



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

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

PROYECTO A DESARROLLAR PREVIO A LA OBTENCION DEL TITULO DE MAGISTER.

**Ing. Andrade Tirira Christian Andrés
Ing. Madrid Ruiz David Alejandro**

**Director:
Ing. Fernando Galárraga Msc.**

Sangolquí - Junio 2014



AGENDA

- Introducción
- Marco Teórico
- Modelo de Gestión
- Conclusiones y Recomendaciones

Introducción

- Planteamiento del problema
- Objetivos
- Alcance

Planteamiento del Problema

DWConsultware 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?

Objetivos

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.

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.

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.

Planteamiento del Problema

DWConsultware 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?

Objetivos

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.

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.

Objetivos

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.

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.

para estandarizar los proyectos de desarrollo de Business Intelligence pertenecientes a la empresa DWConsultware.

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.

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 DWConsulware.

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.

Marco Teórico

- Inteligencia de Negocios
- Modelo de Madurez TDWI
- Calidad de Datos

Inteligencia de Negocios

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.

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



Etapas del Modelo TDWI



Calidad de datos

Para ser considerados de calidad, los datos deben tener las características que se indican en la siguiente figura.

Exactitud	• Precisión • Fidelidad • Veracidad • Integridad
Disponibilidad	• Accesibilidad • Disponibilidad • Usabilidad
Consistencia	• Homogeneidad • Uniformidad • Interoperabilidad • Integridad • Intercambio • Interoperabilidad
Seguridad	• Privacidad • Confidencialidad • Integridad • Disponibilidad • Disponibilidad • Disponibilidad



Inteligencia de Negocios

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.

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

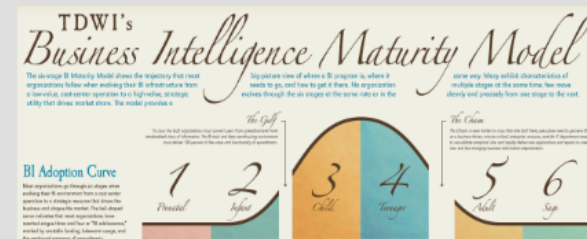


Etapas del Modelo TDWI



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



Etapas del Modelo TDWI

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

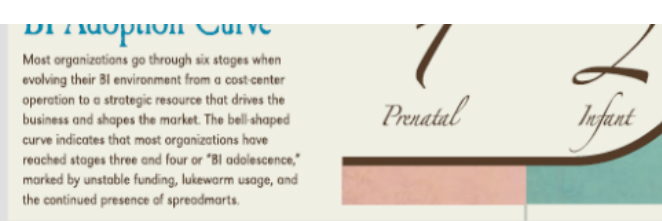
big-picture view of where a BI program is, where it needs to go, and how to get it there. No organization evolves through the six stages at the same rate or in the

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.

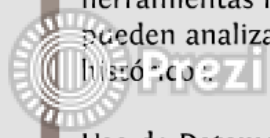




Etapas del Modelo TDWI

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

BI se transforma de una herramienta táctica a una estratégica, es ahora el centro de los sistemas de TI, los procesos son monitoreados usando



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.

1. Pre Natal

2. Infancia

atal

2. Infancia

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 y ensamblados en conjuntos relativamente organizados antes de ser pasados a los usuarios.

Uso de spreadmarts.

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.

ancia

3. Niñez

4. Adole

z

4. Adolescencia

5. Edac

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.

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.

encia

5. Edad Adulta

6. Sab

lta

6. Sabiduría

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

Calidad de datos

Para ser considerados de calidad, los datos deben tener las características que se indican en la siguiente figura.

Libre de defectos	<ul style="list-style-type: none">• Reconocimiento• Medición	<ul style="list-style-type: none">• Validación• Evaluación	<ul style="list-style-type: none">• Verificación
Cumple con las especificaciones	<ul style="list-style-type: none">• Formato• Privacidad	<ul style="list-style-type: none">• Contenido• Seguridad	<ul style="list-style-type: none">• Estructura
Se ajusta a los objetivos	<ul style="list-style-type: none">• Transacciones• Medición• Decisiones	<ul style="list-style-type: none">• Reporteo• Análisis• Descubrimiento	<ul style="list-style-type: none">• Registro Auditoria• Pronosticos
Cumple con las expectativas de los clientes de información	<ul style="list-style-type: none">• Ejecutivos• Analistas	<ul style="list-style-type: none">• Administradores• Auditores	<ul style="list-style-type: none">• Operadores• Reguladores

8 Dimensiones de los Datos de Calidad



Calidad de datos

Para ser considerados de calidad, los datos deben tener las características que se indican en la siguiente figura.



Prezi

- Reconocimiento
- Validación
- Verificación
- Medición
- Evaluación

Libre de defectos

- Reconocimiento
- Validación
- Verificación
- Medición
- Evaluación

Cumple con las especificaciones

- Formato
- Contenido
- Estructura
- Privacidad
- Seguridad

Se ajusta a los objetivos

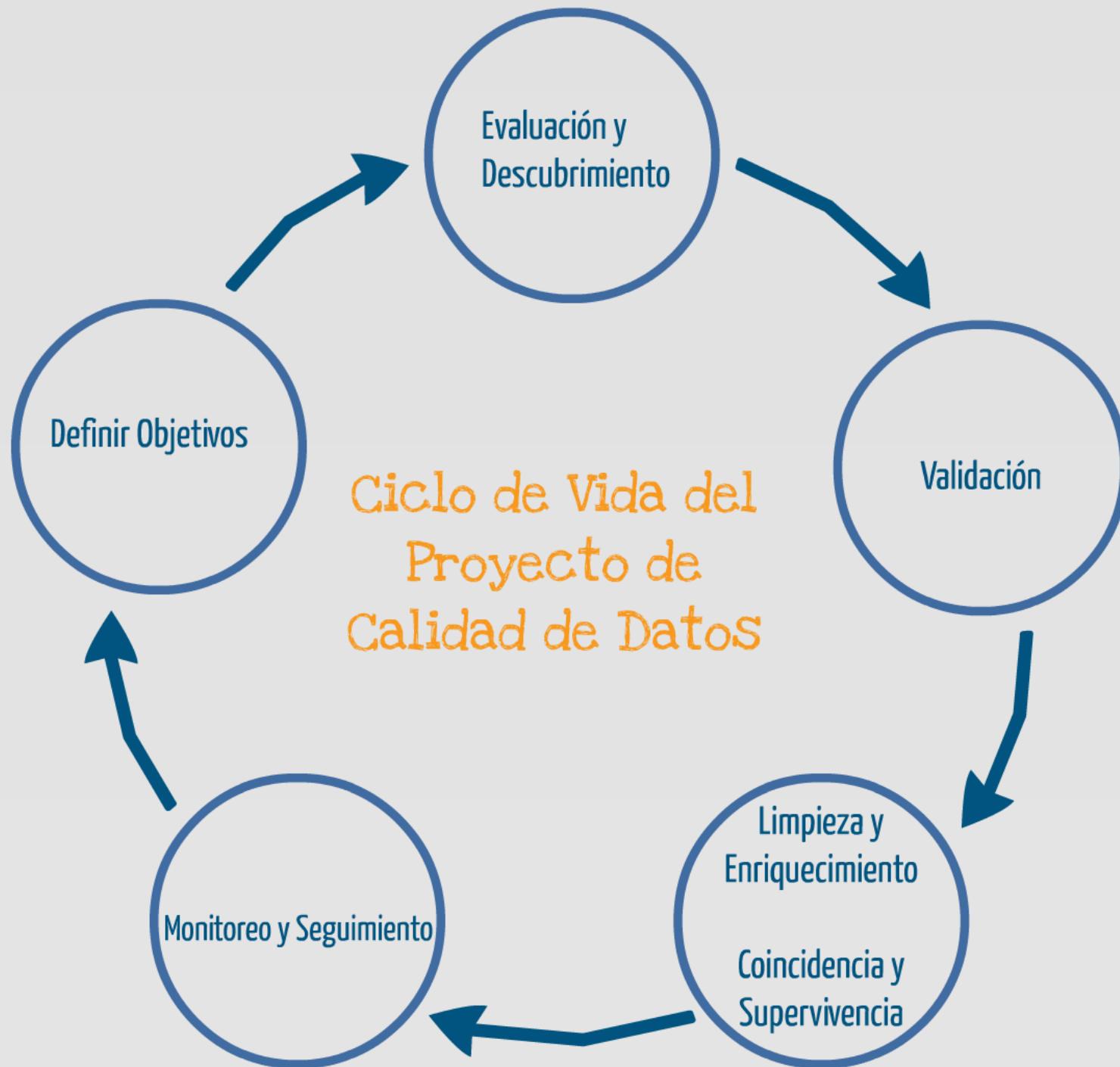
- Transacciones
- Reporteo
- Registro Auditoría
- Medición
- Análisis
- Pronosticos
- Decisiones
- Descubrimiento

Cumple con las expectativas de los clientes de información

- Ejecutivos
- Administradores
- Operadores
- Analistas
- Auditores
- Reguladores

8 Dimensiones de los Datos de Calidad





Modelo de Gestión

- Definición Proyecto BI
- Modelo IBM
- Gestión de Calidad de Datos
- Fases de la Gestión de Calidad de Datos
- Metodología

Modelo IBM

Este modelo se enfoca en responder cuatro preguntas básicas:

- ¿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?



El modelo de gestión de calidad de datos de IBM requiere de tres disciplinas para obtener resultados, sin embargo se soportan en dos habilitadores de la organización, en este caso su representancia, la estructura organizacional y sus políticas aumentan la seguridad y privacidad de los datos, que permiten a futuro la auditoría de los mismos.

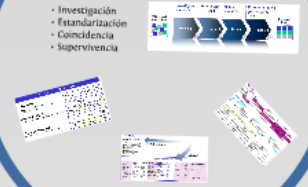
Definición Proyecto de BI

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



Fases de la gestión de calidad de datos

- Investigación
- Estandarización
- Coincidencia
- Supervivencia



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. Llegar 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.

Metodología

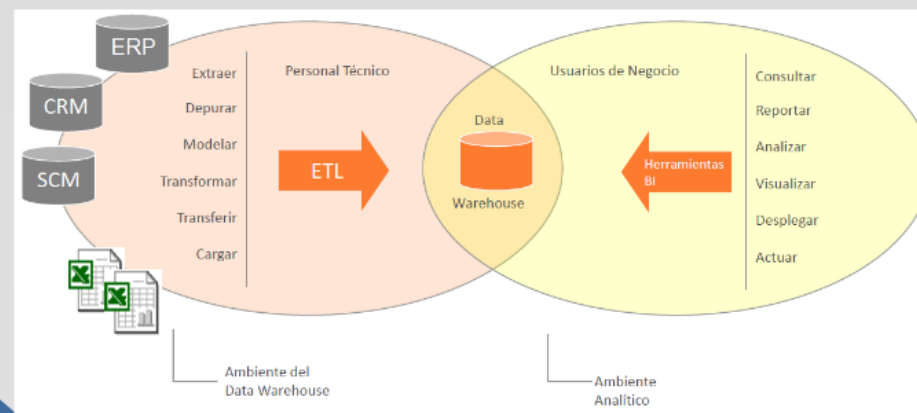
- 1) Análisis y monitoreo de la calidad de los datos
- 2) Estructuración de reglas de datos y conjuntos de reglas
- 3) Estándares de nomenclatura
- 4) Análisis de reglas de datos
- 5) Medidas
- 6) Resultados del monitoreo
- 7) Implementación de reglas, conjuntos de reglas y medidas
- 8) Gestión de un ambiente de reglas de calidad de datos

La metodología de IBM para calidad de datos propone ocho criterios

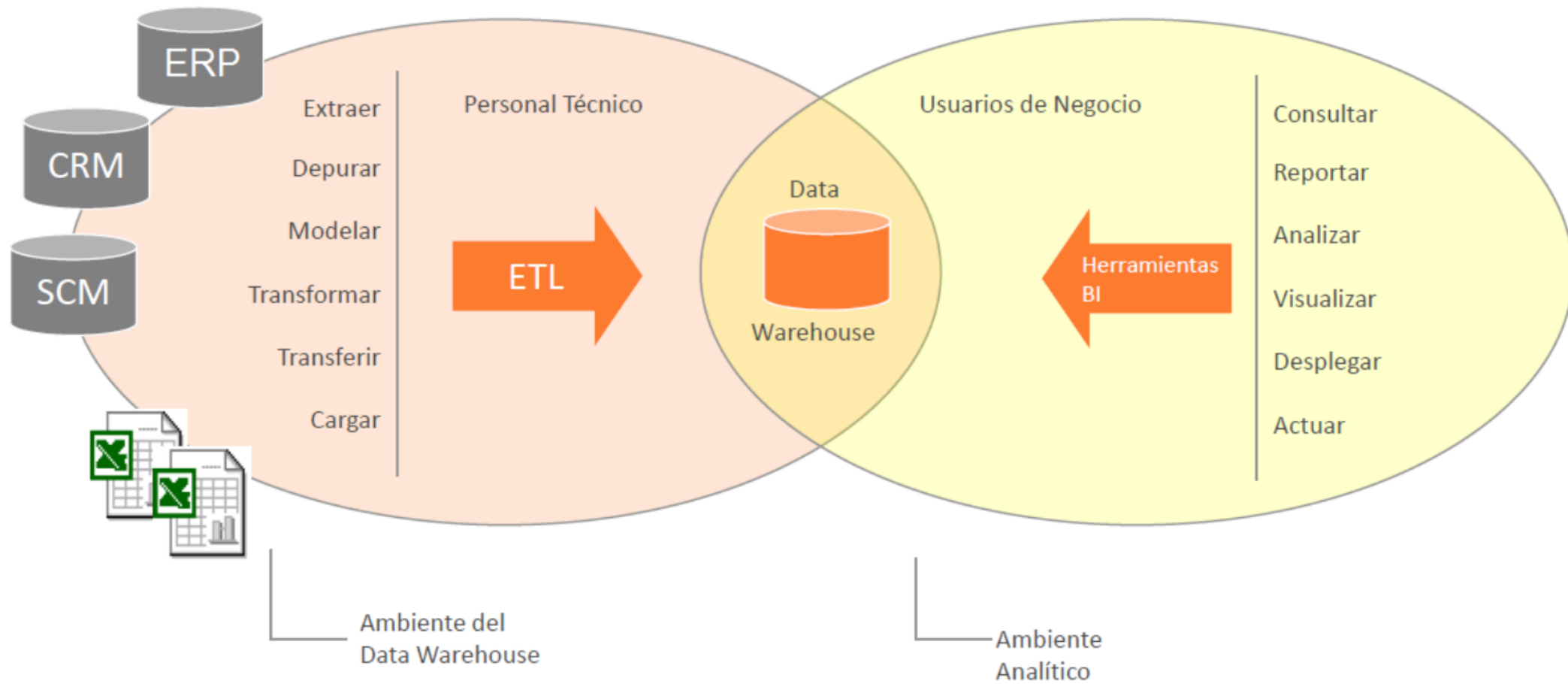
Definición Proyecto de BI

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

Componentes del BI



Componentes del BI



Modelo IBM

- Este modelo se enfoca en responder cuatro preguntas básicas:
- ¿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?



El modelo de gestión de calidad de datos de IBM requiere de tres disciplinas para obtener resultados, sin embargo, se soportan en dos disciplinas 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.

Los proy
caracter
tienen u
dificulta
factores
En concre
aspectos

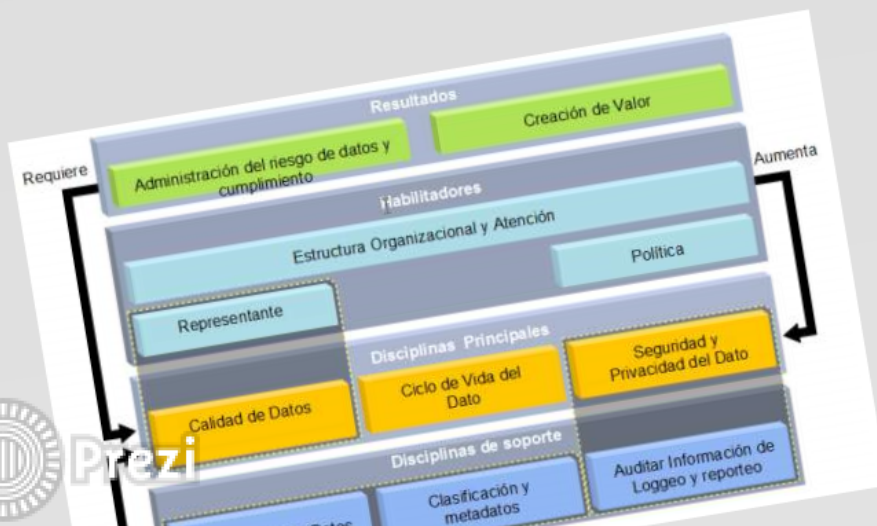
Compon



Modelo IBM

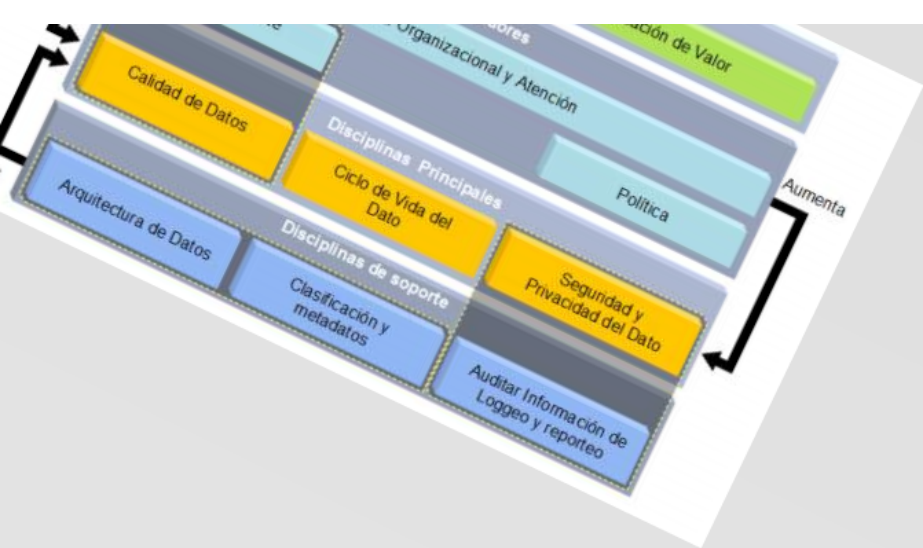
Este modelo se enfoca en responder cuatro preguntas básicas:

- ¿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?



de calidad de datos de IBM requiere de tres resultados, sin embargo, se soportan en dos datos, Clasificación y metadatos. Los en este caso su representante, la técnicas aumentan la seguridad y futuro la auditoría de los





¿Qué disciplinas básicas:
sería para tomar la mejor
ente protegidos?
en base al entendimiento actual?

El modelo de gestión de calidad de datos de IBM requiere de tres disciplinas para obtener resultados, sin embargo, se soportan 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.

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.

Fases de la gestión de calidad de datos

- Investigación
- Estandarización
- Coincidencia
- Supervivencia



Example: 123 St. Virginia St.

123	St.	Virginia	St.
House	Street Type	Appl.	Street Type
123	St.	Virginia	St.

• Parsing:
 • Separates multi-valued fields into individual pieces

Lexical Analysis:
 Determine the relative significance of individual pieces

Context Sensitive:
 Address: 123456 789 1011121314151617181920

House	Street Name	Street Type
123	St. Virginia	St.

Input Data:
 Wash. Trar. Metr. and Metrol
 1234 LINDCOLN PARK 1234
 5678 CHERRYLIT 1234 5678
 9101 2345 6789 1011 1234 5678
 9101 2345 6789 1011 1234 5678
 9101 2345 6789 1011 1234 5678
 9101 2345 6789 1011 1234 5678
 9101 2345 6789 1011 1234 5678

Go from this ...

... to this!

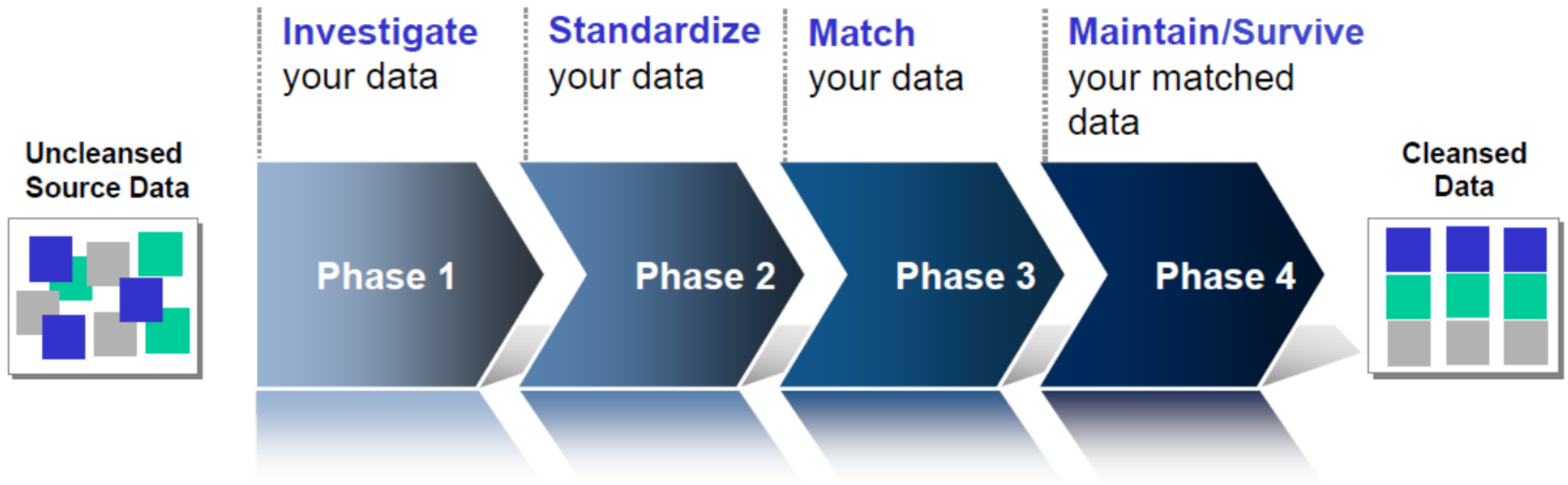
State	ZIPCODE	Block #	Mail	SECTOR	Subsector	Mail #	Mail # Y1	Additional Address Lines
LINCOLN	LINCOLN	1234	MARK	MARK	1234	5	1996	
CHERRYLIT	CHERRYLIT	1234	MARK	MARK	1234	5	1997	
POSSGT	RIAS	1234	RIAS	RIAS	1234	1999	1991	DETST
WESMAN	WESMAN	1234	WESMAN	WESMAN	1234	2000	2000	
POSD	PAD	1234	POSD	POSD	1234	2000	2000	
CHERRYLIT	CHERRYLIT	1234	CHERRYLIT	CHERRYLIT	1234	1999	1999	
WOLVO	WOLVO	1234	C	C	1234	20	2001	CONVERTIBLE

Input Data

State	ZIPCODE	Block #	Mail	SECTOR	Subsector	Mail #	Mail # Y1	Additional Address Lines
LINCOLN	LINCOLN	1234	MARK	MARK	1234	5	1996	
CHERRYLIT	CHERRYLIT	1234	MARK	MARK	1234	5	1997	
POSSGT	RIAS	1234	RIAS	RIAS	1234	1999	1991	DETST
WESMAN	WESMAN	1234	WESMAN	WESMAN	1234	2000	2000	
POSD	PAD	1234	POSD	POSD	1234	2000	2000	
CHERRYLIT	CHERRYLIT	1234	CHERRYLIT	CHERRYLIT	1234	1999	1999	
WOLVO	WOLVO	1234	C	C	1234	20	2001	CONVERTIBLE

Use a special program to "clean" the "raw" data

de datos



Example: *123 St. Virginia St.*

▪ Parsing:

- Separates multi-valued fields into individual pieces

123	St.	Virginia	St.
-----	-----	----------	-----

Number	Street Type	Alpha	Street Type
123	St.	Virginia	St.

Lexical Analysis:

Determines business significance of individual pieces

Number	Street Name	Street Type
123	St. Virginia	St.

Context Sensitive:

Identifies various data structures and content

Input Data:

Vehicle Year, Make and Model

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

Go from this ...

... to this!

Standardized Data:

Make	NYSIIS	RSoundex	Model	NYSIIS	RSoundex	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	EXPLARAR	R641		2002	
CHEVROLET	CAFRALAT	T461	SUBURBAN	SABARBAN	N161	1500	1998	
VOLVO	VALV	0141	C	C	C000	70	2001	CONVERTIBLE

Input Data

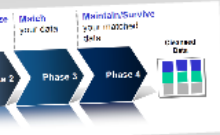
SETID	SSN	Name	Address	City	State	Zip Code
1	NULL	Jerome David Salinger	Holden Caulfield Hwy	Agerstown	PA	19102
1	NULL	J. David Salinger	51 Holden Caulfield Hwy	Agerstown	PA	19102-1919
1	123-45-6789	J.D. Salinger	Holden Caulfield Hwy	Agerstown	PA	19102

Survivorship Output

SETID	SSN	Name	Address	City	State	Zip Code
1	123-45-6789	Jerome David Salinger	51 Holden Caulfield 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	123-45-6789	Jerome David Salinger	51 Holden Caulfield Hwy	Agerstown	PA	19102-1919
1	123-45-6789	Jerome David Salinger	51 Holden Caulfield Hwy	Agerstown	PA	19102-1919
1	123-45-6789	Jerome David Salinger	51 Holden Caulfield Hwy	Agerstown	PA	19102-1919



A table with multiple columns and rows, containing various data points and text. The table is tilted and partially obscured by a blue circular graphic.

Metodología

- 1) Análisis y monitoreo de la calidad de los datos
- 2) Estructuración de reglas de datos y conjuntos de reglas
- 3) Estándares de nomenclatura
- 4) Análisis de reglas de datos
- 5) Medidas
- 6) Resultados del monitoreo
- 7) Implementación de reglas, conjuntos de reglas y medidas
- 8) Gestión de un ambiente de reglas de calidad de datos

La metodología de IBM para calidad de datos
propone ocho criterios

Metodología

- 1) Análisis y monitoreo de la calidad de los datos
 - 2) Estructuración de reglas de datos y conjuntos de reglas
 - 3) Estándares de nomenclatura
 - 4) Análisis de reglas de datos
 - 5) Medidas
 - 6) Resultados del monitoreo
 - 7) Implementación de reglas, conjuntos de

Análisis de reglas de datos

5) Medidas

6) Resultados del monitoreo

7) Implementación de reglas, conjuntos de reglas y medidas

8) Gestión de un ambiente de reglas de calidad de datos

Conclusiones y Recomendaciones

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.

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 de la calidad de datos junto con la Metodología descrita en un caso de estudio.

Muchas Gracias.

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.

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.



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.

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 de la calidad de datos junto con la Metodología descrita en un caso de estudio.

- 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 de la calidad de datos junto con la Metodología descrita en un caso de estudio.

Muchas Gracias.