

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA EN SISTEMAS

DISEÑO E IMPLEMENTACIÓN DE UN FRAMEWORK PARA LA
EVALUACIÓN DE LA CALIDAD DE DATOS EN UN
DATAWAREHOUSE EN EL SISTEMA DE REGISTRO FUENTE
(CLIENTES) DE UNA ENTIDAD FINANCIERA

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL GRADO DE
MAGISTER EN SOFTWARE MENCIÓN CALIDAD

FREDDY RUBEN TORRES TIPANTIZA

freddy.torres@epn.edu.ec

Director: MOLINA BUSTAMANTE MARCO EDUARDO, PhD.

marco.molinab@epn.edu.ec

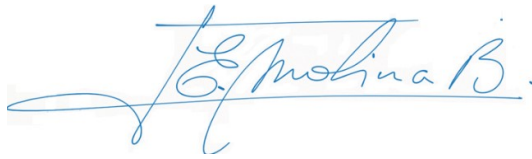
Co-Director: MONTENEGRO ARMAS CARLOS ESTALESMIT WILLAM, MSc.

carlos.montenegro@epn.edu.ec

Quito, octubre 2020

APROBACIÓN DEL DIRECTOR

Como director del trabajo de titulación “DISEÑO E IMPLEMENTACIÓN DE UN FRAMEWORK PARA LA EVALUACIÓN DE LA CALIDAD DE DATOS EN UN DATAWAREHOUSE EN EL SISTEMA DE REGISTRO FUENTE (CLIENTES) DE UNA ENTIDAD FINANCIERA” desarrollado por Freddy Rubén Torres Tipantiza, estudiante de la Maestría en Software, Mención: Calidad, habiendo supervisado la realización de este trabajo y realizado las correcciones correspondientes, doy por aprobada la redacción final del documento escrito para que prosiga con los trámites correspondientes a la sustentación de la Defensa oral.

A handwritten signature in blue ink, reading "M. Molina B.", is positioned above a horizontal line. The signature is stylized and cursive.

MOLINA BUSTAMANTE MARCO EDUARDO, PhD.

DIRECTOR

APROBACIÓN DEL CO-DIRECTOR

Como director del trabajo de titulación “DISEÑO E IMPLEMENTACIÓN DE UN FRAMEWORK PARA LA EVALUACIÓN DE LA CALIDAD DE DATOS EN UN DATAWAREHOUSE EN EL SISTEMA DE REGISTRO FUENTE (CLIENTES) DE UNA ENTIDAD FINANCIERA” desarrollado por Freddy Rubén Torres Tipantiza, estudiante de la Maestría en Software, Mención: Calidad, habiendo supervisado la realización de este trabajo y realizado las correcciones correspondientes, doy por aprobada la redacción final del documento escrito para que prosiga con los trámites correspondientes a la sustentación de la Defensa oral.



MONTENEGRO ARMAS CARLOS ESTALESMIT

WILLAM, MSc.

CO-DIRECTOR

DECLARACIÓN DE AUTORÍA

Yo, Freddy Rubén Torres Tipantiza, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Ing. Freddy Rubén Torres Tipantiza

DEDICATORIA

Dedico este trabajo a Dios por estar conmigo en todo momento forjando mi camino, haberme dado fuerza y valor para culminar esta etapa de mi vida.

A mis padres y hermanas que me enseñaron que con constancia nada es imposible cuando se tiene un sueño.

A mi esposa Mónica e hijos Neithan y Sean por su motivación, apoyo incondicional, amor y ante todo por enseñarme que la vida es hermosa junto a ellos.

AGRADECIMIENTO

Un sincero agradecimiento a mis tutores, Marco Molina y Carlos Montenegro por su contribución y aporte profesional para la realización de este trabajo.

ÍNDICE DE CONTENIDO

LISTA DE FIGURAS.....	I
LISTA DE TABLAS	II
LISTA DE ABREVIATURAS.....	V
RESUMEN	VI
1. INTRODUCCIÓN	1
1.1 ANTECEDENTES	3
1.2 PLANTEAMIENTO DEL PROBLEMA	5
1.3 OBJETIVO GENERAL.....	7
1.4 OBJETIVOS ESPECÍFICOS	7
2. MARCO TEÓRICO	8
2.1 Calidad de datos	8
2.2 Dimensiones de calidad de datos	10
2.3 Categorías de datos	12
2.4 Sistema de registro fuente	13
2.5 Proceso de extracción, transformación y carga (ETL)	14
2.6 Datawarehouse (DWH)	15
2.7 Framework	16
3 METODOS Y HERRAMIENTAS	22
3.1 DESCRIPCIÓN DEL FRAMEWORK.....	22
3.2 IDENTIFICACIÓN DE LOS STAKEHOLDERS Y SUS NECESIDADES.....	22
3.3 IDENTIFICACIÓN INCONVENIENTES DE DQ	23
3.4 IDENTIFICACIÓN REQUISITOS RELEVANTES	24
3.5 RECOPIACIÓN DE DATOS	25
3.6 COMPRENDER LOS DATOS Y SUS ESPECIFICIDADES RELEVANTES	26
3.7 TECNOLOGÍA ASOCIADA CON LOS DATOS	28
3.8 DEFINICIÓN DE LOS ATRIBUTOS DE CALIDAD DE DATOS:.....	29
3.9 SELECCIÓN DE LA HERRAMIENTA DE EVALUACIÓN.....	30
3.10 DEFINIR MECANISMOS Y HERRAMIENTAS DE REPORTE	32
4 RESULTADOS Y DISCUSIÓN	33
4.1 RESULTADOS	33

4.1.1 RESULTADOS DE LA EVALUACIÓN DE CALIDAD	34
4.2 DISCUSIÓN.....	48
5 CONCLUSIONES Y RECOMENDACIONES.....	50
5.1 CONCLUSIONES.....	50
5.2 RECOMENDACIONES.....	51
6 REFERENCIAS	53
7 ANEXOS.....	55
ANEXO I RESOL_SB-2017-602 - SBS ORGANISMO DE CONTROL.....	55
ANEXO II RESOLUCIÓN NO. JB-2010-1782 DE JUNTA BANCARIA.....	57
ANEXO III RESOLUCIÓN NO. SB-2018-771 SBS ORGANISMO DE CONTROL.....	59
ANEXO IV PLAN ESTRATÉGICO DE TECNOLOGÍAS DE LA INFORMACIÓN.....	61
ANEXO V MODELO DE LA ENCUESTA APLICADA	64

LISTA DE FIGURAS

Figura 1 Proceso de toma decisiones	9
Figura 2 Modelo de calidad de datos ISO25012.....	11
Figura 3 Relación entre categoría de datos	13
Figura 4 Proceso ETL.....	15
Figura 5 Cubo tridimensional de los datos de venta	16
Figura 6 Framework de Calidad de Datos.....	21
Figura 7 Stakeholders y sus necesidades.....	23
Figura 8 Requisitos relevantes	24
Figura 9 Modelo del BDWH SSoR	26
Figura 10 Arquitectura de datos	28
Figura 11 Detalle de la tecnología asociada	29
Figura 12 Resultado de la evaluación de calidad de datos	35
Figura 13 Resultado de la evaluación exactitud, factor correctitud sintáctica	37
Figura 14 Resultado de la evaluación exactitud, factor precisión	39
Figura 15 Resultado de la evaluación completitud, factor densidad.....	41
Figura 16 Resultado de la evaluación consistencia, factor integridad de dominio .	42
Figura 17 Resultado de la evaluación consistencia, factor inter-relación	43
Figura 18 Resultado de la evaluación unicidad, factor no duplicación.....	45
Figura 19 Resultado de la evaluación unicidad, factor no contradicción.	47

LISTA DE TABLAS

Tabla 1 Modelo de calidad de datos ISO/IEC 25012	12
Tabla 2 Criterios para realizar la SLR	17
Tabla 3 Cuadro comparativo FDQ.....	19
Tabla 4 Identificación de fases y tareas FDQ	20
Tabla 5 Fases y tareas del FDQ.....	22
Tabla 6 Inconvenientes de calidad de datos	24
Tabla 7 Detalle de la recopilación de la información	26
Tabla 8 Dimensiones identificadas.....	29
Tabla 9 Factor y tipo de error.....	30
Tabla 10 Herramientas comerciales para la evaluación de la calidad de datos	31
Tabla 11 Resumen de herramientas para la evaluación de calidad de datos	32
Tabla 12 Comparación de herramientas de análisis de datos.	32
Tabla 13 Desagregación de la evaluación de calidad de datos	33
Tabla 14 Cualitativa de medición	34
Tabla 15 Dimensiones de calidad	34
Tabla 16 Estructura de la matriz de calidad de datos	34
Tabla 17 Fórmula para calcular la calidad de datos.....	35
Tabla 18 Evaluación de calidad: dimensión exactitud, factor correctitud sintáctica	36
Tabla 19 Ejemplos descripción: dimensión exactitud, factor correctitud sintáctica	36
Tabla 20 Evaluación de calidad: dimensión exactitud, factor precisión	38
Tabla 21 Ejemplos descripción: dimensión exactitud, factor correctitud precisión.	38
Tabla 22 Evaluación de calidad: dimensión completitud, factor correctitud densidad.....	40
Tabla 23 Ejemplos descripción: dimensión completitud, factor correctitud densidad	40
Tabla 24 Evaluación de calidad: dimensión consistencia, factor integridad de dominio.....	42
Tabla 24 Ejemplos descripción: dimensión consistencia, factor integridad de dominio.....	42

Tabla 26 Evaluación de calidad: dimensión consistencia, factor inter-relación.....	43
Tabla 27 Evaluación de calidad: dimensión unicidad, factor no duplicación	44
Tabla 28 Ejemplos de la dimensión unicidad, factor no duplicación.....	44
Tabla 29 Evaluación de calidad: dimensión unicidad, factor no contradicción.....	46
Tabla 30 Ejemplos de la dimensión unicidad, factor no contradicción.....	46
Tabla 31 Resultado final de la evaluación de la calidad de datos.....	47

LISTA DE ANEXOS

Anexo I RESOL_SB-2017-602 - SBS Organismo de Control	55
Anexo II Resolución No. JB-2010-1782 de Junta Bancaria.....	57
Anexo III Resolución No. SB-2018-771 SBS Organismo de Control.....	59
Anexo IV Plan estratégico de tecnologías de la información.....	61
Anexo V Modelo de la encuesta aplicada.....	64

LISTA DE ABREVIATURAS

SSoR	Sistema de registro fuente
ETL	Extraer, transformar y cargar
DQ	Data Quality
IDC	Informe de datos de cotización
DWH	Datawarehouse
SLR	Revisión sistemática literaria
SBS	Superintendencia de Bancos y Seguros
TIC	Tecnologías de la información y comunicación
ISO	International Organization for Standardization
DWH	Datawarehouse

RESUMEN

El presente trabajo tiene como objetivo el diseño e implementación de un framework para la evaluación de calidad de datos en un datawarehouse en el sistema de registro fuente (clientes) de una entidad financiera. El framework propuestos se basó en la revisión bibliográfica de fuentes similares que se ajustan a las necesidades, el cual proporciona un marco de trabajo compuesto por fases que facilitan y garantizan información de calidad.

En función del contenido de la información se seleccionó cuatro dimensiones con sus respectivos siete factores de calidad para la evaluación del datawarehouse, el entendimiento de estos conceptos permitió tener una percepción más amplia del proceso de evaluación, de la misma manera se determinó que la dimensión completitud factor densidad (campos nulos y blancos) obtiene el promedio más baja del análisis 8.27/10.

Con los resultados obtenidos en la evaluación se determina un promedio total de 9,347 que la ubica en un rango satisfactorio; con ello se puede evidenciar que la calidad del datawarehouse es admisible, sin embargo para una entidad financiera se requiere alcanzar un grado altamente satisfactorio debido a su giro de negocio que demanda información de calidad.

En un entorno empresarial como el financiero sus éxitos están basados en la calidad y enriquecimiento de datos, si bien es cierta la entidad financiera evaluada presenta un promedio de calidad de datos de 9,35/10 en el registro fuente no es suficiente, una adecuada estrategia de calidad de datos permitirá que la entidad pueda mejorar su imagen y marca.

Palabras clave: Framework de calidad de datos, evaluación de calidad de datos, entidad financiera, información de calidad.

ABSTRACT

The objective of this work is the design and implementation of a framework for the evaluation of data quality of a data warehouse, in the source registration system (clients) of a financial institution. The proposed framework was based on a bibliographic review of similar sources that fit the needs, which provides a framework composed of phases that facilitate and guarantee quality information.

Based on the content of the information, four dimensions with their respective seven quality factors were selected for the evaluation of the data warehouse. Understanding these concepts allowed for a broader perception of the evaluation process. In the same way, it was determined that the dimension of completeness of the density factor (null and blank fields) obtains the lowest average of the analysis, which corresponds to 8.27/10.

With the results obtained in the evaluation, a total average of 9.347 is determined, which places it in a satisfactory range. With this, it can be seen that the quality of the data warehouse is admissible; however, for a financial institution, it is required to achieve a highly satisfactory degree, because its line of business demands quality information.

In a business environment such as finance, its successes are based on the quality and enrichment of data; however, although it is true that the evaluated financial entity presents an average data quality of 9.35/10 in the source record, this is not enough, since an adequate data quality strategy must also be considered, which will allow the entity to improve its image and brand.

Keywords: data quality framework, data quality assessment, financial institution, quality information.

1. INTRODUCCIÓN

Según el **Anexo I** - Superintendencia de bancos, resolución N# SB-2017-602, indica:

La Superintendencia de Bancos, mediante la resolución N°_SB-2017-602, según el Código Orgánico Monetario y Financiero, se establece como función de la Superintendencia de Bancos ejercer la vigilancia, auditoria, control y supervisión del cumplimiento de las disposiciones de este código y de las regulaciones dictadas en lo que corresponde a las actividades financieras ejercidas por las entidades que conforman los sectores financieros público y privado.

Resuelve expedir la norma de control para la gestión integral y administración de las entidades de los sectores financieros público y privado, denominadas en adelante como entidades controladas.

Entre las disposiciones generales de la norma, seestablece que las entidades controladas deben disponer de un sistema informático capaz de proveer a la administración y a las áreas involucradas, toda la información necesaria para identificar, medir, controlar, mitigar y monitorear las exposiciones de riesgo que están asumiendo y apoyar en la toma de decisiones oportuna y adecuada.

Por otro lado en el **Anexo II**, el Código de Derechos del Usuario Financiero de la RESOLUCIÓN No. JB-2010-1782 de LA JUNTA BANCARIA, establece que las personas tienen derecho a disponer de bienes y servicios de óptima calidad y a elegirlos con libertad, así como a una información precisa y no engañosa sobre su contenido y características.

El usuario tendrá derecho a recibir protección que garanticen la seguridad de las operaciones financieras, del defensor del cliente, de la Superintendencia de Bancos y Seguros o de otras instancias administrativas o judiciales pertinentes.

Recibir protección de los datos personales que las entidades financieras obtengan del usuario para la prestación de productos y servicios financieros prestados por vía electrónica. Las instituciones financieras adoptarán específicamente las medidas de seguridad necesarias para este tipo de operaciones financieras.

Obtener protección de los datos personales sobre su solvencia patrimonial, crediticia, y demás; donde las instituciones financieras respeten las normas relativas a la reserva y sigilo bancario.

Sigilo bancario, es la confidencialidad en la entrega de información que las instituciones del sistema financiero y sus funcionarios deben mantener sobre los datos y demás captaciones de información que se reciban.

Por este régimen limitado de no revelar datos bancarios e información privada de los clientes para beneficio personal o de terceros; da lugar a reservar el nombre de la entidad financiera para este estudio, actualmente la organización maneja gran cantidad de información confidencial, por ello se ha enfocado en la información del sistema de registro fuente denominado SSoR, por sus siglas en inglés, que analizará la información básica de clientes.

Según el **Anexo III**, Superintendencia de Bancos, resolución N° SB-2018-771, indica: Con el objeto de garantizar que la administración de la tecnología de la información soporte los requerimientos de operación actuales y futuros de la entidad, debe contar al menos con lo siguiente:

1. El apoyo y compromiso formal del directorio, a través de la aprobación de un plan estratégico de tecnología de la información alineado con el plan estratégico institucional; y, un plan operativo anual que establezca las actividades a ejecutar, traducido en tareas, cronogramas, personal responsable y presupuesto, de manera que se asegure el logro de los objetivos tecnológicos propuestos.
2. La unidad de Tecnología de la Información elaborará e implementará un plan informático estratégico para administrar y dirigir todos los recursos tecnológicos, el mismo que estará alineado con el plan estratégico institucional y éste, a su vez, con el Plan Nacional de Desarrollo y las políticas públicas de Gobierno.
3. El plan informático estratégico tendrá un nivel de detalle suficiente para permitir la definición de planes operativos de tecnología de información y especificará cómo ésta contribuirá a los objetivos estratégicos de la organización; incluirá un análisis de la situación actual y las propuestas de mejora con la participación de todas las unidades de la organización, se considerará la estructura interna, procesos, infraestructura, comunicaciones, aplicaciones y servicios a brindar, así como la definición de estrategias, riesgos, cronogramas, presupuesto de la inversión y

operativo, fuentes de financiamiento y los requerimientos legales y regulatorios de ser necesario.

4. El plan estratégico y los planes operativos de tecnología de información, así como el presupuesto asociado a éstos serán analizados y aprobados por la máxima autoridad de la organización e incorporados al presupuesto anual; además, se actualizarán de manera permanente y serán monitoreados y evaluados en forma trimestral para determinar su grado de ejecución a fin de tomar las medidas necesarias en caso de desviaciones.

La entidad financiera pretende mejorar sus capacidades de toma de decisiones en el sistema de registro fuente (SSoR), mediante la implementación y diseño de un framework para evaluar la calidad de datos aplicado al datawarehouse.

Actualmente la empresa recopila información del cliente a través del portal, dicha información se registra en el core bancario y posteriormente se almacena en el datawarehouse del sistema de registro fuente, esta información del SSoR presenta inconsistencias en sus reportes, lo cual influye negativamente en la toma de decisiones que influyen en los procesos operativos y transaccionales.

La calidad de los datos es un factor clave para la toma de decisiones. A medida que aumenta la disponibilidad de datos financieros en línea, también aumenta el problema de mantenimiento de la calidad de los datos[1].

Los frameworks o marcos de trabajo definen, modelan o evalúan la calidad de datos;[2][3][4][5]. Para el estudio propuesto se realizará un análisis previo de la información del sistema de registro fuente, que se refiera a la información básica de los clientes, para lo cual se llevarán a cabo los siguientes procesos:

1. Definición de las necesidades y del enfoque empresarial.
2. Estudio del entorno de información.
3. Evaluación de la calidad de los datos.

1.1 Antecedentes

El **Anexo IV**- Plan estratégico de tecnologías de la información, indica lo siguiente:

El Directorio de la Entidad Financiera ha considerado que mediante decreto ejecutivo No. 677, el Presidente de la República creó una entidad financiera con personalidad jurídica

propia y jurisdicción nacional, con patrimonio autónomo; autonomía técnica, administrativa, financiera y presupuestaria; que forma parte del sector financiero del País, cuyo objetivo es el ejercicio de actividades financieras previamente autorizadas por la Superintendencia de Bancos, y la prestación de servicios financieros de crédito, ahorro e inversión, bajo el criterio de intermediación financiera de recursos públicos y privados, atendiendo a personas naturales y jurídicas en sectores de producción, comercio, servicios, entre otros; con el enfoque de desarrollo local, coadyuvando al fortalecimiento de las políticas y planes nacionales de desarrollo, a través de mecanismos de la banca financiera.

La entidad financiera desarrollará productos y servicios financieros especiales y específicos para cubrir las necesidades de financiamiento de cada segmento de negocio y así promover el desarrollo socioeconómico.

La Agenda de Desarrollo del país y su concreción por medio de las políticas públicas de cada una de las instituciones del Estado, se orienta a lograr el bienestar de los ciudadanos y ciudadanas, entendiendo como una apuesta de cambio que se construye en un proceso orientado a promover en el país un modelo de desarrollo incluyente, que tome en cuenta a las poblaciones históricamente excluidas de la dinámica económica y productiva.

En este marco, la economía y en particular los servicios financieros constituyen elementos determinantes que cooperan a la inserción productiva de los diferentes actores.

El análisis de la evolución y de la situación financiera actual del banco es de vital importancia para la identificación de debilidades y fortalezas, así como para el establecimiento de objetivos, estrategias, indicadores, metas ambiciosas y alcanzables, que en el marco de los instrumentos de la planificación nacional, contribuyan a la consolidación y empoderamiento de la entidad financiera.

La Banca emite las siguientes directrices propias del uso de las TICs en el entorno de la prestación de servicios financieros que se encuentran definidas en el marco normativo que las rige:

- La innovación de los productos y servicios financieros se debe alcanzar, sin poner en riesgo la seguridad y conformidad de los datos y garantizando la continuidad de las operaciones y servicios.
- La incorporación de nuevas tecnologías debe garantizar la integración con los otros elementos del entorno tecnológico y la interrelación con sistemas heredados debe ser automática y transparente preservando la seguridad e integridad de la información.

A fin de potenciar su crecimiento y mejorar su posicionamiento, el Banco requiere mantener una gestión financiera eficaz y eficiente que garantice la sostenibilidad institucional; para ello contempla y define ejes clave de la política pública en torno al sistema financiero del Estado, a través de los diferentes programas y proyectos constantes en los Planes Operativos Anuales de la Institución.

La entidad financiera consciente de que los avances tecnológicos y la innovación constituyen la principal ventaja competitiva, para reducir costos y gastos, provocan además un incremento del nivel de satisfacción de los clientes, al tiempo que permiten dar cumplimiento a la normativa en el ámbito de las tecnologías de la Información, emprenderá las acciones necesarias para que se evalúe el sistema de información, del que dispone (SSoR) y que se identifiquen las oportunidades de mejora.

Una de las acciones fundamentales es la de mejorar los procesos internos, en este caso la propuesta se enfoca en el análisis de la información básica de los clientes, con lo que se incorporará un valor agregado a las necesidades de los procesos operativos y transaccionales.

1.2 Planteamiento del problema

Las actividades del día a día y la toma de decisiones en una organización se basan en información obtenida a partir del análisis de datos, pues, esto último proporciona los insumos para construir procesos confiables y eficaces. En vista de que los datos constituyen un recurso importante en toda organización, su calidad es fundamental para que los usuarios empresariales sean capaces de identificar problemas relacionados con su accionar dentro de la empresa[6].

Uno de los objetivos de las organizaciones es usar buenas prácticas para mejorar la calidad de la información y así reducir los errores, a fin de evitar la pérdida de oportunidades de negocio e incrementar la satisfacción de sus clientes[2].

Un marco de trabajo de calidad de datos se define como un conjunto de guías y técnicas, que desde la información de entrada define el proceso para usar la información, para medir y mejorar la calidad de datos de una organización a través de las fases o etapas definidas y puntos de decisión[7]. De igual manera una metodología definida y aplicada en datos financieros define el reconocimiento y clasificación de variables primarias, análisis de técnicas y reglas de calidad de datos[4].

Son muchos los trabajos académicos que definen, modelan, evalúan o proponen mejoras para la calidad de datos. La calidad suele estudiarse en múltiples dimensiones, caracterizadas por los atributos de los datos, por ejemplo, exactitud, accesibilidad e integridad, entre otros[8]. En el presente trabajo se identificarán las dimensiones que mejor se adaptan a las necesidades del caso de estudio.

Cada dimensión de calidad de los datos, requiere diferentes herramientas, técnicas y procesos para ser medida y es importante valorar el tiempo y los recursos disponibles para realizar su evaluación[2].

De igual manera los frameworks de calidad de datos deben satisfacer las necesidades de las organizaciones, mediante la implementación de conceptos clave, así como la selección de pasos, actividades y técnicas apropiadas[2].

La entidad financiera cuenta con varios medios a nivel nacional que reciben los datos de los clientes por medio de sus asesores; esta información se registra y almacena en el core bancario, por medio de herramientas tecnológicas que ejecutan el proceso de extracción, transporte y carga (ETL) de los datos en el datawarehouse del sistema de registro fuente (SSoR).

Este datawarehouse al ser un activo estratégico muy importante en caso de haber datos inconsistentes, todos los procesos se verán afectados ocasionando una toma de decisiones con alta probabilidad de ser errónea.

En la actualidad la información del SSoR, que se extrae del datawarehouse y que es utilizada diariamente por los procesos de negocio, contiene datos inconsistentes debido a

que no cuentan con un framework que garantice una calidad; además no tienen un registro de monitoreo de inconsistencias, ni procesos de validación que ayuden a prevenir o corregir los errores de los datos. Debido a esta problemática, no ha sido posible contar con información de calidad, lo que ha dado lugar a toma de decisiones cuestionables que, a la larga no generar valor por el contrario su aporte para la empresa es negativo.

1.3 Objetivo general

Diseñar e implementar un framework para la evaluación de calidad de los datos de un datawarehouse en el sistema de registro fuente(clientes) en una entidad financiera.

1.4 Objetivosespecíficos

- Identificar las dimensiones de calidad de datos a ser evaluados, en el datawarehouse del sistema de registro fuente (clientes).
- Seleccionar las herramientas y procesos necesarios para realizar la evaluación.
- Definir mecanismo y herramientas de reporte.
- Ejecutar la evaluación de calidad de datos.

2. MARCO TEÓRICO

En la actualidad y en cualquier organización, los datos son activos estratégicos de mucha importancia y su calidad constituye una característica primordial. Trabajar con datos de calidad resulta imprescindible para obtener información confiable, que permita realizar una buena toma de decisiones[8].

En esta sección se enuncian conceptos teóricos para un mejor entendimiento del análisis que se, aborda en los siguientes capítulos del presente estudio.

2.1 Calidad de datos

Calidad de Datos, DQ por sus siglas en inglés, es un área de investigación muy amplia, que implica muchos aspectos, problemas y desafíos diferentes. Además, tiene una enorme relevancia para la industria, debido a su gran impacto en los sistemas de información. El término “Calidad de Datos” se usa para referirse a un conjunto de características que deben poseer los datos, tales como exactitud, accesibilidad, integridad, entre otras. La falta de calidad en los datos a menudo se experimenta en la operación cotidiana de los sistemas de información, por ejemplo: una dirección incorrecta o duplicada[9].

La definición de calidad de datos manifiesta que es “la capacidad de cumplir con los requerimientos necesarios para el uso de los datos”. Por tanto, los datos carecen de calidad en la medida en que no satisfacen los requerimientos. En otras palabras, según los autores, la calidad de datos depende tanto del uso que se les vaya a dar, como de los datos en sí[9].

Una de las razones para abordar los problemas de calidad de datos, es la creciente necesidad de integrarlos a partir de fuentes de datos heterogéneas, ya que la mala calidad de los datos obstaculiza tal integración[9].

El uso de las tecnologías de la información y comunicación (TICs), el crecimiento continuo de los datos junto a la facilidad de acceso a los mismo, han dado lugar a la realización de actividades complejas a partir del uso de herramientas informáticas que permiten recoger, almacenar, analizar, procesar y visualizar grandes cantidades de datos.

El mercado de datos en la actualidad es un negocio de miles de millones de dólares que está en continuo crecimiento. Según el informe de datos de cotización (IDC) del año 2016,

se estima que el mercado de datos en 2020 tendrá un valor aproximado equivalente a ciento treinta y un mil millones de dólares, por lo que la explotación de grandes volúmenes de datos en diversos sectores representará un especial potencial de crecimiento económico. Por ello, es necesario adoptar nuevas tecnologías, aplicaciones, casos de uso y modelos de negocio, entre varios sectores y dominios, que aseguren una rápida adopción por parte de los clientes y que proporcionen un importante crecimiento de la competitividad[8].

Esta evolución es una realidad que las empresas deben asimilar y explotar de una manera estructurada, agresiva y ambiciosa, para crear valor. Las actividades económicas y sociales han dependido por mucho tiempo de los datos, pero el aumento del volumen, velocidad, variedad y su valor social y económico, apuntan a un cambio de paradigma que se presenta como un nuevo modelo de sistema socioeconómico[8].

Por ello, los datos se han transformado en un recurso clave para las organizaciones, razón por la cual su análisis y gestión se han convertido en factores imprescindibles para las instituciones. En un mercado cada vez más competitivo, toda empresa debe tener claro cuál es su principal objetivo (inversiones, reducción de problemas, etc.). Todo ello se puede conseguir con tomas de decisiones acertadas basadas en información completa y de calidad[8].

La toma de decisiones es el proceso por medio del cual se obtiene como resultado una o más decisiones que tienen como objetivo brindar solución a una situación, como se muestra en la Figura 1.

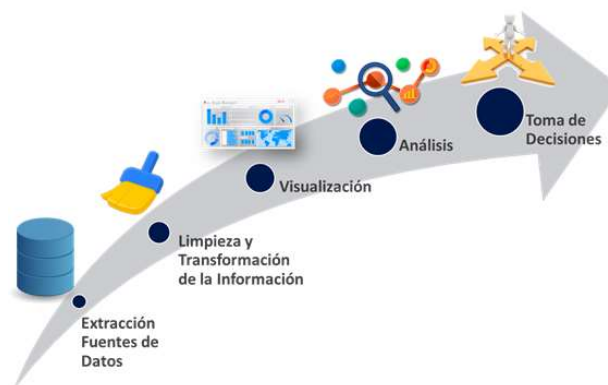


Figura 1Proceso de toma decisiones Tomado de [8].

Cabe destacar que la calidad de los datos depende del objetivo que busque cada organización. Mientras que para una empresa un dato puede resultar de especial interés, para otra podría ser insignificante, por lo que es fundamental que los datos a tratar sean útiles y que se alineen con los objetivos de la organización. De ese modo, los datos serán un activo valioso que puede ser usado como soporte de los procesos y para la toma de decisiones[8].

Uno de los objetivos en un proceso de calidad de datos es determinar su consistencia y homogeneidad. Para que este objetivo se cumpla, los analistas deben encontrar sus inconsistencias, aislarlas, realizar un análisis y proporcionar una solución válida, mediante la utilización de herramientas o frameworks para, de esa forma, usar los datos de forma adecuada[8].

2.2 Dimensiones de calidad de datos

Una dimensión de DQ proporciona una forma de medir y administrar la calidad, tanto de los datos como de la información. Las dimensiones serán seleccionadas de las que mejor se adapten a las necesidades de cada organización[10].

Para medir cada una de las dimensiones de DQ, se requiere del uso de diferentes herramientas, técnicas y procedimientos, por lo que cada una de ellas demandará un cierto presupuesto, un determinado equipo humano y un tiempo específico para efectuar su evaluación[10].

En el modelo de DQ de la Organización Internacional de Normalización ISO 25012 por sus siglas en inglés, establece las características de DQ que deben tomarse en cuenta al momento de evaluar las propiedades de un conjunto de datos determinado. ISO 25012 define quince características que se listan y detallan en la Figura 2 y la Tabla 1 respectivamente[11].

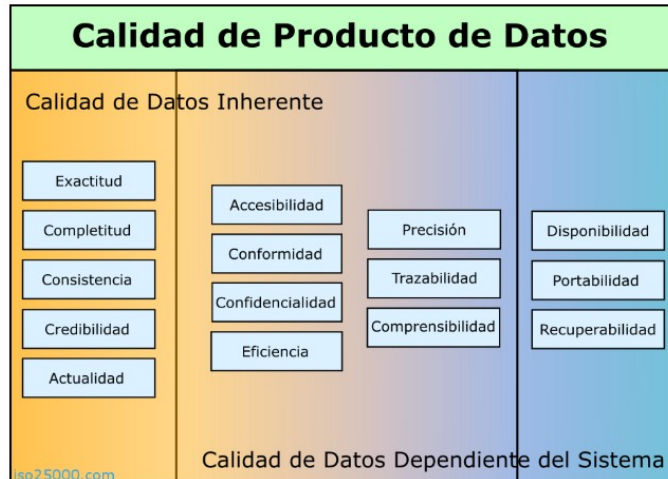


Figura 2 Modelo de calidad de datos ISO25012
Tomado de [11].

CALIDAD DE PRODUCTO DE DATOS		
Calidad de datos inherente	Exactitud	Grado en el que los datos representan correctamente el verdadero valor del atributo deseado de un concepto o evento en un contexto de uso específico.
	Completitud	Grado en el que los datos asociados con una entidad tienen valores para todos los atributos esperados e instancias de entidades relacionadas en un contexto de uso específico.
	Consistencia	Grado en el que los datos están libres de contradicción y son coherentes con otros datos en un contexto de uso específico. Puede ser analizada en datos que se refieran tanto a una como a varias entidades comparables.
	Credibilidad	Grado en el que los datos tienen atributos que se consideran ciertos y creíbles en un contexto de uso específico. La credibilidad incluye el concepto de autenticidad (la veracidad de los orígenes de datos, atribuciones, compromisos).
	Actualidad	Grado en el que los datos tienen atributos con la edad correcta en un contexto de uso específico.
Calidad de datos inherente y dependiente del sistema	Accesibilidad	Grado en el que los datos pueden ser accedidos en un contexto específico, particularmente por personas que necesiten tecnologías de apoyo o una configuración especial por algún tipo de discapacidad.
	Conformidad	Grado en el que los datos tienen atributos que se adhieren a estándares, convenciones o normativas vigentes y reglas similares referentes a la calidad de datos en un contexto de uso específico.
	Confidencialidad	Grado en el que los datos tienen atributos que aseguran que son sólo accedidos e interpretados por usuarios autorizados en un contexto de uso específico. La confidencialidad es un aspecto de la seguridad de la información (junto con la disponibilidad y la integridad) definida como en ISO/IEC 13335-1:2004.
	Eficiencia	Grado en el que los datos tienen atributos que pueden ser procesados y proporcionados con los niveles de rendimiento esperados mediante el uso de cantidades y tipos adecuados de recursos en un contexto de uso específico.
	Precisión	Grado en el que los datos tienen atributos que son exactos o proporcionan discernimiento en un contexto de uso específico.
	Trazabilidad	Grado en el que los datos tienen atributos que proporcionan un camino de acceso auditado a los cambios realizados sobre los datos en un contexto de uso específico.

	Comprensibilidad	Grado en el que los datos tienen atributos que permiten ser leídos e interpretados por los usuarios y son expresados utilizando lenguajes, símbolos y unidades apropiados en un contexto de uso específico. Cierta información sobre la comprensibilidad puede ser expresada mediante metadatos.
Calidad de datos dependiente del sistema	Disponibilidad	Grado en el que los datos tienen atributos que permiten ser obtenidos por usuarios y/o aplicaciones autorizados en un contexto de uso específico.
	Portabilidad	Grado en el que los datos tienen atributos que les permiten ser instalados, reemplazados o eliminados de un sistema a otro, preservando su nivel de calidad en un contexto de uso específico.
	Recuperabilidad	Grado en el que los datos tienen atributos que permiten mantener y preservar un nivel específico de operaciones y de calidad, incluso en caso de fallos, en un contexto de uso específico.

Tabla 1 Modelo de calidad de datos ISO/IEC 25012
Tomado de [11].

2.3 Categorías de datos

Este concepto se refiere a que los datos con características comunes se agrupan formando conjuntos denominados categorías. A la hora de administrar datos, las categorías de datos son útiles, pues la forma de administrarlos depende de su categorización y además, la comprensión de la relación de dependencia entre las categorías ayuda a dirigir los esfuerzos para medir la calidad de datos[2].

Los datos se clasifican en tres categorías que son encontradas en la mayoría de las empresas:

- Datos maestros
- Datos transaccionales
- Datos históricos

Datos Maestros: describen las personas, lugares y cosas que están involucradas en el negocio de una organización. Los ejemplos incluyen personas (clientes), debido a que estos datos tienden a ser utilizados por múltiples procesos de negocios y sistemas de TI, la estandarización de formatos de datos maestros y la sincronización de valores son críticos para la integración exitosa del sistema[2].

Datos transaccionales: describen un evento o transacción interna o externa. Los ejemplos incluyen órdenes de venta, facturas, etc. Estos datos suelen agruparse en registros transaccionales, que incluyen datos maestros de referencia y de referencia asociados[2].

Datos de referencia: son conjuntos de valores o esquemas de clasificación que son referenciados por sistemas, aplicaciones, almacenes de datos, procesos e informes, así como por registros transaccionales y maestros. Los ejemplos incluyen listas de valores válidos, listas de códigos, códigos de estado, abreviaturas de estados, campos demográficos, indicadores, tipos de productos, género, plan de cuentas y jerarquía de productos. Los datos de referencia estandarizados son clave para la integración de datos y la interoperabilidad, adicionalmente facilitan el intercambio de datos entre procesos y sistemas y la composición de informes[2].

Metadatos: significa "datos sobre los datos". Se refiere a describir o caracterizar los datos y facilitar su recuperación, interpretación y su uso adecuado. Los metadatos técnicos son metadatos utilizados para describir la tecnología y la estructura de datos. Ejemplos de metadatos técnicos son nombres de campos, longitud, tipo y diseños de tablas de bases de datos. Se almacenan por separado de los datos que describen[2].

La relación entre categorías de datos se muestra en la Figura 3.

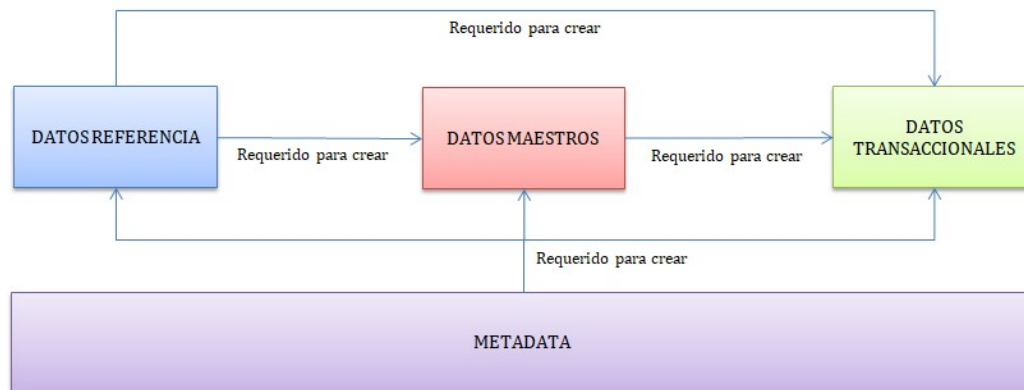


Figura 3 Relación entre categoría de datos
Modificado de [2].

2.4 Sistema de registro fuente

Un sistema de registro fuente es el origen de datos autorizado para un determinado elemento de datos o pieza de información, en otras palabras, es el repositorio de datos considerado como columna vertebral para los procesos de la empresa. El poder de los

sistemas de registro consiste en que son la fuente definitiva y, por lo tanto, se constituyen en el registro de datos críticos.

Los sistemas de registro fuente son el único origen de datos válido referente a un departamento o empresa en particular. Los sistemas de registro se valoran por su capacidad para generar información y conocimiento para el equipo directivo de la empresa.

En la presente tesis, el sistema de registro de datos fuente corresponde al origen de datos de los clientes, que son ingresados por los asesores en el portal y que, posteriormente son almacenados en el datawarehouse.

2.5 Proceso de extracción, transformación y carga (ETL)

El concepto de ETL por las siglas en inglés “Extract, Transform and Load”, se trata de un proceso necesario para la construcción del Datawarehouse DWH por sus siglas. Este proceso es responsable de extraer datos de sistemas fuente y de guardarlos en el almacén de datos. Este proceso permite a las empresas integrar los datos provenientes de varias fuentes, transformarlos y cargarlos en otro repositorio denominado DWH, donde se facilita su análisis, sirviendo de gran soporte para el proceso de negocio[8].

Extracción: Se trata del primer paso de una ETL, consiste en extraer datos de los sistemas origen y hacerlos accesibles para su posterior procesamiento. El objetivo principal es recuperar todos los datos requeridos del sistema fuente con el menor número de recursos posibles; para ello, debe diseñarse de manera que no afecte negativamente al sistema fuente en términos de rendimiento y tiempo de respuesta y que no provoque ningún bloqueo en el procesamiento de transacciones[8].

Transformación: Es el segundo paso del proceso ETL. Aplica un conjunto de reglas de negocio, sobre los datos extraídos en la fase anterior para posteriormente, convertirlos en datos con la estructura común para su procesamiento y análisis[8].

Carga: Este es el último paso de un proceso ETL; recoge los datos transformados en el paso anterior y los carga en el DWH. Este proceso varía en función del fin que persigue la empresa. A fin de agilizar la ejecución de la carga y para garantizar que no se produzcan bloqueos que puedan entorpecerla, es importante que se desactiven las restricciones de

integridad, antes de la realización de la carga y, que se vuelvan a habilitar luego de cumplido este paso[8].

El proceso de extracción, transformación y carga se muestra a continuación en la Figura4.

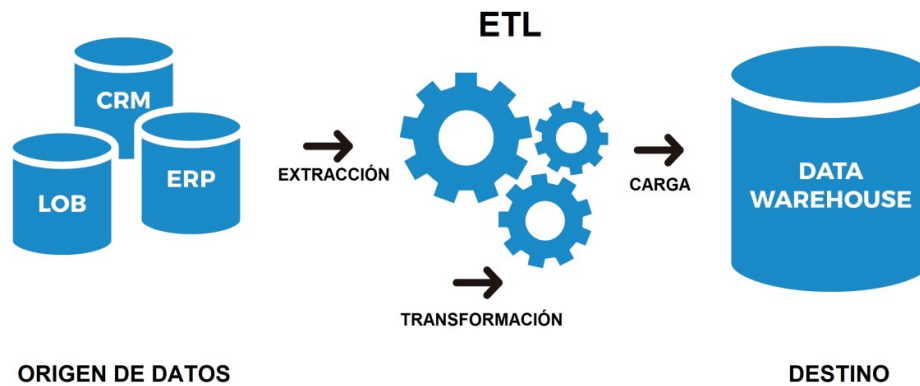


Figura 4Proceso ETL

2.6 Datawarehouse (DWH)

El datawarehouse se ha convertido en una tecnología clave para la integración de fuentes de información distribuidas. Conceptualmente, un DWH es una colección de datos orientados a temas, integrados, no volátiles y variables en el tiempo, organizados de tal forma que facilitan el proceso de la toma de decisiones[14].

Los DWH's están basados en un modelo multidimensional[14]. Este modelo permite una mejor comprensión de los datos y proporciona un mejor rendimiento en consultas complejas. Los datos, en los modelos multidimensionales, se presentan en un espacio n-dimensional, generalmente llamado cubo de datos o hipercubo. Un ejemplo de cubo de datos se muestra en la Figura 5.

Un cubo de datos está definido por dimensiones y hechos. Las dimensiones representan a las distintas perspectivas que se utilizan para analizar los datos. Por ejemplo, el cubo de datos de la Figura 5, se utiliza para analizar ventas y tiene tres dimensiones: tienda, tiempo y producto.

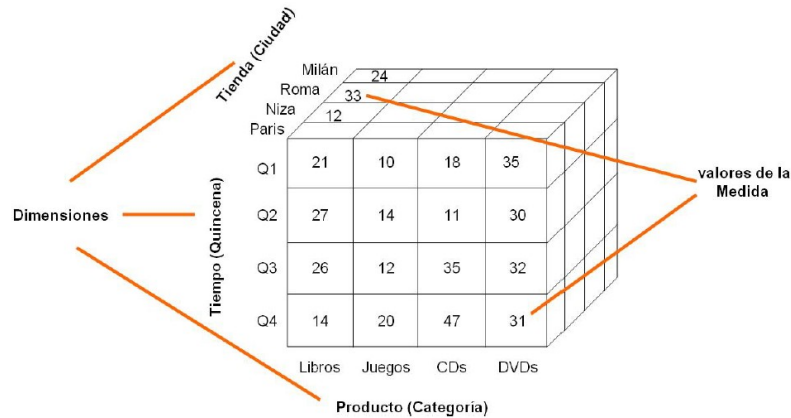


Figura 5 Cubo tridimensional de los datos de venta
Tomado de [9].

2.7 Framework

El término framework o marco de trabajo se emplea en muchos ámbitos de desarrollo de aplicaciones como por ejemplo: [15]

1. Calidad de datos
2. Aplicaciones médicas
3. Desarrollo de juegos
4. Virtualización de escenarios
5. Interfaces de usuario
6. Desarrollo de aplicaciones web, escritorio, móviles, etc.

Un framework de calidad de datos es una estructura conceptual y tecnológica compuesta por fases que facilitan y garantizan la implementación de: información de calidad, desarrollo ágil, seguro y escalable, etc. [16].

Revisión sistemática de la literatura

Se realiza una revisión sistemática de la literatura (SLR) que consiste en un proceso que requiere la exploración de la literatura disponible, que cumpla con los criterios específicos del presente trabajo. Se considera que los resultados obtenidos a partir de la SLR proporcionarán una visión adecuada de la DQ.

Cadena de búsqueda

Para obtener la información necesaria para el diseño e implementación del framework de calidad de datos, se plantea la siguiente pregunta de investigación: ¿Un framework de

calidad de datos en una entidad financiera contribuye a una toma de decisiones adecuada?

En la Tabla 2 se muestra los criterios utilizados para la selección de la revisión literaria.

ASPECTOS	CRITERIO
Población	Financial Entity
Intervención	Framework data quality
	Framework data quality management
	Data quality management
	Data quality
Resultados	Información de calidad
	Adecuada toma de decisiones
Contexto	Academia
	Universidades
Criterios de inclusión	Ciclo de vida de los datos
	Gobierno de datos
Criterios de exclusión	Big Data
	Minería de datos
Criterios de selección	Escritos en inglés
	Haber obtenido resultados
	Presente fases y dimensiones definidas
	Presente ejemplos
	Ser único
	Desde 2008

Tabla 2 Criterios para realizar la SLR

El proceso de selección del estudio con los criterios definidos en la anterior tabla, se describe a continuación:

- Se realiza búsquedas para identificar estudios potencialmente relevantes de las fuentes seleccionadas a partir del año 2008, sobre la base de títulos de lectura.
- Se realiza un muestreo en base a los títulos obtenidos en el ítem anterior y se determina nuevos términos de exclusión no considerados inicialmente.
- Con los resultados obtenidos se realiza la lectura del resumen y de las conclusiones, para establecer si aporta a la investigación y, con ello, se determina la relevancia de los trabajos.
- Se obtiene la lista final de artículos relevantes y se realiza la lectura completa de los documentos, para extraer la información requerida para la investigación.

Luego del proceso mencionado, se selecciona seis trabajos de gran relevancia en lo que referente a la revisión de la literatura de Frameworks de DQ. A continuación, se realiza una breve descripción de los trabajos identificados:

Data quality framework: El framework tiene por objetivo proporcionar una solución a corto y mediano plazo mediante la identificación de iniciativas específicas para posteriormente abordar las prioritarias. El ciclo de vida de los datos es considerado como componente central que consiste en seis etapas[17].

The Development of a Data Quality Framework and Strategy for the New Zealand Ministry of Health: Herramienta de evaluación y medición, integrada en el proceso de organización, que proporciona un punto de referencia para la eficacia de las futuras iniciativas de mejora de calidad de datos y una plantilla estandarizada para obtener información sobre la calidad a partir de los datos[18].

Executing Data Quality Projects: Ten steps to Quality Data and Trusted Information: Proporciona un enfoque sistemático para mejorar la calidad de los datos y producir información de calidad en el contexto de una empresa. Describe una metodología que combina un marco conceptual para comprender la calidad de los datos, junto con las técnicas, herramientas e instrucciones para mejorar y crear información de calidad. El framework para la calidad de la información es el marco conceptual y diez son los pasos del proceso para implementar sus conceptos[2].

Data Quality Management with Semantic: Describe una forma innovadora sobre cómo administrar la calidad de datos, mediante la representación de conocimiento y el uso de tecnologías, mediante un framework de gestión de calidad de datos semánticos [19].

Data Quality Framework: Framework sobre el que se presenta de un amplio material que explica la pertinencia de los datos, el uso de métodos y de fuentes de información, útiles para los usuarios que requieran la comprensión, representación y saber cómo se construyen los datos[20].

BayesWipe A Scalable Probabilistic Framework for Improving Data Quality: Propone un método para corregir valores de atributos individuales en una base de datos usando un modelo generador Bayesiano y un modelo de error estadístico aprendido directamente de la base de datos ruidosa. Por lo tanto, evitar la necesidad de un experto en el dominio o en limpieza de datos maestros [18].

DAMA-DMBOK: La guía proporciona una introducción a la gestión de datos. Define una visión estándar de la industria de las funciones de gestión de datos, la terminología y las mejores prácticas, sin detallar métodos y técnicas específicos [21].

Finalmente, para el diseño del framework que será utilizado en este trabajo de titulación se realiza un cuadro comparativo que se muestra en la Figura 3 con los frameworks seleccionados a partir de la SLR.

FRAMEWORK		FÁCIL DE APLICAR	APLICABLE VARIAS INDUSTRIAS	PRESENTA FASES DEFINIDAS	DIMENSIONES DEFINIDAS	PRESENTA EJEMPLOS	RESULTADOS	CASO DE ESTUDIO	PRESENTA PLANTILLAS
1	DATA QUALITY FRAMEWORK	PAPER	●	●	●	●	●	○	●
2	THE DEVELOPMENT OF A DATA QUALITY FRAMEWORK AND STRATEGY FOR THE NEW ZEALAND MINISTRY OF HEALTH	PAPER	●	●	●	●	●	●	
3	EXECUTING DATA QUALITY PROJECTS: TEN STEPS TO QUALITY DATA AND TRUSTED INFORMATION	LIBRO	●	●	●	●	●	○	●
4	DATA QUALITY MANAGEMENT WITH SEMANTIC	TESIS	●	●	●	●	●	●	●
5	DATA QUALITY FRAMEWORK	PAPER	○	●	●	●	●	●	
6	BAYESWIPE: A SCALABLE PROBABILISTIC FRAMEWORK FOR IMPROVING DATA QUALIT	LIBRO	●	●	●	●	●	○	○
7	DAMA-DMBOK	LIBRO	●	●	●	●	●	○	○

●	PRESENTA
	NO PRESENTE
○	POR VERIFICAR

Tabla 3 Cuadro comparativo FDQ

En la Tabla 4 y Figura 6 se identifica las tareas y fases del framework de calidad de datos.

FRAMEWORK DATA QUALITY		Requisitos de la evaluación			Especificar y diseñar la evaluación							Ejecutar la evaluación		Concluir la evaluación	
		Identificar stakeholders y sus necesidades	Identificar inconvenientes data quality	Identificar negocio	Identificar requisitos relevantes	Recopilación de datos	Comprender datos y especificidades relevantes	Comprender la tecnología, procesos, personas relevantes	Definir ciclo de vida de la información	Definir los atributos de calidad de datos	Plan de evaluación y captura de datos de diseño	Evaluar los niveles de calidad de los datos existentes	Análisis resultados de evaluación	Recomendaciones basadas en los resultados de la evaluación	
1	DATA QUALITY FRAMEWORK	●	●	●		●	●	●	●	●	●	●	●	●	●
2	THE DEVELOPMENT OF A DATA QUALITY FRAMEWORK AND STRATEGY FOR THE NEW ZEALAND MINISTRY OF HEALTH			●	●	●		●	●	●	●	●	●	●	●
3	EXECUTING DATA QUALITY PROJECTS: TEN STEPS TO QUALITY DATA AND TRUSTED INFORMATION	●	●	●	●	●	●	●	●	●	●	●	●	●	●
4	DATA QUALITY MANAGEMENT WITH SEMANTIC TECHNOLOGIES			●		●	●	●		●		●	●	●	●
5	DATA QUALITY FRAMEWORK	●			●	●			●	●	●	●	●	●	●
6	BAYESWIPE: A SCALABLE PROBABILISTIC FRAMEWORK FOR IMPROVING DATA QUALITY		●		●				●		●		●	●	
7	DAMA-DMBOK	●		●		●			●	●	●		●		

Tabla 4 Identificación de fases y tareas FDQ

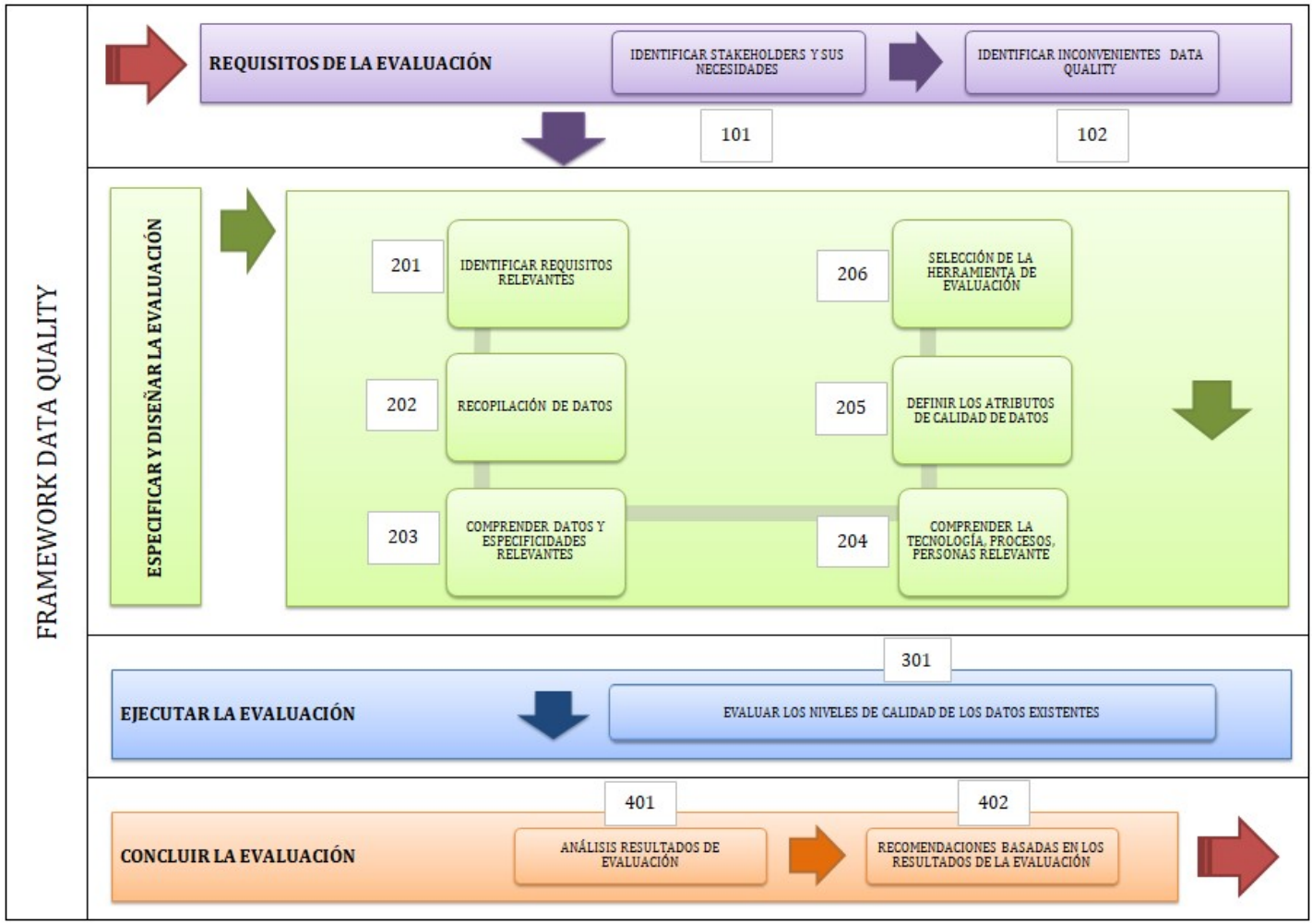


Figura 6 Framework de Calidad de Datos

3 METODOS Y HERRAMIENTAS

En este capítulo se describe el framework diseñado y su posterior aplicación en la calidad de datos de un Datawarehouse en el sistema de registro fuente de una entidad financiera.

3.1 Descripción del framework

Describe cómo implementar conceptos y tareas para mitigar problemas de calidad de datos.

En la Tabla 5 se muestra las fases y tareas del framework de calidad de datos.

ID	Fase	ID	Tarea
100	Requisitos de la evaluación	101	Identificar stakeholders y sus necesidades.
		102	Identificar inconvenientes data quality.
200	Especificar y diseñar la evaluación	201	Identificar requisitos relevantes.
		202	Recopilación de datos.
		203	Comprender datos y especificidades relevantes.
		204	Comprender la tecnología, procesos, personas relevantes.
		206	Definir los atributos de calidad de datos.
		207	Selección de la herramienta de evaluación.
300	Ejecutar la evaluación	301	Evaluar los niveles de calidad de los datos existentes.
400	Concluir la evaluación	401	Análisis resultados de evaluación.
		402	Recomendaciones basadas en los resultados de la evaluación.

Tabla 5 Fases y tareas del FDQ

3.2 Identificación de los stakeholders y sus necesidades

Los sistemas se construyen con el objetivo de satisfacer necesidades de los stakeholders (Figura 7). Los stakeholders también son la fuente clave de requerimientos. Si no se toman en cuenta los requerimientos de todos los stakeholders de un sistema pone en peligro la completitud del mismo[22].

Stakeholders		Macro		Micro	
		Internos	Externos	Internos	Externos
Stakeholders	Internos	Inversionistas	Quiero realizar inversión segura y rentable.	Quiero un retorno de la inversión asegurada.	Quiero realizar inversiones que tengan algún beneficio, aparte de lo económico.
				Quiero recibir mis dividendos de manera oportuna.	Quiero saber dónde se ejecutó mi inversión.
				Quiero anticiparse a lo que quiere el cliente.	Quiero ofertar productos, servicios con valor agregado.
		Gerentes	Quiero una gama de servicios de alta calidad para ofertar.	Quiero personal calificado dentro de la organización.	Quiero tomar las mejores decisiones basadas en el análisis de datos.
				Quiero tomar buenas decisiones.	Quiero mitigar la morosidad de la cartera.
				Quiero recuperación de la cartera.	
	Externos	Empleados	Quiero un buen ambiente laboral.	Quiero estabilidad laboral.	Quiero reconocimiento por mi desempeño.
				Quiero crecimiento profesional.	Quiero que mi esfuerzo académico sea tomado en cuenta.
				Quiero que se tomen buenas decisiones en PRO de los empleados.	
		Clientes	Quiero atención al cliente de calidad.	Quiero que conozcan mis necesidades.	Quiero amplios canales de comunicación.
				Quiero horarios extendidos en fechas específicas.	Quiero mayor seguridad (mecanismos de seguridad, cajeros zonas seguras).
				Quiero alternativas de transacciones donde se utilice la tecnología.	Quiero que mi información sea manejada con carácter privado, con toda la protección y organización necesaria.
Competencia	Quiero captar el mayor mercado bancario del Ecuador.	Quiero ofertar soluciones informáticas novedosas.	Quiero anticipar los negocios de mi competencia.		
		Quiero saber que está haciendo mi competencia y mejorarlo.			
		Quiero velar por la seguridad, estabilidad, transparencia y solidez de los sistemas financieros.	Quiero mediante un eficiente y eficaz proceso de regulación y supervisión proteger los intereses del público y contribuir al fortalecimiento del sistema económico social.		
Organismo de control	Quiero velar por el sistema bancario.				

Figura 7 Stakeholders y sus necesidades

3.3 Identificación inconvenientes de DQ

Para la identificación de los inconvenientes de calidad de datos se ha diseñado un cuestionario con preguntas de selección simple, de cuya aplicación, se ha obtenido el resultado que se muestra en la Tabla 6.

INCONVENIENTE	DESCRIPCIÓN
Información duplicada o faltante	Sistemas diversos con diferentes estructuras, requisitos y agregaciones. A la hora de integrar estos datos, los analistas se encuentran con campos duplicados o faltantes, o bien una falta de coherencia entre las etiquetas
Los datos no son precisos	Los datos no se ajustan a su valor real (errores tipográficos)
La información no es exacta	Los clientes no se encuentran en las direcciones registradas.

Insuficiente documentación	La documentación relativa, la estructura de la BD, tablas, relaciones, atributos, tipos, restricciones, etc., gráfica o contextual, no existe o esta desactualizada.
----------------------------	--

Tabla 6 Inconvenientes de calidad de datos

3.4 Identificación requisitos relevantes

En la actualidad, la estructura a nivel operacional de cómo se capta la información, hasta servir de insumo de toma de decisiones, se presenta de la siguiente manera:

- ✓ **Recepción:** Corresponde a las operaciones de captación de datos de los clientes por diferentes medios o canales.
- ✓ **Business Intelligence:** Corresponde al conjunto de metodologías, aplicaciones y tecnologías que se utilizan para reunir, depurar y transformar los datos en información.
- ✓ **Decisiones:** Provee el mecanismo para la interacción de usuarios finales o sistemas externos con las aplicaciones analíticas, de información y de reportes, que posibilitan el proceso de toma de decisiones en los negocios.

En la Figura 8 se presenta la estructura de los requisitos relevantes a nivel operacional dividido en capas.

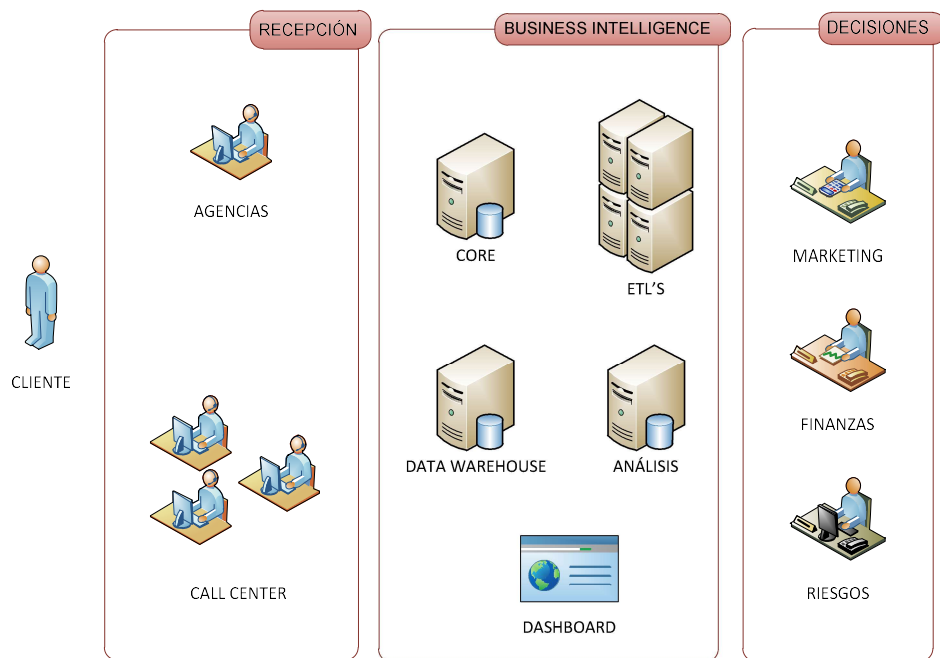


Figura 8 Requisitos relevantes

3.5 Recopilación de datos

La recopilación de datos se refiere al uso de diversas técnicas y herramientas que pueden ser utilizadas por el analista de información, concretamente se trata de la entrevistas, la encuesta, el cuestionario, la observación, el diagrama de flujo y el diccionario de datos. Todos estos instrumentos se aplicarán en un momento en particular, con la finalidad de buscar información que será útil a una investigación en común.

En el presente trabajo se selecciona la técnica del diccionario de datos, el mismo que es considerado como "datos acerca de los datos"; es decir, descripciones de todos los objetos (archivos, programas, informes, sinónimos...) existentes en el sistema. Un diccionario de datos almacena la totalidad de los diversos esquemas y especificaciones de archivos, así como sus ubicaciones. Incluye además, información acerca de qué programas utilizan qué datos, y qué usuarios están interesados en unos u otros informes.

Mediante el diccionario de datos proporcionado se identifica la estructura de la base de datos, lo que incluye sus relaciones que determinan como se almacena los datos, así como el acceso a los mismos.

Más adelante, en el presente trabajo, se mostrarán estructuras de tablas y, por razones de confidencialidad, se nombrará las tablas con nombres ficticios.

Tanto en Figura 9 como en la Tabla 7 se presenta el modelo de la base de datos del sistema de registro fuente, así como el detalle de cada uno respectivamente.

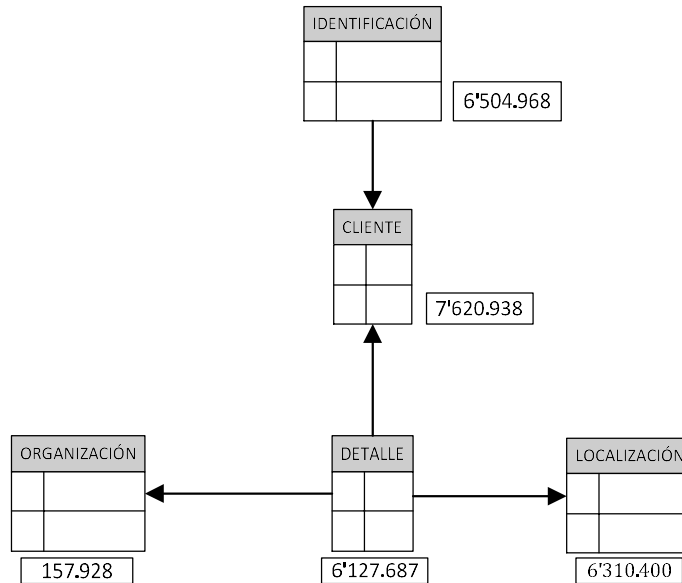


Figura 9 Modelo del BDWH SSoR

BD	Tabla	Campos	Registros
BDWH SSoR	Cliente	17	7620938
	Identificación	8	6504968
	Detalle	32	6127687
	Organización	21	157928
	Localización	10	6310400
Totales	5	88	26721921

Tabla 7Detalle de la recopilación de la información

3.6 Comprender los datos y sus especificidades relevantes

A continuación, se detalla el esquema arquitectónico de los datos de propiedad de la entidad financiera.

FUENTE DE DATOS

- ✓ Esta capa se refiere a las fuentes de información que serán utilizadas para poblar los diferentes repositorios.
- ✓ Las fuentes se encuentran en tecnologías distintas, y se deben establecer estándares.

INTEGRACIÓN

- ✓ Esta capa se refiere a los procesos y ambientes que tratan con las tareas de captura, clasificación, procesamiento y movimiento de datos para preparar la información para su almacenamiento en la capa de repositorios.
- ✓ Los procesos de carga y publicación pueden incluir diferentes estrategias: archivos delta, archivos ftp, servicio Web, conexión directa a bases de datos, etc.

REPOSITORIO

- ✓ Esta capa representa la visión lógica y el repositorio físico de los datos almacenados.
- ✓ Requiere del aporte de arquitectos de datos y administradores de bases de datos, que aseguren su implementación usando las mejores prácticas.

ANALÍTICA

- ✓ Esta capa utiliza datos y modelos que apoyan el proceso de decisión y se constituyen la base para la toma de mejores decisiones.
- ✓ Habilita la construcción de mejores aplicaciones y a los usuarios en la toma de decisiones efectivas al disponer de información completa para su análisis.
- ✓ Esta capa está conformada por varios componentes tecnológicos para cubrir distintas necesidades como modelos matemáticos avanzados, simulaciones, Data Marts, cubos, indicadores y técnicas de optimización que mejoran la eficiencia operacional.

ACCESO

- ✓ Provee el mecanismo para la interacción de usuarios finales o sistemas externos con las aplicaciones analíticas, de información y de reportes.
- ✓ La interacción puede ser humana o automatizada y puede ser directa o remota a través de un servicio web.
- ✓ La capa de acceso debe estar orientada a aplicativos propios de análisis, sobre un web browser y además ofrecer integración con otros aplicativos o visores de información vía Web Services.

En la Figura 10 se muestra la arquitectura de datos que presenta la entidad financiera.

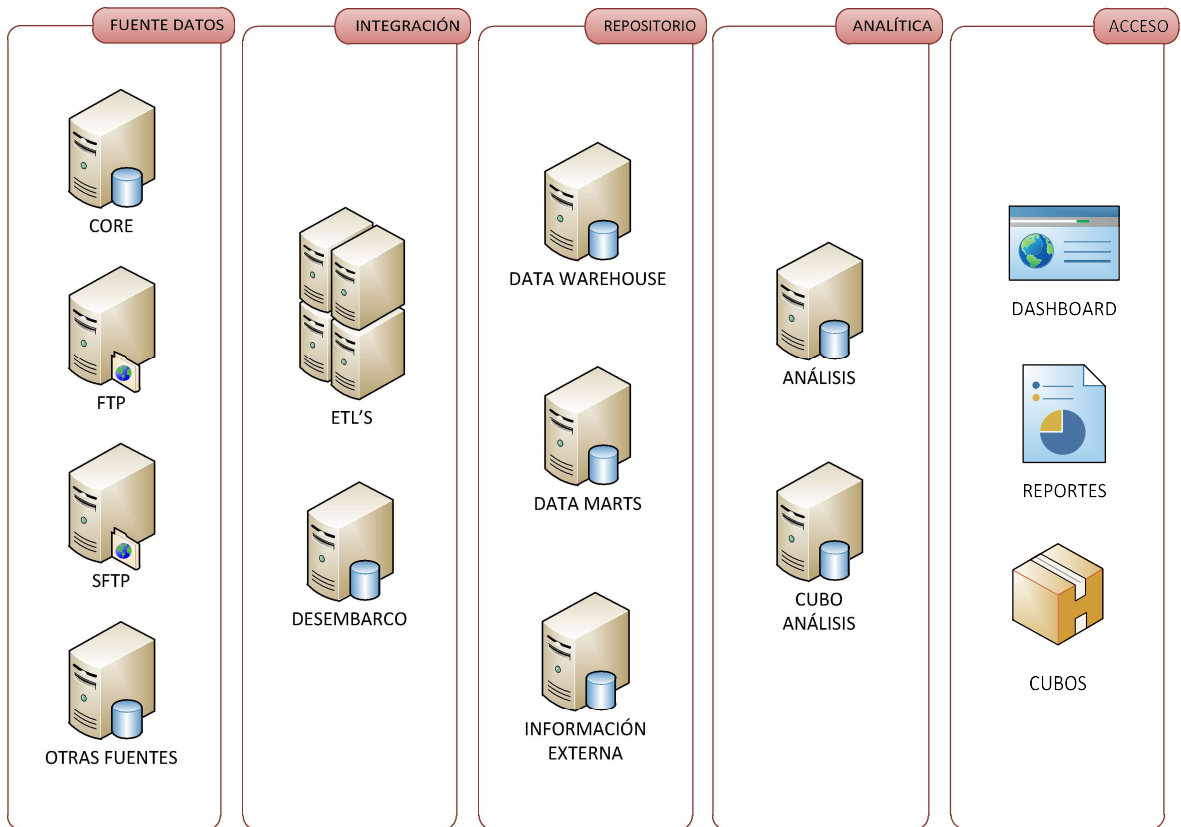


Figura 10Arquitectura de datos

3.7 Tecnología asociada con los datos

La tecnología asociada para la administración y almacenamiento de los datos se muestra en laFigura 11.

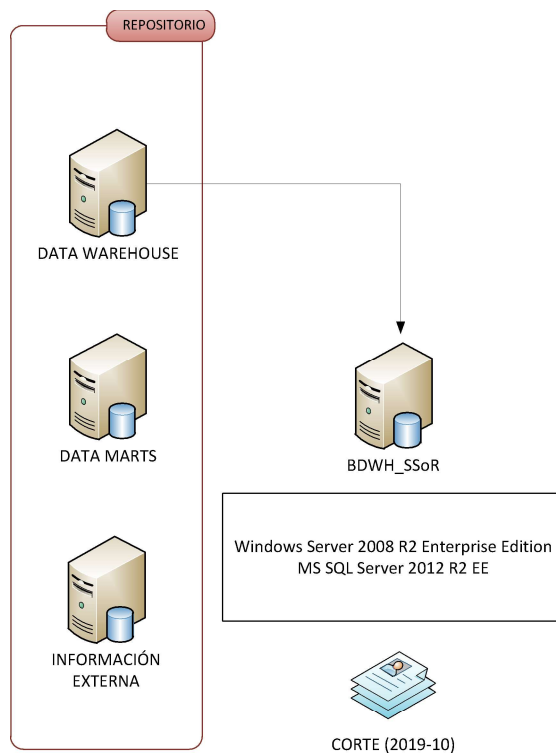


Figura 11Detalle de la tecnología asociada

3.8 Definición de los atributos de calidad de datos:

Para evaluar una dimensión de calidad existen varias métricas, pero la determinación de cuál es la más adecuada, depende del contexto de análisis y los datos existentes en el datawarehouse. En la Tabla 8 y Tabla 9 se muestra las cuatro dimensiones identificadas, así como su factor y tipo de error.

Dimensión	Descripción
Exactitud	Indica que tan precisos, válidos y libres de problemas están los datos. Establece si existe una correcta y precisa asociación entre los estados del sistema de información y los objetos del mundo real.
Compleitud	Indica en qué medida el sistema de información contiene todos los datos de interés, y si los mismos cuentan con el alcance y profundidad requeridos.
Consistencia	Se refiere al cumplimiento de las reglas semánticas que son definidas sobre los datos.
Unicidad	Hace referencia al nivel de redundancia de los datos. La redundancia ocurre cuando un objeto del mundo real se encuentra representado más de una vez en los datos, esto es, varias tuplas, posiblemente diferentes representan el mismo objeto.

Tabla 8Dimensiones identificadas

Dimensión	Factor	Tipo de error
Exactitud	Exactitud sintáctica	Valor fuera de rango.
		Estandarización.
	Exactitud semántica	Registro inexistente.
		Defecto mal registrado.
Completitud	Densidad	Valor fuera de referencial.
		Valor nulo.
Consistencia	Integridad intra-relación	Clasificación de defecto.
		Reglas de integridad intra-relación.
	Integridad referencial	Valor único.
Unicidad	Duplicación	Referencia inválida.
	Contradicción	Registro duplicado.
		Registro contradictorio.

Tabla 9Factor y tipo de error

3.9 Selección de la herramienta de evaluación

En el mercado existen varias herramientas tecnológicas que proveen funcionalidades para la evaluación de la calidad de datos, las mismas que se basan en modelos que manejan diferentes conceptos de calidad, en varios niveles de abstracción.

DataCleaner: Es una herramienta open source de escritorio desarrollada en Java disponible en GitHub con flexibilidad de añadir nuevos componentes. Existe documentación de la aplicación y un manual en formato web, permite conectarse a diferentes orígenes de datos, sin embargo no describe todos los parámetros de configuración. Su comunidad no es muy grande, razón por la cual no es una herramienta de uso masivo [23].

Data Quality Services: Herramienta de escritorio disponible en Microsoft SQL Server desde su edición 2012. Permite detectar duplicados, limpiar, validar, y unir datos utilizando exclusivamente SQL Server [23].

Open Studio for Data Quality de Talend: Aplicación de escritorio open source con licencia apache, soporte conexión a diferentes orígenes de base de datos, permitiendo realizar data profiling y evaluar la calidad de datos. Los análisis retornan las medidas con diferentes granularidades, dependiendo del método utilizado la granularidad puede ser un atributo, un conjunto de atributos o una tabla [23].

Oracle Enterprise Data Quality: herramienta open source donde los análisis ya están predefinidos, el usuario selecciona sobre qué tabla y atributos se aplicará y genera una medida de calidad. La granularidad es predefinida para cada análisis, en algunos casos

retorna el resultado en varios niveles de granularidad y se muestran los registros que cumplen o no las condiciones [23].

La descripción y resumen de las herramientas comerciales analizadas son mostradas a continuación en la Tabla 10 y Tabla 11 respectivamente.

HERRAMIENTA	DIMENSIÓN	FACTOR	GRANULARIDAD
Datacleaner	Exactitud	Correctitud sintáctica	Celda
		Correctitud semántica	Conjunto de celdas
		Precisión	Celda
	Consistencia	Integridad de dominio	Conjunto de celdas
		Intra-relación	Celda
		Inter-relación	Celda
	Compleitud	Densidad	Conjunto de celdas
Unicidad	No-duplicación	Conjunto de celdas	
Frescura	Actualidad	Celda	
Data quality services	Exactitud	Correctitud sintáctica	Celda
	Unicidad	No-duplicación	Conjunto de celdas
Open studio for data quality de talend	Exactitud	Precisión	Conjunto de atributos y celdas
		Correctitud sintáctica	Conjunto de atributos y celdas
	Compleitud	Cobertura	Conjunto de atributos y celdas
		Densidad	Conjunto de atributos y celdas
	Consistencia	Integridad de dominio	Atributo - Celda
		Intra-relación	Conjunto de atributos y celdas
		Intra-relación	Conjunto de atributos y celdas
	Unicidad	No-duplicación	Conjunto de atributos y celdas
No-contradicción		Conjunto de atributos y celdas	
Frescura	Actualidad	Conjunto de atributos y celdas	
Oracle enterprise data quality	Exactitud	Precisión	Celda
		Correctitud sintáctica	Conjunto de atributos y celdas
		Correctitud semántica	Conjunto de atributos y celdas
	Compleitud	Densidad	Conjunto de atributos y celdas
	Consistencia	Integridad de dominio	Conjunto de atributos y celdas
		Intra-relación	Conjunto de atributos y celdas
	Unicidad	No-duplicación	Conjunto de atributos y celdas
		No-contradicción	Conjunto de atributos y celdas
Frescura	Actualidad	Conjunto de celdas	

Tabla 10 Herramientas comerciales para la evaluación de la calidad de datos
Tomado de [23].

	Open source	Documentos	Ejemplos	Soporte	Conexión varias base de datos	Sencillo
Datacleaner	●	○	○		●	●
Data quality		○	○			
Data quality talend	●	●	●	○	●	●
Oracle enterprise data quality	●	●	●	○	●	○

Tabla 11 Resumen de herramientas para la evaluación de calidad de datos

3.10 Definir mecanismos y herramientas de reporte

En el mercado existen varias herramientas tecnológicas que proveen la funcionalidad de análisis y visualización de datos, de las cuales se compararán tres líderes del mercado que se muestran en la Tabla 12.

	POWER BI	TABLEAU DESKTOP	QLIK VIEW
Versión gratuita disponible	●		●
Versiones móviles	●	●	●
Analítica en tiempo real	●	●	●
Analítica predictiva	●	●	
Herramientas de preparación de datos	●	●	●
Herramientas para combinar / unir / integrar datos	●	●	●
Consulta semántica / lenguaje natural	●		
Característica de visualizaciones	●	●	●
Herramienta para compartir / colaborar	●	●	●

Tabla 12 Comparación de herramientas de análisis de datos.

Partiendo de la comparación se selecciona al producto Power BI como herramienta de reporte y análisis de datos que estará alienado al esquema de gobernanza de datos que definirá la periodicidad, prioridad y la toma de decisiones.

4 RESULTADOS Y DISCUSIÓN

De acuerdo con los resultados obtenidos en la evaluación se procede a la obtención del análisis, desagregado por dimensiones y factores descritos en la Tabla 12.

Dimensión	Factor
Exactitud	Correctitud sintáctica
	Precisión
Compleitud	Densidad
Consistencia	Integridad - dominio
	Integridad inter-relación
Unicidad	No duplicación
	No contradicción

Tabla 13 Desagregación de la evaluación de calidad de datos

4.1 Resultados

El sistema que va a ser evaluado se encuentra conformado por cinco tablas, ochenta y ocho campos y más de veinte y seis millones de registros del datawarehouse del sistema de registro fuente (SSoR) de la entidad financiera.

Muchas de las organizaciones gastan enormes cantidades de dinero en campañas de marketing con el objetivo de promocionar sus productos, algunas de ellas incluso generan catálogos muy costosos para llegar a sus clientes dejando a un lado el activo más importante que es la calidad de la información.

Los datos de calidad son elementos vitales de una organización que constituye la base de muchas decisiones empresariales, mucho más si son de carácter financiero donde su línea de negocio gira en base a la toma de decisiones.

Para la parametrización de los rangos cualitativos se propone la valoración que se muestra en la Tabla 13, dichos valores se encuentran en función de los resultados obtenidos del análisis, es decir el resultado más bajo del análisis es considerado como deficiente y el resultado más alto como altamente satisfactorio.

Valoración		
Satisfacción	Rango	
Altamente satisfactorio	9,5	10
Satisfactorio	9	9,4
Poco satisfactorio	8,5	8,9
Deficiente	8	8,4

Tabla 14 Cualitativa de medición

4.1.1 Resultados de la evaluación de calidad

Para la evaluación de calidad de datos se emplea el análisis de información contemplada en el datawarehouse del sistema de registro fuente, así como de las cuatro dimensiones identificadas y descrito en la Tabla 14.

Dimensión	Factor	Evaluación
Exactitud	Correctitud sintáctica	No brindan el suficiente detalle.
		Caracteres especiales.
		Errores de estandarización.
		Valores de múltiples atributos.
	Precisión	No brindan el suficiente detalle.
		No presenta formato.
Complejidad	Densidad	No brindan el suficiente detalle.
		Nulo.
		Blanco.
Consistencia	Integridad - dominio	Regla fuera de dominio.
	Integridad inter-relación	Integridad referencial.
Unicidad	No duplicación	Información duplicada.
	No contradicción	Información contradictoria.

Tabla 15 Dimensiones de calidad

A partir de la identificación de las dimensiones y diferentes factores se presenta la matriz implementada para realizar la evaluación, descrito su estructura en la Tabla 16 y su respectiva implementación en el Anexo VI Matriz de calidad de datos.

CELDA	DESCRIPCIÓN
Dimensión	Descripción de la dimensión evaluada.
Factor	Descripción del factor evaluada.
Tablaalias	Alias de la tabla evaluada.
Columnaalias	Alias de la columna evaluada.
Ejemplo descripción	Ejemplos de la evaluación.
Registros	Número de registros evaluados.
Valor obtenido	Valor obtenido de la evaluación.
Calidad de datos (/10)	Calidad obtenida de las diferentes dimensiones.

Tabla 16 Estructura de la matriz de calidad de datos

La calidad de datos se encuentra en función de la información que no cumple con el factor evaluado sobre el total de información establecida, por ello el valor que se obtiene es el resultado de la siguiente fórmula, descrito en la Tabla 17.

FÓRMULA
$X = A / B$
A = Registros que no cumple con la métrica
B = Registros totales
Dónde $B > 0$
EVALUACIÓN
El más cercano a "0" es lo mejor.

Tabla 17 Fórmula para calcular la calidad de datos

En la Figura 12 muestra gráficamente la desagregación de los resultados cuantitativos obtenidos para cada una de las dimensiones y factores identificados.

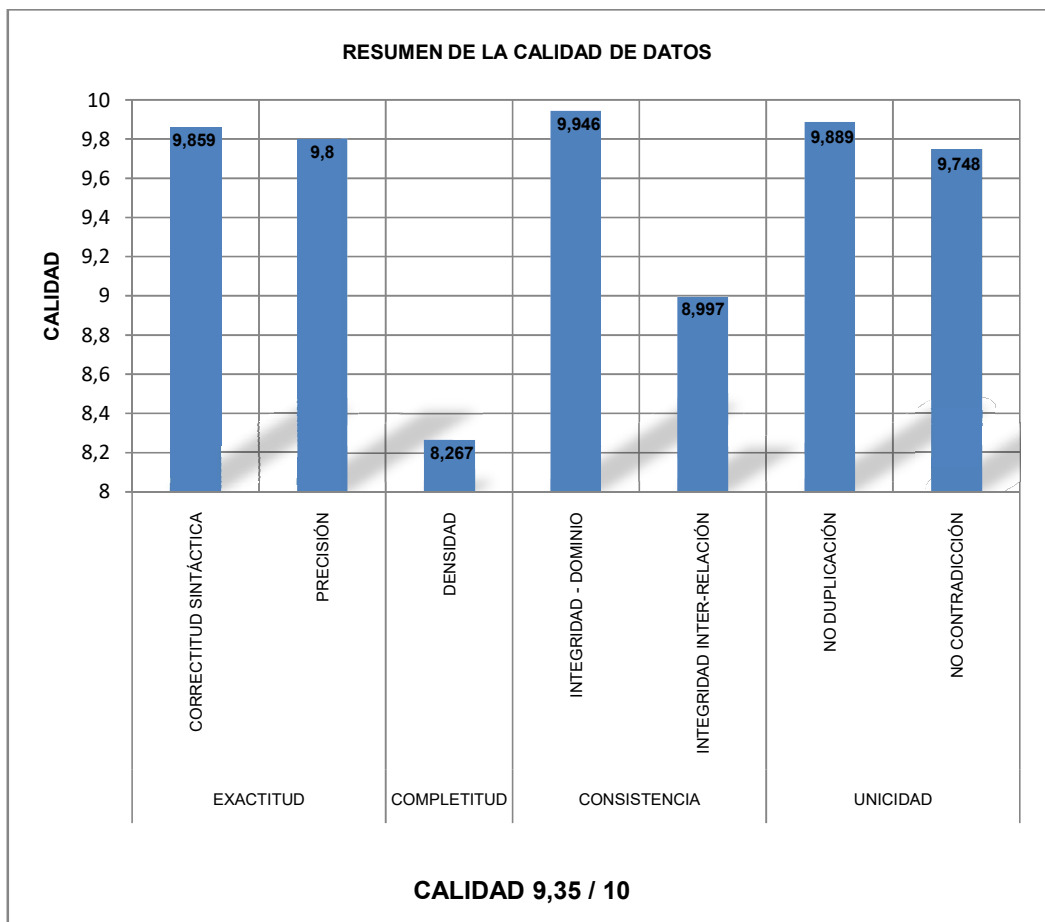


Figura 12 Resultado de la evaluación de calidad de datos

En la Figura 12 se puede determinar el valor promedio de 9,35 que corresponde a la valoración de calidad general, que la ubica en el rango satisfactorio descrito en la Tabla 14.

4.1.2 Resultados de la evaluación exactitud, factor correctitud sintáctica

En la Tabla 18 se presenta los resultados cuantitativos para la dimensión exactitud factor correctitud sintáctica que corresponden a las tablas: detalle, cliente, localización, organización e identificación. Para su cálculo se emplea la función promedio que devuelve la media de una serie de valores, para finalmente generar la medida de calidad que se describe en la Tabla 17.

Tabla alias	Registros	Columna alias	A	X	Calidad / 10
Detalle	6127687	Fecha nacimiento	111502	0,01819643	9,818
		Primer nombre	2081	0,00033961	9,997
		Segundo nombre	3339	0,0005449	9,995
		Primer apellido	3309	0,00054001	9,995
		Segundo apellido	2933	0,00047865	9,995
		Sexo	262185	0,04278694	9,572
Cliente	7620938	Nombre	85474	0,01121568	9,888
Localización	6310400	Nombre	410	6,4972e-05	9,999
		Descripción	379202	0,06009159	9,399
Organización	157928	Nombre	3324	0,02104757	9,79
Identificación	6504968	Identificación	331	5,0884e-05	9,999

Tabla 18 Evaluación de calidad: dimensión exactitud, factor correctitud sintáctica

En la Tabla 19 se presenta varios ejemplos de la información contemplada para el análisis, donde se puede evidenciar que existe información con caracteres especiales; por ejemplo para la tabla detalle el primer nombre registrado es: “_MARGARITA”.

Columna alias	Descripción ejemplo	
Fecha nacimiento	No brindan el suficiente detalle	Fechas superiores a los 90 años
Primer nombre	Caracteres especiales	_MARGARITA
Segundo nombre	Caracteres especiales	JUDITH
Primer apellido	Caracteres especiales	BLANCO.TOVAR
Segundo apellido	Caracteres especiales	CORINNE.**
Sexo	Errores de estandarización	-999
Nombre	Valores que corresponden a múltiples atributos	CASTA?EDA CACHI DELIA GUISELA - YARUQUI
Nombre	Caracteres especiales	AGUARICO**
Descripción	Caracteres especiales	346541/888367/
Nombre	Caracteres especiales	S.T.A.O.A.E.P.A.P.-G
Identificación	Caracteres especiales	00000.2625

Tabla 19 Ejemplos descripción: dimensión exactitud, factor correctitud sintáctica

En la Tabla 18 y Figura 13 se puede determinar el valor promedio de 9,85 que corresponde a la valoración de calidad que la ubica en el rango altamente satisfactorio descrito en la Tabla 14.

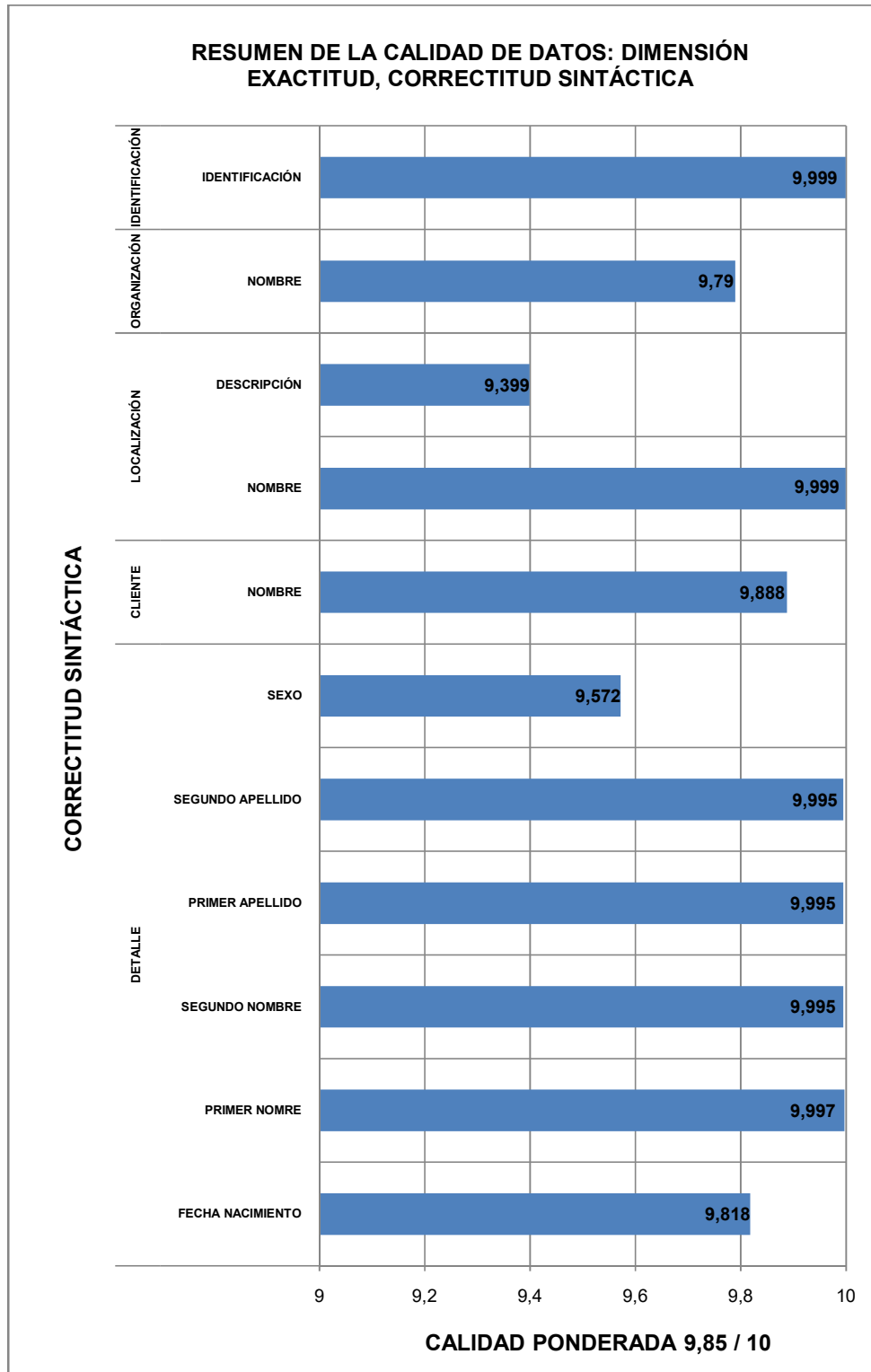


Figura 13 Resultado de la evaluación exactitud, factor correctitud sintáctica

4.1.3 Resultados de la evaluación exactitud, factor precisión

En la Tabla 20 se presenta los resultados cuantitativos para la dimensión exactitud factor precisión que corresponden a las tablas: detalle, cliente, localización, organización e identificación, para el cálculo se emplea la función promedio que devuelve la media de una serie de valores, generando un promedio de calidad descrito en la Tabla 17.

Tabla alias	Registros	Columna alias	A	X	Calidad
Cliente	7620938	Nombre	1270	0,00016665	9,998
		Cadena generada	1032369	0,13546482	8,645
Identificación	6504968	Identificación	75565	0,01161651	9,884
Organización	157928	Descripción	202	0,00127906	9,987
Localización	6310400	Descripción	176104	0,02790695	9,721
Detalle	6127687	Primer nombre	6278	0,00102453	9,99
		Segundo nombre	11728	0,00191394	9,981
		1° y 2° nombre	704	0,00011489	9,999
		1° y 2° apellido	906	0,00014785	9,999

9,8

Tabla 20 Evaluación de calidad: dimensión exactitud, factor precisión

En la Tabla 21 se presenta ejemplos de la información contemplada para el análisis, donde se puede evidenciar que existe información que no brinda suficiente detalle; por ejemplo para la tabla localización la descripción registrada: "SUR".

Columna alias	Descripción ejemplo	
Nombre	No brindan el suficiente detalle	I2E
Cadena generada	No presenta formato	99999/1201441142001
Identificación	No brindan el suficiente detalle	00009449
Descripción	No brindan el suficiente detalle	SUR
Primer nombre	No brindan el suficiente detalle	DE
Segundo nombre	No brindan el suficiente detalle	JO
1° y 2° nombre	No brindan el suficiente detalle	CAICEDO
1° y 2° apellido	No brindan el suficiente detalle	ESTEBAN

Tabla 21 Ejemplos descripción: dimensión exactitud, factor correctitud precisión

En la Tabla 20 y Figura 14 se puede determinar el valor promedio de 9,8 que corresponde a la valoración de calidad que la ubica en el rango altamente satisfactorio descrito en la Tabla 14.

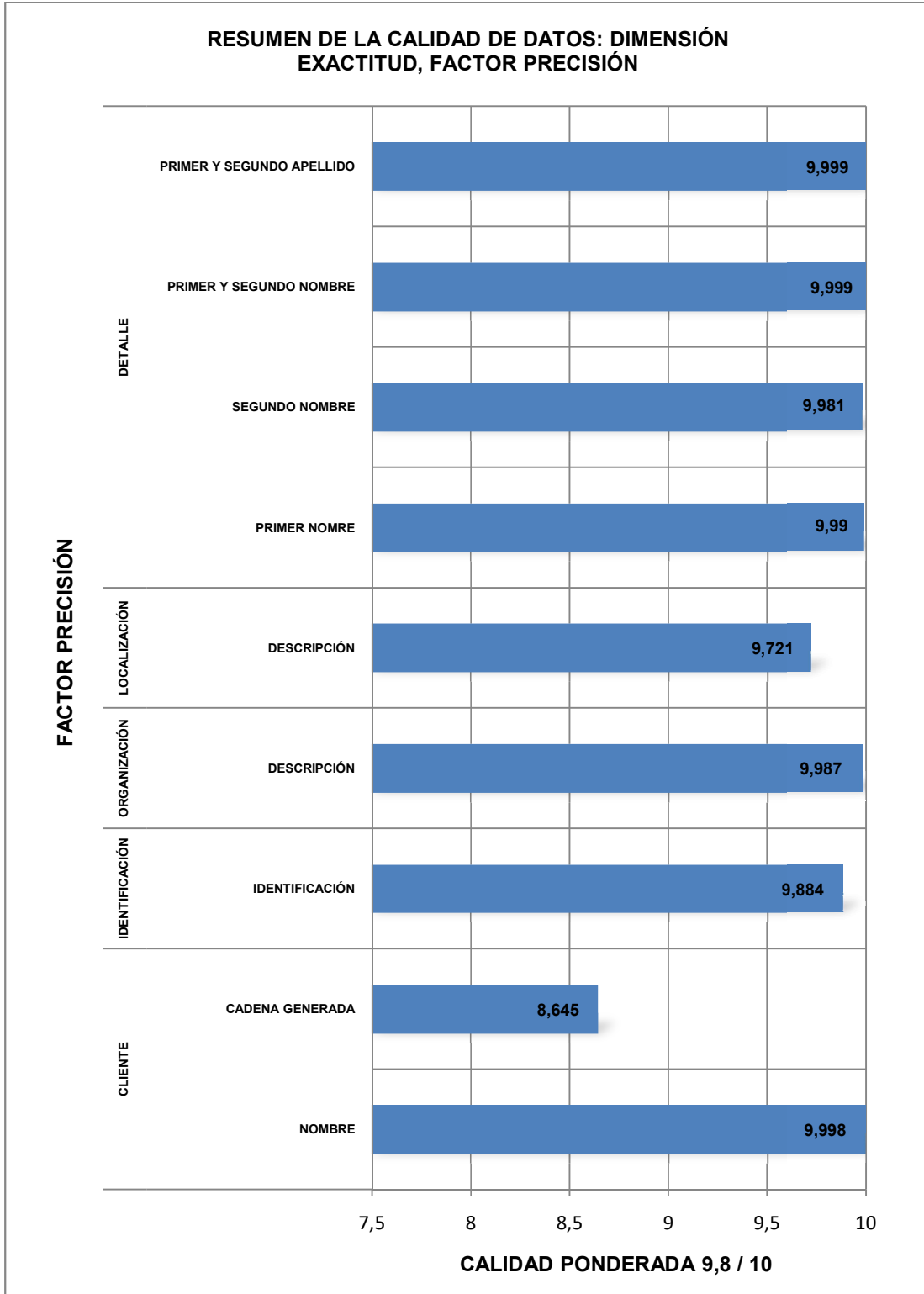


Figura 14 Resultado de la evaluación exactitud, factor precisión

4.1.4 Resultados de la evaluación completitud, factor densidad

En la Tabla 22 se presenta los resultados cuantitativos para la dimensión completitud factor densidad que corresponden a las tablas: detalle, cliente, localización y organización, para el cálculo se emplea la función promedio que devuelve la media de una serie de valores, generando un promedio de calidad descrito en la Tabla 17.

Tabla alias	Registros	Columna alias	A	X	Calidad / 10
Cliente	7620938	Fecha creación	1041	0,000136597	9,999
		Nombre	603206	0,079151149	9,208
		Cadena generada	596	7,82056E-05	9,999
Localización	6310400	Descripción	1511550	0,239533152	7,605
Organización	157928	Nombre	116383	0,736937085	2,631
		Descripción	8311	0,052625247	9,474
Detalle	6127687	Fecha nacimiento	261997	0,042756263	9,572
		Primer nombre	31433	0,005129668	9,949
		Segundo nombre	368720	0,060172786	9,398
		Nombre completo	5075853	0,828347303	1,717
		Primer apellido	30248	0,004936283	9,951
		Segundo apellido	186985	0,030514777	9,695

Tabla 22 Evaluación de calidad: dimensión completitud, factor correctitud densidad

En la Tabla 23 se presenta ejemplos de la información contemplada para el análisis, donde se puede evidenciar la ausencia de valor (Nulos y Blancos).

Columna alias	Descripción ejemplo	
Fecha creación	Nulo	NULL
Nombre	Nulo	NULL
Cadena generada	Blanco	
Descripción	Nulo	NULL
Nombre	Nulo	NULL
Descripción	Nulo	NULL
Fecha nacimiento	Nulo	NULL
Primer nombre	Nulo	NULL
Segundo nombre	Nulo	NULL
Nombre completo	Nulo	NULL
Primer apellido	Nulo	NULL
Segundo apellido	Nulo	NULL

Tabla 23 Ejemplos descripción: dimensión completitud, factor correctitud densidad

En la Tabla 22 y Figura 15 se puede determinar el valor promedio de 8,26 que corresponde a la valoración de calidad que la ubica en el rango deficiente descrito en la Tabla 14.

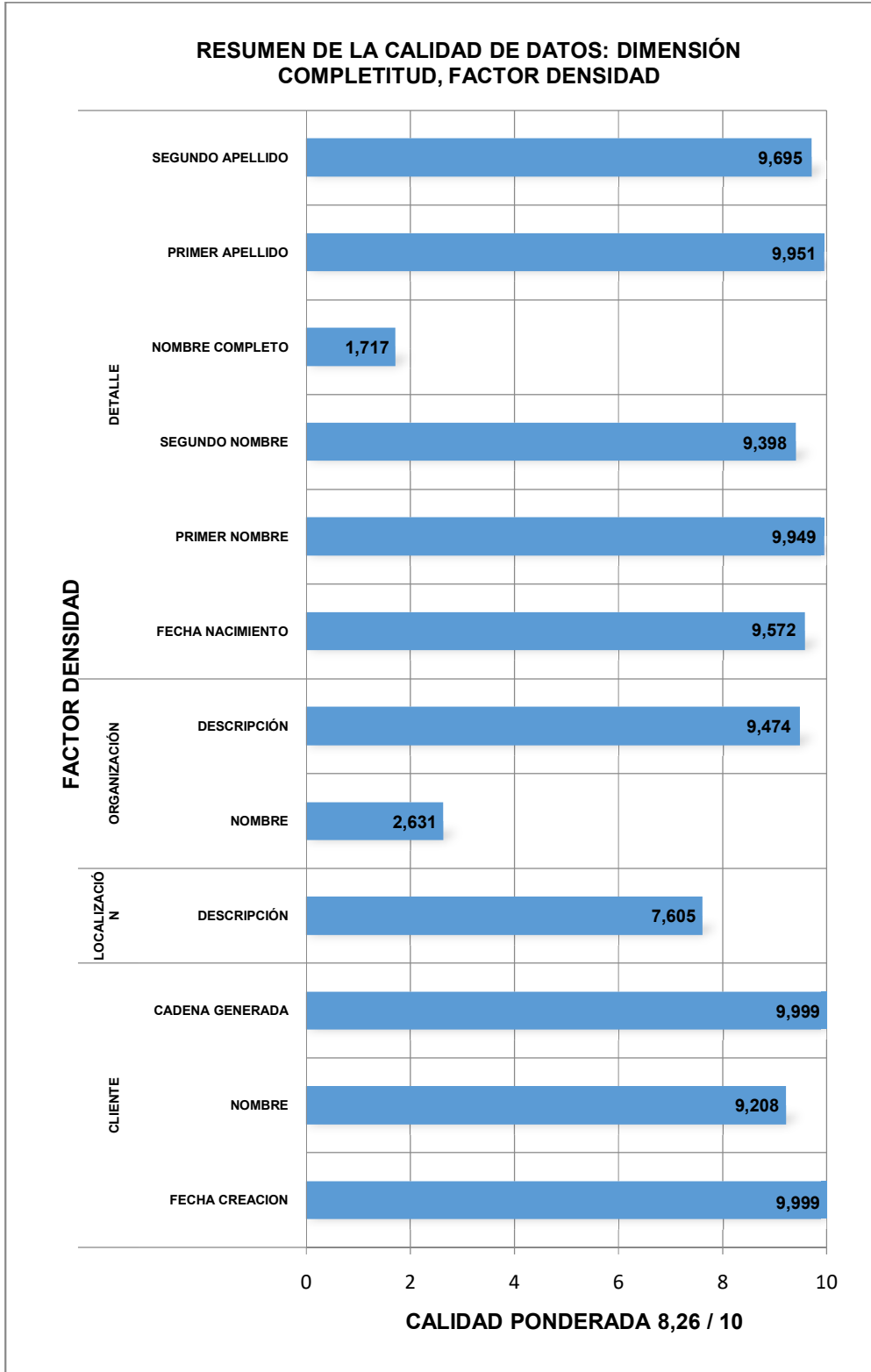


Figura 15 Resultado de la evaluación completitud, factor densidad

4.1.5 Resultados de la evaluación consistencia, factor integridad de dominio

En la Tabla 24 se presenta los resultados cuantitativos para la dimensión consistencia factor integridad de dominio que corresponden a la tabla detalle, para el cálculo se emplea la función promedio que devuelve la media de una serie de valores, generando un promedio de calidad descrito en la Tabla 17.

TABLAALIAS	REGISTROS	COLUMNAALIAS	A	X	CALIDAD / 10	
Detalle	6127687	Fecha de nacimiento	33158	0,005411177	9,946	9,946

Tabla 24 Evaluación de calidad: dimensión consistencia, factor integridad de dominio.

En la Tabla 24 se presenta un ejemplo de la información contemplada para el análisis, donde se puede evidenciar que existe información fuera de rango; por ejemplo para la tabla detalle la fecha de nacimiento registrada corresponde a 110 años.

COLUMNAALIAS	EJEMPLO DESCRIPCIÓN	
Fecha de nacimiento	Regla fuera de dominio	110 años

Tabla 25 Ejemplos descripción: dimensión consistencia, factor integridad de dominio

En la Tabla 24y Figura16 se puede determinar el valor promedio de 9,94 que corresponde a la valoración de calidad que la ubica en el rango altamente satisfactorio descrito en la Tabla 14.

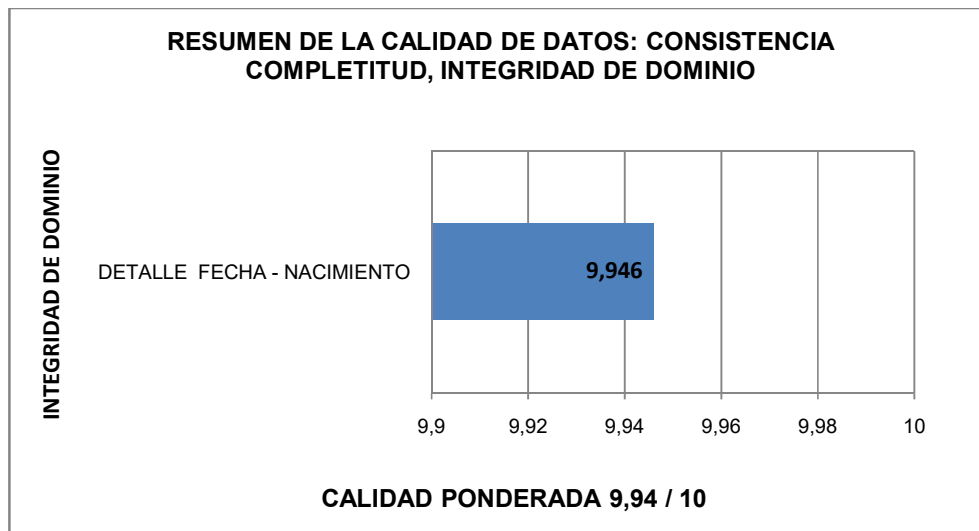


Figura 16 Resultado de la evaluación consistencia, factor integridad de dominio

4.1.6 Resultados de la evaluación consistencia, factor inter - relación

En la Tabla 26 se presenta los resultados cuantitativos para la dimensión consistencia factor inter - relación que corresponden a las tablas: cliente, identificación, localización y detalle, para el cálculo se emplea la función promedio que devuelve la media de una serie de valores, generando una ponderación de calidad descrito en la Tabla17.

TABLAALIAS	REGISTROS	COLUMNAALIAS	A	X	CALIDAD / 10
Cliente - Identificación	7620938	ID CLIENTE	1115970	0,14643473	8,536
Cliente - Detalle		ID CLIENTE	1493251	0,195940578	8,041
Localización - Detalle	6310400	ID CIUDAD	187907	0,029777352	9,702
		ID PAIS	182715	0,028954583	9,71

Tabla 26 Evaluación de calidad: dimensión consistencia, factor inter-relación

En la Tabla 26y Figura 17 se puede determinar el valor promedio de 8,99 que corresponde a la valoración de calidad que la ubica en el rango poco satisfactorio descrito en la Tabla 14.

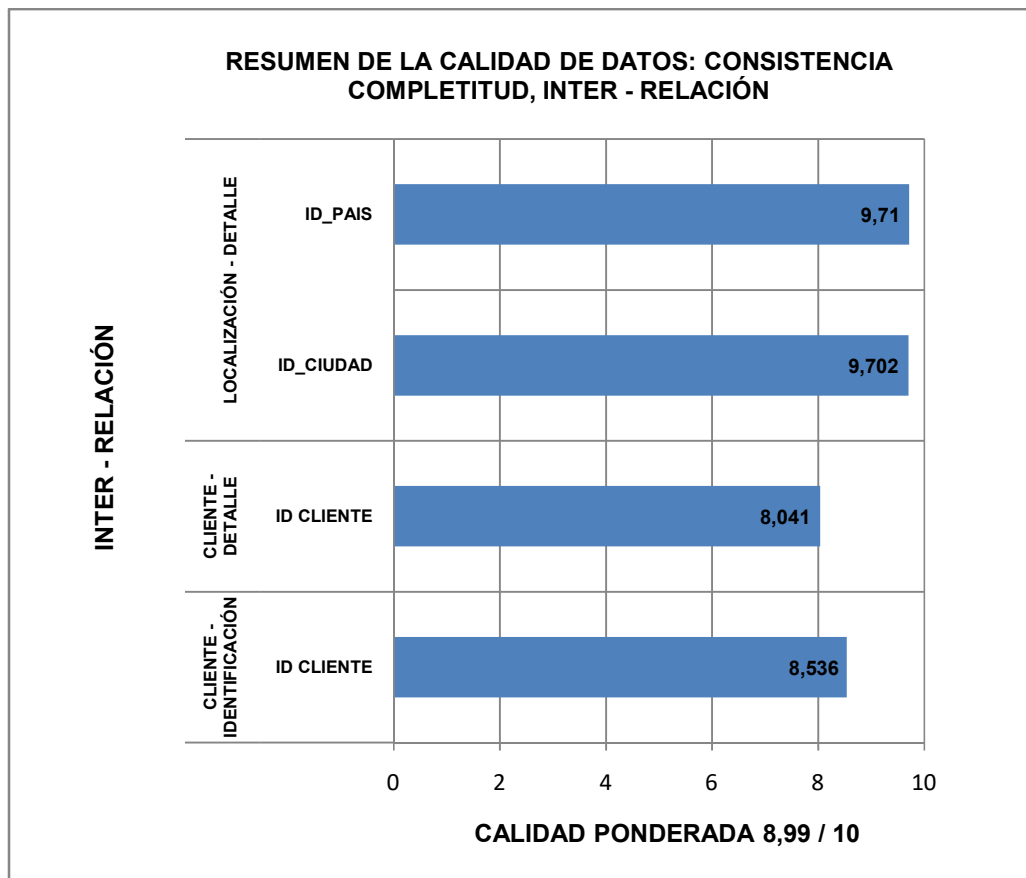


Figura 17 Resultado de la evaluación consistencia, factor inter-relación

4.1.7 Resultados de la evaluación unicidad, factor no duplicación

En la Tabla 27 se presenta los resultados cuantitativos para la dimensión unicidad factor no duplicación que corresponden a las tablas: cliente, identificación, organización, localización y detalle, para el cálculo se emplea la función promedio que devuelve la media de una serie de valores, generando un promedio de calidad descrito en la Tabla 17.

TABLAALIAS	REGISTROS	COLUMNAALIAS	A	X	CALIDAD / 10
Cliente	7620938	Nombre	209321	0,027466566	9,725
		Cadena generada	149339	0,019595882	9,804
Identificación	6504968	Identificación	13044	0,002005237	9,98
Organización	157928	Nombre	4991	0,031603009	9,684
		Descripción	2525	0,015988298	9,84
Localización	6310400	Cadena generada	337	5,34039E-05	9,999
		Descripción	16509	0,002616157	9,974
		Nombre descripción	2256	0,000357505	9,996
Detalle	6127687	Primer y segundo nombre	1736	0,000283304	9,997

Tabla 27 Evaluación de calidad: dimensión unicidad, factor no duplicación

En la Tabla 28 se presenta ejemplos de la información contemplada para el análisis, donde se puede evidenciar que existe información duplicada; por citar el primer ejemplo para la tabla cliente el nombre registrado es “ZAMBRANO VINCES MARIANA DE JESUS” con una duplicación de dos registros.

COLUMNAALIAS	DESCRIPCIÓN	EJEMPLO	DUPLICACIÓN
Nombre	información duplicada	ZAMBRANO VINCES MARIANA DE JESUS	2
Cadena generada	información duplicada	6745873	5
Identificación	información duplicada	1798458946	8
Nombre	información duplicada	VIRTCORPORATION S. A	5
Descripción	información duplicada	COMERCIALIZACION DE ARTICULO ELECTRONICOS	2
Cadena generada	información duplicada	003/000000000064950/01/A/	2
Descripción	información duplicada	EMPRESA AGRICOLA AGRIFEG	23
Nombre descripción	información duplicada	SANTA ROSA - SANTA ROSA	4
Primer y segundo nombre	información duplicada	GABRIEL	9

Tabla 28 Ejemplos de la dimensión unicidad, factor no duplicación.

En la Tabla 27y Figura 18 se puede determinar el valor promedio de 9,88 que corresponde a la valoración de calidad que la ubica en el rango altamente satisfactorio descrito en la Tabla 14.

RESUMEN DE LA CALIDAD DE DATOS: UNICIDAD, NO DUPLICACIÓN

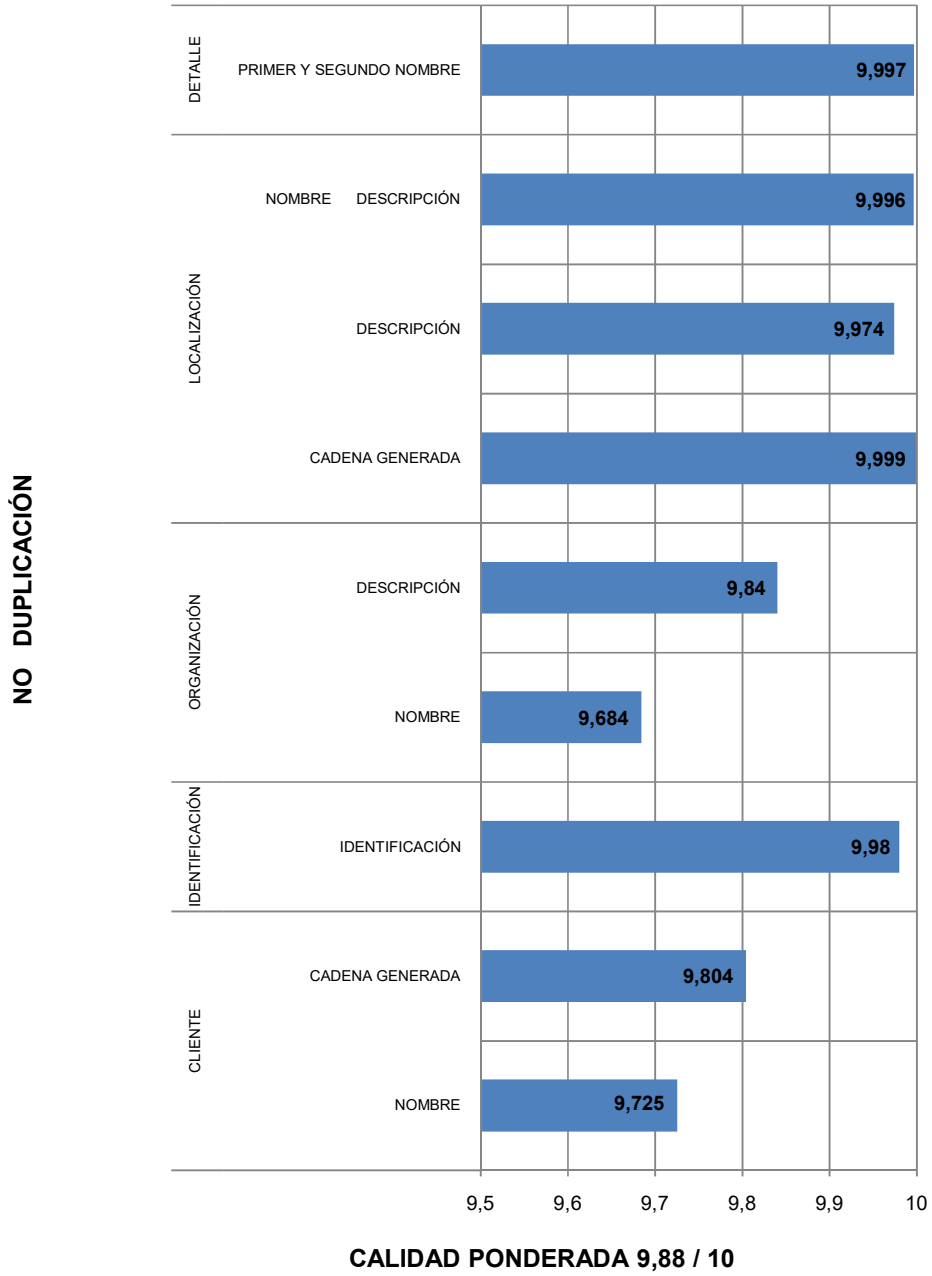


Figura 18 Resultado de la evaluación unicidad, factor no duplicación

4.1.8 Resultados de la evaluación unicidad, factor no contradicción

En la Tabla 29 se presenta los resultados cuantitativos para la dimensión unicidad factor no contradicción que corresponden a las tablas: cliente y detalle, para el cálculo se emplea la función promedio que devuelve la media de una serie de valores, generando una calidad descrito en la Tabla 17.

TABLA ALIAS	REGISTROS	COLUMNA ALIAS	A	X	CALIDAD/ 10
Cliente / Detalle	6127687	Primer nombre	31437	0,005130321	9,949
		Segundo nombre	368720	0,060172786	9,398
		Primer apellido	30248	0,004936283	9,951
		Segundo apellido	186985	0,030514777	9,695
					9,748

Tabla 29 Evaluación de calidad: dimensión unicidad, factor no contradicción.

En la Tabla 30 se presenta ejemplos de la información contemplada para el análisis, donde se puede evidenciar que existe información contradictoria entre las tablas cliente y detalle; por citar el primer ejemplo para la tabla cliente el primer nombre registrado es “Martha” mientras que en la tabla detalle el primer nombre es “Rosario”.

CLIENTE / DETALLE		
COLUMNA ALIAS	DESCRIPCIÓN	EJEMPLO
PRIMER NOMBRE	Información contradictoria	MARTHA - ROSARIO
SEGUNDO NOMBRE	Información contradictoria	ANDRES - EDUARDO
PRIMER APELLIDO	Información contradictoria	SUAREZ - LEON
SEGUNDO APELLIDO	Información contradictoria	FLORES - DE LA CRUZ

Tabla 30 Ejemplos de la dimensión unicidad, factor no contradicción.

En la Tabla 29 y Figura 19 se puede determinar el valor promedio de 9,748 que corresponde a la valoración de calidad que la ubica en el rango altamente satisfactorio descrito en la Tabla 14.

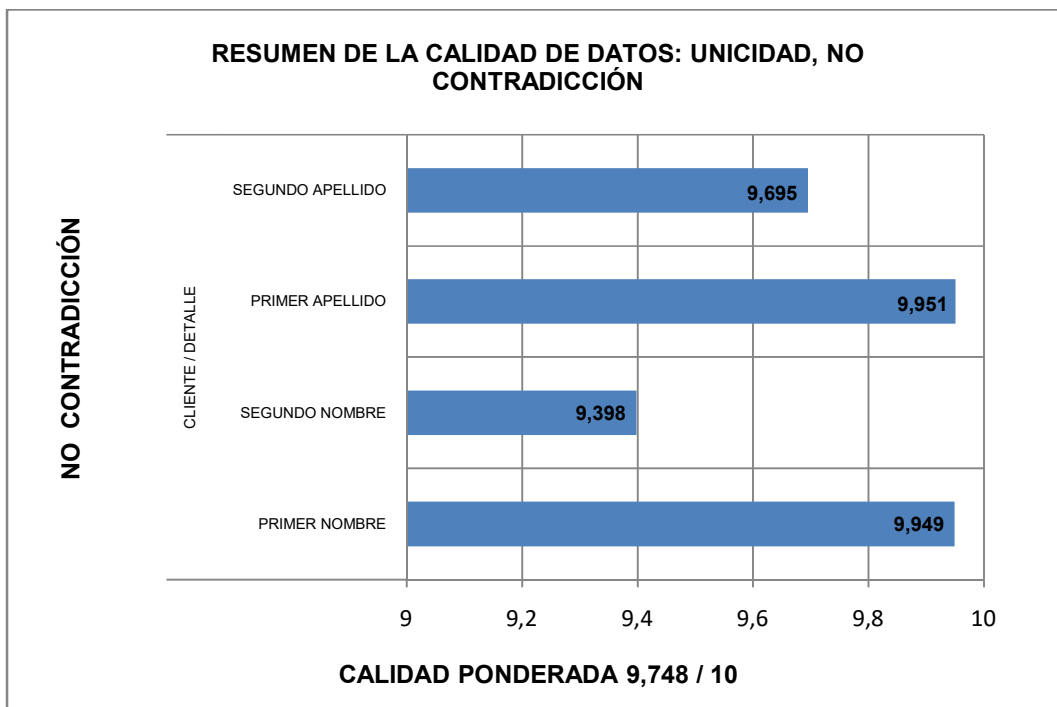


Figura 19 Resultado de la evaluación unicidad, factor no contradicción.

Se presenta el resultado final de la evacuación de las cuatro dimensiones con sus respectivos factores, dicho resultado lo ubica como satisfactorio, como se muestra en la Tabla 14.

Dimensión	Factor	Calidad	Satisfacción
Exactitud	Correctitud sintáctica	9,859	9,83 Altamente satisfactorio
	Precisión	9,8	
Complejidad	Densidad	8,267	8,267 Deficiente
Consistencia	Integridad - dominio	9,946	9,472 Satisfactorio
	Integridad inter-relación	8,997	
Unicidad	No duplicación	9,889	9,819 Altamente satisfactorio
	No contradicción	9,748	
TOTAL		9,347	SATISFACTORIO

Tabla 31 Resultado final de la evaluación de la calidad de datos.

4.2 Discusión

En el presente trabajo se diseñó e implementó un framework de calidad de datos para el datawarehouse de la entidad financiera. Sobre la base de la revisión sistemática de la literatura, se adoptó una estructura compuesta por cinco fases que facilitan y garantizan calidad de datos:

1. Requisitos
2. Especificación
3. Diseño
4. Ejecución
5. Conclusión de la evaluación

De acuerdo con los resultados obtenidos en la encuesta que se realizó, se identificó cuatro inconvenientes de calidad de datos: información duplicada o faltante, datos no precisos, información no exacta e información insuficiente, lo que demuestra que la información se encuentra comprometida, ocasionando insumos inadecuados para las posibles toma de decisiones.

En base a la información recolectada se identificó cuatro dimensiones con sus respectivos factores se realiza el análisis y la recomendación en función del promedio sobre diez de cada dimensión; evaluando la satisfacción en una escala con cuatro niveles: altamente satisfactorio, satisfactorio, poco satisfactorio y deficiente.

Para la dimensión exactitud con los factores correctitud sintáctica y precisión la ponderación de calidad registra 9,83 con un grado altamente satisfactorio. La dimensión completitud, factor densidad tiene una ponderación de calidad de 8,267 con un grado de satisfacción deficiente. En la dimensión consistencia, los factores integridad – dominio e integridad inter – relación con una ponderación de calidad 9,472 con grado satisfactorio. Finalmente para la dimensión unicidad, los factores no duplicación y no contradicción, con una ponderación de calidad de 9,819 con un rango altamente satisfactorio.

Con los resultados de la evaluación de las dimensiones se obtuvo una ponderación total promedio de 9,347 que la ubica en un rango satisfactorio; con ello se puede evidenciar que la calidad del datawarehouse es admisible, sin embargo para una entidad financiera se requiere alcanzar un grado altamente satisfactorio debido a su giro de negocio que demanda información de calidad superior.

Para alcanzar el grado altamente satisfactorio que se sugiere en la presente propuesta, se debe iniciar con la dimensión completitud, factor densidad que presenta un grado de satisfacción deficiente y afecta al resto de dimensiones, para ello se requiere identificar el origen donde se genera los campos blancos y nulos para posteriormente trabajar en los planes de mejora específicos.

5 CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

Una vez realizado el diseño e implementación de un framework de calidad de datos para la entidad financiera, se obtuvo información necesaria y suficiente para el presente caso de estudio cuyo análisis permite llegar a las siguientes conclusiones:

Mediante la revisión bibliográfica y comparación de siete frameworks de calidad de datos existentes se diseñó e implementó un marco de trabajo que proporciona una estructura y tecnología compuesta por cinco fases que facilitan y garantizan una información de calidad:

1. Requisitos
2. Especificación
3. Diseño
4. Ejecución
5. Conclusión de la evaluación

En numerosas ocasiones se toman medidas de vital importancia para la entidad financiera. En la actualidad la entidad no cuenta con un buen asesoramiento y análisis de datos que ha dado lugar a estrategias negativas, evidenciándose con este estudio la importancia de implementar un framework de calidad de datos cuya necesidad fue identificada a través de los stakeholders de la alta gerencia.

Se logró identificar tres requisitos relevantes a nivel operacional: Recepción, Business Intelligence y Decisiones; que se consideraron en el sistema del registro fuente. Esa estructura o arquitectura corresponde a los diferentes niveles de la información, la que debe completar su ciclo hasta llegar a su repositorio y posterior proceso, generando el insumo requerido para la toma de decisiones.

Para el análisis y evaluación del datawarehouse del registro fuente, perteneciente a la entidad financiera se analizó más del veinte seis millones de registros, contemplados en cinco tablas que corresponden al corte 2019-10, contenido en un esquema arquitectónico basado en cinco fases o etapas (fuente de datos, integración, repositorio, analítica y acceso).

De acuerdo con la revisión literaria y contenido de la información se seleccionó cuatro dimensiones y sus respectivos siete factores de calidad para la evaluación del datawarehouse, el entendimiento de estos conceptos permitió tener una percepción más amplia del proceso de evaluación de la entidad financiera, del mismo modo se determinó que la dimensión completitud factor densidad (campos nulos y blancos) obtiene la ponderación más baja del análisis 8.27/10. Siendo esta la dimensión más crítica del análisis, pues su uso podría generar decisiones poco confiables.

En un entorno empresarial como el financiero sus éxitos están basados en la calidad y enriquecimiento de datos. Si bien es cierto la entidad financiera evaluada presenta un promedio de calidad de datos de 9,35/10 en el registro fuente, no es suficiente, es preciso adoptar una estrategia de calidad de datos adecuada que permita que la entidad mejore su imagen y marca.

5.2 Recomendaciones

En el presente trabajo se desarrolló e implementó un framework de calidad de datos orientado a una entidad financiera para el datawarehouse del sistema de registro fuente. Dicho framework está basado en la revisión bibliográfica de fuentes similares que se ajustan a las necesidades. A pesar de haber logrado los objetivos propuestos, aún queda mucho trabajo por realizar.

Aun si se tuviese las mejores metodologías, técnicas o framework para la gestión de calidad de datos, no se lo lograría si no se tuviese una verdadera cultura de calidad de datos. Se recomienda que los proyectos tecnológicos tengan como requerimiento un componente de calidad de datos que valide los siguientes factores: datos blancos, nulos, caracteres especiales, datos fuera de rango, información contradictoria, entre otros.

El alcance del proyecto fue considerado únicamente el datawarehouse del sistema de registro fuente, para obtener mayores beneficios de calidad de datos de la entidad financiera, se recomienda la implementación y evaluación del framework para los sistemas restantes, así como la identificación de sus respectivas dimensiones.

Mediante los resultados obtenidos en la evaluación del datawarehouse, se recomienda la identificación y priorización de las causas que originan los problemas y posteriormente a ello trabajar en planes de mejora que aborden dichos problemas.

Se recomienda el diseño e implementación de soluciones que corrijan los problemas identificados, así como el monitoreo continuo que mitigue dichos problemas.

6 REFERENCIAS

- [1] J. Du y L. Zhou, «Improving financial data quality using ontologies», *Decis. Support Syst.*, vol. 54, n.º 1, pp. 76-86, 2012.
- [2] D. McGilvray, *Executing data quality projects: ten steps to quality data and trusted information*. 2008.
- [3] C. Batini, C. Cappiello, C. Francalanci, y A. Maurino, «Methodologies for data quality assessment and improvement», *ACM Comput. Surv.*, vol. 41, n.º 3, pp. 1-52, 2009.
- [4] J. L. Kulikowski, «Data Quality Assessment», *Int. J. Organ. Collect. Intell.*, vol. 4, n.º 1, pp. 24-36, 2014.
- [5] J. Debattista, S. Auer, y C. Lange, «Luzzu-A Framework for Linked Data Quality Assessment», *Proc. - 2016 IEEE 10th Int. Conf. Semant. Comput. ICSC 2016*, pp. 124-131, 2016.
- [6] F. Sidi, P. H. Shariat Panahy, L. S. Affendey, M. A. Jabar, H. Ibrahim, y A. Mustapha, «Data quality: A survey of data quality dimensions», *Proc. - 2012 Int. Conf. Inf. Retr. Knowl. Manag. CAMP'12*, n.º June 2014, pp. 300-304, 2012.
- [7] S. M. Batini Carlo, *Data Quality: concepts, methodologies and techniques*. Italia, 2006.
- [8] A. J. Martín, «Gestión de la calidad de los datos en la empresa», Universidad Politécnica de Madrid, 2017.
- [9] F. Serra, «Evaluación de la Calidad de Datos en un Sistema de Data Warehousing: Un enfoque basado en Contextos», Universidad de la República Montevideo - Uruguay, 2015.
- [10] G. Marisol y C. Tipantaxi, «APLICACIÓN DE UN FRAMEWORK DE GESTIÓN DE CALIDAD DE DATOS PARA EL EJE ESTRATÉGICO DE DOCENCIA DE LA ESCUELA POLITÉCNICA NACIONAL», ESCUELA POLITÉCNICA NACIONAL, 2018.
- [11] Organización Internacional de Normalización, «ISO/IEC 25012», 2017. [En línea]. Disponible en: <https://iso25000.com/index.php/normas-iso-25000/iso-25012>.
- [12] G. F. Knolmayer y M. Röthlin, «Quality of material master data and its effect on the usefulness of distributed ERP systems», *Lect. Notes Comput. Sci. (including Subser.*

- Lect. Notes Artif. Intell. Lect. Notes Bioinformatics*), vol. 4231 LNCS, pp. 362-371, 2006.
- [13] A. Haug, F. Zachariassen, y D. van Liempd, «The costs of poor data quality», *J. Ind. Eng. Manag.*, vol. 4, n.º 2, pp. 168-193, 2011.
- [14] E. Malinowski, *Advanced Data Warehouse Design*, 1st ed. 20. Berlin, Germany: Springer, Berlin, Heidelberg, 2008.
- [15] J. C. Moyota Chafla, «ANÁLISIS DE FRAMEWORK PHP OPENSOURCE EN EL DESARROLLO DE UNA APLICACIÓN WEB PARA LA GESTIÓN DE REPORTES ESTADÍSTICOS EN EL SISTEMA DE AGUA POTABLE DE LA PROVINCIA DE CHIMBORAZO», UNIVERSIDAD NACIONAL DE CHIMBORAZO, 2015.
- [16] C. B. Espinoza Aguirre, «Comparativa de Frameworks para el desarrollo de aplicaciones con php», Universidad del Azuay, 2012.
- [17] W. Marchment y C. Delahunty, «Data Quality Framework(DQF)», n.º June, pp. 1-10, 2013.
- [18] K. Kerr, «The development of a data quality framework and strategy for the New Zealand Ministry of Health - Part 1», *Healthc. Rev. Online*, vol. 8, n.º 3, pp. 1-14, 2013.
- [19] C. Fürber, «Data Quality Management with Semantic Technologies», *Springer Gabler*, p. 230, 2016.
- [20] Bank of England, *Statistics and Regulatory Data Division Data Quality Framework*, n.º March. London: Business and Communications Group, 2014.
- [21] P. Cupoli, S. Earley, y D. Henderson, «DAMA-DMBOK2: The Data Management Body Of Knowledge», p. 27, 2014.
- [22] A. Oliveros, F. J. Danyans, y M. L. Mastropietro, «Stakeholders en los requerimientos de aplicaciones web», *XLIII Jornadas Argentinas Informática e Investig. Oper. (43JAIIO)-XV Simp. Argentino Ing. Softw. (Buenos Aires, 2014)*, n.º September, pp. 240-252, 2014.
- [23] F. Clerici, B. Fernández, y A. Marotta, «Herramienta para la evaluación de calidad de datos», 2019.

7 ANEXOS

Anexo I RESOL_SB-2017-602 - SBS Organismo de Control



Profesores a la Justicia

RESOLUCIÓN No. SB-2017-602

CHRISTIAN CRUZ RODRÍGUEZ
SUPERINTENDENTE DE BANCOS

CONSIDERANDO:

QUE el Código Orgánico Monetario y Financiero se encuentra en vigencia desde su publicación en el Segundo Suplemento del Registro Oficial No. 332 de 12 de septiembre de 2014;

QUE el numeral 1, del artículo 62 del Código Orgánico Monetario y Financiero establece como función de la Superintendencia de Bancos ejercer la vigilancia, auditoría, control y supervisión del cumplimiento de las disposiciones de este Código y de las regulaciones dictadas por la Junta de Política y Regulación Monetaria y Financiera, en lo que corresponde a las actividades financieras ejercidas por las entidades que conforman los sectores financieros público y privado;

QUE el último inciso del artículo 62 ibídem, dispone que la Superintendencia, para el cumplimiento de sus funciones, podrá expedir las normas en las materias propias de su competencia, sin que puedan alterar o innovar las disposiciones legales ni las regulaciones que expida la Junta de Política y Regulación Monetaria y Financiera;

QUE mediante resolución No. 380-2017-F del 22 de mayo del 2017, la Junta de Política y Regulación Monetaria y Financiera emitió la "Política para la gestión integral y administración de riesgos de las entidades de los sectores financieros público y privado", y dispone que mediante norma de control la Superintendencia de Bancos expedirá las disposiciones que reglamenten la política contenida en dicha resolución; y,

EN ejercicio de sus funciones legales,

RESUELVE:

EXPEDIR LA NORMA DE CONTROL PARA LA GESTIÓN INTEGRAL Y ADMINISTRACIÓN DE RIESGOS DE LAS ENTIDADES DE LOS SECTORES FINANCIEROS PÚBLICO Y PRIVADO

SECCIÓN I.- ÁMBITO Y ALCANCE.

ARTÍCULO 1.- Las disposiciones de la presente norma son aplicables a las entidades de los sectores financieros público y privado, cuyo control le compete a la

Quito: Av. 12 de Octubre N24-185 y Madrid. Teléfonos: (02) 299 76 00 / 299 61 00
Guayaquil: Chimborazo 412 y Aguirre. Teléfono: (04) 370 42 00
Cuenca: Antonio Borrero 710 y Presidente Córdova. Teléfonos: (07) 283 59 61 / 283 57 26
Portoviejo: Calle Olmedo y Alajuela. Teléfonos: (05) 263 49 51 / 263 58 10

www.superbancos.gob.ec
@superbancos.EC
Superintendencia de Bancos

- p. Construir la matriz de riesgo institucional y someterla a evaluación del comité de administración integral de riesgos;
- q. Elaborar y presentar trimestralmente al comité de administración integral de riesgos el informe correspondiente al análisis de cada uno de los riesgos de ese período, el cual debe ser remitido a la Superintendencia de Bancos, previa aprobación del directorio, dentro de los quince (15) días posteriores al término de cada trimestre y cuyo contenido mínimo, será determinado por la Superintendencia de Bancos para cada uno de los riesgos;
- r. Preparar las actas de las sesiones del comité de administración integral de riesgos para su conocimiento y aprobación; y,
- s. Las demás que determine el comité de administración integral de riesgos de la entidad, o las que sean dispuestas por la Superintendencia de Bancos.

ARTÍCULO 5.- El comité de administración integral de riesgos y la unidad de riesgos estarán dotados, de manera permanente, de los recursos humanos, materiales y tecnológicos necesarios para el cumplimiento de sus funciones. Estarán conformados por personas idóneas que deben acreditar un alto conocimiento y experiencia en materia de gestión y control de riesgos y capacidad de comprender las metodologías y procedimientos utilizados en la entidad para identificar, medir, controlar / mitigar y monitorear los riesgos asumidos y por asumir, de manera tal que garanticen el adecuado cumplimiento de sus funciones.

Las entidades controladas podrán crear subunidades de riesgo especializadas cuyo funcionamiento se registrará por las disposiciones contenidas en la presente norma de control y en otras que expida la Superintendencia de Bancos, atendiendo la naturaleza de su función.

SECCIÓN III.- DISPOSICIONES GENERALES

PRIMERA.- Las entidades controladas deben incluir en su marco de administración integral de riesgos, y en forma detallada para cada tipo de riesgo, los lineamientos establecidos en la "Política para la gestión integral y administración de riesgos de las entidades de los sectores financieros público y privado", expedida por la Junta de Política y Regulación Monetaria y Financiera.

SEGUNDA.- Las entidades controladas deben disponer de un sistema informático capaz de proveer a la administración y a las áreas involucradas, toda la información necesaria para identificar, medir, controlar / mitigar y monitorear las exposiciones de riesgo que están asumiendo, y apoyar en la toma de decisiones oportunas y adecuadas.

Estos sistemas incorporarán los procesos definidos para la elaboración de los informes, que involucren todas las variables relacionadas con la medición de los riesgos y la vulnerabilidad institucional, bajo las diversas condiciones del entorno.

Anexo II Resolución No. JB-2010-1782 de Junta Bancaria

Junta Bancaria del Ecuador

RESOLUCIÓN No. JB-2010-1782

LA JUNTA BANCARIA

CONSIDERANDO:

Que el artículo 11 de la Constitución de la República del Ecuador, publicada en el Registro Oficial No. 449 de 20 del octubre de 2008, determina que el ejercicio de los derechos de los ciudadanos y ciudadanas ante las autoridades competentes se regirá, entre otros principios, en base a que ninguna norma jurídica podrá restringir el contenido de los derechos ni de las garantías constitucionales; los principios y los derechos son inalienables, irrenunciables, indivisibles, interdependientes y de igual jerarquía; y que el más alto deber del Estado consiste en respetar y hacer respetar los derechos garantizados en la Constitución;

Que el artículo 52 de la Constitución de la República del Ecuador establece que las personas tienen derecho a disponer de bienes y servicios de óptima calidad y a elegirlos con libertad, así como a una información precisa y no engañosa sobre su contenido y características;

Que el artículo 55 de la Constitución de la República del Ecuador dispone que las personas usuarias y consumidoras podrán constituir asociaciones que promuevan la información y educación sobre sus derechos, y las representen y defiendan ante las autoridades judiciales o administrativas;

Que el artículo 66 de la Constitución de la República del Ecuador reconoce y garantiza a las personas, entre otros, los derechos a desarrollar actividades económicas, en forma individual o colectiva; a la libertad de contratación; a la protección de datos de carácter personal; a dirigir quejas y peticiones individuales y colectivas a las autoridades y a recibir atención o respuestas motivadas, a acceder a bienes y servicios públicos y privados de calidad;

Que el artículo 75 de la Constitución de la República del Ecuador establece que toda persona tiene derecho al acceso gratuito a la justicia y a la tutela efectiva, imparcial y expedita de sus derechos e intereses, con sujeción a los principios de intermediación y celeridad; y que en ningún caso quedará en indefensión;

Que el artículo 213 de la Constitución de la República del Ecuador establece que las superintendencias son organismos técnicos de vigilancia, auditoría, intervención y control de las actividades económicas, sociales y ambientales, y de los servicios que prestan las entidades públicas y privadas, con el propósito de que estas actividades y servicios se sujeten al ordenamiento jurídico y atiendan al interés general;

Que el artículo 308 de la Constitución de la República del Ecuador dispone que las actividades financieras son un servicio de orden público; que el Estado fomentará el acceso a los servicios financieros y a la democratización del crédito; y, que se prohíben las prácticas colusorias, el anatocismo y la usura;

Que el artículo 372 de la Constitución de la República del Ecuador establece que cada entidad integrante del sistema financiero nacional tendrá una defensora o defensor del cliente, que será independiente de la institución y designado de acuerdo con la ley;

Que el artículo 1 de la Ley Orgánica de Defensa del Consumidor establece como objeto de la Ley el normar las relaciones entre proveedores y consumidores promoviendo el conocimiento y protegiendo los derechos de los consumidores, procurando la equidad y la seguridad jurídica en las relaciones entre las partes;

Que el artículo 4 de la citada ley, establece, entre otros derechos fundamentales del consumidor, la seguridad en el consumo de bienes y servicios de óptima calidad y a elegirlos con libertad; el derecho a la información adecuada, veraz, clara, oportuna y completa; a la

Junta Bancaria del Ecuador

Resolución No. JB-2010-1782

Página 7

en caso de consentimiento libre y expreso, específico, inequívoco e informado, por parte del usuario, de disposición judicial o del mandato de la ley;

- 14.3** Recibir protección de los datos personales que las entidades financieras obtengan del usuario para la prestación de productos y servicios financieros prestados por vía electrónica. Las instituciones financieras adoptarán específicamente las medidas de seguridad necesarias para este tipo de operaciones financieras;
- 14.4** Obtener protección de los datos personales sobre su solvencia patrimonial y crediticia, y a que las instituciones financieras respeten las normas relativas a la reserva y sigilo bancario;
- 14.5** Exigir rectificación de la información de los datos personales en las bases de datos cuando ésta sea inexacta o errónea;
- 14.6** Demandar protección cuando las instituciones financieras empleen métodos de cobranza extrajudicial que atenten contra su privacidad, dignidad personal y/o familiar;
- 14.7** Exigir que se mantenga la validez de las ofertas financieras. Las condiciones incluidas en los contratos tendrán fuerza vinculante si llegan a efectuarse con base en ellas;
- 14.8** Formar y participar en asociaciones para la defensa de los derechos del usuario del sistema financiero, y acudir al defensor del cliente en defensa de sus derechos; y,
- 14.9** Demandar la cobertura del Fondo de Garantía de Depósitos, de acuerdo con la ley.

SECCION III.- DE LA DEFENSA DE LOS DERECHOS DEL USUARIO DEL SISTEMA FINANCIERO

PARAGRAFO I.- DERECHO A RECLAMO

ARTÍCULO 15.- El usuario tiene derecho a reclamar por la existencia de cláusulas y/o prácticas abusivas o prohibidas que incluyan y ejecuten las instituciones financieras; y en general de todas aquellas acciones u omisiones que vayan en desmedro de sus derechos, para lo cual podrá ejercer ante las instituciones competentes las acciones que correspondan de acuerdo con la ley. Igualmente, para exigir las indemnizaciones y reparaciones del caso.

ARTÍCULO 16.- El usuario tiene derecho a que su reclamo o queja sea recibido en la institución financiera, a que sea atendido en forma diligente; a que las respuestas que reciba sean escritas, motivadas, oportunas y que tengan firma de responsabilidad.

ARTÍCULO 17.- El usuario tiene derecho a presentar sus quejas y reclamos ante el defensor del cliente, ante la Superintendencia de Bancos y Seguros y las demás instancias que determine la ley, cuando sus derechos han sido vulnerados y/o han recibido productos o servicios indebidos o un trato inapropiado por parte de las instituciones financieras; y, particularmente cuando no estén de acuerdo con lo resuelto por la institución financiera ante la que han presentado sus reclamaciones.

Sin embargo, cuando el reclamo haya sido puesto en consideración de de la justicia ordinaria por parte de los usuarios del sistema financiero, la Superintendencia de Bancos y Seguros se abstendrá de seguir conociéndolo, en atención al principio de independencia de la Función Judicial, consagrado en el numeral 1 del artículo 168 de la Constitución de la República, que concuerda con el artículo 8 del Código Orgánico de la Función Judicial.



SUPERINTENDENCIA
DE BANCOS

Ante la Señal

Resolución No. SB-2018-771
Página No. 13

aspectos jurídicos para llegar al finiquito y a la finalización de la relación laboral.

- ii. Mantener un archivo digital centralizado con información actualizada del capital humano, misma que deberá detallar: formación académica y experiencia; forma y fechas de reclutamiento, selección y contratación; información histórica sobre los eventos de capacitación en los que ha participado; cargos que ha desempeñado en la entidad; resultados de evaluaciones de desempeño realizadas; fechas y causas de separación del personal que se ha desvinculado; con la finalidad de permitir la toma de decisiones por parte de los niveles directivos y la realización de análisis cualitativos y cuantitativos de acuerdo con sus necesidades.
- c. **Tecnología de la información.**- Las entidades controladas deben contar con tecnología de la información que garantice la captura, procesamiento, almacenamiento y transmisión de la información de manera oportuna y confiable; evitar interrupciones del negocio y lograr que la información, inclusive aquella bajo la modalidad de servicios provistos por terceros, esté disponible para la toma de decisiones.

Para considerar la existencia de un apropiado ambiente de gestión de riesgo tecnológico, las entidades controladas deben:

- i. Contar con un área de tecnología de la información en función del tamaño y complejidad de las operaciones, y conformar el comité de tecnología, que es el responsable de evaluar, y supervisar las actividades estratégicas de carácter tecnológico.

Dicho comité estará integrado como mínimo por: un miembro del directorio, quien lo presidirá, el representante legal de la entidad, el funcionario responsable del área de riesgo operativo y el funcionario responsable del área de tecnología, quienes no podrán delegar su participación. Mantendrá un reglamento en donde se establezcan sus funciones y responsabilidades. Las reuniones de este comité se realizarán al menos trimestralmente dejando evidencia de las decisiones adoptadas.

El comité de tecnología sesionará con la mitad más (1) uno de sus integrantes, cuyo quórum no deberá ser menor a tres (3) y las decisiones serán tomadas por mayoría de votos. El presidente del comité tendrá voto dirimente.

- ii. Con el objeto de garantizar que la administración de la tecnología de la información soporte los requerimientos de operación actuales y futuros de la entidad, debe contar al menos con lo siguiente:

- El apoyo y compromiso formal del directorio, a través de la aprobación de un plan estratégico de tecnología de la información alineado con el plan estratégico institucional; y, un plan operativo anual que establezca las actividades a ejecutar en el corto plazo, traducido en

tareas, cronogramas, personal responsable y presupuesto, de manera que se asegure el logro de los objetivos tecnológicos propuestos; y,

- Políticas, procesos, procedimientos y metodologías de tecnología de la información, alineados a los objetivos y actividades de la entidad, así como las consecuencias de su incumplimiento.

Las políticas, procesos, procedimientos y metodologías de tecnología de la información deben ser revisados y aceptados por el comité de tecnología y propuestos para la posterior aprobación del directorio; deben ser difundidos y comunicados a todo el personal involucrado de tal forma que se asegure su cumplimiento.

iii. Con el objeto de garantizar que las operaciones de tecnología de la información satisfagan los requerimientos de las entidades controladas, se debe implementar al menos lo siguiente:

- Procedimientos que establezcan las actividades y responsable de la operación y el uso de los centros de datos, que incluyan controles que eviten accesos no autorizados;
- Procedimientos de gestión de incidentes y problemas de tecnología de la información, que considere al menos su registro, priorización, análisis, escalamiento y solución; y,
- Procedimientos de respaldo de información periódicos, acorde a los requerimientos legales y de continuidad del negocio, que incluyan: la frecuencia de verificación, eliminación y el transporte seguro hacia una ubicación remota, que no debe estar expuesta a los mismos riesgos del sitio principal y mantenga las condiciones físicas y ambientales necesarias para su preservación y posterior recuperación.

iv. Con el objeto de garantizar que el proceso de adquisición, desarrollo, implementación y mantenimiento de las aplicaciones satisfagan los objetivos del negocio, las entidades controladas deben implementar:

- Una metodología que permita la administración y control del ciclo de vida de desarrollo y mantenimiento de aplicaciones, que describa las etapas del proceso, la documentación entregable en cada una de ellas, estándares de desarrollo y aseguramiento de la calidad y considere al menos lo siguiente:
 - Requerimientos funcionales aprobados por el área solicitante;
 - Requerimientos técnicos y el análisis de la relación y afectación a la capacidad de la infraestructura tecnológica actual, aprobados por el área técnica;

Anexo IV Plan estratégico de tecnologías de la información

Plan Estratégico de tecnologías de la Información
PLN-GTIC-00-02 | V01.01 | abril 2019

1. ANTECEDENTES

Mediante Decreto Ejecutivo No.677, suscrito el 13 de mayo del 2015, publicado en el Registro Oficial No512 del 1 de junio del 2015, se crea ##### como una entidad financiera que forma parte del sector financiero público.

El Artículo 3 del mencionado decreto especifica que el objeto del Banco será *"el ejercicio de actividades financieras previamente autorizadas por la Superintendencia de Bancos, y la prestación de servicios financieros de crédito, ahorro e inversión, bajo el criterio de intermediación financiera de recursos públicos y privados, atendiendo a la micro, pequeña y mediana empresa y empresas asociativas en sectores de producción, principalmente de agro negocios, comercio y servicios, con claro enfoque de desarrollo local y con preferencias en áreas rurales y urbano marginales, coadyuvando al fortalecimiento del Plan Nacional del Buen Vivir, a través de mecanismos de banca de primer y segundo piso"*.

cuenta con un marco estratégico de gestión institucional en el que se establecieron los elementos orientadores tales como: Misión, Visión, Objetivos Estratégicos, los segmentos a atender y un plan de negocios que se encuentran recopilados en el documento del Plan Estratégico 2018-2021, aprobado mediante resolución D-2018-101 de fecha 12 de diciembre de 2018. La Constitución, expresa los sentidos básicos para el modelo de gestión de la banca pública: sustentabilidad, eficiencia, accesibilidad y equidad, los cuales se han considerado en la definición de los objetivos estratégicos de

Por otro lado, #####, consciente de que los avances tecnológicos e innovaciones son la principal ventaja competitiva, para reducir costos y gastos e incrementar el nivel de satisfacción de los clientes, así como dar cumplimiento a la normativa y regulatorio vigente en el ámbito de las tecnologías de la Información y Comunicación, se elabora el presente documento de Plan Estratégico de Tecnología Informática - PETI a fin de que, en el marco de las estrategias y objetivos Institucionales, se evalúe la infraestructura tecnológica de que dispone y se identifiquen las oportunidades de mejora de: los sistemas de información, las arquitecturas y plataformas tecnológicas, las capacidades de almacenamiento y procesamiento, las comunicaciones y las seguridades, y la gestión de TI; de manera tal, que resulten alineados a las estrategias de negocio y aporten en el mediano plazo, valor agregado sustantivo a las necesidades de operación y gestión de la Entidad.

1.1 DIRECTRICES DE CONTROL DE ALTO NIVEL

Las directrices de control de alto nivel de TI se enmarcan en:

- La Codificación de las Normas de la Superintendencia de Bancos, en el libro I "Normas de control para las entidades de los sectores financieros público y privado", título IX "De la gestión y administración de riesgos", "CAPÍTULO V.- NORMA DE CONTROL PARA LA GESTIÓN DE RIESGO OPERATIVO",

Pág. 5

“SECCIÓN III.- FACTORES DEL RIESGOS OPERATIVO”, artículo 10, literal c. Tecnología de la Información, acápite ii, (Capítulo sustituido por la Res. Nro. Sb-2018-771 de 30 de julio de 2018; reformado por Res. Nro. Sb-2018-814 de 13 de agosto de 2018) indica:

“Con el objeto de garantizar que la administración de la tecnología de la información soporte los requerimientos de operación actuales y futuros de la entidad, debe contar al menos con lo siguiente:

- *El apoyo y compromiso formal del directorio, a través de la aprobación de un plan estratégico de tecnología de la información alineado con el plan estratégico institucional; y, un plan operativo anual que establezca las actividades a ejecutar en el corto plazo, traducido en tareas, cronogramas, personal responsable y presupuesto, de manera que se asegure el logro de los objetivos tecnológicos propuestos”*
- Las Normas de Control Interno de la Contraloría General del Estado, Nro. 410 “TECNOLOGÍA DE LA INFORMACIÓN”, emitidas mediante oficio Nro. 22784 SGEN del 2 de diciembre de 2009, numeral 410-03 “Plan informático estratégico de tecnología”, que indica:

“La Unidad de Tecnología de la Información elaborará e implementará un plan informático estratégico para administrar y dirigir todos los recursos tecnológicos, el mismo que estará alineado con el plan estratégico institucional y éste con el Plan Nacional de Desarrollo y las políticas públicas de gobierno.

El plan informático estratégico tendrá un nivel de detalle suficiente para permitir la definición de planes operativos de tecnología de Información y especificará como ésta contribuirá a los objetivos estratégicos de la organización; incluirá un análisis de la situación actual y las propuestas de mejora con la participación de todas las unidades de la organización, se considerará la estructura interna, procesos, infraestructura, comunicaciones, aplicaciones y servicios a brindar, así como la definición de estrategias, riesgos, cronogramas, presupuesto de la inversión y operativo, fuentes de financiamiento y los requerimientos legales y regulatorios de ser necesario.

(...)

El plan estratégico y los planes operativos de tecnología de información, así como el presupuesto asociado a éstos serán analizados y aprobados por la máxima autoridad de la organización e incorporados al presupuesto anual de la organización; se actualizarán de manera permanente, además de ser monitoreados y evaluados en forma trimestral para determinar su grado de ejecución y tomar las medidas necesarias en caso de desviaciones.

Adicional, se tomaron en cuenta las siguientes directrices de ##### propias del uso de las TICs en el entorno de las entidades que prestan servicios financieros y que se encuentran definidas en el marco normativo que las rige:

- La innovación de los productos y servicios y financieros se debe alcanzar sin poner en riesgo la seguridad y conformidad de los datos y garantizando la continuidad de las operaciones y servicios
- La incorporación de nuevas tecnologías debe garantizar la integración con los otros elementos del entorno tecnológico y la interrelación con sistemas heredados debe ser automática y transparente preservando la seguridad e integridad de la información.

1.2 IDENTIFICACIÓN DE PERSPECTIVA DE NEGOCIO

La estrategia corporativa es la fuerza que rige la formulación de la estrategia informática. Con este objetivo se identifican los procesos de negocio que tienen mayores posibilidades de lograr una ventaja estratégica significativa sobre la competencia, mediante el uso de sistemas informáticos.

Del análisis de las perspectivas, objetivos y estrategias de negocio, establecidos en el Plan Estratégico de ##### 2018 – 2021, se concluye que las áreas de ventaja estratégica para el fortalecimiento de los servicios de TI se encuentran enmarcados en los procesos relacionados con la cadena de valor y que constituyen todos aquellos que están asociados con la prestación de servicios y productos financieros a los clientes, en el marco de las siguientes perspectivas:

- Inclusión de la microempresa, pequeños y medianos productores y sus organizaciones
- Servicios Financieros innovadores y eficientes
- Mejora de las condiciones de vida de los clientes atendidos

Por otra parte, las tendencias tecnológicas bancarias están orientadas a:

- Mejorar la posición competitiva de las entidades en el mercado mediante la oferta de productos y servicios en múltiples canales
- Profundizar la relación con el cliente diferenciándose por la atención.
- Estandarizar los procesos y reducir costos
- Adoptar los cambios del mercado, organizacionales y regulatorios de manera ágil

Anexo V Modelo de la encuesta aplicada

INCONVENIENTES DE CALIDAD DE DATOS

El objetivo de la encuesta tiene como objetivo identificar los inconvenientes de calidad de datos que afronta de la entidad financiera, el resultado servirá de insumo para la especificación y diseño de la evaluación.

El desarrollo de la encuesta no llevara más de 10 minutos, se pide responder con la mayor sinceridad, subraye o encierre el literal que se ajuste a tu criterio.

DATOS GENERALES

¿Cuál es su rol dentro la entidad financiera?

- Desarrollo de aplicaciones
- Administrador de Base de Datos
- Usuario final

EXACTITUD		
Indica que tan precisos, válidos y libres de errores están los datos.		
EJEMPLOS		
¿Los datos del sistema de información corresponden con la realidad?	Datos de un estudiante	Estudiante inexistente. Estudiante equivocado. Correcto pero con algunos errores.
¿Los datos del sistema de información brindan el suficiente detalle?	No contiene suficiente detalle	Salario: \$10.000 vs. \$10.014 vs. \$10.013,88 Fecha: 1977 vs. julio de 1977 vs. 14/7/1977 vs. 14/7/1977 10:55:32.4 Color: Rojo vs. 204R-51G-0B Cabello: Castaño vs. Castaño claro cobrizo n° 5 Dirección: J.Herrera y Reissig 565, 11300, Montevideo vs. Montevideo

1.- ¿En qué medida los datos que utiliza, se ajustan a su valor real?

- a.- Nunca
- b.- En alguna ocasión
- c.- Frecuentemente
- d.- Muchas ocasiones

Comentarios

COMPLETITUD		
La completitud indica si el sistema de información contiene toda la información de interés.		
EJEMPLOS		
¿Cuánta información se tiene sobre las entidades en el sistema de información?	Varias interpretaciones de la falta de valores (nulos)	Existen pero no los conozco (ej. No conozco el teléfono de Raquel).
		Porque no existe (ej. Raquel no tiene teléfono).
		No sé si existe (ej. No sé si Raquel tiene teléfono).

2.- ¿En qué medida los datos que utiliza presenta toda la información de su interés?

- a.- Nunca
- b.- En alguna ocasión
- c.- Frecuentemente
- d.- Muchas ocasiones

Comentarios

FRESCURA		
La completitud indica qué tan viejos son los datos.		
EJEMPLOS		
¿Qué tan vigentes son los datos del sistema de información?	Interesa medir el desfase entre los datos.	Datos de clientes (direcciones, teléfonos,...)
		Saldo bancarios (casa central, sucursales, estados de cuentas, ...)

3.- ¿En qué medida los datos que utiliza son vigentes?

- a.- Nunca
- b.- En alguna ocasión
- c.- Frecuentemente
- d.- Muchas ocasiones

Comentarios

CONSISTENCIA		
La consistencia captura la satisfacción de reglas semánticas definidas sobre los datos.		
EJEMPLOS		
Interesa medir qué tan bien se satisfacen las reglas de integridad.	Satisfacción de reglas sobre el contenido de un atributo.	Ej. Edad entre 0 y 120 años.
	Satisfacción de reglas entre atributos de una misma tabla.	Dependencias de clave y de unicidad.
		Dependencias de valores. Ej. Edad = Year (now() – FechaNacimiento).
	Satisfacción de reglas entre atributos de varias tablas.	Expresiones condicionales (edits). Ej. EstadoCivil = casado Edad ≥ 14.
Dependencias de inclusión (clave foránea, integridad referencial).		

4.- ¿En qué medida los datos que utiliza presentan consistencia?

- a.- Nunca
- b.- En alguna ocasión
- c.- Frecuentemente
- d.- Muchas ocasiones

Comentarios

UNICIDAD		
La unicidad indica el nivel de duplicación entre los datos.		
EJEMPLOS		
Hay duplicación si la misma entidad aparece repetida en forma exacta.	Interesa medir la cantidad de repetidos	Los valores de la clave y los atributos coinciden (o son nulos en algunas tuplas)
Hay contradicción si la misma entidad aparece repetida con contradicciones		Ej. 1.234.567-8, A. Sosa, 25 años, casado, 1.234.567-8, A. Sosa, NULL, NULL
		Los valores de la clave pueden coincidir o no
		Hay diferencias en valores de algunos atributos (no nulos)
	Ej. 1.234.567-8, A. Sosa, 25 años, casado, 1.234.567-6, Andrés Sosa, 24 años, NULL	

5.- ¿En qué medida los datos que utiliza presentan duplicación?

- a.- Nunca
- b.- En alguna ocasión
- c.- Frecuentemente
- d.- Muchas ocasiones

Comentarios

DOCUMENTACIÓN

Suficiente información gráfica o textual actualizada, sobre la estructura de la BD y de sus objetos.
--

6.- ¿En qué medida mantiene o crea documentación de las BD que utiliza?

- a.- Nunca
- b.- En alguna ocasión
- c.- Frecuentemente
- d.- Muchas ocasiones