Universidad de Manizales
Facultad de Ciencias e Ingeniería
Ingeniería de sistemas y telecomunicaciones

COMPARATIVO DE METODOLOGÍAS Y HERRAMIENTAS PARA EL DESARROLLO DE UN DATA WAREHOUSE

SANTIAGO HERNÁNDEZ MEJÍA

UNIVERSIDAD DE MANIZALES
FACULTAD DE CIENCIAS E INGENIERÍA
INGENIERÍA DE SISTEMAS Y TELECOMUNICACIONES
MANIZALES
2017

COMPARATIVO DE METODOLOGÍAS Y HERRAMIENTAS PARA DATAWAREHOUSING

SANTIAGO HERNÁNDEZ MEJÍA

Trabajo de Grado presentado como requisito parcial para optar al título de Ingeniero de sistemas y telecomunicaciones

Presidente
Carlos Betancourt Correa
Magister en Educación Docencia

UNIVERSIDAD DE MANIZALES
FACULTAD DE CIENCIAS E INGENIERÍA
INGENIERÍA DE SISTEMAS Y TELECOMUNICACIONES
MANIZALES
2017

CRÉDITOS

Las personas que participaron en este proyecto fueron las siguientes**:**

NOMBRE COMPLETO	FUNCIÓN EN EL PROYECTO	DIRECCIÓN DE CONTACTO	CORREO ELECTRÓNICO
Santiago Hernández	Autor	Carrera 9 No. 19 - 03	santiago.hernandez@umanizales.edu.co
Carlos Betancourt	Presidente	Carrera 9 No. 19 - 03	cbc@umanizales.edu.co
Omar Antonio Vega	Asesor metodológico	Carrera 9 No. 19 - 03	oavega@umanizales.edu.co
Luis Carlos Correa	Asesor metodológico	Carrera 9 No. 19 - 03	lcco@umanizales.edu.co

PÁGINA DE ACEPTACIÓN

_____
_____
_____
_____
_____
_____
_____

_____
<NOMBRE COMPLETO>
JURADO

_____
<NOMBRE COMPLETO>
JURADO

_____
<NOMBRE COMPLETO>
JURADO

Manizales, 25 de Octubre de 2017

CONTENIDO

LISTA DE FIGURAS

Pág.

LISTA DE TABLAS

Pág.

LISTA DE ANEXOS

Pág.

RESUMEN

En este documento se estudian, analizan y comparan diversas metodologías y herramientas para el desarrollo de un *Data Warehouse* (DW) que permita la integración de información en caso, o no que dichos datos se encuentren en diferentes motores de bases de datos y/o provengan de diferentes fuentes de datos, esto, con el fin de convertir los datos en información pertinente y para que dichos datos cumplan con características como la calidad y exactitud, entre otras. Con la gran ventaja de que una vez el *Data Warehouse* esté desarrollado, se puedan ejecutar procesos de *Business Intelligence* (BI) para lograr que la información pueda ser usada para la toma de decisiones.

PALABRAS CLAVES: Data Warehouse, Comparativo, Metodologías, Herramientas.

ABSTRACT

This document examines, analyzes and compares various methodologies and tools for the development of a Data Warehouse (DW) that allows the integration of information in case, or not that said data is found in different database engines and / or comes from Of different data sources, in order to convert the data into relevant information and so that such data comply with characteristics such as quality and accuracy, among others. With the great advantage that once the Data Warehouse is developed, you can execute Business Intelligence (BI) processes to make the information can be used for decision making.

KEY WORDS: Data Warehouse, Comparative, Methodologies, Tools.

INTRODUCCIÓN

Las empresas poseen grandes cantidades de datos, que necesitan convertirse en información, para que sea de apoyo en la toma de decisiones en ellas, pero en algunas de estas instituciones estos datos se encuentran en diferentes motores de bases de datos, lo cual hace que la consulta y extracción de la información, no sea tan eficiente, y la visualización de la información no es tan intuitiva para los encargados del área específica que desean acceder a esta información. Actualmente en algunas empresas tienden a usar soluciones de *Business intelligence*, pero la información se encuentra en diferentes motores de bases de datos y son controlados y gestionados por diferentes aplicaciones en múltiples plataformas. En este trabajo se pretende plantear un comparativo de metodologías y herramientas para una solución de integración de datos para la ayuda en el manejo de la información y la toma de decisiones, proponiendo un análisis de metodologías para el diseño de almacenes de datos (*Data Warehouse*), que pueda ser adaptable a cualquier empresa y dejar abierta la posibilidad después de la implementación a gusto de la empresa que pueda ser aplicar soluciones de *Business Intelligence* si así lo desean.

Para realizar el presente trabajo se hizo una recopilación de información, consultando diferentes bases de datos científicas y diferentes fuentes confiables de información para la abstracción de información necesarias para este documento.

1. ÁREA PROBLEMÁTICA

Para la gestión de los datos es pertinente que la información cumpla con unas características como la calidad, exactitud, integridad, confidencialidad, disponibilidad, seguridad, relevancia, entre otras. En la evolución que han tenido los sistemas de información de algunas empresas, las suele llevar a que tengan aplicativos de software que le dan persistencia a sus datos en diferentes motores de base de datos.

Uno de los elementos fundamentales para la gestión de los datos, es tenerlos integrados y existen herramientas libres y de pago que permiten la integración de dicha información y además permiten la visualización de diferentes fuentes de datos (Data Warehouse). La información que se maneja en algunas empresas no logra presentarse de una manera comprensible con el fin de basarse en ella para la gestión de los datos que permita tomar decisiones.

Algunos proyectos suelen estancarse en el análisis de herramientas o metodologías o no suelen referenciar fuentes confiables para la elección de las mismas, llegando incluso a causar retrasos al no elegir correctamente la metodología o herramienta que mejor se adapte al proyecto y tener que replantear requerimientos o adaptar el proyecto a una nueva herramienta o una nueva metodología.

El proyecto pretende llegar hasta una comparación de metodologías y herramientas para la construcción de un Data Warehouse.

2. OBJETIVOS

2.1 OBJETIVO GENERAL

Realizar un comparativo de metodologías y herramientas para el diseño de un Data Warehouse (DW) que permita integrar la información.

2.2 OBJETIVOS ESPECÍFICOS

- Revisar y comparar las metodologías para el diseño y control de *Data Warehouse*.
- Describir los lineamientos de diseño.
- Revisar y comparar las herramientas para el diseño y control de *Data Warehouse*.
- Revisar y comparar las herramientas para la visualización de datos.

3. JUSTIFICACIÓN

Se hace necesaria la proposición de una comparación de metodologías y herramientas para el diseño de un *Data Warehouse* en las empresas, debido a que están abocadas a desarrollar propuestas de *Business Intelligence* a partir de sus datos y fuentes externas, para esto, se deben integrar los datos de la empresa y el modo que se propone, es el de un Data Warehouse que permita la integración de las bases de datos de algunas áreas sin alterar las existentes, con base en otras implementaciones que no ofrecen la posibilidad de visualizar los datos de forma gráfica. Su factor innovador es la capacidad que tiene para adaptarse a cualquier empresa que necesite de su implementación, adicionalmente que puede ser actualizado a medida que cambian la empresa o sus áreas.

Muchas personas se ven involucradas en procesos extensos debido a la separación de las bases de datos en la empresa. Empresas que presenten problemas similares podrían verse beneficiadas al adaptar esta comparación de metodologías y herramientas de *Data Warehouse*.

Haciendo la comparación de metodologías y herramientas para el *Data Warehouse*, tanto los trabajadores como los entes administrativos de la empresa, podrán acceder en solo sitio, a la información proporcionada por las diferentes áreas de la empresa.

4. MARCO TEÓRICO

4.1 MARCO CONCEPTUAL

4.1.1 DATA WAREHOUSE.

> W.H. Inmon, considerado el padre de las bodegas de datos en el 92, define los Data Warehouse como: "Un sistema orientado al usuario final, integrado, con variaciones de tiempo y sobre todo una colección de datos como soporte al proceso de toma de decisiones". Por otra parte, Ralph Kimball, considerado como uno de los más importantes precursores y padre del concepto Data Warehouse, lo define como: "una copia de los datos de la transacción estructurados específicamente para preguntar y divulgar"[1].

La gran importancia de los DW radica en la facilidad en la síntesis de determinados procesos empresariales que repercutan en la toma de decisiones, ya que uno de los principales propósitos de los DW es buscar apoyar la toma de decisiones, además, la información que se muestra en los reportes no se trata únicamente de datos vacíos, los reportes traen consigo un proceso de filtrado de datos que permite convertirlos en información comprensible para los encargados de tomar las decisiones, información de la cual se sacan conclusiones del funcionamiento de la empresa sin poner a los encargados de tomar las decisiones en la tediosa labor de estar analizando dato por dato cada uno de los procesos que ejecuta la empresa.

> La tendencia hacia la que apunta la “inteligencia de negocios” es la divulgación de la información, tanto a nivel gerencial como a todo aquel que la necesite desde diferentes dimensiones y niveles asociados, para lograr obtener informes consolidados o detallados que faciliten la síntesis de determinado proceso empresarial y que repercutan directamente en la toma de decisiones, objeto que en últimas constituye el objetivo mismo de los Data Warehouses[2].

En la actualidad, con el fin de lograr un mejor rendimiento las organizaciones tienden a gastar su esfuerzo en la maximización de ingresos y minimizar los gastos, para conseguir dicha meta se requieren elementos que deben reflejarse desde los empleados de nivel más bajo hasta los altos mandos ejecutivos, esto se logra con unos arduos y tediosos procesos de análisis estratégicos para mejorar los procesos empresariales basados en las decisiones de quienes están al mando, es por esto que el lograr obtener informes consolidados o detallados de la información de la empresa, resulta

[1] SALCEDO PARRA, Octavio J.; GALEANO, Rita Milena & RODRIGUEZ B., Luis G. Metodología Crisp para la implementación Data Warehouse. En: Tecnura. Bogotá: Universidad Distrital Francisco José de Caldas. Vol. 14, No. 26, 2010, p. 37. ISSN: 0123-921X

[2] Ibíd., p.35

vital en la toma de decisiones y es aquí donde entran en juego los DW, debido a que su propósito en simples cuentas es obtener informes detallados de la información de la empresa con el fin de mejorar la toma de decisiones.

4.1.2 CUBOS DE DATOS.

> (OLAP) es una estructura tridimensional aplicada en una base de datos multidimensional[3].

Se necesita indicar cómo se aplican los cubos durante el proceso de creación de un Data Warehouse para lo cual su relación con los procesos del negocio, adicional da a conocer definiciones de los diferentes elementos que aplican en la construcción del DW en este caso la aplicación de un cubo, para ello se basa en la explicación de lo que son los cubos OLAP con todas sus característica y el papel que juega permitiendo así manipular la nueva información de manera dinámica quitando así toda la complejidad anterior que manejaba la organización para el manejo de la información.

4.1.3 DATA MART.

> Las creación de consultas ad-hoc con la herramienta de elaboración de informes es mucho más accesible para los usuarios de negocio al utilizar los data marts como fuente de datos, ya que contienen conjuntos de elementos de información más pequeños y focalizados, y requieren también menos lógica compleja por parte del usuario, puesto que las reglas del negocio están ya incorporadas en los propios datos[4].

Los Data Marts se especializan en el almacenamiento y persistencia de datos de un área específica de la organización, un Data Mart por cada dependencia. Se distinguen porque su estructura se centra en analizar la información con todo el detalle teniendo en cuenta los procesos que llevan a cabo en el departamento o área donde fue implantado. Como se menciona anteriormente, los Data Marts al estar especializados y focalizados en ser pequeños y de fácil manejo para los encargados de monitorear los reportes que genera, esto se debe a que, tanto la lógica del negocio, como la del área donde fue implantado, ya se encuentra incorporada en los datos y en su modelo. Cabe destacar que los Data Warehouse están comprendidos por tantos Data Marts sean necesarios según la cantidad de áreas o dependencias hayan en la empresa donde se vaya a implementar.

---

[3] KIMBALL, Ralph y ROSS, Margy. The Data Warehouse toolkit. 3 ed. Indianapolis: John Wiley & Sons, Inc, 2013. p. 40. ISBN 978-1-118-53080-1

[4] WALZ, Aaron. Caso de estudio Universidad de Illinois. En: EVEREST. Libro blanco inteligencia institucional en universidades. Arequipa: Oficina de Cooperación Universitaria S.A, 2013. p. 317. ISBN: 978-84-695-8892-5

4.1.4 BASES DE DATOS RELACIONALES.

> Una base de datos relacional es un repositorio compartido de datos. Para hacer disponibles los datos de una base de datos relacional a los usuarios hay que considerar varios aspectos. Uno es la forma en que los usuarios solicitan los datos: ¿cuáles son los diferentes lenguajes de consulta que usan? (…) Otro aspecto es la integridad de datos y la seguridad; las bases de datos necesitan proteger los datos del daño provocado por los usuarios, ya sean intencionados o no. El componente de mantenimiento de la integridad de una base de datos asegura que las actualizaciones no violan las restricciones de integridad que hayan especificado sobre los datos. El componente de seguridad de una base de datos incluye la autenticación de usuarios y el control de acceso para restringir las posibles acciones de cada usuario[5].

Las bases de datos relaciones son un concepto bastante importante para abordar, ya que, usualmente se orientan a procesos operativos de la empresa, con el fin de automatizar tareas de la organización y a recolectar información de la misma, además, los cálculos de las bases de datos relacionales usualmente son simples y poco sofisticados, mientras que el DW está orientado a la toma de decisiones.

El DW que se plantea en este proyecto, tiene como primera función integrar algunas bases de datos de la UM, que son de tipo relacionales. Esto con muchos fines debido a la característica de un DW, tales como, almacenar datos categorizados o estructurándolos para el análisis y toma de decisiones por parte de los directivos, proporcionar análisis históricos de la información, sistemas de información que permiten la visualización gráfica de la información almacenada en forma de datos, y además, los DW suelen ser intensivos en cálculos y se prestan para la explotación de herramientas de minería de datos o *data mining*.

Si bien debido a las características de una base de datos relacional se pueden lograr algunas funciones básicas de las características antes mencionadas, no es suficiente para la toma de decisiones por parte de los directivos y por eso se hace necesaria una implementación, porque no hay ningún DW en ella.

INTEGRIDAD DE DATOS Y SEGURIDAD. Otra diferencia notoria entre un DW y una base de datos relacional, es la seguridad de los datos, una base de datos relacional es bastante insegura en comparación a un DW, debido a que estas permiten información de lectura y escritura, en cambio, los DW solo permiten leer los datos. Debido a que la autenticación de los usuarios es un tema a destacar en cuanto a la integridad de datos y seguridad en una base de datos relacional, vale la pena mencionar que como el DW está orientado a la toma de decisiones, solo

[5] SILBERSCHATZ, A., KORTH, H. & SUDARSHAN, S. Fundamentos de bases de datos. 4ed. Madrid: MacGraw-Hill. 2002. p. 86. ISBN: 84-481-3654-3

tendrán acceso quienes las puedan tomar, dándole un control absoluto sobre la información que tiene el DW.

4.1.5 GOBERNANZA DE DATOS.

> El gobierno de datos es una disciplina de control de calidad para añadir nuevo rigor y disciplina al proceso de gestión, utilizando, mejorando y protegiendo la información de la organización. El gobierno de datos eficaz puede mejorar la calidad, disponibilidad e integridad de los datos de una empresa mediante el fomento de la colaboración entre la organización y la formulación de políticas estructuradas. Equilibra silos facciones con interés de las organizaciones[6].

En las diferentes organizaciones se evidencia que el no correcto uso y obtención de la información de la organización afecta directamente en la toma de decisiones, por ejemplo si necesita obtener la información de los movimientos en las ventas de los últimos cinco años, pero tienen los datos dispersos, la obtención de esta información va a ser más difícil y va a tardar más tiempo en conseguirse, dejando de ser eficiente y eficaz. Como se puede apreciar en la información que nos da a conocer IBM la gobernanza de los datos de las organizaciones es la mano derecha en cuanto a la toma de decisiones haciendo un uso efectivo de la información que la organización genera.

En el proyecto se propone que la organización tenga un mejor uso de la información y los datos que esta produce, por medio de este modelo de DW es indispensable que el término “gobernanza de datos” este claro para que así la organización pueda entender como a partir de sus datos mejoran la toma de decisiones.

4.1.6 ETL.

EXTRACCIÓN.

> Dos factores principales diferencian la extracción de datos para un nuevo sistema operativo a partir de la extracción de datos de un almacén de datos. En primer lugar, para un almacén de datos, hay que extraer los datos de muchas fuentes diferentes. A continuación, para un almacén de datos, tiene que extraer datos sobre los cambios de las cargas incrementales en curso, así como para una carga completa inicial de una sola vez. Para los sistemas operativos, todo lo que necesita es extracciones de una sola vez y conversiones de datos.[7]

---

[6] IBM. The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance. 2007; 3 p. ISSN: LO11960-USEN-00

[7] PONNIAH, Paulraj. Data Warehousing: Fundamentals for IT professionals. 2 ed. USA: John Wiley & Sons, Inc, 2010. p. 286. ISBN 978-0-470-46207-2

Los autores del libro indican los pasos para los cuales se extrae la información de relevancia para la organización, en estos pasos tiene una explicación clara de lo que debe hacerse y el por qué, es muy importante en la realización de un DW tener conocimientos claros y precisos para la creación del mismo. Para la creación del diseño del DW es importante conocer todos los procesos de manera que se puedan aplicar de manera secuencial, cada técnica mencionada en el libro van enfocadas a cómo se desarrolla la organización más específicamente en cada sistema de esta.

TRANSFORMACIÓN.

> En primer lugar, todos los datos extraídos deben ser utilizables en el almacén de datos. Tener información que es útil para la toma de decisiones estratégicas es el principio subyacente del almacén de datos. Usted sabe que los datos en los sistemas operativos no son utilizables para este propósito. A continuación, ya que los datos de funcionamiento se extraen de muchos sistemas heredados de edad, la calidad de los datos en estos sistemas es menos probable que sea lo suficientemente bueno para el almacén de datos. Usted tiene que enriquecer y mejorar la calidad de los datos antes de que pueda ser utilizable en el almacén de datos[8].

Se debe identificar cómo la transformación de los datos se aplican a las reglas del negocio o a la manera en la que van a ser funcionales los datos que ya fueron extraídos para luego ser cargados. Este proceso de transformación permite manipular los datos o utilizar cierta información, todo depende de los datos que para la organización o área sean necesarios, de manera que no solo sean datos separados si no, ya unificados en lo estrictamente indicado por la organización, el libro entonces indica una serie de transformaciones predeterminadas y la explicación de cada una de ellas.

La obtención de todos los tipos de transformaciones que se le pueden realizar a los datos y entender en qué casos se utilizan es vital, ya que se tiene que hacer un gran paralelo en los requerimientos y los pasos que se deben realizar para cada uno de ellos, entender su funcionamiento y expresar si se está cumpliendo o no, a medida que se desarrolla el proceso de transformación de los datos.

CARGA Y LIMPIEZA.

> En general se acepta que las funciones de transformación terminan tan pronto como se crean imágenes de carga. La siguiente serie importante de funciones consta de los que toman los datos preparados, que se aplican al almacén de datos, y lo almacenan en la base de datos existente. Se crean imágenes de

[8] Ibíd., p 295

carga que correspondan con los archivos de destino para ser cargados en la base de datos de almacenamiento de datos[9].

En este proceso de carga se tiene en cuenta actividades diferentes que dependen de los requerimientos de la organización para lo cual se cargan en la base de datos correspondientes. El libro también expresa que esta etapa es la última durante el proceso de ETL en la cual se puede resumir todas las transacciones durante un tiempo y devolver una única en el DW, pero también se puede almacenar la información en distintos niveles ya sea por jerarquía o por un periodo de tiempo determinado, como se indicó anteriormente depende de lo que la empresa desea visualizar o de lo que es necesario para ellos.

El proceso de carga seguirá los pasos indicados en el libro, de manera que se pueda obtener nuevos conceptos acerca del proceso, y realizar paso a paso los procedimientos mencionados en este, verificando cuales opciones son más acertadas para ser aplicadas de manera que cumplan con los requerimientos que se han planteado para el desarrollo del proyecto.

4.2 MARCO LEGAL

4.2.1 LEY ESTATUTARIA 1266 DE 2008. La ley 1266 de 2008 al tener por objeto el desarrollo del derecho constitucional[10] que tienen todas las personas a conocer actualizar y rectificar las informaciones que se hayan recogido sobre ellas en bases de datos, y los demás derechos libertades y garantías constitucionales relacionadas con la recolección tratamiento y circulación de datos personales a que se refiere el artículo 15 de la Constitución Política. Por ende, es importante tener en cuenta cada consideración debido a que el desarrollo del proyecto en cuestión necesita el tratamiento de cierta información con la cual no se piensa incumplir ningún mandato de esta ley con respecto al tratamiento o distribución de datos o a cualquier otro punto sobre los datos personales que use el proyecto añadiendo que se rige bajo el mandato público de la Constitución política.

4.2.2 LEY ESTATUTARIA 1581 DE 2012. Las bases de datos utilizadas en la las empresas poseen información personal de administrativos y estudiantes, a su vez, contiene información relevante de la institución, por lo cual se necesita tener protección de estos datos con los límites establecidos por la empresa para su uso, adicional se debe tener reserva total de la información que se va a visualizar. En el

[9] Ibíd., p 302

[10] COLOMBIA. CONGRESO DE LA REPÚBLICA. Ley estatutaria 1266 (31, Diciembre, 2008) por la cual se dictan las disposiciones generales del habeas data y se regula el manejo de la información contenida en bases de datos personales, en especial la financiera, crediticia, comercial, de servicios y la proveniente de terceros países y se dictan otras disposiciones. Diario oficial. Bogotá, 2008. no. 47219. p.1

compromiso con la empresa, se garantiza total integridad de las bases de datos utilizadas durante el proyecto, de igual manera la información solo la podrán visualizar aquellas personas que ha autorizado la empresa quienes se rigen por la ley estatutaria 1581 de 2012 en la que se dictan disposiciones generales para la protección de datos personales y además tiene por objeto[11] desarrollar el derecho constitucional que tienen todas las personas a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en bases de datos.

[11] COLOMBIA. CONGRESO DE LA REPÚBLICA. Ley estatutaria 1581 (17, Octubre, 2012) Por la cual se dictan disposiciones generales para la protección de datos personales. Diario oficial. Bogotá, 2012. no. 48587. p.1

4.3 MARCO REFERENCIAL

4.3.1 SCHEDULING TO MINIMIZE STALENESS AND STRETCH IN REAL-TIME DATA WAREHOUSES

> Los almacenes de datos integran información de múltiples bases de datos operacionales para permitir el análisis de negocio complejas. En las aplicaciones tradicionales, los almacenes se actualizan periódicamente (por ejemplo, cada noche o una vez a la semana) y el análisis de datos se lleva a cabo fuera de línea. Por el contrario, los almacenes en tiempo real, también conocidos como depósitos activos, se cargan continuamente los datos de entrada se alimenta para apoyar el análisis de tiempo crítico[12].

En las empresas se almacenan gran cantidad de datos que provienen de las bases de datos correspondientes a las diferentes áreas de las mismas, el artículo mencionado, permite dar una idea de cómo se puede aplicar los Data Warehouses, un ejemplo es que por medio de todas las transacciones que se realizan en el área financiera, se puede identificar en tiempo real todas las tendencias históricas, y adicionalmente todas las transacciones realizadas, pero la eficacia de este almacén de datos depende del ingreso de los nuevos datos, es decir, la integración de todas las bases de datos.

4.3.2 INTELIGENCIA INSTITUCIONAL EN UNIVERSIDADES

> Un sistema dedicado: Se consolidan en un solo sistema central (Data Warehouse) los datos de las diversas áreas de gestión, unificando el tratamiento y especializando el modelo de almacenamiento para facilitar el análisis. Todos los datos institucionales identificados como relevantes para la producción de información analítica, independientemente de su origen, naturaleza y destino, son recogidos e integrados en este almacén central, lo que facilita su procesamiento y consumo con fines analíticos[13].

Enfocar en cómo los Data Warehouse realizan un análisis de los datos con el objetivo de extraer la información necesaria para las diferentes áreas de una empresa, todos los estudios y capítulos en este libro está enfocado a las universidades, lo cual complementa las ideas que se tienen para la implementación del Data Warehouse.

---

[12] BATENI, Mohammad. Scheduling to Minimize Staleness and Stretch. En: Real-Time Data Warehouses. USA: Theory of Computing Systems Vol. 49, No 4. (2011); p. 758. ISNN: 1432-4350

[13] Oficina de Cooperación Universitaria. Inteligencia institucional en universidades. Arequipa, Perú, 2013. 40 p. ISBN: 978-84-695-8892-5

4.3.3 INCORPORACIÓN DE ELEMENTOS DE INTELIGENCIA DE NEGOCIOS EN EL PROCESO DE ADMISIÓN Y MATRÍCULA DE UNA UNIVERSIDAD CHILENA

> La inteligencia de negocios ha dado impulso notable al mejoramiento de procesos de las organizaciones, la Universidad de Tarapacá indica que "se implementó un data mart (DM) centrado en el área de Admisión y Matrícula de la Vicerrectoría Académica. Su desarrollo requirió de la realización de actividades tales como la obtención de los requerimientos del negocio, la investigación del indicador clave de rendimiento (KPI) del área, el análisis de las distintas fuentes de información interna y el desarrollo de un modelado dimensional basado en el esquema estrella de Kimball[14].

El artículo publicado da a conocer los procesos y conceptos que utilizaron para el análisis de los datos. Este artículo ayuda a comprender cómo la inteligencia de negocios, optimiza los procesos de mejora en un área específica de la organización, en el caso de la Universidad de Chile en los procesos de admisión y matrícula de los estudiantes.

4.3.4 RECURSOS DIGITALES PARA LA EDUCACIÓN Y LA CULTURA

> La reprobación escolar, específicamente en el nivel superior, es un fenómeno altamente indicativo de la crisis por la que atraviesa la sociedad en general y la educación. En este trabajo se llevó a cabo el análisis de los datos que permitirán generar un modelo que ayude a predecir, desde que los alumnos ingresan a la Universidad, las causas que los llevarán a reprobar, así como las materias con mayor riesgo[15].

Este artículo permite relacionar dentro del contexto universitario y los recursos digitales, estudios realizados con los que se podrían realizar para el mejor rendimiento en las diferentes materias que vean los estudiantes en una universidad, esto es importante ya que la universidad está fundamentada por los estudiantes y es fundamental el rendimiento positivo de los estudiantes y abre posibilidades a las empresas para mejorar el rendimiento de sus trabajadores.

4.3.5 MODELO DE UN SISTEMA DE INTELIGENCIA DE NEGOCIOS BASADO EN S-BSC PARA ENTIDADES PRESTADORAS DE SERVICIO DE SALUD DE ALTA COMPLEJIDAD SIN ÁNIMO DE LUCRO

> El uso correcto de la información proporciona a las organizaciones una ventaja competitiva. La revolución de la información está barriendo a través de nuestra

[14] FUENTES TAPIA, Luis y VALDIVIA PINTO, Ricardo .Incorporación de elementos de inteligencia de negocios en el proceso de admisión y matrícula de una universidad chilena. Arica (Chile). (2010); p.383. ISSN: 0718-3291

[15] PRIETO, Manuel; DODERO, Juan y VILLEGAS, David. Recursos digitales para la educación y la cultura. 2010; p.48. ISBN: 978-607-95446-1-4

> economía, Ninguna empresa puede escapar a sus efectos. Reducciones en el costo de obtener el procesamiento y la transmisión de información están cambiando la forma de hacer negocios. El cuadro de mando integral expande el conjunto de objetivos de las unidades de negocios más allá de los indicadores financieros. Los ejecutivos de una empresa pueden, ahora medir la forma en que sus unidades de negocio crean valor para sus clientes presentes y futuros, y la forma en que deben potenciar las capacidades internas y las inversiones en personal, sistemas y procedimientos que son necesarios para mejorar su actuación futura[16].

Las empresas necesitan ir más allá de tener una lista de indicadores, para ello necesita utilizar su información para convertir estos indicadores y estrategias en un plan de acción y ejecutarlo correctamente basados en la información analizada por parte de los directivos.

4.3.6 BUSINESS INTELIGENCE

> Sistemas de BI combinan los datos operacionales con herramientas analíticas para presentar compleja y la información de la competencia a los planificadores y tomadores de decisiones. El objetivo es mejorar la puntualidad y la calidad de los insumos en el proceso de decisión. Inteligencia de Negocios se utiliza para entender las capacidades disponibles en la empresa; el estado de la técnica, las tendencias y direcciones futuras en los mercados, las tecnologías y el entorno regulatorio en el que la empresa compite; y las acciones de los competidores y las consecuencias de estas acciones[17].

La inteligencia de negocios permite entender las capacidades de la empresa; el estado de la técnica, las tendencias y direcciones futuras en los mercados, las tecnologías y el entorno regulatorio en el que la empresa compite; y las acciones de los competidores y las consecuencias de estas acciones. Se podrían tomar gran cantidad de decisiones siendo apoyados por la información procesada en el momento adecuado, el lugar correcto y de la forma adecuada.

4.3.7 ANÁLISIS COMPARATIVO DE MODELOS DE MADUREZ EN INTELIGENCIA DE NEGOCIO

> En el área de Business Intelligence (BI) se han planteado varios modelos de madurez, con distintos orígenes y propósitos. Dentro de esta diversidad, actualmente no se dispone de una comparación cuantitativa o cualitativa de

[16] ROBAYO, Eduard. Modelo de un sistema de inteligencia de negocios basado en S-BSC para entidades prestadoras de servicio de salud de alta complejidad sin ánimo de lucro. 2014. Magister en Gestión de Información. Escuela Colombiana de Ingeniería

[17] NEGASH, Solomon. Business Intelligence. En: Communications of the Association for Information Systems. Vol. 13, Article 15. (2004); p.178. ISSN: 1529-3181

> estos modelos que entregue argumentos para seleccionar y utilizar uno de ellos como referencia en la mejora de la madurez en BI para una organización[18].

Con la intención de tener todos los almacenes de datos interactuando, este artículo da una idea para realizar un análisis más profundo de ellos, con el fin de obtener información de importancia, estableciendo fuentes de datos centralizadas utilizando BI (Business intelligence), con una arquitectura en común para estos almacenes de datos.

4.3.8 APLICACIÓN DE INTELIGENCIA DE NEGOCIOS (BI Y KPI) EN LA ESTRATEGIA DE PERMANENCIA ESTUDIANTIL: CASO FUNDACIÓN UNIVERSITARIA CATÓLICA DEL NORTE (COLOMBIA)

> Un sistema que, en conjunto con todos los procesos constitutivos de la Fundación Universitaria Católica del Norte, facilite el acompañamiento integral de los estudiantes, con el objetivo de generar el efecto contrario a la deserción, es decir, la permanencia. Esta solución hace parte de la cultura de gestión del conocimiento desde el modelo institucional de Inteligencia de Negocios (BI)[19].

Otros retos serían la realización de la depuración de los datos existentes y futuros de tal manera que sean consistentes y confiables para la toma de decisiones. Convertir la información en conocimiento a través de reportes y gráficos dinámicos, cuadros de mando integrales e indicadores que sean claves para el rendimiento de la institución es otro de los retos para la aplicación de BI, y no solo en la institución en cuestión, sino también en cualquier organización.

4.3.9 MICROSOFT'S CORTANA ASSISTS IN UNEARTHING POWER BI INSIGHTS

> La nueva integración de BI Cortana-Power "permite a cualquier persona obtener respuestas directamente desde sus datos clave para la empresa de una forma más amable, proactiva y de manera natural", Marcus Ash, grupo administrador de programas de Microsoft Cortana, anunció el 1 de diciembre en un post en el blog de la empresa. "Mediante el poder del BI capacidades de visualización de datos, Cortana puede proporcionarle las respuestas, que van desde simples valores numéricos ('revenue para el último trimestre'), gráficos ('número de oportunidades por team"), mapas ('cliente promedio de gasto en California por la ciudad'), o incluso informes completos del poder BI[20].

---

[18] PRIETO, Roberto et al. Análisis comparativo de modelos de madurez en inteligencia de negocio. Antofagasta. Chile. En: Ingeniare. Vol 23. No. 3. (2015); p.361. ISSN: 0718-3291. http://www.scielo.cl/scielo.php?pid=S0718-33052015000300005&script=sci_arttext

[19] TÓRRES VELÁSQUEZ, Carlos Fernando. Aplicación de Inteligencia de Negocios (BI y KPI) en la estrategia de permanencia estudiantil: caso Fundación Universitaria Católica del Norte (Colombia). En: Trabajos Conferencias TICAL. (2015)

[20] HERNÁNDEZ, Pedro. Microsoft's Cortana Assists in Unearthing Power BI Insights. eWeek. (2015); p1-1. ISSN: 1530-6283

Volviendo a Microsoft, esta vez, relacionado con Cortana, un asistente de búsqueda, quien recolecta información de búsquedas de los usuarios y al usar BI, le permite a Microsoft obtener una perspectiva rápida caracterizada por poder ejecutar varios algoritmos de análisis sobre los datos y únicamente con un clic. Implementar este tipo de análisis traería enormes ventajas, y si además, se logra una visualización gráfica y estadística de los datos, se podrían analizar aún más fácil, logrando así, facilitar la toma de decisiones por parte de las personas correspondientes.

4.3.10 MICROSOFT IMPROVES POWER BI MANAGEMENT FOR ENTERPRISES

> Para los usuarios finales, Microsoft anunció una nueva adición a la potencia de gráficos BI Gallery, el desplazador visualización del poder BI consultor Fredrik Hedenström. En lugar de los gráficos, el desplazador puede mostrar el poder BI información en un formato ticker, similar a los indicadores que se desplazan a lo largo de la parte inferior de los canales de noticias financieras en la televisión[21].

En caso de Microsoft, quienes implementaron un gateway con BI para ejercer cavar en sus organizaciones y lograr un mejor control sobre los usuarios del sistema. Teniendo en cuenta la exploración de datos y actividades de visualización. Lograr un control interno en cualquier organización facilita y acelera mucho los procesos internos, logrando una eficiencia y eficacia en la toma de decisiones basadas en datos, que es lo que se busca lograr.

[21] HERNÁNDEZ, Pedro. Microsoft Improves Power BI Management for Enterprises. eWeek. (2015); p1-1. ISSN: 1530-6283

4.4 MARCO METODOLÓGICO

4.4.1 SELECCIÓN DE REFERENCIAS. Para la selección de los documentos que sirven de apoyo para este proyecto, se tuvieron en cuenta su confiabilidad y su coincidencia con la temática del proyecto, buscando siempre apoyar tanto la definición de conceptos claves y específicos tanto como para argumentar afirmaciones.

Se hizo una búsqueda rigurosa de información relevante y confiable a lo largo de la duración del proyecto, con el fin de evitar que los conceptos correspondientes a la temática del proyecto estén desactualizados.

Para la selección de las referencias se hizo un filtrado por las temáticas de las fuentes, desechando las que no pudieran aportar ningún concepto, procedimiento o conocimiento para la realización de este proyecto. El filtrado y la selección de fuentes de información se hizo bajo los criterios de los autores.

Al final, se dejaron y citaron las fuentes que aportan a la temática del proyecto ya sea como referente pasado, como fuente de conceptos, como fuente de información legal o como fuente de procedimientos.

4.4.2 ELECCIÓN DE METODOLOGÍA DE DATAWAREHOSING.

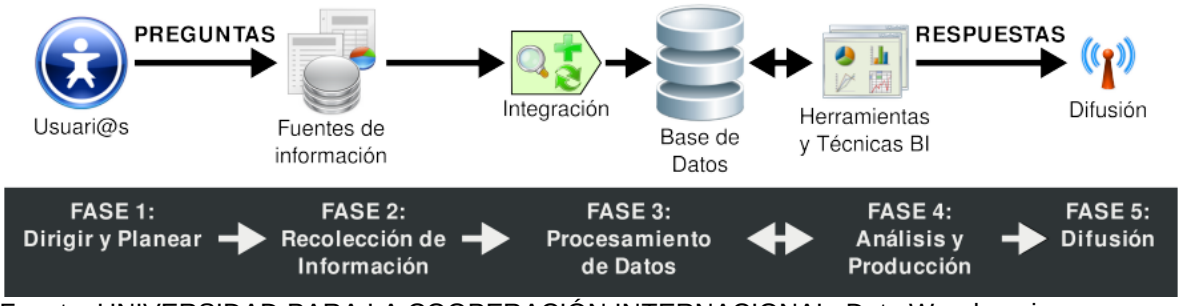
Debido a la cantidad de metodologías existentes para la implementación de un Data Warehouse (DW), es necesario estudiar la mejor opción para la creación del DW, por ende, en este apartado se plantea una comparación entre varias metodologías existentes para la implementación de DW.

Se plantea la descripción de cada una de las metodologías pre-seleccionadas, la comparación de las metodologías y los criterios de selección.

METODOLOGÍA DE HEFESTO

Con base en comprender cómo una organización puede crear inteligencia de negocios de sus datos, La metodología de Hefesto se divide en cinco fases y se sintetiza de la siguiente manera:

Figura 1. Fases de la metodología Hefesto:



Fuente: UNIVERSIDAD PARA LA COOPERACIÓN INTERNACIONAL. Data Warehousing y metodología Hefesto. [en línea]. Córdoba: Argentina. 2007. Sp. Fecha de consulta: 09/10/2016. Disponible en: http://www.ucipfg.com/Repositorio/MATI/MATI-10/BLOQUE-ACADEMICO/Unidad-03/lecturas/bibliografia/Data%20Warehousing%20y%20metodolog%C3%ADa%20Hefesto%201.pdf

- **FASE 1: Dirigir y Planear.** En esta fase inicial es donde se deberán recolectar los requerimientos de información específicos de los diferentes usuarios, así como entender sus diversas necesidades, para que luego en conjunto con ellos se generen las preguntas que les ayudarán a alcanzar sus objetivos.
- **FASE 2: Recolección de Información.** Es aquí en donde se realiza el proceso de extraer desde las diferentes fuentes de información de la empresa, tanto internas como externas, los datos que serán necesarios para encontrar las respuestas a las preguntas planteadas en el paso anterior.
- **FASE 3: Procesamiento de Datos.** En esta fase es donde se integran y cargan los datos en crudo en un formato utilizable para el análisis. Esta actividad puede realizarse mediante la creación de una nueva base de datos, agregando datos a una base de datos ya existente o bien consolidando la información.
- **FASE 4: Análisis y Producción.** Ahora, se procederá a trabajar sobre los datos extraídos integrados, utilizando herramientas y técnicas propias de la tecnología BI, para crear inteligencia. Como resultado final de esta fase se obtendrán las respuestas a las preguntas, mediante la creación de reportes, indicadores de rendimiento, cuadros de mando, gráficos estadísticos, etc.
- **FASE 5: Difusión.** Finalmente, se les entregará a los usuarios que lo requieran las herramientas necesarias, que les permitirán explorar los datos de manera sencilla e intuitiva[22].

[22] BERNEBEU, Ricardo Dario. Data warehousing y metodología Hefesto. [En línea]. Córdoba, Argentina: Universidad para la cooperación internacional. 2010. Sp. [consulta 09/10/2016]. Disponible en: < http://www.ucipfg.com/Repositorio/MATI/MATI-10/BLOQUE-ACADEMICO/Unidad-03/lecturas/bibliografia/Data%20Warehousing%20y%20metodolog%C3%ADa%20Hefesto%201.pdf >

METODOLOGÍA DE RALPH KIMBALL

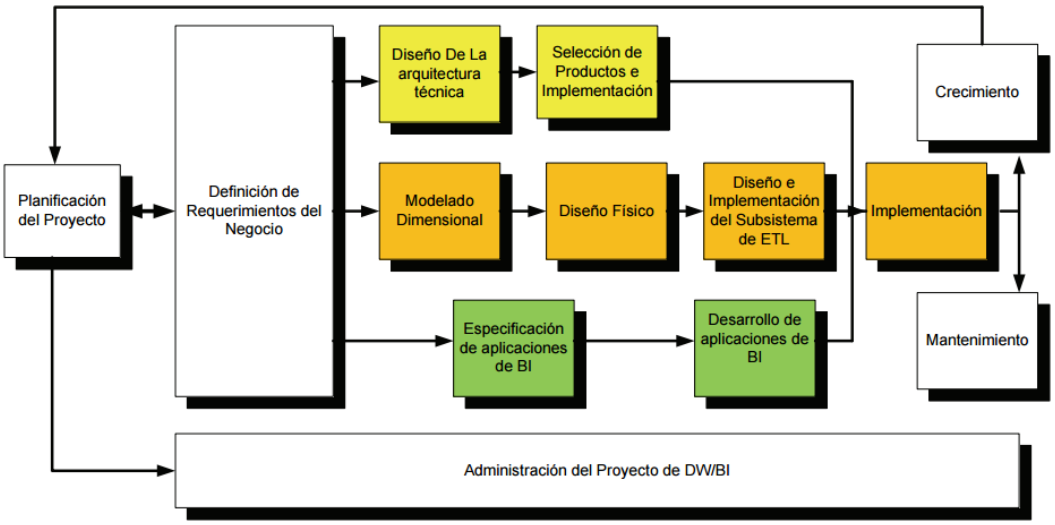
La metodología de Kimball se basa en cuatro principios fundamentales:

- **Centrarse en el negocio:** Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.
- **Construir una infraestructura de información adecuada:** Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.
- **Realizar entregas en incrementos significativos:** crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses. Hay que usa el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos. En esto la metodología se parece a las metodologías ágiles de construcción de software.
- **Ofrecer la solución completa:** proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. También se deberá entregar herramientas de consulta ad hoc, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación[23].

La construcción de una solución de DW es compleja, y Kimball propone una metodología que ayuda a simplificar dicha solución. Las tareas de esta metodología (ciclo de vida) se muestran en la siguiente figura:

[23] RIVADERA, Gustavo. La metodología de Kimball para el diseño de almacenes de datos (Data warehouses). [en línea]. Salta, Argentina: Cuadernos de la Facultad n. 5, 2010. P.58. [consulta: 09/10/2016]. Disponible en: <http://www.ucasal.edu.ar/htm/ingenieria/cuadernos/archivos/5-p56-rivadera-formateado.pdf>

Figura 2. Metodología de Ralph Kimball:



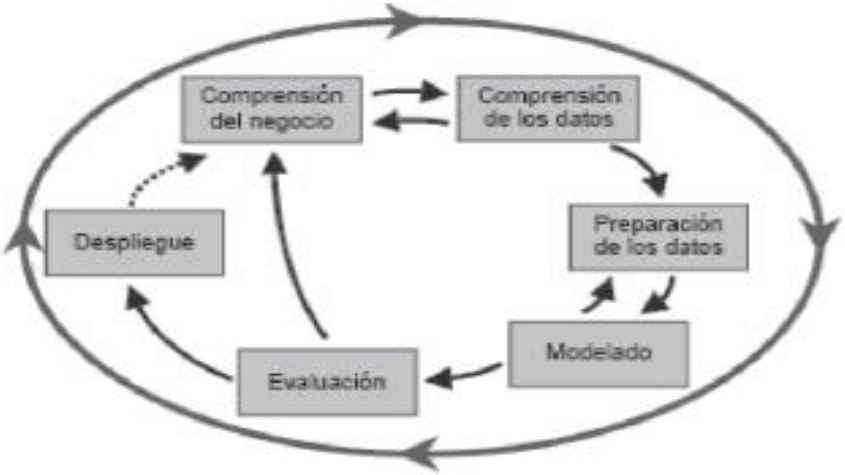
Fuente: CUADERNOS DE LA FACULTAD. La metodología de Kimball para el diseño de almacenes de datos (Data warehouses). [en línea]. Salta, Argentina: 2010. P.59. Fecha de consulta: 09/10/2016. Disponible en: http://www.ucasal.edu.ar/htm/ingenieria/cuadernos/archivos/5-p56-rivadera-formateado.pdf

METODOLOGÍA CRISP-DM (Cross- Industry Standard Pocessfor Data Mining).

La metodología de CRISP es una de las principales metodologías por seguir por los analistas en la inteligencia de negocios, donde se puede rescatar primordialmente Data Warehouse y Data Mining.

La metodología CRISP está sustentada en estándares internacionales que reflejas la robustez de sus procesos y que facilitan la unificación de sus fases en una estructura confiable y amigable para el usuario. Además de ello, esta tecnología interrelaciona las diferentes fases del proceso entre sí, de tal manera que se consolida un proceso iterativo y recíproco. Otro aspecto fundamental de esta tecnología es que es planteada como una metodología imparcial o “neutra respecto a la herramienta que se utilice para el desarrollo del proyecto de Data Warehouse o Data Mining siendo su distribución libre y gratuita.

Figura 3. Ciclo de vida de la metodología CRISP:



Fuente: REVISTA TECNURA. Metodología Crisp para la implementación Data Warehouse. [en línea]. Bogotá, Colombia: 2010. P.40. . Fecha de consulta: 09/10/2016. Disponible en: http://revistas.udistrital.edu.co/ojs/index.php/Tecnura/article/viewFile/6685/8268

El ciclo de vida del proyecto según la metodología según la metodología de CRISP está basado en seis fases cambiantes entre sí y nunca terminantes, lo cual lo postula como un ciclo en constante movimiento.

- **Comprensión del negocio:** Se trata de entender claramente los requerimientos y objetivos del proyecto siempre desde una visión de negocio. Esta fase se subdivide a su vez en las siguientes categorías:

    - Definición de los objetivos del negocio (inicial, objetivos de negocio y criterios de éxito del negocio).
    - Evaluación de la situación (inventario de recursos, requisitos supuestos y requerimientos, riesgos y contingencias, terminología y costes y beneficios).
    - Definición de los objetivos del DW (objetivos y criterios de éxito).
    - Realización del plan del proyecto (plan del proyecto y valoración inicial de herramientas y técnicas).

- **Comprensión de los datos:** Es conseguir y habituarse con los datos, reconocer las dificultades en la calidad de los datos y reconocer también las fortalezas de estos mismos que pueden servir en el proceso de análisis. Sus subdivisiones son:

    - Recolección inicial de datos (informe de recolección).
    - Descubrimiento de los datos (informe descriptivo de los datos).
    - Exploración de los datos (informe de exploración de los datos).
    - Verificación de la calidad de los datos (informes de calidad).

- **Preparación de los datos:** Es analizar los datos realmente importantes en el proceso de selección, depuración y transformación. Sus subdivisiones son:

    - Selección de los datos (motivos para incluirlos o excluirlos).
    - Depuración de los datos (reporte de depuración).
    - Estructuración de los datos (generación de atributos y registros)
    - Integración de los datos (agrupar los datos).
    - Formateo de datos (informe de la calidad de datos formateados).

- **Modelado:** Es la aplicación de técnicas de modelado o de Data Warehouse. Sus subdivisiones son:

    - Selección de la técnica de modelado (técnica y sus supuestos).
    - Generar el plan de pruebas (plan de pruebas).
    - Construcción del modelo (parámetros escogidos, modelos, descripción de los modelos).
    - Evaluación del modelo (evaluar el modelo, revisión de los parámetro elegidos).

- **Evaluación:** Esta fase es muy importante y decisiva, pues corresponde a la evaluación de la escogencia de los modelos anteriores y la toma de decisión respecto a si realmente son útiles en el proceso. Sus subdivisiones son:

    - Evaluar resultados (valoración de los resultados respecto al éxito del negocio, modelos aprobados).
    - Proceso de revisión (revisar el proceso).
    - Determinación de los pasos siguientes (listado de posibles acciones, técnica modelada).

- **Despliegue o divulgación:** Es la fase de implementación o de divulgación de los modelos anteriormente escogidos y evaluados. Sus subdivisiones son:

    - Plan de divulgación o implementación (plan de implementación).
    - Plan de monitoreo y mantenimiento (plan de monitoreo y mantenimiento).
    - Presentación del informe fina (informe final, presentación final).
    - Revisión del proyecto (documentación de la experiencia)[24].

4.4.3 ELECCIÓN DE HERRAMIENTA DE INTEGRACIÓN DE DATOS.

Debido a la cantidad de herramientas existentes para la implementación de un Data Warehouse (DW), es necesario estudiar la mejor opción para la creación del

[24] SALCEDO, Op. Cit., P.40

DW, por ende, en este apartado se plantea una comparación entre varias herramientas existentes para la implementación de DW.

Este apartado contiene la descripción de cada una de las herramientas pre-seleccionadas, la comparación de las herramientas y los criterios de selección. Toda la información de las aplicaciones fue sustraída de las páginas oficiales de los desarrolladores.

PENTAHO COMUNITY. Proporciona la extracción de gran alcance, transformación y capacidades de carga (ETL). Se puede utilizar esta herramienta de manera independiente para visualizar transformaciones de diseño y puestos de trabajo que extraen los datos existentes y ponen a disposición para facilitar la presentación de informes y análisis.

APLICACIONES. Pentaho Community tiene las siguientes aplicaciones:

- **PENTAHO'S DATA INTEGRATION.** Esta aplicación permite:

  - La supervisión del rendimiento del sistema mejorada
  - Mejora de los datos de perfiles
  - Fácil de añadir nuevos plugins
  - Transmitir datos de múltiples fuentes de datos
  - Revertir cambios en las transacciones de bases de datos de empleo

- **Report Designer.** Es una herramienta gráfica que genera informes de datos escuchados a través del motor de integración de datos sin la necesidad de ninguna tabla de etapas intermedias. Puede dar salida a sus informes en varios formatos, incluyendo PDF, Excel, HTML, texto enriquecido en archivos, XML. y CSV.

- **Marketplace.** Permite a los usuarios explorar y probar los plugins que son más relevantes para ellos. Descargar e instalar plugins desarrollados por la Comunidad Pentaho, la ampliación de las capacidades de su plataforma Pentaho.

- **Pentaho Business analytics platform:** Ofrece un enfoque moderno, simplificado, e interactiva de Pentaho permite a los usuarios de negocio acceder, descubrir y mezclar cualquier tipo de datos, independientemente de su tamaño. Con una amplia gama de herramientas de análisis cada vez más avanzadas, a partir de informes básicos a modelos de predicción, los usuarios

pueden ayudarse a sí mismos para analizar y visualizar los datos a través de múltiples medidas y dimensiones sin depender de TI.

Requerimientos. Pentaho Comunity tiene los siguientes requerimientos mínimos para funcionar:

- Procesador: CPUs: 4 cores.
- RAM: 8gb. (4 gb para Pentaho Data Integration Dedicado, 4 Servidor Apache Tomcat).
- Espacio en Disco: 80gb.

TALEND DATA INTEGRATION. Talend ofrece la integración de datos robustos en una arquitectura abierta y escalable para maximizar su valor a su negocio. Como parte de la Talend Data Tela, software de integración de datos de Talend ofrece las herramientas unificadas para integrar, limpiar, la máscara y el perfil de todos sus datos, lo que le permite convertir los datos en decisiones más rápidamente.

Aplicaciones. Talend Data Integration (TDI) posee las siguientes aplicaciones:

- **Integración ágil.** TDI permite responder más rápidamente a las peticiones de negocio sin necesidad de escribir código usando más de 900 conectores fuera de la caja, ricas herramientas gráficas basadas en Eclipse y un generador de código optimizado por rendimiento.

- **La productividad del equipo.** TDI colabora como nunca antes al usar el control de versiones de gran alcance, análisis de impacto, pruebas y depuración, gestión de metadatos y herramientas repositorio compartido.

- **Manejar con facilidad.** TDI está en la cabina del piloto de gestión de uso de las funciones de control y programación avanzada con tableros de instrumentos de integración de datos en tiempo real y control centralizado para el despliegue instantáneo a miles de nodos.

Requerimientos. TDI tiene los siguientes requerimientos:

- **Sistemas operativos.** TDI está soportado en los siguientes sistemas operativos:

    - CentOS Linux
    - OS X
    - Redhat Enterprise Linux
    - Solaris
    - SUSE Linux

    - Ubuntu Linux
    - Microsoft Windows

- **Soporte en bases de datos.** TDI Soporta bases de datos y conectividad de almacenamiento en: Amazon Aurora, Amazon RDS, Amazon Redshift, Amazon S3, AS400, DB2, Derby DB, Exasol, eXist-db, Firebird, Google Storage, Greenplum, H2, HSQLDB, Informix, Ingres, InterBase, JavaDB, JDBC, MariaDB, MaxDB, Microsoft OLE-DB, Microsoft SQL Server, MySQL, Netezza, Oracle, ParAccel, PostgresSQL, PostgresPlus, SAP Business Warehouse, SAS, SQLite, Sybase, Teradata, VectorWise, Vertica, Windows Azure Blob Storage.

INAPLEX INAPORT. Inaport proporciona datos completos de migración, integración de datos y transformación de datos de Microsoft CRM, Infor CRM (anteriormente SalesLogix), Sage CRM, GoldMine, y ACT! por Sage.

Inaport se caracteriza por la facilidad de uso, la capacidad de adaptarse de forma excepcional, y un enfoque no-código. A pesar de su amplitud y poder, sin embargo, es Inaport con un precio accesible y tiene bajos costos de propiedad, lo que garantiza una excelente relación sin compromiso.

Aplicaciones. Inaport posee las siguientes aplicaciones:

- **Data Migration.** Inaport proporciona una migración completa de los sistemas de CRM heredadas, tales como ACT! y GoldMine, a otros sistemas de CRM usando mapas estándar libres. Para las implementaciones no estándar, el poder de las funciones coincidentes y de manipulación de datos de Inaport se puede utilizar para modificar los mapas estándar y creó mapas personalizados, que proporciona una manera rentable y realista para cualquier migración.

- **Data Integration.** Inaport se puede utilizar para integrar su en las instalaciones o sistema de CRM alojado con otra empresa o fuentes externas de datos, y proporciona una transferencia de datos bidireccional. Los ejemplos típicos incluyen:
    - La integración multinacional Microsoft CRM con sistemas ERP en cuatro países
    - Compañía de software utilizando Inaport para importar clientes potenciales de Eloqua en Sage CRM para múltiples unidades de negocio
    - Gran fabricante usando Inaport para tirar de datos ERP en Infor CRM utilizando sdata

POWERCENTER. Una plataforma de integración de datos empresariales líder del mercado, escalable y de alto rendimiento que promueve la automatización, la reutilización y la agilidad. Respalda todo el ciclo de vida de integración de datos, desde el impulso del primer proyecto hasta la garantía de implantaciones empresariales de misión crítica satisfactorias.

Aplicaciones. PowerCenter posee las siguientes aplicaciones:

- **Integración:** PowerCenter constituye la base de todas sus iniciativas de integración de datos, como análisis y data warehousing, migración de aplicaciones o consolidación y gobernanza de datos.

- **Colaboración entre negocio y TI:** Negocio y TI colaboran mediante herramientas basadas en funciones y procesos agiles, lo que facilita a los negocios el autoservicio y les permite distribuir datos oportunos y fiables a su negocio.

- **Creación de prototipos, validación y perfilado rápidos:** Los analistas colaboran con TI en la creación rápida de prototipos y en la validación ágil e iterativa de los resultados.

- **Reutilización, automatización y facilidad de uso:** Los usuarios se benefician de herramientas gráficas y libres de código fáciles de utilizar, lo que favorece la automatización y la reutilización, además de permitir aprovechar un amplio abanico de transformaciones preintegradas.

- **Conectividad universal:** Los usuarios acceden a los datos y los integran sin complicaciones desde todo tipo de fuentes mediante conectores de alto rendimiento listos para usar.

- **Gestión de metadatos:** Las gráficas e intuitivas vistas basadas en los metadatos de los flujos de datos, el análisis de impacto y el linaje mejoran la gobernanza, la auditabilidad y la gestión de cambios.

- **Business Glossary:** PowerCenter ofrece una experiencia de uso adaptada al negocio para crear y gestionar términos de negocio, lo que mejora la colaboración entre negocio e TI.

- **Escalabilidad, rendimiento e inactividad del sistema nula:** Soporte para grid computing, procesamiento distribuido, alta disponibilidad, balanceo de carga adaptable, particionado dinámico y optimización de pushdown.

- **Pruebas de validación de datos automatizadas:** Pruebas de validación de datos movidos o transformados mediante un proceso libre de secuencias

de comandos, automatizado, auditable y repetible durante las fases de desarrollo, pruebas y producción.

- **Control proactivo: operaciones y gobernanza:** Un "sistema de alerta temprana" controla, alerta y gestiona problemas de producción de forma proactiva y promueve mejores prácticas de codificación para evitar un costoso control de daños más adelante.

ORACLE WAREHOUSE BUILDER. Oracle Warehouse Builder (OWB) 11g es una solución de integración de datos centrada en el almacenamiento de datos. OWB 11gR2 está pre-instalado con base de datos Oracle 11gR2, y puede ser instalado y utilizado junto con Oracle Database 10gR2 y 11gR1.

Aplicaciones. OWB se divide en los siguientes grupos de características:

- **Conjunto básico de capacidades ETL:** OWB está adecuado para la construcción de almacenes de datos simples. (También llamado ETL Core. Se corresponde aproximadamente con el conjunto de características de Warehouse Builder 10gR1).

- **Funcionalidad ETL-Advanced Enterprise:** Orientada para proyectos de almacenamiento de datos empresariales e integración de datos.

- **Conectividad:** OWB contiene adaptadores de aplicación para OWB-conectividad a las aplicaciones ERP de Oracle y SAP. (Anteriormente llamada Conectores de Aplicaciones OWB).

SQL SERVER INTEGRATION SERVICES. Microsoft Integration Services es una plataforma para la construcción de integración de datos y los datos de nivel empresarial soluciones transformaciones. Utilizar servicios de integración para resolver problemas complejos de negocios mediante la copia o la descarga de archivos, el envío de mensajes de correo electrónico en respuesta a eventos, la actualización de los almacenes de datos, limpieza y extracción de datos y la gestión de objetos y datos de SQL Server. Los paquetes pueden trabajar solos o en concierto con otros paquetes para hacer frente a las necesidades empresariales complejas. Servicios de integración pueden extraer y transformar datos de una amplia variedad de fuentes, tales como archivos XML de datos, archivos planos, y fuentes de datos relacionales, y luego cargar los datos en uno o más destinos.

Integration Services incluye un amplio conjunto de tareas incorporadas y transformaciones; herramientas para la construcción de paquetes; y el servicio Integration Services para ejecutar y administrar paquetes. Puede utilizar las herramientas gráficas de Integration Services para crear soluciones sin escribir

una sola línea de código; o se puede programar el modelo extensivo de objetos de Integration Services para crear paquetes mediante programación y tareas personalizadas de código y otros objetos de paquete.

4.4.4 ELECCIÓN DE HERRAMIENTA DE VISUALIZACIÓN DE DATOS.

Debido a la cantidad de herramientas existentes para la visualización de datos, es necesario estudiar la mejor opción para la visualización de datos existente, por ende, en este apartado se plantea una comparación entre varias herramientas existentes para la visualización.

Este documento contiene la descripción de cada una de las herramientas pre-seleccionadas, la comparación de las herramientas, los criterios de selección, la decisión final sobre la herramienta a implementar, conclusiones sobre el caso y la bibliografía. Toda la información de las aplicaciones fue sustraída de las páginas oficiales de los desarrolladores.

QLIK VIEW.

QlikView es una aplicación que permite recuperar y analizar datos provenientes de fuentes diversas. Una vez cargados en la aplicación, los datos se presentan de una forma fácil de entender y manejar.

Para efectuar selecciones en QlikView no es necesario tener experiencia previa en bases de datos o en rutinas de búsqueda: sólo hay que hacer un simple clic de ratón sobre el elemento del que se desee obtener más información. Dicho elemento se tornará de color verde, y el programa mostrará al instante todos los datos relacionados con la selección.

Se pueden crear gráficos y tablas que mejoren la visión conjunta de los datos. Se pueden imprimir todos los gráficos y tablas, o exportarlos a otros programas. Es una herramienta muy flexible que tiene las siguientes características:

Control: análisis guiado altamente personalizado y gestionado de manera centralizada.

Analítica guiada: guía a los empleados hacia el descubrimiento de información y a la correcta toma de decisiones.

Seguridad: control vertical de apps de analítica, permisos y gestión de datos.
Flexibilidad: crear las herramientas exactas que necesita para la organización y equipos.

Personalizable: apps personalizadas con scripts de QlikView y ampliar el desarrollo con QlikView Workbench.

Combinable: integrar QlikView en las aplicaciones empresariales y software de gestión de sistemas con amplias API.

Búsqueda global: búsqueda natural para desplazarse por información compleja y acelerar el descubrimiento.

Coherencia: proporciona un conjunto de datos y apps para su uso en toda la organización.

Integración de datos: unificar fuentes de datos para ofrecer una vista completa de la información; los datos y las apps gestionados de manera centralizada facilitan el descubrimiento de información de valor.

Escalabilidad empresarial: Las ventajas de una analítica guiada con gestión centralizada y segura de datos a nivel empresarial.

TABLEU.

Tableau Software ayuda a las personas a ver y comprender los datos. Tableau ayuda a todas las personas a analizar, visualizar y compartir información rápidamente. Más de 46000 cuentas de clientes obtienen resultados rápidos con Tableau en la oficina o en cualquier otro lugar. Además, decenas de miles de personas usan Tableau Public para compartir datos en sus blogs y sitios web.
Tableu ofrece múltiples funciones, entre las cuales están:

- Conexión directa para comenzar a trabajar. Conexión de datos en vivo para realizar cambios minuto a minuto.
- Mezclas perfectas de datos. Combinar varias fuentes de datos en una sola vista
- Extremadamente fácil de usar: arrastrar y soltar para crear visualizaciones enriquecidas
- Cuadros de mandos informativos e interactivos. Combinar varias vistas en un cuadro de mandos.
- Detección de datos rápida y fácil. Trabajar con bases de datos y hojas de cálculo de cualquier tamaño.
- Análisis en la web y en cualquier lugar. Desarrollar en minutos, comparta con miles de personas.

Tableu posee conectividad con:

- Actian Vectorwise
- Amazon Redshift
- Amazon Elastic MapReduce

- Cloudera Hadoop Hive e Impala
- Cisco Information Server
- DataStax
- EXASOL
- Firebird
- Google Analytics
- Google BigQuery
- Hojas de cálculo de Google
- Hortonworks Hadoop Hive
- HP Vertica
- IBM DB2
- IBM Netezza
- Kognitio
- MapR
- Marketo
- Microsoft Access
- Microsoft Excel
- Microsoft PowerPivot
- Microsoft SQL Server
- Microsoft SQL Server Analysis Services
- Microsoft SQL Server PDW
- Microsoft Windows Azure Marketplace DataMarket
- MemSQL
- MySQL
- OData
- Bases de datos Oracle
- Oracle Hyperion Essbase
- ParAccel Analytics Database
- Pivotal Greenplum
- Presto
- PostgreSQL
- Progress OpenEdge
- QuickBooks Online
- Salesforce.com, incluidos Force.com y Database.com
- SAP HANA
- SAP NetWeaver Business Warehouse
- SAP Sybase IQ
- Splunk Enterprise
- Archivos estadísticos

- Extracción de datos de Tableau
- Teradata V2
- Teradata Aster Data nCluster
- Teradata OLAP Connector
- Archivos de texto, archivos de valores separados por comas (.csv)
- Aplicaciones y bases de datos compatibles con ODBC 3.0*
- Multitud de datos web con el Conector de datos web

POWER BI.

Power BI es un conjunto de aplicaciones de análisis de negocios que permite analizar datos y compartir información. Los paneles de Power BI ofrecen a los usuarios una vista de 360 grados con sus métricas más importantes en un mismo lugar. La información se actualiza en tiempo real y está disponible en todos sus dispositivos. Con un solo clic, los usuarios pueden explorar los datos subyacentes del panel mediante herramientas intuitivas que permiten obtener respuestas fácilmente. La creación de un panel es una sencilla operación gracias a las más de 50 conexiones a conocidas aplicaciones empresariales, que se completan con paneles pregenerados y diseñados por expertos para ayudarle a ponerse en marcha rápidamente. Asimismo, puede acceder a sus datos e informes desde cualquier lugar con las aplicaciones móviles de Power BI Mobile, que se actualizan automáticamente con los cambios que se realizan en los datos.

Power BI en su versión de escritorio ofrece:

- Transformar y limpiar los datos. En ocasiones puede llevar mucho tiempo preparar los datos para su análisis. Nos hemos propuesto la misión de facilitar esta tarea. Pruebe las funciones de modelado y forma de datos de Power BI Desktop y no pierda más horas a lo largo de su atareado día.
- Ciclo de vida de análisis completo. Power BI Desktop es una elegante solución integral para la creación de análisis. Desktop tiene todas las funciones necesarias para conectar la información sobre los datos, darle forma, visualizarla y compartirla con total rapidez a través de Power BI.
- Un solo diseño visualizable desde cualquier lugar. Sabemos que debe proporcionar datos a los responsables de la toma de decisiones cuando y donde estos lo soliciten. Power BI facilita la tarea de publicar y compartir bonitos informes interactivos.

Power BI tiene un soporte inmenso de documentación con ejemplos en su página oficial, que además cuenta con ejemplos bastante entendibles y cortos que ayudan a la adaptación rápida de la herramienta.

En cuanto a las fuentes de datos, Power BI organiza los tipos de datos en las categorías siguientes:

- Todos
- Archivo
- Base de datos
- Azure
- Online Services
- Otros

La categoría Todos incluye todos los tipos de conexión de datos de todas las categorías.

La categoría Archivo proporciona las siguientes conexiones de datos:

- Excel
- CSV
- XML
- Texto
- JSON
- Carpeta
- Carpeta de SharePoint

La categoría Base de datos proporciona las siguientes conexiones de datos:

- Base de datos de SQL Server
- Base de datos de Access
- Base de datos de SQL Server Analysis Services
- Base de datos de Oracle
- Base de datos IBM DB2
- Base de datos Informix de IBM (beta)
- Base de datos de MySQL
- Base de datos de PostgreSQL
- Base de datos de Sybase
- Base de datos de Teradata
- Base de datos SAP HANA
- SAP Business Warehouse
- Amazon Redshift (Beta)
- Impala (Beta)
- Snowflake (Beta)

La categoría Azure proporciona las siguientes conexiones de datos:

- Base de datos SQL de Microsoft Azure
- Almacenamiento de datos SQL de Microsoft Azure

- Microsoft Azure Marketplace
- HDInsight de Microsoft Azure
- Almacenamiento de blobs de Microsoft Azure
- Almacenamiento de tabla de Microsoft Azure
- Azure HDInsight Spark (Beta)
- Microsoft Azure DocumentDB (Beta)
- Microsoft Azure Data Lake Store (Beta)

La categoría Online Services proporciona las siguientes conexiones de datos:

- Lista de SharePoint Online
- Microsoft Exchange Online
- Dynamics CRM Online
- Facebook
- Google Analytics
- Objetos de Salesforce
- Informes de Salesforce
- appFigures (Beta)
- Azure Enterprise (Beta)
- comScore Digital Analytix (beta)
- GitHub (Beta)
- MailChimp (Beta)
- Marketo (Beta)
- Planview Enterprise (beta)
- QuickBooks Online (Beta)
- SparkPost (Beta)
- Smartsheet
- SQL Sentry
- Stripe (Beta)
- SweetIQ (Beta)
- Troux (beta)
- Twilio (Beta)
- tyGraph (Beta)
- Webtrends (Beta)
- Zendesk (Beta)

La categoría Otros proporciona las siguientes conexiones de datos:

- Web
- Lista de SharePoint
- Fuente de OData
- Archivo Hadoop (HDFS)

- Active Directory
- Microsoft Exchange
- ODBC
- Script R
- Spark (Beta)
- Consulta en blanco

IBM WATSON ANALYTICS.

Watson Analytics ofrece los beneficios de la analítica avanzada sin la complejidad. Un servicio de detección inteligente de datos disponibles en la nube, que guía a la exploración de datos, automatiza el análisis predictivo y permite crear tableros de mando sin esfuerzo y la creación infográfica.
Puede obtener respuestas y nuevos puntos de vista para tomar decisiones fiables en minutos a todos por su cuenta.
Tanto si necesita detectar rápidamente una tendencia como si tiene un equipo que necesita visualizar los datos de un informe en un panel de control. Permite descubrir fácilmente patrones y significados en sus datos, por su cuenta, con el análisis estadístico y la visualización inteligente de los datos. El descubrimiento guiado de datos, la analítica predictiva automatizada y el diálogo de lenguaje natural permiten interactuar con los datos y obtener respuestas que comprenda. La herramienta permite:

- Realizar descubrimientos con sus propias palabras: permite hacer una pregunta y obtener información de valor que pueda comprender. Añadir sus datos en Watson Analytics y conseguir prácticamente al instante una lista de puntos de partida relevantes.
- Comprenda qué impulsa su negocio: permite conocer aquellos factores que probablemente incidirán en los resultados de negocio y las matemáticas que hay detrás de las conclusiones.
- Utilice paneles de control para presentar sus ideas: elegir una plantilla y añadir las visualizaciones y la información de valor que desea que comprenda su público.

5. METODOLOGÍA

5.1 TIPO DE TRABAJO

El proyecto “Comparativo de metodologías y herramientas para el desarrollo de un Data Warehouse”, corresponde a un proyecto de desarrollo tecnológico por lo cual, está encaminado a resolver problemas prácticos, a través de una evaluación del proyecto en mención. Por la naturaleza es una investigación formativa en razón de que busca analizar el problema mediante la interpretación y comprensión de los procesos y resultados de la aplicación de un DW. El proyecto está avalado por el Grupo de Investigación y desarrollo en informática y telecomunicaciones en su línea de Gestión del conocimiento.

5.2 PROCEDIMIENTO

5.2.1 Fase 1. Identificación Documental. Se Obtuvo una idea del desarrollo y las características de los procesos y también de disponer de información de lo que se desea hacer en el proyecto. Comprende las actividades:

- **Actividad 1. Revisar documentación.** En esta actividad se hizo una recopilación de información confiable, proveniente de diferentes fuentes, usando buscadores científicos y base de datos. Para así obtener la información necesaria como base para este proyecto, haciendo énfasis en que su contenido tenga similitud y conceptos que apoyen el proyecto.

- **Actividad 2. Abstraer documentación.** En esta actividad se hizo una recopilación de información confiable, proveniente de diferentes fuentes, usando buscadores científicos y base de datos. Para así obtener la información necesaria como base para este proyecto, haciendo énfasis en que su contenido tenga similitud y conceptos que apoyen el proyecto Abstraer información. Se realizó un filtrado de información según la concordancia con respecto al proyecto a realizar y además, se identificaron antecedentes para el apoyo bibliográfico al proyecto.

5.2.2 Fase 2. Revisión y comparación de metodologías para el diseño y control de Data Warehouse. Se realizó un estudio de las metodologías para la construcción del DW mediante una comparación de las metodologías existentes. Comprende las actividades:

- **Actividad 1. Seleccionar las metodologías.** En esta actividad se definieron las metodologías para la definición del Data Warehouse, haciendo un filtrado entre diferentes metodologías disponibles en el mercado.

- **Actividad 2. Análisis de metodologías.** En esta actividad se estudiaron las metodologías para la construcción del Data Warehouse, haciendo un análisis comparativo entre diferentes metodologías disponibles en el mercado.

- **Actividad 3. Selección de metodología.** En esta actividad se definió la metodología para la construcción del Data Warehouse, basado en el análisis comparativo entre diferentes metodologías disponibles en el mercado.

5.2.3 Fase 3. Descripción de los lineamientos de diseño. Se obtuvo la descripción de los lineamientos de diseño a través de las restricciones de desempeño y almacenamiento de su sistema. Comprende las actividades:

- **Actividad 1. Diseñar lineamientos.** Fragmentar datos históricos o los que se van a almacenar, el diseñador define el estilo de diseño para el DW e indica requerimientos de desempeño y almacenamiento.

- **Actividad 2. Indicar el conjunto de cubos que serán implementados y almacenados físicamente en el DW.** Para que la materialización corresponda con el esquema conceptual, indicar al menos un cubo que materialice cada relación dimensional.

- **Actividad 3. Indicar niveles de almacenamiento.** Se indicarán para cada dimensión, qué niveles desea almacenar juntos, conformando una fragmentación de los niveles de la dimensión, restricciones de desempeño y almacenamiento de su sistema.

5.2.4 Fase 4. Revisión y comparación de las herramientas para el diseño y control de Data Warehouse. Se realizó un estudio de las herramientas para el control y diseño del DW mediante una comparación de las herramientas existentes. Comprende las actividades:

- **Actividad 1. Seleccionar las herramientas.** En esta actividad se definieron las herramientas para el diseño y control del Data Warehouse, haciendo un filtrado entre diferentes herramientas disponibles en el mercado.

- **Actividad 2. Análisis de herramientas.** En esta actividad se estudiaron las herramientas para la construcción del Data Warehouse, haciendo un análisis comparativo entre diferentes herramientas disponibles en el mercado.

- **Actividad 3. Selección de herramienta.** En esta actividad se definió la herramienta para la construcción del Data Warehouse, basado en el análisis comparativo entre diferentes herramientas disponibles en el mercado.

5.2.5 Fase 5. Revisión y comparación de las herramientas para la visualización de datos. Se realizó un estudio de las herramientas de

visualización de datos mediante una comparación de las herramientas existentes. Comprende las actividades:

- **Actividad 1. Seleccionar las herramientas.** En esta actividad se definieron las herramientas de visualización de datos, haciendo un filtrado entre diferentes herramientas de visualización de datos disponibles en el mercado.

- **Actividad 2. Análisis de herramientas.** En esta actividad se estudiaron las herramientas de visualización de datos, haciendo un análisis comparativo entre diferentes herramientas de visualización de datos disponibles en el mercado.

- **Actividad 3. Selección de herramienta.** En esta actividad se definió la herramienta para la construcción del Data Warehouse, basado en el análisis comparativo entre diferentes herramientas disponibles en el mercado.

6. RESULTADOS

6.1. SELECCIÓN METODOLOGÍA PARA EL DESARROLLO DE UN DATA WAREHOUSE. Hace referencia a los factores que afectan la decisión para elegir la metodología de implementación del DW, con una evaluación de alta, media o bajo según los siguientes criterios:

Nivel de implementación: Hace referencia al porcentaje del proyecto que abarca la metodología, si se trata de solo el diseño del DW o de todo el proceso desde el análisis del negocio.

Documentación: Hace referencia a la cantidad de documentos con información confiable que acredita la metodología.

Uso en el mercado: Hace referencia a la usabilidad de la metodología en el mercado y su popularidad, para identificar si ha satisfecho las necesidades del mercado y tener como referencia experiencias que ya se han tenido con la metodología.

Costos: Hace referencia al nivel de inversión monetaria que implica implementar la metodología para la organización.

Nivel de detalle: Hace referencia a la completitud de la metodología para desarrollar un Data Warehouse.

Metodología	Nivel de implementación	Documentación	Uso en el mercado	Costos	Nivel de detalle
Ralph Kimball	Medio	Alta	Alto	Medio	Alto
CRISP-DM	Bajo	Medio	Medio	Medio	Alto
Hefesto	Alto	Bajo	Bajo	Medio	Alto

Tabla 1 Tabla comparativa de metodologías para el desarrollo de un Data Warehouse

6.1.1 METODOLOGÍA PARA EL DESARROLLO DE UN DATA WAREHOUSE.

Debe ser pertinente haber delimitado y clasificado el proyecto antes de llegar a una elección concreta acerca de la metodología que se piensa implementar. Lo que evidencian las referencias y los antecedentes, sugieren que las tres metodologías pueden usarse para diferentes tipos proyectos que varían por su tamaño y complejidad, aunque dentro de las metodologías se repitan procesos. No hay mejor metodología que otra, existe la más adecuada, el punto de referencia sería el tamaño del proyecto. Sin embargo, es recomendable usar la metodología

de Kimball, esto, debido a la gran cantidad de documentación que hay al respecto y al nivel medio de implementación, que permite abarcar más que la construcción el DW y sin mencionar que es mucho más usada en el mercado, lo cual facilita la búsqueda de antecedentes y proyectos similares que faciliten el proyecto que se quiera realizar, ya sea en pequeña o gran escala.

6.2 DESCRIPCIÓN DE LOS LINEAMIENTOS DE DISEÑO.

Los lineamientos de diseño se pueden definir como la información del diseño lógico que complementan al esquema conceptual y permiten al diseñador dar pautas sobre el esquema lógico deseado para el DW[25].

> La construcción de un DW puede verse como una secuencia de tres etapas fundamentales: diseño conceptual, diseño lógico y diseño físico.
>
> Durante la etapa de diseño conceptual se realiza una abstracción de la realidad basados en los objetos del negocio y los requerimientos de información. El resultado de esta etapa es un esquema conceptual que especifica el problema a resolver.
>
> El esquema lógico es una especificación más detallada que el esquema conceptual donde se modelan los datos según un modelo lógico de DBMS (por ejemplo el modelo relacional), definiéndose por lo tanto, estructuras más cercanas al nivel de almacenamiento. Durante la etapa de diseño lógico se construye el esquema lógico teniendo en cuenta no sólo el esquema conceptual, sino también estrategias para resolver los requerimientos de performance y almacenamiento.
>
> En el caso de diseño de DWs se debe tener en cuenta un componente adicional: las bases de datos fuentes. Un DW se construye con información extraída de un cierto conjunto de bases de datos fuentes. Durante el diseño lógico deben considerarse estas bases y cómo se corresponden con el esquema conceptual.
>
> Durante la etapa de diseño físico se incorporan elementos específicos de almacenamiento y performance, como son la elección de índices, almacenamiento especializado, parámetros de sistemas, etc[26].

DESCRIBIR UN ESQUEMA RELACIONAL.

Debido a las diferencias entre los esquemas existentes para la construcción de un DW y la repercusión que tienen según la lógica del negocio, es necesario estudiar

[25] PERALTA, Verónika. Diseño Lógico de Data Warehouses a partir de Esquemas Conceptuales Multidimensionales. Montevideo, 2001, p. 24. Maestría en informática. Universidad de la república. Facultad de ingeniería.
[26] Ibíd., p. 17.

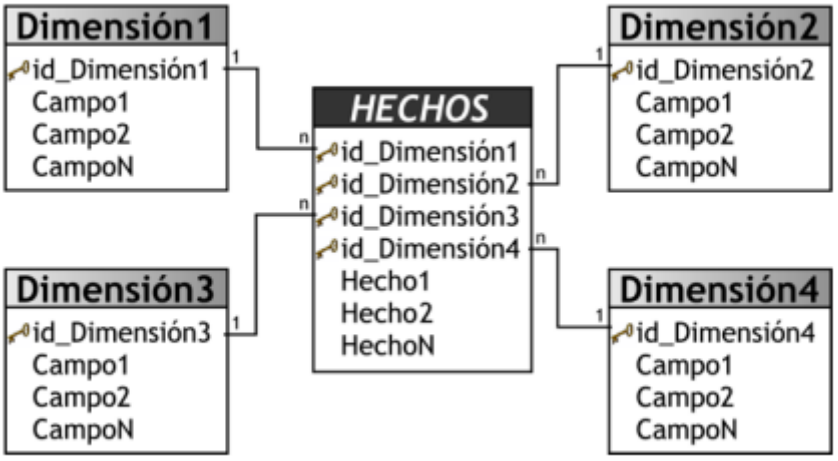
la mejor opción para la creación del DW, por ende, en este documento se plantea una comparación entre varios esquemas existentes para la implementación de DW.

Este apartado contiene la descripción de cada una de los esquemas relacionales, la comparación de estos, los criterios de selección, la decisión final sobre el esquema a implementar en la UM.

ESQUEMA EN ESTRELLA (STAR SCHEME):

Para Bernebeu el esquema en estrella está conformado por dos componentes los cuales son la tabla de hechos y las tablas de dimensiones que están relacionados a esta y que tienen sus claves respectivas[27], para identificar mejor el esquema se muestra la siguiente figura:

Figura 4. Representación esquema estrella.



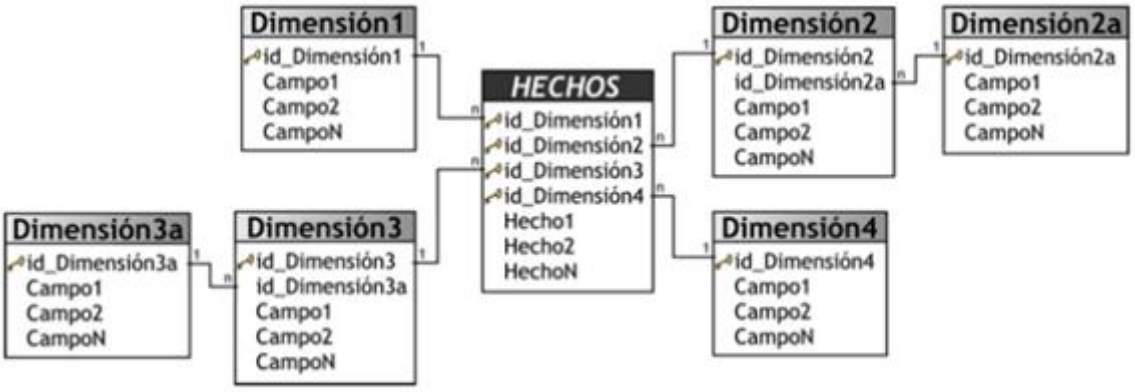
Fuente: UNIVERSIDAD PARA LA COOPERACIÓN INTERNACIONAL. Data Warehousing y metodología Hefesto. [en línea]. Córdoba: Argentina. 2007. Sp. Fecha de consulta: 21/11/2016. Disponible en: http://www.ucipfg.com/Repositorio/MATI/MATI-10/BLOQUE-ACADEMICO/Unidad-03/lecturas/bibliografia/Data%20Warehousing%20y%20metodolog%C3%ADa%20Hefesto%201.pdf

ESQUEMA COPO DE NIEVE (SNOWFLAKE SCHEME):

Este esquema representa la amplificación del modelo en estrella cuando las tablas de dimensiones se organizan por jerarquías de dimensiones[28], en la siguiente figura se muestra la composición de este esquema:

[27] BERNEBEU, Op. Cit., SP.
[28] Ibíd., Sp.

Figura 5. Representación esquema copo de nieve.



Fuente: UNIVERSIDAD PARA LA COOPERACIÓN INTERNACIONAL. Data Warehousing y metodología Hefesto. [en línea]. Córdoba: Argentina. 2007. Sp. Fecha de consulta: 21/11/2016. Disponible en: http://www.ucipfg.com/Repositorio/MATI/MATI-10/BLOQUE-ACADEMICO/Unidad-03/lecturas/bibliografia/Data%20Warehousing%20y%20metodolog%C3%ADa%20Hefesto%201.pdf

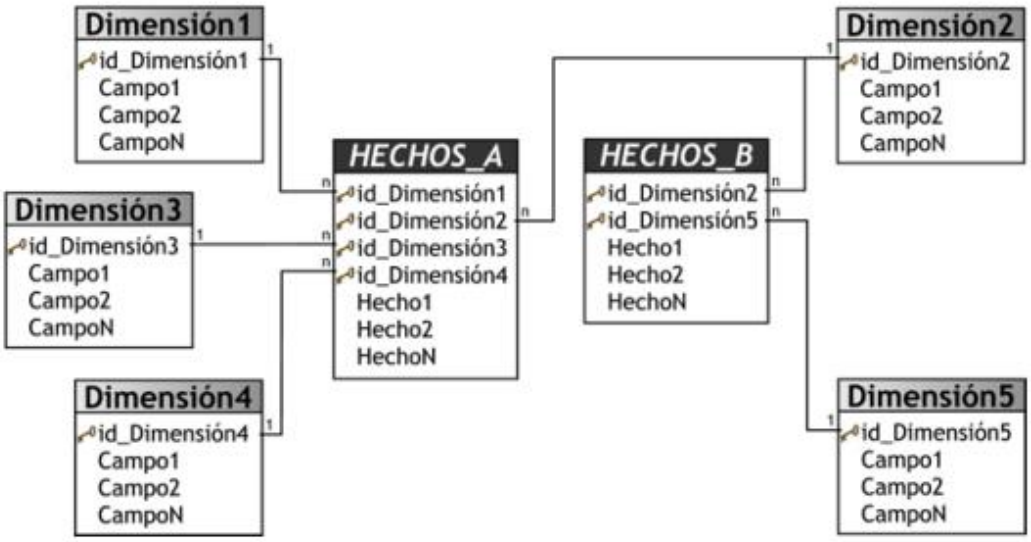
ESQUEMA CONSTELACIÓN O COPO DE ESTRELLAS (STARFLAKE SCHEME):

Este esquema es una composición de esquemas en estrella, comprende una tabla de hechos principal y por tablas de hechos auxiliares, estas tablas tienen sus respectivas tablas de dimensiones.

Las tablas de hechos no necesariamente tienen que compartir las mismas tablas de dimensiones, debido a que las tablas de hechos auxiliares pueden estar vinculas a solo algunas de las tablas de dimensiones que están asignadas a la tabla de hechos principal, y a su vez pueden hacerlo con nuevas tablas de dimensiones[29]. La figura que se muestra a continuación muestra la estructura de un esquema en constelación:

[29] Ibíd., Sp.

Figura 6. Representación esquema en constelación.



Fuente: UNIVERSIDAD PARA LA COOPERACIÓN INTERNACIONAL. Data Warehousing y metodología Hefesto. [en línea]. Córdoba: Argentina. 2007. Sp. Fecha de consulta: 21/11/2016. Disponible en: http://www.ucipfg.com/Repositorio/MATI/MATI-10/BLOQUE-ACADEMICO/Unidad-03/lecturas/bibliografia/Data%20Warehousing%20y%20metodolog%C3%ADa%20Hefesto%201.pdf

6.1.3 CRITERIOS DE SELECCIÓN. Hace referencia a los factores que afectan la decisión para elegir el tipo de esquema para la construcción del DW, con una evaluación de alta, media o bajo según los siguientes criterios:

Nivel de complejidad
Hace referencia al nivel de detalle del proceso para el desarrollo del esquema y su relación con el tiempo.

Utilización de espacio
Hace referencia a la cantidad de almacenamiento físico que ocupa el esquema.

Tiempo de respuesta
Hace referencia al tiempo con el cual el esquema responde a las operaciones.

Soporte para herramientas de consulta y análisis
Hace referencia a la cantidad de compatibilidades con motores de búsqueda y herramientas de análisis.

Esquema	Nivel de complejidad	Utilización de espacio	Tiempo de respuesta	Soporte
Esquema en estrella	Medio	Bajo	Alto	Medio
Esquema copo de nieve	Medio	Medio	Medio	Medio
Esquema en constelación	Alto	Medio	Medio	Alto

Tabla 2 Tabla comparativa de los esquemas para el desarrollo de un Data Warehouse

ELECCIÓN DE ESQUEMA PARA LA CONSTRUCCIÓN DEL DW.

Con base en los criterios anteriormente establecidos, se llega a la selección de usar el esquema en constelación ya que sus características dejan en evidencia que es un esquema diseñado para soluciones medianas o grandes y al tener mayor complejidad y completitud debido al hecho de tener varias tablas de hechos, se adapta mejor al proyecto.

MATERIALIZACIÓN DE RELACIONES

Siguiendo el proceso para la descripción de los lineamientos de diseño, se debe materializar las relaciones. En CMDM (En CMDM se definen los conceptos de nivel, dimensión y relación dimensional y se presenta un lenguaje para especificar restricciones de integridad[30]) una relación dimensional representa un espacio de cubos resultante de cruzar niveles de las dimensiones. Dicho espacio de cubos puede limitarse mediante las restricciones de integridad del propio modelo.

Las restricciones se construyen en base a predicados con cuantificadores (∀ , $, ¬) para indicar que “todos los cubos deben tener”, o “debe existir un cubo que tenga” o “ningún cubo debe tener” dicha estructura. Por ejemplo: debe existir un cubo que cruce el nivel mes con el nivel producto.

Estas restricciones sugieren qué cubos sería interesante tener y cuáles no deberían existir. Sin embargo, la decisión de cuáles de esos cubos se deben materializar debe ser tomada en un tiempo después. En este contexto, materializar un cubo corresponde a precalcular los valores para los cruces de las dimensiones y almacenarlos en una tabla. Luego se pueden obtener otros cruces mediante operaciones efectuadas sobre estos.

[30] CARPANI, Fernando & RUGGIA, Raul. An Integrity Constraints Language for a Conceptual Multidimensional Data Model. [en línea]. Montevideo, 2001, Sp. Universidad de la república. Facultad de ingeniería. [consulta: 23/11/2016] Disponible en: <https://www.fing.edu.uy/inco/grupos/csi/esp/Publicaciones/2001/seke2001-fc.pdf>

En CMDM la definición de un cubo se hace mediante un macro donde se indica los niveles y las medidas del cubo. Un cubo debe tener por lo menos un nivel de cada dimensión, uno de ellos con la función de medida.

Esta característica puede ser demasiado restrictiva y exceptúa algunos casos que pueden ser de interés y que se necesita una definición de cubos más flexible, donde además de los cubos que se pueden especificar en CMDM, se puedan representar:

- Cubos que no tengan medidas. Esto se hace para representar únicamente los cruzamientos.
- Cubos que omitan algunas de las dimensiones de la relación dimensional. Al sumarizar la dimensión, no se obtiene el nivel, sino simplemente el total de la dimensión.

En algunas dimensiones puede interesar mantener más de un nivel de detalle. Esto último tiene sentido si los niveles son de diferentes jerarquías, pero formalmente no se restringe.

Como lineamiento, se debe indicar el conjunto de cubos que serán implementados y almacenados físicamente en el DW. Para que la materialización corresponda con el esquema conceptual, se debe indicar al menos un cubo que materialice cada relación dimensional[31].

FRAGMENTACIÓN VERTICAL DE DIMENSIONES

Se debe indicar el grado de normalización o normalización que se quiere lograr al generar estructuras relacionales para cada dimensión. Se puede tratar de diferente manera por cada dimensión, indicando para cada una si se normaliza, desnormaliza o efectúa una estrategia intermedia y en caso de usar la estrategia intermedia se debe indicar qué niveles quedan en la misma tabla.

Como lineamiento, se debe indicar para cada dimensión qué niveles se desea almacenar juntos, conformando una fragmentación de los niveles de la dimensión. Para que se justifique una fragmentación los niveles de cada fragmento deben estar relacionados jerárquicamente. Es decir, si se considera el grafo que representa a la jerarquía de la dimensión, el subgrafo inducido por los niveles del fragmento debe ser conexo. Si dos niveles de un mismo fragmento no están relacionados, ni directa ni transitivamente, entonces conforman un cruzamiento y se pierde la relación jerárquica de la dimensión[32].

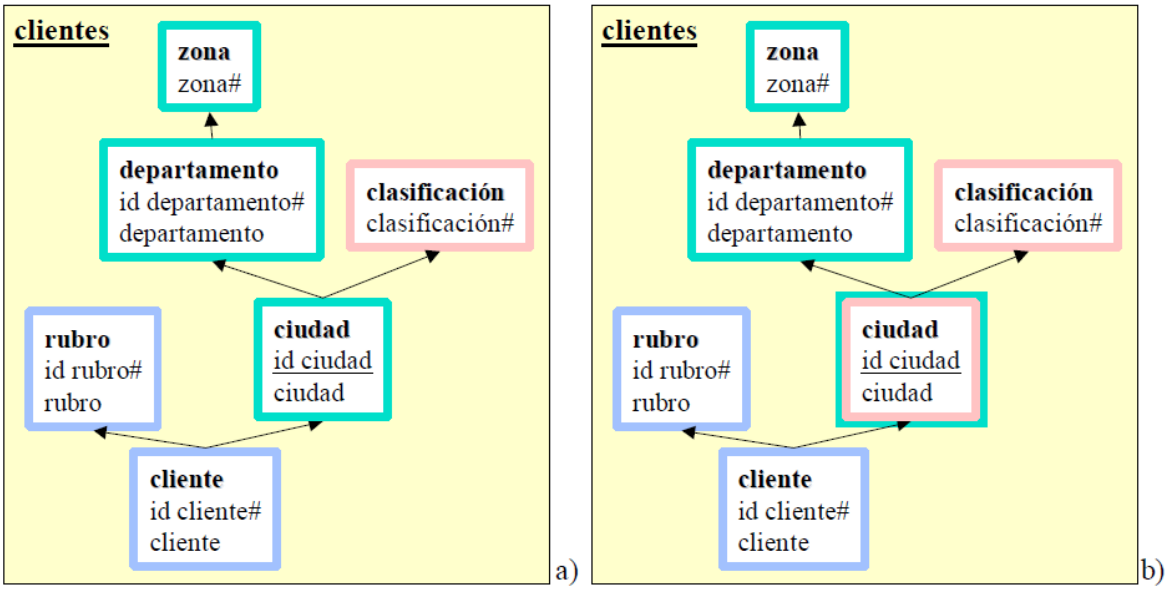
Es importante conocer que:

[31] PERALTA. Op. Cit., p. 24.
[32] Ibíd., p.25.

> Un fragmento es un subconjunto de los niveles de la dimensión, que conforman un sub-grafo conexo de su jerarquía. Un conjunto completo de fragmentos cumple que cada nivel está al menos en uno de los fragmentos.
>
> Una fragmentación dimensional es una función que a cada dimensión le asocia un conjunto completo de fragmentos. La función la define por extensión el diseñador. Un fragmento es un subconjunto de los niveles de la dimensión, que conforman un sub-grafo conexo de su jerarquía. Un conjunto completo de fragmentos cumple que cada nivel está al menos en uno de los fragmentos.
>
> Una fragmentación dimensional es una función que a cada dimensión le asocia un conjunto completo de fragmentos. La función la define por extensión el diseñador[33].

La fragmentación se puede representar gráficamente como una coloración de los niveles. Los niveles de un mismo fragmento se recuadran con el mismo color. Un nivel tiene varios colores si está en más de un fragmento. La fragmentación es completa si todos los niveles tienen color. En la siguiente figura la fragmentación tiene 3 fragmentos disjuntos: rubro y cliente (celeste), zona, departamento y ciudad (verde) y clasificación (rosa). Los fragmentos de la figura no son disjuntos ya que el nivel ciudad pertenece al fragmento verde y al rosa[34].

Figura 7. Fragmentación vertical de dimensiones.



Fuente: UNIVERSIDAD DE LA REPÚBLICA. Diseño Lógico de Data Warehouses a partir de Esquemas Conceptuales Multidimensionales. Montevideo: Uruguay. 2001. p. 26. Fecha de consulta: 23/11/2016.

[33] Ibíd., p. 26.
[34] Ibíd., p.26.

FRAGMENTACIÓN HORIZONTAL DE DIMENSIONES

Representar un cubo en el modelo relacional puede dar como resultado una o varias tablas (fact tables) dependiendo del grado de fragmentación que se quiera lograr.

Fragmentar horizontalmente una tabla relacional corresponde a construir varias tablas con la misma estructura, y dividir las instancias entre ellas. Con esto se logra almacenar juntas las tuplas que son consultadas juntas y tener tablas más pequeñas, lo cual resulta en un aumento de la performance en las consultas.

Para la fragmentación horizontal de bases de datos distribuidas, se proponen dos propiedades que debe cumplir la fragmentación horizontal: completitud y disjuntez. La propiedad de completitud exige que cada tupla esté en alguno de los fragmentos, y la propiedad de disjuntez exige que cada tupla que esté en a lo sumo un fragmento. Como ejemplo se considera la siguiente fragmentación:

mes ³ ene-2000
mes ³ ene-1997 ^ mes < ene-1999
mes < ene-1998

La fragmentación no es completa, ya que las tuplas correspondientes a 1999 no pertenecen a ningún fragmento; y no es disjunta, ya que las tuplas correspondientes a 1997 pertenecen a los dos fragmentos. En el contexto de bases de datos distribuidas, estas propiedades deben cumplirse para asegurar la completitud de los datos consultados, minimizando la redundancia. En el contexto de DW, es importante considerar estas propiedades, pero no es necesario que se cumplan. Si no se cumple la propiedad de disjuntez se obtiene redundancia, pero esto no es necesariamente un problema en un DW. Por ejemplo, es común mantener tablas de hechos con toda la historia, y tablas de hechos para los datos del último año. Respecto a la completitud, si bien interesa evitar la pérdida de información a nivel global, no siempre debe cumplirse localmente en la fragmentación de cada cubo.

Cubo mensual

- mes ≥ ene-2000
- mes < ene-2000

Cubo anual

- mes ≥ ene-2000

La fragmentación del cubo mensual es completa, pero la del cubo anual sólo tiene información de los últimos años. Sin embargo, las tuplas que no pertenecen al fragmento del cubo anual, pueden obtenerse a partir de los fragmentos del cubo mensual, por lo que no hay pérdida de información[35].

[35] Ibíd., p. 27.

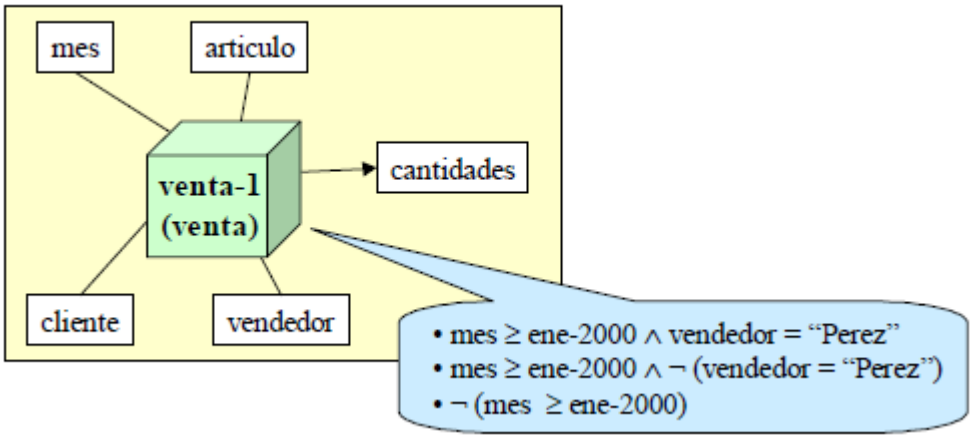
Cuando hay redundancia, la completitud de las fragmentaciones debe ser considerada globalmente, se sugiere tenerlas en cuenta durante la definición de la fragmentación.

Para definir una fragmentación se debe indicar el conjunto de franjas a utilizar. Las franjas se expresan en términos de los ítems de los niveles del cubo, y se expresarán en términos de atributos de tablas en una etapa posterior. Se pueden usar las franjas construidas o construirlas con sus propios criterios.

Cabe mencionar que una fragmentación de cubos es una función que para cada cubo indica el conjunto de franjas en que se fragmenta.

Para representar gráficamente la fragmentación se puede presentar como un bloque de llamada a partir del cubo. Dentro del bloque se escriben las franjas. La Figura 8 muestra un conjunto de franjas asociado al cubo venta-1. Las dos primeras franjas determinan las ventas posteriores a enero de 2000 del vendedor Pérez y del resto de los vendedores (menos Pérez). La tercera franja determina las ventas anteriores a ene de 2000 de cualquier vendedor. La fragmentación es completa y disjunta[36].

Figura 8. Fragmentación horizontal de cubos.



Fuente: UNIVERSIDAD DE LA REPÚBLICA. Diseño Lógico de Data Warehouses a partir de Esquemas Conceptuales Multidimensionales. Montevideo: Uruguay. 2001. p. 28. Fecha de consulta: 23/11/2016.

DEFINICIÓN DE LINEAMIENTOS

Se debe definir en forma explícita:

[36] Ibíd., p.28.

El conjunto SchCubes, que indica qué cubos se desean implementar. (conjunto de cubos)
La función SchDFragmentation, que indica cómo se fragmentan verticalmente las dimensiones. (las funciones que fragmentan las dimensiones)
La función SchCFragmentation, que indica cómo se fragmentan horizontalmente los cubos. (las funciones que fragmentan los cubos)

Para definir los lineamientos el diseñador usualmente se basa en las restricciones de performance y almacenamiento de su sistema.

El conjunto SchCubes debe tener al menos un cubo por cada relación dimensional y con el máximo detalle para no perder información. Si se tienen limitaciones de espacio se puede implementar sólo un cubo de cada relación. Si en cambio el espacio físico no es un problema y se quiere priorizar la performance en las consultas, se puede materializar otros cubos para los cruzamientos más frecuentes o más exigentes. En general es inviable materializar todos los cruzamientos posibles, por lo tanto la elección de los cubos a materializar resulta en un compromiso entre costos de almacenamiento y tiempos de respuesta.

La forma de fragmentar las dimensiones tiene que ver con el estilo de diseño que se quiere dar al DW. Para lograr un esquema estrella se define un único fragmento por cada dimensión (con todos los niveles), en cambio, para un esquema snowflake se define un fragmento por cada nivel. La denormalización logra mejores tiempos de respuesta en las consultas ya que no es necesario realizar el join de varias tablas, pero tiene el máximo grado de redundancia. La redundancia es controlable si los datos de las dimensiones cambian lentamente, pero aumenta el tamaño de las tablas y por tanto degrada la performance cuando los datos son volátiles y se quieren conservar las distintas versiones. La normalización tiene los efectos contrarios. Las estrategias intermedias hacen un compromiso entre tiempos de respuesta y el control de redundancia.

En la fragmentación de los cubos intervienen dos factores: la cantidad de datos y los rangos de datos consultados juntos. Cuantas más franjas se tengan, cada una será de menor tamaño y por tanto se ganará en tiempos de respuesta. Por otro lado, si las consultas involucran datos de varias franjas, deben realizarse operaciones de drill-across (combinar datos de varios cubos) con efectos peores en los tiempos de respuesta. La decisión debe basarse en los requerimientos, observando qué datos son consultados juntos (afinidad), por ejemplo, qué rangos de fechas se consultan más frecuentemente. Las fragmentaciones más comunes se realizan por rangos de fechas o zonificaciones.

Se pueden definir estrategias por defecto que automaticen la definición de los lineamientos. Por ejemplo, una estrategia posible es construir siempre un esquema estrella con un único cubo por relación y una única franja por cubo. Otra estrategia podría consistir en fragmentar las dimensiones con más de 4 niveles, en fragmentos de 3 niveles cada uno, y crear franjas quinquenales para los cubos.

Podrían estudiarse funciones de costos y heurísticas que automaticen la definición de lineamientos de acuerdo a los factores mencionados: limitaciones de espacio físico, performance en las consultas, afinidad, cantidad de datos[37].

6.2.1 DESCRIPCIÓN DE LOS LINEAMIENTOS DE DISEÑO

Los lineamientos de diseño son un proceso necesario en todo proyecto que requiera el desarrollo de un DW, se tienen que dejar muy claras las condiciones en las cuales se desarrolla el DW. En cuanto al tipo de esquema, que hace parte importante en la descripción de los lineamientos de diseño, de acuerdo con la tabla 2, se recomienda bastante que se considere como la opción principal el esquema de constelación, esto debido a que sus características son mucho más específicas, es decir, aunque el proceso sea minucioso, se considera como un esquema muy organizado, fácil de referenciar y de describir, además, permite un tiempo de respuesta adecuado y una óptima utilización de espacio en el disco, aunque esta elección puede variar según el tamaño del proyecto que se quiera realizar.

6.3 CRITERIOS DE SELECCIÓN HERRAMIENTA DE INTEGRACIÓN DE DATOS. Hace referencia a los factores que afectan la decisión para elegir la herramienta de implementación del DW, con una evaluación de alta, media o bajo según los siguientes criterios:

Tipo de licencia: Hace referencia del software para identificar si es de pago, versión de prueba o libre.

Costo: Hace referencia al precio en el mercado que tiene el software por su licencia.

Facilidad de uso: Hace referencia a la usabilidad de la herramienta en el mercado, donde la interfaz sea intuitiva para el usuario.

Soporte: Hace referencia a la documentación existente y manuales que tiene el software.

Requerimientos de hardware: Hace referencia a las características que necesita el software para su correcto funcionamiento.

Conectividad: Hace referencia a la compatibilidad para conectarse con una base de datos en específico.

[37] Ibíd., p.29.

Herramienta	Tipo de licencia	Costo	Facilidad de uso	Soporte	Hardware	Conexión
Pentaho Community	Libre	No aplica	Alta	Alto	Medio	Alta
Talend Data Integration	Libre	No aplica	Medio	Medio	Medio	Medio
Inaplex Inaport	Libre	No aplica	Bajo	Bajo	Bajo	Bajo
PowerCenter	Pago	Alto	Alto	Alto	Alto	Alto
Oracle Warehouse Builder	Pago	Alto	Medio	Alto	Alto	Medio
SQL Server Integration Services	Prueba/ Pago	Alto	Alto	Alto	Alto	Medio

Tabla 3 Tabla comparativa de herramientas de integración de datos para el desarrollo de un Data Warehouse

6.3.1 HERRAMIENTA DE INTEGRACIÓN DE DATOS

De acuerdo a la tabla 3 y asumiendo que el proyecto no cuente con recursos para pagar licencias de software, se recomienda bastante el uso de Pentaho community, debido a su gran facilidad de uso y al gran soporte, tanto a nivel documental, como a nivel de conexiones con motores de bases de datos. Incluso con disponibilidad de recursos para licencias, es bastante recomendable adquirir la licencia de Pentaho para este tipo de proyectos.

6.4 CRITERIOS DE SELECCIÓN HERRAMIENTA DE VISUALIZACIÓN DE DATOS. Hace referencia a los factores que afectan la decisión para elegir la herramienta de visualización de datos, con una evaluación de alta, media o bajo según los siguientes criterios:

Tipo de licencia. Hace referencia del software para identificar si es de pago, versión de prueba o libre.

Costo. Hace referencia al precio en el mercado que tiene el software por su licencia.

Facilidad de uso. Hace referencia a la usabilidad de la herramienta en el mercado, donde la interfaz sea intuitiva para el usuario.

Soporte. Hace referencia a la documentación existente y manuales que tiene el software.

Requerimientos de hardware. Hace referencia a las características que necesita el software para su correcto funcionamiento.

Conectividad. Hace referencia a la compatibilidad para conectarse con fuentes de datos.

Número de visualizaciones. Hace referencia al número de figuras disponibles para graficar datos.

Herramienta	Tipo de licencia	Costo	Facilidad de uso	Soporte	Requerimientos de hardware	Conectividad	Número de visualizaciones
Qlik View	Prueba/ pago	Alto	Media	Medio	Medio	Alto	Alto
Tableu Public	Prueba/ pago	Alto	Baja	Bajo	Medio	Alto	Alto
Power BI	Libre/Pago	No aplica/ Alto	Alta	Alto	Medio	Alto	Alto
IBM Watson Analytics	Libre/Pago	No aplica/ Alto	Baja	Bajo	Medio	Medio	Medio

Tabla 4 Tabla comparativa de herramientas para la visualización de datos

6.4.1 HERRAMIENTA DE VISUALIZACIÓN DE DATOS

Concorde a la tabla 4, se recomienda mucho la utilización de Power BI como herramienta de visualización, es una herramienta con un soporte increíble a nivel documental y práctico brindado por Microsoft. En su modalidad libre, ofrece una cantidad importante de conexiones con motores de bases de datos y es una herramienta flexible en cuanto a la cantidad y diversidad de visualizaciones.

7. CONCLUSIONES

- La metodología de Kimball supera ampliamente a CRISP-DM y a Hefesto en el apartado de documentación, el cual es vital para un correcto desarrollo de un DW. Es una metodología muy descriptiva y se adapta a la mayoría de proyectos, sin importar su naturaleza o tamaño.

- Los lineamientos de diseño son la parte más importante de un proceso para el desarrollo de un DW, con esto bien especificado, el DW difícilmente falle en su funcionamiento y usando el esquema constelación, se tendrá una bonificación a nivel organizativo debido a su cualidad de ser específico.

- Pentaho cumple bastante bien como herramienta de integración de datos y se adapta a cualquier tipo de metodología descrita en este documento, tiene un gran soporte y muchos tutoriales que se pueden seguir para una correcta utilización de esta herramienta.

- Power BI es una herramienta bastante provechosa, con una amplia diferencia sobre las demás referenciadas en este documento, con un soporte excepcional y con una amplia gama de visualizaciones, las cuales serán necesarias para diversificar y adaptar la información que procese un DW.

- Este documento es una guía comparativa que trata de facilitar y simplificar procesos de selección de herramientas y metodologías necesarias para el desarrollo de un DW. Esto se busca mediante un conjunto de comparaciones y análisis a algunas herramientas y metodologías disponibles en el mercado, basados en criterios sencillos para mejor compresión, con esto, se pretende demostrar que casi cualquier proceso para el desarrollo de un DW puede usar estas comparaciones a manera de guía y acelerar decisiones.

8. RECOMENDACIONES

- Se recomienda tener en cuenta que no todas las herramientas y metodologías no fueron estudiadas y comparadas, existen más de estas para ser analizadas y algunas pueden cumplir mejor para el desarrollo de un DW que las mencionadas en este documento.

- Una vez se haya desarrollado un DW usando estas comparaciones, ya se podrán aplicar los procesos de BI que normalmente se aplican.

BIBLIOGRAFÍA

BATENI, Mohammad. Scheduling to Minimize Staleness and Stretch. En: Real-Time Data Warehouses. USA: Theory of Computing Systems Vol. 49, No 4. (2011); p. 758-780. ISNN: 1432-4350

BERNEBEU, Ricardo Dario. Data warehousing y metodología Hefesto. [En línea]. Córdoba, Argentina: Universidad para la cooperación internacional. 2010. Sp. [consulta 09/10/2016]. Disponible en: < http://www.ucipfg.com/Repositorio/MATI/MATI-10/BLOQUE-ACADEMICO/Unidad-03/lecturas/bibliografia/Data%20Warehousing%20y%20metodolog%C3%ADa%20Hefesto%201.pdf>

CARPANI, Fernando & RUGGIA, Raul. An Integrity Constraints Language for a Conceptual Multidimensional Data Model. [en línea]. Montevideo, 2001, Sp. Universidad de la república. Facultad de ingeniería. [consulta: 23/11/2016] Disponible en: <https://www.fing.edu.uy/inco/grupos/csi/esp/Publicaciones/2001/seke2001-fc.pdf>
PERALTA. Op. Cit., p. 24.

COLOMBIA. CONGRESO DE LA REPÚBLICA. Ley estatutaria 1266 (31, Diciembre, 2008) Por la cual se dictan las disposiciones generales del habeas data y se regula el manejo de la información contenida en bases de datos personales, en especial la financiera, crediticia, comercial, de servicios y la proveniente de terceros países y se dictan otras disposiciones. Diario oficial. Bogotá, 2008. no. 47219. 17.p

COLOMBIA. CONGRESO DE LA REPÚBLICA. Ley estatutaria 1581 (17, Octubre, 2012) Por la cual se dictan disposiciones generales para la protección de datos personales. Diario oficial. Bogotá, 2012. no. 48587. 15.p

FUENTES TAPIA, Luis y VALDIVIA PINTO, Ricardo. Incorporación de elementos de inteligencia de negocios en el proceso de admisión y matrícula de una universidad chilena. Arica (Chile): [en línea]. En: Revista chilena de ingeniería. 2010. No. 18 p383-394 ISSN: 0718-3291 [consulta: 18/08/2016]. Disponible en: < http://web.b.ebscohost.com.biblioteca.umanizales.edu.co:2048/ehost/detail/detail?vid=34&sid=3061e044-3445-4895-b091-af4f9b500f07%40sessionmgr111&hid=110&bdata=Jmxhbmc9ZXMmc2l0ZT1laG9zdC1saXZl#AN=60144741&db=aph>

HERNÁNDEZ, Pedro. Microsoft's Cortana Assists in Unearthing Power BI Insights. [en línea]. En: eWeek. (2015); Foster (CA, USA): QuinStreet Enterprise. ISSN: 1530-6283. [Consulta: 18/08/2016]. Disponible en:

<http://search.ebscohost.com.biblioteca.umanizales.edu.co:2048/login.aspx?direct=true&db=aph&AN=111385317&lang=es&site=ehost-live>

HERNÁNDEZ, Pedro. Microsoft Improves Power BI Management for Enterprises. En: eWeek. (2015); p1-1. ISSN: 1530-6283 [Consulta: 18/08/2016]. Disponible en: <http://www.eweek.com/enterprise-apps/microsofts-cortana-assists-in-unearthing-power-bi-insights.html>

KIMBALL, Ralph y ROSS, Margy. The Data Warehouse toolkit. 3 ed. Indianapolis: John Wiley & Sons, Inc, 2013. 564 p. ISBN 978-1-118-53080-1

IBM. The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance. 2007, p. 3. ISSN: LO11960-USEN-00

NEGASH, Solomon. Business Intelligence. [En línea]. En: Communications of the Association for Information Systems. Vol. 13, Article 15. (2004); p.177-195. ISSN: 1529-3181. [Consulta: 18/08/2016]. Disponible en: <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=3234&context=cais>

Oficina de Cooperación Universitaria, S.A. Inteligencia institucional en universidades. Arequipa, Perú. (2013); p.21-704. ISBN: 978-84-695-8892-5

PERALTA, Verónika. Diseño Lógico de Data Warehouses a partir de Esquemas Conceptuales Multidimensionales. Montevideo, 2001, 153 p. Maestría en informática. Universidad de la república. Facultad de ingeniería.

PONNIAH, Paulraj. Data Warehousing: Fundamentals for IT professionals. 2 ed. USA: John Wiley & Sons, Inc, 2010. 571 p. ISBN 978-0-470-46207-2

PRIETO, Manuel, DODERO, Juan y VILLEGAS. Recursos digitales para la educación y la cultura. México: Kaambal. 2010. p.11-235. ISBN Volumen: 978-607-95446-1-4 [Consulta: 18/08/2016] Disponible en: <http://www.itsmotul.edu.mx/ccita2011/documentos/Recursos_digitales.pdf#page=49>

PRIETO, Roberto., et al. Análisis comparativo de modelos de madurez en inteligencia de negocio. [En línea]. En: INGENIARE. Antofagasta, Chile. Vol 23. (2015); p.361-371. ISSN: 0718-3291. [Consulta: 18/08/2016] Disponible en: <http://search.ebscohost.com.biblioteca.umanizales.edu.co:2048/login.aspx?direct=true&db=aph&AN=109143779&lang=es&site=ehost-live>

RIVADERA, Gustavo. La metodología de Kimball para el diseño de almacenes de datos (Data warehouses). [en línea]. Salta, Argentina: Cuadernos de la Facultad n. 5, 2010. 56-71.P. [consulta: 09/10/2016]. Disponible en:

<http://www.ucasal.edu.ar/htm/ingenieria/cuadernos/archivos/5-p56-rivadera-formateado.pdf>

ROBAYO, Eduard. Modelo de un sistema de inteligencia de negocios basado en S-BSC para entidades prestadoras de servicio de salud de alta complejidad sin ánimo de lucro. 2014. Colombia. Maestría en Gestión de Información. Escuela Colombiana de Ingeniería. [Consulta: 18/08/2016] Disponible en: <http://repositorio.escuelaing.edu.co/bitstream/001/303/1/FC-Maestria%20en%20Gestion%20de%20la%20Informaci%C3%B3n-80872862.pdf>

SALCEDO PARRA, Octavio J.; GALEANO, Rita Milena & RODRIGUEZ B., Luis G. Metodología Crisp para la implementación Data Warehouse. En: Tecnura. Bogotá: Universidad Distrital Francisco José de Caldas. Vol. 14, No. 26, 2010, pp. 35-48. ISSN: 0123-921X

SILBERSCHATZ, A., KORTH, H. & SUDARSHAN, S. Fundamentos de bases de datos. 4ed. Madrid: MacGraw-Hill, 2002. 816 p. ISBN: 84-481-3654-3

TÓRRES VELÁSQUEZ, Carlos Fernando. Aplicación de Inteligencia de Negocios (BI y KPI) en la estrategia de permanencia estudiantil: caso Fundación Universitaria Católica del Norte (Colombia). [En línea]. En: Trabajos Conferencias TICAL. (2015). [Consulta: 18/08/2016] Disponible en: <http://documentos.redclara.net/bitstream/10786/988/6/9-Aplicaci%C3%B3n%20de%20Inteligencia%20de%20Negocios%20%28BI%20y%20KPI%29.pdf>

WALZ, Aaron. Caso de estudio Universidad de Illinois. En: EVEREST. Libro blanco inteligencia institucional en universidades. Arequipa: Oficina de Cooperación Universitaria S.A, 2013. p. 311-388. ISBN: 978-84-695-8892-5

ANEXO A
RESUMEN ANALÍTICO

Título del proyecto	Comparativo de metodologías y herramientas para el desarrollo de un Data Warehouse
Presidente	Betancourt Correa, Carlos cbc@umanizales.edu.co Magister en Educación Docencia, Docente, Universidad de Maizales.
Tipo de documento	Trabajo de grado.
Referencia documento	Santiago Hernández Mejía. Comparativo de metodologías y herramientas para el desarrollo de un Data Warehouse. Manizales, 2017, paginación o número de volúmenes. Trabajo de grado para optar por el título de ingeniero de sistemas y telecomunicaciones. Universidad de Manizales. Ciencias e ingeniería.
Institución	Ingeniería de sistemas y telecomunicaciones, Ciencias e ingeniería, Universidad de Manizales.
Palabras claves	Data Warehouse, Comparativo, Herramientas, Metodologías.
Descripción	En este documento se estudian, analizan y comparan diversas metodologías y herramientas para el desarrollo de un Data Warehouse (DW) que permita la integración de información en caso, o no que dichos datos se encuentren en diferentes motores de bases de datos y/o provengan de diferentes fuentes de datos, esto, con el fin de convertir los datos en información pertinente y para que dichos datos cumplan con características como la calidad y exactitud, entre otras. Con la gran ventaja de que una vez el Data Warehouse esté desarrollado, se puedan ejecutar procesos de Business Intelligence (BI) para lograr que la información pueda ser usada para la toma de decisiones.
Fuentes	Bases de datos científicas.
Contenido	El documento está compuesto por: área problemática, objetivos, justificación, marco teórico, descripción de la metodología, resultados, conclusiones y recomendaciones.

Metodología	Fase 1. identificación documental Fase 2. revisión y comparación de metodologías para el diseño y control de Data Warehouse Fase 3. descripción de los lineamientos de diseño Fase 4. revisión y comparación de las herramientas para el diseño y control de Data Warehouse Fase 5. revisión y comparación de las herramientas para la visualización de datos.
Conclusiones	Este documento es una guía comparativa que trata de facilitar y simplificar procesos de selección de herramientas y metodologías necesarias para el desarrollo de un DW. Esto se busca mediante un conjunto de comparaciones y análisis a algunas herramientas y metodologías disponibles en el mercado, basados en criterios sencillos para mejor compresión, con esto, se pretende demostrar que casi cualquier proceso para el desarrollo de un DW puede usar estas comparaciones a manera de guía y acelerar decisiones.
Anexos	ANEXO A. Resumen Analítico