Consultoría para el diseño del Data Warehouse de Fresenius Medical Care Colombia

Camilo Andrés Navarrete Arias, Deicy Paola Santos Morales, Germán David Gutiérrez Barreto, Juan Carlos Ortiz Barrios y Manuel Alberto Rojas Valero

Facultad de Ingeniería, Fundación Universitaria Compensar

Especialización en Big Data

Ingeniero Carlos Alberto Cañón Romero

18 de noviembre de 2022

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Resumen

Se dice que la información es el activo más valioso para las compañías, fundamental en la toma de decisiones para que las organizaciones sean competitivas, puedan anticiparse al futuro y sea el soporte de nuevas estrategias de negocio. Pero no sólo basta con captar información y más información, es necesario saber cómo y en donde se almacena para que sea un apoyo en lo que se conoce hoy como inteligencia de negocios.

El objetivo del presente trabajo es determinar la mejor forma en que la empresa Fresenius Medical Care Colombia, líder en la atención de pacientes con insuficiencia renal crónica, debe almacenar su información, de tal manera que esta facilite y sea un apoyo en la gestión empresarial. Llegaremos a resolver dicha inquietud, a través del análisis de la información de entrevistas realizadas a personas claves en el área de tecnología de Fresenius, para así sugerir la metodología para la Bodega de Datos (Data Warehouse) de Ralph Kimball, infraestructura y presentar un cronograma de implementación.

Palabras clave: Bodega de Datos, Almacén de Datos, Big Data, Inteligencia de Negocios, Modelo Dimensional de Datos.

Abstract

It´s said that the information is the most valuable asset for companies, essential in decision making for the organizations to be competitive, can anticipate the future and be the support of new business strategies. But it´s not enough just to capture information and more information, it's necessary to know where and how store it so that it be a support in what we know today as business intelligence.

The objective of this work is to determinate the way in wich the company Fresenius Medical Care Colombia, leader in the care of patients with cronical renal failure, should store its information so that it facilitates and support the business management. We Will resolve this concern, through the analysis of information from interviews with key people in the Fresenius technology area, in order to suggest the methodology for the Data Warehouse, Ralph Kimball, Infrastructure and present a implementation schedule.

Keyword: Datawarehouse, DWH, Datamart, Big Data, Business Intelligence (BI), Data dimensional modelling.

Tabla de Contenidos

1. Introducción 6
2. Antecedentes y Justificación 8
3. Formulación del Problema 9
4. Objetivos 10
4.1. Objetivo General 10
4.2. Objetivos Específicos 10
5. Alcances y Limitaciones 11
5.1. Alcance: 11
5.2. Limitaciones: 11
6. Marco Teórico 12
6.1. DatawareHouse o Bodega de datos. 12
6.1.1. Definición. 12
6.1.2. Características. 13
6.1.2.1. Orientados al tema. 13
6.1.2.2. Integración. 14
6.1.2.3. Información histórica y no volátil. 14
6.2. Datamart 14
6.2.1 Comparativo entre Data Warehouse y un Datamart 15
6.3. Metodologías 16
6.3.1. Metodología de Ralph Kimball 16
6.3.1.1. Definición 16
6.3.1.2. Características 17
6.3.1.3. Fases 18
6.3.2. Metodología de William Inmon 19
6.4. Arquitectura utilizada en un Data Warehouse. 23
6.4.1. Procesamiento Analítico en Línea OLAP (On-Line Analytical Processing) 23
6.4.2. Análisis del Procesamiento de Transacciones en Línea OLTP (On-Line Transactional Processing) 24
6.5. El papel de la inteligencia de Negocio cuando se crea un Data Warehouse. 24
7. Diseño Metodológico 27
8. Costos de Implementación Data Warehouse 28
9. Cronograma de Implementación 30
10. Resultados 31
11. Conclusiones y recomendaciones 33

11.1. Conclusiones 33
11.2. Recomendaciones 34
12. Anexos 35
- Anexo 1 35
- Anexo 2 35
- Anexo 3 35
Referencias 36

1. Introducción

Los países y las empresas buscan día a día ser más competitivos, tener mejores procesos, anticiparse a errores que puedan conllevar a una crisis en sus operaciones, minimizar costos, aumentar sus ganancias o ser mejor que la competencia y para ello definitivamente es importante tomar decisiones basadas en la información y no en supuestos. Hoy cuando tenemos datos y muchos más datos debido al auge de la tecnología y los sistemas de información algunas de las preguntas son: ¿cómo manejar esa información?, ¿dónde y cómo la almacenamos?, ¿y si se pierde la información?

La misma tecnología ha dado respuesta a estos interrogantes y desde finales de los años 80 se origina el concepto de Bodega de Datos (Data Warehouse en inglés), el cual es un repositorio unificado de los datos que los diferentes sistemas de información de las empresas van recolectando (PoweData, 2022). Al tener los datos centralizados se agiliza el acceso a estos, la integridad de la información y se facilitan los procesos de analítica de información.

El presente documento resume el trabajo realizado en la empresa Fresenius Medical Care Colombia, firma líder en la atención de pacientes con insuficiencia renal crónica. Allí, luego del análisis de información a las entrevistas realizadas al Gerente de Sistemas y el Coordinador de Aplicaciones y Datos, se observará la necesidad o no de la compañía de contar con una bodega de datos, cuál sería la metodología e infraestructura para la construcción de esta.

Será para al lector una aproximación clara y sencilla al concepto de diseño e implementación de una Bodega de Datos. Le invitamos a continuar con este lindo trabajo que

apoya nuestro proceso de formación como estudiantes de la especialización en Big Data y brinda una solución real a una empresa líder de nuestro país.

2. Antecedentes y Justificación

Fresenius Medical Care Colombia (2022) es una empresa de origen alemán fundada en el año 1996, como resultado de la fusión de Fresenius Worldwide Dialysis y National Medical Care, dos compañías del área médica y farmacéutica cuya tradición se extiende hasta el siglo XV. Fresenius se especializa en la producción y desarrollo de suministros médicos, haciendo especial enfoque en productos y servicios para pacientes con insuficiencia renal crónica; en la actualidad atienden a cerca de 345.000 pacientes en todo el mundo en más de 4.000 clínicas de diálisis.

En 1996 llega al país bajo el nombre de Fresenius medical Care Colombia SA y desde entonces y debido a las necesidades propias del sector, Fresenius medical Care se ha caracterizado por su liderazgo e innovación en el mercado, por lo que constantemente invierte en investigación y desarrollo.

Siendo el liderazgo y la innovación dos de los principales pilares de la compañía y a pesar de que actualmente cuenta con varios sistemas de información, se hace indispensable continuar avanzando para estar al día con las nuevas tecnologías, en especial aquellas que le permitan hacer una mejor gestión de sus activos como es la información. Es pertinente entonces para Fresenius Medical Care Colombia, avanzar hacia la implementación de una bodega de datos centralizados (Data Warehouse) que facilite los procesos de análisis de datos para modelos estadísticos e inteligencia de negocios. (Fresenius Medical Care Colombia, 2022)

3. Formulación del Problema

En el mundo actual, sin importar el sector empresarial o el área de negocio, se están generando y capturando grandes volúmenes de datos, por esto se hace cada vez más importante que las empresas cuenten con la posibilidad de hacer un buen proceso de inteligencia de negocios (Business Intelligence). Para ello es indispensable contar con un lugar donde la información repose (Bodega de Datos - Data Warehouse) y este se encuentre estructurado de la manera adecuada para que la organización pueda acceder de manera oportuna y rápida a la información y esta sea útil para el análisis, toma de decisiones y un apoyo real a la consecución de los objetivos empresariales.

Ahora bien, la Empresa Fresenius Medical Care Colombia, no solo genera grandes volúmenes de datos, si no que sus datos son de pacientes y afecta a los mismos, por lo que no es ajena a la necesidad de poder hacer un buen análisis de los datos y tomar decisiones oportunas. Resulta entonces significativo para la organización saber ¿cuál sería la mejor alternativa para centralizar sus datos?

4. Objetivos

4.1. Objetivo General

Analizar el proceso de tratamiento actual de la información con el fin de proponer el diseño de la bodega de datos (Data Warehouse) para Fresenius Medical Care Colombia, lo que permitirá centralizar los datos para la toma de decisiones en la organización.

4.2. Objetivos Específicos

1) Diagnosticar la situación actual de los sistemas de información transaccionales de la empresa.

2) Analizar y definir una metodología para la implementación de la bodega de datos (Data Warehouse).

3) Definir la infraestructura para la implementación de la bodega de datos (Data Warehouse).

5. Alcances y Limitaciones

5.1. Alcance:

Para la empresa Fresenius Medical Care Colombia con sede en Bogotá (Colombia), se presenta un proyecto de consultoría en el diseño de la bodega de datos (DatawareHouse) bajo la metodología de Ralph Kimball. Se implementarán Data Marts para cada una de las siguientes áreas de negocio: Financiera y Administrativa, Care Enablement, Care Delivery, Dirección Médica, logística y distribución.

Los principales sistemas de información para tener en cuenta para la centralización de los datos en una bodega de datos (DatawareHouse) son: SAP – ERP, Sistema Central de Salud – Europe Clinical Database - EUCLID y algunos sistemas intermedios que la empresa considere son fuente de información.

5.2. Limitaciones:

Limitaciones encontradas durante el análisis de información:

- Información general no centralizada.
- Cada área maneja su propia información.
- La información no se cuenta con data histórica ordenada.
- No se tiene un proceso claro para la elaboración de reportes a nivel táctico y estratégico.
- Políticas globales con el tratamiento y la seguridad de la información.

6. Marco Teórico

En el mundo actual las organizaciones tienen la necesidad de empezar a recopilar sus datos de modo que les permita generar valor a través del análisis de información histórica, con el fin de lograr una ventaja competitiva en el mercado; Así mismo, debido a las múltiples y diferentes sistemas de información empleados los datos en muchas ocasiones no se encuentran articulados, centralizados y disponibles para una adecuada gestión de la información. Es aquí donde el concepto de Bodega de Datos (Data Warehouse) cobra relevancia.

6.1. DatawareHouse o Bodega de datos.

6.1.1. Definición.

Una de las definiciones clásicas es la siguiente “Un Data Warehouse, es una colección de datos orientados al tema, integrados, no volátiles e históricos, cuyo objetivo es el de servir de apoyo en el proceso de toma de decisiones gerenciales” (Bernal, 2002).

La finalidad de los Data Warehouse o Almacenes de datos consisten en centralizar los datos, integrarlos y organizarlos de modo que nos permita ofrecer un apoyo en los procesos de toma de decisiones de las empresas y de este mismo modo hace parte central para un sistema de inteligencia de negocios.

Otra forma de comprender el concepto de un Data Warehouse es un lugar centralizado de todos los aplicativos existentes que permite una distribución óptima para la toma de decisiones Gerenciales.

6.1.2. Características.

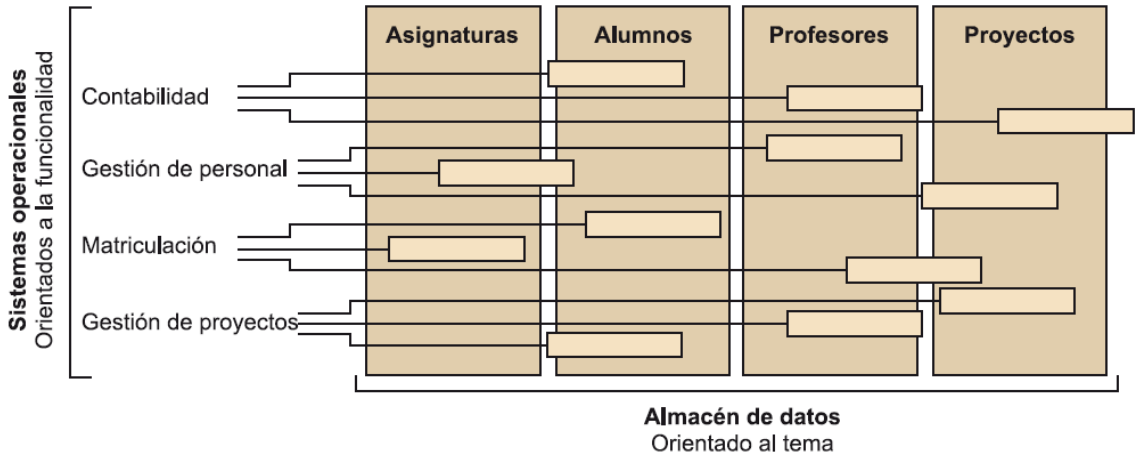
6.1.2.1. Orientados al tema.

Hace referencia a las directrices de los diseñadores de los DWH. Los sistemas operacionales se diseñan por un conjunto de requerimientos que buscan satisfacer una necesidad concreta y conocida, por lo cual, se habla de orientación a la funcionalidad.

Como difícilmente se logrará llegar al detalle de la necesidad de información por parte del cliente final, al momento de crear el data warehouse, la información no se estructura según su funcionalidad o para lo que será utilizada, sino que se divide por áreas o temas principales de interés de la organización, permitiendo a los responsables disponer de los datos que consideren oportunos para el análisis y definición de estrategias. (Rius , Serra , & Curto)

Figura 1

Sistemas operacionales, Orientados por temas



Nota. La figura anterior nos muestra como cada sistema operacional en este caso, de una universidad accede exactamente a los datos que le corresponde, ejemplo Contabilidad acceder a datos tanto de alumnos, profesores y proyectos, pero no accederá a todos. Tomado de *Introducción al almacenamiento de datos* (Rius , Serra , & Curto , pág. 12)

6.1.2.2. Integración.

Integrar todos los sistemas de información a pesar de que el origen de los datos pertenezca a fuentes diferentes, con el fin de que el cliente final lo vea como si proviniera de una única fuente, hay que tener presente que las consultas no serán realizadas por informáticos sino usuarios que desconocen del tema a los cuales se les debe facilitar el trabajo. Este proceso no es simplemente unificar la información en un solo lugar, estos datos deben ser integrados y previamente transformados. (Rius , Serra , & Curto)

6.1.2.3. Información histórica y no volátil.

Estas son las dos últimas características de un DWH: La historicidad permite analizar la evolución, los datos no son actualizados, sino que se almacena el histórico de cada dato; por el contrario, la no volatilidad nos muestra que los datos almacenados no sufren ninguna actualización convirtiéndose en información de solo lectura y se mantiene para consultas futuras. (Rius , Serra , & Curto)

6.2. Datamart

Podemos definir a un datamart (DM) como la versión más reducida de un Data Warehouse, los DM por lo general contienen información específica de un área de la organización permitiendo de este modo responder a la necesidad del negocio más específicamente. (Rodríguez, 2010)

Según Orellana (2013) Los datamart son entonces bodegas de datos con información específica de un área de la organización, se caracterizan por disponer de una estructura optima de datos, que podrá ser consultados para la elaboración de informes y la toma de decisiones. (pág. 25)

Un DataMart posee los siguientes componentes:

- Fuentes de datos
- Procesos de extracción, transformación y carga de los datos (ETL)
- Data Warehouse
- Herramientas de exploración

6.2.1 Comparativo entre Data Warehouse y un Datamart

Tabla 1

Comparativo entre Data Warehouse y un Data mart

	Data Warehouse	**Datamart**
Alcance.	Construido para satisfacer las necesidades de información de toda la organización.	Busca satisfacer las necesidades de un determinado grupo de usuarios.
Objetivo.	Proporcionar un entorno integrado y una imagen coherente de la empresa en un momento determinado.	Utilizado principalmente en una división de negocios a nivel de departamento.
Características.	Administra grandes cantidades de datos históricos a nivel atómico.	Administra resúmenes o datos totalizados.
Pertenencia.	Pertenece a toda la organización.	Pertenece al área de negocio al cual está orientado.
Uso.	Ayuda a tomar una decisión estratégica.	Ayuda a tomar decisiones tácticas para el negocio.

Nota. (ABC del data mart, 2021)

6.3. Metodologías

Antes de implementar un Data Warehouse este debe ser diseñado utilizando diferentes métodos o modelos, estos cuentan con diferentes características e instrucciones a seguir, todo con el fin de entregar un modelo lo más cercano posible y entendible a nuestro cliente final. La información contenida debe ser de calidad y eficiente para que la organización pueda hacer uso de esta de manera óptima a través de diferentes aplicaciones operacionales y que este se convierta en una herramienta que apoye la toma de decisiones.

Actualmente contamos con diferentes metodologías que nos permite definir el paso a paso para el diseño y construcción de un Data Warehouse, aunque cada una de estas proporcionen mecanismos de solución a problemáticas reales en el ámbito de su aplicación, cada autor nos presenta un enfoque de manera diferente, la principal diferencia se precisa en la forma de la construcción del Data Warehouse. (Forero & Sánchez, 2021, pág. 7)

6.3.1. Metodología de Ralph Kimball

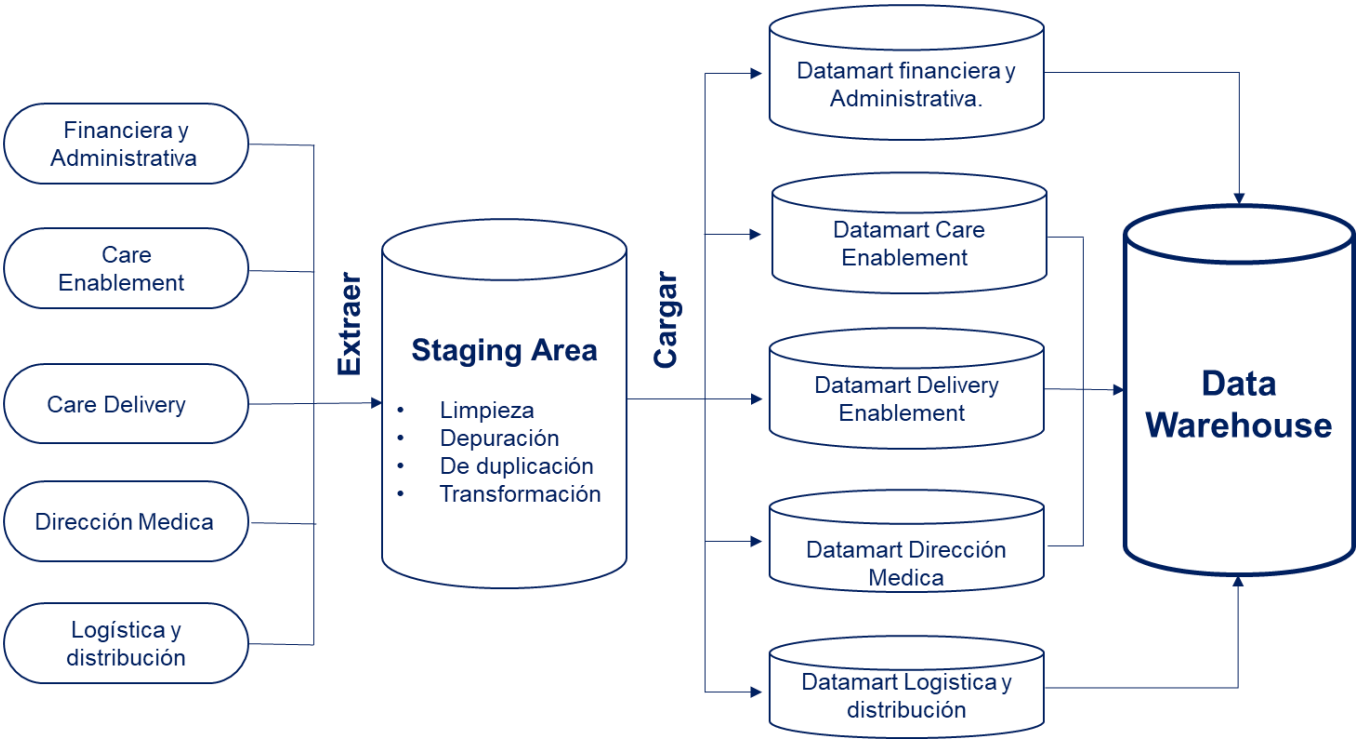
6.3.1.1. Definición

La metodología de Kimball también conocida como modelo dimensional, se basa en el denominado ciclo de vida dimensional del negocio, esta una de las metodologías más usadas al momento de diseñar una bodega de datos, se constituye en un modelo de dimensiones y tablas de hechos con la finalidad de apoyar la toma de decisiones.

El modelo dimensional es una técnica de diseño lógico que busca presentar datos dentro de un marco de trabajo estandarizado e intuitivo, que permita su acceso con un alto rendimiento. (Forero & Sánchez, 2021)

Figura 2

Metodología Multidimensional de Ralph Kimball



Nota. Adaptado de *Metodologías de Data Warehouse* [Imagen], https://gravitar.biz/datawarehouse/metodologias-data-warehouse/).

6.3.1.2. Características

- **Centrado en el negocio.** Conocer a detalle la gestión del negocio para poder identificar las necesidades y/o los diferentes requerimientos detallados permitiéndonos desarrollar e implementar un modelo optimo y satisfactorio.
- **Construir una infraestructura de información adecuada.** Esto implica analizar a detalle la información y diseñar una base única, integrada, fácil de usar y con buen rendimiento donde podamos ver reflejado los requerimientos identificados en la empresa.
- **Realizar entregas en incrementos significativos.** De acuerdo con la metodología las entregas deberán realizarse en los tiempos

- acordados previamente con la organización.
- **Ofrecer una solución completa.** Esto indica presentar la bodega de datos o DatawareHouse funcional, cumpliendo con las expectativas de la organización. A su vez se deberá proporcionar herramientas de consulta, aplicaciones para informes y análisis, capacitación, soportes y documentación.

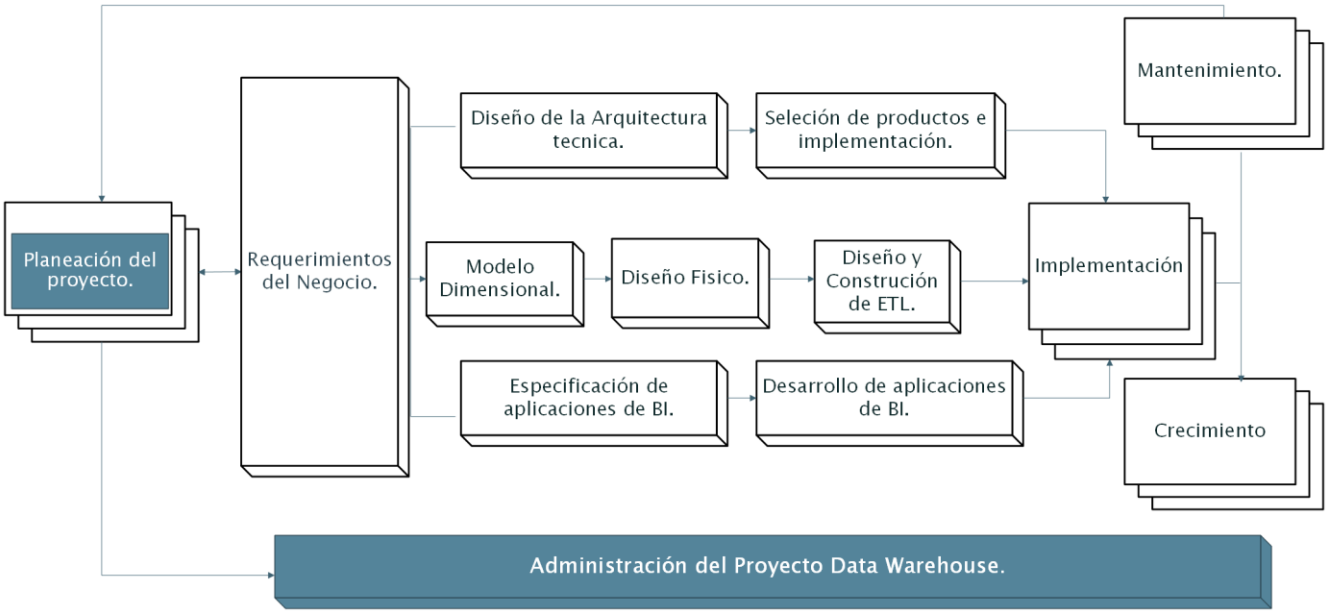
Esta metodología facilita la creación de un DWH ya que de acuerdo con los requerimientos de las diferentes áreas de la organización permite el diseño y la construcción de los Datamart específicos, que darán forma al futuro modelo general. (Forero & Sánchez, 2021, pág. 9)

6.3.1.3. Fases

Para garantizar la calidad en el desarrollo del Data Warehouse las fases establecidas por Ralph Kimball, han sido elaboradas para que se pueda desarrollar en modo paralelo o secuencial.

Figura 3

Fases metodología Ralph Kimball



Nota. Adoptada de (Forero & Sánchez, 2021)

La metodología Kimball propone tres líneas base que unidas entre si conforman el ciclo de vida de esta, Las cuales con las siguientes:

- **Tecnología:** En este punto es importante definir el diseño, arquitectura y herramientas informáticas que se utilizaran durante la implementación del modelo previamente seleccionado.
- **Datos:** Se deberán llevar a cabo los modelos dimensionales y los modelos de transformación de datos.
- **Inteligencia de Negocio:** Esta fase se deberá contar con las aplicaciones elegidas con el fin de poder mostrar los resultados planteados a los usuarios de negocio. (Forero & Sánchez, 2021, pág. 10)

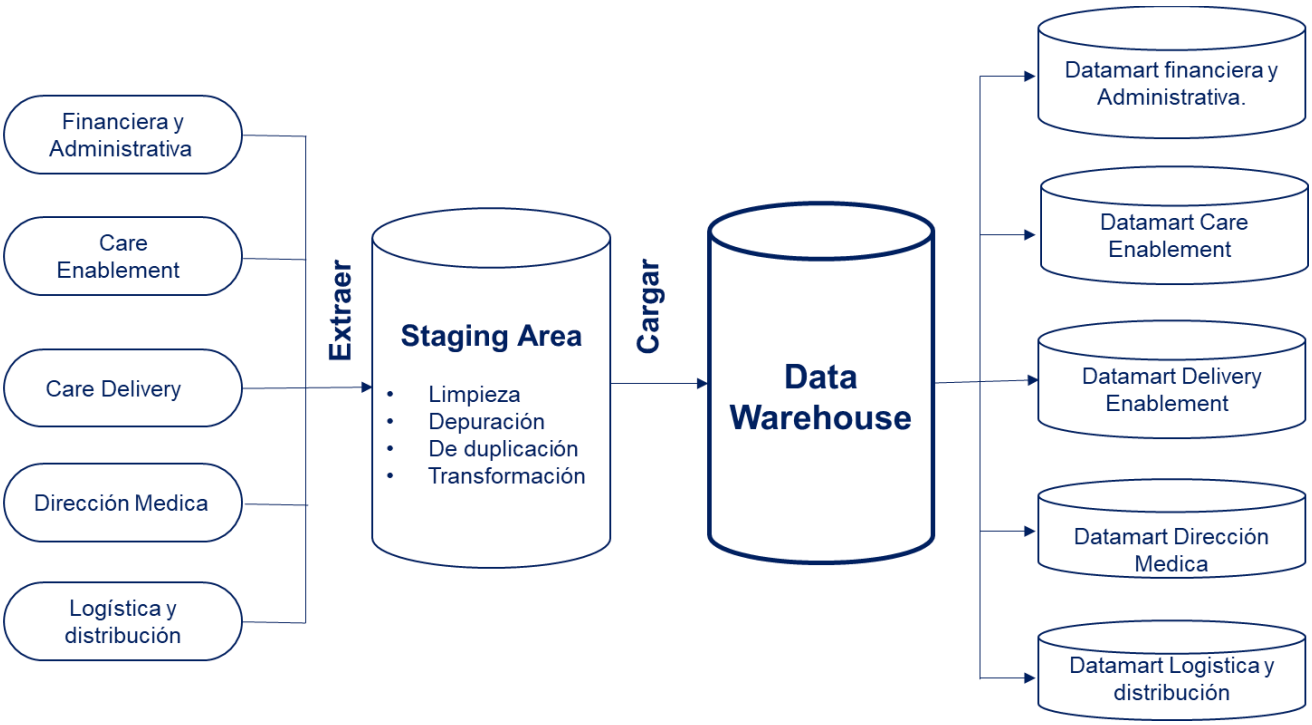
6.3.2. Metodología de William Inmon

En este modelo Bill Inmon, considerado el padre de las bodegas de datos, propuso un enfoque de implementación descendente, en el que se adaptan herramientas de bases de datos relacionales tradicionales al desarrollo de un Data Warehouse corporativo, en el cual, desde la construcción de pequeños Data Marts se busca orientar y apoyar los procesos de toma de decisiones en los diferentes departamentos o áreas del negocio.

Inmon, argumenta que es importante realizar un esfuerzo inicial para la construcción de un Data Warehouse global debido a que este permite crear cualquier cantidad de bases de datos departamentales sin correr el riesgo de que se presenten incompatibilidades entre los datos. (Flórez & Del Mar, 2015)

Figura 4

Metodología Relacional Bill Inmon



Nota. Adaptado de *Metodologías de Data Warehouse* [Imagen], https://gravitar.biz/datawarehouse/metodologias-data-warehouse/).

De acuerdo con lo anterior esta metodología propone un modelo de datos compuesto de tres niveles: nivel superior, también llamado modelo entidad relación (ERD, por sus siglas en inglés Entity Relationship Diagram); nivel intermedio, conocido como conjunto de elementos de datos (DIS, por sus siglas en ingles Data Item Set) y finalmente el nivel inferior, denominado modelo físico. (Flórez & Del Mar, 2015)

Tabla 2

Diferencias entre metodologías

	Multidimensional / *Ralph Kimball*	**Relacional / *Bill Inmon***
Diseño del data warehouse	Esfuerzo Menor / Requiere de menor tiempo.	Esfuerzo Alto / Requiere de una gran cantidad tiempo.
Mantenimiento	Medio-Alto	Simple
Inversión	Baja / Bajo costo inicial y para fases posteriores.	Alta / Alto costo inicial, pero menor para fases posteriores.
Tiempo/Plazo	Menor tiempo para la configuración inicial.	Mayor tiempo para su inicio.
Elementos	Data Mart – Data Warehouse.	Data Warehouse – Data Mart.
Nivel de Especialización Requerida	Media-Baja / No requiere un alto grado de especialización.	Alta / Requiere de un grado elevado de especialización.
Requisitos De Integración De Datos	Áreas individuales de negocio.	Empresarial / Amplia.
Flexibilidad	Menor flexibilidad.	Mayor flexibilidad.

Nota. Tomado de Metodologías de Data Warehouse (Gravitar, s.f.)

6.4. Arquitectura utilizada en un Data Warehouse.

Como requisito en la arquitectura de un data Warehouse es que sus usuarios puedan depender de precisión, prontitud y facilidad en el acceso con la cual se genera la información consultada de acuerdo con las necesidades particulares de cada área o usuario. La información contenida debe ser simple y comprensible y no debe variar en relación con otra operación o con otro usuario dentro de la organización. (Pérez, 2015)

6.4.1. Procesamiento Analítico en Línea OLAP (On-Line Analytical Processing)

Se entiende como proceso analítico en línea, método para organizar y consultar datos sobre una estructura multidimensional, comparando este proceso con las bases de datos relacionales, todas las posibles consultas están calculadas con antelación a su ejecución, lo que hace que sea más ágil y flexible al usuario.

Una herramienta OLAP se conforma por un motor y un visor, el motor es el concepto que definimos anteriormente mientras que el visor es una interfaz que nos permite consultar, manipular, reordenar y filtrar los datos existentes en una estructura OLAP de tal modo que le permita al usuario de negocio resolver dudas que serían mucho más complejas de solucionar por medio de consultas de SQL. (Curto & Conesa)

Con el concepto definido anteriormente frente a los sistemas OLAP podemos inferir que es una herramienta que nos permite trabajar en la transformación de los datos para facilitar posteriormente al usuario final las consultas de información según su necesidad, sin tener que requerir a un personal del área tecnológica. (Reinoso & Ariza, 2014)

6.4.2. Análisis del Procesamiento de Transacciones en Línea OLTP (On-Line Transactional Processing)

De acuerdo con la interpretación dada por el Ing. Bernabéu (2009) “un OLTP representa toda aquella información transaccional que genera la empresa en su accionar diario, además, de las fuentes externas con las que puede llegar a disponer” (pág. 22)

También se relaciona los OLTP más comunes como lo son:

- Archivos de texto.
- Hipertextos.
- Hojas de cálculos.
- Informes semanales.
- Bases de datos transaccionales.

Para el proceso de extracción de datos desde los OLTP, para poder hacer uso de esta información y que estos puedan ser agregados en el DW, se debe hacer uso de un sistema que se encargue de este proceso, como los ETL, que extraen los datos de las diversas fuentes existentes, los transforman y por último después de la depuración se procede con la carga en la bodega de datos. (Ing. Bernabeu, 2009, pág. 23)

6.5. El papel de la inteligencia de Negocio cuando se crea un Data Warehouse.

De acuerdo con la definición dada por el Ing. Bernabeu (2009) la inteligencia de negocios “Integra por un lado el almacenamiento y por el otro el procesamiento de grandes cantidades de datos, con el principal objetivo de transformarlos en conocimiento y en decisiones en tiempo real, a través de un sencillo análisis y exploración” (pág. 6)

La inteligencia de negocios, continuando con la interpretación dada por Ing. Bernabeu (2009), hace hincapie en los procesos de recoleción y utilización correcta de la información con

el fin de promover una mejora en la forma de operar de la organización, brindando a sus usuarios el acceso a la información clave que necesiten en cada area para poder tomar decisiones oportunas basadas en datos correctos y eficaces. Contar con la información exacta y en el menor tiempo, nos permite identificar y actuar de manera preventiva a las situaciones que puedan presentarse evitando que se conviertan en un problema. Cuanto mas relevante y útil sea la inteligencia con la cual la empresa gestiona su informaciónn sobre su gestión, clientes, provedores, socios, operaciones, etc, mayor sera la ventaja competitiva y la contribución en la toma de decisiones (pág. 6).

En la actualidad toda organización genera datos como producto de sus actividades diarias y estos mismos son almacenados y administrados a través de diferentes sistemas transaccionales. Lo que se busca al centralizar los datos por medio de un Data Warehouse, es que estos dejen de ser simples datos, para transformarlos en información que apalanque las decisiones de la organización y el trabajo de la inteligencia de negocios, la cual nos permite que el proceso de toma de decisiones este basado sobre un amplio conocimiento de sí mismo y del entorno, reduciendo el riesgo y la incertidumbre. Además, contribuye a que las organizaciones puedan traducir sus objetivos en indicadores, que puedan ser analizados desde diferentes perspectivas, con la finalidad de encontrar información que no solo se encargue de responder a preguntas de lo que está sucediendo, lo que ya sucedió, sino también predecir escenarios futuros y las acciones a tomar para el cumplimiento de los objetivos deseados. (Ing. Bernabeu, 2009, pág. 5)

6.6. Beneficios de un Data Warehouse en la nube.

- Volver datos operativos en información analítica.
- Facilidad en la generación de reportes dinámicos para el análisis en los cuales la empresa soporta su estrategia.

- Contribuye a la mejora continua de la organización.
- Pago por uso (pay as you go).
- Escalabilidad rápida (incrementa y reduce recursos basados en demanda)
- Ahorro en gastos operacionales como arriendo, personal de seguridad, aire acondicionado, infraestructura civil, cámaras de seguridad, aseo y mantenimiento, servicios públicos.
- Ahorro en Infraestructura tecnológica física: servidores, almacenamiento, redes, routers, switches, backup y archivo.
- Optimización de tiempo en el plan de recuperación y desastres.

7. Diseño Metodológico

Debido a la naturaleza del proyecto y después de realizar una revisión sobre los tipos de investigación a usar, se optó por el modelo cuantitativo descriptivo, el cual se apoya en la recolección de datos que puedan ser procesables y a partir de su análisis se pueda apoyar la toma de decisiones.

El proceso de investigación se llevará a cabo en la compañía Fresenius Medical Care Colombia, ubicada en la ciudad de Bogotá. Para la recolección de datos se realizará una entrevista al gerente de sistemas y proyectos IT para Colombia y con el Coordinador de Aplicaciones y Datos IT, con el fin de determinar las necesidades de la empresa respecto a la centralización de los datos, así como el nivel de madurez de la plataforma tecnológica.

Una vez obtenida la información, se procederá con el análisis de esta y en base a las necesidades detectadas se presentará la alternativa para la bodega de datos.

8. Costos de Implementación Data Warehouse

8.1. Servicio de Azure Synapse Analytics (Data Warehouse)

Tabla 3

Calculadora de servicios Azure

Categoría de Servicio	Bases de datos
Tipo Servicio	Azure Synapse Analytics
Región	East US
Descripción Servicio	Nivel: Optimizado para Compute Gen2, grupos de SQL dedicados: DWU 100, 1 año de reserva, 1 TB de almacenamiento con recuperación ante desastres con redundancia geográfica; Región Este de EE. UU., 100 GB de datos recopilados al día, 7 días de caché de acceso frecuente, 30 días de retención total, Compresión de datos estimada en 7 veces, 1 Mes de 2 x Extrapequeña (2 núcleos virtuales) instancias de motor, 1 Mes de 2 x 1 núcleo virtual instancias de administración de datos
TRM (aprox.)	$ 4.850,00
Grupos de SQL dedicados: DWU 100	Costo Mensual Estimado
USD	$ 694,42
COP	$ 3.367.937,00
Una Tera de Almacenamiento	Costo Mensual Estimado
USD	$ 23,00
COP	$ 111.550,00
Recuperación ante desastres con redundancia geográfica	Costo Mensual Estimado
USD	$ 57,55
COP	$ 279.117,50
Núcleos virtuales del motor	Costo Mensual Estimado
USD	$ 639,48
COP	$ 3.101.478,00
Núcleos virtuales de administrador de datos	Costo Mensual Estimado

USD	$ 159,87
COP	$ 775.369,50
Almacenamiento	Costo Mensual Estimado Mensual
USD	$ 27,54
COP	$ 133.569,00
Total costo mensual USD	**$ 1.601,86**
Total costo mensual COP	**$ 7.769.021,00**
TOTAL COSTO ANUAL USD	**$ 19.222,32**
TOTAL COSTO ANUAL COP	**$ 93.228.252,00**

Nota. Tomado del portal Azure https://azure.microsoft.com/en-au/pricing/calculator/

8.2. Costos Profesionales:

Consultoría	$ 120.000.000,00
Ingenieros de Datos	$ 230.000.000,00

9. Cronograma de Implementación

Aproximadamente 290 días de esfuerzo para la implementación del Data Warehouse:

Figura 5

Cronograma de Implementación del Data Warehouse para Fresenius Medical center.

Proyecto Data WareHouse	289,38 días	lun 16/1/23	jue 23/5/24
Proyecto 1 (Data Mart Financiera y Logistica)	101 días	lun 16/1/23	jue 6/7/23
Fase 1 (Análisis)	15 días	lun 16/1/23	mié 8/2/23
Fase 2 (Diseño)	20 días	mié 8/2/23	vie 10/3/23
Fase 3 (Implementación)	37 días	lun 13/3/23	mar 16/5/23
Fase 4 (Realizar Pruebas)	17 días	mar 16/5/23	mié 14/6/23
Fase 5 (Paso a Producción)	2 días	mié 14/6/23	vie 16/6/23
Fase 6 (Gestion del Cambio)	10 días	mar 20/6/23	jue 6/7/23
Proyecto 2 (Data Mart Care Enablement y Care Delivery)	94,38 días	jue 6/7/23	jue 14/12/23
Fase 1 (Análisis)	15 días	jue 6/7/23	mar 1/8/23
Fase 2 (Diseño)	20 días	mar 1/8/23	mar 5/9/23
Fase 3 (Implementación)	30,38 días	mar 5/9/23	mié 25/10/23
Fase 4 (Realizar Pruebas)	17 días	mié 25/10/2:	jue 23/11/23
Fase 5 (Paso a Producción)	2 días	jue 23/11/23	lun 27/11/23
Fase 6 (Gestion del Cambio)	10 días	lun 27/11/23	jue 14/12/23
Proyecto 3 (Data Mart Direccion Médica)	94 días	jue 14/12/23	jue 23/5/24
Fase 1 (Análisis)	15 días	jue 14/12/23	jue 11/1/24
Fase 2 (Diseño)	20 días	jue 11/1/24	mar 13/2/24
Fase 3 (Implementación)	30 días	mar 13/2/24	mié 3/4/24
Fase 4 (Realizar Pruebas)	17 días	mié 3/4/24	jue 2/5/24
Fase 5 (Paso a Producción)	2 días	vie 3/5/24	mar 7/5/24
Fase 6 (Gestion del Cambio)	10 días	mar 7/5/24	jue 23/5/24

10. Resultados

Al analizar la información recolectada en las entrevistas con el Gerente de Tecnología Colombia y el Coordinador de Aplicaciones y datos, se encontró que:

- La Empresa posee dos principales sistemas de información que apoyan la gestión empresarial SAP y EUCLID y algunos otros que complementan alguna funcionalidad.
- En el manejo de la información de usuario final, de los sistemas transaccionales se extraen reportes ya diseñados los cuales no se usan por completo, ya que para que esta información sea valiosa para la empresa, los usuarios realizan manipulación de dichos reportes y como consecuencia de esto la información se puede adulterar.
- Los usuarios en general manifiestan que la información generada por los sistemas actuales de información, no cumplen con las necesidades locales, de reglamentación o normativas, como consecuencia deben hacer mucho trabajo manual adicional.
- La compañía requiere herramientas que le permita tomar decisiones sustentadas en el valor que tiene el dato y que dicha información ofrezca la orientación necesaria desde el punto de vista estratégico.
- A Fresenius Medical Care Colombia le duele no poder llegar a hacer una predicción, ser más preventivos y no solo reactivos, se termina analizando lo que ya pasó, pero no se puede hacer una orientación basada en los datos hacia el futuro.

- Aunque Fresenius cuenta con una infraestructura tecnológica adecuada y repotenciada, desea alinearse con las políticas globales de casa matriz (Alemania) y por eso requiere la centralización de datos (Bodega de Datos) y esta sea implementada en la nube.
- Fresenius posee varias áreas de gestión, pero las susceptibles para llevar su información a la bodega de datos son 5: Financiera y Administrativa, Care Enablement, Care Delivery, Dirección Médica, Logística y distribución.

11. Conclusiones y recomendaciones

11.1. Conclusiones

Debido a las necesidades de centralización de la información, es pertinente para Fresenius Medical Care Colombia cuente con una bodega de Datos (Data Warehouse)

Una vez expuestas las metodologías con el gerente para el diseño e implementación de la bodega de Datos, esta se llevará a cabo bajo la metodología de Ralph Kimball, es la que más se ajusta a las necesidades de la empresa, ya que permite un diseño de esfuerzo menor, una inversión baja, un menor tiempo de configuración inicial y se implementan por áreas individuales de negocio, trabajo que se puede hacer paralelamente.

La infraestructura de acuerdo con el análisis y deseo de la gerencia de Fresenius Medical Care Colombia será en la nube Microsoft Azure por directriz de Casa Matriz. Ventajas: inversión mínima de infraestructura para la implementación del DWH respecto a Onpremise, si durante la implementación del DWH los recursos en cuanto a almacenamiento y procesamiento se están quedando cortos se pueden aumentar con mínimos tiempos de espera para su despliegue y su costo se encuentra inmerso dentro del valor mensual del servicio, si la empresa así lo considera se puede configurar redundancia del servicio en otra(s) zona(s) y esto permite que el DWH se encuentre en espejo y en determinado momento se pueda pasar de una zona a otra en segundos y no se verá afectado el servicio.

Se centralizan los datos de las fuentes principales de información de Fresenius Medical Care (SAP – Euclid) y de los sistemas intermedios que la empresa considere.

Se llevará a cabo la centralización de la información a 5 áreas de la compañía: Financiera y Administrativa, Care Enablement, Care Delivery, Dirección Médica, logística y

distribución. Esto no es un impedimento para que sea escalable en el futuro e incluir más áreas al DWH.

Las áreas o equipos de analítica diseñarán los Dashboards de una única fuente principal, única, centralizada y homogeneizada, que permita la toma de decisiones a nivel estratégico y de mandos medios.

El tiempo del proyecto será aproximadamente de 290 días hábiles.

El costo de profesionales para la elaboración de la bodega de Datos se estima en $350.000.000,oo.

El costo anual aproximado de la Bodega de Datos en plataforma Azure es de $93.228.500,oo.

11.2. Recomendaciones

Se recomiendan los siguientes aspectos a implementar en un futuro para un mejor proceso de Inteligencia de Negocios (Business Intelligence), emprender proyectos similares o fortalecer el proceso realizado.

- Implementar Data Mining: es un proceso técnico, automático o semiautomático, que analiza grandes cantidades de información dispersa para darle sentido y convertirla en conocimiento. Busca anomalías, patrones o correlaciones entre millones de registros para predecir resultados. (IBERDROLA, 2022)
- Implementar Machine Learning, es una disciplina del campo de la Inteligencia Artificial, que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones (análisis predictivo). (IBERDROLA, 2022)
- Implementar procesos de Análisis Predictivo y Prescriptivo

12. Anexos

- **Anexo 1*:*** Modelo Canvas Bodega de Datos (Data Warehouse) Fresenius Medical Care Colombia.
- **Anexo 2*:*** Matriz DOFA Bodega de Datos (Data Warehouse) Fresenius Medical Care Colombia.
- **Anexo 3*:*** Cronograma de Implementación.

Referencias

ABC del data mart. (13 de Septiembre de 2021). *Evaluando Software*. Obtenido de https://www.evaluandosoftware.com/abc-del-data-mart/

Bernal, B. L. (2002). *Guía para la construcción de un Data Warehouse .*

Curto, J., & Conesa, J. (s.f.). El conocimiento es imprescindible.

Flórez, D. F., & Del Mar, L. M. (2015). *Guía para la implementación de una solución de inteligencia de negocios para pequeñas y medianas empresas.*

Forero, D. A., & Sánchez, J. A. (2021). Introducción a la inteligencia de negocios basada en la metodología Kimball. *Revista TIA*, págs. 1-17.

Fresenius Medical Care Colombia. (2022). *Fresenius Medical Care Colombia*. Obtenido de https://www.freseniusmedicalcare.com.co/es-co/quienes-somos/historia

Gravitar. (s.f.). Obtenido de https://gravitar.biz/datawarehouse/metodologias-data-warehouse/

IBERDROLA. (2022). Obtenido de Descubre cómo el 'data mining' predecirá nuestro comportamiento: https://www.iberdrola.com/innovacion/data-mining-definicion-ejemplos-y-aplicaciones

IBERDROLA. (2022). Obtenido de https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico

Ing. Bernabeu, R. D. (17 de Enero de 2009). *DATA WAREHOUSING Investigación y Sistematización de Conceptos.*

Orellana, F. (2013). *Propuesta de implementación de un Data Warehouse para el área de Soporte de Información, Rabie S.A.*

Pérez, C. (2015). IV. Bill Inmon - Ralph Kimball. *VORTECH*.

PoweData. (2022). *Data Warehouse: todo lo que necesitas saber sobre almacenamiento de datos*.

Reinoso, I., & Ariza, M. T. (Septiembre de 2014). Explotación de un Data Warehouse. Fundamentos y caso práctico para la gestión de proyectos. Sevilla.

Rius , À., Serra , M., & Curto , J. (s.f.). *Introducción al almacenamiento de datos.* Obtenido de https://openaccess.uoc.edu/bitstream/10609/75985/1/Data%20warehouse_M%C3%B3dulo%201_Introducci%C3%B3n%20al%20almacenamiento%20de%20datos.pdf

Rodríguez, M. (22 de Julio de 2010). *Análisis diseño de un data mart para el seguimiento académico de alumnos en un entorno universitario*.