**Universidad Nacional Mayor de San Marcos**
**Universidad del Perú. Decana de América**
Facultad de Ingeniería de Sistemas e Informática
Escuela Profesional de Ingeniería de Sistemas

# Proyecto de migración de datos hacia un Data Lake para una entidad de seguros

**TRABAJO DE SUFICIENCIA PROFESIONAL**

Para optar el Título Profesional de Ingeniero de Sistemas

**AUTOR**

Dante Wenceslao RAMOS CÓRDOVA

**ASESOR**

Arturo Alejandro BARTRA MORE

Lima, Perú

2022

**Referencia bibliográfica**

Ramos, D. (2022). *Proyecto de migración de datos hacia un Data Lake para una entidad de seguros*. [Trabajo de suficiencia profesional de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Escuela Profesional de Ingeniería de Sistemas]. Repositorio institucional Cybertesis UNMSM.

## Metadatos complementarios

| | |
|---|---|
| **Datos de autor** | |
| Nombres y apellidos | DANTE WENCESLAO RAMOS CÓRDOVA |
| Tipo de documento de identidad | DNI |
| Número de documento de identidad | 72739836 |
| URL de ORCID | https://orcid.org/0000-0002-6366-7270 |
| **Datos de asesor** | |
| Nombres y apellidos | Arturo Alejandro Bartra More |
| Tipo de documento de identidad | DNI |
| Número de documento de identidad | 40233946 |
| URL de ORCID | https://orcid.org/**0000-0002-3411-7456** |
| **Datos del jurado** | |
| **Presidente del jurado** | |
| Nombres y apellidos | Jorge Santiago Pantoja Collantes |
| Tipo de documento | DNI |
| Número de documento de identidad | 06254022 |
| **Miembro del jurado 1** | |
| Nombres y apellidos | Raúl Marcelo Armas Calderón |
| Tipo de documento | DNI |
| Número de documento de identidad | 07156168 |
| **Datos de investigación** | |
| Línea de investigación | No aplica |
| Grupo de investigación | No aplica |
| Agencia de financiamiento | Financiamiento Propio |
| Ubicación geográfica de la investigación | Perú, Lima, Lima,<br>Psj. Felipe Alvarado Mz. 24 Lte. 25 – Urb. Previ – Los Olivos<br>Latitud: -11.981607<br>Longitud: -77.067262 |

| | |
|---|---|
| Año o rango de años en que se realizó la investigación | 2021 |
| URL de disciplinas OCDE | 2.02.04 -- Ingeniería de sistemas y comunicaciones<br>https://purl.org/pe-repo/ocde/ford#2.02.04 |

***UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS***
***FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA***
***Escuela Profesional de Ingeniería de Sistemas***

***Acta Virtual de Sustentación***
***del Trabajo de Suficiencia Profesional***

*Siendo las 19:57 horas del día 07 de enero del año 2022, se reunieron virtualmente los docentes designados como Miembros de Jurado del Trabajo de Suficiencia Profesional, presidido por el Lic. Pantoja Collantes Jorge Santiago (Presidente), Ing. Armas Calderón Raúl Marcelo (Miembro) y el Ing. Bartra More Arturo Alejandro (Miembro Asesor), usando la plataforma Meet (https://meet.google.com/jjy-yahj-fza), para la sustentación virtual del Trabajo de Suficiencia Profesional intitulado:* ***"PROYECTO DE MIGRACIÓN DE DATOS HACIA UN DATA LAKE PARA UNA ENTIDAD DE SEGUROS",*** *por el Bachiller* ***Ramos Córdova Dante Wenceslao;*** *para obtener el Título Profesional de Ingeniero de Sistemas.*

*Acto seguido de la exposición del Trabajo de Suficiencia Profesional, el Presidente invitó al Bachiller a dar las respuestas a las preguntas establecidas por los miembros del Jurado.*

*El Bachiller en el curso de sus intervenciones demostró pleno dominio del tema, al responder con acierto y fluidez a las observaciones y preguntas formuladas por los señores miembros del Jurado.*

*Finalmente habiéndose efectuado la calificación correspondiente por los miembros del Jurado, el Bachiller obtuvo la nota de* ***18 DIECIOCHO.***

*A continuación, el Presidente de Jurados el Lic. Pantoja Collantes Jorge Santiago, declara al Bachiller* ***Ingeniero de Sistemas****.*

*Siendo las 20:41 horas, se levantó la sesión.*

***Presidente***
*Lic. Pantoja Collantes Jorge Santiago*

***Miembro***
*Ing. Armas Calderón Raúl Marcelo*

***Miembro Asesor***
*Ing. Bartra More Arturo Alejandro*

**DEDICATORIA:**

A mi familia por su apoyo incondicional y consejos brindados para lograr mis objetivos.

**AGRADECIMIENTO:**

A la Universidad Nacional Mayor de San Marcos, que me brindo los conocimientos necesarios.

A las empresas en las que pude desarrollarme profesionalmente.

A mi asesor, por el tiempo y dedicación brindada para lograr el objetivo de culminar el presente trabajo.

**UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS**

**FACULTAD DE INGENIERIA DE SISTEMAS E INFORMÁTICA**

**ESCUELA PROFESIONAL DE INGENIERIA DE SISTEMAS**

**Proyecto de migración de datos hacia un Data Lake para una entidad de seguros**

Autor: Ramos Córdova Dante Wenceslao

Asesor: Bartra More Arturo Alejandro

Título: Trabajo de Suficiencia Profesional para Optar por el Título Profesional de Ingeniero de Sistemas

Fecha: Enero 2022

---

## RESUMEN

El presente trabajo de suficiencia profesional, describe el proyecto de implementación de un Data Lake en una empresa del sector Asegurador, como una solución analítica de Big Data. Debido al gran aumento de clientes y datos, se tuvieron dificultades dentro de las Bases de Datos de la entidad. Tales como lentitud de procesos, bloqueo de tablas productivas por los tiempos de procesamiento, y la generación de reportes poco confiables. A la vez que se requería ofrecer campañas 100% personalizadas para los clientes, y para esto debía de integrarse la información del cliente que se encontraba dispersa en los múltiples aplicativos y tablas de la entidad Aseguradora. Es por ello que se optó por implementar un repositorio único de datos que será utilizado como una fuente de cálculo del CLV (Customer Lifetime Value[1]), obteniendo el valor de vida de cada cliente, de esta manera poder ofrecer productos específicos y/o promociones para fidelizar al consumidor final.

**Palabras clave:** Data lake, Migración de datos, SCRUM, Big data, Datastage.

---

[1]Según (We are Marketing, 2020) indica: “El Customer Lifetime Value o valor del tiempo de vida de un cliente es un pronóstico sobre la cantidad de dinero que espera recibir la empresa por parte del usuario, durante todo el tiempo en que siga siendo su cliente.”

**NATIONAL MAJOR UNIVERSITY OF SAN MARCOS**

**FACULTY OF SYSTEMS ENGINEERING AND INFORMATIC**

**PROFESSIONAL SCHOOL OF SYSTEMS ENGINEERING**

**Data migration project to a Data lake for an insurance company**

Author: Ramos Córdova Dante Wenceslao

Advisor: Bartra More Arturo Alejandro

Title: Professional Sufficiency Work for opt for the Professional Title of Systems Engineer

Date: January 2022

## ABSTRACT

The present work of professional sufficiency, describes the project of implementation of a Data Lake in a company of the Insurance sector, as a Big Data analytical solution. Due to the large increase in customers and data, difficulties were encountered within the entity's databases. Such as slow processes, blocking of productive tables due to processing times, and the generation of unreliable reports. At the same time, it was required to offer 100% personalized campaigns for clients, and for this the client information that was dispersed in the multiple applications and tables of the Insurance entity had to be integrated. That is why it was decided to implement a single data repository that will be used as a source for calculating the CLV (Customer Lifetime Value[2]), obtaining the lifetime value of each customer, in this way to be able to offer specific products and / or promotions for build loyalty to the end consumer.

**Key words:** Data lake, Data migration, SCRUM, Big data, Datastage.

[2] The Customer Lifetime Value or value of the customer's lifetime is a forecast on the amount of money that the company expects to receive from a user, for as long as this user continues to be its customer.

ÍNDICE GENERAL

# INDICE DE TABLAS

INDICE DE FIGURAS

# INTRODUCCIÓN

El presente trabajo de suficiencia profesional nos describe cómo se abordó la migración de datos de diferentes fuentes hacia un único repositorio centralizado, brindando la solución a una empresa aseguradora del mercado peruano.

La empresa de seguros mencionada es una de las principales aseguradoras dentro del país, contando con una amplia cartera de productos para los distintos tipos de consumidores. Funciona como un respaldo frente a situaciones de riesgo en las que podría verse involucrado el cliente.

Al ser una de las empresas líderes del sector asegurador y pertenecer a un grupo organizacional, que se encuentra en una etapa de innovación tecnológica, conlleva a alinearse a los objetivos y metas de dicho grupo.

El presente trabajo se divide en 5 capítulos:

En el capítulo I; se detalla la experiencia laboral, formación académica recibida y estudios complementarios pertenecientes al autor del informe.

En el capítulo II se menciona detalles importantes sobre la empresa en donde el autor adquirió la experiencia, tanto como misión, visión, organigrama; también las actividades que realizó.

En el capítulo III se especifica la problemática encontrada, la solución desarrollada, los objetivos cumplidos y etapas del desarrollo para implementar la solución.

En el capítulo IV se hace una reflexión sobre cómo se abordó la solución, si fue la más adecuada, qué fue lo más importante a destacar de la experiencia adquirida.

En el capítulo V se mencionan las conclusiones encontradas por el autor, recomendaciones brindadas que pueden servir de ayuda al lector o interesados en el proyecto.

## CAPÍTULO I: TRAYECTORIA PROFESIONAL

El autor actualmente tiene el grado de Bachiller en la carrera de Ingeniería de Sistemas e Informática en la Universidad Nacional Mayor de San Marcos, con un diplomado de Alta Especialización en Inteligencia de Negocios en la ESAN. Presenta facilidad de trabajar en equipo, buenas relaciones interpersonales.

Además, alta capacidad de análisis de información y experiencia en la integración de datos. Manejo de grandes volúmenes de información en proyectos ETL (Extracción, Transformación y Carga por sus siglas en inglés). Participación en el proyecto de Migración en Telefónica Perú hacia el nuevo repositorio centralizado, utilizando la herramienta IBM Datastage para realizar los cruces de información y generación de Validaciones de Integración (Integrity Checks), asegurando la calidad de los datos.

Además, en la consultora Teamsoft participando en la migración de datos a un Data lake, consolidando la información del Datawarehouse. Utilizando como el entorno de trabajo Hadoop, almacenando la información en archivos HDFS (Sistema de ficheros distribuido de Hadoop) y el motor de consultas SQL de Cloudera, para el procesamiento masivo en paralelo.

*Tabla 1: Formación Académica Profesional*

| FORMACIÓN RECIBIDA | INSTITUCIÓN | PERIODO |
|---|---|---|
| Grado Académico de Bachiller en Ingeniería de Sistemas | Universidad Nacional Mayor de San Marcos - Facultad de Ingeniería de Sistemas e Informática | 2011-2016 |

Nota: Elaboración propia

*Tabla 2: Formación Académica Complementaria*

| FORMACIÓN RECIBIDA | INSTITUCIÓN | PERIODO |
|---|---|---|
| Diplomado de Alta Especialización en Business Intelligence | Universidad ESAN | Setiembre 2017 - Junio 2018 |
| Programa de Especialización Big Data | Big Data Academy Perú | Setiembre 2019 |

Nota: Elaboración propia

*Tabla 3: Experiencia Laboral B89*

| B89<br>Octubre 2021 - Actualidad | |
|---|---|
| Cargo | Data Engineer |
| Funciones | Análisis, diseño y desarrollo del modelo de datos para nuevas funcionalidades del aplicativo. Apoyo en tickets de Customer Support, para resolución de problemas con clientes en productos financieros. Uso de herramientas de integración continua Jenkins y Bitbucket para realizar los pases a producción. |

Nota: Elaboración propia

*Tabla 4: Experiencia Laboral Bluetab – BCP*

| Bluetab Perú<br>Marzo 2021 - Actualidad | |
|---|---|
| Cargo | Data Engineer |
| Cliente | BCP |
| Funciones | Realizar mapeo de procesos en Oracle, elaboración de análisis de Impacto para tablas y campos según requerimientos de usuario. Apoyo en revisión de trazabilidad de información, hasta la fuente origen del mismo. |

Nota: Elaboración propia

*Tabla 5: Experiencia Laboral Teamsoft*

| Teamsoft<br>Agosto 2018 - Febrero 2021 | |
|---|---|
| Cargo | Data Engineer |
| Cliente | Empresa Aseguradora |
| Funciones | Diseño, desarrollo, mantenimiento y pruebas del procesamiento de datos desde su origen hasta su carga en el CLV. Carga de entidades a Cloudera HUE en tablas Hive e Impala. |

Nota: Elaboración propia

*Tabla 6: Experiencia Laboral Everis Perú*

| Everis Perú | |
|---|---|
| Cargo | Analista ETL |
| Cliente | Telefónica del Perú - Movistar |

| | |
|---|---|
| Funciones | Elaboración de jobs paralelos, secuenciales, a partir de cruces de información necesarios. Gestionar el proceso de Extracción y Carga a DataStage de las tablas que contienen información de los clientes proporcionados por Telefónica |

Nota: Elaboración propia

# CAPÍTULO II
# CONTEXTO EN EL QUE SE DESARROLLÓ LA EXPERIENCIA

## 2.1 Empresa – Actividad que realiza

En su página web, TeamSoft (TeamSoft, 2021) manifiesta:

Es una empresa nacional conformada por accionistas peruanos que reúnen más de 20 años de experiencia nacional e internacional en el campo de la tecnología de la información, brindando soluciones de valor agregado al negocio en empresas del sector privado y público del país. Para el logro de este objetivo, cuenta con consultores y analistas de la más alta calidad y experiencia en desarrollo de soluciones, bajo una adecuada y estricta metodología. TeamSoft se fundó para garantizar que las soluciones brindadas estén orientadas a la integración con el Cliente y a su total satisfacción.

## 2.2 Visión

De acuerdo a la Visión de (Teamsoft, 2021), manifiesta:

Ser reconocidos por nuestros clientes y por el mercado informático peruano como el mejor Socio Tecnológico, fundamentando la relación con nuestros clientes en confianza mutua. La actitud de servicio y compromiso de nuestro personal, refleja la creatividad, conocimiento y profesionalismo para resolver las necesidades y problemas de nuestros clientes como si fueran propios.

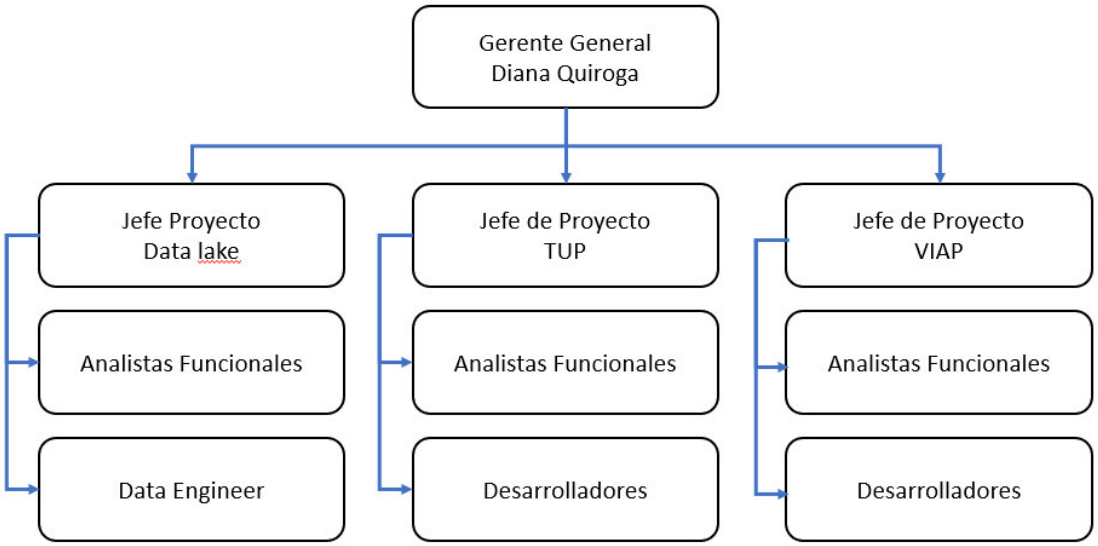
## 2.3 Misión

De acuerdo a la Misión de (Teamsoft, 2021), manifiesta:

Nuestra misión es aplicar inteligentemente todos nuestros recursos en la obtención de soluciones basadas en tecnología de la información para el desarrollo sostenido del sector empresarial de nuestra sociedad. Como resultado indirecto de nuestra misión buscamos el desarrollo de nuestro personal en una línea de principios y ética que contribuya al correcto desarrollo de nuestro país.

## 2.4 Organización de la Empresa

*Figura 1: Organigrama Organizacional*




**Nota: Elaboración propia**[3]

## 2.5 Área, Cargo y funciones desempeñadas

El autor de este informe de experiencia profesional se desempeñó como Data Engineer en el Proyecto del CLV.

Las funciones desempeñadas están alineadas a lo siguiente:

- Diseño, desarrollo, mantenimiento y pruebas del procesamiento de datos desde su origen hasta su carga en el data lake.
- Carga de entidades a Cloudera HUE en tablas Hive e Impala.
- Desarrollo de procesos de transformación y limpieza de Datos.
- Apoyo y desarrollo de mapeos de las entidades utilizadas.
- Realizar validaciones de los procesos desarrollados en el entorno.

## 2.6 Experiencia profesional realizada en la organización

El autor del presente trabajo, dentro de su experiencia laboral en la empresa TeamSoft Perú S.A ha sido parte del equipo de migración de datos a un Data

[3] Para este proyecto, el suscrito se encontraba en el Proyecto CLV.

lake, para consolidar la información de clientes. Elaborando procesos para la migración de datos de clientes con la herramienta ETL IBM Datastage, preparación de ambientes de prueba y posterior carga de tablas en Impala, bajo el sistema de archivos distribuidos de Hadoop (HDFS).

# CAPÍTULO III – ACTIVIDADES DESARROLLADAS

## 3.1 Situación Problemática

### 3.1.1 Definición del problema

La empresa forma parte de un grupo organizacional, quienes han iniciado la implementación y explotación de las tecnologías Big Data; por lo que se le ha invitado a formar parte de esta ola de innovación tecnológica.

Dentro de la empresa Aseguradora; con el pasar de los años, se está teniendo un gran aumento de información de clientes y datos en general. Debido a esto la organización presenta problemas en tiempos de procesamiento, generación y sobre todo el almacenamiento de la data.

Las ejecuciones realizadas durante las noches; tomaba más tiempo del debido; como consecuencia de lo antes expuesto, al día siguiente existía un bloqueo de tablas productivas, la generación de reportes a destiempo y problemas en la toma de decisiones por tales motivos.

## 3.2 Solución

El presente informe de Suficiencia Profesional busca resolver la problemática planteada líneas arriba, mediante la migración de datos con orígenes en tablas SQL Server y Oracle, hacia un repositorio centralizado. De manera que se consolida la información en un Data lake cumpliendo con los siguientes objetivos detallados a continuación.

### 3.2.1 Objetivos

**Objetivo General:**

Gestionar la migración de información hacia un Data Lake que centralice la información de distintos orígenes, como un almacén de datos escalable para datos estructurados de base de datos relacionales, semiestructurados (Xml) y no estructurados (documentos, imágenes, audio), permitiendo la ingesta y recuperación de datos.

El data lake funcionará como repositorio con una diversidad de data que sirva en un futuro cercano para análisis predictivos sobre la base de clientes y pólizas.

Y de esta manera, lograr un buen performance en tiempos de consulta de datos para la generación de reportes.

**Objetivos específicos**

- Implementar una nueva arquitectura informacional por capas.
- Implementar un repositorio centralizado de datos.
- Realizar la extracción de información mediante la herramienta IBM Datastage para la ingesta de datos desde las fuentes origen hacia un repositorio centralizado para la consolidación y estructuración de datos.
- Reducir los tiempos de consulta de datos

### 3.2.2 Alcance

Alcance Funcional:

El alcance del proyecto consiste en la migración tecnológica de la información sin pérdida de datos entre los sistemas origen SQL Server y Oracle, hacia el Data lake sobre tablas Impala. Dicha migración se realizará con la herramienta IBM Datastage.

Se requiere que el modelo del Data lake; contenga la siguiente información para el CLV.

*Tabla 7: Entidades que intervienen en el proceso de migración*

| NRO | ENTIDADES |
|---|---|
| 1 | Pólizas |
| 2 | Pagos |
| 3 | Siniestros |
| 4 | Intermediarios |
| 5 | Asegurados |
| 6 | Beneficiarios |
| 7 | Familiares |
| 8 | Afiliados |
| 9 | Asistencias |
| 10 | Encuestas |

Nota: Elaboración propia

Alcance Organizacional:

El presente proyecto se implementó en el área tecnológica de la Aseguradora, quién solicitó a su proveedor (Teamsoft), la puesta en marcha del proyecto y desarrollo de la solución beneficiando a las áreas usuarias de Vida y Seguros.

### 3.2.3 Etapas y Metodología

Para realizar la migración, se trabajó bajo la metodología SCRUM como marco de trabajo. Ya que permitió realizar entregables parciales del producto final, una estrategia de desarrollo incremental que se da en ciclos cortos de tiempo con duración de 2 semanas cada uno.

Las etapas del proyecto se detallarán a continuación:

Gestión del proyecto: La aseguradora al formar parte del grupo organizacional, debió seguir los lineamientos y decidió por realizar la migración. En esta etapa se escucharon propuestas de distintos proveedores; siendo la elegida la consultora Teamsoft, en la cual estuvo participando el autor del presente informe.

Definición de Requerimientos: En conjunto con el cliente, se establecieron cuáles serían las funcionalidades y requisitos que debían implementarse y/o desarrollar. Se establecieron las tablas y campos a extraer para la primera capa del Data lake, tablas de equivalencias que se usaron para la homologación y limpieza de datos.

Diseño: Se modelaron las tablas finales en la capa final que serán pobladas para el Data lake, se estableció la ruta de ejecución desde la extracción de las fuentes de bases de datos tradicionales, el poblamiento de las capas de extracción y procesamiento de datos.

Despliegue: Se desarrolló la migración de datos, siguiendo el plan de ejecución establecido en la etapa anterior. La herramienta utilizada para la extracción y preparación de la data fue IBM Datastage. Se carga la data desde tablas contenidas en SQL Server y Oracle a las capas definidas en la arquitectura del Data lake. El almacenamiento de la data fue en archivos distribuidos HDFS, y para realizar la consulta de datos se utilizó un Motor SQL Impala, para poder acceder a las tablas y que lo usuarios puedan generar reportes o visualizar la información contenida.

### 3.2.4 Fundamentos Utilizados

A continuación, se detallarán ciertas definiciones que servirán para comprender de una mejor manera la implementación del presente informe de Suficiencia Profesional.

3.2.4.1 ETL: es un proceso informático para integrar datos y consta de 3 fases, las cuales comprenden de Extracción - Transformación - Carga (por sus siglas en inglés 'Load'). En la primera fase se extraen los datos desde los sistemas fuentes desde donde se leerá la información, en un próximo paso se transformarán bajo ciertas reglas de negocio o casos de uso que se requieran para limpiar y homologar la data, depurando inconsistencias o data duplicada que pudiera existir. Una vez que se tiene consolidada la data, pasa por un proceso de carga hacia un repositorio de datos destino.

Se considera de importancia este proceso ETL, debido a que permite la extracción no sólo de una, sino múltiples sistemas origen a la vez. Convirtiéndola en un componente de integración propiamente dicho en una organización que cuenta con varias áreas, varios sistemas con distintas bases de datos y conectores.

De la misma manera, también permite la carga hacia múltiples destinos, mejorando la eficiencia de los ingenieros de datos al permitirle conexión a distintos destinos con una herramienta gráfica que, configurada correctamente, evitará que se requieran habilidades muy técnicas sobre programación de scripts.

3.2.4.2 Big data:

Según (Power Data, 2021), Big Data es un término que describe el gran volumen de datos, tanto estructurados como no estructurados, que inundan los negocios cada día. Pero no es la cantidad de datos lo que es importante. Lo que importa con el Big Data es lo que las organizaciones hacen con los datos. Big Data se puede analizar para obtener ideas que conduzcan a mejores decisiones y movimientos de negocios estratégicos.

Según (Instituto de Ingenieria del conocimiento, 2019) nos relata que, las características más resaltantes se clasifican por magnitudes.

- Volumen: se refiere a la cantidad de datos que son generados cada segundo, minuto y días en nuestro entorno. Es la cualidad que refleja claramente lo que es Big Data, se refiere a la gran cantidad de datos que almacenan con el propósito de procesar toda la información y generar valor para la toma de decisiones.
- Velocidad: debido a la gran cantidad de datos que se almacenan, es indispensable que exista un dinamismo rápido para la creación, almacenamiento y próximamente consultas de los datos.
- Variedad: nos habla sobre la diversidad de los datos, tanto como formas, tipos y diferentes orígenes de las que proviene el dato. Pueden ser estructurados, semiestructurados y no estructurados.
- Veracidad: se refiere a la fiabilidad del dato, ya que, al tener múltiples fuentes, debemos asegurar la calidad del dato y que sea una fuente confiable de información.
- Valor: Los datos por si solos no nos representan un valor. Éste se obtiene de los datos que se transforman en información, y que luego de procesarla se transforma en conocimiento. El valor de los datos se centra en que nos sirva como un plan de acción para la toma decisiones.

#### 3.2.4.3 Scrum:

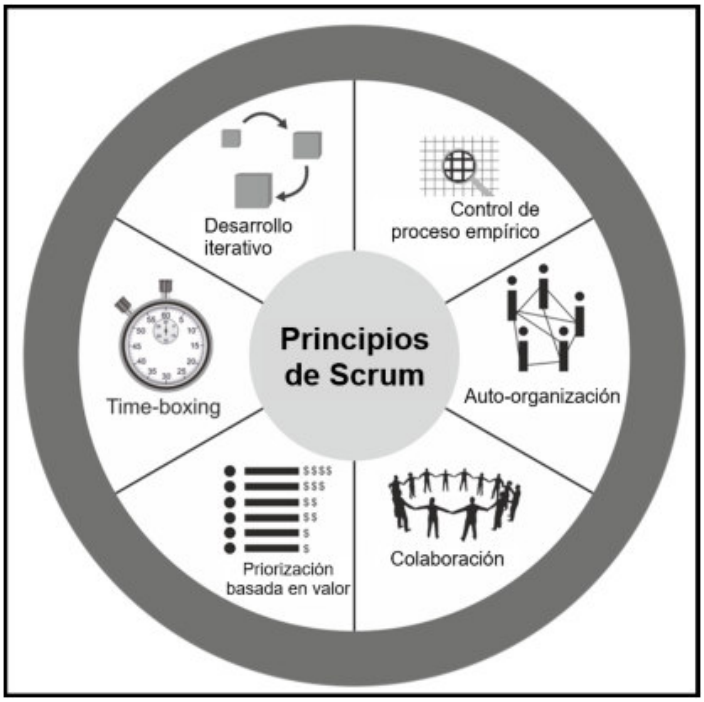
Según (Atlassian, 2021), nos define como un marco de trabajo colaborativo entre equipos. Al igual que un equipo de rugby (de donde proviene su nombre) cuando entrena para un gran partido, scrum anima a los equipos a aprender a través de las experiencias, a autoorganizarse mientras aborda un problema y a reflexionar sobre sus victorias y derrotas para mejorar continuamente.

Según (SCRUMstudy, 2018), menciona que, los principios del Scrum son las pautas básicas para aplicar este marco de trabajo y deben implementarse de forma obligatoria en los proyectos.

- Control de proceso empírico: se basa en 3 ideas principales de transparencia, inspección y adaptación.

- Auto-organización: los equipos poseen un gran sentido de compromiso y responsabilidad, teniendo una organización propia.
- Colaboración: fomenta la gestión de proyectos como un proceso de creación de valor compartido con equipos que trabajan e interactúan conjuntamente.
- Priorización basada en valor: colocar en el foco las actividades que ofrezcan el máximo valor para el negocio.
- Time-boxing: describe que se considera como restricción al tiempo, ayudando a planificar de manera eficaz. Se consideran en este principio a los sprints, dailys, Reuniones de Planificación y las Review de Sprint.
- Desarrollo iterativo: define un desarrollo incremental y se enfoca en organizar de manera eficaz los cambios o crear entregables que sean de valor para el cliente.

Figura 2: Principios del Scrum

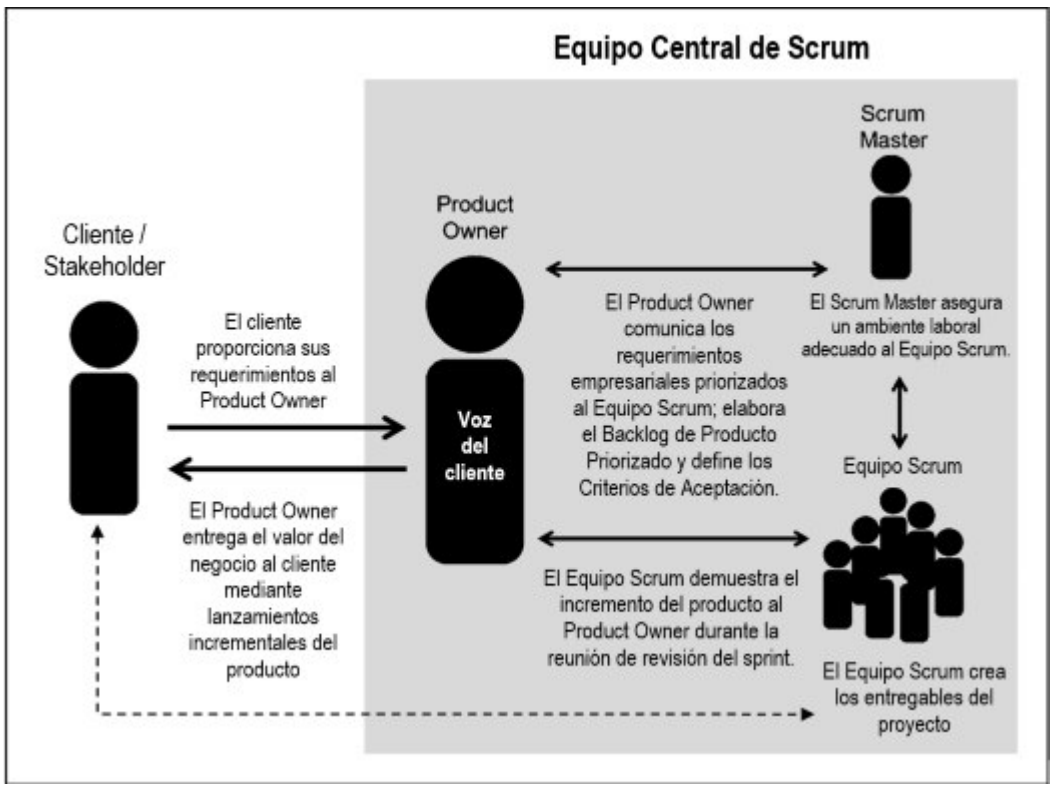



Nota: Extracto de SCRUMstudy-SBOK-Guide-3rd-edition-Spanish.pdf

Los roles que se incluyen este marco de trabajo:

- Product Owner: persona responsable de lograr el máximo valor empresarial para el proyecto. Establece los requerimientos del cliente, y representa al cliente dentro del proyecto.
- Scrum Master: facilita que el equipo cuente con un ambiente propicio para lograr completar el proyecto satisfactoriamente. Elimina los posibles impedimentos y supervisa que se estén cumpliendo los procesos de Scrum.
- Equipo Scrum: grupo de personas que tienen la responsabilidad de comprender los requerimientos del cliente y crear los entregables del proyecto.
- Stakeholders: agrupa a los clientes y usuarios, que influyen en el proyecto durante la implementación. Son las personas a las que va a beneficiar el desarrollo del mismo.

Figura 3: Estructura organización Scrum




Nota: Extracto de SCRUMstudy-SBOK-Guide-3rd-edition-Spanish.pdf

3.2.4.4 Customer Lifetime Value - CLV (Valor de un cliente en un periodo de tiempo)

Según (We are Marketing, 2020) es una métrica que sirve a las empresas para determinar el valor o ganancia que representa un cliente durante un determinado periodo de tiempo. Es una medida muy importante para conocer a nuestro cliente, saber su comportamiento dentro del negocio, establecer el valor que él nos aporta en la organización y así poder crear una campaña totalmente personalizada para el cliente. Algunos valores que podemos considerar para calcular el CLV pueden ser: Valor de sus compras, porcentaje de clientes que hacen efectivas las compras, tamaño de sus redes sociales (seguidores, actividad en redes), porcentaje de clientes que ofrecen una opinión sobre la empresa, etc.

Esto permite a las organizaciones a ofrecer un mejor servicio/producto y lograr la fidelización de clientes con acciones rentables para la empresa.

### 3.2.5 Implementación de las áreas de proceso

#### 3.2.5.1 Gestión

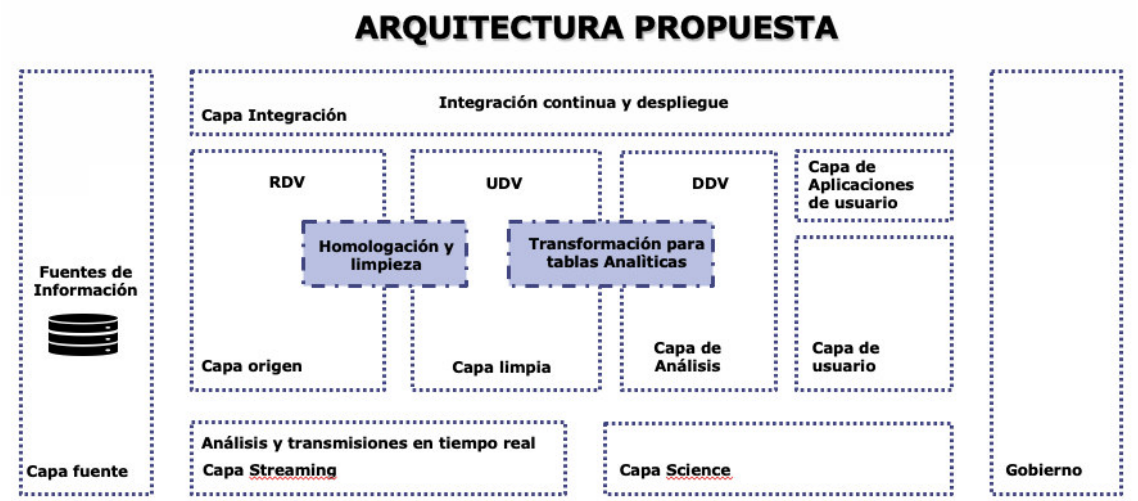
Se realizó una reunión entre el proveedor y la empresa Aseguradora, con el objetivo principal de definir una arquitectura que pueda soportar con garantía las futuras demandas de Big Data que el área usuaria pueda solicitar, haciéndola convivir con la plataforma actual para favorecer la inversión realizada de una manera modular y escalable que garantice el éxito de la implementación.

Se revisaron en conjunto con el cliente los objetivos del proyecto, el alcance, beneficios, y se estableció que la metodología a utilizar sería SCRUM.

En la reunión se indicó lo siguiente, la nueva arquitectura Data lake en el grupo organizacional cuenta con varias áreas de datos que permite la mejor gestión de las mismas.

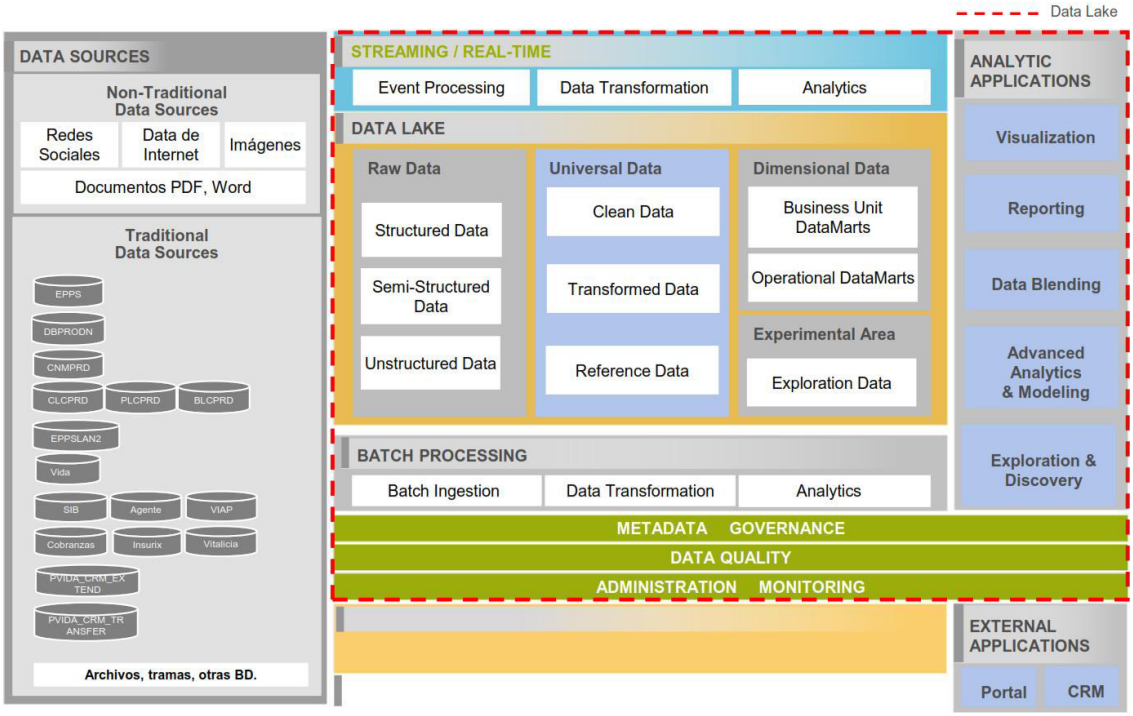
Por lo tanto, se definió que sea una arquitectura informacional, como se detalla en el gráfico:

*Figura 4: Arquitectura Propuesta*




Nota: Elaboración propia

*Figura 5: Arquitectura Informacional*



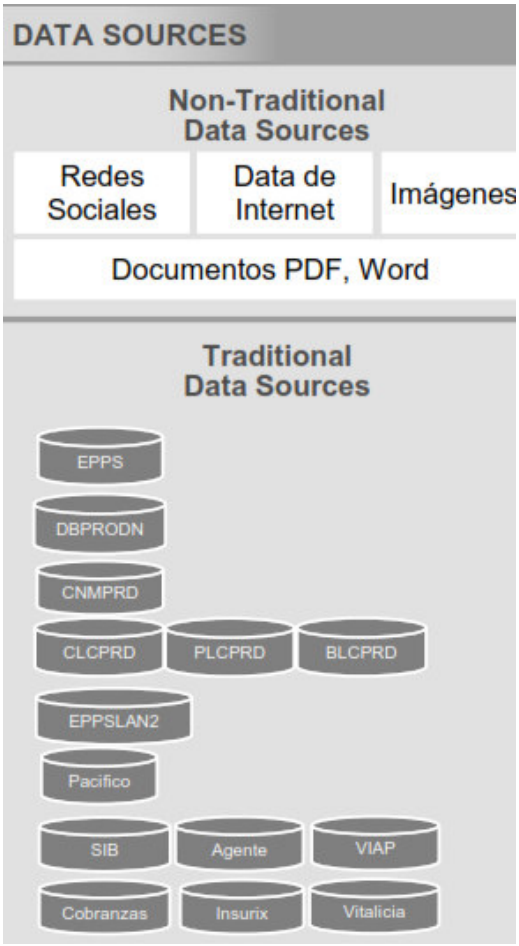

Nota: Extracto de la Propuesta de Arquitectura para el Data lake (Fuente: PGA – Modelo de Arquitectura To Be).

Ahora se dará mayor detalle sobre cada una de las capas.

- Capa Data Source: en esta capa se realiza la extracción de información que es requerida por los casos de uso mediante procesos de extracción para la ingesta de datos de las distintas fuentes hacia el Raw Data.

Se cargará inicialmente al Raw data según la periodicidad definida por cada caso de uso, según sea requerido se realiza una carga inicial la cual se estima maneja una gran cantidad de datos según la data histórica almacenada.

*Figura 6: Data Sources - Fuentes origen*

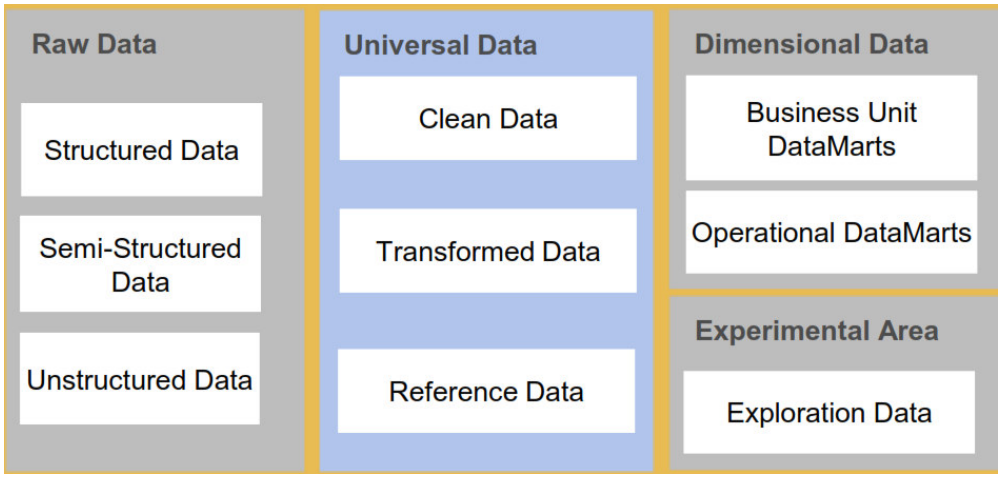



Nota: Extracto de la Propuesta de Arquitectura para el Data lake (Fuente: PGA – Modelo de Arquitectura To Be).

- Capas del Data Lake: se subdividen en Raw Data, Universal Data, Dimensional Data, Experimental Area y Traditional EDW.
    - Raw Data: es la entrada del Data Lake cercana a los sistemas de origen. La Raw Data almacena la información tal cual están en los distintos sistemas de origen, de ser necesario se puede aplicar una ligera transformación a los datos. Se le conoce como capa RDV.

- Universal Data: consume data de RDV, en esta capa los datos pasan por un proceso de transformación y calidad a fin de obtener información lista para el análisis. Conocida como capa UDV.
- Dimensional Data: En esta capa es donde se crean los Data Marts, en donde se permite el análisis de datos “Self-Service”, es usado también para la visualización, reportes y combinación de datos por usuarios finales. Conocida como capa DDV.
- Experimental Area: el área de experimentación será utilizada por los usuarios avanzados (expertos analíticos/ científico de datos) para descubrir nuevos conocimientos y experimentar con nuevos modelos/algoritmos analíticos avanzados.

*Figura 7: Capas del Data Lake*




Nota: Extracto de la Propuesta de Arquitectura para el Data lake (Fuente: PGA – Modelo de Arquitectura To Be).

- Capa Streaming/Real time: en esta capa está el procesamiento de la información que se captura en tiempo real y se tiene una latencia mínima, por lo general segundos o milisegundos.
    - Event Processing: captura los datos de la fuente de eventos (como por ejemplo sensores del clima, tráfico) en real time para su procesamiento, así pueda ser consumido y reflejado en el destino.

- Capa Batch Processing: esta capa está asociada a las ejecuciones programadas por batch [4]del plan de ingesta, transformación y analítica.
    - Batch Processing: ejecución de procesos de ingesta donde se extrae información de las fuentes de origen que prueban el Raw data.
    - Data Transformation: en esta etapa se ejecutan los procesos de limpieza de datos, de transformación y aplicación de reglas de negocio según el caso de uso. Se ubica en la capa del Universal Data.
    - Analytics: esta etapa se ubica en la capa "dimensional data" en la cual se encuentran modelos de datos según la necesidad del negocio para su explotación de usuarios finales, reportería y visualización. También se encuentra en la capa 'Experimental Área', donde se realizan tareas de analítica y ciencia de datos.
- Capa Analytic Applications: en esta capa se ubican las herramientas que permiten visualizar la información procesada en las capas mencionadas anteriormente. Ya sean reportes, tableros de control, etc.

También, en la arquitectura se tiene de forma transversal el gobierno de datos que es el proceso de definir las reglas que los datos deben seguir, y a la vez administrar y controlar el cumplimiento de estas reglas. Nos indica cómo el dato será guardado, cómo o con qué periodos de tiempo se realizarán los respaldos y cómo será protegido el dato. Define los responsables del dato, los accesos y permisos para usuarios autorizados.

En conjunto con el Gobierno de datos, se tiene la Calidad del dato. Porque nos permite implementar actividades técnicas para garantizar que los datos se ajusten a las necesidades de la organización. Es decir que no existan problemas de duplicidad de data, datos incompletos, datos inconsistentes, etc.

Además, se tomaron ciertas consideraciones previas de volumetría. El espacio efectivo para los nodos de información en el clúster es de 6.5 TB con un factor

[4] Batch: se conoce como sistema por lotes a la ejecución de un programa sin supervisión directa del usuario y utilizada para tareas repetitivas sobre grandes volúmenes de información. (Wikipedia, 2021)

de replicación de 3; sobre esta premisa se realizaron todas las estimaciones de volumetría.

La información con la que es poblado el Data Lake para Negocios de Vida, está conformada principalmente por la información perteneciente a los Administradores de Vida y otros aplicativos dentro de la empresa Aseguradora. El volumen de dicha información histórica a la fecha en que se realizó la evaluación oscila sobre los 7.1 TB de información a nivel de bases de datos y archivos relacionados al negocio.

Se estima un crecimiento anual aproximado de un 20%, y se presenta un ejercicio que muestra la cantidad de información y volúmenes en diversos periodos de tiempo, en base al cual se elabora el posible escenario de proyección.

*Tabla 8: Estimación de crecimiento histórico*

| **Volumen (TB)** | **Periodo** | |
|---|---|---|
| | **Meses** | **Años** |
| 7.22 | 1 | |
| 8.52 | 12 | 1 |
| 10.22 | 24 | 2 |
| 12.27 | 36 | 3 |
| 14.72 | 48 | 4 |
| 17.67 | 60 | 5 |
| 21.20 | 72 | 6 |

Nota: Elaboración propia

En el caso de la volumetría para Negocios de Riesgos Generales y Salud, está conformada por los aplicativos de Guidewire, Acsel/X, Novasys y otros aplicativos. El volumen en ese momento oscilaba sobre los 46.2 TB de información a nivel de bases de datos.

### 3.2.5.2 Definición de Requerimientos

Describiremos los requisitos de solicitados por el cliente, requerimientos funcionales y no funcionales.

A. Modelo del Negocio

Al momento de la evaluación, el sistema contaba con un repositorio de base de datos llamado 'Base Unificada', el cual se genera con un proceso mensual de carga, homologación y unificación de clientes naturales y pólizas activas, las mismas que provienen de las siguientes fuentes:

*Tabla 9: Sistemas que generan la Base Unificada*

| Compañía | Sistema | Consideraciones |
|---|---|---|
| **VIDA** | VIAP | Clientes naturales y sus pólizas activas, obtenidas del **ODS** de Aseguradora Vida. |
| **VIDA** | SAM | |
| **VIDA** | INSURIX | |
| **VIDA** | SAM MASIVO | Clientes naturales y sus pólizas activas obtenidas de **SAM Masivo**, que tengan el producto **Seguro Múltiple** y **Desgravamen** (Tarjetas Banco, Préstamos Bancarios y Multired BN). |
| **GENERALES** | AXCEL X | La información de los clientes naturales y sus pólizas activas, proporcionada por Generales, cuyo repositorio de información es una **base de datos** intermedia llamada UnificaClientes_PPS. |
| **GENERALES** | GUIDEWARE | |

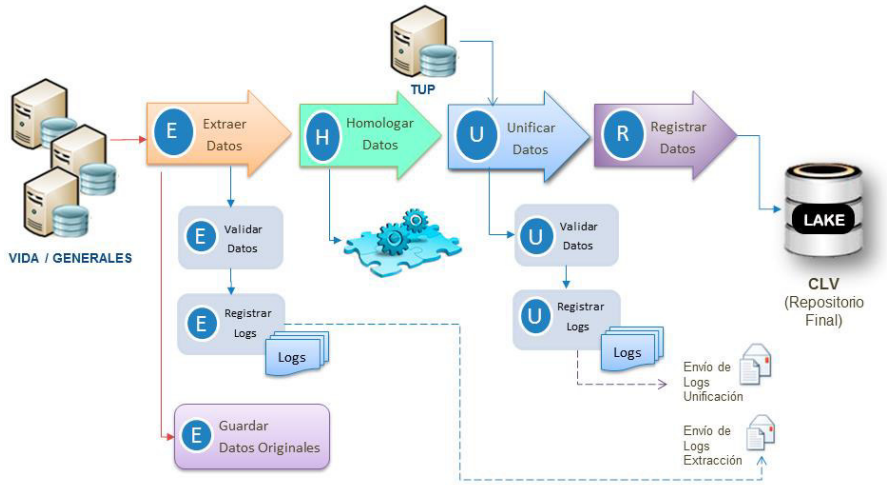
Nota: Elaboración propia

Asimismo, el área cuenta con un proceso mensual de generación de las bases de stock de Vida y Seguros Personales, para la creación de campañas de Sponsor y Tele Ventas.

Con la finalidad de que el área cuente con toda la información requerida para los procesos de gestión del cliente, se proporcionará un repositorio único de datos, que será la fuente para el cálculo del CLV (Customer Lifetime Value). El objetivo de este repositorio es contener la información homologada y unificada de cada cliente, así como de sus pólizas, pagos, siniestros,

interacciones, asistencias y encuestas traídas de distintos sistemas, enriquecida con información de pagos y transacciones financieras.

Figura 8: Diagrama de extracción, homologación y carga.



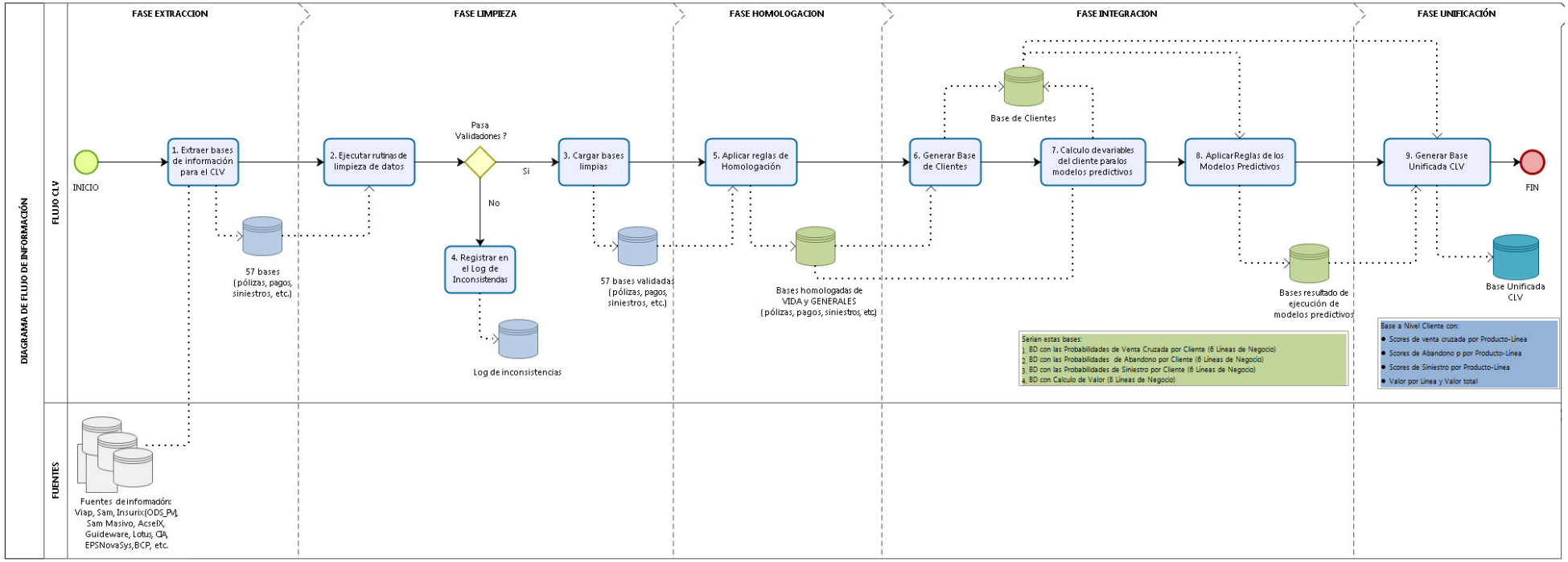

Nota: Extracto del Documento de Análisis de Requerimientos de Teamsoft.

Se unificarán datos complementarios de clientes, relacionados a la información de las pólizas, pagos, siniestros, etc. Los datos generales del cliente se obtendrán de la Tabla Única de Personas (TUP).

Se deberá contar con un Gobierno de Datos, para registrar los datos tal y como llegan de los sistemas orígenes. Permitirá mantener la trazabilidad del dato desde la extracción de los orígenes hasta su ingreso al repositorio CLV.

Además, será necesaria la validación del dato, ya que permite identificar casos de inconsistencias e informar a los usuarios. Asegurando que la plataforma ejecute con datos limpios y precisos. No implican rutinas de modificación de los datos en los orígenes, ni en el repositorio CLV.

Figura 9: Diagrama de flujo de la información




Nota: Extracto del Documento de Análisis de Requerimientos de Teamsoft.

B. Requerimientos Funcionales y No Funcionales

**Requerimientos Funcionales:**

- Tablas y campos para el RDV del Data Lake: Se realizó el análisis de la documentación proporcionada por el área usuaria, para determinar la relación de todas las tablas y campos que deben ser cargados de los sistemas orígenes hacia el RDV.
- Variables de Bases de Líneas: Se realizó el análisis y la consolidación de todas las variables solicitadas en las Bases de Líneas, para que sean incluidas en el modelo del UDV del Data lake de Pacifico.
- Reglas de negocio para Bases de Línea: Se realizó el análisis y la consolidación de todas las reglas de limpieza, homologación y creación de nuevas variables, solicitadas en las Bases de Líneas, para el cálculo del CLV.
- Tablas de equivalencias UDV: Se realizó el análisis de las tablas de equivalencia proporcionadas por el área usuaria en coordinación con Gobierno de la información, para definir aquellas equivalencias que serán incluidas en el UDV.

**Requerimientos No Funcionales:**

- Verificar que se mantengan los flujos actuales de negocio de base unificada.
- Los servidores estarán en un centro de datos con las condiciones de calidad requeridas.
- La plataforma debe contar con manuales de usuario para los usuarios finales que se encuentren definidos correctamente.
- Los accesos a las tablas y vistas, podrán ser solo modificados por el administrador de datos.
- Asegurar que los datos estén protegidos contra el acceso no autorizado por personas externas a la organización o departamento.

- Los tiempos de modificación o actualización del dato deben óptimos.

#### 3.2.5.3 Definición tablas y campos

Se detallarán las tablas más relevantes en la capa UDV del Data Lake, y se coloca una definición de lo que contendrá cada campo.

Figura 10: Campos MD_PERSONA

| Table | Caption | Name | Data Type |
|---|---|---|---|
| MD_PERSONA | Origen Dato | CODORIGEN | String |
| MD_PERSONA | Identificador Persona Unica | NUMID | String |
| MD_PERSONA | Tipo Documento Persona | TIPDOCMTO | String |
| MD_PERSONA | Numero Documento Persona | NRODCMNTO | String |
| MD_PERSONA | Nombres Persona | NOMPERS | String |
| MD_PERSONA | Apellido Paterno Persona | APEPATPERS | String |
| MD_PERSONA | Apellido Materno Persona | APEMATPERS | String |
| MD_PERSONA | Nombre Completo Persona | NOMBCOMP | String |
| MD_PERSONA | Nombre Comercial | NOMBCOMER | String |
| MD_PERSONA | Codigo Tipo de Cliente | CODTIPCLIENTE | String |
| MD_PERSONA | Descripcion Tipo de Cliente | DESCTIPCLIE | String |
| MD_PERSONA | Codigo Tipo de Persona | CODTIPPERSON | String |
| MD_PERSONA | Descripcion Tipo de Persona | DESCTIPPERS | String |
| MD_PERSONA | Fecha de Nacimiento | FECNCMNTO | String |
| MD_PERSONA | Fecha de Fallecimiento | FECDECESO | String |
| MD_PERSONA | Sexo Persona | SEXO | String |
| MD_PERSONA | Codigo Estado Civil | CODESTCIV | String |
| MD_PERSONA | Estado Civil | DESCESTCIV | String |
| MD_PERSONA | Codigo Ubigeo | CODUBIC | String |
| MD_PERSONA | Codigo Departamento | CODDPTO | String |
| MD_PERSONA | Departamento | DESCDPTO | String |
| MD_PERSONA | Codigo Provincia | CODPROV | String |
| MD_PERSONA | Provincia | DESCPROV | String |
| MD_PERSONA | Codigo Distrito | CODDISTR | String |
| MD_PERSONA | Distrito | DESCDISTR | String |
| MD_PERSONA | Direccion | DIRECC | String |

Nota: Elaboración propia

Figura 11: Campos HD SINIESTRO

| Table | Caption | Name |
|---|---|---|
| HD_SINIESTRO | Codigo Origen | CODORIGEN |
| HD_SINIESTRO | Numero de Siniestro | NUMSIN |
| HD_SINIESTRO | Identificador Interno de Siniestro | IDESIN |
| HD_SINIESTRO | Identificador Interno Poliza | IDEPOL |
| HD_SINIESTRO | Numero de Certificado | NUMCERT |
| HD_SINIESTRO | Numero de Poliza | NUMPOL |
| HD_SINIESTRO | Numero de Solicitud Vida | NROSOLVIDA |
| HD_SINIESTRO | Estado-Situacion del Siniestro | STSSIN |
| HD_SINIESTRO | Fecha de Estado Siniestro | FECSTS |
| HD_SINIESTRO | SubEstado del Siniestro (Vida) | SUBSTSSIN |
| HD_SINIESTRO | Identificación Persona Contratante Origen | NUMIDORIG |
| HD_SINIESTRO | Identificador Persona Unica Contratante | NUMID |
| HD_SINIESTRO | Identificador Persona Asegurado Origen | CODASEGORIG |
| HD_SINIESTRO | Identificador Persona Asegurado | CODASEG |
| HD_SINIESTRO | Codigo Producto Origen | CODPRDORIG |
| HD_SINIESTRO | Codigo Producto | CODPRD |
| HD_SINIESTRO | Descripcion Producto | DESCPRD |
| HD_SINIESTRO | Codigo Linea de Negocio Origen | CODLINEGORIG |
| HD_SINIESTRO | Codigo de Linea de Negocio | CODLINNEG |
| HD_SINIESTRO | Descripcion de Linea de Negocio | DESCLINEG |
| HD_SINIESTRO | Moneda Siniestro Origen | CODMONORIG |
| HD_SINIESTRO | Moneda del Siniestro | CODMON |
| HD_SINIESTRO | Fecha Ocurrencia Siniestro | FECOCURR |
| HD_SINIESTRO | Hora Ocurrencia del Siniestro | HOROCURR |
| HD_SINIESTRO | Codigo Motivo Rechazo Origen | CODMTVRECHORIG |
| HD_SINIESTRO | Codigo Motivo Rechazo | CODMTVRECH |
| HD_SINIESTRO | Descripcion Motivo Rechazo | DESCMTVRECH |
| HD_SINIESTRO | Codigo Motivo Anulacion Siniestro Origen | CODMTVANULORIG |
| HD_SINIESTRO | Codigo de Motivo de Anulacion del Siniestro | CODMTVANUL |
| HD_SINIESTRO | Descripcion Motivo de Anulacion Poliza | DESCMTVANUL |
| HD_SINIESTRO | Codigo Causa Siniestro Origen | CODCAUSINORIG |
| HD_SINIESTRO | Codigo Causa Siniestro | CODCAUSIN |
| HD_SINIESTRO | Descripcion Causa Siniestro | DESCCAUSIN |
| HD_SINIESTRO | Monto Pagado Bruto - Moneda Poliza | MTOPAGBRUPOL |
| HD_SINIESTRO | Monto Pagado Bruto - Soles | MTOPAGBRUSOL |
| HD_SINIESTRO | Monto Pagado Bruto - Dolarizado | MTOPAGBRUDOL |
| HD_SINIESTRO | Monto Pagado Neto - Moneda Poliza | MTOPAGNETPOL |
| HD_SINIESTRO | Monto Pagado Neto - Soles | MTOPAGNETSOL |
| HD_SINIESTRO | Monto Pagado Neto - Dolarizado | MTOPAGNETDOL |

Nota: Elaboración propia

Figura 12: Campos HD POLIZA

| Table | Caption | Name | Data Type |
|---|---|---|---|
| HD_POLIZA | Codigo Origen | CODORIGEN | String |
| HD_POLIZA | Identificador Interno Poliza | IDEPOL | String |
| HD_POLIZA | Numero de Poliza | NUMPOL | String |
| HD_POLIZA | Identificador Persona Contratante Origen | NUMIDORIG | String |
| HD_POLIZA | Identificador Persona Unica Contratante | NUMID | String |
| HD_POLIZA | Codigo Producto Origen | CODPRDORIG | String |
| HD_POLIZA | Codigo Producto | CODPRD | String |
| HD_POLIZA | Descripcion Producto | DESCPRD | String |
| HD_POLIZA | Fecha Primera Vigencia - Primera Emision | FECPRIMVIG | String |
| HD_POLIZA | Fecha Inicio Vigencia | FECINIVIG | String |
| HD_POLIZA | Fecha Fin de Vigencia | FECFINVIG | String |
| HD_POLIZA | Numero Renovacion | NUMREN | String |
| HD_POLIZA | Estado Poliza Origen | STSPOLORIG | String |
| HD_POLIZA | Estado Poliza | STSPOL | String |
| HD_POLIZA | Descripción Estado Póliza | DESCSTSPOL | String |
| HD_POLIZA | Moneda Poliza Origen | CODMONORIG | String |
| HD_POLIZA | Moneda Poliza | CODMON | String |
| HD_POLIZA | Fecha Anulacion Poliza | FECANUL | String |
| HD_POLIZA | Codigo Agente-Broker-Intermediario Origen | CODBROKORIG | String |
| HD_POLIZA | Codigo Agente-Broker-Intermediario | CODBROKER | String |
| HD_POLIZA | Nombre de Agente-Broker-Intermediario | NOMBROKER | String |
| HD_POLIZA | Numero Poliza Referencia Acselx | NUMPOLAX | String |
| HD_POLIZA | Segmento Mercado | SGMTOMCDO | String |
| HD_POLIZA | Codigo Canal de Ventas | CODCNLVTA | String |
| HD_POLIZA | Nombre Canal | NOMCNLVTA | String |
| HD_POLIZA | Codigo Grupo Canal | CODGRPCNL | String |
| HD_POLIZA | Codigo Grupo Comercial | CODGRPCOM | String |
| HD_POLIZA | Prima Facturada Bruta Moneda Poliza | MTOPRMBRUPOL | String |
| HD_POLIZA | Prima Facturada Bruta Soles | MTOPRMBRUSOL | String |
| HD_POLIZA | Prima Facturada Bruta Dolarizada | MTOPRMBRUDOL | String |
| HD_POLIZA | Prima Facturada Neta Moneda Poliza | MTOPRMNETPOL | String |
| HD_POLIZA | Prima Facturada Neta Soles | MTOPRIMNETSOL | String |
| HD_POLIZA | Prima Facturada Neta Dolarizada | MTOPRMNETDOL | String |
| HD_POLIZA | Tipo Documento Contratante | TIPODCMTO | String |
| HD_POLIZA | Numero de Documento Contratante | NRODCMTO | String |

Nota: Elaboración propia

Figura 13: Campos HD POLIZA SALUD

| Table | Caption | Name | Data Type |
|---|---|---|---|
| HD_POLIZA_SALUD | Codigo Origen | CODORIGEN | String |
| HD_POLIZA_SALUD | Identificador Interno Poliza | IDEPOL | String |
| HD_POLIZA_SALUD | Numero de Certificado | NUMCERT | String |
| HD_POLIZA_SALUD | Codigo de Asegurado Titular | CODASEGTIT | String |
| HD_POLIZA_SALUD | Identificador Persona Asegurado Origen | IDASEGORIG | String |
| HD_POLIZA_SALUD | Identificador Persona Unica Asegurado | IDASEG | String |
| HD_POLIZA_SALUD | Numero de Poliza | NUMPOL | String |
| HD_POLIZA_SALUD | Identificador de Persona Contratante Origen | NUMIDORIG | String |
| HD_POLIZA_SALUD | Identificador Persona Unica Contratante | NUMID | String |
| HD_POLIZA_SALUD | Codigo Producto Origen | CODPRDORIG | String |
| HD_POLIZA_SALUD | Codigo Producto | CODPRD | String |
| HD_POLIZA_SALUD | Descripcion Producto | DESCPRD | String |
| HD_POLIZA_SALUD | Fecha Primera Vigencia o Afiliacion | FECIPRMVIG | String |
| HD_POLIZA_SALUD | Fecha de Inicio de Vigencia | FECINIVIG | String |
| HD_POLIZA_SALUD | Fecha Fin de Vigencia | FECFINVIG | String |
| HD_POLIZA_SALUD | Numero Renovacion | NUMREN | String |
| HD_POLIZA_SALUD | Estado Poliza Origen | STSPOLORIG | String |
| HD_POLIZA_SALUD | Estado Poliza | STSPOL | String |
| HD_POLIZA_SALUD | Descripcion Estado Poliza | DESCSTSPOL | String |
| HD_POLIZA_SALUD | Moneda Poliza Origen | CODMONORIG | String |
| HD_POLIZA_SALUD | Moneda Poliza | CODMONEDA | String |
| HD_POLIZA_SALUD | Fecha Anulacion Poliza | FECANUL | String |
| HD_POLIZA_SALUD | Codigo Motivo Anulacion Origen | CODMVANUORIG | String |
| HD_POLIZA_SALUD | Codigo Motivo Anulacion | CODMTVANU | String |
| HD_POLIZA_SALUD | Descripcion Motivo Anulacion Poliza | DESCMTVANU | String |
| HD_POLIZA_SALUD | Codigo Agente-Broker-Intermediario Origen | CODBROKORIG | String |
| HD_POLIZA_SALUD | Codigo Agente-Broker-Intermediario | CODBROKER | String |
| HD_POLIZA_SALUD | Nombre Agente-Broker-Intermediario | NOMBROKER | String |

Nota: Elaboración propia

Figura 14: Campos MD PERSONA ROL POLIZA

| Table | Caption | Name | Data Type |
|---|---|---|---|
| MD_PERSONA_ROL_POLIZA | Codigo Origen | CODORIGEN | String |
| MD_PERSONA_ROL_POLIZA | Tipo Rol Persona | TIPOROL | String |
| MD_PERSONA_ROL_POLIZA | Identificador Persona Unica | NUMID | String |
| MD_PERSONA_ROL_POLIZA | Identificador Interno Poliza | IDEPOL | String |
| MD_PERSONA_ROL_POLIZA | Numero de Poliza | NUMPOL | String |
| MD_PERSONA_ROL_POLIZA | Identificador Persona Origen | NUMIDORIG | String |
| MD_PERSONA_ROL_POLIZA | Codigo de Producto Origen | CODPRDORIG | String |
| MD_PERSONA_ROL_POLIZA | Codigo de Producto | CODPRD | String |
| MD_PERSONA_ROL_POLIZA | Descripcion de Producto | DESCPRD | String |
| MD_PERSONA_ROL_POLIZA | Codigo de Parentesco Origen | CODPRNTORIG | String |
| MD_PERSONA_ROL_POLIZA | Codigo de Parentesco | CODPRNT | String |
| MD_PERSONA_ROL_POLIZA | Descripcion Parentesco | DESCPRNT | String |
| MD_PERSONA_ROL_POLIZA | Porcentaje Beneficio | PORCENBENEF | String |
| MD_PERSONA_ROL_POLIZA | Tipo de Beneficiario | TIPOBENFCIARIO | String |
| MD_PERSONA_ROL_POLIZA | Tipo de Documento | TIPODCMTO | String |
| MD_PERSONA_ROL_POLIZA | Numero de Documento | NRODCMNTO | String |
| MD_PERSONA_ROL_POLIZA | Fecha Proceso Log Carga | FECPRCLG | Date |
| MD_PERSONA_ROL_POLIZA | Usuario Proceso Log de Carga | USRPRCLG | String |
| MD_PERSONA_ROL_POLIZA | Hora Proceso Log de Carga | HORPRCLG | Timestamp(%p1) |
| MD_PERSONA_ROL_POLIZA | Flag Estado de Registro | FLGSTSRG | String |

Nota: Elaboración propia.

Figura 15: Campos de Equivalencias UDV

| Table | Caption | Name | Comments |
|---|---|---|---|
| MD_EQUIVAL_UDV | Codigo de Tabla de Homologacion | CODTABLA | Permite Separar y clasificar los diferentes grupos de valores en la tabla, por ej:<br>PRNT=Parentescos<br>ECIV=Estado Civil<br>SPOL=Estados de Póliza |
| MD_EQUIVAL_UDV | Descripcion Tabla de Homologacion | DESCTABLA | Describe referencia de tabla de valores homologados |
| MD_EQUIVAL_UDV | Aplicacion Origen del Dato a Homologar | CODORIGEN | Contiene la Referencia de la aplicaicon de origen del dato:AX(Acselx), GW(GuideWire), VI(Viap), SM(Sam Masivo), IN(Insurix) |
| MD_EQUIVAL_UDV | Valor de Busqueda | VALORIGEN | Puede contener un solo codigo o valor (que llega del origen) o contener una concatenacion de valores para hallar una homologacion |
| MD_EQUIVAL_UDV | Valor Homologado | VALUNIFICA | Valor devuelto como Homologado - Unificado |
| MD_EQUIVAL_UDV | Descripcion Homologada | DESCUNIFICA | Descripción Homologada |
| MD_EQUIVAL_UDV | Agrupación Valores Homologados | AGPUNIFICA | Agrupacion para Valores Homologados |
| MD_EQUIVAL_UDV | Fecha Inicio de Vigencia | FECINIVIG | Fecha de Inicio de Vigencia de Homologacion |
| MD_EQUIVAL_UDV | Fecha fin de Vigencia | FECFINVIG | Fecha de Fin de Vigencia de Homologacion |
| MD_EQUIVAL_UDV | Fecha de Creacion | FECCREACION | Fecha de Creacion del Registro |
| MD_EQUIVAL_UDV | Fecha Actualizacion | FECACTUALIZA | Fecha de Actualizacion del Registro |
| MD_EQUIVAL_UDV | Usuario Creador | USRCREACION | Usuario Creador del Registro |
| MD_EQUIVAL_UDV | Usuario Ultima Actualizacion | USRMODIFICA | Usuario Actualizo Registro |

Nota: Elaboración propia.

Figura 16: Campos de Equivalencias DDV

| Table | Caption | Name | Comments |
|---|---|---|---|
| MD_EQUIVAL_DDV | Caso de Uso | CODCUSO | Describe Caso de Uso, Ejemplo: CLV, |
| MD_EQUIVAL_DDV | Codigo de Tabla de Valores | CODTABLA | Permite Separar y clasificar los diferentes grupos de valores en la tabla, por ej:<br>PRNT=Parentescos<br>ECIV=Estado Civil<br>SPOL=Estados de Póliza |
| MD_EQUIVAL_DDV | Descripcion tabla Homologacion | DESCTABLA | Descripción de la tabla definida para Homologacion, por ej. PARNT,ECIVIL,MONEDA |
| MD_EQUIVAL_DDV | Aplicacion Origen de Dato a Homologar | CODORIGEN | Contiene el Origen del dato a clasificar-homologar (referencia a la fuente del dato) AX(Acselx), GW(GuideWire), VI(Viap), SM(Sam Masivo), IN(Insurix) |
| MD_EQUIVAL_DDV | Valor de Busqueda | VALORIGEN | Puede contener un solo valor (que llega del origen) o contener una concatenacion de valores para hallar una clasificación. |
| MD_EQUIVAL_DDV | Valor Homologado | VALUNIFICA | Valor devuelto como Homologado - Unificado |
| MD_EQUIVAL_DDV | Descripcion Homologada | DESCUNIFICA | Descripción Homologada |
| MD_EQUIVAL_DDV | Agrupación Valores Homologados | AGPUNIFICA | Agrupacion - Homologacion |
| MD_EQUIVAL_DDV | Fecha Inicio de Vigencia | FECINIVIG | Fecha de Inicio de Vigencia de Homologacion |
| MD_EQUIVAL_DDV | Fecha Fin de Vigencia | FECFINVIG | Fecha de Fin de Vigencia de Homologacion |
| MD_EQUIVAL_DDV | Fecha de Creacion | FECCREACION | Fecha de Creación de Registro de Homologación |
| MD_EQUIVAL_DDV | Fecha Actualizacion | FECACTUALIZA | Fecha de Actualizacion de Registro |
| MD_EQUIVAL_DDV | Usuario Creador | USRCREACION | Usuario Creador de registro de Homologación |
| MD_EQUIVAL_DDV | Usuario Ultima Actualizacion | USRACTUALIZA | Usuario actualizacion registro de Homologación |

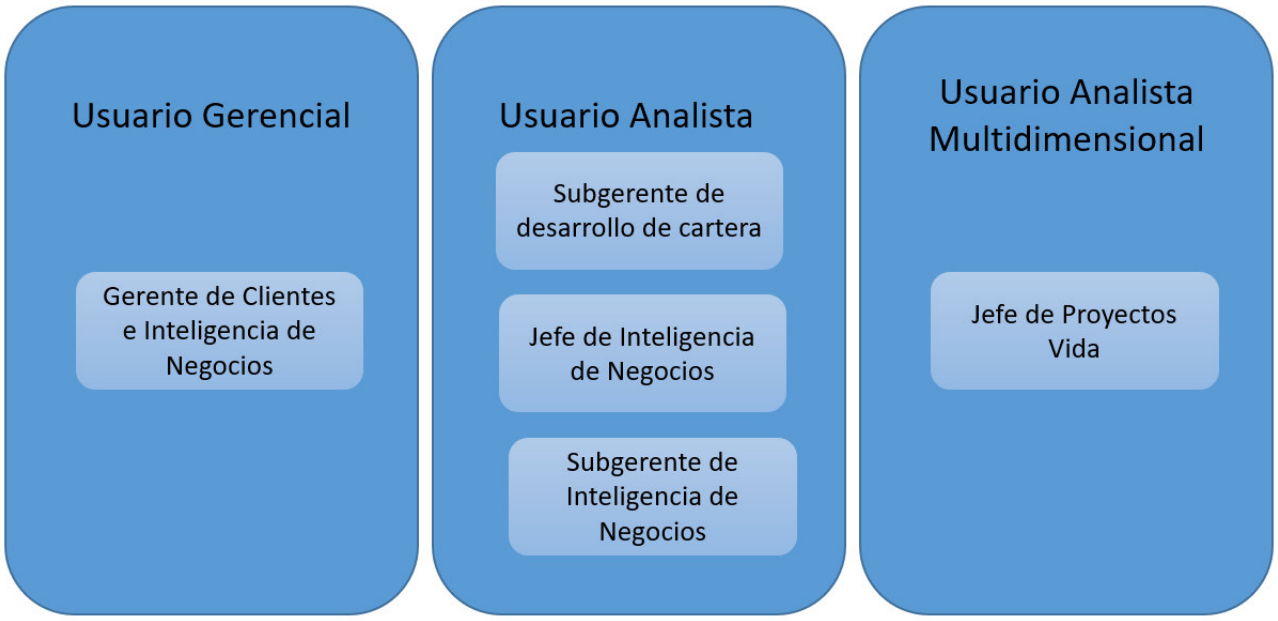
Nota: Elaboración propia

### 3.2.5.4 Despliegue

En esta etapa se indica la ruta de ejecución tomada, siguiendo el plan de trabajo establecido en la primera etapa de Gestión.

Se debe definir a los usuarios/clientes que consumirán la información almacenada en el Data Lake, entre ellos tenemos Usuario Gerencial, Usuario Analista y Usuario Analista Multidimensional.

Figura 17: Consumidores de la información en Data Lake




Nota: Elaboración propia

**Usuario Gerencial:** Son los usuarios corporativos o ejecutivos que ven la información en detalle e históricos, y requieren que las consultas sean construidas previamente. Son usuarios que deben poder utilizar el sistema de una forma fácil e intuitiva. La labor de manipulación de los datos no es parte de su función. Bajo este rol se encuentran los gerentes o ejecutivos del negocio.

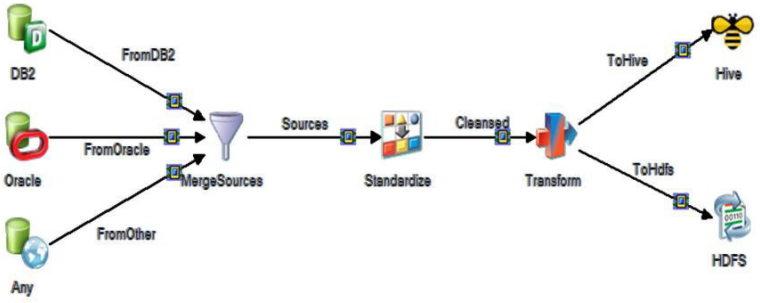
**Usuario Analista:** Es el usuario que estará familiarizado con el ambiente OLAP y con el de Análisis Multidimensional. Este usuario no tiene un patrón definido de consultas, elaborará consultas propias o realizará consultas directamente a los Cubos de Análisis Multidimensional: Estos usuarios tienen control total sobre el universo de indicadores y filtros (metadatos) del modelo OLAP. Pueden crear sus propios análisis y convertirlos en consultas para que puedan ser usados por el usuario gerencial.

**Usuario Analista Multidimensional:** Estos usuarios especializados se encargarán de extender la funcionalidad del sistema, es decir poder generar nuevos cubos, nuevos indicadores y/o nuevas agrupaciones.

Tomar en cuenta que, para poblar el Data Lake, debe seguir un orden de carga. Pasando primero por la capa RDV, luego UDV para la homologación y por último la carga en DDV.

En la capa RDV, se extraen las bases de información de los sistemas origen. Se mantienen los datos almacenados tal como se encuentran en las aplicaciones, un repositorio de los distintos tipos de datos en un único lugar. Los esquemas de las tablas son definidos por la aplicación fuente de la cual se está realizando la carga. En este caso tomamos los datos de BD SQL Server y Oracle, utilizamos flujos de trabajos para el procesamiento de extracción, al cual dentro de la herramienta DataStage se le conoce como “Job Paralelo”. Se leerá la estructura y campos tal cual de origen, utilizando herramientas dentro del programa para hacer la carga a los archivos HDFS en la capa RDV.

Figura 18: Job Extracción para RDV



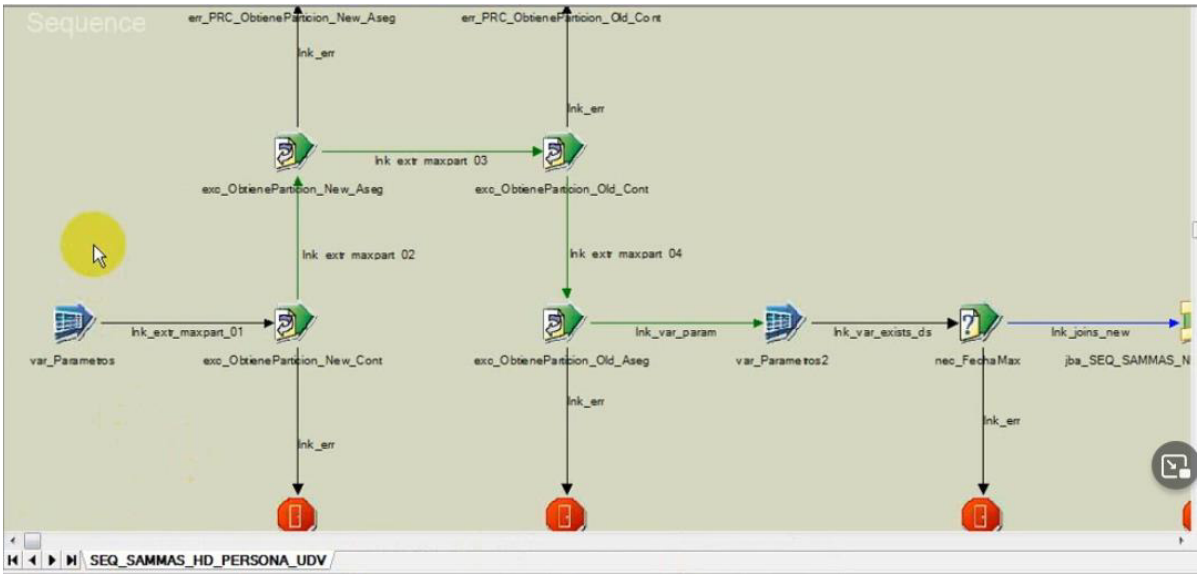

Nota: Fuente de 8.3 IBM ETL.pdf (e-LEARNING and Online Teaching, 2018)

Una vez extraída la información, se ejecutan procesos para la limpieza, homologación y modelado de datos según las reglas de negocio. Se siguen los lineamientos establecidos según una matriz de equivalencias para homologar la data en esta fase (Anexo 01) y también se realizan los cruces de información definidos en los mapeos de las entidades UDV que se listan en el documento de Diseño de Sistemas (Anexo 02 y Anexo 03), generados en conjunto con el cliente.

Aquí se cuenta con un repositorio consistente y el esquema de datos es universal y definido por el negocio.

A continuación, se presentará un proceso de la capa UDV, en este caso para poblar la tabla HD PERSON para el aplicativo SAM MASIVO. Se cuenta con un proceso llamado Job Secuencial que enmalla y agrupa a los Jobs Paralelos, siguiendo un orden de pasos para la ejecución.

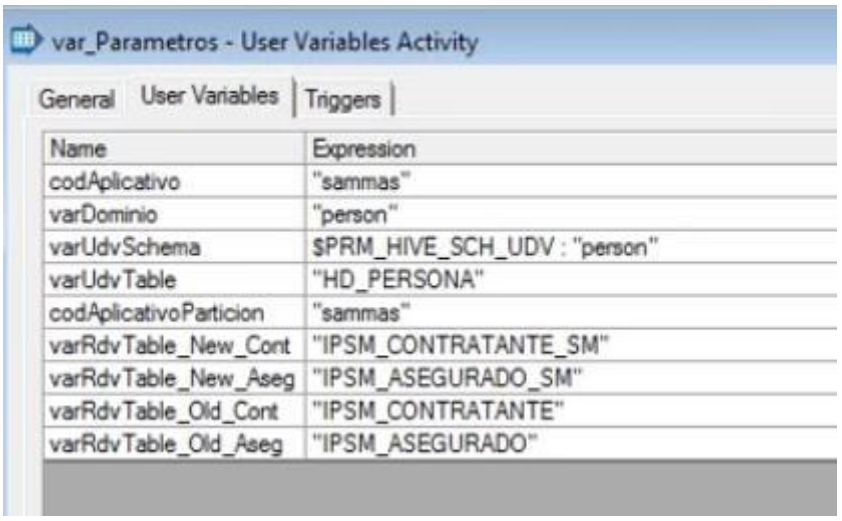
Figura 19: Job Secuencial HD PERSONA SAM MASIVO




Nota: Elaboración propia

Primero se configura los parámetros propios del Job Secuencial, que servirá para definir el aplicativo que está ejecutando e indicar los nombres de las tablas que intervienen en el proceso.

Figura 20: Parámetros del Job Secuencial

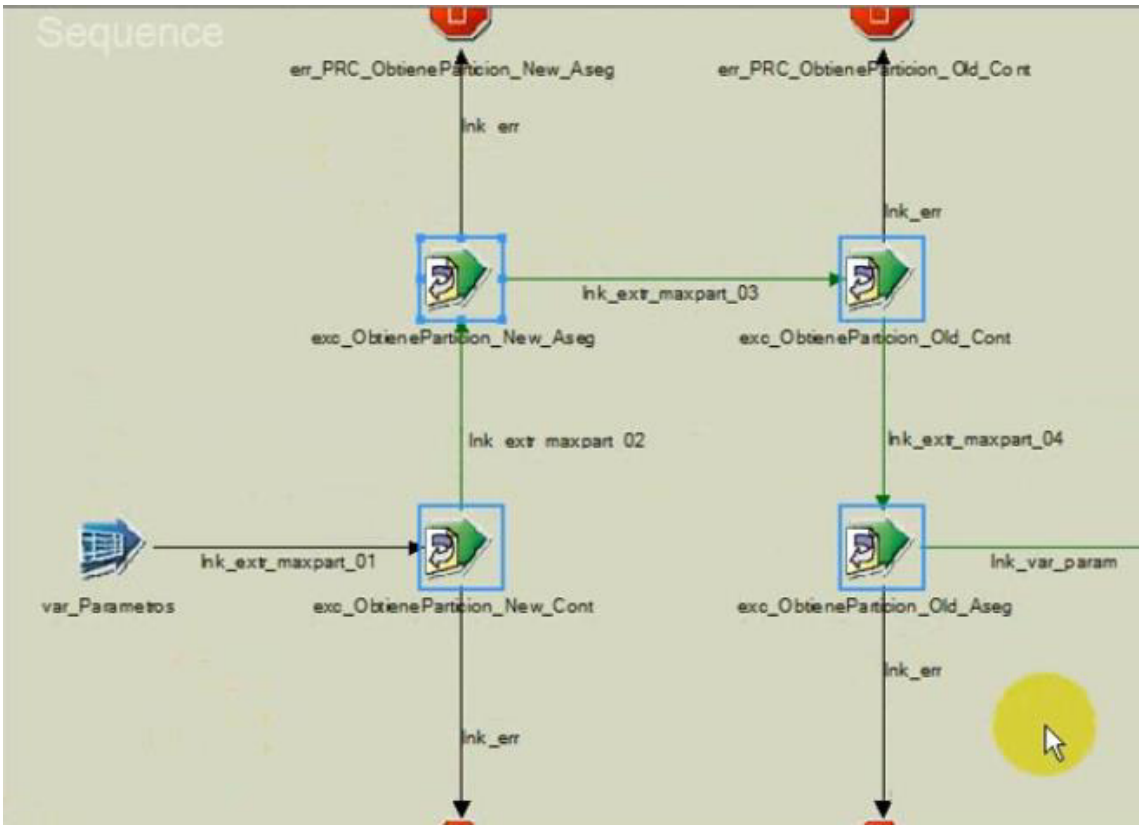



Nota: Elaboración propia

Recordemos que la carga de la data hacia las entidades, se encuentra separado por aplicativos y se registra en particiones diferentes dentro los archivos HDFS. Por lo tanto, dentro de la tabla se debe indicar en un campo, cuál será el origen de la data, así servirá el parámetro Código de aplicativo para establecer que la fuente origen será “sammas”.

Se colocan además que las tablas a consumir serán IPSM_CONTRATANTE_SM, IPSM CONTRATANTE, IPSM_ASEGURADO_SM e IPSM_ASEGURADO. Y también se coloca el esquema y la tabla a poblar, en los parámetros varUdvSchema y varUdvTable respectivamente.

Figura 21: Rutinas para leer particiones en HDFS

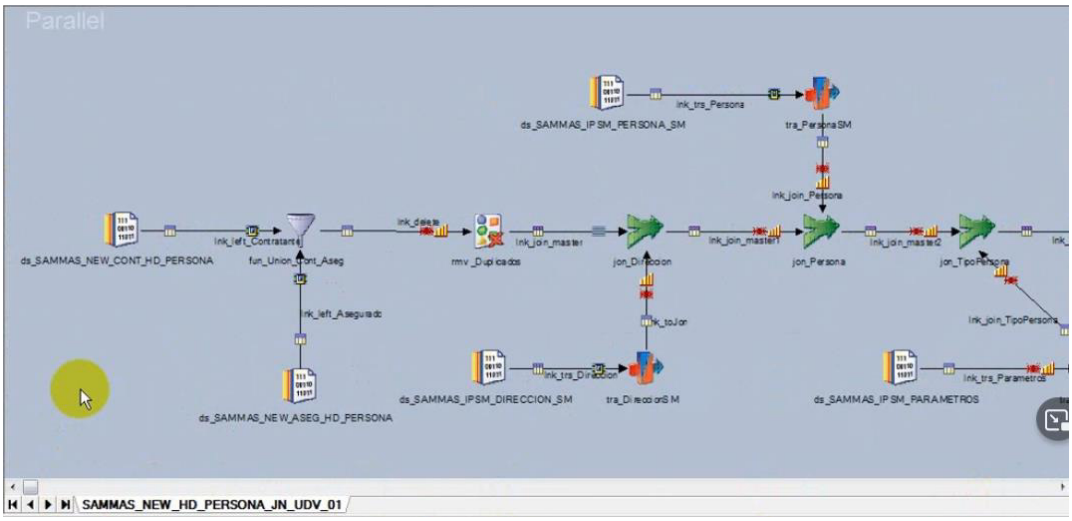



Nota: Elaboración propia

En la figura 20, se tienen rutinas que internamente validan las particiones en los archivos HDFS de las tablas mencionadas anteriormente, para obtener la partición más reciente para el procesamiento.

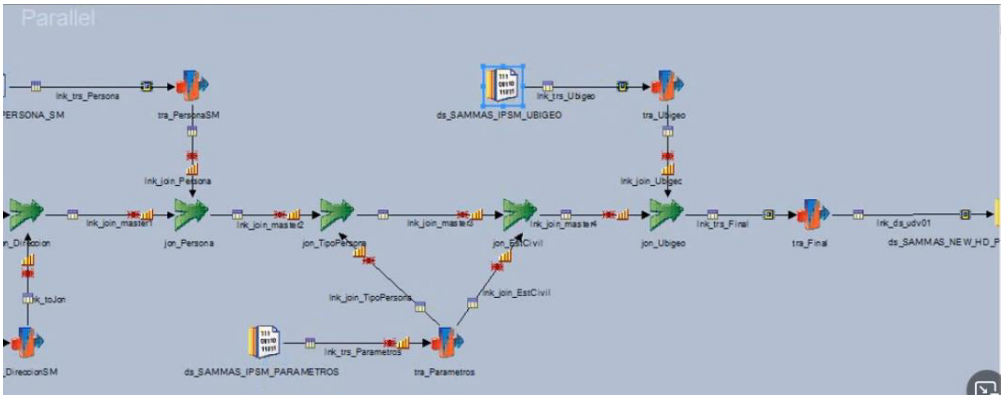
Una vez se tengan preparadas las particiones y las tablas a procesar, se continúa con un proceso para realizar los cruces de información necesarios.

Figura 22: Proceso Join de Persona 01



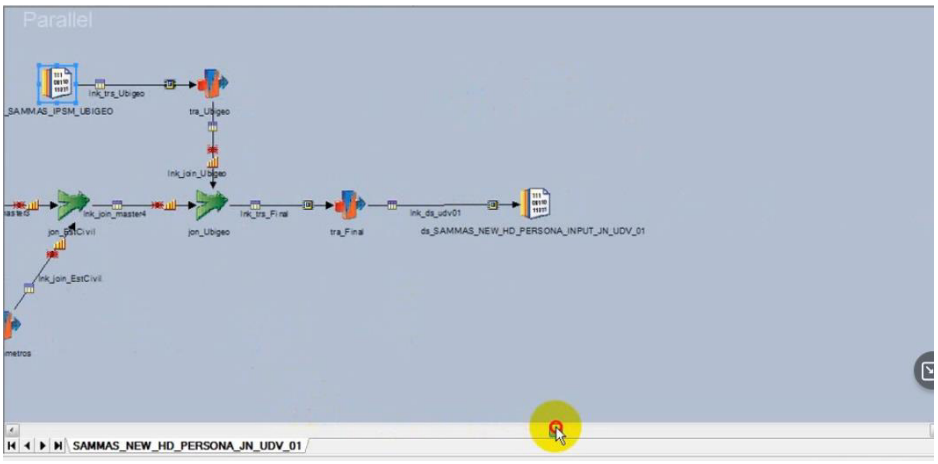

Nota: Elaboración propia

Figura 23: Proceso Join de Persona 02




Nota: Elaboración propia

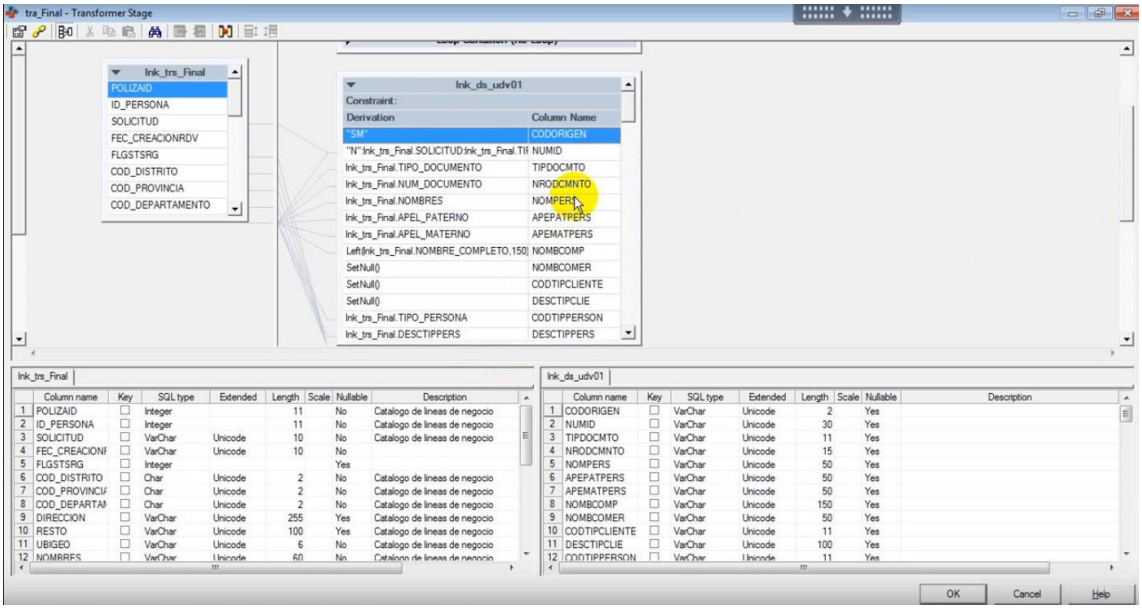
Figura 24: Proceso Join de Persona 03




Nota: Elaboración propia

En las figuras 21,22 y 23; se realizan los joins de tablas involucradas en el proceso. Es aquí donde se traduce el mapeo brindado por los Analistas Funcionales, donde indica los filtros o cruces para formar la entidad Persona.

Figura 25: Transformador de datos para Persona

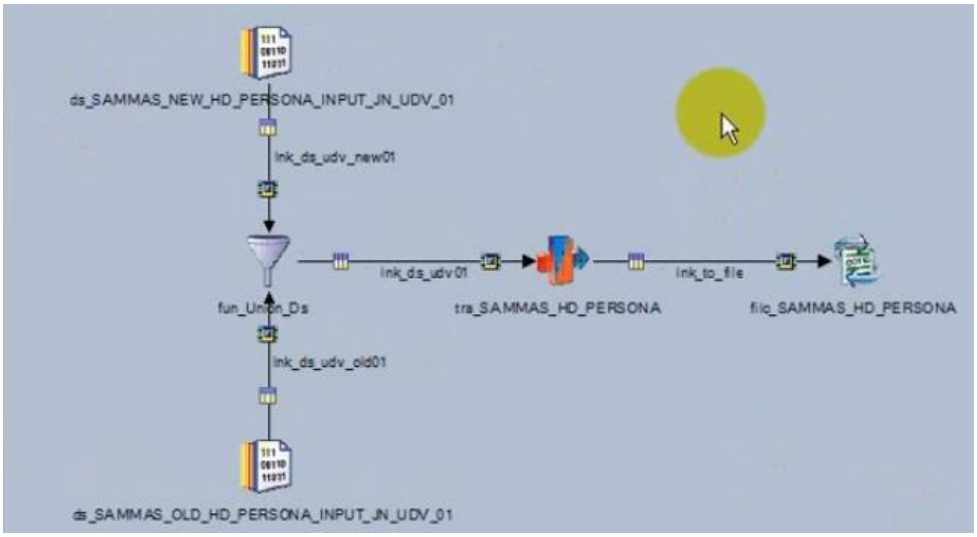



Nota: Elaboración propia

En la figura 24, observamos que colocamos como CODORIGEN, el valor de “SM” que pertenece a Sam Masivo. Esto servirá para identificar de que tabla está viniendo

la información. Recordemos que, dentro de las entidades, se tienen particionadas las tablas por aplicativo.

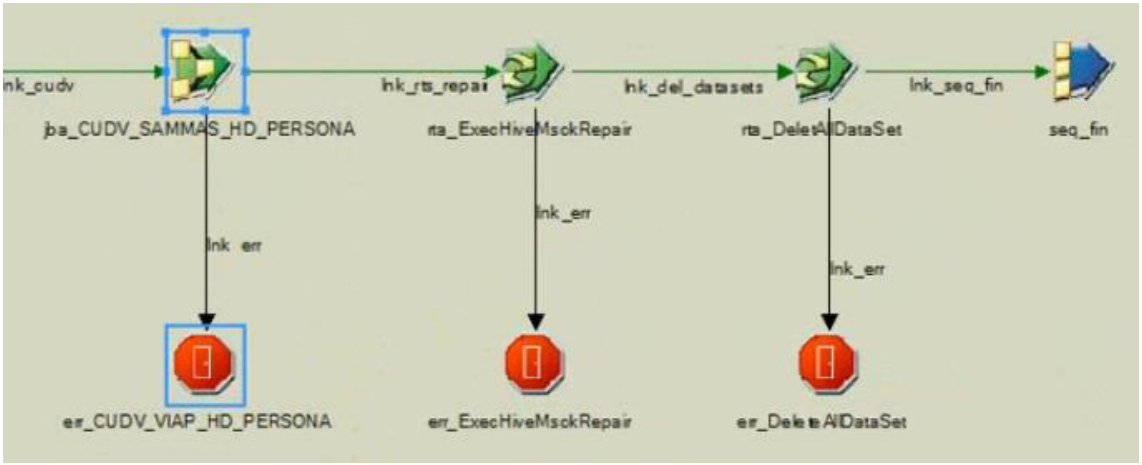
Figura 26: Carga entidad Persona




Nota: Elaboración propia

A continuación, en la figura 25, tenemos el proceso de carga hacia la entidad Persona como archivo HDFS.

Figura 27: Reparar particiones Entidad Persona

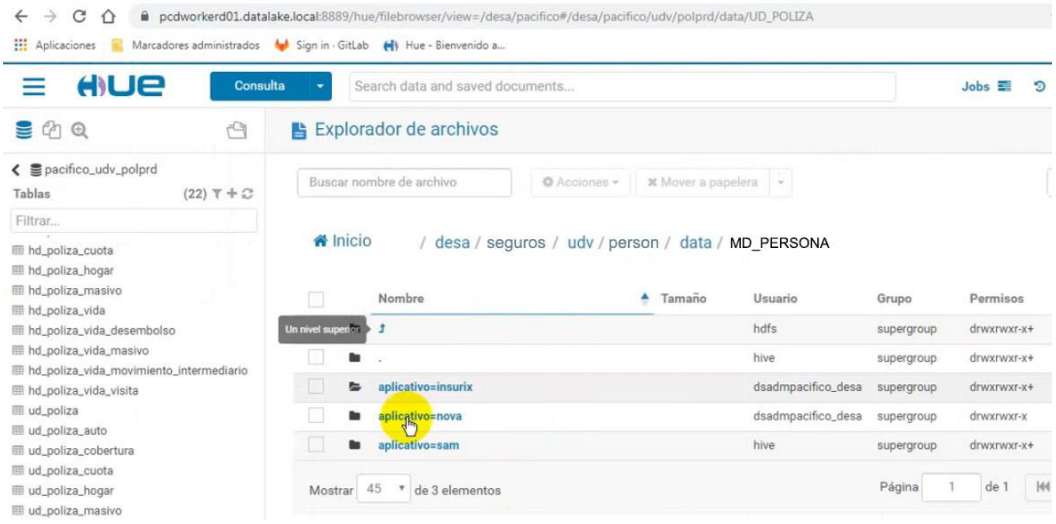



Nota: Elaboración propia

Luego del proceso de Carga en HDFS, se debe ejecutar el comando de MSCK REPAIR TABLE para recuperar todas las particiones que se agregaron al sistema de archivos después de crear la tabla y actualizar la metadata.

Y por último se ejecuta una rutina para eliminar los Datasets temporales utilizados durante la ejecución del proceso, esto con el objetivo de limpiar los nodos y liberar espacio.

Figura 28: Particiones entidad Persona




Nota: Elaboración propia

Para la capa DDV, se cuenta con datos dimensionales modelados según necesidades del negocio. En esta capa se homologan datos maestros que servirán para cálculos de información realizado por otros equipos. También se aplican reglas de limpieza de datos, que discrimine la información extraída del UDV únicamente con la data requerida para el cálculo del CLV. Se implementaron procesos de unificación de datos, los cuales se pueda unificar la información del cliente, tomando como fuente toda la información contenida en esta capa.

## 3.3 Evaluación

### 3.3.1 Evaluación Económica

Se da la siguiente estimación de costos para el desarrollo de la migración de datos.

Tabla 10: Costos de Infraestructura

| Infraestructura | Cantidad | Costo Unitario | Costo Total |
|---|---|---|---|
| Equipos Portátiles | 7 | S/ 3500,00 | S/ 24.500,00 |
| Server Desarrollo | 1 | S/ 20.000,00 | S/ 20.000,00 |
| Server Producción | 1 | S/ 20.000,00 | S/ 20.000,00 |
| Server de archivos | 1 | S/ 20.000,00 | S/ 20.000,00 |
| | | Total | S/ 84.500,00 |

Nota: Elaboración propia

Tabla 11: Costos del Proyecto

| | Fases | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Equipo | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Jefe de Proyecto | S/ 2.000,00 | S/ 2.000,00 | S/ 2.000,00 | S/ 2.000,00 | S/ 2.000,00 | S/ 2.000,00 | S/ 2.000,00 | S/ 2.000,00 |
| Scrum Master | S/ 2.450,00 | S/ 2.450,00 | S/ 2.450,00 | S/ 2.450,00 | S/ 2.450,00 | S/ 2.450,00 | S/ 2.450,00 | S/ 2.450,00 |
| Ing. De Datos ( 4 personas) | S/ 0,00 | S/ 20.800,00 | S/ 20.800,00 | S/ 20.800,00 | S/ 20.800,00 | S/ 20.800,00 | S/ 20.800,00 | S/ 20.800,00 |
| Analista Funcional (2 personas) | S/ 8.000,00 | S/ 8.000,00 | S/ 8.000,00 | S/ 8.000,00 | S/ 8.000,00 | S/ 8.000,00 | S/ 0,00 | S/ 0,00 |
| Infraestructura | S/ 84.500,00 | S/ 0,00 | S/ 0,00 | S/ 0,00 | S/ 0,00 | S/ 0,00 | S/ 0,00 | S/ 0,00 |
| **Total presupuesto** | S/ 96.950,00 | S/ 33.250,00 | S/ 33.250,00 | S/ 33.250,00 | S/ 33.250,00 | S/ 33.250,00 | S/ 25.250,00 | S/ 25.250,00 |
| **Total acumulado** | S/ 96.950,00 | S/ 130.200,00 | S/ 163.450,00 | S/ 196.700,00 | S/ 229.950,00 | S/ 263.200,00 | S/ 288.450,00 | S/ 313.700,00 |

Nota: Elaboración propia

El presente autor no tuvo acceso a la información sobre la viabilidad del proyecto y retornos de ganancias. Por lo general para proyectos de tecnología Big Data, se estima unas ganancias del 35% sobre la inversión total. Para el presente informe teniendo un costo total de S/. 313 700, la empresa obtendría un retorno de S/. 423 495. Generando un beneficio total de la inversión de S/. 109 795.

## CAPÍTULO IV – REFLEXIÓN CRÍTICA DE LA EXPERIENCIA

Dentro del proyecto, el haber trabajado con la metodología ágil permitió dividir el proyecto en tareas más pequeñas e ir realizando entregas en periodos de corto tiempo.

El autor del presente trabajo destaca la realización de videos de aprendizaje, que sirvieron para los nuevos ingresos del equipo. Permitiendo tener primero un plan de capacitación con estos recursos, sin utilizar mucho tiempo del Ing. De datos para la explicación del proyecto o el uso de herramientas durante el mismo. Para que luego de esto, se agendaban reuniones de revisión para culminar con la explicación o resolución de consultas por parte del integrante nuevo del equipo.

Como reflexión, el autor considera que en todo proyecto debe existir una buena documentación de lo que se está desarrollando. Al momento de implementar el proyecto, se dejaron de lado ciertos procesos en la documentación. Ya sea por la premura de terminar el sprint u otro motivo. Pero este paso de la documentación completa y concisa es de suma importancia, tanto para el equipo cuando se tenga que dar algún mantenimiento a la plataforma o para el cliente mismo, en caso cambie de proveedor y se quiera modificar alguna funcionalidad.

# CAPÍTULO V – CONCLUSIONES Y RECOMENDACIONES

## 5.1 Conclusiones

- La carga de data sobre tablas Hive, habilita a los usuarios con poco conocimiento en programación a realizar consultas SQL sobre los archivos HDFS, ya sea para poder generar reportes y/o explotar la información contenida de primera mano. Validando en cada una de las fases por las qué pasó el dato.
- El uso de un Data lake brinda muchos beneficios a la organización, al tener un procesamiento sobre archivos distribuidos permite un tiempo de respuesta óptimo, dando velocidad a la hora de consultar la información. Por lo tanto, la creación de reportes y tableros a tiempo, colaborando para la toma de decisiones con información veraz.
- Al realizar el cálculo del valor para cada cliente, se podrá determinar sobre quién conviene aplicar ofertas o nuevos productos específicos con tal de retener al cliente y no permitir que pueda irse con la competencia.

## 5.2 Recomendaciones

- Se debe de tener especial cuidado con el gobierno del dato, ya que al tener diferentes sistemas origen, los datos deberán integrarse para así evitar incoherencias o duplicidad de la información del cliente. De esta forma aseguramos totalmente la calidad del dato y otorgar información confiable a la plataforma.
- Aprovechando la implementación del Data Lake, se recomienda almacenar data no estructurada como la que proviene de redes sociales y/o las llamadas grabadas con los clientes al momento de su atención. Esto permitirá que en un futuro se pueda procesar inteligencia artificial para realizar un análisis de sentimiento e identificar información subjetiva del mundo digital, para determinar qué opinión tienen nuestros usuarios o posibles clientes hacia la empresa.

- El uso de una metodología ágil es muy recomendable, permite a los equipos poder organizarse y establecer un desarrollo colaborativo entre los miembros del mismo. Así en las reuniones diarias, se cometan las tareas de cada persona, y se verifican si existen dependencias o bloqueantes entre las actividades de cada uno.

Bibliografía

Atlassian. (2021). *Scrum: qué es, cómo funciona*. Obtenido de https://www.atlassian.com/es/agile/scrum

e-LEARNING and Online Teaching. (11 de Enero de 2018). Obtenido de e-LEARNING and Online Teaching: https://elearning.unimib.it/pluginfile.php/551607/mod_folder/content/0/8.3%20IBM%20ETL.pdf

Instituto de Ingenieria del conocimiento. (2019). *Las 7V del Big Data*. Obtenido de https://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/

Power Data. (2021). *Big Data ¿En qué consiste?* Obtenido de https://www.powerdata.es/big-data

SCRUMstudy. (2018). Obtenido de SCRUMstudy: https://online.vmedu.com/Courses/DownloadReferenceMaterial

Teamsoft. (2021). *Misión y Visión*. Obtenido de https://www.teamsoft.com.pe/nosotros/#mision-vision

TeamSoft. (2021). *Nosotros - Teamsoft*. Obtenido de https://www.teamsoft.com.pe/nosotros/#nuestro-proceso

We are Marketing. (2020). *¿Qué es Customer Lifetime Value?* Obtenido de https://www.wearemarketing.com/es/blog/que-es-el-customer-lifetime-value-da-valor-a-tus-clientes-y-mejora-tu-roi.html

Wikipedia, C. (29 de Abril de 2021). *Procesamiento por lotes*. Obtenido de Wikipedia: La enciclopedia libre: https://es.wikipedia.org/wiki/Procesamiento_por_lotes

ANEXO 01: Matriz de Equivalencias UDV

Homologación para Estado Civil:

| Origen | Código Búsqueda | Descripción | Código de Homologación | Descripción codificación propuesta | NUEVA HOMOLOGACION CLV |
|---|---|---|---|---|---|
| AX | C | CASADO | 002 | CASADO/CASADA | CASADO |
| | | | | | |
| NV | C | CASADO | 002 | CASADO/CASADA | CASADO |
| SM | CBS | Casado(a) con Bienes Separados | 002 | CASADO/CASADA | CASADO |
| | | | | | |
| NV | N | CONVIVIENTE | 005 | CONVIVIENTE | CONVIVIENTE |
| NV | N | CONVIVIENTE | 005 | CONVIVIENTE | CONVIVIENTE |
| VI | O | Conviviente | 005 | CONVIVIENTE | CONVIVIENTE |
| SM | D | Divorciado(a) | 004 | DIVORCIADO/DIVORCIADA | DIVORCIADO |
| NV | D | DIVORCIADO | 004 | DIVORCIADO/DIVORCIADA | DIVORCIADO |
| VI | | VACIO | 996 | VACIO | NA_VACIO |
| VI | | VACIO | 996 | VACIO | NA_VACIO |
| VI | null | NULL | 997 | NULL | NA_VACIO |
| GW | N | No Aplica | 007 | NO APLICA | NA_VACIO |
| AX | N | NO APLICA | 007 | NO APLICA | NA_VACIO |
| | | | | | |
| SM | Z | No Aplica | NULL | NO IDENTIFICADO/NO COINCIDE | NA_VACIO |
| VI | E | Separado | 006 | SEPARADO/SEPARADA | SEPARADO |
| AX | P | SEPARADO | 006 | SEPARADO/SEPARADA | SEPARADO |
| | | | | | |
| NV | P | SEPARADO | 006 | SEPARADO/SEPARADA | SEPARADO |
| VI | S | Soltero | 001 | SOLTERO/SOLTERA | SOLTERO |
| | | | | | |
| NV | S | SOLTERO | 001 | SOLTERO/SOLTERA | SOLTERO |
| VI | V | Viudo | 003 | VIUDO/VIUDA | VIUDO |
| | | | | | |
| NV | V | VIUDO | 003 | VIUDO/VIUDA | VIUDO |

Homologación de Tipo de Documento:

| Origen | Código Búsqueda | Descripción | Propuesta Valor Homologado UDV_PGA VARCHAR(20) | Descripción codificación propuesta UDV_PGA VARCHAR(080) | NUEVA HOMOLOGACION CLV |
|---|---|---|---|---|---|
| VI | DNI | DNI | DNI | DOCUMENTO NACIONAL DE IDENTIDAD | D |
| IN | DN | DOCUMENTO NACIONAL DE IDENTIDAD | DNI | DOCUMENTO NACIONAL DE IDENTIDAD | D |
| | | | | | |
| AX | DNI | DNI/Libreta Electoral | DNI | DOCUMENTO NACIONAL DE IDENTIDAD | D |
| | | | | | |
| VI | CE | CE | CE | CARNÉ DE EXTRANJERÍA | E |
| | | | | | |
| NV | 2 | CARNET EXTRANJERIA | CE | CARNÉ DE EXTRANJERÍA | E |
| IN | FA | FAMILIARES | NULL | NO IDENTIFICADO/NO COINCIDE | F |
| AP | CI | C.FUERZAS ARMADAS | CI | CARNÉ DE IDENTIDAD | I |
| AP | CIP | C.FUERZAS POLICIALES | CIP | CARNÉ DE FUERZAS POLICIALES | I |
| AX | OTR | Otro | OTR | OTROS | O |
| VI | OTR | O | OTR | OTROS | O |
| | | | | | |
| | | | | | |
| GW | foreign_RUC_Ext | CUIT | CUIT | CODIGO DE IDENTIFICACIÓN TRIBUTARIA DEL PAIS DE ORIGEN | O |
| VI | PAS | PAS | PAS | PASAPORTE | P |
| | | | | | |
| NV | 6 | PASAPORTE | PAS | PASAPORTE | P |
| IN | PA | PASAPORTE | PAS | PASAPORTE | P |
| VI | RUC | RUC | RUC | REGISTRO UNICO DE CONTRIBUYENTES | R |
| | | | | | |
| AX | RUC | Registro Unico Contribuyente | RUC | REGISTRO UNICO DE CONTRIBUYENTES | R |
| AX | | VACIO | 996 | VACIO | S |
| VI | | VACIO | 996 | VACIO | S |
| NV | | VACIO | 996 | VACIO | S |

Homologación de Parentesco:

| Origen | Código Búsqueda | Descripción | Propuesta Valor Homologado UDV_PGA | Descripción codificación propuesta UDV_PGA VARCHAR(080) | NUEVA HOMOLOGACION CLV |
|---|---|---|---|---|---|
| AX | 26 | CONVIVIENTE FEMENINO | 035 | CONVIVIENTE | CONYUGE |
| AX | 25 | CONVIVIENTE MASCULINO | 035 | CONVIVIENTE | CONYUGE |
| AP | 001 | CONYUGE | 001 | CONYUGE | CONYUGE |
| VI | | Cónyuge (Esposo - Esposa) | 001 | CONYUGE | CONYUGE |
| AX | 3 | ESPOSO | 001 | CONYUGE | CONYUGE |
| AX | 4 | ESPOSA | 001 | CONYUGE | CONYUGE |
| AX | 6 | HIJA | 003 | HIJA | HIJA |
| NV | 12 | HIJA>25 | 003 | HIJA | HIJA |
| AX | 23 | HIJASTRA | 028 | HIJASTRA | HIJA |
| VI | 35 | Hija de cónyuge | 028 | HIJASTRA | HIJA |
| AX | 24 | HIJASTRO | 029 | HIJASTRO | HIJO |
| VI | 34 | Hijo de cónyuge | 029 | HIJASTRO | HIJO |
| AX | 5 | HIJO | 002 | HIJO | HIJO |
| NV | 16 | HIJO>18 | 002 | HIJO | HIJO |
| SM | null | NULL | 997 | NULL | INVALIDO/VACIO |
| VI | | VACIO | 996 | VACIO | INVALIDO/VACIO |
| VI | null | NULL | 997 | NULL | INVALIDO/VACIO |
| AX | 27 | SIN DECLARAR | 996 | VACIO | INVALIDO/VACIO |
| VI | 10 | Abuela | 015 | ABUELA | OTROS |
| AX | 7 | ABUELO | 014 | ABUELO | OTROS |
| IN | 011 | AHIJADA | 019 | AHIJADA | OTROS |
| IN | 012 | AHIJADO | 018 | AHIJADO | OTROS |
| AP | 028 | ASEGURADO | NULL | NO IDENTIFICADO/NO COINCIDE | OTROS |
| VI | 28 | Asegurado de Tutor | NULL | NO IDENTIFICADO/NO COINCIDE | OTROS |
| AX | 33 | ASEGURADO GRATIS | NULL | NO IDENTIFICADO/NO COINCIDE | OTROS |
| AX | 31 | CHOFERES | NULL | NO IDENTIFICADO/NO COINCIDE | OTROS |
| VI | 19 | Concubina(o) | 034 | CONCUBINO/CONCUBINA | CONYUGE |
| NV | 23 | CONVIVIENTE | 035 | CONVIVIENTE | CONYUGE |
| VI | 43 | Cuñada | 027 | CUÑADA | OTROS |
| VI | 42 | Cuñado | 026 | CUÑADO | OTROS |
| VI | 31 | Ex Concubina(o) | NULL | NO IDENTIFICADO/NO COINCIDE | OTROS |

| | | | | | |
|---|---|---|---|---|---|
| VI | 30 | Ex Cónyuge (ExEsposo-ExEsposa) | NULL | NO IDENTIFICADO/NO COINCIDE | OTROS |
| AX | 15 | HEREDEROS LEGALES | NULL | NO IDENTIFICADO/NO COINCIDE | OTROS |
| VI | 29 | Heredero(s) Legal(es) | NULL | NO IDENTIFICADO/NO COINCIDE | OTROS |
| NV | 6 | HERMANO | 006 | HERMANO | OTROS |
| VI | 39 | Hermanastra | 033 | HERMANASTRA | OTROS |
| VI | 38 | Hermanastro | 032 | HERMANASTRO | OTROS |
| VI | 6 | Hermano | 006 | HERMANO | OTROS |
| VI | 33 | Madrastra | 031 | MADRASTRA | OTROS |
| VI | 5 | Madre | 005 | MADRE | OTROS |
| VI | 41 | Madre de hijo(s)(a)(as) | NULL | NO IDENTIFICADO/NO COINCIDE | OTROS |
| VI | 27 | Madrina | 021 | MADRINA | OTROS |
| VI | 37 | Media Hermana | 007 | HERMANA | OTROS |
| VI | 18 | Nieta | 017 | NIETA | OTROS |
| VI | 17 | Nieto | 016 | NIETO | OTROS |
| VI | 45 | Nuera (Hija Política) | 024 | NUERA | OTROS |
| IN | 008 | OTRO | OTR | OTROS | OTROS |
| VI | 8 | Otros | OTR | OTROS | OTROS |
| VI | 32 | Padrastro | 030 | PADRASTRO | OTROS |
| VI | 4 | Padre | 004 | PADRE | OTROS |
| VI | 40 | Padre de hijo(s)(a)(as) | NULL | NO IDENTIFICADO/NO COINCIDE | OTROS |
| VI | 26 | Padrino | 020 | PADRINO | OTROS |
| IN | 013 | PRIMA | 013 | PRIMA | OTROS |
| VI | 12 | Sobrina | 009 | SOBRINA | OTROS |
| VI | 11 | Sobrino | 008 | SOBRINO | OTROS |
| AX | 22 | SUEGRO | 023 | SUEGRA | OTROS |
| VI | 24 | Suegra (Madre Política) | 023 | SUEGRA | OTROS |
| NV | 17 | SUEGRO | 022 | SUEGRO | OTROS |
| VI | 23 | Suegro (Padre Político) | 022 | SUEGRO | OTROS |
| NV | 25 | TIA | 011 | TIA | OTROS |
| AX | 10 | TIA | 011 | TIA | OTROS |
| NV | 24 | TIO | 010 | TIO | OTROS |
| AX | 9 | TIO | 010 | TIO | OTROS |
| AX | 30 | TRABAJADORAS DEL HOGAR | NULL | NO IDENTIFICADO/NO COINCIDE | OTROS |
| VI | 25 | Tutor | NULL | NO IDENTIFICADO/NO COINCIDE | OTROS |
| VI | 44 | Yerno (Hijo Político) | 025 | YERNO | OTROS |

| | | | | | |
|---|---|---|---|---|---|
| NV | 0 | TITULAR | 000 | TITULAR | TITULAR |
| AX | 2 | TITULAR FEMENINO | 000 | TITULAR | TITULAR |
| AX | 1 | TITULAR MASCULINO | 000 | TITULAR | TITULAR |

ANEXO 02: Documento Diseño de Sistemas

MAPEO DE DATOS RDV – UDV PARA VIDA

Se realizó el mapeo de datos de Vida, para el UDV, los mismos que se encuentran en los siguientes archivos:

**EVENT**

- Mapeo_UDV_CLV - HD_EVENTO.xlsx

**INSURIX**

- Mapeo_UDV_CLV - HD_POLIZA - INSURIX.xlsx
- Mapeo_UDV_CLV - HD_POLIZA_COBERTURA - INSURIX.xlsx
- Mapeo_UDV_CLV - HD_POLIZA_CUOTA - INSURIX.xlsx
- Mapeo_UDV_CLV - HD_POLIZA_VIDA_AP - INSURIX.xlsx
- Mapeo_UDV_CLV - HD_SINIESTRO - INSURIX.xlsx
- Mapeo_UDV_CLV - MD_PERSONA-BEN-INSURIX.xlsx
- MAPEO_UDV_CLV- MD_PERSONA_ROL_POLIZA-INSURIX.xlsx

**SAM**

- Mapeo_UDV_CLV - HD_POLIZA - SAM.xlsx
- Mapeo_UDV_CLV - HD_POLIZA_COBERTURA - SAM.xlsx
- Mapeo_UDV_CLV - HD_POLIZA_CUOTA - SAM.xlsx
- Mapeo_UDV_CLV - HD_POLIZA_VIDA_AP - SAM.xlsx
- Mapeo_UDV_CLV - HD_SINIESTRO - SAM.xlsx
- Mapeo_UDV_CLV - MD_PERSONA-FAM-BEN-SAM.xlsx
- MAPEO_UDV_CLV- MD_PERSONA_ROL_POLIZA-SAM.xlsx

**SM**

**ANTIGUO**

- Mapeo_UDV_CLV - HD_POLIZA-SM_ANTIGUO.xlsx
- Mapeo_UDV_CLV - HD_POLIZA_VIDA_DESEMBOLSO.xlsx
- Mapeo_UDV_CLV - HD_POLIZA_CUOTA - SM_ANTIGUO.xlsx
- Mapeo_UDV_CLV - HD_SINIESTRO-SM_ANTIGUO.xlsx

- MAPEO_UDV_CLV-MD_PERSONA_ROL_POLIZA_SM_ANTIGUO.xlsx
- Mapeo_UDV_CLV-HD_POLIZA_VIDA_MASIVO-SM-ANTIGUOS.xlsx
- Mapeo_UDV_CLV-MD_PERSONA-CON-ASE-SM_ANTIGUO.xlsx

**NUEVO**

- Mapeo_UDV_CLV - HD_POLIZA_CUOTA - SM_NUEVO.xlsx
- Mapeo_UDV_CLV-HD_POLIZA_VIDA_MASIVO-SM_NUEVO.xlsx
- Mapeo_UDV_CLV - HD_POLIZA-SM_NUEVO.xlsx
- Mapeo_UDV_CLV - HD_SINIESTRO-SM_NUEVO.xlsx
- MAPEO_UDV_CLV-MD_PERSONA_ROL_POLIZA_SM_NUEVO.xlsx
- Mapeo_UDV_CLV - MD_PERSONA-CON-ASE-SM.xlsx

**TUP**

- Mapeo_UDV_CLV - MD_PERSONA.XLSX

**VIAP**

- Mapeo_UDV_CLV - HD_POLIZA - VIAP.xlsx
- Mapeo_UDV_CLV - HD_POLIZA_COBERTURA - VIAP.xlsx
- Mapeo_UDV_CLV - HD_POLIZA_CUOTA - VIAP.xlsx
- Mapeo_UDV_CLV - HD_POLIZA_VIDA_AP - VIAP.xlsx
- Mapeo_UDV_CLV-hD_POLIZA_VIDA_MOVIMIENTO_INTERMEDIARIO.xlsx
- Mapeo_UDV_CLV-HD_POLIZA_VIDA_SALDO_CUENTA-VIAP.xlsx
- Mapeo_UDV_CLV - HD_POLIZA_VIDA_VISITAS.xlsx
- Mapeo_UDV_CLV - HD_SINIESTRO - VIAP.xlsx
- Mapeo_UDV_CLV-MD_PERSONA_ROL_POLIZA_FAM_BENEF_VIAP.xlsx
- Mapeo_UDV_CLV - MD_PERSONA_ROL_POLIZA_VIAP.xlsx
- Mapeo_UDV_CLV - MD_PERSONA-FAM-BEN-VIAP.xls

ANEXO 03: Mapeo de entidad UDV Persona para aplicativo SAM MASIVO

Mapeo UDV PERSONA – SAM MASIVO:

## MODELO UDV - CLV - SEGUROS

| Descripción de Entidad | |
|---|---|
| N° | 0001 |
| Origen | C/S |
| Categoria | |
| Modelo - CU | CLV |
| Esquema Datalake | UDV |
| Nombre de Entidad | MD_PERSONA |
| Nombre Estructura Archivo | |
| Exposición de Información | Full |
| Almacenamiento | Disco |
| # Registros Iniciales | |
| Periodicidad | L-D |
| Regs. nuevos por periodo | |
| Tipo de Dato provista por la Fuer | Full |

| ¿Qué información contiene esta entidad? | | |
|---|---|---|
| Información de Contratante y Asegurado SM | | |
| **Tipo de información :** | Ad-hoc | *Histórico de Pólizas emitidas por Proc* |

| Historia del Documento | | | | | |
|---|---|---|---|---|---|
| Fecha de Inicio | Fecha de Término | Tipo | Versión | Nombre | Observación |
| 28/10/2019 | 28/10/2019 | Versión Inicial Mapeo | 0.0 | Merydith O. | Generación del documento. |
| 17/04/2020 | 17/04/2020 | | 1.0 | Merydith O. | Actualización Mapeo por Definicion Persona Sam Masivo |
| | | | | | |

| Tablas FS-LDM de Búsqueda | | |
|---|---|---|
| IPSM_PERSONA_SM | | |
| IPSM_DIRECCION_SM | | |
| IPSM_PARAMETROS | | |
| | | |
| | | |

| ENTIDAD EN UDV | | | | | | FUENTE-ORIGEN DATOS | | |
|---|---|---|---|---|---|---|---|---|
| MD_PERSONA | | | | | | BDATOS.TIPO: SQL SERVER / NOMBRE: EPPS / ESQUEMA: EPPS | | |
| Nombre de Columna | Descripción de columna | Tipo | Tamaño | Llave | Observaciones | Tabla | Columna | Transformación |
| CODORIGEN | Aplicación de Origen del Dato | String | 2 | Si | Constante 'SM' | | | Constante 'SM' |
| NUMID | Identificador Persona Unica | String | 12 | Si | Se genera el campo. Verificar la columna de transformación. | | | ---Para los Contratantes:<br>Select 'N' + ltrim(rtrim(C.SOLICITUD)) +<br>CASE WHEN ISNULL(PER.TIPO_DOCUMENTO ,0) = 0 then<br>E.TipoDocumento ELSE PER.TIPO_DOCUMENTO END +<br>CASE WHEN ISNULL(PER.NUM_DOCUMENTO,'0') = '0' then<br>E.NumeroDocumento ELSE PER.NUM_DOCUMENTO END<br>CodigoPersonaUnico<br>from IPSM_POLIZA_SM PO<br>Inner Join IPSM_CONTRATANTE_SM C on C.SOLICITUD = PO.SOLICITUD<br>and C.POLIZAID = PO.ID_POLIZA<br>Inner Join IPSM_PERSONA_SM PER on PER.Id_Persona = C.Personald<br>LEFT JOIN #ENTIDADES E on E.COD_ENTIDAD = PO.COD_ENTIDAD<br><br>---Para los Asegurados:<br>Select 'N' + ltrim(rtrim(C.SOLICITUD) +<br>CASE WHEN ISNULL(PER.TIPO_DOCUMENTO ,0) = 0 then<br>E.TipoDocumento ELSE PER.TIPO_DOCUMENTO END +<br>CASE WHEN ISNULL(PER.NUM_DOCUMENTO,'0') = '0' then<br>E.NumeroDocumento ELSE PER.NUM_DOCUMENTO END<br>CodigoPersonaUnico |
| TIPDOCMTO | Tipo Documento Persona | String | 4 | No | | PSM_PERSONA_SM | Tipo_Documento | |
| NRODCMNTO | Numero Documento Persona | String | 15 | No | | PSM_PERSONA_SM | Num_Documento | |
| NOMPERS | Nombres Persona | String | 50 | No | | PSM_PERSONA_SM | Nombres | |
| APEPATPERS | Apellido Paterno Persona | String | 50 | No | | PSM_PERSONA_SM | Apel_Paterno | |
| APEMATPERS | Apellido Materno Persona | String | 50 | No | | PSM_PERSONA_SM | Apel_Materno | |
| NOMBCOMP | Nombre Completo Persona | String | 150 | No | SE DEBEN CONCATENAR LOS CAMPOS | PSM_PERSONA_SM | Nombre_Completo | |
| NOMBCOMER | Nombre Comercial | String | 50 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| CODTIPCLIENTE | Codigo Tipo de Cliente | String | 2 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DESCTIPCLIE | Descripcion Tipo de Cliente | String | 20 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| CODTIPPERSON | Codigo Tipo de Persona | String | 2 | No | | PSM_PERSONA_SM | Tipo_Persona | |
| DESCTIPPERS | Descripcion Tipo de Persona | String | 20 | No | | PSM_PARAMETROS | Desc_Param | Select P.DESC_PARAM from IPSM_PERSONA_SM PER Inner join IPSM_PA |
| FECNCMNTO | Fecha de Nacimiento | String | 10 | No | | PSM_PERSONA_SM | Fecha_Nacimiento | |
| FECDECESO | Fecha de Fallecimiento | String | 10 | No | NO APLICA PARA FAMILIARES | | | |
| SEXO | Sexo Persona | String | 2 | No | | PSM_PERSONA_SM | Sexo | |
| CODESTCIV | Código Estado Civil Persona | String | 4 | No | | PSM_PERSONA_SM | Est_Civil | |
| DESCESTCIV | Estado Civil Persona | String | 20 | No | | PSM_PARAMETROS | Adic_Param | Select P.ADIC_PARAM from IPSM_PERSONA_SM PER Left join IPSM_PAR |
| CODUBIC | Codigo Ubigeo | String | 8 | No | | PSM_DIRECCION_SM | Ubigeo | SELECT DIR.Ubigeo FROM IPSM_PERSONA_SM PER Left Join IPSM_DIREC |
| CODDPTO | Codigo Departamento | String | 8 | No | | PSM_DIRECCION_SM | Cod_Departamento | SELECT DIR.Cod_Departamento FROM IPSM_PERSONA_SM PER Left Join I |
| DESCDPTO | Departamento | String | 20 | No | | | | Select |
| CODPROV | Codigo Provincia | String | 8 | No | | PSM_DIRECCION_SM | Cod_Provincia | SELECT DIR.Cod_Provincia FROM IPSM_PERSONA_SM PER Left Join IPSM |
| DESCPROV | Provincia | String | 20 | No | | | | Select SUBSTRING (U.DESCRIPCION, CHARINDEX('-',U.DESCRIPCION, 1) + 1, |
| CODDISTR | Codigo Distrito | String | 8 | No | | PSM_DIRECCION_SM | Cod_Distrito | SELECT DIR.Cod_Distrito FROM IPSM_PERSONA_SM PER Left Join IPSM_D |
| DESCDISTR | Distrito | String | 20 | No | | | | Select SUBSTRING (U.DESCRIPCION, 1, CHARINDEX('-',U.DESCRIPCION, 1) - |
| DIRECC | Direccion | String | 250 | No | | PSM_DIRECCION_SM | Direccion | SELECT DIR.Direccion FROM IPSM_PERSONA_SM PER Left Join IPSM_DIR |
| CODTIPCALLE | Codigo Tipo de Calle | String | 4 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DESCTIPCALLE | Descripcion Tipo de Calle | String | 20 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| NOMBCALLE | Direccion - Nombre Calle | String | 50 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| NROCALLE | Direccion - Numero de Calle | String | 10 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DIRECLOTE | Direccion - Lote | String | 10 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DIRECMZNA | Direccion - Manzana | String | 10 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| CODTIPINTER | Direccion - Codigo Tipo de Interior | String | 4 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DESCTIPINTER | Direccion - Descripcion Tipo de Interior | String | 20 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DIRECINTERIOR | Direccion - Interior | String | 8 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DIRECPISO | Direccion - Piso | String | 8 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| CODTIPURB | Direccion - Codigo Tipo de Urbanizacion | String | 4 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DESTIPURB | Direccion - Descripcion Tipo Urbanizacion | String | 20 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DIRECURBAN | Direccion - Urbanizacion | String | 50 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DIRECEDIFIC | Direccion - Edificio | String | 50 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DIRECREFER | Direccion - Referencia | String | 50 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DIRECJTOHAB | Direccion - Conjunto Habitacional | String | 50 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DIRECRESTO | Direccion - Resto Direccion | String | 50 | No | | PSM_DIRECCION_SM | Resto | SELECT DIR.Resto FROM IPSM_PERSONA_SM PER Left Join IPSM_DIREC |
| DIRECCORDX | Coordenada X | String | 15,6 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| DIRECCORDY | Coordenada Y | String | 15,6 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |
| CODCIU | Codigo de Actividad Economia | String | 4 | No | NO APLICA PARA CONTRATANTE MASIVO | | | |