



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN Y VINCULACIÓN CON LA COLECTIVIDAD

MAESTRÍA EN GERENCIA DE SISTEMAS IX PROMOCIÓN

TESIS DE GRADO MAESTRÍA EN GERENCIA DE SISTEMAS

**TEMA: “DESARROLLO DE UN MODELO DE GESTIÓN DE CALIDAD
DE DATOS BASADO EN LA METODOLOGÍA IBM DATA QUALITY
PARA EL PORTAFOLIO DE BUSINESS INTELLIGENCE
DE LA EMPRESA DWCONSULWARE”**

**AUTORES:
ANDRADE TIRIRA, CHRISTIAN ANDRÉS
MADRID RUIZ, DAVID ALEJANDRO**

DIRECTOR: ING. FERNANDO GALÁRRAGA, MSc.

SANGOLQUÍ, ENERO DE 2015

UNIVERSIDAD DE LAS FUERZAS ARMADAS – ESPE
MAESTRÍA EN GERENCIA DE SISTEMAS

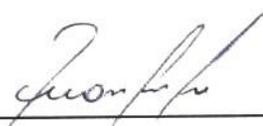
CERTIFICACIÓN

ING. FERNANDO GALÁRRAGA MSc.

CERTIFICA

El proyecto de grado titulado “DESARROLLO DE UN MODELO DE GESTIÓN DE CALIDAD DE DATOS BASADO EN LA METODOLOGÍA IBM DATA QUALITY PARA EL PORTAFOLIO DE BUSINESS INTELLIGENCE DE LA EMPRESA DWCONSULWARE”, realizado por los señores Christian Andrés Andrade Tirira y David Alejandro Madrid Ruiz, ha sido guiado y revisado periódicamente y cumple normas estatutarias establecidas por la Universidad de las Fuerzas Armadas ESPE, por tanto, se autoriza su presentación para los fines legales pertinentes.

Sangolquí, Enero de 2015



Ing. Fernando Galárraga MSc.
Director

UNIVERSIDAD DE LAS FUERZAS ARMADAS – ESPE

MAESTRÍA EN GERENCIA DE SISTEMAS

DECLARACIÓN DE RESPONSABILIDAD

CHRISTIAN ANDRADE Y DAVID MADRID

DECLARAMOS QUE:

El proyecto de grado denominado “DESARROLLO DE UN MODELO DE GESTIÓN DE CALIDAD DE DATOS BASADO EN LA METODOLOGÍA IBM DATA QUALITY PARA EL PORTAFOLIO DE BUSINESS INTELLIGENCE DE LA EMPRESA DWCONSULWARE”, ha sido desarrollado en base a una investigación exhaustiva, respetando los derechos intelectuales de terceros, conforme las citas que constan en cada texto correspondiente, cuyas fuentes se incorporan en la bibliografía.

Consecuentemente este trabajo es nuestra autoría.

En virtud de esta declaración, nos responsabilizamos del contenido, veracidad y alcance científico del proyecto de grado en mención.

Sangolquí, Enero de 2015



Christian Andrade Tirira



David Madrid Ruiz

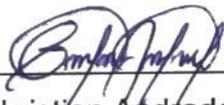
UNIVERSIDAD DE LAS FUERZAS ARMADAS – ESPE
MAESTRÍA EN GERENCIA DE SISTEMAS

AUTORIZACIÓN

Nosotros, Christian Andrade y David Madrid

Autorizamos a la Universidad de las Fuerzas Armadas – ESPE la publicación, en la biblioteca virtual de la Institución del trabajo “DESARROLLO DE UN MODELO DE GESTIÓN DE CALIDAD DE DATOS BASADO EN LA METODOLOGÍA IBM DATA QUALITY PARA EL PORTAFOLIO DE BUSINESS INTELLIGENCE DE LA EMPRESA DWCONSULWARE”, cuyo contenido, ideas y criterios son de nuestra exclusiva responsabilidad y autoría.

Sangolquí, Enero de 2015


Christian Andrade Tirira


David Madrid Ruiz

DEDICATORIA

Este trabajo se lo dedicamos a Dios, quien supo guiarnos en todo momento y darnos fuerzas para continuar y no desmayar a pesar de todas las adversidades que se presentaron en el camino. A nuestras familias por su permanente apoyo y solidaridad, aun cuando la meta parecía inalcanzable.

Christian - David

AGRADECIMIENTO

A nuestros compañeros, profesores y el coordinador de la Novena Promoción de la Maestría en Gerencia de Sistemas, por hacer de cada clase un lugar de aprendizaje y encuentro, donde fue posible compartir conocimientos, experiencias y opiniones, que nos han ayudado a convertirnos en mejores profesionales y personas.

Christian - David

ÍNDICE DE CONTENIDO

CERTIFICACIÓN	ii
DECLARACIÓN DE RESPONSABILIDAD	iii
AUTORIZACIÓN	iv
DEDICATORIA	v
AGRADECIMIENTO	vi
ÍNDICE DE CONTENIDO	vii
ÍNDICE DE TABLAS	x
ÍNDICE DE CUADROS	xi
ÍNDICE DE FIGURAS	xii
RESUMEN	xiii
ABSTRACT	xiv
CAPÍTULO I	3
1. ANTECEDENTES	3
1.1 Justificación e Importancia	3
1.2 Planteamiento del Problema	3
1.3 Objetivos	4
1.3.1 Objetivo General	4
1.3.2 Objetivos Específicos.	4
1.4 Alcance	4
CAPÍTULO II	6
2. MARCO TEÓRICO	6
2.1 Inteligencia de Negocios	6
2.1.1 Información	6
2.1.2 Sistemas de Información	7
2.1.3 Business Intelligence	9
2.1.4 Toma de Decisiones	10
2.1.5 Arquitectura de un SDIN	12
2.2 Modelos de Madurez de BI	12
2.2.1 Modelo de Madurez TDWI	14

2.2.1.1	Etapas del modelo TDWI	15
2.2.1.2	Puntos críticos: el Golfo y el Abismo	16
2.3	Calidad de Datos	16
2.4	Gestión de Calidad de Datos	19
2.4.1	Ciclo de Vida del Proyecto de Calidad de Datos	19
2.4.1.1	Definir Objetivos.....	20
2.4.1.2	Evaluación y descubrimiento	21
2.4.1.3	Validación	22
2.4.1.4	Limpieza y enriquecimiento - Coincidencia y supervivencia	23
2.4.1.5	Monitoreo y Seguimiento	23
2.5	Modelos de Gestión de Calidad de Datos.	24
CAPÍTULO III.		25
3.	MODELO DE GESTIÓN	25
3.1	Modelo de Gestión de Calidad de Datos de IBM	25
3.2	Definición del Proyecto de BI	26
3.2.1	Ciclo de vida del proyecto BI	30
3.3	Evaluación de Niveles de Madurez	31
3.4	Gestión de Calidad de Datos	37
3.4.1	Fases de la Gestión de Calidad de Datos.	38
3.4.1.1	Investigación de los datos	38
3.4.1.2	Estandarización de los datos	40
3.4.1.3	Coincidencia de los datos	41
3.4.1.4	Supervivencia de los datos	42
3.4.2	Metodología de Calidad de Datos de IBM	44
CAPÍTULO IV		49
4.	CONCLUSIONES Y RECOMENDACIONES	49
4.1	CONCLUSIONES	49
4.2	RECOMENDACIONES	50

	ix
BIBLIOGRAFÍA	51
ANEXOS	54
ANEXO A TDWI BI Maturity Model	55
ANEXO B CUESTIONARIO TDWI	56

ÍNDICE DE TABLAS

Tabla 1: Escala de madurez de TDWI.....	35
Tabla 2: Escala de madurez de TDWI con Golfo y Abismo.....	36

ÍNDICE DE CUADROS

Cuadro 1: Tipos de Business Intelligence	9
Cuadro 2: Tipos de Decisiones	11
Cuadro 3: Características de las Decisiones	11
Cuadro 4: Niveles de Madurez por Perspectiva.....	37

ÍNDICE DE FIGURAS

Figura 1: Niveles de Necesidad de la Información	7
Figura 2: Niveles de Sistemas de Información	8
Figura 3: Arquitectura de un Sistema de Inteligencia de Negocios	12
Figura 4: Etapas de Madurez de BI	14
Figura 5: Características de datos de calidad.....	17
Figura 6: Características de la Calidad de Datos	18
Figura 7: Ciclo de Vida de un Proyecto de Calidad de Datos	20
Figura 8: Definición de Reglas.....	22
Figura 9: Ciclo del modelo de gestión de calidad de datos.....	26
Figura 10: Componentes del BI	28
Figura 11: Tipos básicos de Soluciones de BI.....	29
Figura 12: Ciclo de vida de un proyecto BI	30
Figura 13: Ejemplo de evaluación del modelo de madurez TDWI	34
Figura 14: Etapas del modelo de madurez TDWI.....	34
Figura 15: Fases de la Gestión de Calidad de Datos	38
Figura 16: Ejemplo de Análisis de Datos	39
Figura 17: Ejemplo de Estandarización de Datos I.....	40
Figura 18: Ejemplo de Estandarización de Datos II.....	41
Figura 19: Ejemplo de Supervivencia de Datos I.....	44
Figura 20: Ejemplo de Supervivencia de Datos II.....	44

RESUMEN

Los modelos de madurez permiten evaluar la percepción, aprovechamiento y soporte de la inteligencia de negocios (BI) dentro de una organización, por otro lado la Gestión de Calidad de Datos se enfoca en corregir y disminuir los defectos en los datos utilizados para procesos de BI, la combinación de ambos conceptos nos permite efectivamente evaluar las ventajas de la aplicación de la Gestión de Calidad de Datos. Este trabajo presenta el desarrollo de un modelo de Gestión de Calidad de Datos para los proyectos de la empresa DWConsultware, en el modelo se definió el ciclo de vida, las fases y la metodología de un proyecto de calidad de datos, adicionalmente se establecieron los niveles de un modelo de madurez y la manera de evaluar una organización en función de las diferentes perspectivas definidas en el modelo propuesto por TDWI para evaluar el cambio en la organización al desarrollar un proyecto de mejora de la calidad de los datos.

PALABRAS CLAVES:

CALIDAD DE DATOS

MODELOS DE MADUREZ DE BI

PROYECTO DE BI

TDWI

ABSTRACT

Maturity models allow measuring the perception, use and support of Business Intelligence (BI) within an organization, Data Quality Management focuses on correct and reduce defects in the data used for BI processes, the combination of both concepts effectively allows us to evaluate the benefits of implementing Data Quality Management processes. This work presents the development of a Data Quality Management Model to be applied in projects of the company DWConsultware, the model contains a life cycle, phases and methodology for a data quality projects, additionally the maturity levels and the mechanism to evaluate an organization in relation with the different perspectives defined in the maturity model proposed by TDWI in order to evaluate the changes in the organization when a data quality project is developed.

KEY WORDS:

CALIDAD DE DATOS

MODELOS DE MADUREZ DE BI

PROYECTO DE BI

TDWI

DESARROLLO DE UN MODELO DE GESTIÓN DE CALIDAD DE DATOS
BASADO EN LA METODOLOGÍA IBM DATA QUALITY PARA EL
PORTAFOLIO DE BUSINESS INTELLIGENCE DE LA EMPRESA
DWCONSULWARE

La Gestión de Calidad de Datos es el proceso que comprueba que los datos tengan los valores correctos y tipos de datos válidos, comprende una serie de procedimientos y técnicas aplicadas con el fin de asegurar que los datos disponibles cumplan con una serie de requisitos y características que los hagan “apropiados para los usos previstos en las operaciones, toma de decisiones y planificación” como indica Joseph M. Juran (Menéndez, 2013).

Este proyecto de tesis se enfocará en desarrollar un modelo de Gestión de la Calidad de Datos que permita estandarizar los proyectos de desarrollo de Business Intelligence basado en la metodología IBM Data Quality la cual básicamente define a la Calidad de Datos como información completa, válida, coherente y precisa, lo que hace a los datos apropiados para un uso específico.

El Modelo de Gestión desarrollado será aplicado en el portafolio de Business Intelligence de la empresa DWConsultware, la cual es líder en soluciones de administración de desempeño corporativo (CPM). Los productos y servicios de esta empresa permiten que las organizaciones alcancen un alto desempeño planificando, presupuestando, pronosticando, monitoreando con alertas o tableros de control y analizando con BI.

DWConsultware es una empresa con más de 15 años en el mercado, presta sus servicios a las más importantes empresas del país en diferentes industrias. Su alto nivel de servicio le ha permitido destacarse en Latinoamérica recibiendo premios importantes de su asociado principal IBM.

Para la elaboración de este proyecto se analizarán en forma general los Modelos de Gestión de Calidad de Datos y de forma particular el Modelo de

Gestión de IBM, igualmente se analizarán los diferentes Modelos de Madurez de Inteligencia de Negocios y específicamente el que plantea The Data Warehousing Institute (TDWI).

En cuanto a la Metodología y Técnicas de Investigación:

- Es una *investigación Aplicada* ya que a través del trabajo se obtendrá un modelo de gestión de calidad de datos para DWConsulware.
- La investigación se abordará de forma *Cualitativa* esto debido a que el trabajo se basa en estándares definidos por IBM.
- Los objetivos son propios de una *investigación Explicativa* porque el trabajo está dirigido a establecer un procedimiento que permita administrar la calidad de datos en las empresas.
- Para esta investigación se utilizará el procedimiento técnico del Estudio de Caso, el cual se basa en el *método científico inductivo* ya que se va a desarrollar un caso en base al modelo de gestión de calidad de datos.

CAPÍTULO I

1. ANTECEDENTES

1.1 Justificación e Importancia

Fundada en 1998, DWConsulware es líder en el país en cuanto a implementaciones de proyectos de inteligencia de negocios. Su principal base de operación se encuentra en Quito, Ecuador, sin embargo tiene instalaciones en Ecuador y Perú, su portafolio de clientes se extiende a varios países de América Latina. El rápido crecimiento del portafolio de clientes de la empresa hace que DWConsulware quiera incursionar en más líneas de negocio en busca de ofrecer nuevos servicios.

Para seguir innovando y continuar con su crecimiento, DWConsulware está incursionando en nuevas líneas de negocio como la calidad de datos, la cual requiere de un modelo definido para poder llevar a cabo la gestión de este tipo de proyectos.

1.2 Planteamiento del Problema

DWConsulware ha definido una metodología de proyectos de Inteligencia de Negocios cuya implementación se realiza para sus clientes, sin embargo, en cuanto a los proyectos de *calidad de datos*, no se ha establecido todavía una metodología que permita asegurar su éxito.

Las preguntas a ser resueltas en este proyecto son:

- ¿Cómo definir un modelo de gestión a ser aplicado en los proyectos de calidad de datos?

- ¿Cómo identificar el nivel de madurez de BI dentro de una organización, basándose en el Modelo de Madurez de TDWI?

1.3 Objetivos

1.3.1 Objetivo General

Desarrollar un modelo de Gestión de Calidad de Datos basado en la metodología IBM Data Quality para estandarizar los proyectos de desarrollo de Business Intelligence pertenecientes a la empresa DWConsultware.

1.3.2 Objetivos Específicos

- Identificar las etapas o fases para el proceso de mejora de la calidad de datos y su ciclo de vida.
- Establecer el mecanismo para determinar el nivel de madurez de un proyecto de Business Intelligence basado en el modelo TDWI.
- Definir directrices para la implementación de los procesos de mejora y gestión de calidad de datos utilizando IBM Data Quality.

1.4 Alcance

Este proyecto culmina con la presentación de un Modelo de Gestión de Calidad de Datos para ser aplicado en el portafolio de BI de la empresa DWConsultware.

En el Modelo de Gestión se incluye la definición de las fases y la metodología de un proyecto de calidad de datos, adicionalmente se establecen los niveles de un modelo de madurez para evaluar el cambio en la organización al desarrollar un proyecto de mejora de la calidad de los datos.

Este trabajo no incluye una aplicación del Modelo a una empresa como un caso de estudio, sin embargo esto se incluirá en el segundo proyecto.

CAPÍTULO II

2. MARCO TEÓRICO

2.1 Inteligencia de Negocios

Según Larissa T. Moss (Moss, 2003), Business Intelligence (BI) es una arquitectura integrada de aplicaciones y bases de datos que dan soporte a la toma de decisiones y que facilitan a los usuarios el acceso rápido y oportuno a la información sobre el negocio. Según Howard Dresner (Power, 2007), la Inteligencia de Negocios (BI) se define como el término que agrupa conceptos y métodos que mejoran la toma de decisiones al sustentarla con sistemas que aportan información.

2.1.1 Información

Para poder llegar a un concepto de Información es necesario primero revisar dos conceptos aún más básicos como son:

- **Dato:** representación simbólica (numérica, alfabética, etc.) de un atributo o variable cuantitativa.
- **Metadato:** son datos utilizados para describir *otros* datos.

Es entonces que podemos definir **Información** como un conjunto organizado de datos, o un conjunto de datos que en sí mismos tienen sentido.

Una empresa genera una inmensa cantidad de datos, los cuales deben ser procesados para que se conviertan en información de valor, este es el reto de los proyectos de BI. Los proyectos de BI o Data Warehouse tienden a ser corporativos ya que la iniciativa de BI se genera dependiendo del nivel

de necesidad de la información, tal como ilustra la **Figura 1**, esta necesidad (Otake, 2010) puede ser operacional, táctica o estratégica.

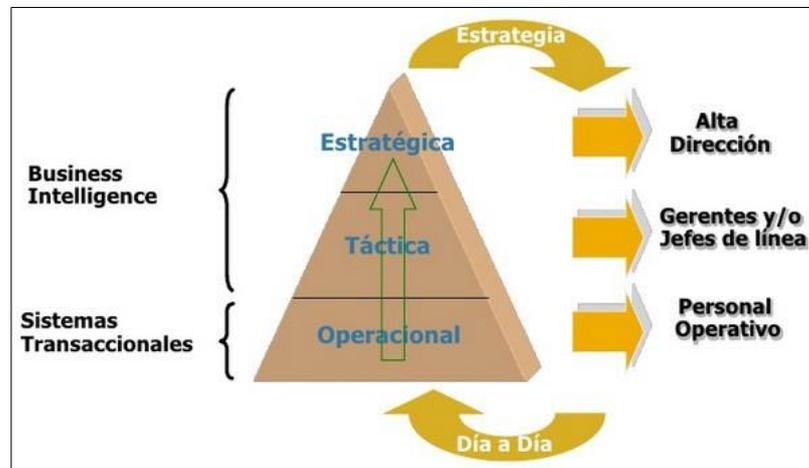


Figura 1: Niveles de Necesidad de la Información

Fuente: (Otake, 2010)

Son tres los niveles de necesidad de la información en una empresa, siendo los dos últimos ámbitos trascendentales para el Business Intelligence:

- **Estratégico:** Desarrolla estrategias y objetivos organizacionales, supervisión estratégica.
- **Táctico:** Desarrolla planes y presupuestos a corto y mediano plazo, especifica procedimientos y objetivos de negocio.

2.1.2 Sistemas de Información

Los Sistemas de Información se ubican de acuerdo al nivel en el que son utilizados dentro de la Organización, la **Figura 2** ilustra los tipos y clasificaciones de los Sistemas de Información; donde estos se clasifican en Sistemas Empresariales (ES) que son utilizados a nivel Operativo y Sistemas de Automatización de Oficinas (OAS) los cuales se usan en los niveles Táctico y Estratégico.



Figura 2: Niveles de Sistemas de Información
Fuente: (Delgadillo, 2011)

En cuanto a los tipos de Sistemas de Información existen los siguientes:

- **OAS:** Sistema de Automatización de Oficinas
 - Hojas de Cálculo.
 - Procesadores de Texto.

- **TPS:** Sistema de Procesamiento de Transacciones.
 - Rol de Pagos
 - Inventarios

- **MIS:** Sistema de Información Gerencial.
 - DataWarehouse
 - Data Mining

- **DSS:** Sistema de Soporte a Decisiones.
 - Modelos de Simulación.
 - Modelos de Optimización.

- **EIS:** Sistema de Información para Ejecutivos.
 - Tableros de control.

2.1.3 Business Intelligence

Los sistemas de Business Intelligence están clasificados de acuerdo a los niveles de una organización, esto se presenta porque los sistemas pueden ayudar a tomar decisiones en cualquier área de una empresa. El **Cuadro 1** indica el detalle de los tipos de BI.

Cuadro 1: Tipos de Business Intelligence

	OBJETIVO	USUARIOS PRINCIPALES	LAPSO DE TIEMPO	DATOS
BI Estratégico	Desarrollar las metas a largo plazo	Ejecutivos y Analistas Negocio	Meses o Años de	Datos Históricos
BI Táctico	Administrar las iniciativas, tácticas para el cumplimiento de las metas estratégicas	Directivos, analistas Gerentes (LOB de managers)	Días, y Semanas (LOB Meses)	Datos o Históricos
BI Operacional	Administrar y optimizar las operaciones diarias del negocio	Analistas, Gerentes (LOB Managers), usuarios procesos operativos	Día de operación (LOB) y	Datos en tiempo real, datos de baja latencia, Datos Históricos

Fuente: (Delgadillo, 2011)

- **El BI Estratégico** es la respuesta al seguimiento de objetivos a largo plazo de las organizaciones. Por su concepción estratégica, su horizonte de análisis es de medio y largo plazo.
- **El BI Táctico** es la aplicación que analiza, compara metas, simula escenarios, buscando información que permita tomar ventaja en el mercado o apoyar decisiones.
- **El BI Operacional** es un enfoque diferente, hacia una nueva generación de sistemas de Inteligencia de Negocios, en las que la toma de decisiones se orienta no solamente hacia decisiones gerenciales sino a la optimización de la operación del negocio, de

manera que la empresa pueda operar más eficientemente y obteniendo mejores beneficios.

El BI requiere aplicar un proceso ETL (Extract, Transform and Load) lo que significa Extraer, Transformar y Cargar información. Dicho proceso permite a las organizaciones manipular los datos desde múltiples fuentes; con el fin de analizar dicha información para apoyar un proceso del negocio. Un proceso ETL consiste en:

- **Extracción:** Obtención de datos de las distintas fuentes tanto internas como externas.
- **Transformación:** Filtrado, limpieza, depuración, homogeneización y agrupación de los datos.
- **Carga:** Organización y actualización de los datos y los metadatos.

2.1.4 Toma de Decisiones

Una empresa busca mantenerse competitiva y marcar la diferencia, por lo que requiere tomar decisiones efectivas, estas decisiones, deben estar orientadas hacia la visión de la empresa y alineadas con los planes estratégicos, basándose en las políticas de la organización. A continuación, se detallan los tipos de decisiones y los niveles organizacionales a los que aplica la toma de decisión.

En el **Cuadro 2** se presentan las características de las decisiones según su nivel de estructuración.

Cuadro 2: Tipos de Decisiones

CARACTERÍSTICA	DECISIÓN ESTRUCTURADA	DECISIÓN SEMI-ESTRUCTURADA	DECISIÓN NO ESTRUCTURADA
<i>Objetivos</i>	Todos definidos con claridad previamente.	Algunos definidos con claridad previamente.	No definidos con anticipación.
<i>Proceso</i>	Definido con claridad previamente.	Definido con claridad previamente pero también intervienen factores externos.	No definido con claridad y detalle de manera previa. Definición muy general.
<i>Criterios</i>	Todos establecidos con claridad previamente.	La mayoría establecidos con claridad previamente.	No establecidos con anterioridad.
<i>Impacto</i>	A corto plazo	A mediano plazo	A largo plazo
<i>Información requerida</i>	estructurada	Semi-estructurada	No estructurada

Fuente: (Delgadillo, 2011)

En el **Cuadro 3** se presentan las características de los niveles de la Organización a los cuales aplica una decisión.

Cuadro 3: Características de las Decisiones

CARACTERÍSTICA	NIVEL OPERATIVO	NIVEL TÁCTICO	NIVEL ESTRATÉGICO
<i>Objetivo</i>	Mantener operación de la empresa	Lograr las metas establecidas por la alta dirección	Definir objetivos, políticas y estrategias
<i>Impacto a largo plazo</i>	Bajo	Medio	Alto
<i>Tipo de decisiones</i>	Estructuradas, a corto plazo, orientadas al presente.	Semi-estructuradas. A mediano plazo, orientadas al presente y al futuro cercano.	No estructuradas, a largo plazo, orientadas al futuro y al largo plazo.
<i>Tipo de información</i>	Detallada, del presente, Estructurada.	Detalle y algunas sumalizaciones, interna y algo de información externa (alcance medio). Semi-estructurada.	Sumarizada, presente e histórica, interna y externa (alcance amplio), herramientas de manipulación y análisis. No estructurada.
<i>Costo</i>	Bajo	Medio	Alto

Fuente: (Delgadillo, 2011)

2.1.5 Arquitectura de un SDIN

Los Sistemas de Inteligencia de Negocios (SDIN) utilizan datos originados en otros sistemas, como sistemas de Gestión de Relación con Clientes (CRM), Planificación Empresarial de Recursos (ERP) y de Administración de Cadena de Abastecimiento (SCM), para una vez procesados mediante ETL alimentar un sistema de Data Warehouse. Finalmente los usuarios acceden a la información generada a través de herramientas de reporte, Data Mining o análisis OLAP.

La **Figura 3** indica la Arquitectura general de un Sistema de Inteligencia de Negocios de Negocios.

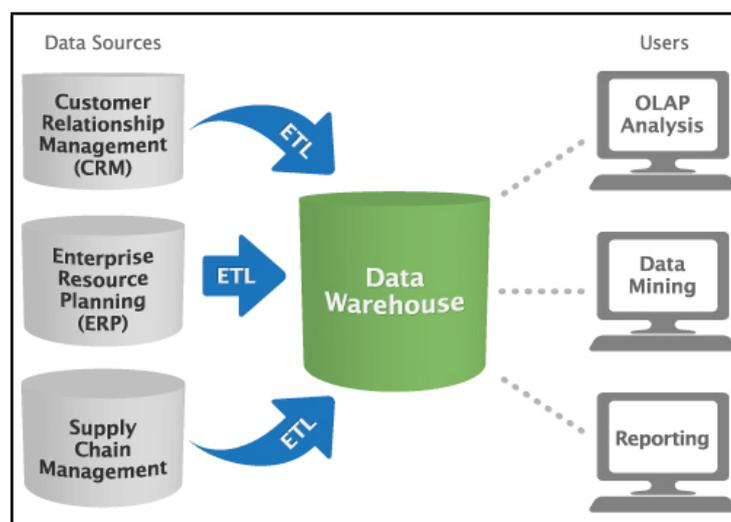


Figura 3: Arquitectura de un Sistema de Inteligencia de Negocios

Fuente: (Trejo, 2011)

2.2 Modelos de Madurez de BI

Los Modelos de Madurez de Inteligencia de Negocios son usados para describir, explicar y evaluar ciclos de vida y crecimiento (Rajterič, 2010) y además tienen los siguientes componentes (Lara, Programa de Business Intelligence, Business Intelligence Strategy: Definición, implementación y seguimiento, 2012):

- Entre 3 y 6 niveles.
- Concepto ilustrativo de cada nivel.
- Definición de áreas de proceso.
- Definición de elementos para áreas de proceso.

Estos modelos describen el uso de BI de acuerdo a la evolución del negocio y la concepción de BI dentro de la organización.

En general, los niveles descritos por estos modelos van de un nivel inicial donde no hay una concepción o uso claro de BI hasta el nivel más alto donde el uso de BI es crucial para el desempeño y desarrollo de la organización.

Existen algunos modelos de madurez, por ejemplo:

- Modelo de madurez de HP:
 - Etapas: Operación, Mejora, Alineación (con objetivos estratégicos), Empoderamiento, Excelencia.
- Modelo de madurez de Gartner:
 - Etapas: Sin conocimiento, Táctico, Enfocado, Estratégico, Compenetrado.
- Modelo de madurez AMR Research's Business Version 2:
 - Etapas: Nivel 1 Reactivo, Nivel 2 Anticipación, Nivel 3 Colaboración, Nivel 4 Organizativo.
- Modelo de madurez TDWI, que será explicado en detalle más adelante en este trabajo.

2.2.1 Modelo de Madurez TDWI

The Data Warehousing Institute, define un modelo de madurez de BI que comprende 3 niveles, 6 etapas y 2 puntos críticos, la **Figura 4** muestra la curva de adopción de BI de acuerdo a este modelo.

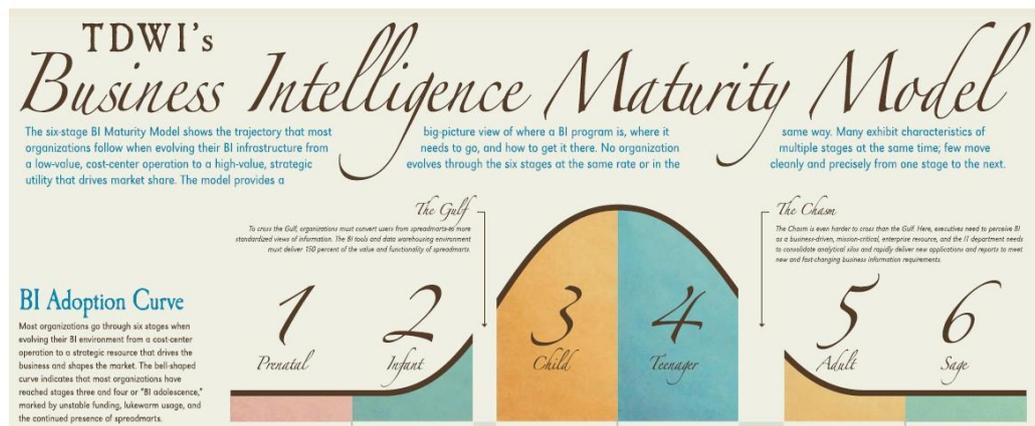


Figura 4: Etapas de Madurez de BI

Fuente: (Eckerson, TDWI Poster: Business Intelligence Maturity Model, 2010)

Este modelo fue originalmente definido por Wayne Eckerson en el 2004 y se enfoca principalmente en los aspectos técnicos a ser evaluados.

Cada una de las etapas comprende además:

- Infraestructura o arquitectura.
- Alcance.
- Tipo de sistema
- Tipo de reportes.
- Usuarios.
- Enfoque de BI.
- Percepción.

2.2.1.1 Etapas del modelo TDWI

Las etapas del modelo TDWI son (Rajterič, 2010):

- 1. Pre-natal:** es la etapa inicial del modelo de madurez, en esta no existe una estructura centralizada de datos, sino más bien una serie de fuentes de datos dispersas enfocadas en tareas o necesidades muy específicas. Puede existir un cierto nivel de estandarización. La administración se hace en base a reporte.
- 2. Infancia:** en esta etapa existe un nivel de estandarización que se basa en los acuerdos entre usuarios, los datos son obtenidos de las fuentes existentes por el departamento de TI (Tecnologías de la Información) y ensamblados en conjuntos relativamente organizados antes de ser pasados a los usuarios. Uso de spreadmarts.
- 3. Niñez:** los trabajadores del área de conocimiento se unen a los usuarios de BI, la información se recolecta y usa a nivel departamental. Las compañías usualmente compran sus primeras herramientas interactivas de reporte y pueden analizar tendencias y datos históricos. Uso de Datamarts.
- 4. Adolescencia:** la necesidad de estándares es reconocida y se empieza a usarlos, la gestión de BI es realizada por miembros de los diferentes departamentos bajo el liderazgo de un jefe de BI, se desarrollan soluciones de software de BI.
- 5. Edad Adulta:** BI se transforma de una herramienta táctica a una estratégica, es ahora el centro de los sistemas de TI, los procesos son monitoreados usando paneles de control.
- 6. Sabiduría:** en este nivel las capacidades de BI se convierten en servicios de negocio, el uso de BI se fomenta y distribuye en la

organización a través de Centros de Excelencia (COE – Centers of Excellence), existe un grupo de gestión de BI centralizado a cargo del data warehouse y el desarrollo de soluciones customizadas se deja a los grupos distribuidos.

2.2.1.2 Puntos críticos: el Golfo y el Abismo

El Modelo TDWI define dos puntos críticos: el Golfo y el Abismo, los cuales separan las etapas de Infancia con Niñez y Adolescencia con Edad Adulta.

El Golfo es el punto crítico que separa las etapas de Infancia y Niñez, el paso de este punto supone un cambio de percepción de BI por parte de la gerencia de la empresa en el cual se toma conciencia de la importancia de BI así como del uso de las herramientas adecuadas para la gestión de la misma. En términos generales se puede decir que el paso de este punto no implica una mayor complejidad, ya que las ventajas de pasar a la siguiente etapa son fácilmente perceptibles.

El segundo punto crítico se denomina el Abismo y presenta una mayor dificultad para ser superado, este se debe a que en esta etapa al ya contar con las herramientas necesarias para la gestión de BI, así como la conciencia en la organización de las ventajas de la adopción de estos conceptos, el paso consiste en aplicar las herramientas de una forma efectiva a fin de realmente utilizar los datos existentes en la toma de decisiones de negocio por parte de usuarios ejecutivos, lo cual es justamente el enfoque de la etapa de la Edad Adulta.

2.3 Calidad de Datos

La calidad se puede definir como el “Grado de Excelencia” o la ausencia de defectos, esto nos indica que la calidad no es algo absoluto sino que

existe en grados o niveles y que esta se obtiene mediante el cumplimiento de especificaciones.

La calidad no reside de forma intrínseca en el producto y solo puede ser juzgada en relación a las expectativas del cliente usando el producto. Es decir un producto es de calidad siempre y cuando cumpla con las expectativas del cliente.

Para ser considerados de calidad, los datos deben tener las características que se indican en la **Figura 5**.



Figura 5: Características de datos de calidad

Fuente: (TDWI T. D., TDWI Data Quality Fundamentals)

Existen 8 características o dimensiones de la calidad de datos:

- **Consistentes:** Los datos están libres de conflictos internos.
- **Reales:** Los datos representan la realidad.
- **Fáciles de usar:** Los datos son accesibles, entendibles y navegables.

- **Granulares:** Los datos son mantenidos y presentados al nivel de detalle correcto para cumplir los requerimientos.
- **Íntegros:** Los datos son correctos y completos.
- **Completos:** Todos los datos requeridos están disponibles.
- **Precisos:** Los datos son tan precisos como requeridos.
- **Oportunos:** Los datos son tan actuales como es necesario y son guardados hasta que ya no se los necesita.

Estas características se muestran en la **Figura 6**.

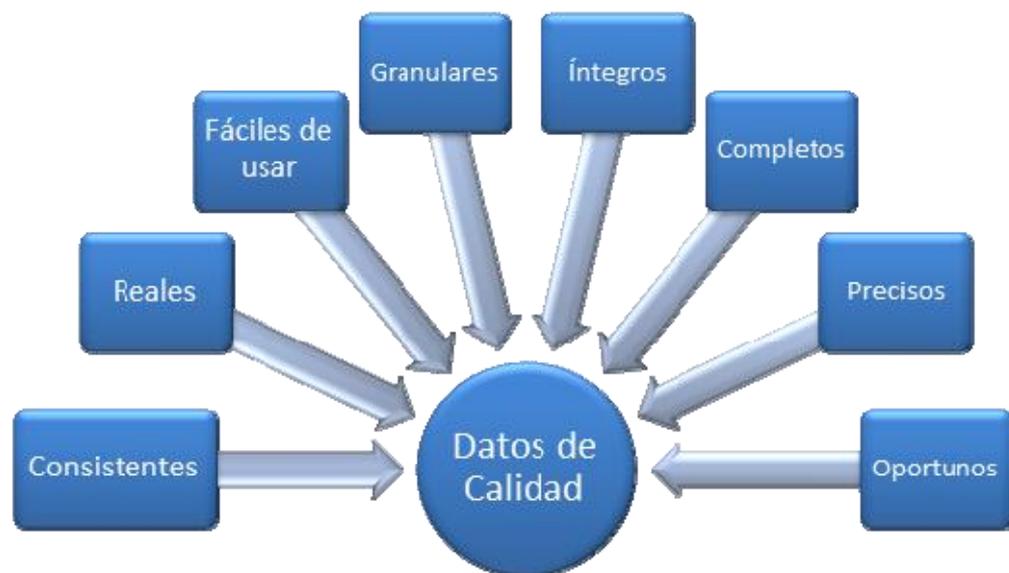


Figura 6: Características de la Calidad de Datos

Fuente: (TDWI T. D., TDWI Data Quality Fundamentals)

Cada objeto de datos, entidad, atributo, elemento de datos, archivos, tablas y columnas- necesita tener una definición completa. Buenas prácticas de definición aseguran consistencia, eliminan la confusión, mejoran la comunicación, permiten la consolidación de datos y mejoran la calidad de datos.

El diseño de datos es una actividad crítica en la gestión de la calidad de datos. Buenas prácticas de diseño ayudan a mejorar la calidad de muchas maneras, mientras un mal diseño crea muchos problemas. Un buen diseño crea estructura de datos que son adecuados para el propósito de los datos.

2.4 Gestión de Calidad de Datos

La Gestión de calidad de datos (DQM) (IBM, Data Quality Management (DQM), 2011) es un proceso que comprueba que los datos tengan los valores necesarios, tipos de datos y códigos válidos. También se puede utilizar la gestión de calidad de datos para corregir los datos proporcionando valores por omisión, añadiendo nuevos códigos y dando formato a números y fechas. La Gestión de calidad de datos realiza, por ejemplo, procesos de limpieza y estandarización de nombres y direcciones, que sirven para optimizar y mejorar la calidad de los datos.

El objetivo de la gestión de calidad es producir un entendimiento de los datos que puede ser traducido en acciones para mejorar la calidad de los mismos.

Los procesos de gestión de calidad de los datos están incluidos dentro del proceso ETL utilizado para transformar los datos y cargarlos al Data Warehouse.

2.4.1 Ciclo de Vida del Proyecto de Calidad de Datos

Para la implementación de un proyecto de gestión de calidad de datos se debe considerar las etapas del ciclo de vida que se presentan en la **Figura 7**.

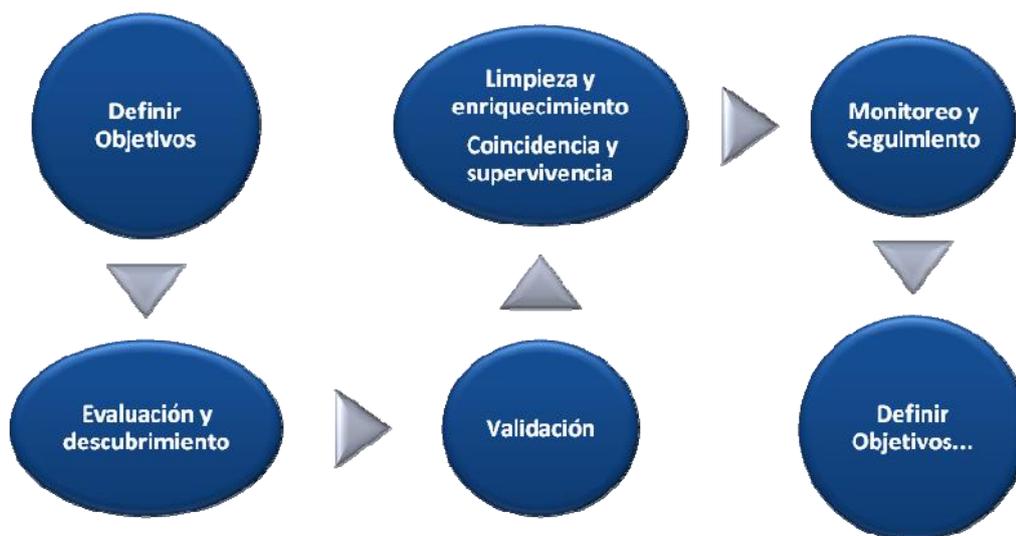


Figura 7: Ciclo de Vida de un Proyecto de Calidad de Datos
Elaborado por: Christian Andrade, David Madrid

2.4.1.1 Definir Objetivos

Los objetivos de calidad de los datos son criterios que definen los tipos apropiados de datos y especifican los niveles tolerables de error. También se utilizan para describir los indicadores de calidad de los datos, tales como los parámetros de precisión, exactitud, representatividad, comparabilidad e integridad.

La finalidad de la definición de objetivos es detallar los mecanismos para obtener los datos en un proceso lógico. Los pasos para establecer los objetivos son los siguientes:

- **Definición del problema:** se describe cada uno de los recursos y procesos que intervienen en el proyecto para analizarlos y así delimitar el alcance de la solución propuesta.
- **Identificación de decisiones:** consiste en identificar las directrices que se aplicarán en cuanto a manejo de los datos iniciales.
- **Identificación de insumos para las decisiones:** identifica las fuentes de datos de entrada y los estándares que se desean aplicar.

- **Definición de los límites del estudio:** plantea el alcance de las correcciones y/o modificaciones que se pueden aplicar a los datos de entrada.
- **Desarrollo de reglas de decisión:** genera las reglas necesarias en base a las decisiones, insumos y límites para alcanzar los objetivos planteados.
- **Especificación de límites de tolerancia para los errores de decisión:** determina los parámetros entre los cuales los datos procesados serán considerados válidos.
- **Optimización del diseño para la obtención de datos:** identifica los puntos en el diseño que son susceptibles de ser modificados para optimizar el proceso de obtención de datos.

2.4.1.2 Evaluación y descubrimiento

En esta etapa se identifican relaciones que se encuentren en la información mediante el análisis automatizado de los datos y sus relaciones, en esta etapa se persiguen los siguientes puntos:

- Identificar los elementos de datos dentro de la fuente de datos.
- Identificar relaciones entre elementos de datos y entidades de negocio dentro de la fuente. (Ej. Cliente, Factura)
- Identificar la lógica compleja que relaciona varias fuentes.

2.4.1.3 Validación

Una vez terminadas las primeras dos etapas la organización está en capacidad de establecer reglas que permitan medir la calidad de los datos, estas reglas se constituyen en los insumos para las siguientes etapas.

La **Figura 8** describe la definición de reglas, luego de la evaluación se obtiene una medida común por ejemplo el campo *tax-id* que tiene varios nulos, en función de esta medida la organización define una regla de datos que no permita que el campo *tax-id* tenga nulos. Mediante esta regla se puede definir un indicador con un umbral de 95% y esta información se presenta en un Tablero o Reporte para el monitoreo de los datos.

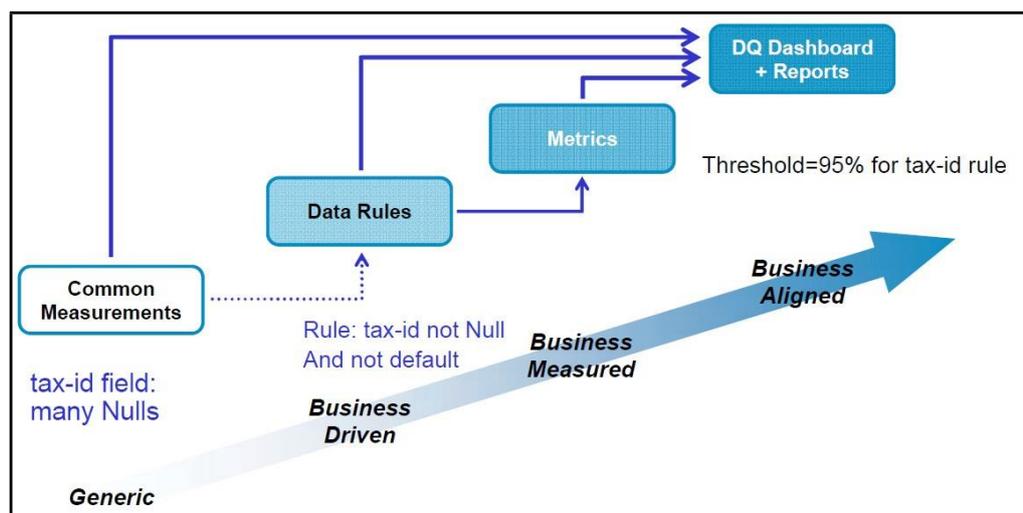


Figura 8: Definición de Reglas

Fuente: (IBM, Data Quality Bootcamp, 2012)

Las reglas de datos son definidas por los responsables del negocio, a continuación, se detallan ejemplos de reglas de datos:

- El campo *Género* debe estar lleno con uno de los valores aceptados de la lista: Masculino, Femenino.
- El *Número de Cédula* debe ser una cadena con el siguiente formato 999999999-9.

2.4.1.4 Limpieza y enriquecimiento - Coincidencia y supervivencia

Consiste en diseñar y probar un conjunto de operaciones de limpieza de datos y búsqueda de coincidencias que se conocen aplicando reglas definidas para mejorar la calidad de los datos. La información se extrae del sistema de origen, se mide, limpia, enriquece, consolida y carga en el sistema de destino. Durante esta etapa se realizan las siguientes tareas:

1. Investigación de datos: Comprender totalmente la información.
2. Estandarización de los datos: Limpiar por completo la información.
3. Coincidencia de los datos: Crear claves para identificar relaciones entre la información.
4. Supervivencia de los datos: Crear la mejor vista disponible de información relacionada.

2.4.1.5 Monitoreo y Seguimiento

En esta etapa se establecen indicadores de calidad de datos alineados a los objetivos del negocio que permitan descubrir problemas en los datos y establecer un plan de mejora.

Gracias a los indicadores y las reglas de calidad de datos desarrolladas y desplegadas desde la etapa de evaluación de la calidad de los datos se puede revisar la calidad de la información y hacer un seguimiento de esta, en el repositorio de datos a lo largo del tiempo. Al finalizar esta etapa, y dependiendo del resultado, se empieza de nuevo el ciclo definiendo objetivos para mejorar la calidad de la información.

2.5 Modelos de Gestión de Calidad de Datos

El objetivo de los Modelos de Gestión de Calidad de Datos es diseñar datos para que estos cumplan las características que se han enunciado previamente en este trabajo, además de permitir que los datos sean usados de una forma efectiva por usuarios que no siempre están familiarizados con estos (Wang, Kon, & Madnick, 1998).

Los Modelos de Gestión de Calidad de Datos se enfocan en cuatro aspectos:

- **Aplicación:** El propósito para el cual los datos son recolectados.
- **Recolección:** Los procesos por medio de los cuales los datos son acumulados.
- **Almacenamiento:** Procesos y sistemas usados para archivar los datos.
- **Análisis:** Los procesos usados para transformar datos en información a ser utilizada por una aplicación.

Cada herramienta de ETL tiene definido un modelo de implementación de proyectos de calidad de datos. A continuación, se listan algunos de los modelos que se pueden encontrar en la red:

- **SAS Data Quality** (SAS Community, 2014)
- **SAP Data Quality** (SAP, 2013)
- **Informatica Data Quality** (Informatica, 2014)

CAPÍTULO III

3. MODELO DE GESTIÓN

3.1 Modelo de Gestión de Calidad de Datos de IBM

Este modelo se enfoca en responder cuatro preguntas básicas (IBM, Information Integration and Governance, 2014):

- ¿Qué datos son correctos?
- ¿Tenemos la información de contexto necesaria para tomar la mejor decisión?
- ¿Están los datos sensibles debidamente protegidos?
- ¿Existe la confianza para actuar en base al entendimiento actual?

Para responder estas preguntas se crea un programa de Gobierno de Datos que consta de 6 pasos (Martínez, 2012):

- 1) Establecer metas: Sentencias principales que guían la operación y desarrollo de la cadena de suministro de información.
- 2) Definir métricas: Conjunto de medidas usadas para evaluar la efectividad del programa y los procesos de gobierno asociados.
- 3) Tomar decisiones: La estructura organizacional y el modelo de cambio ideológico para analizar y crear políticas de decisión.
- 4) Comunicar políticas: Herramientas, habilidades y técnicas usadas para comunicar decisiones políticas a la organización.

- 5) Medir resultados: Comparar resultados de las políticas con las metas, entradas, modelos de decisión y comunicación para proveer constante retroalimentación sobre la efectividad de la política.
- 6) Auditar: Herramienta usada para comprobar permisos, accesos, cambios en los datos.

El modelo de gestión de calidad de datos de IBM requiere de tres disciplinas para obtener resultados, sin embargo, se soporta en dos disciplinas: Arquitectura de datos, Clasificación y metadatos. Los habilitadores de la organización, en este caso su representante, la estructura organizacional y sus políticas aumentan la seguridad y privacidad de los datos, que permiten a futuro la auditoría de los mismos. Todo este ciclo se ilustra en la **Figura 9**.



Figura 9: Ciclo del modelo de gestión de calidad de datos

IBM Fuente: (IBM, Data Quality Bootcamp, 2012)

3.2 Definición del Proyecto de BI

Los proyectos de Business Intelligence se caracterizan por su complejidad y sólo unos pocos tienen un éxito rotundo a corto plazo. Parte

de la dificultad está en la intervención de múltiples factores tanto tecnológicos como no tecnológicos. En concreto, es necesario diferenciar estos tres aspectos o sub-proyectos (Urquiza, 2009):

- **Definición:** Es la etapa en la que se identifican los conceptos de negocio que se desea analizar.
- **Datawarehouse:** El DWH es el repositorio de datos donde se centralizará toda la información corporativa y será la fuente para todas las consultas y necesidades analíticas.
- **Herramientas:** Existen en el mercado diferentes aplicaciones de BI y cada solución tiene características y necesidades diferentes.

Un proyecto de BI tiene dos componentes principales:

- **Ambiente de Data Warehouse:** Donde el equipo de desarrollo pasa de 60 a 80 por ciento del tiempo de proyecto. El trabajo consiste en extraer, limpiar, modelar, transformar, transferir y cargar datos de transacciones de uno o más sistemas operacionales en una base de datos centralizada.
- **Ambiente Analítico:** El entorno de análisis, que es el dominio de los usuarios de negocios, quienes utilizan herramientas analíticas para consulta, reporte, análisis, minería, visualización, y, sobre todo, actuar en función de los datos del DWH.

Estos ambientes se muestran en la **Figura 10**.

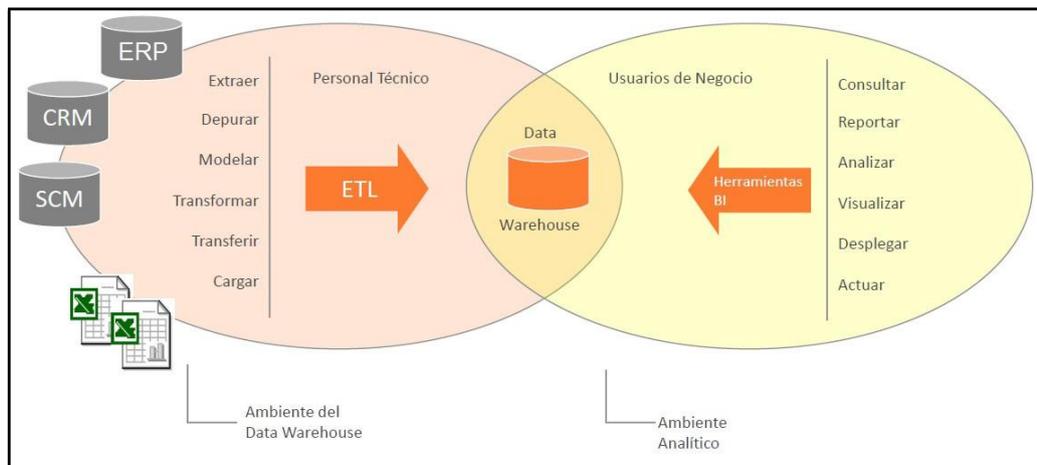


Figura 10: Componentes del BI

Las soluciones de BI o herramientas se agrupan de acuerdo a su función básica:

- a) Reporteo:** Informes predefinidos que se enfocan en la actividad pasada. Este tipo de solución es ideal en situaciones en donde las preguntas están predeterminadas y la estructura de la respuesta es muy bien conocida con anticipación.
- b) Análisis:** Las funcionalidades de este tipo de solución permiten que los usuarios respondan preguntas de manera autónoma conforme éstas van surgiendo espontáneamente al analizar una situación. La posibilidad de responder preguntas ad hoc es ideal para los analistas y otros usuarios de negocio que necesitan facilidad de acceso a los datos, capacidad de autoservicio, interacción rápida y directa y finalmente flexibilidad.
- c) Planeación:** Las funcionalidades de este tipo de solución permiten realizar la planificación corporativa de las organizaciones. Responde a la pregunta ¿Qué debemos hacer?

d) Monitoreo: Seguimiento al desempeño, eventos, tablas, graficas, mapas. Permite que los usuarios puedan analizar y dar seguimiento a los indicadores de desempeño de los cuales ellos son responsables.

Estas soluciones son estructuradas en tableros personalizados para facilitar el análisis visual y gráfico para comunicar información de manera rápida e intuitiva, también permiten identificar con alarmas eventos que demandan la atención del usuario.

e) Analíticas Avanzadas: Análisis estadísticos o de negocio para identificar tendencias, patrones y excepciones.

Estos tipos de soluciones se muestran en la **Figura 11**.



Figura 11: Tipos básicos de Soluciones de BI

Fuente: (Eckerson, What CIO's Should Know to Ensure BI Success, s.f.)

Según los conceptos de negocio que se desee analizar, es decir de acuerdo al resultado de la etapa de Definición es que se debe escoger el tipo de Herramienta necesario de entre los que se han descrito anteriormente.

3.2.1 Ciclo de vida del proyecto BI

Para asegurar el éxito de un proyecto de BI se pueden estructurar en cinco fases, cada una con sus objetivos y procedimientos como se detalla en la **Figura 12**.

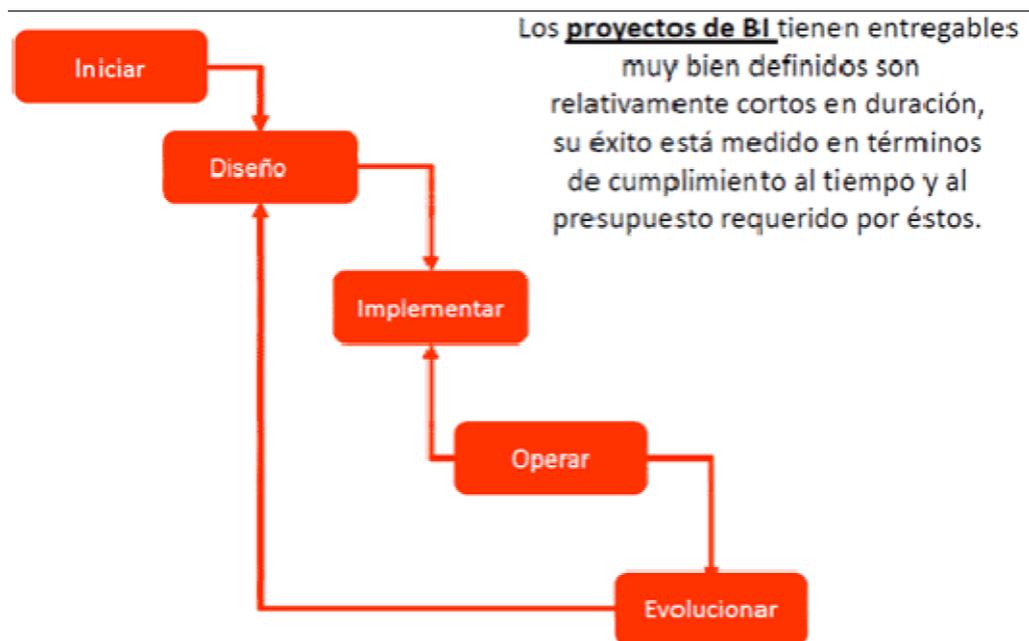


Figura 12: Ciclo de vida de un proyecto BI
Fuente: (Lara, Módulo VII: Business Intelligence Strategy, 2011)

1) Iniciar o Analizar: Es la fase en la cual se determinan y analizan todas las necesidades del proyecto, responde a dos preguntas clave:

- a. ¿Qué información necesitamos para desempeñar nuestro actual trabajo? y,
- b. ¿Qué información nos gustaría tener para poder hacer mejor nuestro trabajo?

2) Definición del modelo de datos (Diseño): El modelo de datos es una representación conceptual de las estructuras que maneja la empresa y sus relaciones.

3) Implementar: consiste en construir la infraestructura lógica que dará cabida al modelo de datos, esta es una fase puramente técnica donde se seleccionan además las herramientas de hardware y software, si es que estas no fueron previamente definidas por requerimientos del negocio.

Parte de esta fase es también la realización o desarrollo de la capa de presentación o de los informes, cuadros de mando y otros elementos de visualización de información.

4) Operar: esta fase se inicia con la capacitación a los usuarios en el manejo del sistema por parte de un equipo conocedor del funcionamiento de la solución de BI implementada, una vez que la solución ha arrancado es necesario dar soporte hasta que se logre el punto de estabilidad de operación.

5) Evolución: consiste en repetir el ciclo de vida modificando el diseño para adaptarse a nuevas necesidades del negocio.

3.3 Evaluación de Niveles de Madurez

El modelo de madurez TDWI de BI pretende evaluar la madurez a través de un cuestionario de negocio y TI, el cuestionario está estructurado en cuatro temas:

1) Control Local vs. Estándares Empresariales: Compartir la información de forma ágil cumpliendo con los estándares de la organización.

2) Uso de BI: Tener una herramienta fácil de usar y flexible que permita monitorear los objetivos estratégicos.

3) Preguntas que contesta BI: Contar con un modelo en la herramienta que automatice la toma de decisiones.

4) Valor de Negocio y ROI: Maximizar el valor del BI, el valor generado excede notablemente los costos derivados de la herramienta.

Como se indicó en el 2, el modelo de madurez TDWI establece que la mayoría de empresas pasan por seis etapas de maduración, desde que el BI se percibe como un gasto que genera poco valor hasta convertirse en un recurso estratégico que dirige al negocio. Los resultados han mostrado que la mayoría de las empresas que tienen BI se encuentran en la etapa 3 o 4, principalmente debido a que tienen:

- Una inversión inestable
- Un bajo o moderado uso de la herramienta
- Una arquitectura técnica basada en spreadmarts.

La etapa actual dentro de la que se ubica la empresa dentro del modelo TDWI se obtiene mediante un cuestionario de 40 preguntas, dividido en 8 perspectivas que son:

1) Alcance (Scope): permite identificar el alcance de los proyectos de BI en relación con la planificación estratégica de la organización.

2) Patrocinio (Sponsorship): analiza el nivel de compromiso por parte de los administradores de negocio (áreas de gerencia media y alta) hacia los proyectos de BI.

3) Financiamiento (Funding): identifica la disponibilidad de las fuentes de financiamiento de los proyectos de BI.

4) Valor (Value): analiza la eficacia de los proyectos de BI para cumplir con los requisitos de las áreas de negocio y gerencia.

5) Arquitectura (Architecture): analiza el cumplimiento de estándares en la arquitectura de BI y su nivel de desarrollo.

6) Datos (Data): verifica el nivel al cual los datos provistos por las soluciones de BI cumplen con los requerimientos de las áreas de negocio y gerencia.

7) Desarrollo (Development): investiga qué tan efectivo es el equipo de BI en el desarrollo y despliegue de las soluciones.

8) Entrega de Información (Delivery): comprueba si los formatos de salida de los proyectos de BI están acorde a los requerimientos de los usuarios.

Las preguntas específicas del cuestionario se detallan en el ANEXO B CUESTIONARIO TDWI. Los resultados de la evaluación se muestran en forma de un gráfico que indican el puntaje en cada perspectiva analizada como se muestra en la **Figura 13**.

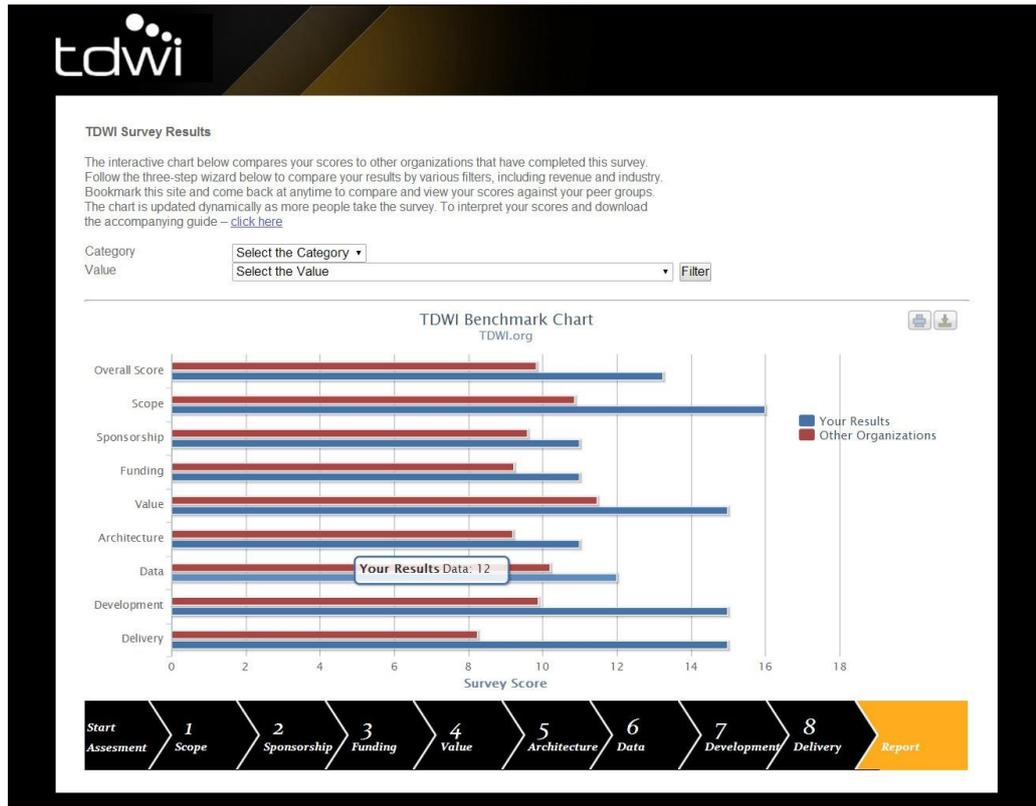


Figura 13: Ejemplo de evaluación del modelo de madurez TDWI
Fuente: (TDWI T. D., TDWI Survey Results, 2013)

La **Figura 14** proporciona una visión general de las 5 etapas de Adopción del Modelo de Madurez de TDWI, las cuales se superponen a las etapas descritas en la **Figura 4** ya que las etapas *Prenatal* e *Infancia* se agrupan en una sola etapa de adopción denominada “Inexistente”. La puntuación es por cada perspectiva y existe un puntaje total.

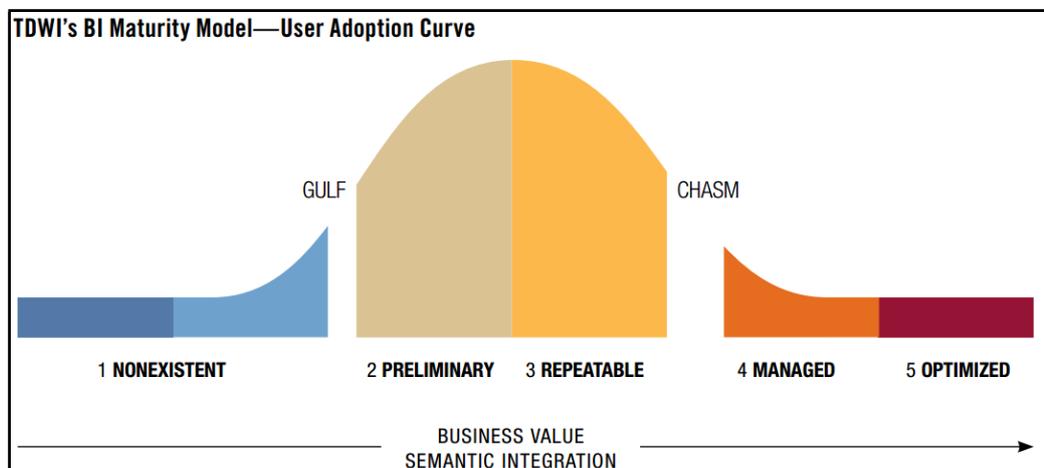


Figura 14: Etapas del modelo de madurez TDWI
Fuente: (TDWI, Evaluación del Modelo de Madurez de BI de TDWI, 2012)

Resultados

Cada pregunta de las ocho categorías cuenta con cinco opciones de respuesta, cada respuesta representa un nivel diferente en el Modelo de madurez de BI de TDWI y se pondera de 1 a 5, donde 1 representa la etapa *Inexistente* y 5 que representa la fase *Optimizada*. Cuando se toma la encuesta se obtiene una puntuación para cada categoría, que representa la suma de las ponderaciones de las cinco preguntas de esa categoría. A continuación, se recibe un puntaje total, que representa la suma de las ponderaciones de todas las categorías, dividido por ocho categorías (TDWI, Evaluación del Modelo de Madurez de BI de TDWI, 2012).

Resultados por perspectiva

Para cada perspectiva, recibirá una puntuación entre 5 y 25 y una designación de su nivel de madurez. La **Tabla 1** muestra las correlaciones entre los puntajes y niveles de madurez para las perspectivas.

Tabla 1: Escala de madurez de TDWI

Puntaje	Etapas
5 a 7	Inexistente
8 a 12	Preliminar
13 a 17	Repetible
18 a 22	Administrado
23 a 25	Optimizado

Fuente: (TDWI, Evaluación del Modelo de Madurez de BI de TDWI, 2012)

Puntaje Total

La madurez de BI se calcula mediante la suma de los valores ponderados de todas las preguntas y dividiendo por ocho. La puntuación total de los encuestados se calcula de la misma manera: se suman las puntuaciones ponderadas de todas las preguntas para todos los encuestados y se divide por el número total de los encuestados, dividido por

ocho. Para determinar el nivel promedio de madurez, se debe asignar calificación a la etapa correspondiente en la **Tabla 2**.

Golfo y Abismo: Se puede superponer el Golfo y Abismo en la escala existente para reforzar la idea de que en ciertos momentos la madurez de BI se moverá a través de estos dos puntos críticos. La **Tabla 2** incluye las calificaciones de las pruebas del Golfo y abismo.

Tabla 2: Escala de madurez de TDWI con Golfo y Abismo

Puntaje	Etapas
5 a 7	Inexistente
6 a 9	<i>El Golfo</i>
8 a 12	Preliminar
13 a 17	Repetible
15 a 19	<i>El Abismo</i>
18 a 22	Administrado
23 a 25	Optimizado

Fuente: (TDWI, Evaluación del Modelo de Madurez de BI de TDWI, 2012)

El **Cuadro 4** proporciona un resumen rápido de los niveles de madurez para cada perspectiva se debe marcar con color Rojo la madurez inicial y con Azul la madurez después de aplicar un proyecto de BI.

Cuadro 4: Niveles de Madurez por Perspectiva

Etapas/ Perspectivas	No Existe	Preliminar	Repetible	Administrado	Optimizado
Alcance (Scope)	Individual	Departamental	Divisional	Empresa	Inter-empresas
Patrocinio (Sponsorship)	No existe	↔	Poco comprometida y responsable	↔	Muy comprometida y responsable
Financiamiento (Funding)	Ninguno	Presupuesto Departamental	Presupuesto Divisional	Presupuesto de TI Corporativo	Autofinanciamiento
Valor (Value)	Spreadmarts	Táctico	Misión Crítica	Estratégico	Competitivo Diferenciador
Arquitectura (Architecture)	Centro de Costo	Datamarts no integrados	Data warehouses no integrados	Data warehouse centralizado	BI a través de arquitectura orientada a servicios
Datos (Data)	No confiables, No oportunos, No completo	↔	Poco confiable, oportuna, completa	↔	Totalmente confiable, oportuna, completa
Desarrollo (Development)	Procesos no estandarizado s	↔	Pocos procesos estandarizado s	↔	Procesos totalmente estandarizados
Entrega de Información (Delivery)	Ver informes estáticos	Analizar tendencias	Procesos monitoreados	Pronostico de resultados	Procesos automatizados

Fuente: (TDWI, Evaluación del Modelo de Madurez de BI de TDWI, 2012)

3.4 Gestión de Calidad de Datos

Los datos que controlan los sistemas empresariales de hoy en día a menudo provienen de distintas fuentes y de estructuras de datos dispares. A medida que las organizaciones van creciendo, retienen sistemas de datos antiguos y los amplían con sistemas nuevos y mejorados. Llega a ser difícil gestionar y utilizar los datos, y es prácticamente imposible determinar una imagen clara de un cliente, un producto o una tendencia de compra. Para resolver este problema las organizaciones deben optar por la gestión de calidad de datos.

Para aplicar la gestión de calidad de datos, es preciso definir reglas de calidad de datos. Estas reglas reparan, limpian y estandarizan valores de datos de entidades de entrada, por ejemplo formatear correctamente los números, identificar y corregir errores e inconsistencias y completar los datos.

3.4.1 Fases de la Gestión de Calidad de Datos

Las Fases de la Gestión de Calidad de Datos se muestran en la Figura 15 y están alineadas con el Ciclo de Vida del Proyecto de Calidad de Datos (IBM, InfoSphere QualityStage, 2012)

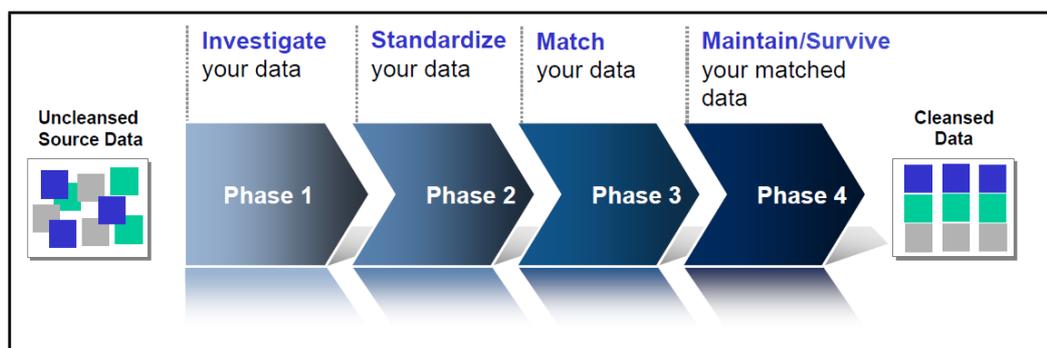


Figura 15: Fases de la Gestión de Calidad de Datos

Fuente: (IBM, Data Quality Bootcamp, 2012)

Se pueden utilizar herramientas como IBM InfoSphere QualityStage para tener un conocimiento claro de los datos y mejorar la calidad de los mismos mediante la ejecución de las etapas tareas:

3.4.1.1 Investigación de los datos

Tiene como objetivo entender la naturaleza y el alcance de las anomalías de datos y permitir una limpieza y una coincidencia de datos más eficaces. Las funciones de investigación dan a la organización una visibilidad completa de la condición de los datos en cualquier momento. Los problemas de datos en orígenes existentes se pueden identificar y corregir antes de que se dañen nuevos sistemas.

En el proceso de investigación se analizan campos de formato libre, se cuentan los valores exclusivos y se clasifica o asigna un significado empresarial a cada ocurrencia de un valor dentro de un campo. Adicionalmente permite descubrir anomalías potenciales, discrepancias de metadatos y prácticas empresariales no documentadas. Con la investigación se consiguen estos objetivos:

- Se descubren tendencias, posibles anomalías, discrepancias en los metadatos y prácticas empresariales no documentadas.
- Se identifican valores predeterminados o valores que no son válidos.
- Se revela la terminología común.
- Se verifica la fiabilidad de los campos propuestos como criterios de coincidencia.

La etapa de investigación toma una única entrada, que puede ser un enlace de cualquier conector de base de datos, de un archivo plano o un conjunto de datos, o de cualquier etapa de proceso. Las entradas de esta etapa pueden ser de longitud fija o variable. En la **Figura 16** se muestra un ejemplo del análisis a un dato efectuado en esta etapa.

Example: 123 St. Virginia St.				
<p>■ Parsing:</p> <ul style="list-style-type: none"> ▪ Separates multi-valued fields into individual pieces 	123	St.	Virginia	St.
<p>Lexical Analysis:</p> <p>Determines business significance of individual pieces</p>	Number	Street Type	Alpha	Street Type
	123	St.	Virginia	St.
<p>Context Sensitive:</p> <p>Identifies various data structures and content</p>	Number	Street Name		Street Type
	123	St. Virginia	St.	

Figura 16: Ejemplo de Análisis de Datos

Fuente: (IBM, Data Quality Bootcamp, 2012)

3.4.1.2 Estandarización de los datos

La creación de una vista estandarizada de datos permite a la organización mantener vistas exactas de entidades clave tales como cliente, socio o producto. Los datos de varios sistemas se reformatean para garantizar que el contenido y el formato especificado son correctos. Las reglas de estandarización se utilizan para crear una representación coherente de los datos, la Figura 17 y la Figura 18 muestran un ejemplo de estandarización de datos.

La etapa de estandarización aplica representaciones coherentes, corrige los errores de escritura e incorpora estándares de la empresa o del sector. Da formato a los datos, coloca cada valor en un único campo de dominio y transforma los datos en un formato estándar.

Input Data:

Vehicle Year, Make and Model

1996 LINCOLN MARK VIII
 CHEVEROLETTE MALIBU '97
 '01 DODGE RAM 25K- DSL
 1999 NISSAN MAXIMA
 EXPLOROR, FORD 02
 98 CHEVY SUBURBAN 1500
 2001 VOLVO C70 CONVERT

Go from this ...

... to this!

Standardized Data:

Make	NYSIIS	RScundex	Model	NYSIIS	RScundex	Model #	Model YR	Additional Vehicle Data
LINCOLN	LANCALN	N425	MARK	MARC	R321	8	1996	
CHEVROLET	CAFRALAT	T461	MALIBU	MALAB	U145		1997	
DODGE	DAG	E233	RAM	RAN	M600	2500	2001	DIESEL
NISSAN	NASAN	N250	MAXIMA	MAXAN	A525		1999	
FORD	FAD	D610	EXPLORER	EXPLARAK	R641		2002	
CHEVROLET	CAFRALAT	T461	SUBURBAN	SABARBAN	N161	1500	1998	
VOLVO	VALV	0141	C	C	C000	70	2001	CONVERTIBLE

Figura 17: Ejemplo de Estandarización de Datos I
 Fuente: (IBM, Data Quality Bootcamp, 2012)

Standardization parts

Input file:

<i>Operation Work Instruction</i> WING ASSY DRILL 4 HOLE USE 5J868A HEXBOLT 1/4 INCH WING ASSEMBLY, USE 5J868-A HEX BOLT .25"- DRILL FOUR HOLES USE 4 5J868A BOLTS (HEX .25) - DRILL HOLES FOR EACH ON WING ASSEM RUDER, TAP 6 WHOLES, SECURE W/KL2301 RIVETS (10 CM)

Result file:

Assembly	Instruction	Qty	Type	Part	Size	Measure	SKU
WING	DRILL	4	HOLES	HEXBOLT	.25	INCH	5J868A
WING	DRILL	4	HOLES	HEXBOLT	.25	INCH	5J868A
WING	DRILL	4	HOLES	HEXBOLT	.25	INCH	5J868A
RUDDER	DRILL	6	HOLES	RIVET	10	CM	KL2301

Figura 18: Ejemplo de Estandarización de Datos II

Fuente: (IBM, Data Quality Bootcamp, 2012)

Para la estandarización de los datos se debe utilizar funciones para mejorar los conjuntos de reglas de estandarización. Puede añadir y modificar clasificaciones, tablas de búsqueda y reglas.

En la etapa de estandarización se pueden aplicar reglas para formatear de nuevo los datos de varios sistemas. Esta etapa facilita el establecimiento efectivo de coincidencias y el formato de la salida.

3.4.1.3 Coincidencia de los datos

Este proceso proporciona una vista precisa de los datos en toda la empresa. Un motor coincidente estático evalúa la probabilidad de que dos o más conjuntos de datos hagan referencia a la misma entidad empresarial.

En el proceso de establecimiento de coincidencias se buscan datos de un único origen de datos o de orígenes de datos independientes que hacen referencia a la misma entidad (por ejemplo, persona, organización, ubicación, producto o material) aun cuando no exista ninguna clave predeterminada.

Después de que se confirme una coincidencia, se debe construir claves de enlace para que los usuarios puedan completar una transacción o cargar un sistema de destino con datos precisos y de calidad.

Para incrementar su utilización y compleción, los datos pueden consolidarse o enlazarse con cualquier relación, como una persona, empresa, lugar, producto, componente o suceso común. También es posible utilizar la coincidencia para detectar entidades duplicadas causadas por variaciones en la entrada de datos o por prácticas empresariales orientadas a cuentas.

Durante la etapa de establecimiento de coincidencias, se debe llevar a cabo las siguientes acciones:

- Identificar las entidades duplicadas (como clientes, proveedores, productos o componentes) en uno o más orígenes de datos.
- Crear una vista consolidada de una entidad de acuerdo con las reglas empresariales.
- Proporcionar una unidad familiar para los individuos (como una familia o un grupo de personas de una ubicación) y una unidad familiar para entidades comerciales (varias empresas de la misma ubicación o de distintas ubicaciones).
- Permitir la creación de grupos de coincidencia en orígenes de datos que pueden tener o no una clave predeterminada.
- Enriquecer los datos existentes con nuevos atributos de orígenes externos, como por ejemplo datos de oficinas de crédito o expedientes de cambio de dirección.

3.4.1.4 Supervivencia de los datos

La supervivencia garantiza que esté creando la mejor vista disponible de información relacionada. Las reglas empresariales y de correlación se implementan para crear las estructuras de salida necesarias para la

aplicación de destino. Los campos que no se ajustan a estándares de carga se identifican y filtran de forma que sólo se cargue la mejor representación de los datos coincidentes en el registro de datos maestro.

Los valores que faltan en un registro se proporcionan con valores de otros registros de la misma entidad. Los valores que faltan también se pueden rellenar con valores de registros correspondientes que se han identificado como grupo en la etapa coincidente.

Esta etapa consolida los registros duplicados, creando de este modo una mejor representación de los datos coincidentes y crea las estructuras de salida necesarias para la aplicación de destino e identifica los campos que no cumplen los estándares de carga.

Durante la etapa de supervivencia, se debe llevar a cabo estas acciones:

- Suministra los valores que faltan en un registro con valores de otros registros de la misma entidad.
- Rellena los valores que faltan en un registro con valores de los registros correspondientes que se han identificado como grupo en la etapa coincidente.
- Enriquece los datos existentes con datos externos.

La **Figura 19** ilustra el paso de los datos de entrada a través de la etapa de supervivencia para obtener registros completos.

Input Data

SETID	SSN	Name	Address	City	State	Zip Code
1	NULL	Jerome David Salinger	Holden Caufield Hwy	Agerstown	PA	19102
1	NULL	J. David Salinger	51 Holden Caufield Hwy	Agerstown	PA	19102-1919
1	123-45-8789	J.D. Salinger	Holden Caufield Hwy	Agerstown	PA	19102

Survivorship Output

SETID	SSN	Name	Address	City	State	Zip Code
1	123-45-8789	Jerome David Salinger	51 Holden Caufield Hwy	Agerstown	PA	19102-1919

Used to update the group with the "best of breed" data

SETID	SSN	Name	Address	City	State	Zip Code
1	123-45-8789	Jerome David Salinger	51 Holden Caufield Hwy	Agerstown	PA	19102-1919
1	123-45-8789	Jerome David Salinger	51 Holden Caufield Hwy	Agerstown	PA	19102-1919
1	123-45-8789	Jerome David Salinger	51 Holden Caufield Hwy	Agerstown	PA	19102-1919

Figura 19: Ejemplo de Supervivencia de Datos I

Fuente: (IBM, Data Quality Bootcamp, 2012)

La **Figura 20** presenta otro un ejemplo de supervivencia de datos.

Source	Legacy Key	Name	Address	Phone	Birth Date	Cust-ID
Life	70328574	John Smith Jr.	10 Main St Boston MA 02110	781-259-9945	02/05/1940	
Home	80328575	Mr. John Smith	10 Main St Unit 10 Boston MA 02111	617-259-9000		
Auto	90238495	J. Smyth	Main St Bostan Mass 02110	781-295-9945	02/05/1941	

Classic transformation: account to customer

Account view

Customer view ↘ Link related records to create cross-reference IDs

Source	Legacy Key	Name	Address	Phone	Birth Date	Cust-ID
Life	70328574	John Smith Jr.	10 Main St Boston MA 02110	781-259-9945	02/05/1940	0001
Home	80328575	Mr. John Smith	10 Main St Unit 10 Boston MA 02111	617-259-9000		0001
Auto	90238495	J. Smyth	Main St Bostan Mass 02110	781-295-9945	02/05/1941	0002

Customer profile ↘ Create a customer profile with the best information from all sources

Source	Legacy Key	Name	Address	Phone	Birth Date	Cust-ID
CP		Mr. John Smith Jr.	10 Main St Unit 10 Boston MA 02111	617-259-9000	02/05/1940	0001
CP		J. Smyth	Main St Bostan Mass 02110	781-295-9945	02/05/1941	0002

Figura 20: Ejemplo de Supervivencia de Datos II

Fuente: (IBM, Data Quality Bootcamp, 2012)

3.4.2 Metodología de Calidad de Datos de IBM

La metodología de calidad de datos de IBM está agrupada por orden analítico y proporciona conocimientos amplios y procedimientos

recomendados para la estrategia de calidad de datos definida en las organizaciones.

Existe una serie de conceptos clave respecto al análisis y monitoreo de calidad de datos que incluyen capacidades para:

- Brindar soporte a la definición y organización de reglas de negocio.
- Aplicar reglas y reutilizarlas de forma coherente en diversas fuentes de datos.
- Aprovechar el análisis de reglas a varios niveles para comprender más a fondo los problemas de calidad de datos.
- Evaluar reglas contra indicadores de rendimiento.
- Evaluar los resultados de la calidad de datos.
- Monitorear las tendencias de la calidad de datos en el tiempo.
- Implementar reglas en el ambiente de calidad de datos.
- Ejecutar actividades de calidad de datos.

La metodología de IBM para calidad de datos propone ocho criterios:

1. Análisis y monitoreo de la calidad de los datos: el primer punto de la metodología es evaluar los datos en base al criterio específico del cliente. Este criterio de evaluación puede utilizarse repetidamente para ver cambios importantes en la calidad de los datos que se validan. Luego del perfilado de datos y en función de las reglas de negocio definidas por el cliente se crean las reglas de calidad de datos.

2. Estructuración de reglas de datos y conjuntos de reglas: el segundo punto es la definición de reglas para asegurar la calidad del dato en base a las definiciones o lógica del negocio. Para esto se utilizan conceptos claves del negocio. Se recomienda la reutilización de reglas en el proceso y también evaluar que reglas pueden ser agrupadas para la ejecución simultánea sobre un conjunto de datos.

Las definiciones de reglas representan una expresión lógica y pueden incluir cualquier número de variables. A partir de las definiciones lógicas pueden generarse una o más reglas de datos físicos. Para la implementación real, estas reglas se almacenan en múltiples repositorios de datos físicos.

3. Estándares de nomenclatura: todas las estructuras que se creen tales como definiciones de reglas y conjuntos de reglas, todas las reglas de datos y conjuntos de reglas ejecutables, todas las medidas, y todas las variables globales deben tener nombre ya que esto puede facilitar o restringir que se reutilicen y se compartan estos componentes, es vital identificar un estándar de denominación efectivo que facilite el desarrollo y el monitoreo de los procesos de calidad.

4. Análisis de reglas de datos: en la base del análisis de reglas de datos, la empresa debe crear reglas de datos para probar y evaluar condiciones de datos específicas.

Para hacer el análisis de reglas de datos eficaz es necesario responder a las siguientes preguntas:

- ¿Qué datos están implicados?
- ¿Hay distintas partes o condiciones en la validación?
- ¿Hay cualidades conocidas de los datos a tener en cuenta?

- ¿Cuáles son los orígenes de los datos (por ejemplo, archivos externos)?
- ¿Hay tipos de datos específicos (por ejemplo, fechas, cantidades y otras clases de datos) a evaluar?
- ¿Hay aspectos de la regla que impliquen a las estadísticas de la validación?
- ¿Hay aspectos de la regla que impliquen comprender qué ha sucedido anteriormente?

5. Medidas: las medidas son objetos definidos por el usuario que no analizan datos pero proporcionan prestaciones de cálculo matemático que pueden convertirse en resultados estadísticos de las reglas de datos, los conjuntos de reglas de datos y las propias medidas.

Las medidas le proporcionan la capacidad de consolidar las mediciones de diversos pasos de análisis de datos en una sola medida con significado para la gestión de la calidad de datos. Las medidas pueden utilizarse para reducir cientos de resultados analíticos detallados a unas cuantas medidas con significado que expliquen con eficacia la condición global de la calidad de los datos.

6. Resultados del monitoreo: las reglas de datos, los conjuntos de reglas y las medidas son objetos que pueden ejecutarse cuando sea necesario o de forma planificada.

Las reglas de datos, los conjuntos de reglas y las medidas generan sucesos históricos, estadísticas y resultados detallados. Dado que estos objetos se ejecutan repetidamente, crean una serie de sucesos de los que puede hacer un seguimiento, crear anotaciones, informes y tendencias con el tiempo.

7. Implementación de reglas, conjuntos de reglas y medidas: las reglas de datos, los conjuntos de reglas y las medidas de un entorno de gestión de calidad de datos suelen estar destinados a datos de producción, lo que requiere un control más explícito sobre las tareas, tales como la definición y el cambio.

Normalmente, el diseño, el desarrollo y las pruebas iniciales se producen en un entorno no de producción, produciéndose la supervisión de la calidad de datos de producción en un entorno de producción aparte. Puede utilizarse el mismo enfoque para compartir objetos entre dos entornos diferenciados.

8. Gestión de un ambiente de reglas de calidad de datos: mientras los usuarios se centran en más áreas de negocio y más sistemas, creando más definiciones de reglas y compartiendo información, se crea una mayor necesidad de gestionar el ambiente de calidad de datos. En esta etapa se incluyen las siguientes tareas:

- Organizar los proyectos en ambientes de producción y no producción.
- Organizar fuentes de información.
- Establecer seguridades, perfiles de acceso y usuarios.
- Definir el seguimiento de uso y auditoría.

CAPÍTULO IV

4. CONCLUSIONES Y RECOMENDACIONES

4.1 CONCLUSIONES

- El desarrollo del presente trabajo permitió cumplir con el objetivo propuesto, el desarrollo de un modelo de Gestión de Calidad de Datos basado en la metodología IBM Data Quality para ser aplicado en los clientes de BI de la empresa DWConsultware.
- Se definió el ciclo de vida el cual consta de cinco etapas para el desarrollo de los proyectos de calidad de datos, este ciclo es similar al ciclo de vida espiral ya que al final de cada etapa se debe volver a fases anteriores, esto permite la mejora continua de la calidad del dato.
- Se determinaron las cuatro fases de la gestión de calidad de datos: Investigar, Estandarizar, Establecer Coincidencias y Supervivencia de los Datos. Estas fases son la base para el desarrollo de los procesos en las herramientas de IBM.
- El modelo TDWI consta de 6 niveles de madurez: Pre-natal, Infancia, Niñez, Adolescencia, Edad Adulta y Sabiduría. Cada nivel describe la realidad del Business Intelligence de una organización. Adicionalmente se identificaron 2 puntos críticos en el modelo el Golfo y el Abismo, estos dividen las etapas y marcan el grado de dificultad de pasar de un nivel a otro.
- Se detalló la metodología de IBM para el desarrollo de un proyecto de calidad de datos. La metodología pone a consideración 8

directrices que deben ser tomados en cuenta en la implementación de los procesos de mejora y gestión de calidad de datos.

4.2 RECOMENDACIONES

- En base al presente trabajo, desarrollar un caso de estudio aplicando el Modelo de Gestión de Calidad de Datos, utilizando la última versión de la herramienta de calidad de datos de IBM.
- En el caso de estudio, utilizar el cuestionario de evaluación del nivel de madurez TDWI para identificar la fase de adopción de BI y la cercanía con los puntos críticos.
- Aplicar la Gestión de Calidad de Datos, es decir las Fases descritas en la sección 3.4.1 junto con la Metodología descrita en la sección 3.4.2 del presente trabajo en el caso de estudio a ser desarrollado.

BIBLIOGRAFÍA

1. Delgadillo, G. B. (2011). Módulo I: Toma de Decisiones Efectiva. Quito: Tecnológico Monterrey Quito.
2. Eckerson, W. (2010). TDWI Poster: Business Intelligence Maturity Model. Obtenido de http://www.tdwi.eu/wissen/whitepaper/?no_cache=1&tx_mwknowledgebase_pi1%5BshowUid%5D=46
3. Eckerson, W. (s.f.). What CIO's Should Know to Ensure BI Success. Obtenido de ftp://public.dhe.ibm.com/software/data/sw-library/cognos/pdfs/analystreports/ar_what_cio_should_know_to_ensure_bi_success.pdf
4. IBM. (2011). Data Quality Management (DQM). (IBM) Recuperado el 19 de Octubre de 2013, de http://pic.dhe.ibm.com/infocenter/easrr/v8r0m0/index.jsp?topic=%2Fcom.ibm.iis.ii.overview.doc%2Ftopics%2Ffeas_con_dqm.html
5. IBM. (2012). Data Quality Bootcamp.
6. IBM. (20 de 09 de 2012). InfoSphere QualityStage. (IBM) Recuperado el 19 de 10 de 2013, de <http://pic.dhe.ibm.com/infocenter/iisinfsv/v8r5/index.jsp?topic=%2Fcom.ibm.swg.im.iis.productization.iisinfsv.overview.doc%2Ftopics%2Fcisoqsusing.html>
7. IBM. (2014). Information Integration and Governance. (IBM) Recuperado el 21 de Septiembre de 2013, de <http://www-01.ibm.com/software/data/information-integration-governance/info-governance.html>
8. Informatica. (2014). Data Quality. Recuperado el 21 de Septiembre de 2013, de <http://www.informatica.com/es/products/data-quality/data-quality/>
9. Lara, G. M. (2011). Módulo VII: Business Intelligence Strategy. Quito: Tecnológico Monterrey Quito.
10. Lara, G. M. (2012). Programa de Business Intelligence, Business Intelligence Strategy: Definición, implementación y seguimiento. Tecnológico de Monterrey.
11. Martínez, J. (23 de Julio de 2012). Seis pasos para el Gobierno de Datos. (IBM) Recuperado el 21 de Septiembre de 2013, de <http://www.ibm.com/developerworks/ssa/data/library/techarticle/gobierno-datos/>
12. Menéndez, C. P. (13 de 08 de 2013). En los datos, la calidad importa. Obtenido de <http://liberix.es/blog/en-los-datos-la-calidad-importa/>
13. Moss, L. T. (2003). Business Intelligence RoadMap. Addison Wesley.

14. Otake, L. A. (2010). Fundamentos de Business Intelligence – Parte I. Obtenido de <http://es.slideshare.net/sandraceciliagg1979/fundamentos-de-bi-parte-1>
15. Power, D. J. (Marzo de 2007). A Brief History of Decision Support Systems. Obtenido de <http://dssresources.com/history/dsshhistory.html>
16. Rajterič, I. H. (2010). OVERVIEW OF BUSINESS INTELLIGENCE MATURITY MODELS. *Management*, 15, 47-67.
17. SAP. (24 de Enero de 2013). Data Quality Management for SAP. Recuperado el 21 de Septiembre de 2013, de <http://wiki.scn.sap.com/wiki/display/EIM/Data+Quality+Management+for+SAP>
18. SAS Community. (17 de Enero de 2014). Data Quality for Analytics. Recuperado el 21 de Septiembre de 2013, de http://www.sascommunity.org/wiki/Data_Quality_for_Analytics
19. TDWI. (2012). Evaluación del Modelo de Madurez de BI de TDWI. Obtenido de http://tdwiorg0000.web711.discountasp.net/Content/TDWI_Benchmark_Final.pdf
20. TDWI, T. D. (2013). TDWI Survey Results. Obtenido de http://tdwiorg0000.web711.discountasp.net/Home/Report/1oR90tRzn0CqW2yk_X3swg
21. TDWI, T. D. (s.f.). TDWI Data Quality Fundamentals. En TDWI, TDWI Data Quality Fundamentals (págs. 9-11).
22. Trejo, R. A. (2011). Modulo III: Data Warehouses. Quito: Tecnológico Monterrey Quito.
23. Urquizi, P. (10 de Junio de 2009). Business Intelligence fácil. Recuperado el 28 de Septiembre de 2013, de <http://www.businessintelligence.info/definiciones/antes-empezar-proyecto-bi.html>
24. Wang, R. Y., Kon, H. B., & Madnick, S. E. (1998). Data Quality Requirements Analysis and Modeling. Ninth International Conference of Data Engineering. Vienna, Austria.

ANEXOS

ANEXO A TDWI BI Maturity Model

TDWI's Business Intelligence Maturity Model

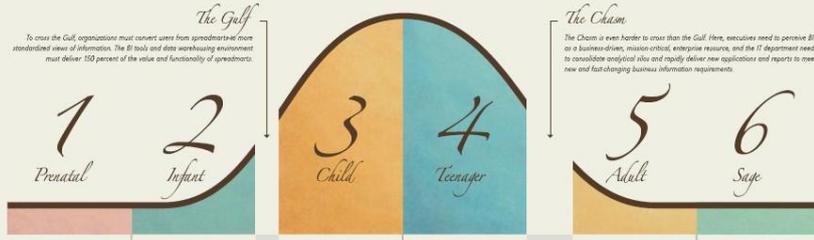
The six-stage BI Maturity Model shows the trajectory that most organizations follow when evolving their BI infrastructure from a low-value, cost-center operation to a high-value, strategic utility that drives market share. The model provides a

big-picture view of where a BI program is, where it needs to go, and how to get there. No organization evolves through the six stages at the same rate or in the

same way. Many exhibit characteristics of multiple stages at the same time; few move cleanly and precisely from one stage to the next.

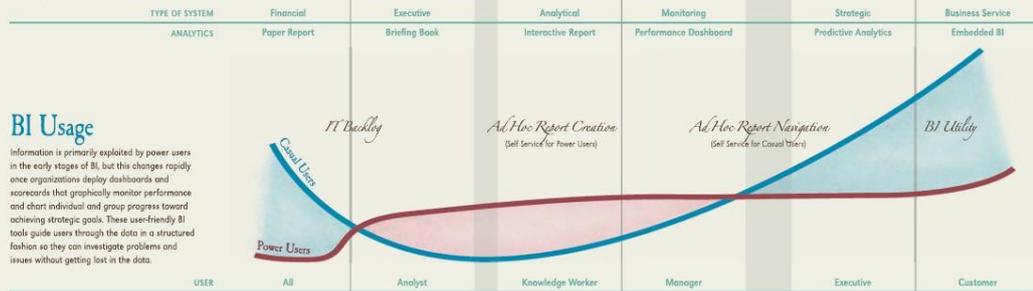
BI Adoption Curve

Most organizations go through six stages when evolving their BI environment from a cost-center operation to a strategic resource that drives the business and shapes the market. The bell-shaped curve indicates that most organizations have reached stages three and four or "BI adolescence," marked by unstable funding, lukewarm usage, and the continued presence of spreadsheets.



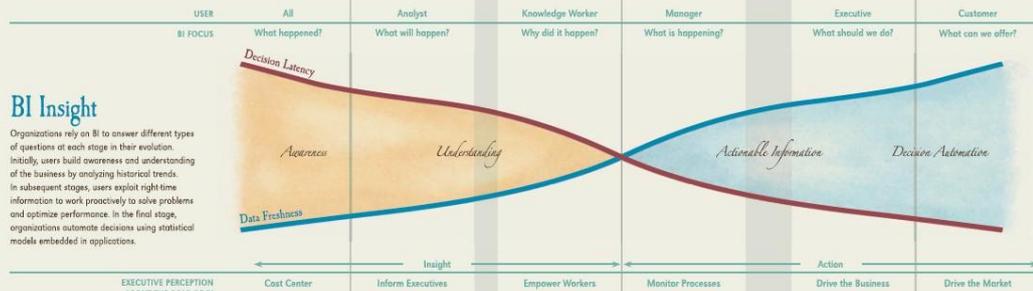
Local Control versus Enterprise Standards

In the early stages of BI, power users equipped with spreadsheets exert tremendous control over the content and display of information. This control declines in subsequent stages as departments, divisions, and finally, the enterprise renegotiate information standards. Once all data is modeled and managed in the BI environment, business users equipped with powerful BI tools create new views, analyses, and reports more quickly than ever, finally balancing the dual need for standards and flexibility.



BI Usage

Information is primarily exploited by power users in the early stages of BI, but this changes rapidly once organizations deploy dashboards and scorecards that graphically monitor performance and chart individual and group progress toward achieving strategic goals. These user-friendly BI tools guide users through the data in a structured fashion so they can investigate problems and issues without getting lost in the data.



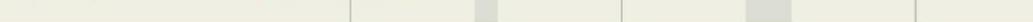
BI Insight

Organizations rely on BI to answer different types of questions at each stage in their evolution. Initially, users build awareness and understanding of the business by analyzing historical trends. In subsequent stages, users exploit real-time information to work proactively to solve problems and optimize performance. In the final stage, organizations automate decisions using statistical models embedded in applications.



Business Value and ROI

The business value of a BI environment grows slowly during the first three stages as organizations struggle to consolidate analytical silos, deliver a consistent view of information, and empower users with BI tools that conform to the way they work. Business value and ROI accelerate in the last three stages as BI becomes an enterprise resource that drives mission-critical processes and delivers rich insights for a competitive advantage.



ANEXO B CUESTIONARIO TDWI

TDWI Benchmark Assessment To Determine Your BI/DW Maturity

Purpose

This 20- to 25-minute assessment is designed to help BI/DW professionals-determine the maturity of their BI/DW initiative and compare it to those at other companies in an objective way.

Who Should Take the Assessment?

The assessment is geared to BI/DW professionals in user organizations or consultants who work closely with at least one user organization that has an active BI/DW initiative.

Taking the Assessment

If you've enabled cookies, you can return within seven days to complete the assessment. Or click the "Save" button and you have 30 days to return to it.

Benchmark Results

When you finish the assessment, you will receive a score indicating the maturity of your BI/DW program based on the TDWI Business Intelligence Maturity Model. An interactive chart will enable you to visually compare your scores in eight categories to others who have taken the assessment. You will be able to filter the results by various variables such as industry, company size, and BI/DW budget. You will also be able to download a PDF of the results and a Benchmark Guide to help you interpret the results.

Acknowledgments

TDWI would like to thank the people who helped write and review this assessment: Wayne Eckerson, Michael Gonzales, Hugh Watson, Barbara Wixom, Jill Dyche, Sid Adelman, Joyce Norris-Montanari, Donna Corrigan, Bill Lay, Steve Hoberman, and Vivian Chan.

Definitions

BI/DW Environment - *Any architecture or system* that integrates data from multiple sources and gives business users the ability to query, report, analyze, monitor, or mine the data to make decisions, develop plans, or take action. The diagram below depicts a classic BI/DW environment.

DW - The back-end or IT-centric portion of a BI/DW environment. The tasks and processes involved in collecting, moving, cleaning, transforming, and loading data into a data warehouse or data mart.

BI - The front-end or user-centric portion of a BI/DW environment. The tools and processes that enable users to query, report, analyze, monitor, and mine integrated information to make decisions, develop plans, or take action.

Your Information

Which best describes your POSITION?

- IT professional
- Business sponsor, driver, or user
- Systems integrator or consultant
- Vendor representative (i.e. sales, marketing, consultant)
- Professor or student

Which best describes WHICH SIDE of the business you're on?

- Business
- Information technology
- I straddle both sides

Which best describes your ROLE?

- BI/DW sponsor
- Business analyst or subject expert
- BI/DW director or manager
- BI/DW architect or developer
- BI/DW administrator or systems analyst
- Other

What is your email address? **used to send you the assessment results*

Your Company

Which best describes your organization's primary INDUSTRY?

What are the annual revenues of your organization?

Where are you LOCATED?

*****Consultants: Please respond to the REMAINING ASSESSMENT questions with a SINGLE client in mind.*****

How many distinct BI/DW groups exist in your ORGANIZATION as a whole?

Your Group

*****Please answer all remaining questions in this assessment based on the BI/DW capabilities deployed by YOUR GROUP ONLY, unless otherwise specified.*****

Which best describes the SCOPE of the group in which you work?

- Corporate
- Business unit
- Department or functional area (e.g. finance, sales, marketing)
- Other

Which best describes the ORIENTATION of the group in which you work?

- Business
- IT
- Other

How LONG AGO was your group's BI/DW initiative started?

Which best describes the BI/DW TASKS your group performs?

- DW only
- BI only

- Both BI and DW
- Other

How many full-time equivalent BI/DW STAFF MEMBERS does your group maintain (including contractors)?

- 5 or fewer
- 6 to 10
- 11 to 20
- 21 to 50
- 51 or more

What is your group's annual BI/DW budget, including staff, software, hardware, and services?

Scope

The goals for BI/DW systems are defined before building a system

- Entirely Unfulfilled
- Unfulfilled
- Undecided or Uncertain
- Filled
- Entirely Filled

BI/DW strategy is aligned with the strategic plan of the organization

- Entirely Unfulfilled
- Unfulfilled
- Undecided or Uncertain

Filled

Entirely Filled

BI/DW objectives adapt to the changing objectives of the organization

Entirely Unfulfilled

Unfulfilled

Undecided or Uncertain

Filled

Entirely Filled

How many APPLICATIONS does your BI/DW environment support?

0

1 to 2

3 to 5

6 to 10

11 to 20

20+

A BI/DW application consists of a distinct set of related reports, dashboards, briefing books, or scorecards designed to support a specific set of tasks within a business domain or process, such as revenue management, customer churn analysis, supplier performance management, etc.

Users are assigned full-time tasks/roles to BI/DW projects

Strongly Disagree

Disagree

Undecided or Uncertain

Agree

Strongly Agree

IT professional	Information techn	BI/DW architect o	christianandrade	Food / Beverage
\$Less than \$25	Other	1	Business unit	IT
2.5 to 5 years	Both BI and DW	6 to 10	\$100k	3
3	3	4	3	

Sponsorship

Which best describes how **EXECUTIVES** perceive the **PURPOSE** of your group's **BI/DW** environment?

- Operational cost center** - An IT system needed to run the business.
- Tactical resource** - Tools to assist decision making.
- Mission-critical resource** - A system that is critical to running business operations.
- Strategic resource** - Key to achieving performance objectives and goals.
- Competitive differentiator** - Key to gaining or keeping customers and/or market share.

Which best describes the **SPONSOR** of your **BI/DW** group?

- None to speak of right now
- CIO or an IT director
- Single sponsor from a business unit or department
- Multiple individual sponsors from multiple business units or departments
- Multiple levels of business-driven, cross-departmental steering committees

A committed sponsor evangelizes the BI/DW program to the entire company, secures funding, leads political interference, appoints a trusted advisor to oversee the project on a part-time basis, and sees the project through to completion (i.e., doesn't leave to pursue personal ambitions).

To what degree is your sponsor **COMMITTED** to the **BI/DW** program?

- Very low
- Low
- Moderate

- High
- Very high

To what degree is the BI sponsor held ACCOUNTABLE for the outcome of the BI/DW solution?

- Very low** - Only project or program managers are held accountable.
- Low** - Between very low and moderate.
- Moderate** - Sponsor assumes accountability but is not incented for performance.
- High** - Between moderate and very high.
- Very high** - Sponsor assumes accountability and is incented for performance.

Senior management is involved in the BI/DW through steering committee/governance

- Strongly Disagree
- Disagree
- Undecided or Uncertain
- Agree
- Strongly Agree

IT professional	Information techn	BI/DW architect o	christianandrade	Food / Beverage
\$Less than \$25	Other	1	Business unit	IT
2.5 to 5 years	Both BI and DW	6 to 10	\$100k	3
3	3	4	3	3
1	2	2	3	

Funding

How easy is it to get **FUNDING** for your group's annual BI/DW budget?

- Very hard** - Our budget gets cut before other IT projects.
- Hard** - Between very hard and moderate.
- Moderate** - We are usually funded at a rate comparable to the rest of IT.
- Easy** - Between moderate and very easy.
- Very easy** - With our demonstrated track record, we usually get what we ask for.

Our organization's level of investment in BI/DW is ...

- Much lower than competitors
- Lower than competitors
- Undecided or Uncertain
- Higher than competitors
- Much higher than competitors

The annual BI/DW **BUDGET** for your BI/DW group represents approximately what percent of the annual IT budget for your group?

4% to 5%

Which best describes the current degree of **CAPITAL INVESTMENT** in your BI/DW system?

- We're bootstrapping our system.
- We've received initial capital funding to purchase software and systems.
- We've received additional capital funding to build on our initial success.

- We've received significant and sustained capital investment to pursue full deployment of our vision.
- We're now fully deployed with less need for additional capital.

Which best describes the current MAINTENANCE BUDGET for your group's BI/DW system?

- Very low** - Limited or no funds to deliver and support requested business projects.
- Low** - Between very low and moderate.
- Moderate** - We have sufficient funds to deliver and support some but not all requested business projects.
- High** - Between moderate and very high.
- Very high** - We're fully funded to deliver and support most or all requested business projects.

IT professional	Information techn	BI/DW architect o	christianandrade	Food / Beverage
\$Less than \$25	Other	1	Business unit	IT
2.5 to 5 years	Both BI and DW	6 to 10	\$100k	3
3	3	4	3	3
1	2	2	3	2
2	3	2	2	

Value

BI/DW reduces the costs for many business processes

- Strongly Disagree
- Disagree
- Undecided or Uncertain
- Agree
- Strongly Agree

BI/DW enhances the value of our products and/or services

- Strongly Disagree
- Disagree
- Undecided or Uncertain
- Agree
- Strongly Agree

BI/DW assists in identifying the most appropriate customers/clients for our organizations

- Strongly Disagree
- Disagree
- Undecided or Uncertain
- Agree
- Strongly Agree

BI/DW assists with the monetizing of information for customers or suppliers

- Strongly Disagree
- Disagree
- Undecided or Uncertain
- Agree
- Strongly Agree

BI/DW projects always contain an assessment of risk

- Strongly Disagree
- Disagree
- Undecided or Uncertain
- Agree
- Strongly Agree

IT professional	Information techn	BI/DW architect o	christianandrade	Food / Beverage
\$Less than \$25	Other	1	Business unit	IT
2.5 to 5 years	Both BI and DW	6 to 10	\$100k	3
3	3	4	3	3
1	2	2	3	2
2	3	2	2	3
3	3	3	3	

Architecture

What is the PREDOMINANT ARCHITECTURE of your DW environment?

- Desktop- or user-generated reports (i.e., spreadmarts)
- Multiple, non-integrated data marts or packaged solutions
- Multiple, non-integrated data warehouses
- A single, central DW with multiple dependent marts (Inmon), conformed data marts (Kimball), or no data marts
- A BI/DW service that federates a central DW and other data sources via a standard interface

To what degree can users directly ACCESS the data they need to make decisions from a single user interface?

- Very low** - Users can access virtually none of the data they need.
- Low** - Between very low and moderate.
- Moderate** - Users can access a moderate amount of the data they need.
- High** - Between moderate and very high.
- Very high** - Users can access all the data they need.

To what degree have you established standards for TECHNOLOGY AND TOOLS in your BI/DW environment?

- Very low** - We have not defined standards
- Low** - We have defined some standards
- Moderate** - We have defined about half the standards

High - We have defined most of the standards

Very high - We have defined all standards

To what degree do individuals and groups ADHERE to the technology and tool standards that your group has established?

Very low - No one adheres (or standards don't yet exist).

Low - Few people and groups adhere.

Moderate - Some people and groups adhere.

High - Most people and groups adhere.

Very high - All people and groups adhere.

To what degree has your group defined, documented, and implemented DEFINITIONS AND RULES for key terms and metrics?

Very low - None defined.

Low - Some defined.

Moderate - About half defined.

High - Most defined.

Very high - All defined.

IT professional	Information techn	BI/DW architect o	christianandrade	Food / Beverage
\$Less than \$25	Other	1	Business unit	IT
2.5 to 5 years	Both BI and DW	6 to 10	\$100k	3
3	3	4	3	3
1	2	2	3	2
2	3	2	2	3
3	3	3	3	3
2	2	2	2	

Data

To what degree do end users TRUST THE DATA in your BI/DW environment?

- Very low** - Users don't trust the data. They find too many errors, exceptions, or omissions. They rely on other sources of data to make critical decisions.
- Low** - Between very low and moderate.
- Moderate** - Users somewhat trust the data. They reconcile it with more trusted sources before using.
- High** - Between moderate and very high.
- Very high** - Users trust the data. They view it as the system of record for the company and rely on it for critical decisions.

How many unique DATA SOURCES does your BI/DW environment draw from?

On average, how often are the MAJORITY of data elements in your BI/DW environment REFRESHED?

- Either annually or quarterly
- Either quarterly or monthly
- Either monthly or weekly
- Either weekly or daily
- Either daily or intra-day

Which best describes the degree of synchronization among the DATA MODELS below that your group maintains?

- ETL source and target models
- Data warehouse and data mart models
- BI semantic or query object models

- Very low** - We do not synchronize models.
- Low** - We manually synchronize some of the models.
- Moderate** - We manually synchronize some models, others automatically.
- High** - We automatically synchronize most models, but not all.
- Very high** - We automatically synchronize all models.

To what degree has your group integrated UNSTRUCTURED DATA (i.e., text or documents) in the BI/DW environment?

- Very low** - No plans or approach.
- Low** - Between very low and moderate.
- Moderate** - Users can search text and documents via a separate (non-BI) application.
- High** - Between moderate and very high.
- Very high** - We return related structured and unstructured data via a single BI query.

IT professional	Information techn	BI/DW architect o	christianandrade	Food / Beverage
\$Less than \$25	Other	1	Business unit	IT
2.5 to 5 years	Both BI and DW	6 to 10	\$100k	3
3	3	4	3	3
1	2	2	3	2
2	3	2	2	3
3	3	3	3	3
2	2	2	2	2
3	3	2	2	

Development

Which best describes your BI/DW group's approach to DEVELOPING BI/DW solutions?

- Independent** - Business groups develop their own BI/DW solutions without our assistance or guidance.
- Ad hoc** - We develop one-off BI/DW solutions as business groups fund them using whatever tools and techniques make sense.
- Aligned** - We develop one-off BI/DW solutions, but use a common set of tools and techniques to ensure long-term alignment.
- Standardized** - We develop a portfolio of integrated BI applications using a common set of tools and techniques, working within a common architecture and process framework.
- Federated** - We let business units develop their own BI/DW solutions as long as they adhere to our architecture and process framework.

To what degree has your group defined, documented, and implemented standards for DEVELOPING, TESTING, and DEPLOYING BI/DW functionality (i.e., ETL code and BI reports)?

- Very low** - Our group has not established BI/DW standards.
- Low** - Our group has established some BI/DW standards.
- Moderate** - Our group has established about half of the BI/DW standards needed.
- High** - Our group has established most BI/DW standards.
- Very high** - Our group has established a comprehensive set of BI/DW standards.

A standardized process for prioritizing BI/DW projects has been established

- Entirely Unfulfilled
- Unfulfilled
- Undecided or Uncertain

Filled

Entirely Filledk

On average, how many BI/DW PROJECTS that last three or more months does your group run concurrently?

How long does it take your team to add a new SUBJECT AREA to the BI/DW environment?

12+ months

9 to 12 months

6 to 9 months

3 to 6 months

Three months or less

Creating a subject area usually involves the following:

- 1) Define user requirements
- 2) Analyze source systems
- 3) Model/revise target model
- 4) Develop extract, transform, load, and validation routines
- 5) Create/revise reports
- 6) Test
- 7) Deploy and train users

IT professional	Information techn	BI/DW architect o	christianandrade	Food / Beverage
\$Less than \$25	Other	1	Business unit	IT
2.5 to 5 years	Both BI and DW	6 to 10	\$100k	3
3	3	4	3	3
1	2	2	3	2
2	3	2	2	3
3	3	3	3	3
2	2	2	2	2
3	3	2	2	3
2	3	3	4	

Delivery

Of people who use BI on a regular basis, most have a strong understanding of business functions

- Strongly Disagree
- Disagree
- Undecided or Uncertain
- Agree
- Strongly Agree

There is a well-organized availability of technical training

- Strongly Disagree
- Disagree
- Undecided or Uncertain
- Agree
- Strongly Agree

Which best describes how users access BUSINESS METADATA?

- Level 1** - There is no business metadata to access.
- Level 2** - Users consult metadata reports that we periodically distribute.
- Level 3** - Users query various repositories to access business metadata.
- Level 4** - Users query a central repository to access up-to-date business metadata.
- Level 5** - Users click once to view integrated, contextual, up-to-date business metadata.

Formal measurement of training is done to improve courses

- Strongly Disagree
- Disagree
- Undecided or Uncertain

Agree

Strongly Agree

There exists a well-organized availability of business training

Strongly Disagree

Disagree

Undecided or Uncertain

Agree

Strongly Agree