

DEFINICIÓN DE ESTRATEGIAS PARA LA IMPLEMENTACIÓN DE UNA
BODEGA DE DATOS EN LA SECRETARIA DE SALUD DE MEDELLIN

LILIANA MARCELA VALENCIA VALDERRAMA

ASESOR

JUAN CAMILO GIRADO MEJIA

UNIVERSIDAD DE SAN BUENAVENTURA SECCIONAL MEDELLIN

FACULTAD DE INGENIERÍAS

ESPECIALIZACION EN LA GESTION DE INFORMACION Y BASES DE DATOS

MEDELLIN

2013

DEDICATORIA

Para las personas que han pasado por mi vida y han aportado para enseñarme quien soy (DIOS, mis padres, mis hermanas, mi amado esposo) y para quien con tanto cariño, dedicación y paciencia a pesar de sus compromisos me apoyo en las dificultades (JG) nunca olvidare sus palabras porque por fin aquí estoy estudiando no para saber más ni ser mejor que otros sino para superarme a mí misma y ayudar a los otros, tremendo regalo de la vida son ustedes para mí.

RESUMEN

La gestión de la información y el conocimiento son actualmente una de las actividades más estratégicas y claves para el éxito de las organizaciones. Optimizar el manejo de la información se ha convertido en un elemento competitivo y de gran poder, el lograr convertir dicha información en conocimiento, será la principal clave para la toma de decisiones **(Tejada, 2010)**.

Las entidades de salud como organizaciones prestadoras de servicios no escapan a estas situaciones, ya que son entidades con grandes volúmenes de datos que requieren ser atendidos y procesados de forma rápida y eficaz para garantizar la excelencia en la prestación del servicio, estas entidades en especial las del sector público deben tomar decisiones orientadas a satisfacer la demanda de servicios de los pacientes que acuden a centros de salud y es por esto que se busca mejorar los sistemas de información ligados a estos procesos de decisión **(Ojeda, 2011)**

Para poder subsanar las dificultades que trae el procesamiento de cantidades de información almacenada en los sistemas de la Secretaría De Salud de la ciudad de Medellín, se presenta en este trabajo la definición de estrategias para la implementación de herramientas de Inteligencia de Negocios (BI) como lo es la bodega de datos o Data Warehouse (DW – Bodega de datos) que utiliza como eje articulador la identificación de las necesidades de información por parte de la dirección, para el soporte de los procesos de control de la Secretaría de Salud del Municipio de Medellín; promoviendo esta herramienta como una posible solución para la gestión y toma de decisiones proactiva, necesaria para provechar oportunidades futuras.

Palabras claves: Información, Datos, base de datos, Inteligencia de Negocios (BI), Data Warehouse.

TABLA DE CONTENIDO

1	INTRODUCCION	9
2	JUSTIFICACION	10
3	PLANTEAMIENTO DEL PROBLEMA	11
4	OBJETIVO GENERAL.....	13
5	OBJETIVOS ESPECIFICOS	13
6	MARCO TEORICO.....	14
6.1	CONCEPTOS Y FUNDAMENTOS	14
6.1.1	Datos	14
6.1.2	Información.....	14
6.1.3	Conocimiento.....	15
6.1.4	Inteligencia de negocios	16
6.1.4.1	Proceso de inteligencia de negocios.....	19
6.1.5	Data Warehouse.....	21
6.1.5.1	Definición	21
6.1.5.2	Objetivos fundamentales de un Data WareHouse	24
6.1.5.3	Elementos básicos de un Data Warehouse	25
6.1.5.4	Arquitectura del Data Warehouse	26
6.1.5.5	Elementos que constituyen la Arquitectura de un DW	27
6.1.5.5.1	Base de datos operacional y externa	27
6.1.5.5.2	Nivel de acceso a la información.....	28
6.1.5.5.3	Nivel de acceso a datos.....	28
6.1.5.5.4	Nivel de directorio de datos (Metadatos).....	28
6.1.5.5.5	Nivel de gestión de procesos.....	28
6.1.5.5.6	Nivel de mensaje de la aplicación.....	29
6.1.5.5.7	Nivel Data Warehouse (físico).....	29
6.1.5.5.8	Nivel de organización de datos.....	29
6.1.5.6	Estructura de un Data Warehouse.....	30

6.1.5.7 Base de datos transaccional (OLTP) vs Data Warehouse	31
6.1.6 Metodología Hefesto	32
6.1.6.1 Introducción	32
6.1.6.2 Principales características	33
6.1.6.3 Fases de la metodología.....	34
6.1.6.3.1 Fase1: Análisis de requerimientos.....	36
<input type="checkbox"/> Indicar preguntas.....	36
<input type="checkbox"/> Identificar indicadores.....	36
<input type="checkbox"/> Modelo Conceptual.....	36
6.1.6.3.2 Fase 2: Análisis de los OLTP	37
6.1.6.3.3 Fase 3: Modelo lógico de los DW	38
<input type="checkbox"/> Diseñar tablas de dimensiones.....	38
<input type="checkbox"/> Diseñar tablas de hechos	38
<input type="checkbox"/> Determinar jerarquías	39
<input type="checkbox"/> Realizar uniones.....	39
6.1.6.3.4 Fase 4: Integración de Datos.....	39
6.1.7 Otras metodologías	40
6.1.7.1 Metodología Kimball	40
6.1.7.2 Metodología CRISP	43
7 DEFINICION DE ESTRATEGIAS A CONSIDERAR EN LA IMPLEMENTACIÓN DE UN DW	47
8 METODOLOGÍA.....	55
9 PRESENTACIÓN DE LA SOLUCIÓN	58
9.1 PASO 1. ANÁLISIS DE REQUERIMIENTOS	58
9.1.1.1 Identificación de las preguntas.....	59
9.1.1.1.1 SAS – Sistema Aseguramiento en Salud	61
9.1.1.1.2 RIPS – Registros Individuales de prestación de servicios.	64
9.1.1.1.3 RUAF – Estadísticas Vitales	67
9.1.1.2 Identificación de las necesidades de información gerencial (Indicadores y Perspectivas)	68

9.1.1.3	Modelo Conceptual	72
9.2	PASO 2. ANALISIS DE LOS OLTP	73
9.2.1	Identificación De Las Fuentes De Datos	73
9.2.1.1	Indicar el servidor, base de datos, sistema operativo, tipo de lenguaje de programación y de consulta.	75
9.2.2	Conformar Indicadores	75
9.2.3	Establecer correspondencias	76
9.2.4	Indicar Tablas y Campos de las Fuentes o Bases de Datos	80
9.2.4.1	SAS (Sistema de Aseguramiento en Salud)	80
9.2.4.2	RUAF (Registro Único de Afiliados)	88
9.2.4.3	RIPS (Registro Individual de prestación de Servicios)	100
9.2.5	Nivel de Granularidad	112
9.2.6	Modelo Conceptual Ampliado	115
9.3	PASO3. MODELO LOGICO DEL DW	116
9.3.1	Tipo de Modelo Lógico del Data Warehouse	116
9.3.2	Tablas de Dimensiones	116
9.3.2.1	La Población afiliada al Sistema General de Seguridad Social en Salud: 117	
9.3.2.2	Nacimientos atendidos en instituciones de salud de la población Medellín:.....	117
9.3.2.3	10 primeros motivos más frecuentes de Morbilidad población Medellín:.....	118
9.3.3	Tablas de Hechos.....	118
9.3.4	Uniones	119
10	CONCLUSIONES.....	122
11	RECOMENDACIONES	123
12	REFERENCIAS BIBLIOGRAFICAS	124

INDICE DE TABLAS

Tabla 1. Definiciones de conocimiento.	15
Tabla 2. Tipos de conocimiento.	16
Tabla 3. Definiciones de Inteligencia de Negocios (BI).....	18
Tabla 4. Fases del Proceso de Inteligencia de Negocios (BI).....	19
Tabla 5. Definiciones de Data Warehouse.....	22
Tabla 6. Características de Data Warehouse	23
Tabla 7. Objetivos del Data Warehouse.....	24
Tabla 8. Elementos Básicos del Data Warehouse	25
Tabla 9. Estructura de un Data Warehouse.....	30
Tabla 10. OLTP vs DW.....	31
Tabla 11. Elementos para la Implementación de un Data Warehouse	47
Tabla 12. Factores de Planificación	50
Tabla 13. Factores de desarrollo	52
Tabla 14. Factores Para el Diseño.....	53
Tabla 15. Factores para la Gestión.....	54
Tabla 16. Actividades de la Metodología de la Investigación.....	56
Tabla 17. Identificación de Indicadores y Perspectivas	70
Tabla 18. Datos que quedan Excluidos del Análisis de esta Propuesta Teórica de Bodega de Datos (DW) y Motivos.	71
Tabla 19. Descripción Fuentes o Base de Datos de la Secretaria de Salud de Medellín.....	74
Tabla 20. Especificaciones Técnicas De Las Fuentes Y Sistemas De Información de La Secretaria De Salud	75
Tabla 21. Calculo de los Indicadores de la Secretaria de Salud de Medellín	76
Tabla 22. Perspectivas, Campos y Descripción.....	113

INDICE DE FIGURAS

Figura 1. Inteligencia de Negocios.....	17
Figura 2. Fases del proceso BI	21
Figura 3. Arquitectura de Data Warehouse.....	27
Figura 4. Data Warehouse Estructura.....	30
Figura 5. Figura Metodología Hefesto.....	33
Figura 6. Metodología Hefesto, pasos.	35
Figura 7. Clico Hefesto.	40
Figura 8. Tareas de la Metodología de Kimball, denominada Business Dimensional lifecycle (Kimball et al 98, 08, Mundy & Thornthwaite 06).	42
Figura 9. Ciclo CRISP.....	44
Figura 10. Fuentes de Información y Proceso del Sistema SAS.....	63
Figura 11. Proceso de Interacción para los Registros Individuales de prestación de servicios	66
Figura 12. Proceso Nacido vivo	68
Figura 13. Modelo Conceptual Información Secretaria de Salud.....	72
Figura 14. Diagrama Entidad Relación de los afiliados al SGSSS vs Modelo Conceptual	77
Figura 15. Diagrama Entidad Relación de Nacimientos vs Modelo Conceptual	78
Figura 16. Diagrama Entidad Relación de RIPS vs Modelo Conceptual.....	79
Figura 17. Modelo Conceptual Ampliado Información Secretaria de Salud	115
Figura 18. Población afiliada al Sistema General de Seguridad Social en Salud	119
Figura 19. Nacimientos atendidos en instituciones de salud de la población Medellín.....	120
Figura 20. Diez (10) primeros motivos más frecuentes de Morbilidad población Medellín.....	121

1 INTRODUCCION

Los avances tecnológicos y los altos volúmenes de datos han llevado a las organizaciones a reestructurarse, creándoles la necesidad de contar con herramientas que faciliten la obtención, procesamiento, análisis y consolidación de la información para la toma de decisiones; herramientas tales como sistemas de información integrados y articulados.

Las instituciones de salud generan con su actividad una enorme cantidad de datos, generalmente no estructurados, dispersos y de mala calidad, que no están a disposición, o que por lo menos no lo están oportunamente. Estos datos no constituyen información, y menos aún son herramientas de utilidad para la adopción de decisiones **(Vasquez, 2003)**.

Los sistemas de información para la gestión consisten en estructuras dedicadas al procesamiento sistemático de datos para transformarlos en información, sostenidas por herramientas informáticas, con el objetivo de apoyar los procesos organizacionales, en sus diferentes sectores (tanto asistenciales como económico-financieros). Están constituidos por un conjunto de personas, datos, formularios, equipamiento, programas de computación, procesos y procedimientos que funcionan articulados. Los “almacenes de datos” o data warehouses y las herramientas para el análisis de la información, llamadas herramientas OLAP, son sólo una parte, aunque importante, de estos sistemas. Un aspecto poco considerado inicialmente es la necesidad de contar con un equipo humano para el análisis de la información **(Vásquez, 2003)**.

Es por esto que la Secretaria de Salud de Medellín deben incursionar en la utilización de herramientas de Inteligencia de Negocios (BI) que permitan a los directivos poder acceder directamente la información que requieran, que faciliten la toma de decisiones, reportes en tiempo real, la implementación de estrategias y análisis de resultados de las mismas.

2 JUSTIFICACION

Entender y manejar la información es crucial para la organización, para hacer que su proceso de toma de decisiones sea efectivo y responda a las condiciones cambiantes que la afectan. Actualmente en la Secretaría de Salud de Medellín se cuenta con desarrollos aceptables en sus sistemas de información que les permiten atender los requerimientos básicos de información necesarios para el cumplimiento de sus competencias. Todas estas aplicaciones y módulos suministran datos e informes que es necesario consolidar y analizar en forma independiente pues hoy existen debilidades en la integración de dichas aplicaciones, lo que dificulta la construcción de diagnósticos y el análisis mismo de la información, haciéndose más lenta e insegura la toma de decisiones.

Con herramientas de Inteligencia de Negocios (BI) como el Data warehouse (bodega de datos) se mejora la productividad de la información que fluye en todos los procesos de la organización, promoviendo una gestión y una toma de decisiones proactiva, necesaria para aprovechar oportunidades futuras. Por tal motivo, esta propuesta pretende presentar algunas estrategias para la implementación de un modelo de Data Warehouse (Bodega de Datos) y particularmente varios Data Mart, que utiliza como eje articulador la identificación de las necesidades de información por parte de la dirección, para el soporte de los procesos de control y de toma de decisiones de la Secretaria de Salud del Municipio de Medellín.

3 PLANTEAMIENTO DEL PROBLEMA

La falta de eficiencia, equidad y calidad de la atención y los programas médicos constituyen uno de los principales limitantes del impacto de los programas de salud y principalmente en las ciudades. Estas carencias están presentes tanto en las ofertas públicas como en las privadas **(Arriada, 2005)**.

El Sistema de Seguridad Social en Salud, como uno de los sectores más importantes para lograr altos niveles de calidad de vida de la población debe suministrar permanentemente información en la que se dé cuenta del estado de salud y de los procesos de enfermedad de la población en general(ley 1438) Aunque la Secretaría de Salud de Medellín cuenta con desarrollos aceptables en sus sistemas de información que le permite atender los requerimientos básicos, el procesamiento y suministro de información se dificulta debido a que hoy cuenta con varias aplicaciones informáticas desarrolladas para cubrir las necesidades de los diferentes subprocesos.

Adicionalmente se le suma a esta dificultad los limitantes internos y externos que obligan a repensar permanentemente los aplicativos existentes, en la medida en que las normas que se expiden a nivel nacional cambian las obligaciones, los derechos, las estructuras de las bases de datos, los flujos de la información, los controles, entre otros; e internos por cuanto las políticas municipales frente a tecnología de información, suministro de recursos de software, hardware, y provisión de personal especializado está centralizado en la Secretaría de Servicios Administrativos, lo que dificulta y demora la solución de los problemas para la operación, mantenimiento y actualización de los sistemas de información.

Es por lo anterior que surge la necesidad imperiosa para la organización de incursionar en la Inteligencia de Negocios (BI) en herramientas como el Data Warehouse (Bodega de datos) que facilite la obtención, procesamiento, análisis y

consolidación de la información necesaria para la toma de decisiones, haciéndose necesario un diseño del modelo de Data Warehouse (bodega de datos) para la información que es manejada en la secretaria de Salud de Medellín, que visualice la necesidad de la herramienta a través de las ventajas que ofrecería a esta dependencia en cuanto a la estandarización y normalización del registro de la información, el almacenamiento, flujo, transferencia y disposición de la información para la difusión y utilización.

4 OBJETIVO GENERAL

Definición de estrategias para la implementación de una bodega de datos en la Secretaria de Salud que dé cuenta de la población y el panorama epidemiológico del municipio de Medellín para el año 2013.

5 OBJETIVOS ESPECIFICOS

- Identificar las diferentes fuentes de información y los diferentes sistemas utilizados actualmente por la secretaria de salud del municipio de Medellín para la toma de decisiones.
- Revisar en la literatura los diferentes enfoques y modelos propuestos para el diseño e implementación de un Data Warehouse en las organizaciones.
- Definir las estrategias para la implementación de la bodega de datos en la secretaria de salud de la ciudad de Medellín utilizando la metodología de Hefesto.
- Diseñar una bodega de datos que ofrezca información unificada y oficial sobre la situación de salud de los residentes de la ciudad que apoye la toma de decisiones en la Secretaria de Salud y que permita dar cuenta del panorama epidemiológico de Medellín.

6 MARCO TEORICO

6.1 CONCEPTOS Y FUNDAMENTOS

6.1.1 Datos

Los datos son la mínima unidad semántica, se corresponden con elementos primarios de información que por sí solos son irrelevantes como apoyo a la toma de decisiones, también son un conjunto de hechos objetivos y discretos acerca de ciertos acontecimientos. Los datos pueden provenir de fuentes externas o internas a la organización, pueden ser de carácter objetivo o subjetivo y de tipo cualitativo o cuantitativo **(Davenport & Prusak, 1998)**.

6.1.2 Información

La información se puede definir como un conjunto de datos procesados y que tienen significado, relevancia, propósito y contexto, por lo tanto son de utilidad para la toma de decisiones **(Davenport & Prusak, 1998)**, también es una interpretación de los datos basada en un cambio de condiciones y en el paso del tiempo **(Beazley et al, 2003)**. Por su parte **Martínez (2011)**, indica que la información está constituida por los datos y algunos hechos organizados, siendo necesario que sean puestos en un contexto para tener valor o relevancia.

6.1.3 Conocimiento

El conocimiento es una mezcla de experiencias, valores e información que sirven como marco para la incorporación de nuevas experiencias e información y con utilidad para la acción. Este conocimiento es derivado de la información así como estas de los datos. **Davenport & Prusak (1998)**, definen el conocimiento de manera práctica como una mezcla flexible de experiencias reflexionadas, de valores, de información contextual, que proporciona una estructura para evaluar e interiorizar nueva información y nuevas experiencias.

Algunas otras definiciones se pueden observar en la Tabla 1.

Tabla 1. Definiciones de conocimiento.

Autores	Año	Definición
Purser y Pasmore	1992	Hechos, modelos, conceptos, ideas e intuiciones que influyen en la toma de decisiones.
Davenport y Prusak	2000	Flujo mixto de experiencias, valores, información de contexto, percepciones de expertos y “saber hacer” que proporcionan un marco para la evaluación e incorporación de nuevas experiencias e información y es, por lo tanto, útil para la
Leonard y Sensiper	1998	Información relevante, procesable y basada, al menos parcialmente, en la experiencia (en el contexto empresarial)
Probst al et.	2001	Conjunto de cogniciones y habilidades con las cuales los individuos suelen solucionar problemas.

Fuente: Arceo (2009).

El conocimiento lo podemos clasificar en tácito y explícito.

Tabla 2. Tipos de conocimiento.

Tipos de conocimiento.	
Tácito	Aprendido gracias a la experiencia personal e involucra factores intangibles como las creencias, el punto de vista propio y los valores.
Explícito	Es aquel conocimiento que ha sido o puede ser articulado, codificado y almacenado en algún tipo de medio. Es un conocimiento formal y sistemático, que puede ser fácilmente comunicado y compartido.

Fuente: Nonaka & Takeuchi, (1995).

6.1.4 Inteligencia de negocios

Business Intelligence (BI) es el proceso de integración y tratamiento de los datos para convertirlos en información que permita apoyar a los tomadores de decisiones en la organización.

Davenport & Harris (2007) indican que la Inteligencia de Negocios consiste en la obtención, administración y reporte de la data orientada a la toma de decisiones, y las técnicas analíticas y procesos computarizados que se usan para el análisis de la misma.

Jourdan, Rainer & Marshall (2008), indican que la Inteligencia de Negocios es un proceso y un producto. El primero compuesto de métodos que las organizaciones usan para desarrollar información aplicable o inteligencia de

negocios, que les permita a las organizaciones salir adelante en un mundo muy competitivo y globalizado. Como producto es información que les permitirá a las organizaciones predecir el comportamiento de competidores, proveedores, clientes, tecnologías, adquisiciones, mercados, productos y servicios y el comportamiento en general del ambiente de negocios, con un cierto grado de precisión.

En la figura 1 se ilustra el porqué de la Inteligencia de Negocios.

Figura 1. Inteligencia de Negocios



Fuente: Inteligencia de Negocios (Introducción) por Elizabeht Leon Guzman – Profesora Universidad Nacional.

Algunas otras definiciones se pueden observar en la tabla 1.

Tabla 3. Definiciones de Inteligencia de Negocios (BI)

Tomado de:	Definición
Elizabeth León Guzmán - Profesora Universidad Nacional de Colombia	Es una alternativa tecnológica que permite generar conocimiento a la empresa. Es el uso de tecnologías para descubrir conocimiento a partir de los datos que posee la empresa
Gartner (consultora internacional especializada en Tecnologías de Información y Comunicación. www.gartner.com)	BI es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un datawarehouse), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones.
The datawarehouse Institute	Business Intelligence (BI) es un término paraguas que abarca los procesos, las herramientas, y las tecnologías para convertir datos en información, información en conocimiento y planes para conducir de forma eficaz las actividades de los negocios. BI abarca las tecnologías de datawarehousing los procesos en el 'back end', consultas, informes, análisis y las herramientas para mostrar información (estas son las herramientas de BI) y los procesos en el 'front end'.

Ing. Bernabeu Ricardo Dario	Se puede describir BI, como un concepto que integra por un lado el almacenamiento y por el otro el procesamiento de grandes cantidades de datos, con el principal objetivo de transformarlos en conocimiento y en decisiones en tiempo real, a través de un sencillo análisis y exploración.
-----------------------------	--

6.1.4.1 Proceso de inteligencia de negocios

Para comprender como una organización puede crear inteligencia de sus datos, se describirá a continuación el proceso de BI el cual está dividido en cinco fases:

Tabla 4. Fases del Proceso de Inteligencia de Negocios (BI)

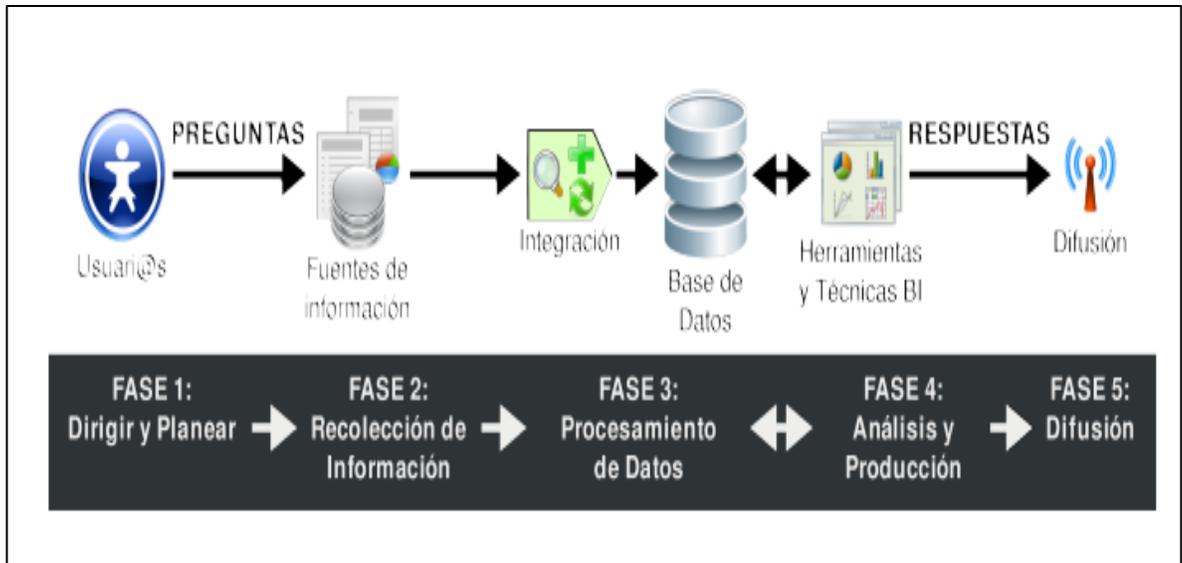
Fases de BI	
FASE 1	Dirigir y Planear. En esta fase inicial es donde se deberán recolectar los requerimientos de información específicos de los diferentes usuarios, entender sus diversas necesidades, para posteriormente generar las preguntas que les ayudarán a alcanzar sus objetivos.
FASE 2	Recolección de Información. Es aquí en donde se realiza el proceso de extraer desde las diferentes fuentes de información (internas, externas), los datos que serán necesarios para encontrar las respuestas a las preguntas planteadas en el paso anterior.
FASE 3	Procesamiento de Datos. En esta fase es donde se integran y cargan los datos en crudo en un formato utilizable para el análisis.

	Esta actividad puede realizarse mediante la creación de una nueva base de datos, agregando datos a una base de datos ya existente o bien consolidando la información.
FASE 4	Análisis y Producción. Se trabaja sobre los datos extraídos e integrados, utilizando herramientas y técnicas propias de la tecnología BI, para crear inteligencia. Como resultado final de esta fase se obtendrán las respuestas a las preguntas, mediante la creación de reportes, indicadores de rendimiento, cuadros de mando, gráficos estadísticos, etc.
FASE 5	Difusión. Se entrega a los usuarios que lo requieran las herramientas necesarias, que les permitirán explorar los datos de manera sencilla e intuitiva.

Fuente: Elaboración propia

En la figura 2 se puede ver el proceso de BI según las fases explicadas.

Figura 2. Fases del proceso BI



Fuente: (Bernabéu, 2010)

6.1.5 Data Warehouse

6.1.5.1 Definición

Es un conjunto de datos integrados u orientados a un objetivo específico, que varían con el tiempo (datos históricos) y que no son transitorios.

El concepto de Data Warehouse nace en la década de los 80 en la empresa IBM. Los investigadores que la desarrollaron buscaban establecer un sistema que permitiera un manejo fluido y permanente de datos que fuera al mismo tiempo organizado de acuerdo a necesidades específicas. Estos datos podían ser de muy variado tipo y abarcar diferentes áreas de la empresa. De este modo, no sólo se

ahorraría tiempo y energía sino también los altos costos monetarios de los sistemas previamente utilizados.

El sistema Data Warehouse o de almacén de datos se maneja a través de dos conceptos centrales:

- El de integración y combinación de diferentes tipos de datos que son utilizados en diferentes áreas y espacios de la organización.
- El de separación y selección de esa misma información de acuerdo a las necesidades específicas de cada usuario o sección de la empresa.

Normalmente, cuando se habla de Data Warehouse se hace referencia a un sistema que está organizado en base a temas o asignaturas especiales, que permite entonces que los datos y la información de mismo tipo quede siempre conectada. Del mismo modo, es un sistema que puede evolucionar con el tiempo y asimilar los cambios en la información de manera tal que cada nuevo acceso refleje las diferencias necesarias. También se define como un sistema de tipo volátil porque no se pierde nunca ningún tipo de información, lo cual hace más fácil recurrir a datos antiguos o que no estaban en uso.

Tabla 5. Definiciones de Data Warehouse

Autor	Definición
NCR	“Es un proceso donde está integrada y estandarizada la información de la operación de la empresa en una bodega de datos centralizada con el fin de hacer real el sueño de tener decisiones basados en datos y hechos, en forma autónoma y confiable”.

Robert Orfali	“Es un almacén de datos inteligente que puede almacenar información de varias fuentes, distribuirla cuando se requiera y activar políticas de negocios”.
---------------	--

Los Data Warehouse son herramientas poderosas para la toma de decisiones en las organizaciones, permiten la organización y consolidación de la información de las empresas, en la tabla 3 se listan las principales características de estos sistemas

Tabla 6. Características de Data Warehouse

Características de Data Warehouse	
Integra información	Permite consolidar información de muy diversas fuentes con la ventaja de datos históricos que ya no son útiles en los sistemas operacionales pero sirven para análisis de escenarios.
Está dirigido al usuario	Estas aplicaciones están dirigidas la alta gerencia o dueños de negocio y a todas la personas encargadas de tomar decisiones
Evoluciona con el tiempo	Son herramientas que tendrán que cambiar con el tiempo, tendrán que adaptarse a requerimientos más precisos y mayores volúmenes de información
Está orientado a la toma de decisiones	Un data warehouse se construye para la consolidación de información de distintas fuentes en una sola y con ello apoyar la toma de decisiones
Su uso está enfocado a la alta gerencia	No está enfocado a los procesos operativos sino a la toma de decisiones por parte de los lideres o dueños del negocio

Fuente: Inteligencia de negocios (vega, 2010)

6.1.5.2 Objetivos fundamentales de un Data Warehouse

Un Data Warehouse es un seguro para proteger toda la información de la organización, de forma que esta quede accesible, entendible, estructurada y completa, la modelización de toda la información dispersa en una organización en un DW centralizado, ofrece información consistente y homogénea para la toma de decisiones, en la siguiente tabla se listara las principales características que tiene los Data Warehouse en las organizaciones.

Tabla 7. Objetivos del Data Warehouse

Hacer que la información de la organización sea accesible	Sus contenidos son entendibles y navegables, y de acceso y desempeño rápido.
Hacer que la información sea consistente	La información debe ser consistente y de alta calidad, debe ser contabilizada y completada.
Información adaptable y elástica	Diseñado para cambios continuos, debe ser adaptable a nuevos requerimientos y a nuevos bloques de datos.
Lugar seguro para proteger la información	Controlar el acceso efectivo a los datos, visibilidad de la información y seguridad de la información mostrada, con perfiles y roles
Toma de decisiones	Tiene los datos correctos para soportar la toma de decisiones

Fuente: Elaboración propia

6.1.5.3 Elementos básicos de un Data Warehouse

Reunir los elementos de datos apropiados desde diversas fuentes de aplicación en un ambiente integral centralizado, simplifica el problema de acceso a la información y en consecuencia, acelera el proceso de análisis, consultas y el menor tiempo de uso de la información.

A continuación se lista algunos elementos básicos del Data Warehouse

Tabla 8. Elementos Básicos del Data Warehouse

Sistema fuente	Sus funciones son capturar las transacciones del negocio
Área de tráfico de datos	En esta área se limpian, transforman, combinan. Remueven duplicados, guardan , archivan y se preparan los datos fuentes para ser usados en el Data Warehouse
Servidor de presentación	El lugar físico donde los datos del Data Warehouse son organizados y almacenados para queries directos del usuario final
Modelo dimensional	Modelado de los datos
Procesos de negocio	Grupo de actividades de negocio aplicados al Data Warehouse
Data Mart	Es un subgrupo lógico de un Data Warehouse
Data warehouse	Almacén de datos de la empresa
Almacenamiento operacional de datos	Punto de integración para los sistemas operacionales del negocio
OLAP	Estructura multidimensional que contiene datos

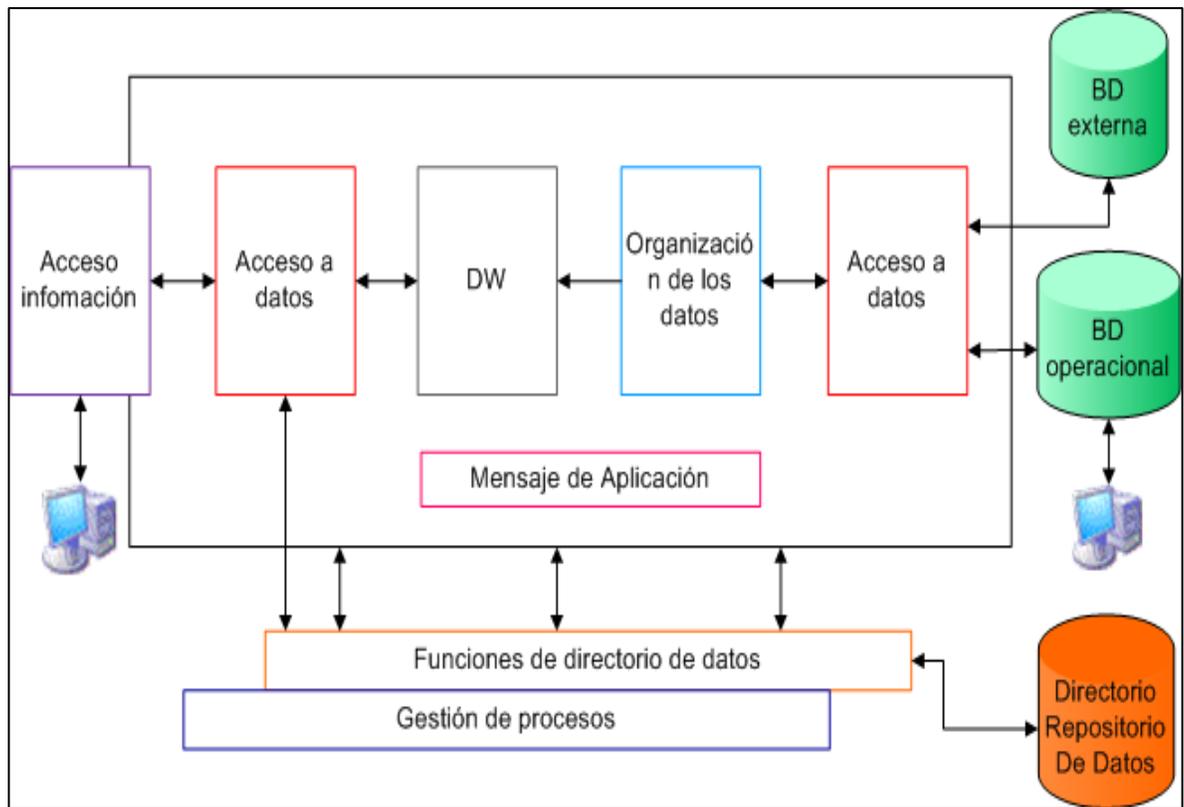
	resumidos de grandes base de datos con el objetivo de agilizar la consulta de grandes cantidades de datos
ROLAP	Sistema OLAP construido sobre un sistema relacional clásico
MOLAP	A diferencia del ROLAP este necesita un almacenamiento previo en un sistemas multidimensional
Aplicación para usuarios finales	Una colección de herramientas que permiten la consulta, el análisis y la presentación de la información para el apoyo a la toma de decisiones
Modelado de aplicaciones	Un cliente de Data Warehouse con capacidades analíticas que transforma y digiere las salidas del Data Warehouse

Fuente: Elaboración propia

6.1.5.4 Arquitectura del Data Warehouse

Una arquitectura de Data Warehouse es una forma de representar la estructura global de los datos, las comunicaciones, los procesos y la presentación al usuario final. La arquitectura está constituida por las siguientes partes interconectadas.

Figura 3. Arquitectura de Data Warehouse



Fuente: (Bernabéu, 2010)

6.1.5.5 Elementos que constituyen la Arquitectura de un DW

6.1.5.5.1 Base de datos operacional y externa

Las organizaciones adquieren datos de bases de datos externas a la propia organización, que incluyen datos demográficos, económicos, datos sobre la competencia.

6.1.5.5.2 Nivel de acceso a la información.

ES la capa con la que trata el usuario final. La información almacenada se convierte en información fácil y transparente para las herramientas que utilizan los usuarios. Se obtienen informes, gráficos, diagramas etc.

6.1.5.5.3 Nivel de acceso a datos.

Comunica el nivel de acceso a la información con el nivel operacional, es el responsable de la interfaz entre las herramientas de acceso a la información y las bases de datos.

6.1.5.5.4 Nivel de directorio de datos (Metadatos).

Directorio de datos o repositorio de información de metadato que ayuda a mantener el control del dato. El metadato aporta información sobre los datos de la organización, de donde proviene, que formato tenía, cuál era el significado y si se trata de un agregado y como se ha calculado este.

6.1.5.5.5 Nivel de gestión de procesos.

Este nivel tiene que ver con la planificación de las tareas que se deben realizar, no solo para construir, sino también para mantener el Data Warehouse y la información del directorio de datos, mantiene control sobre los procesos que actualizan el Data Warehouse.

6.1.5.5.6 Nivel de mensaje de la aplicación.

Este nivel es el encargado del transporte de la información a lo largo del entorno, se puede pensar en el como un middleware.

6.1.5.5.7 Nivel Data Warehouse (físico).

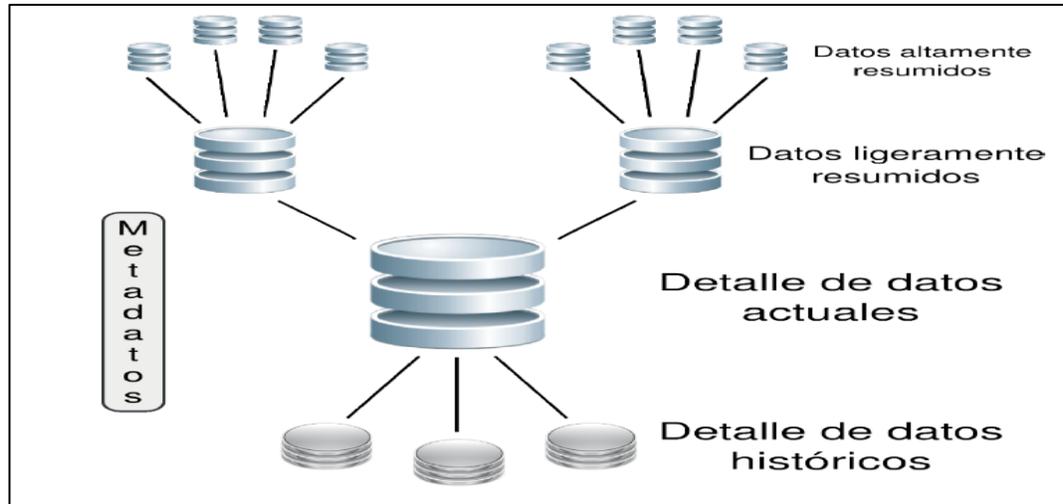
Es el núcleo del sistema, el repositorio central de información donde los datos actuales usados principalmente con fines informacionales residen. En este se almacenan copias de datos operacionales y/o externos, en una estructura que optimiza su acceso para la consulta y que es muy flexible.

6.1.5.5.8 Nivel de organización de datos.

Incluye todos los procesos necesarios para seleccionar, editar, resumir, combinar y cargar en el DW y en la capa de acceso a la información los datos operacionales y/o externos.

6.1.5.6 Estructura de un Data Warehouse

Figura 4. Data Warehouse Estructura



Fuente: (Buiges, 2010)

La estructura de un DW se caracteriza por los diferentes niveles de esquematización y detalle de los datos que se encuentran en él.

Tabla 9. Estructura de un Data Warehouse

Detalle de los datos actuales	Reflejan los acontecimientos más recientes, las últimas informaciones generadas por los sistemas de producción de la organización.
Detalle de datos antiguos	Contiene los datos almacenados anteriormente,

	como repositorios de datos.
Datos resumidos	Son datos obtenidos como resultado de un proceso de síntesis de los datos actuales, datos agregados o resumidos.
Metadatos	Ofrecen información descriptiva sobre el contexto, la calidad, la condición y las características de los datos.

Fuente: Elaboración propia

6.1.5.7 Base de datos transaccional (OLTP) vs Data Warehouse

Los OLTP son diseñadas para soportar el procesamiento de información diaria de las empresas, y el énfasis recae en máxima la capacidad transaccional de sus datos, son de estructura altamente normalizada, para brindar eficiencia a sistemas de muchas transacciones, en cambio las DW están diseñadas para poder llevar a cabo procesos de consulta y análisis para luego tomar decisiones estratégicas y tácticas de alto nivel.

Tabla 10. OLTP vs DW

	OLTP	Data Warehouse
Objetivo	Soportar actividades transaccionales diarias	Consultar y analizar información estratégica y táctica
Tipo de datos	Operacionales.	Para la toma de decisiones.
Modelo de datos.	Normalizado.	Desnormalizado.

Consulta	SQL.	SQL más extensiones.
Datos consultados	Actuales	Actuales e históricos.
Horizonte de tiempo	60-90 días.	5-10 años.
Tipos de consultas	Repetitivas, predefinidas	No previsibles, dinámicas
Nivel de almacenamiento	Nivel de detalle	Nivel de detalle y diferentes niveles de sumarización
Acciones disponibles	Alta, baja, modificación y consulta	Carga y consulta
Número de transacciones	elevado	Medio o bajo
Tamaño	Pequeño – Mediano.	Grande.
Tiempo de respuesta	Minutos- segundos	Minutos- horas
Orientación	Orientado a las aplicaciones	Orientado al negocio
Sello de tiempo	Puede o no depender del tiempo	Depende totalmente del tiempo
Estructura	estable	Varía de acuerdo con su evolución y utilización

Fuente: Elaboración propia

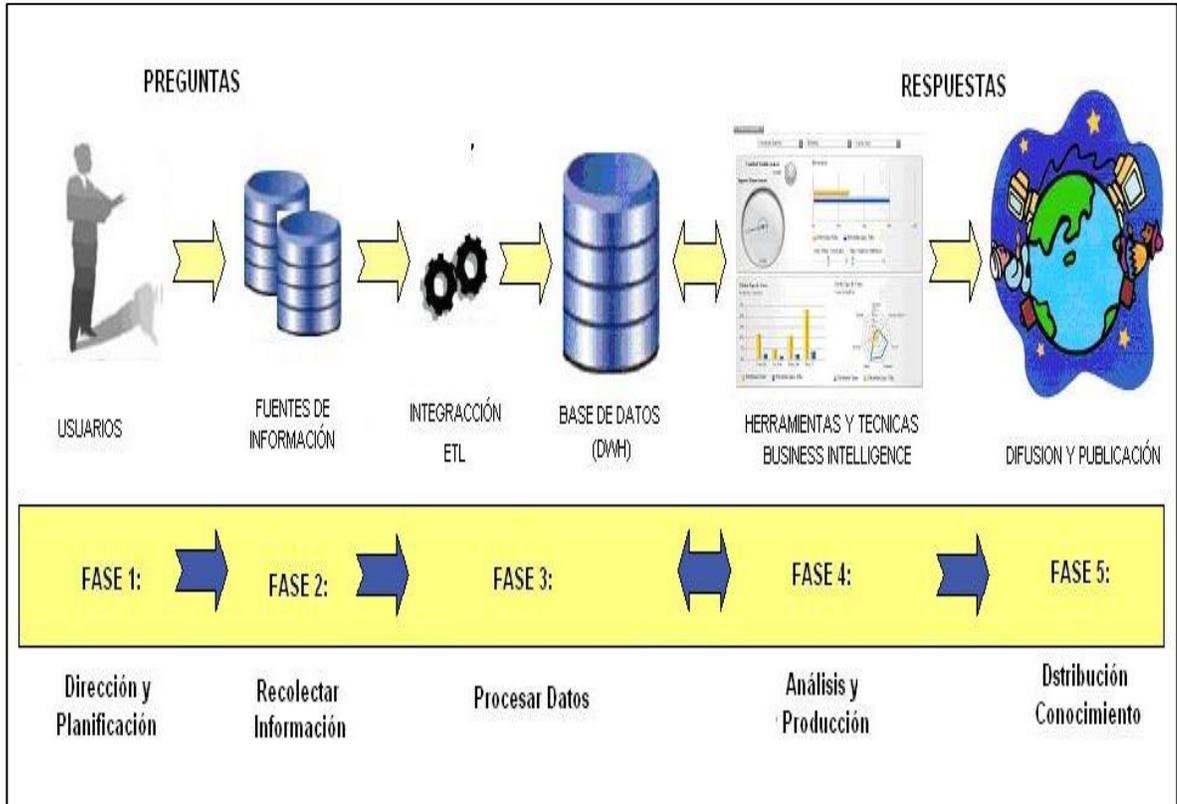
6.1.6 Metodología Hefesto

6.1.6.1 Introducción

Hefesto es una metodología cuya propuesta está fundamentada en una amplia investigación, comparación de metodologías existentes, experiencias propias en procesos de confección de almacenes de datos (Bernabéu, 2010).

La metodología Hefesto puede resumirse a través de la siguiente figura:

Figura 5. Figura Metodología Hefesto



Fuente: (Bernabéu, 2010)

6.1.6.2 Principales características

El principal objetivo de HEFESTO es facilitar el arduo trabajo que significa construir un Data WareHouse desde cero, aportando información que permitirá aumentar el rendimiento del mismo. En adición a ello, esta nueva metodología estará orientada a evitar el tedio que provoca el tener que seguir pasos sin terminar de comprender el porqué de los mismos.

A continuación se listan las principales características que tiene la metodología:

Características
Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender
Se basa en requerimientos de los usuarios, sus estructura es adaptable a los cambios del negocio
Reduce la resistencia al cambio, ya que involucra a los usuarios finales en cada etapa para que tomen decisiones sobre el DW
Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
Es independiente del tipo de ciclo de vida en el que se emplee la metodología
Es independiente de la herramienta utilizada
Es independiente de las estructuras físicas que contengan el Data Warehouse
Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para el paso siguiente
Es aplicable tanto para Data Warehouse como para Data Marts

6.1.6.3 Fases de la metodología

La metodología está orientada a la construcción de DW para análisis dimensional (OLAP), esta propone cuatro pasos fundamentales para su aplicación, cada uno con una serie de puntos que se deben realizar y que se resumen en el siguiente gráfico.

Figura 6. Metodología Hefesto, pasos.



Fuente: (Bernabéu, 2010)

6.1.6.3.1 Fase1: Análisis de requerimientos

- Indicar preguntas

El primer paso comienza con el estudio de las necesidades de información, el cual puede llevarse a cabo a través de varias y diferentes técnicas, cada una de las cuales poseen características inherentes y específicas, como entrevistas, cuestionarios, observaciones etc.

El objetivo principal es obtener e identificar las necesidades de información de alto nivel, que servirán para lograr las metas y estrategias y apoyaran la toma de decisiones.

- Identificar indicadores

Una vez que se han establecido las preguntas claves, se debe proceder a su descomposición para descubrir los indicadores que se utilizaran y las perspectivas de análisis que intervendrán. Para ello, se debe tener en cuenta los indicadores, para que sean realmente efectivos y deben representar lo que se desea analizar concretamente.

- Modelo Conceptual

Se construirá un modelo conceptual a partir de los indicadores y perspectivas obtenidas en el paso anterior. A través de este modelo, se podrá observar con la claridad cuáles son los alcances del proyecto y trabajar sobre ellos, y gracias a su alto nivel de definición puede ser explicado y presentado a los usuarios

6.1.6.3.2 Fase 2: Análisis de los OLTP

- Establecer correspondencias con los requerimientos

El objetivo de este análisis, es el de examinar los OLTP disponibles que contengan la información, como así también sus características, para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos. En el caso de los indicadores, deben explicitarse como se calcularán, y más aún si son fórmulas u operaciones complejas.

- Establecer correspondencia

Una vez que se han establecido las relaciones con los OLTP, se examinarán y seleccionarán los campos que contendrá cada perspectiva, ya que será a través de estos por los que se manipularán y filtrarán los indicadores.

Para ello, basándose en las correspondencias establecidas en el paso anterior, se debe presentar al usuario los datos de análisis disponibles para cada perspectiva. Es muy importante conocer en detalle que significa cada campo y/o valor de los datos encontrados en los OLTP, por lo cual, es conveniente investigar su sentido, ya sea a través de diccionarios de datos, reuniones con los encargados del sistema, análisis de los datos propiamente dichos, etc.

6.1.6.3.3 Fase 3: Modelo lógico de los DW

- **Diseñar tablas de dimensiones**

Este paso, se aplicará por igual a todos los tipos de esquemas lógicos.

Lo primero que se hará será crear las dimensiones del mismo, para ello se tomará cada perspectiva con sus atributos relacionados y se les realizará el siguiente proceso:

- Se elegirá un nombre que identifique la dimensión.
- Se añadirá un campo que represente su clave principal.
- Se redefinirán los nombres de los atributos si es que no son lo bastante explicativos.

- **Diseñar tablas de hechos**

En este paso, se definirán las tablas de hechos, que son las que contendrán los indicadores de estudio. Para los esquemas en estrella y copo de nieve, se realizará lo siguiente:

- Al igual que las dimensiones, se le deberá asignar un nombre a la tabla de hechos que en este caso represente la información analizada, área de investigación, negocio enfocado, etc.
- Se definirá su clave primaria, que se compone de la combinación de las claves primarias de cada dimensión que se utilizará para generar las consultas.
- Se renombrarán los hechos o indicadores si es que no llegasen a ser lo suficientemente explícitos.

- **Determinar jerarquías**

Para los esquemas en estrella y constelación, se deberán especificar las jerarquías que existirán dentro de cada tabla de dimensión, teniendo siempre presente cual es el objetivo de las mismas. Para representar las jerarquías en el modelo lógico, se deberán colocar los atributos pertenecientes a las jerarquías en sus respectivas tablas, en orden descendente y acompañado con un número ordinal encerrado entre corchetes.

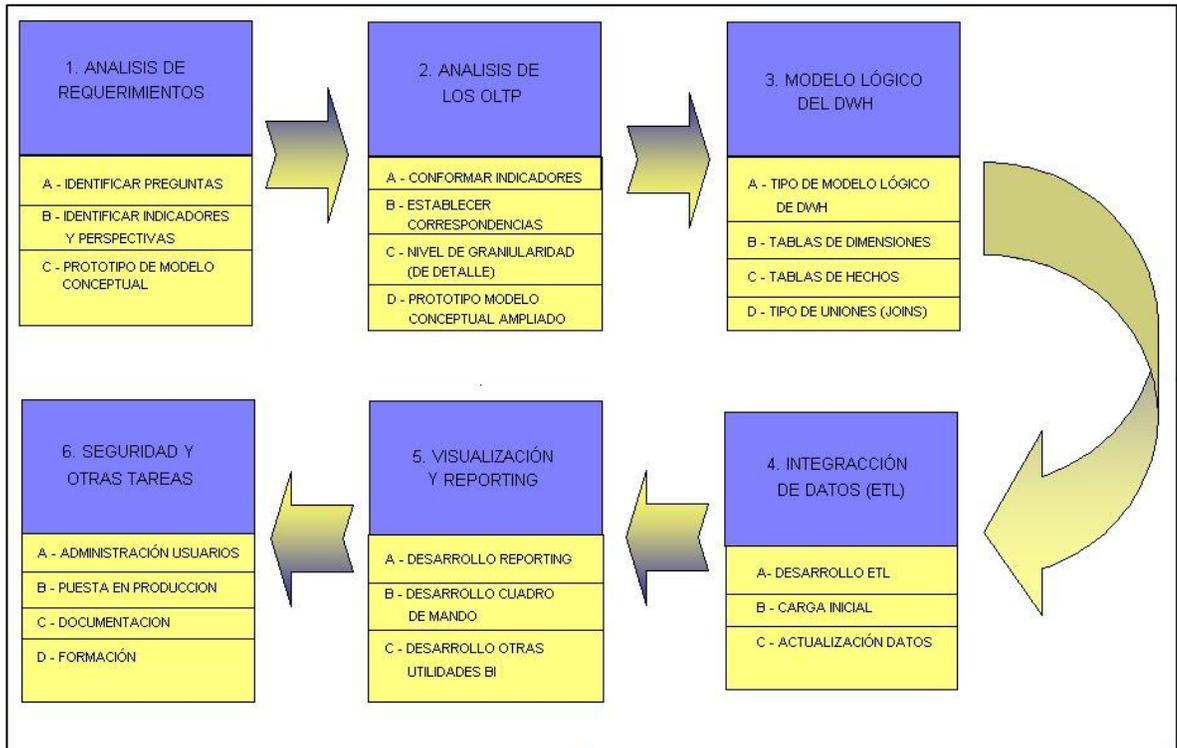
- **Realizar uniones**

Para los tres tipos de esquemas, se realizarán las uniones correspondientes entre sus tablas de dimensiones y sus tablas de hechos.

6.1.6.3.4 Fase 4: Integración de Datos

Una vez construido el modelo lógico, se deberá proceder a probarlo con datos, a través de procesos ETL. Para realizar la compleja actividad de extraer datos de diferentes fuentes, para luego integrarlos, filtrarlos y depurarlos, existen varios software que facilitan estas tareas, por lo cual este paso se centrará solo en la generación de las sentencias SQL que contendrán los datos que serán de interés.

Figura 7. Clico Hefesto.



Fuente: (Bernabéu, 2010)

6.1.7 Otras metodologías

6.1.7.1 Metodología Kimball

La metodología se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle) (Kimball et al 98, 08, Mundy & Thornthwaite 06). Este ciclo de vida del proyecto de DW, se basa en cuatro principios básicos:

- ✓ Centrarse en el negocio: Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.

- ✓ Construir una infraestructura de información adecuada: Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.

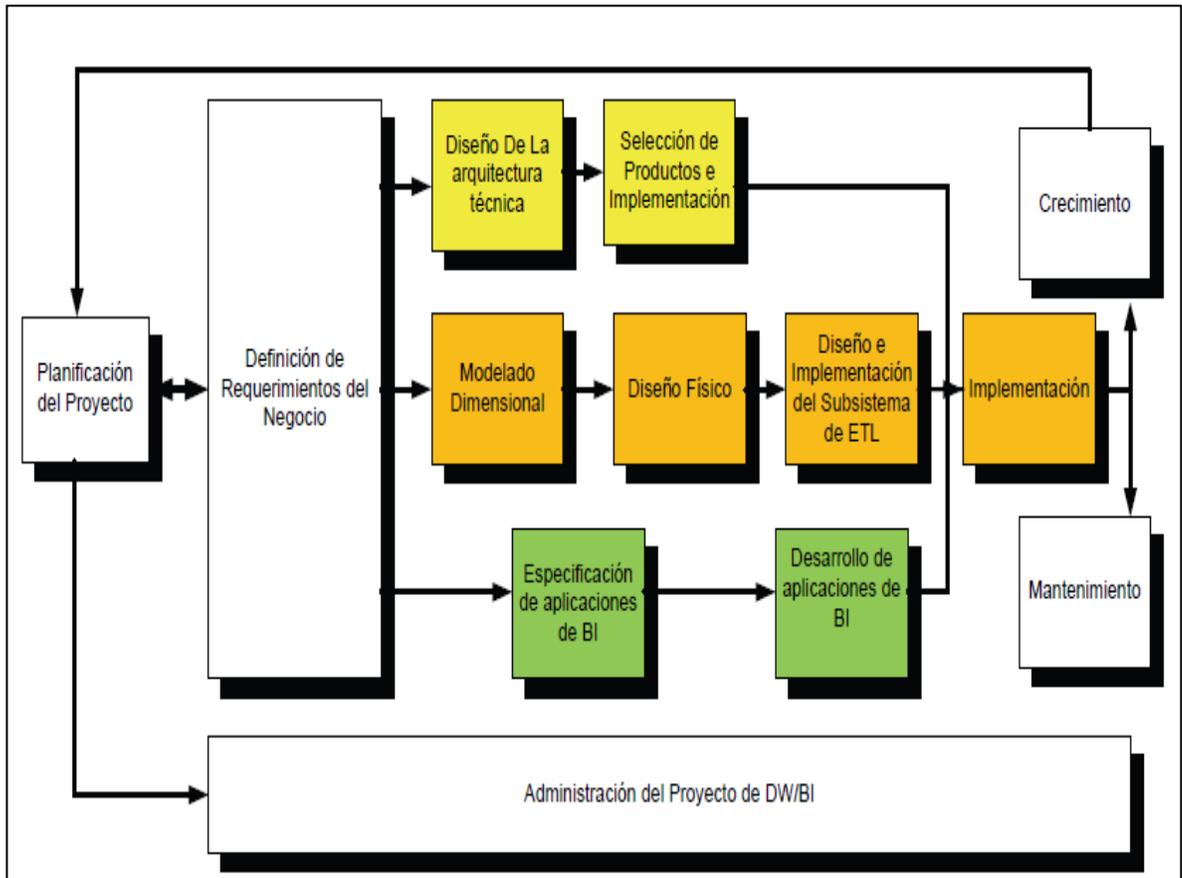
- ✓ Realizar entregas en incrementos significativos: crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos. En esto la metodología se parece a las metodologías ágiles de construcción de software.

- ✓ Ofrecer la solución completa: proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. También se deberá entregar herramientas de consulta ad hoc, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.

La construcción de una solución de DW/BI (Datawarehouse/Business Intelligence) es sumamente compleja, y Kimball nos propone una metodología que nos ayuda a simplificar esa complejidad.

Las tareas de esta metodología (ciclo de vida) se muestran en el siguiente gráfico:

Figura 8. Tareas de la Metodología de Kimball, denominada Business Dimensional lifecycle (Kimball et al 98, 08, Mundy & Thornthwaite 06).



Fuente: Artículo Kimball y inmon (Gustavo R. Rivadera 2010)

Como se puede apreciar se presentan dos cuestiones. Primero, resalta el rol central de la tarea de definición de requerimientos. Los requerimientos del negocio son el soporte inicial de las tareas subsiguientes. También tiene influencia en el plan de proyecto (nótese la doble flecha entre la caja de definición de requerimientos y la de planificación).

En segundo lugar podemos ver tres rutas o caminos que se enfocan en tres diferentes áreas:

- ✓ Tecnología (Camino Superior). Implica tareas relacionadas con software específico, por ejemplo, Microsoft SQL Analysis Services.
- ✓ Datos (Camino del medio). En la misma se diseña el modelo dimensional, y se desarrolla el subsistema de Extracción, Transformación y Carga (Extract, Transformation, and Load - ETL) para cargar el DW.
- ✓ Aplicaciones de Inteligencia de Negocios (Camino Inferior). En esta ruta se encuentran tareas en las que se diseña y desarrolla las aplicaciones de negocios para los usuarios finales.

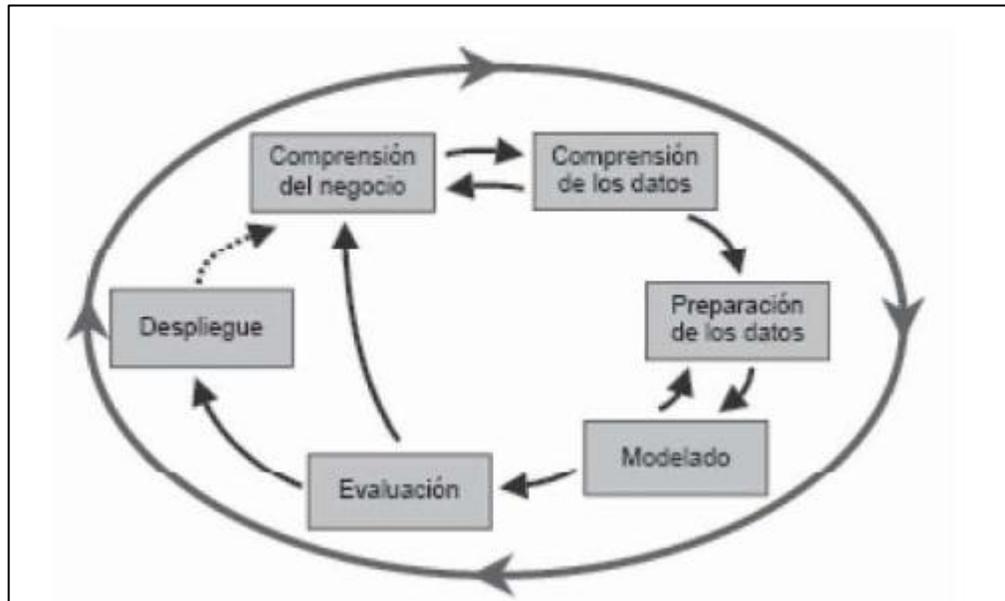
Estas rutas se combinan cuando se instala finalmente el sistema. En la parte de debajo de la figura se muestra la actividad general de administración del proyecto. La metodología de Kimball proporciona una base empírica y metodológica adecuada para las implementaciones de almacenes de datos pequeños y medianos, dada su gran versatilidad y su enfoque ascendente, que permite construir los almacenes en forma escalonada. Además presenta una serie de herramientas, tales como planillas, gráficos y documentos, que proporcionan una gran ayuda para iniciarse en el ámbito de la construcción de un Data warehouse.

6.1.7.2 Metodología CRISP

La Metodología CRISP está sustentada en estándares internacionales que reflejan la robustez de sus procesos y que facilitan la unificación de sus fases en una estructura confiable y amigable para el usuario. Además de ello, esta tecnología interrelaciona las diferentes fases del proceso entre sí, de tal manera que se consolida un proceso interactivo y recíproco. Otro aspecto fundamental de esta metodología es que es planteada como una metodología imparcial o “neutra

respecto a la herramienta que se utilice para el desarrollo del proyecto de Data Warehouse o Data Mining siendo su distribución libre y gratuita”.

Figura 9. Ciclo CRISP



Fuente: Metodología CRIPS para la implementación Data Warehouse (Octavio J. Salcedo Parra/ Rita Milena Galeano/ Luis G Rodriguez. 2009)

Como se puede apreciar el ciclo de vida según la metodología CRISP está basado en seis fases cambiantes entre si y nunca terminantes, lo cual lo postula como un ciclo en constante movimiento.

Se describe a continuación de manera general las fases de este ciclo:

- **Comprensión del negocio:** se trata de entender claramente los requerimientos y objetivos del proyecto siempre desde una visión de negocio.
 - Definición de los objetivos de negocio
 - Evaluación de la situación

- Definición de los objetivos del Data warehouse
 - Realización del plan del proyecto (plan del proyecto y valoración inicial de herramientas y técnicas).
- Comprensión de los datos: es conseguir y habituarse con los datos, reconocer las dificultades en la calidad de los datos y reconocer también las fortalezas de estos mismos que puede servir en el proceso de análisis.
- Recolección inicial de datos
 - Descubrimiento de los datos
 - Exploración de los datos
 - Verificación de la calidad de los datos.
- Preparación de los datos: es analizar los datos realmente importantes en el proceso de selección, depuración y transformación.
- Selección de los datos
 - Depuración de los datos
 - Estructuración de los datos
 - Integración de los datos
 - Formateo de datos
- Modelado: es la aplicación de técnicas de modelado o de Data Warehouse.
- Selección de la técnica de modelado
 - Generar el plan de pruebas
 - Construcción del modelo
 - Evaluación del modelo
- Evaluación: esta fase es muy importante y decisiva, pues corresponde a la evaluación de la escogencia de los modelos anteriores y la toma de decisión respecto a si realmente son útiles en el proceso.
- Evaluar resultados

- Proceso de revisión
- Determinación de los pasos siguientes

- Despliegue o divulgación: es la fase de implementación o de divulgación de los modelos anteriormente escogidos y evaluados.
 - Plan de divulgación o implementación
 - Plan de monitoreo y mantenimiento
 - Presentación del informe final
 - Revisión del proyecto

7 DEFINICION DE ESTRATEGIAS A CONSIDERAR EN LA IMPLEMENTACIÓN DE UN DW

En este capítulo se evalúan aspectos como el desarrollo del proyecto, el cual para este trabajo se ha definido por fases según las necesidades de información y el manejo de los procesos en la Secretaria de Salud de Medellín. Toda vez, que la Secretaria cuenta hoy con varias aplicaciones informáticas desarrolladas para cubrir las necesidades de los diferentes subprocesos, la gran mayoría en lenguaje Visual Fox manejadas en motor de base de datos de SQL server, existen también aplicaciones en Access y se manejan herramientas de escritorio que facilitan el manejo de la información y de archivos para la generación de informes como el Excel.

Por lo anterior se muestra en la siguiente tabla los elementos y estrategias que se deben considerar en el proceso de implementación de un Data Warehouse:

Tabla 11. Elementos para la Implementación de un Data Warehouse	
Proyecto total o Proyecto en Fases	Es más viable el desarrollo de un proyecto en fases que produzca resultados a corto plazo que desarrollar un proyecto que entregue resultados al término de varios años. Por tanto el proyecto debe centrarse en un área o un proceso.
Modelo lógico de datos	El modelo lógico de datos debe tener un alcance más alto y cubrir las áreas de interés, así como los procesos más estratégicos de cada una de ellas.
Proyecto especializado o Proyecto Base	El proyecto especializado soporta un proceso específico. Un proyecto base entrega capacidad genérica de análisis a los usuarios que

	<p>tengan acceso al data warehouse, pero no tiene entre sus funcionalidades, la solución de un problema específico o el soporte especializado de un proceso específico.</p> <p>El proyecto base es más económico y fácil de acabar que uno especializado el cual es más costoso y difícil de terminar.</p>
Estrategias para el proceso de implementación de un DW	
Identificación del Problema	Identificar el problema en el cual el uso de la información detallada, permita conseguir una solución que genere una ventaja competitiva o un ahorro de costos.
Definir el Modelo Lógico	<p>Definir el Modelo Lógico de datos a implementar para resolver el problema planteado.</p> <p>El Modelo Lógico puede darse cuando se presenta al usuario la información en términos de dimensiones y hechos que se registrarán para estas dimensiones.</p>
Reunir Datos	Reunir los datos para poblar el modelo lógico de datos.
Complementación de Información	Tomar iniciativas de complementación de información para asegurar la calidad de los datos requeridos para poblar el modelo de datos.
Estrategias para la Implementación de un DW	
Definir el diseño físico para el modelo de datos	El diseño físico debe estar orientado a generar buen rendimiento en el proceso de consultas, a diferencia del modelo lógico que está orientado al

	usuario y a la facilidad de consulta.
Definir procesos ETL	Definir los procesos de extracción, filtro, transformación de información y carga de datos que se deben implementar para poblar el modelo de datos.
Definir procesos de administración	Definir los procesos de administración de la información que permanece en el data warehouse.
Definir consultas	Definir las formas de consultas a la información del data warehouse que se le proporcionara al usuario. Debe considerarse la necesidad de resolver un problema y la potencia de consulta.
Completar el modelo	Completar el modelo de consulta base relativo al área seleccionada.
Implementar los proceso estratégicos	Implementar los procesos estratégicos del área de trabajo, es decir implementar herramientas especializadas de scoring, herramientas especializadas para inducción de conocimiento (data mining).
Completar áreas de interés	Completar las áreas de interés en forma similar a lo descrito en el punto anterior.

Fuente: Elaboración propia

En la actualidad no existe una manera sencilla de diseñar e implementar una bodega de datos en una organización y garantizar el éxito en su construcción inmediatamente, pero hay muchos aspectos a tener en cuenta para que contribuyan a una implementación, menos difícil y se pueda llevar a feliz término su construcción.

Como parte de las estrategias que se deben tener en cuenta para la implementación, se mencionan en la siguiente tabla puntos clave que se deben considerar en la planificación de un Data Warehouse.

Tabla 12. Factores de Planificación	
Asociaciones entre usuarios, gestión y grupos	Para establecer estas relaciones, es esencial involucrar a los usuarios como a la gestión para asegurar que el Data Warehouse contenga información que satisfaga todos los requerimientos de la organización, esto ayudaría a la priorización del proyecto en la fase de implementación.
Aplicación de un piloto con probabilidad de éxito	Se debe tener en cuenta la realización de pruebas piloto, que contengan un alcance limitado y medible, para ser representadas en la organización, de esta manera poder establecer el Data Warehouse como una tecnología esencial en la organización. Estos mismos criterios (alcance, costos, beneficios) serán aplicados a cada fase de la implementación del Data Warehouse.
Construcción de entregables sencillos y periódicamente	El objetivo principal del DATA WAREHOUSE es que reúna las necesidades de los usuarios, para asegurar esto debe garantizarse la elaboración de prototipos a los largo de todo el procesos de implementación, el

	trabajo continuo entre usuarios y la gestión, es el aspecto clave.
Implementación por fases incrementales	Es importante que la implementación de la Data Warehouse se haga por fases y de manera incremental, ya que así se reducen los riesgos y se asegura que el proyecto permanezca manejable en cada una de estas fases.
Reporte y publicación de casos de éxito	La retroalimentación constante de los usuarios permitirá la publicación de los eventos de éxito en la organización, las experiencias de los usuarios sobre el cómo han sido ayudados por el Data Warehouse, puede apoyar la construcción del Data Warehouse a lo largo de toda la organización.

Fuente: Elaboración propia

Antes de desarrollar un Data warehouse dentro de la organización es necesario el planteamiento de factores que sirvan de estrategia equilibrada y apropiada para los usuarios y las necesidades de esta. De estos factores surgen preguntas como el público objetivo, el alcance y el tipo de Data Warehouse que debe construirse como se muestra en la tabla.

Tabla 13. Factores de desarrollo	
Usuarios	Seleccionar los usuarios basados en la importancia de la información que manejan para la organización y hacer un análisis de sus preguntas, requerimientos y necesidades de acceso a datos
Prototipos	Con las necesidades recopiladas con los usuarios se construyen prototipos para que los usuarios finales puedan experimentar y modificar sus requerimientos.
Datos	Partiendo de las necesidades recopiladas, se consiguen datos provenientes de los sistemas operacionales existentes en la empresa o desde fuentes externas y se cargan a la DATA WAREHOUSE .
Acceso a la información	Se utilizaran herramientas de acceso a la información utilizando sus herramienta favoritas o proporcionando sistemas de acceso a la información multidimensional y de alto performance, usando el núcleo de la Data Warehouse como base

Fuente: Elaboración propia

Actualmente no se tiene un enfoque único, ni un único grupo de estrategias para la construcción de un Data Warehouse que se adapte a las necesidades de la

organización, debido a que las necesidades de cada una de ellas son diferentes, al igual que su contexto.

El diseño de los Data Warehouse es muy diferente al diseño de los sistemas operacionales que normalmente funcionan en una organización, es por eso que se deben tener en cuenta varios factores como se muestra en la tabla que son parte de las estrategias de diseño.

Tabla 14. Factores Para el Diseño	
Usuario	Los usuarios de la Data Warehouse usualmente no conocen ni comparten los requerimientos y necesidades de los usuarios operacionales, es por esto que se deben crear grupos de trabajo entre estos usuarios para socializar y agrupar las necesidades en el Data Warehouse
Diseño	El diseño del Data Warehouse, involucra con frecuencia ámbitos globales y visiones macro de la organización, diferente a lo que hacen los sistemas operacionales. Un Data Warehouse está bastante cerca de la reingeniería de los procesos de negocio. Por eso la socialización y la formación de la cultura empresarial sobre el proyecto es fundamental.
Estrategia	A pesar que el diseño de la Data warehouse es diferente al usado en los diseños tradicionales, no es menos importante. El hecho que los usuarios finales tengan dificultad en definir lo que ellos necesitan, no lo hace menos necesario. En la práctica, los diseñadores de Data Warehouse deben poder ayudar a sus usuarios a visualizar sus necesidades y es por eso que son importantes los prototipos de trabajo.

Fuente: Elaboración propia

Después de que un Data Warehouse sea diseñado e implementado en la organización, viene una etapa no menos importante que la anterior, la cual es la gestión del Data Warehouse y su mantenimiento en el tiempo, solo un Data Warehouse será rentable si está constantemente en depuración y en mantenimiento, y sus datos siempre son de calidad, para esto se debe contar con los siguientes factores que se relacionan en la tabla de las estrategias implementadas para su gestión.

Tabla 15. Factores para la Gestión
Un Data Warehouse es una inversión buena solo si los usuarios finales realmente pueden conseguir información vital más rápida y barata de lo que la obtienen con la tecnología y sus temas actuales con los que cuentan.
La administración debe reconocer que el mantenimiento de la estructura del Data Warehouse es tan crítico como el mantenimiento de cualquier otro sistema de la organización, de hecho los Data Warehouse llegarán a ser unos de los sistemas más usados en las organizaciones.
<p>En el futuro sus sistemas operacionales generará</p> <ul style="list-style-type: none"> • Demandas para mejora de datos • Demanda para datos consistentes • Demanda para tipos diferentes de datos <p>Estos deberá ser importante para mantener con la mayor calidad la información que reside en el Data Warehouse.</p>

Fuente: Elaboración propia

8 METODOLOGÍA

Etapa 1. Análisis de Requerimientos.

- i. Entrevista con cada uno de los responsables del proceso y manejo de los diferentes sistemas de la Secretaria de Salud de Medellín.
- ii. Diagnóstico de requisitos funcionales y no funcionales de los Sistema de información de la Secretaria de salud de Medellín.
- iii. Identificación de las necesidades de información gerencial. (indicadores y perspectivas).
- iv. Modelo conceptual.

Etapa 2. Análisis de los OLTP.

- i. Identificación de las fuentes de datos.
- ii. Conformar Indicadores.
- iii. Establecer correspondencia
- iv. Nivel de granularidad
- v. Modelo conceptual ampliado

Etapa 3. Elaboración del Modelo lógico - conceptual del DW de la Secretaria de Salud de Medellín.

- i. Tipo del modelo lógico del DW
- ii. Tablas de dimensiones
- iii. Tablas de hechos
- iv. Uniones o correspondencias

Tabla 16. Actividades de la Metodología de la Investigación

Actividades que conforman la Metodología de la Investigación				
Actividad	Objetivo	Técnicas	Fuentes	Hito
Recolección de información.	Definir los instrumentos apropiados para la recolección y análisis de la información.	Recolección de información	Internet Entrevista a Ingenieros Libros	Marco Teórico
Identificación de las necesidades de información gerencial, desde la perspectiva del negocio.	Levantar un diagnóstico de requisitos funcionales y no funcionales.	Entrevista	Ingeniero de Salud y TI.	Documento con el diagnóstico de los sistemas de información de la Secretaria de Salud.
Elaboración del modelo lógico – conceptual de la estructura de la bodega de datos.	Establecer las actividades que se deben realizar para el diseño conceptual de la Bodega de datos que se ajusten a las necesidades y características de la Secretaria de Salud.	Documentación	Diagnóstico de requisitos	Documento

Definir las tablas de hechos o las variables de la estructura y las dimensiones que la referencian.	Definir las tablas de hechos y dimensiones de la Bodega de datos.	Documentación	Modelo lógico conceptual.	Documento
Establecer el nivel de granulación y los niveles de agregación.	Establecer el nivel de granulación y los niveles de agregación.	Documentación	Ingenieros del Proyecto	Documento
Elaborar el diagrama en estrella que representa la estructura de la bodega.	Diseñar los datamart necesarios para la Bodega de Datos de la Secretaria de Salud.	Documentación	Ingenieros del Proyecto	Documento

9 PRESENTACIÓN DE LA SOLUCIÓN

Esta sección tiene como finalidad presentar teóricamente el desarrollo de la propuesta de lo que sería el diseño de una bodega de datos para la Secretaria de Salud de Medellín basada en la metodología HEFESTO. La cual se eligió, como la más adecuada para las necesidades de información de acuerdo al análisis e investigación de conceptos plasmados en el marco teórico del proyecto, sobre las principales técnicas o metodologías de diseño de Data Warehouse más utilizadas.

El objetivo de la presente investigación es conocer y determinar las fuentes, sistemas y/o aplicaciones de la Secretaria de Salud de Medellín que intervienen en los diferentes procesos, que permitirán identificar cuáles serán los indicadores y perspectivas que serán tomadas en cuenta para la construcción de la propuesta del diseño teórico de la bodega de datos (DW).

Esta investigación se lleva a cabo a través de un diseño no experimental, porque se limita a describir paso a paso la Metodología Hefesto aplicable a la situación de la Secretaria de salud de Medellín que tiene gran cantidad de fuentes de información o sistemas origen y que puede ser adoptada para facilitar el desarrollo de Data Warehouse en la Alcaldía de Medellín.

9.1 PASO 1. ANÁLISIS DE REQUERIMIENTOS

Esta etapa de definición de requerimientos es el punto de partida del ciclo de vida de la metodología “Hefesto”, por lo que es de gran importancia contar con la participación de los usuarios responsables de los procesos de la Secretaria de Salud, ya que son ellos los que en cierto modo guiarán la investigación hacia la

entrega de una estrategia en desarrollo que reflejara claramente lo que se espera de la bodega de datos, en relación a sus funciones y cualidades.

9.1.1.1 Identificación de las preguntas

En este punto se definirá las necesidades de información de la Secretaria de Salud de Medellín eligiendo una de las técnicas aplicables en esta Metodología. Se determina entonces que la información analizada será recopilada utilizando como técnica la entrevista; por tanto, el análisis de requerimientos para la solución presentada, es producto del trabajo realizado a través de diferentes entrevistas con los analistas y operadores del área de Sistemas de Información (SIS) de la Secretaria de Salud (Alcaldía de Medellín) y de la información que se ha extraído de diferentes servidores con previa autorización de los funcionarios, con la finalidad de realizar un análisis minucioso de las bases de datos como fuentes de información.

Se presenta a continuación la información recopilada a través de las entrevistas realizadas a los funcionarios involucrados en los procesos, iniciando por un diagnóstico o panorama general y posteriormente a nivel de detalle cada uno de sus sistemas de información más relevantes para esta investigación:

Para el manejo de la información del sector Salud en la ciudad de Medellín, la Secretaría cuenta hoy con varias aplicaciones informáticas desarrolladas para cubrir las necesidades de los diferentes subprocesos, la gran mayoría en lenguaje Visual Fox y manejadas con motor de bases de datos SQL Server, existen también algunas aplicaciones en ACCESS y se manejan permanentemente herramientas de escritorio que facilitan el manejo de bases de datos y de archivos necesarios para la generación de informes como el Excel. Igualmente se emplean aplicaciones entregadas por la Dirección Seccional de Salud de Antioquia, que por

políticas a nivel departamental y nacional deben manejarse unificadas y bajo los mismos estándares. El Ministerio de Salud y Protección Social es quien da los parámetros nacionales y por su parte viene implementando también diferentes aplicativos para uso de todo el sector salud, algunos vía web, sin embargo el proceso viene siendo lento y sin claridad a futuro; esta situación le suma complejidad al manejo de los procesos que es particular para cada Municipio. La unificación que pretende el Ministerio se basa en un mínimo de variables, esto debido a que debe ser unificado todo el país y por la particularidad y diversidad de entes territoriales, su capacidad operativa diversa hace que estos sistemas no consideren todas las variables que requieren ciudades como Medellín.

La interacción de la información al interior de los procesos de la Secretaría de Salud se realiza en forma independiente, debido a que no se cuenta con aplicativos informáticos integrados, cada módulo o sistema suministra sus propios datos sin que sea posible relacionarlos o consultarlos desde una misma estación o puesto de trabajo, ocasionando retrasos y dificultando la oportunidad y la integralidad de la información que tiene esta Secretaría.

Las diferentes comisiones de funcionarios que realizan trabajo de campo con el fin de realizar visitas, controles e interventorías, consideran que realizan acciones duplicadas pues levantan manualmente la información que luego tendrá que ser procesada en los diferentes sistemas de información existentes. Por otro lado cada proyecto realiza grandes esfuerzos para recuperar información que puede contener otro aplicativo siendo en muchos casos la misma información. Esta situación genera grandes costos y posibilidades altas de error en los datos.

En la Secretaría de Salud los sistemas de información no almacenan datos de ubicación geográfica, pero en muchos casos se cuenta con los datos de la dirección. En este momento no se tiene acceso a las bondades brindadas por el sistema de información Geográfico de la Alcaldía.

En el diseño de una bodega de datos para este tipo de entidades se debe tener en cuenta que el acceso a la información es muy restringido, la cantidad de datos e información almacenados día a día es muy grande, y los datos históricos tienen grandes problemas de calidad, por lo que se requiere de proyectos de depuración de información.

En esta sección se presenta una descripción detallada de los Sistemas y/o aplicaciones y las fuentes de información que se han identificado para el propósito de este proyecto de investigación:

9.1.1.1.1 SAS – Sistema Aseguramiento en Salud

El sistema SAS tiene como propósito asegurar en el Régimen Subsidiado en salud a toda la población pobre y vulnerable de Medellín, se encarga de identificar personas que no están aseguradas en el Régimen contributivo o de Excepción, los cuales son llamados potenciales beneficiarios para luego postularlos como beneficiarios para la afiliación al Régimen de Subsidiado.

El sistema SAS permite el registro de los afiliados al Régimen Subsidiado y el almacenamiento de información correspondiente a las personas que se encuentran aseguradas bajo algún régimen: contributivo, de excepción o especial. Estos últimos según la información que provean los regímenes especiales.

A partir de la información recibida del SISBEN (planeación nacional y municipal), los listados censales (En el artículo 5 Acuerdo 415 se encuentran las poblaciones especiales de donde provienen las encuestas, por ejemplo: ICBF, desplazados, cabildo indígena, etc) y los afiliados al régimen subsidiado (EPS-S y BDUA), que se cargan en el sistema, se logra determinar los potenciales beneficiarios mediante un cruce, depuración y validación de dicha información.

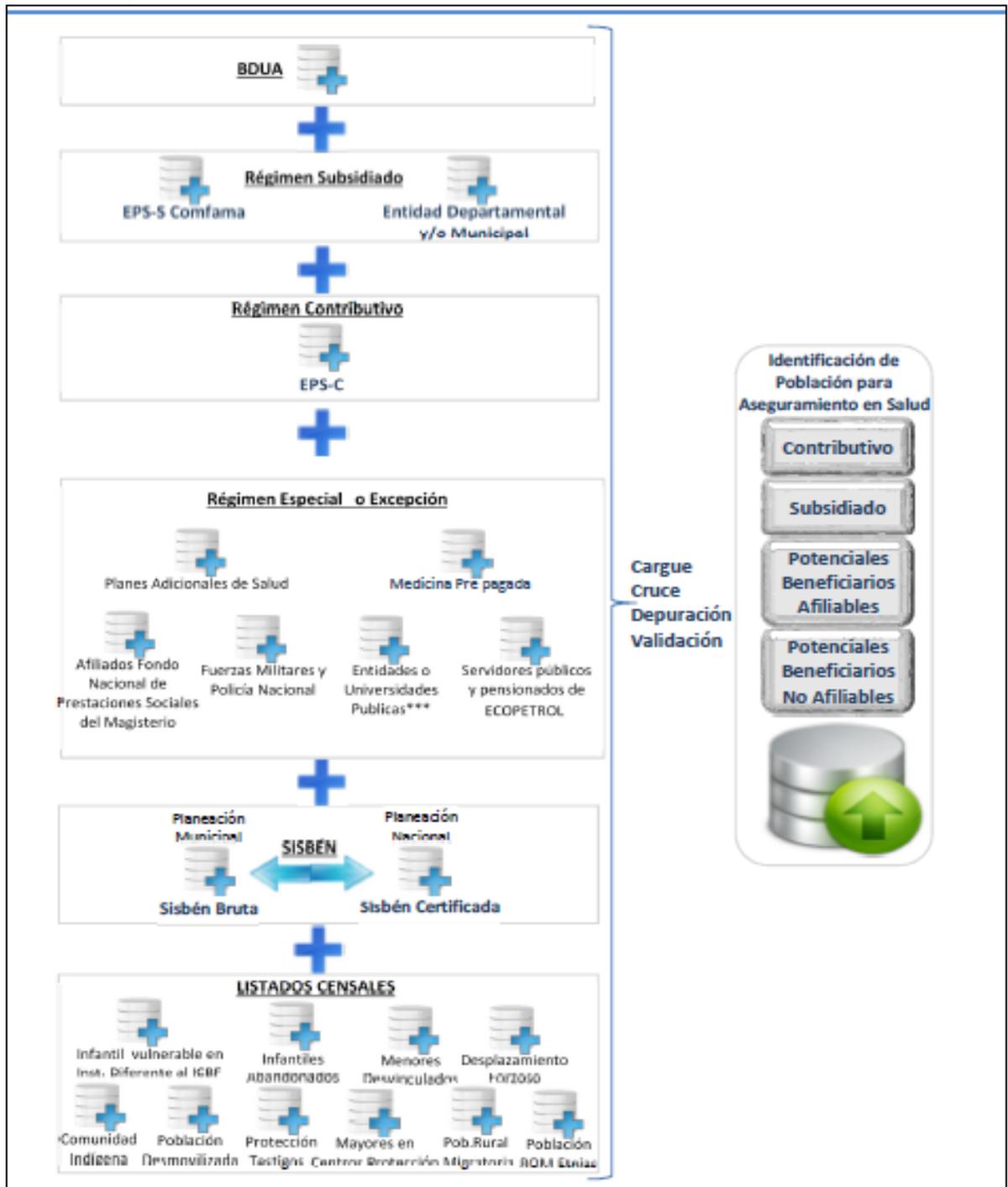
Estos potenciales beneficiarios son caracterizados como “Vinculados” o “No afiliados” y son estos los que se encuentran bajo el cubrimiento de salud por la Entidad territorial departamental y local (Seccional de Antioquia y Secretaria de salud). Los “Vinculados” son preseleccionados para ser ingresados o beneficiarios afiliables al Régimen Subsidiado.

Toda persona registrada al sistema debe pertenecer a una categoría de “Afiliación Régimen” como: contributivo, subsidiado, vinculado, particular, otro, desplazado afiliado régimen contributivo, desplazados afiliado régimen subsidiado y desplazado no asegurado. Adicionalmente se debe identificar si el tipo de afiliado es Cotizante o Beneficiario, así como denotarse si está Activo, Retirado, Desafiliado o Suspendido; condiciones que pueden variar en el tiempo, de acuerdo a las condiciones que se le presente a cada persona. No obstante, los movimientos que sucedan entre Regímenes deben quedar consignados en el sistema de información como el histórico de cada afiliado.

El almacenamiento de los cambios que presenta cada registro (persona), facilita la fidelidad y la trazabilidad de la información. Es importante resaltar que si una persona que habitaba en Medellín se traslada a otro Municipio, o bien si una persona de otro Municipio requiere un servicio o atención en salud, esto queda en movimientos como un histórico de información.

En la figura 7 se ilustra las diferentes fuentes de donde provienen los datos que son procesados en la operación del sistema SAS para la identificación de personas que aún no están aseguradas:

Figura 10. Fuentes de Información y Proceso del Sistema SAS



Fuente: Documento técnico cadena de valor del proceso SAS. Secretaria de Salud de Medellín.

9.1.1.1.2 RIPS – Registros Individuales de prestación de servicios.

El RIPS provee los datos que se requieren para hacer seguimiento al Sistema de Prestaciones de Salud en el Sistema General de Seguridad Social en salud (SGSSS), en relación con el paquete obligatorio de servicios (POS y POSS). Igualmente el objetivo del Registro es facilitar las relaciones comerciales entre las entidades administradoras (pagadores) y las instituciones y profesionales independientes (prestadores del servicio de salud) mediante la presentación del detalle de la factura de venta de servicios en medio magnético, con una estructura, que al ser estándar, facilita la comunicación y los procesos de transferencia de datos y revisión de cuentas, independientemente de las soluciones informáticas que cada prestador utilice.

Los datos de estos registros se refieren a la identificación del prestador del servicio de salud (IPS), del usuario que lo recibe, de la prestación del servicio propiamente dicho y del motivo que origino su prestación: diagnóstico y causa externa.

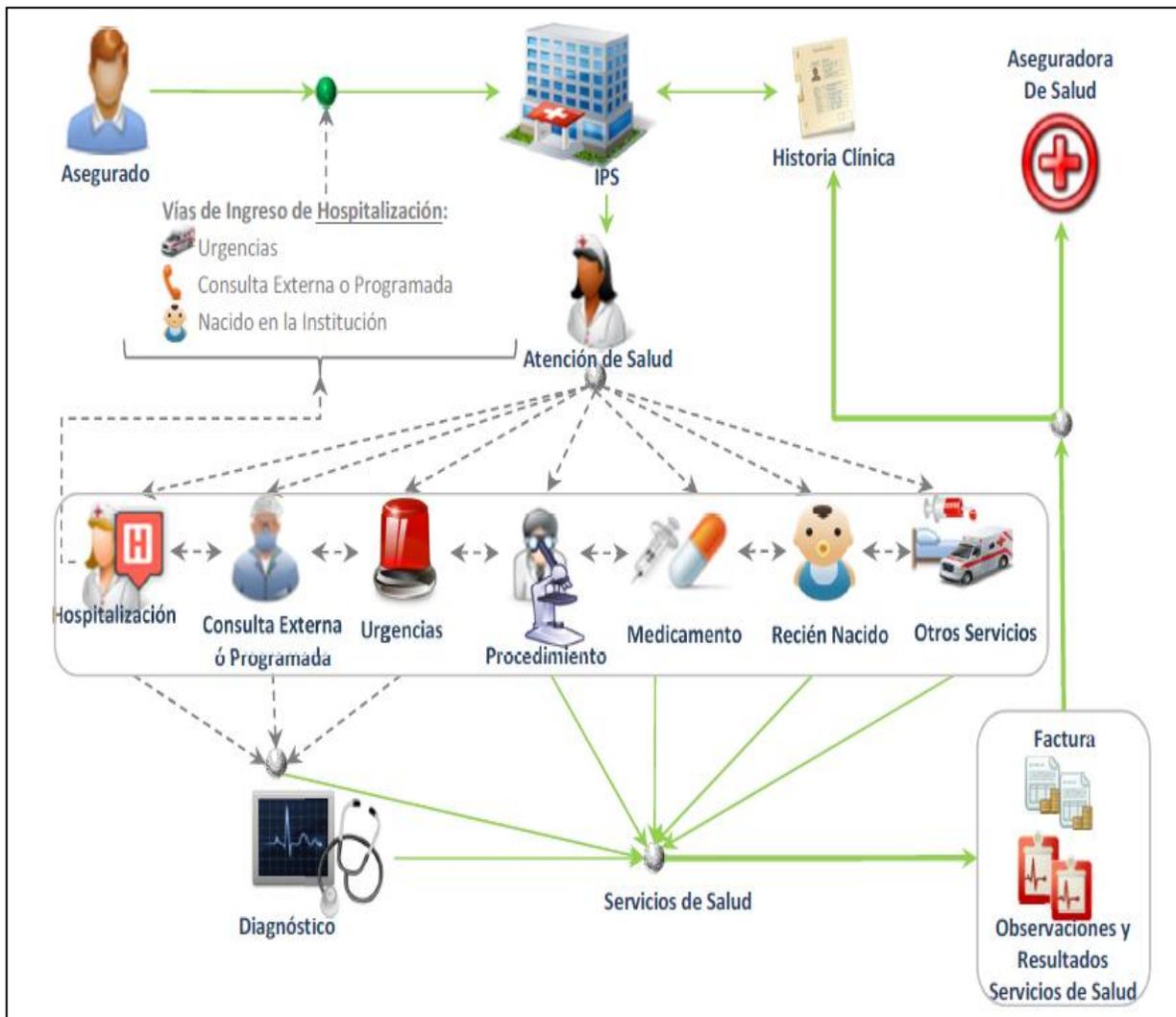
Cuando una persona “Asegurada” o “No Asegurada” ingresa a una IPS solicitando ser atendida por un requerimiento de salud, en la IPS le solicitan el número de identificación (si la persona ya existe dentro del sistema tendrá una historia clínica de todas las atenciones de salud que en esa IPS le han prestado a la persona, en caso de no estar en el sistema se le genera una nueva a partir de ese momento). A este conjunto de registros se le conoce como RIPS “Registros Individuales de Prestación de Servicios” y estos en el sistema se encuentran estrechamente relacionados ya que es importante verificar la información que se tiene en SAS como es principalmente: la identificación de personas Aseguradas y no Aseguradas (“Afiliado”), las Aseguradoras (EPS-C, EPS-S, planes adicionales de salud, etc) y los regímenes a los que pertenece cada Aseguradora (Contributivo, Subsidiado, Especial).

El personal Administrativo de la IPS se encarga de validar la Aseguradora a la que pertenezca la persona; y en caso de ser No Asegurado (“Vinculado o no afiliable”) se genera el RIPS con la Aseguradora “Estado” (Entidad Departamental y/o Municipal) hasta que se genere en SAS el cruce, la depuración y validación de esta información y se ejecute el proceso de afiliación con los “potenciales beneficiarios” hasta que estas personas se encuentren Aseguradas bajo algún Régimen.

Las IPS (Instituciones Prestadoras de Servicio) una vez reciben a la persona, le prestan el/los servicio(s) de salud dependiendo del acontecimiento y nivel de gravedad de salud que presenten las personas, como: “Urgencias”, “Consulta Externa o Programada”, “Hospitalización”, “Procedimiento”, “Medicamento”, “Recién Nacido” (partos) y/u “otros servicios”. Existen algunos servicios de salud que ocasionalmente son necesarios complementarlos con “Diagnósticos”, esos servicios son: “Hospitalización”, “Consulta Externa o Programada” y “Urgencias”, en menor demanda el servicio “Recién Nacido” también requiere de algún tipo de “Diagnostico”; así como el “Diagnostico” de algún problema de salud puede requerir de un “Procedimiento” para un mejor análisis médico. Cada uno de los servicios especifica detalladamente desde el origen del servicio, hasta los resultados que se obtengan durante la atención en salud de cada persona.

En la figura 8 se representa gráficamente el proceso de interacción de las personas con la IPS, los servicios de salud que se atienden:

Figura 11. Proceso de Interacción para los Registros Individuales de prestación de servicios



Fuente: Documento técnico cadena de valor del proceso RIPS. Secretaria de Salud de Medellín.

9.1.1.1.3 RUAF – Estadísticas Vitales

Son las estadísticas recopiladas semanalmente en todas las Registradurías del país para ser entregadas a la entidad del orden Nacional que las requieran para la gestión en tiempo real de las certificaciones en medios magnéticos, buscando reducir de manera significativa el tiempo de obtención, análisis y reproducción estadística de datos de los hechos vitales de nacimientos y defunciones.

“Estadísticas Vitales - RUAF” en el sistema de información se relaciona con la entidad “persona” y con la entidad “IPS “ como una de las principales fuentes de donde proviene los registros para obtener los resultados estadísticos de los nacimientos y fallecimientos que se producen semanal y mensualmente en cada Municipio debido a que son estas las que reciben diariamente mujeres gestantes y todo tipo de emergencias fatales; sin embargo es importante aclarar que estas novedades también provienen de diferentes fuentes externas que consultan y notifican estos sucesos, como son: las Aseguradoras, Registradurías, Medicina Legal, Cementerio Universal, entre otros.

Figura 12. Proceso Nacido vivo



Fuente: Documento técnico cadena de valor del proceso Estadísticas Vitales.
Secretaría de Salud de Medellín

9.1.1.2 Identificación de las necesidades de información gerencial (Indicadores y Perspectivas)

A través de esta investigación se ha recopilado información y estudiado las metodologías más utilizadas para el diseño de Data Warehouse, por lo que se ha presentado como propuesta desarrollar la Metodología "Hefesto" que es fácil de adaptar a cualquier ciclo de vida y busca ser simplificada en la reunión de requerimientos y análisis.

Se ha iniciado entonces con los requerimientos en la identificación de las fuentes y de los diferentes sistemas de información de la Secretaria de Salud de Medellín que son utilizados en los procesos, y partiendo de aquí se identificarán las necesidades de información gerencial (Indicadores y Perspectivas) que permitan realizar la propuesta teórica de lo que sería la Bodega de Datos de esta Secretaria.

La población considerada para esta investigación está constituida e identificada como la población del Municipio de Medellín, de manera que a nivel gerencial se dé cuenta de la situación de salud en Medellín: Régimen de aseguramiento en salud, morbilidad de la población del municipio y la fecundidad o embarazos en la población según nacimientos registrados.

Para esto se consideran datos en los siguientes niveles de análisis:

- La población de Medellín en el Sistema General de Seguridad Social en Salud.
- La población de Medellín con servicios prestados
- La fecundidad en la población de Medellín

La anterior información según los funcionarios que manejan los procesos (SAS, RUAF, RIPS) y los sistemas de información que los soportan, apuntan a los indicadores que se convierten en una herramienta fundamental para la planificación y la toma de decisiones para dar cuenta del panorama epidemiológico del Municipio.

Posteriormente se preguntó a los funcionarios cuales serían las variables o perspectivas desde las cuales se consultarán dichos indicadores. Las preguntas del negocio obtenidas fueron las siguientes:

- Se desea conocer cuántos afiliados hay en el SGSSS en el total de la población de Medellín según comuna y corregimiento de residencia habitual y nivel SISBEN en un instante de tiempo determinado.

- Se desea conocer el total de Nacimientos atendidos en Instituciones de Salud de la ciudad de Medellín en un tiempo determinado.
- Se desea conocer el total de casos de Morbilidad de la población de Medellín por consulta externa (10 primeros motivos más frecuentes) en un tiempo determinado.
- Se desea conocer el total de casos de Morbilidad de la población de Medellín por consulta de urgencias (10 primeros motivos más frecuentes) en un tiempo determinado.

Debido a que la dimensión Tiempo es un elemento fundamental en el DW, se hizo hincapié en él a los funcionarios dueños de los procesos. Además se hizo mucho énfasis en que es este componente el que les permitirá tener varias versiones de los datos a fin de que les permita realizar un correcto análisis posterior en caso de que se tome la decisión del alto nivel de implementar el Data Warehouse.

En la siguiente tabla se visualizan los indicadores y las perspectivas obtenidos de acuerdo a las preguntas del negocio:

Tabla 17. Identificación de Indicadores y Perspectivas

Indicadores	Perspectivas
<ul style="list-style-type: none"> ▪ Total afiliados al SGSSS población Medellín ▪ Total Nacimientos atendidos en instituciones de salud de Medellín ▪ Total de los 10 primeros motivos más frecuentes de Morbilidad población Medellín 	<ul style="list-style-type: none"> ▪ Comuna ▪ Corregimiento ▪ Nivel ▪ Sexo ▪ Instituciones de salud ▪ Causas CIE10

Tabla 18. Datos que quedan Excluidos del Análisis de esta Propuesta Teórica de Bodega de Datos (DW) y Motivos.

Datos Excluidos Del Sistema De Información	Descripción del Motivo
<ul style="list-style-type: none"> ▪ SAS (Sistema de Aseguramiento en Salud) 	<p>Se tiene información de los regímenes Subsidiado, contributivo, vinculado y alguna del régimen especial o de excepción como lo son: EPM, UDEA, UNAL, ECOPETROL. Para el régimen especial o de Excepción no se tiene incluida el total de su población afiliada ya que no existe en la normatividad vigente (Ley 1438 – Acuerdo 415) la obligatoriedad de que reporten la información a los Entes territoriales o a la BDUA (Base de Datos Única de Afiliados), por tanto no se cuenta de este régimen con la información de los afiliados a la Policía o Fuerzas Militares, Medico preventiva, SENA y/u otras entidades de excepción.</p>
<ul style="list-style-type: none"> ▪ RUAF (Estadísticas Vitales) 	<p>se tiene información para difusión estadística, fue diseñado con el propósito de mantener actualizadas las estadísticas de nacimientos y defunciones para permitir un flujo de información que revele los cambios ocurridos en los niveles y patrones de mortalidad, fecundidad y brindar una visión dinámica de la población como complemento al enfoque estático que proveen los censos. El sistema maneja dos bases de datos una de nacimientos y otra de defunciones, para el caso de este proyecto no se trabajó la parte de fallecimientos por ser información de gran reserva y confidencialidad de datos.</p>

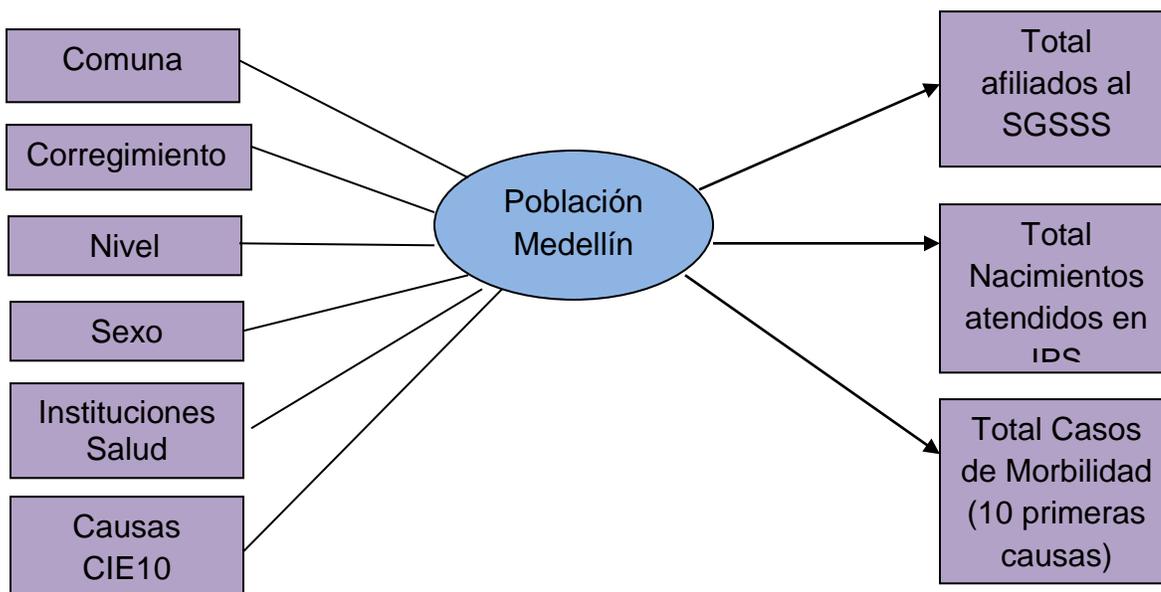
Para el desarrollo de la investigación se incluyó solo los procesos descritos en el paso 1 de Análisis de Requerimientos para hacer la relación e incluirlos en la bodega de datos. Sin embargo, este repositorio podrá crecer y ser modificado luego de la finalización del presente proyecto, ya sea porque la Secretaria de Salud decida tomarlo para su estudio e implementación.

9.1.1.3 Modelo Conceptual

En esta etapa, se construye el modelo conceptual a partir de los indicadores y perspectivas obtenidas en el paso anterior. El Modelo conceptual representa la descripción a alto nivel de la estructura de la base de datos de la información que se recopiló de la Secretaria de Salud.

El modelo conceptual resultante de los datos que se han recolectado, es el que se muestra en la siguiente figura:

Figura 13. Modelo Conceptual Información Secretaria de Salud



A la izquierda se representan las perspectivas seleccionadas en la Tabla 12 las cuales van unidas a un ovalo central que representa y lleva el nombre de “Población Medellín” el cual es la relación que existen entre ellas y constituye el proceso o área de análisis elegida. De dicha relación y entrelazadas con flechas, se desprenden los indicadores, los cuales se ubican a la derecha del esquema. Como puede observarse, la relación mediante la cual se unen las diferentes perspectivas, para obtener como resultado los indicadores requeridos por los usuarios, es “Población Medellín”.

9.2 PASO 2. ANALISIS DE LOS OLTP

Esta etapa consiste en analizar las fuentes OLTP de la Secretaria de Salud de Medellín las cuales fueron extractadas de las fuentes de información en el desarrollo de las entrevistas a cada funcionario responsable de los procesos. Estas fuentes OLTP permitirán determinar cómo serán calculados los indicadores y se podrá establecer la correspondencia entre el Modelo Conceptual representado en la figura 13 y las fuentes de datos. Para ello iniciaremos el desarrollo de este paso partiendo de la:

9.2.1 Identificación De Las Fuentes De Datos

Las fuentes de datos parten de los sistemas de información de la Secretaria de Salud (SAS – Sistema de Aseguramiento en Salud, RUAF – Estadísticas Vitales, RIPS – Registro Individuales de Prestación de Servicios) estos sistemas son alimentados por tres (3) esquemas de bases de datos llamadas **SECSALUD195**, **RUAF**, **SISMASTER** respectivamente, los cuales se describen detalladamente en la siguiente tabla:

Tabla 19. Descripción Fuentes o Base de Datos de la Secretaria de Salud de Medellín.

Fuente Base de datos	Descripción
SECSALUD195	Contiene las tablas maestras y de tipos necesarias para consolidar la información de los afiliados al Régimen Subsidiado y Potenciales Beneficiarios (vinculados) de Medellín, Régimen Contributivo (departamental) y del Régimen de Excepción o especial (solo de los que notifican actualmente). Además de guardar un histórico de la trazabilidad de los registros en sus movimientos de traslados de regímenes o actualización de datos.
RUAF	Contiene la tabla maestra y de tipos necesarias para la información consolidada de los nacimientos registrados por las IPS o entidades prestadoras de servicio.
SISMASTER	Contiene las tablas maestras y de tipos necesarias para el registro de los datos básicos como la transacción, el servicio sobre los servicios individuales de salud.

9.2.1.1 Indicar el servidor, base de datos, sistema operativo, tipo de lenguaje de programación y de consulta.

La tabla que se presenta a continuación contiene la información referente al nombre del servidor donde se encuentran alojadas las fuentes o bases de datos de la Secretaria de Salud; así como las especificaciones de los nombres, del sistema operativo, el lenguaje de programación y de consulta que se utiliza:

Tabla 20. Especificaciones Técnicas De Las Fuentes Y Sistemas De Información de La Secretaria De Salud

Especificaciones Técnicas				
Base de datos	Servidor	Sistema Operativo	Lenguaje de Programación	Lenguaje de consulta
RUAF	Tauro	Windows XP y/o Vista	Visual FoxPro 5.0, 6.0 y/o 7.0	SQL
Secaslud195	Tauro	Windows XP y/o Vista	Visual FoxPro 9.0	SQL
Sismaster	Tauro	Windows XP	Visual FoxPro 5.0	SQL

9.2.2 Conformar Indicadores

En este paso se muestra cómo se calculan los indicadores obtenidos para la Secretaria de Salud de Medellín, los cuales podrán dar cuenta del panorama epidemiológico del Municipio y servirán como herramienta fundamental para la planificación y la toma de decisiones. Además, la conformación de los mismos pretende ofrecer información unificada sobre la situación de salud de los residentes de la ciudad.

Tabla 21. Calculo de los Indicadores de la Secretaria de Salud de Medellín

Indicador	Hechos	Función
Total afiliados al SGSSS población Medellín	Afiliados al SGSSS (Sistema General de seguridad Social en Salud) población Medellín	COUNT (Conteo) Aclaración: el indicador representa el conteo del total de la población de Medellín por régimen de seguridad social.
Total Nacimientos atendidos en instituciones de salud de la población Medellín	Nacimientos atendidos en instituciones de salud de Medellín	SUM (sumatoria) Aclaración: el indicador representa la sumatoria de los nacimientos que se atendieron por institución de salud de Medellín. Unidad de medida: porcentaje (%)
Total de los 10 primeros motivos más frecuentes de Morbilidad población Medellín	10 primeros motivos de morbilidad de la población Medellín	SUM (sumatoria) Aclaración: el indicador representa la sumatoria de cada uno de los casos registrados de los servicios prestados según la clasificación de la CIE10. Unidad de medida: porcentaje (%)

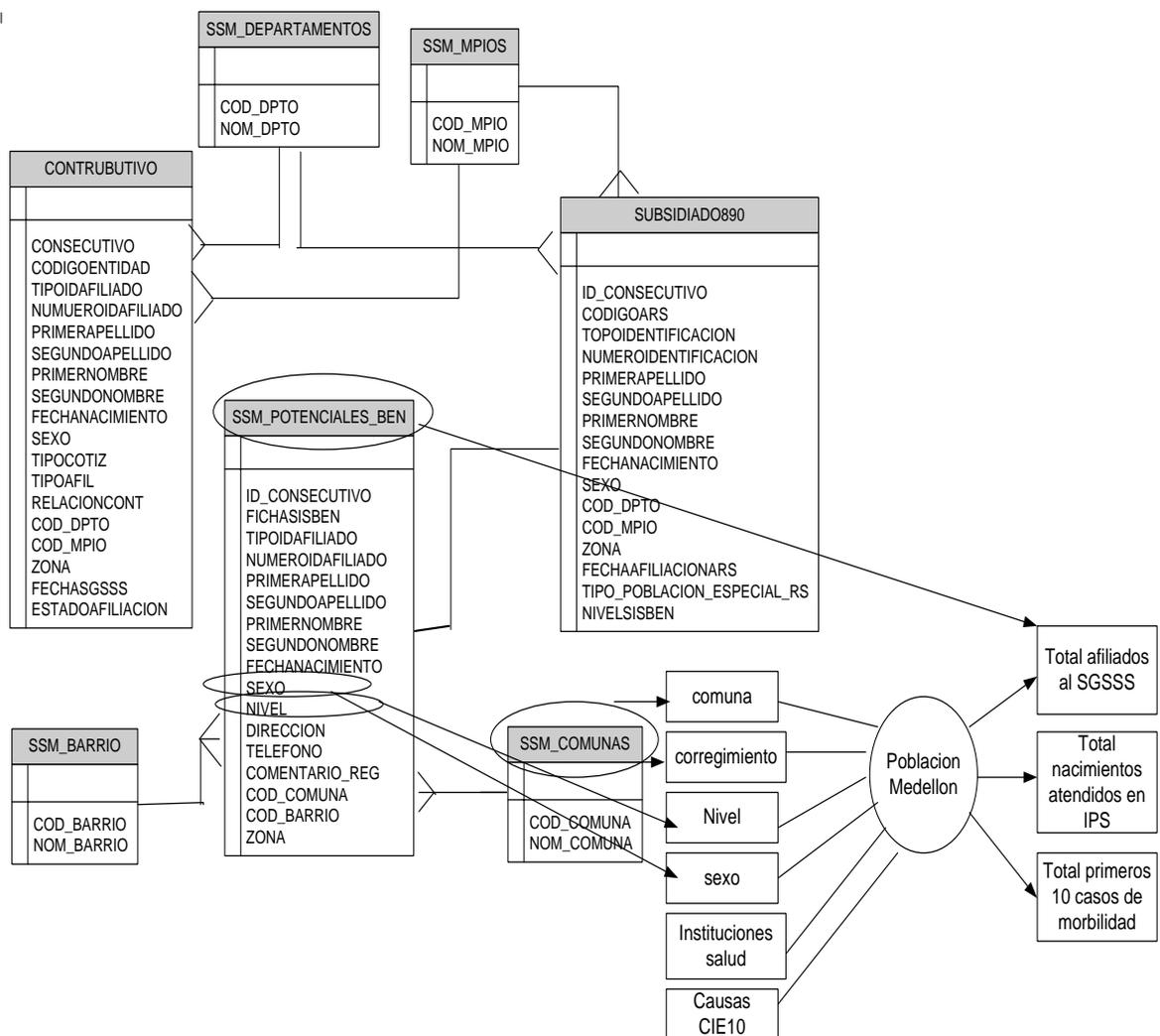
9.2.3 Establecer correspondencias

El objetivo en este paso es el de explorar los OLTP disponibles de la Secretaria de Salud de Medellín que contienen la información, como así también sus

características, de manera que se puedan identificar las correspondencias entre el Modelo Conceptual y las fuentes de datos.

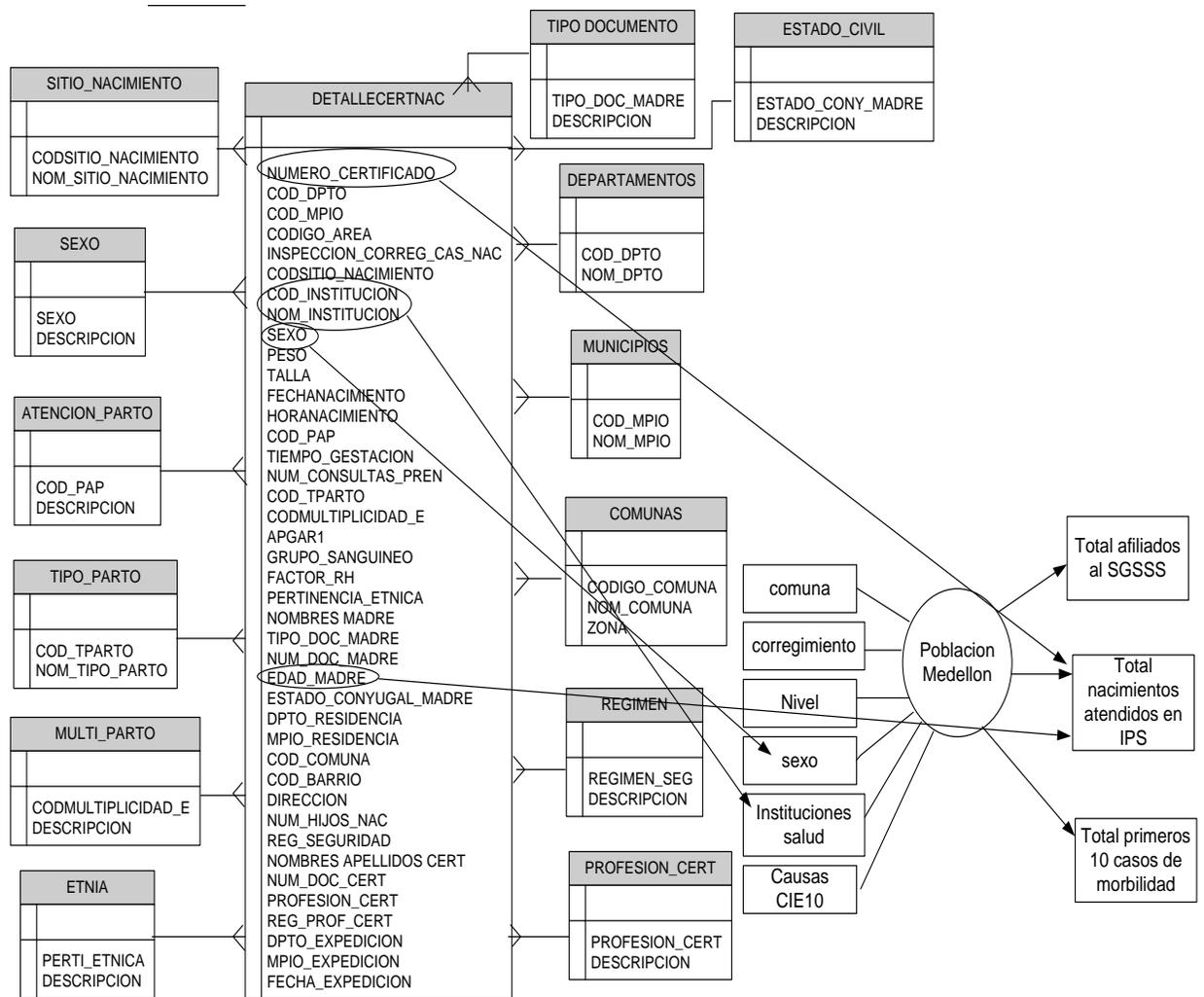
En el OLTP de la Secretaria de salud, el proceso de afiliados al SGSSS (Sistema General de seguridad Social en Salud) está representado por el diagrama de entidad relación de la siguiente figura, además su correspondencia con el Modelo conceptual:

Figura 14. Diagrama Entidad Relación de los afiliados al SGSSS vs Modelo Conceptual



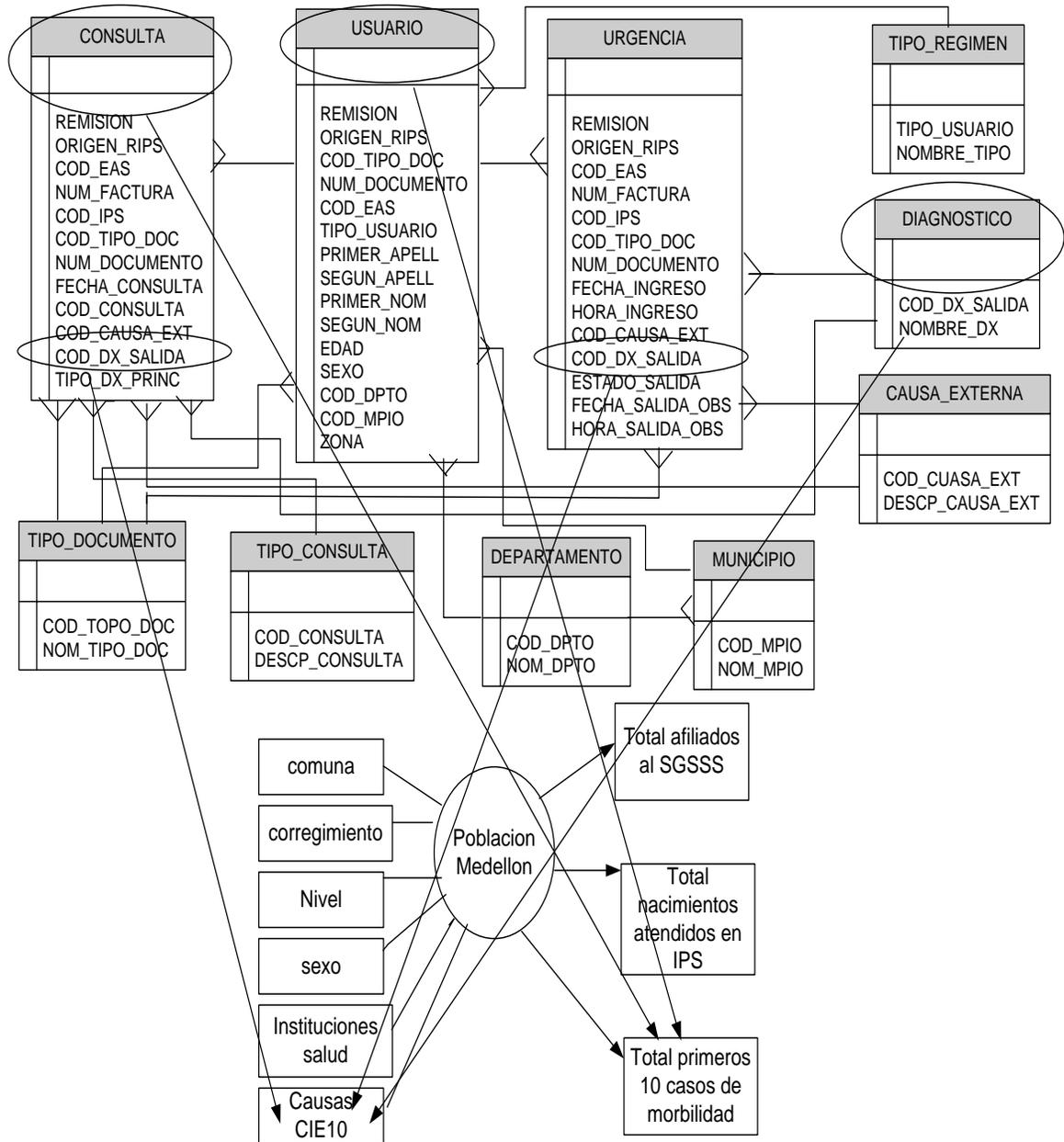
En el OLTP de la Secretaría de salud, el proceso de Estadísticas vitales (Nacimientos) está representado por el diagrama de entidad relación de la siguiente figura, además su correspondencia con el Modelo Conceptual:

Figura 15. Diagrama Entidad Relación de Nacimientos vs Modelo Conceptual



En el OLTP de la Secretaría de salud, el proceso de RIPS – Registro Individual de Prestación de Servicios está representado por el diagrama de entidad relación de la siguiente figura, además su correspondencia con el Modelo Conceptual:

Figura 16. Diagrama Entidad Relación de RIPS vs Modelo Conceptual



9.2.4 Indicar Tablas y Campos de las Fuentes o Bases de Datos

9.2.4.1 SAS (Sistema de Aseguramiento en Salud)

Las siguientes son las tablas y campos de la base de datos Secsalud195 del sistema SAS (Sistema de Aseguramiento en Salud) las cuales corresponden al modelo entidad relación representado en la figura 14:

Tabla			
SUBSIDIADO890			
Campo	Null	Tipo Campo	Descripción
ID_CONSECUTIVO	NOT NULL	NUMERIC(18,0)	Identificador único de la persona
CODIGOARS	NOT NULL	VARCHAR(6)	Indica el código de la EPS subsidiada a la que pertenece el usuario el usuario
TIPOIDENTIFICACION	NOT NULL	CHAR(2)	Indica el tipo de identificación del afiliado: RC – registro civil; CC – cedula de ciudadanía; CE – cedula extranjería; CN – certificado nacido; PA – pasaporte; TI –

			tarjeta de identidad.
NUMEROIDENTIFICACION	NOT NULL	VARCHAR(16)	Identificación del afiliado
PRIMERAPELLIDO	NOT NULL	VARCHAR(20)	Primer apellido de la persona
SEGUNDOAPELLIDO	NULL	VARCHAR(30)	Segundo apellido de la persona
PRIMERNOMBRE	NOT NULL	VARCHAR(20)	Primer nombre de la persona
SEGUNDONOMBRE	NULL	VARCHAR(30)	Segundo nombre de la persona
FECHANACIMIENTO	NOT NULL	DATE TIME	Fecha de nacimiento de la persona formato dd/mm/aa
SEXO	NOT NULL	CHAR(1)	Indica el sexo de la persona: Hombre - Mujer
COD_DPTO	NULL	CHAR(2)	Indica el código de departamento de afiliación (codificación DANE)
COD_MPIO	NULL	CHAR(3)	Indica el código de Municipio de afiliación (codificación DANE)

ZONA	NULL	CHAR(1)	Indica la Zona de afiliación: Urbana; Rural; Cabecera
FECHAAFILIACIONARS	NULL	DATE TIME	Indica la Fecha de afiliación a la Entidad Formato dd/mm/aa
TIPO_POBLACIÓN_ESPECIAL_RS	NULL	CHAR(2)	Indica el tipo de grupo poblacional al que pertenece el usuario
NIVEL SISBEN	NULL	CHAR(1)	Indica el nivel de estratificación del SISBEN

Tabla CONTRIBUTIVO			
Campo	Null	Tipo Campo	Descripción
CONSECUTIVO	NOT NULL	NUMERIC(18,0)	Identificador único de la persona
CODIGOENTIDAD	NOT NULL	VARCHAR(6)	Indica el código de la EPS subsidiada a la que pertenece el usuario el usuario
TIPOIDAFILIADO	NOT NULL	CHAR(2)	Indica el tipo de identificación del afiliado: RC – registro

			civil; CC – cedula de ciudadanía; CE – cedula extranjera; CN – certificado nacido; PA – pasaporte; TI – tarjeta de identidad.
NUMEROIDAFILIADO	NOT NULL	VARCHAR(20)	Identificación del afiliado
PRIMERAPELLIDO	NOT NULL	VARCHAR(20)	Primer apellido de la persona
SEGUNDOAPELLIDO	NULL	VARCHAR(20)	Segundo apellido de la persona
PRIMERNOMBRE	NOT NULL	VARCHAR(20)	Primer nombre de la persona
SEGUNDONOMBRE	NULL	VARCHAR(20)	Segundo nombre de la persona
FECHANACIMIENTO	NOT NULL	VARCHAR(10)	Fecha de nacimiento de la persona formato dd/mm/aa
SEXO	NOT NULL	VARCHAR(1)	Indica el sexo de la persona: Hombre - Mujer
TIPOCOTIZ	NULL	VARCHAR(1)	Indica el tipo de cotizante
TIPOAFIL	NULL	VARCHAR(1)	Indica el tipo de afiliado
RELACIONCONT	NULL	VARCHAR(1)	Indica la relación del usuario con el cotizante
COD_DPTO	NULL	CHAR(2)	Indica el código de departamento de

			afiliación (codificación DANE)
COD_MPIO	NULL	CHAR(3)	Indica el código de Municipio de afiliación (codificación DANE)
ZONA	NULL	VARCHAR(1)	Indica la Zona de afiliación: Urbana; Rural; Cabecera
FECHASGSSS	NULL	VARCHAR(10)	Indica la Fecha de afiliación al Sistema General de Seguridad Social en Salud Formato dd/mm/aa
ESTADOAFILIACION	NOT NULL	CHAR(2)	Indica estado de afiliación del usuario en la EPS

Tabla SSM_POTENCIALES_BENEFICIARIOS			
Campo	Null	Tipo Campo	Descripción
ID_CONSECUTIVO	NOT NULL	NUMERIC(18,0)	Identificador único de la persona
FICHASISBEN	NOT NULL	VARCHAR(14)	Indica el número de ficha asignado por el SISBEN
TIPOIDAFILIADO	NOT NULL	CHAR(2)	Indica el tipo de identificación del

			afiliado: RC – registro civil; CC – cedula de ciudadanía; CE – cedula extranjera; CN – certificado nacido; PA – pasaporte; TI – tarjeta de identidad.
NUMEROIDAFILIADO	NOT NULL	VARCHAR(20)	Identificación del afiliado
PRIMERAPELLIDO	NOT NULL	VARCHAR(20)	Primer apellido de la persona
SEGUNDOAPELLIDO	NULL	VARCHAR(20)	Segundo apellido de la persona
PRIMERNOMBRE	NOT NULL	VARCHAR(20)	Primer nombre de la persona
SEGUNDONOMBRE	NULL	VARCHAR(20)	Segundo nombre de la persona
FECHANACIMIENTO	NOT NULL	VARCHAR(10)	Fecha de nacimiento de la persona formato dd/mm/aa
SEXO	NOT NULL	VARCHAR(1)	Indica el sexo de la persona: Hombre – Mujer
NIVEL	NOT NULL	VARCHAR(1)	Indica el nivel de estratificación del SISBEN
DIRECCION	NOT NULL	VARCHAR(255)	Indica la dirección de residencia de la persona

TELEFONO	NOT NULL	VARCHAR(100)	Indica número de teléfono de la persona
COMENTARIO_REG	NOT NULL	VARCHAR(200)	Indica o hace referencia a la observación donde se ubica la persona en el SGSSS
COD_COMUNA	NOT NULL	CHAR(10)	Indica el código de la comuna a la cual pertenece la persona
COD_BARRIO	NOT NULL	CHAR(10)	Indica el código del barrio al cual pertenece la persona
ZONA	NULL	VARCHAR(1)	Indica la Zona de afiliación: Urbana; Rural; Cabecera

Tabla SSM_DEPARTAMENTOS			
Campo	Null	Tipo Campo	Descripción
COD_DPTO	NOT NULL	CHAR(2)	Indica el código de departamento de afiliación (codificación DANE)
NOM_DPTO	NOT NULL	VARCHAR(50)	Indica el nombre del departamento

Tabla SSM_MUNICIPIOS			
Campo	Null	Tipo Campo	Descripción
COD_MPIO	NOT NULL	CHAR(2)	Indica el código del Municipio de afiliación (codificación DANE)
NOM_MPIO	NOT NULL	VARCHAR(50)	Indica el nombre del municipio

Tabla SSM_COMUNAS			
Campo	Null	Tipo Campo	Descripción
COD_COMUNA	NOT NULL	CHAR(28)	Indica el código de la Comuna
NOM_COMUNA	NOT NULL	CHAR(10)	Indica el nombre de la Comuna
ZONA	NOT NULL	CHAR(20)	Indica la zona

Tabla SSM_BARRIOS			
Campo	Null	Tipo Campo	Descripción
COD_BARRIO	NOT NULL	CHAR(2)	Indica el código del barrio
NOM_BARRIO	NOT NULL	VARCHAR(50)	Indica el nombre del barrio

9.2.4.2 RUAF (Registro Único de Afiliados)

Las siguientes son las tablas y campos de la base de datos RUAF – Nacimientos sistema RUAF (Registro Único de Afiliados) las cuales corresponden al modelo entidad relación representado en la figura 15:

Tabla DETALLECERTNAC			
Campo	Null	Tipo Campo	Descripción
NUMERO_CERTIFICADO	NOT NULL	NUMERIC(18, 0)	Indicador único del certificado
COD_DPTO	NULL	CHAR(5)	Indica el código de departamento de afiliación (codificación DANE)
COD_MPIO	NULL	CHAR(5)	Indica el código de Municipio de afiliación (codificación DANE)
CODIGO_AREA	NULL	CHAR(1)	Indica el área de nacimiento: 1; 2; 3
INSPECCION_CORREG_CAS_NAC	NULL	VARCHAR(50)	Indica el nombre de la inspección; corregimiento o caserío del nacimiento
CODSITIO_NACIMIENTO	NULL	CHAR(1)	Indica el lugar donde ocurrió el nacimiento: 1; 2; 3;
CÓD_INSTITUCIÓN	NULL	VARCHAR(12)	Indica el código de la

			institución de salud donde se dio el nacimiento
NOM_INSTITUCIÓN	NULL	VARCHAR(120)	Indica el nombre de la institución de salud
SEXO	NULL	CHAR(1)	Indica el sexo del nacido: 1; 2
PESO	NULL	NUMERIC(18, 0)	Indica el peso del nacido vivo al nacer (gramos)
TALLA	NULL	TINYINT	Indica la talla del nacido vivo al nacer (centímetros)
FECHANACIMIENTO	NULL	DATETIME	Fecha de nacimiento del nacido vivo formato aa/mm/dd
HORANACIMIENTO	NULL	TIME	Indica la hora en que ocurrió el nacimiento Hora – Minutos
COD_PAP	NULL	CHAR(1)	Indica el código que corresponde a quien atendió el parto: 1;2;3;4;5;6
TIEMPO_GESTACIÓN	NULL	TINYINT	Indica el número de semanas completas de gestación
NÚM_CONSULTAS_PREN	NULL	TINYINT	Indica el número de consultas prenatales que tuvo la madre del

			nacido vivo
COD_TPARTO	NULL	CHAR(1)	Indica el código del tipo de parto del nacimiento 1;2;3;4
CODMULTIPLICIDAD_E	NULL	CHAR(1)	Indica el código de la multiplicidad del embarazo (1;2;3;4)
APGAR1	NULL	CHAR(2)	Indica la valoración que se le hace el recién nacido al minuto y a los 5 minutos de nacimiento
GRUPO_SANGUINEO	NULL	CHAR(2)	Indica la hemoclasificación del nacido vivo – grupo sanguíneo
FACTOR RH	NULL	VARCHAR(10)	Indica la hemoclasificación del nacido vivo – factor RH (positivo; negativo)
PERTENENCIA ÉTNICA	NULL	CHAR(1)	Indica el código de la pertinencia étnica a la que pertenece de acuerdo con la cultura; pueblo o rasgos físicos. 1;2;3;4;5;6

NOMBRES MADRE	NULL	VARCHAR(60)	Indica los nombres y apellidos de la madre como figuran en el documento de identidad
TIPO_DOC_MADRE	NULL	CHAR(1)	Indica tipo de documento de la madre 1;2;3;4; 5;9
NÚM_ DOC_MADRE	NULL	VARCHAR(20)	Indica el número de documento de la madre
EDAD MADRE	NULL	NUMERIC(18, 0)	Indica la edad de la madre en años cumplidos
ESTADO CONYUGAL MADRE	NULL	CHAR(1)	Indica el código que denomina el estado civil de la madre 1;2;3;4;5;6;9
DPTO_RESIDENCIA	NULL	CHAR(5)	Indica el código de departamento de afiliación (codificación DANE)
MPIO_RESIDENCIA	NULL	CHAR(5)	Indica el código de Municipio de afiliación (codificación DANE)
COD_COMUNA	NULL	CHAR(10)	Indica el código de la comuna a la cual pertenece la persona
COD_BARRIO	NULL	CHAR(10)	Indica el código del

			barrio al cual pertenece la persona
DIRECCIÓN	NULL	VARCHAR(100)	Indica la dirección de residencia de la madre del nacido
NÚMERO HIJOS NACIDOS VIVOS	NULL	TINYINT	Indica el número de hijos nacidos vivos que ha tenido la madre incluido el presente
RÉGIMEN SEGURIDAD	NULL	CHAR(1)	Indica el régimen de seguridad social en salud de la madre: 1;2;3;4;5
NOMBRES Y APELLIDOS CERTIFICADOR	NULL	VARCHAR(60)	Indica los nombres y apellidos de quien certifica el nacimiento tal como aparece en el documento de identidad
NÚM_DOC_CERT	NULL	VARCHAR(20)	Indica el número de documento de quien certifica el nacimiento
TIPO DOCUMENTO CERTIFICADOR	NULL	CHAR(2)	Indica el tipo de documento de quien certifica el documento (CC – cedula de ciudadanía; CE – cedula de extranjería;

			PA – pasaporte)
PROFESION_CERT	NULL	CHAR(1)	Indica el código que hace referencia a la profesión de quien certifica el nacimiento: 1;2;3;4;5
REGISTRO PROFESIONAL CERTIFICADOR	NULL	VARCHAR(20)	Indica el número de registro profesional del certificador
DEPARTAMENTO EXPEDICIÓN	NULL	CHAR(5)	Indica el código de departamento de afiliación (codificación DANE)
MUNICIPIO EXPEDICIÓN	NULL	CHAR(5)	Indica el código de Municipio de afiliación (codificación DANE)
FECHA EXPEDICIÓN	NULL	DATETIME	Indica la fecha de expedición del certificado formato aa/mm/dd

Tabla AREA_NACIMIENTO			
Campo	Null	Tipo Campo	Descripción
CODIGO_AREA	NOT NULL	CHAR(1)	Indica el código del área de nacimiento
NOMBRE_AREA	NOT NULL	VARCHAR(100)	Indica el nombre del área de nacimiento (cabecera municipal; centro

			poblado; Rural disperso)
--	--	--	--------------------------

Tabla SITIO_NACIMIENTO			
Campo	Null	Tipo Campo	Descripción
CODSITIO_NACIMIENTO	NOT NULL	CHAR(1)	Indica el código del sitio de nacimiento
NOM_SITIO_NACIMIENTO	NOT NULL	VARCHAR(100)	Indica el nombre del sitio de nacimiento (Institución de salud; el domicilio; otro sitio)

Tabla SEXO			
Campo	Null	Tipo Campo	Descripción
SEXO	NOT NULL	CHAR(1)	Indica el código del sexo
DESCRIPCION	NOT NULL	VARCHAR(100)	Indica el nombre al que corresponde el código Masculino – Femenino

Tabla ATENCION_PARTO			
Campo	Null	Tipo Campo	Descripción
COD_PAP	NOT	CHAR(1)	Indica el código de quien

	NULL		atendió el parto
DESCRIPCION	NOT NULL	VARCHAR(100)	Indica el nombre al que corresponde el código de atención del parto (medico; enfermero; Auxiliar de enfermería; Promotor de salud; partera; otra persona)

Tabla TIPO_PARTO			
Campo	Null	Tipo Campo	Descripción
COD_TPARTO	NOT NULL	CHAR(1)	Indica el código del tipo de parto del nacimiento
NOMBRE TIPO PARTO	NOT NULL	VARCHAR(100)	Indica el nombre del tipo de parto del nacimiento (espontaneo; Cesárea; instrumentado; Ignorado)

Tabla MULTI_PARTO			
Campo	Null	Tipo Campo	Descripción
CODMULTIPLICIDAD_E	NOT NULL	CHAR(1)	Indica el código de la multiplicidad del embarazo
DESCRIPCION	NOT NULL	VARCHAR(100)	Indica el nombre que define la multiplicidad del embarazo (simple;

			doble; triple; cuádruple o más)
--	--	--	---------------------------------

Tabla ETNIA			
Campo	Null	Tipo Campo	Descripción
PERTENENCIA ÉTNICA	NOT NULL	CHAR(1)	Indica el código de la pertinencia étnica a la que pertenece de acuerdo con la cultura; pueblo o rasgos físicos.
DESCRIPCION	NOT NULL	VARCHAR(100)	Indica el nombre que define la pertinencia étnica (Indígena; ROM; Raizal; Palanquero; Negro; Ninguno de los anteriores)

Tabla TIPO_DOCUMENTO			
Campo	Null	Tipo Campo	Descripción
TIPO DOCUMENTO MADRE	NOT NULL	CHAR(1)	Indica tipo de documento de la madre
DESCRIPCION	NOT NULL	VARCHAR(100)	Indica el nombre del tipo de documento de la madre (Registro civil; Tarjeta de identidad; Cedula de ciudadanía;

			Cedula Extranjería; Pasaporte; Sin información)
--	--	--	---

Tabla ESTADO_CIVIL			
Campo	Null	Tipo Campo	Descripción
ESTADO CONYUGAL MADRE	NOT NULL	CHAR(1)	Indica el código que denomina el estado civil de la madre 1;2;3;4;5;6;9
DESCRIPCION	NOT NULL	VARCHAR(100)	Indica el nombre del estado civil (No está casada y lleva más de 2 años viviendo con su pareja; No está casada y lleva menos de 2 años viviendo con su pareja; está separada o divorciada; esta viuda; esta soltera; está casada; sin información)

Tabla DEPARTAMENTOS			
Campo	Null	Tipo Campo	Descripción
COD_DPTO	NOT NULL	CHAR(2)	Indica el código de departamento de afiliación (codificación DANE)

NOM_DPTO	NOT NULL	VARCHAR(50)	Indica el nombre del departamento
----------	-------------	-------------	-----------------------------------

Tabla MUNICIPIOS			
Campo	Null	Tipo Campo	Descripción
COD_MPIO	NOT NULL	CHAR(2)	Indica el código del Municipio de afiliación (codificación DANE)
NOM_MPIO	NOT NULL	VARCHAR(50)	Indica el nombre del municipio

Tabla COMUNAS			
Campo	Null	Tipo Campo	Descripción
CODIGO_COMUNA	NOT NULL	CHAR(28)	Indica el código de la Comuna
NOM_COMUNA	NOT NULL	CHAR(10)	Indica el nombre de la Comuna
ZONA	NOT NULL	CHAR(20)	Indica la zona

Tabla REGIMEN			
Campo	Null	Tipo Campo	Descripción
REGIMEN SEGURIDAD	NOT NULL	CHAR(1)	Indica el código del régimen de seguridad social
DESCRIPCION	NOT NULL	CHAR(10)	Indica el nombre del régimen de seguridad social (contributivo; subsidiado: excepción; especial; no asegurado)

Tabla PROFESION_CERTIFICADOR			
Campo	Null	Tipo Campo	Descripción
PROFESION CERTIFICADOR	NOT NULL	CHAR(1)	Indica el código que hace referencia a la profesión de quien certifica el nacimiento
DESCRIPCION	NOT NULL	CHAR(10)	Indica el nombre de la profesión del certificador (Medico; Enfermero: Auxiliar de enfermería; Promotor de salud; Funcionario de registro civil)

9.2.4.3 RIPS (Registro Individual de prestación de Servicios)

Las siguientes son las tablas y campos de la base de datos Sismaster sistema RIPS (Registro Individual de prestación de Servicios), las cuales corresponden al modelo entidad relación representado en la figura 16:

Tabla RIPS_USUARIOS			
Campo	Null	Tipo Campo	Descripción
REMISION	NOT NULL	VARCHAR(20)	Identificador único de la tabla
ORIGEN_RIPS	NULL	VARCHAR(1)	Indica quien reporta el RIPS si es la IPS o Administradora
COD_TIPO_DOCUMENTO	NULL	VARCHAR(2)	Indica el tipo de identificación de la persona: RC – registro civil; CC – cedula de ciudadanía; CE – cedula extranjera; PA – pasaporte; TI – tarjeta de identidad; AS – adulto sin identificación; MS – menor sin identificación.
NUM_DOCUMENTO	NULL	VARCHAR(20)	Identificación de la persona
COD_EAS	NULL	VARCHAR(10)	Indica el Código asignado al

			administrador del plan de beneficios en el SGSSS. En blanco cuando hay servicios de salud particulares
TIPO_USUARIO	NULL	INT	Indica el código que corresponde al tipo de régimen de la persona: 1;2;3;4;5.
PRIMER_APELLIDO	NULL	VARCHAR(30)	Primer apellido de la persona
SEGUNDO_APELLIDO	NULL	VARCHAR(30)	Segundo apellido de la persona
PRIMER_NOMBRE	NULL	VARCHAR(30)	Primer nombre de la persona
SEGUNDO_NOMBRE	NULL	VARCHAR(30)	Segundo nombre de la persona
EDAD	NULL	numeric(3, 0)	Indica la Edad de la persona al momento de la prestación del servicio.
SEXO	NULL	CHAR(1)	Indica el sexo de la persona: Hombre – Mujer
COD_DEPARTAMENTO	NULL	VARCHAR(2)	Indica el código de departamento de afiliación (codificación DANE)
COD_MUNICIPIO	NULL	VARCHAR(3)	Indica el código de

			Municipio de afiliación (codificación DANE)
ZONA	NULL	VARCHAR(1)	Indica la Zona de afiliación: Urbana; Rural; Cabecera

Tabla RIPS_URGENCIA			
Campo	Null	Tipo Campo	Descripción
REMISION	NOT NULL	VARCHAR(20)	Identificador único de la tabla
ORIGEN_RIPS	NULL	VARCHAR(1)	Indica quien reporta el RIPS si es la IPS o Administradora
COD_EAS	NULL	VARCHAR(10)	Indica el Código asignado al administrador del plan de beneficios en el SGSSS. En blanco cuando hay servicios de salud particulares
NUM_FACTURA	NOT NULL	VARCHAR(20)	Número que corresponda al sistema de numeración consecutiva según las disposiciones de la DIAN
COD_IPS	NOT	VACRHAR(15)	Código asignado en el

	NULL		SGSSS a los prestadores de servicios de salud que se hayan registrado en el "Registro Especial de Prestadores de Servicios de Salud"
COD_TIPO_DOCUMENTO	NULL	VARCHAR(2)	Indica el tipo de identificación de la persona: RC – registro civil; CC – cedula de ciudadanía; CE – cedula extranjera; PA – pasaporte; TI – tarjeta de identidad; AS – adulto sin identificación; MS – menor sin identificación.
NUM_DOCUMENTO	NULL	VARCHAR(20)	Identificación de la persona
FECHA_INGRESO	NULL	DATETIME	Fecha de ingreso del usuario a observación formato dd/mm/aa
HORA_INGRESO	NULL	DATETIME	Hora de ingreso del usuario a observación formato hh/mm
COD_CAUSA_EXTERNA	NOT NULL	INT	Indica cual es el Identificador de la causa que origina el servicio de salud desde 01 hasta 15.
COD_DX_SALIDA	NULL	VARCHAR(5)	Indica el Código del

			diagnóstico a la salida del usuario, según la Clasificación Internacional de Enfermedades vigente
DESTINO_SALIDA	NULL	INT	Indica el destino de la persona a la salida de la observación: (Alta de urgencias; Remisión a otro nivel de complejidad; Hospitalización)
ESTADO_SALIDA	NULL	INT	Indicador para determinar la condición de salida de la persona: (Vivo (a) - Muerto (a))
FECHA_SALIDA_OBS	NULL	DATETIME	Indica la fecha de salida de la persona de observación formato dd/mm/aa
HORA_SALIDA_OBS	NULL	DATETIME	Indica la hora de salida de la persona de observación formato hh:mm

Tabla RIPS_CONSULTA			
Campo	Null	Tipo Campo	Descripción
REMISION	NOT NULL	VARCHAR(20)	Identificador único de la tabla
ORIGEN_RIPS	NULL	VARCHAR(1)	Indica quien reporta el RIPS si es la IPS o Administradora
COD_EAS	NULL	VARCHAR(10)	Indica el Código asignado al administrador del plan de beneficios en el SGSSS. En blanco cuando hay servicios de salud particulares
NUM_FACTURA	NOT NULL	VARCHAR(20)	Número que corresponda al sistema de numeración consecutiva según las disposiciones de la DIAN
COD_IPS	NOT NULL	VACRHAR(15)	Código asignado en el SGSSS a los prestadores de servicios de salud que se hayan registrado en el "Registro Especial de Prestadores de Servicios de Salud"

COD_TIPO_DOCUMENTO	NULL	VARCHAR(2)	Indica el tipo de identificación de la persona: RC – registro civil; CC – cedula de ciudadanía; CE – cedula extranjera; PA – pasaporte; TI – tarjeta de identidad; AS – adulto sin identificación; MS – menor sin identificación.
NUM_DOCUMENTO	NULL	VARCHAR(20)	Identificación de la persona
FECHA_CONSULTA	NULL	DATETIME	Fecha de ingreso del usuario a observación formato dd/mm/aa
COD_CONSULTA		VARCHAR(8)	Indica el Código de la consulta definido en el Sistema, según los codificadores vigentes.
COD_CAUSA_EXTERNA	NOT NULL	INT	Indica cual es el Identificador de la causa que origina el servicio de salud desde 01 hasta 15.
COD_DX_SALIDA	NULL	VARCHAR(5)	Indica el Código del diagnóstico a la salida del usuario, según la

			Clasificación Internacional de Enfermedades vigente
TIPO_DX_PRINCIPAL	NULL	INT	Identificador para determinar si el diagnóstico es confirmado o presuntivo: 1;2;3.

Tabla TIPO_DOCUMENTO			
Campo	Null	Tipo Campo	Descripción
COD_TIPO_DOCUMENTO	NOT NULL	CHAR(28)	Indica tipo de documento de la persona
NOMBRE_TIPO_DOCUMENTO	NOT NULL	CHAR(10)	Indica el nombre del tipo de documento de la persona RC – registro civil; CC – cedula de ciudadanía; CE – cedula extranjera; PA – pasaporte; TI – tarjeta de identidad; AS – adulto sin identificación; MS – menor sin identificación.

Tabla TIPO_REGIMEN			
Campo	Null	Tipo Campo	Descripción
TIPO_USUARIO	NOT NULL	INT	Indica el código que corresponde al tipo de régimen de la persona
NOMBRE_TIPO	NOT NULL	CHAR(10)	Indica el nombre que determina la condición del usuario en relación con el SGSSS (Contributivo; subsidiado; vinculado; particular; otro.

Tabla DEPARTAMENTOS			
Campo	Null	Tipo Campo	Descripción
COD_DPTO	NOT NULL	CHAR(2)	Indica el código de departamento de afiliación (codificación DANE)
NOM_DPTO	NOT NULL	VARCHAR(50)	Indica el nombre del departamento

Tabla MUNICIPIOS			
Campo	Null	Tipo Campo	Descripción
COD_MPIO	NOT	CHAR(2)	Indica el código del

	NULL		Municipio de afiliación (codificación DANE)
NOM_MPIO	NOT NULL	VARCHAR(50)	Indica el nombre del municipio

Tabla CAUSA_EXTERNA			
Campo	Null	Tipo Campo	Descripción
COD_CUASA_EXTERNA	NOT NULL	CHAR(2)	Indica cual es el Identificador de la causa que origina el servicio de salud desde 01 hasta 15.
DESCRIPCION_CAUSA_EXTERNA	NOT NULL	VARCHAR(50)	Indica el nombre del tipo de causa externa: (1 Accidente de trabajo; 2 Accidente de tránsito; 3 Accidente rábico; 4 Accidente ofídico; 5 Otro de tipo de accidente; 6 Evento catastrófico; 7 Lesión por agresión; 8 Lesión auto infligida; 9 Sospecha de maltrato físico; 10 Sospecha de abuso sexual; 11 Sospecha de violencia sexual; 12 Sospecha de maltrato emocional; 13 Enfermedad

			general; 14 Enfermedad profesional; 15 Otra).
--	--	--	---

Tabla DIAGNOSTICO			
Campo	Null	Tipo Campo	Descripción
COD_DX_SALIDA	NOT NULL	VARCHAR(5)	Indica el Código del diagnóstico a la salida del usuario, según la Clasificación Internacional de Enfermedades vigente
NOMBRE_DX	NOT NULL	VARCHAR(255)	Indica el nombre del diagnóstico según la Clasificación Internacional de Enfermedades vigente

Tabla TIPO_CONSULTA			
Campo	Null	Tipo Campo	Descripción
COD_CONSULTA	NOT NULL	VARCHAR(5)	Indica el Código de la consulta definido en el Sistema, según los codificadores vigentes: de 1 hasta 10.
DESCRIPCION_CONSULTA	NOT NULL	VARCHAR(255)	Indica el nombre o la descripción de la consulta: Atención del parto (Pueperio) Atención del recién nacido Atención en planificación Familiar Detección de alteraciones de crecimiento y dllo Detección de alteración del desarrollo joven Detección de alteraciones del embarazo Detección de

			alteraciones del adulto
			Detección de alteraciones de agudeza visual
			Detección de enfermedad profesional
			No Aplica

9.2.5 Nivel de Granularidad

Habiéndose establecido las relaciones con los OLTP representados en las figuras 14 – 15 y 16, se seleccionaran los campos que contendrá cada perspectiva, ya que es a través de estos que se examinaran y filtraran los indicadores. Por tanto, se dará a conocer en detalle que significa cada campo residente en cada tabla de las diferentes fuentes o bases de datos.

Esta información fue recolectada primero examinando cada base de datos y segundo en las diferentes entrevistas y reuniones realizadas se consultó a los funcionarios encargados de los sistemas sobre algunos aspectos para comprender completamente su sentido.

Como puede apreciarse en los diagramas de entidad relación antes expuestos, los nombres de los campos son explícitos lo que permite que se deduzcan con facilidad, pero aun así se detallaran en la siguiente tabla:

Tabla 22. Perspectivas, Campos y Descripción

Perspectiva	Datos Disponibles	Descripción
Comuna Corregimiento	Cod_comuna	Indica el código de la Comuna
	Nom_Comuna	Indica el nombre de la Comuna
nivel – Sexo	Tipoidafiliado	Indica el tipo de identificación del afiliado: RC – registro civil; CC – cedula de ciudadanía; CE – cedula extranjería; CN – certificado nacido; PA – pasaporte; TI – tarjeta de identidad.
	Numeroidafiliado	Identificación del afiliado
	Sexo	Indica el sexo de la persona: Hombre – Mujer
	Nivel	Indica el nivel de estratificación del SISBEN
Instituciones Salud	Numero_certificado	Indicador único del certificado
	Cod_institucion	Indica el código de la institución de salud donde se dio el nacimiento
	Nom_institucion	Indica el nombre de la institución de salud
	Sexo	Indica el sexo de la persona: Hombre - Mujer
	Edad_madre	Indica la edad de la madre en años cumplidos
Causas CIE10	Cod_dx_salida	Indica cual es el Identificador de la causa que origina el

		servicio de salud desde 01 hasta 15.
	Nombre_dx_salida	Indica el nombre del diagnóstico según la Clasificación Internacional de Enfermedades vigente

Recolectada la información pertinente presentada en la tabla 10, fue determinado por los usuarios los datos que consideraban de interés para analizar los indicadores, obteniendo los siguientes resultados:

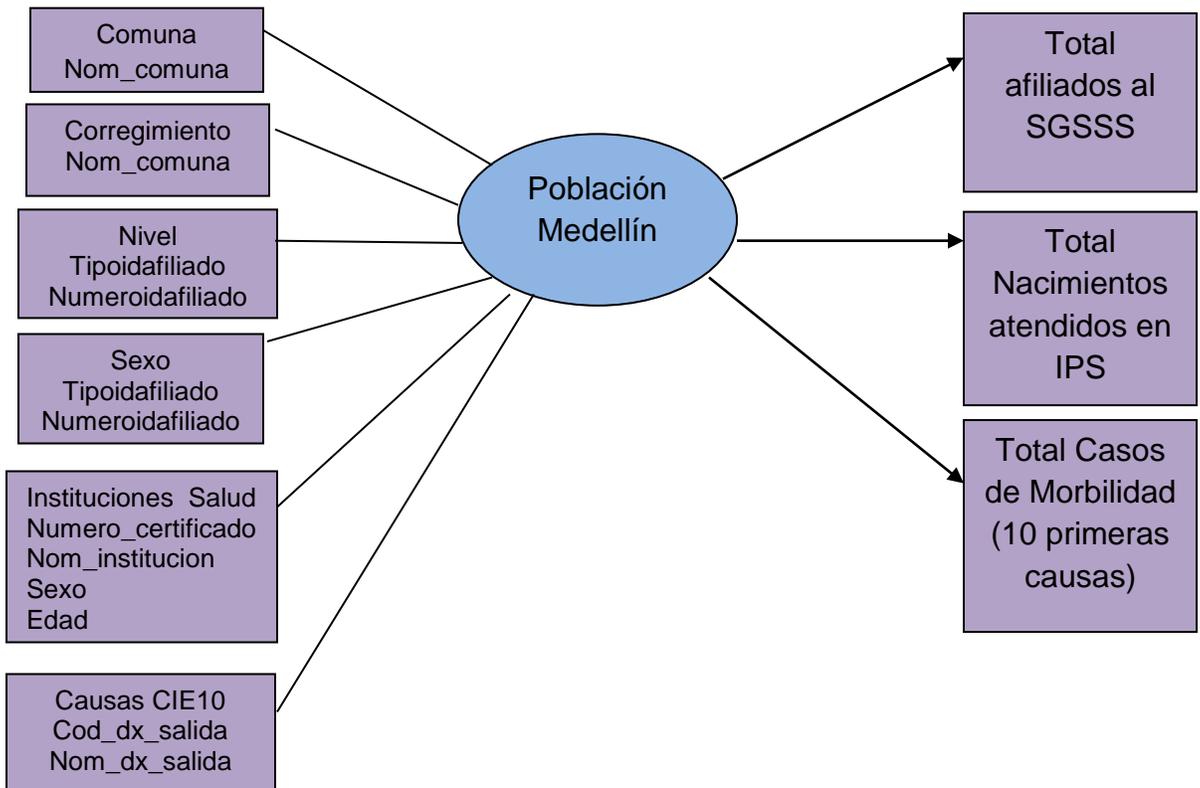
- Perspectiva “Comuna- Corregimiento” :
 - “Nom_comuna” de la tabla “comunas” ya que esta hace referencia al nombre de la comuna.
- Perspectiva “Nivel” – “Sexo” :
 - “Tipoidafiliado” – “Numeroidafiliado” de la tabla “Potenciales Beneficiarios” ya que estos hacen referencia al dato del afiliado.
- Perspectiva “Instituciones salud” :
 - “Numero_certificado” de la tabla “Detallecertnac” ya que este hace referencia al registro del nacido.
 - “Nom_institucion” de la tabla “detallecertnac” ya que este hace referencia al nombre de la institución.
 - “Sexo” y “Edad de la Madre” de la tabla “Detallecertnac” ya que este hace referencia al sexo del nacido y la edad de la madre para definir los grupos etareos.

- Perspectiva “CIE10”:
 - “Cod_dx_salida” – “Nombre_dx_salida” de la tabla “Diagnostico” la cual contiene el código y nombre de los diagnostico según la clasificación internacional de enfermedades vigentes y se relaciona con las tablas “RIPS_Urgencia” y “RIPS_Consulta” a través del campo “Cod_dx_salida”.

9.2.6 Modelo Conceptual Ampliado

Se presenta en la siguiente figura el Modelo Conceptual Ampliado con el fin de graficar los resultados obtenidos en los pasos anteriores, se colocara debajo de cada perspectiva los campos seleccionados y bajo cada indicador su respectiva formula o función

Figura 17. Modelo Conceptual Ampliado Información Secretaria de Salud



9.3 PASO3. MODELO LOGICO DEL DW

En este paso se confeccionará el modelo lógico de la estructura del Data Warehouse teniendo como base el modelo conceptual ampliado que ya fue creado (figura 17). Para ello se define el tipo de modelo que se utilizara y posteriormente se llevaran a cabo las acciones propias para diseñar las tablas de dimensiones y de hechos y las uniones pertinentes entre las tablas.

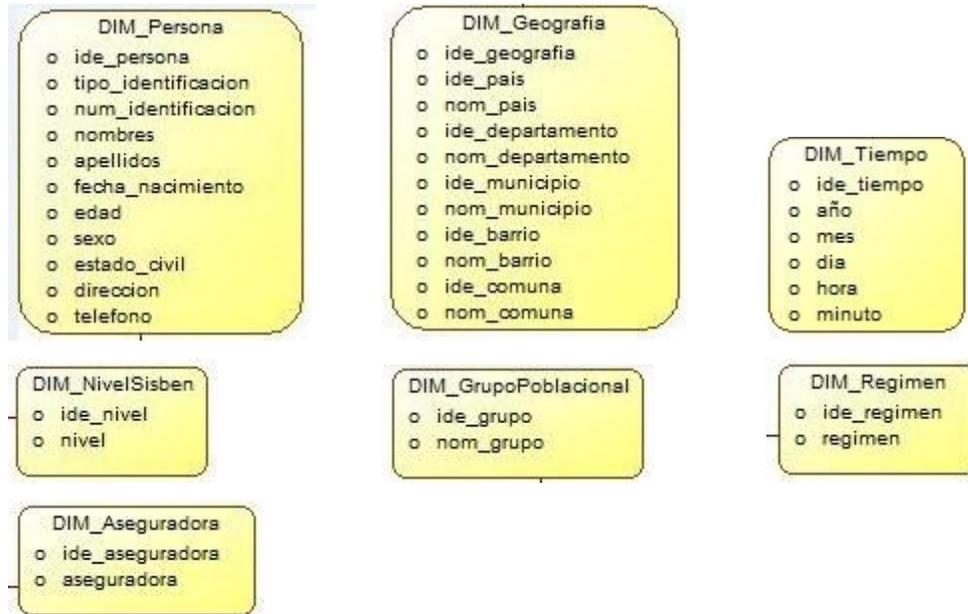
9.3.1 Tipo de Modelo Lógico del Data Warehouse

El esquema que se utilizará será en estrella, debido a sus características y ventajas. Este esquema será el que contendrá la estructura del depósito de datos, ya que se adapta mejor a los requerimientos y necesidades de los usuarios de la Secretaria de Salud. Es ideal por su simplicidad, rendimiento y velocidad para ser usado para análisis, pues permite indexar de forma individualizada sin que repercuta en el rendimiento de la base de datos. Además, Permite acceder tanto a datos agregados como de detalle, establecer jerarquías y niveles entre las dimensiones y reducir el número de joins entre tablas.

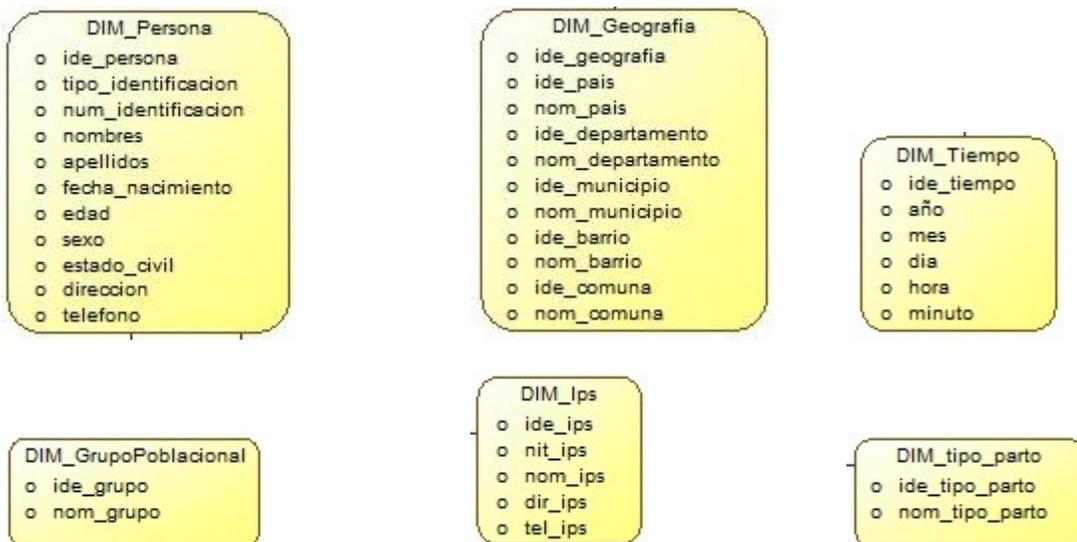
9.3.2 Tablas de Dimensiones

Cada una de las perspectivas definidas en el modelo conceptual (figura 17) constituirá una tabla de dimensión. Las siguientes son las tablas de dimensiones del Data Warehouse para:

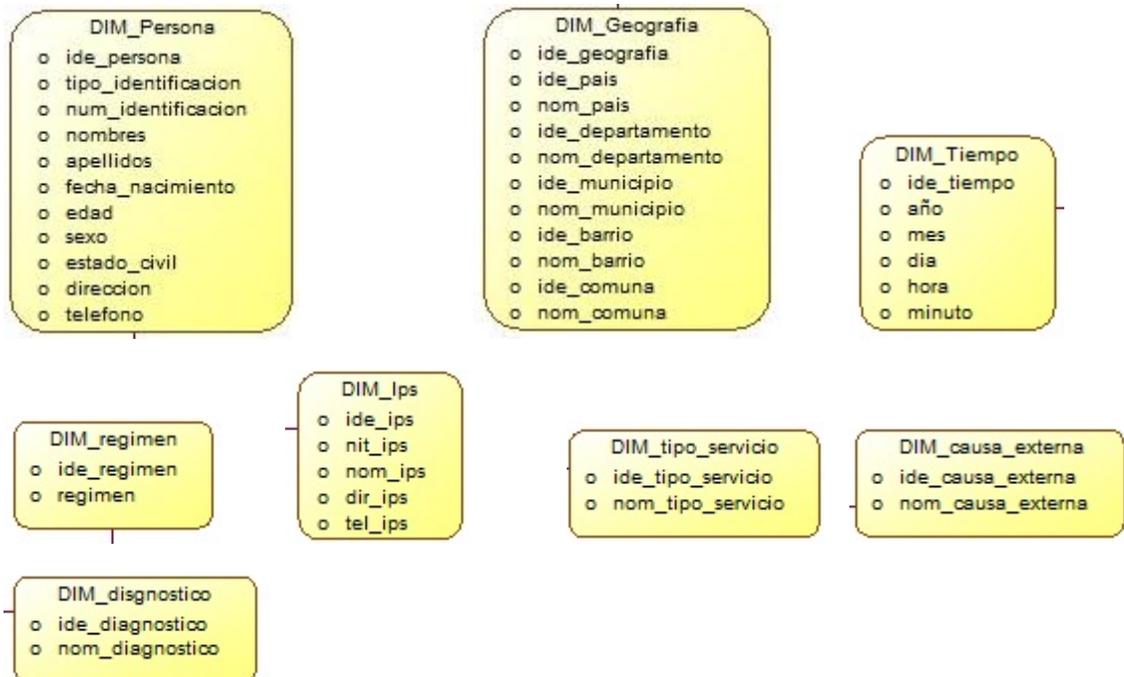
9.3.2.1 La Población afiliada al Sistema General de Seguridad Social en Salud:



9.3.2.2 Nacimientos atendidos en instituciones de salud de la población Medellín:



9.3.2.3 10 primeros motivos más frecuentes de Morbilidad población Medellín:



9.3.3 Tablas de Hechos

A continuación se definirán las tablas de hechos que son las que contendrán los hechos a través de los cuales se constituirán los indicadores.



9.3.4 Uniones

Se realizan las uniones para el esquema estrella entre las tablas de dimensiones y hechos correspondientes:

Figura 18. Población afiliada al Sistema General de Seguridad Social en Salud

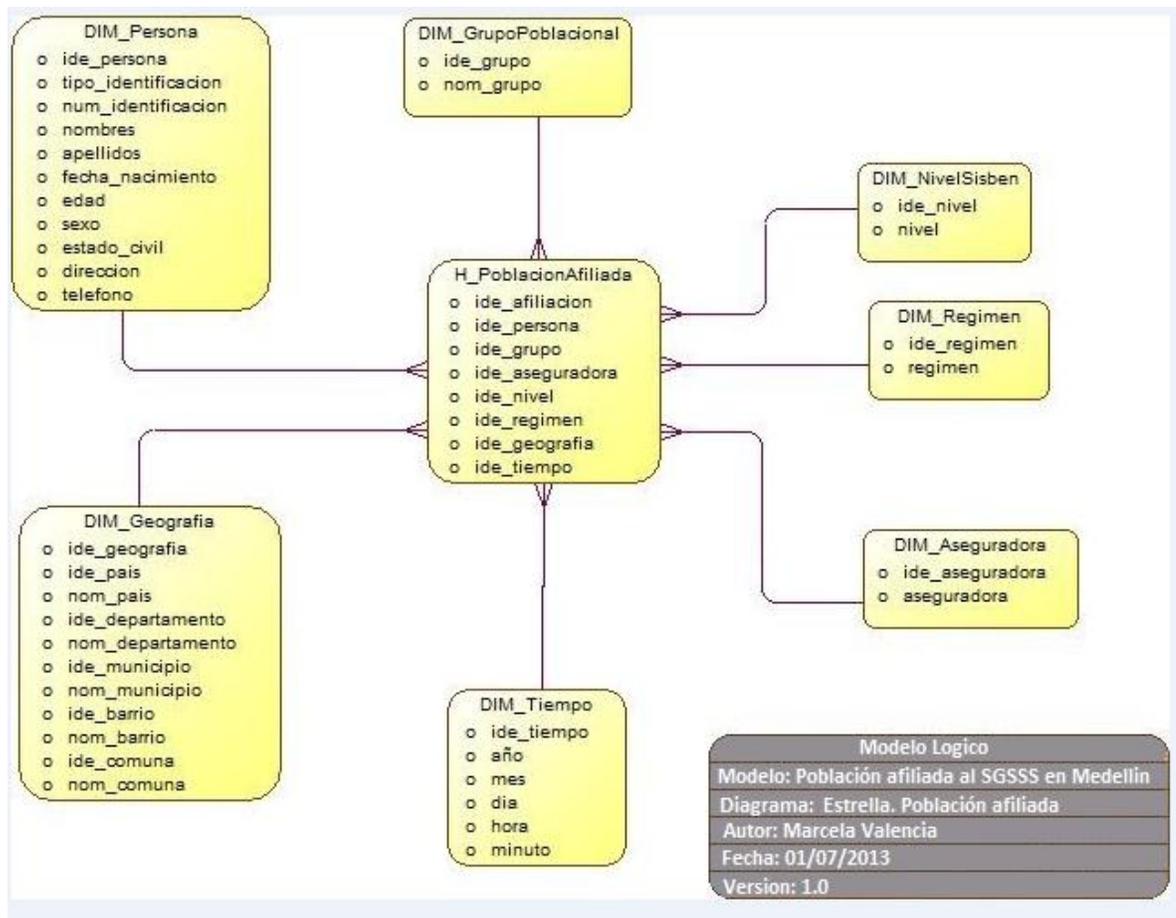


Figura 19. Nacimientos atendidos en instituciones de salud de la población Medellín.

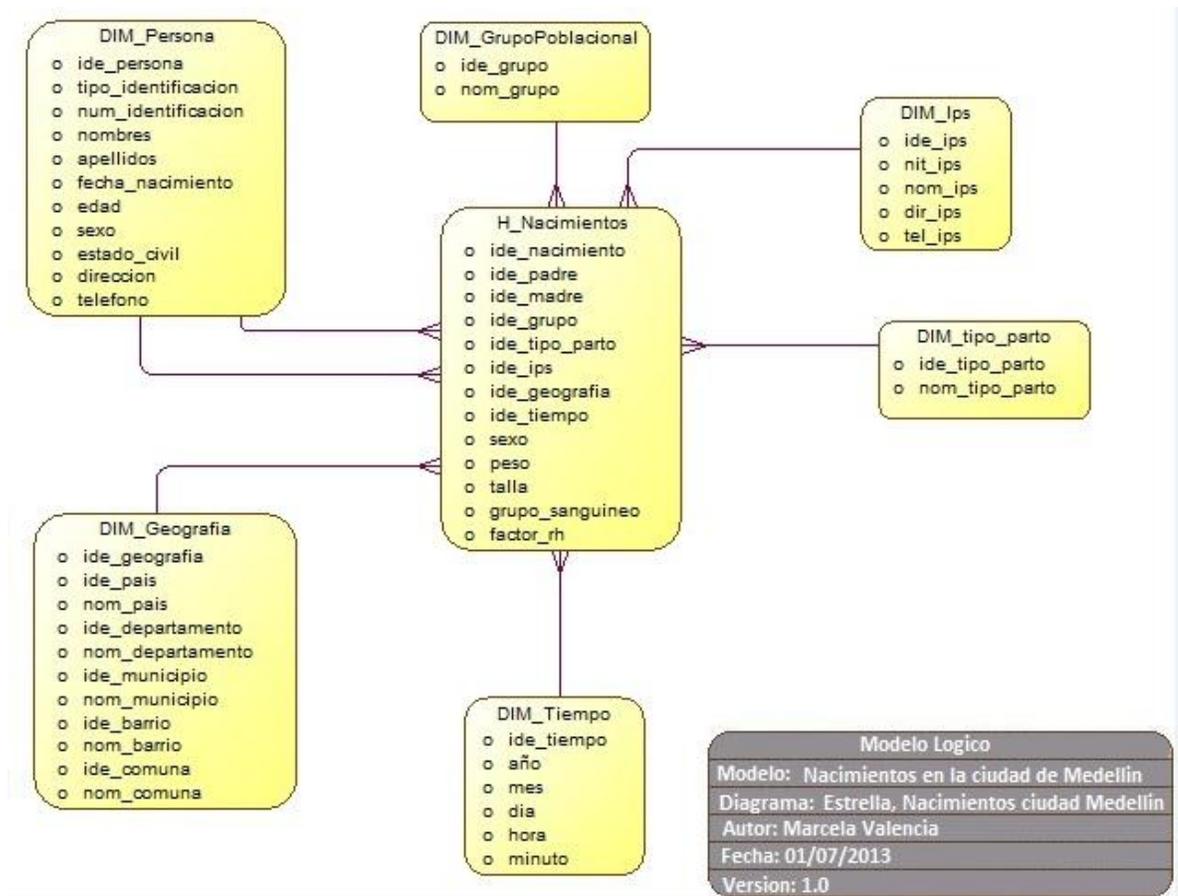
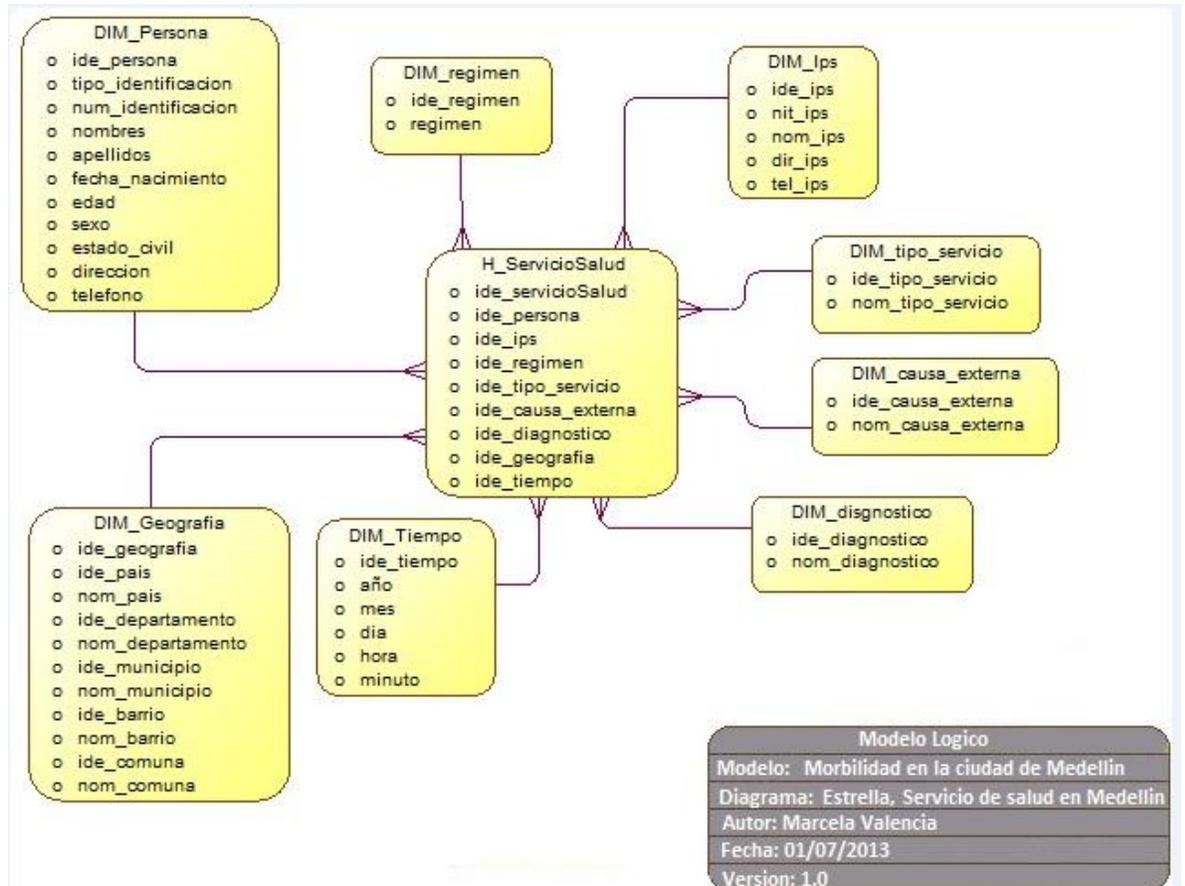


Figura 20. Diez (10) primeros motivos más frecuentes de Morbilidad población Medellín.



10 CONCLUSIONES

- En este proyecto se han presentado algunas estrategias para la implementación de una bodega de datos, de manera que basados en la revisión de la literatura se facilitara la elección de una de las metodologías que más se ajustara a las necesidades de información de la Secretaria de Salud de Medellín.
- A lo largo de este proyecto se presentan varios enfoques y modelos propuestos para la implementación de una bodega de datos, para posteriormente plantear la estrategia “Hefesto” como la metodología de utilidad para la Secretaria de Salud de Medellín que permitirá mejorar la toma de decisiones.
- Este estudio contribuye a la Secretaria de Salud de Medellín a conocer e incursionar en la Inteligencia de Negocios con herramientas como el Data Warehouse ampliamente mencionadas en la lectura referenciada; presentando las ventajas competitivas que ofrece y la relación que existe entre dato, información, conocimiento y toma de decisiones.
- La implementación de soluciones de Data Warehouse como las presentadas en este trabajo genera beneficios a la organización como: el fácil y rápido acceso a los datos, independencia del personal técnico para la obtención de reportes o generación de indicadores, información consolidada.
- En síntesis y conclusión del presente trabajo se puede decir que estas herramientas de Inteligencia de Negocios como el Data Warehouse (DW) convierten datos crudos en información valiosa y ayudan a los directivos a tomar decisiones.

11 RECOMENDACIONES

Este trabajo es un acercamiento inicial para organizaciones como la Secretaria de Salud de Medellín que requieren conocer las ventajas competitivas que ofrece la inteligencia de Negocios (BI), sirve como base para la consideración de una futura implementación como solución a las necesidades y dificultades que actualmente presenta la organización por el manejo individual que tiene de sus procesos y sistemas.

La principal recomendación de esta investigación es que la Secretaria de Salud de Medellín considere la metodología presentada de Data Warehouse como la solución más adecuada a las necesidades que demanda actualmente de manejo de información consistente para la toma de decisiones.

Se recomienda hacer énfasis en las ventajas competitivas que ofrece este tipo de herramientas para la organización de manera que se pueda estimular la innovación y la integración de los procesos y la información.

12 REFERENCIAS BIBLIOGRAFICAS

- Arriagada, I.; Aranda, V.; Miranda, F. (2005) Políticas y programas de salud en América Latina.
- Bernabeu, R. (2010) HEFESTO-Metodología para la construcción de un Data Warehouse.
- Beazley, H.; Boenisch, J. & Harden, D. (2003) La Continuidad del Conocimiento en las Empresas. Bogotá: Grupo Editorial Norma, p. 38.
- Davenport, T.H. & Harris, J.G., (2007). Competing on Analytics, The New Science of Winning. Boston : Harvard Business School Press.
- Devenport, T.; Prusak, L. (1998). Working knowledge: How organizations manage what they know.
- Jourdan, Z., Rainer R.K. & Marshall, T.E. (2008). Business Intelligence: An Analysis of Literature. Information System Management, 25(2), p. 121-131.
- Ley 1438 Reforma Del Sistema De Seguridad Social En Salud, Republica de Colombia
- Martínez, M.E. (2011). Desarrollo de un modelo de gestión del conocimiento en la cadena de suministros de la cadena agroalimentaria, Tesis Doctoral. Universidad politécnica de Madrid. Departamento de ingeniería de organización, administración de empresas y estadística.

- Ojeda, A.,(2011) Análisis, diseño e implementación de un Data warehouse de soporte de decisiones para un hospital del sistema de salud público.
- Tejada, A. (2010) Data Warehousing con procesamiento de datos textuales.
- Vásquez, R. (2003) Sistemas de Información para la Gestión de Instituciones de Salud.