



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

**DESARROLLO DE UN CASO DE ESTUDIO APLICANDO
EL 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 - Julio 2014



AGENDA

- Introducción
- Infosphere Information Server
- Implementación del 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:

- ¿Se ha desarrollado en DWConsultware un caso de estudio aplicando el Modelo de Gestión de Calidad de Datos?
- En el marco de un proyecto específico de Calidad de Datos: ¿Cómo identificar y aplicar las etapas de desarrollo para este tipo de proyectos según el modelo de gestión IBM Data Quality?

Objetivos

Objetivo General: Desarrollar un caso de estudio aplicando el modelo de Gestión de Calidad de Datos basado en la metodología IBM Data Quality en un proyecto de Business Intelligence de la empresa DWConsultware..

Objetivos Específicos:

- Implementar los procesos del Modelo de Gestión de Calidad de Datos en la herramienta IBM Infosphere Information Server para el caso de estudio.
- Determinar el nivel de madurez del caso de estudio mediante la guía de evaluación del Modelo de Madurez de Business Intelligence de TDWI .
- Aplicar las directrices de implementación de los procesos de mejora y Gestión de Calidad de datos para el caso de estudio.

Alcance

Este proyecto culmina con la presentación de un caso de estudio aplicando el Modelo de Gestión de Calidad de Datos desarrollado en el proyecto previamente presentado como primera parte de esta tesis.

En este caso de estudio se incluye el desarrollo de los procesos de Gestión de Calidad de Datos mediante el uso de la herramienta IBM Infosphere Information Server, estos son:

- Perfilamiento
- Investigación
- Estandarización
- Coincidencia
- Supervivencia
- Monitoreo

En este trabajo no se incluye ningún otro proceso de Inteligencia de Negocios, tales como: carga de tabla de hecho y dimensiones, verificación de integridad, etc.

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:

- ¿Se ha desarrollado en DWConsulware un caso de estudio aplicando el Modelo de Gestión de Calidad de Datos?
- En el marco de un proyecto específico de Calidad de Datos: ¿Cómo identificar y aplicar las etapas de desarrollo para este tipo de proyectos según el modelo de gestión IBM Data Quality?

Objetivos

Objetivo General: Desarrollar un caso de estudio aplicando el modelo de Gestión de Calidad de Datos basado en la metodología IBM Data Quality en un proyecto de Business Intelligence de la empresa DWConsultware..

Objetivos Específicos:

- Implementar los procesos del Modelo de Gestión de Calidad de Datos en la herramienta IBM Infosphere Information Server para el caso de estudio.
- Determinar el nivel de madurez del caso de estudio mediante la guía de evaluación del Modelo de Madurez de Business Intelligence de TDWI .
- Aplicar las directrices de implementación de los procesos de mejora y Gestión de Calidad de datos para el caso de estudio.

Objetivos

Objetivo General: Desarrollar un caso de estudio aplicando el modelo de Gestión de Calidad de Datos basado en la metodología IBM Data Quality en un proyecto de Business Intelligence de la empresa DWConsultware..

Objetivos Específicos:

- Implementar los procesos del Modelo de Gestión de Calidad de Datos en la herramienta IBM InfoSphere Information Server para el caso de estudio.
- Determinar el nivel de madurez del caso de estudio mediante la guía de evaluación del Modelo de Madurez de Business Intelligence de TDWI .



Intelligence de la empresa DWConsulware..

Objetivos Específicos:

- Implementar los procesos del Modelo de Gestión de Calidad de Datos en la herramienta IBM Infosphere Information Server para el caso de estudio.
- Determinar el nivel de madurez del caso de estudio mediante la guía de evaluación del Modelo de Madurez de Business Intelligence de TDWI .
- Aplicar las directrices de implementación de los procesos de mejora y Gestión de Calidad de datos para el caso de estudio.

Alcance

Este proyecto culmina con la presentación de un caso de estudio aplicando el Modelo de Gestión de Calidad de Datos desarrollado en el proyecto previamente presentado como primera parte de esta tesis.

En este caso de estudio se incluye el desarrollo de los procesos de Gestión de Calidad de Datos mediante el uso de la herramienta IBM Infosphere Information Server, estos son:

- Perfilamiento
- Investigación
- Estandarización
- Coincidencia
- Supervivencia
- Monitoreo

En este trabajo no se incluye ningún otro proceso de Inteligencia de Negocios, tales como: carga de tabla de hecho y dimensiones, verificación de integridad, etc.

InfoSphere Information Server

- Introducción InfoSphere Information Server
- InfoSphere Information Server for Data Quality
- Arquitectura de la Solución

Introducción InfoSphere Information Server

IBM InfoSphere Information Server para la Calidad de los Datos es una suite de aplicaciones que permiten mejorar y monitorar la calidad de los datos. Proporciona altas capacidades de procesamiento que permiten ejecutar las tareas antes mencionadas, convirtiendo datos en información de calidad. A través del análisis, la limpieza, el control y la gestión de calidad de datos, se puede tomar mejores decisiones y optimizar la ejecución de procesos de negocio.



1. InfoSphere Information Server Business Information Exchange proporciona funciones que le ayudan a establecer y mantener conocimientos de la empresa a medida que cambia el proyecto.
2. InfoSphere Information Server for Data Integration proporciona funciones que permiten la transformación y entrega de datos a diferentes aplicaciones.
3. InfoSphere Information Server for Data Quality proporciona amplias funciones que permiten gestionar la calidad de los datos.

COMPONENTE	DESCRIPCIÓN	ESTADO	ACTIVADO
InfoSphere Information Server
InfoSphere Information Server for Data Quality
InfoSphere Information Server for Data Integration
InfoSphere Information Server Business Information Exchange
InfoSphere Information Server for Data Governance
InfoSphere Information Server for Data Protection
InfoSphere Information Server for Data Migration
InfoSphere Information Server for Data Archiving
InfoSphere Information Server for Data Backup
InfoSphere Information Server for Data Recovery
InfoSphere Information Server for Data Purification
InfoSphere Information Server for Data Deduplication
InfoSphere Information Server for Data Compression
InfoSphere Information Server for Data Encryption
InfoSphere Information Server for Data Decryption
InfoSphere Information Server for Data Masking
InfoSphere Information Server for Data Anonymization
InfoSphere Information Server for Data Archiving
InfoSphere Information Server for Data Backup
InfoSphere Information Server for Data Recovery
InfoSphere Information Server for Data Purification
InfoSphere Information Server for Data Deduplication
InfoSphere Information Server for Data Compression
InfoSphere Information Server for Data Encryption
InfoSphere Information Server for Data Decryption
InfoSphere Information Server for Data Masking
InfoSphere Information Server for Data Anonymization

InfoSphere Information Server for Data Quality

InfoSphere Information Server for Data Quality es la suite que va a ser utilizada en el proyecto de Caso de Estudio, la suite proporciona funcionalidades que permiten limpiar los datos y controlar la calidad de datos, convirtiendo los datos en información confiable.

COMPONENTE	DESCRIPCIÓN	ESTADO	ACTIVADO
InfoSphere Information Server
InfoSphere Information Server for Data Quality
InfoSphere Information Server for Data Integration
InfoSphere Information Server Business Information Exchange
InfoSphere Information Server for Data Governance
InfoSphere Information Server for Data Protection
InfoSphere Information Server for Data Migration
InfoSphere Information Server for Data Archiving
InfoSphere Information Server for Data Backup
InfoSphere Information Server for Data Recovery
InfoSphere Information Server for Data Purification
InfoSphere Information Server for Data Deduplication
InfoSphere Information Server for Data Compression
InfoSphere Information Server for Data Encryption
InfoSphere Information Server for Data Decryption
InfoSphere Information Server for Data Masking
InfoSphere Information Server for Data Anonymization



Arquitectura de la Solución

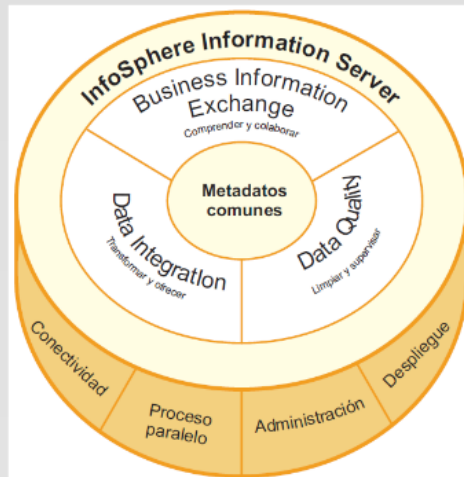
IBM InfoSphere Information Server proporciona una arquitectura unificada que funciona con todos los tipos de integración de información. Los servicios compartidos, los procesos paralelos unificados y los metadatos unificados son la base de la arquitectura del servidor, la cual está orientada a los servicios, lo que permite utilizar IBM InfoSphere Information Server en las arquitecturas en evolución como las orientadas a servicios.



Introducción

InfoSphere Information Server

IBM InfoSphere Information Server para la Calidad de los Datos es una suite de aplicaciones que permiten mejorar y monitorear la calidad de los datos. Proporciona altas capacidades de procesamiento que permiten ejecutar las tareas antes mencionadas, convirtiendo datos en información de calidad. A través del análisis, la limpieza, el control y la gestión de calidad de datos, se puede tomar mejores decisiones y optimizar la ejecución de procesos de negocio



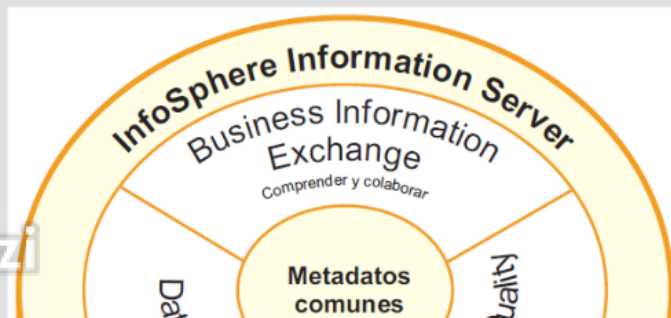
- 1. InfoSphere Information Server Business Information Exchange** proporciona funciones que le ayudan a establecer y mantener conocimientos de la empresa a medida que cambia el proyecto.
- 2. InfoSphere Information Server for Data Integration** proporciona funciones que permiten la transformación y entrega de datos a diferentes aplicaciones.
- 3. InfoSphere Information Server for Data Quality** proporciona amplias funciones que permiten gestionar la calidad de los datos.

COMPONENTE	BUSINESS INFORMATION EXCHANGE	DATA INTEGRATION	DATA QUALITY
InfoSphere Blueprint Director	X	X	X
InfoSphere Diacoverly	X	X	
InfoSphere Metadata Workbench	X		
InfoSphere Data Architect	X		
InfoSphere Business Glossary		X	X
InfoSphere Business Glossary Anywhere		X	X
InfoSphere Information Analyzer		X	X
InfoSphere QualityStage		X	X
InfoSphere Information Services Director		X	
InfoSphere DataStage & QualityStage		X	
Designer		X	
InfoSphere FastTrack		X	

Introducción

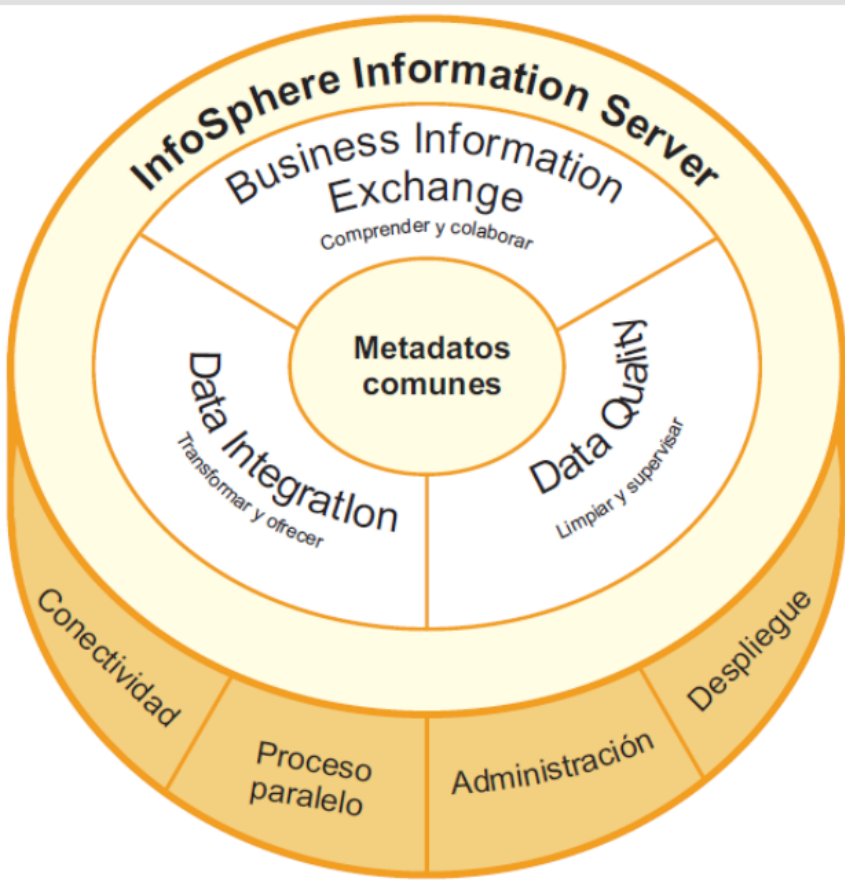
InfoSphere Information Server

IBM InfoSphere Information Server para la Calidad de los Datos es una suite de aplicaciones que permiten mejorar y monitorear la calidad de los datos. Proporciona altas capacidades de procesamiento que permiten ejecutar las tareas antes mencionadas, convirtiendo datos en información de calidad. A través del análisis, la limpieza, el control y la gestión de calidad de datos, se puede tomar mejores decisiones y optimizar la ejecución de procesos de negocio



1. **InfoSphere Information Server Business Information Exchange** proporciona funciones que le ayudan a establecer y mantener conocimientos de la empresa a medida que cambia el proyecto.
2. **InfoSphere Information Server for Data**

permiten ejecutar las tareas antes mencionadas, convirtiendo datos en información de calidad. A través del análisis, la limpieza, el control y la gestión de calidad de datos, se puede tomar mejores decisiones y optimizar la ejecución de procesos de negocio



1. **InfoSphere Information Server Business Information Exchange** proporciona funciones que le ayudan a establecer y mantener conocimientos de la empresa a medida que cambia el proyecto.

2. **InfoSphere Information Server for Data Integration** proporciona funciones que permiten la transformación y entrega de datos a diferentes aplicaciones.

3. **InfoSphere Information Server for Data Quality** proporciona amplias funciones que permiten gestionar la calidad de los datos.

COMPONENTE	BUSINESS INFORMATION EXCHANGE	DATA INTEGRATION	DATA QUALITY
InfoSphere Blueprint Director	X	X	X
InfoSphere Discovery	X	X	
InfoSphere Metadata Workbench	X		
InfoSphere Data Architect	X		
InfoSphere Business Glossary		X	X
InfoSphere Business Glossary Anywhere		X	X

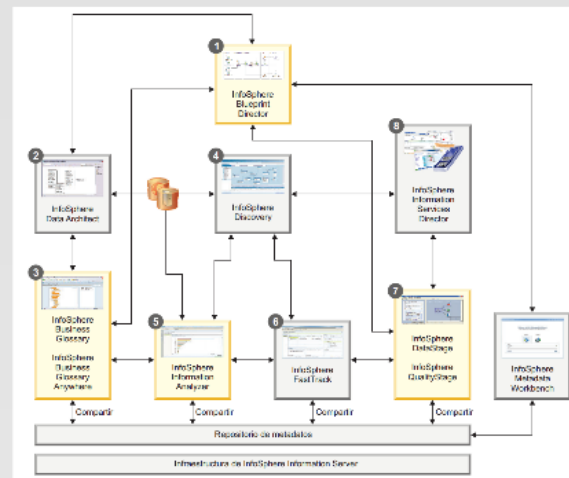
permiten gestionar la calidad de

COMPONENTE	BUSINESS INFORMATION EXCHANGE	DATA INTEGRATION	DATA QUALITY
<u>InfoSphere Blueprint Director</u>	X	X	X
<u>InfoSphere Discovery</u>	X	X	
<u>InfoSphere Metadata Workbench</u>	X		
<u>InfoSphere Data Architect</u>	X		
<u>InfoSphere Business Glossary</u>		X	X
<u>InfoSphere Business Glossary Anywhere</u>		X	X
<u>InfoSphere Information Analyzer</u>		X	X
<u>InfoSphere QualityStage</u>		X	X
<u>InfoSphere Information Services Director</u>		X	
<u>InfoSphere DataStage & QualityStage Designer</u>		X	
<u>InfoSphere FastTrack</u>		X	

InfoSphere Information Server for Data Quality

InfoSphere Information Server for Data Quality es la suite que va a ser utilizada en el proyecto de Caso de Estudio, la suite proporciona funcionalidades que permiten limpiar los datos y controlar la calidad de datos, convirtiendo los datos en información confiable.

COMPONENTE	INFOSPHERE INFORMATION SERVER FOR DATA QUALITY
InfoSphere Blueprint Director	X
InfoSphere Business Glossary	X
InfoSphere Business Glossary Anywhere	X
InfoSphere Information Analyzer	X
InfoSphere QualityStage	X



InfoSphere Information Server for Data Quality

InfoSphere Information Server for Data Quality es la suite que va a ser utilizada en el proyecto de Caso de Estudio, la suite proporciona funcionalidades que permiten limpiar los datos y controlar la calidad de datos, convirtiendo los datos en información confiable.

COMPONENTE	INFOSPHERE INFORMATION SERVER FOR DATA QUALITY
InfoSphere Blueprint Director	X
InfoSphere Business Glossary	X
InfoSphere Business Glossary Anywhere	X
InfoSphere Information Analyzer	X
InfoSphere QualityStage	X

Información confiable.

COMPONENTE

INFOSPHERE INFORMATION SERVER FOR DATA QUALITY

InfoSphere Blueprint Director

X

InfoSphere Business Glossary

X

InfoSphere Business Glossary Anywhere

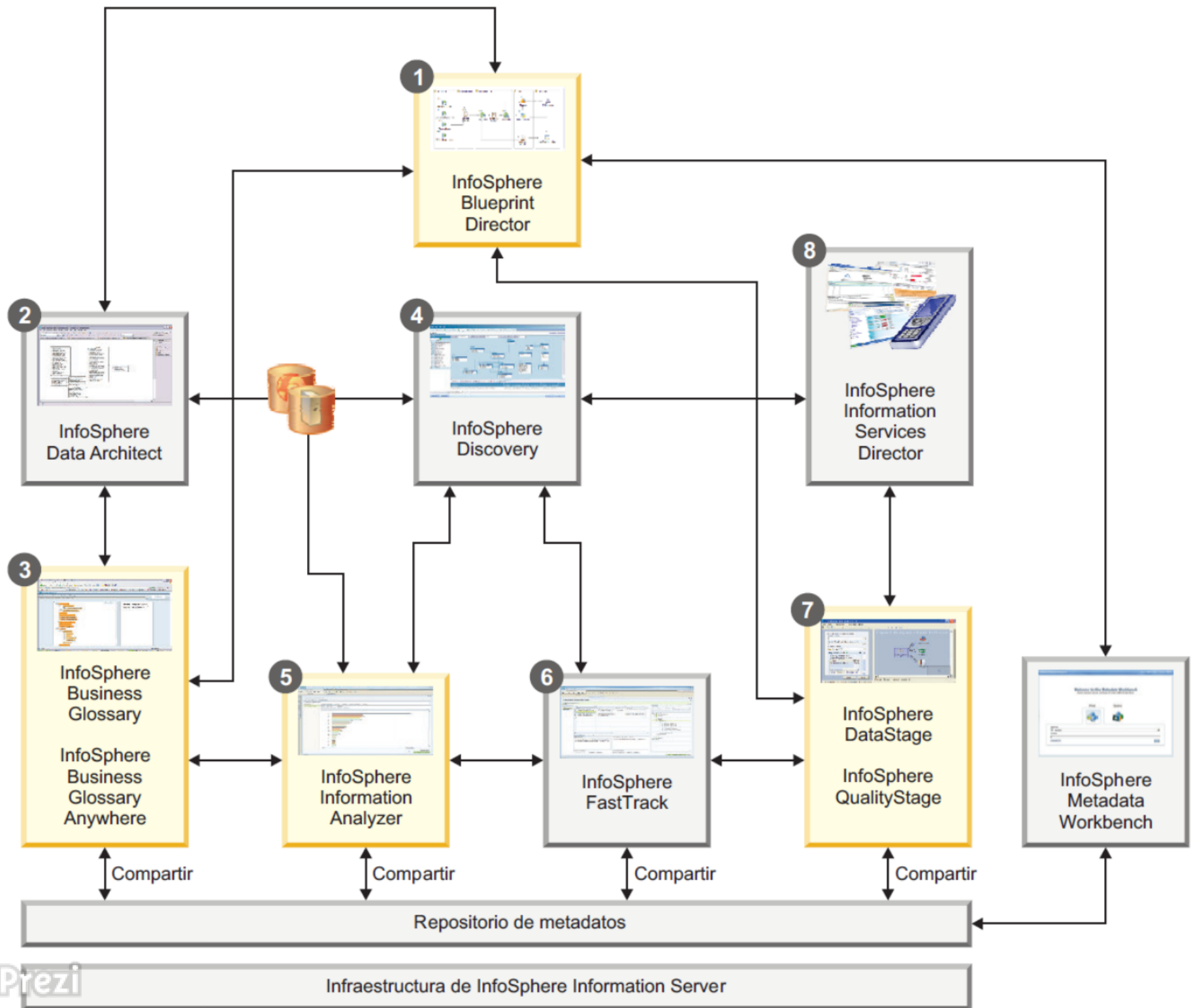
X

InfoSphere Information Analyzer

X

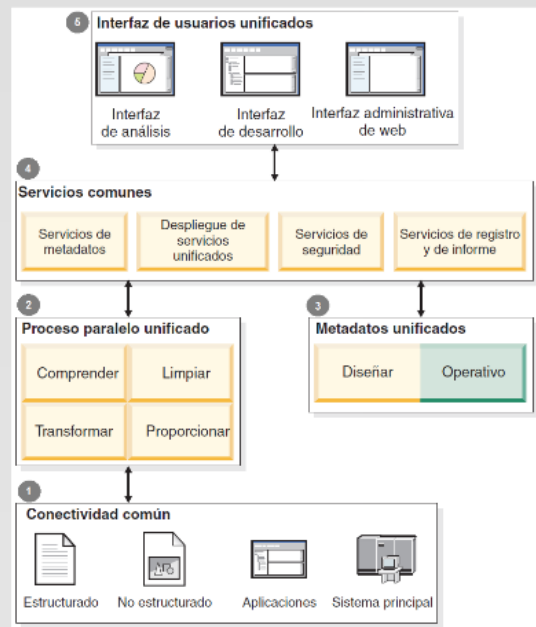
InfoSphere QualityStage

X



Arquitectura de la Solución

IBM InfoSphere Information Server proporciona una arquitectura unificada que funciona con todos los tipos de integración de información. Los servicios compartidos, los procesos paralelos unificados y los metadatos unificados son la base de la arquitectura del servidor, la cual está orientada a los servicios, lo que permite utilizar IBM InfoSphere Information Server en las arquitecturas en evolución como las orientadas a servicios.



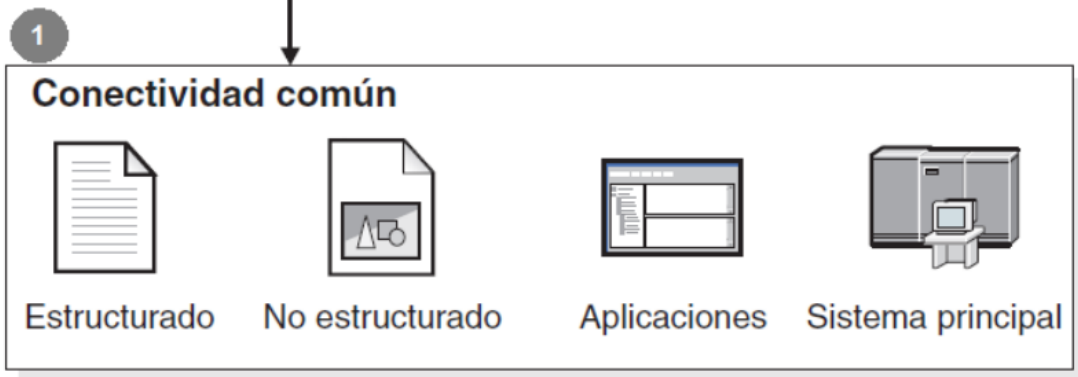
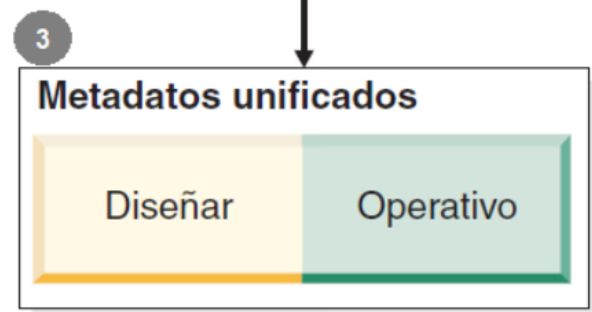
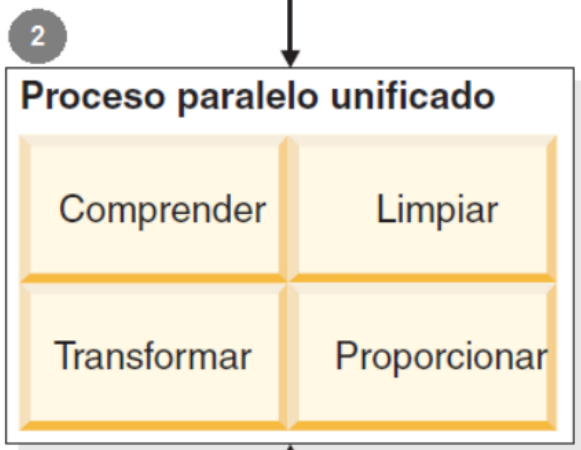
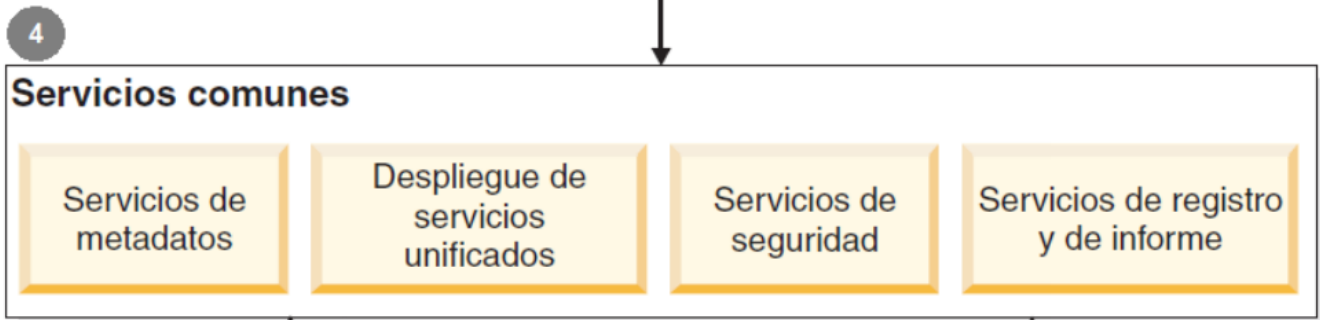
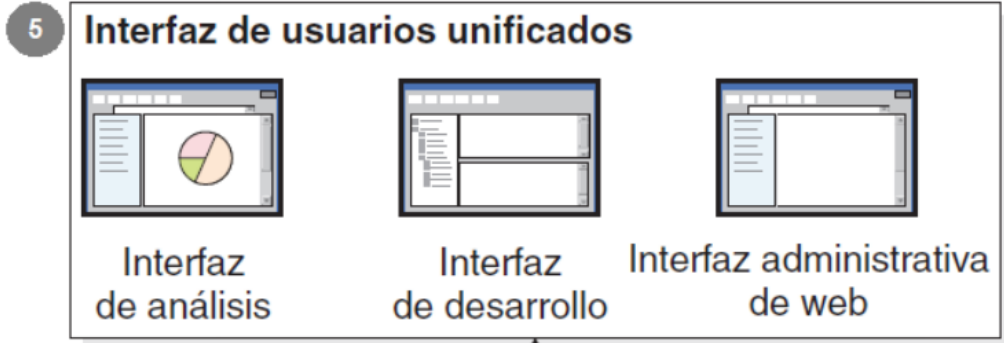
Arquitectura de la Solución

IBM InfoSphere Information Server proporciona una arquitectura unificada que funciona con todos los tipos de integración de información. Los servicios compartidos, los procesos paralelos unificados y los metadatos unificados son la base de la arquitectura del servidor, la cual está orientada a los servicios, lo que permite utilizar IBM InfoSphere Information Server en las arquitecturas en evolución como las orientadas a servicios.

5

Interfaz de usuarios unificados



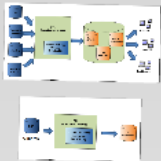


Implementación del Modelo de Gestión

- Situación Actual y Alcance
- Desarrollo del Proyecto
- Ejemplos
- Resumen de Resultados

Situación Actual y Alcance

La empresa actualmente posee bases de datos centralizadas con la información de las demandas, facturas de datos, entre otras en un Sistema de Maestros, un Sistema de Transacciones y un Base de Datos. El proceso de EPC se ha realizado mediante la herramienta de Integrate Database, como proceso manual en la exportación de datos de los sistemas fuente y transferir a los respectivos centralizados de la empresa aplicando una lógica de negocio. Este información luego es cargada a través de herramientas para la carga de datos. La figura muestra la arquitectura de la organización.



Planificación del Proyecto

Utilizando IBM InfoSphere Migration Director se obtiene la Actualización de Datos del Proyecto, como consiste en generar un diseño de la arquitectura del proyecto para ser usada como guía para el desarrollo de los casos particulares.



Desarrollo del Proyecto

Una vez que se ha definido la situación actual de la empresa y se ha determinado como la arquitectura actual de BI de la organización como la arquitectura centralizada con la Gestión de Calidad de Datos se puede iniciar el desarrollo del proyecto para se compare los evaluaciones del nivel de Maestros de la aplicación el Modelo de Maestros TDM y la etapa de Gestión de Calidad de Datos. Posteriormente, las evaluaciones se realizan antes y después de aplicar el Modelo de Gestión de Calidad de Datos y permite determinar el espacio de la Gestión de Calidad de Datos en la organización.

- Modelo de Gestión de Calidad de Datos
- Modelo de Maestros TDM

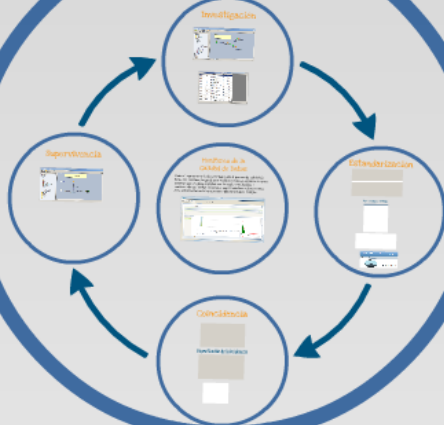
Perfilamiento

En esta etapa del ciclo de perfilado se realiza mediante la condición actual de los datos del cliente el contenido y la calidad de los datos generados en el sistema de gestión de datos.

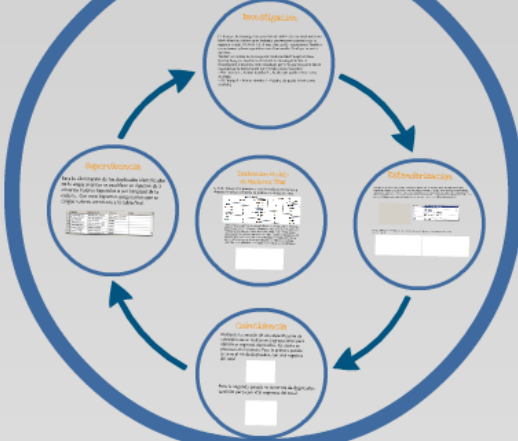
- Validación de datos de origen
- Validación de destino
- Análisis de calidad



Fase de la Calidad de Datos



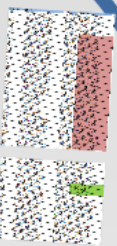
Resultados



Ejemplos

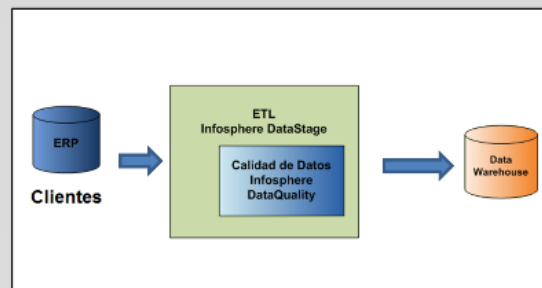
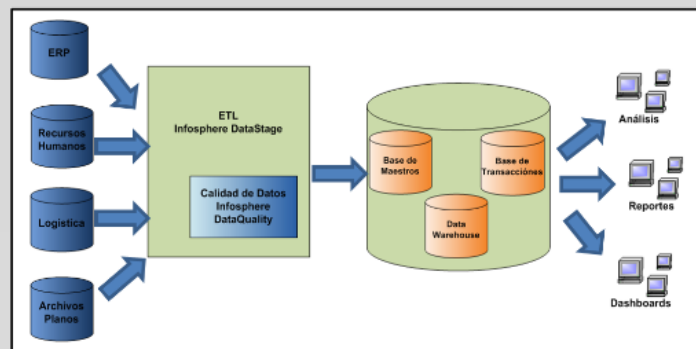
Existen varios casos en los que se presentan duplicados en la información, a continuación se describe un ejemplo de un cliente el que se identificó duplicados. Antes del proceso se tenía el cliente CIRUC 09A00754001 con duplicados en cuanto a dirección tal como indica la Figura marcada en verde y rojo.

Una vez ejecutado el proceso se puede identificar que los clientes que estaban duplicados se simplificaron a uno solo como indica la Figura en los colores respectivos.

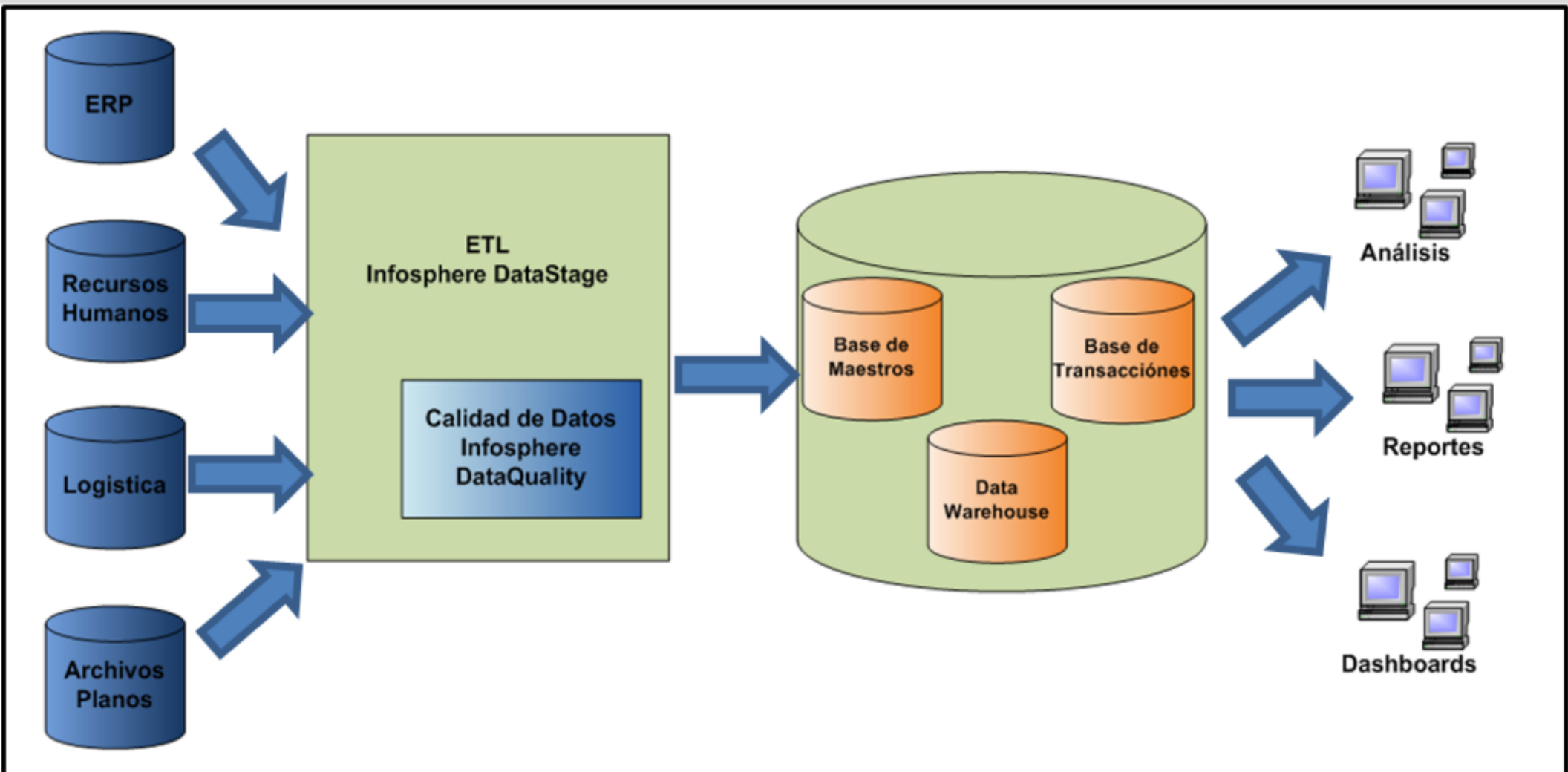


Situación Actual y Alcance

La empresa actualmente posee bases de datos centralizadas con la información de las diversas fuentes de datos, entre ellas están un Sistema de Maestros, un Sistema de Transacciones y un Data Warehouse. El proceso de ETL se lo realiza mediante la herramienta de Infosphere DataStage, este proceso consiste en tomar los datos de los sistemas fuente y llevarlos a los repositorios centralizados de la empresa aplicando una lógica de negocio. Esta información luego es consumida a través de herramientas para la toma de decisiones. La Figura describe la arquitectura de la organización.



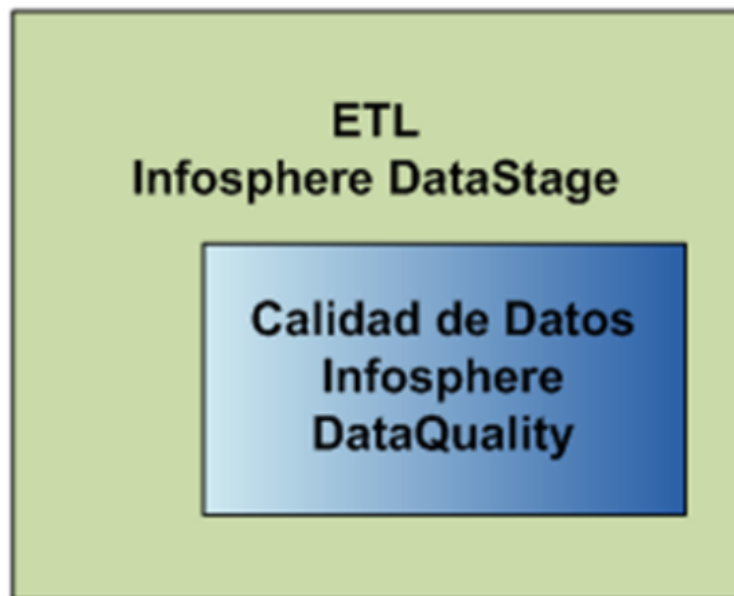
para la toma de decisiones. La Figura describe la a





ERP

Clientes



Data
Warehouse

Desarrollo del Proyecto

Una vez que se ha definido la situación actual de la empresa y se ha determinado tanto la arquitectura actual de BI de la organización como la arquitectura combinada con la Gestión de Calidad de Datos es posible iniciar el desarrollo del proyecto, este se compone dos evaluaciones del nivel de Madurez de BI utilizando el Modelo de Madurez TDWI y la etapa de Gestión de Calidad de Datos propiamente, las evaluaciones se realizan antes y después de aplicar el Modelo de Gestión de Calidad de Datos y permiten determinar el impacto de la Gestión de Calidad de Datos en la organización.

- Modelo de Gestión de Calidad de Datos
- Modelo de Madurez TDWI

Planificación del Proyecto

Utilizando IBM Infosphere Blueprint Director se obtiene la documentación técnica del proyecto, esto consiste en generar un diseño de la arquitectura del proyecto para ser usada como guía para el desarrollo de las etapas posteriores.

El diseño de la arquitectura facilita la comunicación entre los miembros del proyecto la siguiente figura indica la arquitectura de alto nivel para el proyecto de caso de estudio.



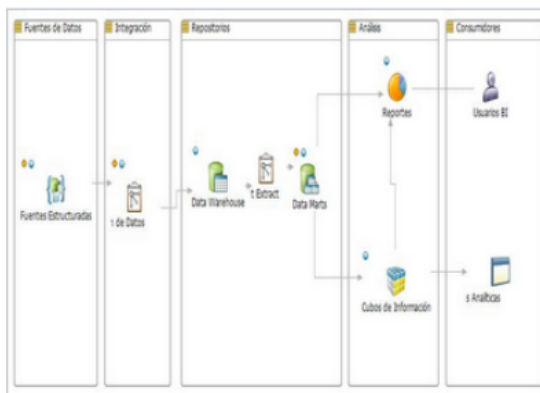
Plan Detalles



Editar Suprimir

Proyecto Caso de Estudio DQ

Imagen



Pulse para iniciar el visor de planes

Descripción Proyecto de Calidad de Datos

Versión 1

▼ Información general

Created By dqa dqa

Fecha de creación 06/04/2014 22:44:06



Cliente

Search



Category Details



Dimensiones Compartidas

Short Description Se incluyen las entidades que forman parte de varios modelos

▼ General Information

Created By dqa dqa

Created On Apr 6, 2014 10:18:52 PM

Last Modified By dqa dqa

Last Modified On Apr 7, 2014 1:12:45 AM

▶ Subcategories (1)

▼ Terms (3)

 Cliente

 Producto

 Proveedor

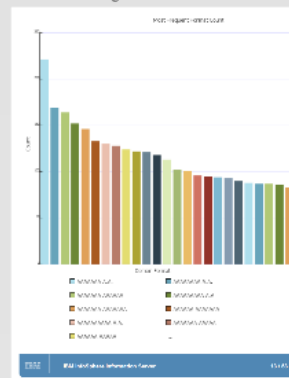


Perfilamiento

En esta etapa del caso de estudio se requiere entender la condición actual de los datos así como el contenido y la estructura de los mismos, para esto se debe evaluar los siguientes aspectos:

- Validez de los tipos de datos.
- Consistencia de formatos.
- Estructura y contenido

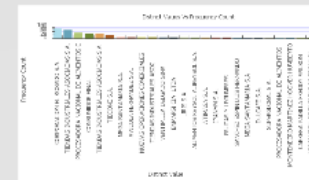
En el gráfico de barras se puede observar que existen varios formatos para el nombre y no solo de personas sino de negocios también. El reporte también presenta los datos de manera tabular, donde indica uno de los formatos más utilizados con 221 repeticiones que indica el nombre de un negocio.



Nombre	Count	Formato	Formato	Formato
...	221

En el gráfico de barras se puede observar que existen algunos valores repetidos que debemos depurar en las siguientes etapas de la calidad de datos.

El reporte también indica los clientes que fueron reconocidos como repetidos.



ID Cliente	Count	Formato	Formato
...

Perfilamiento

En esta etapa del caso de estudio se requiere entender la condición actual de los datos así como el contenido y la estructura de los mismos, para esto se debe evaluar los siguientes aspectos:

Validez de los tipos de datos.

Consistencia de formatos.

Estructura y contenido

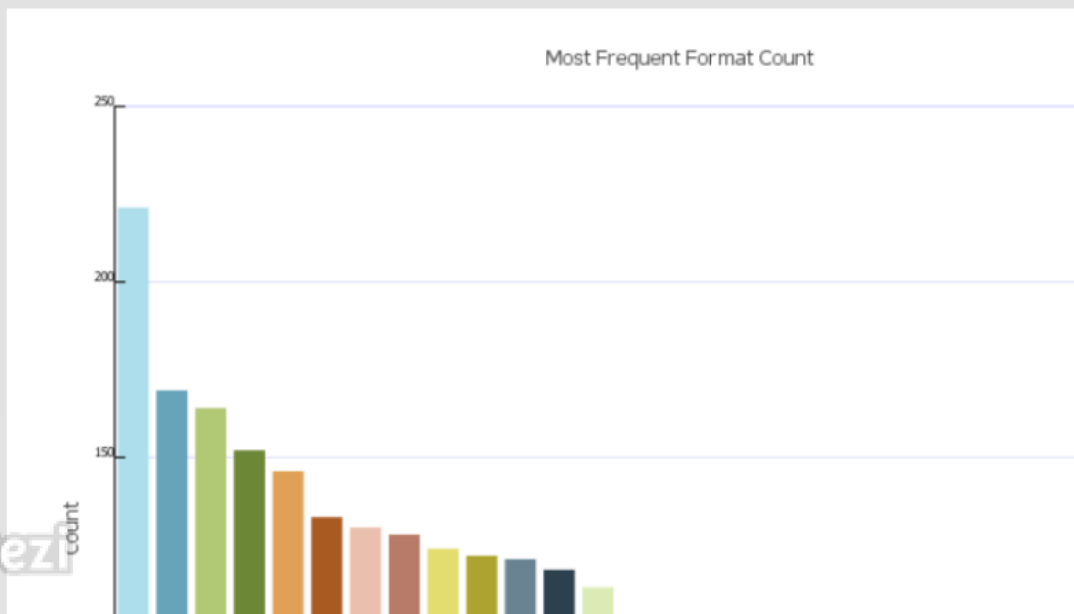
En el gráfico de barras se puede observar que existen varios formatos para el nombre y no solo de personas sino de negocios también.

El reporte también presenta los datos de manera tabular, donde indica uno de los formatos más utilizados con 221 repeticiones que indica el nombre de un negocio.

En el gráfico de barras se puede observar que existen valores repetidos que debemos depurar en las siguientes etapas de la calidad de datos.

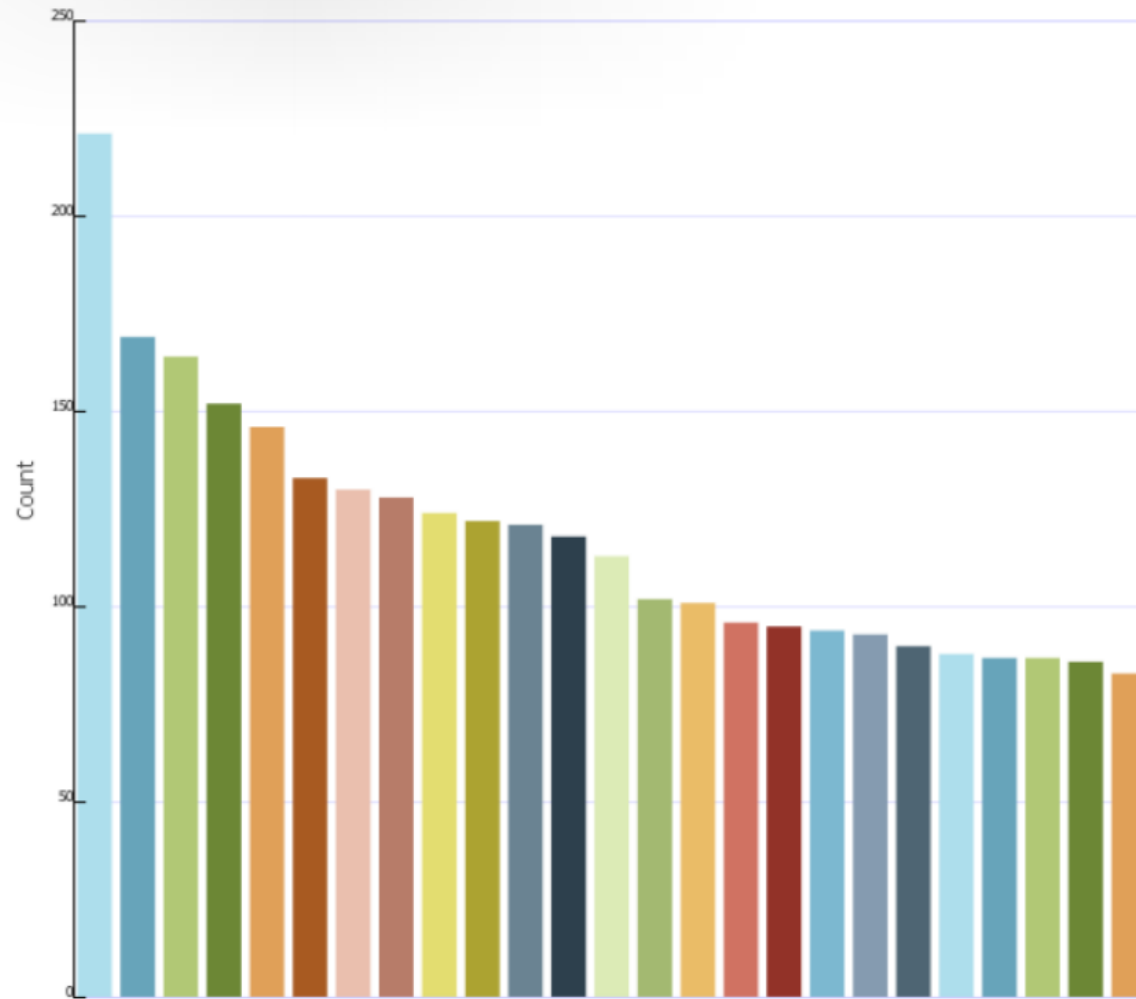
El reporte también indica los clientes que fueron recibidos como repetidos.

En el gráfico de barras se puede observar que existen varios formatos para el nombre y no solo de personas sino de negocios también. El reporte también presenta los datos de manera tabular, donde indica uno de los formatos más utilizados con 221 repeticiones que indica el nombre de un negocio.



Column Level Details		
Format	Count	Total Rows %
AAAAAA A.A.	221	0.66245017

Most Frequent Format Count



- AAAAAAA A.A.
- AAAAAAA AAAAAA
- AAAAAAA AAAAAA
- AAAAAAAA A.A.
- AAAAAA AAAAA
- AAAAAAA A.A.
- AAAAAAAA A.A.
- AAAAAA AAAAAA
- AAAAAA AAAAA
- ...

Column Level D
Format
AAAAAAA A.A.
AAAAAAA A.A.
AAAAAAA AAAAA



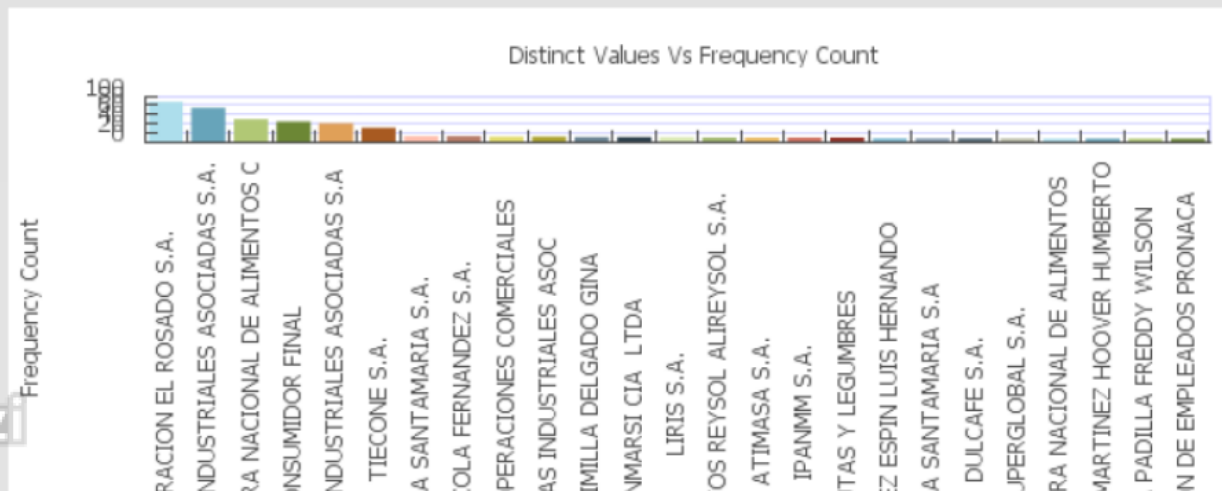
IBM, the IBM logo, and IBM InfoSphere Information Server are trademarks of International Business Machines Corporation in the United States, other countries or both.

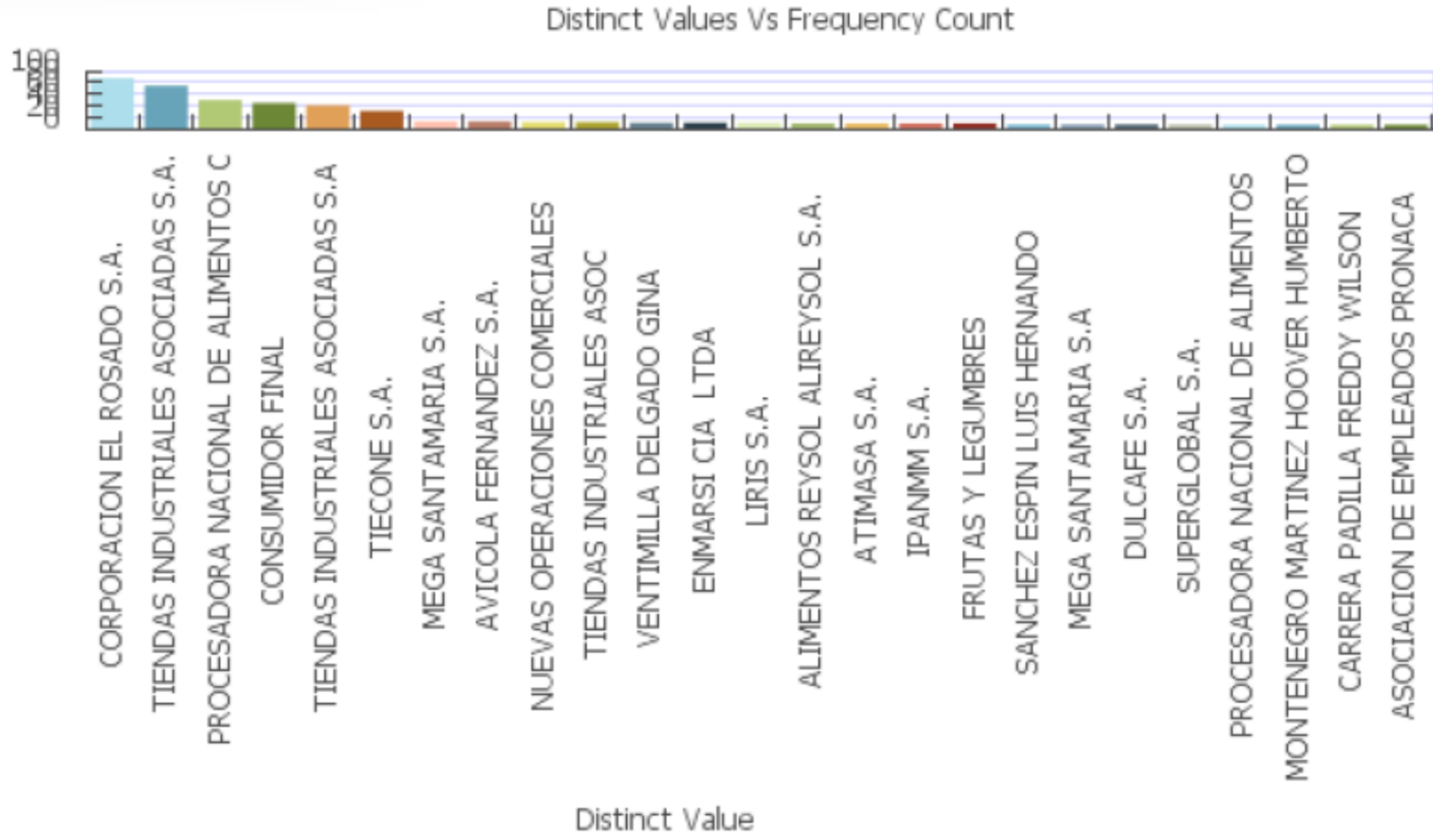
Column Level Details

Format	Count	Total Rows %	Total Rows Cumulative %	Example Values
AAAAAAA A.A.	221	0.66245017	0.66245017	TIECONE S.A. ATIMASA S.A. DULCAFE S.A. DOELDOS S.A. AQUAMAR S.A. NIKAMAR S.A. MISAGRO S.A. MARRIOT S.A. LADYBUS S.A. JERILEX S.A.
AAAAAAA A.A.	169	0.50657954	1.16902971	ENMARDOS S.A. JUDISPRO S.A. BECROMAL S.A. INVERNEG S.A. CASABACA S.A. BORDAINS S.A. QUIFATEX S.A. NAPORTEC S.A. INTERDIN S.A. CONCLINA C.A.
AAAAAAA AAAAA	164	0.49159198	1.66062169	VIVERES ROSITA VIVERES YOLITA SALINAS LIBANO TERCENA CANELA BALSECA MANUEL AVICOLA PIANGO

En el gráfico de barras de puede observar que existen algunos valores repetidos que debemos depurar en las siguientes etapas de la calidad de datos.

El reporte también indica los clientes que fueron reconocidos como repetidos.





Frequency Distribution Data

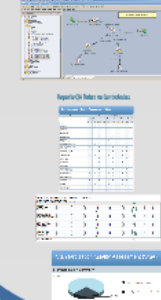
Distinct Value	Frequency Count	Frequency %	Cumulative %
CORPORACION EL ROSADO S.A.	87	0.26078355	0.26078355
TIENDAS INDUSTRIALES ASOCIADAS S.A.	74	0.22181589	0.48259944
PROCESADORA NACIONAL DE ALIMENTOS C	49	0.14687809	0.62947753
CONSUMIDOR FINAL	44	0.13189053	0.76136806
TIENDAS INDUSTRIALES ASOCIADAS S.A	40	0.11990048	0.88126854
TIECONE S.A.	30	0.08992536	0.97119390
MEGA SANTAMARIA S.A.	12	0.03597014	1.00716404
AVICOLA FERNANDEZ S.A.	12	0.03597014	1.04313418
NUEVAS OPERACIONES COMERCIALES	11	0.03297263	1.07610681
TIENDAS INDUSTRIALES ASOC	11	0.03297263	1.10907944
VENTIMILLA DELGADO GINA	10	0.02997512	1.13905456
ENMARSI CIA LTDA	10	0.02997512	1.16902968

Fase de la Calidad de Datos

Investigación

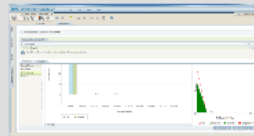


Estandarización

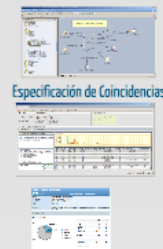


Monitoreo de la Calidad de Datos

Hasta el momento se ha desarrollado todo el proceso de calidad de datos con sus fases, luego de esto se debe realizar el seguimiento para asegurar que los datos cumplen con las reglas establecidas. A continuación se detallan los pasos a seguir para hacer el monitoreo de la información en la herramienta IBM Information Analyzer.



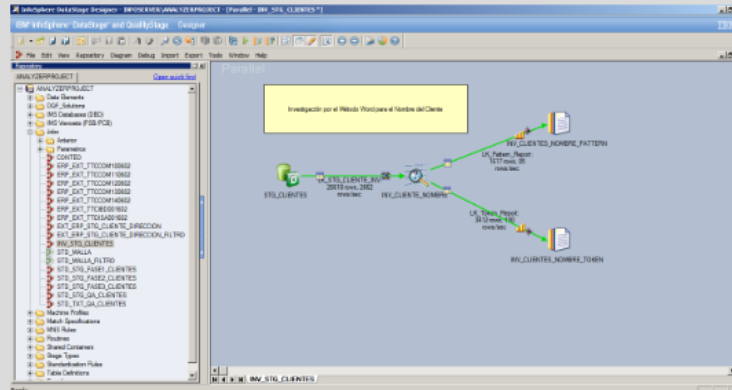
Coincidencia



Supervivencia



Investigación

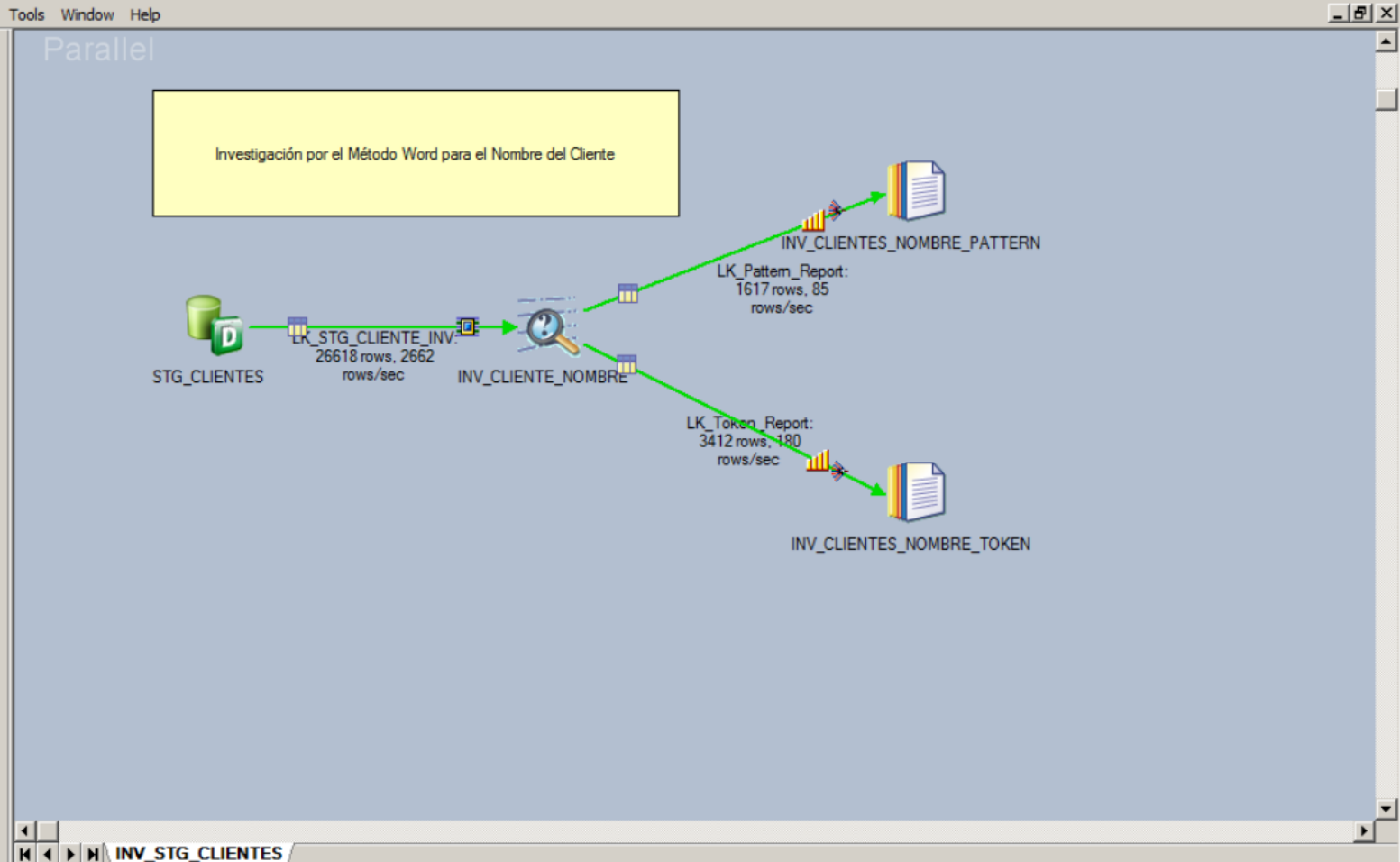


qsTextColumnId	qsTextColumn	qsTextColumnId	qsTextColumn	qsTextColumn	qsTextColumn
1	1111	1	AGUSTO ELIAS JIMENA ELIZABETH	1111	10.8111
1	1112	1	ABAD ABAD CARMEN	2454	0.04020
1	1113	1	ABAD ESTER	2389	0.21207
1	1114	1	ABAD HADACHE SELVA ALEJANDRA	1803	4.88018
1	1115	1	ABACERIA ESITA	1109	4.4609
1	1116	1	ACELERON VELAZQUEZ	662	1.12944
1	1117	1	ABRIL ALBERTO	692	3.1207
1	1118	1	ACOSTA ARTAZA LUCIA ARGENTINA	407	2.28042
1	1119	1	ABEL CASTILLO SIMON	424	1.99291
1	1120	1	ABAROTOS ALBERTO ROSA	404	1.82777
1	1121	1	ACOSTA ELBA MARIA	349	1.31114
1	1122	1	ACOSTA PABLO HELYIA DEL CARMEN	300	1.12756
1	1123	1	ACEROSO	282	1.09743
1	1124	1	ABAROTOS CAMPOVERDE	241	0.98054
1	1125	1	ADA MARLENE CARMEN ORTA	237	0.87075
1	1126	1	ACEROSO EIA. LITA.	209	0.790289
1	1127	1	ACOSTA SIVALLA ROBERTAN ALEJANDRO	190	0.740988
1	1128	1	ACOSTA ZOLA ISABEL MARCELA	142	0.408411
1	1129	1	ALFARO DEL CASTILLO	147	0.582098
1	1130	1	ALFARO TOME AGUIRRE	139	0.522059
1	1131	1	AGROPECUARIA GARZARUA AGRICOLA	138	0.518445
1	1132	1	ABAD AMOROS ANIBAL RICARDO	136	0.510982
1	1133	1	ABAROTOS ROSA	122	0.480206
1	1134	1	ALBA MARLENE CROZINGHARA	120	0.450929
1	1135	1	ABAROTOS EL VECINO	119	0.414925
1	1136	1	AGRICOLA DANIELA DALER	111	0.417011
1	1137	1	AGRICOLA EL HANARRO S.A.	106	0.390217
1	1138	1	ABAD MATE KRISTINA ALEJANDRO	100	0.375484

Repository

ANALYZERPROJECT [Open quick find](#)

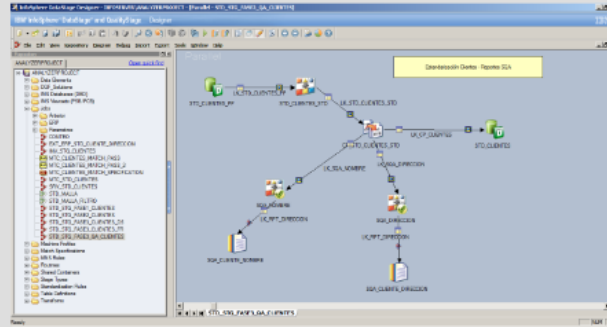
- ANALYZERPROJECT
 - Data Elements
 - DQF_Solutions
 - IMS Databases (DBD)
 - IMS Viewsets (PSB/PCB)
 - Jobs
 - Anterior
 - Parametros
 - CONTEO
 - ERP_EXT_TTCCOM100602
 - ERP_EXT_TTCCOM110602
 - ERP_EXT_TTCCOM120602
 - ERP_EXT_TTCCOM130602
 - ERP_EXT_TTCCOM140602
 - ERP_EXT_TTCIBD001602
 - ERP_EXT_TTDISA001602
 - EXT_ERP_STG_CLIENTE_DIRECCION
 - EXT_ERP_STG_CLIENTE_DIRECCION_FILTRO
 - INV_STG_CLIENTES
 - STD_MALLA
 - STD_MALLA_FILTRO
 - STD_STG_FASE1_CLIENTES
 - STD_STG_FASE2_CLIENTES
 - STD_STG_FASE3_CLIENTES
 - STD_STG_QA_CLIENTES
 - STD_TXT_QA_CLIENTES
 - Machine Profiles
 - Match Specifications
 - MNS Rules
 - Routines
 - Shared Containers
 - Stage Types
 - Standardization Rules
 - Table Definitions



qsInvColumnName	qsInvPattern	qsInvSample	qsInvCount	qsInvPercent
NOMBRE	??FF	?ACATO DIAZ JIMENA ELIZABETH	8119	30.5019
NOMBRE	??F	ABAD ABAD CARMEN	2414	9.06905
NOMBRE	F?	ABDUL SATTAR	1389	5.21827
NOMBRE	??F?	ABAD MASACHE SELVA ALEMANIA	1303	4.89518
NOMBRE	?F	ABACERIA ELSITA	1189	4.4669
NOMBRE	??	ACELINDA VELASQUEZ	833	3.12946
NOMBRE	?II	ABRIDACORP S.A.	832	3.1257
NOMBRE	???F	ACOSTA ARTEAGA LAURA ARGENTINA	607	2.28041
NOMBRE	F??	ABEL CAJILEMA GUAMAN	424	1.59291
NOMBRE	???	ABASTOS ALEXEN MORA	404	1.51777
NOMBRE	?FF	ACURIO ELBA MARIA	349	1.31114
NOMBRE	??FLF	ACARO PEREZ NELVIA DEL CARMEN	300	1.12706
NOMBRE	?	ADESUREG	282	1.05943
NOMBRE	W?	ABARROTES CAMPOVERDE	261	0.98054
NOMBRE	FF??	ADA MARLENE CARRERA ORNA	237	0.890375
NOMBRE	?OS	ACLIMATIC CIA. LTDA.	209	0.785183
NOMBRE	????	ACOSTA ZAVALA SEBASTIAN ALEJAN	198	0.743858
NOMBRE	?FFF	ACOSTA SILA IRENE NARCISA	162	0.608611
NOMBRE	?L?	ALITAS DEL CADILAC	147	0.552258
NOMBRE	FF?	ALFREDO IVAN ALVARADO	139	0.522203
NOMBRE	WW?II	AGROPECUARIA GANADERA ACUICOLA	138	0.518446
NOMBRE	F?FF	ABDO ANDINO ANGEL RICARDO	136	0.510932
NOMBRE	WF	ABARROTES JESSICA	122	0.458336
NOMBRE	?F?	ALBA MARLENE CHUQUIMARCA	120	0.450823
NOMBRE	WL?	ABARROTES EL VECINO	113	0.424525
NOMBRE	W??	AGRICOLA BANANERA TALEB	111	0.417011
NOMBRE	WL?II	AGRICOLA EL NARANJO S.A.	106	0.398227
NOMBRE	???FF	AHMED NU?EZ KRISHNA ALEJANDRO	100	0.375686



Estandarización



Reporte QA Datos no Controlados

Standardization Quality Assessment (SQA)

Comparison Set

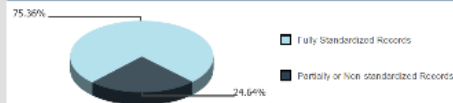
	Standardized with comparison set (100% of the unstandardized records)					
	WFL	WFL	WFL	WFL	WFL	WFL
WFL_CLIENTES	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_3TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_2TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_1TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_4TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_5TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_6TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_7TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_8TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_9TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_10TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_11TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_12TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_13TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_14TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_15TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_16TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_17TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_18TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_19TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_20TO	100%	100%	100%	100%	100%	100%

Standardization Quality Assessment (SQA) - Detailed Examples

Comparison Set	WFL	WFL	WFL	WFL	WFL	WFL
WFL_CLIENTES	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_3TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_2TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_1TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_4TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_5TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_6TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_7TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_8TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_9TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_10TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_11TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_12TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_13TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_14TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_15TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_16TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_17TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_18TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_19TO	100%	100%	100%	100%	100%	100%
WFL_CLIENTES_20TO	100%	100%	100%	100%	100%	100%

Standardization Quality Assessment (SQA)

Standardization Summary



Standardization

InfoSphere DataStage Designer - INFOSERVER\ANALYZERPROJECT - [Parallel - STD_STG_FASE3_QA_CLIENTES]

IBM InfoSphere DataStage and QualityStage Designer

Repository: ANALYZERPROJECT

- ANALYZERPROJECT
 - Data Elements
 - DQF Solutions
 - IMS Databases (DBD)
 - IMS Viewsets (PSB/PCB)
 - Jobs
 - Anterior
 - ERP
 - Parametros
 - CONTEO
 - EXT_ERP_STG_CLIENTE_DIRECCION
 - INV_STG_CLIENTES
 - MTC_CLIENTES_MATCH_PASS
 - MTC_CLIENTES_MATCH_PASS_2
 - MTC_CLIENTES_MATCH_SPECIFICATION
 - MTC_STD_CLIENTES
 - SRV_STD_CLIENTES
 - STD_MALLA
 - STD_MALLA_FILTRO
 - STD_STG_FASE1_CLIENTES
 - STD_STG_FASE2_CLIENTES
 - STD_STG_FASE3_CLIENTES_DS
 - STD_STG_FASE3_CLIENTES_FR
 - STD_STG_FASE3_QA_CLIENTES
 - Machine Profiles
 - Match Specifications
 - MNS Rules
 - Routines
 - Shared Containers
 - Stage Types
 - Standardization Rules
 - Table Definitions
 - Transforms

Parallel

Estandarización Clientes - Reportes SQA

```
graph TD; STD_CLIENTES_PP[STD_CLIENTES_PP] -- LK_STD_CLIENTES_PP --> STD_CLIENTES_STD[STD_CLIENTES_STD]; STD_CLIENTES_STD -- LK_STD_CLIENTES_STD --> JOIN[011 0110]; JOIN -- LK_SQA_NOMBRE --> SQA_NOMBRE[SQA_NOMBRE]; SQA_NOMBRE -- LK_RPT_DIRECCION --> SQA_CLIENTE_NOMBRE[SQA_CLIENTE_NOMBRE]; JOIN -- LK_SQA_DIRECCION --> SQA_DIRECCION[SQA_DIRECCION]; SQA_DIRECCION -- LK_RPT_DIRECCION --> SQA_CLIENTE_DIRECCION[SQA_CLIENTE_DIRECCION]; JOIN -- LK_CP_CLIENTES --> STD_CLIENTES[STD_CLIENTES];
```

STD_CLIENTES_PP

STD_CLIENTES_STD

LK_STD_CLIENTES_PP

LK_STD_CLIENTES_STD

011 0110

LK_CP_CLIENTES

STD_CLIENTES

LK_SQA_NOMBRE

LK_SQA_DIRECCION

SQA_NOMBRE

SQA_DIRECCION

LK_RPT_DIRECCION

SQA_CLIENTE_NOMBRE

SQA_CLIENTE_DIRECCION

STD_STG_FASE3_QA_CLIENTES

Ready



Reporte QA Datos no Controlados

Reporte QA Datos no Controlados

Standardization Quality Assessment (SQA)

Composition Sets

Displayed sets comprise 97.44% of the processed records

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10
	63.17%	4.93%	4.39%	3.90%	3.35%	2.54%	2.27%	1.81%	1.75%	1.68%
TipoNombre		✓	✓	✓	✓	✓	✓	✓	✓	✓
CodigoGenero		✓							✓	
PrefijoNombre										
PrimerNombre		✓							✓	
SegundoNombre										
Apellido		✓	✓	✓	✓	✓	✓	✓	✓	✓
Nombre_Adicional										
SufijoNombre					✓		✓			✓
MatchPrimerNombre		✓							✓	
MatchPrimerNombreNYSIIS		✓							✓	
MatchPrimerNombreRVSNDX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MatchApellido		✓	✓	✓	✓	✓	✓	✓	✓	✓
MatchApellidoHashKey		✓	✓	✓	✓	✓	✓	✓	✓	✓
MatchApellidoPackKey		✓	✓	✓	✓	✓	✓	✓	✓	✓
NumofMatchApellido	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MatchApellido1		✓	✓	✓	✓	✓	✓	✓	✓	✓
MatchApellido2			✓	✓		✓	✓	✓	✓	✓
MatchApellido3			✓			✓		✓		✓
MatchApellido4						✓		✓		
MatchApellido5								✓		
MatchApellido1NYSIIS		✓	✓	✓	✓	✓	✓	✓	✓	✓
MatchApellido1RVSNDX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MatchApellido1SNDX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MatchApellido2NYSIIS			✓	✓		✓	✓	✓	✓	✓
MatchApellido2RVSNDX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MatchApellido2SNDX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
UnhandledPattern	✓									
InputPattern	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ExceptionData										
UserOverrideFlag	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

MatchApellido1SNDX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MatchApellido2NYSIIS			✓	✓		✓	✓	✓	✓	✓
MatchApellido2RVSNDX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MatchApellido2SNDX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
UnhandledPattern	✓									
InputPattern	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ExceptionData										
UserOverrideFlag	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Standardization Quality Assessment (SQA) Record Examples

Displayed sets comprise 97.44% of the processed records

Set 1 of 44 (63.17%, 14182 of 22452 total records) - 20 record examples shown

Input Record	MatchPrimerNombreRVSNDX	NumofMatchApellido	MatchApellido1RVSNDX	MatchApellido1SNDX	MatchApellido2RVSNDX	MatchApellido2SNDX	UnhandledPattern	InputPattern	UserOverrideFlag
GUILLEN NICOLAS	0000	0	0000	0000	0000	0000	+F	+F	NO
LIZARRALDE ALFONSO	0000	0	0000	0000	0000	0000	+F	+F	NO
ORDONIEZ MARIO	0000	0	0000	0000	0000	0000	+F	+F	NO
PROANIO JOSE	0000	0	0000	0000	0000	0000	+F	+F	NO
RODAS JULIO ALFREDO	0000	0	0000	0000	0000	0000	++F	++F	NO
VERGARA MORALES JUANA	0000	0	0000	0000	0000	0000	++F	++F	NO
AGUIRRE ARIAS GENOVEVA	0000	0	0000	0000	0000	0000	++F	++F	NO
ALVARADO APOLO GALO PATRICIO	0000	0	0000	0000	0000	0000	++F+	++F+	NO
CUEVA ZAMBRANO ROQUE	0000	0	0000	0000	0000	0000	+++	+++	NO
DIAZ MORENO JAIME	0000	0	0000	0000	0000	0000	++F	++F	NO
ENRIQUEZ MAZA CARLOS	0000	0	0000	0000	0000	0000	++F	++F	NO
ESPINOSA GUERRON HUGO	0000	0	0000	0000	0000	0000	++F	++F	NO
GODOY SILVA FABIOLA	0000	0	0000	0000	0000	0000	++F	++F	NO
GUEVARA URRUTIA JOSE	0000	0	0000	0000	0000	0000	++F	++F	NO
GUTIERREZ AGURTO RENE	0000	0	0000	0000	0000	0000	++F	++F	NO
LOPEZ ROMERO ARTURO	0000	0	0000	0000	0000	0000	++F	++F	NO
LOPEZ ROMERO GUSTAVO	0000	0	0000	0000	0000	0000	++F	++F	NO
LOPEZ CALDERON VICTOR	0000	0	0000	0000	0000	0000	++F	++F	NO
NARVAEZ GARZON WILLIAM	0000	0	0000	0000	0000	0000	++F	++F	NO
NEGRETE ONTANEDA JORGE	0000	0	0000	0000	0000	0000	++F	++F	NO

IBM InfoSphere Information Server

IBM, the IBM logo, and IBM InfoSphere Information Server are trademarks of International Business Machines Corporation in the United States, other countries or both.

Standardization Quality Assessment (SQA)

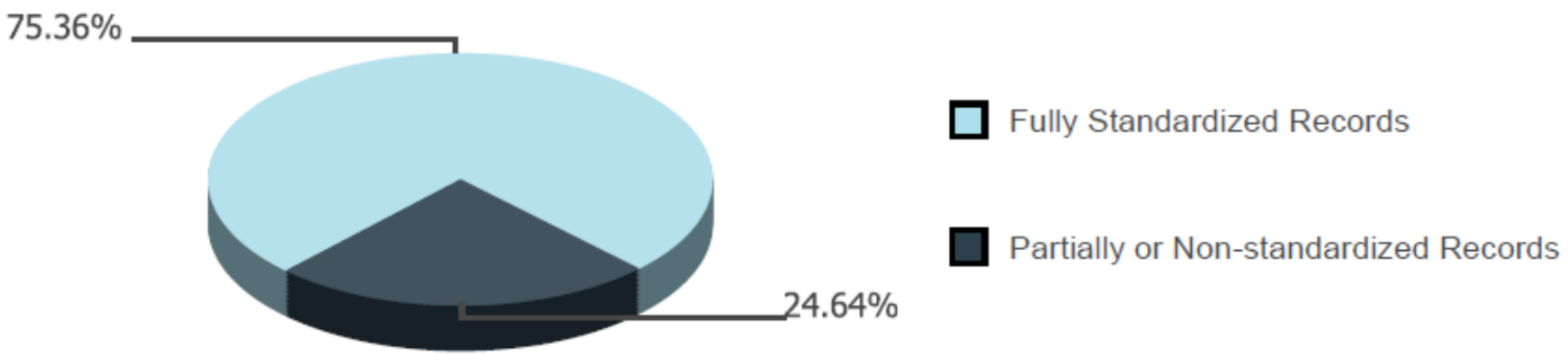


0000	0	0000	0000	0000	0000	0000	+++	+++	NO
0000	0	0000	0000	0000	0000	0000	+++	+++	NO
0000	0	0000	0000	0000	0000	0000	+++	+++	NO

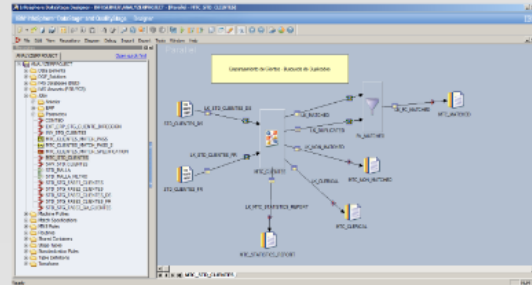
er
e Information Server are trademarks of International Business Machines Corporation in the United States, other countries or both.

Standardization Quality Assessment (SQA)

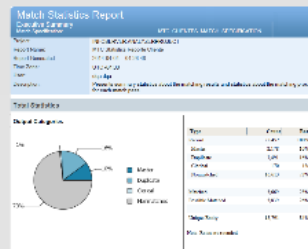
Standardization Summary



Coincidencia



Especificación de Coincidencias



InfoSphere DataStage Designer - INFOSERVER\ANALYZERPROJECT - [Parallel - MTC_STD_CLIENTES]

IBM InfoSphere DataStage and QualityStage Designer

Repository: ANALYZERPROJECT

- ANALYZERPROJECT
 - Data Elements
 - DQF_Solutions
 - IMS Databases (DBD)
 - IMS Viewsets (PSB/PCB)
 - Jobs
 - Anterior
 - ERP
 - Parametros
 - CONTEO
 - EXT_ERP_STG_CLIENTE_DIRECCION
 - INV_STG_CLIENTES
 - MTC_CLIENTES_MATCH_PASS
 - MTC_CLIENTES_MATCH_PASS_2
 - MTC_CLIENTES_MATCH_SPECIFICATION
 - MTC_STD_CLIENTES**
 - SRV_STD_CLIENTES
 - STD_MALLA
 - STD_MALLA_FILTRO
 - STD_STG_FASE1_CLIENTES
 - STD_STG_FASE2_CLIENTES
 - STD_STG_FASE3_CLIENTES_DS
 - STD_STG_FASE3_CLIENTES_FR
 - STD_STG_FASE3_QA_CLIENTES
 - Machine Profiles
 - Match Specifications
 - MNS Rules
 - Routines
 - Shared Containers
 - Stage Types
 - Standardization Rules
 - Table Definitions
 - Transforms

Parallel

Emparejamiento de Clientes - Búsqueda de Duplicados

Ready

Prezi Especificación de Coincidencia

Especificación de Coincidencias

Match Designer - Specification: MTC_CLIENTES_MATCH_SPECIFICATION

Compose | Total Statistics

Save | Add Pass | Remove Pass | Test All Passes | Configure Specification | Show Pass Holding Area

Match Type: One-source Dependent

Match Pass Holding Area

This area holds match passes that are not executed as part of the Match job. To add a pass, press the Ctrl key and drag the pass from the 'Match Flow' area.

Match Pass: MTC_CLIENTES_MATCH_PASS_2

Pass Definition | Pass Statistics

Save Pass | Test Pass | Grouping

Blocking Columns:

- MatchApellido1NYSIIS_PENAME

Match Commands:

- PrimerNombre_PENAME
- SegundoNombre_PENAME
- Apellido_PENAME
- TipoNombre_PENAME
- Calle_PEADDR
- MatchPalabraCalle1NYSIIS_PEADDR

Cutoff Values:

- Match: 8
- Clerical: 7

Test Results:

Frequency

Weight

SetID	Record Type	Weight	DataID	PrimerNombre_PENAMI	SegundoNombre_PEN/	Apellido_PENAME	TipoNombre_PENAME	Calle_PEADDR	MatchPalabraCalle1NY
126	MP	55.87	126	ANA	CECILIA	MONTENEGRO	I	ALVEZ Y JUAN	ALV
126	CP	7.39	127	NELSON		MONTENEGRO	I	ALVEZ YJUAN	ALV
1263	MP	56.03	1263	ANA	LUCIA	HURTADO ARROBA	I	STA LUCIA Y DE	ST
1263	DA	23.18	1351	ANA	LUCIA	HURTADO ARROBA	I	AVIGIRIAS Y	AVAGAR
3465	MP	21.41	3465	ANA		QUISPE LEGUMBRES	I		
3465	DA	21.41	5851	ANA		QUISPE LEGUMBRES	I		
4732	MP	33.77	4732	ANA		FERNANDEZ	I	MERCADO CENTRAL	MARCAD
4732	DA	21.41	227	ANA		FERNANDEZ	I		
8697	MP	38.08	8697	ANA		YAUTIBUG	I	MERC GRAN	MARC
8697	CP	7.11	12643	ANA		YAUTIBUG GUAGCHA	I	MERC LAS ESCLUSAS	MARC
10718	MP	39.60	10718	ANA	MARIA	MULLO NAULA	I	SAUCES	SAC
10718	DA	10.21	888	ANA	MARIA	MII A SINGO	I		

OK | Cancel | Help

Match Statistics Report

Executive Summary

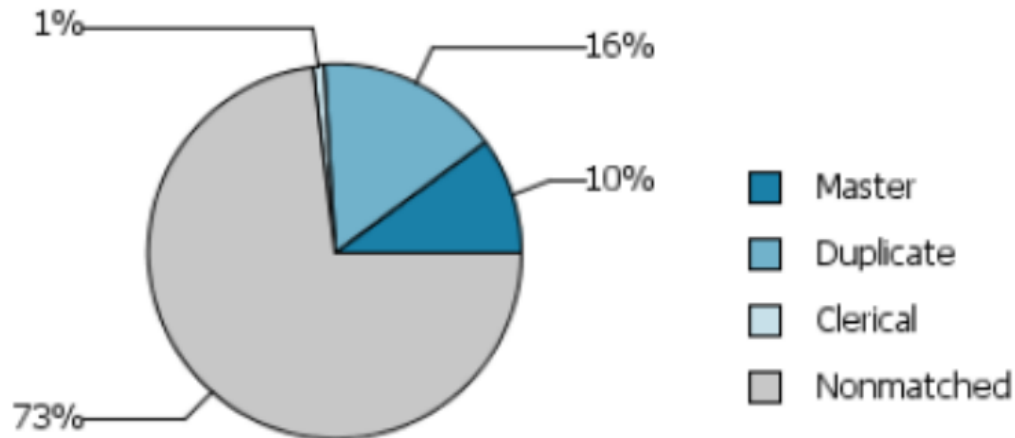
Match Specification:

MTC_CLIENTES_MATCH_SPECIFICATION

Project:	INFOSERVER:ANALYZERPROJECT
Report Name:	MTC Statistics Reporte Cliente
Report Generated:	2014-04-04 04:26:46
Time Zone:	UTC -04:00
User:	dqa dqa
Description:	Presents summary statistics about the matching results and statistics about the matching process for each match pass.

Total Statistics

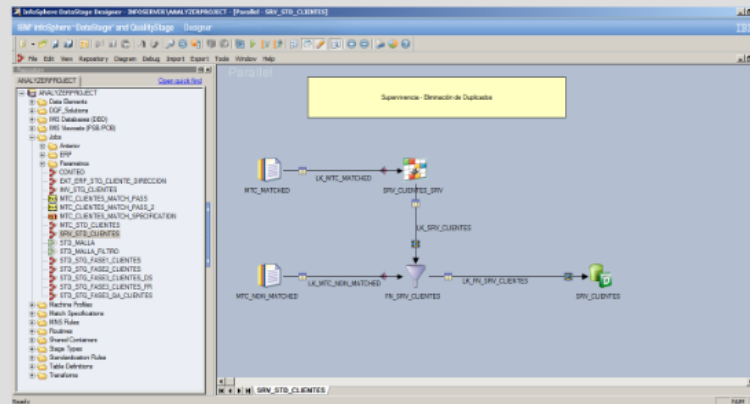
Output Categories



Type	Count	Rate
Record	22,452	100%
Master	2,178	10%
Duplicate	3,491	16%
Clerical	170	1%
Nonmatched	16,613	73%
Matched	5,669	25%
Possible Matched	5,839	26%
Unique Entity	18,791	84%

Note: Rates are rounded.

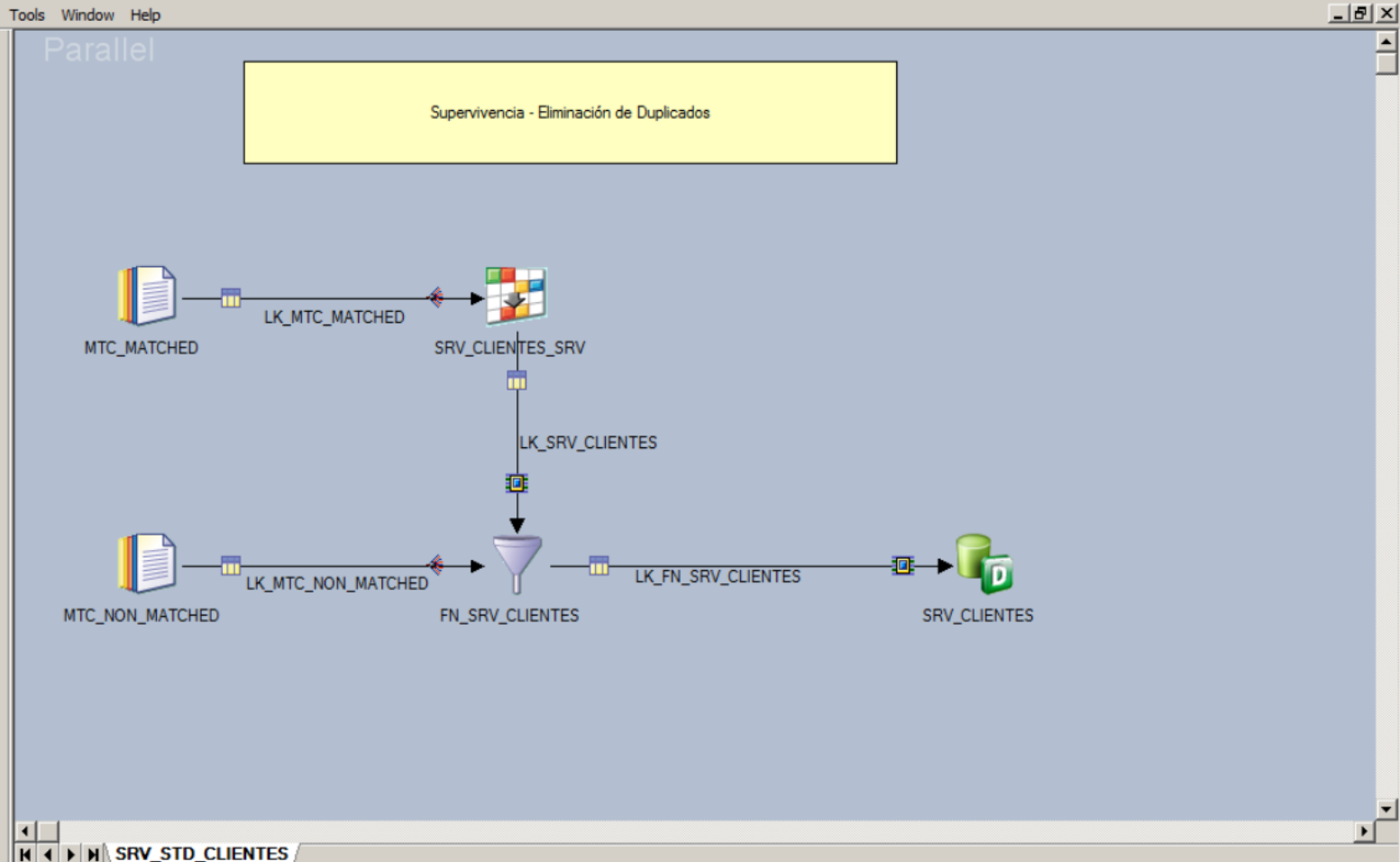
Supervivencia



Repository

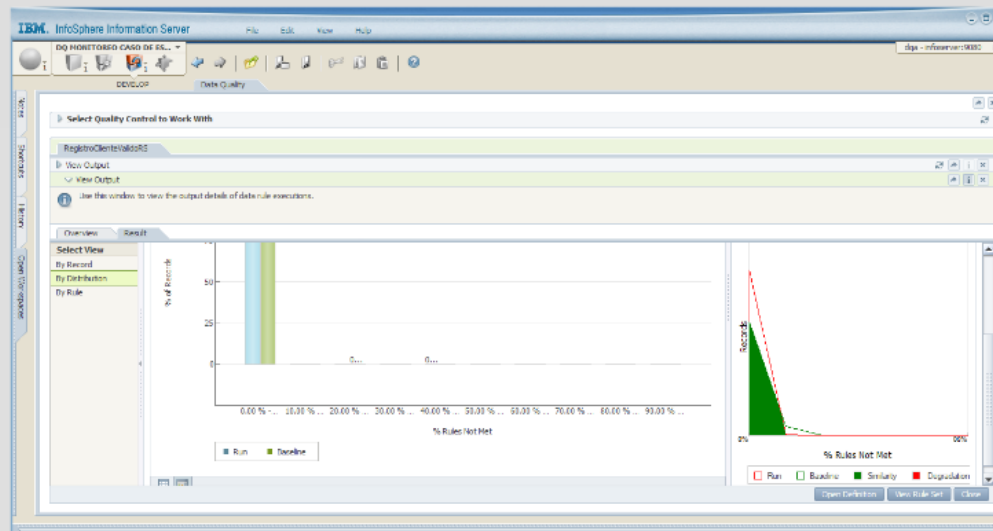
ANALYZERPROJECT [Open quick find](#)

- ANALYZERPROJECT
 - Data Elements
 - DQF_Solutions
 - IMS Databases (DBD)
 - IMS Viewsets (PSB/PCB)
 - Jobs
 - Anterior
 - ERP
 - Parametros
 - CONTEO
 - EXT_ERP_STG_CLIENTE_DIRECCION
 - INV_STG_CLIENTES
 - MTC_CLIENTES_MATCH_PASS
 - MTC_CLIENTES_MATCH_PASS_2
 - MTC_CLIENTES_MATCH_SPECIFICATION
 - MTC_STD_CLIENTES
 - SRV_STD_CLIENTES
 - STD_MALLA
 - STD_MALLA_FILTRO
 - STD_STG_FASE1_CLIENTES
 - STD_STG_FASE2_CLIENTES
 - STD_STG_FASE3_CLIENTES_DS
 - STD_STG_FASE3_CLIENTES_FR
 - STD_STG_FASE3_QA_CLIENTES
 - Machine Profiles
 - Match Specifications
 - MNS Rules
 - Routines
 - Shared Containers
 - Stage Types
 - Standardization Rules
 - Table Definitions
 - Transforms

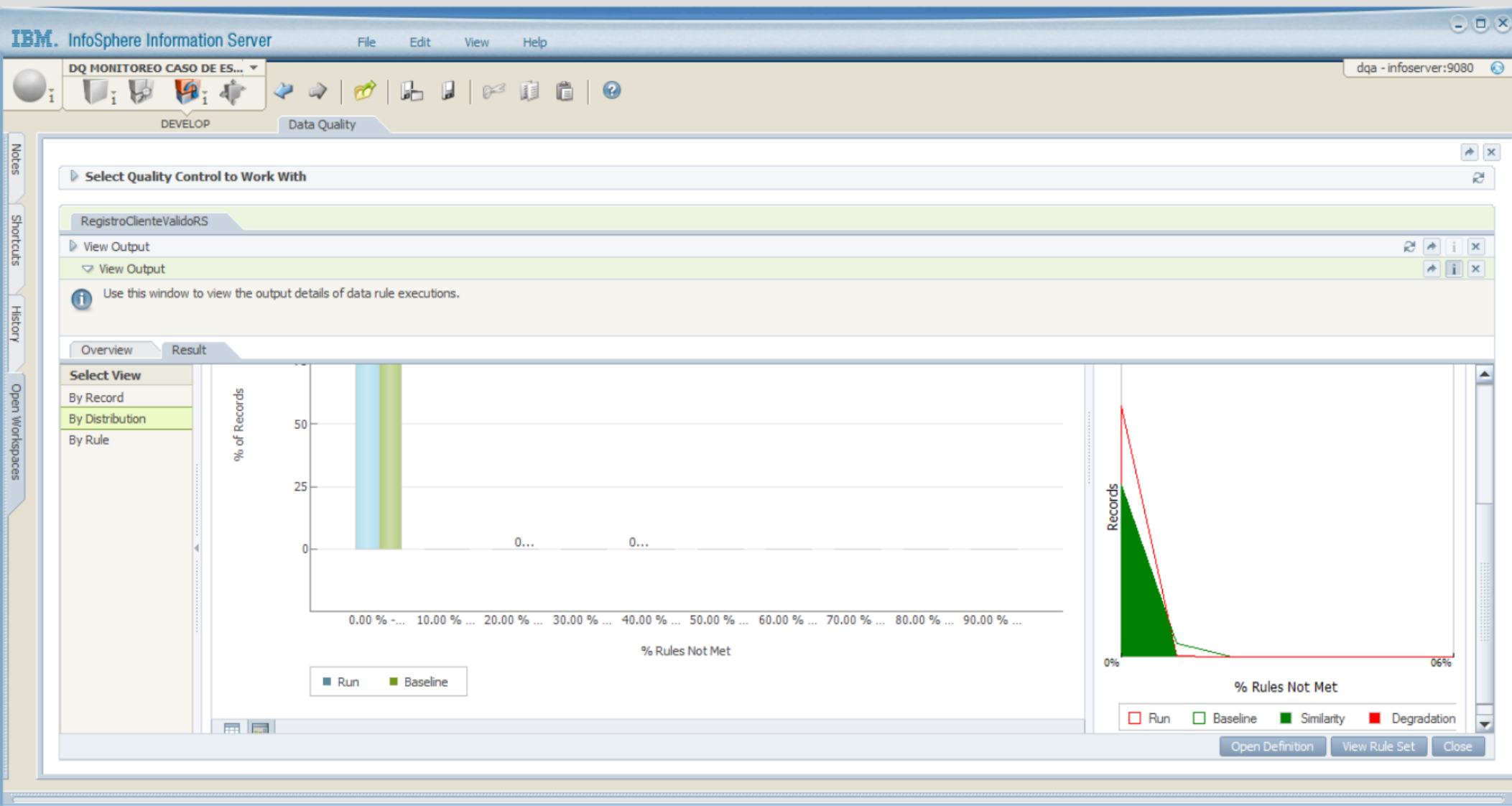


Monitoreo de la Calidad de Datos

Hasta el momento se ha desarrollado todo el proceso de calidad de datos con sus fases, luego de esto se debe realizar el seguimiento para asegurar que los datos cumplen con las reglas establecidas. A continuación se detallan los pasos a seguir para hacer el monitoreo de la información en la herramienta IBM Information Analyzer.



...ación se detallan los pasos a seguir para hacer el monitoreo de información en la herramienta IBM Information Analyzer.



Resultados

Investigación

En la etapa de investigación se realizó el análisis de alto nivel mediante IBM InfoSphere Information Analyzer, encontramos patrones que se repetían siendo AAAAAA A.A. el más alto con 771 repeticiones. También encontramos valores repetidos como Consumidor Final que se repite 44 veces.

También se realizó la investigación mediante IBM InfoSphere Data Quality Designer, mediante el método de investigación Word Investigation obtuvimos como resultado patrones que nos permitieron estandarizar la información por ejemplo estos 3 patrones:

- ???F. Donde F = Primer Número ? = Palabra (Se puede inferir como Apellido)
- ???F. Donde F = Primer Número ? = Palabra (Se puede inferir como Apellido)
- ???F. Donde F = Primer Número ? = Palabra (Se puede inferir como Apellido)

Supervivencia

Para la eliminación de los duplicados identificados en la etapa anterior se establece en función de 2 criterios Valores Repetidos o por longitud de la cadena. Con esto logramos asegurarnos que se cargue valores correctos a la tabla final.

Apellido	Apellido_Cadena	Apellido_Repeticiones	Apellido_Repeticiones
Apellido_Repeticiones_41	Apellido_Repeticiones_41	Apellido_Repeticiones_41	Apellido_Repeticiones_41
Apellido_Repeticiones_42	Apellido_Repeticiones_42	Apellido_Repeticiones_42	Apellido_Repeticiones_42
Apellido_Repeticiones_43	Apellido_Repeticiones_43	Apellido_Repeticiones_43	Apellido_Repeticiones_43
Apellido_Repeticiones_44	Apellido_Repeticiones_44	Apellido_Repeticiones_44	Apellido_Repeticiones_44
Apellido_Repeticiones_45	Apellido_Repeticiones_45	Apellido_Repeticiones_45	Apellido_Repeticiones_45
Apellido_Repeticiones_46	Apellido_Repeticiones_46	Apellido_Repeticiones_46	Apellido_Repeticiones_46
Apellido_Repeticiones_47	Apellido_Repeticiones_47	Apellido_Repeticiones_47	Apellido_Repeticiones_47
Apellido_Repeticiones_48	Apellido_Repeticiones_48	Apellido_Repeticiones_48	Apellido_Repeticiones_48
Apellido_Repeticiones_49	Apellido_Repeticiones_49	Apellido_Repeticiones_49	Apellido_Repeticiones_49
Apellido_Repeticiones_50	Apellido_Repeticiones_50	Apellido_Repeticiones_50	Apellido_Repeticiones_50

Evaluación Modelo de Madurez IDMI

La Tabla muestra los puntajes y nivel de madurez de 11 antes y después de aplicar el modelo de gestión de calidad de datos.

Indicador	Madurez Inicial	Madurez Final
1. Datos	1	2
2. Procesos	1	2
3. Personas	1	2
4. Políticas	1	2
5. Herramientas	1	2
6. Roles	1	2
7. Métricas	1	2
8. Cultura	1	2
9. Gobernanza	1	2
10. Seguridad	1	2
11. Otros	1	2

Como se puede apreciar, la estrategia de madurez se calificó en un punto y además se le agregó información al punto cinco del IDMI, un indicador que no se encontraba en el modelo IDMI original. En los datos categorizados se se midieron en cantidad o dimensión de los parámetros, el aumento de la calificación de los categorías IDMI, así como el hecho de que los datos categorizados fueron parametrizados en el resultado después de la aplicación del Modelo de Gestión de Calidad de Datos.



Estandarización

El proceso de estandarización de datos mediante el uso de reglas de transformación es fundamental para garantizar la calidad y la integridad de los datos en un sistema de datos. Este proceso permite estandarizar los datos en un formato común, lo que facilita la integración y el análisis de los datos.

Se muestra un ejemplo de cómo se estandarizan los datos de una tabla de datos antes y después de aplicar las reglas de transformación.



Coincidencia

Mediante la creación de una especificación de coincidencias se realizaron 2 agrupaciones para identificar registros duplicados, las cuales se procesan en secuencia. Para la primera pasada se tiene el 8% de duplicados, con 1858 registros del total.



Para la segunda pasada se tiene 8% de duplicados también pero con 1633 registros del total.



Investigación

En la etapa de investigación se realizó el análisis de alto nivel mediante IBM Infosphere Information Analyzer, encontramos patrones que se repetían siendo AAAAAAA A.A. el más alto con 221 repeticiones. También encontramos valores repetidos como Consumidor Final que se repite 44 veces.

También se realizó la investigación mediante IBM Infosphere Data Quality Designer, mediante el método de investigación Word Investigation obtuvimos como resultado patrones que nos permitieron estandarizar la información por ejemplo estos 2 patrones:

- ??FF: Donde F = Primer Nombre ? = Palabra (Se puede inferir como Apellido)
- ??F: Donde F = Primer Nombre ? = Palabra (Se puede inferir como Apellido)

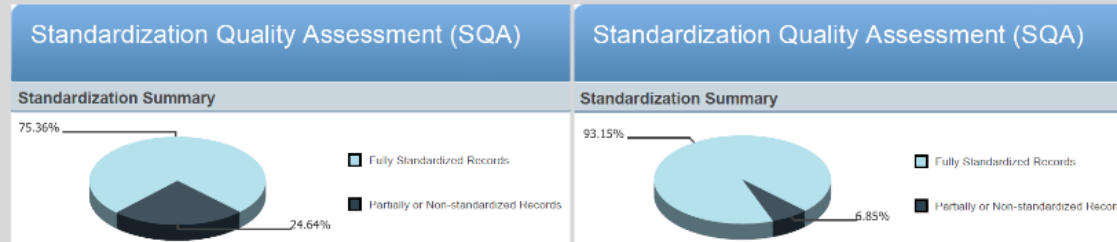
Estandarización

El proceso de estandarización se creó mediante un conjunto de reglas predefinidas en la herramienta para para el nombre 36.83% de datos que cumplían las reglas y para la dirección 35.9%, en base a estos resultados se sobre-escribieron las reglas predefinidas añadiendo patrones que permitan un mayor porcentaje de datos estandarizados. Los siguientes son los patrones añadidos para el nombre. Podemos notar que se incluyen los patrones que notamos en la etapa de investigación.

Standard Pattern	División Codes
***	Calle1
*** **	Calle1 Calle2 Calle3 Calle4 Calle5 Calle6 Calle7 Calle8 Calle9 Calle10 Calle11 Calle12 Calle13 Calle14 Calle15 Calle16 Calle17 Calle18 Calle19 Calle20 Calle21 Calle22 Calle23 Calle24 Calle25 Calle26 Calle27 Calle28 Calle29 Calle30 Calle31 Calle32 Calle33 Calle34 Calle35 Calle36 Calle37 Calle38 Calle39 Calle40 Calle41 Calle42 Calle43 Calle44 Calle45 Calle46 Calle47 Calle48 Calle49 Calle50 Calle51 Calle52 Calle53 Calle54 Calle55 Calle56 Calle57 Calle58 Calle59 Calle60 Calle61 Calle62 Calle63 Calle64 Calle65 Calle66 Calle67 Calle68 Calle69 Calle70 Calle71 Calle72 Calle73 Calle74 Calle75 Calle76 Calle77 Calle78 Calle79 Calle80 Calle81 Calle82 Calle83 Calle84 Calle85 Calle86 Calle87 Calle88 Calle89 Calle90 Calle91 Calle92 Calle93 Calle94 Calle95 Calle96 Calle97 Calle98 Calle99 Calle100
*** **	Calle1 Calle2 Calle3 Calle4 Calle5 Calle6 Calle7 Calle8 Calle9 Calle10 Calle11 Calle12 Calle13 Calle14 Calle15 Calle16 Calle17 Calle18 Calle19 Calle20 Calle21 Calle22 Calle23 Calle24 Calle25 Calle26 Calle27 Calle28 Calle29 Calle30 Calle31 Calle32 Calle33 Calle34 Calle35 Calle36 Calle37 Calle38 Calle39 Calle40 Calle41 Calle42 Calle43 Calle44 Calle45 Calle46 Calle47 Calle48 Calle49 Calle50 Calle51 Calle52 Calle53 Calle54 Calle55 Calle56 Calle57 Calle58 Calle59 Calle60 Calle61 Calle62 Calle63 Calle64 Calle65 Calle66 Calle67 Calle68 Calle69 Calle70 Calle71 Calle72 Calle73 Calle74 Calle75 Calle76 Calle77 Calle78 Calle79 Calle80 Calle81 Calle82 Calle83 Calle84 Calle85 Calle86 Calle87 Calle88 Calle89 Calle90 Calle91 Calle92 Calle93 Calle94 Calle95 Calle96 Calle97 Calle98 Calle99 Calle100
*** **	Calle1 Calle2 Calle3 Calle4 Calle5 Calle6 Calle7 Calle8 Calle9 Calle10 Calle11 Calle12 Calle13 Calle14 Calle15 Calle16 Calle17 Calle18 Calle19 Calle20 Calle21 Calle22 Calle23 Calle24 Calle25 Calle26 Calle27 Calle28 Calle29 Calle30 Calle31 Calle32 Calle33 Calle34 Calle35 Calle36 Calle37 Calle38 Calle39 Calle40 Calle41 Calle42 Calle43 Calle44 Calle45 Calle46 Calle47 Calle48 Calle49 Calle50 Calle51 Calle52 Calle53 Calle54 Calle55 Calle56 Calle57 Calle58 Calle59 Calle60 Calle61 Calle62 Calle63 Calle64 Calle65 Calle66 Calle67 Calle68 Calle69 Calle70 Calle71 Calle72 Calle73 Calle74 Calle75 Calle76 Calle77 Calle78 Calle79 Calle80 Calle81 Calle82 Calle83 Calle84 Calle85 Calle86 Calle87 Calle88 Calle89 Calle90 Calle91 Calle92 Calle93 Calle94 Calle95 Calle96 Calle97 Calle98 Calle99 Calle100
*** **	Calle1 Calle2 Calle3 Calle4 Calle5 Calle6 Calle7 Calle8 Calle9 Calle10 Calle11 Calle12 Calle13 Calle14 Calle15 Calle16 Calle17 Calle18 Calle19 Calle20 Calle21 Calle22 Calle23 Calle24 Calle25 Calle26 Calle27 Calle28 Calle29 Calle30 Calle31 Calle32 Calle33 Calle34 Calle35 Calle36 Calle37 Calle38 Calle39 Calle40 Calle41 Calle42 Calle43 Calle44 Calle45 Calle46 Calle47 Calle48 Calle49 Calle50 Calle51 Calle52 Calle53 Calle54 Calle55 Calle56 Calle57 Calle58 Calle59 Calle60 Calle61 Calle62 Calle63 Calle64 Calle65 Calle66 Calle67 Calle68 Calle69 Calle70 Calle71 Calle72 Calle73 Calle74 Calle75 Calle76 Calle77 Calle78 Calle79 Calle80 Calle81 Calle82 Calle83 Calle84 Calle85 Calle86 Calle87 Calle88 Calle89 Calle90 Calle91 Calle92 Calle93 Calle94 Calle95 Calle96 Calle97 Calle98 Calle99 Calle100

Input Pattern	División Codes
***FF	Apellido1 Apellido1 PrimerNombre2 SegundoNombre5
***F*	Apellido1 Apellido1 PrimerNombre2 SegundoNombre5
***F	Apellido1 Apellido1 PrimerNombre5
F	Apellido1 PrimerNombre6
***	Apellido1 PrimerNombre6
***F	Apellido1 Apellido1 PrimerNombre1 SegundoNombre6
***	Apellido1 Apellido1 PrimerNombre5
*FF	Apellido1 PrimerNombre2 SegundoNombre6
***LF	Apellido1 Apellido1 PrimerNombre2 Apellido2 SegundoNombre6
*	Apellido6
***	Apellido1 Apellido1 PrimerNombre1 SegundoNombre1
F	Apellido1 Apellido1 PrimerNombre2 SegundoNombre6
***FF	Apellido1 Apellido1 Apellido1 Apellido2 Apellido5

Se realizó nuevamente el análisis y se obtuvo para el nombre el 93.15% de datos estandarizados, para la dirección se mejoró a 75.36%.



Unhandled Pattern:	Override Codes:
^	Calle1
++<^++	CalleInterseccion11 CalleInterseccion11 Casa1 Casa1 CalleInterseccion21 CalleInterseccion21
+<^++	CalleInterseccion11 Casa1 Casa1 CalleInterseccion21 CalleInterseccion21
+^	UrbanizacionZona1 UrbanizacionZona1
+++	Sector1 Sector1 Sector1
+++++	Sector1 Sector1 Sector1 Sector1 Sector1
++++	Sector1 Sector1 Sector1 Sector1
++++++	Sector1 Sector1 Sector1 Sector1 Sector1 Sector1
++	Sector1 Sector1
++<^+++	CalleInterseccion11 CalleInterseccion11 Casa1 Casa1 CalleInterseccion21 CalleInterseccion21
+	UrbanizacionZona1
++^	Calle1 Calle1 Casa1
+<^+++	Calle1 Casa1 Casa1 Calle1 Calle1 Calle1
<	Casa1
+++^	Calle1 Calle1 Calle1 Casa1
+++<^++	Calle1 Calle1 Calle1 Casa1 Casa1 Calle1 Calle1
T++++	TipoCalle1 Calle1 Calle1 Calle1 Calle1
T+++	TipoCalle1 Calle1 Calle1 Calle1
T+++++	TipoCalle1 Calle1 Calle1 Calle1 Calle1 Calle1
+^++	Calle1 Casa1 Calle1 Calle1
+T+	Calle1 TipoCalle1 Calle1
++<^	Calle1 Calle1 Casa1 Casa1
+^+	Calle1 Casa1 Calle1
++S	Calle1 Calle1 TipoCalle1
+++<^+	Calle1 Calle1 Casa1 Casa1 Calle1

patrones que notamos en la etapa de investigación.

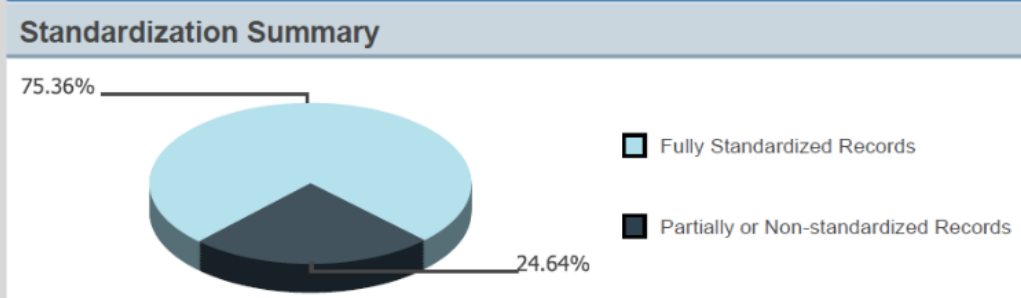
Input Patten:	Override Codes:	Other (▲)
++FF	Apellido 1 Apellido 1 PrimerNombre2 SegundoNombre5	
++F+	Apellido 1 Apellido 1 PrimerNombre2 SegundoNombre5	
++F	Apellido 1 Apellido 1 PrimerNombre6	
+F	Apellido 1 PrimerNombre6	
++	Apellido 1 PrimerNombre5	
+++F	Apellido 1 Apellido 1 PrimerNombre 1 SegundoNombre6	
+++	Apellido 1 Apellido 1 PrimerNombre5	
+FF	Apellido 1 PrimerNombre2 SegundoNombre6	
++FLF	Apellido 1 Apellido 1 PrimerNombre2 Apellido2 SegundoNombre5	
+	Apellido5	
++++	Apellido 1 Apellido 1 PrimerNombre 1 SegundoNombre 1	
F+FF	Apellido 1 Apellido 1 PrimerNombre2 SegundoNombre6	
+++FF	Apellido 1 Apellido 1 Apellido 1 Apellido2 Apellido6	

++++	Sector1 Sector1 Sector1 Sector1
+++++	Sector1 Sector1 Sector1 Sector1 Sector1
++	Sector1 Sector1
++^+++	CalleInterseccion11 CalleInterseccion11 Casa1 Casa1 CalleInterseccion21 CalleInterseccion21
+	UrbanizacionZona1
++^	Calle1 Calle1 Casa1
++^+++	Calle1 Casa1 Casa1 Calle1 Calle1 Calle1
c	Casa1
+++^	Calle1 Calle1 Calle1 Casa1
+++^+++	Calle1 Calle1 Calle1 Casa1 Casa1 Calle1 Calle1
T++++	TipoCalle1 Calle1 Calle1 Calle1 Calle1
T+++	TipoCalle1 Calle1 Calle1 Calle1
T+++++	TipoCalle1 Calle1 Calle1 Calle1 Calle1 Calle1
++^++	Calle1 Casa1 Calle1 Calle1
+T+	Calle1 TipoCalle1 Calle1
++^	Calle1 Calle1 Casa1 Casa1
++^	Calle1 Casa1 Calle1
++S	Calle1 Calle1 TipoCalle1
+++^+	Calle1 Calle1 Casa1 Casa1 Calle1

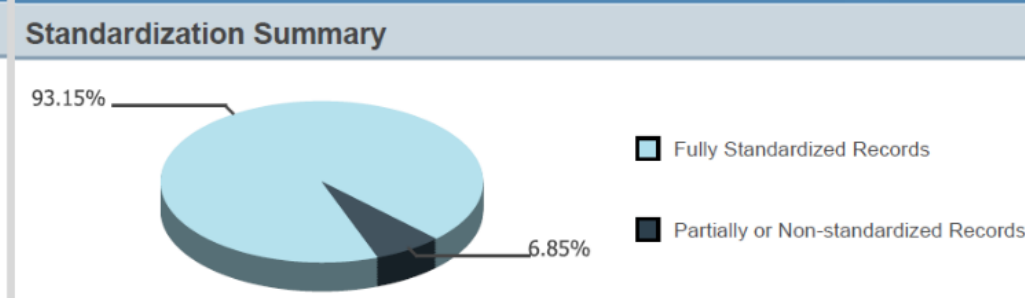
+++	Apellido1 Apellido1 PrimerNombre6
+F	Apellido1 PrimerNombre6
++	Apellido1 PrimerNombre5
+++F	Apellido1 Apellido1 PrimerNombre1 SegundoNombre6
+++	Apellido1 Apellido1 PrimerNombre5
+FF	Apellido1 PrimerNombre2 SegundoNombre6
++FLF	Apellido1 Apellido1 PrimerNombre2 Apellido2 SegundoNombre5
+	Apellido5
++++	Apellido1 Apellido1 PrimerNombre1 SegundoNombre1
F+FF	Apellido1 Apellido1 PrimerNombre2 SegundoNombre6
+++FF	Apellido1 Apellido1 Apellido1 Apellido2 Apellido6

Se realizó nuevamente el análisis y se obtuvo para el nombre el 93.15% de datos estandarizados, para la dirección se mejoró a 75.36%.

Standardization Quality Assessment (SQA)

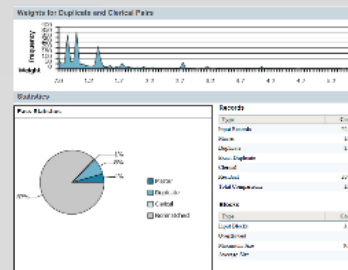


Standardization Quality Assessment (SQA)

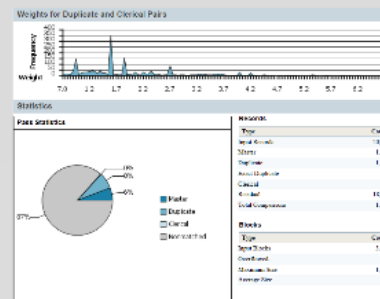


Coincidencia

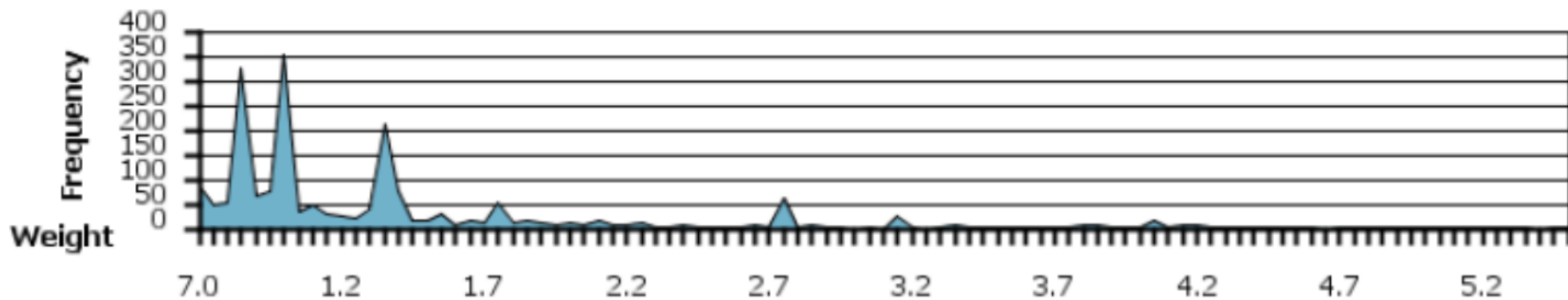
Mediante la creación de una especificación de coincidencias se realizaron 2 agrupaciones para identificar registros duplicados., las cuales se procesan en secuencia. Para la primera pasada se tiene el 8% de duplicados, con 1858 registros del total



Para la segunda pasada se tiene 8% de duplicados también pero con 1633 registros del total.

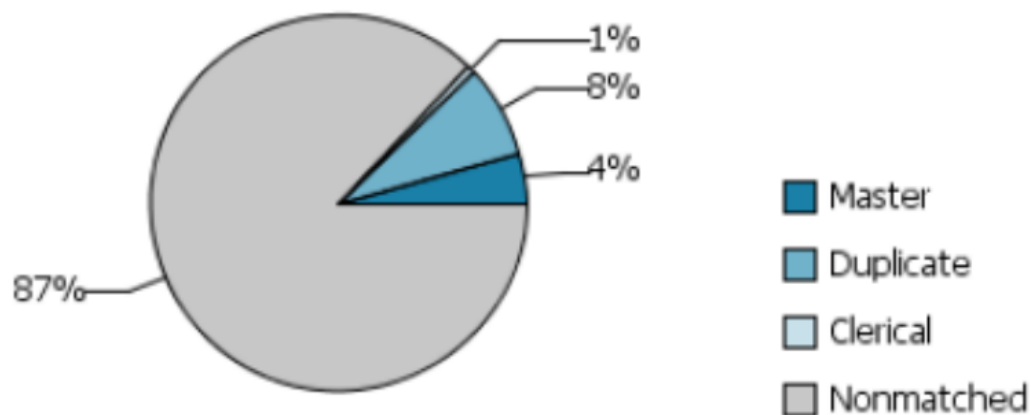


Weights for Duplicate and Clerical Pairs



Statistics

Pass Statistics

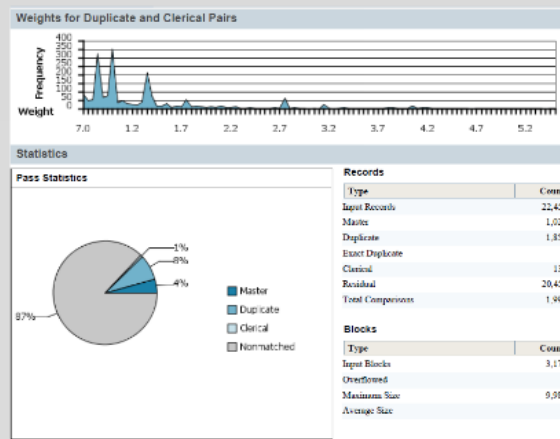


Records

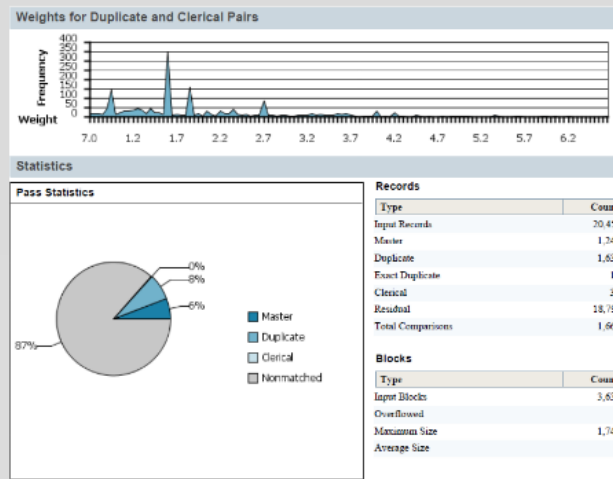
Type	Count
Input Records	22,452
Master	1,021
Duplicate	1,858
Exact Duplicate	9
Clerical	135
Residual	20,459
Total Comparisons	1,993

Blocks

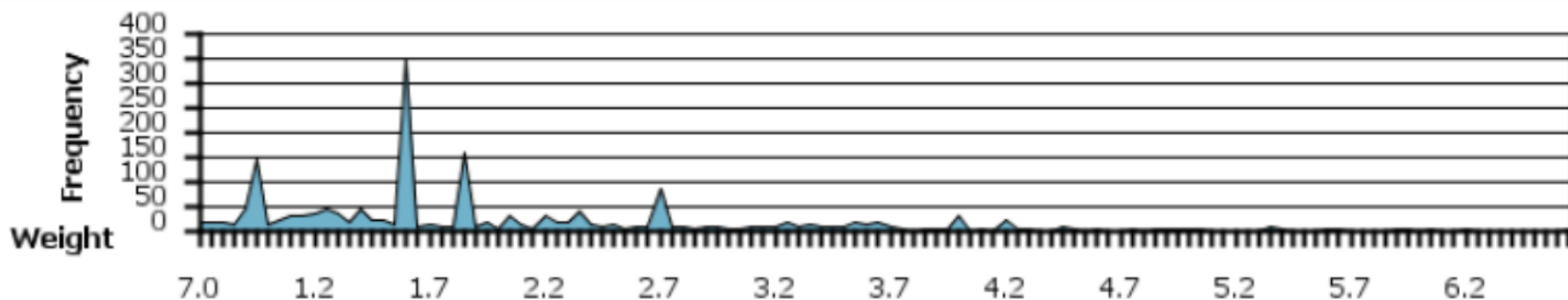
Type	Count
Input Blocks	3,174
Overflowed	0
Maximum Size	9,987
Average Size	3



Para la segunda pasada se tiene 8% de duplicados también pero con 1633 registros del total.

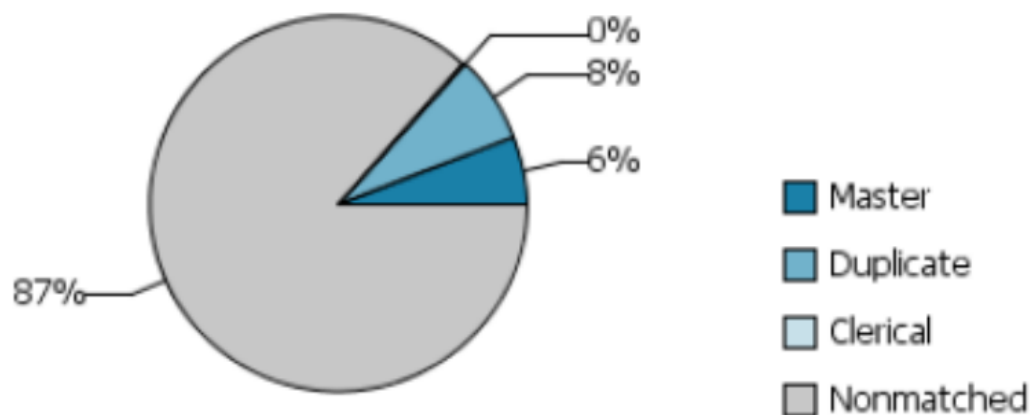


Weights for Duplicate and Clerical Pairs



Statistics

Pass Statistics



Records

Type	Count
Input Records	20,459
Master	1,242
Duplicate	1,633
Exact Duplicate	14
Clerical	35
Residual	18,791
Total Comparisons	1,668

Blocks

Type	Count
Input Blocks	3,638
Overflowed	0
Maximum Size	1,748
Average Size	5

Supervivencia

Para la eliminación de los duplicados identificados en la etapa anterior se establece en función de 2 criterios Valores Repetidos o por longitud de la cadena. Con esto logramos asegurarnos que se cargue valores correctos a la tabla final.

Target(s):	Analyze Column:	Technique:	Data:
TipoDireccion_PE/	TipoDireccion_PEADD	Most Frequent (Non-blan	
TipoCalle_PEADDf	TipoCalle_PEADDR	Most Frequent (Non-blan	
Calle_PEADDR	Calle_PEADDR	Longest	
CalleInterseccion1_	CalleInterseccion1_PE/	Longest	
CalleInterseccion2_	CalleInterseccion2_PE/	Longest	
Numero_PEADDR	Numero_PEADDR	Longest	
Sector_PEADDR	Sector_PEADDR	Longest	
UrbanizacionZona_	UrbanizacionZona_PE/	Longest	

tenios valores repetidos o por longitud de cadena. Con esto logramos asegurarnos que entregue valores correctos a la tabla final.

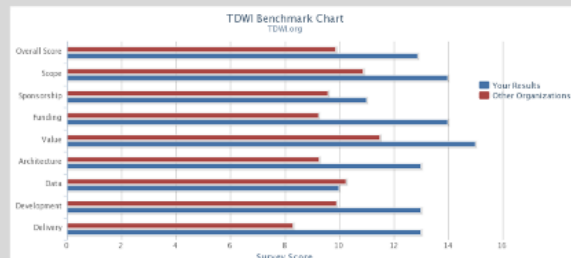
Target(s):	Analyze Column:	Technique:	Data:
TipoDireccion_PE/	TipoDireccion_PEADDR	Most Frequent (Non-blank)	
TipoCalle_PEADDR	TipoCalle_PEADDR	Most Frequent (Non-blank)	
Calle_PEADDR	Calle_PEADDR	Longest	
CalleInterseccion1_	CalleInterseccion1_PE/	Longest	
CalleInterseccion2_	CalleInterseccion2_PE/	Longest	
Numero_PEADDR	Numero_PEADDR	Longest	
Sector_PEADDR	Sector_PEADDR	Longest	
UrbanizacionZona_	UrbanizacionZona_PE/	Longest	

Evaluación Modelo de Madurez TDWI

La Tabla muestra los puntajes y nivel de madurez de BI antes y después de aplicar el modelo de gestión de calidad de datos

Categoría	Evaluación Inicial		Evaluación Final	
	Puntaje	Nivel de Madurez Punto Crítico	Puntaje	Nivel de Madurez Punto Crítico
Ámbito	13	Repetible	13	Repetible
Patrocinio	11	Preliminar	11	Preliminar
Presupuesto	14	Repetible	14	Repetible
Valor	15	Repetible / Abismo	15	Repetible / Abismo
Arquitectura	13	Repetible	13	Repetible
Datos	9	Preliminar/Golfo	10	Preliminar
Desarrollo	13	Repetible	13	Repetible
Entrega	13	Repetible	13	Repetible

Como se puede apreciar, la categoría Datos mejora su calificación en un punto y además eso le permite sobrepasar el punto crítico del Golfo, sin embargo permanece en el nivel de madurez Preliminar. En las demás categorías no se evidencia un aumento o disminución de los puntajes. El aumento de la calificación de la categoría Datos, así como el hecho de que las otras categorías hayan permanecido estables es el resultado esperado de la aplicación del Modelo de Gestión de Calidad de Datos



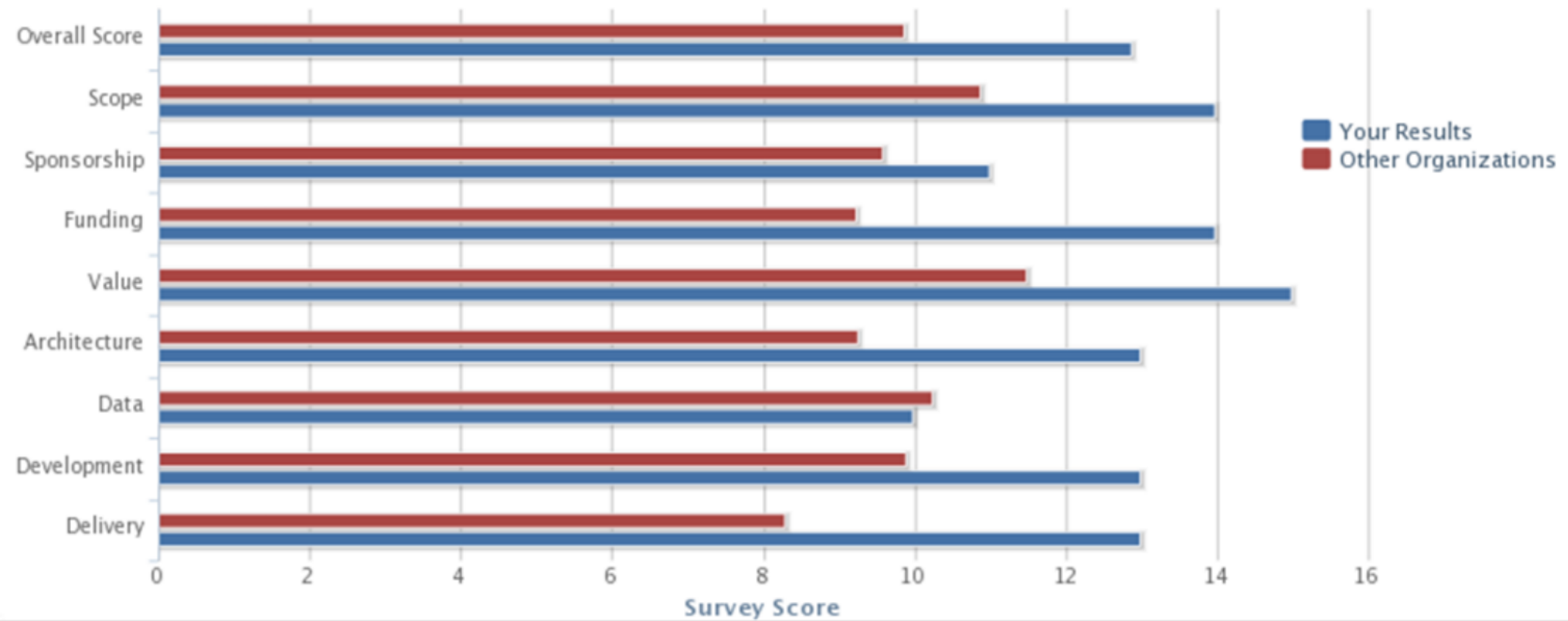
Es de aplicar el modelo de gestión de calidad de datos

Categoría	Evaluación Inicial		Evaluación Final	
	Puntaje	Nivel de Madurez Punto Crítico	Puntaje	Nivel de Madurez Punto Crítico
Ámbito	13	Repetible	13	Repetible
Patrocinio	11	Preliminar	11	Preliminar
Presupuesto	14	Repetible	14	Repetible
Valor	15	Repetible / Abismo	15	Repetible / Abismo
Arquitectura	13	Repetible	13	Repetible
Datos	9	Preliminar/Golfo	10	Preliminar
Desarrollo	13	Repetible	13	Repetible
Entrega	13	Repetible	13	Repetible

Como se puede apreciar, la categoría Datos mejora su calificación en un punto y además eso le permite sobrepasar el punto crítico del Golfo, sin embargo permanece en el nivel de madurez Preliminar. En las demás categorías no se evidencia un aumento o disminución de los puntajes. El aumento de la calificación de la categoría Datos, así como el hecho de que las otras categorías hayan permanecido estables es el resultado esperado de la aplicación del Modelo de Gestión de Calidad de Datos

le la aplicación del Modelo de Gestión de Calidad

TDWI Benchmark Chart
TDWI.org



Ejemplos

Existen varios casos en los que se presentan duplicados en la información, a continuación se describe un ejemplo de un cliente el que se identificó duplicados. Antes del proceso se tenía que el cliente CI/RUC 0990017514001 con duplicados en cuanto a dirección tal como indica la Figura marcado en verde y rojo

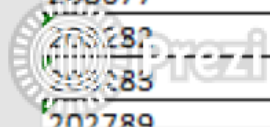
CODIGO_C
203690
203691
144974
203694
105742
205318
286591
289878
131485
205313
205314
205315
205521
259023
259024
259025
285005
288654
288672
401151
104745
104746
104935
176183
185605
205258
205526
205529
206105
273408
286580
288074
288083
288098
288099
289041
289075
300766
300767
300776
300777
300778
300782
300829
300884
300887
300978
301137
301183
401102
401113
401824
401259
401277
289865
300841
301101
301189
301197
288675

n se
enía
ndica

CODIGO_CLIENTE	NOMBRE	CEDULA_RUC	DIRECCION_1
203690	TIENDAS INDUSTRIALES ASOCIADAS TIA	0990017514001	10 DE AGOSTO CALLE PASTAZA
203691	TIENDAS INDUSTRIALES ASOCIADAS TIA	0990017514001	10 DE AGOSTO CALLE PASTAZA
144974	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	10 DE AGOSTO Y BOTA
203694	TIENDAS INDUSTRIALES ASOCIADAS TIA	0990017514001	10 DE AGOSTO Y CIRCUNVALACION
105742	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	CALLE TOCAZO Y MULALILLO ELOY
205318	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	CALLE VELEZ Y LORENZO DE
286591	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CC GARZOCENTRO 2000
289878	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CDLA ACACIAS CALLE G. MORENO
131485	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	CDLA IBARRA AV MARTHA BUCARAM
205313	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE
205314	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE
205315	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE
205521	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE
259023	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	CHIMBORAZO 217 Y LUQUE
259024	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE
259025	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE
285005	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	CHIMBORAZO 217 Y LUQUE
288654	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE
288672	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE
401151	TIENDAS INDUSTRIALES ASOCIADAS S A	0990017514001	CHIMBORAZO 217 Y LUQUE
104745	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
104746	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
104935	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
176183	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
185605	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
205258	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
205526	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
205529	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
206105	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
273408	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
286580	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
288074	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
288083	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
288098	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
288099	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
289041	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
289075	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
300766	TIENDAS INDUSTRIALES ASOC	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
300767	TIENDAS INDUSTRIALES ASOC	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
300776	TIENDAS INDUSTRIALES ASOC	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
300777	TIENDAS INDUSTRIALES ASOC	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
300778	TIENDAS INDUSTRIALES ASOC	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
300782	TIENDAS INDUSTRIALES ASOC	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
300829	TIENDAS INDUSTRIALES ASOC	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
300884	TIENDAS INDUSTRIALES ASOC	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
300887	TIENDAS INDUSTRIALES ASOC	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
300978	TIENDAS INDUSTRIALES ASOC	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
301137	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
301183	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
401102	TIENDAS INDUSTRIALES ASOCIADOS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
401113	TIENDAS INDUSTRIALES ASOCIADOS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
401824	TIENDAS INDUSTRIALES ASOCIADOS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
401259	TIENDAS INDUSTRIALES ASOCIADOS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
401277	TIENDAS INDUSTRIALES ASOCIADOS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQ
289865	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQUINA
300841	TIENDAS INDUSTRIALES ASOC	0990017514001	CHIMBORAZO 217 Y LUQUE ESQUINA
301101	TIENDAS INDUSTRIALES ASOCIADAS	0990017514001	CHIMBORAZO 217 Y LUQUE ESQUINA
301189	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUE ESQUINA
301197	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	CHIMBORAZO 217 Y LUQUE ESQUINA
288675	TIENDAS INDUSTRIALES ASOCIADAS S.A	0990017514001	CHIMBORAZO 217 Y LUQUES ESQ



131295	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	CUSUBAMBA Y APUELA 2343
131381	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	ECUATORIANA E IGNACIO NOBOA
185603	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	ELIZARIO QUEVEDO 111 FELIX VAL
131323	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	EQUINOCCIAL Y 13 DE JUNIO
105829	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	GARCIA MORENO Y BUNINE
273603	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	GERRERO VALENZ Y LA A 1 G VALE
185607	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	GRAN COLOMBIA 127 J MARTIN
104925	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	GUAYAQUIL 958 Y ESPEJO
203080	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	JOSE RODRIGUEZ BONIN URB.RENAC
131347	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	JUAN MONTALVO Y FRAY GONZALO
203707	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	KM 11 1/2 VIA DAULE LOT.
203411	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	KM 11 1/2 VIA DAULE LOTIZ.
203709	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	KM 11 1/2 VIA DAULE LOTZ.
203710	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	KM 11 1/2 VIA DAULE LOTZ.
203345	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	KM.11 1/2 VIA DAULE LOTIZ.
203559	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	LA 17 Y PORTETE
203414	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	LA MANA
185801	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	LIZARDO RUIZ Y ALFAREROS 1 LIZ
505270	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	LOS ESTEROS AVENIDA 103
131374	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	MACHACHI AMAZONAS 3604 11NBRE
203695	TIENDAS INDUSTRIALES ASOCIADAS TIA	0990017514001	MACHALA 169 E/ROCAFUERTE 10
202790	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	MUCHO LOTE 1ERA.ETP. AV.ISIDRO
203332	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	MUCHO LOTE 5TA ETA.CALLION 23
131453	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	PANAMERICANA NORTE ENTRADA A
105393	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	PEDRO V. MALDONADO Y LAS LAJAS
131311	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	PEDRO VICENTE MALDONADO Y GRAL
203231	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	PORTETE Y LA 9NA.
203348	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	REC.2 SALEM ENT.PRINC.1 ETP.
203347	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	RECREO 1 AV.PRINC.9 OCT.PEAT
131350	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	ROCAFUERTE E/BOLIVAR-LIBERTAD
203575	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	RODOLFO BAQUERIZO NAZUR Y CJON
203628	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	SAN FCO. DE MILAGRO Y ERNESTO
288676	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	SAUCES 5 CALLE RODRIGO ICAZA Y
288677	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	SAUCES 6 CALLE GABRIEL ROLDOS
203787	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	TRINITARIA 1 COOP.POLO SUR
203785	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	TRINITARIA 2 COOP. JACOBO
203789	TIENDAS INDUSTRIALES ASOCIADAS S.A.	0990017514001	VIA 12 1/2 DAULE ETE C COMERC



Una vez ejecutado el proceso se puede identificar que los clientes que estaban duplicados se simplificaron a uno solo como indica la Figura en los colores respectivos

CODIGO_CLIENTE	CEDULA_RUC	APELLIDO_PENOME	TIPOCALLE	CALLE_PEADDR	NUMERO_I	KILOMETRO_I	SECTOR_PEADDR	CASA_P	URBANIZACIONZONA
104925	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	GUAYAQUIL	958	NULL	NULL	NULL	Y ESPEJO
105742	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	CALLE	TOACAZO Y MULALILLO ELOY	NULL	NULL	NULL	NULL	NULL
105829	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	NULL	NULL	NULL	NULL	NULL	NULL
131295	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	CUSUBAMBA Y APUELA	2343	NULL	NULL	NULL	NULL
131401	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	AV	GUAYASAMIN UNA	NULL	NULL	NULL	NULL	NULL
170116	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	AV	MARISCAL SUCRE ENTRE OSORIO	NULL	NULL	NULL	NULL	NULL
185603	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	ELIZARIO QUEVEDO	111	NULL	NULL	NULL	FELIX VAL
185801	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	LIZARDO RUIZ Y ALFAREROS	1	NULL	NULL	NULL	LIZ
203080	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	JOSE FERNANDEZ DAVILA	NULL	NULL	NULL	NULL	URB RENAC
185613	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	AMAZONAS Y DE NOVIEMBRE	360	NULL	NULL	NULL	NOVIEMBRE
202789	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	NULL	NULL	NULL	NULL	NULL	NULL
202791	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	AV	ARQUITECTO MODESTO LUQUE	NULL	NULL	NULL	NULL	NULL
202874	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	AV	D COMIN FTE PARADA MET	NULL	NULL	NULL	NULL	VIA
203347	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	RECREO	1	NULL	NULL	NULL	NULL
203348	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	REC	2	NULL	NULL	NULL	NULL
203575	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	NULL	NULL	NULL	NULL	NULL	NULL
203631	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	CALLE	25 ENTRE LA L Y LA LL	NULL	NULL	NULL	NULL	NULL
203681	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	AV	MACHALILLA E ALEJO LASCAN	NULL	NULL	NULL	NULL	NULL
203695	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	MACHALA E ROCAFUERTE	169	NULL	NULL	10	NULL
205318	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	CALLE	VELEZ Y LORENZO DE	NULL	NULL	NULL	NULL	NULL
203332	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	MUCHO ETA CALLION	NULL	NULL	NULL	NULL	NULL
203282	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	TRINITARIA	1	NULL	NULL	NULL	COOP POLO SUR
570091	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	NULL	NULL	KM 11	SAN FCO DE MILA	NULL	VIA DAULE LOTIZ
203692	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	BOLIVAR Y GONZALO CORDOV	NULL	NULL	SECTOR 6	NULL	DE NOVIEMBR
258636	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	LA	11	NULL	NULL	NULL	Y CAMILO DESTRUGE
288676	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	CALLE	SAUCES RODRIGO ICAZA Y	NULL	NULL	NULL	NULL	NULL
401824	0990017514001	TIENDAS INDUSTRIALES ASOCIADOS	ESQ	CHIMBORAZO Y LUQUES	217	NULL	NULL	NULL	Y LUQUE
273603	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	GERRERO VALENZ Y LA A	1	NULL	NULL	NULL	G VALE
286591	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	CC GARZOCENTRO	2000	NULL	NULL	NULL	NULL
500578	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	CALLE	ALAJUELA ENTRE PRIMERO	NULL	NULL	NULL	NULL	NULL
570386	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	CALLE	CHILE ENTRE	9	NULL	NULL	NULL	OCTUBRE

CODIGO_CLIENTI	CEDULA_RUC	APELLIDO_PENOME	TIPOCALLE	CALLE_PEADDR	NUMERO_F	KILOMETRO_I	SECTOR_PEADDR	CASA_P	URBANIZACIONZONA
104925	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	GUAYAQUIL	958	NULL	NULL	NULL	Y ESPEJO
105742	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	CALLE	TOACAZO Y MULALILLO ELOY	NULL	NULL	NULL	NULL	NULL
105829	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	NULL	NULL	NULL	NULL	NULL	NULL
131295	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	CUSUBAMBA Y APUELA	2343	NULL	NULL	NULL	NULL
131401	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	AV	GUAYASAMIN UNA	NULL	NULL	NULL	NULL	NULL
170116	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	AV	MARISCAL SUCRE ENTRE OSORIO	NULL	NULL	NULL	NULL	NULL
185603	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	ELIZARIO QUEVEDO	111	NULL	NULL	NULL	FELIX VAL
185801	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	LIZARDO RUIZ Y ALFAREROS	1	NULL	NULL	NULL	LIZ
203080	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	JOSE FERNANDEZ DAVILA	NULL	NULL	NULL	NULL	URB RENAC
185613	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	AMAZONAS Y DE NOVIEMBRE	360	NULL	NULL	NULL	NOVIEMBRE
202789	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	NULL	NULL	NULL	NULL	NULL	NULL
202791	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	AV	ARQUITECTO MODESTO LUQUE	NULL	NULL	NULL	NULL	NULL
202874	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	AV	D COMIN FTE PARADA MET	NULL	NULL	NULL	NULL	VIA
203347	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	RECREO	1	NULL	NULL	NULL	NULL
203348	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	REC	2	NULL	NULL	NULL	NULL
203575	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	NULL	NULL	NULL	NULL	NULL	NULL
203631	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	CALLE	25 ENTRE LA L Y LA LL	NULL	NULL	NULL	NULL	NULL
203681	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	AV	MACHALILLA E ALEJO LASCAN	NULL	NULL	NULL	NULL	NULL
203695	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	MACHALA E ROCAFUERTE	169	NULL	NULL	10	NULL
205318	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	CALLE	VELEZ Y LORENZO DE	NULL	NULL	NULL	NULL	NULL
203332	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	MUCHO ETA CALLJON	NULL	NULL	NULL	NULL	NULL
203282	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	TRINITARIA	1	NULL	NULL	NULL	COOP POLO SUR
570091	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	NULL	NULL	KM 11	SAN FCO DE MILA	NULL	VIA DAULE LOTIZ
203692	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	BOLIVAR Y GONZALO CORDOV	NULL	NULL	SECTOR 6	NULL	DE NOVIEMBR
258636	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	LA	11	NULL	NULL	NULL	Y CAMILO DESTRUGE
288676	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	CALLE	SAUCES RODRIGO ICAZA Y	NULL	NULL	NULL	NULL	NULL
401824	0990017514001	TIENDAS INDUSTRIALES ASOCIADOS	ESQ	CHIMBORAZO Y LUQUES	217	NULL	NULL	NULL	Y LUQUE
273603	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	GERRERO VALENZ Y LA A	1	NULL	NULL	NULL	G VALE
286591	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	NULL	CC GARZOCENTRO	2000	NULL	NULL	NULL	NULL
500578	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	CALLE	ALAJUELA ENTRE PRIMERO	NULL	NULL	NULL	NULL	NULL
570386	0990017514001	TIENDAS INDUSTRIALES ASOCIADAS	CALLE	CHILE ENTRE	9	NULL	NULL	NULL	OCTUBRE

Conclusiones y Recomendaciones

Conclusiones

- El Modelo de Gestión de Calidad de Datos basado en la metodología Data Quality de IBM, propuesto en la primera parte de este trabajo, fue utilizado en el desarrollo de un caso de estudio práctico, basándose en una fuente de datos real obtenida del ambiente de producción de una reconocida empresa nacional. Esto comprueba que el Modelo de Gestión de Calidad de Datos propuesto es aplicable en los proyectos de Calidad de Datos que son una nueva línea de negocio de la empresa auspiciante DWConsultware.
- La Gestión de Calidad de Datos fue aplicada mediante el uso de la herramienta IBM InfoSphere Information Server, el uso de la herramienta permitió mejorar la calidad de los datos iniciales desarrollando las etapas planteadas por la Metodología IBM DataQuality.
- La mejora en la calidad de datos proporcionada a los usuarios finales una visión más clara de los mismos, efectivamente convirtiendo los datos en información, lo cual indica la conveniencia de generar reglas de negocio para la captura de datos en el campo de trabajo.
- El nivel de madurez de BI del caso de estudio analizado se incrementa únicamente en la categoría de Datos del Modelo de Madurez TDWI lo cual indica que:
 - La Gestión de Calidad de Datos por sí misma no es suficiente para elevar el nivel de madurez de BI de la organización.
 - Las categorías del Modelo de Madurez TDWI que se encuentran en los puntos críticos (el Golfo y el Abismo) requieren más iteraciones de un modelo de gestión, en este caso de estudio Gestión de Calidad de Datos, para avanzar hacia las siguientes etapas del modelo.
 - La Gestión de Calidad de Datos es un proceso de ejecución a corto plazo, el cual influye en la categoría correspondiente en los Modelos de Madurez de BI como la categoría Datos en el Modelo TDWI, por lo tanto se requieren varias iteraciones de este proceso para influir de forma efectiva en otras áreas que por su naturaleza tienen ciclos de ejecución a mediano y largo plazo, como es la categoría de Presupuesto y Patrocinio. Esto se debe a que la Gestión de Calidad de Datos es un proceso enfocado casi exclusivamente hacia el área técnica de BI, mientras que las categorías de Presupuesto y Patrocinio se enfocan hacia el área de negocios y dirección de las organizaciones.

Recomendaciones

- Aplicar el Modelo de Gestión de Calidad de Datos propuesto en la primera parte de este trabajo y probado en un caso de estudio práctico en la segunda parte del mismo, en los proyectos de Gestión de Calidad de Datos de la empresa DWConsultware.
- Integrar el Modelo de Gestión de Calidad de Datos con la metodología de proyectos de Inteligencia de Negocios para el portafolio de clientes de la empresa DWConsultware.
- Desarrollar reglas de estandarización de datos adaptadas a los formatos, leyes y reglamentos utilizados en el país para objetos comunes de negocio como: clientes, direcciones, personas naturales y jurídicas, que se puedan integrar con la plataforma IBM InfoSphere Information Server.
- Considerar el desarrollo de Modelos de Gestión aplicables a otras áreas de los Modelos de Madurez de BI como posibles temas para la elaboración de proyectos de graduación del programa de Maestría en Gerencia de Sistemas.
- Considerar el desarrollo de un estudio comparativo entre los diferentes Modelos de Madurez de BI como posible tema para la elaboración de proyectos de graduación de pregrado.

Muchas Gracias.

Conclusiones

- El Modelo de Gestión de Calidad de Datos basado en la metodología Data Quality de IBM, propuesto en la primera parte de este trabajo, fue utilizado en el desarrollo de un caso de estudio práctico, basándose en una fuente de datos real obtenida del ambiente de producción de una reconocida empresa nacional. Esto comprueba que el Modelo de Gestión de Calidad de Datos propuesto es aplicable en los proyectos de Calidad de Datos que son una nueva línea de negocio de la empresa auspiciante DWConsultware.
 - La Gestión de Calidad de Datos fue aplicada mediante el uso de la herramienta IBM Infosphere Information Server, el uso de la herramienta permitió mejorar la calidad de los datos iniciales desarrollando las etapas planteadas por la Metodología IBM DataQuality.
 - La mejora en la calidad de datos proporciona a los usuarios finales una visión más clara de los mismos, efectivamente convirtiendo los datos en información, lo cual indica la conveniencia de generar reglas de negocio para la captura de datos en el campo de trabajo.
-
- El nivel de madurez de BI del caso de estudio analizado se incrementa únicamente en la categoría de Datos del Modelo de Madurez TDWI lo cual indica que:
 - La Gestión de Calidad de Datos por sí misma no es suficiente para elevar el nivel de madurez de BI de la organización.
 - Las categorías del Modelo de Madurez TDWI que se encuentran en los puntos críticos (el Golfo y el Abismo) requieren más iteraciones de un modelo de gestión, en este caso de estudio Gestión de Calidad de Datos, para avanzar hacia las siguientes etapas del modelo.
 - La Gestión de Calidad de Datos es un proceso de ejecución a corto plazo, el cual influye en la categoría correspondiente en los Modelos de Madurez de BI como la categoría Datos en el Modelo TDWI, por lo tanto se requieren varias iteraciones de este proceso para influir de forma efectiva en otras áreas que por su naturaleza tienen ciclos de ejecución a mediano y largo plazo, como es la categoría de Presupuesto y Patrocinio. Esto se debe a que la Gestión de Calidad de Datos es un proceso enfocado casi exclusivamente hacia el área técnica de BI, mientras que las categorías de Presupuesto y Patrocinio se enfocan hacia el área de negocios y dirección de las organizaciones.

Conclusiones

- El Modelo de Gestión de Calidad de Datos basado en la metodología Data Quality de IBM, propuesto en la primera parte de este trabajo, fue utilizado en el desarrollo de un caso de estudio práctico, basándose en una fuente de datos real obtenida del ambiente de producción de una reconocida empresa nacional. Esto comprueba que el Modelo de Gestión de Calidad de Datos propuesto es aplicable en los proyectos de Calidad de Datos que son una nueva línea de negocio de la empresa auspiciante DWConsultware.
- La Gestión de Calidad de Datos fue aplicada mediante el uso de la herramienta IBM Infosphere Information Server, el uso de la herramienta permitió mejorar la calidad de los datos iniciales desarrollando las etapas planteadas por la Metodología IBM DataQuality.
- La mejora en la calidad de datos proporciona a los usuarios finales una visión más clara de los mismos, efectivamente convirtiendo los datos en información, lo cual indica la conveniencia de generar reglas de negocio para la captura de datos en el campo de trabajo.
- El nivel de madurez de BI del caso de estudio analizado se incrementa únicamente en la categoría de Datos del Modelo de Madurez TDWI lo cual indica que:
 - La Gestión de Calidad de Datos por sí misma no es suficiente para elevar el nivel de madurez de BI de la organización.
 - Las categorías del Modelo de Madurez TDWI que se encuentran en los puntos críticos (Data Governance, Data Architecture, Data Integration, Data Security, Data Analytics)



de los datos iniciales desarrollando las etapas planteadas por la metodología TDWI DataQuality.

- La mejora en la calidad de datos proporciona a los usuarios finales una visión más clara de los mismos, efectivamente convirtiendo los datos en información, lo cual indica la conveniencia de generar reglas de negocio para la captura de datos en el campo de trabajo.
- El nivel de madurez de BI del caso de estudio analizado se incrementa únicamente en la categoría de Datos del Modelo de Madurez TDWI lo cual indica que:
 - La Gestión de Calidad de Datos por no es suficiente por sí misma para elevar el nivel de madurez de BI de la organización.
 - Las categorías del Modelo de Madurez TDWI que se encuentran en los puntos críticos (el Golfo y el Abismo) requieren más iteraciones de un modelo de gestión, en este caso de estudio Gestión de Calidad de Datos, para avanzar hacia las siguientes etapas del modelo.
 - La Gestión de Calidad de Datos es un proceso de ejecución a corto plazo, el cual influye en la categoría correspondiente en los Modelos de Madurez de BI como la categoría Datos en el Modelo TDWI, por lo tanto se requieren varias iteraciones de este proceso para influir de forma efectiva en otras áreas que por su naturaleza tienen ciclos de ejecución a mediano y largo plazo, como es la categoría de Presupuesto y Patrocinio. Esto se debe a que la Gestión de Calidad de Datos es un proceso enfocado casi exclusivamente hacia el área técnica de BI, mientras que las categorías de Presupuesto y Patrocinio se enfocan hacia el área de negocios y dirección de las organizaciones.

Recomendaciones

- Aplicar el Modelo de Gestión de Calidad de Datos propuesto en la primera parte de este trabajo y probado en un caso de estudio práctico en la segunda parte del mismo, en los proyectos de Gestión de Calidad de Datos de la empresa DWConsultware.
- Integrar el Modelo de Gestión de Calidad de Datos con la metodología de proyectos de Inteligencia de Negocios para el portafolio de clientes de la empresa DWConsultware.
- Desarrollar reglas de estandarización de datos adaptadas a los formatos, leyes y reglamentos utilizados en el país para objetos comunes de negocio como: clientes, direcciones, personas naturales y jurídicas, que se puedan integrar con la plataforma IBM Infosphere Information Server.

- Considerar el desarrollo de Modelos de Gestión aplicables a otras áreas de los Modelos de Madurez de BI como posibles temas para la elaboración de proyectos de graduación del programa de Maestría en Gerencia de Sistemas.
- Considerar el desarrollo de un estudio comparativo entre los diferentes Modelos de Madurez de BI como posible tema para la elaboración de proyectos de graduación de pregrado.

Recomendaciones

- Aplicar el Modelo de Gestión de Calidad de Datos propuesto en la primera parte de este trabajo y probado en un caso de estudio práctico en la segunda parte del mismo, en los proyectos de Gestión de Calidad de Datos de la empresa DWConsultware.
- Integrar el Modelo de Gestión de Calidad de Datos con la metodología de proyectos de Inteligencia de Negocios para el portafolio de clientes de la empresa DWConsultware.
- Desarrollar reglas de estandarización de datos adaptadas a los formatos, leyes y reglamentos utilizados en el país para objetos comunes de negocio como: clientes, direcciones, personas naturales y jurídicas, que se puedan integrar con la plataforma IBM Infosphere Information Server.
- Considerar el desarrollo de Modelos de Gestión aplicables a otras áreas de los Modelos de Madurez de BI como posibles temas para la elaboración de proyectos de graduación del programa de Maestría en Gerencia de

formatos, leyes y reglamentos utilizados en el país para objetos comunes de negocio como: clientes, direcciones, personas naturales y jurídicas, que se puedan integrar con la plataforma IBM Infosphere Information Server.

- Considerar el desarrollo de Modelos de Gestión aplicables a otras áreas de los Modelos de Madurez de BI como posibles temas para la elaboración de proyectos de graduación del programa de Maestría en Gerencia de Sistemas.
- Considerar el desarrollo de un estudio comparativo entre los diferentes Modelos de Madurez de BI como posible tema para la elaboración de proyectos de graduación de pregrado.

Muchas Gracias.