

GUÍA PARA LA IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE
NEGOCIOS PARA PEQUEÑAS Y MEDIANAS EMPRESAS

DANIEL FELIPE FLÓREZ SALGADO
LAURA MARGARITA DEL MAR TEJADA CARDONA

UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA, FÍSICA Y CIENCIAS
DE LA COMPUTACIÓN
INGENIERÍA DE SISTEMAS Y COMPUTACIÓN
PEREIRA
2015

GUÍA PARA LA IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE
NEGOCIOS PARA PEQUEÑAS Y MEDIANAS EMPRESAS

DANIEL FELIPE FLÓREZ SALGADO
LAURA MARGARITA DEL MAR TEJADA CARDONA

TRABAJO DE GRADO REALIZADO PARA OPTAR POR EL TÍTULO DE
INGENIERO DE SISTEMAS Y COMPUTACIÓN

ASESOR
OMAR IVÁN TREJOS BURITICÁ
INGENIERO DE SISTEMAS

UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA, FÍSICA Y CIENCIAS
DE LA COMPUTACIÓN
INGENIERÍA DE SISTEMAS Y COMPUTACIÓN
PEREIRA
2015

CONTENIDO

GLOSARIO.....	6
INTRODUCCIÓN	8
1. DEFINICIÓN DEL PROBLEMA.....	10
2. JUSTIFICACIÓN.....	12
3. OBJETIVOS.....	13
3.1 OBJETIVO GENERAL.....	13
3.2 OBJETIVOS ESPECÍFICOS	13
4. MARCO REFERENCIAL	14
4.1 MARCO TEÓRICO.....	14
4.1.1 Inteligencia de Negocios	14
4.1.2 Razones para la Inteligencia de Negocios	15
4.2 MARCO CONCEPTUAL.....	16
4.2.1 Tecnologías de Inteligencia de Negocios.....	16
4.2.2 ETL	16
4.2.3 <i>Data Warehouse</i>	17
4.2.4 <i>Data Mart</i>	17
4.2.5 OLAP	17
5. ESTRATEGIA DE INTELIGENCIA DE NEGOCIOS	19
5.1 CONTEXTO DEL NEGOCIO.....	20
5.2 INDICADORES CLAVE DE RENDIMIENTO	20
5.3 CALIDAD DE LOS DATOS.....	21
5.4 PLAN DE INTELIGENCIA DE NEGOCIOS.....	23
5.5 INTELIGENCIA DE NEGOCIOS EN LAS PYMES.....	24
5.5.1 Factores clave para la adopción de BI en la PYME	26
5.5.2 Niveles de madurez de BI.....	31
6. COMPONENTES DE UN SISTEMA DE INTELIGENCIA DE NEGOCIOS	34
7. BODEGAS DE DATOS.....	38
7.1 DEFINICIÓN.....	38

7.2 ¿POR QUÉ UTILIZAR UN DATA WAREHOUSE?	39
7.3 CARACTERÍSTICAS DE LAS BODEGAS DE DATOS.....	41
7.4 DATA MART	43
7.5 PROCESOS ETL.....	44
7.6 ARQUITECTURAS DE DATA WAREHOUSE	45
7.6.1 <i>Data Warehouse</i> global o empresarial.....	46
7.6.2 <i>Data Marts</i> dependientes.....	47
7.6.3 <i>Data Marts</i> independientes.....	48
7.6.4 <i>Data Marts</i> interconectados.....	49
8. ANÁLISIS DE DATOS.....	51
8.1 HERRAMIENTAS DE CONSULTA Y REPORTE.....	52
8.2 SPREADSHEET	53
8.3 OLAP	54
8.3.1 Implementaciones de OLAP	55
8.3.2 Operaciones sobre OLAP	58
8.4 DASHBOARDS.....	64
9. METODOLOGÍAS PARA LA CONSTRUCCIÓN DE UNA BODEGA DE DATOS	65
9.1 ENFOQUES DE IMPLEMENTACIÓN.....	65
9.1.1 <i>Top-down</i>	66
9.1.2 <i>Bottom-up</i>	66
9.2 METODOLOGÍA DE INMON	67
9.2.1 Modelo de datos	67
9.2.2 Metodología de desarrollo - Meth2.....	70
9.3 METODOLOGÍA DE KIMBALL.....	73
9.3.1 Modelo dimensional.....	73
9.3.2 Ciclo de vida de Kimball.....	78
9.4 COMPARACIÓN DE LAS METODOLOGÍAS	81
10. HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS	86
11. CONCLUSIONES.....	98

TABLA DE FIGURAS

Figura 1. Arquitectura de un sistema de BI.....	34
Figura 2: Ventajas e inconvenientes de las bodegas de datos.	41
Figura 3: <i>Data Warehouse</i> empresarial.....	47
Figura 4: <i>Data Marts</i> dependientes	48
Figura 5: <i>Data Marts</i> independientes.....	49
Figura 6: <i>Data Marts</i> interconectados.....	50
Figura 7: Ejemplo de operación <i>Slice</i> sobre la dimensión producto	59
Figura 8: Ejemplo de operación <i>Dice</i> sobre las dimensiones tienda y producto.....	60
Figura 9: <i>Drill-down</i>	60
Figura 10: <i>Roll-up</i>	61
Figura 11: Ejemplo de operación <i>Drill-across</i> desde la dimensión tienda hacia la dimensión producto.....	62
Figura 12: Ejemplo de operación <i>Pivoting</i> entre las dimensiones tienda y producto.....	63
Figura 13: Modelo entidad relación	68
Figura 14: Relación entre ERD y DIS	69
Figura 15: <i>Data Item Set</i>	70
Figura 16: Meth2	71
Figura 17: Ejemplo tabla de hechos.....	74
Figura 18: Ejemplo tabla de dimensión	75
Figura 19: Esquema de estrella	76
Figura 20: Esquema de copo de nieve	77
Figura 21: Ciclo de vida de Kimball.....	78
Figura 22: Comparación de las características de los modelos de Inmon y Kimball	83
Figura 23: Factores a favor de cada modelo	84
Figura 24: Costos de implementación de un proyecto de BI.....	89
Figura 25: Cuadrante para la clasificación de los beneficios del BI.....	91
Figura 26: Prioridades tecnológicas de las PYMEs frente a las grandes empresas.....	92
Figura 27: Parámetros de comparación de las suites de BI.....	95
Figura 28: Comparación de las soluciones de BI para PYMES (A-L)	96
Figura 29: Comparación de las soluciones de BI para PYMES (M-T).....	97

GLOSARIO

BIG DATA: término utilizado para referirse a grandes volúmenes de datos que no pueden ser procesados por las aplicaciones tradicionales debido a su gran complejidad y tamaño. Los desafíos de Big Data incluyen el análisis, captura, almacenamiento y visualización de los datos.

CIO (CHIEF INFORMATION OFFICER): persona encargada de los sistemas informáticos y tecnológicos que contribuyen a la obtención de los objetivos empresariales.

COSTEO BASADO EN ACTIVIDADES: metodología de costeo que identifica las actividades de la organización y asigna el costo de cada actividad en relación con los recursos y productos que se empleen en cada una de ellas.

CRM (CUSTOMER RELATIONSHIP MANAGEMENT): estrategia de negocio para la gestión de las interacciones de la organización con los actuales y futuros clientes. Usualmente implica el uso de tecnología para automatizar, integrar y gestionar las ventas, el mercadeo, el servicio al cliente y el soporte técnico.

ERP (ENTERPRISE RESOURCE PLANNING): software de gestión empresarial, normalmente compuesto de varias aplicaciones integradas, utilizado para recopilar, almacenar, gestionar e interpretar los datos de actividades del negocio como producción, distribución, facturación, nómina, contabilidad, entre otras.

INDICADOR: medida para evaluar el éxito o resultado obtenido en la ejecución de una actividad, proyecto o programa.

MINERÍA DE DATOS: es el proceso de análisis con el objetivo de descubrir patrones en grandes conjuntos de datos que implica métodos estadísticos, de aprendizaje de máquinas e inteligencia artificial.

PYMES: Pequeñas y medianas empresas.

RETORNO DE LA INVERSIÓN: indicador financiero que mide el beneficio o rentabilidad de una inversión, es decir, la relación entre la utilidad neta y el monto invertido.

SISTEMAS TRANSACCIONALES: sistema de información diseñado para capturar, almacenar y recuperar cualquier tipo de información generada por una organización.

SQL (STRUCTURED QUERY LANGUAGE): lenguaje declarativo para acceder a bases de datos relacionales.

TCO (TOTAL COST OF OWNERSHIP): evaluación financiera que permite a los compradores determinar los costos directos e indirectos de un producto o sistema.

TI: Tecnologías de la información.

VALOR ACTUAL NETO: procedimiento que permite calcular el valor presente de un determinado número de flujos de caja futuros, originados por una inversión.

INTRODUCCIÓN

Durante la última década, la importancia de la información y el conocimiento para la actividad empresarial ha incrementado de manera significativa. La globalización se ha convertido en una realidad, generando nuevos mercados cada vez más competitivos y dinámicos. La información entonces se constituye como un recurso fundamental para la gestión y la creación de valor para el negocio. El conocimiento sobre los competidores y clientes es esencial para determinar las directrices estratégicas que regirán el desarrollo de la organización. El aprovechamiento de este recurso es importante para el crecimiento de las empresas; una forma de hacer frente a esta necesidad es a través del desarrollo de sistemas de información organizacionales que facilitan la manipulación de los datos del negocio y su respectiva difusión entre los diferentes usuarios de la información.

Un sistema de Inteligencia de Negocios se enfoca en la gestión de la información interna y externa, generando conocimiento que puede ser utilizado de manera proactiva para obtener una ventaja competitiva que permita a las organizaciones monitorear su desempeño y alcanzar sus objetivos estratégicos. Este tipo de soluciones disponen de un conjunto de herramientas y aplicaciones que permiten a los ejecutivos reunir, organizar, analizar, distribuir y, finalmente, tomar acciones sobre los temas críticos del negocio, con el objetivo de ayudar a las empresas a tomar mejores decisiones, de forma oportuna e informada.

La utilización de la tecnología de BI (por sus siglas en inglés, *Business Intelligence*) no está limitada solo a las grandes organizaciones. Esta también puede ser adoptada por pequeñas y medianas empresas (PYMES) para obtener mejores resultados en sus negocios. Sin embargo, enfocarla apropiadamente es una tarea bastante difícil para cualquier tipo de compañía. En términos generales, BI se encarga de tomar los datos generados diariamente en las actividades del negocio y presentarlos de forma que puedan ser fácilmente utilizados por los administradores. La capacidad de analizar información histórica para tomar decisiones sobre el negocio constituye la base fundamental para el crecimiento empresarial.

La transformación de los datos en conocimiento y la integración de diferentes fuentes de información puede permitir a las PYMES el acceso oportuno a datos empresariales de alta calidad y métricas confiables. Esto les posibilita tomar acciones acertadas sobre clientes, proveedores o procesos internos. Así, la

Inteligencia de Negocios facilita a las compañías la identificación y reducción de los costos de operación, eliminando también la incertidumbre en las actividades de toma de decisiones.

En cualquier economía, en particular para el caso de los países en vía de desarrollo como Colombia, las PYMES desempeñan un papel vital. Actualmente, estas también generan grandes volúmenes de datos y se están viendo obligadas a utilizar tecnologías y herramientas avanzadas, como los sistemas de Inteligencia de Negocios y visualización de los datos, para apoyar los procesos de toma de decisiones.

Aunque se han realizado muchos estudios sobre la necesidad de los sistemas de soporte a la decisión en las pequeñas y medianas empresas, estos están orientados a la adaptación de soluciones y enfoques existentes, que son apropiados para las grandes organizaciones. Sin embargo, la mayor parte de estos resultan inadecuados para los pequeños negocios. Además, debido a que las filosofías de trabajo en las PYMES y en las grandes compañías difieren considerablemente, no es recomendable utilizar herramientas destinadas para empresas de gran tamaño.

En el presente trabajo de grado se pretende exponer las principales consideraciones, conceptos, metodologías y herramientas que se deben tener en cuenta antes de emprender este tipo de iniciativas. En el capítulo 5 se presenta un análisis sobre los elementos que se deben contemplar en la definición de una estrategia de Inteligencia de Negocios. A continuación se exponen de manera general los componentes de una solución de BI, seguido de un conjunto de conceptos sobre las bodegas de datos. En el capítulo 8 se presentan las tecnologías más utilizadas en la actualidad para realizar análisis sobre los datos. Posteriormente, se realiza un estudio de las metodologías para la construcción de bodegas de datos más representativas en la literatura. Finalmente, se presentan una serie de recomendaciones para la selección de las herramientas de BI, las conclusiones sobre los temas tratados en este documento y las referencias bibliográficas que fueron consultadas para su elaboración.

1. DEFINICIÓN DEL PROBLEMA

Actualmente, es común encontrar en las pequeñas y medianas empresas del país, la utilización de sistemas de gestión empresarial del tipo ERP (por sus siglas en inglés, *Enterprise Resource Planning*), los cuales permiten apoyar los procesos operativos de una compañía. Estos sistemas generalmente incluyen las actividades de producción, distribución, facturación, nómina, contabilidad, entre otras, integrando diferentes funciones de la organización en una plataforma homogénea. Como resultado de estas actividades operacionales, organizaciones de todo tipo poseen una inmensa cantidad de información residente en sus sistemas de procesamiento de transacciones, la cual aumenta en tamaño progresivamente, dificultando las labores de análisis, particularmente cuando las relaciones entre los datos se tornan más complejas.

La mayoría de los ERP modernos proporcionan reportes estándares preconfigurados que satisfacen los requerimientos básicos de presentación de la información. Estos informes proporcionan indicadores de los procesos operativos y administrativos de manera estática, que deben ser consultados manualmente, de forma independiente. Sin embargo, el mercado cambiante exige mejores herramientas que permitan a los ejecutivos realizar un análisis exhaustivo de los datos disponibles para administrar sus negocios.

La información es un factor de valor incalculable para el éxito empresarial; si esta no es orientada adecuadamente y no está disponible para su uso en el momento oportuno, puede perder todo valor ante el proceso de toma de decisiones. Esto hace evidente la necesidad de procesos y herramientas que faciliten el manejo e interpretación de los datos que arrojan los sistemas de operación y producción. Por lo tanto, las iniciativas corporativas deben estar orientadas a proveer a la gerencia información relevante para responder a los problemas del negocio.

Las razones expuestas, muestran causas del por qué la información proporcionada por estos sistemas, no resulta fácilmente utilizable por los altos directivos para la toma de decisiones. Además, los niveles estratégicos de las organizaciones requieren tener conocimiento del entorno que les permita evaluar las condiciones del negocio y vaticinar futuros eventos que afecten sus intereses.

Según un estudio realizado por la firma Aberdeen Group¹, la capacidad de estandarizar un proceso de recopilación e integración de los datos para su uso en la presentación de informes y análisis, es crucial para su transformación en elementos más útiles y cómodos de examinar [1]. Las soluciones de BI se presentan como un medio para realizar análisis a profundidad sobre los datos, proporcionando a los ejecutivos una única versión de la información crítica necesaria para tomar decisiones sobre el negocio.

Se observa entonces la necesidad de implementar sistemas que permitan proporcionar a la gerencia de las empresas de la región, la información y herramientas suficientes para realizar pronósticos en sus operaciones y mejorar su proceso de toma de decisiones. Por lo cual, el presente proyecto se centrará en la elaboración de una guía que contemple los principales conceptos que se deben tener en cuenta al momento de abordar un proyecto de Inteligencia de Negocios, que van desde los procesos necesarios para la extracción, transformación y carga de los datos, la construcción de bodegas y almacenes de datos, hasta la selección de las herramientas de análisis y presentación de la información, enfocado a los requerimientos de las pequeñas y medianas empresas.

¹ Aberdeen Group, Investigación y análisis de empresas y productos de tecnología de la información.

2. JUSTIFICACIÓN

Toda empresa, en el desarrollo de su actividad económica, debe tomar decisiones acertadas que le permitan posicionarse en un mercado altamente competitivo. Generalmente estas decisiones se toman con base en información que procede de fuentes como la intuición personal, el criterio empresarial, las opiniones de compañeros, o bien datos internos y externos de la organización. Sin embargo, las empresas deben disponer de información de calidad que permita conocer el comportamiento de la organización y monitorear los indicadores críticos de gestión, de manera que los ejecutivos dispongan de bases sólidas para generar una ventaja competitiva derivada del conocimiento obtenido de los datos operacionales provenientes de los diferentes procesos de la organización.

Es fundamental entonces que las empresas dispongan de herramientas que permitan diagnosticar situaciones que se presenten durante la ejecución de las actividades del negocio, disponiendo de una mayor cantidad de información útil en el menor tiempo posible. Además, estas deben presentar los datos críticos de la organización de forma fácil e intuitiva, de tal manera que pueda ser manipulada por los gerentes y que el acceso a la información no requiera de la intervención de un experto en tecnología.

Cada vez más, las organizaciones han notado que los beneficios provenientes de las inversiones en soluciones ERP pueden aumentar dramáticamente con la extracción de información adicional, a partir de los datos del negocio almacenados en estos sistemas [1]. En esta medida, las tecnologías de BI han sido ampliamente utilizadas como un medio para obtener valor agregado de estos datos.

La intención de este proyecto es proporcionar un marco de trabajo para la implementación de soluciones de Inteligencia de Negocios que pueda ser utilizado por organizaciones e individuos para el desarrollo de este tipo de iniciativas en las pequeñas y medianas empresas de la región.

3. OBJETIVOS

3.1 OBJETIVO GENERAL

Proponer una guía que reúna un conjunto de recomendaciones para la implementación de una solución de Inteligencia de Negocios para pequeñas y medianas empresas.

3.2 OBJETIVOS ESPECÍFICOS

- Revisar los principales conceptos que se deben tener en cuenta al momento de abordar un proyecto de BI.
- Realizar una revisión documental sobre las metodologías más representativas para el desarrollo de soluciones de Inteligencia de Negocios.
- Recomendar la metodología más adecuada para las condiciones y características de las PYMES.
- Realizar una comparación de las herramientas más representativas para la implementación de sistemas de BI en las PYMES.

4. MARCO REFERENCIAL

4.1 MARCO TEÓRICO

4.1.1 Inteligencia de Negocios

La primera referencia sobre BI fue realizada en 1958 por el investigador de IBM Hans Peter Luhn. El definió la Inteligencia de Negocios como “la habilidad de encontrar las relaciones que se presentan en un conjunto de hechos, de tal manera que permita encaminar acciones hacia un objetivo deseado” [2].

De acuerdo con Revelli, la Inteligencia de Negocios “es el proceso de recolección, tratamiento y difusión de la información que tiene un objetivo: la reducción de incertidumbre en el proceso de toma de decisiones estratégicas” [3]. Por otra parte, Cui et al. describen el BI como “una forma y método de mejorar el rendimiento del negocio proporcionando asistencia de gran alcance para los ejecutivos encargados de la toma de decisiones y brindarles información útil a la mano” [4].

Una definición más acorde con los objetivos de este proyecto fue propuesta por Stackowiak et al. en la que se define el *Business Intelligence* como “el proceso de tomar grandes cantidades de datos, analizarlos y presentarlo en un conjunto de reportes de alto nivel que condensan la esencia de esta información en la base de las acciones del negocio, lo que permite a la gerencia tomar decisiones fundamentales sobre el negocio” [5].

Los sistemas de Inteligencia de Negocios son un conjunto de herramientas y tecnologías utilizadas para apoyar el proceso de toma de decisiones de los niveles tácticos y estratégicos de una organización, mejorar las operaciones del negocio e incrementar la competitividad de las empresas. Estas soluciones combinan datos operativos e históricos con aplicaciones analíticas para presentar información valiosa que pueda ser utilizada por la gerencia para orientar las acciones del negocio.

El principal objetivo de la Inteligencia de Negocios es mejorar la oportunidad y calidad de la información y permitir a los administradores un mejor entendimiento de la situación de su organización. Las tecnologías de BI ayudan a los

empresarios a analizar las tendencias del mercado, los cambios en el comportamiento de los clientes y los patrones de gasto, así como las preferencias de los clientes y las capacidades de la empresa. Esta información puede ser utilizada por analistas y gerentes para el desarrollo de nuevas estrategias que representen oportunidades de crecimiento.

4.1.2 Razones para la Inteligencia de Negocios

La Inteligencia de Negocios concede a las organizaciones la posibilidad de tomar decisiones informadas sobre el negocio y generar estrategias competitivas que impacten positivamente su desempeño. Esto es especialmente cierto cuando las empresas son capaces de extrapolar la información de indicadores sobre el entorno para realizar pronósticos más precisos sobre tendencias futuras y condiciones económicas.

El fin último de la Inteligencia de Negocios es el de mejorar la oportunidad y la calidad de la información. Las empresas se dan cuenta de que en un entorno empresarial altamente competitivo uno de los factores más determinantes para el éxito es la rapidez con que se responde y se adapta al cambio. El BI surge entonces como una herramienta para ayudar a las organizaciones a integrar y analizar los datos, permitiendo desarrollar los negocios de manera efectiva.

El BI proporciona múltiples beneficios a las empresas que lo utilizan. Por ejemplo, se puede eliminar una gran cantidad de decisiones basadas en conjeturas, facilitar la comunicación entre los departamentos mediante la centralización de la información y coordinación de las actividades, además de permitir a las compañías responder rápidamente a los cambios en las condiciones del entorno, de tal manera que contribuye a mejorar el rendimiento general de las compañías en que se utiliza.

Algunos de los beneficios más representativos del BI para las empresas definidos por Ranjan en [6] se presentan a continuación:

- Con BI, las empresas pueden identificar a sus clientes más rentables y las razones subyacentes de la lealtad de los clientes, así como identificar a sus clientes potenciales.
- Detectar rápidamente problemas sobre las garantías reportadas para minimizar el impacto de las deficiencias en el diseño de sus productos.

- Analizar la rentabilidad potencial del cliente y reducir la exposición al riesgo a través de una puntuación de crédito financiero más precisa para sus clientes.
- Determinar qué combinaciones de productos y líneas de servicio son propensas a ser adquiridas por los clientes y en qué momentos.
- Determinar los niveles de deserción y rotación de clientes.
- Detectar comportamientos fraudulentos.

4.2 MARCO CONCEPTUAL

4.2.1 Tecnologías de Inteligencia de Negocios

Aunque el BI es adaptado por las organizaciones a sus propias necesidades, procesos y contextos, existe un enfoque general que tiene que ver con la agregación de datos, análisis del negocio y visualización de los datos. De acuerdo con este enfoque el BI abarca una variedad de herramientas tecnológicas para la elaboración de informes y pronósticos.

Según Khan et al. en [7], el concepto de BI se puede descomponer en tres elementos: captura de datos, almacenamiento de datos y acceso y análisis de los datos. Los datos se recolectan de fuentes tanto internas como externas. Los datos internos provienen de sistemas operacionales de la organización y los datos externos se obtienen por medio de clientes, proveedores, agencias del gobierno, competidores, internet, entre otros. Los datos heterogéneos recopilados se almacenan generalmente en un *Data Warehouse* después de su respectivo tratamiento. Finalmente los datos almacenados en el repositorio son analizados para la toma de decisiones.

A continuación se presentarán brevemente algunas de las tecnologías más representativas para el desarrollo de una solución de BI.

4.2.2 ETL

Consiste de herramientas que poseen interfaces con los sistemas operacionales de la organización para cargar los datos en el *Data Warehouse*. Los datos son capturados desde múltiples fuentes, por lo que resulta necesario realizar un proceso de transformación en el que los datos son depurados, filtrados y homogenizados con el objetivo de convertirlos a formatos consistentes. Esta

limpieza es indispensable debido a que los formatos de los datos pueden variar entre las diferentes fuentes transaccionales. Por último, la información es cargada a la bodega de datos, donde estarán a disposición de los usuarios del negocio.

4.2.3 *Data Warehouse*

Es una colección de información corporativa derivada directamente desde los sistemas operacionales y algunas fuentes externas. Su propósito específico es soportar las decisiones del negocio. Pueden contener información estructurada como tablas u hojas de cálculo o información no estructurada como archivos de texto plano, imágenes y demás recursos multimedia.

La arquitectura de un *Data Warehouse* puede tomar variedad de formas en la práctica pero antes de iniciar el proceso de diseño, se deben tener en consideración los requerimientos y recursos de la organización. Sin embargo, algunas de las opciones de arquitectura que las organizaciones pueden elegir en diferentes circunstancias incluyen: *Data Mart*, *Data Warehouse* empresarial, *Data Warehouse* distribuido, *Data Warehouse* virtual, entre otros que serán descritos con mayor detalle durante el desarrollo del proyecto.

4.2.4 *Data Mart*

Un *Data Mart* es un repositorio de datos de menor tamaño que un *Data Warehouse* dedicado a un departamento o área del negocio para facilitar sus propias actividades de apoyo a la toma de decisiones.

De forma similar a los *Data Warehouse*, los *Data Mart* contienen datos operacionales que permiten a los administradores del negocio crear estrategias basadas en el análisis de las tendencias y experiencias pasadas. La diferencia clave es que la creación de un *Data Mart* está enfocada a un proceso o unidad del negocio y pretende solucionar una necesidad específica.

4.2.5 OLAP

OLAP (por sus siglas en inglés, *On-Line Analytical Processing*) es un sistema que proporciona a los usuarios del negocio un conjunto de herramientas avanzadas que permiten examinar minuciosamente los datos proporcionando un acceso rápido y flexible a la información almacenada en los repositorios de datos.

OLAP brinda los medios para explorar y analizar grandes cantidades de datos, involucrando cálculos complejos, sus relaciones y la presentación visual de los resultados en diferentes perspectivas. Las herramientas de OLAP son una combinación de procedimientos para el procesamiento analítico con interfaces gráficas de usuario. Las principales características de este tipo de aplicaciones son: vistas multidimensionales de datos, capacidades intensivas de cálculo y tiempos de respuesta relativamente cortos en contraste con la cantidad de registros analizados.

El procesamiento analítico en línea ofrece vistas multidimensionales que resumen los datos del negocio y son utilizadas para la generación de reportes, análisis, modelos y planes para la optimización del negocio. Estas técnicas y herramientas son utilizadas generalmente para trabajar sobre *Data Warehouses* o *Data Marts* para el desarrollo de sistemas de inteligencia empresarial sofisticados.

5. ESTRATEGIA DE INTELIGENCIA DE NEGOCIOS

La definición de Inteligencia de Negocios propuesta por Gartner² se enfoca en el problema del análisis de datos, principalmente a través de herramientas de reportes, consultas ad hoc y procesamiento analítico en línea, de forma que estos puedan ser utilizados por los administradores y analistas del negocio. Aunque esta definición sigue siendo precisa, refleja una visión limitada del potencial del *Business Intelligence*.

En una encuesta realizada por Gartner en 2013 entre 2053 CIOs (por sus siglas inglés, *Chief Information Officers*) que incluía 36 industrias en 41 países, se encontró que las soluciones de BI son a menudo una prioridad de las inversiones en tecnología [8]. Esto se debe a que la información es considerada como un recurso estratégico de gran valor para el éxito de las empresas. Hoy en día, la mayoría de las organizaciones conciben la Inteligencia de Negocios como una iniciativa más amplia que puede afectar positivamente en los objetivos clave del negocio, tales como crecimiento financiero, fidelización de clientes, relaciones con los socios y ampliación de mercados.

Así, un proyecto de Business Intelligence no se limita solo a la selección de las tecnologías a implementar, sino que debe estar alineado con los objetivos del negocio y proporcionar a los diferentes actores de la organización la información necesaria para tomar mejores decisiones que contribuyan a la obtención de las metas deseadas.

Es vital establecer una estrategia de BI para asegurar que la implementación de componentes específicos se asocie con el plan de trabajo definido. La estrategia de BI debe abarcar un conjunto de procesos y tecnologías para la recolección, integración, acceso y análisis de los datos, con el propósito de proveer información a los interesados en todos los niveles de la organización.

La estrategia de BI debe establecer y documentar las necesidades identificadas por las partes interesadas, destacando cómo encaja en la visión de la empresa. Además, debe tener en consideración un marco de trabajo apropiado, metodologías, procesos, sistemas y tecnologías necesarios para entregar valor al negocio.

² Empresa de consultoría e investigación en tecnologías de la información.

5.1 CONTEXTO DEL NEGOCIO

La estrategia de BI debe abordar el cómo contribuirá al logro de las metas del negocio, cómo permitirá tomar mejores decisiones para incrementar el rendimiento de la organización y cómo se incorporará en los diferentes procesos de administración. Una estrategia de BI exitosa debe estar alineada con los objetivos empresariales, realizar un mejor uso de la información y permitir la integración de BI en las actividades estratégicas, tácticas y operacionales.

Los objetivos, las decisiones y los procesos empresariales deben promover la estrategia de Inteligencia de Negocios para asegurar la adopción exitosa de esta tecnología. Conducir la estrategia de BI teniendo en cuenta el contexto empresarial asegura la alineación de la visión organizacional, las metas y los procesos con las iniciativas de BI.

Se debe determinar y cuantificar los beneficios que el negocio desea obtener y qué necesidades se están afrontando. Algunos ejemplos de los objetivos que pueden ser abordados por la estrategia de BI incluyen: cómo incrementar la lealtad del cliente, aumentar el número de clientes, elevar la participación en el mercado, promocionar efectivamente los productos, analizar los segmentos de mercado, entre otros. Estos son los objetivos finales que se deben considerar en la definición de la estrategia de BI más no las herramientas o tecnologías a implementar [9].

5.2 INDICADORES CLAVE DE RENDIMIENTO

La Inteligencia de Negocios utiliza comúnmente los indicadores clave de rendimiento para evaluar el estado actual del negocio y establecer un plan de acción. Estos pueden ofrecer información valiosa para la toma de decisiones por lo que es importante asegurarse de que los procesos críticos están siendo monitoreados [10].

Los KPIs (por sus siglas en inglés, *Key Performance Indicators*) son un método para determinar el éxito del negocio utilizando métricas objetivas y directrices cuantificables. Pueden ser utilizados para evaluar el estado global de la

organización, comparar el plan de negocio con el desempeño actual o hacer seguimiento al progreso de una actividad en particular, meta, departamento o producto.

Cuando una organización establece los KPIs para definir los límites y resultados que determinan el éxito, estos pueden ser utilizados para determinar qué estrategias, procesos y actividades contribuyen al logro de los objetivos de la compañía y cuales tareas deben ser ajustadas para corregir resultados adversos y mejorar el rendimiento.

Los KPIs no se conciben como metas de desempeño sino como un mecanismo para determinar los avances en la obtención de los objetivos definidos. Estos proporcionan a los usuarios de la organización información clave sobre el rendimiento del negocio en periodos muy cortos. Sin el uso de estos indicadores tomaría mucho tiempo y esfuerzo realizar la recopilación y procesamiento de los datos para obtener información sobre el estado general de la empresa; en esta medida se presentarían demoras en la identificación y tratamiento de los problemas que se puedan presentar.

Los KPIs incluyen información estadística vital como tendencias de ventas, margen de rentabilidad, grado de satisfacción del cliente, desempeño de los departamentos, inventario en tiempo real o cualquier otro factor que se considere crítico para el éxito de la empresa. Los KPIs son parte integral de una solución de BI puesto que contribuyen a una correcta implementación de esta tecnología en las organizaciones.

5.3 CALIDAD DE LOS DATOS

En un estudio de Gartner realizado a más de 600 usuarios de BI, se encontró que más del 35% identificaron la calidad de los datos como uno de los principales problemas que debe abordar una organización al momento de implementar una solución de BI [9]. Asegurar la completitud de los datos, su calidad y coherencia es una actividad fundamental para la implementación de un entorno de BI, debido a que generalmente determina el éxito de un proyecto de Inteligencia de Negocios y puede obstaculizar las decisiones empresariales en todos los niveles de la organización.

La estrategia de BI debe hacer énfasis en la calidad de los datos, proceso que se debe realizar de manera permanente, durante las diferentes etapas en el transcurso de su adopción. Una iniciativa para asegurar la calidad de los datos debe incluir procedimientos de acceso a los datos, la definición de los roles de las personas que gestionan los datos en la organización, la definición correcta de los datos y establecer un marco de trabajo para resolver inconvenientes con los datos. Si no se tiene en cuenta la calidad de los datos esto resultará en una aceptación limitada o rechazo de la tecnología, ya que la información entregada al final del proceso no será confiable.

En general, existen principalmente dos fuentes de problemas en la calidad de los datos al momento de emprender un proyecto de Inteligencia de Negocios: los sistemas operacionales de la organización y los procesos derivados de la implementación de BI [11].

En algunos casos se detectan errores que se originan por defectos en los sistemas transaccionales, lo cual debería provocar proyectos de mejora en los mismos. Muchos de estos casos se presentan cuando los usuarios pueden introducir datos sin ningún tipo de control. Siempre que se pueda, es recomendable que los usuarios elijan entre distintos valores, en lugar de introducirlos libremente. No es buena opción corregirlos en el proceso ETL, sin modificar las aplicaciones de origen. Esta alternativa es mucho más rápida inicialmente, pero mucho más costosa a largo plazo [12]. Además se recomienda realizar comprobaciones sobre la estructura de los datos introducidos por el usuario para asegurar que los valores son ingresados en un formato específico.

Por otra parte, producto de la integración de los datos que provienen de diferentes fuentes, los proyectos de BI generan nuevos requerimientos para los datos existentes. Ahora los datos deben ajustarse no solo a los requisitos del sistema de origen, sino también a las necesidades del negocio que se definieron para la implementación de BI [11].

Asumir que la calidad de los datos es correcta puede ser un error fatal en los proyectos de *Business Intelligence*. Normalmente, cuando se construye un *Data Warehouse* la mayoría de las organizaciones se focalizan en identificar los datos que necesitan analizar, los extraen y los cargan en la bodega de datos. Debería, por lo tanto, establecerse un conjunto de controles en el proyecto que localizara los errores en los datos y no permitiera la carga de los mismos. Las comprobaciones se deberán llevar a cabo, de forma manual o automatizada,

teniendo en cuenta distintos niveles de detalle y variando los periodos de tiempo, comprobando que los datos cargados coinciden con los de las fuentes de datos origen [11].

5.4 PLAN DE INTELIGENCIA DE NEGOCIOS

Previo a la definición de la estrategia de BI, se debe realizar una evaluación completa de los procesos y las tecnologías actuales, ya que estos serán críticos para determinar las acciones a emprender en el proceso de implementación de una solución de Inteligencia de Negocios para la organización. Es necesario revisar y documentar los sistemas de información y los procedimientos de gobernanza de datos que están siendo utilizados, al igual que detallar cómo está siendo manipulada la información actualmente, lo cual debe incluir un inventario de todas las plataformas, tecnologías y herramientas que hacen parte de este proceso. La identificación del estado actual también incluye documentar los usuarios con sus respectivos perfiles y cómo dichos usuarios están haciendo uso de la información organizacional. Todo esto, es fundamental para establecer los objetivos de la estrategia de BI y la viabilidad de su ejecución.

Durante este proceso, es importante comprobar si los usuarios están recibiendo la información en formatos que les permitan aprovechar este recurso de manera efectiva; esto también permite conocer el uso que se le da en las diferentes áreas y cargos de la organización. Es necesario verificar la existencia de procedimientos de gobernanza y gestión de los datos para asegurar la correcta definición y calidad de los mismos.

El análisis del estado actual ayuda a identificar áreas problemáticas que pueden ser abordadas en la definición del plan. Una vez realizado este proceso se conocen los puntos neurálgicos que deben ser tratados, lo que permite tener claridad sobre los objetivos de la estrategia de BI y la ruta de acción a emprender para llegar al estado deseado.

Tras el análisis de la situación actual se entenderá mejor el entorno de la organización y se tendrá entonces que decidir las actividades y recursos que se requieren para alcanzar la visión de BI. En el análisis del estado, se identifica de qué forma los usuarios están accediendo y usando la información. Posteriormente, se debe optimizar la forma en que dichos usuarios acceden y consumen la

información obtenida, cómo podrán compartirla y cómo se logrará un entorno colaborativo que permita la correcta utilización de las tecnologías de BI.

Después de esta evaluación, se debe construir un plan a largo plazo; el estado objetivo debe reflejar la visión de combinar la Inteligencia de Negocios con las áreas de administración de la organización, la mejora del rendimiento, el servicio al cliente, la gestión del conocimiento, entre otros. Las iniciativas de *Business Intelligence* proporcionan oportunidades para incrementar el desempeño general de la empresa al facilitar la gestión de la información en los diferentes niveles organizacionales.

El plan de transformación debe establecer una propuesta de implementación de la infraestructura de BI y las tecnologías relacionadas. Es necesario incluir actividades operativas: definir un banco de proyectos, establecer los procesos de gestión de la información, especificar la arquitectura conceptual, lógica y física de BI, determinar la infraestructura tecnológica requerida e incorporar las necesidades de información de las diferentes áreas de la organización, identificando cómo los usuarios quieren recibir y consumir la información.

En la estrategia de Inteligencia de Negocios es necesario formalizar la identificación y planeación de las iteraciones de BI. El plan debe considerar la capacidad, el rendimiento, la calidad de los datos, las políticas de seguridad, el control de los metadatos, la política de gestión de los datos, etc. Es importante también considerar el tiempo de implementación de BI, la priorización de los proyectos individuales y la disponibilidad de recursos para la ejecución de iniciativas enfocadas en BI.

5.5 INTELIGENCIA DE NEGOCIOS EN LAS PYMES

El procesamiento de la información se ha convertido gradualmente en un factor fundamental para generar ventajas competitivas en el mercado actual. Con los avances en las tecnologías de la información, la existencia de una gran variedad de productos y mayores demandas por parte de los clientes, las empresas están obligadas a operar en entornos altamente complejos y dinámicos. Las organizaciones que sobreviven y tienen éxito en estas condiciones de mercado necesitan tomar decisiones de manera oportuna, eficaz y apropiada [13].

Con el objetivo de incrementar la eficiencia, muchas organizaciones han implementado sistemas de información en las operaciones de sus negocios para recolectar, combinar, acceder y analizar grandes cantidades de datos. Algunas de estas herramientas de análisis son proporcionadas por las tecnologías de BI, que conducen no solo a la mejora del rendimiento general de la empresa sino que también ayudan en la predicción de tendencias del mercado mediante el análisis de datos históricos.

A pesar de que las tecnologías de BI normalmente se consideran reservadas para las grandes empresas, la demanda actual de BI no se limita al tamaño de las organizaciones [14]. Durante un tiempo considerable, las grandes compañías se han dado cuenta de la importancia de los datos existentes en sus sistemas de información y han realizado grandes inversiones en sistemas que les proporcionen capacidad de análisis y predicción. La oferta de soluciones de Inteligencia de Negocios para las PYME es muy reducida, ya que el público objetivo de los proveedores de herramientas de BI siguen siendo las grandes empresas, debido a que son este tipo de organizaciones quienes demuestran una mayor receptividad para acoger proyectos de esta naturaleza y también debido a su capacidad financiera [15]. Solo recientemente algunas PYMEs comenzaron a reunir los requisitos para adquirir este tipo de soluciones. De hecho, tanto las pequeñas como las medianas empresas observan la necesidad de incorporar este tipo de herramientas en la misma medida que las grandes compañías.

En un entorno empresarial en constante cambio, los requisitos a nivel de tecnologías de la información, los desafíos del mercado y las presiones empresariales que enfrentan las PYMEs no son muy diferentes de los que afrontan las grandes empresas. La combinación de datos históricos con la tecnología de BI permite a las pequeñas y medianas empresas tomar decisiones estratégicas pertinentes a través de técnicas de exploración de grandes volúmenes de datos, potenciando el análisis de patrones desconocidos para generar una ventaja importante en el entendimiento de los procesos del negocio.

La gestión de grandes cantidades de datos y su transformación en información útil representa un gran desafío para las PYMEs. Generalmente las organizaciones requieren información para el análisis de situaciones desfavorables; este análisis es un enfoque reactivo a una situación negativa o adversa. En una perspectiva de gestión proactiva, si los resultados son positivos, las organizaciones deben continuar desarrollando acciones para aprovechar y optimizar los beneficios obtenidos. Una empresa proactiva se enfoca en identificar nuevas oportunidades y

prepararse para anticipar posibles fallos. Por otra parte, las organizaciones reactivas desgastan toda su energía y todos los recursos en revertir los resultados negativos, a expensas de la mejora de sus habilidades y la exploración de oportunidades. Las compañías deben entonces decidir si adoptar un enfoque proactivo o reactivo ante el comportamiento del negocio y del entorno.

La explotación de las tecnologías de BI es importante para el desarrollo de las pequeñas y medianas empresas. En una investigación realizada por Gibson y Arnott en 2003, se concluyó que si las grandes organizaciones continúan aprovechando las últimas tecnologías de soporte a la decisión, y las pequeñas empresas siguen sin correr el riesgo de adoptar inteligencia empresarial, la brecha de poder solo se mantendrá en aumento. Si este desequilibrio persiste, a las pequeñas organizaciones les resultará cada vez más difícil competir en la economía moderna [16]. Sin embargo, existen muchas consideraciones y riesgos que las organizaciones deben tener en cuenta para evaluar la adopción de soluciones de BI.

La falta de recursos es uno de los principales factores que se aborda en el estudio de las PYMEs. Algunos de estos recursos incluyen las finanzas, la tecnología, los conocimientos y los recursos humanos. Debido a su presupuesto financiero restringido y el poco número de trabajadores, la mayoría de los empleados deben realizar múltiples tareas y no se especializan en un área en particular [17]. Por otra parte, la mano de obra no calificada sumada a la falta de especialización técnica resulta en que los gerentes de las PYMEs sean conservadores al adoptar innovaciones de tecnología. Esta puede ser la razón por la que la adopción de nuevas tecnologías en las PYMEs es un proceso lento que se realiza con especial precaución. Por ejemplo, Fuller-Love encontró que los propietarios o gerentes acceden a la adopción de TI solo cuando perciben la promesa de éxito, ya que no quieren correr riesgos [18]. Este es especialmente el caso de BI, debido a que los costos de mantenimiento y de ejecución son muy altos, así como la probabilidad de fracaso en la implementación en comparación con otras tecnologías.

5.5.1 Factores clave para la adopción de BI en la PYME

En un estudio realizado por Boonsiritomachai, McGrath y Burgess se identificaron los posibles factores que impactan la implementación de nuevas tecnologías de la información. Como BI se considera una innovación tecnológica, los factores que pueden llegar a afectar su adopción se basaron en esta propuesta. Estos

elementos fueron clasificados en cuatro categorías que agrupan un conjunto de características tecnológicas, organizacionales, del entorno y de los propietarios o gerentes [19].

5.5.1.1 Características tecnológicas.

Estas características fueron definidas teniendo en cuenta la teoría de la difusión de la innovación, propuesta por Rogers en 1995. De acuerdo con sus planteamientos, los atributos que afectan la adopción de innovación tecnológica son: ventaja relativa, complejidad, compatibilidad y la capacidad de experimentación y observación.

- Ventaja relativa: Se refiere al grado en que una innovación se percibe como mejor que las ideas o los sistemas existentes. Según Lönnqvist, las herramientas de BI proporcionan múltiples ventajas para los negocios, entre las que resalta: la reducción de la dispersión de la información, mayor posibilidad de interacción entre los usuarios, facilidad de acceso a la información, disponibilidad de la información en tiempo real y utilidad en el proceso de toma de decisiones [20].
- Complejidad: Es el grado en el que una innovación se percibe como difícil de entender o utilizar. Varios investigadores han encontrado que la complejidad es una barrera para la adopción de soluciones innovadoras y que entre menos compleja sea una tecnología mayor es la posibilidad de que sea acogida exitosamente. Por lo tanto, debido a la gran complejidad de la tecnología de BI, los empleados pueden resistirse a su adopción y generalmente continuarán utilizando las herramientas tradicionales. Voicu et al. confirmaron que los modelos de BI son bastante complejos debido a que integran funciones matemáticas para predecir tendencias en el desempeño de una organización con el objetivo de ofrecer alternativas de acción en diferentes situaciones [21]. Así, los usuarios con pocos conocimientos en TI y computación requieren una solución simple y estable que satisfaga sus necesidades en el menor tiempo posible.
- Compatibilidad: Es la medida en que una innovación se considera fácilmente adaptable al entorno y las necesidades de la organización. La guía de Inteligencia de Negocios reportó en 2009 que el 40% de los costos asociados a los proyectos de BI están relacionados con el análisis y transformación de los datos que residen en los diferentes sistemas de la organización [22]. Si los

sistemas existentes no son compatibles con la tecnología de BI, se necesitará una importante inversión de tiempo y recursos para migrar e integrar los datos.

- Capacidad de experimentación: Evalúa hasta qué punto los posibles adoptantes tienen la oportunidad de experimentar con una innovación. Cuanto mayor sea esta capacidad, los adoptantes potenciales de esta tecnología se sentirán más cómodos y será más probable su aceptación. Por lo tanto, si los proveedores de BI proporcionan a los usuarios potenciales la oportunidad de experimentar con sus soluciones antes de su implementación, se reducirán las dudas relacionadas con su funcionamiento.
- Capacidad de observación: Representa el grado en el que los adoptantes potenciales de una innovación pueden percibir los resultados de su utilización por parte de otros usuarios que la han acogido previamente. Lundblad afirmó que los resultados visibles de una innovación afectan la percepción de su valor por parte de individuos y comunidades. Por lo tanto, los efectos visibles de una innovación normalmente conducen a una rápida aceptación [23].

5.5.1.2 Características organizacionales.

Las características organizacionales se definieron con base en el modelo Tecnología-Organización-Entorno propuesto por Tornatzky y Fleischer en 1990. La capacidad de las organizaciones para implementar innovaciones tecnológicas es un aspecto importante que afecta la decisión de su adopción. En la dimensión organizacional, existen cuatro factores que permiten determinar si una compañía debe o no adoptar una innovación tecnológica: el tamaño de la organización, la antigüedad de la organización, la capacidad de absorción y la disponibilidad de recursos organizacionales.

- Tamaño de la organización: Numerosos estudios se han realizado para investigar sobre la adopción de tecnologías de la información y el impacto del tamaño de la compañía, encontrando que este factor tienen una gran influencia en su implementación. Sin embargo, estos resultados son variados e inconsistentes; algunos puntos de vista exponen que las pequeñas empresas son más eficientes que las grandes compañías al momento de adoptar innovaciones tecnológicas. Por otra parte, hay quienes argumentan que las grandes organizaciones cuentan con más recursos financieros y humanos y de esta forma pueden satisfacer los requerimientos para la innovación.

En un estudio realizado por Lee y Xia en 2006 se concluyó que el tamaño de la organización y las innovaciones tecnológicas tienen una relación directa. No obstante, esta relación también depende de elementos como el tipo de innovación, el tipo de organización, la etapa de adopción, el alcance de la innovación, entre otros [24]. Algunos autores sugieren que debido al costo y la complejidad de la arquitectura de BI, el tamaño adecuado de una organización para su adopción debe ser superior a los 100 empleados. Por el contrario, desde la perspectiva de los fabricantes de herramientas y soluciones de BI, las aplicaciones en la actualidad son más diversas, flexibles, económicas y menos complejas que en el pasado, por lo que existe una amplia gama de ofertas especialmente dirigidas a las PYMES y sus limitaciones financieras.

- Antigüedad de la organización: Se refiere al tiempo que lleva la compañía en la operación del negocio. En un estudio realizado por Flanagin se encontró que las organizaciones nuevas son más propensas a la adopción de innovación [25]. Adicionalmente, Rogers en 1995 concluyó que a medida que las organizaciones nuevas surgen en un ambiente saturado de tecnologías de la información avanzadas, es natural que se apoyen en dichas herramientas para lograr una ventaja competitiva [26]. En un estudio posterior realizado por Daniel y Myers en el año 2000 se encontró que a mayor antigüedad de una organización, es más probable que se resista a la participación en un proceso de innovación, ya que no puede cambiar fácilmente las actividades rutinarias y es incapaz de adaptar los procesos innovadores en el entorno actual de la organización [27].
- Capacidad de absorción: Griffith et al. en 2003 definieron la capacidad de absorción de una organización como la habilidad de sus miembros para hacer uso de los conocimientos y la información existente [28]. Esta facultad ayuda a las organizaciones a reconocer el valor de la información interna y externa; como resultado, esto se puede traducir en beneficios económicos para la compañía. Las tecnologías de BI requieren del conocimiento y comprensión por parte de los usuarios. Estos requisitos pueden potenciar el desarrollo de las innovaciones de TI en el contexto de las empresas.
- Disponibilidad de recursos organizacionales: Es otro de los factores que se ha identificado como influyentes al momento de adoptar innovaciones tecnológicas. Los directivos apoyan la adopción de nuevas tecnologías cuando los recursos financieros, la infraestructura y el capital humano están disponibles.

5.5.1.3 Características del entorno.

Las características del entorno se basan en el modelo Tecnología-Organización-Entorno. Los factores del entorno son clave en la adopción de innovaciones tecnológicas. Es necesario examinar la influencia de los factores del medio antes de adoptar una innovación; elementos como la presión competitiva o la selección de los proveedores pueden determinar el éxito de su implementación.

- Presión competitiva: Este factor tiende a estimular a las empresas a buscar nuevos enfoques con el objetivo de incrementar su eficiencia y productividad, lo que lleva a las organizaciones a lograr una ventaja competitiva. Waarts et al. en 2002 encontraron que la competencia fue uno de los principales impulsores en la adopción de tecnologías innovadoras [29]. Por ejemplo, en los estudios más recientes sobre las PYMEs, Alshawi et al. en 2011 hallaron que la presión de la competencia era una influencia importante en la adopción de los sistemas CRM (por sus siglas en inglés, *Costumer Relationship Management*) en las organizaciones [30]. Otro estudio sobre la adopción de tecnologías de *Data Warehouse* realizado por Hwang en 2004 reveló que los factores del entorno, incluyendo la presión competitiva y la selección de proveedores, fueron elementos importantes en la implementación de tecnologías de almacenamiento de datos [31].
- Selección de proveedores: En general, los proveedores son responsables de proporcionar el software, hardware, capacitación de usuarios y soporte técnico a los clientes con el fin de mantener un rendimiento óptimo. La selección de los proveedores debe ser abordada teniendo en cuenta factores como: la reputación y experiencia del proveedor, la capacidad técnica y las competencias profesionales de la firma consultora. En el estudio realizado por Hwang et al. los autores encontraron una relación entre la selección de proveedores de BI y la adopción de dicha tecnología [31]. Como BI es diferente de otras tecnologías de la información, se requiere de una solución a la medida para adaptarse a cada empresa en particular y a la industria, y no sólo suites estandarizadas.

5.5.1.4 Características de los propietarios o gerentes.

Las características de los propietarios son fundamentales para el modelo de adopción de sistemas de información en pequeñas empresas propuesto por Thong y Yap en 1996. Ghobakhloo et al. afirmaron en 2011 que las PYMEs suelen tener estructuras simples, con la autoridad centralizada principalmente en el gerente [32]. Por lo tanto, esta persona es quien toma las decisiones que tienen un efecto directo en los procesos, que van desde las funciones diarias a las inversiones futuras. Otro estudio realizado por Thong en 1999 reveló que los propietarios o gerentes que tienen capacidad de innovación y conocimientos de TI poseen mayor potencial de éxito en la adopción de tecnologías innovadoras [33].

- **Innovación:** Se refiere a la voluntad de introducir novedades al negocio a través de la experimentación y el proceso creativo destinado al desarrollo de nuevos productos, servicios y procesos. En un estudio realizado por Chang et al. en 2010, se encontró que, por ejemplo, la capacidad de innovación que poseen los gerentes o propietarios de una organización son un factor determinante para la adopción de sistemas ERP en las pequeñas y medianas empresas [34].
- **Conocimientos de TI:** Se refiere a la experiencia o conocimientos en tecnologías de la información. Este factor involucra los antecedentes de los gerentes en actividades de TI, el reconocimiento de su potencial, así como su capacidad de planificar estratégicamente.

5.5.2 Niveles de madurez de BI.

El mercado moderno, altamente dinámico y competitivo, exige a las pequeñas y medianas empresas contar con tecnologías de Inteligencia de Negocios que les permitan obtener mayor valor de los datos y utilizar la información generada para posicionarse en entornos cada vez más desafiantes. Sin embargo, la probabilidad de fracaso y la gran cantidad de recursos que demandan estas innovaciones tecnológicas, hacen que la implementación de este tipo de sistemas represente un alto riesgo para las PYMEs. En consecuencia, es necesario adoptar un modelo de madurez que permita a las organizaciones evaluar su nivel de preparación e identificar los aspectos a mejorar de cara a la incorporación de soluciones de BI. Estos modelos pueden ayudar a determinar los avances en la incorporación de esta tecnología y por lo tanto, contribuyen a mejorar la tasa de éxito de su implementación en las PYMEs [35].

Los modelos de madurez definen niveles de progreso, eficiencia, capacidad de gestión y medición de los procesos de la organización. Un modelo de madurez de BI se constituye como una herramienta muy valiosa en el proceso de implementación, ya que describe una ruta de acción y apoya a las organizaciones en la definición de una estrategia de BI que contribuya al logro de los objetivos del negocio.

Gartner propone un modelo de madurez en el que se definen cinco niveles para medir el progreso de los esfuerzos que las empresas realizan en pro de la adopción de tecnologías de BI y en ese sentido, determina las características requeridas para alcanzar el máximo nivel de acoplamiento en sus procesos organizacionales. Este modelo se describe a continuación [36]:

- Nivel 1 - inconsciente: En este nivel la organización no cuenta con capacidad de BI. A menudo, es descrito por los autores como “la anarquía de la información”, debido a que los datos no son consistentes en todos los departamentos y no hay métricas identificadas o bien definidas. En este nivel de madurez, se presenta una alta utilización de hojas de cálculo, mientras que el manejo de herramientas de análisis o reportes es limitado. Los principales desafíos en este nivel son identificar los factores relevantes para el negocio, comprender la estructura de gestión de la información actual, el origen y la calidad de los datos, la arquitectura y los sistemas transaccionales.
- Nivel 2 - táctico: En este nivel las organizaciones empiezan a invertir en BI. Las métricas se utilizan generalmente solo a nivel de departamento. Los indicadores generales o comunes en todos los departamentos no existen o son inconsistentes. Los administradores y gerentes carecen de confianza en la calidad o fiabilidad de los datos. En esta etapa, aún se cuenta con poca financiación para los proyectos de Inteligencia de Negocios. En este nivel es importante obtener financiación para iniciativas de BI y definir métricas para analizar el rendimiento departamental y de tareas específicas.
- Nivel 3 - focalizado: En esta fase, la organización empieza a obtener algunos beneficios de la utilización de esta tecnología. Sin embargo, su acción aún es muy limitada. Las métricas se encuentran formalmente definidas y permiten el análisis del desempeño departamental, aunque no están vinculadas a los objetivos organizacionales y sólo se utilizan para evaluar el rendimiento de áreas específicas del negocio. Los datos aún no se encuentran integrados y empiezan a utilizarse los tableros de control. En esta etapa es necesario

aumentar el alcance de la iniciativa de BI y buscar la integración de los datos para la consolidación de herramientas analíticas.

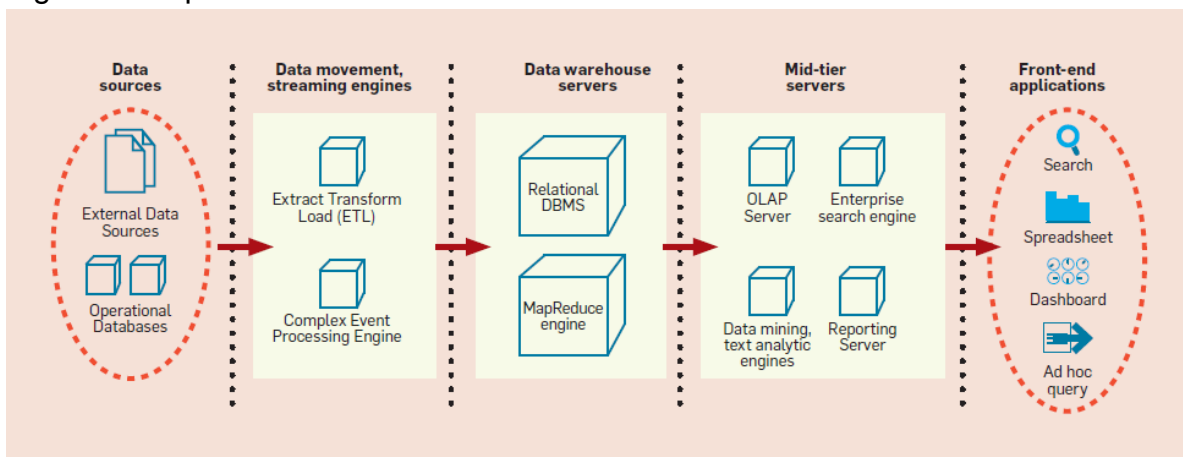
- Nivel 4 - estratégico: En este nivel las empresas definen su estrategia de BI de acuerdo a los objetivos generales de la organización. Integran la Inteligencia de Negocios en los procesos críticos de la compañía, poniendo la información a disposición de los administradores y gerentes. Los datos en el nivel estratégico son de confianza y se utilizan para la toma de decisiones. El desafío principal para esta organización es desarrollar una estructura corporativa equilibrada, alineada con los objetivos del negocio y la estrategia de la compañía.
- Nivel 5 - omnipresente: En este nivel, el BI está presente en todas las áreas de la organización y forma parte de la cultura organizacional y de los procesos del negocio. Los usuarios tienen acceso a la información en múltiples niveles de la organización, lo cual les permite analizar, administrar, innovar y tomar decisiones para mejorar el rendimiento. Los resultados están vinculados a objetivos concretos y medibles.

6. COMPONENTES DE UN SISTEMA DE INTELIGENCIA DE NEGOCIOS

Un sistema de BI está conformado por una serie de elementos que en conjunto proporcionan análisis de datos para tomar decisiones sobre el negocio y permiten explorar nuevas oportunidades a partir del conocimiento obtenido, tales como realizar ofertas, ofrecer servicios personalizados a los mejores clientes, lanzar nuevos productos en mercados potenciales, entre otros.

La arquitectura típica de una plataforma de BI propuesta por Chaudhuri et al. posee cinco capas de aplicación: fuentes de datos, carga de los datos, repositorios de datos, servidores intermedios y herramientas de explotación y visualización de la información [37]. Cada capa contempla diferentes tecnologías que pueden ser implementadas o no dependiendo del alcance de la solución.

Figura 1. Arquitectura de un sistema de BI



Fuente: CHAUDHURI, Surajit; DAYAL, Umeshwar y NARASAYYA, Vivek. An Overview of Business Intelligence Technology. En: Communications of the ACM. Vol. 54. No. 8, p. 90. Agosto, 2011.

Los datos sobre los que se ejecutan las tareas de Inteligencia de Negocios, usualmente provienen de diferentes fuentes, típicamente desde múltiples bases de datos operacionales que se utilizan en los departamentos de la organización, así como de entidades externas. Estos sistemas operacionales también son conocidos como sistemas de procesamiento de transacciones en línea (OLTP, por sus siglas en inglés, *On-Line Transaction Processing*), debido a que procesan un

gran número de transacciones en tiempo real y actualizan los datos cada vez que es necesario. Este tipo de sistemas almacenan solo información actual que se utiliza para apoyar las operaciones diarias del negocio.

Es importante que las organizaciones identifiquen claramente sus fuentes de datos. Saber dónde se pueden obtener los datos requeridos es esencial para el tratamiento de las cuestiones y necesidades específicas del negocio, lo que resulta en ahorro de tiempo significativo y mayor rapidez en la entrega de información.

Estas fuentes pueden diferir en la calidad de la información almacenada y utilizar representaciones inconsistentes, códigos o formatos que se deben homogenizar. Es fundamental que los problemas de integración, limpieza y estandarización de los datos se solucionen correctamente, ya que la carga eficiente de estos es esencial para la implementación de un proyecto de *Business Intelligence*. Además, las tareas de BI necesitan ser ejecutadas constantemente a medida que llegan nuevos datos como, por ejemplo, las ventas del último mes. Las tecnologías de *back-end* utilizadas en la preparación de los datos para la Inteligencia de Negocios son normalmente conocidas como herramientas ETL.

Por otra parte, se ha incrementado la necesidad de soportar las tareas de BI en tiempo real, esto es, tomar decisiones sobre el negocio basadas en los datos operacionales. Los motores especializados conocidos como software de procesamiento de eventos complejos (CEP, por sus siglas en inglés, *Complex Event Processing*) surgieron para soportar dichos escenarios.

Los datos sobre los cuales se realizan las tareas de BI son típicamente cargados a un repositorio llamado *Data Warehouse*. Para el almacenamiento y consulta de estas bodegas de datos se suelen utilizar sistemas de gestión de bases de datos relacionales. Durante las últimas dos décadas se han desarrollado múltiples estructuras de datos, optimizaciones y técnicas para el procesamiento de consultas, principalmente para ejecutar consultas SQL muy complejas sobre grandes volúmenes de datos [37].

Debido a que gran parte de los datos se encuentran almacenados en medios digitales, existe un aumento en la implementación de sistemas que puedan soportar mayores volúmenes de información que los almacenados actualmente en los sistemas relacionales, conocidos como *Big Data*. Con este objetivo en mente, los motores basados en el paradigma *Map Reduce*, que fueron originalmente

construidos para el análisis de documentos web y para el registro de consultas de búsqueda en la web, son ahora objeto del análisis empresarial. Estos motores son actualmente extendidos para soportar consultas complejas similares al SQL, que son esenciales para los entornos de *Data Warehouse* tradicionalmente utilizados.

Los servidores de *Data Warehouse* se complementan con un conjunto de servidores intermedios, que proporcionan funcionalidades especializadas para diferentes escenarios de BI. Las herramientas para el procesamiento OLAP proporcionan una vista multidimensional de los datos para las aplicaciones o usuarios y permiten realizar operaciones comunes de BI tales como *drill-down*, *roll-up*, *slice*, *dice* y *pivot*, las cuales serán descritas en capítulos posteriores.

Los servidores de reportes son ampliamente utilizados para apoyar la toma de decisiones y medir el rendimiento. Las empresas los utilizan para la consolidación financiera, para la evaluación de las estrategias y políticas, o únicamente para la presentación de informes simples.

Adicionalmente el incremento y la disponibilidad de datos de texto ha ocasionado que los motores de búsqueda empresariales soporten el paradigma de búsqueda por palabras clave sobre texto y datos estructurados en la bodega de datos para, por ejemplo, encontrar mensajes de correo, documentos, historial de compras y llamadas de soporte relacionadas a un cliente en particular. Asimismo, los motores de minería de datos permiten realizar un análisis de profundidad en los datos, que va más allá de lo ofrecido por las herramientas OLAP o los servidores de reportes y proporcionan la capacidad de construir modelos predictivos para ayudar a responder cuestiones del negocio.

El proceso de minería de datos se puede realizar con la integración de las bodegas de datos y los servidores OLAP. Debido a que la cantidad de datos en una organización presenta un rápido crecimiento, resulta necesario implementar soluciones de *Data Mining* para tomar decisiones de forma más eficiente. En pocas palabras, se trata de un método que identifica automáticamente información útil tal como patrones inusuales, tendencias y relaciones que se encuentran ocultas en grandes cantidades de datos. Esto se puede lograr mediante la aplicación de técnicas estadísticas como clasificación, análisis de series de tiempo y agrupación [38].

Los motores de análisis de texto pueden procesar grandes cantidades de datos para extraer información valiosa que de otra manera requeriría un esfuerzo

manual significativo; por ejemplo, cuáles productos son mencionados en las respuestas de las encuestas y qué temas son frecuentemente discutidos en relación con esos productos.

Existen muchas aplicaciones populares de *front-end*, a través de las cuales los usuarios realizan análisis de la información, entre las que se pueden encontrar hojas de cálculo, portales empresariales para realizar búsquedas, aplicaciones de gestión del rendimiento que permiten a los directivos hacer seguimiento a los indicadores claves de desempeño del negocio, herramientas para la definición de consultas *ad hoc*, visualizadores de modelos de minería de datos, entre otras.

Además se pueden mencionar otras tecnologías como *Web Analytics* que permite analizar el comportamiento de los visitantes de la página web de una compañía y su interacción con ella, dando a conocer, por ejemplo, qué vistas son las que animan al visitante a realizar compras; igualmente, las aplicaciones modulares como los CRM son ampliamente utilizadas.

En síntesis, para que una solución de BI funcione eficientemente, las cinco capas descritas previamente deben estar enlazadas de manera sistemática. Los datos que se originan de fuentes externas e internas deben ser extraídos, transformados y cargados en el *Data Warehouse*. Cuando la bodega de datos se desarrolla para la utilización por parte de toda la organización, los datos son enviados a repositorios de información más pequeños (*Data Marts*) para satisfacer las necesidades operacionales de un área del negocio específica. En la capa de usuario final, la información que reside en las bodegas de datos y *Data Marts* se puede acceder utilizando una variedad de tecnologías tales como aplicaciones analíticas, herramientas de visualización de los datos y generación de reportes.

Para la ejecución de un proyecto de Inteligencia de Negocios, también se pueden incorporar otras tecnologías que no han sido contempladas en esta revisión teórica debido a que no han sido ampliamente utilizadas o se trata de tecnologías nuevas que aún se encuentran en proceso de investigación. La selección de las herramientas a incorporar en la solución de BI depende única y exclusivamente de las características específicas del negocio y de los objetivos que se pretenden alcanzar con su implementación.

7. BODEGAS DE DATOS

La construcción del repositorio de datos es un proceso fundamental para la implementación de una solución de Inteligencia de Negocios; por esta razón, es necesario realizar una revisión de ciertos conceptos antes de abordar este tipo de proyectos.

Las bodegas de datos han sido incorporadas en la arquitectura de los sistemas de BI desde la década de los 90's, actuando como repositorios de datos para la recuperación de información útil a partir de grandes cantidades de datos almacenados en plataformas heterogéneas, para la ejecución de consultas y análisis complejos sobre los datos relevantes, sin necesidad de sobrecargar las bases de datos operacionales.

La utilización del *Data Warehouse* es un fenómeno que se incrementó debido a la enorme cantidad de datos generados en los últimos años, además de la urgente necesidad de utilizarlos para obtener un mayor conocimiento sobre el comportamiento del negocio y tomar decisiones más acertadas.

7.1 DEFINICIÓN

El término *Data Warehouse* fue acuñado por primera vez en 1992 por W. H. Inmon³ en su libro *Building the Data Warehouse*, donde menciona que: “Un *Data Warehouse* es un conjunto de datos integrados, orientados a un tema específico, que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de la administración” [39].

Ralph Kimball⁴ propone una definición mucho más simple en su libro *The Data Warehouse Toolkit*: “Un *Data Warehouse* es una copia de los datos transaccionales, especialmente estructurada para la consulta y análisis” [40].

Una de las definiciones que se enmarca en términos de la finalidad de un *Data Warehouse*, aparece en el documento Hefesto: Metodología para la Construcción

³ Considerado el padre del concepto de bodegas de datos.

⁴ Es el principal proponente del enfoque dimensional para el diseño de bodegas de datos.

de un *Data Warehouse*: “El *Data Warehousing*, es el encargado de extraer, transformar, consolidar, integrar y centralizar los datos que una organización genera en todos los ámbitos de su actividad diaria (compras, ventas, producción, etc.) y/o información externa relacionada” [41].

El *Data Warehousing* es un procedimiento que posibilita la extracción de datos desde múltiples fuentes operacionales, la integración y homogeneización de información almacenada en tipos y formatos diferentes, para la consolidación de un repositorio central que contiene datos actuales, históricos y sumariados que permiten soportar los procesos de toma de decisiones en los niveles tácticos y estratégicos de la organización.

7.2 ¿POR QUÉ UTILIZAR UN DATA WAREHOUSE?

Una vez que las organizaciones han desarrollado los sistemas de información necesarios para el manejo de sus procesos básicos de operación, estas exigen nuevas prestaciones de sus herramientas que permitan realizar análisis sobre los datos almacenados.

Normalmente, la información que se requiere estudiar sobre una determinada área de la organización se encuentra en bases de datos y otros recursos muy diversos, tanto internos como externos. Algunas de estas fuentes se obtienen a partir de las bases de datos transaccionales que soportan el trabajo diario de los usuarios. Sobre estas ya es posible extraer conocimiento, sin embargo esta práctica también genera una serie de inconvenientes.

En un escenario típico, una gran corporación tiene múltiples áreas de trabajo, y los altos directivos necesitan cuantificar y evaluar el desempeño de cada una de ellas. Los sistemas operacionales proporcionan información detallada sobre las tareas realizadas por estas unidades. Para satisfacer las necesidades del nivel estratégico de la compañía, se diseñan consultas a la medida. Los administradores de bases de datos deben formular la consulta deseada (normalmente en lenguaje SQL), después de estudiar de cerca los catálogos de datos. Luego se procesa la consulta, que puede tardar un par de horas debido a su complejidad, la enorme cantidad de datos y los efectos de otras consultas concurrentes sobre la base de datos. Por último, se genera un reporte con la información solicitada.

Los expertos se percataron de que este enfoque es poco eficiente, ya que es muy exigente en términos de tiempo y recursos, y no siempre logra los resultados esperados. Por otra parte, la ejecución de consultas analíticas y consultas transaccionales rutinarias retrasa inevitablemente el sistema generando incomodidades para ambos tipos de usuarios. Además, las bases de datos operacionales están diseñadas para optimizar los procesos de inserción, actualización y eliminación de registros. Por lo que si estas son utilizadas para el análisis de datos, se obtendrán tiempos de respuesta poco favorables y el rendimiento de los sistemas se verá afectado.

Las bodegas de datos son diseñadas con un objetivo de explotación distinto al de las bases de datos transaccionales. Su construcción está orientada a la consulta de información sobre las actividades de la organización para operar eficientemente sobre los datos almacenados y facilitar el análisis en los niveles tácticos y estratégicos.

Debido a que los costos de almacenamiento masivo y conectividad se han reducido drásticamente en los últimos años, resulta razonable integrar los datos en un sistema separado, de uso específico. El almacenamiento de datos utilizado en la actualidad realiza el procesamiento analítico en línea (OLAP), separado del procesamiento transaccional en línea (OLTP), mediante la creación de un nuevo repositorio de información que reúne los datos básicos de diversas fuentes, organiza correctamente los formatos de estos datos, y los pone a disposición para el análisis y evaluación orientados a los procesos de planificación y toma de decisiones.

Aunque la implementación de un almacén de datos no es imprescindible para realizar extracción de conocimiento a partir de un conjunto de datos, las ventajas de su desarrollo se pueden percibir de manera significativa en el mediano y largo plazo cuando:

- Se cuenta con grandes volúmenes de datos.
- El tamaño de los datos incrementa con el tiempo.
- Los datos provienen de fuentes heterogéneas.

Actualmente, los almacenes de datos y las herramientas OLAP hacen parte de las tecnologías más utilizadas para integrar, transformar y combinar los datos con el objetivo de facilitar al usuario o a otros sistemas el análisis de la información. El

almacén de datos pretende dar un soporte a la organización para proporcionarle una buena gestión de sus datos. A continuación se presentan algunos de los beneficios e inconvenientes asociados a la implementación de proyectos de *Data Warehouse*.

Figura 2: Ventajas e inconvenientes de las bodegas de datos.

Ventajas
Reúne y consolida las diferentes bases de datos que se mantienen en los diferentes departamentos o áreas funcionales de la empresa.
Proporciona un acceso fácil y flexible a la información.
Asegura la calidad y eficiencia de las decisiones que se toman en las organizaciones.
Mejora la productividad de las empresas.
Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar patrones ocultos entre los datos de la bodega, obteniendo un valor agregado para el negocio.
Permite una rápida reacción frente a los cambios del mercado.
Inconvenientes
Se requiere una gran inversión para la implementación de este tipo de proyectos.
Infravaloración del esfuerzo necesario para su diseño y creación.
Infravaloración de los recursos necesarios para la captura, carga y almacenamiento de los datos.
Incremento continuo de los requerimientos de los usuarios.

Fuente: Elaboración propia.

7.3 CARACTERÍSTICAS DE LAS BODEGAS DE DATOS

Un *Data Warehouse* debe incorporar una serie de características con el objetivo de ofrecer a los usuarios información útil para la toma de decisiones. Las principales características de una bodega de datos referidas a continuación, se obtienen de la definición propuesta por W.H Inmon:

- Orientada al tema: La organización de los datos en un *Data Warehouse* se basa en áreas de interés, para proporcionar información entorno a los principales temas del negocio: Clientes, productos, proveedores, ventas, etc. Su principal objetivo es soportar los procesos de toma de decisiones, enfocándose en los asuntos críticos de la empresa.

Las bodegas de datos incluyen solo la información que debe ser empleada en el procesamiento de datos, utilizando un modelo orientado a la definición de variables de análisis entorno a los conceptos de la empresa, que ofrece una mayor flexibilidad y rapidez de acceso, siempre que se requiera manipular grandes volúmenes de información.

- Integración: Las bases de datos operacionales almacenan la información de maneras disímiles, utilizando diferentes formatos, estilos y nomenclaturas. Por lo tanto, desde un punto de vista funcional, estas no pueden ser empleadas con el propósito de realizar análisis y generar reportes. Un *Data Warehouse* es un proyecto empresarial, que reúne datos de gran parte de estas fuentes operacionales de la organización, que deben ser procesados para generar una versión unificada y consistente de la información, para permitir a los analistas concentrarse en la utilización de los datos, sin tener que cuestionarse respecto a su confiabilidad.
- Variante en el tiempo: La información almacenada en las bases de datos operacionales es actualizada periódicamente para presentar el estado actual del negocio. Por otra parte, las bodegas de datos mantienen datos históricos, que permiten el análisis de tendencias y patrones para realizar pronósticos más exactos.

En los sistemas operacionales, la información es sustituida en cada transacción y los valores antiguos usualmente se pierden. Por el contrario, las bodegas de datos no son actualizadas sino que la información es cargada regularmente para mostrar un registro histórico. Los datos en un *Data Warehouse* muestran con precisión el estado del negocio en diferentes momentos, lo cual permite realizar comparaciones entre diferentes instantes de tiempo.

- No volátil: Los datos en un *Data Warehouse* son estáticos, es decir, no se actualizan una vez están almacenados. Las bodegas de datos se almacenan de forma independiente a los sistemas transaccionales, de manera que los cambios realizados en dicho entorno no alteran el estado de la bodega.

En los repositorios de datos, la manipulación básica de la información se realiza a través de dos tipos de operaciones: la carga de datos y el acceso a los mismos. Mientras que en los sistemas operacionales, la información es

volátil puesto que las transacciones procesadas implican la actualización de los datos: la adición de nuevos registros, la modificación o eliminación de otros existentes.

7.4 DATA MART

Un *Data Mart* se define como: “la implementación de un *Data Warehouse* con un alcance restringido a un área funcional, problema en particular, departamento, tema o grupo de necesidades” [41].

La implementación de un entorno de *Data Warehousing* puede ser una tarea desafiante, que conlleva un largo período de tiempo. Muchos departamentos y áreas de negocio ansían obtener los beneficios de las bodegas de datos y se reúsan a esperar la evolución normal de su construcción. El concepto de *Data Mart* existe para satisfacer las necesidades de este tipo de actores. Uno de los propósitos para la construcción de un *Data Mart* es desarrollar un prototipo tan pronto como sea posible sin la necesidad de esperar la implementación de un *Data Warehouse* corporativo, debido a que estos son más pequeños y fáciles de construir.

Los almacenes de datos son una forma simple de *Data Warehouse* enfocados en un solo tema o área funcional, tal como ventas, finanzas, marketing, entre otros. Estos son contruidos y controlados comúnmente por un único departamento dentro de la organización.

Los *Data Marts* por lo general contienen un subconjunto de información que resulta importante para una unidad específica del negocio, un departamento o un conjunto de usuarios. Pueden contener datos históricos, sumarizados o detallados, provenientes de los sistemas de procesamiento de transacciones (*Data Marts* independientes), o de bodegas de datos existentes (*Data Marts* dependientes). Estos conceptos serán presentados con mayor detalle en este capítulo.

7.5 PROCESOS ETL

Una de las etapas fundamentales durante la construcción de un *Data Warehouse* es el poblado de los datos. Factores como la calidad de la información, la falta de datos de origen, la poca documentación de los sistemas operacionales, las múltiples fuentes heterogéneas o la información redundante pueden complicar significativamente este proceso.

La información almacenada en un *Data Warehouse* reúne los datos provenientes de diferentes sistemas transaccionales de la organización, que deben ser extraídos, para su posterior transformación, filtrado y finalmente ser cargados al repositorio. Este tipo de operaciones se conocen como procesos ETL.

Para cargar los datos al *Data Warehouse* se deben realizar los siguientes procedimientos:

- **Extracción:** Los datos relevantes son obtenidos desde las fuentes operacionales durante la fase de extracción. Se denomina extracción estática cuando la bodega de datos necesita ser poblada por primera vez. La extracción incremental se utiliza para actualizar la información del repositorio con regularidad, almacenando los cambios aplicados a los datos de origen desde la última extracción.

Si la información procede de una base de datos relacional, el proceso de extracción se puede limitar a consultas SQL o procedimientos almacenados. Por el contrario, si se encuentra en otro tipo de fuentes, ya sean sistemas externos, archivos de texto u hojas de cálculo, la obtención de los datos puede ser más compleja, debido a que se tendrán que realizar conversiones de formato o utilizar aplicaciones especializadas.

- **Transformación:** Durante el proceso de transformación de los datos se garantizará en gran medida la calidad de la información que será almacenada en el *Data Warehouse*. En esta etapa se pueden identificar dos actividades: la limpieza y la transformación de los datos.

Con la limpieza de los datos se corrigen errores como registros duplicados, información faltante, valores inconsistentes, y en general todos los errores que son frecuentes cuando se manejan múltiples fuentes o se agregan datos de forma manual.

Durante la transformación se convierten los datos de forma que sean compatibles y congruentes, para que puedan ser cargados al *Data Warehouse*. Esta actividad se realiza porque generalmente existen diferentes fuentes de información y es necesario definir estándares que garanticen la consistencia de la información almacenada.

- Carga: La carga de los datos al *Data Warehouse* es el último paso del proceso, este se puede realizar de dos formas diferentes:
 - Reemplazo: La información en la bodega de datos es completamente re-escrita, esto significa que los datos antiguos son reemplazados. Generalmente se utiliza en combinación con la extracción estática para el poblado inicial del *Data Warehouse*.
 - Actualización: Solo los cambios realizados en los datos fuentes son aplicados al *Data Warehouse*. La actualización se realiza generalmente sin modificar o eliminar los datos existentes. Esta técnica se utiliza en conjunto con la extracción incremental para actualizar las bodegas de datos con regularidad.

7.6 ARQUITECTURAS DE DATA WAREHOUSE

La elección de la arquitectura es una actividad que se debe realizar previo a la construcción de un *Data Warehouse*. La arquitectura puede ser modificada en el transcurso de su implementación. Sin embargo, la definición de nuevas actividades genera retrasos en la ejecución del proyecto.

La selección de la arquitectura es una decisión administrativa que será determinada por factores como la infraestructura actual, el entorno del negocio, la estructura de control deseada, el alcance de la implementación, la capacidad técnica que posea la organización y los recursos disponibles [42].

La arquitectura del *Data Warehouse* determinará la ubicación del mismo y su política de control. Por ejemplo, los datos pueden residir en un único punto que es administrado de forma centralizada. Alternativamente, la información puede estar distribuida localmente o en lugares remotos que son controlados de forma central o por unidades independientes.

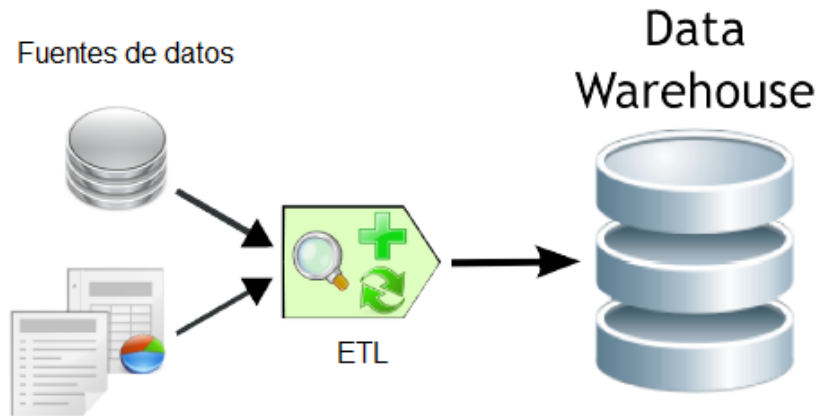
7.6.1 *Data Warehouse* global o empresarial

Un *Data Warehouse* empresarial es aquel que soporta todos los requerimientos del negocio, o gran parte de ellos, para un entorno de almacenamiento de datos más integrado, con un alto grado de acceso y utilización de los datos por parte de todos los departamentos o líneas del negocio. Este ha sido diseñado y construido en función de las necesidades de la empresa en su conjunto. Se considera como un repositorio de datos común para la toma de decisiones que está disponible a través de toda la organización.

Una percepción errónea que se presenta con frecuencia es que un *Data Warehouse* global se encuentra físicamente centralizado. El término global es utilizado para reflejar el alcance del acceso y utilización de los datos, más no la estructura física. Este tipo de *Data Warehouse* se caracteriza por tener todos los datos bajo una administración central. Sin embargo, la bodega de datos puede estar físicamente centralizada o distribuida en toda la organización. Esto es, se encuentra centralizada lógicamente más que físicamente. Cuando se presenta este caso se conoce también como una implementación del tipo *hub and spoke*. El punto clave es que el entorno se gestiona como una única entidad integrada [43].

Una arquitectura de almacén global permite a los usuarios finales tener una vista general de los datos de toda la organización. Se debe tener la certeza de que este es un requerimiento indispensable, ya que el desarrollo de este tipo de entornos puede demandar mucho tiempo y su implementación puede resultar muy costosa.

Figura 3: *Data Warehouse* empresarial



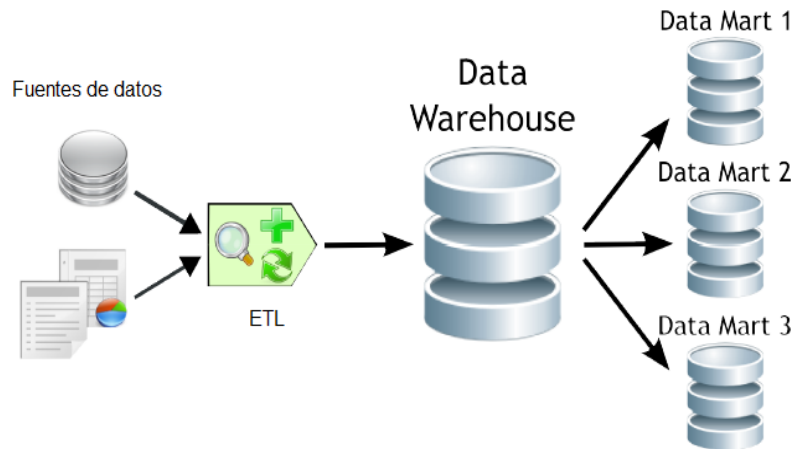
Fuente: Adaptado de BERNABÉU, Ricardo Darío. Hefesto. DATA WAREHOUSING: Investigación y Sistematización de Conceptos. HEFESTO: Metodología para la Construcción de un Data Warehouse. p. 74. Córdoba, Argentina. Julio, 2010.

7.6.2 *Data Marts* dependientes

En esta arquitectura, la información es almacenada en un *Data Warehouse* global que puede ser accedido directamente o mediante la creación de pequeños *Data Marts*, en los cuales se replican los datos de manera parcial para el uso específico de los departamentos.

Las bodegas de datos corporativas actúan como un sistema de almacenamiento centralizado. Por otra parte, los *Data Marts* pueden ser vistos como repositorios de menor tamaño, que almacenan los datos que provienen de un *Data Warehouse* primario, y son requeridos para un dominio de aplicación específico.

Figura 4: *Data Marts* dependientes



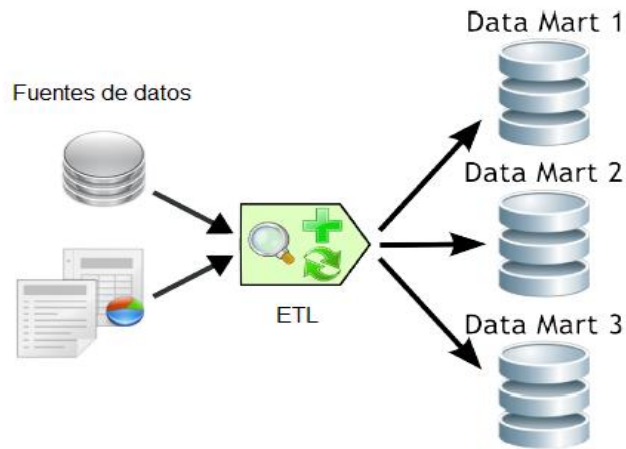
Fuente: Adaptado de BERNABÉU, Ricardo Darío. Hefesto. DATA WAREHOUSING: Investigación y Sistematización de Conceptos. HEFESTO: Metodología para la Construcción de un Data Warehouse. p. 74. Córdoba, Argentina. Julio, 2010.

7.6.3 *Data Marts* independientes

En esta arquitectura, los almacenes de datos son administrados por una determinada área de trabajo, departamento o proceso del negocio y son contruidos exclusivamente para satisfacer sus necesidades específicas. Típicamente, no existe ningún tipo de conexión o conectividad entre los *Data Marts* de departamentos o áreas del negocio diferentes; es decir, los datos son generados y administrados internamente [43]. Por lo tanto, estos *Data Marts* generalmente no poseen dimensiones y hechos comunes entre ellos, lo cual puede resultar en fuentes de datos inconsistentes que conducen a una toma de decisiones inexacta.

En la arquitectura de *Data Mart* independiente, los almacenes son contruidos de manera separada sin ningún tipo de integración. Esta iniciativa puede ser muy beneficiosa para aquellas empresas que no poseen una alta financiación para este tipo de proyectos y por lo tanto presentan limitaciones para la construcción de un *Data Warehouse* corporativo. También es una alternativa muy utilizada cuando las diferentes áreas o departamentos del negocio exteriorizan un débil acoplamiento.

Figura 5: *Data Marts* independientes



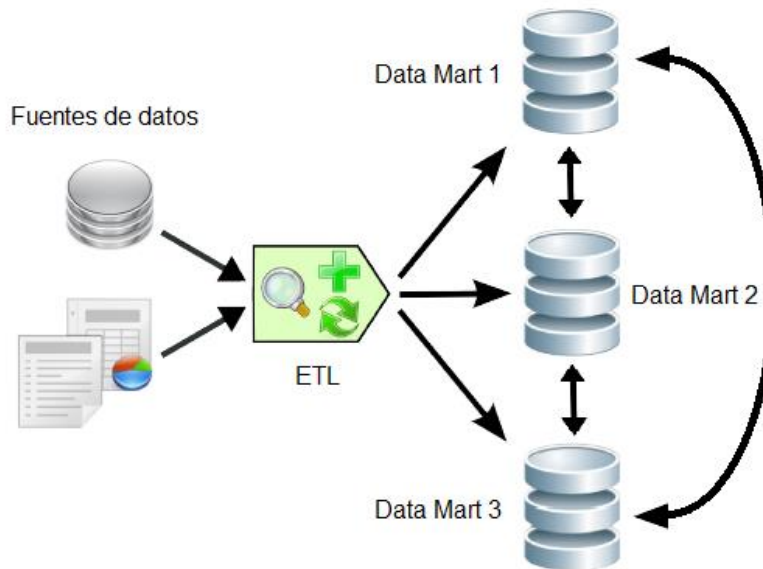
Fuente: Adaptado de BERNABÉU, Ricardo Darío. Hefesto. DATA WAREHOUSING: Investigación y Sistematización de Conceptos. HEFESTO: Metodología para la Construcción de un Data Warehouse. p. 74. Córdoba, Argentina. Julio, 2010.

Este tipo de arquitectura por lo general tiene un impacto mínimo sobre los sistemas operacionales existentes y su implementación puede resultar muy ágil. Sin embargo, la integración de los datos y la falta de una perspectiva global pueden representar una limitante para su construcción, pues los datos de un *Data Mart* en particular solo podrán ser accedidos para los usuarios que hacen parte del grupo de trabajo, departamento o línea de negocio encargado de su implementación [42].

7.6.4 *Data Marts* interconectados

La arquitectura de *Data Marts* interconectados es básicamente una implementación distribuida. Aunque los *Data Marts* son construidos de manera independiente por cada área de negocio, estos son integrados o interconectados para proporcionar una visión general de todos los datos de la organización. En el máximo nivel de integración, la combinación de todos los *Data Marts* interconectados pueden generar un almacén de datos empresarial distribuido. Como consecuencia, los usuarios finales de un determinado departamento pueden acceder y utilizar los datos almacenados en un *Data Mart* de otro departamento o área del negocio.

Figura 6: *Data Marts* interconectados



Fuente: Adaptado de BERNABÉU, Ricardo Darío. Hefesto. DATA WAREHOUSING: Investigación y Sistematización de Conceptos. HEFESTO: Metodología para la Construcción de un Data Warehouse. p. 74. Córdoba, Argentina. Julio, 2010.

Los almacenes de datos interconectados pueden ser administrados de forma individual por cada departamento. Cada área de negocio decide de forma independiente cuáles serán sus fuentes de datos para cargar en el *Data Mart*, cuándo serán actualizados, quién puede acceder a ellos y dónde residirán. También pueden optar por proporcionar las herramientas y habilidades necesarias para implementar el *Data Mart* por sus propios medios.

En este caso, se requeriría un mínimo de recursos del departamento de TI; este podría, por ejemplo, proporcionar ayuda en la implementación de un esquema de seguridad transversal a todas las áreas, en la generación de copias de seguridad y definición de los procesos de recuperación, así como los aspectos de conectividad de red de la aplicación. En contraste, los *Data Marts* interconectados podrían ser controlados y gestionados por el área de TI. Cada grupo de trabajo, área o línea de negocio tendría su propio repositorio, pero las herramientas, habilidades y recursos necesarios para implementar los *Data Marts* serían proporcionados por el departamento de TI [42].

8. ANÁLISIS DE DATOS

El análisis de los datos es típicamente el último nivel de un sistema de Inteligencia de Negocios. Después de la limpieza, integración y transformación de los datos se debe determinar una forma efectiva para generar valor a partir de la información almacenada. Un *Data Warehouse* está diseñado para permitir un fácil acceso a las fuentes de datos; es decir, es un medio que facilita la obtención de información de calidad, pero no puede proporcionarla por sí solo. En este nivel, existe la necesidad de realizar análisis a las fuentes de datos para obtener información que pueda ser útil en el proceso de toma de decisiones. El análisis de datos es utilizado para formular consultas, presentar sus resultados y para estudiar el contenido de los datos mediante su visualización desde diferentes perspectivas.

Todas las organizaciones producen reportes regulatorios, legales e internos de manera frecuente. Los reportes o informes son documentos predefinidos compuestos de tablas, resúmenes y gráficas, que contienen información empresarial representada mediante indicadores organizados en filas y columnas. Los usuarios objetivo de este tipo de reportes son auditores, entidades reguladoras gubernamentales, accionistas, personal del área de finanzas o de la alta gerencia de la organización. Existe entonces una necesidad de información que sea presentada en formatos que puedan ser fácilmente interpretados por los usuarios en todos los niveles de la organización.

La planeación y gestión de las operaciones del negocio por medio de informes mensuales quedó atrás. Hoy en día las organizaciones necesitan desarrollar aplicaciones de información de manera ágil para proporcionar a los usuarios del negocio acceso a la información que refleje el entorno cambiante de la organización.

Los diferentes proveedores comerciales de herramientas de BI tienen diversas opiniones de lo que esta debería incluir. Sin embargo, existen algunos componentes claves que están presentes en todas las distribuciones comerciales. En un estudio realizado por Aliyev en 2008, la mayor parte de las funcionalidades propuestas por el autor fueron consideradas por los entrevistados como parte esencial de una solución de Inteligencia de Negocios [44]. En esta sección se presentarán las herramientas de análisis más destacadas en la literatura, que a su vez coinciden con las que fueron planteadas por Aliyev en su investigación.

8.1 HERRAMIENTAS DE CONSULTA Y REPORTE

El análisis de consultas y la generación de reportes son procesos en los que se plantean cuestiones del negocio a resolver, recuperando datos relevantes desde el *Data Warehouse*, transformándolos en el contexto adecuado y presentándolos en formatos legibles. Estos son generalmente requeridos por los analistas del negocio quienes están muy familiarizados con la definición de este tipo de sentencias para obtener las respuestas a sus preguntas.

Un reporte está definido por una consulta y un formato [45]. Una consulta generalmente implica una restricción y una agregación de los datos multidimensionales. La definición de una consulta es el proceso de tomar una pregunta o hipótesis del negocio y traducirlo en un lenguaje de consulta que puede ser interpretado por una herramienta de soporte a la decisión. Cuando se ejecuta la consulta, la herramienta genera los comandos apropiados para acceder y recuperar los datos solicitados [43]. A continuación, el analista realiza los cálculos y las manipulaciones necesarias en la respuesta para obtener los resultados deseados. Estos resultados se formatean y son organizados en una plantilla para facilitar la comprensión por parte del usuario. Estas plantillas pueden consistir de texto, graficas, video y audio [42]. Por último, el informe se entrega al usuario en el medio requerido, ya sea impreso o en un medio digital.

Este enfoque está orientado a aquellos usuarios que necesitan tener acceso regular a la información de una manera estática. El formato para este tipo de reportes es predeterminado y puede variar cuando se presentan cambios en la legislación o leyes vigentes. Los diseñadores pueden realizar consultas para generar informes con el formato deseado y “congelarlos” en una aplicación [45]. De esta manera, los usuarios finales pueden consultar los datos actuales siempre que lo necesiten.

Los usuarios de reportes están interesados principalmente en el procesamiento de valores numéricos, los cuales son utilizados para analizar el comportamiento de los procesos del negocio, tales como ingresos por ventas, cantidades de envío, entre otros. También pueden monitorear medidas de calidad tales como el nivel de satisfacción del cliente, retrasos en los procesos del negocio y envíos tardíos o erróneos. Pueden evaluar los efectos de las transacciones del negocio, analizar tendencias o extrapolar sus predicciones para el futuro. A menudo, los datos

mostrados harán que el usuario formule otra pregunta para aclarar las repuestas o para obtener información más detallada. Este proceso continua hasta que se obtienen los resultados deseados.

Los reportes existen desde mucho antes que los *Data Warehouses* y han sido siempre la principal herramienta utilizada por los administradores para la evaluación y planificación de las tareas desde la invención de las bases de datos [45]. Sin embargo, la introducción de las bodegas de datos resulta muy útil por dos razones principales: en primer lugar, los datos presentados en los reportes son coherentes y confiables debido a que los proyectos de implementación de un *Data Warehouse* deben incorporar procesos para asegurar la consistencia de la información almacenada. Además, las bodegas de datos agilizan el proceso de generación de reportes porque la separación de la arquitectura de procesamiento de transacciones y el entorno de análisis mejora significativamente el rendimiento.

8.2 SPREADSHEET

Una de las herramientas más utilizadas para el análisis son las hojas de cálculo. Estas aplicaciones fueron inicialmente desarrolladas para los contadores, pero en la actualidad son utilizadas por cualquier persona que requiera preparar un presupuesto, pronosticar los datos de las ventas, presentar estados financieros y en general aplicaciones que requieran cálculos matemáticos.

Las hojas de cálculo son una herramienta muy flexible y poderosa y son especialmente útiles para el procesamiento de números. Por lo tanto, estas aplicaciones están presentes en casi todas las empresas del mundo [43]. Las hojas de cálculo también son ampliamente utilizadas en aplicaciones científicas y financieras. Por ejemplo, una hoja de cálculo puede almacenar datos de una cuenta bancaria, incluyendo los estados financieros e información de interés. Una columna que almacena los saldos de las cuentas de varios clientes puede ser fácilmente resumida para calcular el valor total de los saldos de todos los clientes. Estas cantidades pueden multiplicarse por la tasa de interés para ver cuál será el valor de las cuentas al cabo de un año. Una vez que la formula ha sido creada, la modificación del valor de una celda cambiará el valor proyectado en todas las cuentas.

La utilización de hojas de cálculo trae buenas y malas consecuencias para la organización. Por una parte, permite a los usuarios ser más autosuficientes. Sin embargo, puede dar lugar a una multitud de fuentes de datos independientes, no integrados y no compartidos que existen en cualquier empresa.

Algunos ejemplos del uso de hojas de cálculo son [43]:

- Reportes financieros, como listas de precios o inventarios.
- Funciones analíticas y matemáticas.
- Control estadístico de procesos, que se utiliza a menudo para supervisar y controlar la calidad.

8.3 OLAP

OLAP es quizás la forma principal para explotar la información en una bodega de datos. Es la más popular y ofrece a los usuarios finales, cuyas necesidades de análisis no son fáciles de definir de antemano, la oportunidad de analizar y explotar los datos de forma interactiva sobre la base del modelo dimensional [45].

El análisis multidimensional se ha convertido en una forma muy popular para ampliar las capacidades de consulta y presentación de informes. Es decir, en lugar de ejecutar varias consultas, los datos son estructurados para permitir un acceso fácil y rápido a la información que resuelve las cuestiones realizadas por los usuarios. Por ejemplo, los datos estarán estructurados para incluir la respuesta a la pregunta: “¿Cuántas unidades de determinado producto se venden en un día determinado, por un vendedor en particular, en una tienda específica?”. Cada parte separada de esa consulta se denomina dimensión.

Mientras los usuarios de las herramientas de reportes usualmente desempeñan un papel pasivo, los usuarios OLAP son capaces de iniciar una sesión de análisis complejo de forma activa, donde cada paso es el resultado de los pasos anteriores [45]. Las propiedades en tiempo real de las sesiones de OLAP requieren un conocimiento profundo de los datos y las consultas que pueden ser ejecutadas; el diseño de las herramientas para las personas que no están familiarizados con la tecnología juega un papel crucial. La interfaz gráfica de usuario de estas herramientas debe ser flexible, intuitiva y eficaz.

8.3.1 Implementaciones de OLAP

En la literatura generalmente se describen dos modelos de bases de datos diferentes en los cuales se basa la arquitectura de las herramientas OLAP [46]. El primer modelo descrito es el modelo de bases de datos relacionales estándar. Este modelo es utilizado con los sistemas de gestión de bases de datos relacionales contemporáneos (RDBMS, por sus siglas en inglés *Relational Database Management System*) que se utilizan para implementar gran parte de las bases de datos corporativas actuales. En el modelo de bases de datos relacional, la base de datos es una colección de tablas de dos dimensiones, donde cada fila de la tabla representa un registro de la base de datos.

Por otra parte, un modelo de bases de datos multidimensional puede ser visualizado como un cubo con un determinado número de dimensiones. Por lo general, un modelo dimensional consiste de más de tres dimensiones y se conoce como un hipercubo. Sin embargo, un hipercubo es difícil de visualizar, por lo que un cubo es el término más comúnmente utilizado.

La principal diferencia entre los dos modelos es el método de búsqueda. En el modelo relacional, con el fin de localizar un registro, se debe realizar una búsqueda en la tabla de hechos. La velocidad de la búsqueda depende de factores como la forma en que están ordenados los registros o si la tabla se encuentra indexada. En el cubo multidimensional cada registro puede ser localizado directamente, eliminando la necesidad de una búsqueda. Esto se debe a que cada celda tiene una dirección exacta, compuesta de los valores de los atributos de cada dimensión.

Existen diferentes categorías de implementación de las herramientas OLAP, dependiendo del modelo de bases de datos utilizado, los cuales serán descritos a continuación.

8.3.1.1 ROLAP.

Una categoría importante dentro de la tecnología OLAP son las herramientas OLAP relacionales, generalmente conocidas como herramientas ROLAP. La implementación de ROLAP ofrece las mismas funcionalidades comunes en OLAP. Las consultas son creadas de una manera estándar, simplemente a través de un clic. El servidor ROLAP traduce las consultas a lenguaje SQL y esta se envía a la bodega de datos alojada en el RDBMS. Allí la consulta es ejecutada y el conjunto

de datos resultante se envía al servidor ROLAP y luego a la herramienta OLAP para su presentación al usuario final.

La adopción de tecnología relacional para las bodegas de datos tiene una base sólida si se tiene en cuenta la cantidad de literatura escrita sobre este modelo, la amplia experiencia corporativa en gestión de bases de datos relacionales y el rendimiento y flexibilidad de los RDBMS. El poder expresivo del modelo relacional, sin embargo, no incluye los conceptos de dimensión, medida y jerarquía, por lo que se deben crear esquemas específicos para representar el modelo multidimensional en términos de elementos relacionales básicos como atributos, relaciones y restricciones de integridad.

8.3.1.2 MOLAP.

A diferencia de un sistema ROLAP, un sistema MOLAP se basa en un modelo lógico *ad hoc* que puede ser utilizado para representar directamente los datos y las operaciones multidimensionales. El motor MOLAP toma los datos de la bodega o de las fuentes operativas. Luego almacena los datos en estructuras adecuadas llamadas cubos multidimensionales. La complejidad de los datos subyacentes se oculta al usuario de la herramienta MOLAP. En otras palabras, el usuario ejecuta funciones estándar de OLAP sin necesidad de entender cómo se forman los cubos y en qué se diferencian de las tablas relacionales.

Generalmente, se utiliza un servidor MOLAP independiente que contiene una cantidad limitada de datos. La principal característica de MOLAP es que proporciona un análisis muy rápido. La forma en que un servidor MOLAP logra su objetivo es que pre-calcula el mayor número posible de resultados y los almacena en cubos. Se pre-calculan las jerarquías dentro de las dimensiones individuales, así como las intersecciones entre los atributos de múltiples dimensiones [46].

La gran ventaja de MOLAP en comparación con ROLAP es que las operaciones multidimensionales se pueden realizar de una manera fácil y natural sin necesidad de generar instrucciones complejas. Por esta razón, el rendimiento de MOLAP es excelente. Sin embargo, las implementaciones de sistemas MOLAP tienen muy poco en común, porque aún no se ha establecido una normatividad para el modelo lógico multidimensional [45].

Es importante tener en cuenta que mientras los cubos MOLAP funcionan muy bien en el análisis de datos agregados, estos no son apropiados para el análisis de

datos a nivel transaccional. Los datos transaccionales son datos de grano fino, donde cada registro corresponde a una operación realizada en el mundo real. Los cubos poseen un espacio de almacenamiento limitado y en una organización típica, la enorme cantidad de datos a nivel de transacción superarían la capacidad del cubo. Incluso si estos encajaran de alguna manera en el cubo, cuestiones como el *data sparsity*, que se presenta cuando existen celdas vacías producto de realizar una intersección de los atributos de dimensión a nivel transaccional, ocasionan que los cubos resulten inadecuados para hacer frente a los datos no agregados.

La desventaja de MOLAP frente a ROLAP, es una cuestión de rendimiento *versus* almacenamiento. Las consultas se ejecutan más rápido con las herramientas MOLAP, pero ROLAP es capaz de manejar cantidades de datos mucho más grandes, por lo cual resulta adecuado para el procesamiento de datos detallados a nivel de transacción. Además, los continuos avances en la velocidad de procesamiento de las consultas sobre los RDBMS están reduciendo la diferencia de rendimiento entre las herramientas MOLAP y ROLAP [46].

8.3.1.3 HOLAP.

La arquitectura híbrida combina los enfoques de MOLAP y ROLAP. HOLAP apunta a combinar las ventajas de ambas implementaciones. Aprovecha la capacidad de manejar grandes cantidades de datos de ROLAP, y la velocidad de consulta típica de los sistemas MOLAP [45].

HOLAP implica que la mayor cantidad de datos se debe almacenar en un RDBMS para evitar los problemas causados por el fenómeno del *data sparsity*, y en el sistema multidimensional se almacena solamente la información que necesita ser accedida con mayor frecuencia. Si esa información no es suficiente para resolver las consultas, el sistema accede, de manera transparente, a los datos gestionados por el sistema relacional.

8.3.1.4 DOLAP.

DOLAP es otro término que se utiliza a menudo cuando se discuten las implementaciones de OLAP. DOLAP significa OLAP de escritorio. Al igual que las herramientas MOLAP, DOLAP también utiliza cubos multidimensionales. La diferencia es que los cubos utilizados por DOLAP se descargan a los equipos individuales de los usuarios finales, donde se lleva a cabo el procesamiento de las

consultas. Estos cubos son mucho más pequeños que los utilizados por las herramientas MOLAP [46].

Estos pequeños cubos, también conocidos como micro-cubos, se pueden replicar fácilmente para ser distribuidos a múltiples usuarios. Son fáciles y cómodos de usar, sin embargo, tienen una funcionalidad limitada. Esto no solo es debido a la pequeña cantidad de datos que estos cubos pueden almacenar, sino también al hecho de que la información almacenada en estos cubos es estática, pues no se realizan actualizaciones desde la bodega de datos una vez que el cubo ha sido creado.

8.3.2 Operaciones sobre OLAP


El análisis multidimensional permite a los usuarios examinar un gran número de factores interdependientes involucrados en el análisis de un problema del negocio y visualizar relaciones complejas entre los datos. Los usuarios están interesados en explorar los datos en diferentes niveles de detalle, los cuales se determinan de forma dinámica. Las relaciones pueden ser analizadas a través de un proceso iterativo que incluye la utilización de operadores como: *slice-and-dice*, *drill-down* y *roll-up*, *drill-across* y finalmente *pivoting*.

8.3.2.1 *Slice-and-dice*.

El operador *slice* en la terminología multidimensional es utilizado para definir un miembro o un grupo de miembros y evaluarlos a través de todas las dimensiones. Con el fin de entender este concepto, considere un ejemplo de modelo dimensional planteado por Ballard et al. en [43]. Se tienen tres dimensiones llamadas producto, tienda y fecha en un modelo tridimensional simple. Este modelo solo tiene una tabla de hechos conformada por un hecho llamado ventas.

Suponga que se separan tres miembros de la dimensión producto; estos tres miembros son refresco, leche y jugo. Si se calcula la suma de la cantidad de ventas para todas las tiendas en todas las fechas a través de uno o más miembros de una dimensión (producto en este caso), estaríamos haciendo lo que se conoce como *slice*.

Figura 7: Ejemplo de operación *Slice* sobre la dimensión producto

 (For ALL Stores and Dates)	
Product	Sales in USD
Soda	2,530
Milk	3,858
Juice	15,396
Total	21,784

Fuente: BALLARD, Chuck; FARRELL, Daniel M.; GUPTA, Amit; MAZUELA, Carlos y VOHNÍK Stanislav. Dimensional Modeling: In a Business Intelligence Environment. 1 ed. Redbooks, IBM's International Technical Support Organization. p. 88. Marzo, 2006.

Realizar *slice* sobre la dimensión producto permite enfocarse en los miembros refresco, leche y jugo a través de las otras dimensiones.

El operador *dice* permite seleccionar varios miembros de una dimensión y ubicarlos en un eje y posteriormente ubicar varios miembros de una dimensión diferente en otro eje. Esto posibilita observar la interrelación de los miembros de diferentes dimensiones.

Dicing es el análisis de las interrelaciones entre las diferentes dimensiones o sus miembros. En la siguiente figura se puede observar la interrelación entre los miembros de la dimensión tienda y la dimensión producto. Se puede analizar, por ejemplo:

- Cómo contribuye cada tienda a las cantidades totales de venta de cada producto.
- Cómo un producto en particular contribuye a las ventas totales de cada tienda.

Figura 8: Ejemplo de operación *Dice* sobre las dimensiones tienda y producto

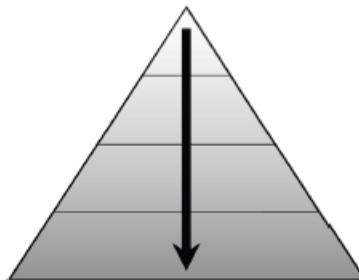
	Metrics	PRODUCT			
		Milk	Coke	Juice	Total
STORE		Sales in USD	Sales in USD	Sales in USD	Sales in USD
	CA	40	60	80	180
	OR	60	1,452	6,210	7,722
	LA	2,430	2,346	9,106	13,882
	Total	2,530	3,858	15,396	21,784

Fuente: BALLARD, Chuck; FARRELL, Daniel M.; GUPTA, Amit; MAZUELA, Carlos y VOHNÍK Stanislav. Dimensional Modeling: In a Business Intelligence Environment. 1 ed. Redbooks, IBM's International Technical Support Organization. p. 89. Marzo, 2006.

8.3.2.2 Drill-down y roll-up.

Drill, es la operación multidimensional que permite pasar de un nivel jerárquico a otro. En otras palabras, *drill* se puede definir como la capacidad para navegar a través de la información, siguiendo una estructura jerárquica.

Figura 9: *Drill-down*

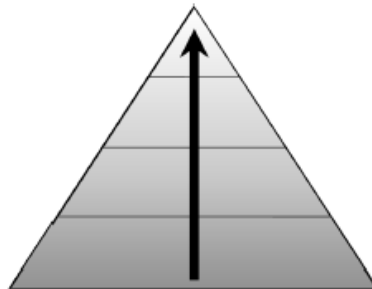


Fuente: Adaptado de BERNABÉU, Ricardo Darío. Hefesto. DATA WAREHOUSING: Investigación y Sistematización de Conceptos. HEFESTO: Metodología para la Construcción de un Data Warehouse. p. 54. Córdoba, Argentina. Julio, 2010.

Drill-down permite apreciar los datos en un mayor detalle, bajando por una determinada jerarquía definida en un cubo. Esto brinda la posibilidad de introducir un nuevo nivel o criterio de agregación en el análisis, disgregando los grupos actuales. En conclusión, *drill-down* es ir de lo general a lo específico [41].

Por otra parte, *roll-up* permite apreciar los datos en menor nivel de detalle, subiendo por la jerarquía definida en un cubo. Esto brinda la posibilidad de quitar un nivel o criterio de agregación en el análisis, agregando los grupos actuales. *Roll-up* es ir de lo específico a lo general [41].

Figura 10: *Roll-up*



Fuente: Adaptado de BERNABÉU, Ricardo Darío. Hefesto. DATA WAREHOUSING: Investigación y Sistematización de Conceptos. HEFESTO: Metodología para la Construcción de un Data Warehouse. p. 55. Córdoba, Argentina. Julio, 2010.

8.3.2.3 *Drill-across*.

Drill-across es una operación en la que se profundiza de una dimensión a otra. Funciona de forma similar a *drill-down*, con la diferencia de que *drill-across* no se realiza sobre una jerarquía, sino que su forma de ir de lo general a lo específico es analizando el comportamiento de un miembro específico de una dimensión contrastando sus valores a lo largo de otras dimensiones.

En la siguiente figura se observa el resultado del operador *drill-across* desde la tienda CA (California) hacia la dimensión producto. En la primera gráfica se

describen las ventas en las diferentes tiendas; posteriormente se realiza un análisis más detallado sobre las ventas de la tienda CA.

Figura 11: Ejemplo de operación *Drill-across* desde la dimensión tienda hacia la dimensión producto

		DATE	1/1/2005	1/2/2005	1/3/2005	Total
Metrics			Sales in USD	Sales in USD	Sales in USD	Sales in USD
STORE						
CA			40	50	90	180
OR			3,115	3,340	1,267	7,722
LA			1,583	7,418	4,881	13,882
Total			4,738	10,808	6,238	21,784

CA → Product

		DATE	1/1/2005	1/2/2005	1/3/2005	Total
Metrics			Sales in USD	Sales in USD	Sales in USD	Sales in USD
PRODUCT						
Soda			10	10	20	40
Milk			20	10	30	60
Juice			10	30	40	80
Total			40	50	90	180

Fuente: BALLARD, Chuck; FARRELL, Daniel M.; GUPTA, Amit; MAZUELA, Carlos y VOHNICK Stanislav. *Dimensional Modeling: In a Business Intelligence Environment*. 1 ed. Redbooks, IBM's International Technical Support Organization. p. 92. Marzo, 2006.

8.3.2.4 *Pivoting*.

El operador *pivot* consiste en intercambiar las filas con las columnas y viceversa. Esta es simplemente una forma rápida de ver los datos desde una vista diferente.

En otras palabras, utilizar el operador *pivot* significa girar el cubo multidimensional para reorganizar las celdas y observar los datos a través de nuevas perspectivas.

Figura 12: Ejemplo de operación *Pivoting* entre las dimensiones tienda y producto

STORE	PRODUCT				Total
	Metrics	Sales in USD	Sales in USD	Sales in USD	Sales in USD
CA		40	60	80	180
OR		60	1,452	6,210	7,722
LA		2,430	2,346	9,106	13,882
Total		2,530	3,858	15,396	21,784

Pivot

PRODUCT	STORE				Total
	Metrics	Sales in USD	Sales in USD	Sales in USD	Sales in USD
Soda		40	60	2,430	2,530
Milk		60	1,452	2,346	3,858
Juice		80	6,210	9,106	15,396
Total		180	7,722	13,882	21,784

Fuente: BALLARD, Chuck; FARRELL, Daniel M.; GUPTA, Amit; MAZUELA, Carlos y VOHNIK Stanislav. Dimensional Modeling: In a Business Intelligence Environment. 1 ed. Redbooks, IBM's International Technical Support Organization. p. 90. Marzo, 2006.

8.4 DASHBOARDS

Son otro método utilizado para la visualización de la información almacenada en las bodegas de datos. El término *dashboard* se refiere a un panel de control que muestra una cantidad limitada de datos relevantes en un formato breve y fácil de interpretar. Pueden proporcionar una visión en tiempo real de las tendencias para un fenómeno específico o para múltiples fenómenos que están relacionados entre sí [45].

Los niveles tácticos y estratégicos de la organización por lo general no constituyen el público objetivo de las herramientas analíticas debido a que sus requerimientos se enfocan en el monitoreo del rendimiento general de la organización. Por lo tanto, estas herramientas son de uso frecuente por los altos directivos que requieren de una forma rápida para visualizar e interpretar la información.

Los tableros de control no son más que indicadores de rendimiento, representados mediante interfaces gráficas. Su eficacia se debe a la selección de los indicadores adecuados, mediante la utilización de estándares de calidad para el uso de las bodegas de datos. Por esta razón, los *dashboards* deben ser vistos como un componente eficaz en los sistemas de almacenamiento de datos, más no como su objetivo principal. De hecho, el objetivo primordial de estos sistemas debe ser la definición de un proceso apropiado para la transformación de los datos en información.

9. METODOLOGÍAS PARA LA CONSTRUCCIÓN DE UNA BODEGA DE DATOS

Los sistemas de Inteligencia de Negocios han sido adoptados con gran popularidad en los últimos años. Compañías de diferentes industrias han notado la importancia de la adopción de este tipo de sistemas en los diferentes niveles de la organización [47]. Sin embargo, un gran número de estas empresas no cuentan con la experiencia y los conocimientos requeridos para enfrentar los desafíos que conllevan los proyectos de bodegas de datos que, como se identificó en capítulos previos, es uno de los componentes principales en la implementación de este tipo de tecnologías. En particular, Golfarelli et al. plantean que la falta de un enfoque metodológico impide que los proyectos de *Data Warehouse* sean finalizados con éxito [45].

En este capítulo se describirán los dos enfoques que generalmente se exponen en la literatura, así como las metodologías más representativas de cada uno. Además, se plantearán una serie de consideraciones importantes a la hora de optar por un determinado enfoque de implementación.

9.1 ENFOQUES DE IMPLEMENTACIÓN

La manera como se aborda el proceso de construcción de una bodega de datos es una decisión de tipo organizacional que puede impactar de forma significativa en el éxito de la implementación de un proyecto de *Data Warehouse*. Las variables que se ven afectadas por el enfoque adoptado son: el tiempo de finalización, el retorno de la inversión, la velocidad con que se obtienen beneficios, la satisfacción del usuario final, los recursos necesarios en cualquier etapa del proceso y la arquitectura de bodegas de datos seleccionada [42].

Los enfoques de implementación considerados comúnmente son: *top-down*, *bottom-up* o una combinación de ambos. Es importante revisar esta fundamentación teórica ya que le proporciona al lector una base de conocimiento para adecuar la tecnología de BI a los requerimientos de la organización y abordar situaciones específicas durante la implementación de una bodega de datos.

9.1.1 *Top-down*

El enfoque *top-down* o descendente concibe el *Data Warehouse* como el principal elemento del entorno analítico. Las bodegas de datos contienen datos atómicos o transaccionales que se extraen de los sistemas fuente para su integración en un único modelo de datos empresarial. Posteriormente, los datos son sumariados y estructurados en diferentes dimensiones para su distribución en uno o varios *Data Marts*.

Con este enfoque, típicamente se construye una bodega de datos global o empresarial. Si los *Data Marts* se incluyen en la arquitectura, por lo general estos son implementados posteriormente y son poblados con los datos que provienen del *Data Warehouse*, en lugar de utilizar directamente los datos que se extraen de las fuentes operacionales [42].

El enfoque *top-down* es un modelo maduro que se utiliza cuando las empresas cuentan con la experiencia y los recursos necesarios para el desarrollo de este tipo de iniciativas. Este consigue alinear todos los procesos del negocio para consolidar un único repositorio de información. Se trata de un método sistemático que minimiza los problemas de integración, pero es costoso, de larga duración y poco flexible.

La implementación descendente se adapta la visión de Bill Inmon, quien considera que el *Data Warehouse* debe responder a las necesidades de todos los usuarios de la organización y no solamente a las de un determinado grupo de usuarios.

9.1.2 *Bottom-up*

En un enfoque *bottom-up* o ascendente, las bodegas de datos son construidas de manera incremental y los *Data Marts* se desarrollan iterativamente. Cada *Data Mart* contiene un conjunto de hechos relacionados con un área específica del negocio para satisfacer los requerimientos de análisis de un departamento o grupo de trabajo.

El objetivo de esta metodología es proporcionar valor al negocio mediante el despliegue de los *Data Marts* en el menor tiempo posible. Una implementación *bottom-up* implica realizar la planificación y diseño de los *Data Marts*, sin necesidad de que una infraestructura global sea desarrollada previamente. Esto

no quiere decir que la plataforma global no vaya a ser implementada, sino que se construirá incrementalmente a partir de la consolidación de los *Data Marts* iniciales [42].

Este modelo se adapta a la visión de Ralph Kimball, quien considera que las bodegas de datos deben ser fácilmente entendidas por los usuarios y desarrolladas con prontitud para proporcionar respuestas a las cuestiones del negocio. Este enfoque se concentra en la definición de los requerimientos para cada proceso del negocio, mientras que el método descendente se orienta a la integración de los datos a nivel de toda la organización.

9.2 METODOLOGÍA DE INMON

Bill Inmon, considerado el padre de las bodegas de datos, propugna un enfoque de implementación descendente, en el que se adaptan las herramientas de bases de datos relacionales tradicionales al desarrollo de un *Data Warehouse* corporativo, a partir del cual se construyen pequeños *Data Marts* para orientar los procesos de toma de decisiones en los diferentes departamentos o áreas del negocio.

Inmon, en su libro *Building the Data Warehouse*, argumenta que es importante realizar un esfuerzo inicial para la construcción de un *Data Warehouse* global debido a que este permite crear cualquier cantidad de bases de datos departamentales sin correr el riesgo de que se presenten incompatibilidades entre los datos. Para esto propone utilizar un modelo de datos compuesto de tres niveles, los cuales serán descritos a continuación [39].

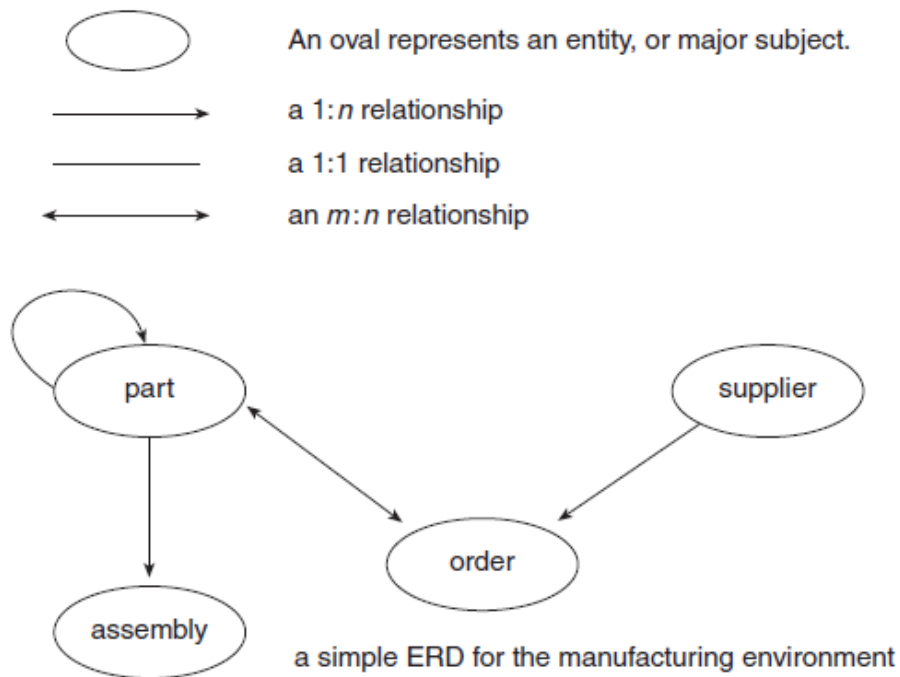
9.2.1 Modelo de datos

Los tres niveles propuestos por Inmon son: nivel superior, también llamado modelo entidad relación (ERD, por sus siglas en inglés *Entity Relationship Diagram*); nivel intermedio, conocido como conjunto de elementos de datos (DIS, por sus siglas en inglés *Data Item Set*) y finalmente el nivel inferior, denominado modelo físico.

Al igual que en el desarrollo de bases de datos operacionales, ERD se utiliza para definir las entidades, sus atributos y las relaciones entre entidades. El nombre de

la entidad aparece rodeado por un ovalo. Las relaciones entre las entidades se representan con flechas. Las puntas de flecha de la relación indican la cardinalidad de la misma y solo se indican las relaciones directas entre entidades. Las entidades representadas en el modelo ERD se encuentran en el máximo nivel de abstracción.

Figura 13: Modelo entidad relación

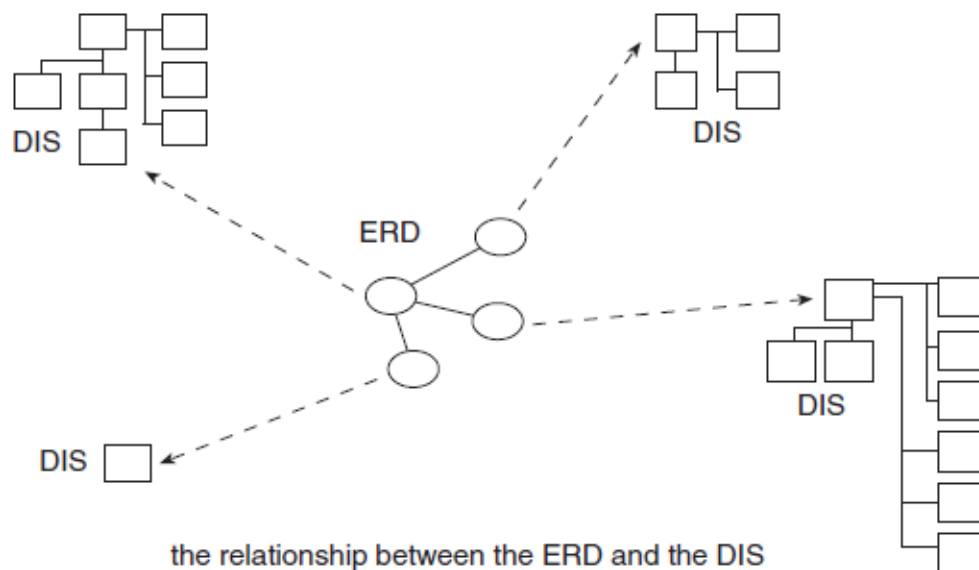


Fuente: INMON, W.H. Building the Data Warehouse. 3 ed. p. 92. New York. Wiley and Sons, 2002.

El equipo de desarrollo genera un conjunto de ERDs para cada departamento que se espera que utilice el *Data Warehouse*. El ERD corporativo es entonces la suma de todos los ERD departamentales, que reflejan los diferentes puntos de vista de las personas a través de la organización.

Posteriormente, se debe realizar el modelo intermedio. Para cada área temática o entidad identificada en el nivel superior se debe crear un DIS. Este modelo se utiliza para identificar los atributos de los datos y las relaciones entre estos atributos.

Figura 14: Relación entre ERD y DIS

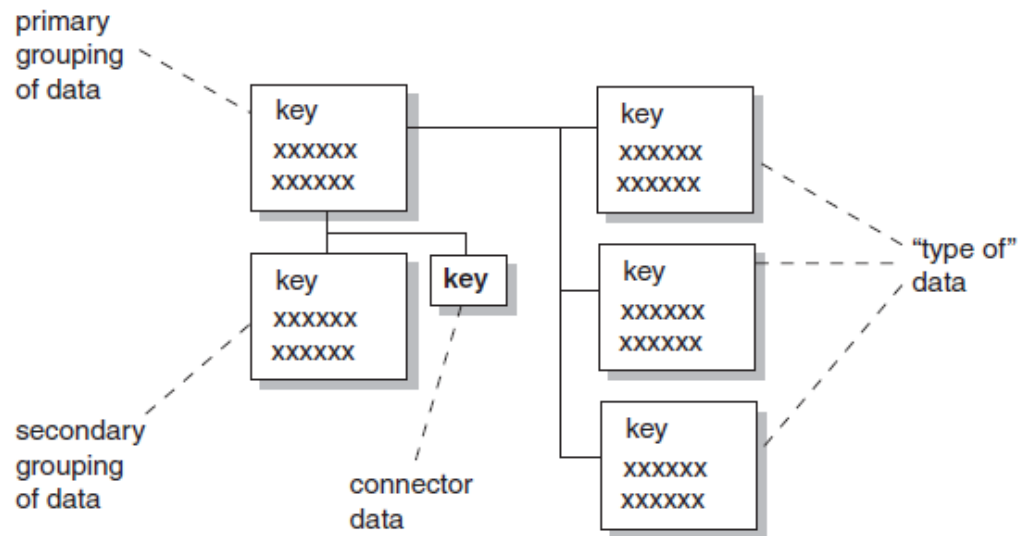


Fuente: INMON, W.H. Building the Data Warehouse. 3 ed. p. 95. New York. Wiley and Sons, 2002.

Un DIS está compuesto por cuatro elementos: una agrupación primaria de datos, una agrupación secundaria, datos conectores y tipos de datos; todos los grupos contienen atributos y claves para cada entidad. El grupo primario de datos incluye los atributos que existen solo una vez para cada entidad. La agrupación secundaria comprende los atributos que pueden aparecer varias veces en cada entidad; es posible tener múltiples agrupaciones de este tipo. El conector representa las relaciones identificadas en el ERD. Finalmente, la agrupación de tipos de datos identifica los supertipos y subtipos de datos.

En un DIS, ya sea departamental o corporativo, cada rectángulo representa una tabla lógica. Las conexiones entre estas tablas son semejantes a las que se utilizan para relacionar las entidades en un ERD. Los rectángulos a la derecha de un determinado DIS representan la agrupación secundaria de datos. Finalmente, los tipos de datos son representados por una línea que conduce a la derecha de una agrupación de datos.

Figura 15: *Data Item Set*



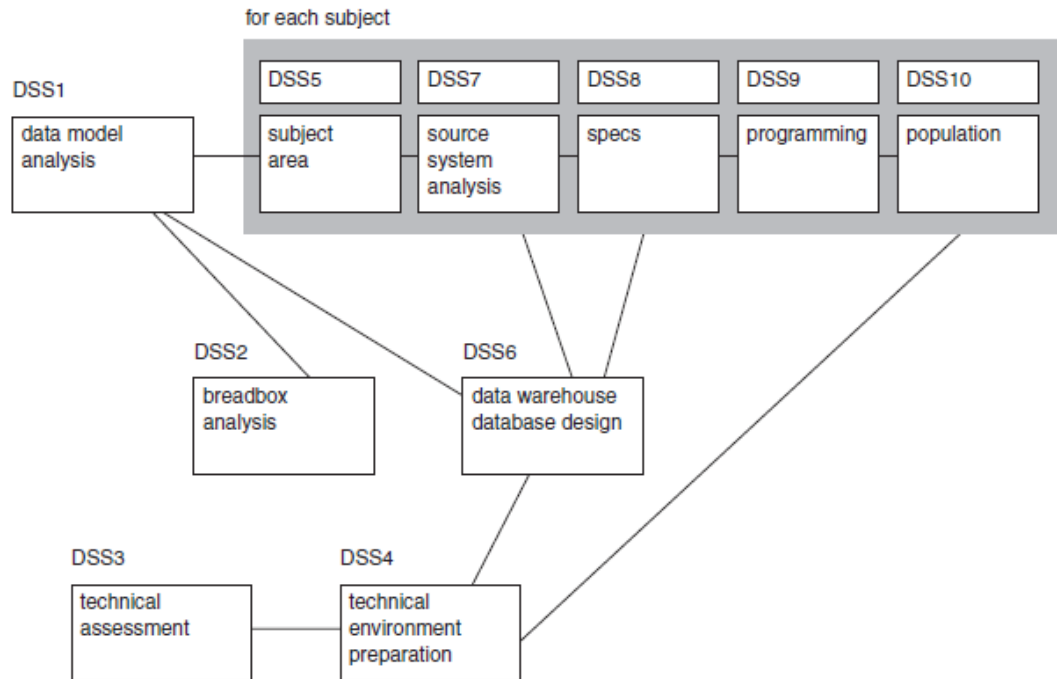
Fuente: INMON, W.H. Building the Data Warehouse. 3 ed. p. 95. New York. Wiley and Sons, 2002.

El último nivel propuesto por Inmon es el modelo de datos físico, el cual se crea a partir de los DIS generados en el nivel intermedio, simplemente extendiéndolo para incluir las claves y las características físicas de las entidades. En este punto, el modelo de datos físico se visualiza como una serie de tablas, también llamadas tablas relacionales.

9.2.2 Metodología de desarrollo - Meth2

Esta es una adaptación especial de la metodología de desarrollo en espiral denominada por Inmon como Meth2. Meth1 es utilizada para la implementación de sistemas operacionales y Meth3 se emplea para el ajuste de bodegas de datos existentes. Las actividades de esta metodología se describen a continuación.

Figura 16: Meth2



Fuente: INMON, W.H. Building the Data Warehouse. 3 ed. p. 365. New York. Wiley and Sons, 2002.

- DSS1. Análisis del modelo de datos: En esta actividad se debe realizar el modelado lógico de los datos descrito anteriormente. En esta fase se deben identificar las áreas temáticas principales, sus atributos, claves y relaciones. La salida de este paso es una confirmación de que la organización ha construido un modelo de datos sólido.
- DSS2. Análisis *breadbox*: Finalizado el modelo de datos de tres niveles, el siguiente paso es llevar a cabo un análisis del tamaño y la granularidad del entorno de soporte a la decisión. La granularidad es una medida del nivel de detalle de los datos. Por ejemplo, los datos transaccionales tienen el nivel de granularidad más bajo debido a que presentan un mayor grado de detalle. Inmon denomina esta fase como el análisis *breadbox*. Si el volumen de los datos es grande, el equipo de desarrollo deberá considerar múltiples niveles de granularidad en los datos. Esto podría implicar almacenar algunos datos en un nivel transaccional y otros de forma sumariada.

- DSS3. Evaluación técnica: Los requisitos técnicos para la gestión de una bodega de datos difieren considerablemente de los necesarios para el manejo y procesamiento de datos en un entorno operacional. La definición técnica del *Data Warehouse* debe satisfacer los siguientes criterios: habilidad para administrar grandes cantidades de datos, flexibilidad para el acceso a los datos, organización de los datos de acuerdo a un modelo lógico, capacidad de interacción entre diferentes tecnologías, carga periódica de cantidades masivas de datos y acceso a un conjunto de datos o a un registro específico de manera instantánea.
- DSS4. Preparación del entorno técnico: Una vez que se ha establecido la disposición de la arquitectura para la bodega de datos, el siguiente paso es identificar cómo se van a satisfacer estos requerimientos técnicos. Se debe definir, por ejemplo, la configuración de red necesaria, cómo proporcionar acceso a los sistemas de almacenamiento, cómo satisfacer el volumen de procesamiento esperado, entre otros.
- DSS5. Análisis de las áreas temáticas: En esta fase se selecciona la primera área temática que será poblada. Esta debe ser lo suficientemente grande para que sea significativa y lo suficientemente pequeña para que su implementación no resulte compleja.
- DSS6. Diseño de la bodega de datos: La bodega de datos es diseñada con base en el modelo de datos. Algunas de las características de esta etapa de diseño incluyen: la especificación de los diferentes niveles de granularidad, la definición de los datos de las áreas temáticas de la organización, la eliminación de los datos no relevantes para el proceso de toma de decisiones, definir la periodicidad de cada registro de información, entre otros.
- DSS7. Análisis de los sistemas fuente: Una vez que las áreas temáticas han sido definidas, la siguiente actividad consiste en identificar las fuentes de los datos en el entorno operacional. En este punto se deben abordar los problemas de integración de los datos.
- DSS8. Especificaciones: Después de definir las interfaces entre los sistemas operacionales y el entorno de soporte a la toma de decisiones, el siguiente paso consiste en especificar los requerimientos de los programas a utilizar.

- DSS9. Programación: Esta etapa consiste en realizar todas las actividades estándares de la fase de desarrollo, tales como realizar el pseudocódigo, codificación, pruebas unitarias, entre otros.
- DSS10. Poblado de datos: Este paso implica la ejecución de los programas desarrollados previamente. Los temas abordados en esta actividad son los siguientes: frecuencia con que se realiza el poblado, limpieza de los datos poblados, gestión de los diferentes niveles de granularidad, entre otros. El resultado de esta fase es una bodega de datos funcional.

9.3 METODOLOGÍA DE KIMBALL

Ralph Kimball sugiere un enfoque ascendente que utiliza el modelo dimensional para describir la organización de los datos en un *Data Warehouse*. En lugar de construir una bodega de datos para toda la empresa, Kimball sugiere que se implementen repositorios de datos más pequeños para los principales procesos del negocio.

La historia del modelo entidad relación surgió de la mano de la tecnología de bases de datos relacionales. Este ha proporcionado un desempeño óptimo para la manipulación de los datos en los entornos transaccionales. Sin embargo, para generar un reporte de las actividades de la organización generalmente se requiere reunir información de múltiples tablas, lo cual se considera demasiado complejo para un usuario cotidiano. Además, para superar los problemas de rendimiento producto de las consultas de gran tamaño realizadas a las bodegas de datos se empezó a utilizar el modelo dimensional. Este enfoque proporciona una mejora en el rendimiento de las consultas sin afectar la integridad de los datos.

9.3.1 Modelo dimensional

Es una técnica de diseño utilizada para definir la estructura lógica de los datos en un *Data Warehouse*. Este modelo se constituye con el objetivo de optimizar el desempeño de las consultas sobre los sistemas de soporte a la toma de decisiones en bases de datos relacionales, a partir de un conjunto de medidas y dimensiones para la caracterización de los procesos del negocio que serán incorporados en los repositorios de información. Por el contrario de los modelos

entidad-relación convencionales, que fueron concebidos para la eliminación de redundancia en los datos y facilitar la recuperación de los registros individuales, logrando de esta manera optimizar el rendimiento de las consultas.

El modelo dimensional es una técnica utilizada para la conceptualización y presentación de los datos como un conjunto de medidas descritas por aspectos comunes del negocio. Resulta especialmente útil para resumir y ordenar los datos, además de que facilita su visualización para soportar las actividades de análisis.

Un modelo dimensional, se compone generalmente de dos tipos de elementos: tablas de hechos y tablas de dimensiones.

- Hechos: Un hecho es una colección de datos que tienen relación entre sí, y consiste de medidas e información del contexto. Normalmente, los hechos representan algún proceso del negocio, una transacción o un evento que puede ser analizado para evaluar el estado del negocio. La tabla de hechos es la tabla más importante del modelo dimensional. En esta se almacenan las mediciones sobre el desempeño de los procesos críticos de la empresa.

Figura 17: Ejemplo tabla de hechos

Daily Sales Fact Table
Date Key (FK) Product Key (FK) Store Key (FK) Quantity Sold Dollar Sales Amount

Fuente: KIMBALL, Ralph y ROSS, Margy. The Data Warehouse Toolkit. The complete guide to Dimensional Modeling. 2 ed. p. 17. Wiley and Sons, 2002.

Por otra parte, una medida es el atributo numérico de un hecho, utilizado para representar el rendimiento de un proceso del negocio. Algunos ejemplos de medidas son: la cantidad de dinero generado por las ventas, el volumen de ventas, la cantidad suministrada, el costo de las provisiones, el monto de la transacción, entre otros. Una medida está determinada por la combinación de los miembros de un conjunto de dimensiones.

- Dimensiones: Una dimensión está conformada por una serie de miembros o unidades del mismo tipo. En un modelo dimensional, cada dato o registro de la tabla de hechos se asocia con solo un miembro de cada una de las dimensiones. Es decir, las dimensiones determinan el contexto de los hechos. Las tablas de dimensiones contienen los descriptores textuales del negocio. En un modelo dimensional bien definido, estas tablas contienen un gran número de columnas o atributos.

Figura 18: Ejemplo tabla de dimensión

Product Dimension Table
Product Key (PK)
Product Description
SKU Number (Natural Key)
Brand Description
Category Description
Department Description
Package Type Description
Package Size
Fat Content Description
Diet Type Description
Weight
Weight Units of Measure
Storage Type
Shelf Life Type
Shelf Width
Shelf Height
Shelf Depth
... and many more

Fuente: KIMBALL, Ralph y ROSS, Margy. The Data Warehouse Toolkit. The complete guide to Dimensional Modeling. 2 ed. p. 20. Wiley and Sons, 2002.

Una dimensión contiene múltiples miembros. Cada miembro de dimensión posee un nombre distintivo o un identificador que es utilizado para determinar la posición de un elemento dentro de la tabla. Por ejemplo, todos los meses, trimestres y años constituyen la dimensión tiempo; y todas las ciudades, regiones y países conforman la dimensión geografía.

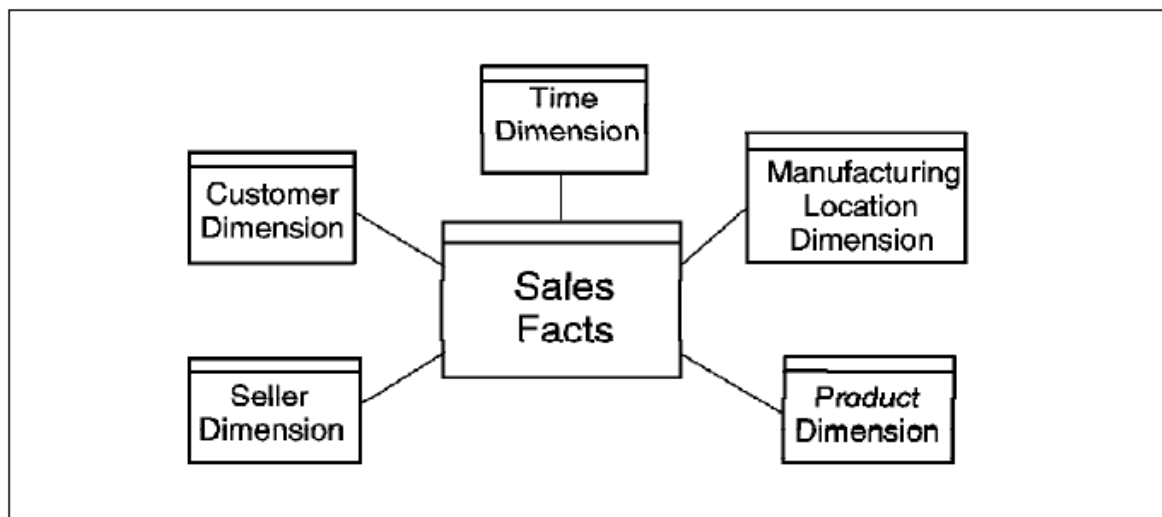
Finalmente es posible organizar los miembros de una dimensión en una o más jerarquías. Cada jerarquía también puede contener varios niveles de jerarquía. Cada miembro de una dimensión no necesariamente debe estar ubicado en una estructura jerárquica.

9.3.1.1 Esquemas de datos dimensionales

Existen diferentes esquemas que pueden ser utilizados para representar un modelo dimensional; entre los más destacados en la literatura se encuentran el modelo de estrella y el modelo de copo de nieve.

- Esquema de estrella: Este representa la estructura básica para el modelo dimensional. Está compuesto, típicamente, por una gran tabla central, llamada tabla de hechos y por un conjunto de tablas, denominadas tablas de dimensiones, organizadas alrededor de la tabla de hechos de manera circular. En la siguiente figura se observa un ejemplo de esquema de estrella. La tabla de hechos ubicada en la parte central del modelo representa el hecho ventas. Ubicadas alrededor de la tabla de hechos se tienen las tablas de dimensiones tiempo, cliente, vendedor, ubicación de la fábrica y producto.

Figura 19: Esquema de estrella

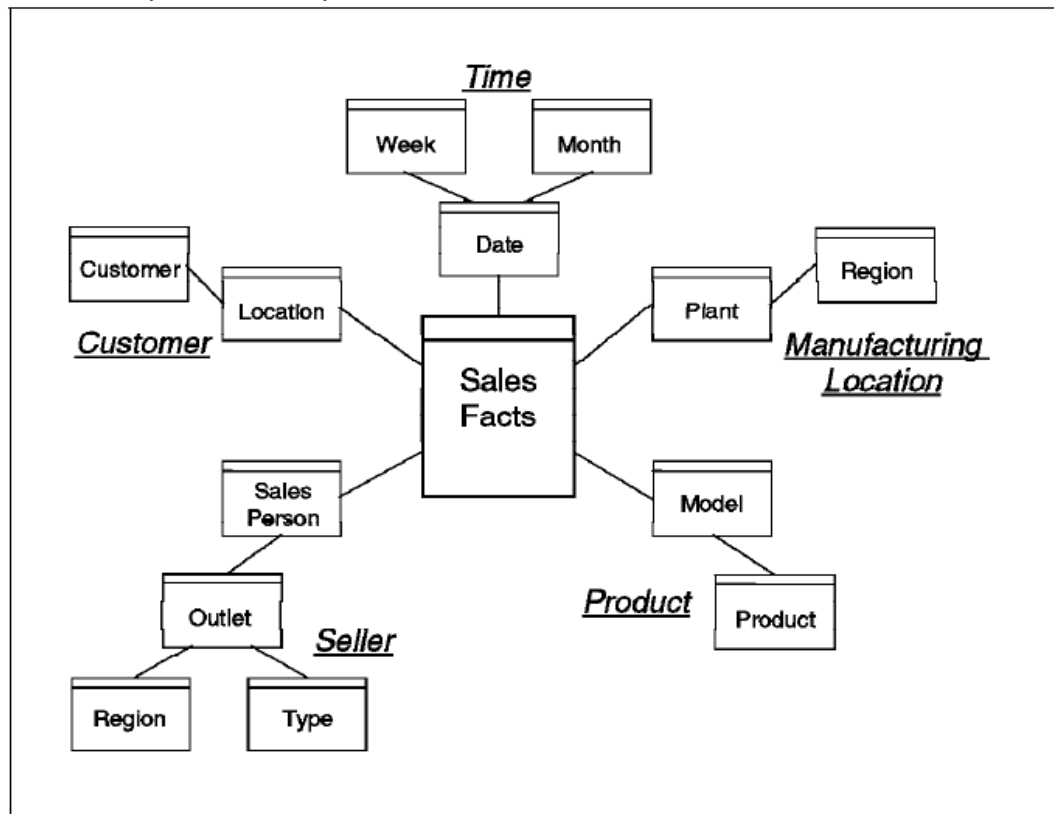


Fuente: BALLARD, Chuck; HERREMAN, Dirk; SCHAU, Don; BELL, Rhonda; KIM, Eunsang y VALENCIC, Ann. Data Modeling Techniques for Data Warehousing. IBM's International Technical Support Organization. p. 47. Febrero, 1998.

- Esquema de copo de nieve: El modelamiento dimensional comienza con la identificación de los hechos y las dimensiones; es por esto que usualmente tiene una apariencia parecida al esquema de estrella, con un hecho en el centro y varias dimensiones a su alrededor.

El esquema de copo de nieve es el resultado de descomponer una o más dimensiones, las cuales tienen algún nivel de jerarquía. Este modelo permite visualizar la estructura jerárquica de las dimensiones, para realizar un análisis más detallado de las dimensiones involucradas en un determinado hecho. Sin embargo, su complejidad puede ocasionar que los usuarios se sientan más cómodos trabajando con el esquema de estrella, que resulta un poco más simple.

Figura 20: Esquema de copo de nieve



Fuente: BALLARD, Chuck; HERREMAN, Dirk; SCHAU, Don; BELL, Rhonda; KIM, Eunsang y VALENCIC, Ann. Data Modeling Techniques for Data Warehousing. IBM's International Technical Support Organization. p. 48. Febrero, 1998.

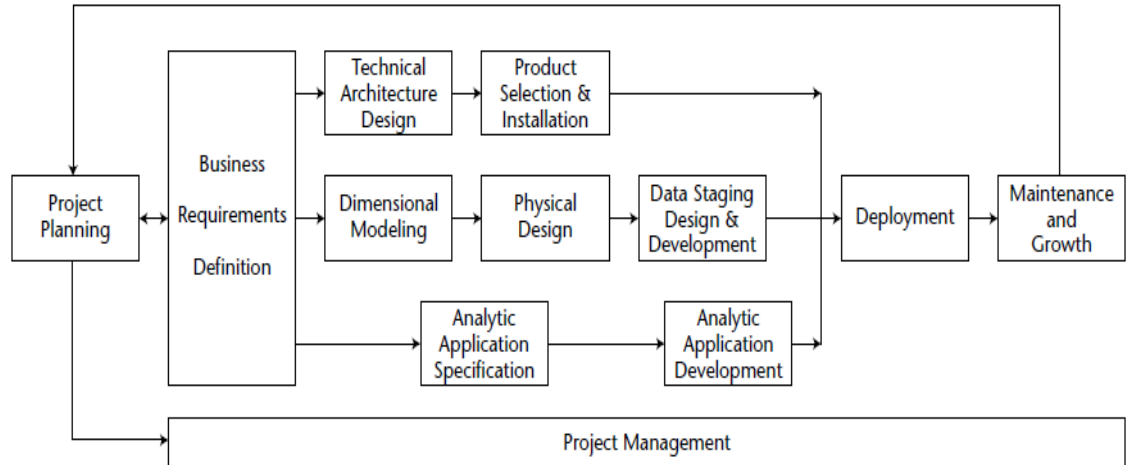
9.3.2 Ciclo de vida de Kimball

La metodología de Kimball, descrita en el libro *The Data Warehouse Lifecycle Toolkit*, se basa en lo que él denomina el ciclo de vida dimensional del negocio. Este método se fundamenta en tres principios elementales:

- Enfoque hacia las necesidades del negocio y en la generación de valor para la organización.
- Datos estructurados de forma dimensional.
- Desarrollo de un *Data Warehouse* de forma iterativa mediante incrementos graduales en lugar de una perspectiva global.

En la siguiente figura se presenta el ciclo de vida propuesto por Kimball. Las actividades para el desarrollo de esta metodología se describen a continuación.

Figura 21: Ciclo de vida de Kimball



Fuente: KIMBALL, Ralph y ROSS, Margy. *The Data Warehouse Toolkit. The complete guide to Dimensional Modeling*. 2 ed. p. 332. Wiley and Sons, 2002.

- Planeación del proyecto: El ciclo de vida dimensional del negocio inicia con la planificación del proyecto. Esta debe abordar la definición del alcance del proyecto de *Data Warehouse*, así como evaluar la preparación de la organización para emprender este tipo de innovaciones tecnológicas. Estas

tareas son críticas debido a los altos costos asociados a la mayoría de los proyectos de bodegas de datos. A partir de este punto, la planificación del proyecto se enfoca en la identificación de los recursos necesarios, el personal requerido, la asignación de tareas, su duración y secuenciación.

- Definición de los requerimientos del negocio: La probabilidad de que un proyecto de *Data Warehouse* sea exitoso depende en gran medida del entendimiento de los requerimientos del negocio y la identificación de los usuarios de la información. Los diseñadores de la bodega de datos deben entender los factores clave que impulsan el negocio para determinar efectivamente los requerimientos y finalmente traducirlos en las consideraciones de diseño.
- Diseño de la arquitectura técnica: Los entornos de *Data Warehouse* requieren la integración de diversas tecnologías. El diseño de la arquitectura técnica establece el marco de trabajo y la visión de la arquitectura global. Para la definición de la arquitectura técnica es necesario tener en cuenta factores como los requerimientos del negocio, el entorno técnico actual y las directrices estratégicas planeadas en el área técnica.
- Selección e instalación del producto: A partir del diseño de la arquitectura técnica, se deben evaluar y seleccionar los componentes arquitecturales específicos, tales como: la plataforma de hardware, el sistema de gestión de bases de datos, las herramientas ETL y las herramientas de acceso a los datos. Una vez que los productos han sido evaluados y seleccionados, se instalan y prueban a fondo para garantizar la integración con el entorno de *Data Warehouse*.
- Modelamiento dimensional: La definición de las características del negocio determina los datos necesarios para hacer frente a los requerimientos analíticos de los usuarios. En esta fase se identifican los procesos clave del negocio, las dimensiones, los niveles jerárquicos y los hechos.
- Diseño físico: Esta fase se concentra en la definición de las estructuras físicas necesarias para soportar el diseño lógico de la base de datos. Los elementos principales de este proceso incluyen la definición de estándares de nomenclatura, la configuración del entorno de bases de datos, las estrategias preliminares de indexación y particionamiento.

- **Diseño y desarrollo de la preparación de los datos:** El proceso de la preparación de los datos tienen tres tareas principales: la extracción, transformación y carga. El proceso de extracción de los datos siempre expone los problemas de calidad ocultos en los sistemas operacionales. Además, en esta fase es necesario diseñar y construir dos procesos esenciales: uno para el poblado inicial del *Data Warehouse*, y otro para la cargas de datos regulares e incrementales.
- **Especificación de aplicaciones analíticas:** Es recomendable definir un conjunto de aplicaciones de usuarios, pues no todas las personas de la organización tienen las mismas necesidades de información y no todas requieren acceso al *Data Warehouse*. Estas especificaciones aseguran que el equipo de desarrollo y los usuarios del negocio tengan una comprensión clara de las aplicaciones que serán entregadas.
- **Desarrollo de aplicaciones analíticas:** Se configuran las herramientas especificadas en el punto anterior. En condiciones óptimas, estas aplicaciones se construyen utilizando una herramienta avanzada para el acceso a los datos, que proporciona una mayor productividad para el equipo de desarrollo de aplicaciones. Además, ofrecen mecanismos para que los usuarios del negocio modifiquen fácilmente las plantillas de los informes existentes.
- **Implementación:** Este proceso implica la convergencia de las tecnologías, datos y aplicaciones analíticas que podrán ser accedidas por los usuarios del negocio. Se requiere de una planificación rigurosa para asegurarse de que todos los elementos se integren correctamente.
- **Mantenimiento y crecimiento:** Es necesario brindar apoyo continuo y educación a los usuarios del negocio. Además, es fundamental asegurarse de que los procesos y procedimientos están funcionando de la forma esperada.
- **Gestión del proyecto:** La gestión del proyecto aparece a lo largo de todo el ciclo de vida, y su función es asegurar que todas las actividades se desarrollan y permanecen correctamente sincronizadas. Esta fase se enfoca en la monitorización del estado del proyecto, el seguimiento a los problemas, el control de los cambios y la gestión del alcance.

9.4 COMPARACIÓN DE LAS METODOLOGÍAS

En la comparación que realiza Breslin entre ambos enfoques, plantea que las similitudes más representativas son: el manejo del atributo tiempo en los datos y los procesos ETL [49]. Independientemente de si se adopta uno u otro enfoque, los usuarios finales podrán consultar la información de la bodega de datos por año, trimestre, mes, día, etc. Además, el entorno de *Data Warehouse* en ambos casos inicia con los procesos ETL. Los datos se extraen desde las fuentes operacionales, son transformados para cumplir con los estándares de almacenamiento y finalmente son cargados a los repositorios de información; en el caso de Inmon se cargan a un *Data Warehouse* global y en el caso de Kimball, los datos se cargan a unidades más pequeñas o *Data Marts*.

Por otra parte, Hugh J. Watson afirma que, cuando se desarrollan correctamente, las dos estrategias son válidas [12]. Ambas aproximaciones abogan por construir una arquitectura robusta que se adapte fácilmente a los cambios en las necesidades del negocio y consolidar una única fuente de información que garantice la consistencia de los datos a través de toda la organización.

Sin embargo, la arquitectura de Inmon se compone de todos los sistemas de información y sus respectivas bases de datos. Inmon concibe las bodegas de datos como parte integral de la visión corporativa. Esto significa, entre otras cosas, que el *Data Warehouse* y los sistemas transaccionales constituyen una gran fuente de información. Este método se basa en los principios y prácticas que han sido utilizados en el entorno de las bases de datos operacionales [49].

Una consecuencia de este enfoque evolutivo es que la audiencia principal de Inmon son los profesionales de TI, ya que estos cuentan con el nivel de entendimiento para utilizar activamente sus herramientas y metodologías de desarrollo. En este caso, los usuarios finales tendrán un papel pasivo en la implementación de la bodega de datos, únicamente revisando el trabajo realizado por los profesionales de TI.

Asimismo, una implementación descendente requiere un esfuerzo mayor en planificación y diseño al inicio del proyecto. Esto trae consigo la necesidad de involucrar a las personas de cada uno de los grupos de trabajo, departamentos o áreas del negocio que van a participar en la consolidación del *Data Warehouse*. Las decisiones respecto a las fuentes de información que serán utilizadas, la estructura, calidad y estándares de los datos, las políticas de seguridad y, en

general la definición del modelo de datos organizacional, necesitan ser completados antes de iniciar con la implementación del proyecto [42].

Este proceso puede resultar en una definición de los datos más consistente. Sin embargo, el costo de la planificación y el diseño inicial es bastante significativo, ya que esta es una actividad que consume mucho tiempo, que puede retrasar los beneficios percibidos con la implementación y el retorno de la inversión. En algunas organizaciones, la gerencia está cada vez menos dispuesta a aceptar este tipo de demoras.

El modelo de Kimball difiere en varios aspectos del enfoque tradicional de bases de datos relacionales. Una diferencia importante radica en que el *Data Warehouse* se crea utilizando un modelo que facilita el entendimiento de los datos y mejora el rendimiento de las consultas sobre el sistema. El modelo dimensional permite a los usuarios finales realizar consultas sobre las bodegas de datos de manera intuitiva debido a que la información se encuentra estructurada en términos de los hechos y las dimensiones de los procesos del negocio. Kimball defiende que los usuarios y los profesionales de TI deben compartir tareas por igual. Al asegurar la participación activa de los actores del negocio en todo el proceso de desarrollo, la probabilidad de aceptación de los usuarios de las bodegas de datos tiende a ser mayor [49].

El enfoque de Kimball ha sido ampliamente aceptado en la actualidad ya que con este es posible percibir resultados inmediatos a partir de los *Data Marts*, lo cual puede ser utilizado como justificación para expandir la implementación a un *Data Warehouse* corporativo. El método *bottom-up* se ha convertido en la elección de un gran número de organizaciones debido a que permite recuperar la inversión en poco tiempo [42]. La construcción de los *Data Marts* se desarrolla de manera más rápida puesto que estos requieren un diseño de menor complejidad que una bodega de datos global. Además, la implementación inicial suele ser menos costosa en términos de hardware y otros recursos.

A pesar de los aspectos positivos de este enfoque, se deben tener en cuenta algunas consideraciones. Por ejemplo, se puede presentar redundancia e inconsistencia en los datos en la medida en que se crean nuevos *Data Marts*. Estos problemas se pueden minimizar con una planeación, supervisión y directrices de diseño efectivas.

La arquitectura de bus propuesta por Kimball permite la integración de los datos en los *Data Marts* para la consolidación de una bodega de datos corporativa. En esta se plantea que todos los *Data Marts* deben utilizar dimensiones estandarizadas. Los requerimientos básicos de esta estandarización son que las claves, nombres de columnas, definición de los atributos y valores de los atributos sean consistentes en todos los procesos del negocio [49].

En la siguiente figura se presenta un resumen de la comparación entre las características más representativas de ambas metodologías.

Figura 22: Comparación de las características de los modelos de Inmon y Kimball

	Inmon	Kimball
Metodología y arquitectura		
Enfoque	<i>Top-down</i>	<i>Bottom-up</i>
Arquitectura	Se construye un único <i>Data Warehouse</i> corporativo, el cual es utilizado para poblar los <i>Data Marts</i> departamentales.	Se construyen múltiples <i>Data Marts</i> , uno por cada proceso del negocio. La consistencia empresarial se logra a través del bus de datos y las dimensiones comunes.
Complejidad	Alta	Simple
Diseño físico	Robusto	Ligero
Modelo de datos		
Orientación de los datos	Orientado a los datos	Orientado a los procesos
Herramientas	Tradicionales (ERD, DIS)	Modelo dimensional
Accesibilidad	Baja	Alta
Filosofía		
Audiencia principal	Profesionales de TI	Usuarios finales
Objetivo	Entregar una solución técnica completa, fundamentada en métodos y tecnologías de bases de datos comprobados.	Entregar una solución que facilite al usuario final la realización de consultas sobre los datos y obtener tiempos de respuesta razonables.

Fuente: Adaptado de BRESLIN, Mary. Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models. Business Intelligence Journal. p. 16. Invierno de 2004.

Idealmente, las organizaciones deben tomar conceptos y tácticas de cada enfoque de implementación para crear un entorno que se ajuste a sus necesidades. Estas requieren comprender las fortalezas y limitaciones de cada metodología para determinar el enfoque más conveniente, de acuerdo a las características particulares del negocio. Contar con una base de mejores prácticas asegura un resultado exitoso en la implementación de este tipo de iniciativas.

Es más probable que una organización tenga éxito en la implementación del *Data Warehouse* utilizando el enfoque de Inmon si cuenta con un gran número de profesionales de TI, planea un proyecto que involucre a todas las áreas del negocio y puede esperar para ver los resultados en lapsos de tiempo considerables, generalmente entre cuatro y nueve meses.

Por otro lado, el enfoque de Kimball proporciona una implementación más rápida, al menos para la entrega del repositorio de información inicial. Por lo general, el desarrollo del primer *Data Mart* funcional para un área específica del negocio puede tardar 90 días y los *Data Marts* posteriores entre 60 y 90 días cada uno [50]. Este modelo resulta favorable cuando la organización puede conformar equipos de trabajo interdisciplinarios, más pequeños que los requeridos por el enfoque de Inmon; además, cuando se desea almacenar en mayor medida información sobre las métricas del negocio.

Figura 23: Factores a favor de cada modelo

Característica	A favor de Kimball	A favor de Inmon
Orientado al soporte de decisiones en el nivel organizacional.	Táctico	Estratégico
Requerimientos de integración de los datos	Orientado a las áreas del negocio de manera individual.	Integración empresarial
Estructura de los datos	Métricas del negocio, medidas de rendimiento y tableros de control.	Datos que se aplican a múltiples necesidades de información.
Escalabilidad	Se adapta fácilmente al cambio de los requerimientos.	Redefinir el alcance y los requerimientos impacta la ejecución del proyecto de forma significativa.
Requerimientos de habilidades y personal	Pequeños grupos de trabajo interdisciplinarios.	Equipos de trabajo conformados por especialistas.
Tiempo de entrega	Se requiere entregar una aplicación de bodegas de datos de forma inmediata.	La organización se puede permitir un mayor tiempo de implementación.

Fuente: Adaptado de BRESLIN, Mary. Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models. Business Intelligence Journal. p. 6-20. Invierno de 2004.

Finalmente, algunos estudios han demostrado que, ya sea utilizando el enfoque de Inmon o de Kimball, es fundamental que el equipo de implementación del *Data Warehouse* cuente con habilidades de carácter no técnico tales como liderazgo, comunicación, planificación y relaciones interpersonales [49]. Estas deben ser empleadas de manera eficaz para asegurar que la organización tenga una visión bien articulada del funcionamiento y la utilización del *Data Warehouse* y de esa forma, garantizar que el proyecto disponga de los recursos suficientes para su culminación. Estas no suelen ser las responsabilidades que un equipo de desarrollo de proyectos de TI debe asumir. Sin embargo, son esenciales para el éxito en la implementación de una solución de Inteligencia de Negocios.

10. HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS

La Inteligencia de Negocios proporciona a las organizaciones la perspectiva necesaria para tomar decisiones críticas del negocio respecto a, por ejemplo, cómo comercializar sus productos, gestionar sus clientes, administrar bienes y servicios o gestionar sus resultados financieros. Desafortunadamente, la implementación de este tipo de soluciones en ocasiones resulta problemática y no satisface por completo las expectativas del proyecto y del usuario final. A pesar de las múltiples ventajas, los sistemas de soporte a la decisión existentes a menudo permanecen inaccesibles o son insuficientes para las pequeñas y medianas empresas. Además, las diferencias en la filosofía de trabajo de las PYMES y las grandes empresas son bastantes significativas, por lo que no es recomendable utilizar soluciones destinadas a grandes compañías en las PYMES. De manera que estas requieren una mayor flexibilidad, menor inversión en infraestructura y herramientas que resulten sencillas y eficientes. Bernardino sugiere que los sistemas de *Business Intelligence* en la actualidad no son lo suficientemente asequibles y adecuados para las pequeñas y medianas empresas debido a factores como [51]:

- Costo.
- Infraestructura de gran alcance.
- Complejidad.
- Funcionalidad inadecuada.
- Poca flexibilidad para hacer frente a un entorno empresarial dinámico y en constante cambio.
- Poca atención a las diferencias en las necesidades de acceso a los datos entre las pequeñas y las grandes empresas.

En este capítulo se explorarán las principales consideraciones a tener en cuenta al momento de elegir una herramienta de BI. En primer lugar, es recomendable escoger y adquirir las herramientas más adecuadas a las necesidades de la organización y a los requerimientos técnicos definidos por la arquitectura del repositorio de datos. Además, en este proceso de selección se debe prestar especial atención a la infraestructura técnica de la compañía.

La comparación entre productos de BI es significativamente difícil debido a la gran variedad de suites que se ofrecen en el mercado y a sus características. La selección de estas herramientas depende de las prioridades del proyecto, es decir, de los requerimientos específicos de cada organización: tableros de control, automatización de informes, creación de cubos, entornos colaborativos, independencia de TI, entre otras, son funcionalidades que deben ser evaluadas en cada herramienta. En el nivel de interacción básico se encuentran los tableros, la ejecución de informes programados y la clasificación y filtrado de los resultados de las consultas. Mientras que los usuarios expertos y los desarrolladores profesionales requieren la capacidad de crear objetos en el entorno de BI y realizar actividades ad hoc. Resulta entonces necesario evaluar este tipo de requerimientos y que las herramientas estén alineadas y proporcionen las funcionalidades que el usuario precisa.

Además, los factores utilizados para determinar el precio de las herramientas son otro asunto que resulta bastante desafiante. Los presupuestos limitados de las pequeñas organizaciones llevaron a los patrocinadores de los proyectos a exigir una estimación de los beneficios económicos antes de aprobar la financiación de este tipo de iniciativas de TI. La inversión en este tipo de proyectos está diseñada para recuperar el costo de implementación y mantenimiento de la respectiva tecnología [44].

Las organizaciones a menudo utilizan métodos tradicionales como el valor actual neto, el retorno de la inversión y el costeo basado en actividades para la evaluación de las inversiones en TI. Sin embargo, algunos investigadores argumentan que las técnicas tradicionales no son adecuadas para la evaluación de proyectos que conllevan beneficios estratégicos significativos, como los sistemas de BI [52].

También se suelen utilizar métodos de análisis costo-beneficio que comparan el valor total de los beneficios con los costos asociados. Por lo tanto, para el cálculo de los beneficios netos y la rentabilidad se debe realizar una aproximación de los costos derivados de una implementación de BI y los beneficios percibidos con la utilización de este tipo de tecnologías.

El proceso de convertir datos en bruto en Inteligencia de Negocios útil para la toma de decisiones empresariales implica realizar una inversión significativa durante todo el proceso de implementación de BI, desde la fase de desarrollo, hasta el despliegue y mantenimiento. Para calcular el costo de un proyecto de BI

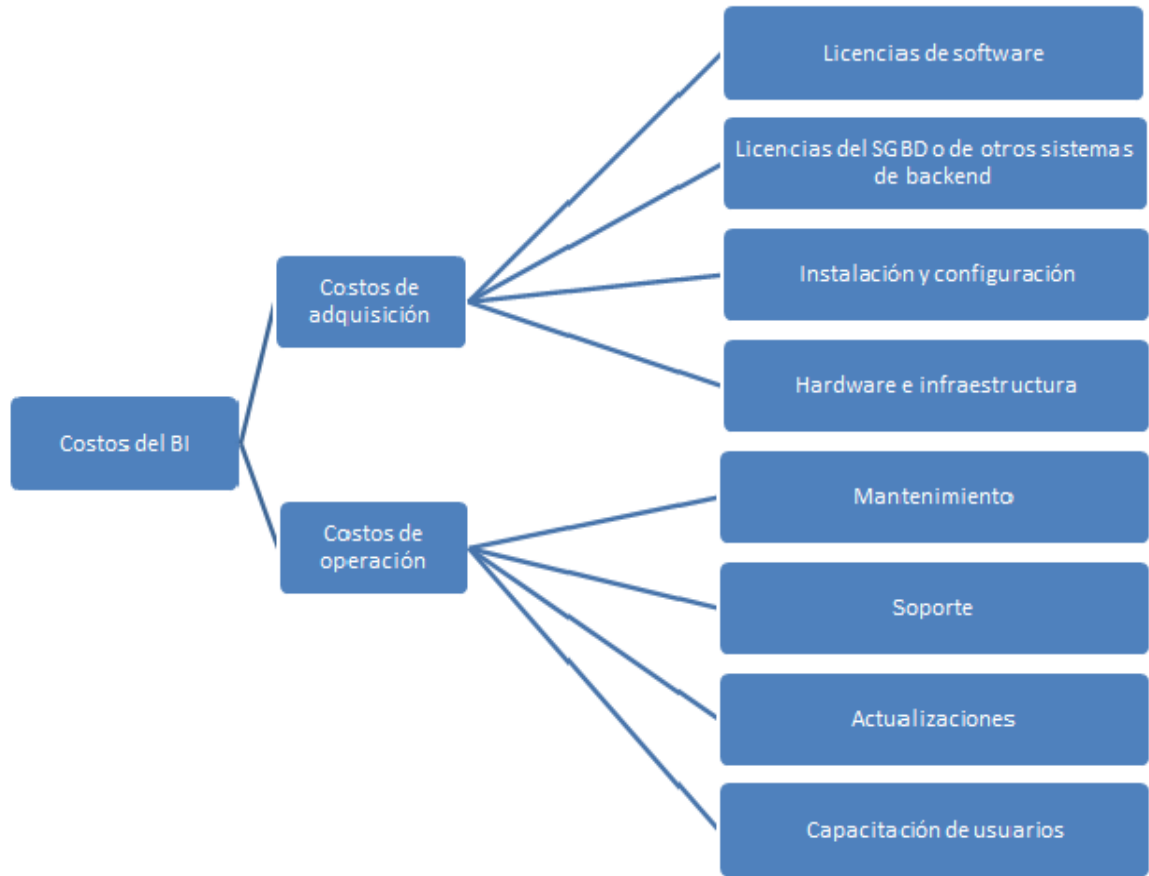
comúnmente se utiliza un método conocido como costo total de propiedad (TCO, por sus siglas en inglés *Total Cost of Ownership*). Esta técnica fue introducida por Gartner Group a finales de los 80's y se utiliza para calcular con mayor precisión el costo de los sistemas de información.

Además de los costos iniciales y de mantenimiento, el TCO incluye los valores asociados con el soporte técnico, la administración del sistema y la capacitación de usuarios. Este método propone calcular el costo general del sistema de BI para un determinado periodo de explotación. Comúnmente se utiliza un lapso de tiempo de entre 3 y 5 años, ya que ese tiende a ser el rango manejado por las empresas para depreciar los activos capitalizados [44].

Los costos de un proyecto de Inteligencia de Negocios se pueden agrupar en tres categorías: los costos de adquisición, los costos de operación y los costos de control [53]. En el mercado comúnmente se ofrecen suites de BI integradas. Por esta razón, si se realiza una evaluación de este tipo de software, los costos relacionados con el control, derivados de la consolidación y estandarización de las herramientas, no son aplicables y debido a esto generalmente son excluidos en el cálculo de los costos. Así, los costos típicos en la implementación de un proyecto de BI planteados por Aliyev se presentan en la figura 23.

De igual forma se debe tener en cuenta que, en muchos casos, los proyectos de BI también generan beneficios intangibles como, por ejemplo, rapidez en la comercialización de nuevos productos, satisfacción de empleados y clientes, mayor agilidad organizacional y un mejor control del negocio. Aliyev asevera que ignorar este tipo de beneficios intangibles puede llevar a la organización a rechazar las inversiones en TI [44]. Por lo tanto, el análisis financiero debe contemplar tanto los beneficios tangibles como los intangibles.

Figura 24: Costos de implementación de un proyecto de BI



Fuente: Adaptado de ALIYEV, Rustam. Business Intelligence for Small Enterprises. An Open Source Approach. Royal Institute of Technology. p. 10. Mayo, 2008.

Knightsbridge Solutions⁵ propone simplificar el proceso de identificación de los beneficios considerándolos por separado en dos categorías principales: mejoras en los ingresos y reducción de costos [54]. Las mejoras en los ingresos se definen como las actividades provechosas que resultan de las decisiones que los individuos toman mediante el uso de la información que proviene de la solución de BI. Mientras que el ahorro de costos se define como la diferencia entre los costos asociados a la nueva implementación de BI comparado con los costos asociados al mantenimiento del entorno de información existente.

⁵ Knightsbridge Solutions, LLC es una firma consultora que se enfoca en las disciplinas de gestión de la información de inteligencia empresarial, almacenamiento de datos, integración de datos y calidad de la información. Esta fue adquirida por la empresa Hewlett-Packard en el año 2007.

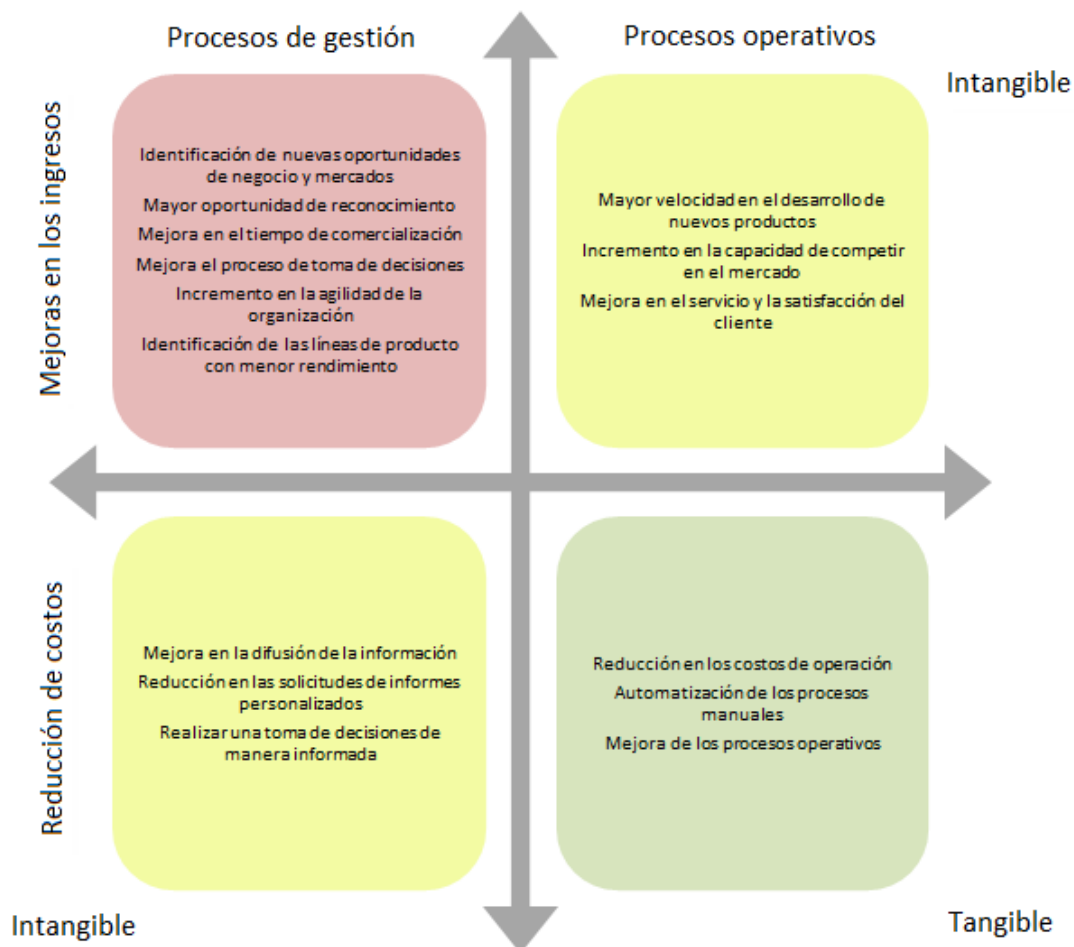
Steve y Nancy Williams identificaron dos dimensiones adicionales para la clasificación de los beneficios del BI [55]. Ellos afirman que la Inteligencia de Negocios se utiliza en primera instancia dentro de los procesos de gestión de la organización. Las decisiones que se toman en el nivel administrativo impactan en los procesos operativos, lo cual contribuye finalmente a que se incrementen los ingresos o se reduzcan los costos.

Estas cuatro dimensiones para los beneficios del BI (mejoras en los ingresos, reducción de costos, mejora de los procesos de gestión y mejora de los procesos operativos) son combinadas por Aliyev en el cuadrante de la figura 24, con la finalidad de clasificar y determinar el nivel de tangibilidad de los beneficios típicos que se obtienen con la implementación de BI.

En la parte superior izquierda de la figura se identifica el impacto de la solución de BI en la mejora de los procesos de gestión. Este es un beneficio que resulta difícil de medir porque generalmente no hay una retribución directa. La parte inferior derecha abarca los beneficios derivados de la mejora de los procesos operativos asociados con el ahorro de costos. Finalmente, en la parte superior derecha e inferior izquierda se resumen los beneficios intangibles como, por ejemplo, la mejora en la satisfacción de los clientes, los cuales pueden ser estimados a través de métodos de evaluación más complejos.

Estos beneficios, presentados por Moss y Atre en [56], pueden ayudar a dimensionar las mejoras que se obtienen con la implementación de este tipo de iniciativas. Sin embargo, realizar una valoración precisa tarda más tiempo del razonable para tomar una decisión de inversión. Esto se debe a que este tipo de estimaciones son complicadas y requieren de la asignación de un valor monetario a los beneficios intangibles [44]. Por lo que, generalmente, el rendimiento de este tipo de inversiones se calcula según las características de las suites de BI y su adecuación a los requerimientos específicos de las empresas.

Figura 25: Cuadrante para la clasificación de los beneficios del BI



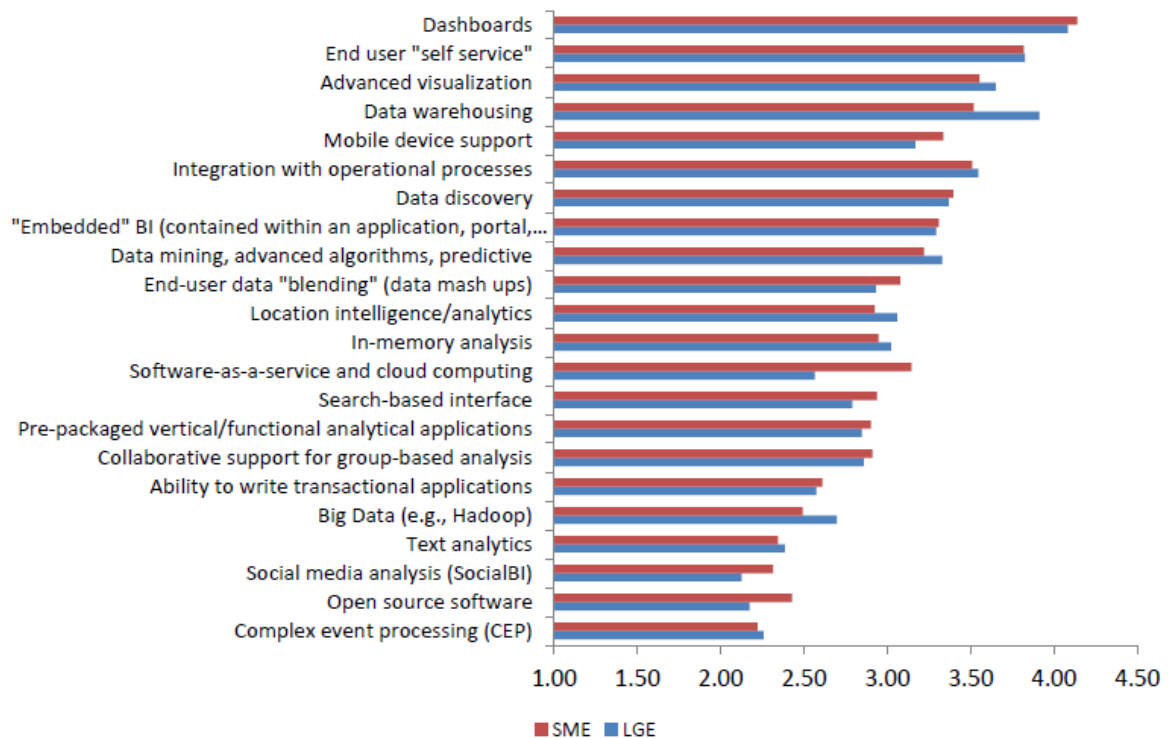
Fuente: Adaptado de ALIYEV, Rustam. Business Intelligence for Small Enterprises. An Open Source Approach. Royal Institute of Technology. p. 11. Mayo, 2008.

Según un estudio realizado por Dresner Advisory Services⁶, las PYMES adoptan una amplia variedad de tecnologías y características de BI, de la misma manera en que lo hacen las grandes organizaciones. Las preferencias comunes en ambos tipos de empresas son los cuadros de mando y el auto-servicio para los usuarios finales. Pero las organizaciones pequeñas no dan prioridad a los *Data Warehouses* como lo hacen a menudo las grandes organizaciones, además de que demuestran un mayor interés en el *Software as a Service* y las aplicaciones

⁶ Dresner Advisory Services, LLC es una firma de asesoría independiente, reconocida por su autoría en las áreas de Business Intelligence y Performance Management.

en la nube. Por otra parte, son menos propensas a adoptar *Big Data* y minería de datos, pero están más interesadas en el software *Open Source* y en el BI social [57]. En la siguiente figura se realiza una evaluación de las características de BI más solicitadas para ambos tipos de organizaciones.

Figura 26: Prioridades tecnológicas de las PYMEs frente a las grandes empresas



Fuente: DRESNER. Small and Mid-Sized Enterprise Business Intelligence Market Study. p. 21. Dresner Advisory Services LLC, 2014.

Acorde con este estudio, entre todas las prioridades de BI, el interés de las PYMEs en el software *Open Source* se incrementó para el año 2014. El software *Open Source* y el *Software as a Service (Cloud)* fueron las únicas categorías que incrementaron notablemente su preferencia en las organizaciones [57].

Por otra parte, ocurre que los proveedores de software licenciado tradicional venden el software para su uso inicial, en el cual se cobra una tasa de licencia por adelantado y requiere una cuota anual de mantenimiento. Además, es responsabilidad de los usuarios del software efectuar la compra del hardware

necesario para soportar el sistema, instalar el producto, realizar la integración requerida de los componentes para su funcionamiento y configurar las herramientas para su uso. Debido a que los precios de estas licencias pueden llegar a ser bastante altos, el *Open Source* ha surgido en los últimos años como una forma de disminuir los costes de implementación de estas tecnologías.

Con el *Open Source*, surgieron comunidades de desarrolladores dispuestos a colaborar de forma gratuita en proyectos y a presentar el código en diferentes foros en línea. Este tipo de software es libre para descargar e instalar. Sin embargo, requiere realizar inversiones adicionales, de la misma manera que ocurre con el software licenciado tradicional. Aunque con el *Open Source* se puede ahorrar en los costos de la licencia del software, no se logra la reducción esperada en el costo total de propiedad. Esto debido a que no se ocupa de los gastos del personal asociados con la implementación y mantenimiento de una solución de BI. Así, aunque los costos de licenciamiento son menores, las soluciones *Open Source* requieren que los clientes compren y administren el hardware, además de que deben instalar, integrar, configurar y gestionar los diferentes componentes de software.

A diferencia de las soluciones licenciadas tradicionales y las herramientas *Open Source*, el *Software as a Service* concibe un modelo de prestación del servicio completamente diferente. Las tecnologías orientadas a la nube han aumentado su popularidad en el mercado actual ya que proporcionan servicios de BI y de *Data Warehouse* a través de internet. Los proveedores de esta clase de aplicaciones venden el software, pero también tienen la responsabilidad de proveer el hardware y de realizar su respectiva instalación e integración en el entorno de alojamiento. Los clientes de las soluciones orientadas a la nube generalmente no requieren más que un navegador para su uso.

Los servicios de computación en la nube son principalmente utilizados por pequeñas organizaciones y *start-ups* que no tienen gran experiencia en la administración de este tipo de inversiones en TI [58]. Las herramientas de BI orientadas a la nube son mucho más baratas para las pequeñas y medianas empresas debido a que no existe la necesidad de realizar mantenimiento tanto del hardware como del software, además de que los precios pueden variar dependiendo de los requerimientos de almacenamiento de datos. Sin embargo, este tipo de aplicaciones no permiten que los usuarios tengan los datos almacenados físicamente, lo que ocasiona una gran dependencia en el proveedor y cierta pérdida de control y seguridad de los datos.

De acuerdo a un estudio realizado en 2012, las tecnologías de *Business Intelligence* orientadas a la nube son las herramientas de administración del negocio que mayor crecimiento han experimentado. A finales del 2011, solamente el 13% de las organizaciones contaban con este tipo de aplicaciones en funcionamiento. Mientras que en 2012, alcanzaron una tasa de crecimiento anual del 84% [58]. La nube provee mejores soluciones de BI, más rápidas, menos costosas, además de que proporciona una buena relación costo-beneficio. Sin embargo, la seguridad no deja de ser una preocupación para las organizaciones, ya que en la medida en que estas orientan sus procesos a la toma de decisiones con base en los datos, más importante resulta asegurar la calidad y, por ende, la seguridad de los mismos. A pesar de esto, las herramientas de BI orientadas a la nube se han convertido en una tendencia para las PYMES. Si opta por la elección y ejecución de las tecnologías de Inteligencia de Negocios en la nube, se debe realizar una evaluación de las ventajas e inconvenientes que podrían presentarse en el desarrollo de las actividades diarias de su organización, teniendo en mente seleccionar la que más se ajuste a los requerimientos de información definidos previamente.

Generalmente, las soluciones en cada una de estas categorías cubren toda la gama de funcionalidades prioritarias de las organizaciones. En ocasiones, las suites de BI comerciales poseen funcionalidades más avanzadas. Sin embargo, por lo menos una característica adicional existente en un módulo requerido representa un incremento en los costos. Además, las funcionalidades del software de BI *Open Source* son lo suficientemente robustas para satisfacer las necesidades básicas de las pequeñas empresas. En la evaluación del software de BI es recomendable limitarse estrictamente a las funcionalidades que se van a utilizar y tratar de no incluir características únicas, que generan un valor añadido, pero que habitualmente resulta en una inversión más costosa.

A continuación se presentará un cuadro comparativo entre las herramientas de BI más representativas del mercado para proporcionarle al lector un marco general de las características que deben ser evaluadas al momento de realizar la selección. En la siguiente tabla se exponen los parámetros utilizados para la comparación entre las 12 suites seleccionadas.

Figura 27: Parámetros de comparación de las suites de BI

Característica	Descripción
Herramientas de integración de datos (ETL)	Con este criterio se evalúa si la herramienta permite realizar la Extracción, Transformación y Carga de los datos desde múltiples fuentes operacionales
Análisis Ad hoc	Define si la herramienta permite diseñar consultas para preguntas específicas del negocio
Análisis OLAP	Indica si la herramienta permite realizar análisis OLAP sobre los datos con el fin de estudiar su comportamiento desde diferentes perspectivas
Análisis de tendencias	Evalúa si el sistema proporciona las herramientas para realizar análisis que permita descubrir patrones o tendencias en los datos
Generación de reportes	Define si la herramienta permite diseñar consultas sobre información específica del negocio y enseñar los resultados en formatos o plantillas predefinidos
Dashboards	Indica si la herramienta permite la generación y utilización de tableros de control
Indicadores de rendimiento (KPI)	Evalúa si la herramienta permite configurar y monitorear los indicadores clave de rendimiento o KPIs
Servicios en la nube	Indica si es posible acceder a las funcionalidades que ofrece la herramienta a través de la nube
Facilidad de uso	Indica si la herramienta proporciona al usuario una interfaz gráfica que facilite el acceso a las características del sistema
Licencia	Define el tipo de licenciamiento de software utilizado por cada herramienta
Relación calidad/precio	Evalúa la relación entre el precio del producto y los beneficios que proporciona a la organización. Esta valoración se obtuvo del estudio de mercado sobre Inteligencia de Negocios en las pequeñas y medianas empresas realizado por Dresner Advisory Services en el año 2014

Fuente: Elaboración propia.

Figura 28: Comparación de las soluciones de BI para PYMES (A-L)

Característica	Actuate BIRT Designer	BIRST	GoodData	IBM Cognos Express	Jaspersoft	Logi Analytics
Herramientas de integración de datos (ETL)	X	X	X	X	X	X
Análisis Ad hoc	X	X	X	X	X	X
Análisis OLAP	X		X	X	X	X
Análisis de tendencias	X	X	X		X	
Generación de reportes	X	X	X	X	X	X
Dashboards	X		X		X	X
Indicadores de rendimiento (KPI)		X	X		X	
Servicios en la nube	X	X	X	X	X	X
Facilidad de uso	X	X	X	X	X	
Licencia	Libre	SaaS	SaaS	Comercial	Comercial ⁷	Comercial
Relación calidad/precio	4.14	4.36	3.90	3.47	4.37	4.10

Fuente: Elaboración propia.

⁷ Jaspersoft también cuenta con una edición Community para uso libre que cuenta con funcionalidades más reducidas que la edición empresarial.

Figura 29: Comparación de las soluciones de BI para PYMES (M-T)

Característica	Microsoft Share Point	Pentaho Community	QlikView	SAP Business Objects Edge	Tableau	Tibco Spotfire
Herramientas de integración de datos (ETL)	X	X	X	X	X	X
Análisis Ad hoc	X	X	X	X	X	X
Análisis OLAP		X		X	X	X
Análisis de tendencias				X	X	X
Generación de reportes	X	X	X	X	X	X
Dashboards	X	X	X	X	X	X
Indicadores de rendimiento (KPI)	X	X			X	
Servicios en la nube	X		X	X	X	X
Facilidad de uso	X		X	X	X	X
Licencia	Comercial	Libre	Comercial	Comercial	Comercial	Comercial
Relación calidad/precio	3.73	4.19	4.17	3.14	4.41	4.69

Fuente: Elaboración propia.

11. CONCLUSIONES

- En la actualidad es común encontrar que las PYMES utilicen sistemas de gestión del tipo ERP. Sin embargo, este tipo de sistemas se limitan a apoyar actividades de la organización como producción, distribución, facturación, nómina, contabilidad, entre otros y están diseñados para satisfacer los requerimientos básicos de información de las empresas.
- Para que una empresa pueda sobrevivir y posicionarse en los mercados modernos, altamente exigentes, debe tomar decisiones acertadas que le permitan obtener ventajas competitivas en su entorno. En este sentido, el *Business Intelligence* se presenta como una alternativa para que las pequeñas y medianas empresas cuenten con las herramientas requeridas para administrar sus actividades de manera más efectiva.
- La información es un activo de valor incalculable para el éxito empresarial y debe ser utilizada oportunamente en el proceso de toma de decisiones. Por lo tanto, las organizaciones deben contar con iniciativas corporativas orientadas a proveer a la gerencia la información relevante para responder a los desafíos del negocio.
- A menudo, se tiene una visión sesgada de lo que implica la Inteligencia de Negocios, asociándola exclusivamente con el análisis de datos y la generación de reportes. Hoy en día, la mayoría de las organizaciones conciben la Inteligencia de Negocios como una iniciativa más amplia que puede afectar positivamente los objetivos claves del negocio.
- Una estrategia de BI exitosa debe estar alineada con los objetivos empresariales, realizar un mejor uso de la información y permitir la integración de BI en las actividades estratégicas, tácticas y operacionales. Conducir la estrategia de BI teniendo en cuenta el contexto empresarial asegura la alineación de la visión organizacional, las metas y los procesos con las iniciativas de BI.
- Al momento de iniciar un proyecto de BI se debe asegurar la completitud de los datos, su calidad y coherencia. Esta es una actividad fundamental para la implementación de un entorno de BI, debido a que generalmente determina el

éxito del proyecto y puede obstaculizar las decisiones empresariales en todos los niveles de la organización.

- Se debe realizar una evaluación completa de los procesos y las tecnologías utilizados por la organización, ya que estos serán críticos para determinar las acciones a emprender en el proceso de implementación de una solución de Inteligencia de Negocios. Es preciso revisar y documentar los sistemas de información y los procedimientos de gobernanza de datos que están siendo utilizados, al igual que detallar cómo está siendo manipulada la información actualmente, lo cual debe incluir un inventario de todas las plataformas, tecnologías y herramientas que hacen parte de este proceso. También es importante documentar los usuarios con sus respectivos perfiles y cómo dichos usuarios están haciendo uso de la información organizacional. Todo esto, es necesario para establecer los objetivos de la estrategia de BI y la viabilidad de su ejecución.
- De manera general, se puede decir que una arquitectura típica de un sistema de Inteligencia de Negocios posee cinco capas de aplicación: fuentes de datos, carga de los datos, repositorios de datos, servidores intermedios y herramientas de explotación y visualización de la información. Cada capa está compuesta de diferentes tecnologías que pueden ser implementadas o no dependiendo de las características específicas del negocio, de los objetivos que se pretenden alcanzar con su implementación y del alcance del proyecto.
- Una de las etapas fundamentales durante la implementación de un sistema de Inteligencia de Negocios es la construcción del repositorio de datos. El Data Warehousing es un procedimiento que posibilita la extracción de datos desde múltiples fuentes operacionales, la integración y homogeneización de información almacenada en tipos y formatos diferentes, para la consolidación de un repositorio central que contiene datos actuales, históricos y resumizados que permiten soportar los procesos de toma de decisiones en los niveles tácticos y estratégicos de la organización.
- Las bodegas de datos son diseñadas con un objetivo de explotación distinto al de las bases de datos transaccionales. Su construcción está orientada a la consulta de información sobre las actividades de la organización para operar eficientemente sobre los datos almacenados y facilitar el análisis en los niveles tácticos y estratégicos.

- La manera como se aborda el proceso de construcción de una bodega de datos es una decisión que puede impactar de forma significativa en el éxito de la implementación de un proyecto de *Data Warehouse*. Las variables que se ven afectadas por el enfoque adoptado son: el tiempo de finalización, el retorno de la inversión, la velocidad con que se obtienen beneficios, la satisfacción del usuario final, los recursos necesarios en cualquier etapa del proceso y la arquitectura de bodegas de datos seleccionada.
- Es más probable que una organización tenga éxito en la implementación del *Data Warehouse* utilizando el enfoque de Inmon si cuenta con un gran número de profesionales de TI, planea un proyecto que involucre a todas las áreas del negocio y puede esperar para ver los resultados en lapsos de tiempo considerables, generalmente entre cuatro y nueve meses.
- El enfoque de Kimball proporciona una implementación más rápida, al menos para la entrega del repositorio de información inicial. Por lo general, el desarrollo del primer *Data Mart* funcional para un área específica del negocio puede tardar 90 días y los *Data Marts* posteriores entre 60 y 90 días cada uno. Este modelo resulta favorable cuando la organización puede conformar equipos de trabajo interdisciplinarios más pequeños que los requeridos por el enfoque de Inmon; además, cuando se desea almacenar en mayor medida información sobre las métricas del negocio.
- Las herramientas tecnológicas adecuadas para las grandes compañías, no siempre resultan convenientes para las PYMES. Esto debido a que sus condiciones y características resultan significativamente diferentes. En ese sentido las PYMES requieren de soluciones tecnológicas que proporcionen una mayor flexibilidad, menor inversión en infraestructura y herramientas que resulten sencillas y eficientes.
- La selección de una herramienta u otra depende de las prioridades del proyecto, es decir, de los requerimientos específicos de cada organización: tableros de control, automatización de informes, creación de cubos, entornos colaborativos, independencia de TI, entre otras, son funcionalidades que deben ser evaluadas en cada herramienta.

- Debido a la naturaleza de los proyectos de Inteligencia de Negocios, las técnicas tradicionales para la evaluación de la inversión como el valor actual neto, el retorno de la inversión y el costeo basado en actividades no resultan adecuadas para proyectos tecnológicos que generan beneficios estratégicos intangibles para la organización. La necesidad de una técnica que permitiera calcular con mayor precisión el costo de los sistemas de información llevo a que Gartner Group en los años 80 propusiera un método conocido como costo total de propiedad (TCO, por sus siglas en inglés *Total Cost of Ownership*).
- Existen diferentes alternativas al momento de implementar un sistema de Inteligencia de Negocios, que van desde el software licenciado tradicional, el software *Open Source* o las soluciones enfocadas a ofrecer servicios a través de la nube. La selección de uno u otro modelo dependerá de las características y necesidades de la organización.
- El alto costo de las licencias para la utilización de las plataformas de BI tradicionales puede ser una barrera para su implementación en las PYMES, además se debe tener en cuenta que la adquisición de la infraestructura tecnológica y la instalación, configuración e integración de los componentes es responsabilidad del usuario, lo que puede significar un incremento en los costos o la necesidad de recurso humano con habilidades técnicas específicas.
- Por otra parte, se tiene el software *Open Source*, libre para descargar e instalar sin ningún tipo de costo por licencia. Aunque con el *Open Source* se puede ahorrar en los costos de la licencia del software, no se logra la reducción esperada en el costo total de propiedad. Esto debido a que no se ocupa de los gastos del personal asociados con la implementación y mantenimiento de una solución de BI ni de la infraestructura tecnológica que requiere.
- Por último, se tiene el *Software as a Service*. Los proveedores de esta clase de aplicaciones venden el software, pero también tienen la responsabilidad de proveer el hardware y de realizar su respectiva instalación e integración en el entorno de alojamiento. Sin embargo, este tipo de aplicaciones no permiten que los usuarios tengan los datos almacenados físicamente, lo que ocasiona una gran dependencia en el proveedor y cierta pérdida de control y seguridad de los datos.

- Las soluciones en cada una de estas categorías cubren una amplia gama de funcionalidades prioritarias para las organizaciones y en ocasiones, las suites de BI comerciales incorporan herramientas más avanzadas. Sin embargo, las características del software de BI Open Source son lo suficientemente robustas para satisfacer las necesidades básicas de las pequeñas empresas. En la evaluación de la solución de BI es recomendable limitarse estrictamente a las funcionalidades que se van a utilizar y tratar de no incluir características únicas, que generan un valor añadido, pero que habitualmente resulta en una inversión más costosa.

BIBLIOGRAFÍA

- [1] HATCH, David y JUTRAS, Cindy. The ERP/BI Connection. Adding Value through Actionable Intelligence. Julio, 2009. 26 p.
- [2] LUHN, Hans Peter. A Business Intelligence system. IBM Journal of Research and Development. 1958.Vol. 2.4, p. 314-319.
- [3] REVELLI, Carlo. Intelligence stratégique sur Internet. Paris. 1998: Dunod.
- [4] CUI, Zhan; DAMIANI, Ernesto y LEIDA, Marcello. "Benefits of ontologies in real time data access." Digital EcoSystems and Technologies Conference, 2007. DEST'07. Inaugural IEEE-IES. IEEE, p. 392-397. 2007.
- [5] STACKOWIAK, Robert; RAYMAN, Joseph y GREENWALD, Rick. Oracle Data Warehousing & Business Intelligence. John Wiley & Sons, Wiley Publishing, Inc, Indianapolis, 2007.
- [6] RANJAN, Jayanthi. Business Intelligence: concepts, components, techniques and benefits. Journal of Theoretical and Applied Information Technology. Vol. 9, No 1, p. 60-70. 2009.
- [7] KHAN, Rafi Ahmad y QUADRI, S.M. K. Business Intelligence: an integrated approach. Business Intelligence Journal. Vol. 5, No.1, p. 64-70. Enero, 2012.
- [8] Gartner Research. Gartner Executive Program Survey of More Than 2,000 CIOs Shows Digital Technologies Are Top Priorities in 2013 [En línea]. STAMFORD, Conn. Enero 16, 2013. [Citado en 7 de Abril de 2015]. Disponible en: <<http://www.gartner.com/newsroom/id/2304615>>.
- [9] PANT, Prashant. Business intelligence (BI). How to build successful BI strategy. Deloitte Touche Tohmatsu Limited. 19 p. 2009.
- [10] LAW, Jeff. Business Intelligence: Good performance indicators. Sage ERP X3. White Paper.

- [11] VERBITSKIY, Yuriy y YEOH, William. Data quality management in a Business Intelligence environment: from the lens of metadata. 16th International Conference on Information Quality, 2011.
- [12] CANO, Josep Lluís. Business Intelligence: competir con información. 393 p. 2008.
- [13] HABJAN, A. y POPOVIC, A. Achieving Business Process Change With Improved Business Intelligence Systems: A Case Of Slovenian Company. 7th WSEAS International Conference on Applied Computer Science. Venecia, Italia. 2007.
- [14] Pegasus Software. 2008. Using Business Intelligence to navigate in stormy times: an SME's Guide.
- [15] GUARDA, Teresa; SANTOS, Manuel; PINTO, Filipe; AUGUSTO, Maria y SILVA, Carlos. Business Intelligence as a Competitive Advantage for SMEs. International Journal of Trade, Economics and Finance, Vol. 4, No. 4. Agosto, 2013.
- [16] GIBSON, Marcus y ARNOTT, David. "BI for Small Business: Assessment, Framework & Agenda." 7th Pacific Asia Conference on Information Systems. 2003.
- [17] KIRCHMER, M. Small and Medium Enterprises Also Benefit from MPE. High Performance Through Process Excellence, p. 147-157. 2011.
- [18] FULLER-LOVE, N. Management development in small firms. International Journal of Management Reviews, 8, p. 175-190. 2006.
- [19] BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.
- [20] LÖNNQVIST, A. y PIIRTIMÄKI, V. The measurement of business intelligence. Information Systems Management, p. 32-40, 2006.
- [21] VOICU, V.; ZIRRA, D. y CIOCIRLAN, D. Business intelligence effective solutions of management. Proceedings of the 10th WSEAS international

conference on Mathematics and computers in business and economics. Praga, República Checa. World Scientific and Engineering Academy and Society (WSEAS), 2009. Citado por: BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.

[22] LA GROUW, Gail. The Business Intelligence Guide. Drivers of Business Intelligence [BI], 2009. [En línea]. [Citado en 22 de Abril de 2015]. Disponible en: <http://www.thebusinessintelligenceguide.com/bi_strategy/Drivers_Of_BI.php>.

[23] LUNDBLAD, J. P. A review and critique of Rogers' diffusion of innovation theory as it applies to organisations. Organization Development Journal. p. 50-64. 2003. Citado por: BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.

[24] LEE, G. y XIA, W. Organizational size and IT innovation adoption: A meta-analysis. Information & Management, 43, p. 975-985. 2006. Citado por: BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.

[25] FLANAGIN, J. Social pressures on organizational website adoption. Human Communication Research, 26, p. 618-646. 2000. Citado por: BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.

[26] ROGERS, E. M. Diffusion of Innovations, New York, Free Press. 1995. Citado por: BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.

[27] DANIEL, E. y MYERS, A. Levelling the playing field: electronic commerce in SMEs. Capstone Publishing Limited. 2000. Citado por: BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.

[28] GRIFFITH, T. L.; SAWYER, J. E. y NEALE, M. A. Virtualness and knowledge in teams: managing the love triangle of organisations, individuals, and information technology. MIS Quarterly, 27, p. 265-287. 2003. Citado por: BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.

[29] WAARTS, E.; EVERDINGEN, Y. M. y HILLEGERSBERG, J. The dynamics of factors affecting the adoption of innovations. Journal of Product Innovation Management, 19, p. 412-423. 2002. Citado por: BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.

[30] ALSHAWI, S.; MISSI, F. e IRANI, Z. Organisational, technical and data quality factors in CRM adoption--SMEs perspective. Industrial Marketing Management, 40, p. 376-383. 2011. Citado por: BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.

[31] HWANG, H. G.; KU, C. Y.; YEN, D. V. y CHENG, C. C. Critical factors influencing the adoption of data warehouse technology: A study of the banking industry in Taiwan. Decision Support Systems, p. 1-21. 2004. Citado por: BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.

[32] GHOBAKHLOO, M.; ARIAS-ARANDA, D. y BENITEZ-AMADO, J. Adoption of e-commerce applications in SMEs. *Industrial Management & Data Systems*, 111, p. 1238-1269. 2011. Citado por: BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.

[33] THONG, J. 1999. An integrated model of information systems adoption in small business. *Journal of Management Information Systems*, 15 p. 187-214. Citado por: BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.

[34] CHANG, S. I.; HUNG, S. Y.; YEN, D. C. y LEE, P. J. Critical Factors of ERP Adoption for Small-and Medium-Sized Enterprises: An Empirical Study. *Journal of Global Information Management (JGIM)*. p. 82-106. 2010. Citado por: BOONSIRITOMACHAI, Waranpong; MCGRATH, Michael y BURGESS, Stephen. A research framework for the adoption of Business Intelligence by Small and Medium-sized enterprises. Small Enterprise Association of Australia and New Zealand. 27th Annual SEAANZ Conference Proceedings. 22 p. Julio, 2014.

[35] FEDOUAKI, Faycal; OKAR, Chafik y EL ALAMI, Semma. A maturity model for Business Intelligence System Project in Small and Medium-sized Enterprises: an empirical investigation. *IJCSI - International Journal of Computer Science Issues*, Vol. 10, No. 1, p. 61-69. Noviembre, 2013.

[36] SHAABAN, Essam; HELMY, Yehia; KHEDR, Ayman y NASR, Mona. Business Intelligence Maturity Models: Toward New Integrated Model. 2012.

[37] CHAUDHURI, Surajit; DAYAL, Umeshwar y NARASAYYA, Vivek. An overview of Business Intelligence Technology. *Communications of the ACM*. Vol. 54, No. 8. p. 88-98. Agosto, 2011.

[38] ONG, In Lih; SIEW, Pei Hwa y WONG, Siew Fan. A Five-Layered Business Intelligence Architecture. *Communications of the IBIMA*. Vol. 2011. 11 p.

- [39] INMON, W.H. Building the Data Warehouse. 3 ed. 412 p. New York. Wiley and Sons, 2002.
- [40] KIMBALL, Ralph. The Data Warehouse Toolkit. Wiley and Sons, New York, 1996. p. 310.
- [41] BERNABÉU, Ricardo Darío. Hefesto. DATA WAREHOUSING: Investigación y Sistematización de Conceptos. HEFESTO: Metodología para la Construcción de un Data Warehouse. 148 p. Córdoba, Argentina. Julio, 2010.
- [42] BALLARD, Chuck; HERREMAN, Dirk; SCHAU, Don; BELL, Rhonda; KIM, Eunsang y VALENCIC, Ann. Data Modeling Techniques for Data Warehousing. IBM's International Technical Support Organization. 200 p. Febrero, 1998.
- [43] BALLARD, Chuck; FARRELL, Daniel M.; GUPTA, Amit; MAZUELA, Carlos y VOHNIK Stanislav. Dimensional Modeling: In a Business Intelligence Environment. 1 ed. Redbooks, IBM's International Technical Support Organization. 648 p. Marzo, 2006.
- [44] ALIYEV, Rustam. Business Intelligence for Small Enterprises. An Open Source Approach. Royal Institute of Technology. 46 p. Mayo, 2008.
- [45] GOLFARELLI, Matteo y RIZZI, Stefano. Data Warehouse Design: Modern Principles and Methodologies. Traducido por Claudio Pagliarani. New Delhi. 458 p. Tata McGraw Hill, 2009.
- [46] JUKIC, Nenad; JUKIC, Boris y MALLIARIS, Mary. Online Analytical Processing (OLAP) for Decision Support. 24 p. 2008.
- [47] MEULEN, Ron Van Der y RIVERA, Janessa. Gartner Says Worldwide Business Intelligence Software Revenue to Grow 7 Percent in 2013. [En línea]. STAMFORD, Conn. Febrero 19, 2013. [Citado en 16 de Mayo de 2015]. Disponible en: <<http://www.gartner.com/newsroom/id/2340216>>
- [48] KIMBALL, Ralph; REEVES, Laura; ROSS, Margy y THORNTHWAITE, Warren. The Data Warehouse Lifecycle Toolkit. 2 ed. 672 p. Wiley and Sons, 2008.

[49] BRESLIN, Mary. Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models. Business Intelligence Journal. p. 6-20. Invierno de 2004.

[50] MIMNO, P. Project Plan for Bottom-Up Development. TDWI FlashPoint, TDWI Web site. Agosto 28, 2002. Citado por: BRESLIN, Mary. Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models. Business Intelligence Journal, invierno de 2004.

[51] BERNARDINO, Jorge. Emerging Business Intelligence Technologies for SMEs. 29 p. Institute Polytechnic of Coimbra – ISEC, Portugal. IGI Global, 2014.

[52] IRANI, Z. y P. E. D.; Love. "The Propagation of Technology Management Taxonomies for Evaluating Investments in Information Systems." Journal of Management Information Systems 17(3), 2001. Citado por: ALIYEV, Rustam. Business Intelligence for Small Enterprises. An Open Source Approach. Royal Institute of Technology. 46 p. Mayo, 2008.

[53] TURBAN, Efraim; LEIDNER, Dorothy; MCLEAN, Ephraim y WETHERBE, James. "Information Technology for Management: Transforming Organizations in the Digital Economy." In Information Technology for Management: Transforming Organizations in the Digital Economy, 2006. Citado por: ALIYEV, Rustam. Business Intelligence for Small Enterprises. An Open Source Approach. Royal Institute of Technology. 46 p. Mayo, 2008.

[54] KNIGHTSBRIDGE. Measuring the value of business intelligence and data warehousing initiatives. Knightsbridge Solutions LLC, 2005. Citado por: ALIYEV, Rustam. Business Intelligence for Small Enterprises. An Open Source Approach. Royal Institute of Technology. 46 p. Mayo, 2008.

[55] WILLIAMS, Steve y WILLIAMS, Nancy. "The Business Value of Business Intelligence." p. 32-43. Business Intelligence Journal, 2003. Citado por: ALIYEV, Rustam. Business Intelligence for Small Enterprises. An Open Source Approach. Royal Institute of Technology. 46 p. Mayo, 2008.

[56] MOSS, Larissa T. y ATRE, Shaku. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. 2003. Citado por: ALIYEV, Rustam. Business Intelligence for Small Enterprises. An Open Source Approach. Royal Institute of Technology. 46 p. Mayo, 2008.

[57] DRESNER. Small and Mid-Sized Enterprise Business Intelligence Market Study. Dresner Advisory Services LLC, 2014.

[58] LANKE, Govinda rajulu y BHUVANESWARI, T. Giving Intelligence to SMEs. SCSVMV University, Enathur, Kanchipuram, India. International Conference on Computing and Intelligence Systems, Vol. 4. Marzo, 2015.