



UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE COMPUTACIÓN

**DESARROLLO DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIO,
PARA LA OBTENCIÓN DE INDICADORES DEL PROCESO DE VISITAS
CLÍNICAS PARA LAS COMPAÑÍAS ASEGURADORAS.**

Trabajo especial de grado presentado ante la ilustre
UNIVERSIDAD CENTRAL DE VENEZUELA por la
Br. Diannellys Maria Marcano Rivas
para optar al título de Licenciado en Computación

Tutor: Franky Uzcátegui

2017

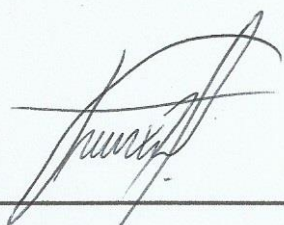
UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE COMPUTACIÓN

ACTA

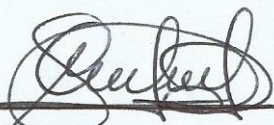
Quienes suscriben, miembros del jurado designado por el Consejo de la Escuela de Computación, para examinar el Trabajo Especial de Grado titulado **“Desarrollo de una Solución de Inteligencia de Negocio, para la obtención de indicadores del proceso de Visitas Clínicas para las compañías aseguradoras”** y presentado por la bachiller: Br. Diannellys Maria Marcano Rivas C.I.: 17053587, a los fines de optar al título de Licenciado en Computación, dejamos constancia de lo siguiente:

Leído como fue dicho trabajo, por cada uno de los miembros del jurado, se fijó el día 14 de Diciembre de 2017, a las 9:30 AM horas, para que la autora lo defendió en forma pública, lo que estos hicieron en la Sala 16 de la Escuela de Computación, mediante una presentación oral de su contenido, luego de lo cual respondieron a las preguntas formuladas. Finalizada la defensa pública del Trabajo Especial de Grado, el jurado decidió aprobar con la nota de 19 puntos.


En fe de lo cual se levanta la presente Acta, en Caracas el día 14 de Dic de 2017.



Prof. Franky Uzcátegui
(Tutor)



Profa. Di Vasta Concettina
(Jurado Principal)



Prof. Mirabal Jose
(Jurado Principal)

DEDICATORIAS

A mi padre Jesús Marcano
por mostrarme a partir del trabajo duro y responsable que cualquier
cosa que YO decida realizar, es posible. Te amo papi.

A mi madre Nellys de Marcano
por tu amor, tu paciencia, tu dedicación y esfuerzo para que yo culminara
cada una de mis metas. Mami tú eres mi sol.

A mi hermana Marinellys,
por darme un ejemplo fiel de que es la fuerza, la responsabilidad,
gracias a tu amor sé que nunca estaré sola. Te amo.

A mi hermana Diana,
por enseñarme la pasión por el trabajo, por tu amor infinito,
gracias por defenderme y cuidarme en cada segundo de mi vida.
Te amo morocha.

A mi hermano Jesús,
gracias por mostrarme día a día, que NADA es imposible. Te amo

A mi sobrinita Ema Paola.
por ser nuestra pequeña fuente de alegría.

Diannellys Marcano

AGRADECIMIENTOS

Primeramente, quiero agradecer a Dios Todopoderoso, por convertirme en la mujer que soy hoy, por siempre escuchar mis oraciones y ser el único testigo de mis sueños. Gracias por ayudarme a cumplir cada uno de ellos. Sigue iluminando mi camino y mis pasos.

En primer lugar quiero agradecer a mis Padres Nellys y Jesus, todo lo que soy es gracias a sus esfuerzos y sus sacrificios. Les agradezco por todo el amor que siempre me han dado, gracias a ustedes soy feliz. Los AMO inmensamente.

A mis hermanos, gracias por tanto: amor, alegría, sacrificio, por el constante apoyo y por la motivación para luchar contra todo obstáculo que se me presente. Mil palabras no bastarían para agradecerles su apoyo y su comprensión en los momentos difíciles. Los AMO mi equipo del Alma, mi corazón les pertenece.

Pero quienes han supuesto un verdadero apoyo en estos años han sido aquellos amigos que sin acabar de entender, en algunos momentos, la dedicación exclusiva a este proyecto, siempre han estado allí, muchas gracias: Maria E, Gustavo, Yesenia, Annielys, Evelyn, Monica, Marisela, Keina, Nelson, Fernando, Alejandra, Andreina, Luis J, Jorman, Estefanía, Vanesa, Angélica y en general a todos mis amigos que me acompañaron durante la carrera, empezamos muchos, terminamos pocos, pero igual siempre estuvimos juntos y seguimos en contacto aunque por diversas circunstancias ya no estemos todos sentados en el mismo lugar conversando y sonriéndole a la vida, espero que pronto nos reunamos nuevamente.

A mis Angelitos protectores que están en cielo: Sixta Pastora, Jesús M, Izaida M, Rosario R y Rafael R.

A mi tutor Franky Uzcátegui, por ser un apoyo incondicional y orientarme a lo largo del desarrollo de este proyecto, pero sobre todo por creer en mí!.; Gracias por guiarme, corregir mis errores y educarme para convertirme en un mejor profesional.

Y por último pero no menos importante, le agradezco enormemente a la *UNIVERSIDAD CENTRAL DE VENEZUELA*, que gracias a sus métodos propios y únicos, nos ha dado las herramientas necesarias para que nos formáramos como profesionales capaces de enfrentar y superar diferentes retos.

UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE COMPUTACIÓN
CENTRO DE INVESTIGACIÓN EN SISTEMAS DE INFORMACIÓN

Desarrollo de una solución de inteligencia de negocio, para la obtención de indicadores del proceso de visitas clínicas para las compañías aseguradoras.

Autor: Diannellys Maria Marcano Rivas.

Tutor: Prof. Franky Uzcátegui.

Fecha: 7/12/2017.

RESUMEN

El presente Trabajo Especial de Grado describe el desarrollo de una solución de Inteligencia de Negocio (BI, siglas en inglés), orientada hacia el proceso de Visitas Clínicas. Este proceso consiste en visitar clínicas privadas y proceder a la verificación de datos de presupuesto y de nivel de atención al paciente, después que la compañía de seguros se compromete a cubrir sus gastos con una Carta Aval. Para ello se construye una arquitectura de componentes tecnológicos, que permiten la obtención de indicadores clave para este proceso. Esta solución aprovecha las ventajas que ofrece la arquitectura de Inteligencia de Negocio, dando soporte a la generación de indicadores y reportes que permiten apoyar la toma de decisiones. Para lograr los objetivos definidos, se utilizó la metodología ascendente propuesta por Ralph Kimball, la herramienta Pentaho Data Integration para la construcción de los ETLs y la herramienta denominada Tableau para visualizar los indicadores y reportes.

Palabras Claves: Base de Datos, Sistemas de Información, Almacenes de Datos, Inteligencia de Negocio, Planificación de Proyectos, Indicadores de gestión.

Índice General

<i>INTRODUCCIÓN.....</i>	<i>15</i>
<i>CAPÍTULO 1.....</i>	<i>17</i>
<i>PROBLEMA DE INVESTIGACIÓN.....</i>	<i>17</i>
1.1 Situación actual.....	17
1.2 Planteamiento del problema.....	18
1.3 Objetivo general.....	22
1.4 Objetivo específicos.....	22
1.5 Justificación.....	22
1.6 Alcance.....	23
<i>CAPITULO 2.....</i>	<i>24</i>
<i>MARCO CONCEPTUAL.....</i>	<i>24</i>
2.1 Procesamiento de Transacciones en Línea (OLTP) vs Procesamiento Analítico en Línea (OLAP).....	24
2.1.1.- On-Line Transaction Processing.....	24
2.1.2.- On-Line Analytical Processing.....	24
2.1.3.- OLTP vs. OLAP.....	25
2.1. Inteligencia de Negocio.....	26
2.1.2 Inteligencia.....	26

2.1.2 Negocio	26
2.1.3 Definición.....	26
2.1.3 Niveles de Soluciones de Inteligencia de Negocios	28
2.1.4 Arquitectura	29
2.2 Almacén de Datos	31
2.2.1 Definición.....	31
2.2.2 Características.....	32
2.2.3 Arquitectura	34
2.2.4 Ventajas y desventajas de los almacenes de datos	36
2.2.5 Bodega de datos.....	37
2.3 Modelo dimensional.....	39
2.3.1 Definición.....	39
2.3.2 Dimensión.....	40
2.3.3 Tabla de dimensión	40
2.3.4 Hecho	41
2.3.4 Tabla de hechos.....	41
2.3.5 Esquema estrella.....	42
2.3.6 Granularidad	43
2.3.7 Jerarquía.....	43
2.3.8 Agregación.....	44
2.3 Indicadores de gestión	44
2.3.1 Clases de indicadores.....	45
2.3.2 Objetivos de los indicadores	46
2.3.3 Utilidad de los indicadores	47
2.4 Herramientas de Inteligencia de Negocio	48
2.4.1. Herramientas para la programación de procesos ETL y construcción de cubos.	48
2.4.2 Sistemas manejadores de base de datos	49

2.4.3 Herramienta de acceso y análisis de datos	50
<i>CAPÍTULO 3.....</i>	58
<i>MARCO METODOLÓGICO.....</i>	58
3.1 Método ascendente (Ralph Kimball).....	58
3.1.1 Ciclo de vida del modelo de negocio.	59
3.1.2.- Definición de los requerimientos del negocio.....	61
3.1.3 Diseño técnico de la arquitectura.....	61
3.1.4.- Selección de productos e instalación.	62
3.1.5.- Diseño del modelo dimensional.	63
3.1.6.- Diseño físico.....	67
3.1.7.- Diseño y construcción de procesos ETC.....	67
3.1.8.- Especificación y desarrollo de aplicaciones analíticas.	67
3.1.9.- Integración y despliegue.	68
3.1.10.- Mantenimiento y crecimiento.	68
<i>CAPÍTULO 4.....</i>	70
<i>MARCO APLICATIVO.....</i>	70
4.1 Planificación del proyecto.....	70
4.2 Definición de requerimientos.....	71
4.3 Diseño de la Arquitectura Técnica.....	74
4.4 Selección de Producto e Instalación.....	75
4.5 Modelo Dimensional.....	76
4.6 Diseño físico.....	84

4.7. Diseño de procesos ETL y desarrollo	87
4.7.1. Diseño de Procesos de ETL y Desarrollo (Almacén de datos)	87
4.7.2. Detallar los requerimientos según el modelo.....	89
4.8 Desarrollo de los requerimientos de Información.....	90
4.8.1 Verificar la calidad de los datos	90
4.8.2. Desarrollo y distribución de consultas	96
4.8.3 Generación de indicadores de gestión	98
4.8.3 Pruebas de aceptación de usuario.....	105
<i>CONCLUSIONES</i>	<i>110</i>
<i>RECOMENDACIONES.....</i>	<i>112</i>
<i>REFERENCIAS BIBLIOGRÁFICAS Y DIGITALES</i>	<i>113</i>

Tabla de ilustraciones

Ilustración 1: Proceso de visita clínica.....	20
Ilustración 2: Arquitectura de la solución.	21
Ilustración 3: Niveles Jerárquicos de la Organización	28
Ilustración 4: Arquitectura de una Solución de Inteligencia de Negocios	29
Ilustración 5: Características de un almacén de datos	32
Ilustración 6: Arquitectura de una solución BI.....	36
Ilustración 7: Tabla Dimensión.....	40
Ilustración 8: Modelo dimensional	42
Ilustración 9. Esquema estrella	43
Ilustración 10: Principales clases de indicadores.....	46
Ilustración 11: Arquitectura Tableau Server	53
Ilustración 12: Método ascendente.....	58
Ilustración 13: Tareas de la metodología de Kimball.....	59
Ilustración 14: Arquitectura de Diseño con sus herramientas.	74
Ilustración 15: Modelo Dimensional.....	77
Ilustración 16: Jerarquía de las dimensiones.....	78
Ilustración 17: Dimensión Tiempo	85

Ilustración 18: Diseño lógico de las dimensiones	86
Ilustración 19: ETL de la Dimensión Presupuesto con los campos adicionales.....	86
Ilustración 20: ETL de la Dimensión Visitador Clínico.	87
Ilustración 21: Trabajo para implementar el modelo dimensional.....	87
Ilustración 22: ETL Tabla de hecho Visita Clínica	88
Ilustración 23: Verificación de la dimensión visita clínica.....	91
Ilustración 24: Verificación de la cantidad de los datos de la dimensión visitador clínico.	92
Ilustración 25: Verificación de la tabla dimensión paciente beneficiario.....	92
Ilustración 26: Verificación de la cantidad de los datos de la dimensión paciente beneficiario.....	93
Ilustración 27: Verificación de fuente con la dimensión presupuesto.....	93
Ilustración 28: Verificación de la cantidad de los datos de la dimensión presupuesto.	94
Ilustración 29: Verificación de fuente con la dimensión carta aval.....	94
Ilustración 30: Verificación de la cantidad de los datos de la dimensión carta aval.	95
Ilustración 31: Verificación de fuente con la Dimensión Estado.....	95
Ilustración 32: Verificación de la cantidad de los datos de la Dimensión Estado.....	96
Ilustración 33: Tabla de hechos con sus dimensiones asociadas.....	97
Ilustración 34: Cuadro de Mando de Visita Clínica.....	97
Ilustración 35: Cantidad de visitantes clínicos por estatus	98

Ilustración 36: Porcentaje de visitantes clínicos por estatus.....	99
Ilustración 37: Cantidad de visitantes clínicos por estatus en un periodo de tiempo (Año, mes, día, trimestre).....	99
Ilustración 38: Cantidad de pacientes beneficiarios según el estado en el que se realizó la solicitud de visita clínica	100
Ilustración 39: Cantidad de pacientes Beneficiarios por periodo de tiempo (Año, Mes, Trimestre, Día de la Semana)	101
Ilustración 40: Monto pendiente de pagos por cartas avales activas por periodo (Año, Trimestre, Mes, Día de la Semana) de tiempo y clínica.	101
Ilustración 41: Cantidad pendiente de cartas avales activas por periodo de tiempo y clínica.....	102
Ilustración 42: Variación de cantidad pendiente de presupuesto en porcentaje.	102
Ilustración 43: Varianza de gastos de presupuestos por año	103
Ilustración 44: Cantidad de cartas avales por estatus.	103
Ilustración 45: Porcentaje de cartas avales por estatus.	104
Ilustración 46: Cantidad de cartas avales por estatus por periodo de tiempo (Año, trimestre, mes, día de la semana).	104
Ilustración 47: Interfaz encuesta de aceptación.....	105
Ilustración 48: Resultados de la pregunta N° 1. Fuente: Prueba de aceptación con usuarios finales	106
Ilustración 49: Resultados de la pregunta N° 2. Fuente: Prueba de aceptación con usuarios finales	106
Ilustración 50: Resultados de la pregunta N° 3. Fuente: Prueba de aceptación con usuarios finales	107

Ilustración 51: Resultados de la pregunta N° 4. Fuente: Prueba de aceptación con usuarios finales 107

Ilustración 52: Resultados de la pregunta N° 5. Fuente: Prueba de aceptación con usuarios finales 108

Ilustración 53: Resultados de la pregunta N° 6. Fuente: Prueba de aceptación con usuarios finales 108

Índice de tablas

Tabla 1: OLTP Vs OLAP.....	25
Tabla 2: Diferencias entre sistema transaccional y almacén de Datos.	33
Tabla 3: Diferencias entre Almacén de Datos y Bodega de Datos.....	39
Tabla 4: Cuadro comparativo entre manejadores de bases de datos.....	50
Tabla 5: Comparaciones entre las soluciones de inteligencia de negocios.....	57
Tabla 6: Plan de actividades.....	70
Tabla 7 Tabla de especificación de los indicadores propuesto	72
Tabla 8: Dimensión Tiempo	79
Tabla 9: Dimensión Visitador Clínico	79
Tabla 10: Dimensión Paciente Beneficiario.....	80
Tabla 11: Dimensión Carta Aval.....	81
Tabla 12: Dimensión Presupuesto.	82
Tabla 13: Dimensión Estado.....	83
Tabla 14: Tabla resumen de encuesta de aceptación.....	106

INTRODUCCIÓN

En los últimos años, las empresas han estado experimentando una gran presión competitiva que los obliga a revisar sus métodos y herramientas utilizadas, y si es necesario, realizar inversiones en tecnología para desarrollar nuevos instrumentos que mejoren la prestación de sus servicios. La globalización, que ha introducido a los mercados nuevos competidores, junto con la presión de nuevas normativas por parte del gobierno nacional, son factores que ejercen presión para bajar los costos y brindar, a la vez, mejores servicios.

Algunas de las organizaciones que se ha sumado a la implantación de herramientas tecnológicas son las compañías aseguradoras, las cuales ofrecen un respaldo financiero a sus clientes mediante un conjunto de opciones de planes de cobertura. Existe un tipo de póliza médica la cual permite a sus clientes planificar la hospitalización de su asegurado; este servicio es denominado visita clínica. Consiste en verificar el presupuesto de hospitalización entregado por el asegurado en solicitud de la carta aval, con la clínica prestadora del servicio.

Es por lo antes mencionado que este Trabajo Especial de Grado propone estudiar esta la información descriptiva de los datos correspondiente al proceso de visita clínica, utilizando las metodologías, esquemas y arquitecturas que proporcionan una solución a nivel de Inteligencia de Negocios para la obtención de indicadores claves para el monitoreo y planificación de la supervisión de clínicas.

Para el desarrollo del sistema es necesario definir los conceptos y características, en el área de Computación, tales como: Base de Datos, Sistema Manejador de Base

de Datos, sistema de información, conceptos relacionados al proceso de visita clínica, indicadores de gestión, los componentes de la arquitectura de una solución de inteligencia de negocios como el almacén de datos, las fuentes de datos y las herramientas de Acceso a los Datos. Abarcando para cada tópico las ventajas, desventajas y procesos más relevantes en el desarrollo una solución eficiente y funcional. Finalmente se presenta la solución de Inteligencia de Negocios con la arquitectura, las herramientas e indicadores seleccionados para dar solución al problema planteado. De acuerdo con lo señalado, el trabajo se estructuró en cuatro capítulos, a saber:

En el Capítulo 1, se plantea formalmente el problema a abarcar, dándole un contexto completo y justificándolo, así como se define el alcance y los objetivos de este trabajo.

En el Capítulo 2, se definen todos los conceptos necesarios para poder desarrollar la solución al problema presentado en el capítulo anterior y las técnicas necesarias que ayudarán a alcanzar los objetivos.

En el Capítulo 3, se explica la metodología a utilizar para el desarrollo del trabajo.

En el Capítulo 4, se explica formalmente el desarrollo de la aplicación, siguiendo las actividades planteadas en la metodología seleccionada, y se expone el producto final.

Para finalizar, se ofrecen las conclusiones alcanzadas y algunas propuestas para posibles trabajos futuro.

CAPÍTULO 1

PROBLEMA DE INVESTIGACIÓN

1.1 Situación actual

Las compañías aseguradoras que ofrecen entre sus servicios una póliza de hospitalización o de salud, deben despachar a los asegurados que vayan a hacer uso de su póliza de salud, un documento conocido como carta aval, el cual es garantía del cubrimiento del monto que en esta carta se especifica. La compañía aseguradora tramita la carta aval en base al presupuesto del tratamiento que se va a practicar el asegurado. Una vez que el asegurado se ha practicado el tratamiento se lleva a cabo un proceso conocido, en la compañía aseguradora sobre la cual se centra el presente estudio, como Visita Clínica.

Este proceso es realizado entre la compañía aseguradora y la institución prestadora del servicio (generalmente una clínica) y consiste en verificar el total de gastos que debe cubrir la compañía aseguradora en cuanto al tratamiento del asegurado. El proceso requiere que el personal de la compañía involucrado en el mismo, tenga que mediar con varios documentos, entre ellos la carta aval y el presupuesto del tratamiento en cuestión, para así totalizar los gastos según los cambios que hayan surgido en los mismos luego de practicado el tratamiento.

1.2 Planteamiento del problema

El proceso de visita clínica inicia cuando algunos de los asegurados de la compañía aseguradora deciden hacer uso del seguro de hospitalización al cual se encuentra suscrito, por tanto debe entregar a la clínica en la cual se haría el tratamiento, una carta aval por parte de la compañía aseguradora que garantice el cubrimiento de los gastos amparados de la hospitalización. El asegurado debe solicitar esta carta a la compañía aseguradora y para ello debe entregar ciertos recaudos, entre ellos la evaluación médica del paciente y el presupuesto de gastos de hospitalización debidamente desglosado.

Una vez que la carta aval es emitida, esta es entregada por el asegurado a la clínica afiliada, la cual posteriormente se comunicara con la compañía aseguradora para solicitar la activación de la carta aval. Cuando la clínica afiliada se comunica con la compañía aseguradora para activar la carta aval, un analista de seguimiento de carta aval se encarga de solicitar una visita clínica para dicho caso. La solicitud hecha por el analista de seguimiento de carta aval está conformada por una planilla, la cual es llenada por este analista con los datos requeridos del asegurado y del beneficiario del seguro, además de la carta aval y el presupuesto de gastos de hospitalización entregado por el asegurado. Esta solicitud es enviada al coordinador de visitas clínicas, el cual es encargado de asignar dicha solicitud a algún visitador clínico de la compañía. Para ello, debe consultar un registro de visitadores clínicos y asignar dicha visita clínica a algún visitador que se encuentre disponible. Esta asignación se hace llenando ciertos campos, en la misma planilla que recibe con la solicitud de visita, con los datos del visitador clínico asignado y luego se envía esta asignación al visitador junto con el resto de la información que conforma la solicitud de visita clínica.

Esta asignación es recibida por el visitador, el cual debe dirigirse al departamento de presupuesto de la clínica para verificar el presupuesto entregado en solicitud de la carta aval, y luego debe realizar una encuesta al beneficiario del

seguro, con la intención de recaudar información acerca de la calidad de los servicios prestados por la clínica afiliada. El visitador debe ir a la compañía de seguros y hacer entrega de la documentación al coordinador de visitas, el cual se encargaría de registrar la realización de la visita clínica y posteriormente entregara la documentación de la visita al analista de seguimiento de carta aval, el cual a su vez, en base a la información recibida, prosigue con el protocolo llevado a cabo en la compañía aseguradora.

El proceso se automatizó a partir de un portal web en donde se ven involucrado los roles de analista de seguimiento de carta aval, coordinador de visita clínica, visitador clínico, Administrador. En primer lugar se encuentra el analista, el cual luego de la activación de la carta aval solicita una visita clínica para dicha carta aval. Luego, se registra en el sistema la solicitud de esta visita clínica y se envía al coordinador de visitas una notificación indicando que tiene una solicitud de visita pendiente por asignar. Una vez recibida la notificación, el coordinador puede consultar la lista de visitadores clínicos que coordina y hace la asignación de la visita a alguno de estos. Después de haber realizado la asignación, el sistema envía una notificación al visitador clínico en cuestión indicando que tiene una visita por realizar.

El visitador clínico luego de recibir la notificación, puede visualizar e imprimir la documentación necesaria para llevar a cabo el proceso de visita clínica y también cargar la información obtenida en la visita clínica. Posteriormente, esta información cargada es visualizada por el coordinador de visitas, quien se encarga de verificar la información cargada y autoriza el envío de esta información (en caso de que no haya problema) al analista de seguimiento de carta aval para que la pueda visualizar y continuar con el protocolo seguido por la compañía de seguros. En caso de que el coordinador detecte o visualice alguna incoherencia con la información cargada por el visitador clínico, puede ordenar una revisión sobre esta información, lo cual envía al visitador clínico una notificación indicando que debe hacer una revisión sobre la información cargada en el proceso de visita clínica. Una

vez hecha la corrección, se notifica al coordinador y se continúa con el flujo del proceso. Una vez almacenada toda la información de la visita clínica, los analistas deben obtener indicadores de gestión para posteriormente registrarlos en hojas de cálculo como se muestra en la ilustración 1.

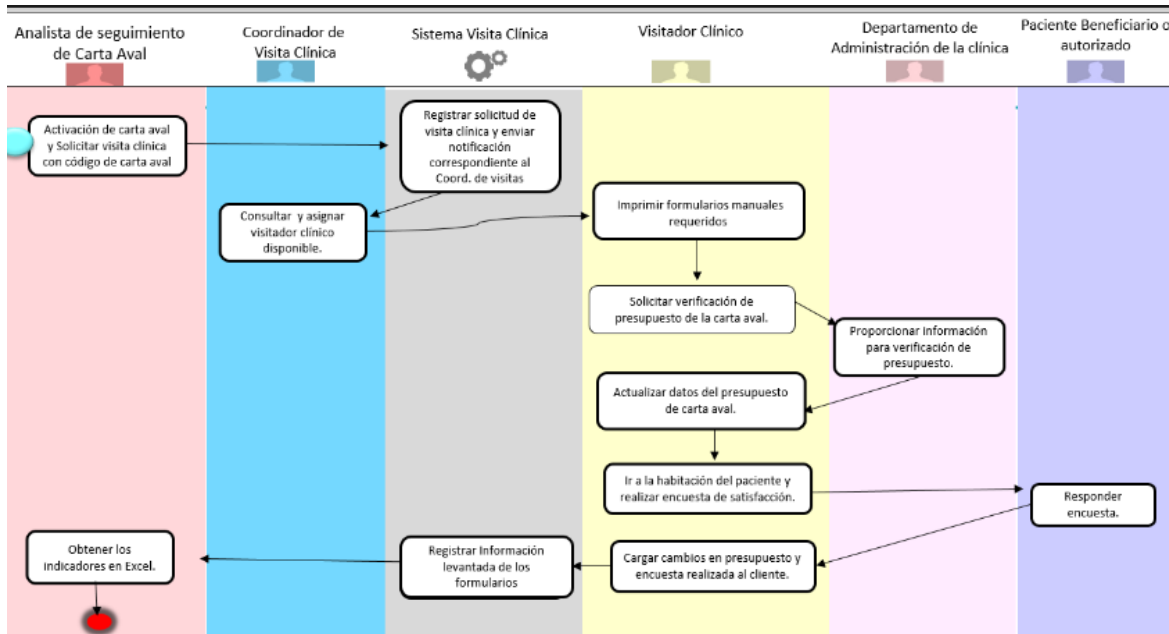


Ilustración 1: Proceso de visita clínica.

El proceso de las visitas a clínicas de una compañía aseguradora de forma convencional, siempre ha conllevado una serie de problemas, entre los cuales se destacan el incumplimiento en los tiempos de entrega, gastos mayores al presupuesto asignado inicialmente y en general el paso por alto de acciones importantes del negocio. Las compañías aseguradoras manejan un gran volumen de datos almacenados y en constante aumento. Sin embargo, este incremento no se traduce en un mayor conocimiento real, puesto que es verdaderamente difícil poder procesar tal magnitud de datos con los métodos clásicos y extraer información útil para el planteo de estrategias y toma de decisiones. La alta gerencia de las empresas encuentra dificultades al momento de elaborar reportes gerenciales y analizarlos.

Diariamente se enfrentan a la problemática de obtener más información y menos tiempo para analizarla, aunado a esto, existe una alta dependencia con el departamento de informática, ya que en la mayoría de los casos son los encargados de elaborar los reportes operativos y gerenciales requeridos por la alta gerencia de la empresa, lo que trae como consecuencia demora en la entrega de la información. El problema es que el procesamiento de transacciones solo responde a las necesidades de la operación de visita clínica y no a las necesidades de la alta gerencia la cual requiere responder preguntas asociadas a la situación del proceso, para poder tomar decisiones de manera rápida y eficiente.

Dada la situación actual, con sus respectivos problemas anteriormente señalados, se plantea como solución la creación de un modelo dimensional que permite obtener indicadores de gestión de visita clínica, para posteriormente acceder de forma rápida y flexible a la información relevante y oportuna para la toma de decisiones. El levantamiento de requerimientos de la solución vendrá dada por el modelo transaccional que utiliza la organización para manejar el proceso de cartas avales.



Ilustración 2: Arquitectura de la solución.

A partir del modelo transaccional se construyó un modelo dimensional que posee toda la información necesaria para el estudio de los indicadores generados a partir de las hojas de cálculo construidas por el analista de carta aval. Tales Indicadores fueron construidos con la finalidad de satisfacer los requerimientos del negocio, en tal sentido estas medidas se apoyaran en una herramienta de visualización las cuales estarán conectadas con el repositorio de datos, dando lugar a una solución de inteligencia de negocios, la cual brindara la capacidad de que el usuario manipule las variables del indicador de forma que pueda obtener el requerimiento solicitado.

1.3 Objetivo general

Desarrollar una solución de inteligencia de negocio, para la obtención de indicadores del proceso de visitas clínicas para las compañías aseguradoras.

1.4 Objetivo específicos

- Levantar requerimientos de indicadores del proceso.
- Diseñar el modelo dimensional que soporte los indicadores de gestión. propuestos utilizando la metodología de Ralph Kimball.
- Construir el Almacén de Datos.
- Desarrollar los programas de extracción, transformación y carga desde la estructura de datos operativa al Almacén de Datos.
- Construir los indicadores.
- Realizar pruebas funcionales, no funcionales y de calidad de datos.

1.5 Justificación

Se pretende desarrollar una solución de Inteligencia de Negocios que se adapte a los requerimientos de la organización con respecto a la gestión de la visita clínica de una compañía aseguradora, el cual permitirá automatizar los indicadores que

se encuentran en documentos generados de forma manual a través de los analistas de seguimiento de carta aval. Estos documentos son creados a medida que son solicitados por los directivos de la compañía aseguradora, es por ello que se pueden presentar inconvenientes como: carga manual de la información con errores y retrasos en la creación de los reportes. La utilización de una solución de inteligencia de negocio permitirá a la organización disponer de una plataforma computarizada, orientada a la toma de decisiones utilizando tecnología de software libre como PostgreSQL para la construcción de la base de datos, Pentaho para el desarrollo de los distintos ETL's y Tableau como herramienta generadora de reportes la cual a pesar de no ser un software libre tiene una versión gratuita anual. A partir de las siguientes herramientas será posible la creación de un modelo de datos orientado al análisis del negocio, la integración de los datos transaccionales, la creación del Almacén de Datos y la creación de consultas al repositorio que muestren la información que será analizada.

1.6 Alcance

La solución presentada implementa un conjunto de indicadores asociados a los presupuesto, las cartas avals, los pacientes beneficiarios y los visitantes clínicos para hacer seguimiento y asegurar el uso eficiente del tiempo en cuanto a la disponibilidad de los recursos humanos disponibles en el proceso de visita clínica. Adicionalmente, la visualización de dichos indicadores será bajo la plataforma de inteligencia de negocios de un Software libre. Se realizará el análisis, diseño e implementación de dicha solución que abarcará, desde diseño y extracción de los datos, además del Almacén de Datos (DWH), y la creación de los diferentes indicadores de gestión para el apoyo de la toma de decisiones, las consultas en tiempo real, de la información almacenada en el repositorio y los cuadros de mando parametrizables con la información necesaria en el almacén. No es una actividad de una sola vez, es un proceso de mejoramiento continuo.

CAPITULO 2

MARCO CONCEPTUAL

2.1 Procesamiento de Transacciones en Línea (OLTP) vs Procesamiento Analítico en Línea (OLAP)

Para el desarrollo de una solución de inteligencia de negocio es importante tener claro los conceptos y diferencias entre los sistemas OLTP y las tecnologías OLAP. Es por esto que a continuación se definen estos conceptos y se hace una breve comparación entre ellos.

2.1.1.- On-Line Transaction Processing

Sistemas diseñados para el manejo de datos, tanto para su almacenamiento como para su rápida recuperación. Comúnmente asociados con los sistemas manejadores de base de datos. Estos sistemas suelen estar diseñados bajo la arquitectura cliente - servidor. 21

2.1.2.- On-Line Analytical Processing

OLAP son tecnologías para el análisis de grandes volúmenes de datos con el fin de generar información útil para el usuario. Son usados frecuentemente para la toma de decisiones en las empresas a través de la manipulación de los datos corporativos.

Según (Ponniah, 2001), OLAP es “una categoría de tecnología de software, que permite a los analistas, gerentes y ejecutivos, obtener una mejor percepción de los datos, a través de un acceso rápido, consistente e interactivo, en una variedad de vistas posibles de la información que ha sido transformada, para reflejar la dimensionalidad real de la empresa de una manera que entienda el usuario”.

2.1.3.- OLTP vs. OLAP

La diferencia entre OLTP y OLAP se pueden observar en características tales como la integración de datos, el acceso y manipulación de los datos por parte de los usuarios, las tareas de los administradores en estos sistemas, las características de las transacciones y como se ven los datos en el tiempo. En la Tabla 2, se muestra una comparación bajo ciertos criterios:

Tabla 1: OLTP Vs OLAP

	OLTP	OLAP
Integración	Comúnmente no integradas	Debe ser integrada
	Cada tema de negocios puede tener información en diferentes sistemas	Toda la información referente a un tema, tiene una única fuente
	Diferentes plataformas de hardware	Posee un solo servidor lógico warehouse
Acceso y Manipulación de datos por parte de usuarios	Los usuarios son los que insertan, actualizan y borran los datos	Los usuarios solo realizan consultas
Administradores	Manipulan los datos a nivel de registros	Cargan y acceden a los datos en forma masiva
	Transacciones y procedimientos almacenados de validación a nivel de registros	Validaciones antes de la carga
Transacciones	Se manejan gran cantidad de transacciones por día	Pocas transacciones al día
	Si la transacción se realizo con éxito se asegura consistencia de los datos que intervinieron	Si la carga termina exitosamente se tiene consistencia asegurada de todo el conjunto de datos.
Tiempo	No existe un historial explícitamente	Está compuesta por fotografías de las bases de datos transaccionales

2.1. Inteligencia de Negocio

Antes de definir formalmente que es Inteligencia de Negocio, se presentan a continuación los siguientes términos que la conceptualizan.

2.1.2 Inteligencia

La *inteligencia* es información procesada y explotada a un muy alto nivel (Prior, 2007). Es la habilidad que se tiene para pensar abstractamente sobre un problema y la capacidad de organizar grandes cantidades de información para luego razonar y pensar sobre un aspecto en específico

2.1.2 Negocio

Un *negocio*, es la acción y el efecto de articular, disponer y hacer operativos un conjunto de medios, factores o elementos para conseguir un fin concreto (Andrade, 2005), en otras palabras, es un grupo de personas que trabajan en conjunto desempeñando diversas actividades para lograr un objetivo en común.

2.1.3 Definición

“La *Inteligencia de Negocios* son procesos, tecnologías y herramientas necesarias para transformar datos en información, información en conocimiento y conocimiento en planes de negocios rentables” (Loshin, 2003).

“Inteligencia de negocio es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un almacén de datos), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones.

El proceso de inteligencia de negocio incluye la comunicación de los descubrimientos y efectuar los cambios. Las áreas incluyen clientes, proveedores,

productos, servicios y competidores.” (Cano, 2007). Se puede descomponer detalladamente la definición:

- **Proceso Interactivo:** al hablar de inteligencia de negocio estamos suponiendo que se trata de un análisis de información continuado en el tiempo, no sólo en un momento puntual. Aunque evidentemente este último tipo de análisis nos puede aportar valor, es incomparable con lo que nos puede aportar un proceso continuado de análisis de información, en el que por ejemplo podemos ver tendencias, cambios, variabilidades, etc.
- **Explorar:** En todo proyecto de inteligencia de negocios hay un momento inicial en el que por primera vez accedemos a información que nos facilita su interpretación. En esta primera fase, lo que hacemos es “explorar” para comprender qué sucede en nuestro negocio; es posible incluso que descubramos nuevas relaciones que hasta el momento desconocíamos.
- **Analizar:** Pretendemos descubrir relaciones entre variables, tendencias, es decir, cuál puede ser la evolución de la variable, o patrones. Si un cliente tiene una serie de características, cuál es la probabilidad que otro con similares características actúe igual que el anterior.
- **Área de análisis:** Todo proyecto de inteligencia de negocio debe tener un objeto de análisis concreto. Nos podemos centrar en los clientes, los productos, las ventas, la cobranza, etc. que pretendemos analizar con detalle y con un objetivo concreto: por ejemplo, la reducción de costes, el incremento de ventas, el aumento de la participación de mercado, el ajuste de previsiones de venta, el cumplimiento los objetivos de venta presupuestados, etc.
- **Comunicar los resultados y efectuar los cambios:** Un objetivo fundamental del BI es que, una vez descubierto algo, sea comunicado a aquellas personas que tengan que realizar los cambios pertinentes en la organización para mejorar nuestra competitividad.

2.1.3 Niveles de Soluciones de Inteligencia de Negocios

La Inteligencia de Negocios muestra resultados, de acuerdo a las necesidades de los distintos niveles jerárquicos de la organización mostrados en la ilustración 3 se presenta una pirámide con los distintos tipos de BI, mostrando de izquierda a derecha el personal y a la derecha las herramientas o instrumentos que utilizan para realizar las tareas de la organización.



Ilustración 3: Niveles Jerárquicos de la Organización

Fuente: (Inteligencia de Negocios, 2017)

Apoyan a los distintos niveles de la organización:

- Nivel Operativo. Se permite a los empleados de la empresa recibir de forma oportuna, exacta y adecuada la información operativa, basándose en herramientas de trabajos como reportes, hojas de cálculos, manteniendo siempre un formato fijo cuya información se actualiza cada cierto tiempo.
- Nivel Táctico. Permite a los analistas de datos y a la gerencia media de una empresa, utilizar herramientas de análisis y consultas con el fin de obtener acceso a la información sin intervención de terceros.
- Nivel Estratégico. Permite que los directivos de la empresa pueda analizar y seguir día a día las tendencias, patrones, metas y objetivos estratégicos de la empresa. Un ejemplo de ello, lo constituye el cuadro de mando integral, entre otros.

2.1.4 Arquitectura

La arquitectura de una solución de Inteligencia de Negocios parte de los sistemas que contiene la empresa permite aplicar un cambio estructural, logrando la optimización de un proceso analítico. La integración de todos los elementos se encuentra expresada en la Ilustración 14, en donde se aprecian los diversos niveles y las diversas interacciones que se toman en cuenta en una solución de Inteligencia de Negocios. Una implementación completa de Inteligencia de Negocios trabaja con grandes cantidades de datos cambiantes, módulos, procesos y componentes.

Para una empresa u organización, una solución de Inteligencia de Negocios tiende a ser una de las piezas más complejas de software, debido a que se necesita de la integración de diversos sistemas que no tienen relación entre sí. Por esto, se puede decir que una solución de Inteligencia de Negocios va más allá de un software solamente.



Ilustración 4: Arquitectura de una Solución de Inteligencia de Negocios

Fuente: (Cano, 2007).

Una organización que maneja grandes cantidades de datos necesita, para maximizar sus ingresos e incrementar su eficiencia, monitorear una serie de indicadores claves que informan el funcionamiento de la organización en tiempo

real. Las piezas que forman parte de una solución de Inteligencia de Negocios completa y bien implementada, serán mostradas a continuación. Así se aprecia claramente la complejidad de una implementación apropiada y como estos componentes interactúan entre ellos.

2.1.4.1 Fuentes de Datos

Son los datos que se obtienen de los procesos diarios y que se requieren para el análisis del negocio, estos pueden ser extraídos de diferentes lugares: de procedencia externa a la organización o pueden ser los mismos datos arrojados por los procesos operacionales de la organización.

2.1.4.2 Proceso de Extracción, Transformación y Carga (ETL)

Consiste en la extracción de los datos provenientes de las distintas fuentes disponibles, la adaptación de los mismos según un formato o estándar requerido para su almacenamiento y la carga en otro repositorio de almacenamiento.

2.1.4.3 Metadata o Diccionario de datos

Se busca almacenar los datos de una forma que maximice su flexibilidad, facilidad de acceso y administración.

2.1.4.4 Motor OLAP

Provee capacidad de cálculo, consultas, funciones de planeamiento, pronóstico y análisis de escenarios en grandes volúmenes de datos, sin embargo, en la actualidad existen otras alternativas tecnológicas al OLAP.

2.1.4.5 Las herramientas de acceso.

Permiten el análisis y la navegación de la información a través de los mismos.

2.2 Almacén de Datos

2.2.1 Definición

Inmon (2000), define los *Almacenes de Datos* (Data Warehouse) como una colección de datos orientados a temas, integrados, no volátiles y variantes en el tiempo que apoyan a las decisiones de la gerencia.

Los almacenes de datos dan acceso a datos para análisis complejos dando respuesta a las demandas de alto rendimiento de datos e información de una organización.

2.2.3 Objetivos

Los objetivos principales que tiene un Almacén de Datos son:

- Lograr que la información de la organización sea accesible: permitiendo que el contenido del Almacén de Datos sea entendible, manejable, rápido de acceder y navegable, facilitándole al personal obtener esa información para poder lograr un objetivo en específico.
- Lograr que la información de la organización sea consistente: permitiendo que la información que se trabaja en la organización tenga el mismo nombre y el mismo significado, lógica, coherencia y solidez entre ellos.
- Proporcionar información adaptable y elástica: es decir, que el Almacén de Datos está hecho para cambios continuos en la información, cuando se le hacen consultas o se agregan datos, tanto los datos como la tecnología existentes no cambian en absoluto.
- Proporcionar un seguro que protege los valores de la información: se le brinda al usuario el derecho de poder manipular los datos o hacer uso de ellos dependiendo de una permisología que se le asigna, logrando brindar una seguridad en la información ya que no todos los usuarios pueden hacer y deshacer información.

- Fundamentar la toma de decisiones: el Almacén de Datos tiene los datos correctos, como el volumen de datos necesarios para apoyar la toma de decisiones de una organización.

2.2.2 Características

Según Inmon, un almacén de datos debe tener las siguientes características como se muestra en la Ilustración 5.

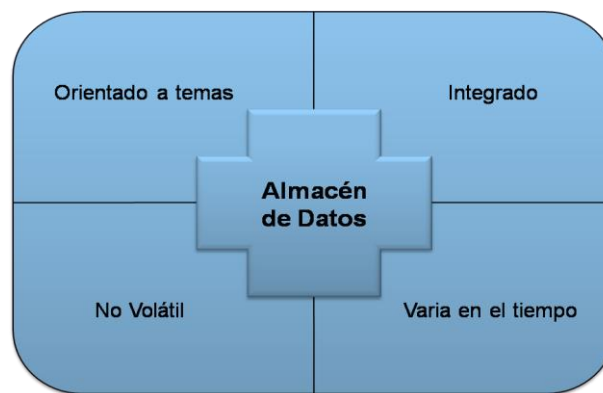


Ilustración 5: Características de un almacén de datos

Fuente: (Inmon I. , 1996)

- **Orientado a temas:** La información en un almacén de datos se clasifica por áreas temáticas en base a los aspectos que son de interés para la empresa. Es por esto que el diseño se orienta a realizar consultas eficientes en relación a la información de las actividades básicas de la organización, como por ejemplo ventas, compras o reclamos. Estos temas pueden verse como un conjunto de indicadores o medidas que son de interés para la empresa.
- **Integrado:** En un almacén de datos los datos provienen de distintas fuentes y son almacenados en un mismo repositorio para así lograr la integración en aspectos como: convenciones de nombres, medidas uniformes de variables, codificación de estructuras.
- **No volátil:** Los datos cargados en un almacén de datos, debe mantenerse invariables. En un sistema operacional, los datos se actualizan regularmente

mediante procesos que incluyen tareas como borrar, insertar o modificar registros. En un almacén de datos sólo se realizan dos tipos de operaciones: la carga inicial de los datos y el acceso a los mismos. Es por esto que los datos operacionales son movidos hacia el almacén de datos en intervalos de tiempo específicos, y dependiendo de los requerimientos del negocio, estos intervalos pueden ser establecidos una vez al día o una vez a la semana, según los requerimientos de los usuarios.

- **Varía en el tiempo:** La información almacenada en un almacén de datos se considera de tiempo variante, porque los datos que son extraídos desde los sistemas transaccionales son archivados, y por consiguiente históricos. Cada vez que se hace una nueva carga del almacén de datos, los datos anteriores no son eliminados, se mantienen en el tiempo para así hacer comparaciones y generar conocimientos sobre el negocio. En la Tabla 1 se muestra una comparación entre un sistema operacional y un almacén de datos basándose en la característica de Tiempo Variante.

Tabla 2: Diferencias entre sistema transaccional y almacén de Datos.

Sistema transaccional	Almacén de datos
Horizonte de tiempo: 60 - 90 días.	Horizonte de tiempo: 5 - 10 años.
La clave puede o no, tener un elemento de tiempo.	La clave contiene un elemento de tiempo.
Los datos pueden ser actualizados.	Una vez que se ha obtenido la información, el registro no puede ser modificado.

Esta comparación denota que la información que es representada mediante los datos almacenados en un almacén de datos, corresponde a un período largo de tiempo (de 5 a 10 años), en cambio en el ambiente operacional sólo se representan valores actuales, es decir, a períodos más cortos como podrían ser de 60 a 90 días. Por otra parte, se puede ver el tiempo variante en el almacén de datos por la estructura clave, ya que la misma contiene, implícita o explícitamente, un elemento

de tiempo como lo puede ser el día, el mes, el año, etc. En conjunto a esto, podríamos decir que la información del almacén de datos es, para todos los propósitos prácticos, una serie larga de vistas instantáneas que una vez registradas correctamente, no deberían ser actualizables. Sin embargo, se permite hacer modificaciones en caso de que los datos sean tomados incorrectamente.

2.2.3 Arquitectura

Una arquitectura de Almacén de Datos es una forma de representar la estructura global de los datos, la comunicación, los procesos y la presentación al usuario final. La arquitectura está constituida por las siguientes partes interconectadas Ilustración 6 (DWH, 2010): Elementos que constituyen la arquitectura de un Almacén de Datos:

- **Nivel de base de datos externos, Base de Datos operacional.** Las organizaciones adquieren datos de Bases de Datos externas a la propia organización, que incluyen datos demográficos, económicos, datos sobre la competencia, etc. Mediante el proceso de DWH se extrae la información que está en la bases de datos operacionales y se mezcla con otras fuentes de datos. Enriquecemos la información.
- **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.
- **Nivel de acceso a los 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. La clave de este nivel está en proveer al usuario de un acceso universal a los datos, es decir, que los usuarios sin tener en cuenta la ubicación de los datos o la herramienta de acceso a la información, deberían ser capaces de acceder a cualquier dato del DWH que les fuera necesario para realizar su trabajo.

-
- **Nivel de directorio de datos (metadatos).** Para proveer de un acceso universal, es absolutamente necesario mantener alguna clase de directorio de datos o repositorio de información de metadato que ayude a mantener un control sobre los datos. El metadato aporta información sobre los datos de la organización, de dónde proviene, qué formato tenía, cuál era su significado y si se trata de un agregado, cómo se ha calculado éste. Para mantener un almacén completamente funcional, es necesario disponer de una amplia variedad de metadatos, información sobre las vistas de datos para los usuarios finales y sobre las Bases de Datos operacionales.
 - **Nivel de gestión de proceso.** Este nivel tiene que ver con la planificación de las tareas que se deben realizar, no sólo para construir, sino también para mantener el DWH y la información del directorio de datos. Es o el controlador de alto nivel de los procesos que se han de llevar a cabo para que el Almacén de Datos permanezca actualizado.
 - **Nivel de mensaje de la aplicación.** El nivel de mensaje de la aplicación tiene que ver con el transporte de información alrededor de la red de la empresa. Puede usarse para aislar aplicaciones operacionales o estratégicas a partir de un formato de datos exacto, recolectar transacciones o los mensajes y entregarlos a una ubicación segura en un tiempo específico.
 - **Nivel Almacén de Datos (físico).** Es el repositorio central, altamente flexible de información, donde residen las copias de datos operacionales usados. En un DWH físico las copias de datos, operaciones y/o externos se almacenan de forma que sea fácil de acceder. En la actualidad un DWH se almacena en plataformas cliente/servidor, pero también existen configuraciones sobre mainframes, equipos externos para su rápido acceso mediante consulta. Nivel de organización de datos.

El componente final de la arquitectura Almacén de Datos es la organización de los datos. Incluye todos los procesos necesarios para seleccionar, editar, resumir, combinar y cargar datos en el Almacén de Datos y para acceder a la información desde bases de datos operacionales y/o externas. La forma de

implementar un Almacén de Datos está sujeta a la forma en la que se va a estructurar el almacenamiento de los datos dentro del mismo. Independientemente del modelo que se escoja el objetivo principal es escoger uno que satisfaga las necesidades empresariales.

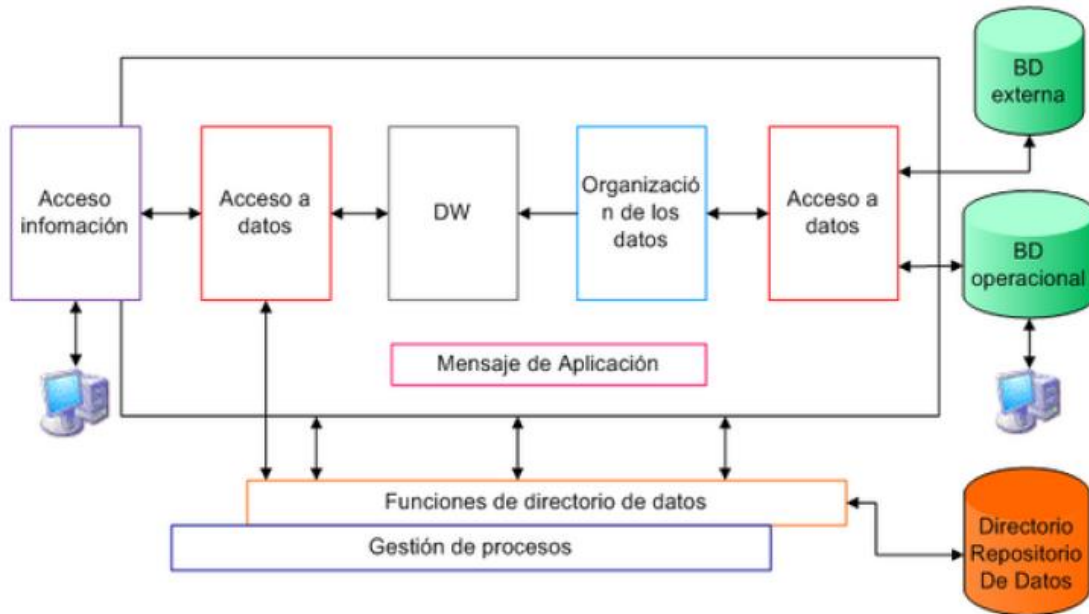


Ilustración 6: Arquitectura de una solución BI.

(Laudon K., 2008)

2.2.4 Ventajas y desventajas de los almacenes de datos

Algunas ventajas son:

- Facilitan el funcionamiento de las aplicaciones de los sistemas de apoyo a la toma de decisiones tales como informes de tendencia, informes de excepción, entre otros.
- Los Almacenes de Datos hacen más fácil el acceso a una gran variedad de datos por parte de los usuarios finales.
- Los almacenes de datos pueden trabajar en conjunto y por lo tanto, aumentar el valor operacional de las aplicaciones empresariales, en especial la gestión de relaciones con clientes.

- Facilitan el uso de técnicas estadísticas, para identificar relaciones o patrones ocultos.
- Ayuda a reaccionar rápido al cambio.

Utilizar un Almacén de Datos también plantea algunas desventajas como son:

- A lo largo de su vida los Almacenes de Datos pueden suponer altos costos. Ya que el Almacén de Datos no suele ser estático, los costos de mantenimiento son elevados.
- Los Almacenes de Datos se pueden quedar obsoletos relativamente en un futuro.
- A veces, ante una petición de información estos devuelven una información no del todo correcta, que genera pérdida para la organización. Esto implica que debe haber un monitoreo constante de la calidad de los datos.
- A menudo existe una delgada línea entre los almacenes de datos y los sistemas operacionales. Hay que determinar qué funcionalidades de estos se pueden aprovechar y cuáles se deben implementar en el Almacén de Datos, resultaría costoso implementar operaciones no necesarias o dejar de implementar algunas que sí vayan a necesitarse.

2.2.5 Bodega de datos

Según (Inmon, 2000), una Bodega de Datos (Data Mart) es una estructura departamental de suministro de datos desde el Almacén de Datos donde los datos están desnormalizados, basado en la información que necesite un departamento.

En un Almacén de Datos se almacena toda la información de la empresa u organización, y en una *Bodega de Datos* se almacena un subconjunto de datos, con el propósito de ayudar a un área específica dentro de un negocio para tomar mejores decisiones. Las *Bodegas de Datos*, suelen ser cargadas con datos de sólo

unas pocas fuentes. Las fuentes pueden ser sistemas internos de funcionamiento, almacenes de datos centrales, o de datos externos.

2.2.2.1 Características de las bodegas de datos

- Trata con una cantidad de usuarios mucho más limitada.
- Trabaja con un área específica de una empresa u organización.
- Tiene un propósito u objetivo específico.
- Tiene una función de apoyo con respecto al Almacén de Datos.
- Tiene una mayor rapidez de consulta.
- Facilidad para la historización de los datos.
- Validación directa de los datos.

2.2.2.2 Razones para crear una bodega de datos

- Fácil acceso a los datos que se necesitan frecuentemente.
- Crea vista colectiva para grupo de usuarios.
- Mejora el tiempo de respuesta del usuario final.
- Facilidad de creación.
- Costo inferior al de la aplicación de un completo Almacén de Datos.
- Los usuarios potenciales son claramente más identificables que en un Almacén de Datos completo.

2.2.2.3 Diferencias entre una bodega de datos y un almacén de datos

Un Almacén de Datos, a diferencia de una Bodega de Datos, se ocupa de varios temas y es típicamente controlado e implementado por una unidad central organizacional, se entiende como central lo que es conocido como el Almacén de Datos de la empresa. Por lo general un Almacén de Datos reúne información de múltiples sistemas. Ninguna de las definiciones básicas de una Bodega de Datos limita el tamaño como la complejidad del apoyo y la toma de decisiones que puede

contener. Sin embargo, las Bodegas de Datos son más pequeñas y mucho menos complejas que los almacenes de datos, por lo que en general son más fáciles de construir y mantener. A continuación, se tiene una comparación entre ellos en la Tabla 2:

Tabla 3: Diferencias entre Almacén de Datos y Bodega de Datos

Categoría	Almacén de datos	Bodega de datos
Alcance	Corporativo	Línea de negocio
Tema	Temas	Único tema
Orígenes de datos	Mucha	Pocos
Tamaño (relativo)	100 GB-TR+	<100 GB
Tiempo de implementación	Meses o años	Meses

2.3 Modelo dimensional

2.3.1 Definición

El modelo dimensional es una técnica de diseño para bases de datos (a nivel lógico) que pretende representar los hechos del negocio. Este modelo está optimizado para llevar a cabo consultas con un alto rendimiento.

Todo diseño de un Almacén de Datos debe comenzar con un modelo dimensional, debido a que con este se identifican cuáles son los hechos que se desean medir y desde cuáles perspectivas se desean agrupar los mismos. Por tal motivo, es necesario definir y tener claros los conceptos básicos concernientes al modelo dimensional, los tipos de tablas involucradas, y los esquemas de trabajo que pueden emplearse en una solución.

2.3.2 Dimensión

“La *Dimensión* es una entidad independiente en el modelo dimensional que sirve como un punto de entrada ó como un mecanismo de reordenamiento y fraccionamiento de las medidas sumarizadas, localizadas en la tabla de hechos del modelo” (Kimball & Caserta, 2008).

Son criterios puntuales y conocidos por el área de negocio, para la presentación de los datos al usuario final, como por ejemplo: producto, zonas, tiempo, entre otros. Las *dimensiones* son la información que define a cada uno de los registros de la tabla de hechos.

2.3.3 Tabla de dimensión

“La *tabla de dimensión* es una tabla en el modelo dimensional con una clave primaria simple y columnas de atributos descriptivos” (Kimball & Caserta, 2008).

Dimensión 1
Nivel 1
Atributos
Nivel 2
Atributos
Nivel n
Atributos

Ilustración 7: Tabla Dimensión

Fuente: (Kimball R. &, 2008)

Cuando se crea un cubo OLAP, se requiere una tabla de hechos y se crean varias *tablas de dimensiones* (su forma se muestra en la Ilustración 8), estas *tablas de dimensiones* acompañan a la tabla de hecho para definir los parámetros de los que dependen los hechos registrados en ella, para eso se tiene que entender que una

tabla de dimensión son ciertos elementos que contiene atributos (campos), que se utilizan para agrupar o restringir los datos que se almacenan en una tabla de hechos que son consultadas en un entorno de Bodega de Datos o Almacén de Datos Ilustración 7.

2.3.4 Hecho

“Un *hecho* es una medida del desempeño empresarial, usualmente son valores numéricos y aditivos que son almacenados en una tabla de hechos” (Kimball & Caserta, 2008).

Generalmente los *hechos* que se almacenan son números enteros (ejemplo: cantidad vendida de un producto, cantidad solicitada de un producto, entre otros) ó números reales (ejemplo: costo de producción de un producto, monto obtenido por venta, entre otros).

2.3.4 Tabla de hechos

“En el modelo dimensional, la *tabla de hechos* es la tabla principal con las mediciones de rendimiento numéricas que son caracterizadas por una clave compuesta, donde cada elemento de la misma es una clave foránea que corresponde a una tabla de dimensiones” (Kimball & Caserta, 2008).

Se entiende como *tabla de hechos*, la tabla principal en un modelo numérico, donde las mediciones de una organización o empresa se almacenan en ellas. Cada medida se toma de la intersección de las dimensiones que la definen, estas dimensiones son las que van a estar alrededor de ella, es decir, relacionada directamente con ella. Una fila de la tabla de hechos corresponde a la medida. Una medida es una fila de la tabla de hechos. Todas las medidas deben estar basadas en un mismo nivel de granularidad (grano). Los hechos más útiles en una tabla de hechos son tanto los numéricos como los aditivos. La tabla de hechos representa esa relación de muchos a muchos, entre las distintas dimensiones que se relacionan

con ella. Se puede ver un ejemplo en la Ilustración 9 donde se tiene una tabla de hechos llamada en el centro, tiene un esquema de tipo estrella y se tiene los que son las tablas de dimensión que están alrededor de la tabla de hechos para así formar un modelo dimensional.



Ilustración 8: Modelo dimensional

Fuente: (Kimball R. &, 2008)

2.3.5 Esquema estrella

“El *esquema de estrella* es la representación genérica de un modelo dimensional en una base de datos relacional en donde una tabla de hechos con una clave compuesta es unida a un número de tablas de dimensiones, cada una con una clave primaria simple” (Kimball & Caserta, 2008).

El *esquema en estrella* consiste en estructurar la información en procesos, vistas y métricas a modo de estrella. En la tabla de hechos encontramos los atributos destinados al hecho que constituye el proceso de negocio a medir, es decir, sus métricas. Mientras, en las tablas de dimensión, los atributos se destinan a elementos de nivel (que representan los distintos niveles de las jerarquías de dimensión) y a atributos de dimensión (encargados de la descripción de estos elementos de nivel). Podemos ver una representación de su estructura en la Ilustración 9.

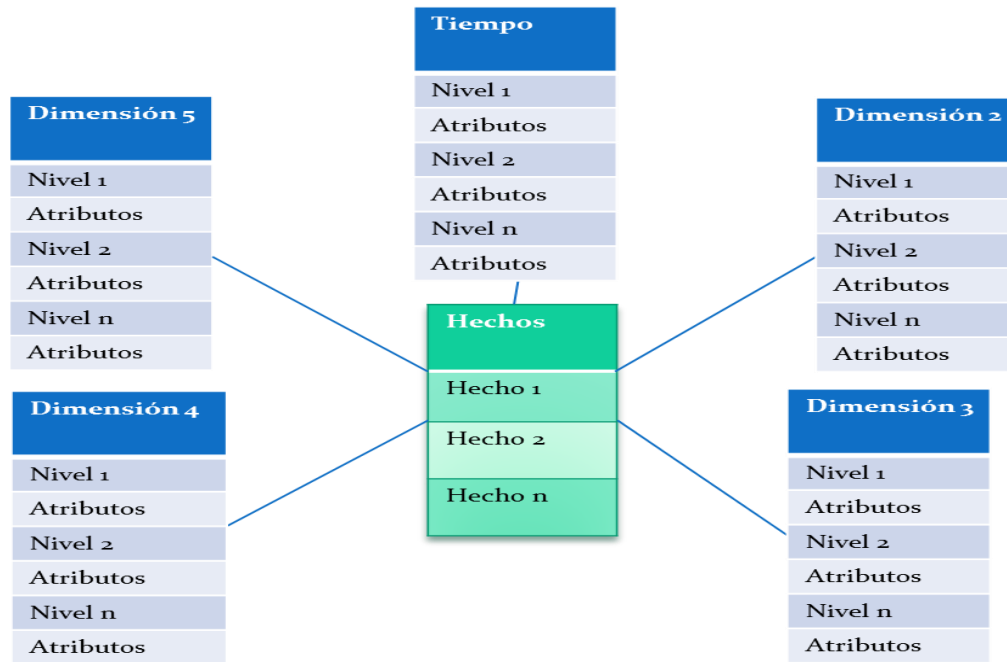


Ilustración 9. Esquema estrella

Fuente: (Kimball R. &, 2008)

2.3.6 Granularidad

Presenta el nivel de detalle al que se desea almacenar la información que posee cada registro de una tabla de hechos en un DWH [Kimball, 2002]. Es decir deberemos decidir cuál es el nivel de granularidad necesario para poder construir un modelo que nos permita responder a aquellas preguntas que nos hemos formulado a determinar un nivel de granularidad podemos responder unas preguntas pero no otras. Por ejemplo, los datos referentes a ventas o compras realizadas por una empresa, pueden registrarse día a día, en cambio, los datos pertinentes a pagos de sueldos o cuotas de socios, podrán almacenarse a nivel de mes.

2.3.7 Jerarquía

“La jerarquía es una serie de relaciones en cascada de uno a muchos” (Kimball & Caserta, 2008).

Después de que ya se sabe cuál sería la granularidad, se parte a realizar la jerarquía. Esta jerarquía corresponde con las tablas de dimensión que se definen como los niveles de asociación que se tienen de los datos. Una dimensión debe contener al menos una jerarquía, de la cual la jerarquía puede tener varios niveles. Por ejemplo, puede tener en una jerarquía que organiza a los clientes por ubicación geográfica, en la que se puede incluir los niveles Región, País, Ciudad y el nivel cliente.

2.3.8 Agregación

“La *agregación* es una medida, es decir un dato contable” (William, 2007).

Por ejemplo, si se quiere tener el total del salario de una empresa, lo que se hace es hacerle un proceso de *agregación*, definiendo como operación la suma de los sueldos de los empleados, logrando que cada vez que entre más información en el atributo sueldo, se va a ir mostrando sumalizando. Otro ejemplo es si se tiene la dimensión Producto se puede incluir un valor llamado "Total Productos", que será padre de todos los demás niveles de jerarquía y que contendrá el acumulado de todos ellos, logrando que cuando se introduzca más productos, lo que hace es sumarizar y actualizar el resultado.

2.3 Indicadores de gestión

“Los indicadores de gestión son elementos del sistema de control de gestión que proporcionan información significativa sobre aspectos críticos o claves de una organización mediante la relación de dos o más datos” (Pacheco J.P., 2002).

Los indicadores son uno de los agentes determinantes para que todo proceso de producción se lleve a cabo con eficiencia y eficacia, con ellos se puede implementar en un sistema adecuado de indicadores para calcular la gestión o la administración de los mismos, con el fin de que se puedan efectuar y realizar en posiciones estratégicas que muestren un efecto óptimo en el mediano y largo plazo, mediante

un sistema de información que permita comprobar las diferentes etapas del proceso logístico.

Se tiene que tener en cuenta que medir es comparar una magnitud con un patrón prestablecido, la idea es escoger las variables críticas para el éxito del proceso. Para una gestión eficaz y eficiente es conveniente diseñar un sistema de control de gestión de soporte administrativo que le permite evaluar el desempeño de la empresa. Se puede ver como esa relación cuantitativa y cualitativa que por medio de ellas se puede analizar y estudiar una situación, como la tendencia de cambios generados por algún fenómeno determinado, con respecto a unos objetivos y metas previstas o ya indicadas.

A cada uno de los usuarios, el sistema debería facilitar información oportuna y efectiva sobre el comportamiento de las variables críticas, para el éxito a través de los indicadores de gestión que hayan sido previamente definidos. Solo de esta manera se garantiza que la información que se genera en el sistema de control tenga efectos en los procesos de toma de decisiones y se logre así mejorar los niveles de aprendizaje de la organización.

De tal manera se entiende que los indicadores de gestión pueden ser valores, unidades, índices, series, estadísticas, entre otros. Es decir, es la expresión cuantitativa del comportamiento o del desempeño de toda una organización o de sus departamentos, cuya magnitud al ser comparada con un nivel de referencia, puede estar señalando una desviación en la cual se toma acciones correctivas o preventivas según el caso.

2.3.1 Clases de indicadores

- **Indicador de utilización:** Consiente entre la capacidad utilizada y la disponibilidad.
- **Indicador de eficiencia:** Cociente entre producción real y la esperada.

- **Indicador de productividad:** Consiste entre los valores reales de la producción y los esperados.
- **Indicadores de eficacia:** están relacionados con los ratios que indican capacidad o acierto en la consecución de tareas y/o trabajos.
- **Indicadores de gestión:** teniendo en cuenta que gestión tiene que ver con administrar y/o establecer acciones concretas para hacer realidad las tareas y/o trabajos programados y planificados.

Los indicadores más usados son los de eficacia y los de eficiencia, los cuales engloban a los llamados factores claves del éxito, que son: calidad, Satisfacción del cliente, y resultados que se generalizan en eficacia, los de tiempos de proceso, costos operativos, y desperdicios, que se refieren a eficiencia (ilustración 10).

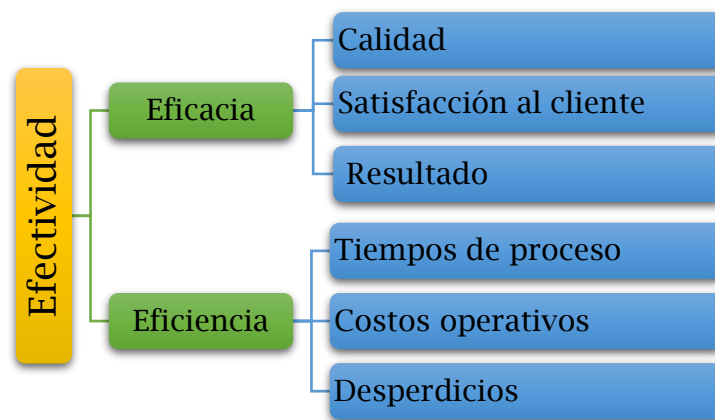


Ilustración 10: Principales clases de indicadores.

2.3.2 Objetivos de los indicadores

- Identificar y tomar acciones sobre los problemas operativos.
- Medir el grado de competencia de la empresa con respecto a las otras empresas.
- Reducir gastos y aumentar la eficiencia operativa.

- Poder hacer comparaciones.
- Mejorar el uso de los recursos y activos asignados, para aumentar la productividad y efectividad en las diferentes actividades hacia el cliente final.

2.3.3 Utilidad de los indicadores

Las organizaciones actualmente se ven afectadas por diferentes factores en su desempeño, de los departamentos que conformen dicha organización. El comportamiento de estos factores son probabilísticos ya que estos pueden reducir esa duda que se tiene por medio de la información registrada o captada. Algunas de las causas de por qué se quiere obtener esta información son estas:

- La exposición de la información.
- La creciente complejidad de la administración.
- El ritmo rápido del cambio.
- La interdependencia de las unidades que conforman la organización.
- El reconocimiento de la información como recursos.
- La evaluación y la disponibilidad de la tecnología y de las telecomunicaciones.
- La necesidad de desarrollo de las organizaciones y de las personas.
- Incrementar la competitividad y lograr la diferenciación.

Los indicadores se utilizan ya que trae una ventaja fundamental para la organización, debido a que el uso de los indicadores tiene consigo una reducción drástica de la duda, de la angustia y subjetividad, con el consecuente incremento del bienestar de todos los trabajadores. Estas son algunas ventajas que trae el uso del manejo de los indicadores de gestión:

- Estimular y promover el trabajo en equipo.
- Contribuir con el desarrollo y el crecimiento tanto personal como el equipo dentro de la organización.

- Generar un proceso de innovación y enriquecimientos del trabajo diario.
- Impulsar la eficiencia, la eficacia y la productividad de cada uno de los negocios.

2.4 Herramientas de Inteligencia de Negocio

En el mercado se tiene una cierta cantidad de herramientas BI (Business intelligence/ Inteligencia de Negocios), solo que se hace difícil o compleja la elección de ellos (Oracle BI, Pentaho BI, Report Portal BI Solution, Buisness Objects, IBM Cognos, entre otros) ya que tienen una gran gama de funcionalidades que brinda cada herramienta. El proceso de selección de una herramienta tecnológica puede llegar a ser complejo debido a las múltiples alternativas existentes en el mercado (Oracle BI, Pentaho BI, IBM Cognos, MicroStrategy, entre otros) y a la diversa gama de funcionalidades que cada herramienta brinda. Independientemente de la herramienta a seleccionar, ésta debe proveer un repositorio centralizado, visualización de cuadros de mando y la posibilidad de construir consultas a la medida.

2.4.1. Herramientas para la programación de procesos ETL y construcción de cubos.

2.4.1.1 Pentaho

Es una plataforma de inteligencia de negocio, la cual está orientada a soluciones y centrada en procesos que incluye todos los principales componentes requeridos para implementar soluciones basadas en procesos, para la gestión y toma de decisiones empresariales. Es una plataforma compuesta de diferentes programas que satisfacen los requisitos fundamentales para una solución de inteligencia de negocio. Ofrece soluciones para la gestión y análisis de la información, incluyendo el análisis multidimensional OLAP, presentación de informes, minería de datos, creación de cuadros de mando para el usuario, entre otros. Actualmente, existe una versión comercial y una versión de software libre desarrollada por la comunidad. La plataforma ha sido desarrollada bajo el lenguaje de programación

Java y tiene un ambiente de implementación, también basado en Java, haciendo de Pentaho una solución flexible que cubre una gran cantidad de necesidades empresariales.

Data Integration: Pentaho Data Integration extrae, limpia e integra la información disponible en aplicaciones y base de datos separadas, y la pone en manos del usuario, proveyendo consistencia. Esto se debe a que centraliza una versión de todos los recursos de información, que es uno de los más grandes desafíos para las organizaciones de hoy en día. Pentaho Data Integration, anteriormente conocida como Kettle, incluye cinco (5) herramientas, las cuales son:

- Spoon: herramienta gráfica para diseñar ETLs.
- Pan: ejecuta transformaciones diseñadas en el Spoon.
- Chef: herramienta para ejecutar trabajos complejos, que automatizan los procesos de actualización de la base datos.
- Kitchen: herramienta que ayuda a ejecutar los trabajos por lotes, permitiendo iniciar y controlar fácilmente el proceso ETL.
- Carter: servidor web que permite la supervisión remota del proceso ETL.

2.4.2 Sistemas manejadores de base de datos

2.4.2.1 PostgreSQL

Es un sistema manejador de bases de datos objeto-relacional, distribuido bajo licencia BSD y desarrollado bajo la filosofía de código abierto. Es el sistema manejador de bases de datos (SMBD) de código abierto más potente del mercado y, en sus últimas versiones, no tiene nada que envidiarle a otras bases de datos comerciales. (Martínez, 2010). PostgreSQL utiliza un modelo cliente/servidor y usa multiprocesos, en vez de multihilos, para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto, y el sistema continuará

funcionando. PostgreSQL funciona muy bien con grandes volúmenes de datos y una alta concurrencia de usuarios en el sistema. (Martínez, 2010).

PostgreSQL posee grandes ventajas frente a la competencia como se describe a continuación:

Oracle	Postgre SQL
Puede revertir a punto de retorno. También existe la posibilidad de iniciar una nueva transacción, independiente de la transacción actual.	A diferencia del Oracle este no puede realizar estas operaciones
Más de mil personas. Capacidad para dar servicio a una gran cantidad de conexiones en paralelo a una base de datos - el uso de multi-threaded servidor.	N / D
Idiomas para escribir procedimientos almacenados: PL / SQL y Java.	Idiomas para escribir procedimientos almacenados: PL / pgSQL, PL / Tcl, PL / Perl, SQL, C, posibilidad de crear un nuevo lenguaje - CREATE LANGUAGE . Las funciones escritas en C se pueden compilar en objetos cargables de forma dinámica, y se utiliza para implementar funciones definidas por el usuario de SQL.

Tabla 4: Cuadro comparativo entre manejadores de bases de datos.

2.4.3 Herramienta de acceso y análisis de datos

2.4.3.1 Tableau

Tableau es una herramienta o software de Inteligencia de Negocios que permite visualizar de una manera reducida grandes volúmenes de información, en forma rápida, flexible y gran usabilidad. Según (Kamkolkar, 2016) Tableau es en la actualidad la herramienta de BI líder y de mayor velocidad de crecimiento según

Gartner, destacando por su facilidad de uso, potencialidad para generar visualizaciones y capacidad de manejo de grandes volúmenes de Datos.

A diferencia de las herramientas tradicionales de Inteligencia de Negocios (BI) desarrolladas pensando en el usuario técnico del área de sistemas, Tableau está orientado a que personas de todos los ámbitos puedan manejar información fácilmente y presentarla en forma atractiva. Así, abogados, periodistas, ingenieros, médicos, entre otros, que trabajen en una organización o en forma independiente, encontrarán en Tableau un poderoso aliado analítico. Tableau Software se fundó sobre la idea de que el análisis de datos y los informes subsiguientes no deben ser actividades aisladas, sino que deben integrarse en un proceso único de análisis visual: uno que les permita a los usuarios ver rápidamente patrones en sus datos y cambiar las vistas al instante para seguir su línea de pensamiento.

Tableau combina la exploración de los datos y la visualización de estos en una aplicación fácil de usar que todos pueden aprender rápidamente. Cualquier persona acostumbrada al uso de Excel puede crear análisis interactivos y enriquecidos, eficaces para compartirlos de manera segura en la empresa. Los equipos de TI pueden administrar los datos y metadatos de manera centralizada, controlar los permisos y escalar hasta implementaciones en toda la empresa. Esta descripción general se diseñó para responder las preguntas comunes de los gerentes y administradores de TI y ayudarles a admitir las implementaciones de software de análisis visual de cualquier tamaño.

Arquitectura Tableau Server cuenta con una arquitectura de cliente-servidor de n niveles altamente escalable que presta servicios a clientes móviles, clientes web y software instalado en equipos de escritorio. Las soluciones de Tableau tienen 2 componentes principales: Tableau Desktop y Tableau Server. Tableau Server es una plataforma de análisis de negocios de clase empresarial que puede escalar verticalmente hasta cientos de miles de usuarios. Ofrece poderosos análisis móviles y basados en navegador, y funciona con la arquitectura de datos, la administración del ciclo de vida y las restricciones de seguridad y gestión

existentes en la empresa. Tableau Server cumple con requisitos empresariales, entre los que se incluyen:

- Escalabilidad Tableau Server puede escalar vertical y horizontalmente para satisfacer las necesidades de su empresa. El servidor puede escalar verticalmente con la incorporación de CPU y RAM adicionales. Todos los componentes de Tableau
- Server son de multiproceso y pueden configurarse en función de sus patrones de uso. Se puede escalar aún más agregando nodos adicionales que pueden configurarse de manera que cumplan con los requisitos de la organización.
- Alta disponibilidad con administración de clústeres interna y admite equilibradores de carga externos.
- Seguridad, cifra el tráfico interno, admite integración con Active Directory, SAML y OAuth.
- Facilidad de administración directa, desde la gestión de usuarios hasta las actualizaciones.
- Extensibilidad ofrece API eficaces.
- Capa de datos. Una de las características fundamentales de Tableau es que admite la arquitectura de datos que usted elija. Tableau no requiere que sus datos se almacenen en un solo sistema, propietario o de otro tipo. La mayoría de las organizaciones tiene un entorno de datos heterogéneo: almacenes de datos conviven con bases de datos, aunque sean locales o se encuentren en la nube.

Los cubos y los archivos planos como los de Excel se siguen usando mucho. Tableau puede trabajar con todos ellos de manera simultánea. No es necesario que reúna todos los datos en la memoria, a menos que así lo decida. Si sus plataformas de datos actuales son rápidas y escalables, Tableau le permite beneficiarse de manera directa de su inversión aprovechando la eficacia de la base de datos para responder preguntas. Si este no es el caso, Tableau proporciona opciones simples para actualizar sus datos aun de que sean

rápidos y responden con nuestro motor de datos en memoria. El siguiente diagrama muestra la arquitectura de Tableau Server:

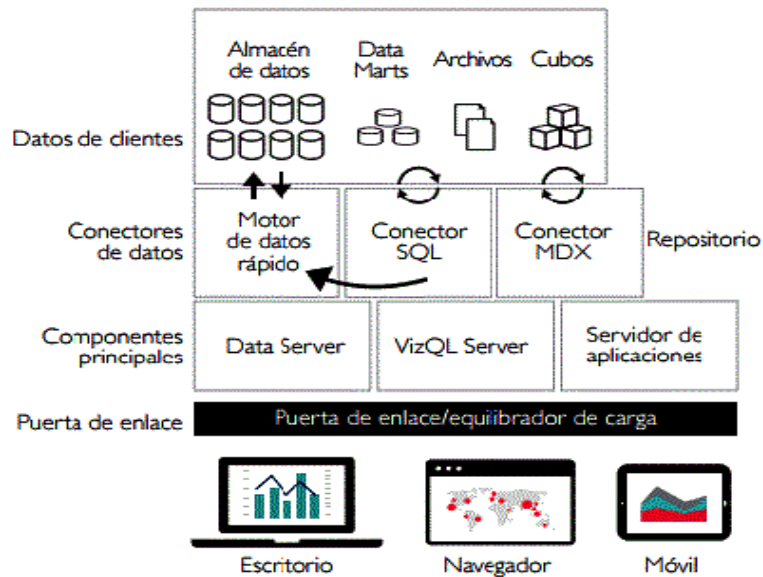


Ilustración 11: Arquitectura Tableau Server

Fuente: (Kamkolkar, 2016).

- **Conectores de datos.** Tableau incluye más de 40 conectores de datos optimizados para fuentes de datos como Microsoft Excel, SQL Server, Google BigQuery, Amazon Redshift, Oracle, SAP HANA, Salesforce.com, Teradata, Vertica, Cloudera y Hadoop, y se agregan conectores de datos nuevos regularmente. También hay un conector ODBC genérico para cualquier sistema sin conector nativo. Tableau proporciona dos modos de interacción con los datos: conexión en vivo o en memoria. Los usuarios pueden cambiar entre la conexión en vivo y en memoria, según lo deseen.
- **Componentes de Tableau Server.** El trabajo de Tableau Server se realiza mediante los siguientes cuatro procesos de servicio:
 - Servidor de aplicaciones. Los procesos del servidor de aplicaciones (wgserv.exe) controlan la exploración de contenido, la administración del servidor y los permisos para las interfaces web y móvil de Tableau Server. Cuando un usuario abre una vista en un

dispositivo cliente, ese usuario inicia una sesión (workgroup_session_id) en Tableau Server. El administrador puede configurar fácilmente el tiempo de expiración predeterminado de esta sesión. El usuario puede ejecutar dos o más procesos de servidor de aplicaciones para satisfacer sus necesidades de escalabilidad y disponibilidad. VizQL Server: una vez que el usuario recibe la autenticación del servidor de aplicaciones, puede abrir una vista. El cliente envía una solicitud al proceso VizQL (vizqlserver.exe). A continuación, el proceso VizQL envía las consultas directamente a la fuente de datos y devuelve un conjunto de resultados que se expresa en imágenes y se presenta al usuario. En muchos casos, Tableau Server aprovecha las representaciones y el almacenamiento en caché del cliente para reducir la carga del servidor. Además, cada VizQL Server tiene su propia memoria caché que pueden compartir varios usuarios de manera segura. El usuario puede ejecutar dos o más procesos de VizQL Server para satisfacer sus necesidades de escalabilidad y disponibilidad.

- Data Server. A diferencia de los enfoques tradicionales de administración de metadatos, el Data Server de Tableau es un componente clave que permite a los administradores de TI habilitar el monitoreo, la administración de metadatos y el control para los equipos de TI, a la vez que se habilitan los análisis de autoservicio para usuarios profesionales. Permite administrar y almacenar fuentes de datos de Tableau de manera centralizada y proporciona a los usuarios finales acceso seguro a datos confiables por medio de implementaciones de análisis de autoservicio. Usted puede administrar centralizadamente metadatos, como conexiones, controladores y filtros de fuentes de datos, para acceder a los datos. Puede asignar permisos específicos a las fuentes de datos de manera que permita al equipo de TI administrar los permisos a las fuentes de datos en función de grupos de identificación específica. En un entorno administrado, los usuarios que conocen mejor sus datos

también cuentan con la flexibilidad de definir y publicar definiciones, cálculos y grupos. Estos se pueden compartir para que los usen todos los miembros de la organización o los usuarios de Tableau Desktop para crear y aprovisionar sus propios cálculos, definiciones y grupos. La fuente de datos publicada se puede basar en:

- Una extracción del motor de datos de Tableau;
 - Una conexión en vivo (los cubos no se admiten como conexiones en vivo).
- Componente de segundo plano. El componente de segundo plano actualiza las extracciones programadas, entrega notificaciones y administra otras tareas de segundo plano. El componente de segundo plano está diseñado para consumir todos los recursos de CPU disponibles afín de concluir la actividad de segundo plano tan pronto como sea posible
- **Puerta de enlace/equilibrador de carga.** La puerta de enlace dirige solicitudes a otros componentes. Las solicitudes de los clientes, primero, se encuentran con un equilibrador de carga externo, si se configura uno, o la puerta de enlace y, de allí, se los dirige al proceso adecuado. En caso de que no haya un equilibrador de carga externo, si varios procesos se configuran para cualquier componente, la puerta de enlace actúa como un equilibrador de carga y distribuye las solicitudes hacia los procesos. En una configuración de un solo servidor, todos los procesos residen en la puerta de enlace o en el servidor primario. Cuando se trabaja en un entorno distribuido, se designa una máquina física como servidor primario, y las otras se designan como servidores de trabajo y pueden ejecutar cualquier cantidad de procesos adicionales. Tableau Server siempre usa una sola máquina como servidor primario.
 - **Clientes:** navegadores web y aplicaciones móviles. Tableau Server proporciona dashboards interactivos a los usuarios mediante HTML5 que no deja rastro en navegadores web o móviles, o de manera nativa mediante una aplicación móvil. No se necesita ActiveX, Java ni Flash para

ejecutar informes o visualizaciones. No se requieren complementos ni aplicaciones auxiliares. Tableau Server admite:

- Navegadores web: Internet Explorer, Firefox, Chrome y Safari.
 - Safari para móviles: las vistas optimizadas para la función táctil se presentan automáticamente en Safari para móviles.
 - Aplicación para iPad: aplicación nativa para iPad que proporciona vistas, contenido, navegación y edición optimizados para la función táctil.
 - Navegador de Android: vistas optimizadas para la función táctil que se ofrecen automáticamente en el navegador de Android.
 - Aplicación para Android: aplicación nativa para Android que proporciona vistas, contenido, navegación y edición optimizados para la función táctil.
- **Clientes: Tableau Desktop.** Tableau Desktop es el entorno de creación rápida de análisis de negocios que se usa para crear y publicar vistas, informes y dashboards en Tableau Server. Mediante Tableau Desktop, un autor de informes puede conectarse a varias fuentes de datos, explorar relaciones, crear dashboards, modificar metadatos y, por último, publicar un libro de trabajo completo o una fuente de datos en Tableau Server. Tableau Desktop también puede abrir cualquier libro de trabajo publicado en Tableau Server o conectarse con cualquier fuente de datos publicada, ya sea que se haya publicado como una extracción o como una conexión en vivo. Tableau Desktop es compatible con escritorios de Windows y de Mac.

En la tabla 1 se describe herramientas utilizadas para el manejo y acceso de los datos, en una solución de inteligencia de negocios:

SOLUCIÓN	LICENCIA	ORIENTACIÓN	HERRAMIENTAS DE MANIPULACIÓN DE DATOS	COSTOS	HABILIDAD DE EJECUCIÓN	COMPLEJIDAD DE LA VISIÓN DEL NEGOCIO
OLIK	Comunitaria y privada	No	Si	Moderados	Alta	Alta
TABLEAU	Comunitaria y privada	No	Si	Moderados	Alta	Alta
PENTAHO	Comunitaria y privada	Si	Si	Bajos	Limitada	Moderada
ORACLE	Privada	Si	Si	Altos	Limitada	Alta

Tabla 5: Comparaciones entre las soluciones de inteligencia de negocios.

Fuente: (Gartner, 2014)

CAPÍTULO 3

MARCO METODOLÓGICO

3.1 Método ascendente (Ralph Kimball)

La metodología de Ralph Kimball define un Almacén de Datos como “una copia de los datos transaccionales específicamente estructurados para consultas y análisis” (Kimball & Caserta, *The Data Warehouse ETL Toolkit* (2nd edition), 2008), también da una segunda definición como la unión de todas las Bodegas de Datos de una entidad. Kimball defiende su metodología mencionando como se ve en la Ilustración 12, que es mejor usar un método ascendente a la hora de diseñar un

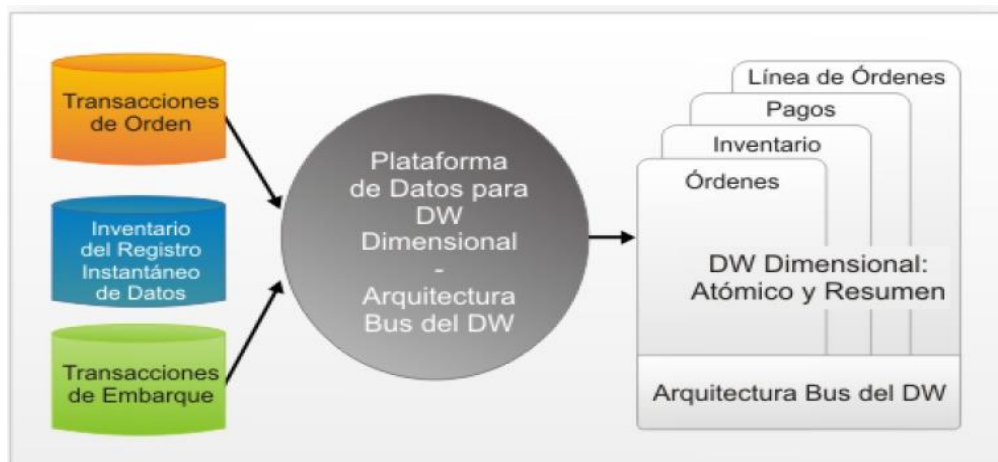


Ilustración 12: Método ascendente
Fuente: (Kimball R. &, 2008)

Almacén de Datos, es decir primero crear las Bodegas de Datos y después el Almacén de Datos.

3.1.1 Ciclo de vida del modelo de negocio.

La metodología de Ralph Kimball se compone por una serie de fases como se muestra en la ilustración 13, las cuales se desarrollan como sigue:

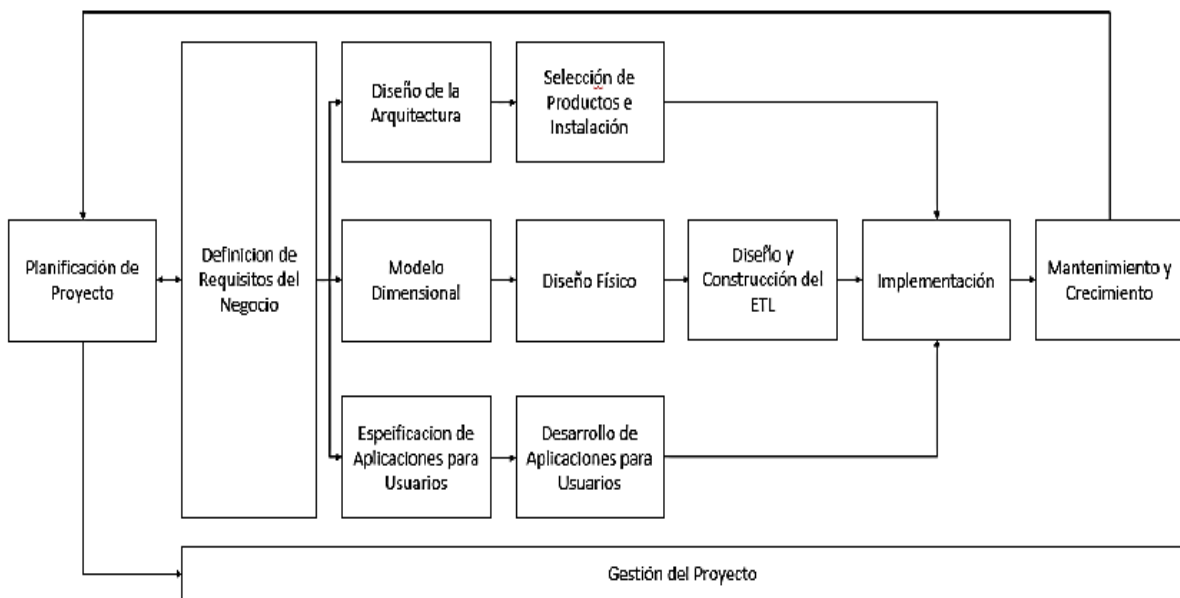


Ilustración 13: Tareas de la metodología de Kimball

Fuente: (Kimball R. &, 2008)

3.1.1.1 Planificación del proyecto.

La planificación del proyecto es la primera etapa dentro del ciclo de vida para el desarrollo de un almacén de datos, dando pie a una serie de actividades como lo son la definición, alcance, evaluación y la justificación del proyecto. En el caso de la definición, existen factores identificados por el autor, que diferencian a los proyectos fallidos de los que han logrado seguir en pie. Estos factores son:

-
- Poseer un aliado empresarial comprometido con el proyecto, su papel principal será de conciliador entre el proyecto y la organización. Esta persona, debe ser un líder políticamente astuto que pueda convencer a sus compañeros para que estos también apoyen la solución.
 - La existencia de una fuerte motivación organizacional por la construcción del almacén de datos.
 - La factibilidad y disponibilidad de datos.
 - Uso de tecnologías de información. Algunas organizaciones no usan tecnología de información para realizar sus labores cotidianas y desean adentrarse en este tipo de proyectos, dedicándole así más atención en mantener y manejar los mismos objetivos.
 - La preparación de la cultura actual hacia el nuevo enfoque.

Definición del Alcance: Una vez que se ha logrado un acuerdo con la organización, es el momento para poner los límites del proyecto. La determinación del alcance requiere la participación conjunta tanto del departamento de tecnología de la información, como de la gestión empresarial.

Justificación: Estimación de los beneficios y costos asociados al desarrollo e implementación de este tipo de solución. Es necesario estimar los costos aproximados para el hardware y software que serán usados. Debido a que los almacenes de datos tienden a expandirse rápidamente, es recomendable realizar estimaciones a corto plazo.

Adicionalmente, en la fase de planificación se debe seleccionar una aplicación piloto que posea alta probabilidad de éxito, así como la construcción de prototipos rápidos y frecuentes que apoyen el desarrollo del proyecto, reportar y publicar casos exitosos y finalmente tener a disposición herramientas que sean fáciles de usar y que den posibles visualizaciones del proyecto.

3.1.2.- Definición de los requerimientos del negocio.

La definición de los requerimientos está constituida por la definición de las necesidades del negocio que se desean medir, así como el planteamiento de cómo recolectar los requerimientos, la recolección de documentos y el seguimiento de los mismos. Se centra principalmente en reunir el conocimiento de los analistas de la organización, entendiendo los factores claves que determinan los requerimientos del negocio y establecer los fundamentos de tecnología, datos y aplicaciones de usuario final, para así obtener los posibles factores a medir o indicadores. Adicionalmente en esta etapa, se fijan políticas de seguridad de datos, posibles roles y usuarios.

3.1.3 Diseño técnico de la arquitectura.

En el desarrollo del proyecto del almacén de datos, el diseño de la arquitectura es un aspecto que puede generar conflictos, debido a que algunos grupos de desarrollo no prestan la atención debida a este punto, le restan importancia y descuidan elementos que más adelante generaran problemas en el desarrollo del proyecto. Todo almacén de datos posee una arquitectura técnica, y por esto (Kimball & Ross, 2002) ha identificado una serie de pasos a seguir, que ayudarán a conseguir los objetivos deseados en esta etapa.

El primer punto es el establecimiento de un grupo de trabajo especial, conformado por dos o tres personas centradas en el diseño de la arquitectura. Por lo general, lo conforma un arquitecto técnico que maneja aspectos de diseño y desarrollo de aplicaciones analíticas, y a su vez es el coordinador del grupo, quien garantizará el buen funcionamiento del mismo.

El segundo punto es la recolección de requerimientos de la arquitectura, donde se debe identificar las implicaciones arquitectónicas asociadas con las necesidades críticas del negocio. Además de aprovechar los requerimientos del negocio,

también se llevan a cabo entrevistas adicionales dentro de la organización. Dichas sesiones, están enfocadas en la tecnología, para comprender normativas vigentes, instrucciones técnicas previstas y los límites no negociables.

El siguiente punto trata de la necesidad de documentar hallazgos, donde simplemente se listen todos los requisitos del negocio que tienen un impacto en la arquitectura, en conjunto a los aspectos que resultan de la construcción de la arquitectura del almacén de datos, dando pie al siguiente punto, donde se utiliza esta documentación para realizar el modelo de la arquitectura a un nivel más alto, porque se ilustran los componentes principales como son la puesta en escena de los datos, el acceso a los datos, metadatos y la infraestructura. Se puede decir que el modelo generado es similar a los planos de una vivienda.

Seguidamente, se realiza un estudio preliminar de sistemas que den apoyo a los requerimientos antes mencionados, considerando adicionalmente las necesidades de seguridad, infraestructura física y las necesidades de configuración. En algunos casos las decisiones de infraestructura, como el hardware del servidor y el software de base de datos ya están predeterminadas. El estudio de la implementación por fases es el siguiente paso, y va de la mano con la documentación técnica de la arquitectura.

Finalmente, la culminación y constante revisión de la arquitectura técnica es el último paso a seguir, anunciando el plan de la arquitectura al equipo de proyecto, los colegas, empresas patrocinadoras y demás entes del negocio involucrados en el proyecto del almacén de datos. En caso de una revisión, la documentación debe ser actualizada y puesta en práctica de inmediato en el proceso de selección de productos.

3.1.4.- Selección de productos e instalación.

La selección de los productos está estrechamente relacionada al plan de arquitectura planteado en el paso anterior, e incluye una serie de tareas específicas

para una buena elección de productos, los cuales vienen dados por la comprensión de los procesos de negocio, el establecimiento de un plan de evaluación del producto, donde se coloquen ponderaciones para indicar la importancia de los elementos y cuyos criterios comunes tomados en esta fase sean funcionalidad, la arquitectura técnica, las características de software, el impacto de las infraestructuras y la viabilidad de los proveedores. Adicionalmente, se debe realizar un estudio de las herramientas disponibles en el mercado, que puedan dar soporte a las necesidades del proyecto. Este estudio arrojará puntuaciones que serán tomadas en cuenta para limitar la lista de proveedores, y comenzar a realizar una evaluación más detallada, donde podría surgir la creación de un prototipo con no más de dos productos y así demostrar cuál de las herramientas satisface mejor las expectativas de la organización. Una vez seleccionado el producto, comienzan las negociaciones directas con el proveedor, a quien se debe confrontar constantemente con la finalidad de seguir en un periodo de prueba, donde este tiene la oportunidad de utilizar su producto en el entorno real de la empresa. Y finalmente a medida que el juicio llega a su fin, se procede a la negociar el proceso de compra.

3.1.5.- Diseño del modelo dimensional.

Para la elaboración del diseño del modelo dimensional se plantean los siguientes pasos:

Paso 1: Definir del nivel de granularidad: Una vez que en el paso anterior se determina el proceso del negocio, el paso siguiente es determinar la granularidad, es decir, el nivel de detalle al que se desea almacenar información del proceso a modelar. Para la determinación del nivel de granularidad se debe responder a la pregunta “¿Cuál es el nivel de detalle con que deseo medir? o ¿Cómo se representaría una fila en la tabla de hechos?”. Ejemplos: Si nuestro negocio consiste en ventas de productos a clientes y al realizar estas ventas se genera una factura

por cliente, debemos decidir que queremos almacenar, si el monto total de la factura o el detalle de los montos por productos.

Paso 2: Determinar las dimensiones: Las tablas dimensiones son compañeros integrales de una tabla de hechos. Los atributos de la dimensión sirven como fuente principal para limitaciones de consulta, agrupaciones y etiquetas dentro del reporte. En una solicitud de consulta o de reporte, los atributos son identificados por los requerimientos de los usuarios. Por ejemplo, cuando un usuario declara que quiere ver los montos de venta por semana y por la marca, donde la semana y la marca deben de estar disponibles como atributos de la dimensión.

Los atributos de la tabla de dimensiones juegan un rol vital dentro del almacén de datos, puesto que son la fuente de las restricciones de búsqueda y campos de los reportes. El poder del almacén de datos es directamente proporcional a la calidad de la profundidad lograda por los atributos de las dimensiones.

Las tablas de dimensiones son los puntos de entrada en la tabla de hechos. Si los atributos de las dimensiones son robustos entonces las capacidades analíticas también lo serán. Los mejores atributos son textuales y discretos, deben estar identificados por palabras reales, no abreviaturas. Los atributos típicos de una dimensión producto por ejemplo incluyen una breve descripción compuesta máximo por 30 o 50 caracteres, una marca, un nombre de categoría, tipo de empaque, tamaño y otras características del producto.

En ocasiones, cuando se diseña el modelo dimensional no se tiene claro si un campo de datos numérico extraído de la fuente de datos es un hecho o un atributo de dimensión. A menudo se toma la decisión, preguntando si el campo es una medida que toma diferentes valores y que participa en los cálculos (lo que define al hecho) o básicamente es una descripción discreta, posiblemente constante y que participa en las limitaciones (lo que define a un atributo dimensional). Por ejemplo, el costo puede cambiar tan a menudo que finalmente se decida que es más como

un hecho medible. Las tablas dimensiones a menudo representan las relaciones jerárquicas de la organización, por ejemplo dentro de la tabla producto las dimensiones podrían identificar a los productos según la marca o según su categoría, almacenando en cada fila de la tabla dimensión la descripción de la marca y la categoría asociada a cada producto, facilitando el uso y rendimiento de las consultas a pesar de la evidente redundancia de datos, proceso que define la desnormalización de las tablas.

Paso 3: Determinar el hecho. La tabla de hecho es la tabla principal en un modelo dimensional donde el desempeño medible del negocio es almacenado. Nos esforzamos para almacenar los datos medibles resultantes de los procesos del negocio de un almacén.

Se usa el término hecho para representar la medida del negocio. Ejemplo: Estar en un mercado observando los productos que son vendidos y anotar la cantidad vendida y el monto en dinero vendido por cada día de cada producto en esa tienda. Una medición se realiza entonces, con la intersección de todas las dimensiones (día, producto y tienda). Esta lista de dimensiones define el grano de la tabla de hecho y el nivel de alcance de la medida.

Una fila en la tabla de hecho corresponde a una medida por lo tanto todas las medidas en la tabla de hecho debe tener el mismo nivel de granularidad.

El hecho más usable en una tabla de hechos son los numéricos y los sumarizables. Es teóricamente posible para un hecho medible ser textual. Sin embargo, esto se presenta raramente. En la mayoría de los casos, una medida textuales una descripción de algo y se extrae de una lista de valores discretos.

Todas las tablas de hechos tienen dos o más claves foráneas, que conectan con las tablas dimensión a través de sus claves primarias. Por ejemplo, una clave de producto en la tabla de hecho siempre coincidirá con un producto específico de la

dimensión producto. Cuando todas las claves foráneas de la tabla de hecho hacen referencia a sus respectivas claves primarias en las dimensiones correspondientes, entonces se dice que las tablas satisfacen la integridad referencial. Se accede al hecho como tal a través de la intersección de las dimensiones combinadas.

La tabla de hecho por sí misma tiene su propia clave primaria formada por un subconjunto de claves foráneas. Esta clave es comúnmente llamada clave compuesta o clave concatenada. Cada tabla de hecho en un modelo dimensional tiene una clave compuesta, por lo que, cada tabla que contenga una clave compuesta es una tabla de hecho. Las tablas de hechos expresan relaciones de muchos a muchos entre dimensiones en los modelos dimensionales

Finalizado el esquema se procede a la validación del mismo y para esto se requiere de una evaluación mucho más detallada, centralizada en la realización de preguntas que puedan ser respondidas por este modelo. Una vez que el grupo de trabajo encuentre que el modelo cumple con los objetivos, es presentado al grupo de tecnologías de información y a los encargados del almacén de datos y finalmente, luego de varias pruebas ser presentado a la comunidad del negocio. Una vez que se tienen definidas las tablas de hechos y dimensión, se deben relacionar a través de las claves primarias y foráneas correspondientes.

En los modelos dimensionales, podemos añadir nuevas dimensiones en el esquema siempre y cuando al menos un valor de la dimensión clasifique o defina cada uno de los hechos existente en la tabla de hechos. Del mismo modo, podemos añadir nuevos hechos imprevistos a la tabla de hechos, asumiendo que el nivel de detalle es coherente con los hechos existentes.

Existe una relación natural entre los modelos dimensionales y los normalizados. La clave es entender que la relación se encuentra en el desglose del modelo entidad relación dentro de los múltiples esquemas dimensionales posibles de modelar Finalmente se realiza la documentación en esta fase que debe contener los nombres

de las tablas y columnas, las definiciones y normas de cálculo, ya sea para el hecho o para un atributo de alguna dimensión. Adicionalmente, debe contener ciertos estándares con respecto a las convenciones de nomenclaturas para la identificación de los elementos.

3.1.6.- Diseño físico.

Dado que se posee el modelo dimensional lógico, se procede a su implementación física. En esta fase se debe seleccionar las estructuras necesarias que puedan dar soporte al diseño lógico, así como definir algunos aspectos importantes como son las convenciones estándares de nombres, tipos de datos, declaraciones claves y permisividad de valores nulos. Aspectos referentes al ambiente de bases de datos así como indexación, estrategias de particionamiento, ajustes de rendimiento y distribución de archivos.

3.1.7.- Diseño y construcción de procesos ETC.

En el diseño de los proceso de extracción, transformación y carga (ETC), como ha sido expuesto anteriormente, implica la extracción de los datos generalmente de los sistemas operacionales y la preparación de los estos para su puesta en escena en el área de presentación de los datos. En esta etapa se deben formular planes, parecidos a los elaborados en el diseño de la arquitectura, así como analizar el planteamiento de compra de una herramienta que realice los procesos ETC o la elaboración de una herramienta propia. Otra decisión fundamental es con respecto a la estructura de los almacenes de datos, donde la normalización de los datos de origen, es lo más apropiado aunque no muy necesario pero aumenta la credibilidad los datos.

3.1.8.- Especificación y desarrollo de aplicaciones analíticas.

El primer paso para la especificación de las aplicaciones analíticas, es el estudio de un conjunto de herramientas posibles, que puedan ayudar a solventar las

necesidades del negocio. Luego de haber definido la herramienta, se procede a establecer estándares de diseño como características del menú y la apariencia del entorno. En la fase de desarrollo de la aplicación, también se deben establecer ciertas normas. Convenciones de nomenclatura a usar, los cálculos que se realizarán, las bibliotecas y hasta la forma de codificación. La actividad de desarrollo de aplicaciones puede comenzar una vez que el diseño de la base de datos está completa.

Las herramientas de acceso a los datos y metadatos están instalados, y un subconjunto de los datos históricos se han cargado. Mientras la aplicación está siendo desarrollada, varios beneficios secundarios resultan. Los desarrolladores de aplicaciones, equipados con una herramienta robusta de acceso a los datos, rápidamente podrán encontrar respuestas a los diferentes problemas. Esta es una razón por la que los desarrolladores prefieren comenzar las actividades de desarrollo de las aplicaciones lo antes posible.

3.1.9.- Integración y despliegue.

La tecnología, datos y aplicaciones analíticas convergen en el despliegue, donde se logra la unificación de las tecnologías, datos y la aplicación de usuario final. Adicionalmente, la educación es fundamental en esta etapa, debido a que es la vía para la aceptación exitosa del almacén por parte de la comunidad de la organización. El programa de adiestramiento debe estar enfocado en aspectos como, estándares de datos, exploración sobre las aplicaciones analíticas, y finalmente uso de la aplicación para el acceso de los datos.

3.1.10.- Mantenimiento y crecimiento.

Una vez terminado el desarrollo completo, no se debe olvidar los siguientes puntos:

- **Soporte:** Se deberá hacer mantenimiento tanto de soporte técnico como de programación periódicamente al sistema, y estos deben estar programados de tal manera que no interrumpan con el desarrollo del negocio. Se debe mantener vigilado el funcionamiento del negocio para poder estar preparado a realizar cualquier tipo de cambio necesario para adaptar el sistema a las nuevas necesidades.
- **Capacitación de los usuarios:** se debe capacitar a los usuarios para que realicen un buen uso del sistema y así, se logre mantener una calidad estable en los datos.

CAPÍTULO 4

MARCO APLICATIVO

En este capítulo presenta cómo la metodología de Ralph Kimball fue aplicada para el desarrollo de una solución de inteligencia de negocio en el área de visita clínica de la compañía aseguradora. Se explican aspectos propios de la arquitectura de la solución planteada como el diseño, construcción del almacén de datos, elaboración de procesos ETC, diseño y construcción de las consultas analíticas para el despliegue de los indicadores de gestión relevantes en el área visita clínica de una compañía aseguradora.

4.1 Planificación del proyecto

Como en todo proyecto software, es necesario realizar una planificación del mismo para tratar de garantizar que el desarrollo, se realizara según las siguientes actividades previstas:

Tabla 6: Plan de actividades.

Actividades	Tiempo
Levantar requerimientos de indicadores del proceso.	2 Semanas
Diseñar el modelo dimensional que soporte los indicadores de gestión.	1 Semana
Construir el Almacén de Datos.	3 Semanas
Desarrollar los programas de extracción, transformación y carga desde la estructura de datos operativa al Almacén de Datos.	2 Semanas

Construir los indicadores.	2 Semanas
Realizar pruebas funcionales, no funcionales y de calidad de datos.	1 Semanas

4.2 Definición de requerimientos

Bajo la perspectiva de seguimiento de proyectos de investigación, abstrayéndonos de las variantes que pudieran implementarse en el mismo, para poder generar beneficios a la organización necesitan poder monitorear el presupuesto, las cartas avales, los visitadores clínicos, los beneficiarios y el nivel de compañías aseguradoras en términos generales, entre otros. Los indicadores fueron proporcionados por parte del analista de seguimiento de carta aval, el cual almacenaba la información relacionada a los indicadores en lapsos de tiempo que dependían del tamaño de la solicitud. Alguno de los indicadores que fueron solicitados, estaban relacionados con los requerimientos que se describen a continuación:

- Cantidad de Visitadores Clínicos por carta aval, estado en un periodo de tiempo.
- Cantidad de Pacientes beneficiarios por Carta aval, visitador clínico, estatus en un periodo de tiempo.
- Variación del Presupuesto por carta aval y clínica.
- Cantidad de Cartas Avales según estatus, presupuestos.
- Monto pendiente pendientes por cartas avales.

Para poder desarrollar los indicadores y reportes referentes a los requerimientos contemplados para la solución de inteligencia de negocios propuesta para el proceso de visita clínica, se realizaron entrevistas con los usuarios finales con el propósito de obtener información de las necesidades del negocio. En segundo lugar se estudiaron los requerimientos expuestos en las hojas de cálculo expuestas por los actores del negocio y finalmente es importante resaltar que para esta primera versión del sistema, no se van a definir metas de los indicadores, ya que se requiere

previamente acumular información de varios meses. A continuación se describe los indicadores planteados:

Tabla 7 Tabla de especificación de los indicadores propuesto

Fuente: Elaboración propia.

No	Formula	Unidad	Criterios de Clasificación	Representación
1	Cantidad de visitadores clínicos según estatus.	#	Por visitador clínico, por estatus.	Tablas de texto.
2	Porcentaje de visitadores clínicos por estatus.	%	Por visitador clínico, por estatus.	Gráfico histórico (de torta).
3	Cantidad de visitadores clínicos según estatus en un periodo de tiempo.	#	Por visitador clínico, por estatus, por periodo de tiempo (Año, mes, día, trimestre).	Tablas de texto
4	Cantidad de pacientes beneficiarios según el estado en el que se realizó la solicitud de visita clínica.	#	Por paciente beneficiario y estado.	Tablas de texto.
5	Cantidad de pacientes beneficiarios según estatus en un periodo de tiempo.	#	Por paciente beneficiario, por estatus, por periodo de tiempo (Año, mes, día, trimestre).	Gráfico histórico (barra/aplicada).

6	Monto pendiente de pagos por cartas avales activas por periodo de tiempo y clínica.	BsF	Por cartas avales, presupuestos, clínica por periodo de tiempo (Año, mes, día, trimestre).	Grafico histórico (Barra apilada).
7	Cantidad pendiente de cartas avales activas por periodo de tiempo y clínica.	#	Por cartas avales, presupuestos, clínica por periodo de tiempo (Año, mes, día, trimestre).	Grafico histórico (Barra apilada).
8	Variación de cantidad pendiente de presupuesto en porcentaje.	Δ	Por presupuesto, cartas avales.	Gráfico histórico (barra/aplicada).
9	Variación de monto pendiente de presupuesto en porcentaje.	Δ	Por presupuesto y cartas avales.	Gráfico histórico (barra/aplicada) .
10	Porcentaje de cartas avales según el estatus.	%	Por paciente beneficiario, por estatus, por carta aval, visitador clínico.	Gráfico histórico (de torta).
11	Cantidad de cartas avales según estatus.	#	Por paciente beneficiario, por estatus, por carta aval, visitador clínico.	Gráfico histórico (barra/línea).
12	Cantidad de cartas avales según estatus en un periodo de	#	Por paciente beneficiario, por estatus, por carta.	Gráfico histórico (barra/aplicada).

	tiempo (Año, mes, día, trimestre).		aval, visitador clínico, por periodo de tiempo (Año, mes, día, trimestre).	
--	------------------------------------	--	----------------------------------------------------------------------------	--

4.3 Diseño de la Arquitectura Técnica

El diseño de la arquitectura implementado para la solución de inteligencia de negocios de un modelo genérico se expone en la ilustración 14, la cual está conformada por cuatro (4) componentes y diversos procesos y elementos que permiten llevar a cabo el flujo de trabajo de la misma:



Ilustración 14: Arquitectura de Diseño con sus herramientas.

Fuente: Elaboración propia.

1. La fuente de Datos consistía en una base de datos transaccional denominada Base de datos de Visita Clínica, la cual fue suministrada de manera documental, es por ello que fue necesario, crear la BD física.
2. Es el proceso de extracción, transformación y carga que se ejecutará periódicamente, se procede a almacenar los datos de las fuentes de datos en el Almacén de Datos.

3. La estructura del Almacén de datos me permite llevar a cabo la obtención de los datos requeridos por el usuario, manteniendo un alto rendimiento.
4. El último componente se refiere a la capa de presentación, para la cual es necesario estudiar las herramientas de acceso a los datos que permite el uso del Almacén de Datos para que puedan visualizar y analizar los distintos indicadores y reportes contemplados para la solución de inteligencia de negocios.

4.4 Selección de Producto e Instalación

El proceso de selección de una herramienta tecnológica puede llegar a ser complejo debido a las múltiples alternativas existentes en el mercado (Oracle BI, Pentaho BI, IBM Cognos, MicroStrategy, entre otros) y a la diversa gama de funcionalidades que cada herramienta brinda, he decido utilizar para la creación de la fuente de datos de visita clínica y del Almacén de Datos, Postgres SQL, pues es una herramienta Multiplataforma de software libre, con la fortaleza de trabajar con multiproceso, lo cual garantiza la estabilidad del sistema.

La herramienta a utilizar para realizar la extracción, transformación y carga de los datos (ETL) es Data Integration (kettle), la cual utiliza una interfaz gráfica denominada Spoon. Las ventajas de esta herramienta PDI, es que posee una amplia comunidad, gran documentación en las redes, es de software libre y es muy intuitiva. Para el acceso y visualización de los datos he seleccionado Tableau como podemos ver en la ilustración 14. Entre las razones por las que se utilizó Tableau para el desarrollo de la solución de Inteligencia de negocios se enumeran a continuación:

- Usabilidad de la herramienta enfocada a los usuarios finales e inexpertos. No es necesario tener conocimientos técnicos ni de programación como para poder usar la herramienta.

- La instalación y configuración de la herramienta es sencilla y no requiere de conocimientos técnicos.
- Se cuenta con expertos en el área técnica dispuestos a dar asesoramiento y soporte.
- Existe una amplia documentación sobre Tableau.

4.5 Modelo Dimensional

El diseño del modelo dimensional se realizó basándose en los requerimientos definidos con anterioridad para un modelo de proceso de visita clínica de una compañía aseguradora específica, éstos fueron analizados para determinar qué elementos podrían considerarse como hechos medibles o dimensiones que pudieran formar parte de la solución de inteligencia de negocios. Para poder saber que hechos y dimensiones eran necesarios para poder realizar la solución planteada, se desglosó uno a uno los requerimientos haciendo hincapié en determinar lo que se quería medir y bajo qué perspectivas se deseaba hacer. Los hechos medibles que se tomaron en cuenta para la solución planteada son los elementos o indicadores que se quieren saber en proceso de visita clínica son:

- Cantidad de Visitadores Clínicos.
- Cantidad de Pacientes Beneficiarios.
- Variación de Presupuestos.
- Cantidad de Cartas Avales.

Así como se contemplaron los hechos, el análisis indico las diversas perspectivas con los cuales se desearía o se pudiera agrupar los datos. Así como se identificaron las distintas dimensiones que se contemplan en la solución y se mencionan a continuación:

- Tiempo.
- Paciente Beneficiario.
- Visitador Clínico
- Carta Aval

- Presupuesto
- Estado

Con los elementos expuestos anteriormente se construyó el almacén de datos, Ilustración 15.

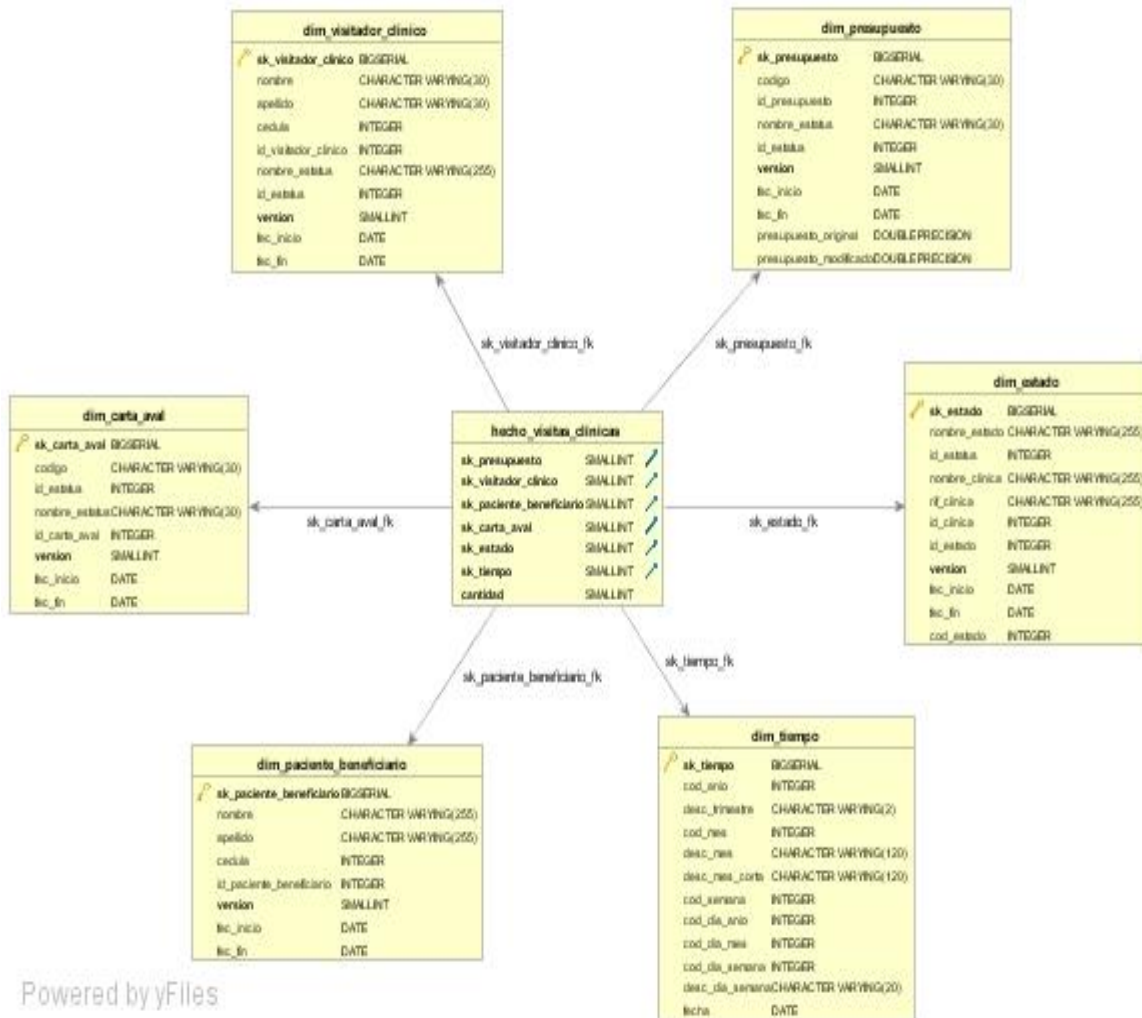


Ilustración 15: Modelo Dimensional.

Una vez definidas las perspectivas a utilizar y los hechos a medir en la solución propuesta, se procedió a identificar por cada dimensión los atributos y las jerarquías que se consideraron para poder cumplir con los objetivos planteados. Una vez establecida las dimensiones, puede establecerse las distintas jerarquías

asociadas a ella. Entendiendo que las jerarquías van a permitir navegar sobre los distintos niveles lógicos existentes. En la Ilustración 16 se observan las jerarquías establecidas para cada una de las dimensiones.

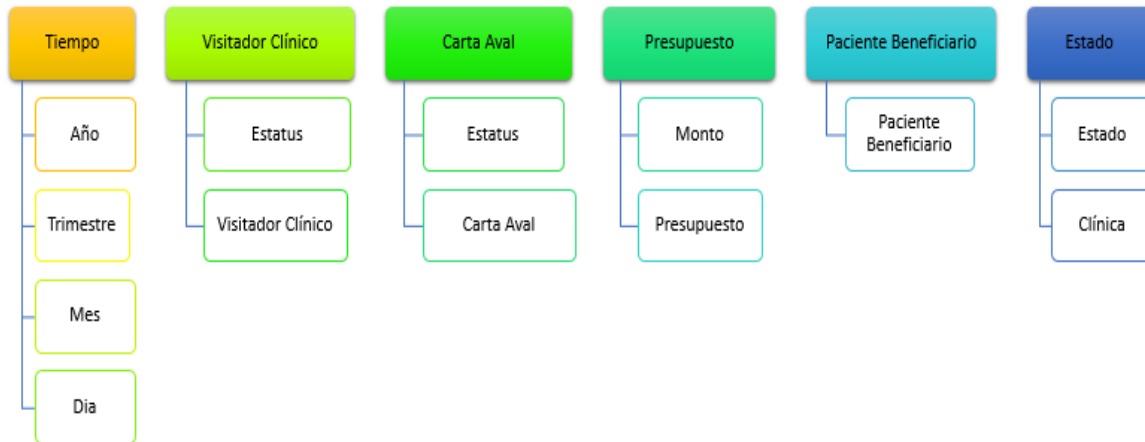


Ilustración 16: Jerarquía de las dimensiones.

La jerarquía en cada dimensión se define dependiendo de lógica del negocio. Una vez definidos los niveles jerárquicos para cada dimensión contemplada para la solución de inteligencia de negocios propuesta, se definen los atributos necesarios para cumplir con los requerimientos planteados. A continuación se describirán todas las dimensiones y la tabla de hecho contempladas para la solución de inteligencia de negocios para un proceso de visita clínica:

- Dimensión Tiempo:** permite analizar la gestión de visita clínica en función del momento en que se activa, se estima o finaliza el proceso. Por esto, dicha dimensión tomara distintos roles en el modelo dimensional en función del criterio a utilizar. Una característica particular de la dimensión tiempo es que son generadas todas las fechas acorde a un período de tiempo en el proceso de ETC, para luego asociar las fechas específicas a las ocurrencias en las instancias. Esta forma de trabajo permite generar en los cuadros de mandos calendarios porque se poseen todas las fechas, para posteriormente solo obtener las cartas avales, los visitantes clínicos, los pacientes y presupuestos a partir de un rango de fechas establecido. La tabla 6 muestra una descripción de la dimensión tiempo.

Tabla 8: Dimensión Tiempo

Atributo	Tipo de Dato	Descripción	Precedencia
SK_TIEMPO	SERIAL()	Identificador univoco de una instancia.	Generado en el proceso ETL.
FECHA	DATE	Una fecha del calendario.	Generado en el proceso ETL.
FECHA_NOMBRE	VARCHAR (255)	Fecha del calendario valida.	Generado en el proceso ETL.
ANO	INTEGER	Año de la fecha a partir del 2017.	Generado en el proceso ETL.
ANO_NOMBRE	VARCHAR (255)	Año de la fecha	Generado en el proceso ETL.
MES	INTEGER	Mes del año del 1 al 12.	Generado en el proceso ETL.
MES_NOMBRE	VARCHAR (255)	Nombre de los meses	Generado en el proceso ETL.
DIA_SEMANA	INTEGER	Día de la semana	Generado en el proceso ETL.
DIA_SEMANA_NOMBRE	VARCHAR (255)	Nombre de la semana, Domingo hasta Sábado.	Generado en el proceso ETL.

- **Dimensión Visitador Clínico:** almacena la información relacionada a los visitadores clínicos. El visitador representa un recurso crítico, en el desarrollo de este proceso, es por ello que debe vigilarse el estatus del mismo como se muestra tabla 7.

Tabla 9: Dimensión Visitador Clínico

Atributo	Tipo de Dato	Descripción	Precedencia
----------	--------------	-------------	-------------

SK_VISITADOR_CLINICO	INTEGER	Identificador univoco de una instancia.	Identificador univoco de una instancia.
VISITADOR_CLINICO_ID	INTEGER	Identificador univoco utilizado en el ambiente transaccional de solo los visitantes clínicos.	Base de datos visita_clinica.persona_id where persona.perfil='Visitador clinico'
NOMBRE	VARCHAR (30)	Primer nombre del visitador clínico.	Base de datos visita_clinica.persona.nombre where persona.perfil='Visitador clinico'
CEDULA	INTEGER	Numero de cedula o pasaporte del visitador clínico.	Base de datos visita_clinica.persona.cedula where persona.perfil='Visitador clinico'
APELLIDO	VARCHAR (30)	Primer apellido del visitador clínico.	Base de datos visita_clinica.persona.apellido Where persona.perfil='Visitador clinico'
ID_ESTATUS	INTEGER	Identificador univoco utilizado en el ambiente transaccional de los estatus de los visitantes clínicos.	Base de datos visita_clinica.estatus_visitador_clinico
NOMBRE_ESTATUS	VARCHAR (255)	Nombre descriptivo del estatus	Base de datos visita_clinica.estatus_visitador_clinico

- **Dimensión Paciente Beneficiario:** registra la información relacionada a los pacientes beneficiarios los cuales solicitan una carta aval a su compañía aseguradora. La tabla 6 muestra una descripción de la dimensión paciente beneficiario.

Tabla 10: Dimensión Paciente Beneficiario.

Atributo	Tipo de Dato	Descripción	Precedencia
SK_PACIENTE_BENEFICARIO	INTEGER	Identificador univoco de una instancia.	Identificador univoco de una instancia.
ID_PACIENTE	INTEGER	Identificador univoco utilizado en el ambiente transaccional de solo los pacientes beneficiarios.	Base de datos visita_clinica.persona.id where persona.perfil=Visitador clínico
NOMBRE	VARCHAR (255)	Primer nombre del paciente beneficiario.	Base de datos visita_clinica.persona.nombre where persona.perfil=Paciente
CEDULA	INTEGER	Numero de cedula o pasaporte del paciente beneficiario.	Base de datos visita_clinica.persona.cedula where persona.perfil=Paciente
APELLIDO	VARCHAR (255)	Primer apellido del paciente beneficiario.	Base de datos visita_clinica.persona.apellido where persona.perfil=Paciente

- **Dimensión Carta Aval:** permite gestionar las cartas avaluas asociados a la visita clínica. La tabla 9 muestra una descripción de la dimensión carta aval.

Tabla 11: Dimensión Carta Aval.

Atributo	Tipo de Dato	Descripción	Precedencia
SK_CARTA_AVAL	INTEGER	Identificador univoco de una instancia.	Generado en el proceso ETL.
ID	INTEGER	Identificador univoco utilizado en el ambiente transaccional de solo las cartas avaluas	Base de datos visita_clinica.carta_aval_id

CODIGO	VARCHAR (30)	Código de solicitud de carta aval.	Base de datos visita_clinica. carta_aval.codigo
ID_ESTATUS	INTEGER	Representa el id del estatus de la carta aval.	Base de datos visita_clinica.estatus_carta
NOMBRE_ESTATUS	VARCHAR (30)	Estatus de la carta aval. Puede ser activada o desactivada.	Base de datos visita_clinica.estatus_carta_aval.id

- **Dimensión Presupuesto:** permite gestionar los presupuestos asociados a la carta aval. La tabla 8 muestra una descripción de la dimensión presupuesto.

Tabla 12: Dimensión Presupuesto.

Atributo	Tipo de Dato	Descripción	Precedencia
SK_PRESUPUESTO	INTEGER	Identificador univoco de una instancia.	Generado en el proceso ETL.
PRESUPUESTO_ID	INTEGER	Identificador univoco utilizado en el ambiente transaccional de solo los presupuestos	Base de datos visita_clinica. presupuesto.id_presupuesto
CODIGO	VARCHAR (30)	Código del presupuesto	Base de datos visita_clinica. presupuesto.codigo
MONTO	DOUBLE	Representa el valor monetario que necesita el paciente beneficiario para realizar su operación.	Generado en el proceso ETL.
ESTATUS_ID	INTEGER	Identificador univoco utilizado en el ambiente transaccional de solo los estatus de los presupuestos y visitantes clínicos	Base de datos visita_clinica. presupuesto

ESTATUS_NOMBRE	VARCHAR (30)	Campo que indica el nombre que describe el estatus	Base de datos visita_clinica.estatus
----------------	-----------------	----------------------------------------------------	-----------------------------------------

- **Dimensión Estado:** permite gestionar los estados en donde se solicitó la carta aval La tabla 9 muestra una descripción de la dimensión Estado.

Tabla 13: Dimensión Estado.

Atributo	Tipo de Dato	Descripción	Precedencia
SK_ESTADO	INTEGER	Identificador univoco de una instancia.	Generado en el proceso ETL.
ESTADO_ID	INTEGER	Identificador univoco utilizado en el ambiente transaccional de solo los estados a los que pertenecen los pacientes y visitantes clínicos.	Base de datos visita_clinica. estado.estado_id
ESTADO_NOMBRE	VARCHAR (255)	Campo que indica el estado del país en donde se realizó la solicitud de visita clínica.	Base de datos visita_clinica. estado.nombre
CLINICA_ID	INTEGER	Identificador univoco utilizado en el ambiente transaccional de las clínicas	Base de datos visita_clinica. clínica.id
NOMBRE_CLINICA	VARCHAR (255)	Describe el nombre institucional de la clínica	Base de datos visita_clinica. clínica.nombre

- **Tabla de Hechos Gestión de Visita clínica:** El modelo dimensional busca la representación de la visita clínica caracterizada por las dimensiones. Por tanto la medida en el modelo dimensional descrito anteriormente es la incidencia de un hecho (tabla de hechos sin hechos) y no medidas específicas. Por tanto la tabla de

hechos, está conformada por un conjunto de claves foráneas que enlazan al hecho con las dimensiones, una clave que identifica unívocamente ese hecho y un atributo medida con valor por defecto 1, que ayuda a comprender que cada instancia tiene el mismo valor. La tabla 10 muestra una descripción de la tabla de hechos de visita clínica.

Tabla 10: Tabla de Hechos del Visita Clínica.

Atributo	Tipo de Dato	Descripción	Precedencia
SK_VISITADOR_CLINICO	INTEGER	Asociación con la Dimensión visitador clínico	Generado en el proceso ETL.
SK_ESTADO	INTEGER	Asociación con la Dimensión estado	Generado en el proceso ETL.
SK_PACIENTE_BENEFICARIO	INTEGER	Asociación con la Dimensión Paciente Beneficiario	Generado en el proceso ETL.
SK_CARTA_AVAL	INTEGER	Asociación con la Dimensión Carta Aval	Generado en el proceso ETL.
SK_PRESUPUESTO	INTEGER	Asociación con la Dimensión Presupuesto	Generado en el proceso ETL.
CANTIDAD	INTEGER	Forma de esclarecer que la medida de un evento es igual para cualquier instancia (1 por defecto).	Generado en el proceso de ETL:

4.6 Diseño físico

Una vez que se definió la arquitectura, las herramientas y el modelo dimensional a utilizar en la solución de inteligencia de negocios se extendieron con la realización del diseño físico. El Almacén de Datos se desarrolló físicamente usando PostgreSQL y basándose en el modelo dimensional realizado y descrito con

anterioridad. En la primera instancia, usando PostgreSQL, se definieron cada una de las dimensiones contempladas en el modelo dimensional, estableciendo inicialmente el nombre de la dimensión, luego sus atributos y por último sus niveles jerárquicos. Para crear la dimensión tiempo se utilizó Data Integration de Pentaho que permite generar esta dimensión. En este caso se especificó el nombre que se deseaba para dicha dimensión y los atributos a considerar (año, mes y día) a partir del año 2007, de esta forma la herramienta se generó la fecha con su respectiva jerarquía como vemos ilustración 17.

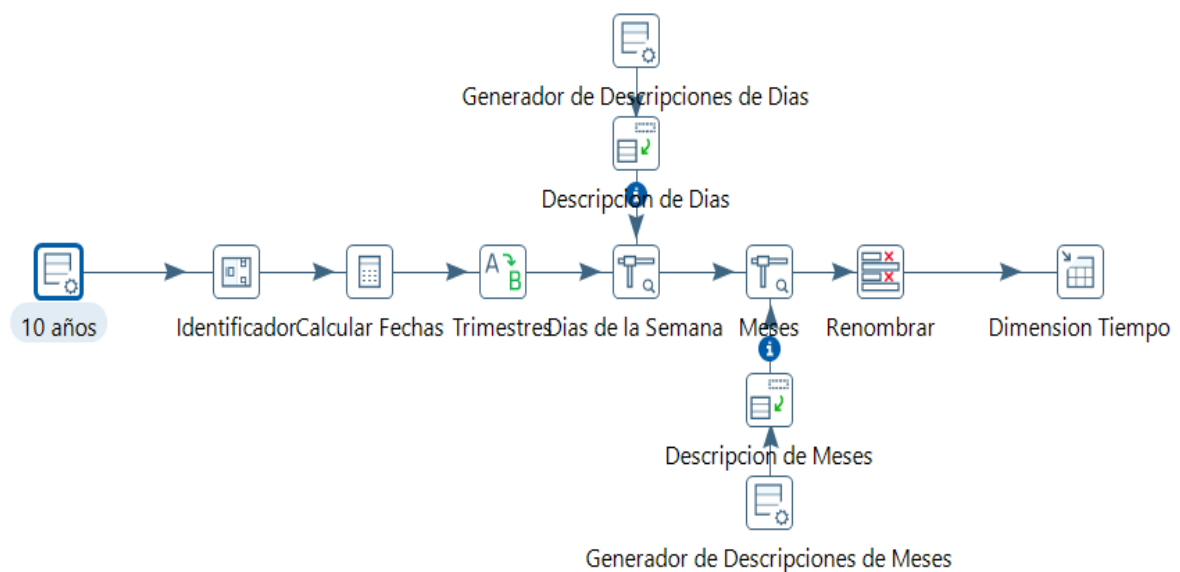


Ilustración 17: Dimensión Tiempo

Una vez completado dicho diseño se realizó el despliegue del mismo usando PostgreSQL herramienta seleccionada, de esta forma se crearon todas tablas y elementos que componen el almacén de datos propuesto para la solución de inteligencia de negocios como se muestran en la ilustración 18.

Cabe destacar que al crear las tablas para el almacén de datos, se agregaron una serie de campos o columnas necesarias para el correcto funcionamiento de las distintas aplicaciones. Estos campos adicionales no interfieren de ninguna manera con el modelo propuesto para la solución de inteligencia de negocios, tanto los requerimientos como los objetivos establecidos pueden cumplirse aún con las

modificaciones de las dimensiones, que son los atributo que tienen por nombre (versión, fecha inicio, fecha_fin), como vemos en la ilustración 19.

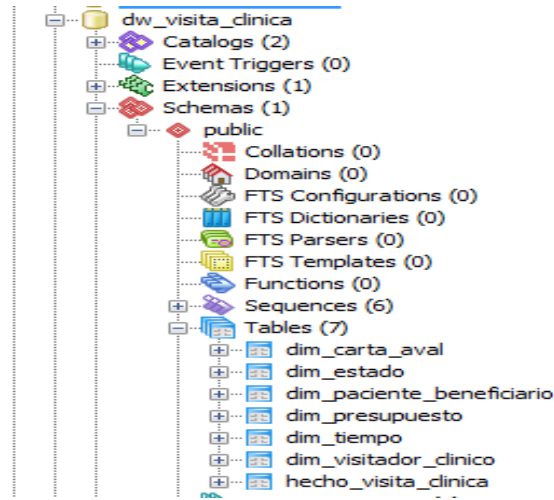


Ilustración 18: Diseño lógico de las dimensiones

Para la construcción del almacén de datos se ha utilizado el modelo dimensional planteado previamente, utilizando la base datos Visita Clínica Fuente y utilizando el manejador de bases de datos PostgreSQL, se perpetuó la construcción del almacén de datos la cual ha sido desarrollada utilizando la misma herramienta para el modelo relacional.

#	cod_presupuesto	codigo	total_presupuesto	id_estado_presupuesto	nombre_estado_p	sk_presupuesto
1	1	PP1	24000,0	1	APROBADO	1
2	10	PP10	3360000,0	1	APROBADO	2
3	11	PP11	240000,0	1	APROBADO	3
4	12	PP12	709900,0	1	APROBADO	4
5	13	PP13	60000,0	1	APROBADO	5
6	14	PP14	120000,0	1	APROBADO	6
7	15	PP15	615000,0	1	APROBADO	7
8	16	PP16	280000,0	1	APROBADO	8
9	17	PP17	4450000,0	1	APROBADO	9
10	18	PP18	24000,0	1	APROBADO	10
11	19	PP19	1134000,0	2	RECHAZADO	11
12	2	PP2	60000,0	1	APROBADO	12
13	20	PP20	47000,0	1	APROBADO	13

Ilustración 19: ETL de la Dimensión Presupuesto con los campos adicionales

4.7. Diseño de procesos ETL y desarrollo

Dado a la arquitectura propuesta se contempla (la base de datos, y el almacén de datos), fue necesario elaborar procesos ETL para las extraer los datos de la base de datos visita clínica. Un proceso capaz de permitir el llenado del al almacén de datos usando como fuente de datos la base de datos. Dicho proceso se muestra a continuación

4.7.1. Diseño de Procesos de ETL y Desarrollo (Almacén de datos)



Ilustración 20: ETL de la Dimensión Visitador Clínico.

El diseño de los procesos de extracción, transformación y carga para el almacén de datos se realizaron utilizando la herramienta Data Integration de Pentaho (kettle). La realización de un proceso ETL para cada una de la dimensiones es similar al mostrado en la Ilustración 20.

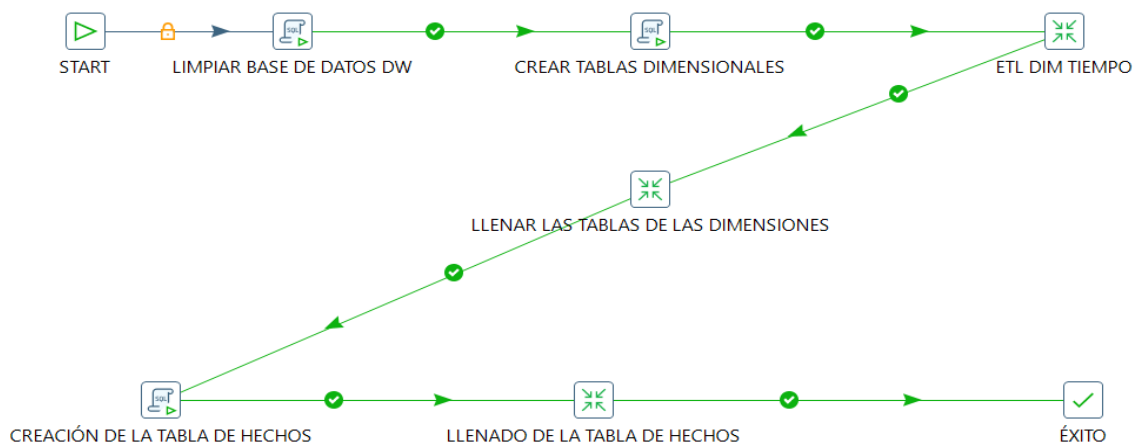


Ilustración 21: Trabajo para implementar el modelo dimensional.

Y para la tabla de hechos conllevan un conjunto de elementos adicionales que serán descritos a continuación. Para la realización del proceso de la tabla de hecho

es requerido previamente ejecutar un ETL asociado a las tablas de dimensiones que se relaciona ese hecho como se muestra en la ilustración 21. Para la realización de todas las transformaciones fueron utilizados elementos de diseño, específicamente elementos de entrada (table input), de salida (table output), distintos elementos para transformar, como secuencias, concatenación, ordenamiento, etc, elementos de tipo unión.

A continuación se describen las tareas necesarias para la construcción del modelo dimensional:

1. Limpieza de tablas dimensionales y tabla de hechos, lo cual nos permite generar el proceso desde 0, una vez ejecutado.
2. Se crea un SQL para la creación de las tablas dimensionales con los atributos mencionados anteriormente.
3. Para la Dimensión Tiempo se ha creado una Transformación adicional que permite general el tiempo a partir del año 2007 al 2017.

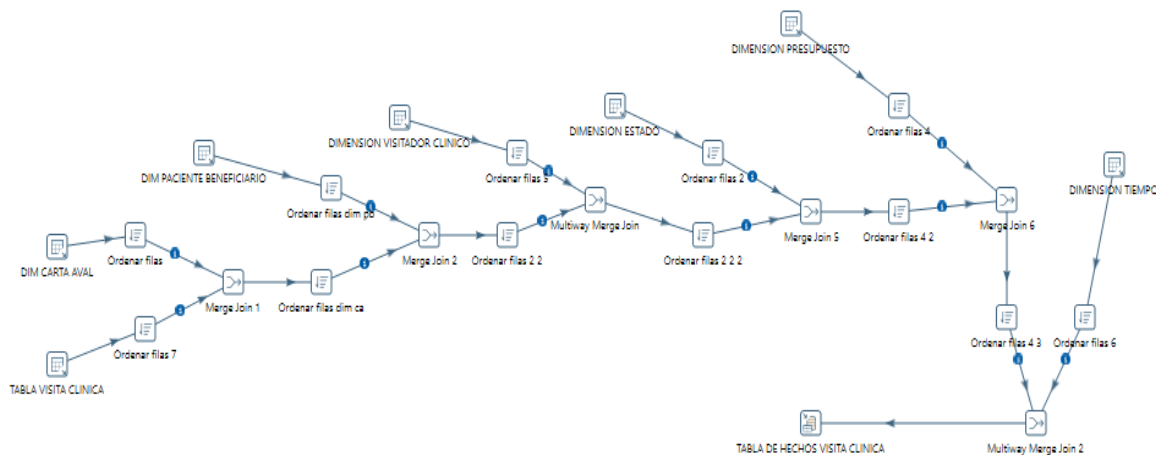


Ilustración 22: ETL Tabla de hecho Visita Clínica

4. Por medio de una serie de consultas en las distintas tablas del modelo relacional, se procede a cargar las tablas del modelo dimensional.
5. Se crea un SQL para la creación física de la tabla de hechos.

6. Se carga la tabla de hechos a partir de una serie de consultas que son extraídas del modelo relacional (ilustración 22).

4.7.2. Detallar los requerimientos según el modelo

Desde el estudio del modelo dimensional, que plasma los datos relevantes de la base de datos, específicamente las cartas avals, presupuestos, visitantes clínicos, pacientes beneficiario, y visitantes clínicos utilizando los gráficos que se estaban generando en la base de datos para buscar satisfacer las necesidades analíticas, se han propuesto un conjunto de requerimientos analíticos presentados a continuación.

- Cantidad de visitantes clínicos por estatus.
- Porcentaje de visitantes clínicos por estatus.
- Cantidad de visitantes clínicos por estatus en un periodo de tiempo (Año, mes, día, trimestre).
- Cantidad de pacientes beneficiarios.
- Cantidad de pacientes beneficiarios según estatus, en un periodo de tiempo (Año, mes, día, trimestre).
- Variación de gastos totales anuales de presupuesto por paciente beneficiario
- Cantidad de cartas avals según el estatus.
- Porcentaje de cartas avals según estatus.
- Cantidad de cartas avals según estatus en un periodo de tiempo (Año, mes, día, trimestre).
- Monto pendiente de pagos por cartas avals activas por periodo de tiempo (Año, mes, día, trimestre) y clínica.
- Cantidad pendiente de cartas avals activas por periodo de tiempo (Año, mes, día, trimestre) y clínica.
- Variación de cantidad pendiente de presupuesto en porcentaje.
- Variación de monto pendiente de presupuesto en porcentaje.

4.8 Desarrollo de los requerimientos de Información

Para el desarrollo de la solución, el primer paso fue definir como se visualizarían los indicadores que la aplicación se encargará de calcular, basándonos en el usuario que se mencionan en el Capítulo 1, especialmente el visitador clínico. Como planteó en el Capítulo 3, específicamente en la actividad correspondiente al Desarrollo de la Aplicación de Inteligencia de Negocio, ésta se realizó siguiendo un diseño basado en prototipos, ágil, iterativo e incremental. Para el funcionamiento de la solución, fue necesario el desarrollo de varios procesos ET (extracción y transformación de los datos), dado que la herramienta a utilizar para el desarrollo de la solución requiere que los datos tengan un formato estructurado, tomando en cuenta que partimos de una base de datos Fuente en donde la información se encuentra distribuida en múltiples tablas que necesitan ser organizadas.

Es necesario recalcar que los ET están fuertemente vinculados a los datos que se visualizan en la interfaz, por lo tanto, cada vez que se definen o modifican los filtros o indicadores, el ET asociado debe modificarse o crearse igualmente. Para crear los ET, se utilizó *Pentaho Data Integration*. Para la creación de los indicadores de gestión usando la herramienta de inteligencia de negocios (Tableau), de forma tal que puedan satisfacerse los requerimientos analíticos planteados previamente. Para esto se debe configurar desde la herramienta a utilizar la conexión a la base de datos en este caso es la conexión a PostgreSQL, realizar las consultas que satisfacen los requerimientos analíticos para luego de forma sencilla poder visualizar los distintos indicadores dispuesto de una forma ordenada. Por último se realiza una verificación de los datos para confirmar la exactitud de los indicadores.

4.8.1 Verificar la calidad de los datos

Una vez generado la visualización de los datos, se realiza la verificación de los datos. En este caso fueron analizadas las distintas dimensiones lo cual implica la realización de consultas sobre la base de datos relacional ajustando los indicadores

de gestión a partir de los filtros planteados las fuentes de datos para comprobar su igualdad. Dicha verificación consiste en la aplicación de un conjunto de consultas en ambos ambientes y evaluar los resultados para establecer si no existen diferencias. A continuación se presenta la verificación por cada elemento del modelo dimensional.

- Dimensión Visitadores Clínicos:** Se extraen los datos de las personas que solo son visitantes clínicos, para ello es necesario revisar el perfil de las personas y en caso de que sean iguales a 3 luego será visitantes clínicos. Para obtener el estatus del visitante será necesario consultar la tabla estatus. En el caso del modelo dimensional, solo es necesario contar los registros de visitantes clínicos registrados, como se muestra en la Ilustración 23.

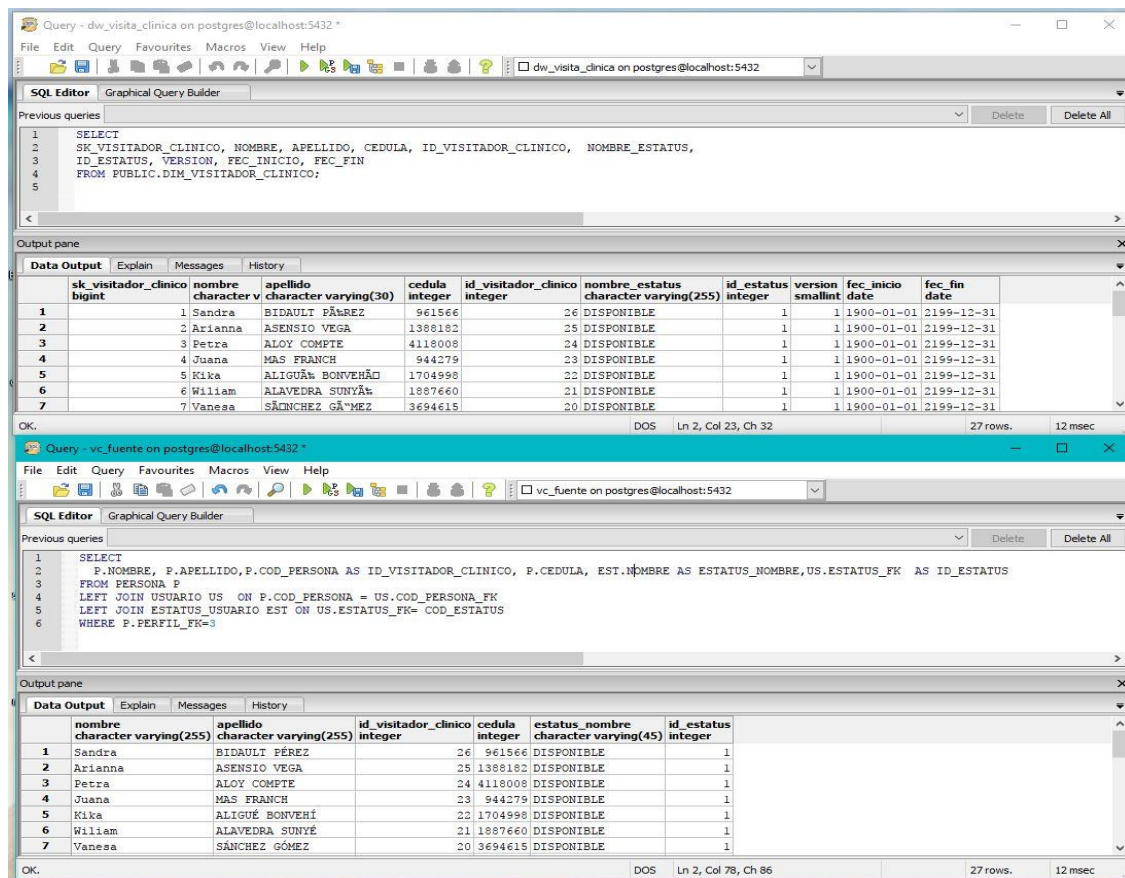


Ilustración 23: Verificación de la dimensión visita clínica.

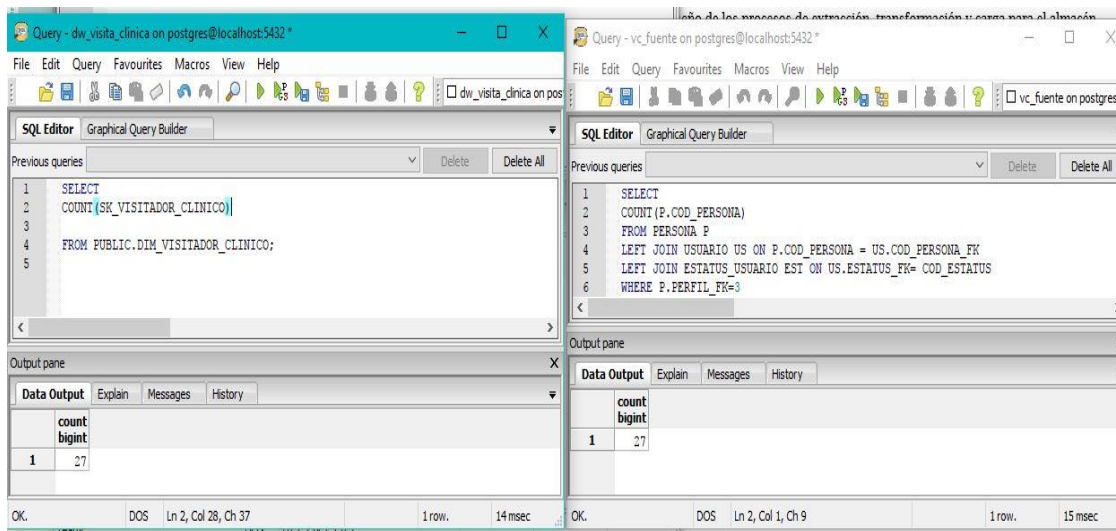


Ilustración 24: Verificación de la cantidad de los datos de la dimensión visitador clínico.

- Dimensión Paciente Beneficiario:** Se extraen los datos de las personas que solo son pacientes beneficiarios de la base de datos relacional, para ello es necesario revisar el perfil de las personas y en caso de que sean iguales a 4 luego será pacientes registrados. Para obtener el estatus del paciente será necesario consultar la tabla estatus de su carta aval. En el caso del modelo dimensional, solo es necesario contar los registros de pacientes registrados como se muestra en la ilustración 25.

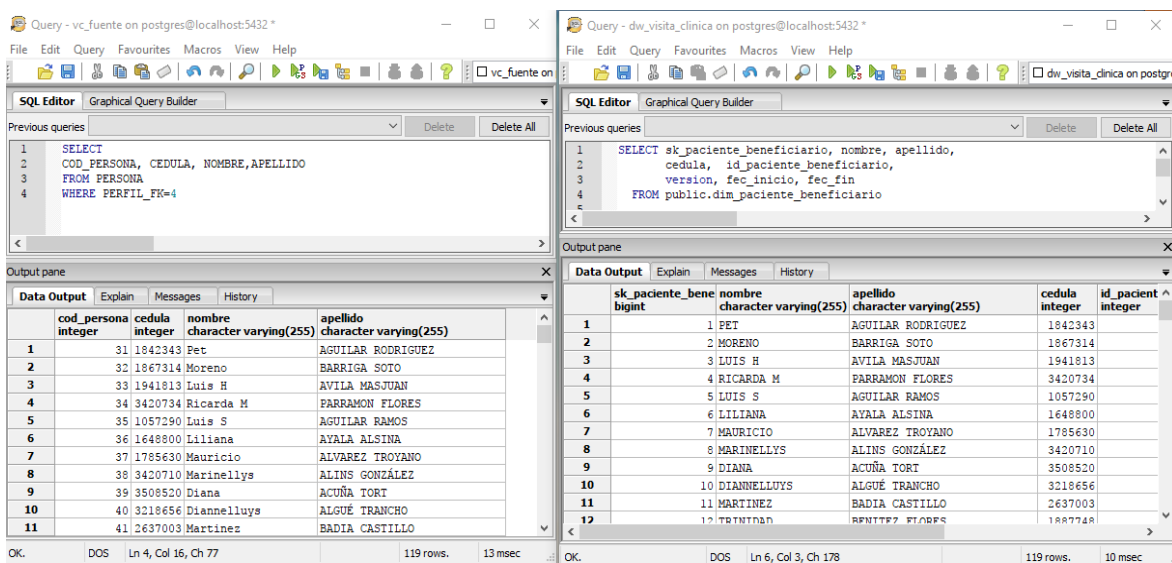


Ilustración 25: Verificación de la tabla dimensión paciente beneficiario.

Luego de ejecutar la consulta, se verifico que existen 119 pacientes beneficiarios almacenados en el repositorio y en la fuente de datos.

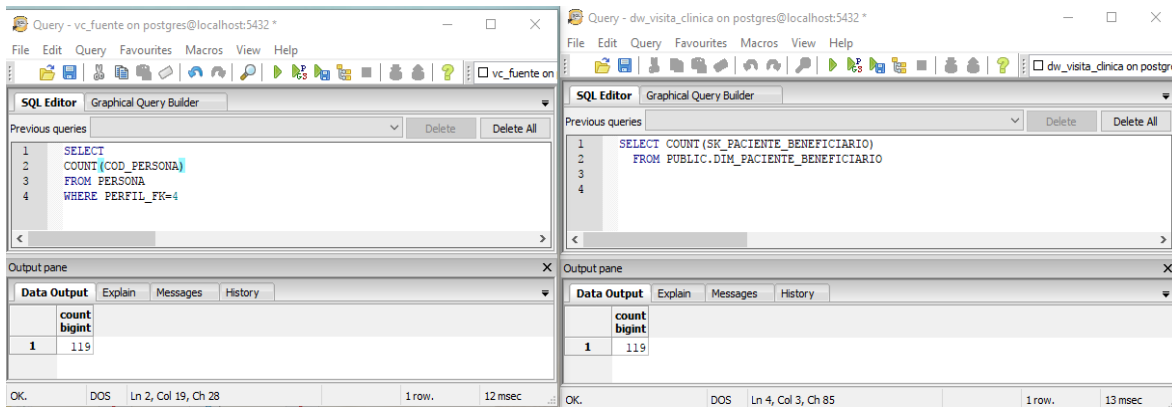


Ilustración 26: Verificación de la cantidad de los datos de la dimensión paciente beneficiario.

- Dimensión presupuesto.** Se extraer los datos de los presupuestos de los pacientes, para ello es necesario revisar el id de cada paciente e interceptar las tablas estatus y gastos de presupuestos. En el caso del modelo dimensional, solo es necesario contar los registros de los presupuestos como se muestra en la ilustración 27. Para obtener los gastos de los presupuestos asociados a un cliente, fue necesario crear un ETL, que sumarice los gatos del asegurado y de esta manera obtener el presupuesto original y el presupuesto modificado, es por esta razón que existe muchos más registros en la fuente.

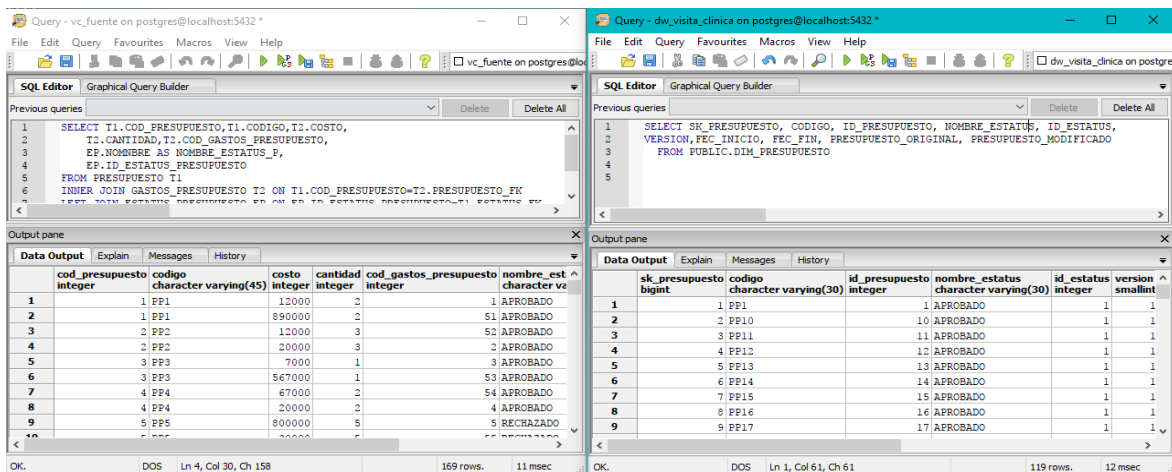


Ilustración 27: Verificación de fuente con la dimensión presupuesto.

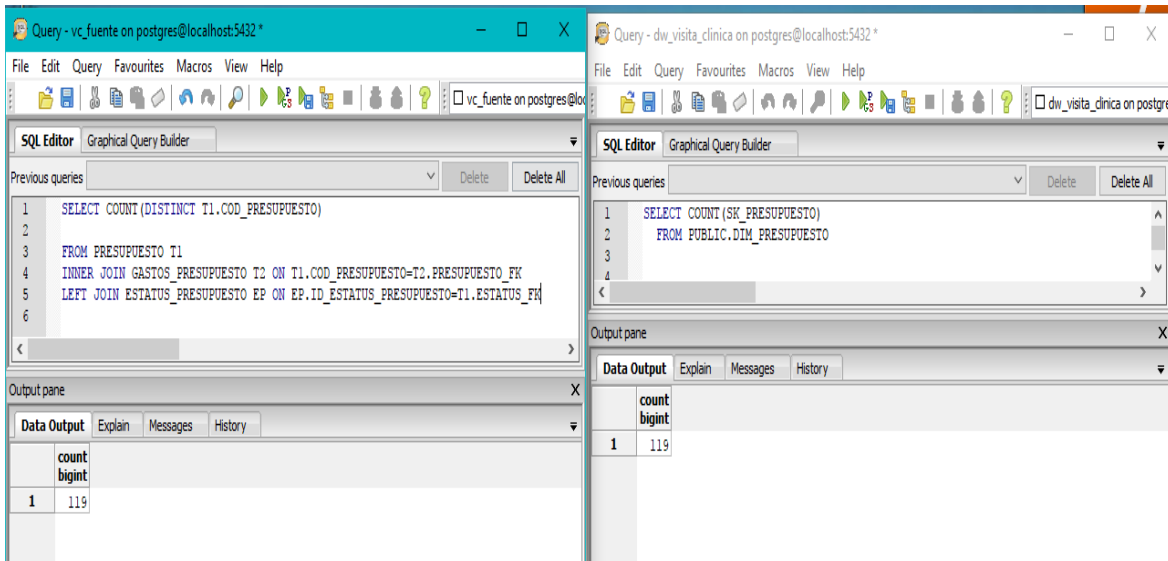


Ilustración 28: Verificación de la cantidad de los datos de la dimensión presupuesto.

- **Dimensión carta aval.** Se extraer los datos de las cartas avales relacionadas a un paciente y gestionada por un visitador clínico en particular, por tanto es necesario interceptar las tablas de estatus y carta aval como se muestra en la ilustración 29.

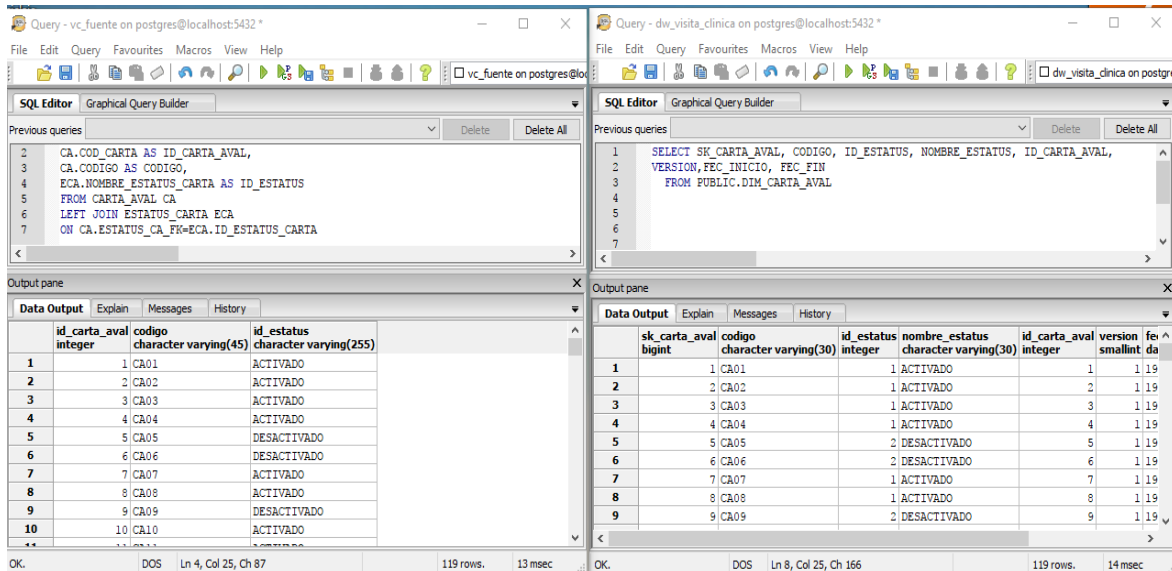


Ilustración 29: Verificación de fuente con la dimensión carta aval.

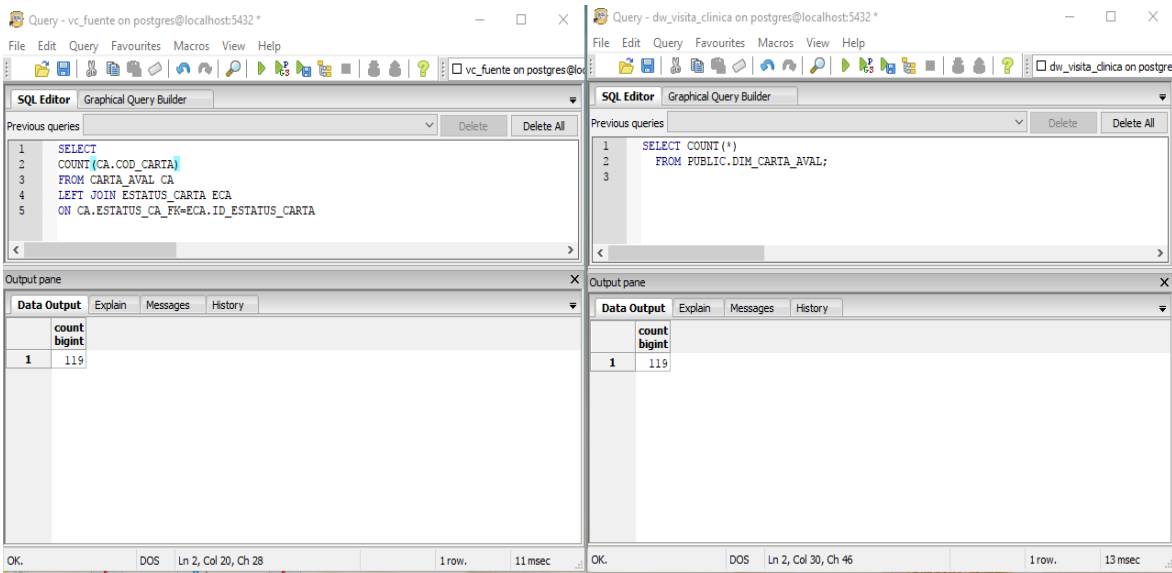


Ilustración 30: Verificación de la cantidad de los datos de la dimensión carta aval.

- **Dimensión Estado.** Se extraen los datos de los estados relacionados a la solicitud de visita clínica, por tanto es necesario interceptar las tablas de estados con las clínicas asociadas a la compañía aseguradora (ilustración 31).

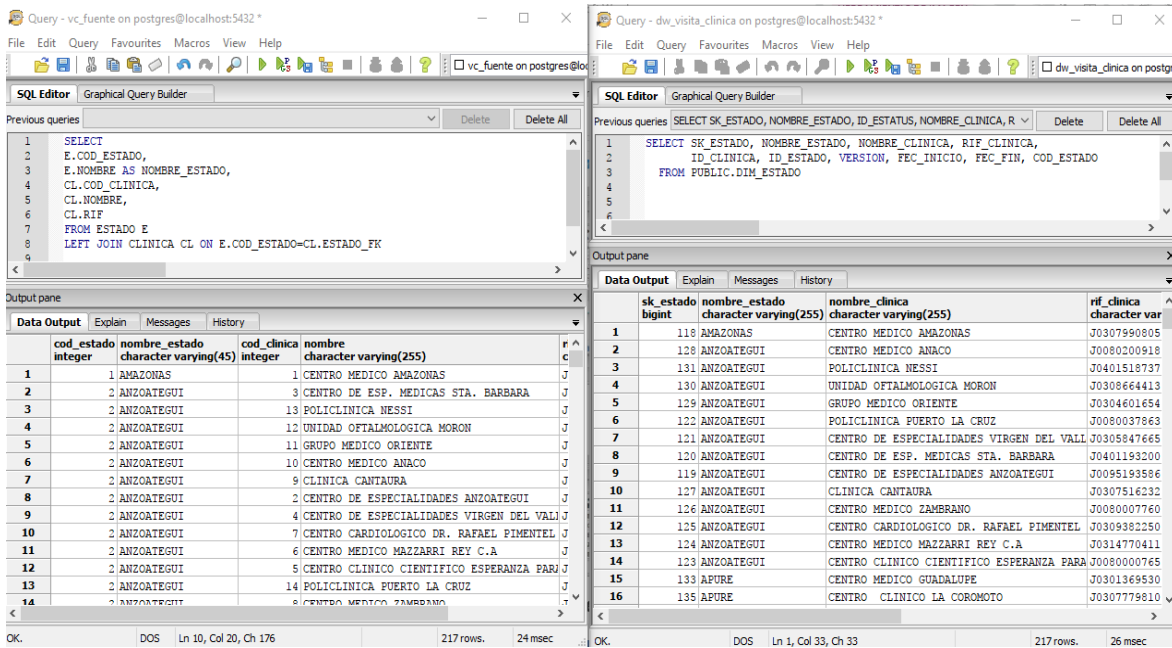


Ilustración 31: Verificación de fuente con la Dimensión Estado.

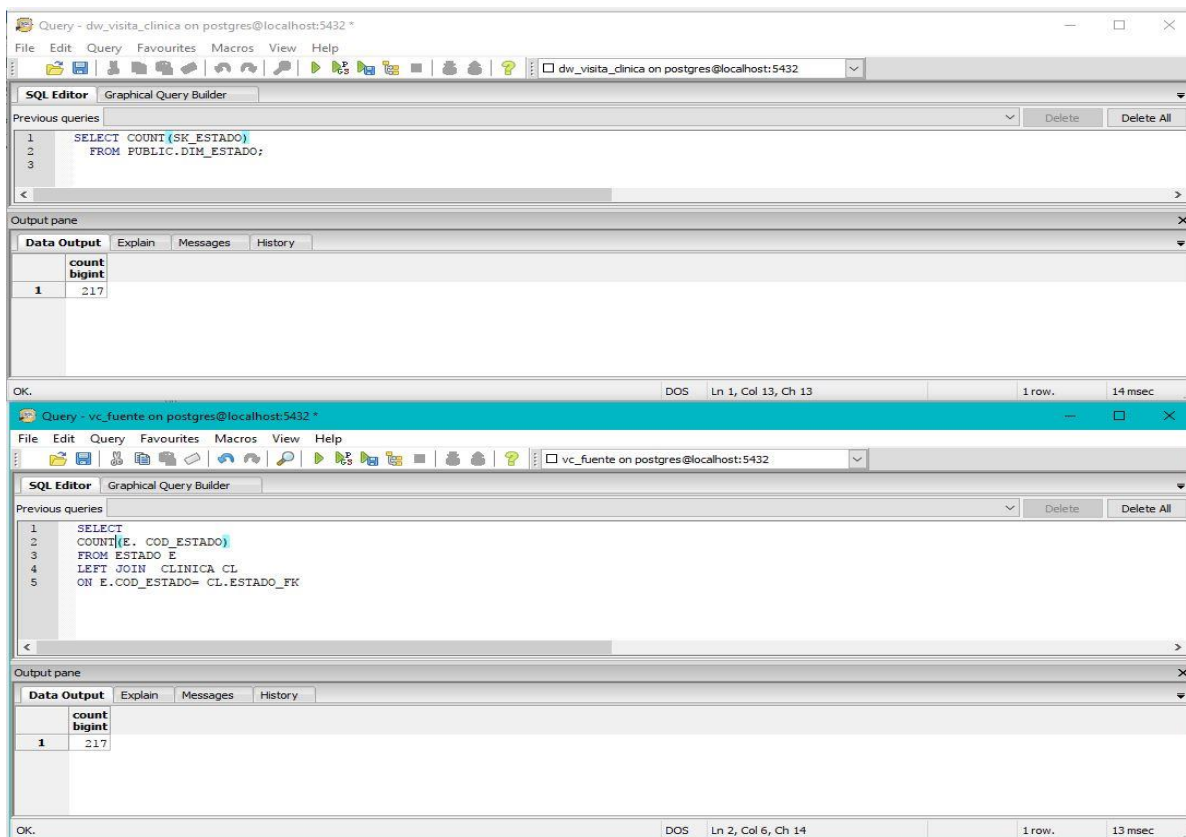


Ilustración 32: Verificación de la cantidad de los datos de la Dimensión Estado.

- **Dimensión Tiempo.** La dimensión no requiere una verificación de datos con respecto a la base de datos, sin embargo puede confirmarse que los datos son consistentes en función de una fecha.

4.8.2. Desarrollo y distribución de consultas

Para la realización de las consultas, se selecciona la tabla de hecho con sus dimensiones asociadas a ella para luego ir analizar dichas tabla de hecho. Se detallan las tablas de hechos con sus dimensiones. Una de las ventajas que ofrece Tableau es que permite las agrupaciones de las dimensiones a la tabla de hecho siempre que esta tenga una relación con esas tablas. Se muestran en la ilustración 33:

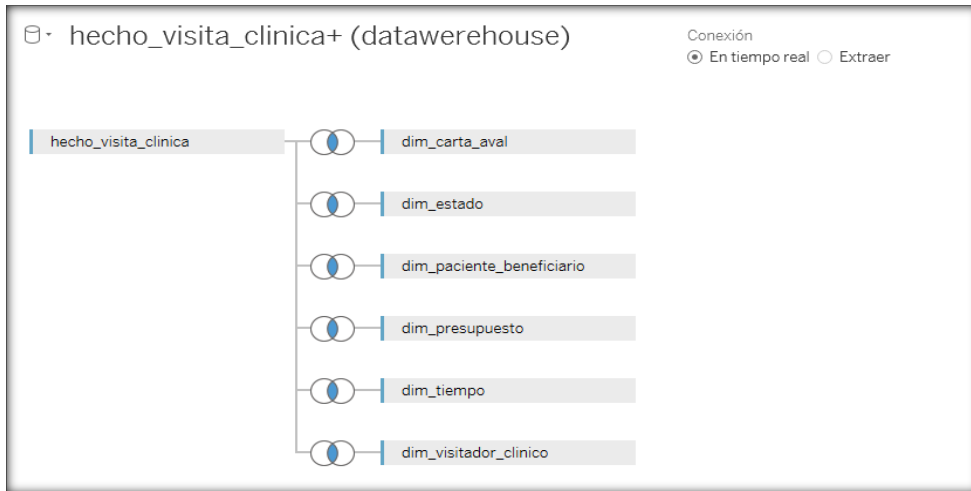


Ilustración 33: Tabla de hechos con sus dimensiones asociadas

Dentro de Tableau se pueden generar distintos cuadros de mando o sheets (hojas en inglés) dentro de un proyecto o archivo .tde. De esta forma se puede organizar los datos de un modelo dimensional en bloques separados. Para la realización de las consultas, se ha optado por generar cuatro cuadros de mando (cada hoja corresponde a un cuadro de mando dentro del proyecto). El primer cuadro de mando muestra un vistazo general del modelo implementado como muestra la figura 34.

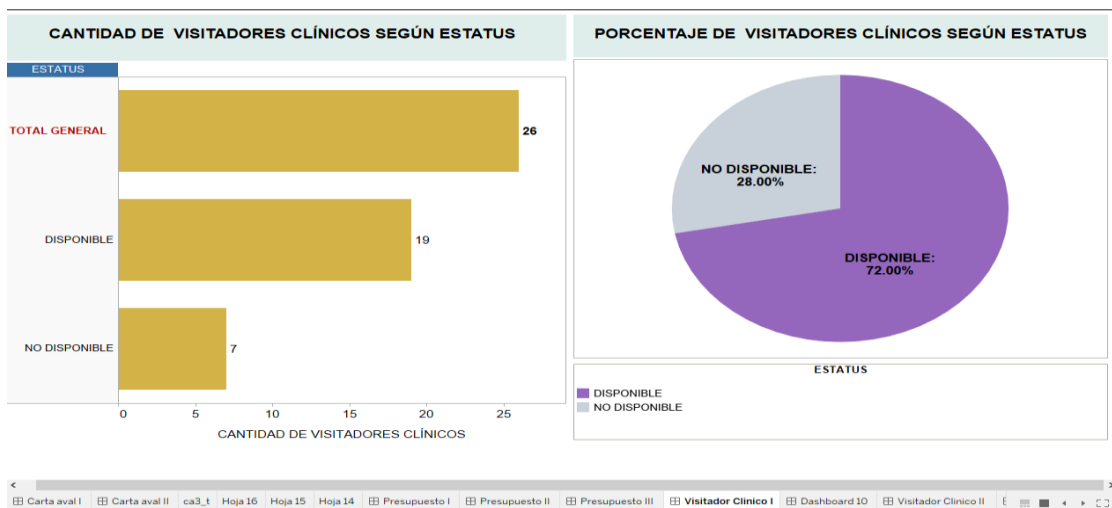


Ilustración 34: Cuadro de Mando de Visita Clínica

Se puede navegar dentro de las distintas dimensiones observando el máximo nivel de agregación en cada una de ellas.

4.8.3 Generación de indicadores de gestión

Se presentan todas las tablas con que se relaciona la tabla de hecho visita clínica que se desea analizar para responder los indicadores nombrado anteriormente.

- **Cantidad de visitantes clínicos por estatus:** se puede observar en la ilustración 35 la disponibilidad de todos los visitantes clínicos involucrados en la visita clínica en tiempo real, lo cual permite a la institución manejar de forma eficiente sus recursos. La ventaja de utilizar una tabla descriptiva es posibilidad de ubicar un determinado recurso de forma automática.

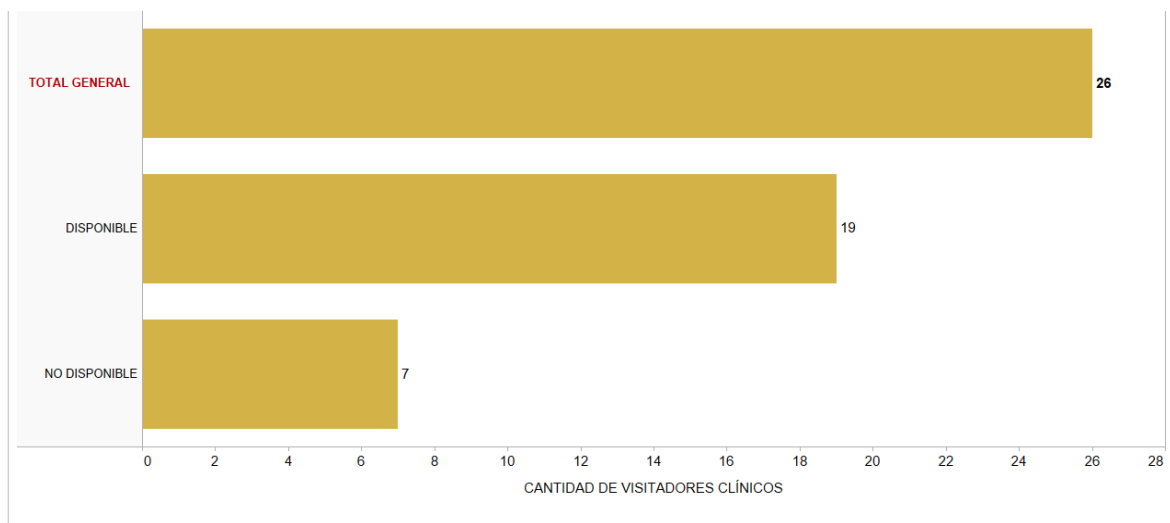


Ilustración 35: Cantidad de visitantes clínicos por estatus

- **Porcentaje de visitantes clínicos por estatus:** La ilustración 36 nos permite verificar la disponibilidad de los visitantes clínicos de la institución de manera gráfica, permitiendo vigilar la eficiencia de sus recursos de forma porcentual.

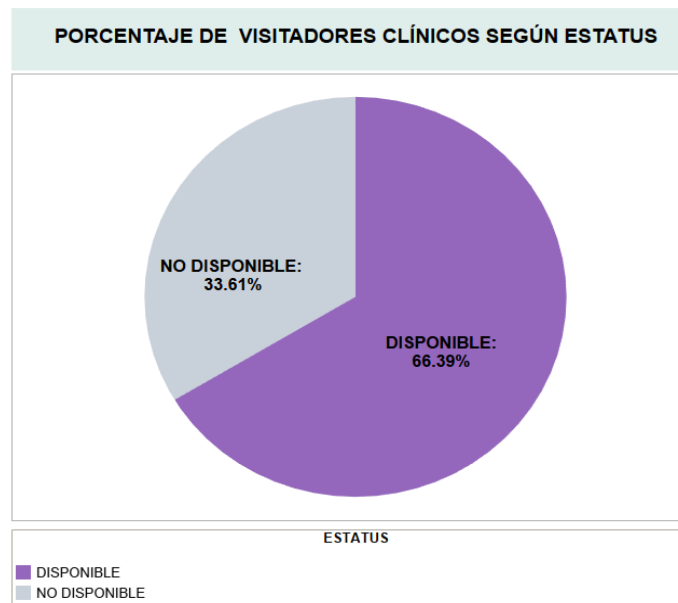


Ilustración 36: Porcentaje de visitantes clínicos por estatus.

- **Cantidad de *visitadores clínicos por estatus* en un periodo de *tiempo* (Año, mes, día, trimestre):** Por medio del cuadro de mando es posible visualizar la información asociada a los visitantes clínicos en los distintos niveles de jerarquías definidos en la dimensión tiempo (Ilustración 37).

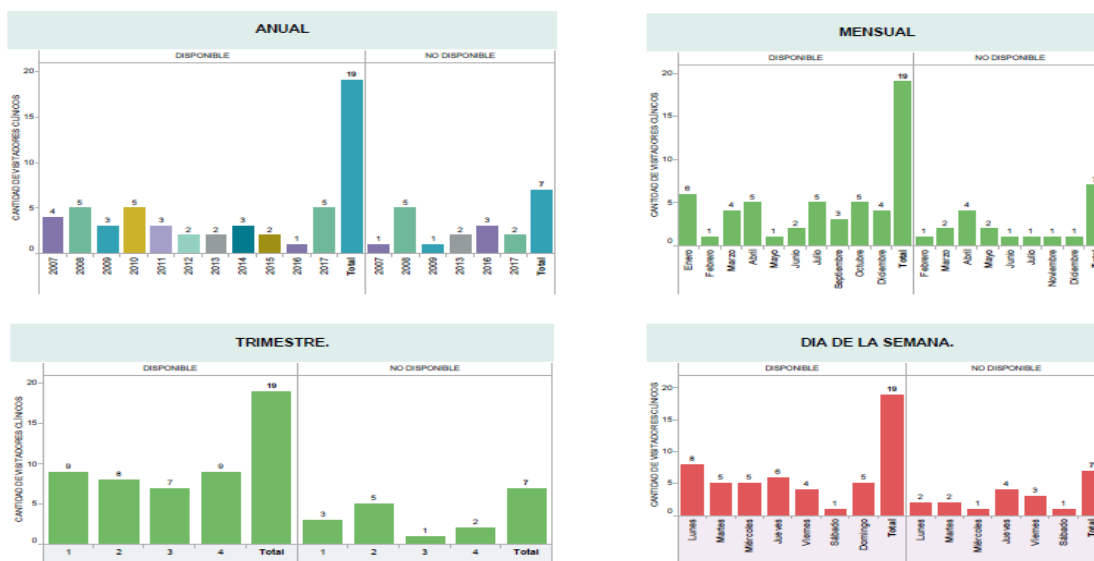


Ilustración 37: Cantidad de visitantes clínicos por estatus en un periodo de tiempo (Año, mes, día, trimestre).

- **Cantidad de pacientes beneficiarios según el estado en el que se realizó la solicitud de visita clínica:** nos permite monitorear las solicitudes de visitas clínicas realizadas por los clientes de la compañía aseguradora, tomando como referencia la procedencia de la solicitud y el estatus que posee el paciente beneficiario en tiempo presente. (Ilustración 38).

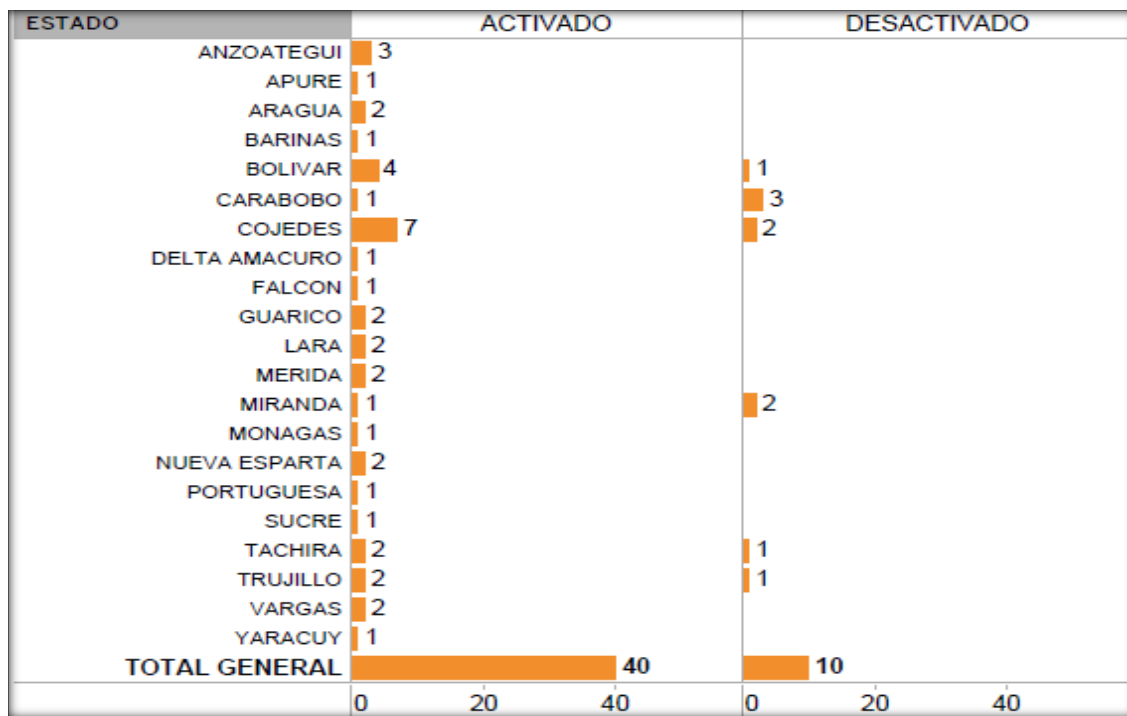


Ilustración 38: Cantidad de pacientes beneficiarios según el estado en el que se realizó la solicitud de visita clínica

- **Cantidad de *pacientes beneficiarios* según estatus, en un periodo de tiempo (Año, mes, día, trimestre):** Un elemento importante dentro de las herramientas de inteligencia de negocios es poder representar un indicador de gestión de distintas formas. Aunque la representación de los visitantes clínicos por estatus está presente en el cuadro de mando general, se ha optado por incluir dicho indicador dentro de este cuadro de mando para facilitar la visualización de los visitantes clínicos según una fecha determinada y mostrar la capacidad de la herramienta de generar un indicador de múltiples formas (Ilustración 39).

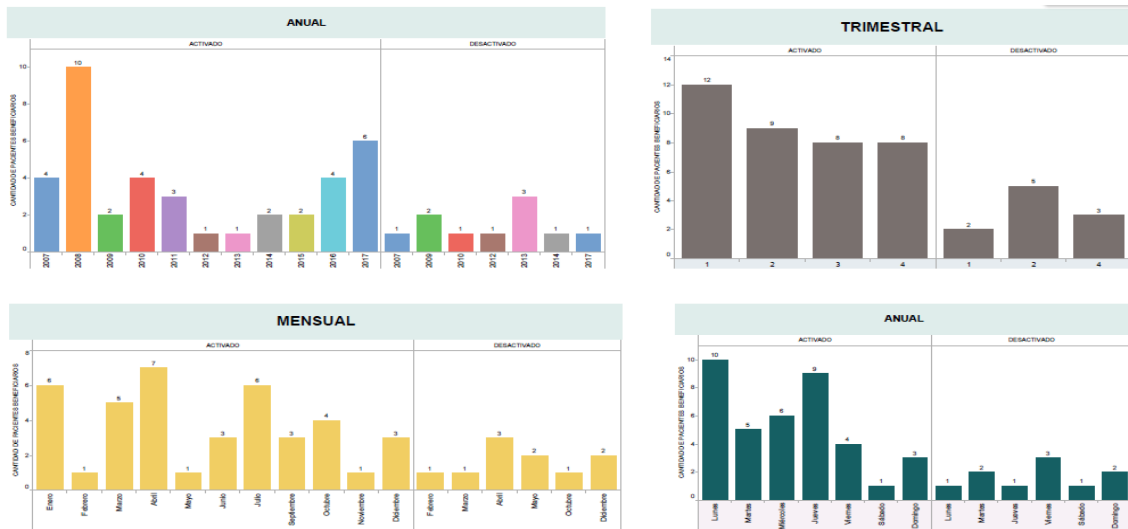


Ilustración 39: Cantidad de pacientes Beneficiarios por periodo de tiempo (Año, Mes, Trimestre, Día de la Semana)

- **Monto pendiente de pagos por cartas avales activas por periodo de tiempo (Año, Trimestre, Mes, Día de la Semana) y clínica:** El siguiente indicador, representa la sumatoria de todas las cartas avales agrupados por clínica en donde se ha actualizado el presupuesto en periodos de tiempo de año, trimestre, mes y día de la semana, como se describe en la Ilustración 40.

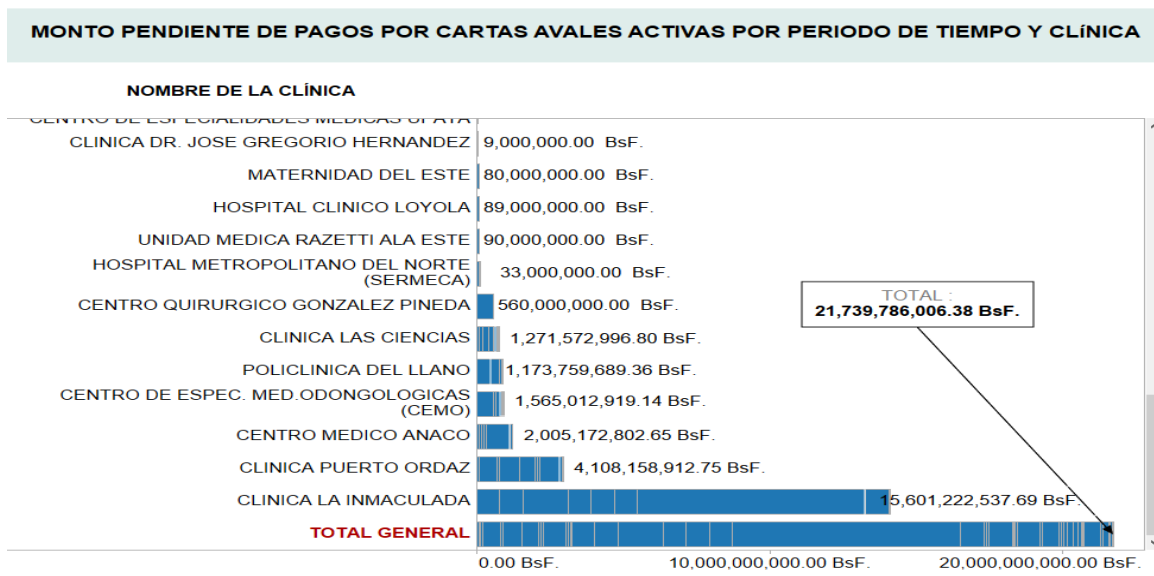


Ilustración 40: Monto pendiente de pagos por cartas avales activas por periodo (Año, Trimestre, Mes, Día de la Semana) de tiempo y clínica.

- **Cantidad pendiente de cartas avales activos por periodo de tiempo (Año, Trimestre, Mes, Día de la Semana) y clínica:** En la ilustración 41 se describe la cantidad de cartas avales que han sido actualizadas y tienen un monto pendiente por solventar. Las cantidades asociadas, se encuentran agrupadas por periodos de tiempo y clínica asociada.

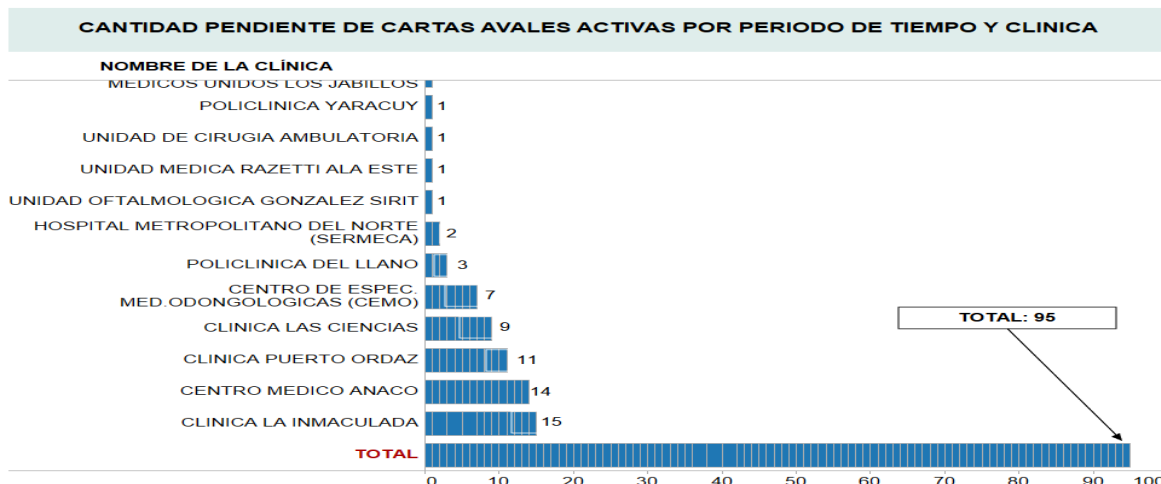


Ilustración 41: Cantidad pendiente de cartas avales activas por periodo de tiempo y clínica

- **Variación de cantidad pendiente de presupuesto en porcentaje:** En la ilustración 42 se describe el indicador de variación de cantidades de presupuestos pendientes por cancelar, los cuales se encuentran de forma porcentual. A partir de este indicador, es posible monitorear los presupuestos que han sido alterados al menos una vez, a lo largo del proceso.

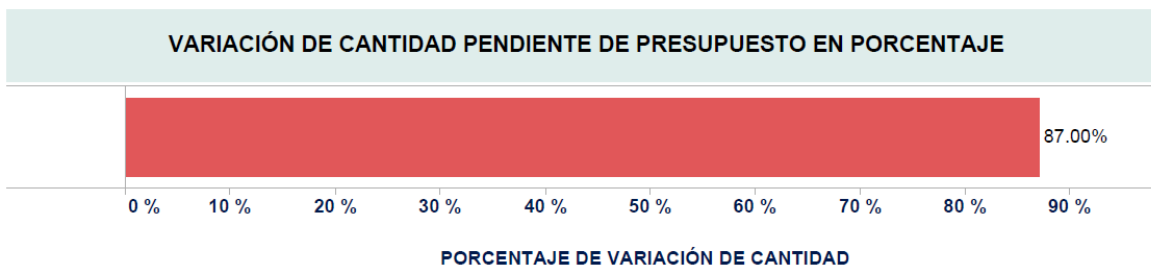


Ilustración 42: Variación de cantidad pendiente de presupuesto en porcentaje.

- **Variación de monto pendiente de presupuesto en porcentaje:** En la ilustración 43 se describe la alteración porcentual de la sumatoria de todos los presupuestos actualizados con respecto al monto original del presupuesto.

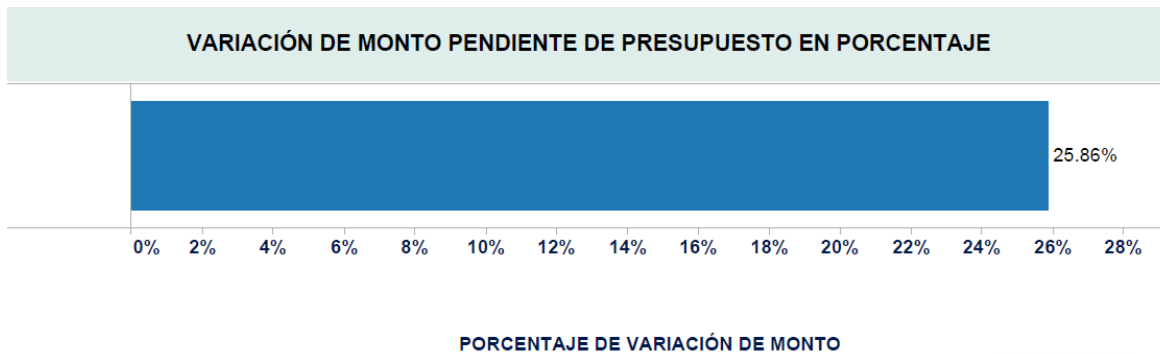


Ilustración 43: Varianza de gastos de presupuestos por año

- **Cantidad de cartas avals según el estatus:** en la ilustración 44, se describe el estatus de la visita clínica, pues es a partir del documento de carta aval, que es posible gestionar el proceso solicitado.

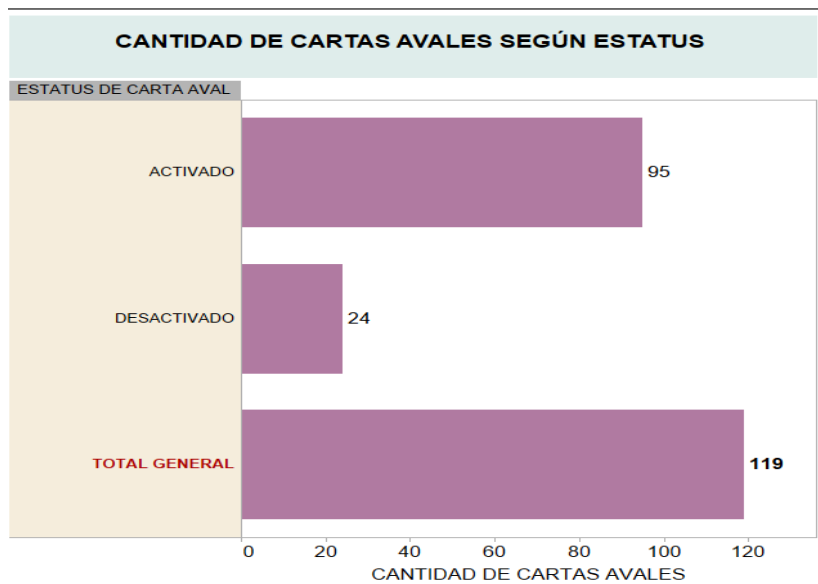


Ilustración 44: Cantidad de cartas avals por estatus.

- **Porcentaje de cartas avals según estatus:** La ilustración 45 representa la cantidad de cartas avals según el estatus del documento de forma porcentual.

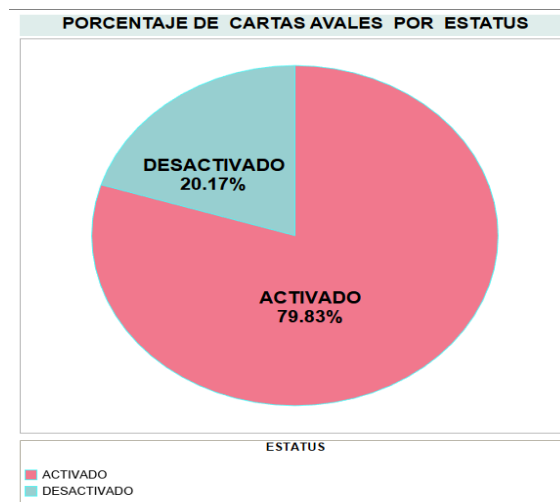


Ilustración 45: Porcentaje de cartas avales por estatus.

- Cantidad de cartas avales según estatus en un periodo de tiempo (Año, mes, día, trimestre):** Por medio del cuadro del siguiente cuadro de mando es posible visualizar la información asociada a las cartas avales en los distintos niveles de jerarquías definidos en la dimensión tiempo. Vigilar el estatus de las cartas avales en los distintos periodos de tiempo, resulta eficiente para valorar la cantidad de solicitudes realizadas por los pacientes (Ilustración 46).

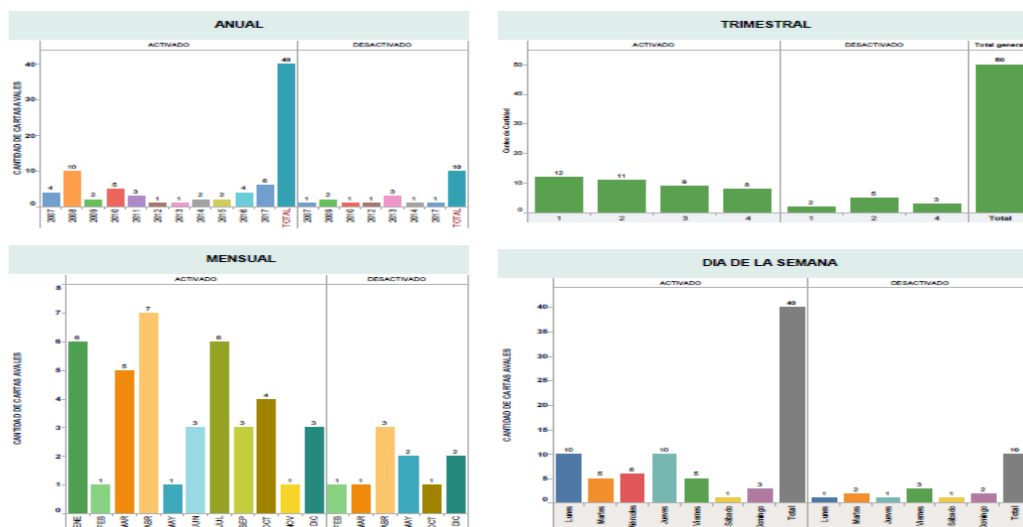


Ilustración 46: Cantidad de cartas avales por estatus por periodo de tiempo (Año, trimestre, mes, día de la semana).

4.8.3 Pruebas de aceptación de usuario

Para verificar el cumplimiento de los requerimientos no funcionales especificados para el sistema y el nivel de aceptación de los usuarios finales en cuanto a las interfaces de usuario, se elaboró una encuesta de cuatro preguntas presentada en la tabla 11. Esta encuesta fue aplicada a un grupo de 12 entre los cuales se encontraban: 4 especialistas en Inteligencia de negocio, 3 personas sin conocimientos en el área tecnológica, 2 especialistas en sistemas de información y 3 usuarios finales los cuales se encargaron de analizar la retroalimentación obtenida por parte de estos, y detectar en que aspectos se pueden mejorar las interfaces con la intención de mejorar la experiencia del usuario.

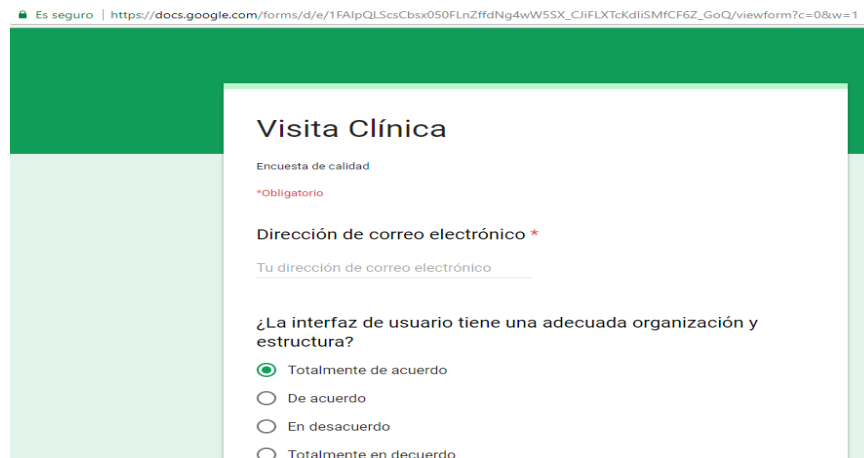


Ilustración 47: Interfaz encuesta de aceptación.

Preguntas	Totalmente de acuerdo	De acuerdo	En desacuerdo	Totalmente en desacuerdo
¿La interfaz de usuario tiene una adecuada organización y estructura?	9	3	0	0
¿El diseño de la interfaz es consistente?	10	1	1	0
¿Los reportes le brindaron un apoyo en la toma de decisiones?	8	2	1	1
¿El tamaño de los gráficos y fuentes son legibles?	9	1	1	1

¿Considera que el sistema es de fácil uso?	12	0	0	0
¿El sistema es de navegación intuitiva?	10	2	0	0

Tabla 14: Tabla resumen de encuesta de aceptación.

A continuación se presentan los resultados obtenidos para cada pregunta de la encuesta:



Ilustración 48: Resultados de la pregunta N° 1. Fuente: Prueba de aceptación con usuarios finales

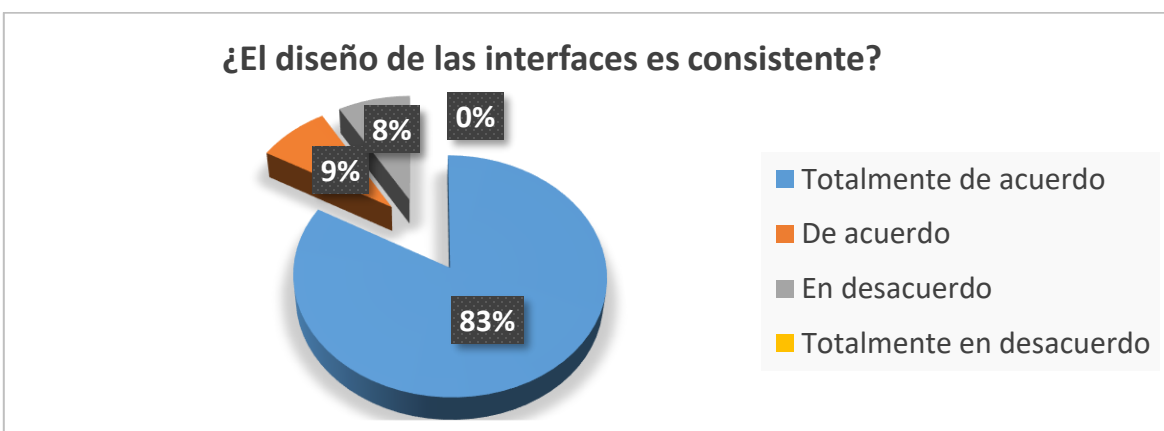


Ilustración 49: Resultados de la pregunta N° 2. Fuente: Prueba de aceptación con usuarios finales.

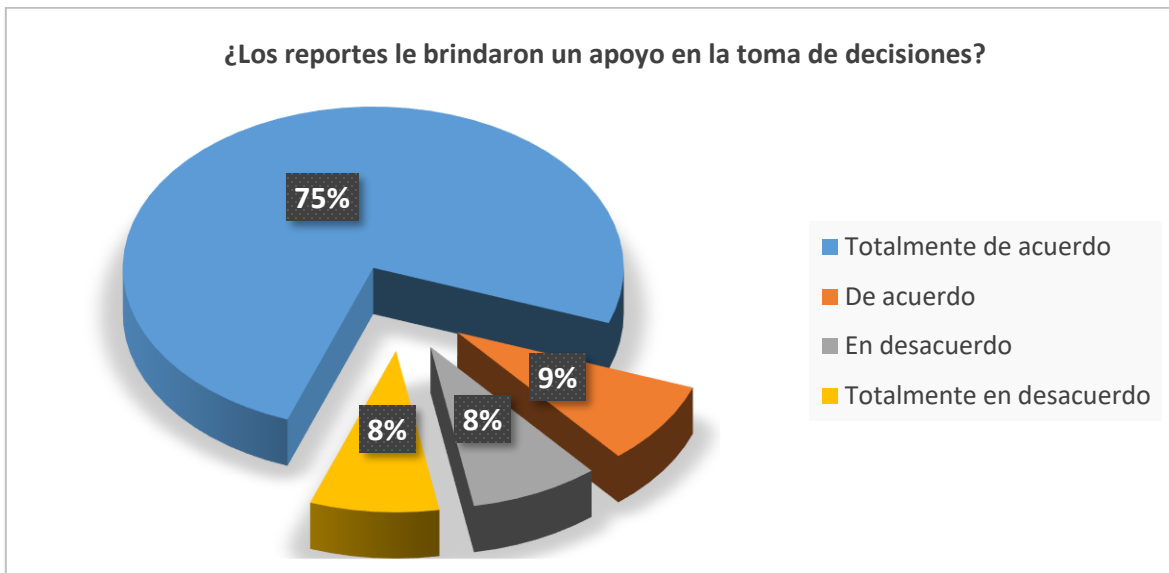


Ilustración 50: Resultados de la pregunta N° 3. Fuente: Prueba de aceptación con usuarios finales.

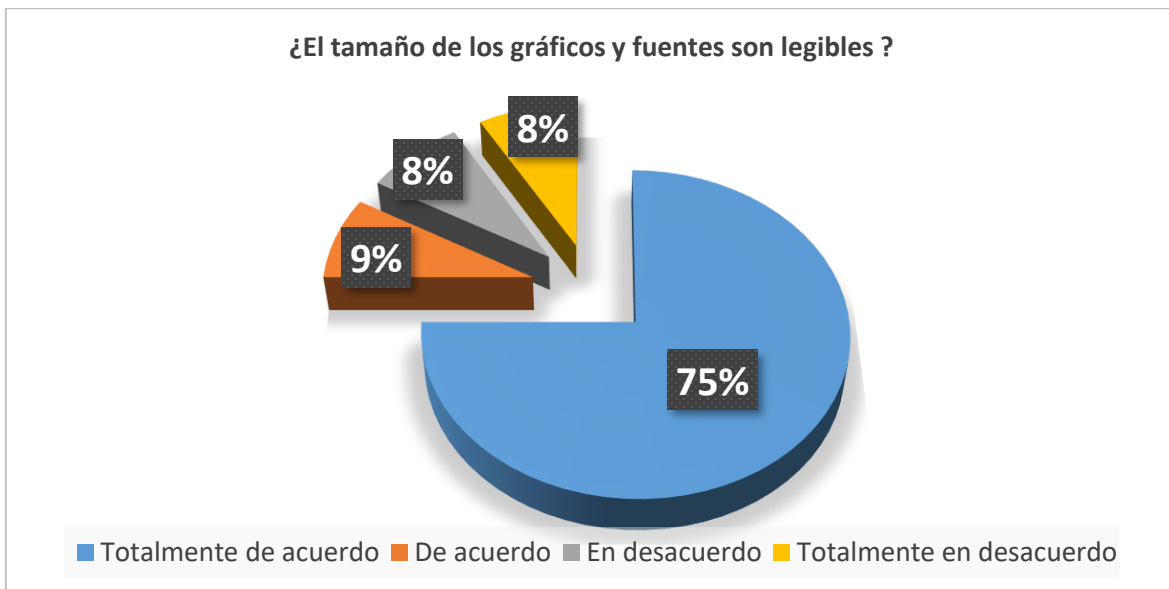


Ilustración 51: Resultados de la pregunta N° 4. Fuente: Prueba de aceptación con usuarios finales.

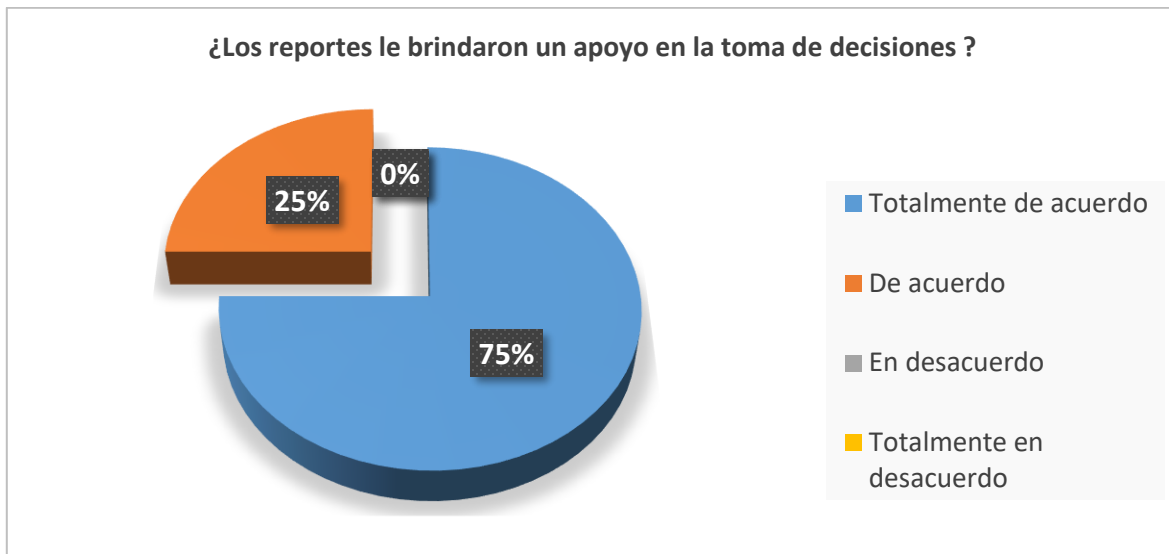


Ilustración 52: Resultados de la pregunta N° 5. Fuente: Prueba de aceptación con usuarios finales.

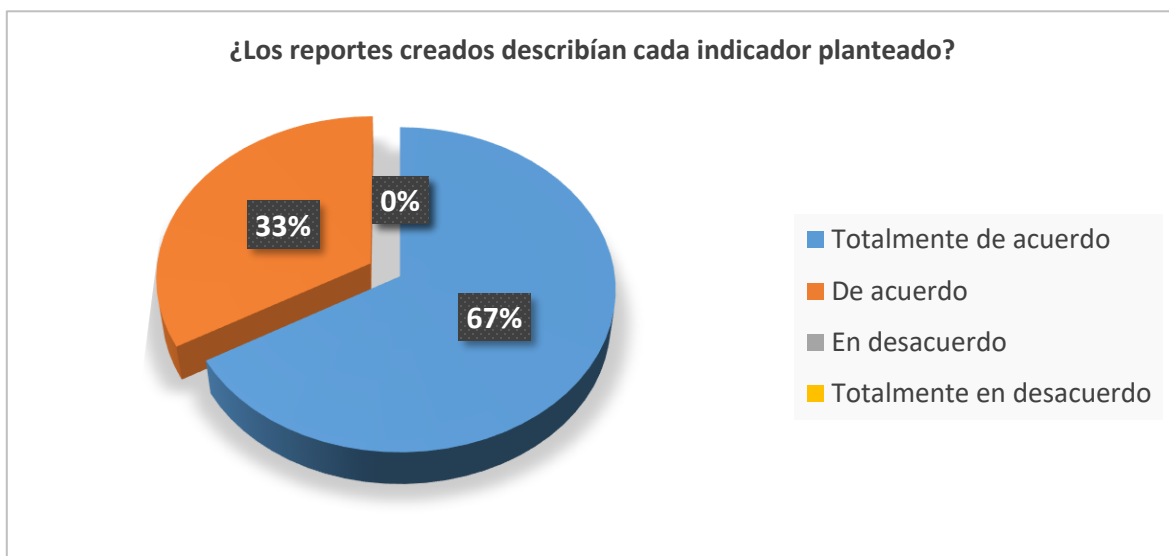


Ilustración 53: Resultados de la pregunta N° 6. Fuente: Prueba de aceptación con usuarios finales.

Según los resultados obtenidos a través de la encuesta realizada, se puede concluir que los usuarios perciben el sistema como fácil de usar, de navegación intuitiva, interfaces con un contraste adecuado, interfaces organizadas y consistentes, además de utilizar un lenguaje claro, conciso e informar claramente

acerca de los eventos que están ocurriendo en el sistema. Es importante destacar que en estas pruebas participaron tres (3) usuarios finales especializados en el proceso de visita clínica, cuatro (4) especialistas en Inteligencia de Negocio, tres (3) personas sin conocimientos en el área tecnológica, dos (2) especialistas en sistemas de información, por lo que se consideró que un total de doce (12) participantes suficientes para medir el grado de entendimiento de los reportes creados. Además, a través de la realización de estas pruebas los usuarios pudieron manifestar sus opiniones sobre otros aspectos del sistema, como son: el tamaño de letra, color de la letra en algunos títulos, el tamaño de los gráficos, la el desplazamiento entre pestañas y opinar acerca del posicionamiento de elementos de las interfaz. Estas opiniones permitieron ajustar más las interfaces de usuarios de acuerdo a las necesidades de los mismos.

CONCLUSIONES

El objetivo principal de este Trabajo Especial de Grado fue alcanzado exitosamente al lograr diseñar y desarrollar una solución de inteligencia de negocio orientada a la gestión de visita clínica con el fin de apoyar la toma de decisiones precisas y en corto tiempo. Utilizando la metodología Kimball fue posible ejecutar una serie de actividades que me guiaron a través del diseño y construcción de la solución de inteligencia de negocio, pudiendo avanzar rápida y ordenadamente durante el desarrollo, dando como resultado un producto usable y que satisface las necesidades de los diferentes usuarios definidos. A partir de los indicadores asociados a los visitantes clínicos es posible monitorear el estatus del recurso, el cual representa un factor crítico en el desarrollo de proceso, de esta manera es posible aumentar la eficiencia de las actividades necesarias para la continuidad del proceso de visita clínica.

Durante el desarrollo de la solución, se presentaron algunas limitantes al momento de construir la solución. Los requerimientos aportados por los usuarios finales no eran precisos, por tanto fue necesario varias reuniones que me permitiera capturar la información de forma concreta y precisa. A su vez, la fuente de datos fue proporcionada en forma documental, por medio de un esquema; en este sentido fue necesario construir la Base de Datos de Visita Clínica para posteriormente proceder a cargar la data y realizar el modelo dimensional.

A través de pruebas de aceptación con el usuario, realizadas durante el desarrollo de la solución, se obtuvieron diversos comentarios y sugerencias sobre la aplicación, así como también un gran interés por parte de los usuarios, lo cual generó motivación adicional para entregar un producto con altos estándares y que agregara valor al novedoso a la compañía aseguradora.

La finalidad del siguiente trabajo especial de grado, ayudara a los usuarios directores y administradores a crear estrategias, definir objetivos y métodos para cumplir el objetivo de desarrollar una solución de inteligencia de negocio, para la obtención de indicadores del proceso de visitas clínicas para las compañías aseguradoras.

RECOMENDACIONES

A pesar de que gracias a esta solución de inteligencia de negocio, se cuenta con una opción de poder conocer más a fondo los procesos de visita clínica en el sector de compañías aseguradoras, se recomienda que para trabajos futuros en este tipo de soluciones o para la mejora de esta misma lo siguiente:

- Implementar la solución para dispositivos móviles, así los usuarios de la solución podrán acceder a la información desde cualquier parte y podrán tomar decisiones soportadas por los resultados obtenidos de los indicadores, además de poder supervisar en todo momento el funcionamiento de sus procesos.
- Integrar otras áreas temáticas de la empresa como: recursos humanos, contabilidad, entre otros.
- Estudiar que otros indicadores de gestión se puedan agregar a la lista propuesta para brindarle más información a los usuarios de la solución y así estos puedan tomar mejores decisiones.
- Adiestrar a los usuarios para que aprendan a construir ellos mismos las consultas analíticas y así agilizar el proceso de los nuevos requerimientos en cuanto a indicadores de gestión.
- Incorporar los indicadores asociados a la encuesta de satisfacción del cliente.

REFERENCIAS BIBLIOGRÁFICAS Y DIGITALES

Cano, J. L. (2007). *Business Intelligence: Competir con Información*.

Conceptos., V. (s.f.). *Diccionario de la Real Academia Española*. Obtenido de <http://www.rae.es/rae.html/>.

Dunbar. (1998). *The Essential Guide to Oracle Warehousing*. Oracle Corporation.

DWH,2010 (2010). *Buigues, A. Arquitectura de un Data*. Obtenido de <http://anabuigues.com/2010/03/05/>

Gartner. (2014). Obtenido de http://www.gartner.com/technology/research/methodologies/research_mq.js.

Inmon, I. (1996). *Building the Operational Data Store*. New York: John Wiley and Sons.

Inmon, W. (2000). *Data Mart Does Not Equal Data Warehouse*. DMReview.com.

Inteligencia de Negocios. (2017). Obtenido de <http://www.idensa.com>.

Kamkolkar, N. &. (2016). *Tableau para la empresa: descripción general de TI*. Tableau software.

Kimball, R. &. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (2nd edition ed.)*. Wiley.

Kimball, R. &. (2008). *The Data Warehouse ETL Toolkit (2nd edition)*. New York: Wiley.

Laudon K., L. J. (2008). *Sistemas de Información Gerencial, 8va. Edición. Pearson Education.*

Loshin. (2003). *Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph.*

Pacheco J.P., C. W. (2002). *Indicadores Integrales de Gestión. McGraw-Hill Interamericana S.A. Bogotá, Colombia.*

Pentaho. (2017). Obtenido de <http://www.pentaho.com/>

Sheldon, T. (2001). *linktionary.* Obtenido de <http://www.linktionary.com/m/metadata.html>.

William, I. (2007). *The Father of Data Warehousing. San Diego: Inmon Consulting Services.*