



Tesis de Maestría

**MARCO DE TRABAJO PARA LA IMPLEMENTACIÓN DE BIG DATA
ANALYTICS EN EL CONTEXTO ESPECÍFICO DEL ÁREA DE SALUD**

Carolina Paula Blanco

FACULTAD DE INGENIERÍA
Maestría en Tecnología de la Información

Director de Tesis
Prof. Dr. Nicolás D'Ippolito

Buenos Aires, Argentina 2015



**MARCO DE TRABAJO PARA LA IMPLEMENTACIÓN DE BIG DATA
ANALYTICS EN EL CONTEXTO ESPECÍFICO DEL ÁREA DE SALUD**

Carolina Paula Blanco

Director de Tesis

Presidente del Jurado

Jurado

Jurado Externo

Defensa de la Tesis

Ciudad Autónoma de Buenos Aires, a los ____ días del mes de _____

de _____.

*A mi madre, por su apoyo incondicional
en cada una de las decisiones de mi vida.*

*A mi querido James, por recordarme que todo es posible
cuando ponemos las ganas y el empeño suficientes.*

*A mis amigos, por ser el respiro y el impulso
en tiempos difíciles.*

RECONOCIMIENTOS

Cuando pienso en el camino recorrido durante el desarrollo de mi tesis, veo mezclados momentos de interminables hojas en blanco con algunos de breve inspiración que lograron facilitar el avance. Quiero aprovechar esta oportunidad para reconocer a aquellas personas que me acompañaron durante todos esos momentos.

Mi más profundo agradecimiento a mi director de tesis, Nicolás D'Ippolito, por tener tan claro el camino de la investigación y compartirlo conmigo, por ser mi guía en el proceso de aprendizaje y devolverme el rumbo cuando pensé que lo había perdido.

Mi mayor gratitud a mis profesores de la Maestría en Tecnología de la Información de la Universidad de Palermo por haberme abierto las puertas del mundo de la ingeniería del software y haberme dado nuevas perspectivas que permitieron este trabajo interdisciplinario que conjuga las TI y el área de salud.

Agradezco a mis compañeros de la maestría, Andrés, Daniela, Javier, Mónica y Pablo, porque cada trabajo en equipo fue una experiencia enriquecedora.

Finalmente, para mis compañeros de trabajo, pasados y actuales; al staff de la Clínica ODEM en Buenos Aires por su paciencia durante la cursada de mi maestría; y a la gente de RN Engineering en Canadá, por ser fuente inagotable para mi aprendizaje.

RESUMEN DE LA TESIS

Este trabajo se propone establecer lineamientos para el aprovechamiento de los grandes volúmenes y tipos de datos disponibles en el área de la salud mediante la creación de un marco de trabajo para la implementación de *big data analytics* adaptado para el sector. Su desarrollo está basado en la comparación de marcos de trabajo para la implementación de *business intelligence* y en la investigación de los factores que definen una implementación de *big data analytics* exitosa. Además, se analizan las complejidades específicas del sector de salud.

Durante esta investigación, se encontraron factores críticos de éxito para la implementación de *business intelligence* que resultaron ser comunes entre varios autores y se integraron nuevos factores propios de *big data* y del sector de salud. Se espera que el marco de trabajo propuesto ayude a implementar *big data analytics* en el campo de la salud para facilitar la gestión y la toma de decisiones mediante la integración de datos clínicos y administrativos.

Palabras claves: *Big data analytics*, *big data*, *business analytics*, *business intelligence*, marco de trabajo, salud, implementación, factores críticos de éxito.

ABSTRACT

The purpose of this study is to establish a set of guidelines needed to benefit from the increasingly high volumes and types of data available in the healthcare field, by creating a tailored big data analytics framework. The construction of the framework is based on the comparison of business intelligence implementation frameworks and a literature review on the key success factors for the implementation of big data analytics initiatives. Besides, specific healthcare field complexities are analyzed in order to tailor the framework.

Several common key success factors for the implementation of business intelligence were found by comparing research studies. Also, new factors related to big data and sector- specific challenges were discussed. It is expected that the new framework will help the healthcare sector on the implementation of big data analytics solutions in order to improve management practices and decision making through clinical and administrative data integration.

Keywords: Big data analytics, big data, business analytics, business intelligence, framework, healthcare, implementation, key success factors.

TABLA DE CONTENIDOS

1. INTRODUCCIÓN	1
1.1. Definición del Problema.....	2
1.2. Justificación del Estudio.....	6
1.3. Objetivos	8
1.3.1. Objetivo General.....	8
1.3.2. Objetivos Específicos.....	9
1.4. Estructura de la Tesis	9
2. FUNDAMENTOS TEÓRICOS.....	11
2.1. Business Intelligence.....	11
2.1.1. OLTP vs. OLAP.....	12
2.1.2. Data Warehouse	13
2.1.3. Business Intelligence vs. Business Analytics	20
2.2. Big Data y Big Data Analytics.....	25
2.3. Marcos de Trabajo para la Implementación de BI	28
2.3.1. Marco de Trabajo LOBI.....	30
2.3.2. Marco de Trabajo sobre los Componentes de BI de TDWI.....	34
2.3.3. Marco de Trabajo para BI basado en Factores Críticos de Éxito	38
2.3.4. Marco de Trabajo de Gartner	44
2.4. BI en el Contexto de Salud.....	49
3. DESARROLLO DEL MARCO PROPUESTO	52
3.1. Generalidades del Marco de Trabajo	52
3.2. Primera Etapa: Factores para la Implementación de BI.....	53
3.2.1. Comparación de los Marcos de Trabajo para BI	53
3.2.2. Agrupamiento de los Factores de BI.....	55
3.3. Segunda Etapa: Factores para la Implementación de BDA	57
3.3.1. Análisis de las Diferencias entre la Implementación de BI y BDA.....	57
3.3.2. Incorporación de Factores Inherentes a BDA	60
3.4. Tercera Etapa: Factores Específicos del Área de Salud.....	65
3.4.1. Estudio de BDA Aplicada a la Calidad del Servicio de Salud	65
3.4.2. Análisis del Valor para el Área de Salud mediante Big Data	68
3.4.3. Adaptación del Marco al Área de Salud	70
3.5. Marco de Trabajo para la Implementación de BDA en el Área de Salud.....	74
4. CONCLUSIONES	77

5. FUTURAS LÍNEAS DE INVESTIGACIÓN.....	81
BIBLIOGRAFÍA.....	82
APÉNDICES	86

LISTA DE TABLAS

Tabla 1. Abordaje de la Investigación	5
Tabla 2. Diferencias entre Sistemas Operacionales e Informacionales	13
Tabla 3. Diferencias Principales entre BI y BA.....	22
Tabla 4. Comparación de Capacidades de BA según IBM y SAS Institute Inc.	24
Tabla 5. Factores Críticos de Éxito de BI Agrupados para el Nuevo Marco.....	56
Tabla 6. Desafíos de Big Data y sus Soluciones.....	61
Tabla 7. Factores Inherentes a BDA y su Relación con los Factores de BI	63

LISTA DE FIGURAS

Figura 1. Esquema Estrella vs. Copo de Nieve.....	16
Figura 2. Slice, Dice, Drill Down/Up, Roll-up y Pivot.....	18
Figura 3. Las 4 Vs de Big Data.....	27
Figura 4. Niveles LOBI.....	30
Figura 5. Diez Pasos para una Implementación LOBI Exitosa	32
Figura 6. Marco de Trabajo sobre los Componentes de BI de TDWI.....	36
Figura 7. Marco de Trabajo para BI basado en Factores Críticos de Éxito	42
Figura 8. Factores Críticos de Éxito para la Implementación de BI.....	43
Figura 9. Marco de Trabajo para BI y Gestión del Rendimiento de Gartner.....	46
Figura 10. Marco para BI, Analytics y Gestión del Rendimiento de Gartner	47
Figura 11. Marco de Trabajo para Business Analytics de Gartner	48
Figura 12. Factores Comunes dentro de los Marcos de BI.....	54
Figura 13. Primera Etapa: Factores para la Implementación de BI.....	57
Figura 14. Triple Plataforma de Big Data.....	60
Figura 15. Segunda Etapa: Factores Inherentes a Big Data Analytics	65
Figura 16. Caminos hacia el Valor para el Área de la Salud	69
Figura 17. Marco de Trabajo Propuesto	76

LISTA DE SIGLAS

SIGLA	SIGNIFICADO
BA	Business Analytics
BAM	Business Activity Monitoring
BDA	Big Data Analytics
BI	Business Intelligence
CIO	Chief Information Officer
DW	Data Warehouse
EDW	Enterprise Data Warehouse
ERP	Enterprise Resource Planning
ETL	Extract, Transform and Load
IBM	International Business Machines
KPIs	Key Performance Indicators
LOBI	Ladder of Business Intelligence
OLAP	On-Line Analytical Processing
OLTP	On-line Transaction Processing
RFID	Radio-Frequency IDentification
RSS	Rich Site Summary
s.f.	sin fecha
TB	Terabyte
TDWI	The Data Warehousing Institute
TI	Tecnología de la Información
XML	eXtensible Markup Language

1. INTRODUCCIÓN

Learn why the world wags and what wags it. That is the only thing which the mind can never exhaust, never alienate, never be tortured by, never fear or distrust, and never dream of regretting.

-T.H. White, *The Once and Future King*, 1958.

Aprender qué mueve al mundo es una tarea fascinante e inagotable, como lo expresa White en esta cita. En nuestros tiempos, la disponibilidad de datos y la capacidad de construir conocimiento presenta una oferta cada vez más amplia. Pero ¿cómo pueden estos datos asistir en lo que las organizaciones resuelven cada día? La optimización en la toma de decisiones ha sido objeto de estudio desde hace más de medio siglo y el término *business intelligence*, como método basado en datos para lograr la optimización, fue popularizado en 1989 (Power, 2007).

Brynjolfsson, Hit y Kim (2011) encontraron que las compañías que basan sus decisiones en datos son más productivas y mostraron una correlación entre este tipo de práctica al momento de tomar decisiones y un aumento en la rentabilidad, en el valor de mercado, y una optimización en la utilización de bienes dentro de estas compañías.

Tras la experiencia de implementación de soluciones de *business intelligence*, se han creado marcos de trabajo que reúnen los factores críticos de éxito con el fin de asistir a las empresas al momento de emprender proyectos de este tipo.

Sin embargo, año tras año los datos digitales existentes a nivel mundial crecen de forma exponencial llegando a niveles nunca antes vistos. Este volumen ya no puede ser manejado utilizando bases de datos relacionales. Los entornos tradicionales de datos y los métodos de *business intelligence* presentan limitaciones para manipular múltiples formatos y grandes volúmenes. Los tiempos de procesamiento se vuelven insostenibles. Esta situación ha llevado a la creación de nuevas soluciones que dan origen al concepto de *big data analytics*.

El área de salud en particular es uno de los dominios en los cuales la explosión de datos digitales disponibles ha sido notoria, especialmente desde el advenimiento de la historia clínica electrónica y los dispositivos inteligentes que registran distintos parámetros del usuario relacionados con la salud, como por ejemplo la distancia recorrida, la calorías gastadas, el ritmo cardíaco, etc.

Este trabajo se propone establecer lineamientos necesarios para el aprovechamiento de los grandes volúmenes y tipos de datos disponibles, en constante crecimiento, mediante la creación de un marco de trabajo para la implementación de *big data analytics* en el área de la salud.

1.1. Definición del Problema

Frente a las limitaciones de los entornos tradicionales de datos, se genera la necesidad de un nuevo concepto para el análisis y procesamiento de datos. El término *big data* ha incrementado su popularidad y su utilización en las empresas está en constante crecimiento. El concepto se basa en grandes volúmenes de datos, de una gran

diversidad, recolectados a distinta frecuencia y velocidad. Pero mayor cantidad de datos no significa más información. Para poder explorar *big data* y aprovechar al máximo las posibilidades que estos datos representan es necesario utilizar un conjunto de técnicas y herramientas que la soporten.

Nuevas capacidades de procesamiento y escalabilidad combinadas con el volumen y la diversidad de datos que provee *big data*, permiten descubrir nuevas realidades de negocio que hasta hoy no han podido ser reveladas. Esta combinación es denominada *big data analytics*.

A pesar de la creciente disponibilidad de datos, el sector de salud se enfrenta con grandes dificultades. Una de las más importantes, se relaciona con lograr mantener la calidad de la prestación del servicio, con costos que son cada vez más altos. Estos costos no siempre están relacionados con una población que necesita mayores cuidados médicos, sino con ineficiencias en la administración del servicio. Berwick y Hackbarth (como se cita en Lallemand, 2012) encontraron que más de un tercio del gasto anual en servicios de salud en los Estados Unidos se debe a prácticas redundantes o ineficientes; un resumen de sus resultados se presenta en el Apéndice A. Además, identifican seis áreas principales de malgasto de recursos que incluyen:

- *Fallas en la provisión del tratamiento:* Falta de adopción de las mejores prácticas o pobre ejecución de la terapéutica indicada.
- *Fallas en la coordinación del tratamiento:* Tratamiento fragmentado en la transición entre profesionales.

- *Fallas en el nomenclador de honorarios profesionales:* El precio del servicio excede ampliamente el costo más un margen de ganancia razonable.
- *Sobre prestación:* Tratamientos basados en hábitos del profesional en lugar de la preferencia del paciente.
- *Complejidad administrativa:* Reglas y procedimientos ineficientes.
- *Fraude y abuso del sistema:* Facturación de prestaciones clasificadas de manera inexacta o no realizadas (Lallemand, 2012).

El aprovechamiento de los grandes volúmenes de datos disponibles mediante la utilización de *big data analytics* puede ser la llave para resolver, desde la tecnología de la información, problemas históricos en el campo de la salud, como por ejemplo, cómo mantener y mejorar la calidad de la prestación del servicio.

Si bien existen marcos de trabajo para la implementación de *business intelligence*, después de una minuciosa revisión bibliográfica, no se han encontrado marcos de trabajo para la implementación de *big data analytics* específicamente adaptados al área de salud.

Entonces, ¿cómo se relaciona *big data analytics* con *business intelligence*? ¿En qué áreas puede *big data analytics* ayudar al sector de salud? ¿Qué factores críticos de éxito se deberían tener en cuenta para una implementación de *big data analytics*?

¿Qué factores específicos para el área de salud se deben tener en cuenta al momento de implementar una solución de *big data analytics*?

Se presenta un resumen del abordaje de la investigación en la Tabla 1.

Tabla 1. Abordaje de la Investigación

Instancia	Descripción
Observación	<p>Los datos digitales que se generan a nivel mundial crecen a un ritmo cada vez más acelerado y la tecnología disponible presenta limitaciones para el almacenamiento y procesamiento de estos datos.</p> <p>El área de salud en particular es uno de los dominios en los cuales la explosión de datos digitales disponibles ha sido notoria.</p> <p>A pesar de la creciente disponibilidad de datos, el sector de salud se enfrenta con grandes dificultades, como por ejemplo lograr mantener la calidad de la prestación del servicio, con costos cada vez más altos.</p>
Bases para la investigación	<p>Las organizaciones que toman decisiones basadas en datos mejoran su productividad (Brynjolfsson et al., 2011).</p> <p>Existen marcos de trabajo con el fin de asistir a las empresas en la implementación de <i>business intelligence</i>.</p> <p><i>Big data analytics</i> trae nuevas capacidades de procesamiento y escalabilidad para aprovechar los grandes volúmenes y la diversidad de los datos digitales, permitiendo descubrir nuevas realidades de negocio que hasta hoy no han podido ser reveladas.</p>

Definición del problema	No se han encontrado marcos de trabajo para la implementación de <i>big data analytics</i> adaptados al área de salud.
Objetivo	Construir un marco de trabajo para la implementación de <i>big data analytics</i> en el área de salud.
Beneficios esperados	Ayudar al sector de salud a mejorar la gestión y la toma de decisiones tanto clínica como administrativa mediante la integración de datos de ambos dominios.
Preguntas	<p>¿Cómo se relaciona <i>big data analytics</i> con <i>business intelligence</i>?</p> <p>¿En qué áreas puede <i>big data analytics</i> ayudar al sector de salud?</p> <p>¿Qué factores críticos de éxito se deberían tener en cuenta para una implementación de <i>big data analytics</i>?</p> <p>¿Qué factores específicos para el área de salud se deben tener en cuenta al momento de implementar una solución de <i>big data analytics</i>?</p>

Fuente: Elaboración propia

1.2. Justificación del Estudio

Davenport (2009) dice que una de las formas de mejorar la toma de decisiones es refinando el análisis de datos. Además, destaca que las organizaciones que utilizan el análisis de datos como ventaja competitiva aplican los datos de manera de optimizar operaciones en grados sin precedencia y transforman la tecnología de una herramienta de soporte en un arma estratégica (Davenport, 2006). Por otro lado, resalta:

En un momento en que las compañías en muchas industrias ofrecen productos similares y utilizan tecnologías comparables, los procesos de negocio son los puntos de diferenciación fundamentales. Y las organizaciones que compiten utilizando el análisis de datos extraen cada gota de valor de esos procesos. Entonces, como otras compañías, ellas saben qué productos sus clientes quieren, pero también saben qué precios sus clientes pagarán, cuántos productos cada cliente comprará en su vida y qué desencadena que el público compre más (Davenport, 2006, p. 1).

Weill y Aral (2006) realizaron un estudio en el cual recopilaban información sobre inversiones de TI de 147 compañías norteamericanas en un período de 5 años y encontraron, entre otros resultados, que las organizaciones de mayor desempeño en la industria de ventas por mayor y menor y en la industria de transporte gastan 11% más en TI que el promedio, con portfolios que se orientan a la inversión para la obtención de información, indicando que la ventaja competitiva está en el uso efectivo de la información.

Para el dominio de la salud, investigadores del McKinsey Global Institute¹ hablan de *big data* como una revolución que acelera el valor y la innovación en el sector, e identifican cinco áreas de obtención de valor mediante la utilización de *big data*: estilo de vida favorable, cuidados apropiados, prestador más idóneo, valor agregado e innovación apropiada (Groves, Kayyali, Knott, y Van Kuiken, 2013). Estas áreas serán estudiadas en la Sección 3.4.2.

¹ El McKinsey Global Institute es una firma internacional compuesta por más de 9.000 consultores y casi 2.000 profesionales de la rama de la investigación y la información, que proveen desde consultoría para la alta gerencia hasta coaching para empleados de primera línea.

Una investigación sobre *big data* en Estados Unidos, estima que es posible agregar valor al sector de salud norteamericano en unos 300 mil millones de dólares al año, casi el doble del gasto total anual de España en el mismo sector (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh y Byers, 2011). Los resultados de la investigación se resumen en el Apéndice B.

Entonces podemos decir que existe valor en los datos para el área de salud. Este trabajo está orientado a ayudar al campo de la salud a aprovechar el valor potencial de los datos ante la ausencia de marcos de trabajo para la implementación de *big data analytics* que se adapten a las problemáticas específicas de este sector.

1.3. Objetivos

1.3.1. Objetivo General

El objetivo general de esta tesis es construir un marco de trabajo para la implementación de *big data analytics* en el área de salud, mediante una visión de alto nivel que permita definir e integrar conceptos y criterios para llevar adelante iniciativas de implementación de este tipo, de modo de permitir desplegar los recursos de manera eficiente y, en última instancia, poder mejorar la calidad del servicio de atención de salud.

1.3.2. Objetivos Específicos

El recorrido a seguir para alcanzar el objetivo general de este estudio está basado en los siguientes objetivos específicos:

- Comparar marcos de trabajo existentes para la implementación de *business intelligence*.
- Definir la relación existente entre *business intelligence* y *big data analytics*.
- Identificar aquellos conceptos que dentro de los marcos de trabajo para *business intelligence* puedan ser utilizados en la creación del nuevo marco.
- Reconocer y enumerar los conceptos faltantes dentro de los marcos existentes para la creación del nuevo marco para *big data analytics*.
- Desarrollar un nuevo marco de trabajo para la implementación de *big data analytics*.
- Adaptar el nuevo marco de trabajo al área de salud.

1.4. Estructura de la Tesis

La tesis está organizada en cinco partes: Introducción, Fundamentos Teóricos, Desarrollo del Marco Propuesto, Conclusiones y Futuras Líneas de Investigación.

En la primera parte se ha presentado la introducción, que se centró en la definición del problema y en la determinación del objetivo general y de los objetivos específicos que dirigen esta investigación.

La segunda parte describe el marco teórico donde se definen los conceptos generales que incluyen *business intelligence*, *data warehouse*, *business analytics*, *big data* y *big data analytics*. Además se estudian distintos marcos de trabajo para la implementación de *business intelligence* y se ubica *business intelligence* en el contexto de la salud.

En la tercera parte se desarrolla el marco de trabajo propuesto, basado en la comparación de los marcos de trabajo para la implementación de *business intelligence* y se investigan las diferencias entre la implementación de soluciones de *business intelligence* y *big data analytics*. También se exploran las posibilidades que ofrece *big data analytics* específicamente para el área de salud, se estudia el concepto de valor para el sector, y se distinguen factores propios que deben tenerse en cuenta para la adaptación del marco.

Finalmente, la cuarta parte presenta las conclusiones obtenidas como resultado de esta investigación y la quinta parte menciona las líneas de investigación futuras.

2. FUNDAMENTOS TEÓRICOS

2.1. Business Intelligence

Según Power (2007), la optimización en la toma de decisiones ha sido objeto de estudio desde la década de los sesenta, con la implementación de los primeros sistemas de soporte de decisión.

En su investigación sobre la historia de los sistemas de soporte de decisión, Power (2007) relata que el término *business intelligence* (BI) es popularizado recién en 1989 por un analista de Gartner² llamado Howard Dresner, y describe BI como:

Un conjunto de conceptos y métodos para mejorar la toma de decisiones utilizando sistemas de soporte basados en información. Los términos BI y compendios de información, reportes, herramientas de consulta y sistemas ejecutivos de información suelen utilizarse indistintamente. En general los sistemas de BI son sistemas de soporte de decisión basados en datos (Power, 2007).

Por otro lado, The Data Warehousing Institute (TDWI)³, destaca:

² Gartner es una organización dedicada a la investigación de las tecnologías de la información, fundada en 1979 con sede en Stamford, Connecticut, USA. Provee consultoría y programas ejecutivos entre otros servicios.

³ TDWI es el principal instituto de educación en el área de BI y Data Warehousing, fundado en 1995. Se dedica a la investigación, educación y certificación de profesionales dentro del sector de tecnología de la información.

BI unifica datos, tecnología, análisis y conocimiento humano para optimizar las necesidades del negocio y en última instancia conducir el éxito de la empresa. Los programas de BI generalmente combinan un *data warehouse* empresarial y una plataforma de BI o un conjunto de herramientas para transformar datos en información comercial que pueda ser utilizada y que dirija un plan de acción (TDWI, s.f.).

A continuación, las Secciones 2.1.1 a 2.1.3 describen algunos de los términos que serán empleados en esta tesis. El objetivo no es el estudio profundo de los mismos, sino ubicarlos en relación a BI para facilitar su comprensión en el contexto en que son utilizados.

2.1.1. OLTP vs. OLAP

La transformación de datos en información que provee BI lleva a distinguir dos tipos de sistemas: los sistemas operacionales y los informacionales.

On-line transaction processing (OLTP) se refiere a los sistemas que se utilizan en las operaciones diarias de una empresa, que facilitan el manejo de las aplicaciones basadas en transacciones, son sistemas de tipo operacionales. OLTP permite crear, actualizar y borrar registros como por ejemplo en un sistema de órdenes de compra o de transacciones financieras. Estas tareas se basan generalmente en sistemas de bases de datos relacionales. Algunas de sus características son tiempo de respuesta corto, alta concurrencia y alta disponibilidad (Oracle, 2011).

Por otro lado, *on-line analytical processing* (OLAP) permite el análisis de datos transformándolos en un reflejo de las dimensiones reales de la compañía para la toma de decisiones, son sistemas de tipo informacionales. Estas tareas están basadas generalmente en un *data warehouse*, que es el utilizado por los programas de BI.

Las diferencias entre los sistemas operacionales e informacionales se resumen en la Tabla 2.

Tabla 2. Diferencias entre Sistemas Operacionales e Informacionales

	Operacional	Informacional
Contenido de Datos	Valores actuales	Archivados, derivados, resumidos
Estructura de Datos	Optimizada para transacciones	Optimizada para consultas complejas
Frecuencia de Acceso	Alta	Mediana a baja
Tipo de Acceso	Lectura, Actualización, Borrado	Sólo lectura
Utilización	Predecible, repetitiva	Aleatoria, para situaciones específicas, para descubrimiento
Tiempo de Respuesta	Milisegundos	Segundos a minutos
Usuarios	Gran número	Relativamente pequeño

Fuente: Adaptado de Ponniah, P. (2004). *Data warehousing fundamentals: a comprehensive guide for IT professionals*. John Wiley & Sons.

2.1.2. Data Warehouse

Toda iniciativa de BI requiere de una infraestructura: el *data warehouse* (DW). Según Inmon (2005), considerado el padre del DW, la definición dice que “un *data warehouse* es una colección de datos orientada a un tema, integrada, no volátil y

variante en el tiempo, utilizada para avalar las decisiones de la gerencia. El *data warehouse* contiene datos corporativos granulares” (p. 31).

En contraste con los sistemas que se utilizan en las operaciones diarias de una empresa, el DW almacena datos orientados a un tema. Por ejemplo, un negocio de ventas puede organizar su información para procesar órdenes de compra, facturar a sus clientes, ordenar cuentas a cobrar y cuentas a pagar, todos procesos que soportan sus actividades del día a día. Pero su DW estaría organizada por tema, de acuerdo a las prioridades del negocio, por ejemplo, ventas, productos, clientes, etc.

Los datos son integrados porque provienen de distintas aplicaciones que soportan los procesos del negocio. Para integrar los datos es necesario eliminar las inconsistencias que puedan existir entre las distintas aplicaciones y fijar estándares. Continuando con el ejemplo anterior, para el tema *cliente* se obtendrían datos de los sistemas de órdenes de compra, de facturación, de cuentas a cobrar, etc., relacionadas con ese cliente en particular. La integración implica estandarizar códigos, medidas, convenciones para nombres, etc.

El DW es no volátil porque almacena los datos en masa y no se actualiza con cada transacción que realiza la empresa. Por ejemplo, la aplicación de órdenes de compra debe permitir leer, insertar, editar y borrar cada transacción mientras que el DW sólo permitirá leer los datos de las órdenes de compra para un determinado momento.

La característica de variante en el tiempo se relaciona con la manera de almacenar los datos: se crea una foto que representa un período de tiempo. Por ejemplo, la base de

datos de las cuentas a cobrar mantiene la información actualizada con cada transacción mientras que el DW contiene una secuencia histórica de cada registro de las cuentas a cobrar por cliente para el período de tiempo estipulado en esa foto.

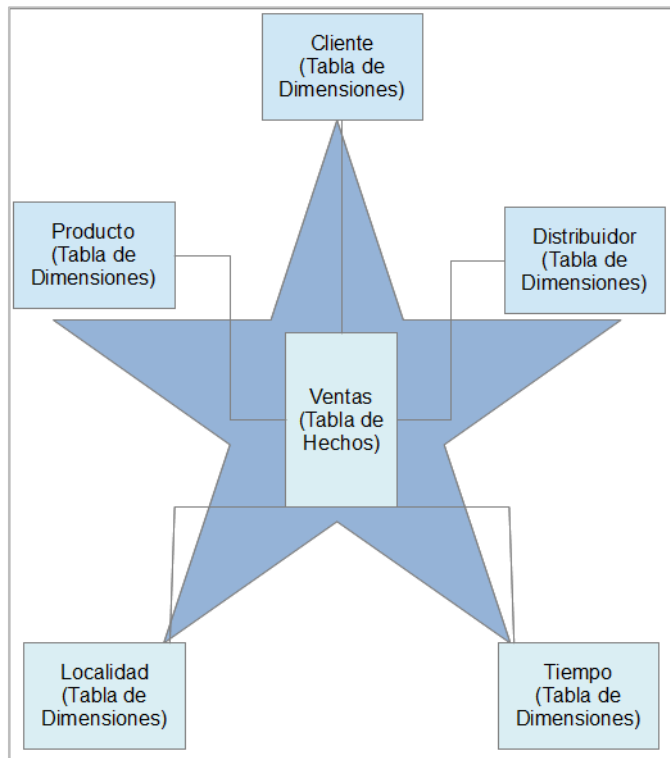
La granularidad se refiere al nivel de detalle: a mayor nivel de detalle, mayor granularidad. Por ejemplo, se podría evaluar el resumen de las compras realizadas por un cliente en los últimos seis meses o cada compra realizada por un cliente en los últimos seis meses, incrementando de esta manera el nivel de detalle.

Como este enfoque alcanza a toda la organización se lo conoce como *enterprise data warehouse*. Existen otros enfoques que se basan en la creación de *data marts*, cuyo alcance se limita a un departamento de la empresa (Inmon, 2005).

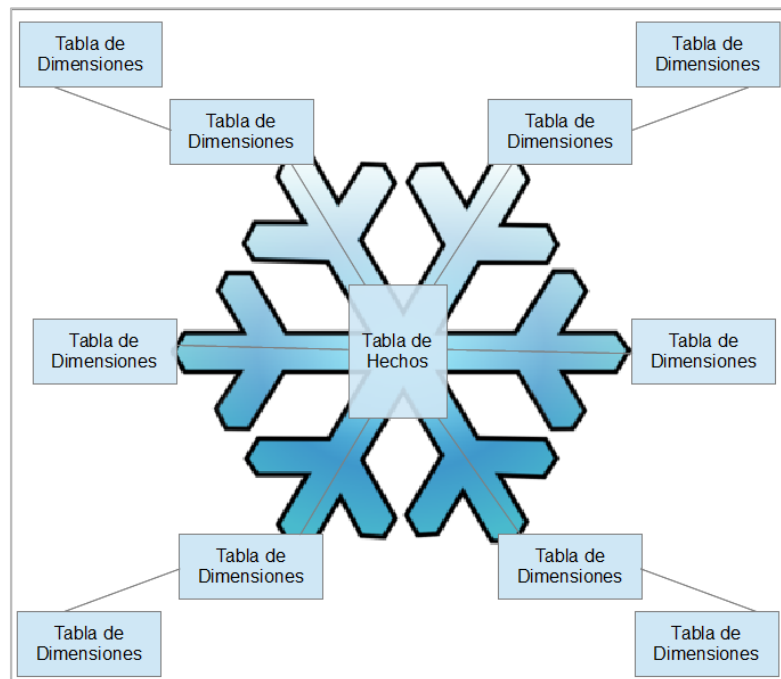
La estructura de datos de un DW difiere de la de una base de datos relacional. Las dos dimensiones de una tabla en el modelo relacional se extienden a un modelo multidimensional que puede ser del tipo estrella o copo de nieve (Figura 1).

La relación entre los datos suele representarse como un cubo en el cual los datos o medidas se categorizan por dimensiones. Los cubos suelen estar resumidos por dimensiones para, de esta manera, mejorar drásticamente el tiempo de respuesta en comparación con una base de datos relacional (Rouse, 2010).

Figura 1. Esquema Estrella vs. Copo de Nieve



Fuente: Adaptado de Pandre, A. (s.f.). OLAP Cubes. Recuperado el 24 de mayo de 2015 del sitio Web de Data Visualization: <https://apandre.wordpress.com/data/datacube/>.



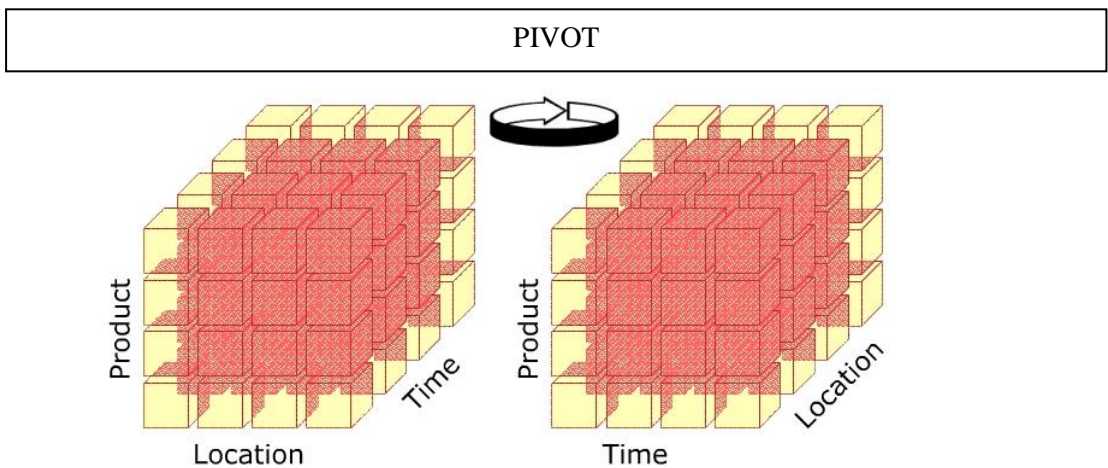
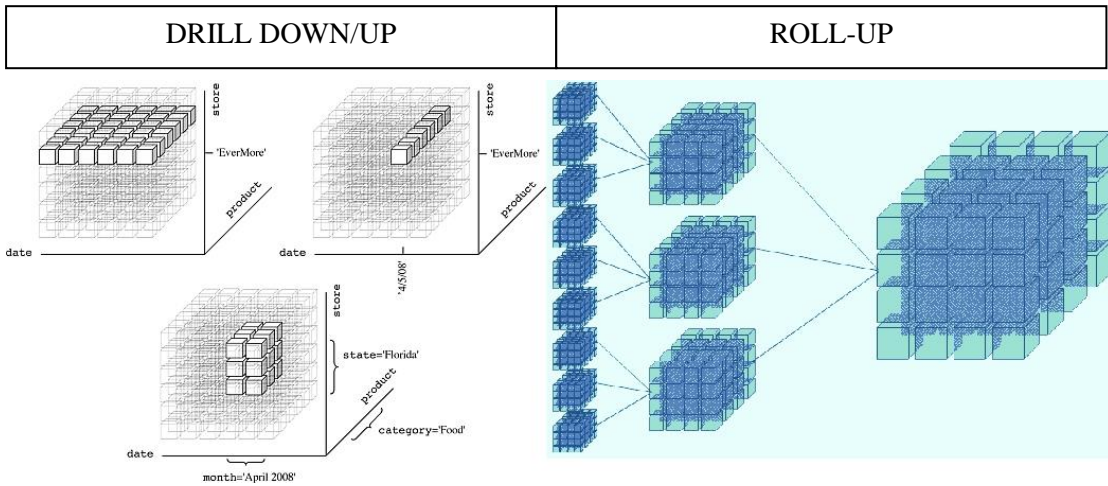
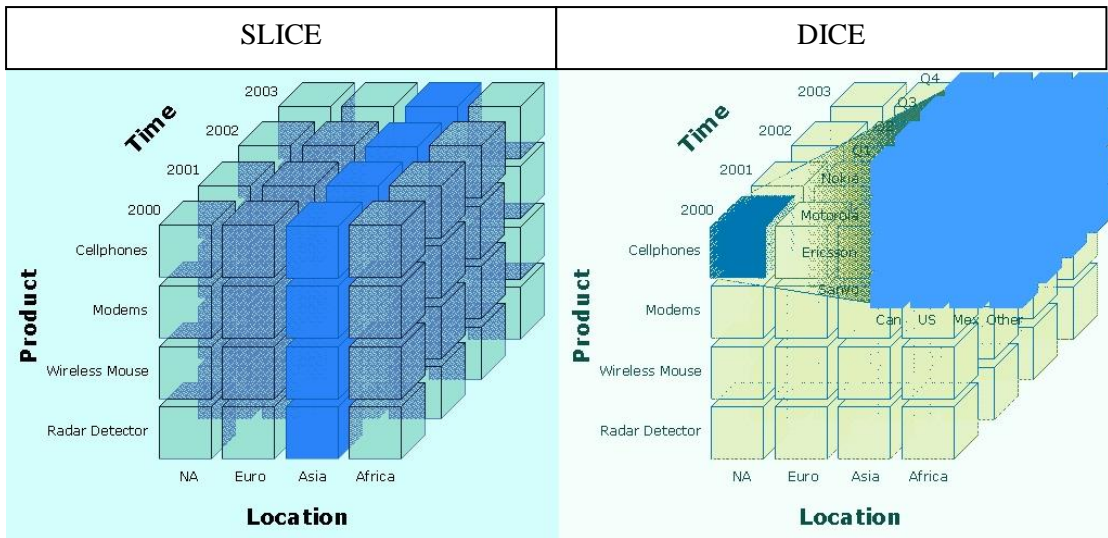
Fuente: Elaboración propia

El cubo y sus vistas se navegan para un subconjunto de datos en particular, cambiando la orientación de los datos y definiendo cálculos analíticos. El proceso que inicia el usuario de navegar páginas interactivamente a través de la especificación de cortes vía rotaciones y de distintos niveles de granularidad se conoce como *slice and dice* (Pandre, s.f.). Las operaciones más comunes incluyen (Figura 2):

- *Slice*: Un *slice* o corte es un rango dentro de un conjunto multidimensional que corresponde a único valor para una de sus dimensiones.
- *Dice*: Un *dice* o corte en dado es la operación que corta más de dos dimensiones consecutivas, dando como resultado un cubo más pequeño que incluye un menor número de valores.
- *Drill Down/Up*: Es la técnica donde el usuario navega los distintos niveles de granularidad de los datos.
- *Roll-up*: Implica consolidar las relaciones de los datos para una o más dimensiones.
- *Pivot*: Es una operación de rotación de los datos que provee distintas orientaciones dimensionales, mostrando presentaciones alternativas de los datos (Pandre, s.f.).

La arquitectura del DW está dividida en tres áreas funcionales: adquisición de datos, almacenamiento y presentación de la información. Los datos se obtienen de

Figura 2. Slice, Dice, Drill Down/Up, Roll-up y Pivot



Fuente: Pandre, A. (s.f.). OLAP Cubes. Recuperado el 24 de mayo de 2015 del sitio Web de Data Visualization: <https://apandre.wordpress.com/data/datacube/>.

distintas fuentes operacionales y son limpiados y transformados antes de ser cargados en el DW. Esta limpieza se realiza en un sector ubicado entre los sistemas operacionales y el DW, y se conoce como área de *staging*. El área de *staging* es necesaria ya que el DW es orientado a un tema y atraviesa varias de las aplicaciones operacionales (Ponniah, 2004).

La adquisición y el almacenamiento de los datos está soportado por tres procesos de soporte que se conocen como ETL, *extract-transform and load* o extraer-transformar y cargar. Estos procesos incluyen la extracción de datos de los distintos sistemas fuentes, su transformación a los formatos y estructuras apropiadas para su almacenamiento en la base de datos del DW y el traslado de los datos al repositorio del DW. La carga de los datos es una función que requiere tiempo, por ello el negocio deberá evaluar los ciclos de recarga según sus necesidades (Ponniah, 2004).

La presentación de la información incluye diferentes métodos que se adaptan a las necesidades de los distintos tipos de usuarios, desde los más novatos hasta los más avanzados. Las consultas *ad hoc*⁴ permiten al usuario definir la información que necesita y componer sus propios *queries*, resultando en reportes online. Existen además reportes con formatos predefinidos donde el usuario ingresa los parámetros necesarios; el reporte puede ser programado para un determinado horario o bien ser ejecutado cuando el usuario lo requiera (Ponniah, 2004).

Por otro lado, existen representaciones gráficas de la información de negocio llamadas visualizaciones, que pretenden mejorar la comprensión del significado de los datos,

⁴ Locución latina, literalmente “para esto”.

exponiendo patrones, tendencias y correlaciones que podrían no ser detectados en un texto. Las visualizaciones son especialmente útiles cuando se manejan grandes cantidades de información (Rouse, 2010).

Otra forma de presentación es mediante *scorecards*. Los *scorecards* son herramientas que muestran el progreso de la compañía a través del tiempo hacia un objetivo determinado. Se basan en KPIs o *key performance indicators*, definidos como factores críticos para el éxito de la organización que permiten estimar el progreso hacia los objetivos de negocio que han sido establecidos por la compañía (Rouse, 2010).

El DW también puede sumar el contexto histórico a los eventos detectados en tiempo real por un sistema BAM. Los sistemas BAM o *Business Activity Monitoring* realizan un monitoreo automatizado de indicadores críticos del negocio en tiempo real, los comparan con reglas preestablecidas y generan alertas cuando los indicadores cumplen con las reglas (Cavalheiroa, Dahanayakea, y Welkeb, 2006). La información puede presentarse en *dashboards*, herramientas que muestran el estado actual de las métricas y los KPIs. Las métricas son medidas que se utilizan para evaluar un componente cuantificable del rendimiento de la compañía, por ejemplo: cantidad de unidades vendidas, monto de ventas, etc. (Rouse, 2010).

2.1.3. Business Intelligence vs. Business Analytics

Los términos BI y *business analytics* (BA) son utilizados indistintamente en algunos casos. En otros, las opiniones con respecto a su definición son dispares.

Davenport y Harris (2007) definen BA como una subcategoría dentro de BI, basada en estadísticas, predicción y optimización. Además consideran a BI más enfocado hacia las capacidades de reporte con respecto a BA.

Hostmann, Rayner y Herschel (2009) definen BI como “la habilidad general de organizar, acceder y analizar información con el fin de aprender y comprender el negocio” y consideran que esta habilidad puede ser aplicada a procesos de negocio, decisiones y distintas áreas. Este BI aplicado es lo que ellos denominan *analytics*.

IBM (s.f.), por otro lado, define BA como capacidades que se extienden más allá de los sistemas tradicionales de BI y que permiten descubrir qué está pasando, determinar el por qué, predecir lo que va a ocurrir y prescribir un plan de acción.

Rouse (2010) reconoce la utilización en general indistinta de los términos BI y BA, pero remarca las diferencias principales que se muestran en la Tabla 3, y destaca que BA se enfoca hacia el análisis estadístico, el *data mining*, los modelos predictivos y el análisis multivariante.

El análisis estadístico en el contexto de BA requiere la recolección y el análisis de todos y cada uno de los datos en el conjunto del cual se puedan obtener muestras. En contraste, en la estadística, una muestra es una selección representativa de la población total. El objetivo del análisis estadístico para BA es identificar tendencias y mediante la utilización de análisis predictivo, probar escenarios que guíen acciones futuras (Rouse, 2010).

Tabla 3. Diferencias Principales entre BI y BA

BI vs. BA	BUSINESS INTELLIGENCE	BUSINESS ANALYTICS
Contesta las siguientes preguntas:	<ul style="list-style-type: none"> • ¿Qué paso? • ¿Cuándo? • ¿Quién? • ¿Cuánto? 	<ul style="list-style-type: none"> • ¿Por qué paso? • ¿Ocurrirá nuevamente? • ¿Qué pasará si cambiamos x? • ¿Qué más nos dicen los datos que nunca antes habíamos pensado preguntar?
Incluye:	<ul style="list-style-type: none"> • Reportes (métricas, KPIs) • Monitoreo Automatizado/ Alertas (umbrales) • <i>Dashboards</i> • <i>Scorecards</i> • OLAP (Cubos, <i>Slice & Dice</i>, <i>Drilling</i>) • Consultas <i>Ad Hoc</i> 	<ul style="list-style-type: none"> • Análisis Estadístico/ Cuantitativo • <i>Data Mining</i> • Modelos Predictivos • Análisis Multivariante

Fuente: Adaptado de Rouse, M. (2010). Search Business Analytics. Recuperado el 6 de marzo de 2014, del sitio Web de TechTarget: <http://searchbusinessanalytics.techtarget.com>

El *data mining* se centra en el descubrimiento automatizado de nuevos hechos y relaciones entre los datos. Las herramientas tradicionales de consulta permiten la búsqueda de información conocida, mientras que las herramientas de *data mining* permiten descubrir información que no es obvia. OLAP responde a preguntas del pasado, y mientras que es posible estimar el futuro basado en el rendimiento pasado, el *data mining* trabaja de otra manera: descubre patrones específicos y nuevas relaciones para predecir el futuro (Ponniiah, 2004).

Los modelos predictivos son patrones estadísticos de comportamiento futuro basados en factores variables que puedan influenciar los resultados, es decir se ocupan de

pronósticos y tendencias. Los modelos predictivos se utilizan entre otras aplicaciones, para mejorar la experiencia de compra de los clientes, para planificar la capacidad de producción y para el manejo del cambio (Rouse, 2010).

El análisis multivariante implica la observación y el análisis de más de una variable al mismo tiempo, permitiendo el estudio de los datos a través de las distintas dimensiones, considerando los efectos de todas las variables con respecto a un hecho (Rouse, 2010).

Con respecto a las herramientas y capacidades que incluye BA, IBM (s.f.) propone cinco categorías: descriptivas, diagnósticas, predictivas, prescriptivas y cognitivas.

Las herramientas descriptivas permiten explicar qué pasó y lo reflejan en *dashboards* y reportes. Las herramientas diagnósticas buscan distinguir el por qué y se representan en visualizaciones. Las predictivas, en cambio, se basan en análisis estadístico, *data mining* y modelos predictivos. Las herramientas prescriptivas indican la acción a seguir. Finalmente, las herramientas cognitivas integran todas las herramientas anteriores y utilizan correlaciones e hipótesis además de recordar y aprender de cada evento.

Por otro lado, SAS Institute Inc. (2008) sugiere ocho niveles de capacidades analíticas: reportes standard, reportes *ad hoc*, consultas *drilldown/ OLAP*, alertas, análisis estadístico, pronósticos, modelos predictivos y optimización. Los cuatro primeros niveles se basan en el reporte de datos históricos mientras que los últimos

cuatro se orientan hacia una visión predictiva y otorgan respuestas a preguntas más complejas.

En la Tabla 4 se presenta un cuadro comparativo entre las capacidades de BA propuestas por IBM (s.f.) y SAS Institute Inc. (2008).

Tabla 4. Comparación de Capacidades de BA según IBM y SAS Institute Inc.

CAPACIDADES SEGÚN IBM	PREGUNTAS QUE RESPONDEN	CAPACIDADES SEGÚN SAS	PREGUNTAS QUE RESPONDEN
Descriptivas	¿Qué pasó?	Reportes Standard	¿Qué pasó? ¿Cuándo pasó?
		Reportes <i>Ad Hoc</i>	¿Cuánto? ¿Con qué frecuencia? ¿Dónde?
Diagnósticas	¿Por qué?	Consultas <i>Drilldown/OLAP</i>	¿Dónde está el problema? ¿Cómo encuentro las respuestas?
Prescriptivas	¿Qué acción se debe tomar?	Alertas	¿Cuándo reaccionar? ¿Qué acciones son necesarias ahora?
Predictivas	¿Qué va a pasar?	Análisis Estadístico	¿Por qué pasa esto? ¿Qué oportunidades estoy perdiendo?
		Pronósticos	¿Qué pasaría si estas tendencias continúan? ¿Cuánto se necesita? ¿Cuándo se necesitará?
		Modelos Predictivos	¿Cómo sigue? ¿Cómo se verá afectado el negocio?
Cognitivas	¿Qué va a pasar y cómo actuar basado en lecciones aprendidas?	Optimización	¿Cómo hacer las cosas mejor? ¿Cuál es la mejor decisión frente a un problema complejo?

Fuente: Elaboración propia

2.2. Big Data y Big Data Analytics

La dificultad para manejar el crecimiento exponencial y los distintos tipos de datos que presentan las bases de datos relacionales ha promovido el desarrollo de un nuevo concepto que ha generado un interés creciente para investigadores y organizaciones: *big data*.

Según Barranco Fragoso (2012), *big data* es el concepto aplicado al análisis y procesamiento de enormes cantidades de datos que no podrían ser procesados o analizados utilizando herramientas existentes, ya que tomaría demasiado tiempo y sería muy costoso hacerlo utilizando bases de datos tradicionales, como por ejemplo las bases de datos relacionales.

Russom (2011) manifiesta que la mayoría de las definiciones de *big data* se basan en el volumen de los datos almacenados y destaca que si bien el volumen es un factor importante, existen otros dos atributos a tener en cuenta: la velocidad y la variedad de los datos.

Con respecto al volumen, la mayoría de las definiciones hablan de terabytes o petabytes⁵. Sin embargo, también puede ser cuantificado mediante cantidad de transacciones, registros, tablas o archivos. Otras organizaciones cuantifican *big data* en términos de tiempo, por ejemplo 7 años de datos (Russom, 2011).

⁵ Un petabyte equivale a 10^{15} bytes, 1.000 terabytes o 1.000.000 de gigabytes.

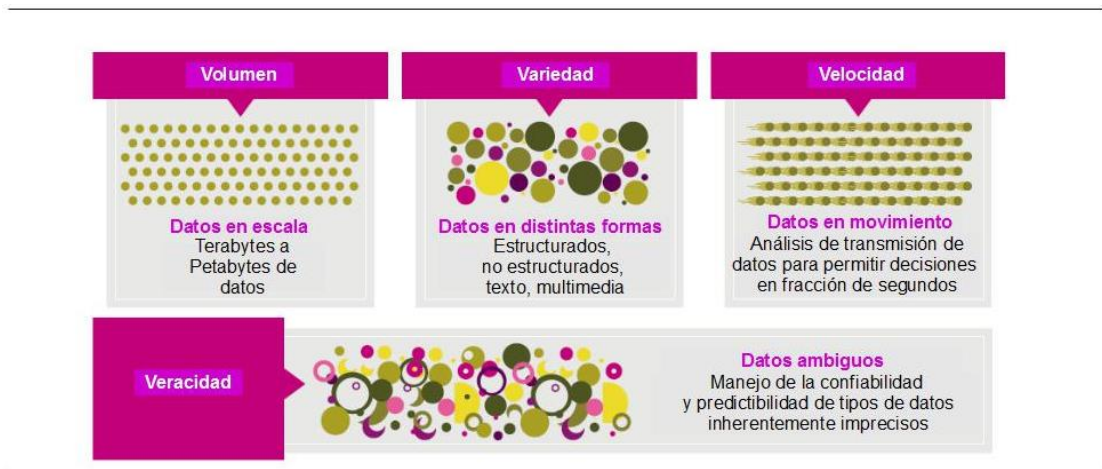
La velocidad se refiere a la frecuencia con la cual se crean, se procesan y se obtienen los datos. Hoy en día ciertos tipos de datos deben ser analizados en tiempo real para agregar valor al negocio, como por ejemplo para la detección de fraude (Barranco Fragoso, 2012).

La variedad se relaciona con la diversidad de fuentes y tipos de datos. Las fuentes incluyen páginas web, datos de RFID de aplicaciones logísticas y datos obtenidos de aplicaciones de *call-center*, entre otras. Los tipos de datos se clasifican en estructurados -como por ejemplo los obtenidos de las bases de datos relacionales-, no estructurados -como texto y lenguaje humano-, y semiestructurados -como por ejemplo XLM y RSS- (Russom, 2011).

Schroeck, Shockley, Smart, Romero-Morales, y Tufano (2012) agregan una cuarta dimensión: la veracidad:

La veracidad se refiere al nivel de confiabilidad asociado a ciertos tipos de datos. El esfuerzo por lograr datos de alta calidad es un requerimiento importante de *big data* y un desafío, pero ni aun los mejores métodos de limpieza de datos pueden remover la falta de predictibilidad inherente de algunos tipos de datos, como el estado del tiempo, la economía o la decisión de compra futura de un cliente existente. La necesidad de reconocer y planificar sobre la veracidad es una dimensión de *big data* que ha sido incorporada a medida que los ejecutivos buscan una mejor comprensión del mundo incierto que los rodea (Schroek et al., 2012, p. 5).

Figura 3. Las 4 Vs de Big Data



Fuente: Adaptado de Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). Analytics: The real-world use of Big Data. *IBM Global Business Services*.

Pero solamente la capacidad de almacenar *big data* no agrega valor al negocio. Una vez resueltas las dificultades de escalabilidad, almacenamiento y velocidad de procesamiento es necesario transformar los datos en información útil que permita tomar decisiones. Para ello se emplean herramientas de BA.

Russom (2011) define *big data analytics* como “la aplicación de técnicas analíticas avanzadas para verificar grandes conjuntos de datos” (p. 8), y la clasifica como una de las tendencias más importantes dentro del BI de hoy. Además:

- Resalta la unión de dos entidades: *big data* y *analytics*.
- Destaca grandes volúmenes de datos con un alto nivel de detalle.
- Describe las herramientas analíticas con el término *discovery analytics*, teniendo en cuenta su capacidad para descubrir nuevas realidades de negocio.

- Incluye análisis predictivo, *data mining*, estadística e inteligencia artificial entre otros ejemplos de herramientas analíticas.

2.3. Marcos de Trabajo para la Implementación de BI

Si bien las organizaciones tienen cada vez mayor cantidad de datos disponibles, la capacidad de convertir los datos en información y obtener valor de ella sigue siendo un desafío. Davenport et al. destacan:

La construcción de capacidades analíticas profundas requiere mucho más que el solo hecho de recolectar y almacenar datos en grandes cantidades. Existen muchas piezas que deben ser puestas en su lugar tales como software, tecnología, datos, procesos, métricas, incentivos, capacidades, cultura y *sponsorship* (Davenport et al., 2007).

Yeoh y Koronios (2010) consideran que la implementación de BI no es un proyecto de TI convencional, es decir no es una actividad que sólo implica la compra de una combinación de hardware y software, sino que es un emprendimiento complejo que requiere infraestructura y recursos apropiados durante un período de tiempo prolongado. Además, reconocen la existencia de una gran cantidad de líneas de orientación para la implementación de BI creadas por la industria de TI, pero en su mayoría basadas en reportes anecdóticos.

Chandler, Hostmann, Rayner y Friedman (2011) remarcan que no existe un único proveedor que pueda cumplir con todas las necesidades tecnológicas, las aplicaciones y los servicios requeridos para una implementación de BI exitosa. El gran portfollio de soluciones disponibles con capacidad, escala y alcance funcional divergentes incrementa la necesidad de adecuar el foco de TI para lograr óptimos beneficios de sus inversiones.

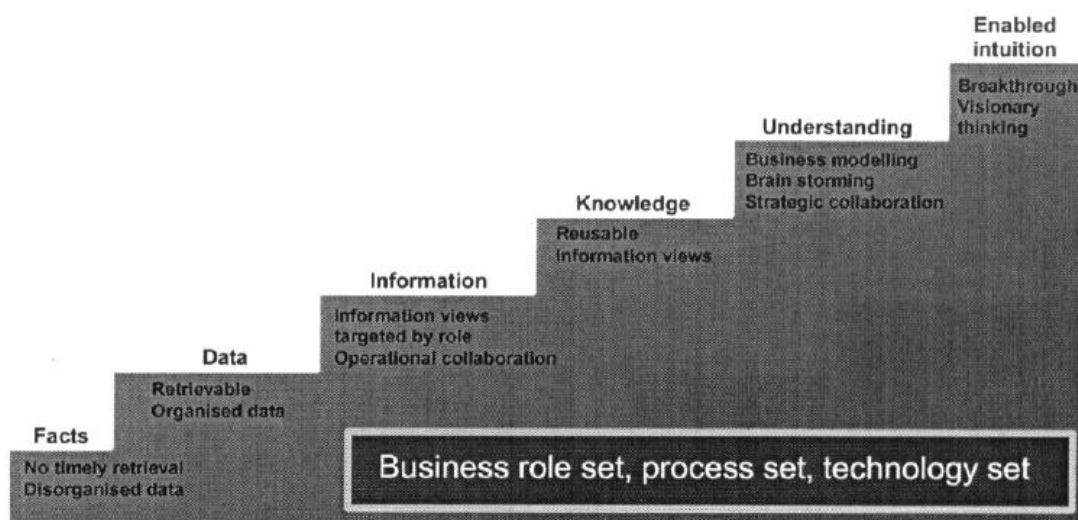
Frente a la complejidad que el proyecto de implementación de BI implica, la disponibilidad de guías de implementación dirigidas por la industria que no se basan en investigación formal y la gran cantidad de soluciones divergentes existentes en el mercado, se genera la necesidad de creación de un marco de trabajo que ordene e integre los componentes necesarios para expresar una visión estratégica y un plan de implementación para las iniciativas de BI.

A continuación se describen cuatro marcos de trabajo para la implementación de BI, los cuales exploran los pasos a seguir para una implementación exitosa, los componentes, los factores críticos de éxito, y las dimensiones a tener en cuenta frente a una iniciativa de BI, tales como el personal necesario, los procesos y las plataformas. Para el caso de aquellos marcos donde se encontraron distintas versiones, se estudia su evolución. El orden de presentación es cronológico, según el año de su última actualización.

2.3.1. Marco de Trabajo LOBI

El LOBI (Ladder of Business Intelligence) es un marco de trabajo creado por Cates, Gill y Zeituny, que fue desarrollado en un período de 30 años de trabajo, y presentado en 2005. Su investigación muestra nueve casos de implementación de su marco y los beneficios obtenidos: alineamiento de TI con el negocio, definición de inteligencia a nivel operativo, capacidad de medición del valor para el negocio, contribución a la efectividad, y mapeo de los niveles de inteligencia. El marco LOBI define una metodología que facilita la creación de un plan de TI y el diseño de una arquitectura de TI para el negocio. La metodología incluye diez pasos para una implementación exitosa. Además, incluye seis niveles de madurez que representan niveles crecientes de inteligencia: a mayor nivel, mejor comprensión del entorno del negocio. Los niveles LOBI se muestran en la Figura 4.

Figura 4. Niveles LOBI



Fuente: Cates, J. E., Gill, S. S., & Zeituny, N. (2005). The Ladder of Business Intelligence (LOBI): A framework for enterprise IT planning and architecture. *International Journal of Business Information Systems*, 1(1), p. 227.

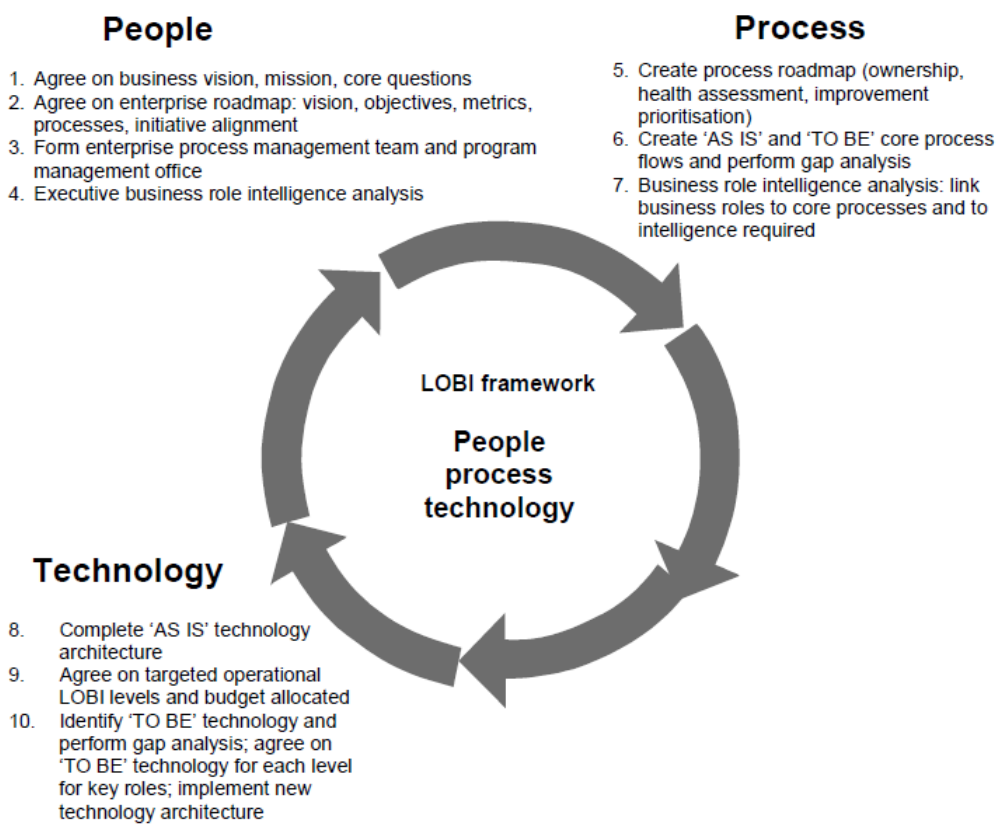
Los pasos que Cates et al. (2005) recomiendan para una implementación exitosa se dividen en tres grupos. Los cuatro primeros pasos se relacionan con el personal, los siguientes tres pasos con los procesos y los últimos tres con la tecnología.

Los diez pasos se muestran en la Figura 5 y se detallan a continuación:

- *Acordar la visión y misión del negocio y abordar cuestiones esenciales:* La planificación estratégica por parte del equipo ejecutivo es el primer paso. Además de determinar la visión y misión del negocio es importante discutir cuestiones esenciales tales como cuáles son las fortalezas, oportunidades y desafíos que enfrenta el negocio, si existe una metodología de éxito, si la entrega de productos y servicios es consistente, si es posible sostener el crecimiento, si es posible ser efectivo e innovador al mismo tiempo, etc.
- *Acordar el rumbo de la empresa:* La planificación se articula en un plan de negocio para la empresa y uno para cada unidad de negocio. Estos planes incluyen objetivos alineados con la visión de la empresa, métricas para evaluar los objetivos, procesos necesarios para la entrega de productos y servicios, y la identificación de iniciativas relacionadas con los procesos principales que permitan la alineación de TI con el negocio.
- *Formar un equipo corporativo de gestión de procesos y una oficina de gestión de programas:* Para la implementación del marco LOBI se requiere una estructura para el gobierno de TI. La estructura está compuesta por un equipo corporativo de gestión de procesos y una oficina de gestión de programas. El equipo de gestión de procesos es responsable de la mejora de los procesos para

lograr operaciones efectivas y más eficientes. La oficina de gestión de programas se encarga de evaluar, priorizar, planificar, guiar y reportar las iniciativas que conducen los procesos principales de negocio.

Figura 5. Diez Pasos para una Implementación LOBI Exitosa



Fuente: Cates, J. E., Gill, S. S., & Zeituny, N. (2005). The Ladder of Business Intelligence (LOBI): A framework for enterprise IT planning and architecture. *International Journal of Business Information Systems*, 1(1), p. 230.

- *Analizar los roles ejecutivos del negocio:* Durante esta etapa se identifican los roles de los ejecutivos de negocio y se responden las preguntas claves para

llevar adelante el rol de manera exitosa; se determinan la inteligencia, las fuentes, la frecuencia, el formato y la forma en que se transmitirán los datos; se reconocen los procesos claves que crean entradas y salidas para este rol de negocio y aquellos otros roles que intervienen en esos procesos; se reconoce el nivel de LOBI existente y se define el nivel a lograr.

- *Crear una guía de procesos:* La guía de procesos permite identificar las actividades claves, asignar responsabilidades, definir los criterios para la evaluación y priorización de mejoras necesarias para los procesos.
- *Crear un flujo de procesos claves existentes y a lograr, desarrollar un análisis de las necesidades:* En esta etapa se realiza un mapeo de los procesos claves existentes y aquellos a lograr, indicando la brecha a cubrir para alcanzar el estado deseado. Es fundamental la colaboración interdisciplinaria para asignar responsabilidades en los procesos y luego manejar y medir los resultados.
- *Analizar el resto de los roles de negocio:* Una vez que se identifican los procesos claves, se relacionan los roles de negocio con los procesos principales y la inteligencia requerida.
- *Determinar la arquitectura de la tecnología existente:* En este paso se documenta la arquitectura que soporta los procesos claves existentes como por ejemplo servidores web, portales, seguridad, aplicaciones de negocio, BI, correo electrónico, acceso a datos, redes, plataformas, etc.
- *Acordar el nivel operativo de LOBI a ser alcanzado y asignar un presupuesto:* La gerencia revisa las distintas iniciativas e identifica el nivel LOBI que desea

alcanzar. Además, una vez acordado el nivel LOBI, es necesario el soporte de un presupuesto asignado. En esta etapa se utiliza el *balance scorecard* para plantear objetivos operativos y financieros.

- *Identificar la tecnología a implementar y desarrollar un análisis de las necesidades; acordar la tecnología a implementar para cada nivel para los roles claves; implementar la nueva arquitectura:* TI propone la tecnología necesaria para cubrir los objetivos de negocio y alcanzar el nivel LOBI acordado. Finalmente se crea un plan de implementación que incluye el presupuesto detallado, el plan de adquisiciones y los detalles de la implementación para alcanzar el estado deseado (Cates et al., 2005).

2.3.2. Marco de Trabajo sobre los Componentes de BI de TDWI

The Data Warehousing Institute (como se cita en Moss, 2013) explica claramente los componentes de BI en su marco de trabajo creado en 2006, jerarquizando los conceptos para permitir estructurar un proyecto de implementación.

El marco de trabajo está compuesto por tres capas: la capa de negocio, la capa de administración y operaciones, y la capa de implementación. Según detalla Moss (2013), la capa de negocio enmarca el proyecto de BI involucrando al negocio como dueño de la iniciativa e incluye cuatro componentes:

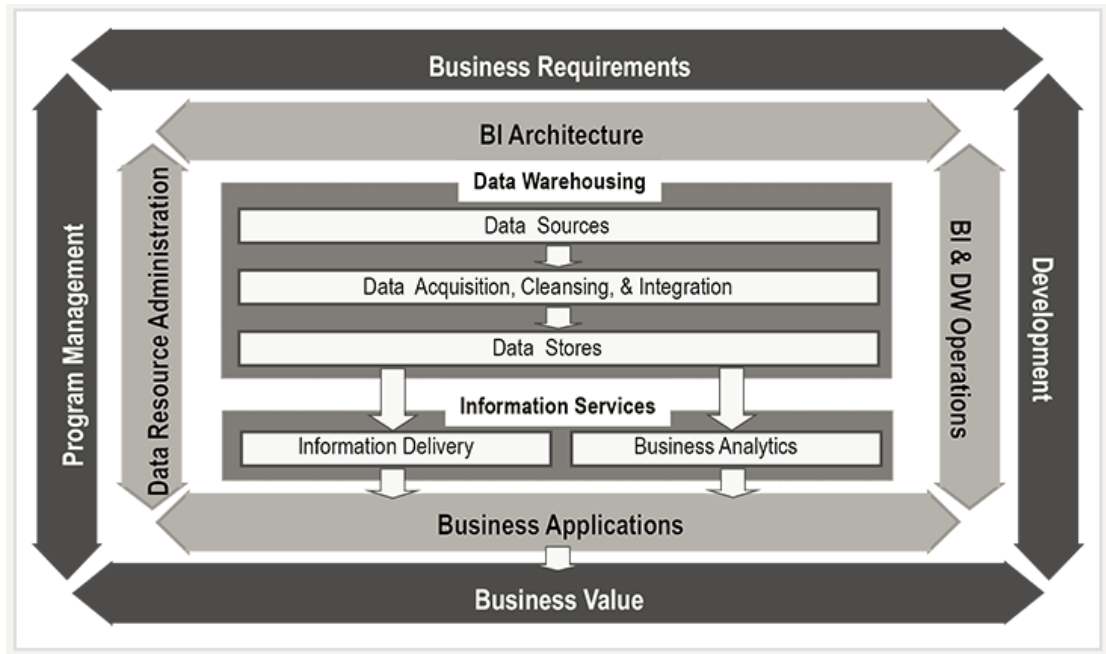
- *Requerimientos del negocio:* El objetivo principal del proyecto de BI es el de proveer tanto la información como las capacidades analíticas necesarias para

el negocio. La capacidad de priorizar los requerimientos se relaciona con el segundo componente, que es el valor para el negocio.

- *Valor para el negocio:* Este componente se define como el beneficio que se anticipa de la utilización de la iniciativa de BI, como por ejemplo incremento en los resultados, mayor margen de ganancia, riesgos mitigados o evitados, reducción de costos, etc.
- *Gestión de Programas:* La gestión de programas requiere actividad continua en el manejo de los componentes de BI para lograr optimizar el valor para el negocio. Este componente incluye el manejo de la infraestructura, el trabajo interdisciplinario, la arquitectura empresarial y los estándares, de manera de sincronizar múltiples proyectos, reajustándolos según las necesidades del negocio que suelen cambiar frecuentemente.
- *Desarrollo:* El desarrollo se refiere a las actividades del proyecto para crear aplicaciones y bases de datos para BI. Se relaciona con la utilización de metodologías comunes, descomposición de los servicios en entregas múltiples, obtención de los recursos para el proyecto, identificación de las métricas para el éxito del proyecto, etc. (Moss, 2013).

Con respecto a las responsabilidades del negocio y las responsabilidades de TI, Moss (2013) destaca que tanto los requerimientos como su priorización son dominio del negocio, mientras que la gestión de programas y el desarrollo son dominios comunes entre TI y el negocio.

Figura 6. Marco de Trabajo sobre los Componentes de BI de TDWI



Fuente: Moss, L. (2013). *Extreme scoping*. Bradley Beach, NJ: Technics Publications.

La capa de administración y operaciones está compuesta por cuatro componentes que conectan los aspectos técnicos de BI con el negocio. Estos componentes son: la arquitectura de BI, las aplicaciones de negocio, la administración de los recursos de datos, y las operaciones de BI y *data warehousing* (Moss, 2013).

- *Arquitectura de BI:* Este componente incluye marcos de trabajo, modelos, estándares, políticas, procedimientos y convenciones que describen los componentes de BI y sus relaciones.
- *Aplicaciones de negocio:* Se refiere a la forma de presentación de la información. Son procesos de negocio que recuperan datos del entorno de BI a

través de reportes, consultas *ad hoc*, aplicaciones analíticas, *scorecards* y *dashboards*. La información obtenida es utilizada para la toma de decisiones en la empresa.

- *Administración de los recursos de datos*: Este componente se encarga de la creación y gestión de políticas, procedimientos y procesos para el gobierno de los datos, lo cual incluye modelado de datos, estructura de metadatos, colaboración con los dueños y los custodios de los datos, mapeo, validación y limpieza de los datos, etc.
- *Operaciones de BI y data warehousing*: Se refiere a la extracción, limpieza, transformación y carga de los datos, así como también a la ejecución y el monitoreo de las aplicaciones de BI. Una operación eficiente asegura el mantenimiento de la calidad, disponibilidad y seguridad del entorno de BI (Moss, 2013).

Tanto el negocio como TI colaboran en la arquitectura de BI, mientras que TI es responsable de las operaciones de BI y el negocio de utilizar las aplicaciones (Moss, 2013).

La tercera capa, llamada capa de implementación, está integrada por los componentes técnicos necesarios para extraer, limpiar, transformar, cargar y entregar los datos al negocio (Moss, 2013). Las funciones principales de esta capa son el almacenamiento, el reporte y el análisis de los datos. Los componentes de la capa de implementación son:

- *Data warehousing*: Es el conjunto de programas, procesos y procedimientos que se utilizan para extraer, integrar y almacenar los datos.
- *Servicios de información*: Son las aplicaciones, los procesos y los procedimientos que se utilizan para convertir datos en información (Moss, 2013).

TI carga la mayor responsabilidad de la capa técnica ya que desarrolla las soluciones. El negocio, por su parte, participa en el análisis, diseño, prototipado y testeado del software (Moss, 2013).

2.3.3. Marco de Trabajo para BI basado en Factores Críticos de Éxito

Yeoh, Koronios y Gao (2008) desarrollaron el *Marco de Trabajo para la Implementación de Sistemas de Business Intelligence basado en Factores Críticos de Éxito*. Para la construcción de su marco, utilizaron el método Delphi⁶ y condujeron tres rondas de estudios con 15 expertos en sistemas de BI. En su estudio encontraron siete factores críticos de éxito y elementos contextuales asociados que influyen en una implementación de BI exitosa. Los factores que identificaron en su investigación se resumen a continuación:

⁶ Método Delphi: Procedimiento que busca lograr el consenso entre expertos sobre determinados temas, utilizando una serie de cuestionarios repetitivos y obteniendo conclusiones mediante la exploración estadística de los datos.

- *Compromiso de la gerencia y sponsorship:* El respaldo de los ejecutivos de negocio facilita la obtención de los recursos operativos necesarios como por ejemplo el presupuesto, los recursos humanos y otros requerimientos, especialmente teniendo en cuenta que la implementación de BI es un proyecto complejo y prolongado.
- *Gestión del cambio orientado al usuario de negocio:* Cuanto mayor es la participación de los usuarios, mayor las posibilidades de comunicar sus necesidades, lo cual conlleva en última instancia a una implementación exitosa. El involucramiento de los usuarios es particularmente importante cuando los requerimientos del sistema no están claros. El entrenamiento es otro factor importante para que los usuarios comprendan tanto la tecnología como las reglas de negocio para las aplicaciones de BI.
- *Visión clara y caso de negocio bien establecido:* Para dirigir la implementación es necesaria una visión estratégica del negocio. Una visión a largo plazo, principalmente estratégica e institucional, es indispensable para establecer un caso de negocio de BI. El caso de negocio debe estar alineado con la visión corporativa porque eventualmente impacta en la adopción y en el resultado del sistema de BI.
- *Metodología y gestión de proyectos impulsada por el negocio:* Definir el alcance adecuado del proyecto permite definir parámetros claros y ayuda al equipo del proyecto a enfocarse en los hechos importantes, descartando aquellos innecesarios. Comenzar con pequeños pasos y adoptar una entrega

incremental aseguran un impacto inmediato y un aumento en la visibilidad, lo cual incrementa el apoyo de la gerencia.

- *Champion centrado en el negocio y composición balanceada del equipo del proyecto:* Es importante contar con un *champion* que tenga profundo conocimiento del negocio, que pueda prever los desafíos que enfrenta la organización y cambiar el rumbo según sea necesario; alguien que pueda ver el sistema desde una perspectiva estratégica e institucional más que técnica.

También la composición y las capacidades del equipo influyen en el éxito de la implementación. El equipo del proyecto debe ser interdisciplinario y sus integrantes deben poseer una combinación de conocimientos técnicos con una sólida base del negocio. Los consultores externos pueden completar el equipo, con una visión objetiva del problema.

- *Marco de trabajo técnico estratégico y escalable:* Otro factor crucial en la implementación de BI es la estabilidad de los sistemas de fuente de datos. La actualización de datos en los sistemas de gestión debe funcionar correctamente para que los procesos de extracción, transformación y carga sean posibles en el sistema de BI. Es fundamental un sistema de gestión confiable, consistente y estable antes de embarcarse en una iniciativa de BI.

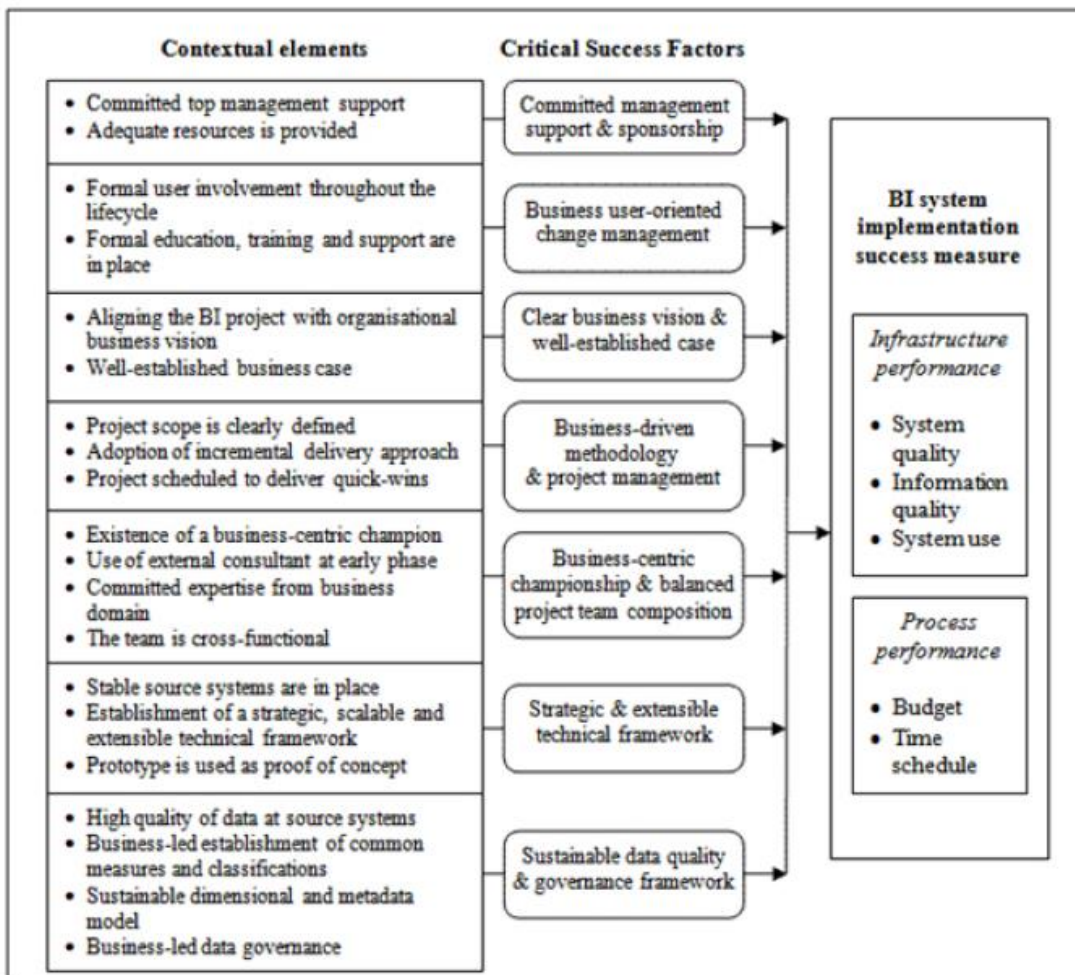
Por otro lado, el marco técnico del sistema de BI debe ser escalable. De esta manera, debe permitir incorporar nuevas fuentes de datos, atributos y modelos dimensionales para que la solución de BI cubra las necesidades del negocio a largo plazo. Los prototipos pueden ser de gran ayuda para persuadir a los

usuarios y a los *sponsors* sobre los beneficios de la implementación del sistema de BI.

- *Calidad de datos sustentable y marco de gobierno*: El propósito del sistema de BI es integrar silos de datos para llevar adelante un análisis avanzado y mejorar la toma de decisiones. Por ello, la calidad de los datos en la fuente es crucial para la implementación exitosa del sistema de BI. El gobierno de los datos es fundamental para asegurar la calidad, ya que datos poco confiables tendrían un efecto multiplicador en el sistema de BI alterando por consiguiente el resultado en la toma de decisiones. El compromiso del negocio para establecer consenso con respecto a medidas y clasificaciones es importante para mantener la consistencia a lo largo y a lo ancho de toda la organización. Además, los modelos dimensionales y de metadatos deben ser sustentables, con la suficiente flexibilidad para permitir escalabilidad (Yeoh et al., 2008).

Yeoh et al. (2008) consideran que estos factores críticos influyen el éxito de la implementación tomando en cuenta dos medidas claves: el rendimiento de la infraestructura y el rendimiento del proceso. El rendimiento de la infraestructura se relaciona con *la calidad del sistema* -facilidad de uso, confiabilidad, funcionalidad, flexibilidad, integración, tiempo de respuesta-, *la calidad de la información* -precisión, relevancia, consistencia, oportunidad, utilidad, nivel de detalle- y *la utilización del sistema*, mientras que el rendimiento del proceso se puede evaluar en términos de *tiempo y presupuesto*.

Figura 7. Marco de Trabajo para BI basado en Factores Críticos de Éxito



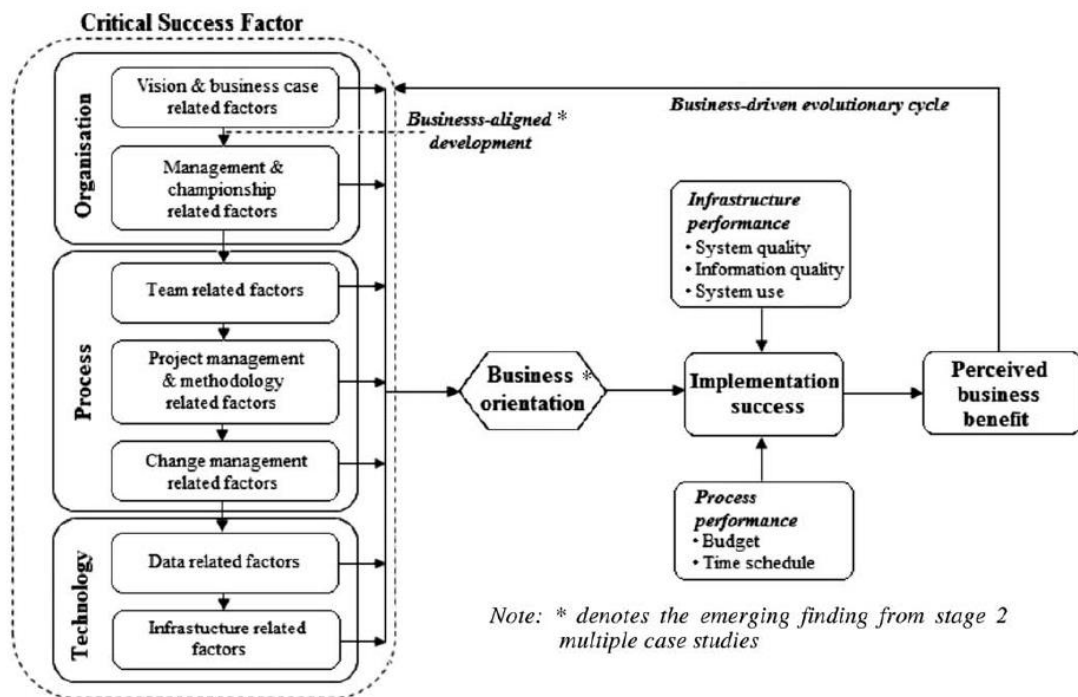
Fuente: Yeoh, W., Koronios, A., & Gao, J. (2008). Managing the implementation of business intelligence systems: a critical success factors framework. *International Journal of Enterprise Information Systems (IJEIS)*, 4(3), p. 86.

En una segunda etapa, el marco de trabajo fue verificado contra cinco casos de estudio basados en grandes empresas y en 2010, Yeoh et al. presentaron una nueva versión del marco.

Las principales diferencias encontradas en el marco de 2010 con respecto al de 2008 son las siguientes:

- Los elementos contextuales y los factores críticos de éxito del marco de 2008, fueron revisados y como resultado fueron condensados en siete nuevos factores que se relacionan con las siguientes áreas: visión y caso de negocio, gestión y *championship*, equipo de trabajo, gestión de proyectos y metodología, gestión del cambio, datos e infraestructura.

Figura 8. Factores Críticos de Éxito para la Implementación de BI



Fuente: Yeoh, W., & Koronios, A. (2010). Critical success factors for business intelligence systems. *Journal of computer information systems*, 50(3), p. 25.

- Los siete nuevos factores críticos de éxito fueron agrupados en tres capas: organización, procesos y tecnología.

- El enfoque de orientación al negocio fue incorporado para resaltar la necesidad de encontrar un propósito específico de negocio para la iniciativa de BI, ya que sin este propósito no se lograría el impacto buscado.
- La percepción de los beneficios para el negocio fue creada como parte de un ciclo interactivo que evoluciona en el tiempo, mediante la evaluación continua de la información y las sugerencias del usuario, y que refuerza la idea del desarrollo alineado con el negocio.

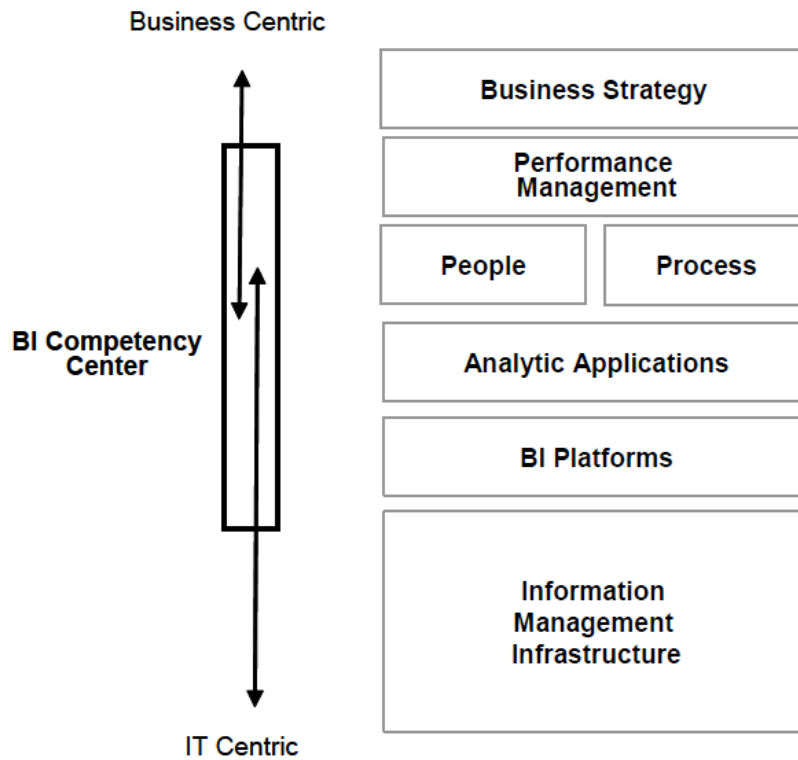
2.3.4. Marco de Trabajo de Gartner

En 2006, Gartner publicó su marco de trabajo para BI y gestión del rendimiento, creado por Hostmann, Rayner y Friedman. El marco está dividido en seis capas:

- *Capa de estrategia de negocio:* Los programas de BI y gestión del rendimiento deben ayudar al negocio a alcanzar sus objetivos estratégicos y operativos. En esta capa se definen la estrategia, los objetivos, y las métricas y se establece el presupuesto para la iniciativa.
- *Capa de gestión del rendimiento:* La gestión del rendimiento se encarga de conectar las actividades operativas con la estrategia de negocio, más específicamente encuentra la relación entre métricas asociadas a lo largo de todo el negocio para acrecentar el proceso de creación de valor.

- *Capa de personal y procesos:* En esta capa se determinan los tipos de usuarios y la información necesaria para llevar a cabo los procesos. También se estudia su relación con otros factores de negocio que pueden afectar el cambio.
- *Capa de aplicaciones analíticas:* Las aplicaciones analíticas presentan un amplio rango de capacidades que deben ser definidas durante esta etapa. Para ello, se priorizan los usuarios y procesos que presentan mayores beneficios para el negocio y se definen la relación y el nivel de integración para alcanzar la consistencia, relevancia y precisión necesarias para un análisis interdisciplinario.
- *Capa de plataforma de BI:* Durante esta etapa se prioriza y se define la combinación de capacidades de BI necesarias para los distintos usuarios, así como también se busca minimizar la redundancia.
- *Capa de infraestructura de gestión de la información:* Más del 70% del tiempo y el esfuerzo en una implementación de BI tiene que ver con la infraestructura de la gestión de la información. La arquitectura e integración de datos aseguran eficiencia y agilidad para reaccionar frente a los requerimientos cambiantes del negocio. Además, en esta etapa se busca asegurar la consistencia de datos utilizados en otras capas. El objetivo es definir una infraestructura sólida pero flexible, para lograr soportar el cambio de manera exitosa (Hostmann et al., 2006).

Figura 9. Marco de Trabajo para BI y Gestión del Rendimiento de Gartner



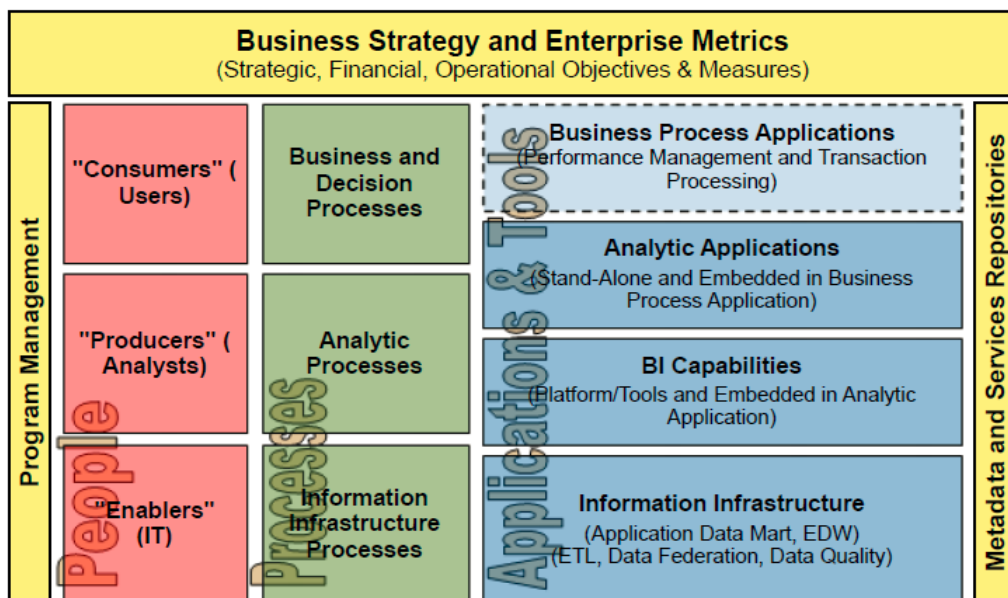
Fuente: Hostmann, B., Rayner, N., & Friedman, T. (2006). Gartner's business intelligence and performance management framework. *Gartner Inc.*, p.3.

En 2009, Hostmann et al. revisaron el marco de Gartner y publicaron una nueva versión titulada *Marco de Trabajo para BI, Analytics y Gestión del Rendimiento*. Las principales diferencias observadas en la versión de 2009 incluyen:

- La inclusión explícita del término *analytics* para definir el alineamiento y la integración del personal, los procesos y herramientas necesarias para las actividades de BI, *analytics* y gestión del rendimiento, que presentan varias áreas de superposición.

- La división de la capa de personal y procesos para lograr mayor foco en ambas áreas.

Figura 10. Marco para BI, Analytics y Gestión del Rendimiento de Gartner



BI = business intelligence

EDW = enterprise data warehouse

ETL = extraction, transformation and loading

Fuente: IBM (2010). A practical framework for business intelligence and planning in midsize companies IBM. Recuperado de <http://www.ibm.com/midmarket/uk/en/att/pdf/practicalframework.pdf>.

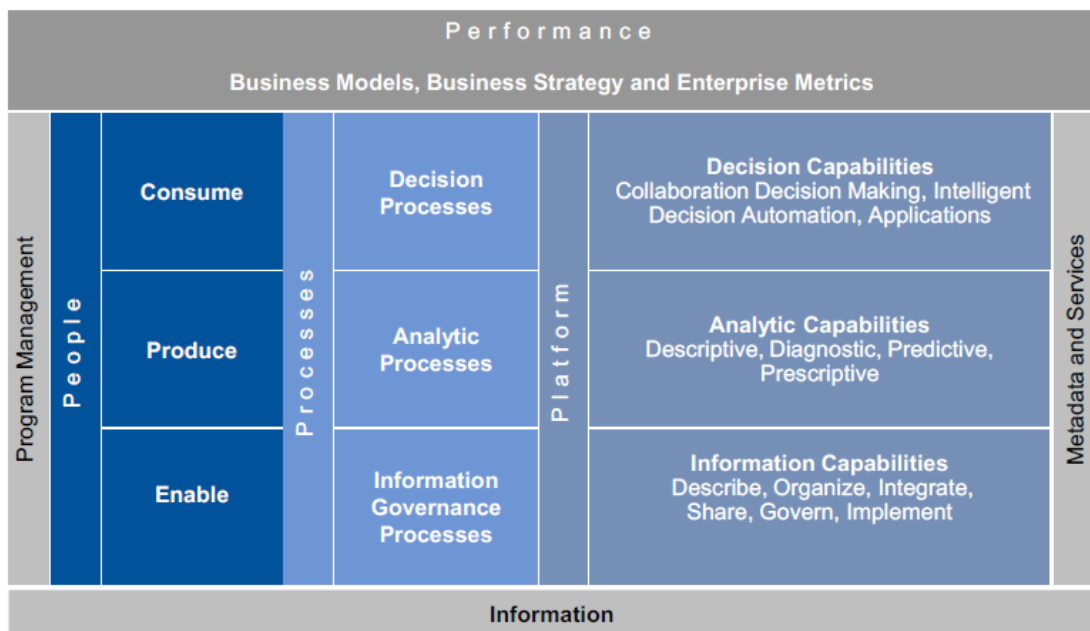
- Heterogeneidad para permitir una arquitectura que provea consistencia y eficiencia coexistiendo con una variedad de herramientas, aplicaciones y modelos de información integrados para cumplir con los requerimientos del negocio (Hostmann et al., 2009).
- La incorporación de metadatos y repositorio de servicios para permitir la integración con otros marcos de trabajo. La capa superior del marco,

Estrategia de negocio y métricas para la empresa, asegura el alineamiento con otros marcos de trabajo de negocio.

En 2011, Chandler, Hostmann, Rayner y Friedman revisaron el marco nuevamente y Gartner publicó su *Marco de Trabajo para Business Analytics*. La evolución del marco muestra las siguientes diferencias con respecto a la versión de 2009:

- La inclusión explícita del término rendimiento, un factor muchas veces desestimado, para reforzar la idea de que el marco está compuesto por cuatro aspectos: el personal, los procesos, las plataformas y el rendimiento.

Figura 11. Marco de Trabajo para Business Analytics de Gartner



Fuente: Chandler, N., Hostmann, B., Rayner, N., & Friedman, T. (2011). Gartner's business analytics framework. *Gartner Inc*, p. 3.

- La representación de las actividades del personal como tareas en lugar de roles, ya que un usuario de BA puede consumir, producir o facilitar ciertas actividades.
- La incorporación de la información como un fundamento subyacente que refleja la conexión y coexistencia de todas las fuentes de datos que BA utiliza, por ejemplo, datos estructurados y no estructurados, datos en la nube, *big data*, etc.
- La expansión de la definición del término *plataforma* hacia tres grupos de capacidades: toma de decisiones, analíticas y centradas en la información (Chandler et al., 2011).

2.4. BI en el Contexto de Salud

De acuerdo con los analistas de Gartner, Shaffer y Beyer (2014), la mayoría de los proveedores de servicio de TI que trabajan para el área de salud y para otras industrias coinciden en que los datos que maneja el campo de la salud son los más complejos, posiblemente excluyendo los utilizados por los servicios de inteligencia gubernamental. El potencial guardado en los datos generados bajo el dominio de la salud no ha sido desarrollado completamente.

Después de haber evaluado tecnologías para soluciones de *data warehousing* y definido mejores prácticas y errores comunes durante muchos años, Shaffer et al. (2014) consideran que la mayor deficiencia para el área de salud radica en la falta de

una estrategia de BI documentada o la pobreza en su desarrollo. Además predicen que la utilización de *analytics* será un factor fundamental de éxito para el sistema de salud durante el resto de esta década y es un componente en crecimiento dentro del trabajo del CIO.

En su marco de trabajo para BI en salud, Mettler y Vimarlund (2009), observan la adopción de TI en salud como una oportunidad para mejorar la efectividad, la eficiencia y la calidad del servicio de salud así como también la transparencia de las actividades económicas y la disponibilidad de la información en tiempo real.

El área de salud presenta sus propios desafíos al momento de considerar una solución de BI. Por ello, es necesario analizar las semejanzas y diferencia entre el sector de salud y otros sectores. Dentro de las semejanzas es posible mencionar:

- Todos los sectores buscan mejorar costos, calidad y tiempo de espera a través de procesos integrados.
- El foco de atención en el área de salud debería ser el paciente, sin descuidar la variedad de actores intervinientes, teniendo en cuenta el éxito que significa el enfoque centrado en el cliente que utilizan otros sectores.
- La integración de los sistemas trae beneficios para todos los sectores (Mettler et al., 2009).

Las diferencias entre el sector de salud y otros sectores al momento de considerar una solución de BI incluyen:

- El sector de salud necesita reportes clínicos y administrativos.
- La mayoría de los sectores presenta un grupo de clientes definido con algunas variantes en sus productos, mientras que en salud se involucran una multiplicidad de actores con distintas necesidades (pacientes, médicos, organismos gubernamentales, compañías aseguradoras, etc.).
- La mayoría de los sistemas industriales utiliza métricas duras, el área de salud debe además tener en cuenta métricas blandas, como la elección y el sentimiento del paciente (Mettler et al., 2009).

BI en el contexto de salud debería entonces facilitar la gestión y la toma de decisiones tanto clínica como administrativa mediante la integración de datos de ambos dominios (Mettler et al., 2009; Shaffer et al., 2014).

3. DESARROLLO DEL MARCO PROPUESTO

3.1. Generalidades del Marco de Trabajo

Según lo estudiado en el Capítulo 2 “Fundamentos Teóricos”, consideramos que BA es una extensión de las capacidades que ofrece BI, y BDA es la conjunción de *big data* y BA. Entonces ¿la implementación de un proyecto de BDA comparte los mismos factores de éxito de los cuales depende una implementación de BI bien lograda? *Big data* posee características propias como el volumen, la variedad, la velocidad, y la necesidad de manejar la veracidad de los datos. Estas características deben ser tomadas en cuenta y nuevos factores asociados deben ser incorporados para alcanzar el éxito.

La creación del nuevo marco de trabajo propuesto en esta tesis está basada en los factores críticos de éxito para la implementación de BI, estudiados en la Sección 2.3, a los cuales se sumarán los nuevos factores inherentes a *big data analytics*. Luego el marco será adaptado al contexto específico del área de la salud. El avance del nuevo marco será presentado en cada una de estas etapas.

3.2. Primera Etapa: Factores para la Implementación de BI

3.2.1. Comparación de los Marcos de Trabajo para BI

Los cuatro marcos de trabajo estudiados en la Sección 2.3 con sus evoluciones, observan la implementación de BI desde distintos lugares pero con el mismo objetivo final: obtener valor para el negocio mediante BI.

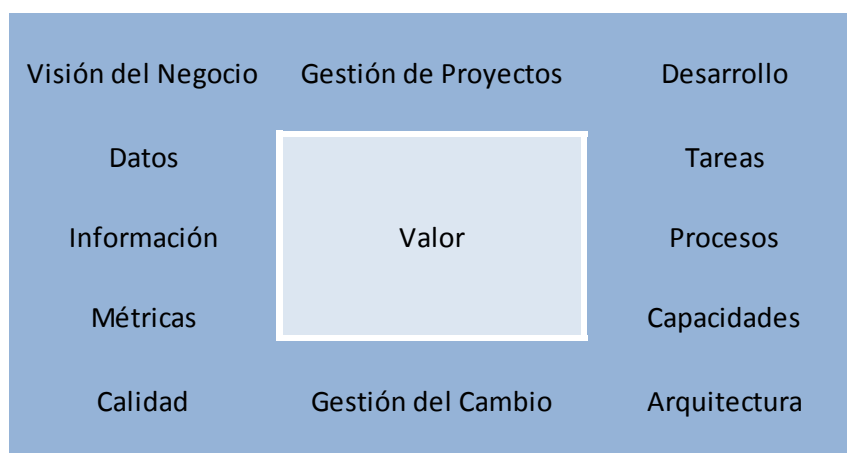
Cates et al. (2005) con su marco LOBI, proponen una metodología que incluye no sólo los pasos a seguir para la implementación, sino los niveles de madurez de inteligencia que una empresa puede alcanzar. TDWI (como se cita en Moss, 2013) jerarquiza conceptos en capas y detalla los componentes para dar estructura a la implementación de BI. Yeoh et al. (2008) se enfocan en los factores críticos de éxito, y el marco de 2010 evoluciona destacando el enfoque de orientación al negocio. Gartner propone en 2006 un marco de seis capas, que evoluciona en 2009 incluyendo el término *analytics* por primera vez entre los marcos estudiados. Además, divide la capa de personal y procesos, y destaca la necesidad de integración. Finalmente en 2011, Gartner cambia el nombre de su marco de BI a BA, girando su enfoque de la tecnología hacia las capacidades.

El Apéndice C muestra un cuadro comparativo de los conceptos presentados por cada autor. Los cuatro marcos mencionan en algún momento de su evolución factores relacionados con la visión del negocio, la gestión de proyectos, y la arquitectura e integración de datos. Tres de los cuatro marcos consideran el personal con sus roles y tareas, los procesos, los datos, y la información. Dos de los marcos hablan de valor para el negocio, aplicaciones y capacidades de BI, y del desarrollo. En contraste, sólo

uno de los cuatro marcos habla de performance y métricas, calidad, gestión del cambio, y nivel operativo y presupuesto. Sin embargo, la performance y métricas mencionadas en el marco de Gartner, y la calidad y la gestión del cambio mencionadas en el marco de Yeoh et al., se mantienen durante la evolución de los respectivos marcos.

Los factores encontrados en la comparación fueron dispuestos paralelamente para detectar la existencia de factores comunes, resultando en catorce conceptos abreviados: visión del negocio, gestión de proyectos, tareas, métricas, arquitectura, procesos, capacidades, desarrollo, datos, información, calidad, gestión del cambio, nivel operativo y presupuesto, y el factor central que es el valor para el negocio.

Figura 12. Factores Comunes dentro de los Marcos de BI



Fuente: Elaboración propia

El factor nivel operativo y presupuesto, fue desarrollado solamente en uno de los marcos estudiados, el marco LOBI presentado en la Sección 2.3.1. Este factor no será incluido en el nuevo marco de trabajo, ya que es un vasto tema de estudio que escapa los alcances de esta tesis.

3.2.2. Agrupamiento de los Factores de BI

Para facilitar su incorporación al nuevo marco de trabajo, los factores resultantes de la comparación presentada en la Sección 3.2.1 podrían ser reunidos en cuatro grupos: factores relacionados con la planificación, con la evolución, con los datos y con las operaciones (Tabla 5):

- *Factores relacionados con la planificación:* Son aquellos que reúnen los esfuerzos requeridos para alcanzar el objetivo de la obtención de valor: visión del negocio, gestión de proyectos y desarrollo.
- *Factores relacionados con la evolución:* Se enfocan en la capacidad de adaptación y crecimiento en un entorno dinámico como son las necesidades del negocio: calidad, gestión del cambio y arquitectura.
- *Factores relacionados con los datos:* Incluyen los datos y la información útil derivada de ellos que permite mejorar la toma de decisiones: datos, información y métricas.

- *Factores relacionados con las operaciones:* Estos factores se refieren a las actividades que definen el día a día del negocio: tareas, procesos y capacidades.

Tabla 5. Factores Críticos de Éxito de BI Agrupados para el Nuevo Marco

Eje Central: Valor para el Negocio			
Factores relacionados con:			
La Planificación	La Evolución	Los Datos	Las Operaciones
Visión del Negocio	Calidad	Datos	Tareas
Gestión de Proyectos	Gestión del Cambio	Información	Procesos
Desarrollo	Arquitectura	Métricas	Capacidades

Fuente: Elaboración propia

De esta manera, se completa la primera etapa de comparación y agrupamiento de los factores para la implementación de BI que se muestra en la Figura 13.

Figura 13. Primera Etapa: Factores para la Implementación de BI



Fuente: Elaboración propia

3.3. Segunda Etapa: Factores para la Implementación de BDA

3.3.1. Análisis de las Diferencias entre la Implementación de BI y BDA

Según Kobielus (como se cita en Zicari, 2014), existen cinco factores fundamentales que deben ser tomados en cuenta al momento de emprender una iniciativa de BDA.

Los factores son detallados a continuación:

- *Requerimientos de los interesados:* Las iniciativas de *big data* deben estar alineadas con los requerimientos de los interesados. Los criterios para declarar la iniciativa lista para la producción deben estar de acuerdo con los requerimientos y ello depende en gran medida en los casos de uso y las aplicaciones que se pretenden para *big data*.

- *Componentes tecnológicos:* La plataforma de *big data* debe cumplir con la disponibilidad, seguridad y otros requerimientos de robustez esperados de la mayoría de las infraestructuras empresariales.
- *Escalabilidad:* La provisión, el aumento y la reasignación de las capacidades de almacenamiento, procesamiento y transferencia para una plataforma de *big data* deben ser ágiles, costo-efectivas y modulares para así poder sostener nuevos requerimientos.
- *Capacidades del personal:* El personal debe ser capacitado en base de datos, integración, análisis de datos y otras herramientas que soporten la iniciativa de *big data*.
- *Servicio de TI:* La resolución de problemas, el entrenamiento de los usuarios y otras funciones de soporte deben ser provistos de manera eficiente y confiable, consistente con las operaciones existentes (Zicari, 2014).

Sin embargo, una empresa con un programa maduro de *enterprise data warehousing* en producción, no debería redefinir la puesta en producción para una iniciativa de *big data* (Zicari, 2014).

Entonces podría decirse que el éxito de la implementación de un proyecto de BDA depende de los mismos factores de los que depende una implementación de BI, a los cuales se le suman factores relacionados con las características particulares de *big data*: el volumen, la variedad, la velocidad, y la necesidad de manejar la veracidad de los datos.

En su exploración de cómo extraer valor de *big data* de manera exitosa, Brobst (2013) destaca los siguientes factores:

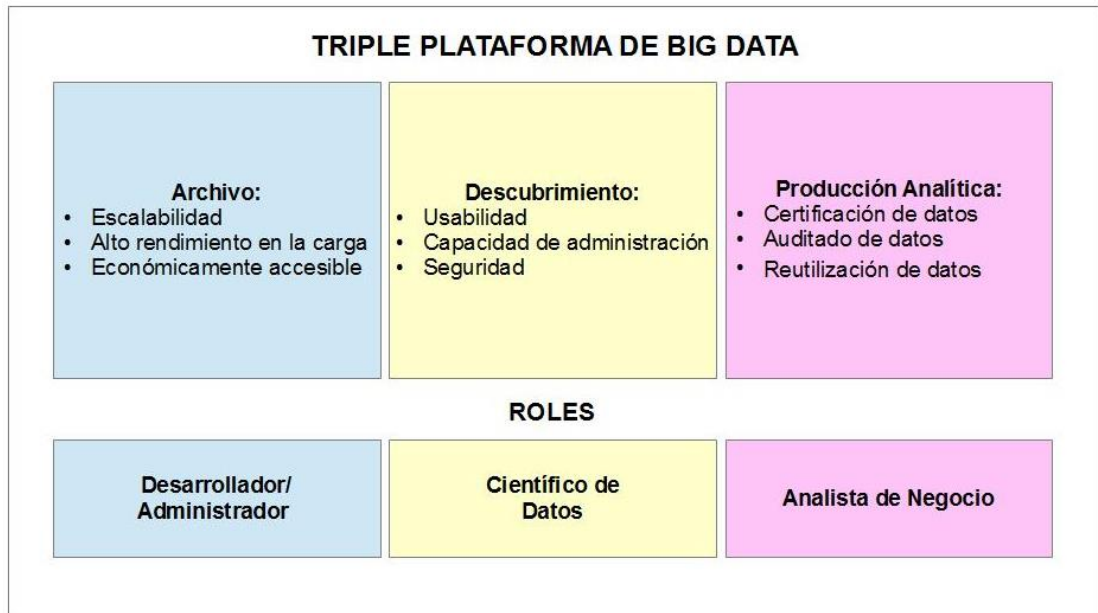
- *La nueva función de los científicos de datos:* Si bien el analista de negocio utiliza datos para responder preguntas del negocio, el científico de datos se encarga de encontrar nuevas preguntas, trabajando con visualización de datos y *data mining* para encontrar patrones y relaciones no reconocidos previamente entre los datos.
- *Triple plataforma: archivo, descubrimiento y producción analítica:* El almacenamiento y suministro de datos se realiza desde la plataforma de archivo que debe cumplir con las características de escalabilidad y alto rendimiento en la carga. Además debe ser económicamente accesible.

Los datos en la plataforma de archivo no son procesados y debido a ello, su manipulación se dificulta. Por esta razón, surge la plataforma de descubrimiento que provee usabilidad, capacidad de administración y seguridad. La plataforma de descubrimiento permite a los científicos de datos trabajar con SQL y No-SQL sobre datos relacionales y no relacionales, diseñando y ejecutando experimentos que llevan a nuevas perspectivas.

Finalmente, los datos se promueven a la plataforma analítica si se encuentra valor en ellos dentro de la plataforma de descubrimiento. La plataforma analítica consiste en un *enterprise data warehouse*, donde los datos son

certificados, auditados y reutilizados para la producción analítica (Brobst, 2013).

Figura 14. Triple Plataforma de Big Data



Fuente: Elaboración propia

Tanto Brobst (2013) como Kobiellus (como se cita en Zicari, 2014) destacan que *big data* extiende las capacidades de las tecnologías de gestión y análisis de datos existentes y que se ha generado la necesidad de incorporar tanto las tecnologías existentes como las nuevas dentro de una arquitectura unificada.

3.3.2. Incorporación de Factores Inherentes a BDA

Repasando las dimensiones de *big data*, se evidencia que estas características constituyen desafíos frente a los cuales se han logrado distintas soluciones, que

permanecen en estudio para su mejoramiento. Estos desafíos y sus soluciones se muestran en la Tabla 6.

El procesamiento de datos en tiempo real se relaciona con el concepto de *data in motion* y es una de las posibilidades que ofrece *big data*. El DW tradicional se basa en la recolección y almacenamiento de una imagen de los datos para luego ser analizados. Estos datos se denominan *data at rest* o datos en reposo. Por otro lado, el procesamiento en tiempo real implica el análisis sobre la marcha, sin un almacenamiento previo. Estos datos se denominan *data in motion* o datos en movimiento (Ebbbers, 2013).

Tabla 6. Desafíos de Big Data y sus Soluciones

Big Data	Solución
Volumen	Escalabilidad
Velocidad	Procesamiento en tiempo real
Variedad	Procesamiento de Datos Estructurados y No Estructurados
Veracidad	Confiabilidad de las Fuentes y Gobierno de los Datos

Fuente: Elaboración propia

La posibilidad de procesamiento de *data in motion* que brinda BDA incrementa el impacto en la toma de decisiones, por ejemplo, la capacidad de personalizar la experiencia de un cliente de acuerdo a sus necesidades durante el proceso de compra o la capacidad de ajustar un tratamiento médico para mejorar los resultados.

Por otro lado, BDA trae consigo nuevos roles y nuevas capacidades que se desprenden de su modelo de análisis mediante el descubrimiento, citando a Russom (2011) que utiliza el término *discovery analytics* para definir BDA.

El científico de datos es un nuevo rol que emerge con el *discovery analytics*. La posibilidad de encontrar nuevas preguntas que generen nuevas perspectivas sobre el negocio y que develen nuevos patrones requiere de profesionales que posean capacidades que exceden las del perfil tradicional del analista, e incluyen programación, matemática, estadística, y además conocimiento del negocio y habilidad de comunicación (Davenport, Barth y Bean, 2012).

Por otro lado, BDA incorpora nuevas capacidades dentro de los procesos analíticos. Esas capacidades son la predicción y la optimización. Según Halper (2014), las técnicas analíticas predictivas extienden las capacidades tradicionales reactivas hacia lo proactivo, del reporte histórico hacia el futuro.

Finalmente, dentro de las capacidades y tecnologías, es necesario incluir la triple plataforma de *big data*: archivo, descubrimiento y producción analítica (Brobst, 2013); la plataforma de archivo con sus requerimientos de velocidad y escalabilidad para trabajar con las características particulares de *big data*, la plataforma de descubrimiento para encontrar nuevos patrones y perspectivas dentro del negocio y la plataforma de producción analítica donde se realiza el análisis en tiempo real.

Los factores inherentes a BDA se muestran como una extensión de los factores críticos de éxito para la implementación de BI, estudiados en la Sección 2.3. BDA trae consigo nuevos roles, nuevas capacidades y nuevas tecnologías necesarias para maximizar las oportunidades que presenta para el negocio.

En la Tabla 7 se presenta un resumen de los factores inherentes a BDA y su relación con los factores de BI.

Tabla 7. Factores Inherentes a BDA y su Relación con los Factores de BI

Factores de BI	Factores de BDA
Roles	Científico de Datos
Capacidades	Predicción y Optimización
Tecnologías	Archivo, Descubrimiento y Producción Analítica

Fuente: Elaboración propia

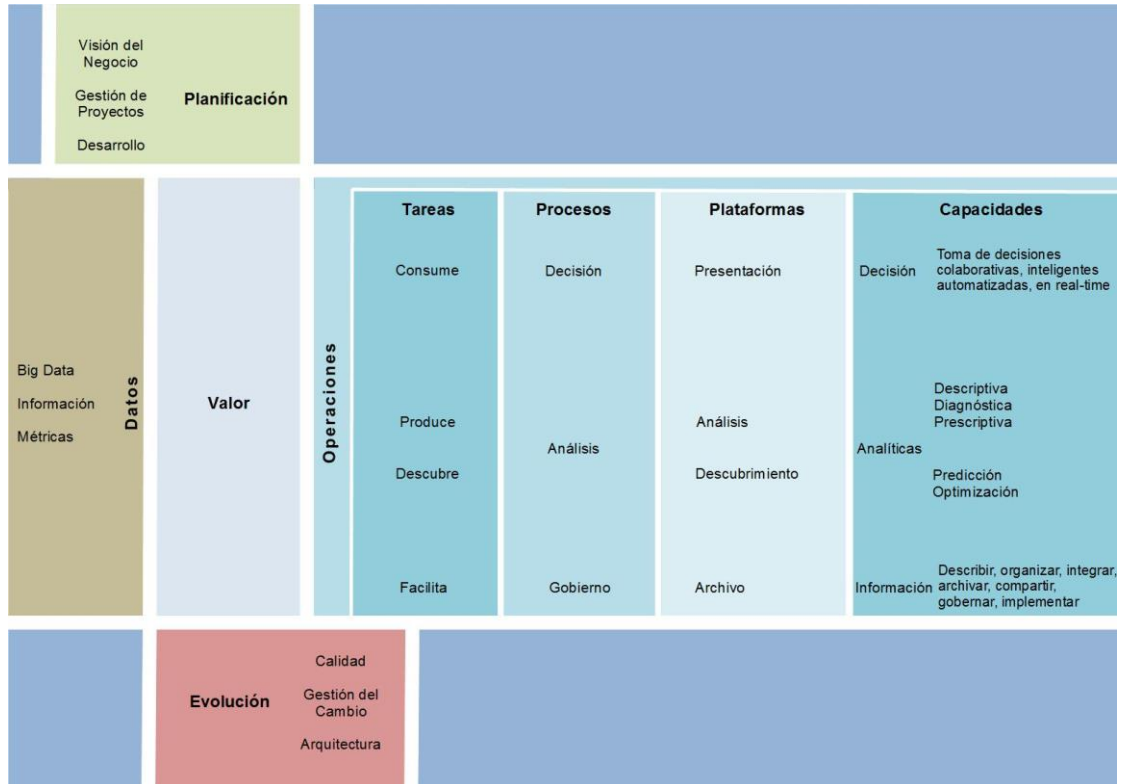
Continuando con el desarrollo del marco, es necesario incorporar los factores inherentes a BDA. El nuevo marco se muestra en la Figura 15 y es completado según los siguientes criterios:

- Los factores relacionados con la planificación y la evolución mantienen los mismos criterios que los empleados para la implementación de un proyecto de

BI como fue discutido en la Sección 3.3.1: “Análisis de las Diferencias entre la Implementación de BI y BDA”.

- El factor Datos es reemplazado por *big data*, es decir los datos que presentan ciertas características de volumen, velocidad y variedad, que son el objeto de estudio para el nuevo marco.
- Se incorporan las tareas, los procesos y las capacidades citadas en el *Marco de Trabajo para Business Analytics de Gartner* (Chandler et al., 2011), que es el que describe en detalle estos factores.
- Se suma el rol del científico de datos que se conceptualiza como la tarea del descubrimiento.
- Se describe la triple plataforma de *big data* que se suma a la plataforma de presentación para el usuario.
- Se incorpora la toma de decisiones en tiempo real, otra posibilidad que ofrece BDA.
- Se describen las capacidades analíticas según fueron discutidas en la Sección 2.1.3. Estas capacidades son predicción y optimización en la plataforma de descubrimiento, y descriptiva, diagnóstica y prescriptiva para la plataforma de análisis.
- Dentro de las capacidades de información, se suma la de archivo, teniendo en cuenta la importancia que ella implica frente a los grandes volúmenes de datos que maneja *big data*.

Figura 15. Segunda Etapa: Factores Inherentes a Big Data Analytics



Fuente: Elaboración propia

3.4. Tercera Etapa: Factores Específicos del Área de Salud

3.4.1. Estudio de BDA Aplicada a la Calidad del Servicio de Salud

Según Díaz, Juan, Lucas y Ryuga (2012):

El ser humano es quien genera la mayor cantidad de información de mayor valor crítico. Una persona puede llegar a generar 1TB de información en un día, millones de datos que podrían ser utilizados en varias disciplinas y

algunos comienzan a ser utilizados en el área de marketing o en las redes sociales, sin embargo dentro del campo de *e-health* las posibilidades ofrecidas por estos datos aún no han sido explotadas (Díaz et al., 2012, p. 898).

O'Reilly, Loukides, Steele y Hill (2012) consideran que frecuentemente cuando un profesional de la salud indica un tratamiento, ya sea cirugía o medicación sin receta, está aplicando ciertas normas de atención o alguna variante basada en su propia intuición, esperando los mejores resultados. Sin embargo, destacan que la realidad de la medicina es que la relación entre tratamiento y resultado no es siempre bien entendida.

Los trabajos de investigación muestran que determinados tratamientos funcionan más a menudo que un placebo y se sabe que gran parte de la medicina de hoy no funciona para un grupo de pacientes, sólo que no se puede predecir para qué grupo. Una de las promesas de *big data* es que la capacidad de recolectar suficientes datos sobre los tratamientos médicos y la utilización efectiva de esos datos, puede permitir la predicción más precisa sobre cuáles son los tratamientos más efectivos para qué grupo de pacientes y cuáles no lo son (O'Reilly et al., 2012).

Existen dos factores subyacentes detrás de este nuevo enfoque para la salud: el cambio en la forma de utilización de los datos y la disponibilidad de nuevos tipos de datos. No se trata solamente de indicar la efectividad de una droga en la mayoría de los pacientes sino de la utilización de técnicas de BI para agrupar pacientes y determinar la diferencia entre los distintos grupos. La pregunta cambia de: *¿la droga es efectiva?*, a: *¿para qué pacientes la droga es efectiva?* El foco se redirige de los

tratamientos hacia el paciente. Esto significa un cambio de paradigma, ya que los tratamientos específicos para un paciente no existen en la medicina de hoy. No hay posibilidades de realizar ensayos clínicos para una droga que sólo será utilizada por única vez (O'Reilly et al., 2012).

En el área del manejo clínico, los beneficios principales que puede proveer *big data* incluyen:

- Información detallada sobre distintos tipos de tratamiento, determinando qué tratamiento es más efectivo para cada paciente o grupo de pacientes.
- Toma de decisiones con el soporte del análisis de datos y detección de errores en los tratamientos médicos.
- Nivel de desempeño medible de los profesionales, detección de procesos óptimos y nivel de las instituciones.
- Nuevos modelos predictivos y de segmentación de acuerdo a los perfiles de los pacientes.
- Automatización del sistema de pagos y control de costos en la administración del sistema.
- Mejoras en la transmisión de la información, que se realizará a la persona correcta, en el momento indicado y de manera de enriquecer el proceso de toma de decisiones (Díaz et al., 2012).

3.4.2. Análisis del Valor para el Área de Salud mediante Big Data

De acuerdo con los marcos de trabajo estudiados en la Sección 2.3, el objetivo final de la implementación de sistemas de BI es la obtención de valor para el negocio. Esto lleva a investigar el concepto de *big data* como un medio para la obtención de valor en el área de salud.

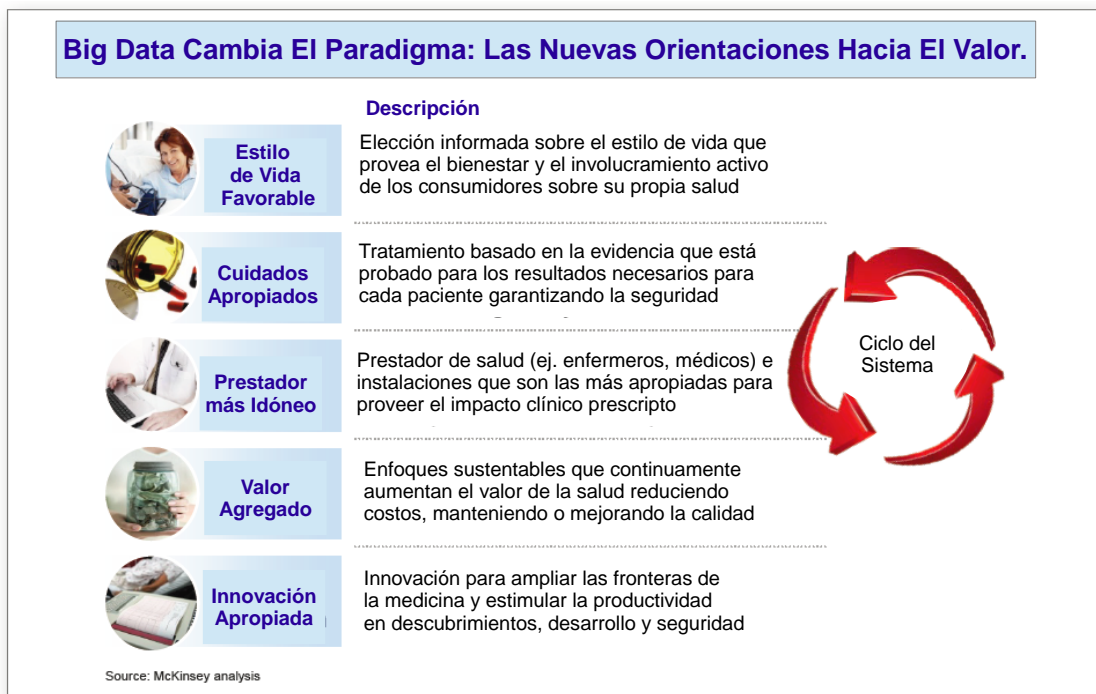
La introducción de *big data* en los sistemas de salud ha planteado la discusión entre qué es lo apropiado para el paciente y qué lo apropiado para el sistema. Acompañando estos cambios, Groves et al. (2013) han diseñado un marco de trabajo centrado en el paciente que considera cinco caminos claves para obtener valor, basado en el concepto derivado del equilibrio entre el costo del servicio y el impacto para el paciente. Los nuevos caminos para obtener valor en el área de salud utilizando *big data* son:

- *Estilo de vida favorable*: Los pacientes pueden construir valor participando activamente en su tratamiento, incluyendo prevención de enfermedades. El camino hacia un estilo de vida favorable fomenta en los pacientes toma de decisiones saludables sobre su estilo de vida, por ejemplo dieta apropiada y ejercicio físico.
- *Cuidados apropiados*: Este camino se refiere a todo lo relacionado con asegurar que el paciente obtenga el tratamiento más oportuno y apropiado que haya disponible. Además de basarse fuertemente en protocolos, los cuidados

apropiados requieren un enfoque coordinado. Tanto instituciones como prestadores deben contar con la misma información y trabajar con el mismo objetivo de manera de evitar duplicar esfuerzos o aplicar estrategias sub óptimas.

- *Prestador más idóneo:* Este camino propone que le paciente sea siempre tratado por profesionales de alto rendimiento, que están mejor relacionados con los tratamientos y que lograrán los mejores resultados. Por lo tanto, el prestador más idóneo tiene dos significados: el conjunto correcto de habilidades necesarias para afrontar la complejidad del tratamiento, así como también la selección específica del prestador con los mejores resultados probados.

Figura 16. Caminos hacia el Valor para el Área de la Salud



Fuente: Adaptado de Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The 'Big Data' revolution in healthcare. *McKinsey & Company*.

- *Valor agregado:* Para cumplir con los objetivos de este camino, los prestadores y coberturas de salud deberán acrecentar la calidad. Podría incluir medidas para asegurar costo-efectividad del tratamiento, como vincular el reintegro de los prestadores con el resultado del tratamiento, o eliminar fraude, malgasto o abuso del sistema.
- *Innovación apropiada:* Este camino incluye la identificación de nuevas terapéuticas y perspectivas de tratamiento, atravesando todos los aspectos del sistema y mejorando los motores mismos de innovación, avanzando en el campo de la medicina y apoyando la productividad del sector de investigación y desarrollo. Para capturar valor en este camino, los interesados deben utilizar de mejor manera los datos de ensayos previos, tales como buscar moléculas de alto potencial en farmacología. También podrían utilizar los datos para encontrar oportunidades de mejorar ensayos clínicos y protocolos tradicionales de tratamiento, como en el caso de las cirugías (Groves et al., 2013).

3.4.3. Adaptación del Marco al Área de Salud

Para adaptar el marco de trabajo para la implementación de BDA, factores particulares que se aplican a la salud deben ser considerados y son detallados a continuación:

- *Integración clínica y administrativa:* Varios autores coinciden en que la gestión y la toma de decisiones tanto clínica como administrativa es facilitada

mediante la integración de datos de ambos dominios (Mettler et al., 2009; Shaffer et al., 2014). Esta integración amplía el espectro de procesos (Mettler et al., 2009).

- *Procesos clínicos, administrativos y de gestión:* Además de los procesos clínicos y administrativos, es necesario distinguir los procesos de gestión, que tienen una visión global sobre la organización. Por ejemplo, el profesional de la salud analiza y toma decisiones con respecto a los tratamientos, el asistente administrativo cita pacientes de acuerdo con los requerimientos del tratamiento guiado por el clínico, y el director de la institución toma decisiones basadas en la interacción de ambos dominios. Es decir hay procesos clínicos, administrativos y de gestión que son interdependientes.
- *Componentes mínimos del repositorio integrado de datos y analytics en salud:* Shaffer et al. (2014) recomiendan la integración, como mínimo, de las siguientes plataformas: historia clínica electrónica, liquidación de prestaciones y ciclo de facturación, planificación de recursos empresariales, contabilidad de costos, y datos sobre la satisfacción del paciente. Además detallan que la nueva generación de *analytics* en salud se enfoca hacia: análisis de riesgo y predicción, estudios de cohorte y flexibilización del registro de pacientes, identificación de brechas en la asistencia sanitaria, priorización de la participación del paciente con distintas opciones de tratamiento, evaluación de la efectividad, e investigación e innovación.
- *El profesional de la salud como consumidor y productor de la información:* En la última versión del marco de trabajo de Gartner (Chandler et al., 2011)

los roles son reemplazados por tareas ya que el mismo usuario puede producir, es decir analizar, y utilizar la información para la toma de decisiones. Esto se ve claramente en el trabajo del profesional de la salud que debe monitorear valores clínicos para hacer un diagnóstico y planificar un tratamiento en base a ello.

- *Transparencia en la representación del procesamiento del conocimiento clínico:* Si bien Wright, Sittig, Ash, Sharma, Pang, y Middleton (2009) consideran que los sistemas de soporte de decisión clínicos ayudan a mejorar la calidad de la atención y a prevenir errores y omisiones, otros autores consideran que existen posibles desventajas relacionadas con su utilización (Open Clinical, 2005). La inteligencia artificial y el procesamiento complejo que muchos de estos sistemas utilizan se convierte en una caja negra para el profesional de la salud. Las desventajas de la utilización de estos sistemas incluyen la posibilidad de provocar un efecto de *de-skilling* o desprofesionalización del usuario que no entiende cómo se generan los resultados, la aparente poca flexibilidad de estos sistemas que dirigen o dan órdenes al usuario con respecto al proceder en el tratamiento, la tendencia a depender del software limitando la capacidad de pensamiento del profesional de la salud y la dificultad de evaluar los sistemas al no existir aun estándares (Open Clinical, 2005).

En contraste con los sistemas que operan como caja negra, existen otros sistemas que proveen los medios para instanciar el conocimiento clínico dentro del camino del procesamiento de la información. Un ejemplo es el proyecto Artemis, implementado en Agosto de 2009, en el departamento de

Terapia Intensiva de Neonatología del *Hospital for Sick Children* en la ciudad de Toronto, Canadá. El proyecto utiliza BDA para monitorear datos asincrónicos obtenidos del sistema de historia clínica electrónica y valores clínicos sincrónicos en tiempo real, como ser ritmo cardíaco derivado del electrocardiograma, frecuencia respiratoria, saturación de oxígeno, respiración por impedancia, y presión arterial en bebés prematuros. El análisis de los datos genera un resumen temporal e integrado de eventos en lugar de alertas frente a eventos aislados (McGregor, Catley, James, y Padbury, 2010). Esto permite la detección temprana de ciertas afecciones de alto índice de mortalidad en el recién nacido como la sepsis neonatal de aparición tardía (McGregor, Catley y James, 2012) y la apnea del neonato (Thommandram, Eklund, McGregor, Pugh y James, 2014).

El proyecto Artemis opera mediante reglas clínicas, lo que permite transparencia en la representación del procesamiento del conocimiento (McGregor et al., 2010).

- *Valor para la salud:* Basado en los cinco caminos que tiene el área de salud para obtener valor de *big data* según lo descrito por Groves et al. (2013), se puede decir que el valor para la salud radica en: la elección informada del paciente, el tratamiento probado, el profesional más idóneo, el mejoramiento de la calidad y la incorporación de la innovación.
- *Métricas blandas:* Mettler et al. (2009) consideran que una de las diferencias entre el sector de salud y otros sectores es la necesidad de incorporar métricas blandas, como la elección y el sentimiento del paciente.

3.5. Marco de Trabajo para la Implementación de BDA en el Área de Salud

Teniendo en cuenta los factores de éxito comunes encontrados en los marcos de trabajo para la implementación de BI, que fueron agrupados en la Sección 3.2.2, y la incorporación de los factores inherentes a BDA desarrollada en la Sección 3.3.2, el *Marco de Trabajo para la Implementación de Big Data Analytics en el Contexto Específico del Área de Salud* se completa considerando los siguientes criterios:

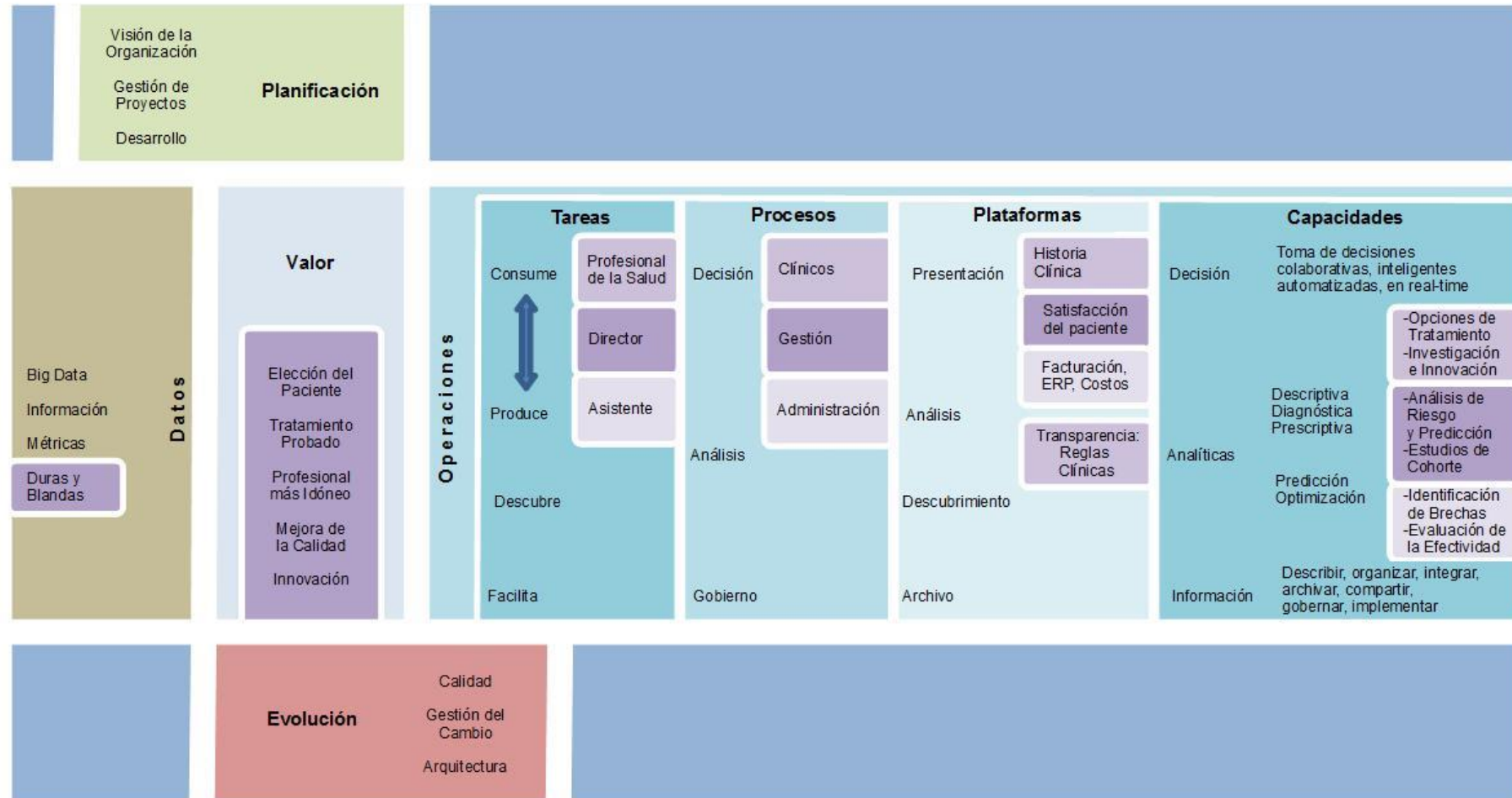
- *Factores relacionados con la Planificación:* La visión del negocio se reemplaza por visión de la organización, ya que si bien son sinónimos, organización es un término más adecuado cuando se habla de instituciones de salud.
- *Factores relacionados con los Datos:* Dentro de las métricas, se incorporan las métricas blandas necesarias para el sector de salud.
- *Factores relacionados con la Evolución:* Estos factores no presentan cambios con respecto a aquellos mencionados en el marco de trabajo para la implementación de BI.
- *Factores relacionados con las Operaciones:*
 - *Tareas:* La tarea de producir se acerca más a la de consumir, lo cual se evidencia claramente en el área de salud, y se distinguen roles específicos del sector como son el profesional de la salud en el área

clínica, el director en el área institucional y el asistente en el área administrativa.

- *Procesos*: Dentro de los procesos de decisión y análisis, se detallan los procesos específicos del área de salud: procesos clínicos, de gestión y de administración.
- *Plataforma*: Dentro de la plataforma de presentación se incorporan los componentes mínimos discutidos en la Sección 3.4.3: la historia clínica electrónica, la satisfacción del paciente, la facturación, el ERP, y los costos. Dentro de la plataforma de análisis y descubrimiento se detalla la necesidad de creación de reglas clínicas que provean transparencia en la representación del conocimiento clínico.
- *Capacidades*: Dentro de las capacidades analíticas se detallan la específicas para el área de salud: opciones de tratamiento, investigación e innovación, análisis de riesgo y predicción, estudios de cohorte, identificación de brechas y evaluación de la efectividad.
- *Valor para el área de Salud*: Se incorporan las áreas en las cuales *big data* puede colaborar en la incorporación de valor para el sector: elección del paciente, tratamiento probado, profesional más idóneo, mejora de la calidad del servicio e innovación.

El Marco de Trabajo para la Implementación de Big Data Analytics en el Contexto Específico del Área de Salud se presenta en la Figura 17.

Figura 17. Marco de Trabajo Propuesto



Fuente: Elaboración propia

4. CONCLUSIONES

Dando los primeros pasos en el aprendizaje de todo profesional de la salud, aparece una frase de Hipócrates que dice que no hay enfermedades sino enfermos. En un principio es difícil comprender la magnitud de su significado; la complejidad que implica el diagnóstico de la enfermedad teniendo en cuenta parámetros fisiológicos, genéticos, ambientales, psicológicos y la definición del rumbo del tratamiento ponderando no sólo las causas que la provocan, sino también respetando las preferencias y los sentimientos del paciente. Esta complejidad es la que puede verse beneficiada con el apoyo de las tecnologías de la información.

A continuación, se detallan las conclusiones que surgen de la investigación de esta tesis:

- La optimización en la toma de decisiones tiene una historia de más de cincuenta años, que comienza con los primeros sistemas de soporte de decisión y continúa con la introducción del término *business intelligence*.
- Se define el objetivo de *business intelligence* como la búsqueda de información oportuna que pueda ser puesta en marcha y así conducir al éxito del negocio.
- Los datos se fueron complejizando en volumen, variedad, velocidad y en la necesidad de verificar su veracidad, dando origen al término *big data*.

- Se señala que mayor cantidad de datos no significa más información, por lo cual es necesario explorar *big data* con nuevas técnicas y herramientas que la soporten.
- Con la complejidad de los datos también evoluciona la capacidad de análisis y nace el concepto de *business analytics*.
- Se encontró que si bien no todos los autores estudiados coinciden en si *business analytics* está incluido o es una extensión de *business intelligence*, hay consenso sobre los alcances de ambos dominios: *business intelligence* se enfoca en el reporte mientras que *business analytics* se orienta hacia la predicción y optimización.
- De la unión de dos entidades, *big data* y *analytics*, surge *big data analytics*, que incorpora la nueva capacidad de descubrir nuevas realidades de negocio.
- A pesar del creciente volumen de datos del cual disponen hoy las organizaciones, la capacidad de convertir los datos en información y obtener valor de ella sigue siendo un desafío.
- La implementación de *business intelligence* es un proyecto complejo que requiere la administración de recursos de software, tecnologías, datos, procesos, capacidades, cultura de la organización y apoyo de los niveles gerenciales.
- Se destaca la necesidad de un marco de trabajo que ordene el proceso de implementación debido a la gran cantidad de soluciones divergentes existentes

en el mercado, sumado a la disponibilidad de guías de implementación dirigidas por la industria que no se basan en investigación formal.

- Durante la investigación para la construcción del *Marco de Trabajo para la Implementación de Big Data Analytics en el Contexto Específico del Área de Salud*, se compararon cuatro marcos de trabajo para la implementación de *business intelligence* que fueron elaborados por destacados investigadores, respaldados por organizaciones dedicadas a la investigación de las tecnologías de la información. Estos marcos revelaron factores críticos de éxito en común.
- Los criterios resultantes de la comparación de los marcos de trabajo para la implementación de *business intelligence* fueron sintetizados en catorce nuevos conceptos: visión del negocio, gestión de proyectos, tareas, métricas, arquitectura, procesos, capacidades, desarrollo, datos, información, calidad, gestión del cambio, nivel operativo y presupuesto, y el factor central que es el valor para el negocio.
- Se encontró que debido a la relación existente entre ambos dominios, el éxito de la implementación de un proyecto de *big data analytics* depende de los mismos factores de los que depende una implementación de *business intelligence*, a los cuales se le suman factores relacionados con las características particulares de *big data*. Estos factores inherentes a *big data* fueron incorporados durante la creación del nuevo marco de trabajo presentado en esta tesis.

- El marco de trabajo para la implementación de *big data analytics* fue adaptado al campo de la salud, teniendo en cuenta las similitudes y diferencias con otros sectores.
- Las principales diferencia encontradas entre el área de salud y otros sectores incluyen la necesidad de reportes tanto clínicos como administrativos, la multiplicidad de actores involucrados con distintas necesidades (pacientes, médicos, organismos gubernamentales, compañías aseguradoras, etc.), y la utilización de métricas blandas (elección y sentimiento del paciente).
- Se exploraron los nuevos caminos para obtener valor en el área de salud a través de *big data*: estilo de vida favorable, cuidados apropiados, prestador más idóneo, valor agregado e innovación apropiada.
- Y por sobre todo se resalta que se necesita más que tecnología para crear valor. Se espera que este marco de trabajo ayude a implementar *big data analytics* en el campo de la salud. Este nuevo marco considera las complejidades propias del área sin perder de vista los factores críticos de éxito que provienen de la vasta experiencia de implementación de *business intelligence* en otros sectores. De esta manera, el campo de la salud debería ver facilitada la gestión y la toma de decisiones tanto clínica como administrativa mediante la integración de datos de ambos dominios.

5. FUTURAS LÍNEAS DE INVESTIGACIÓN

A partir de esta tesis surgen otras líneas de investigación que pueden enriquecer el *Marco de Trabajo para la Implementación de Big Data Analytics en el Contexto Específico del Área de Salud*. Estas líneas de investigación incluyen:

- Utilizar el marco de trabajo propuesto exponiéndolo a un entorno real y proponiendo mejoras sobre el mismo.
- Definir niveles de madurez de inteligencia para el sector de salud teniendo en cuenta tanto el dominio clínico como el administrativo.
- Profundizar el marco especificando pautas para el gobierno de los datos en el área de salud.
- Profundizar el marco detallando las medidas de seguridad necesarias para compartir datos clínicos respetando la confidencialidad del paciente.
- Profundizar el marco especificando los lineamientos necesarios para implementaciones en la nube.
- En el caso de la utilización de datos clínicos en tiempo real, investigar el impacto de los requerimientos cambiantes del sistema frente a nuevos descubrimientos médicos, evaluar la necesidad de regulación, responsabilidad de actualización, etc.

BIBLIOGRAFÍA

- Barranco Fragoso, R. (2012). ¿Qué es Big Data? *IBM DeveloperWorks*. Recuperado el 1 de octubre de 2013 del sitio Web de IBM: <http://www.ibm.com/developerworks/ssa/local/im/que-esbig-data/>
- Brobst, S. (5 de Marzo de 2013). Why Big Data projects fail. Recuperado el 15 de septiembre de 2014 del sitio Web de DataInformed: <http://data-informed.com/why-big-data-projects-fail/>
- Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? Recuperado el 4 de abril de 2015 del sitio Web de SSRN: <http://ssrn.com/abstract=1819486>
- Cates, J. E., Gill, S. S., & Zeituny, N. (2005). The Ladder of Business Intelligence (LOBI): a framework for enterprise IT planning and architecture. *International Journal of Business Information Systems*, 1(1), 220-238.
- Cavalheiroa, G., Dahanayakea, A., & Welkeb, R. (2006). Combining business activity monitoring with the data warehouse for event-context correlation. En *Proceedings of the 8th International Conference on Enterprise Information Systems*.
- Chandler, N., Hostmann, B., Rayner, N., & Friedman, T. (2011). Gartner's business analytics framework. *Gartner Inc*.
- Davenport, T. H. (2006). Competing on analytics. *Harvard business review*, (84), 98-107.
- Davenport, T. H. (2009). Make better decisions. *Harvard business review*, 87(11), 117-123.
- Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.
- Davenport, T. H., Barth, P., & Bean, R. (2012). How 'Big Data' is different. *MIT Sloan Management Review*, 54(1), 22-24.
- Díaz, M., Juan, G., Lucas, O., & Ryuga, A. (2012, Julio). Big Data on the Internet of Things: An example for the E-health. En *Proceedings of the 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing* (pp. 898-900). IEEE Computer Society.
- Ebbers, M. (2013). 5 things to know about Big Data in motion. *IBM DeveloperWorks*. Recuperado el 2 de enero de 2015 del sitio Web de IBM: https://www.ibm.com/developerworks/community/blogs/5things/entry/5_things_to_know_about_big_data_in_motion?lang=en

- Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The 'Big Data' revolution in healthcare. *McKinsey & Company*.
- Halper, F. (2014). Predictive Analytics for business advantages. *TDWI Best Practices Report, First Quarter*.
- Hostmann, B., Rayner, N. & Friedman, T. (2006). Gartner's business intelligence and performance management framework. *Gartner Inc.*
- Hostmann, B., Rayner, N. & Herschel, G. (2009). Gartner's business, analytics and performance management framework. *Gartner Inc.*
- IBM (2010). A practical framework for business intelligence and planning in midsize companies. Recuperado el 6 de marzo de 2014, del sitio Web de IBM: <http://www.ibm.com/midmarket/uk/en/att/pdf/practicalframework.pdf>.
- IBM (s.f.). Analytics. Analyze all of your data for a fast ROI. Recuperado el 29 de junio de 2014, del sitio Web de IBM: <http://www.ibm.com/analytics/us/en/>
- Inmon, W. H. (2005). *Building the Data Warehouse*. John Wiley & Sons.
- Lallemant, N. C. (2012). Health policy brief: reducing waste in health care. *Health Affairs*, 13.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big Data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*.
- McGregor, C., Catley, C., & James, A. (2012, Junio). Variability analysis with analytics applied to physiological data streams from the neonatal intensive care unit. En *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on* (pp. 1-5). IEEE.
- McGregor, C., Catley, C., James, A., & Padbury, J. (2010). Next generation neonatal health informatics with Artemis. *Studies in health technology and informatics*, 169, 115-119.
- Mettler, T., & Vimarlund, V. (2009). Understanding business intelligence in the context of healthcare. *Health informatics journal*, 15(3), 254-264.
- Moss, L. (2013). *Extreme scoping*. Bradley Beach, NJ: Technics Publications
- Open Clinical (2005). DSS Success Factors. Recuperado el 17 de Enero de 2015, del sitio Web de Open Clinical: <http://www.openclinical.org/dssSuccessFactors.html>
- Oracle (2011). Oracle Database VLDB and Partitioning Guide 11g Release 2 (11.2). Recuperado el 24 de mayo de 2015, del sitio Web de Oracle: http://docs.oracle.com/cd/E11882_01/server.112/e25523.pdf

- O'Reilly, T., Loukides, M., Steele, J., & Hill, C. (2012). *How data science is transforming health care*. O'Reilly Media Inc.
- Pandre, A. (s.f.). OLAP Cubes. Recuperado el 24 de mayo de 2015 del sitio Web de Data Visualization: <https://apandre.wordpress.com/data/datacube/>.
- Ponniah, P. (2004). *Data warehousing fundamentals: a comprehensive guide for IT professionals*. John Wiley & Sons.
- Power, D. J. (2007). A brief history of decision support systems. Recuperado el 8 de marzo de 2014, del sitio Web de DSSResources: <http://dssresources.com/history/dsshhistory.html>, versión 4.
- Rouse, M. (2010). Search Business Analytics. Recuperado el 6 de marzo de 2014, del sitio Web de TechTarget: <http://searchbusinessanalytics.techtarget.com>
- Russom, P. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter*.
- SAS Institute Inc. (2008). Eight levels of analytics. Recuperado el 5 de Julio de 2014, del sitio Web de SAS: http://www.sas.com/news/sascom/analytics_levels.pdf
- Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). Analytics: The real-world use of big data. *IBM Global Business Services*.
- Shaffer, V & Beyer, M. (10 de Febrero de 2014). Top actions for healthcare delivery organization CIOs, 2014; Avoid 25 years of mistakes in Enterprise Data Warehousing. Recuperado el 10 de enero de 2015 del sitio Web de Gartner: <http://www.gartner.com/technology/reprints.do?id=1-1R0QJVG&ct=140221&st=sb>
- TDWI (s.f.). Business Intelligence. Recuperado el 5 de Julio de 2014, del sitio Web de TWDI: <http://tdwi.org/portals/business-intelligence.aspx>
- Thommandram, A., Eklund, J. M., McGregor, C., Pugh, J. E., & James, A. G. (2014, June). A rule-based temporal analysis method for online health analytics and its application for real-time detection of neonatal spells. In *Big Data (BigData Congress), 2014 IEEE International Congress on* (pp. 470-477). IEEE.
- Weill, P., & Aral, S. (2006). Generating premium returns on your IT investments. *Sloan Management Review*, 47(2).
- Wright, A., Sittig, D. F., Ash, J. S., Sharma, S., Pang, J. E., & Middleton, B. (2009). Clinical decision support capabilities of commercially-available clinical information systems. *Journal of the American Medical Informatics Association*, 16(5), 637-644.
- Yeoh, W., & Koronios, A. (2010). Critical success factors for business intelligence systems. *Journal of computer information systems*, 50(3), 23-32.

Yeoh, W., Koronios, A., & Gao, J. (2008). Managing the implementation of business intelligence systems: a critical success factors framework. *International Journal of Enterprise Information Systems (IJEIS)*, 4(3), 79-94.

Zicari, R. V. (15 de Mayo de 2014). How to run a Big Data project. Interview with James Kobielus. *ODBMS Industry Watch*. Recuperado el 28 de septiembre de 2014 del sitio Web de ODBMS: <http://www.odbms.org/blog/2014/05/james-kobielus/>

APÉNDICES

Apéndice A. Estimación del gasto ineficiente en salud en los Estados Unidos en 2011

Estimates of Waste in US Health Care Spending in 2011, by Category

	Cost to Medicare and Medicaid ^a			Total cost to US health care ^b		
	Low	Midpoint	High	Low	Midpoint	High
Failures of care delivery	\$26	\$36	\$45	\$102	\$128	\$154
Failures of care coordination	21	30	39	25	35	45
Overtreatment	67	77	87	158	192	226
Administrative complexity	16	36	56	107	248	389
Pricing failures	36	56	77	84	131	178
Subtotal (excluding fraud and abuse)	166	235	304	476	734	992
Percentage of total health care spending	6%	9%	11%	18%	27%	37%
Fraud and abuse	30	64	98	82	177	272
Total (including fraud and abuse)	197	300	402	558	910	1,263
Percentage of total health care spending				21%	34%	47%

SOURCE Donald M. Berwick and Andrew D. Hackbarth, "Eliminating Waste in US Health Care," *JAMA* 307, no. 14 (April 11, 2012):1513-6. Copyright © 2012 American Medical Association. All rights reserved.

NOTES Dollars in billions. Totals may not match the sum of components due to rounding. ^aIncludes state portion of Medicaid. ^bTotal US health care spending estimated at \$2.687 trillion.

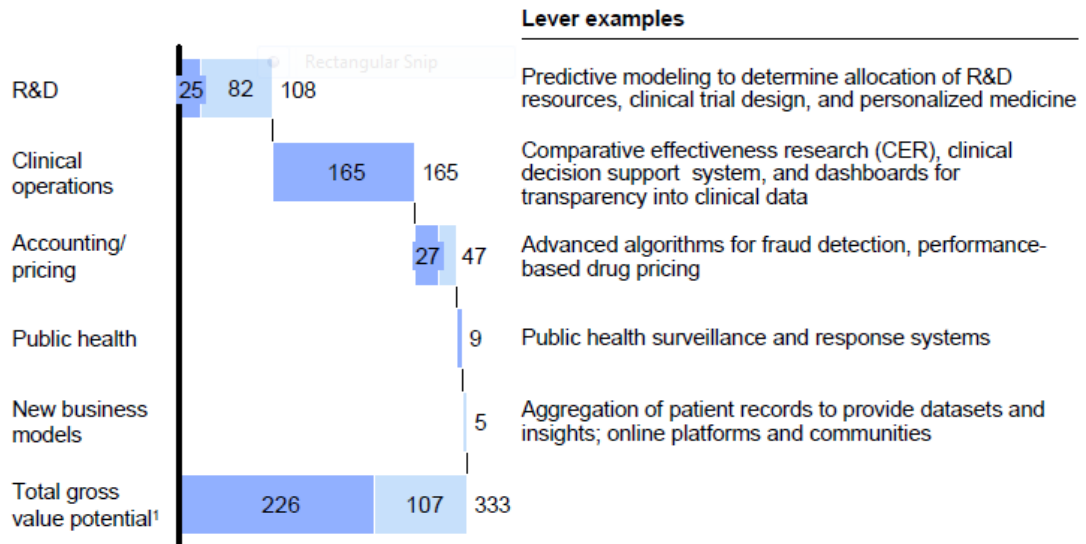
Fuente: Lallemand, N. C. (2012). Health policy brief: reducing waste in health care. *Health Affairs*, 13, p. 2.

Apéndice B. Valor Potencial de la Utilización de Big Data en Salud en los Estados Unidos

The estimated long-term value of identified levers is more than \$300 billion, with potentially more than \$200 billion savings on national health care spending

Value potential from use of big data
\$ billion per year

■ Direct reduction on national health care expenditure
■ Unclear impact on national health care expenditure



¹ Excluding initial IT investments (~\$120 billion–\$200 billion) and annual operating costs (~\$20 billion per annum).

SOURCE: Expert interviews; press and literature search; McKinsey Global Institute analysis

Fuente: Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*, p.50.

Apéndice C. Comparación de los Marcos de Trabajo para la Implementación de BI

	Cates et al. (2005)	TDWI (2006)	Yeoh et al. (2008)	Yeoh et al. (2010)	Gartner (2006)	Gartner (2009)	Gartner (2011)
Valor para el negocio		Valor para el negocio	Apoyo gerencial y sponsorship	Percepción de beneficios para el negocio			
			Championship centrado en el negocio y composición balanceada del equipo del proyecto	Gerencia y Championship			
Visión del negocio	Visión, misión y preguntas claves para el negocio	Requerimientos del negocio		Orientación al negocio	Estrategia de negocio		
	Rumbo de la empresa		Visión clara del negocio y caso bien establecido	Visión y caso de negocio			
Gestión de Proyectos	Equipo corporativo de gestión de procesos y Oficina de Gestión de Programas	Gestión de Programas	Metodología y gestión de proyectos dirigida por el negocio	Gestión de proyectos y metodología		Gestión de Programas	Gestión de Programas
Personal: Roles y Tareas	Análisis de roles ejecutivos						
	Análisis del resto de roles de negocio						
	Personal			Equipo	Personal		
						Consumidores (Usuarios)	Personal: Consumidores
					Productores (Analistas)	Personal: Productores	
					Facilitadores (TI)	Personal: Facilitadores	
Performance y Métricas					Gestión de la performance (métricas)	Estrategia y métricas de negocio	Performance: Modelo, estrategia y métricas de negocio
Arquitectura e Integración de Datos	Determinar la arquitectura de la tecnología existente	Arquitectura de BI	Marco de trabajo técnico estratégico y extensible		Infraestructura del manejo de la información	Repositorios de metada y servicios	Servicios y metadata
	Identificar la tecnología a implementar e implementar la nueva arquitectura Tecnología		Calidad de datos sustentable y marco de gobierno				

(cont.)	Cates et al. (2005)	TDWI (2006)	Yeoh et al. (2008)	Yeoh et al. (2010)	Gartner (2006)	Gartner (2009)	Gartner (2011)
Procesos	Procesos		Performance del proceso (presupuesto y tiempo)	Performance del proceso (presupuesto y tiempo)	Procesos	Procesos de negocio y decisión	Procesos de decisión
	Guía de procesos					Procesos Analíticos	Procesos Analíticos
	Procesos claves existentes y a lograr						
						Procesos de infraestructura de la información	Procesos de gobierno de la información
Aplicaciones y Capacidades de BI		Data Warehousing					
		Aplicaciones de negocio				Aplicaciones de Procesos de Negocio	Plataforma: capacidades de decisión
					Aplicaciones analíticas	Aplicaciones analíticas	Plataforma: capacidades analíticas
			Operaciones de BI y data warehousing				Plataforma: capacidades de Información
Desarrollo		BI			Plataforma de BI	Capacidades de BI	
		Desarrollo		Desarrollo alineado con el negocio			
Datos		Recursos de Datos		Datos		Administración de los recursos de datos	
		Adquisición, limpieza e integración de Datos					
		Almacenamiento de Datos					
Información		Suministro de Información		Infraestructura		Infraestructura de la información	Información
		Servicios de Información					
Calidad			Rendimiento de la infraestructura (calidad del sistema, calidad de la información y utilización del sistema)	Rendimiento de la infraestructura (calidad del sistema, calidad de la información y utilización del sistema)			

(cont.)	Cates et al. (2005)	TDWI (2006)	Yeoh et al. (2008)	Yeoh et al. (2010)	Gartner (2006)	Gartner (2009)	Gartner (2011)
Gestión del Cambio			Gestión del cambio orientado al usuario de negocio	Gestión del cambio			
Nivel Operativo y Presupuesto	Acordar el nivel operativo de LOBI a ser alcanzado y asignar un presupuesto						

Criterios mencionados por:	
Cuatro marcos	
Tres marcos	
Dos marcos	
Un marco	

Fuente: Elaboración propia