 1
C23	GRUPO 11 - 13	Urgencia	60 y mas	Magdalena Caldense	CONTRIBUTIVO	F	Soltero	Ninguna	Estrato 1
C23	GRUPO 11 - 13	Urgencia	60 y mas	Magdalena Caldense	CONTRIBUTIVO	F	Soltero	Ninguna	Estrato 1
C61	GRUPO 11 - 13	Programada	60 y mas	Magdalena Caldense	SUBSIDIADO	M	Casado	Ninguna	Estrato 0
C61	GRUPO 11 - 13	Programada	60 y mas	Magdalena Caldense	SUBSIDIADO	M	Casado	Ninguna	Estrato 0
C32	GRUPO 7 - 10	Urgencia	De 50 - 54	Magdalena Caldense	VINCULADO	M	Casado	Ninguna	Estrato 1
C32	GRUPO 11 - 13	Urgencia	De 50 - 54	Magdalena Caldense	VINCULADO	M	Casado	Ninguna	Estrato 1
C32	GRUPO 11 - 13	Urgencia	De 50 - 54	Magdalena Caldense	VINCULADO	M	Casado	Ninguna	Estrato 1
C32	GRUPO 11 - 13	Urgencia	De 50 - 54	Magdalena Caldense	VINCULADO	M	Casado	Ninguna	Estrato 1
C38	GRUPO 11 - 13	Urgencia	De 50 - 54	Magdalena Caldense	SUBSIDIADO	M	Soltero	Ninguna	Estrato 1
C38	GRUPO 11 - 13	Urgencia	De 50 - 54	Magdalena Caldense	SUBSIDIADO	M	Soltero	Ninguna	Estrato 1
C38	GRUPO 11 - 13	Urgencia	De 50 - 54	Magdalena Caldense	SUBSIDIADO	M	Soltero	Ninguna	Estrato 1

Figura 74. Vista Minable para aplicación de Clustering.

Se procesa el juego de datos para ser explorado a través del software de minería de datos WEKA, versión 3.8.0 incluyendo 2268 registros y 11 atributos.

```

DM_SSOPIA.arff: Bloc de notas
Archivo Edición Formato Ver Ayuda
@relation DM_SSOPIA
@attribute 'Edad' {'81', '77', '78', '54', '53', '52', '58', '57', '55', '35', '59', '22', '65', '51', '47', '48', '67'}
@attribute 'Grupo_CIE10' {'C00', 'C01', 'C02', 'C03', 'C04', 'C07', 'C09', 'C13', 'C14', 'C15', 'C16', 'C17', 'C18'}
@attribute 'Grupo_Cirugia' {'GRUPO 11-13', 'GRUPO 7-10', 'GRUPO 2-6', 'GRUPO 20-23'}
@attribute 'TipoAtencion' {'Programada', 'Urgencia'}
@attribute 'Etareo' {'60 y mas', '45 a 49', '55 a 59', '50 a 54', '20 a 24', '40 a 44', '35 a 39', '30 a 34'}
@attribute 'ZonaResidencia' {'AltoOccidente', 'CentroSur', 'AltoOriente', 'BajoOccidente', 'MagdalenaCaldense'}
@attribute 'Regimen' {'SUBSIDIADO', 'PARTICULAR', 'CONTRIBUTIVO', 'VINCULADO', 'OTRO'}
@attribute 'Genero' {'F', 'M'}
@attribute 'Estado_Civil' {'Casado', 'Soltero', 'Viudo', 'Unionlibre'}
@attribute 'Etnia' {'Ninguna', 'Indigena'}
@attribute 'Estrato' {'Estrato2', 'Estrato3', 'Estrato1', 'Estrato0', 'Estrato4'}

@data
'81' 'C44' 'GRUPO 2-6' 'Programada' '60 y mas' 'MagdalenaCaldense' 'SUBSIDIADO'
'81' 'C44' 'GRUPO 2-6' 'Programada' '60 y mas' 'MagdalenaCaldense' 'SUBSIDIADO'
'77' 'C23' 'GRUPO 7-10' 'Urgencia' '60 y mas' 'MagdalenaCaldense' 'CONTRIBUTIVO'
'77' 'C23' 'GRUPO 7-10' 'Urgencia' '60 y mas' 'MagdalenaCaldense' 'CONTRIBUTIVO'
'77' 'C23' 'GRUPO 11-13' 'Urgencia' '60 y mas' 'MagdalenaCaldense' 'CONTRIBUTIVO'
'77' 'C23' 'GRUPO 11-13' 'Urgencia' '60 y mas' 'MagdalenaCaldense' 'CONTRIBUTIVO'
'78' 'C61' 'GRUPO 11-13' 'Programada' '60 y mas' 'MagdalenaCaldense' 'SUBSIDIADO'
'78' 'C61' 'GRUPO 11-13' 'Programada' '60 y mas' 'MagdalenaCaldense' 'SUBSIDIADO'
'54' 'C32' 'GRUPO 7-10' 'Urgencia' '50 a 54' 'MagdalenaCaldense' 'VINCULADO'

```

Figura 75. Archivo Arff para minería en WEKA

## ANALISIS E INTERPRETACION DE LOS RESULTADOS A PARTIR DEL SET DE DATOS N= 2268.

Comenzamos describiendo la edad de los pacientes en una escala discreta en años cumplidos.

Tabla 20.  
Exploración de datos atributo Edad.

		Edad			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	13	4	,2	,2	,2
	14	14	,6	,6	,8
	16	2	,1	,1	,9
	18	3	,1	,1	1,0
	19	14	,6	,6	1,6
	21	6	,3	,3	1,9
	22	13	,6	,6	2,5
	23	13	,6	,6	3,0
	24	10	,4	,4	3,5
	25	1	,0	,0	3,5
	26	19	,8	,8	4,4
	27	6	,3	,3	4,6
	28	11	,5	,5	5,1
	29	4	,2	,2	5,3
	30	10	,4	,4	5,7
	31	15	,7	,7	6,4
	34	20	,9	,9	7,3
	35	27	1,2	1,2	8,5
	36	7	,3	,3	8,8
	37	10	,4	,4	9,2
	38	3	,1	,1	9,3
	39	17	,7	,7	10,1
	40	4	,2	,2	10,3
	41	7	,3	,3	10,6
	42	57	2,5	2,5	13,1
	43	6	,3	,3	13,4
	44	43	1,9	1,9	15,3
	45	51	2,2	2,2	17,5
	46	35	1,5	1,5	19,0
	47	59	2,6	2,6	21,6
48	57	2,5	2,5	24,2	
49	20	,9	,9	25,0	
50	27	1,2	1,2	26,2	
51	53	2,3	2,3	28,6	
52	56	2,5	2,5	31,0	
53	56	2,5	2,5	33,5	
54	39	1,7	1,7	35,2	
55	60	2,6	2,6	37,9	
56	41	1,8	1,8	39,7	

57	44	1,9	1,9	41,6
58	35	1,5	1,5	43,2
59	50	2,2	2,2	45,4
60	53	2,3	2,3	47,7
61	58	2,6	2,6	50,3
62	85	3,7	3,7	54,0
63	60	2,6	2,6	56,7
64	75	3,3	3,3	60,0
65	61	2,7	2,7	62,7
66	65	2,9	2,9	65,5
67	80	3,5	3,5	69,0
68	47	2,1	2,1	71,1
69	54	2,4	2,4	73,5
70	61	2,7	2,7	76,2
71	71	3,1	3,1	79,3
72	17	,7	,7	80,1
73	23	1,0	1,0	81,1
74	33	1,5	1,5	82,5
75	59	2,6	2,6	85,1
76	73	3,2	3,2	88,4
77	69	3,0	3,0	91,4
78	67	3,0	3,0	94,4
79	7	,3	,3	94,7
81	39	1,7	1,7	96,4
82	28	1,2	1,2	97,6
84	15	,7	,7	98,3
85	21	,9	,9	99,2
86	9	,4	,4	99,6
87	6	,3	,3	99,9
89	3	,1	,1	100,0
Total	2268	100,0	100,0	

La edad mínima fue de 13 años y la máxima de 89, la edad más observada en estos procedimientos fue de 62 años que equivale al 3,7% del total de la población en ese término de tiempo referenciado para este proyecto.

Tabla 21.  
Exploración de datos atributo Grupo\_CIE10

Grupo_CIE10					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	C00	9	,4	,4	,4
	C01	1	,0	,0	,4
	C13	4	,2	,2	,6
	C14	6	,3	,3	,9
	C15	74	3,3	3,3	4,1
	C16	396	17,5	17,5	21,6
	C17	4	,2	,2	21,8
	C18	204	9,0	9,0	30,8
	C19	54	2,4	2,4	33,2

C02	14	,6	,6	33,8
C20	74	3,3	3,3	37,0
C21	7	,3	,3	37,3
C23	10	,4	,4	37,8
C24	40	1,8	1,8	39,6
C25	124	5,5	5,5	45,0
C26	15	,7	,7	45,7
C03	4	,2	,2	45,9
C31	16	,7	,7	46,6
C32	81	3,6	3,6	50,1
C33	2	,1	,1	50,2
C34	95	4,2	4,2	54,4
C38	52	2,3	2,3	56,7
C04	1	,0	,0	56,7
C41	84	3,7	3,7	60,4
C43	17	,7	,7	61,2
C44	21	,9	,9	62,1
C45	1	,0	,0	62,2
C47	4	,2	,2	62,3
C48	52	2,3	2,3	64,6
C49	3	,1	,1	64,8
C50	17	,7	,7	65,5
C53	32	1,4	1,4	66,9
C54	9	,4	,4	67,3
56	47	2,1	2,1	69,4
C57	1	,0	,0	69,4
C60	7	,3	,3	69,8
C61	25	1,1	1,1	70,9
C62	9	,4	,4	71,3
C64	18	,8	,8	72,0
C67	11	,5	,5	72,5
C07	24	1,1	1,1	73,6
C71	117	5,2	5,2	78,7
C72	14	,6	,6	79,4
C73	78	3,4	3,4	82,8
C75	6	,3	,3	83,1
C76	127	5,6	5,6	88,7
C77	52	2,3	2,3	91,0
C78	64	2,8	2,8	93,8
C79	30	1,3	1,3	95,1
C82	5	,2	,2	95,3
C83	18	,8	,8	96,1
C84	13	,6	,6	96,7
C85	12	,5	,5	97,2
D01	9	,4	,4	97,6
D02	5	,2	,2	97,8
D04	23	1,0	1,0	98,9
D05	6	,3	,3	99,1
C09	5	,2	,2	99,3
D06	6	,3	,3	99,6
D09	9	,4	,4	100,0
Total	2268	100,0	100,0	

Con relación al grupo de clasificación de enfermedades de acuerdo al Código Internacional, las más representativas fueron; C16, que equivale al 17,5% del total de la información, en seguida, C18, al 9%, C76, al 5,6%, C25, al 5,5% y C71, al 5,2% respectivamente.

Tabla 22.  
Exploración de datos atributo Grupo\_Cirugía.

Grupo Cirugía					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Grupo 2-6	186	8,2	8,2	8,2
	Grupo 7-10	1044	46,0	46,0	54,2
	Grupo 11-13	581	25,6	25,6	79,9
	Grupo 20-23	457	20,1	20,1	100,0
	Total	2268	100,0	100,0	



Figura 76. Exploración atenciones por grupos de cirugía.

Con relación Grupo quirúrgico al que pertenece la cirugía. Útil para clasificar de acuerdo a la complejidad del procedimiento. Se observó con un 46% al grupo al representativo Grupo 7-10.

Tabla 23.  
Exploración de datos atributo Tipo\_Atención.

Tipo Atención					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Programada	1643	72,4	72,4	72,4
	Urgencia	625	27,6	27,6	100,0
	Total	2268	100,0	100,0	



Figura 77. Exploración registros por tipo de atención.

El tipo de atención determina si la cirugía corresponde a una urgencia o una cirugía programada, la cita más utilizada es la programada con un 72,4%.

Tabla 24.  
Exploración de datos atributo Grupo Etareo.

Etareo					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	10 a 14	18	,8	,8	,8
	55 a 59	230	10,1	10,1	10,9
	60 y mas	1239	54,6	54,6	65,6
	15 a 19	19	,8	,8	66,4
	20 a 24	42	1,9	1,9	68,3
	25 a 29	41	1,8	1,8	70,1
	30 a 34	45	2,0	2,0	72,0
	35 a 39	64	2,8	2,8	74,9
	40 a 44	117	5,2	5,2	80,0
	45 a 49	222	9,8	9,8	89,8
	50 a 54	231	10,2	10,2	100,0
	Total	2268	100,0	100,0	

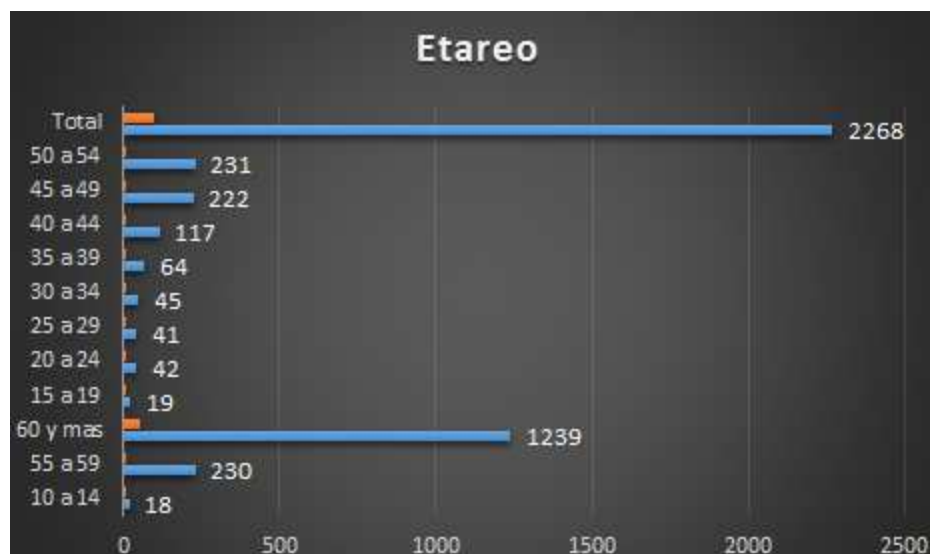


Figura 78. Exploración atenciones por grupos etareos.

El Grupo etáreo de acuerdo a la edad del paciente (por quinquenios). El grupo más representativo más representativo en edad fue el de “60 y más” con un 54,6% del total de la Información.

Tabla 25.  
Exploración de datos atributo Zona\_residencia.

		Zona Residencia			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Alto occidente	166	7,3	7,3	7,3
	Alto oriente	114	5,0	5,0	12,3
	Bajo occidente	105	4,6	4,6	17,0
	Centro sur	1625	71,6	71,6	88,6
	Magdalena caldense	141	6,2	6,2	94,8
	Norte	117	5,2	5,2	100,0
	Total	2268	100,0	100,0	

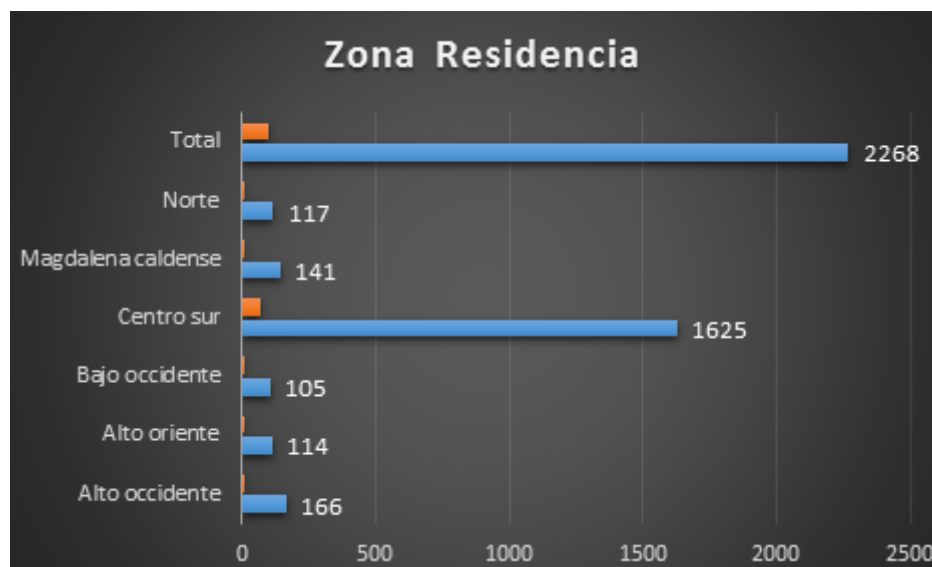


Figura 79. Exploración atenciones por Zona de Residencia.

Zona del departamento de Caldas, de donde proviene el paciente, la zona más predominante fue la de centro sur con un 71,6%.

Tabla 26.  
Exploración de datos atributo Régimen.

		Régimen			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Contributivo	684	30,2	30,2	30,2
	Otro	82	3,6	3,6	33,8
	Particular	117	5,2	5,2	38,9
	Subsidiado	1205	53,1	53,1	92,1
	Vinculado	180	7,9	7,9	100,0
	Total	2268	100,0	100,0	

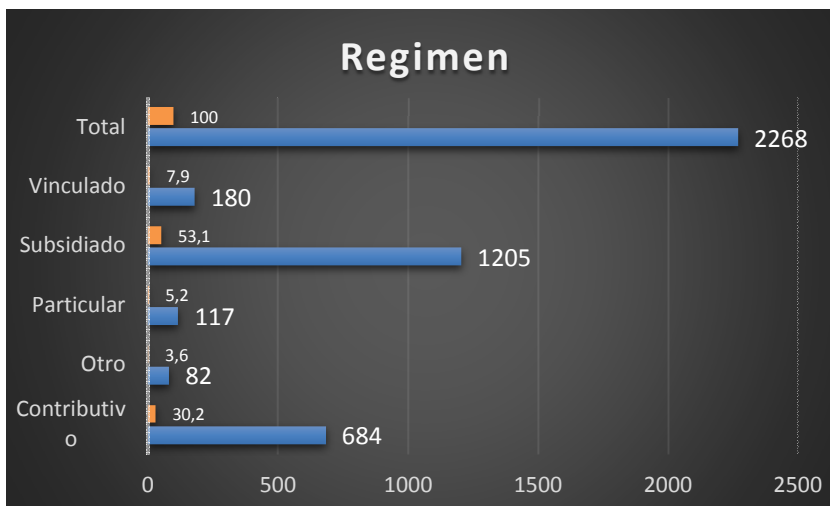


Figura 80. Exploración atenciones por Régimen.

En cuanto al régimen de salud al que pertenecía el paciente al momento de la cirugía, el más observado fue el de subsidiado con un 53,1%, seguido del contributivo con 30,2% del total de examinados.

Tabla 27.  
Exploración de datos atributo Género.

		Genero			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	F	1151	50,7	50,7	50,7
	M	1117	49,3	49,3	100,0
	Total	2268	100,0	100,0	

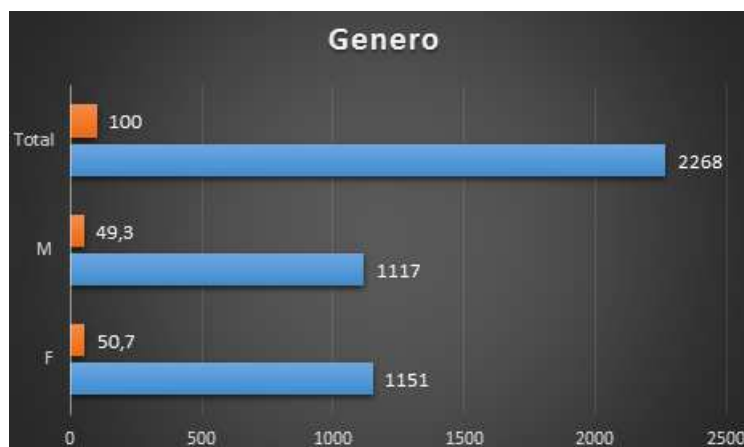


Figura 81. Exploración atenciones por Género.

Género del paciente, se podría afirmar que no existe una diferencia significativa, pero los hombres tienen un 50,7 %, frente a un 48,3% de las mujeres.

Tabla 28.  
Exploración de datos atributo Estado\_Civil.

Estado Civil					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Casado	1706	75,2	75,2	75,2
	Soltero	434	19,1	19,1	94,4
	Unionlibre	33	1,5	1,5	95,8
	Viudo	95	4,2	4,2	100,0
	Total	2268	100,0	100,0	

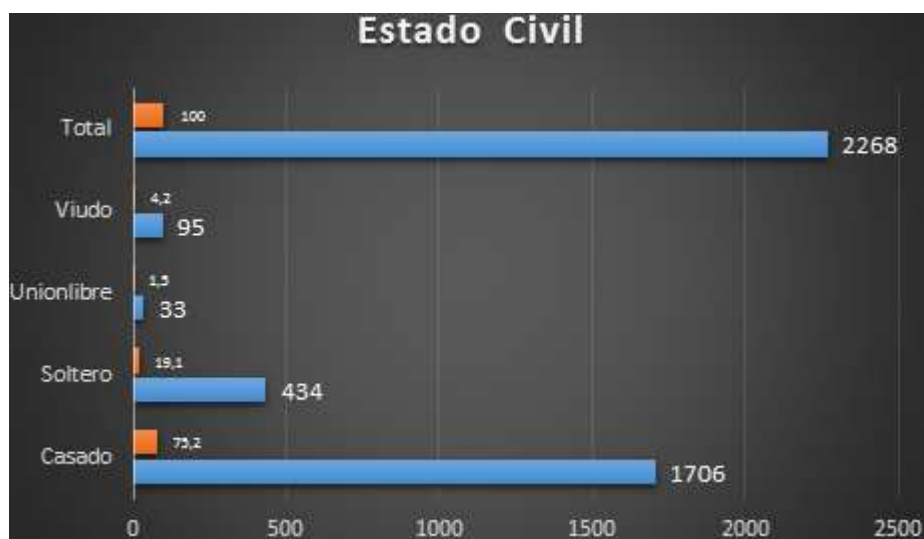


Figura 82. Exploración atenciones por Estado Civil.

Con relación al estado civil del paciente, el más atendido casado con un 75,2%

Tabla 29.  
Exploración de datos atributo Etnia.

Etnia					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Indígena	50	2,2	2,2	2,2
	Ninguna	2218	97,8	97,8	100,0
	Total	2268	100,0	100,0	

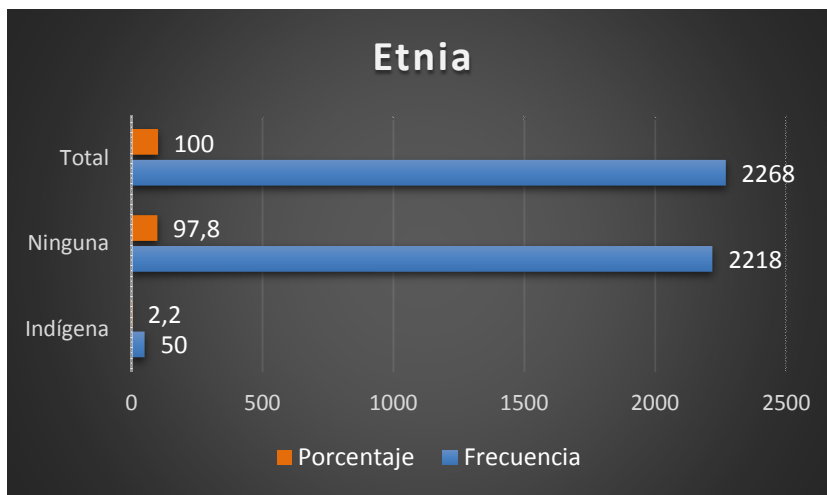


Figura 83. Exploración atenciones por Etnias.  
Elaboración Propia

Aunque en la zona del Alto Occidente del departamento de Caldas existen asentamientos indígenas, sólo el 2.2% de los observados pertenecen a dicha etnia, diferencia de un 97,8% de atención que registra ninguna etnia.

Tabla 30.  
Exploración de datos atributo Estrato.

		Estrato			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Estrato0	607	26,8	26,8	26,8
	Estrato1	912	40,2	40,2	67,0
	Estrato2	378	16,7	16,7	83,6
	Estrato3	298	13,1	13,1	96,8
	Estrato4	73	3,2	3,2	100,0
Total		2268	100,0	100,0	

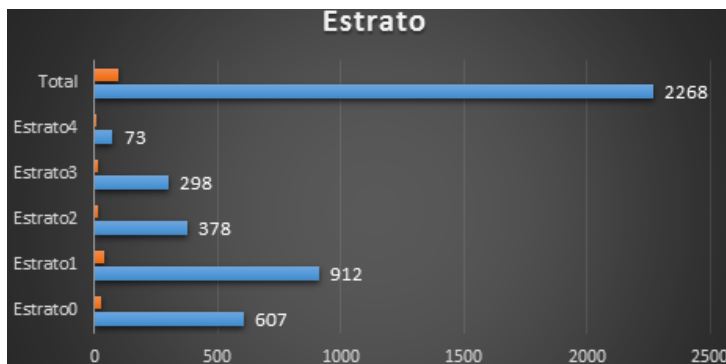


Figura 84. Exploración atenciones por Estratos.

Para la investigación, Nivel socio económico del paciente al momento de la atención es importante; el estrato más atendido fue el estrato 1 con 40,2%, seguido estrato 0 con un 26,8%.

## ANÁLISIS DE CLUSTER

La técnica conocida como clustering divide la base de datos en diferentes grupos. La meta del clustering es encontrar grupos que son diferentes entre sí y cuales miembros son similares el uno del otro. A diferencia de la clasificación, no se sabe dónde habrá clúster o con que atributos de los datos se harán los clústeres. Consecuentemente, alguien con gran conocimiento del negocio debe interpretar los clúster. Con frecuencia si es necesario modificar el clustering excluyendo variables que han sido usadas para agrupar, porque mediante este examen el usuario identifica que datos son irrelevantes o sin significado. Después de haber encontrado clúster que segmenten razonablemente la base de datos, estos pueden ser usados para clasificar nuevos datos.

No hay que confundir clustering con segmentación. La segmentación se usa para identificar grupos que tienen características comunes. El clustering es un modo de segmentar datos en grupos que no están previamente definidos, mientras que la clasificación es un modo de segmentar datos asignándolos a grupos que están previamente definidos.

The screenshot shows the Weka Explorer interface with the SimpleKMeans clustering algorithm selected. The 'Clusterer output' window displays the following data:

Regimen	Subsidiado	Subsidiado	Contributivo
	(2268.0)	(1520.0)	(748.0)
Edad	62	62	67
Grupo_CIE10	C16	C16	C16
Grupo_Cirugia	GRUPO 7-10	GRUPO 7-10	GRUPO 7-10
TipoAtencion	Programada	Programada	Programada
Etareo	60 y mas	60 y mas	60 y mas
ZonaResidencia	CentroSur	CentroSur	CentroSur
Regimen	SUBSIDIADO	SUBSIDIADO	CONTRIBUTIVO
Genero	F	F	F
Estado_Civil	Casado	Casado	Casado
Etnia	Ninguna	Ninguna	Ninguna
Estrato	Estrato1	Estrato0	Estrato1

Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	1520 ( 67%)
1	748 ( 33%)

Figura 85. Centroides y Clústeres generados por K-Means en WEKA.

Utilizando el Weka, y aplicando el método de simple k-means, se obtuvieron los anteriores resultados.

A simple vista se observa que los dos centroides son prácticamente idénticos, luego el análisis clúster no jerárquico viene a confirmar que la solución que obtuvimos con anterioridad es fiable. De todas formas, un buen indicador de la calidad de los clúster obtenidos es si las medias de las variables en cada clúster son significativamente distintas. En su momento indicamos que el objetivo es que cada conglomerado sea cuanto más distinto mejor de los demás.

El programa efectúa un análisis de varianza que, en síntesis, viene a contrastar la hipótesis nula de que las medias de cada variable en los dos conglomerados son las mismas. Esta hipótesis debe poder rechazarse, por lo menos, para la mayoría de variables.

Se obtuvieron dos clústeres, el primero con un 67% de la población, en el cual se agrupan alrededor de la enfermedad bajo el código C16, grupo de cirugía 7-10, tipo de atención programada, régimen subsidiado, hombres casados mayores de 60 años, sin etnia y de estrato 0.

El segundo clúster con un 33 % de la población, de más de 60 años, de sexo femenino, se agrupan alrededor de la enfermedad bajo el código C16, grupo de cirugía 7-10, tipo de atención programada, régimen contributivo, sin etnia y de estrato 1.

En suma, lo más destacado en los clúster fueron los siguientes atributos:

Tabla 31.  
Resumen Exploración de datos. Atributos relevantes.

Variable	Valor Relevante
Edad	62
Grupo_CIE10	C16
Grupo Cirugía	GRUPO 7-10
Tipo Atención	Programada
Etareo	60 y mas
Zona Residencia	Centro sur
Régimen	Subsidiado
Genero	F
Estado Civil	Casado
Etnia	Ninguna
Estrato	1

Al obtener los Clústeres anteriores, se pudo validar la información estadística sobre la patología de cáncer de estómago<sup>21</sup> (C16) la cual de acuerdo a la información epidemiológica, tiene presencia importante en la zona de influencia.

Se procedió entonces a excluir el grupo CIE-10 (C16), analizando 1872 registros y obteniendo dos nuevos clústeres, que ratifican la prevalencia de tumores malignos en personas mayores de 60 años, pero con intervenciones clasificadas en grupos quirúrgicos del 11 al 13. Estos nuevos clústeres hacen énfasis en tumores del colon (C18) y presentan un clúster con el 56% y el otro con el 44% de la población.

Attribute	Full Data (1872.0)	Cluster#	
		0 (1045.0)	1 (827.0)
Edad	62	62	66
Grupo_CIE10	C18	C18	C18
Grupo_Cirugia	GRUPO 7-10	GRUPO 11-13	GRUPO 7-10
TipoAtencion	Programada	Programada	Programada
Etareo	60 y mas	60 y mas	60 y mas
ZonaResidencia	CentroSur	CentroSur	CentroSur
Regimen	SUBSIDIADO	CONTRIBUTIVO	SUBSIDIADO
Genero	F	F	M
Estado_Civil	Casado	Casado	Casado
Etnia	Ninguna	Ninguna	Ninguna
Estrato	Estrato1	Estrato1	Estrato0

Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	1045 ( 56%)
1	827 ( 44%)

Figura 86. Resultados Clustering, set de datos 2  
Instantánea tomada de WEKA

Notando entonces que los grupos de CIE-10 detectados ya se encuentran identificados por los clústeres anteriores, se procede a excluir toda la población del grupo etario de más de 60 años, buscando agrupar personas de otros rangos de edad y el tipo de tumores con mayor prevalencia.

<sup>21</sup> Sistema de información para la Calidad. Hospital.

En este caso se obtienen 861 registros y se dos clústeres, el primero con un 62% de la población, con diagnóstico de tumor maligno del encéfalo (C71), grupo de edad entre 45 y 49 años del régimen contributivo y cuyas cirugías pertenecieron al grupo 7 al 10, de sexo femenino, sin etnia y de estrato 1. Para el segundo clúster se asignó el 38% de la población con presencia de tumores del colon (C18), grupo de cirugías entre el 7 y el 10, rango de edad entre 55 y 59 años, en su mayoría del sexo masculino y del estrato 0.

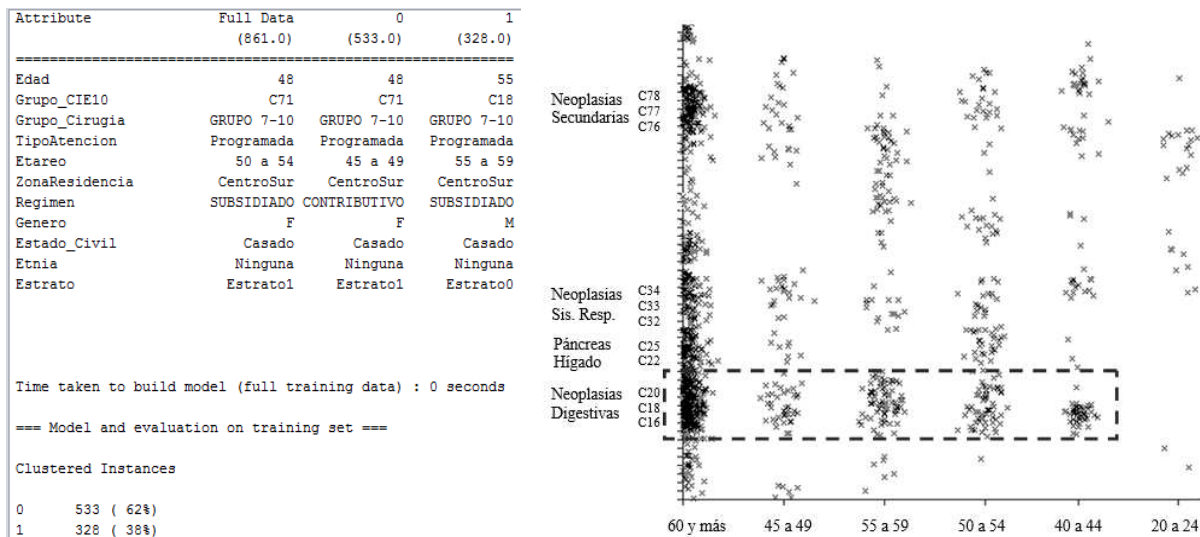


Figura 87. Resultados Clustering, set de datos 3 Instantánea tomada de WEKA

## CONCLUSIONES CLUSTERING

La edad mínima fue de 13 años y la máxima de 89, la edad más observada en estos procedimientos fue de 62 años que equivale al 3,7% del total de la población en ese término de tiempo referenciado para este proyecto.

Con relación al grupo de clasificación de enfermedades de acuerdo al Código Internacional, las más representativas fueron; C16, que equivale al 17,5% del total de la información, en seguida, C18, al 9%, C76, al 5,6%, C25, al 5,5% y C71, al 5,2% respectivamente.

Tabla 32.  
Resumen Grupos de diagnóstico relevantes identificados.

RESULTADOS TECNICA DE MINERIA CLUSTERING RESUMEN DE GRUPOS DE DIAGNÓSTICO CIE-10		
Grupo CIE-10	Descripción diagnóstico	Porcentaje
C16	TUMOR MALIGNO DEL ESTOMAGO	17.5%
C18	TUMOR MALIGNO DEL COLON	9%
C76	TUMOR MALIGNO DE OTROS SITIOS Y DE SITIOS MAL DEFINIDOS	5.6%
C25	TUMOR MALIGNO DEL PANCREAS	5.5%
C71	TUMOR MALIGNO DEL ENCEFALO	5.2%

Con relación Grupo quirúrgico al que pertenece la cirugía. Útil para clasificar de acuerdo a la complejidad del procedimiento. Se observó con un 46% al grupo al representativo Grupo 7-10.

El tipo de atención determina si la cirugía corresponde a una urgencia o una cirugía programada, la cita más utilizada es la programada con un 72,4%.

El Grupo etario de acuerdo a la edad del paciente (por quinquenios). El grupo más representativo más representativo en edad fue el de 60 y más con un 54,6 % del total de la Información.

Zona del departamento de Caldas, de donde proviene el paciente, la zona más predominante fue la de centro sur con un 71,6%.

En cuanto al régimen de salud al que pertenecía el paciente al momento de la cirugía, el más observado fue el de subsidiado con un 53,1%, seguido del contributivo con 30,2% del total de examinados.

Género del paciente, se podría afirmar que no existe una diferencia significativa, pero los hombres tienen un 50,7 %, frente a un 48,3% de las mujeres.

Con relación al estado civil del paciente, el más atendido casado con un 75,2% y según los registros explorados el 97,8% de las atenciones no pertenecen a ninguna etnia, mientras que el 2,2% de los observados es de etnia indígena.

Para la investigación, Nivel socio económico del paciente al momento de la atención es importante; el estrato más atendido fue el estrato 1 con 40,2%, seguido estrato 0 con un 26,8%.

La anterior información ratifica elementos conceptuales que se han venido tratando al interior de la Institución frente al incremento de atenciones relacionadas con cáncer. En particular el Cáncer de Estómago tiene un alto impacto en la región y el *Clustering* anterior lo ratifica, así como la prevalencia de dicho diagnóstico para pacientes mayores de 60 años.

### 3.6.5. Fase V. Evaluación

#### 3.6.5.1. Objetivo

Evaluar la solución de inteligencia de negocios y minería de datos, para determinar si son útiles a las necesidades del proceso de Atención Integral en el Servicio de Quirófanos del Hospital Departamental Universitario Santa Sofía de Caldas E.S.E.

El proceso de evaluación y validación por parte de líderes de la institución, incluyó perfiles como:

- Director Científico
- Líder del Proceso de Garantía de Calidad
- Líder de Servicio de Quirófanos
- Jefe de Talento Humano
- Jefe Financiera
- Líder Procesos de Facturación y Cartera.

Luego de la respectiva socialización de los resultados obtenidos, se utilizó el **Formato Evaluación de Satisfacción en Requerimientos**, en el cual se calificaron en una escala de 0 a 5 cada uno de los once (11) requerimientos funcionales propuestos en la Fase I “Comprensión del Negocio”, al final se obtuvo una calificación de cinco (5) para cada uno de ellos, alcanzando un 100% de satisfacción con la solución presentada.

Por otro lado se aplicó el **Formato Evaluación de Resultados Obtenidos**, con el objetivo de validar otros aspectos relacionados con: facilidad de uso, cumplimiento de expectativas, beneficio potencial de la solución, rendimiento en la ejecución, consistencia, confiabilidad y pertinencia de implementación de soluciones de BI en otros procesos.

Para esto se realizaron sesiones de trabajo personalizadas con cada uno de los interesados en el proyecto, documentando la apreciación en cuanto a preguntas como:

- ¿El proyecto cumple con expectativas del Proceso de Cirugía?
- ¿El proyecto ha cumplido los requerimientos planteados?
- ¿Se observa un beneficio potencial de la solución desarrollada?
- ¿Los gráficos, tablas y tableros utilizados, son apropiados para mostrar la información?
- ¿El rendimiento (tiempo) de ejecución del sistema fue el adecuado?
- ¿La transferencia de conocimiento al equipo de trabajo fue correcta?
- ¿La información obtenida de la solución de (BI) es consistente?
- ¿La solución de (BI) ofrece herramientas fáciles de usar y entendibles?
- ¿Considera que los datos generados por la solución de (BI) sirven para tomar decisiones?
- ¿Recomendaría implementar solución de (BI) para otros procesos institucionales?

En este caso, se obtuvieron cinco (5) evaluaciones con un 100% de satisfacción y una (1) con un 97%, para un promedio de 99.5% de satisfacción con los resultados obtenidos.


Las siguientes imágenes muestran algunos de los formatos aplicados y en el **Anexo 5. Documentación Adicional Fase V. Evaluación**, se pueden apreciar los formatos mencionados con sus respectivas firmas.

E.S.E HOSPITAL DEPARTAMENTAL UNIVERSITARIO SANTA SOFIA DE CALDAS UNIVERSIDAD AUTÓNOMA DE MANIZALES SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS BASADA EN MINERÍA DE DATOS MAESTRIA EN GESTIÓN Y DESARROLLO DE PROYECTOS DE SOFTWARE		Santa Sofia			
METODOLOGIA CRISP-DM FASE V. EVALUACIÓN					
FORMATO EVALUACIÓN DE SATISFACCIÓN EN REQUERIMIENTOS					
Nro	Requerimiento	Calificación	Ponderación	Total	Evaluador
RF-1	La solución de inteligencia de Negocios permitirá generar informes dinámicos mediante herramienta OLAP, para analizar datos asistenciales.	5!	10%	0,5	[Firma]
RF-2	La solución de inteligencia de Negocios permitirá generar informes dinámicos mediante herramienta OLAP, para analizar datos financieros.	5	9%	0,45	[Firma]
RF-3	La solución de inteligencia de negocios incluirá un tablero para el monitoreo de tiempos quirúrgicos, por periodo, especialidad, especialista y procedimiento quirúrgico.	5	9%	0,45	[Firma]
RF-4	La solución de inteligencia de negocios incluirá tablero para el seguimiento a indicadores de proceso de cirugía.	5!	9%	0,45	[Firma]
RF-5	La solución de inteligencia de negocios incluirá tablero para el seguimiento a valores facturados, glosados y soportados, por periodos de tiempo.	5	9%	0,45	[Firma]
RF-6	La solución de inteligencia de negocios incluirá reporte para efectuar seguimiento a re- intervenciones quirúrgicas.	5	9%	0,45	[Firma]
RF-7	La solución de inteligencia de negocios permitirá visualizar los indicadores de producción reportados históricamente en el Decreto 2193 de 2004.	5	9%	0,45	[Firma]
RF-8	La solución permitirá el uso de aplicativo móvil para la visualización de informes	5	9%	0,45	[Firma]
RF-9	La solución permitirá que los usuarios puedan personalizar los informes y exportarlos en diferentes formatos (xls y pdf), así como generar reportes activos que puedan ser usados en modo off line]	5	9%	0,45	[Firma]
RF-10	Se podrán personalizar y guardar informes que permitan un acceso rápido de acuerdo a las necesidades de los líderes.	5	9%	0,45	[Firma]
RF-11	Aplicar técnica Clustering de minería de datos, para agrupar pacientes que han sido sometidos a procedimientos quirúrgicos durante los años 2007-2015 y cuyos diagnósticos estaban asociados a cáncer.	5	9%	0,45	[Firma]
		Total		5,0	
		Máximo Puntaje		5	
		Satisfacción de Requerimientos		100%	


  

$\sum_{i=1}^n (Ponderación_i) = 100\% = 1$ $\sum_{i=1}^n (Ponderación_i + Calificación_i) = Total$ <p>Donde:                      n= cantidad de requerimientos.                      i= requerimiento específico evaluado                      Máximo Puntaje = 5                      Satisfacción requerimientos = (Total/Máximo Puntaje) *100%</p>		<table border="1"> <tr> <th colspan="2">Valores Válidos</th> </tr> <tr> <th colspan="2">Calificación</th> </tr> <tr> <td>0</td> <td>Nada</td> </tr> <tr> <td>1</td> <td>Muy Poco</td> </tr> <tr> <td>2</td> <td>Poco</td> </tr> <tr> <td>3</td> <td>Satisfecho</td> </tr> <tr> <td>4</td> <td>Muy Satisfecho</td> </tr> <tr> <td>5</td> <td>Totalmente Satisfecho</td> </tr> </table>	Valores Válidos		Calificación		0	Nada	1	Muy Poco	2	Poco	3	Satisfecho	4	Muy Satisfecho	5	Totalmente Satisfecho
Valores Válidos																		
Calificación																		
0	Nada																	
1	Muy Poco																	
2	Poco																	
3	Satisfecho																	
4	Muy Satisfecho																	
5	Totalmente Satisfecho																	


Figura 88. Evaluación de Satisfacción de Requerimientos – Calificado  
Elaboración propia



E.S.E HOSPITAL DEPARTAMENTAL UNIVERSITARIO SANTA SOFIA DE CALDAS  
UNIVERSIDAD AUTÓNOMA DE MANIZALES  
SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS BASADA EN MINERÍA DE DATOS  
MAESTRIA EN GESTIÓN Y DESARROLLO DE PROYECTOS DE SOFTWARE



E.S.E HOSPITAL DEPARTAMENTAL UNIVERSITARIO SANTA SOFIA DE CALDAS  
UNIVERSIDAD AUTÓNOMA DE MANIZALES  
SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS BASADA EN MINERÍA DE DATOS  
MAESTRIA EN GESTIÓN Y DESARROLLO DE PROYECTOS DE SOFTWARE



¿El proyecto cumple con expectativas del Proceso de Cirugía?  
¿El proyecto ha cumplido los requerimientos planteados?  
¿Se observa un beneficio potencial de la solución desarrollada?  
¿Los gráficos, tablas y tableros utilizados, son apropiados para mostrar la información?  
¿El rendimiento (tiempo) de ejecución del sistema fue el adecuado?  
¿La transferencia de conocimiento al equipo de trabajo fue correcta?  
¿La información obtenida de la solución de (BI) es consistente?  
¿La solución de (BI) ofrece herramientas fáciles de usar y entendibles?  
¿Considera que los datos generados por la solución de (BI) sirven para tomar decisiones?  
¿Recomendaría implementar solución de (BI) para otros procesos institucionales?

METODOLOGIA CRISP-DM FASE V. EVALUACIÓN			
FORMATO EVALUACIÓN DE RESULTADOS OBTENIDOS			
Criterio Evaluado	Calificación	Ponderación	Total
¿El proyecto cumple con expectativas del Proceso de Cirugía?	5	10%	0,5
¿El proyecto ha cumplido los requerimientos planteados?	5	10%	0,5
¿Se observa un beneficio potencial de la solución desarrollada?	5	10%	0,5
¿Los gráficos, tablas y tableros utilizados, son apropiados para mostrar la información?	5	10%	0,5
¿El rendimiento (tiempo) de ejecución del sistema fue el adecuado?	5	10%	0,5
¿La transferencia de conocimiento al equipo de trabajo fue correcta?	5	10%	0,5
¿La información obtenida de la solución de (BI) es consistente?	5	10%	0,5
¿La solución de (BI) ofrece herramientas fáciles de usar y entendibles?	5	10%	0,5
¿Considera que los datos generados por la solución de (BI) sirven para tomar decisiones?	5	10%	0,5
¿Recomendaría implementar solución de (BI) para otros procesos institucionales?	5	10%	0,5
<b>Total</b>			<b>5</b>
<b>Máximo Puntaje</b>			<b>5</b>
<b>Satisfacción de Requerimientos</b>			<b>100%</b>

Math

Jefe I  
Lider S

Dr. Gonzalo Calle Hoyos  
Director Científico

Donde:

- n = cantidad de requerimientos
- Ponderación = Máximo Puntaje = 5
- Satisfacción requerimientos

$$\sum_{i=1}^n (Ponderación_i) = 100\% = 1$$

Donde:

- n = cantidad de requerimientos
- Ponderación = Máximo Puntaje = 5
- Satisfacción requerimientos

$$\sum_{i=1}^n (Ponderación_i \cdot Calificación_i) = Total$$

Donde:

- n = cantidad de requerimientos
- Ponderación = Máximo Puntaje = 5
- Satisfacción requerimientos

Valores Válidos
Calificación
0 Nada (En desacuerdo)
1 Muy Poco (Muy poco de acuerdo)
2 Poco (Poco de acuerdo)
3 Satisfecho (de acuerdo)
4 Muy Satisfecho (Muy de acuerdo)
5 Totalmente (Satisfecho - de acuerdo)

Figura 89. Evaluación de Resultados obtenidos – Calificado  
Elaboración propia

### 3.6.5.2. Evaluación de los resultados

Se concluye de las evaluaciones obtenidas, que los interesados estuvieron **“Totalmente de acuerdo”** con los requerimientos cubiertos y **“Totalmente satisfechos”** al momento de evaluar los resultados obtenidos.

Cada uno de los líderes e integrantes de la Institución que conocieron la solución, coincidieron en establecer que existe un gran potencial para el aprovechamiento de inteligencia de negocios en otros servicios y procesos. Se resalta el *“impacto que podría tener un sistema de información de este tipo para impactar de manera positiva los estándares en el proceso de acreditación, toda vez que son muy relevantes la toma de decisiones y los escenarios para el mejoramiento continuo”*<sup>22</sup>.

Los gráficos, tablas y tipos de reporte utilizados en la solución, fueron los suministrados por la herramienta Web Query. Permitieron de manera concreta visualizar los datos del proceso, de acuerdo a la calificación satisfactoria obtenida.

Los tiempos de ejecución de los reportes, se encuentra supeditados a diversos factores como los son: cantidad de campos a recuperar, cantidad de registros a consultar, horarios de ejecución de los informes. Sin embargo, durante las pruebas realizadas y al momento de la evaluación de los resultados, se obtuvo una calificación satisfactoria de este aspecto.

Se llevó a cabo transferencia de conocimiento a los líderes del proceso y a los integrantes del proceso de Administración de TICS, con el objetivo de que continúen brindando apoyo en el despliegue de la solución. De igual forma se elaboró un Manual de Usuario de Solución de Inteligencia de Negocios, bajo el código Institucional **TI080-R9-M02 Manual de Solución de Inteligencia de Negocios**, el cual se concibe como herramienta de despliegue, inducción y re-inducción de los colaboradores.

En las pruebas realizadas se encontró consistencia en los datos reportados, obedeciendo siempre a los filtros y parámetros que fueron suministrados.

Los interesados calificaron de manera satisfactoria la facilidad de uso de los reportes y ratifican la utilidad que puede brindar la solución para la toma de decisiones en todos los niveles de la institución.

---

<sup>22</sup> Coordinadora de Talento Humano – Equipo Estratégico para la Acreditación Institucional.

### 3.6.5.3. Proceso de revisión

Esta sección permite documentar el ciclo de mejoramiento continuo de la metodología aplicada, permitiendo expresar la utilidad de cada una de las fases llevadas a cabo, así como registrar aquellos aspectos que generaron dificultades o afectaron el desarrollo del proyecto.

#### Fase I. Comprensión del Negocio

Esta fase agrega valor al desarrollo de la solución de inteligencia de negocios, puesto que permitió conocer a fondo la interdependencia del proceso de **Atención Integral en el Servicio de Quirófanos**, sus objetivos, procedimientos e indicadores. Gracias a esta fase, se pudo identificar los participantes en el proyecto, los requerimientos que debería suplirse y la estructura de la solución.

Fue importante analizar otra información o requerimientos de ley (Sistema Único de Acreditación, Sistema de Información para la Calidad, Requerimientos de reporte de información al Ministerio de Salud y Protección Social) para ofrecer resultados acordes con las necesidades del Hospital.

#### Fase II. Comprensión de los datos

Fue crucial para el proyecto, haber realizado una revisión a fondo de los orígenes de datos para la integración del modelo. Debido a la cantidad de tablas que conforman la base de datos relacional del ERP HOSVITAL utilizado en el Hospital, fue necesario contar con la participación activa de los profesionales en soporte y el experto en el sistema de información.

Gracias a la documentación obtenida se logró realizar una revisión de cada atributo que podría integrarse en las dimensiones. Aunque se obtuvo apoyo de profesionales en salud para la interpretación de algunos atributos, se ratificó la necesidad de trabajar siempre con el acompañamiento de expertos en atención en salud. La terminología y la relevancia de los datos sólo pueden ser definidas por el personal asistencial, o administrativo con alto nivel de experiencia en presentación de informes de Alta Dirección.

La estructura de los datos que se conoció en esta fase, fue de gran relevancia para dar continuidad al proyecto.

#### Fase III. Preparación de los datos

Esta fase fue clave para el proceso de desarrollo de la solución de inteligencia de negocios. Integrar los datos provenientes de los orígenes de datos, definir la estructura de

las tablas de hechos y dimensionales, establecer la relación correcta entre dichas tablas, especificar los atributos de cada una de ellas y conocer su contenido, permitieron obtener al final un modelo de datos coherente con los requerimientos y objetivos de minería.

Aunque la herramienta DB2 Web Query ofrece herramientas para la documentación de proyectos de inteligencia de negocios, fue necesario recurrir a otras aplicaciones para la diagramación de los modelos.

El modelo dimensional podrá ser mejorado a partir del aprendizaje luego del despliegue, el mantenimiento de la solución, deberá implicar el análisis de nuevos requerimientos y objetivos de minería de datos.

En la preparación de datos se encontraron atributos vacíos, nulos o con datos estándar del tipo (1900-00-00). Sin embargo, dada la naturaleza de los datos clínicos y las limitantes documentadas en la sección 1.6.2 Confidencialidad de los Datos, no fue posible realizar transformación a dichos datos. Esto no fue una barrera de calidad, pues en la atención hospitalaria se conocen perfectamente las prohibiciones de ley para modificar datos de la historia clínica.

Para la integración de datos, los usuarios usaban denominaciones distintas para referirse a los mismos conceptos, lo cual dificultó el análisis

El modelo se construyó de manera correcta, incorporando los atributos necesarios para dar respuesta a los requerimientos.

#### **Fase IV. Modelado**

En esta fase se materializa todo el conocimiento obtenido de la Institución en las fases anteriores y se aprovechan los datos integrados. El Recurso Humano asociado al proyecto fue crucial para el logro de los objetivos.

Gracias a las numerosas sesiones de trabajo articulado, se pudo presentar al final la documentación sobre la ejecución del proyecto. La importancia de los resultados obtenidos debe ser analizada desde dos aspectos: el primero, relacionado con la respuesta dada a los requerimientos y cuya calificación por parte de los interesados mostró un alto nivel de satisfacción. El segundo, la documentación que ofrece al Hospital Departamental Universitario Santa Sofía de Caldas la posibilidad de implementar soluciones similares en otros procesos.

Esta fase contribuyó plenamente a la consecución de los resultados finales, puesto que aunque la Fase III fue útil para preparar y construir el modelo de datos, es en la fase de modelado donde se establece la “visualización” conforme a las necesidades de los usuarios.

La manera de simplificar esta fase, es incrementar las sesiones de trabajo con el usuario final, para evitar reprocesos en estructuración de los cubos dimensionales y tableros de monitoreo.

Aunque los profesionales en salud disponen de poco tiempo y sus agendas están muy comprometidas, se concluye que con una mayor participación en todas las fases se pueden lograr resultados en menor tiempo, y, por supuesto de alta calidad.

La base de datos clínica y financiera del hospital, incluye centenares de tablas con datos valiosos de todos los procesos de atención. En el mantenimiento de la solución se podrán incorporar datos adicionales, incluso es factible la aparición de nuevas dimensiones o tablas de hechos.

### **Fase V. Evaluación**

Los interesados en la solución tuvieron la oportunidad no sólo de participar en la construcción de la solución, sino también la validación de los resultados obtenidos y la calificación del nivel de satisfacción con la solución.

Esta fase permitió cuantificar el nivel de aceptación y la percepción por parte de líderes de diferentes procesos.

Luego de revisar las apreciaciones de los evaluadores, se resalta la recomendación que ratifican de implementar soluciones de este tipo en otros procesos institucionales.

Esta fase puede ser mejorada, si se repite meses después del despliegue, puesto que los usuarios podrán aportar desde la experiencia que hayan tenido durante cierto período de tiempo.

#### **3.6.5.4. Determinación de los pasos siguientes**

Luego de haber modelado la solución de inteligencia basada en minería de datos y su correspondiente evaluación por parte de los líderes interesados, se recomienda proceder a su despliegue y uso.

### 3.6.6. Fase VI. Despliegue

#### 3.6.6.1. Objetivo

Explotar la utilidad de los modelos, integrándolos en las tareas de toma de decisiones del proceso de Cirugía del Hospital Departamental Universitario Santa Sofía de Caldas

#### Planificación del despliegue

Tabla 33.  
Planificación del despliegue de la Solución de BI.

<b>Acción</b>	<b>Descripción</b>	<b>Responsable</b>
Capacitar a los líderes en el uso de la herramienta desarrollada	Capacitación a los líderes en el uso de la herramienta de BI para la visualización y personalización de reportes.	Autor de Tesis
Capacitar al equipo de TICS en la administración de la solución de BI.	Capacitar a los integrantes del equipo de TICS en el uso de la solución de BI, para que estén en capacidad de ofrecer soporte técnico a los líderes.	Autor de Tesis
Soporte y acompañamiento a los usuarios finales en el uso de la solución desarrollada	El proceso de Administración de Tics de la Institución, brinda soporte a los líderes en cuanto al uso y adaptación de nuevos informes en la plataforma de inteligencia de negocios desarrollada.	Administración de TICS
Esquema de Actualización de los datos	Cada día se registran datos en el sistema de información transaccional del Hospital. Por ello es necesario definir en el administrador de esquemas de bases de datos DB2 System iNavigator de IBM, la agenda de actualización de las tablas dimensionales y de hechos.	Experto Asesor en Hosvital
Mantenimiento de la solución	Identificación e implementación de requerimientos adicionales que surjan durante el aprovechamiento de la solución.	Administración de TICS

**Ver Anexo 6 Documentación Adicional Fase VI. Despliegue** para otros formatos asociados a esta sección.

### 3.7. Pruebas

Se llevaron a cabo las pruebas planeadas, integrantes del proceso de Administración de Tecnologías de la Información y las Comunicaciones (TICS) efectuaron pruebas unitarias, comparando los resultados obtenidos en la solución de inteligencia de negocios, con los resultados de scripts de SQL ejecutados sobre los orígenes de datos. En cada “Formato de Caso de Prueba” se consignaron los resultados obtenidos, los parámetros y filtros de cada una de las pruebas.

Se incluyeron también aspectos asociados a la facilidad de uso, interfaz, contenido y aceptación. Se integraron a los casos de prueba líderes interesados en el proyecto.

**Ver Anexo 7. Documentación Adicional Pruebas** para evidencias adicionales

#### 3.7.1. Plan de Pruebas

Tabla 34.  
Planeación y Control de Casos de Prueba

PLANEACIÓN Y CONTROL CASOS DE PRUEBA						
Nro. Caso	Nombre del Caso	Responsable	Realizado		Aprobado	
			SI	NO	SI	NO
1	Prueba Informe Decreto 2193	Ingeniera Soporte	X		X	
2	Prueba OLAP Cirugía	Ingeniera Soporte	X		X	
3	Prueba Informe Monitoreo de Recursos	Ingeniera Soporte	X		X	
4	Prueba OLAP Financiero	Ingeniera Soporte	X		X	
5	Prueba Tablero de Contratación	Ingeniera Soporte	X		X	
6	Prueba Informe Indicadores	Autor Tesis	X		X	
7	Prueba Tablero de Facturación	Autor Tesis	X		X	
8	Prueba Informe 2193	Talento Humano	X		X	
9	Prueba Tablero de Facturación	Jefe Facturación	X		X	
10	Prueba OLAP Financiero	Jefe Facturación	X		X	
11	Prueba OLAP Cirugía	Jefe De Cirugía	X		X	
12	Prueba Informes Dispositivos Móviles	Experto Hosvital	X		X	

A continuación imágenes correspondientes a las pruebas realizadas.

FORMATO DE CASO DE PRUEBA																
Fecha	<table border="1"> <tr> <td>Día</td> <td>Mes</td> <td>Año</td> </tr> <tr> <td>23</td> <td>5</td> <td>2016</td> </tr> </table>	Día	Mes	Año	23	5	2016	Duración	<table border="1"> <tr> <td>HH</td> <td>MM</td> </tr> <tr> <td>8</td> <td>00</td> </tr> </table>	HH	MM	8	00	Responsable	LUZ ADRIANA SANCHEZ	
Día	Mes	Año														
23	5	2016														
HH	MM															
8	00															
Nombre Caso	PRUEBA OLAP CIRUGIA. SOLUCION INTELIGENCIA DE NEGOCIOS			Nro. Caso	2											
Objetivo	Verificar mediante SQL la exactitud de los datos visualizados por la consulta, utilizando la función TRACE de SQL y la funcionalidad de guardar informes personalizados y exportación de resultados en otros formatos															
Lugar	Oficina de TICS	Técnica	Caja Blanca	<input checked="" type="checkbox"/>	Caja Negra	<input type="checkbox"/>										
Pre - requisitos																
Se ejecuta el OLAP Cirugía con los parámetros básicos																
DETALLE CASO DE PRUEBA																
Tipo de Prueba	Contenido	<input checked="" type="checkbox"/>	Interfaz	<input type="checkbox"/>	Facilidad Uso	<input type="checkbox"/>										
	Unitaria	<input checked="" type="checkbox"/>	Integración	<input type="checkbox"/>	Sistema	<input type="checkbox"/>										
					Aceptación	<input checked="" type="checkbox"/>										
Descripción Hardware Requerido			Descripción Software Requerido													
PLATAFORMA BLADE CENTER H. EQUIPO CLIENTE			IBMI, iNavigator (DB2)													
Descripción del Procedimiento a seguir																
Se ejecuta el informe OLAP Cirugía para garantizar que muestre los resultados básicos, se procede a personalizar el informe utilizando los atributos de las dimensiones (tipo de agenda, código de cirugía, grupo de procedimiento, código de procedimiento, nombre procedimiento, diagnósticos de entrada, salida y complicación, hora de inicio y fin, valores de facturación, glosa, soportado y aceptado).																
Resultados Esperados																
Los datos arrojados por ambos informes, coincida, logrando un alto nivel de confiabilidad. Así mismo que el informe pueda ser guardado (personalizado) y exportado en pdf, excel y Active report																
Resultados Obtenidos																
Luego de ejecutar el código SQL se puede evidenciar que coinciden los valores arrojados por la solución de inteligencia de negocios y los arrojados por la ejecución del script SQL. Así mismo se verifica la funcionalidad de exportar a diversos formatos, guardar un informe personalizado y utilizar la opción de informe activo (active report), el cual se deja abrir en un navegador de manera offline																
Facilidad de Uso		Contenido		Interfaz												
Difícil de entender	<input type="checkbox"/>	Confuso	<input type="checkbox"/>	Ilustraciones	<input type="checkbox"/>											
Complejo	<input type="checkbox"/>	Un poco ambiguo	<input type="checkbox"/>	Sin Uniformidad	<input type="checkbox"/>											
Simple	<input type="checkbox"/>	Informativa	<input type="checkbox"/>	Generalmente Uniforme	<input type="checkbox"/>											
Efectiva	<input checked="" type="checkbox"/>	Clara	<input checked="" type="checkbox"/>	Uniforme	<input checked="" type="checkbox"/>											
Fácil de aprender	<input type="checkbox"/>															
Prueba superada?	Observaciones															
SI	<input checked="" type="checkbox"/>															
NO	<input type="checkbox"/>															
Se utilizaron parámetros básicos de ejecución del informe, asociadas a cirugías en estado "facturadas y Realizadas" durante los años 2014 y 2015, para la especialidad "Cirugía Cardiovascular"																
Realizado Por	<u>LUZ ADRIANA SANCHEZ CAMO</u>			Revisado Por	<u>WILSON ALEJANDRO ROJAS</u>											
Cargo	INGENIERA DE SOPORTE HOSVITAL			Cargo	AUTOR TESIS											

Figura 90. Caso de Prueba 2. Requerimiento Guardar Informes Personalizados en Web Query  
Elaboración propia

Se muestran también imágenes sobre el código SQL utilizado para la ejecución de la prueba anterior. DB2 permite efectuar el seguimiento a la ejecución de órdenes a través de la funcionalidad Visual Explain.

CODIGO SQL UTILIZADO

```

SELECT T1."TIPO_
T8."MPNOMC", T
T1."DX_COMPLIC
T1."COD_CIRUGIA
SUM(T2."VLR_AC

FROM ((( (BI_SS
T1."IDENTIFICACI
T1."FECHA_CX" )
BI_SSOFIA/DIM_F

WHERE (YEAR(T3
2014) AND (T1."E
(T4."MES_ABREV

GROUP BY T1."TI
T8."MPNOMC", T
T1."DX_COMPLIC
T7."GRUPO_PROCC
T1."TIPO_HERIDA
FETCH ONLY;

```

**Información sobre el plan.**

Condición ordenación	SI
Desplazable	SI
Nombre de plan	Scan Sorted Temporary
Tipo de paso de planificación	Logis
Código de razón	Índice para agrupamiento u ord...
Nombre de paso de planificación	Node_18976
Texto de sentencia	SELECT Concat(Rep(Cast(PROCPR_1 PR OCPRCOD AS Char(1)) OCPRC 254), Trailing , Rep(PROCPR_3 CROCOD, Trailing ' '), CaseWhen DESCRIMED_5HORMICR<='19 0 0.00' Then 'Nocturno' Else Case When DESCRIMED_5HORMICR<='08 0 0.00' Then 'Nocturno' Else CASE(DIUM) AS VRLNRM) (L34) 254) End End, MAESTP_3MENOME, D_PROFESIO_DIM_P000R_19 N OM_MEDICO FROM CREATE SORTED LIST QdsSortePS(Node_18920)SELE CT Case When DESCRIMED_5HORMICR<='19 0 0.00' Then 'Nocturno' Else Case When DESCRIMED_5HORMICR<='08 0 0.00' Then 'Nocturno' Else

Figura 91. Script SQL utilizado y su Visual Explain - Caso de Prueba 2  
Elaboración pruebas

FORMATO DE CASO DE PRUEBA										
Fecha	Día	Mes	Año	Duración	HH	MM	Responsable			
	27	5	2016		8	00	LUISA FERNANDA RAMIREZ			
Nombre Caso	PRUEBA OLAP FINANCIERO. SOLUCION DE INTELIGENCIA DE NEGOCIOS DEL PROCESO DE CIRUGIA						Nro. Caso	4		
Objetivo	Ejecutar informe generado a través de OLAP Financiero y verificar sus resultados a través de scripts SQL sobre el origen de los datos, garantizando la coherencia de la información									
Lugar	Oficina de TICS	Técnica	Caja Blanca	<input checked="" type="checkbox"/>	Caja Negra	<input type="checkbox"/>				
Pre - requisitos										
DETALLE CASO DE PRUEBA										
Tipo de Prueba	Contenido	<input checked="" type="checkbox"/>	Interfaz	<input type="checkbox"/>	Facilidad Uso	<input type="checkbox"/>				
	Unitaria	<input checked="" type="checkbox"/>	Integración	<input type="checkbox"/>	Sistema	<input type="checkbox"/>	Aceptación	<input checked="" type="checkbox"/>		
Descripción Hardware Requerido					Descripción Software Requerido					
PLATAFORMA BLADE CENTER H. EQUIPO CLIENTE					IBMi, iNavigator (DB2)					
Descripción del Procedimiento a seguir										
Se ejecuta un informe personalizado a través del cubo OLAP financiero incluido en la solución de BI para cirugía. Se construye script SQL cuyas variables y parámetros sean los mismos del cubo y se ejecutan en el origen de datos con el objetivo de comparar los resultados. Las variables a mostrar son: tipo_agenda, cod_cirugia, cod_proc, nomc_cups, identificacion, MPhomc, mpsexo, tip_anestesia, tipo_herida, dx_entrada, dx_salida, dx_complicacion, sala, hora_ini, hora_fin, valor_unitario, vlr_glosa, vlr_Aceptado, vlr_soportado.										
Resultados Esperados										
Que los datos generados tanto en el cubo OLAP como en los scripts de SQL sean iguales.										
Resultados Obtenidos										
Luego de ejecutar los Scripts de SQL y compararlos con los datos visualizados en el cubo OLAP, se aprecia coherencia entre ellos.										
Facilidad de Uso			Contenido			Interfaz				
Difícil de entender <input type="checkbox"/>			Confusa <input type="checkbox"/>			Inconsistente <input type="checkbox"/>				
Complicada <input type="checkbox"/>			Un poco ambigua <input type="checkbox"/>			Sin Uniformidad <input type="checkbox"/>				
Simple <input type="checkbox"/>			Informativa <input checked="" type="checkbox"/>			Generalmente Uniforme <input checked="" type="checkbox"/>				
Efectiva <input checked="" type="checkbox"/>			Clara <input type="checkbox"/>			Uniforme <input type="checkbox"/>				
Fácil de aprender <input type="checkbox"/>										
Prueba superada?			Observaciones							
SI <input checked="" type="checkbox"/>			Parámetros utilizados para ambos informes: año (2014), trimestre (3), mes (Julio), día (24), cirugías (facturadas y realizadas), tipo de cirugía (cirugía cardiovascular), fin de semana (NO).							
NO <input type="checkbox"/>										
Realizado Por	<u>LUISA FERNANDA RAMIREZ PATIÑO</u>				Revisado Por	<u>FABIO LÓPEZ FORERO</u>				
Cargo	INGENIERA DE SOPORTE HOSVITAL				Cargo	EXPERTO HOSVITAL				

Figura 92. Caso de prueba 4. Funcionamiento OLAP Financiero. Elaboración propia.

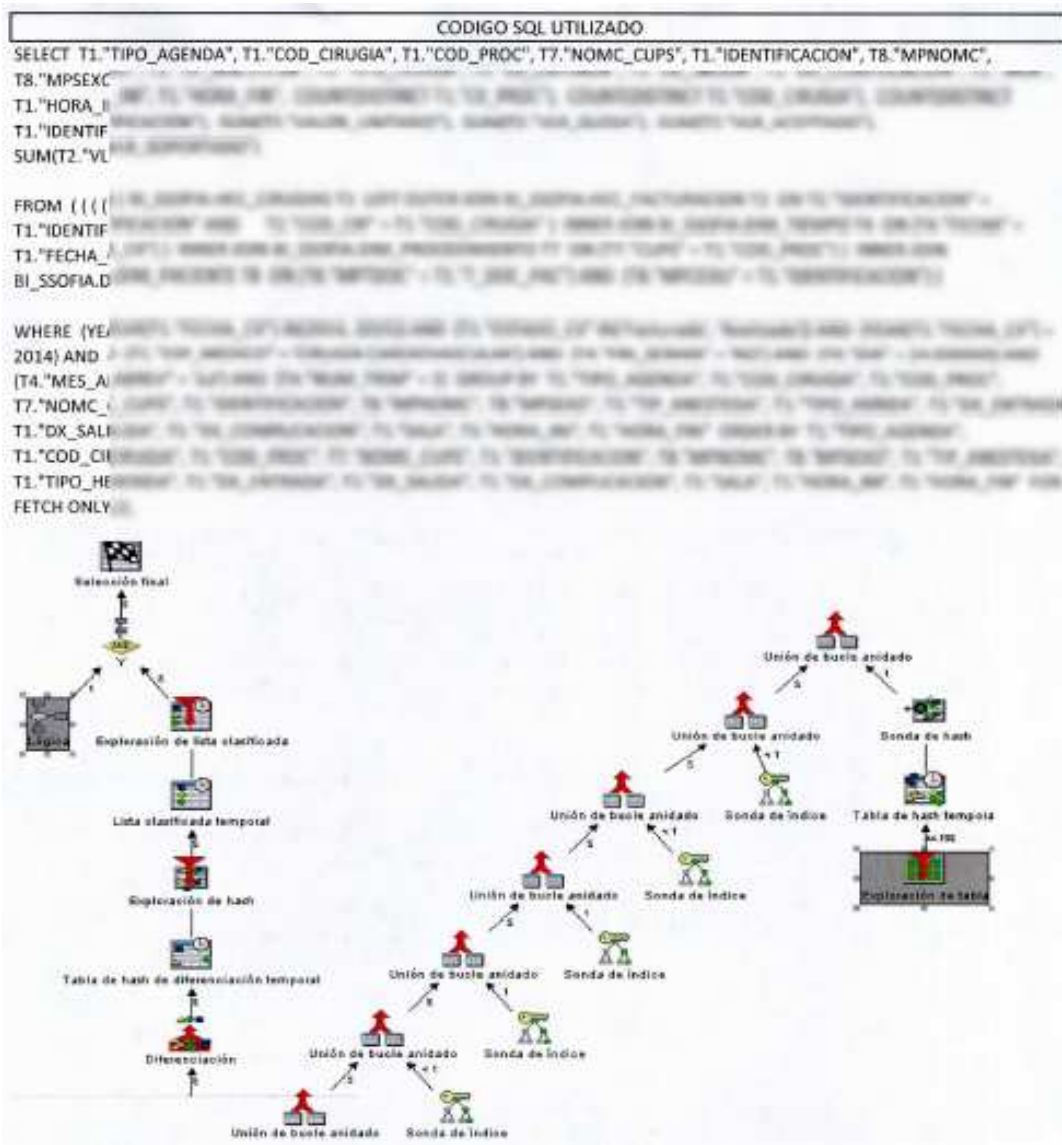


Figura 93. Script SQL utilizado y su Visual Explain - Caso de Prueba 4  
Elaboración propia.

### 3.8. Análisis de resultados

Tabla 35.  
Resultados esperados vs Resultados obtenidos en el proyecto de investigación

Descripción y Beneficios	Indicador	Resultado obtenido
<p>Estructura coherente de datos, después de su respectivo análisis, especificación, transformación y almacenamiento.</p> <p>Elemento fundamental para el despliegue seguro, eficiente y confiable de una solución de inteligencia de negocios, así como la fuente de datos de calidad para aplicación de técnicas de minería de datos <b>clustering</b></p>	<p>Porcentaje de modelo de datos especificado e implementado</p> <p>(Tablas de hechos y dimensiones implementadas/Tablas de hechos y dimensiones especificadas en el modelo) x 100</p> <p>Meta: 100%</p>	<p>Doce (12) tablas implementadas en el modelo dimensional, de 12 tablas que fueron especificadas en el modelo de datos</p> <p>Resultado 100%</p>
<p>Solución de inteligencia que incluye las siguientes funcionalidades:</p> <ul style="list-style-type: none"> <li>• Módulo de reportes estáticos y activos</li> <li>• Módulo de Procesamiento analítico en línea (OLAP) como interfaz de usuario final que emplea una técnica llamada análisis multidimensional para presentar informes gráficos y permitir el análisis flexible e interactivo de datos. (Wang et al, 2012).</li> <li>• Dashboards para la visualización y monitoreo de los indicadores definidos por el proceso</li> <li>• Despliegue de informes a través de dispositivos móviles.</li> <li>• Resultados de aplicación de <b>clustering</b> como técnica de agrupamiento en minería de datos.</li> </ul>	<p>Porcentaje de implementación de funcionalidades de inteligencia de negocios propuestos</p> <p>(N° funcionalidades implementadas en la solución de inteligencia de negocios / N° funcionalidades propuestas (5)) x 100</p> <p>Meta: 100%</p>	<p>11 requerimientos funcionales definidos e implementados, en los cuales se usan todas las funcionalidades mencionadas</p> <p>Resultado: 100%</p>
<p>Evaluar el porcentaje de aceptación sobre los requerimientos propuestos por el proyecto.</p>	<p>Porcentaje obtenido de la evaluación por pares académicos, medido desde dos aspectos:</p> <p>a. Porcentaje de Satisfacción con requerimientos. Meta &gt;=80%</p> <p>b. Porcentaje de Evaluación de Resultados Meta &gt;=80%</p>	<p>a. Porcentaje de Satisfacción con requerimientos. Resultado = 100%</p> <p>b. Porcentaje de Evaluación de Resultados Resultado = 99.5%</p>

## 4. CONCLUSIONES

Los proyectos de BI en el sector de la salud son diferentes a los del sector industrial, esto está determinado por la magnitud, complejidad y subjetividad de los datos clínicos, así como por el tipo de información y reportes solicitados. Para que un proyecto de BI sea exitoso en un Hospital, es necesario integrar personal especializado en varias áreas que le brinden un enfoque holístico y combinen el conocimiento con los recursos tecnológicos.

Los bajos niveles de calidad y completitud de los registros clínicos puede generar dificultades en la integración y transformación de datos, así como en interpretación de resultados y su aprovechamiento para generar conocimiento.

La fase Comprensión del Negocio incluida en la metodología CRISP-DM, permitió que la solución se integrara con el Direccionamiento Estratégico y se alineara con el enfoque a riesgos y resultados que requería el proceso de Cirugía.

La fase Comprensión de los Datos propuesta por la metodología CRISP-DM fue crucial, puesto que permitió realizar una revisión a fondo de los orígenes de datos para la integración del modelo.

De acuerdo con las evaluaciones de requerimientos y de resultados obtenidas, este proyecto ofrece una hoja de ruta, un marco de trabajo, un procedimiento práctico para el desarrollo de proyectos de minería de datos e inteligencia de negocios en el Hospital Departamental Universitario Santa Sofía de Caldas, permitiendo el uso de información y recursos tecnológicos en la toma de decisiones. Se espera que la metodología utilizada, así como la documentación de sus fases, sirva para la implementación en otros procesos.

Los resultados obtenidos con la ejecución del proyecto, realizan aportes importantes en cuanto a la implementación de proyectos de inteligencia de negocios para la toma de decisiones en la atención hospitalaria y en cuanto a la aplicación de técnicas de minería de datos (*Clustering*) para la identificación de grupos de interés en el proceso de cirugía.

Aplicar una metodología ágil y dinámica para la implementación de proyectos de BI, permite mantener control sobre el logro de los resultados propuestos, facilita la gestión del conocimiento para desarrollo futuro de soluciones y reduce los tiempos de ejecución.

La comparación realizada entre las diferentes metodologías para la implementación de proyectos de minería de datos, permitió adoptar y adaptar formatos que facilitaron la documentación de las fases. La metodología CRISP-DM resultó ser abierta y flexible para trabajar con cualquier herramienta de explotación de datos e incluye una fase dedicada íntegramente al entendimiento del negocio.

La fase de "Comprensión del Negocio" de la metodología CRISP-DM permitió alinear los objetivos del negocio, la participación del equipo de salud en el diseño y lograr el apoyo de los diversos procesos de la Institución, factores que son críticos para el desarrollo de proyectos de inteligencia de negocios.

El modelo de datos es la base fundamental para lograr óptimos niveles de rendimiento, disponibilidad de los datos y resultados de BI satisfactorios, el uso de modelo conceptual, lógico y físico, permitió conocer desde la etapa de planeación los retos a ser resueltos y la definición del alcance de la solución.

Los *Data Marts* propuestos por (Kimball & Ross, 2013) a través de su arquitectura de bus, permitieron al proyecto construir un modelo de datos bajo el esquema de constelación de hechos, lo que permite adherir trabajos futuros dentro de la institución sin que ello represente cambios en la estructura actual.

La inclusión de informes como datos generales de cirugía, total de procedimientos por sala, producción bruta por sala, número de cirugías, uso de horas de cirugía, cuadros comparativos entre salas, semaforización de indicadores, indicadores de eficiencia, comparaciones por semanas, meses, trimestres, semestres y años, permite apoyar la toma de decisiones en quirófanos para obtener mejores resultados.

La visualización de reportes a través de dispositivos móviles, representó uno de los factores de mayor impacto en los resultados.

No se evidenciaron dificultades metodológicas, sin embargo si hubo dificultades relacionadas con la completitud de los datos en la historia clínica de los pacientes y la imposibilidad de hacer transformación durante el proceso de *ETL*, debido a aspectos normativos.

Aunque la herramienta *DB2 Web Query* ofrece herramientas para la documentación de proyectos de inteligencia de negocios, fue necesario recurrir a otras aplicaciones para la diagramación de los modelos.

De acuerdo con la evaluación realizada con los líderes, la solución implementada permite realizar inferencias válidas, lo cual es sumamente beneficioso para la institución, tanto desde lo estratégico, como gerencial y financiero.

La técnica de *Clustering* aplicada al agrupamiento de pacientes, permitió ratificar conceptos institucionales frente al tipo de patologías asociadas a cáncer detectadas en pacientes de edad avanzada, ayudando a soportar la necesidad de fortalecer servicios de naturaleza oncológica en la red pública de salud.

Validar los resultados obtenidos a través de la participación activa de los interesados, no solo permite evidenciar el logro de los resultados, sino que también compromete a los usuarios con el éxito en la ejecución del proyecto.

## 5. RECOMENDACIONES

Promover la mayor cantidad de interacciones entre el equipo desarrollador de la solución, con el personal relacionado con atención en salud y los líderes que usarán el conocimiento para la toma de decisiones, más aun teniendo en cuenta el tipo de datos gestionado, su complejidad técnica y susceptibilidad a interpretaciones.

Aplicar los formatos y documentación propuesta en proyectos futuros de minería de datos en la institución, con el objetivo de refinar la metodología y encontrar un modelo ágil, práctico y efectivo.

Definir un modelo de datos dinámico que pueda ser escalable y sistémico, más aun cuando la razón de ser de los proyectos de BI son los usuarios, por ello es clave mantener su interés en el uso del producto, generando conjuntos de datos innovadores, dinámicos, prácticos, móviles e ir más allá de los reportes estáticos (Madsen, 2011)

Socializar a nivel estratégico, táctico y operativo, los resultados obtenidos, procurando que todos los niveles de la organización se vean beneficiados con la solución.

Aplicar técnicas de *Clustering* diferentes, permite validar los resultados obtenidos y establecer técnicas, algoritmos y distancias de similitud apropiadas para este tipo de obtención de conocimiento.

Debido a que los campos vacíos no pueden ser modificados por la prohibición normativa de modificar el contenido de la historia clínica y por sus posibles efectos en la interpretación clínica, es necesario efectuar revisión detallada de los orígenes de datos en compañía de profesionales de la salud.

En la fase de modelado, es necesario disponer de varias sesiones de trabajo con el usuario final, para evitar reprocesos en estructuración de los cubos dimensionales y tableros de monitoreo.

## 6. TRABAJO FUTURO

Durante el desarrollo de esta tesis ha sido posible identificar problemas relacionados con la implementación de proyectos de inteligencia de negocios y minería de datos, surgiendo la importancia de plantearlos como trabajos futuros de investigación:

- I. Si bien es cierto que se aplicó la técnica de minería de datos *Clustering* y particularmente el algoritmo *K-Means*, su campo de acción se limitó a la identificación de grupos de pacientes sometidos a procedimientos quirúrgicos relacionados con diagnósticos de cáncer a partir de variables demográficas y tiempos de estancia. Podría aprovecharse el alcance de la técnica mencionada, para integrar atributos clínicos derivados de las atenciones hospitalarias que permita a los profesionales identificar similitudes clínicas entre grupos de pacientes atendidos.
- II. Aunque se encontró en la revisión documental la aplicación de diversas técnicas de minería de datos en el sector salud, fue difícil hallar literatura asociada a la minería de datos en la gestión administrativa de los hospitales. Sería importante identificar diferentes algoritmos de minería de datos y su posible asociación a la gerencia de entidades de salud.
- III. Luego de aplicar las fases propuestas por la metodología *CRISP-DM* se logró implementar la solución de inteligencia de negocios basadas en minería de datos planteada inicialmente, sin embargo sería recomendable ejecutar posteriormente, monitoreo y análisis de los resultados obtenidos en la práctica y la identificación de posibles problemas, riesgos y barreras que puedan ser intervenidos previamente en proyectos futuros.
- IV. El proyecto actual aplicó las fases propuestas por la metodología *CRISP-DM*, la cual fue elegida entre varias opciones a partir de algunas comparaciones conceptuales. Un trabajo futuro podría implementar productos similares con otra(s) metodología(s), ofreciendo un estudio comparativo y facilitando la decisión al momento de elegir el marco de trabajo.
- V. La solución de inteligencia de negocios desarrollada, se enfocó en un único proceso. Mediante trabajos futuros, es posible validar el método utilizado y su efectividad en otros procesos o servicios asistenciales.
- VI. En el proyecto actual se utilizó la herramienta *DB2 Web Query* para el modelamiento de datos y visualización de reportes de *BI* y *ROAMBI Analytics* para dispositivos móviles. Como trabajo futuro podría realizarse un experimento que permita comparar los resultados con otro tipo de herramientas y ayudar a definir el *framework* que más se asimile a las necesidades institucionales.
- VII. La Fase de “Comprensión del Negocio” planteó posibles riesgos a los cuales se podría ver sometido el proyecto de inteligencia de negocios. Dicho análisis

partió de la información institucional, pero no obedece a un enfoque amplio. Sería de gran importancia contar con un estudio que, a través de *bigdata* pueda recopilar información sobre los riesgos a los cuales se han visto expuestos este tipo de proyectos y recomendar elementos de control.

- VIII. Durante la fase de evaluación, se diseñaron formatos para medir el nivel de satisfacción de los interesados, tras los resultados obtenidos. Este tipo de evaluación podría ser sometido a un estudio de investigación que ratifique su validez, la cobertura necesaria para garantizar su objetividad y la formalización de sus mediciones.

## 7. DIVULGACIÓN DE RESULTADOS

Como resultado de este proyecto de investigación se generaron dos artículos científicos:

### Artículo 1

**A Business Intelligence solution for a surgical service in a high complexity public hospital.** Presentado y aceptado para revisión en la revista virtual de la Universidad Católica del Norte. ISSN 0124-5821, indexación pubindex categoría B.

### Artículo 2

**Clustering technique based on k- means algorithm for the identification of clusters of surgical patients.** Presentado en el IX Seminario Internacional de Ciencias de la Computación – SICC 2016 – organizado por la Universidad de Medellín y la Universidad del Quindío en la ciudad de Medellín. 27 y 28 de octubre de 2016.

## BIBLIOGRAFÍA

Abril D., Pérez J. (2007). Estado actual de las tecnologías de bodega de datos y OLAP aplicadas a bases de datos espaciales. [Fecha de consulta: 23 de junio de 2015]. Disponible en <http://www.scielo.org.co/pdf/iei/v27n1/v27n1a08.pdf>

Aghdaie, M. H., Zolfani, S. H., & Zavadskas, E. K. (2014). Synergies of Data Mining and Multiple Attribute Decision Making. *Procedia - Social and Behavioral Sciences*, 110, 767–776. Doi:10.1016/j.sbspro.2013.

Agnoletti, V., Bucciolini, M., Padovani, E., Corso, R. M., Perger, P., Piraccini, E., Gambale, G. (2013). Operating room data management: improving efficiency and safety in a surgical block. *BMC Surgery*, 13(1), 7. <http://doi.org/10.1186/1471-2482-13-7>

Albarrán Trujillo SE, Salgado Gallegos M. (2013) La Inteligencia Analítica y la Competitividad en las Empresas. *RECAI Revista de Estudios en Contaduría, Administración e Informática* [Internet]. 30–4 [cited 2015 Apr 22]; 0(0). Available from: <http://www.revistarecai.mx/index.php/recai/article/view/18>

Aldás Manzano, Joaquin. (2002) El análisis clúster. Universitat de València Dpto. de Direcció de Empreses “Juan José Renau Piqueras”.

Aldás Manzano, Joaquin. (2000) El análisis Multivariable: Conceptos Básicos. Universitat de València Dpto. de Direcció de Empreses “Juan José Renau Piqueras”.

Azevedo, A. and Santos, (2008) M. F. KDD, SEMMA and CRISP DM: a parallel overview. In *Proceedings of the IADIS European Conference on Data Mining 2008*, pp 182 185.

Azma, F., & Mostafapour, M. A. (2012). Business intelligence as a key strategy for development organizations. *Procedia Technology*, 1, 102–106. Doi:10.1016/j.protcy.2012.02.020

Bedoya, Hernando. Jansen, Jackie. Cobb, Gene. Andrews, Robert. Swan, Linda. Even Tyler. (2010). *Getting Started with DB2 Web Query for i*. International Technical Support Organization. IBM RedBooks. ISBN 0738433985 <http://www.redbooks.ibm.com/redbooks/pdfs/sg247214.pdf>

Bedoya Hernando, Andrews Robert, Bar Theo, Cobb Gene, Falgler Rick, Fujii Kei, Gongora Hector, Mack Doug, Pacchiarini Simona, Sukariningrum Septia, Wang Yu Yu, Zhi Yu Zhe, Zhu Yu. 2014. *IBM DB2 Web Query for i. Version 2.1. Implementation Guide*. <http://www.redbooks.ibm.com/redbooks/pdfs/sg248063.pdf>

Bellazzi R, Abu-Hanna A. (2009) Data Mining Technologies for Blood Glucose and Diabetes Management. *J Diabetes Sci Technol* [Internet]. [Cited 2014 Aug 16]; 3(3):603–12. Available from:<http://dst.sagepub.com/content/3/3/603>

Bellucci Emilia, Yee Ling Boo, Nguyen Lemai. (2012) Data Visualisation in Healthcare Analytics | Visual Analytics [Internet]. [Cited 2014 Aug 17]. Available from: <http://visanalytics.org/info/projects/data-visualisation-in-health/>

Bernal-Acevedo O, Forero-Camacho JC. (2011) Information systems in health sector in Colombia. *Revista Gerencia y Políticas de Salud*. [cited 2015 Jul 3];10(21):85–100. Available from: [http://www.scielo.org.co/scielo.php?script=sci\\_abstract&pid=S1657-70272011000200006&lng=en&nrm=iso&tlng=en](http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S1657-70272011000200006&lng=en&nrm=iso&tlng=en)

Berry, M.J.A.; Linoff, G. (1997): *Data mining techniques: for marketing, sales and customer support*. Ed. John Wiley & Sons, Inc. Nueva York. ISBN: 978-0-471-17980-1

Bonney, W. (2013). Applicability of Business Intelligence in Electronic Health Record. *Procedia - Social and Behavioral Sciences*, 73, 257–262. Doi:10.1016/j.sbspro.2013.02.050

Boyer, Jonn. Bill, Frank. Green, Brian. Tracy, Harris. Van de Vanter, Kay. (2011) *Business Intelligence Strategy. A Practical Guide for Achieving (BI) Excellence*. McPress. ISBN: 978-158347-362-7

Britos, Paola. (2008). *Procesos de Explotación de Información basados en Sistemas Inteligentes*. Tesis Doctoral. Universidad Nacional de La Plata, Facultad de Informática. La Plata, Argentina. <http://www.iidia.com.ar/rgm/tesistas/td-pb-fi-unlp.pdf>. 13/03/2015

Bustamante Martínez A, Galvis Lista E.A, Gómez Flórez LC. 2013. Técnicas de modelado de procesos de ETL: una revisión de alternativas y su aplicación en un proyecto de desarrollo de una solución de BI. (Spanish). *Scientia et Technica* [cited 18(1):185–91.

Caron, F., Vanthienen, J., & Baesens, B. (2013). Healthcare Analytics: Examining the Diagnosis–treatment Cycle. *Procedia Technology*, 9, 996–1004. Doi:10.1016/j.protcy.2013.12.111

Castellani B, Castellani J. (2003) Data Mining: Qualitative Analysis with Health Informatics Data. *Qual Health Res.* 9–1 [cited 2014 Aug 16]; 13(7):1005–18. Available from:<http://qhr.sagepub.com/content/13/7/1005>

Cattaneo MF, Amatriain H, Rodríguez D, Pytel P, Ciccolella E, Vegega C, et Al. (2010) Ingeniería de proyectos de explotación de la información. [cited 2015 Apr 22]. Available from:<http://hdl.handle.net/10915/19463>

Chaffey Dave, Wood Steve. (2005) *Business Information Management. Improving Performance Using Information Systems*. Prentice Hall. Cap 5. Knowledge Management Strategy. 220-270 p. ISBN: 978-0-273-68655-2

Chapman, Pete. Clinton, Julian. Kerber, Randy. Khabaza Thomas. Reinartz, Thomas. Shearer, Colin. Wirth Rüdiger. SPSS, NCR, DaimlerChrysler. (2000) *Guía paso a paso de Minería de Datos*. Copyright © 1999, 2000. <http://the-modeling-agency.com/crisp-dm.pdf> 12/02/2015

Cheung RB, Moody LE, Cockram C. (2002) Data Mining Strategies for Shaping Nursing and Health Policy Agendas. *Policy Politics Nursing Practice* [Internet]. 8–1 [cited 2014 Aug 16];3(3):248–60. Available from:<http://ppn.sagepub.com/content/3/3/248>

Clark Karen, Normile Loretta. (2012). Nursing informatics and data collection from the electronic medical record: Study of characteristics, factors and occupancy impacting outcomes of critical care admissions from the Emergency Department *Health Informatics Journal*; vol. 18, 4: pp. 309-319

Cleven, Anne. Winter Robert, Wortmann, Feliz. Mettler, Tobias. (2014) *Process Management in Hospitals – An Empirically Grounded Maturity Model*. <https://www.alexandria.unisg.ch/Publikationen/233581> 08/07/2014

Cobos, Carlos. Zuñiga, Jhon. Guarín, Juan. León, Elizabeth. Mendoza, Marha. (2010) CMIN - herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos. *Ingeniería e Investigación*. <http://www.scielo.org.co/pdf/iei/v30n3/v30n3a04.pdf> 20/04/2015

Combes C, Meskens N, Rivat C, Vandamme J-P. (2008) Using a KDD process to forecast the duration of surgery. *International Journal of Production Economics* [Internet]. [Cited 2014 Nov 14]; 279–93. Available from: <http://www.sciencedirect.com/science/article/pii/S0925527307001466>

Curto, Josep. Conesa, Jordi. (2011). *Introducción al Business Intelligence*. Editorial UOC. 2011. Primera Edición. ISBN: 978-84-9788-886-8

Dan Vesset, Brian McDonough, David Schubmehl, and Mary Wardley. (2012) *IDC Market Analysis: Worldwide Business Analytics Software 2013–2017 Forecast and 2012 Vendor Shares*. [http://idcdocserv.com/241689e\\_sas](http://idcdocserv.com/241689e_sas) 15/06/2014

Davenport, Thomas; Harris, Jeanne (2007). *Competing on analytics – The new science of winning*. Boston: Harvard Business Press.

Davenport, Thomas H., HARRIS, Jeanne G., y MORISON, Robert (2010). *Analytics at Work: Smarter Decisions, Better Results*. Harvard Business Press. School Publishing Corp. ISBN: 978-1-4221-7769-3

Dorda W, Gall W, Duftschmid G. (2002) Clinical data retrieval: 25 years of temporal query management at the University of Vienna Medical School. *Methods Inf Med* 2002; 41 (2): 89–97

Dresner, Howard. Buytendijk, Frank. Linden, Alexander. Friedman, Ted. Strange, Kevin. Knox, Mary. Camm, Mark. (2002). *The Business Intelligence Competency Center: An Essential Business Strategy*. Retrieved July 30, 2015, from <https://www.gartner.com/doc/358967/business-intelligence-competency-center-essential>

Dwivedi, A. N., Bali, R. K., & Naguib, R. N. G. (2007). Building New Healthcare Management Paradigms: A Case for Healthcare Knowledge Management. In R. K. Bali & A. N. Dwivedi (Eds), *Healthcare Knowledge Management* (pp. 3–10). Springer New York. Retrieved from [http://link.springer.com/chapter/10.1007/978-0-387-49009-0\\_1](http://link.springer.com/chapter/10.1007/978-0-387-49009-0_1)

Epstein I. (2014) Building a Bridge or Digging a Pipeline? Clinical Data Mining in Evidence-Informed Knowledge Building. *Research on Social Work Practice* [Internet]. [Cited 2014 Aug 16]; Available from:<http://rsw.sagepub.com/content/early/2014/06/17/1049731514536475.full.pdf+html>

Fayyad, Usama. Piatesky-Shapiro, Gregory. Smyth, Padhraic. (1996). "Advances From Data Mining to Knowledge Discovery in Databases". American Association for Artificial Intelligence. <http://www.csd.uwo.ca/faculty/ling/cs435/fayyad.pdf> 15/02/2015.

Fernández, Enrique J. (2006) Asistente para la Gestión de Documentos De Proyectos de Explotación de Datos. Tesis de Maestría. Universidad Politécnica de Madrid. Buenos Aires. <http://iidia.com.ar/rgm/tesistas/fernandez-tesisdemagister.pdf> 13/03/2015

Flexer, A. (2000). Data mining and electroencephalography. *Statistical Methods in Medical Research*, 9(4), 395–413. Doi: 10.1177/096228020000900405

Flower, J. (2006). Who owns health care's most valuable information? *PhysicianExecutive*, 32(5), 54–55. <http://www.freepatentsonline.com/article/Physician-xecutive/152420393.html> ISSN: 0898-2759. [Cited 13 Nov 14]

Foshay, N., & Kuziemsky, C. (2014). Towards an implementation framework for business intelligence in healthcare. *International Journal of Information Management*, 34(1), 20–27. <http://www.sciencedirect.com/science/article/pii/S0268401213001126>

Glover, S., Rivers, P. A., Asoh, D. A., Piper, C. N., & Murph, K. (2010). Data mining for health executive decision support: an imperative with a daunting future! *Health Services Management Research*, 23(1), 42–46. Doi:10.1258/hsmr.2009.009029

Gamarra Barrios, Mónica. (2006) Minería de Datos aplicada a problemas de Investigación Operativa. Universidad de la República. Uruguay. Instituto de Computación. [cited 18 Abr 2015] <https://www.colibri.udelar.edu.uy/bitstream/123456789/3092/1/tg-gamarra.pdf>

Ghazanfari M, Jafari M, Rouhani S. (2011) A tool to evaluate the business intelligence of enterprise systems. *Scientia Iranica* [Internet]. [cited 2015 Mar 16]; 18(6):1579–90. Available from:<http://www.sciencedirect.com/science/article/pii/S102630981100215X>

Hernández Orallo, José. Ramírez Quintana, M José. Ferri Ramírez, César. (2005) *Introducción a la Minería de Datos*. Editorial Pearson, ISBN: 84 205 4091 9. p. 575.

Hernández Ramírez, Denia Martha. (2011) Procedimiento para el desarrollo de un sistema de inteligencia de negocios en la gestión de ensayos clínicos en el Centro de Inmunología Molecular. *ACIMED* [Internet]. [Cited 2014 Nov 26]; 22(4):351–63. Available from: [http://scielo.sld.cu/scielo.php?script=sci\\_abstract&pid=S1024-94352011000400006&lng=es&nrm=iso&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S1024-94352011000400006&lng=es&nrm=iso&tlng=es)

Hochberg, A. M., Reisinger, S. J., Pearson, R. K., O'Hara, D. J., & Hall, K. (2007). Using Data Mining to Predict Safety Actions from FDA Adverse Event Reporting System Data. *Drug Information Journal*, 41(5), 633–643. doi: 10.1177/009286150704100510

HOWE, Bill. (2014) Introduction to Data Science. Retrieved November 8. Universidad de Washington. ([https://www.coursera.org/course/datasci?from\\_restricted\\_preview=1&course\\_id=972772&r=https%3A%2F%2Fclass.coursera.org%2Fdatasci-002%2Fclass](https://www.coursera.org/course/datasci?from_restricted_preview=1&course_id=972772&r=https%3A%2F%2Fclass.coursera.org%2Fdatasci-002%2Fclass)).

IBM. 2010. Manual CRISP-DM. IBM SPSS Modeler.

Inmon, W. H., (1997) *Building the Data Warehouse.*, 2nd edition, John Wiley & Sons.

Inmon, W. H., (2005) *Building the Data Warehouse.*, 4 edition, John Wiley Publishing.

Isik, O., Jones, M. C., & Sidorova, A. (2011). Business Intelligence (bi) Success and the Role of Bi Capabilities. *Intelligent Systems in Accounting, Finance and Management*, 18(4), 161–176. Doi:10.1002/isaf.329

Jones AR. (2009) Data mining can support quality assurance. *J R Soc Med [Internet]*. 9–1 [cited 2014 Aug 16]; 102(9):358–9. Available from:<http://jrs.sagepub.com/content/102/9/358.2>

Kahn MG, Kaplan D, Sokol RJ, DiLaura RP. (2007) Configuration challenges: implementing translational research policies in electronic medical records. *Acad Med*; 82 (7): 661–669

Kaplan, R S; Norton, D P (1992). "The balanced scorecard - Measures that drive performance". En: *Harvard Business Review*. pp. 72

Khademolqorani, S., & Hamadani, A. Z. (2013). An Adjusted Decision Support System through Data Mining and Multiple Criteria Decision Making. *Procedia - Social and Behavioral Sciences*, 73, 388–395. Doi:10.1016/j.sbspro.2013.02.066

Kimball, R. and Ross, M., (2002). *The data warehouse toolkit: the complete guide to dimensional modelling*. John Wiley & Sons, New York.

Kimball, Ralph. Ross, Margy. (2013). *The Data Warehouse Toolkit. The definitive guide to Dimensional Modeling*. Third Edition. Kimball Group. Editorial Wiley. ISBN: 978-1-118-53080-1.

Kudyba, S., & Gregorio, T. (2010). Identifying factors that impact patient length of stay metrics for healthcare providers with advanced analytics. *Health Informatics Journal*, 16(4), 235–245. Doi:10.1177/1460458210380529

Luhn, Hans P. (1958). *A Business Intelligence System*. New York: IBM. [Fecha de consulta: 23 de junio de 2015] <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5392644>

Luminița, Ș., & Magdalena, R. (2012). Optimizing Time in Business with Business Intelligence Solution. *Procedia - Social and Behavioral Sciences*, 62, 638–648. Doi:10.1016/j.sbspro.2012.09.106

Luzón Calderón, Valentina. (2004) *Implantación de un proyecto de Knowledge Center con una herramienta comercial (Synera)*. Universitat Oberta de Catalunya. [Cited 2014 Aug 16]; From: <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/442/1/27899tfc.pdf>

MANTILLA, Adriana. (2000) *Empresas Sociales del Estado*. Tesis de Grado. Universidad Sergio Arboleda. Bogotá.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Presented at the Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, The Regents of the University of California. Retrieved from <http://projecteuclid.org/euclid.bsmmsp/1200512992>

Madsen L. (2011) Optimize your BI: five reasons healthcare business intelligence differs from other industries. *Health Manag Technol*. <http://www.ncbi.nlm.nih.gov/pubmed/22141241> 10/03/2015

Marcano, Yelitza. Talavera, Rosalba. (2007) Minería de Datos como soporte a la toma de decisiones empresariales, vol. 23, núm. 52, abril, pp. 104-118, Universidad de Zulia. Venezuela. Disponible en: <http://www.redalyc.org/articulo.oa?id=31005208>  
30 de julio de 2014

Martínez Álvarez, Clemente Antonio. (2012). Aplicación de Técnicas de Minería de Datos para Mejorar el Proceso de Control de Gestión en Entel [Internet]. Universidad De Chile; Available from: [http://repositorio.uchile.cl/bitstream/handle/2250/112065/cf-martinez\\_ca.pdf?sequence=1](http://repositorio.uchile.cl/bitstream/handle/2250/112065/cf-martinez_ca.pdf?sequence=1)

Martínez García, J. H. (2010). La inteligencia de negocios como herramienta para la toma de decisiones estratégicas en las empresas. Análisis de su aplicabilidad en el contexto corporativo colombiano. Universidad Nacional de Colombia. [Cited 2015 Jul 7] Retrieved from <http://www.bdigital.unal.edu.co/3098/>

Megalooikonomou, V., Ford, J., Shen, L., Makedon, F., & Saykin, A. (2000). Data mining in brain imaging. *Statistical Methods in Medical Research*, 9(4), 359–394. Doi:10.1177/09622802000900404

MINSALUD. (2008) Ministerio de Salud y Protección Social. Colombia. Alianza mundial para la seguridad del paciente. Seguridad en los procedimientos quirúrgicos. Versión 1.0. 2008. [http://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/CA/SEGURIDAD\\_EN\\_PROCESOS\\_QUIRURGICOS.pdf](http://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/CA/SEGURIDAD_EN_PROCESOS_QUIRURGICOS.pdf) 04/11/2014

-----, -----, (2013) Oferta y demanda del recurso médico en Colombia. Boletín Electrónico para los Actores del Sistema de Salud En Colombia No. 34 – JULIO 2 DE 2013. [http://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/COM/Enlace\\_MinSalud\\_34.pdf](http://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/COM/Enlace_MinSalud_34.pdf)

-----, -----, (2014) Dirección de Epidemiología y Demografía. Gestión del Conocimiento y Fuentes de Información. Número de atenciones por procedimientos e intervenciones quirúrgicas, según códigos CUPS en Colombia para los años 2009 a 2012. Consultado Bodega de Datos SISPRO 30 julio 2014.

-----, -----, (2014) Informe de Actividades 2010 – 2014. Sector Administrativo de Salud y Protección Social. Al Honorable Congreso de la República. Bogotá, D.C. Plan Decenal de Salud Pública 2012-2021

Mans, Ronny. Schonenberg, Helen, Leonardi, B. Panzarasa, Silvia. Cavallini, Anna. Quaglini Silvana. Aalsta, Wil Van Der. (2008). Process Mining Techniques: an Application to Stroke Care. Eindhoven University of Technology, the Netherlands. Dept. Computer and System science, University of Pavia, Italy. *EHealth Beyond the Horizon*. <http://schoenberg.info/files/SHT1136-0573.pdf> 31 de julio de 2014

Martínez García JH. (2010) La Inteligencia de Negocios como Herramienta para la Toma de Decisiones Estratégicas en las Empresas. Análisis de su aplicabilidad en el contexto corporativo colombiano. Universidad Nacional de Colombia; [cited 2015 Jun 23]. Available from: <http://www.bdigital.unal.edu.co/3098/>

Mclaren, Christine. Pin Chen, Wen. Nie, Ke. Su, Ming-Ying. (2009) Prediction of malignant breast lesions from MRI features: a comparison of artificial neural network and logistic regression techniques. Department of Epidemiology, University of California. *Academic radiology (Impact Factor: 2.09)*; 16(7):842-51. Elsevier. ISSN: 1878-4046.

Méndez, P., & Rodríguez, A. (2009). Herramienta de estudio de viabilidad para proyectos que utilizan la metodología P3TQ. Universidad de Buenos Aires, Buenos Aires, Argentina. <http://laboratorios.fi.uba.ar/lisi/mendez-rodriquez-trabajoprofesional.pdf> 10/05/2015

Mettler, T., & Vimarlund, V. (2009). Understanding business intelligence in the con-text of healthcare. *Health Informatics Journal*, 15(3), 254–264. Doi:10.1177/1460458209337446

Migut, M., & Worrying, M. (2012). Visual exploration of classification models for various data types in risk assessment. *Information Visualization*, 11(3), 237–251. Doi:10.1177/1473871611433715

M. Neves, (2011) "Os Médicos vão ter de ser os motores da reforma do sistema," Revista Portuguesa de Gestão & Saúde, no. 5.

Moine, Juan Miguel. Haedo, Silvia. Gordillo, Silvia. (2010). Estudio comparativo de metodologías para minería de datos. UTN Rosario Facultad de Ciencias Exactas, Universidad Nacional de Buenos Aires Facultad de Informática, Universidad Nacional de La Plata. [http://sedici.unlp.edu.ar/bitstream/handle/10915/20034/Documento\\_completo.pdf%3Fsequence%3D112/02/2015](http://sedici.unlp.edu.ar/bitstream/handle/10915/20034/Documento_completo.pdf%3Fsequence%3D112/02/2015).

Molina L, Juan. García H, Jesús. (2004). Técnicas de análisis de datos aplicaciones prácticas utilizando Microsoft Excel y Weka. Retrieved Cited [July 22, 2015] from <http://myslide.es/documents/tecnicas-de-analisis-de-datos.html>

Montaño Moreno, Juan J.; Gervilla García, Elena; Cajal Blasco, Berta; Palmer, Alfonso. (2014) Técnicas de clasificación de data mining: una aplicación al consumo de tabaco en adolescentes Anales de Psicología, vol. 30, núm. 2, mayo-agosto, pp. 633-641 Universidad de Murcia Murcia, España <http://www.redalyc.org/pdf/167/16731188027.pdf>

Negash S, Gray P. (2008) Business Intelligence. Handbook on Decision Support Systems 2 [Internet]. Springer Berlin Heidelberg; [cited 2015 Jun 23]. p. 175–93. Available from: [http://link.springer.com/chapter/10.1007/978-3-540-48716-6\\_9](http://link.springer.com/chapter/10.1007/978-3-540-48716-6_9)

Nofal MI, Yusof ZM. (2013) Integration of Business Intelligence and Enterprise Resource Planning within Organizations. Procedia Technology [Internet]. [Cited 2014 Jul 21];11:658–65. Available from: <http://www.sciencedirect.com/science/article/pii/S2212017313003964>

O'Brien, James. Marakas, George. (2006) Sistemas de Información Gerencial. McGraw-Hill Séptima Edición. p 12-15. ISBN 978-970-10-5630-1

Olaleye, D., & Tardiff, B. E. (2001). Practical Issues in and Applications of Clinical Data Mining. Drug Information Journal, 35(3), 791–808. Doi:10.1177/009286150103500317

Oliveria, Sérgio. Portela, Filipe. Manuel F. Santos. José Machado. António Abelha. (2013) Predictive Models for Hospital Bed Management. Using Data Mining Techniques. University of Minho, Guimarães, Portugal.

OMS. (2004) Organización Mundial de la Salud. Developing health management information systems : a practical guide for developing countries [Internet]. Manila : WHO Regional Office for the Western Pacific; [cited 2015 Jul 3]. Available from: <http://iris.wpro.who.int/handle/10665.1/5498>

Parker SH, Rosen M. (2012) Developing methods to measure healthcare team performance in acute and chronic care settings. Proceedings of the Human Factors and Ergonomics Society Annual Meeting [Internet]. [Cited 2014 Aug 17];56(1):834–5. Available from:<http://pro.sagepub.com/content/56/1/834>

Pogue JM, Devereaux PJ, Thorlund K, Yusuf S. (2013) Central statistical monitoring: Detecting fraud in. Clin Trials [Internet]. [Cited 2014 Aug 17];10(2):225–35. Available from:<http://ctj.sagepub.com/content/10/2/225>

Popovič, A., Hackney, R., Coelho, P. S., & Jaklič, J. (2012). Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. Decision Support Systems, 54(1), 729–739. Doi:10.1016/j.dss.2012.08.017

Porter, Michael E. (1980). Competitive Strategy: Techniques for Analyzing Industries and Competitors. Originally published: New York Free Press. ISBN 0-684-84148-7

Pressman, Roger S. (2010) Ingeniería Del Software. Un Enfoque Práctico. 7ª Edición. McGraw Hill. ISBN. 978-607-15-0314-5

Prokosch, H.-U., & Ganslandt, T. (2009). Perspectives for Medical Informatics: Reusing the Electronic Medical Record for Clinical Research. Methods of Information in Medicine. <http://doi.org/10.3414/ME9132>

Pyle, Dorian. (2003) Business Modeling and Data Mining. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann. ISBN 0080500455, 9780080500454.

Rezaei, Mansour. Eghbal, ZanKarimi. Amirhossein Hasheiman. (2013) Comparison of Artificial Neural Network, Logistic Regression and Discriminant Analysis Efficiency in Determining Risk Factors of Type 2 Diabetes. World Applied Sciences Journal 07/2013; 23(11):1522-1529. DOI: 10.5829/idosi.wasj.2013.23.11.1119

Rivas IT, Rivera MR, Lizama ER. (2007) Una metodología para sectorizar pacientes en el consumo de medicamentos aplicando datamart y datamining en un hospital [Internet]. Industrial Data. [Cited 2015 Jul 16]. Available from: <http://www.redalyc.org/articulo.oa?id=81610114>

Rodríguez Ahumada Rosa Janeth. (2011) Diseño de un Sistema de Información Gerencial Alineado con la Orientación Estratégica de la Empresa para el Soporte en la Toma de Decisiones a Nivel Estratégico. Repositorio institucional [Internet] [masters]. Universidad Nacional de Colombia; [cited 2015 Jul 7]. Available from: <http://www.bdigital.unal.edu.co/6433/>

Rodríguez, Ma Teresa. Álvarez, Valeriano. Mesa, José. (2003) Metodología para el Desarrollo de Proyectos de Data Mining. Universidad de Oviedo. [http://aeipro.com/files/congresos/2003pamplona/ciip03\\_0257\\_0265.2134.pdf](http://aeipro.com/files/congresos/2003pamplona/ciip03_0257_0265.2134.pdf)

Rohanizadeh, Seyyed. Moghadama, Mohammad. (2009). A Proposed Data Mining Methodology and its Application to Industrial Procedures. Journal of Industrial Engineering 4 37-50.

Rohloff, R. (2011). Healthcare (BI): a tool for meaningful analysis. Healthcare Financial Management: Journal of the Healthcare Financial Management Association, 65(5), 100–6, 108. Retrieved from <http://europepmc.org/abstract/MED/21634274>

Salcedo, Octavio. Galeano, Rita. Rodríguez, Luis (2010). Metodología crisp para la implementación Data Warehouse / Methodology crisp for data warehouse implementation. Tecnura, (26), 35. Citado [26/08/2015] url: [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0123-921X2010000100005&lng=en&tlng=en](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-921X2010000100005&lng=en&tlng=en)

Sarmiento, Xavier. Guardiola, Juan. Roca, J. Soler, M. Toboso, J.M. Klamburg, J. Artigas, Antonio. (2013) Servicio Medicina Intensiva, Hospital Universitario Germans Trias i Pujol, Badalona, Barcelona, España. Medicina Intensiva (Impact Factor: 1.32). 37(3):132–141. DOI: 10.1016/j.medin.2012.03.006

Singhal, A. (2007). An Overview of Data Warehouse, OLAP and Data Mining Technology. In Data Warehousing and Data Mining Techniques for Cyber Security (pp. 1–23). Springer US. Retrieved from [http://link.springer.com/chapter/10.1007/978-0-387-47653-7\\_1](http://link.springer.com/chapter/10.1007/978-0-387-47653-7_1)

Shneiderman B. (2002). Inventing Discovery Tools: Combining Information Visualization with Data Mining1. Information Visualization [Internet]. [Cited 2014 Aug 16];1(1):5–12. Available from: <http://ivi.sagepub.com/content/1/1/5>

Tamayo, Marysol. Moreno A, Francisco. J. (2010). Repositorio institucional UN. Ingeniería E Investigación; Vol. 26, Núm. 3 (2006); 135-142 Ingeniería E Investigación; Vol. 26, Núm. 3 (2006); 135-142 2248-8723 0120-5609. Retrieved from <http://revistas.unal.edu.co/index.php/ingeninv/article/view/14764>

Taié Armando. (2008). Desarrollo de una metodología de extracción de conocimientos a partir de datos de micromatrices de DNA basada en ontologías genéticas. Universidad de Buenos Aires. [Cited 15/07/2015]. Available from: <http://inta.gob.ar/personas/taie.armando>

Tomasello M, Pytel P, Rodríguez D, Arboleya H, Pollo Cattaneo MF, Britos PV, et al. (2011) Estimación de proyectos de explotación de información. [Cited 2015 Apr 22]. Available from: <http://hdl.handle.net/10915/20041>

Tremblay, M. C., Hevner, A. R., & Berndt, D. J. (2012). Design of an information volatility measure for health care decision making. Decision Support Systems, 52(2), 331–341. Doi:10.1016/j.dss.2011.08.009

Tsumoto, S, Iwata, H. Hirano, S. Tsumoto, Y. (2012) Exploratory temporal data mining process in hospital information systems Cognitive Informatics & Cognitive Computing (ICCI\*CC), 2012 IEEE 11th International Conference on; 01/2012

Two Crows Corporation. (2005). Introduction to Data Mining and Knowledge Discovery. Third Edition. ISBN: 1-892095-02-5. [Cited 2015 Jul 22] Available from: <http://www.twocrows.com/intro-dm.pdf>

Van Manen, Robbert P. Identifying Patients at Risk in Clinical Trials using Data Mining and Data Visualization Technologies. Conference: Drug Information Association 26th Annual EuroMeeting, At Vienna, Austria. 2014.

Vanrell JÁ, Bertone RA, García Martínez R. (2010) Un modelo de procesos de explotación de la información. [Cited 2015 Apr 22]. Available from: <http://hdl.handle.net/10915/19462>

Van Oostrum, J. M., Parlevliet, T., Wagelmans, A. P. M., & Kazemier, G. (2011). A Method for Clustering Surgical Cases to Allow Master Surgical Scheduling. *INFOR: Information Systems and Operational Research*, 49(4), 254–260. <http://doi.org/10.3138/infor.49.4.254>

Valcárcel Asencios. (2004). Data Mining y el descubrimiento del conocimiento *Industrial Data*, vol. 7, núm. 2, julio-diciembre, pp. 83-86, Universidad Nacional Mayor de San Marcos. Perú. <http://www.redalyc.org/pdf/816/81670213.pdf> (Consultado 16/07/2014)

Velasquez, J D. Palade, V. (2008). Adaptive Web Sites: A Knowledge Extraction from Web Data Approach - Volume 170 *Frontiers in Artificial Intelligence and Applications*. Amsterdam : Washington, D.C: IOS Press.

Wang M. (2013). Predefined Three Tier Business Intelligence Architecture in Healthcare Enterprise. *Journal of Medical Systems [Internet]*. [Cited 2015 Jul 7]; 37(2):1–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23334840>

Wang, F. Cheng, Y. Xie, Y. (2012). Research on Intelligent Power Consumption Business Intelligence System based on Cloud Computing. *Proceedings of the 2012 International Conference on Computer Science and Service System (CSSS)*. 11-13, pp. 1524; Washington DC, USA: IEEE; 2012

West, Suzanne L, Blake, C., Liu, Z., McKoy, J. N., Oertel, M. D., & Carey, T. S. (2009). Reflections on the use of electronic health record data for clinical research. *Health Informatics Journal*, 15(2), 108–121. Doi:10.1177/1460458209102972

Wirth, Rüdiger. Hipp, Jochen. (2000) CRISP-DM: Towards a Standard Process Model for Data Mining. DaimlerChrysler Research & Technology. University of Tübingen. Germany. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.198.5133&rep=rep1&type=pdf> 21/04/2015

Xu, Rui. Donald, Wunsch. (2005). "Survey of Clustering Algorithms." *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council* 16(3):645–78.

Xu, Rui. Donald, Wunsch. (2009). *Clustering*. Wiley-IEEE Press. ISBN 9780470276808