

UNIVERSIDAD NACIONAL DE INGENIERÍA FACULTAD
DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS



ANALISIS DE CUMPLIMIENTO DE METAS DE TRADING EN INTERBANK

Curso: SISTEMA DE INTELIGENCIA DE NEGOCIOS
Sección: SI 807-U

Alumnos:

Gamboa Checnes Joel Fernando	20212635D
Aymachoque Aymachoque Luis Jairo	20191144G
Laureano Hidalgo Jordan Cesar	20212559F

Ciclo: 25-2

Lima, 13 de octubre de 2025

ÍNDICE

Introducción	3
1. Generalidades de la Empresa	3
2. Problema de Negocio	4
3. Necesidades de Información y Decisiones Críticas	5
4. KPIs	6
5. Implementación del Hortonworks	7
6. Arquitectura Inicial	10
7. Preguntas de negocio	11
8. KPIs Claves	12
9. Modelo Conceptual Preliminar	14
10. Inventario de Fuentes OLTP	15
11. Implementación del Hortonworks	16

Introducción

El Banco Internacional del Perú S.A.A. – Interbank es una de las instituciones financieras más reconocidas del país, dedicada a ofrecer productos y servicios que abarcan desde cuentas de ahorro, créditos y tarjetas, hasta soluciones digitales e inversiones. Con sede en Lima, Interbank se ha consolidado como un banco universal que busca mejorar la calidad de vida de sus clientes mediante un servicio ágil, accesible y cercano. No obstante, como toda organización en un entorno altamente competitivo, enfrenta desafíos en áreas críticas como el trading, donde factores externos e internos generan incertidumbre respecto al cumplimiento de las metas propuestas. Analizar estas dificultades resulta esencial para diseñar estrategias que permitan optimizar la gestión, motivar al equipo y fortalecer la posición de la entidad en el mercado.

1. Generalidades de la Empresa

Nombre o razón social	Banco Internacional del Perú S.A.A. – Interbank
Giro de la empresa	Institución financiera universal, banco comercial

	que ofrece productos y servicios financieros para personas naturales, empresas e instituciones, incluyendo cuentas, créditos, tarjetas, inversiones, seguros, medios digitales.
Ubicación	Av. Carlos Villarán N° 140, La Victoria, Lima, Perú
Misión y visión	Mejorar la calidad de vida de nuestros clientes, brindando un servicio ágil y amigable en todo momento y en todo lugar. Ser el mejor banco a partir de las mejores personas.
Productos y clientes	<p>Productos destacados:</p> <ul style="list-style-type: none"> • Cuentas de ahorro, corrientes, CTS, cuentas digitales. • Depósitos a plazo. • Créditos personales, hipotecarios, con garantía, por convenio. • Tarjetas de crédito y débito. • Programas de beneficios, banca digital. <p>Clientes principales:</p> <ul style="list-style-type: none"> • Personas naturales. • Empresas. • Usuarios de canales digitales. • Público con necesidades de ahorro, crédito e inversión.

2. Problema de Negocio

Dificultades para alcanzar las metas propuestas en utilidades en ciertos meses en el área de trading.

Descripción

En el área Transaccional, se dividen en 4 partes, las cuales son trading, saldos, Izipay y principalidad, de las cuales, por orden de la gerencia comercial, cada uno tiene una meta proyectada a la cual debe llegar para poder seguir en el mercado competitivo y aspirar a la misión que se tiene en general como banco. Es aquí donde encontramos el área de Trading, que por cada mes debe intentar lograr el reto que se le imponga para poder seguir creciendo como área y como banco en general. Sin embargo, se sabe que esta meta es muy impredecible de saber si se va o no a poder lograr, ya que hay diversos factores que influyen sobre el poder llegar a la meta, estos pueden ser desde la cantidad de días útiles que tiene el mes hasta el comportamiento mismo del mes en el cual se intente llegar a la meta. Por estos motivos de incertidumbre, hay veces en donde el área no ha podido llegar a la meta propuesta por la gerencia comercial, raspando a veces el presupuesto generado para trading (mínimo de utilidad que deben hacer para que el área no tenga pérdidas monetarias)

Si se pudiera saber si un mes tiene altas o pocas probabilidades de obtener una gran utilidad, se podrían armar mejores estrategias para poder maximizar esa utilidad, es por ello por lo que se desea analizar este problema.

Analizando los dolores del área, encontramos los siguientes:

- Creación lenta de textos para las comunicaciones: Los textos que van en las comunicaciones que se le envía al cliente para animarlo a que cambie su dinero, suelen hacerse el mismo día, perjudicando la hora de programación de envío
- Envío de cards de comunicación lenta: El área de trading no es la única área que envía comunicaciones, generándose así un cuello de botella
- Poca motivación de los ejecutivos: Existen campañas en las cuales se premia al ejecutivo que haga más utilidad en trading, junto con sus desembolsos, pero si en la campaña (que dura 2 meses) no cumplen con su meta en 1 mes, estos ya no intentan hacer más trading en sus desembolsos porque saben que ya no están en competencia.
- Mes cuyo día útil final no coincide con el último día de mes: El día donde se suele hacer más utilidad en el mes, normalmente es el último día hábil, pero si este cae en un día no laboral, lo que se genere en ese día pasará a contarse como utilidad para el siguiente mes.
- Fugas de clientes: Se han reportado clientes cuyo promedio de utilidad en los últimos 3 meses es mayor a 1500 pero en el último, la utilidad que generan es menor a 500, lo cual genera una pérdida de oportunidad para hacer más utilidad.
- Información de la performance de comunicaciones desactualizada: Normalmente no se puede hacer un estudio de la performance de una comunicación ya que la información suele tardar en actualizarse y en muchos casos, si deseas obtenerla, debes hacerlo manualmente.
- Limitaciones de envío de base de clientes por presupuestos: El área trading comparte presupuesto con otras áreas en la VP comercial, es por ello que el envío de cierto tipo de comunicación se ve perjudicada (por ejemplo, por cada comunicación de tipo SMS, solo podemos enviar hasta 10K clientes por cada comunicación)
- Demora en la creación de artes para ciertas comunicaciones: Existe un área llamada make en la cual se editan todos los flyers que se quieran comunicar, pero estas normalmente demoran aproximadamente 2 semanas en elaborarlo por motivos de capacidad.

3. Necesidades de Información y Decisiones Críticas

Nivel	Tipo de Decisión	Necesidad de información
Estratégico	Definir la estrategia de crecimiento y rentabilidad a largo plazo. Fijar metas anuales de utilidades. Aprobar presupuestos de marketing y comunicaciones.	Información de Mercado y Tendencias: Tendencias macroeconómicas que afectan el trading. Análisis de la competencia (tasas, productos, estrategias). Análisis de fuga de clientes a nivel macro.

Táctico	Planificar y ejecutar las campañas de trading. Asignar presupuestos de comunicaciones por tipo. Diseñar promociones y campañas para alcanzar metas mensuales. Definir el mensaje y el arte de las comunicaciones.	Información de Rendimiento y Operativa: Performance de comunicaciones pasadas (CTR, conversión, utilidades generadas). Análisis de la base de clientes por segmento para una segmentación efectiva. Costos de envío de comunicaciones para optimizar el presupuesto.
Operativo	Gestionar las operaciones diarias del equipo. Priorizar la creación de artes y textos. Enviar comunicaciones a clientes en el momento adecuado. Motivar al equipo de ejecutivos.	Información de Actividad y Flujo de Trabajo: Estado actual del inventario de artes y textos para comunicaciones. Registro de las comunicaciones enviadas y sus resultados inmediatos.

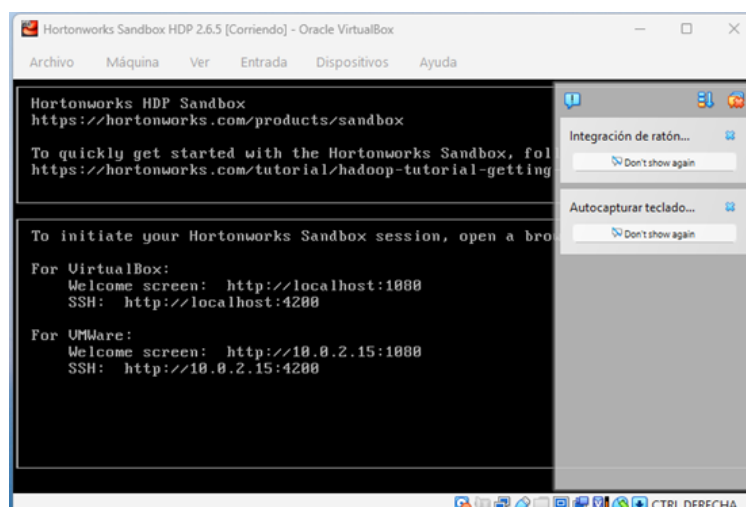
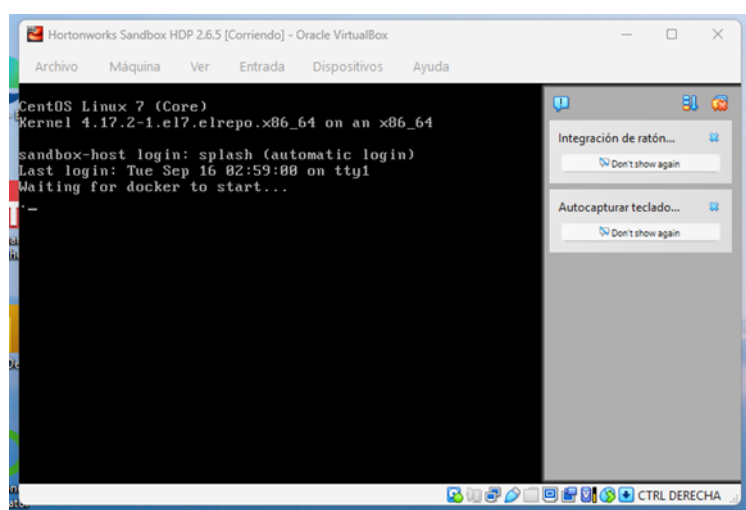
4. KPIs

Nombre del KPI	Descripción	Fórmula	Unidad de medida	Frecuencia
Cumplimiento de Meta de Utilidades Mensual	Mide el porcentaje de cumplimiento de la meta de utilidades establecida para el área de trading en un mes específico.	$(\text{Utilidad Real del Mes} / \text{Meta de Utilidad del Mes}) \times 100$	%	Mensual
Utilidad Promedio Generada por Ejecutivo	Evalúa la productividad promedio de cada ejecutivo de trading, permitiendo identificar desempeños excepcionales o problemas específicos con ejecutivos particulares.	$\text{Utilidad Total del Mes} / \text{Número de Ejecutivos Activos}$	Dólares	Mensual
Tasa de Conversión por Comunicación	Mide la efectividad de las comunicaciones enviadas a clientes en términos de generar la acción deseada	$(\text{Número de clientes que realizaron la acción deseada} / \text{Total de clientes que recibieron la comunicación}) \times 100$	%	Por campaña o semanalmente
Tasa de Fuga de Clientes Rentables	Cuantifica el porcentaje de clientes rentables que dejaron de operar con el banco durante un período determinado	$(\text{Número de clientes rentables que cerraron sus posiciones o se fueron} / \text{Total de clientes rentables al inicio del período}) \times 100$	%	Mensual
Tiempo Medio de Creación de una Comunicación	Mide el tiempo promedio que toma crear y aprobar el contenido textual de una comunicación, desde la recepción del brief inicial hasta la aprobación final del texto	$\Sigma (\text{Tiempo de creación por comunicación}) / \text{Número total de comunicaciones}$	horas	Semanal
Tiempo Medio de Producción de Arte	Cuantifica el tiempo promedio requerido para producir material gráfico una vez aprobado el texto.	$\Sigma (\text{Tiempo de producción de arte por comunicación}) / \text{Número total de comunicaciones}$	horas	semanal
Ciclo Total de Lanzamiento de una Campaña	Mide el tiempo completo desde la concepción de la idea de una campaña hasta su envío final a los clientes.	$\Sigma (\text{Fecha de envío} - \text{Fecha de ideación}) / \text{Número total de campañas}$	Días	Por campaña

Tasa de Apertura / Click-Through Rate (CTR) por Comunicación	Mide el porcentaje de clientes que abren una comunicación (email, app notificación) o hacen clic en los enlaces incluidos.	(Número de aperturas o clicks / Total de comunicaciones enviadas) × 100	%	Semanalmente
Costo por Adquisición/Retención de Cliente (CAC/CRC)	Calcula la inversión promedio requerida para adquirir un nuevo cliente o retener uno existente.	Costo total de adquisición/retención / Número de clientes adquiridos y/o retenidos	dólares	Mensual
% de Días Hábiles Efectivos vs. Mes Natural	Calcula el porcentaje que representan los días hábiles efectivos frente al total de días del mes.	(Número de días hábiles en el mes / Total de días del mes) × 100	%	Mensual

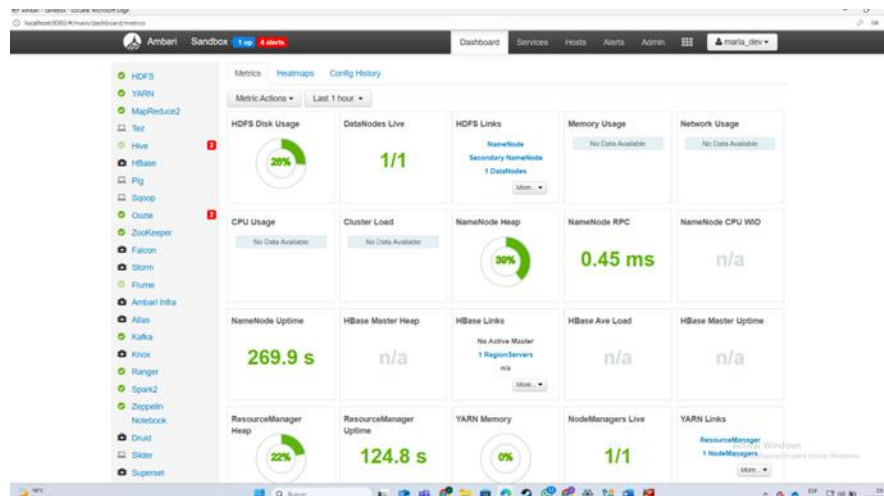
5. Implementación del Hortonworks

5.1 Arranque Inicial del Hortonworks:

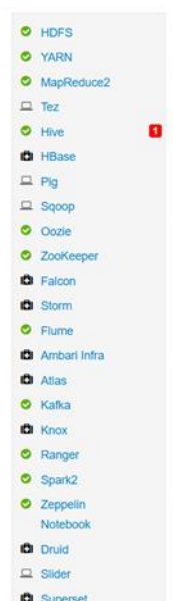


Con el ip dado por el hortonworks iniciar en local host el Ambari.

5.2 Local Host:



5.3 Servicios Disponibles:



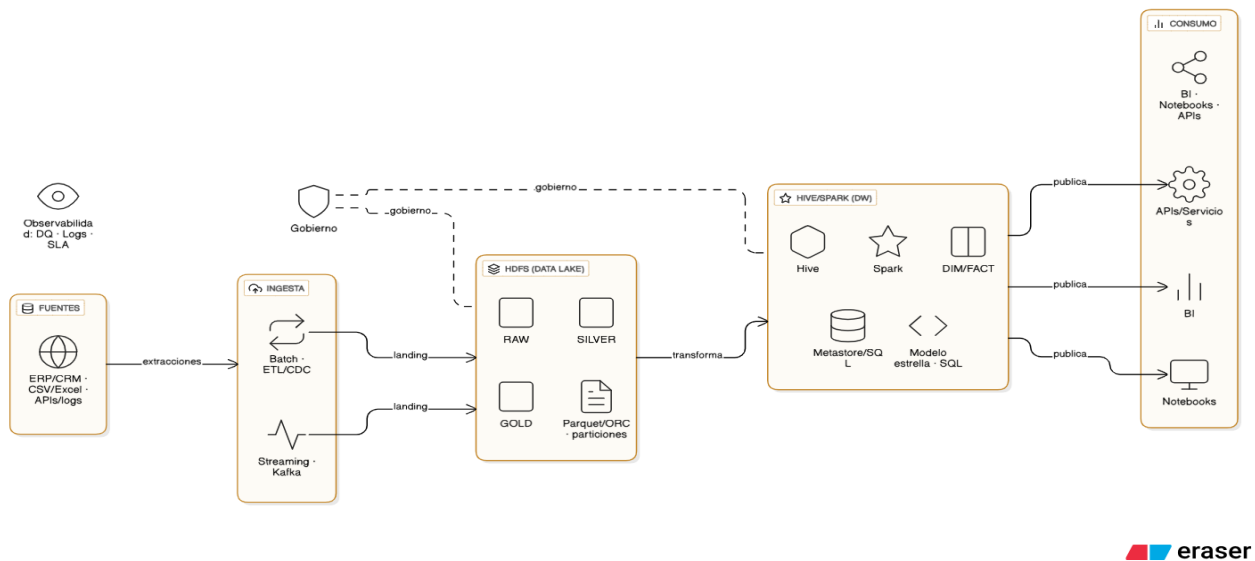
5.4 Listar Archivos HDFS:

HDFS, que significa Hadoop Distributed File System, es un sistema de archivos distribuido de código abierto que se utiliza para almacenar grandes volúmenes de datos. Es un componente central del ecosistema de Apache Hadoop y está diseñado para funcionar en hardware de bajo costo, lo que lo hace muy escalable y económico.

Para listar los archivos se utiliza el comando: `hdfs dfs -ls /`

6. Arquitectura Inicial

Después de todo lo mostrado, se procede a mostrar una arquitectura inicial del tema:



6.1 FUENTES (Sources)

Este es el punto de partida de los datos. La arquitectura puede extraer información de una variedad de sistemas, como:

ERP/CRM: Sistemas de planificación de recursos empresariales y gestión de relaciones con clientes.

CSV/Excel: Archivos planos.

6.2 INGESTA (Ingestion)

Es la fase donde se capturan los datos de las fuentes y se preparan para su almacenamiento. Hay dos métodos principales mostrados:

Batch - ETL/CDC: Para procesar datos en lotes. Se usan herramientas de ETL (Extract, Transform, Load) para extraer, transformar y cargar datos de manera periódica, o de CDC (Change Data Capture) para capturar solo los cambios en la fuente.

Streaming - Kafka: Para datos en tiempo real. Kafka es una plataforma de streaming que maneja flujos de datos continuos.

6.3 HDFS (DATA LAKE)

Los datos ingeridos se almacenan en un Data Lake utilizando HDFS (Hadoop Distributed File System). Este lago de datos está estructurado en diferentes "capas" de calidad:

RAW: La primera capa, donde los datos se almacenan tal como vienen de las fuentes, sin ninguna modificación.

SILVER: La segunda capa, donde los datos son limpiados, estandarizados y enriquecidos. Aquí se resuelven problemas de calidad de datos.

GOLD: La capa final, con datos de alta calidad. Contiene tablas agregadas y optimizadas, listas para el consumo. Aquí se usan formatos de archivo como Parquet u ORC y particiones para mejorar el rendimiento de las consultas.

6.4 HIVE/SPARK (DW)

Desde la capa GOLD, los datos transformados se mueven a un Data Warehouse (DW). Aquí se utilizan herramientas para procesar y consultar los datos:

6.5 Hive: Permite realizar consultas SQL sobre los datos almacenados en el Data Lake.

Spark: Un motor de procesamiento distribuido para transformaciones complejas.

Modelo estrella - SQL: Los datos se organizan en un esquema de estrella, que es un diseño optimizado para consultas analíticas, con tablas de hechos y tablas de dimensiones.

6.6 GOBIERNO (Governance) y OBSERVABILIDAD (Observability)

Estas son capas transversales que aseguran la calidad y el monitoreo del sistema:

Gobierno: Asegura que los datos sean consistentes, seguros y cumplan con las políticas de la organización a lo largo de todo el proceso de ingesta y almacenamiento en el Data Lake.

Observabilidad: Se refiere al monitoreo del pipeline. Incluye la calidad de los datos (D.Q.), el registro de eventos (Logs) y el cumplimiento de los acuerdos de nivel de servicio (SLA).

6.7 CONSUMO (Consumption)

La etapa final, donde los datos están disponibles para los usuarios finales y las aplicaciones.

BI: Herramientas de Inteligencia de Negocios para crear reportes y dashboards.

Notebooks: Para que los científicos e ingenieros de datos realicen análisis exploratorios y desarrollen modelos.

APIs/Servicios: Permite que otras aplicaciones accedan a los datos del Data Warehouse de manera controlada.

7. Preguntas de negocio

Las preguntas de negocio constituyen el punto de partida para la construcción del sistema de inteligencia de negocios, ya que permiten identificar las principales necesidades de información de la organización. A través de ellas se busca dar respuesta a problemáticas clave como la variación en los días hábiles de cada mes, el cumplimiento de metas de utilidad, la productividad de los ejecutivos y la retención de clientes rentables. Estas preguntas orientan la definición de indicadores y el diseño del modelo conceptual, garantizando que la solución propuesta no solo se centre en medir resultados, sino también en anticipar riesgos y apoyar la toma de decisiones estratégicas.

Area	Rol de usuario	Pregunta de negocio	Nivel de prioridad	Fuente de datos
Dirección / Trading	Director de Trading	Cuando el mes trae menos días hábiles que el promedio, ¿qué ajuste de volumen y timing de contactos/comunicaciones debemos ejecutar para mantener el run-rate de utilidades?	Alta	Datos históricos de utilidades por día útil
Comercial	Gerente Comercial	¿Qué ejecutivos están por debajo del umbral de utilidad promedio?	Alta	Utilidad generada por ejecutivo por canal
Cientes / CRM	Jefe de fidelización	¿En qué segmentos se concentra la fuga de clientes rentables y qué acción de retención priorizamos para proteger la utilidad del mes?	Alta	Segmentación de clientes por período

Dirección / Trading	Director de Trading	¿Cómo se compara el avance de la utilidad de este mes con meses recientes a la misma altura de días hábiles?	Media	Datos históricos de utilidades por día útil
Marketing / Comunicaciones	Jefe de Comunicaciones	¿En qué momentos del mes la conversión suele ubicarse por encima o por debajo de su promedio reciente?	Media	Número de comunicaciones por mes

8. KPIs Claves

Campo	Valor Completado
Nombre del KPI	Cumplimiento de Meta de Utilidades Mensual
Descripción	Mide el porcentaje de cumplimiento de la meta de utilidades establecida para el área de trading en un mes específico
Fórmula	$(\text{Utilidad Real del Mes} / \text{Meta de Utilidad del Mes}) * 100$
Unidad de medida	%
Frecuencia	Mensual
Responsable	Gerente de Trading
Fuente de Datos	Sistema de Trading
Meta	100%
Semaforo	● < 85% - Crítico
	● 85% - 95% - En riesgo
	● > 95% - Cumplido

- Cumplimiento de Meta de Utilidades Mensual**
 Por qué es clave: Es el indicador de resultado final que quieren predecir
 Importancia: Establece la línea base para análisis predictivos

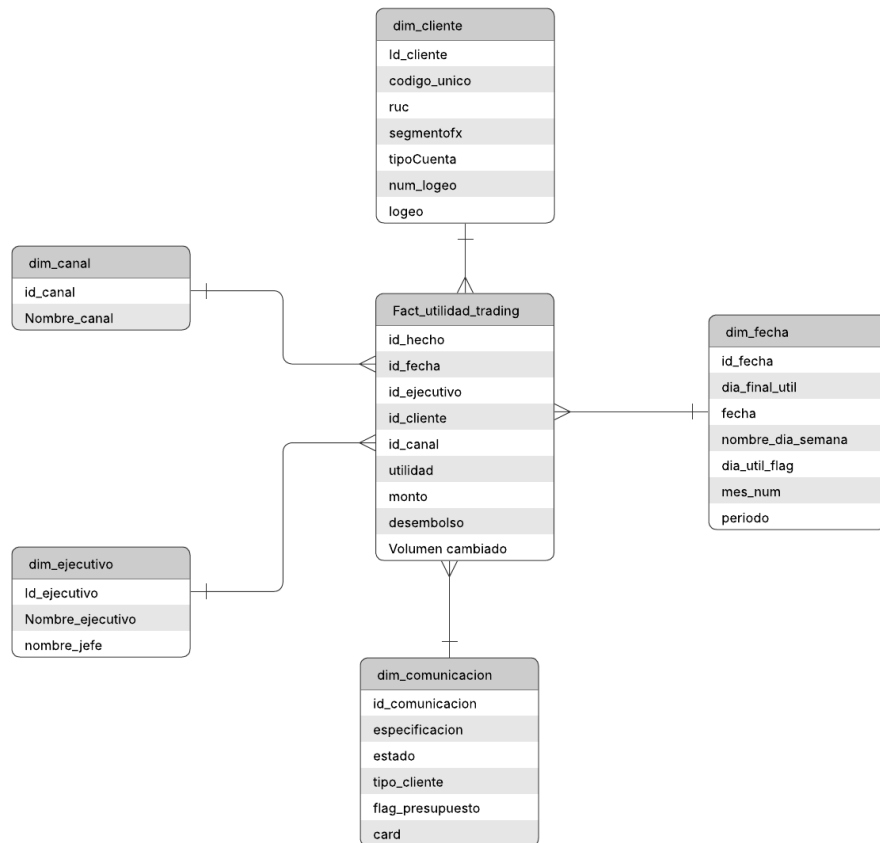
Campo	Valor Completado
Nombre del KPI	Utilidad Promedio Generada por Ejecutivo
Descripción	Evalúa la productividad promedio de cada ejecutivo de trading, permitiendo identificar desempeños excepcionales o problemas específicos con ejecutivos particulares
Fórmula	Utilidad Total del Mes / Número de Ejecutivos Activos
Unidad de medida	Dólares
Frecuencia	Mensual
Responsable	Supervisor de Trading
Fuente de Datos	Sistema de Trading + Nómina/HR
Meta	\$15,000 USD
Semaforo	● < \$10,000 - Bajo desempeño
	● \$10,000 - \$14,999 - Medio desempeño
	● ≥ \$15,000 - Alto desempeño

- Utilidad Promedio por Ejecutivo
Por qué es clave: Indica la capacidad productiva del equipo
Valor predictivo: Un promedio bajo anticipa dificultades para alcanzar la meta

Campo	Valor Completado
Nombre del KPI	Tasa de Fuga de Clientes Rentables
Descripción	Cuantifica el porcentaje de clientes rentables que dejaron de operar con el banco durante un período determinado
Fórmula	$\left(\frac{\text{Número de clientes rentables que cerraron sus posiciones o se fueron}}{\text{Total de clientes rentables al inicio del período}} \times 100 \right)$
Unidad de medida	%
Frecuencia	Mensual
Responsable	Jefe Comercial / Gerente de Clientes
Fuente de Datos	CRM + Sistema de Trading + Base de Datos de Clientes
Meta	≤ 5%
Semaforo	● > 8% - Crítico
	● 5.1% - 8% - En riesgo
	● ≤ 5% - Aceptable

- Tasa de Fuga de Clientes Rentables
Por qué es clave: Identifica pérdida de ingresos recurrentes
Impacto predictivo: Alta tasa de fuga = menor probabilidad de alcanzar meta

9. Modelo Conceptual Preliminar



- TABLA DE HECHOS: fact_utilidad_trading

¿Por qué existe?

Es el corazón del análisis - contiene todas las métricas medibles

Registra utilidad diaria para detectar patrones temporales

Conecta todos los factores que influyen en el resultado

utilidad	¿CUÁNTO ganamos?
volumen_usd	¿CUÁNTO volumen manejamos?
monto	¿A QUÉ debemos llegar?
porcentaje_meta	¿QUÉ TAN cerca estamos?

- DIMENSIÓN FECHA: dim_fecha

Problema que resuelve:

"Mes cuyo día útil final no coincide con el último día de mes"

Análisis estacionalidad y patrones temporales

- DIMENSIÓN COMUNICACIÓN: dim_comunicacion

Problemas que resuelve:

"Creación lenta de textos para comunicaciones"

"Información de performance desactualizada"

"Limitaciones de presupuesto en envíos"

- DIMENSIÓN EJECUTIVO: dim_ejecutivo

Problema que resuelve:

"Poca motivación de los ejecutivos"

Seguimiento de campañas y ranking

- DIMENSIÓN CLIENTE: dim_cliente

Problema que resuelve:

"Fugas de clientes"

Segmentación para comunicaciones efectivas

- DIMENSIÓN CANAL: dim_canal

Problema que resuelve:

Definir los canales que me generan mayores utilidades.

10. Inventario de Fuentes OLTP

Sistema	Área Usuaria	Tipo	Tecnología	Frecuencia Actualización	Observaciones
CRM Comercial	Área Comercial	Datos de desempeño	CRM interno	Semanal	Ranking de utilidad por ejecutivo y por campaña
Sistema de Trading	Área de Trading	Datos transaccionales	Base de datos SQL	Diaria	Datos históricos de utilidades por día útil
Plataforma de Comunicaciones	Marketing Digital	Datos operativos	Marketing Cloud/Herramientas de emailing	En tiempo real	Número y tipo de comunicaciones por mes
Sistema de Clientes	Área Comercial	Datos maestros	Base de datos cliente	Mensual	Segmentación de clientes por periodo y comportamiento

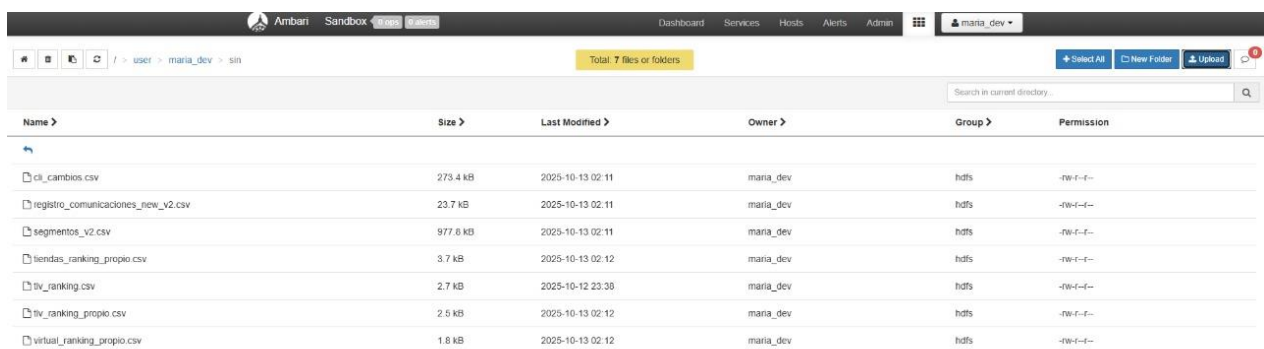
11. Implementación del Hortonworks

11.1 Dashboard: Cuando hayamos accedido a Ambari y queramos ingresar información para empezar a manejar, lo primero que debemos hacer es dirigirnos a los cuadritos pequeños que se encuentran en la parte superior derecha de la pantalla del dashboard.

11.2 Archivos: Una vez accedido a los archivos, deberemos crear una carpeta para subir nuestros archivos a manejar.

11.3 Folder: Una vez creemos el nombre del folder donde subiremos nuestros archivos de trabajo, le damos a “Add” y listo.

11.4 Subida de Archivos: Ahora que creamos la carpeta podremos subir los archivos que deseamos para su posterior analisis.



Name	Size	Last Modified	Owner	Group	Permission
ci_cambios.csv	273.4 kB	2025-10-13 02:11	maria_dev	hdfs	-rw-r--r--
registro_comunicaciones_new_v2.csv	23.7 kB	2025-10-13 02:11	maria_dev	hdfs	-rw-r--r--
segmentos_v2.csv	977.6 kB	2025-10-13 02:11	maria_dev	hdfs	-rw-r--r--
tiendas_ranking_propio.csv	3.7 kB	2025-10-13 02:12	maria_dev	hdfs	-rw-r--r--
tv_ranking.csv	2.7 kB	2025-10-12 23:38	maria_dev	hdfs	-rw-r--r--
tv_ranking_propio.csv	2.5 kB	2025-10-13 02:12	maria_dev	hdfs	-rw-r--r--
virtual_ranking_propio.csv	1.8 kB	2025-10-13 02:12	maria_dev	hdfs	-rw-r--r--

11.5 Creación de Tabla: Ya tenemos la información subida, ahora podremos crear una tabla que contenga la información que subimos y así poder manejarla a nuestro antojo.

11.6 Query: Una vez creada la tabla, procedemos a hacer las consultas que deseamos y poder manejar la

información a nuestro antojo.

Worksheet

```
1 SELECT * FROM cli_cambios;
2 WHERE FECHA >= '2020-01-01';
```

Execute Explain Upload Save as... New Worksheet

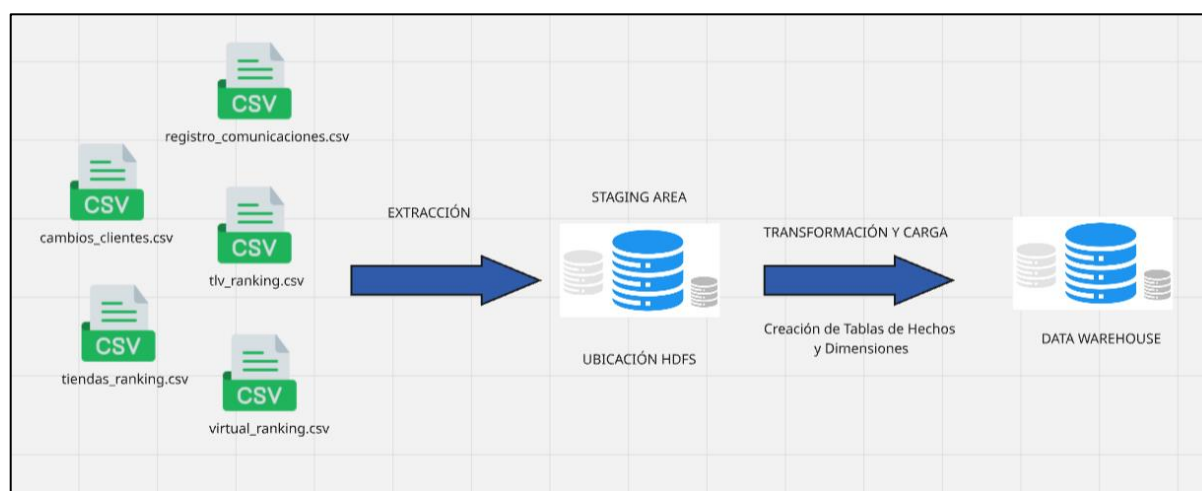
Query Process Results (Status: SUCCEEDED) Save results...

Logs Results

Filter columns:

cli_cambios.fecha	cli_cambios.cu	cli_cambios.ruc	cli_cambios.nombre	cli_cambios.tipo	cli_cambios.utilidad	cli_cambios.vol_usd
2020-01-08	77077553	24414966791	Charlie SAC	V	0.0	5903.8
2020-01-21	13617344	20616249531	Trujillo SA	V	0.0	5463.17
2020-01-29	15977290	00530618078	Echo SA	V	2525.47	10437.64
2020-02-08	91606767	31212002311	Golf EIRL	V	300.74	881.28

12. Diseño del Proceso ETL



Etapa de Fuentes (Source) y Staging (Preparación):

En esta etapa, los archivos CSV que contienen registros de clientes, ejecutivos, rankings, registros de comunicaciones se cargan directamente en Hive como tablas de preparación (staging) para ser limpiadas y transformadas.

Etapa de Transformación (Reglas de Negocio y Modelado):

Esta etapa aplica la lógica de negocio y el modelado dimensional, generando las Claves Sustitutas (ID) y consolidando los atributos.

Se crean tabla de Hechos y Dimensiones (Se muestran en la tabla siguiente).

Etapa de Carga Final (Target):

Para crear las tablas de hechos y dimensiones se usa el formato de archivo ORC para mayor eficiencia en las consultas.

Las tablas finales de destino se almacenan en Hive, optimizadas para el análisis dimensional.

Tipo	Tabla Final (Hive - ORC)	Propósito de Análisis
Tabla de Hechos	FACT_UTILIDAD_TRADING	Mide el rendimiento: Utilidad por ejecutivo, canal y fecha.
Dimensión	DIM_CANAL	Análisis de performance por canal (TLV, VIRTUAL, TIENDAS).
	DIM_EJECUTIVO	Análisis de motivación y metas por ejecutivo y su jefe.
	DIM_CLIENTE	Análisis de segmentación, comportamiento digital (logeo) y fugas (flag_fuga).
	DIM_FECHA	Análisis de tendencia, metas mensuales, y el impacto del "último día útil".
	DIM_COMUNICACION	Análisis del impacto de las especificaciones y presupuestos de comunicación en la utilidad.

13. Desarrollo paso a paso de lo realizado

Paso 1:

Subir los archivos csv al HDFS

En este caso los archivos que se subió (Raw) son cambios_clientes, registro_comunicaciones, segmentos, tiendas_ranking, tlv_ranking, virtual_ranking, todos son csv.

Todo esto se guardó en una dirección llamado /user/maria_dev/sin.

Paso 2:

Crear las tablas desde los archivos csv hacia el hive.

Para ello se utilizó el HiveQL:

```
CREATE EXTERNAL TABLE IF NOT EXISTS TIENDAS_RANKING (
    canal                STRING,
    posicion             INT,
    jefe                 STRING,
    ejecutivo            STRING,
    monto                DOUBLE,
    desembolsado         INT,
    volumen_cambiado     DOUBLE,
    utilidad             DOUBLE
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
TBLPROPERTIES("skip.header.line.count"="1");

LOAD DATA INPATH '/user/maria_dev/sin/tiendas_ranking.csv'
OVERWRITE INTO TABLE TIENDAS_RANKING;
```

```

CREATE EXTERNAL TABLE IF NOT EXISTS SEGMENTOS (
    periodo                INT,
    codigo_unico            BIGINT,
    ruc                    BIGINT,
    jefe                   STRING,
    ejecutivo              STRING,
    segmento_fx            STRING,
    tipo_cuenta            STRING,
    logeo                  INT,
    num_logeos             INT,
    flg_dig                INT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
TBLPROPERTIES("skip.header.line.count"="1");

LOAD DATA INPATH '/user/maria_dev/sin/segmentos_repeated.csv'
OVERWRITE INTO TABLE SEGMENTOS;

```

Nota: Estos son algunos Scripts utilizados, los demás Scripts están en el GitHub del grupo.

The screenshot shows the Hive web interface. The top navigation bar includes 'QUERY', 'JOBS', 'TABLES', 'SAVED QUERIES', 'UDFs', and 'SETTINGS'. The 'TABLES' tab is active, showing a list of tables on the left and the details of the 'CAMBIOS_CLIENTES' table on the right. The table details view shows the following columns:

COLUMN NAME	COLUMN TYPE	COMMENT	CLUSTERED
fecha	string		
codigo_unico	int		
ruc	bigint		
nombre	string		
tipo	string		
utilidad	double		
vol_usd	double		

Acá se muestra las tablas cargadas listas para realizar su transformación.

Paso 3:

Crear las tablas de hechos y dimensiones

Para ello también se utilizó el HiveQL y además se determinó en base al esquema OLAP definido anteriormente.

Algunos Scripts utilizados:

```
--DIMENSIÓN CLIENTE
CREATE EXTERNAL TABLE IF NOT EXISTS DIM_CLIENTE (
  id_cliente      BIGINT, -- Clave Primaria (Sustituta)
  codigo_unico    BIGINT,
  ruc             BIGINT,
  nombre_cliente  STRING,
  segmento_fx     STRING,
  tipo_cuenta     STRING,
  flag_fuga       INT,

  -- Atributos de comportamiento del cliente
  logeo           INT,
  num_logeos      INT,
  flg_dig         INT
)
STORED AS ORC;
```

```
CREATE EXTERNAL TABLE IF NOT EXISTS DIM_COMUNICACION (
  id_comunicacion INT, -- Clave Primaria (Sustituta)
  tipo_cliente    STRING,
  especificacion  STRING,
  estado          STRING,
  card            STRING,
  flag_presupuesto INT
)
STORED AS ORC;
```

The screenshot shows the Hive web interface with the 'TABLES' tab selected. The 'CAMBIOS_CLIENTES' table is highlighted in the left sidebar. The main panel shows the 'COLUMNS' tab for this table, displaying a list of columns with their names, types, and comments.

COLUMN NAME	COLUMN TYPE	COMMENT	CLUSTERED
fecha	string		
codigo_unico	int		
ruc	bigint		
nombre	string		
tipo	string		
utilidad	double		
vol_usd	double		

En este caso se crea las tablas de dimensiones: dim_canal, dim_cliente, dim_ejecutivo, dim_fecha, dim_comunicación, y luego se crea la tabla de hechos fact_utilidad_trading.

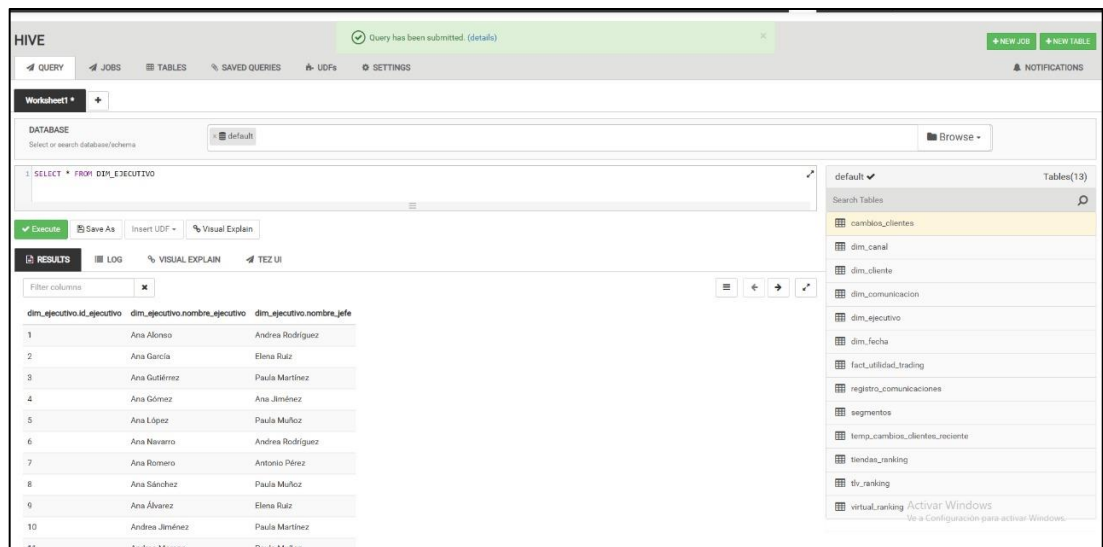
Paso 4:

Insertar datos en las tablas de dimensiones

Algunos Scripts Utilizados:

```
--DIMENSION EJECUTIVO
INSERT OVERWRITE TABLE DIM_EJECUTIVO
SELECT
    ROW_NUMBER() OVER (ORDER BY t.ejecutivo, t.jefe) AS id_ejecutivo,
    t.ejecutivo AS nombre_ejecutivo,
    t.jefe AS nombre_jefe
FROM (
    -- Combina la lista única de Ejecutivos y Jefes de todas las fuentes
    SELECT DISTINCT ejecutivo, jefe FROM TLV_RANKING
    UNION
    SELECT DISTINCT ejecutivo, jefe FROM VIRTUAL_RANKING
    UNION
    SELECT DISTINCT ejecutivo, jefe FROM TIENDAS_RANKING
) AS t;
```

```
--DIMENSION CANAL
INSERT OVERWRITE TABLE DIM_CANAL
SELECT
    ROW_NUMBER() OVER (ORDER BY canal) AS id_canal,
    canal AS nombre_canal
FROM (
    SELECT DISTINCT canal FROM TLV_RANKING
    UNION ALL
    SELECT DISTINCT canal FROM VIRTUAL_RANKING
    UNION ALL
    SELECT DISTINCT canal FROM TIENDAS_RANKING
) t;
```



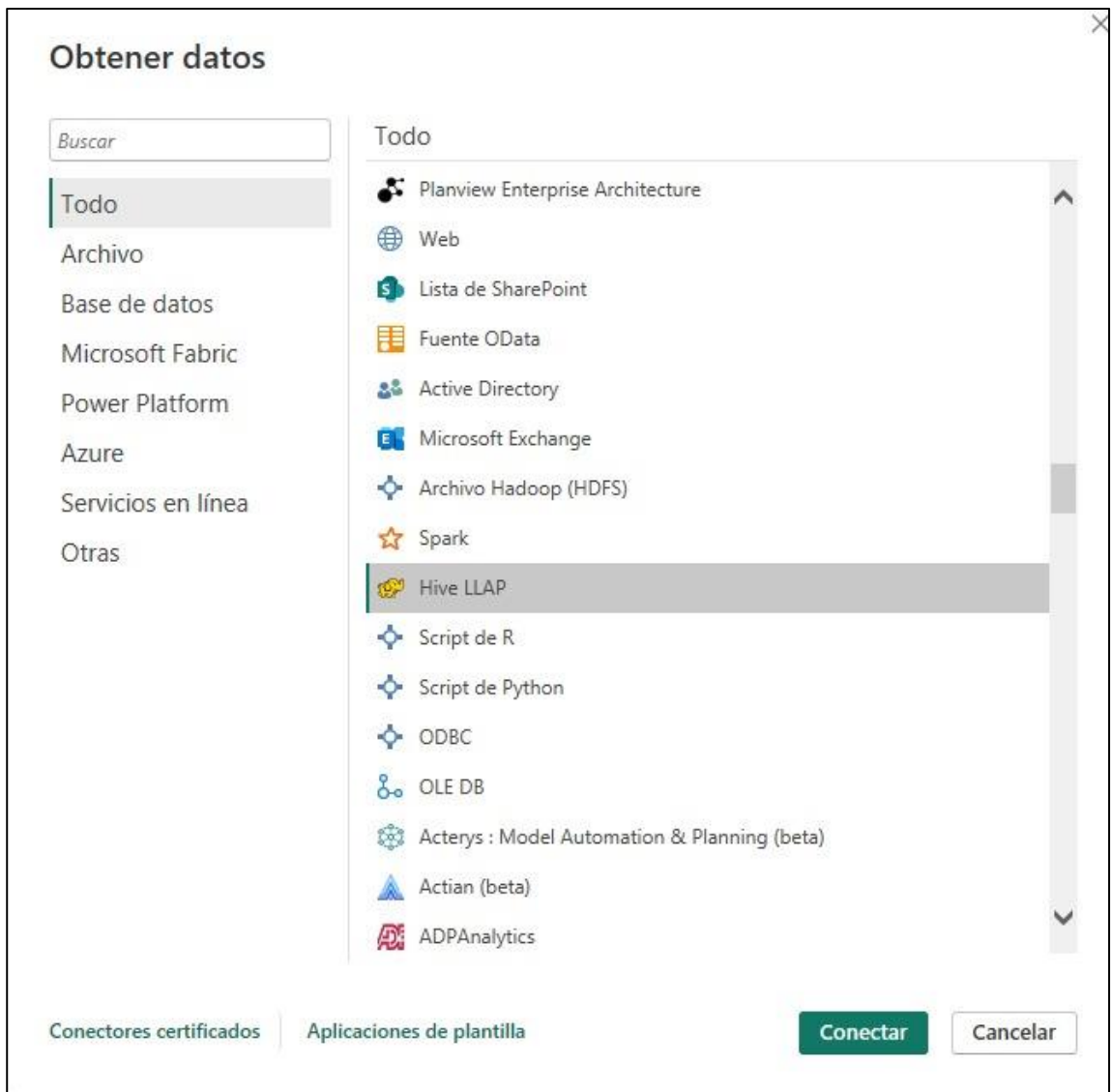
Al final probamos algunas consultas a las tablas creadas, en este caso una consulta simple: `SELECT * FROM DIM_EJECUTIVO`.

Tablas en Hive:

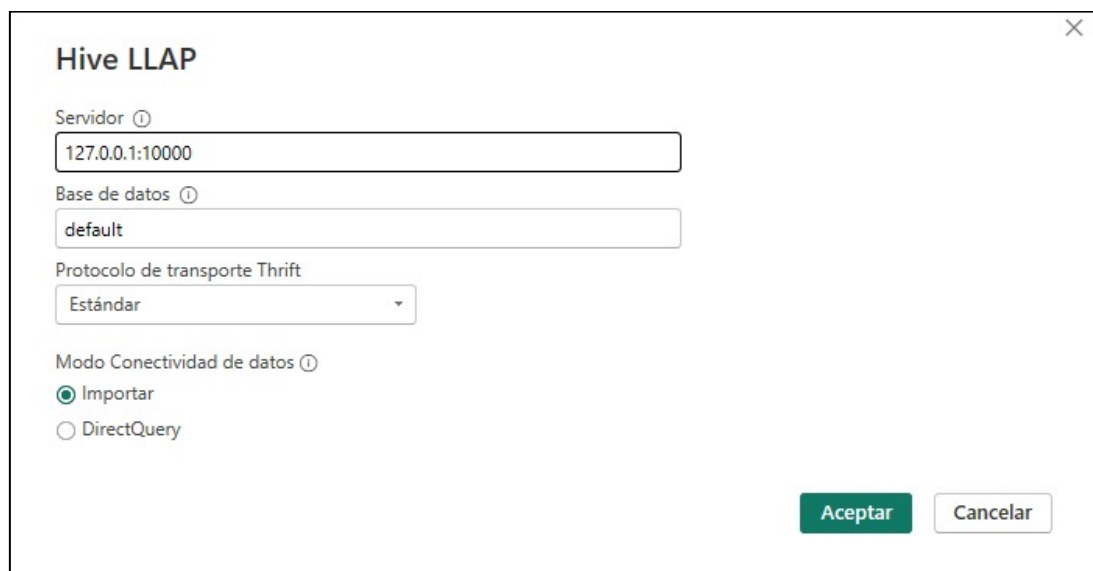
Nombre de la Tabla	Tipo de Tabla	Formato de Archivo	Partición
DIM_CANAL	Curated (Curada)	ORC	No Aplica
DIM_EJECUTIVO	Curated (Curada)	ORC	No Aplica
DIM_CLIENTE	Curated (Curada)	ORC	No Aplica
DIM_COMUNICACION	Curated (Curada)	ORC	No Aplica
DIM_FECHA	Curated (Curada)	ORC	No Aplica
FACT_UTILIDAD_TRADING	Curated (Curada)	ORC	PARTITIONED BY (periodo)

Paso 5:

Conectar PowerBI con el hive para crear el cubo OLAP y poder consumirlo para crear visualizaciones



Nos dirigimos a la pestaña de Obtener datos en PowerBI, luego buscamos el origen de los datos que sería Hive LLAP.



Para hacer la conexión del PowerBI con el Hive e importar las tablas, sería mediante LocalHost en el

puerto 10000.

Navegador

Opciones de presentación ▾

127.0.0.1:10000: default: 1 [13]

- ☐ cambios_clientes
- ☒ dim_canal
- ☒ dim_cliente
- ☒ dim_comunicacion
- ☒ dim_ejecutivo
- ☒ dim_fecha
- ☐ fact_utilidad_trading
- ☐ registro_comunicaciones
- ☐ segmentos
- ☐ temp_cambios_clientes_reciente
- ☐ tiendas_ranking
- ☐ tlh_ranking
- ☐ virtual_ranking

fact_utilidad_trading

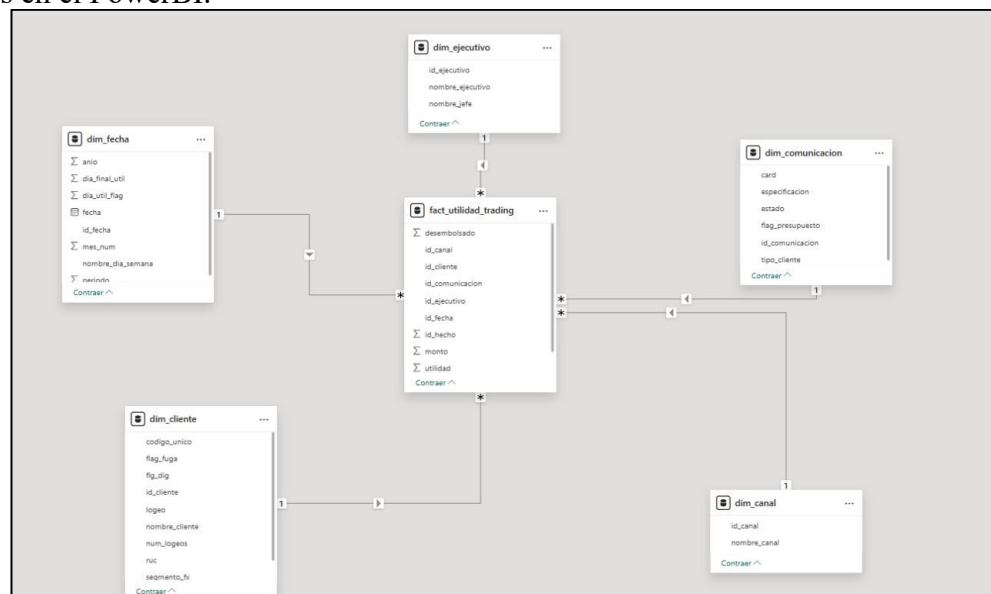
id_hecho	id_fecha	id_ejecutivo	id_cliente	id_canal	id_comunicacion
1	20251013	25	0	0	1
2	20251013	70	0	0	2
3	20251013	58	0	0	2
4	20251013	112	0	0	2
5	20251013	66	0	0	1
6	20251013	101	0	0	1
7	20251013	92	0	0	2
8	20251013	115	0	0	2
9	20251013	62	0	0	2
10	20251013	67	0	0	1
11	20251013	63	0	0	2
12	20251013	27	0	0	3
13	20251013	6	0	0	1
14	20251013	119	0	0	2
15	20251013	134	0	0	2
16	20251013	44	0	0	1
17	20251013	61	0	0	3
18	20251013	97	0	0	1
19	20251013	129	0	0	2
20	20251013	24	0	0	2
21	20251013	36	0	0	3
22	20251013	56	0	0	2
23	20251013	4	0	0	1

Cargar **Transformar datos** **Cancelar**

Y aquí se puede ver la importación de las tablas de dimensiones y hecho del HIVE.

Cubo OLAP

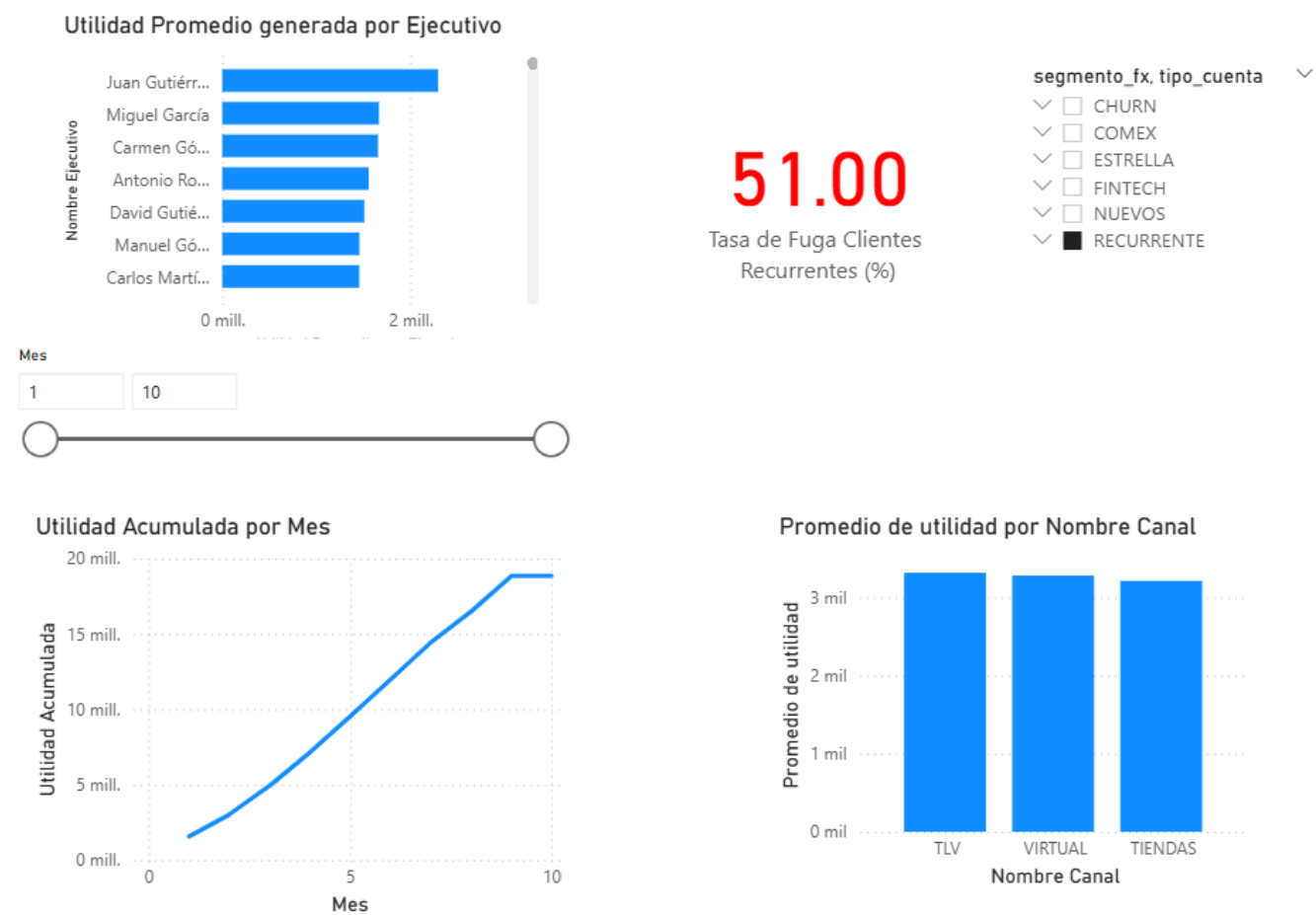
Se realiza el esquema estrella definido anteriormente y ya estaría listo para consumirse para realizar visualizaciones en el PowerBI.



Paso 6:

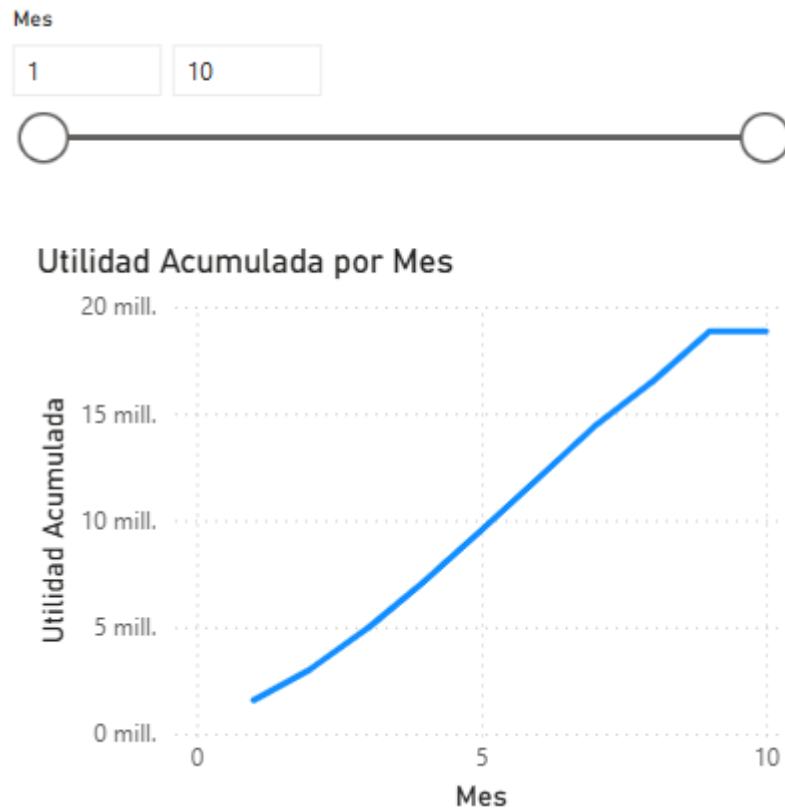
Realizar el dashboard

Para la elaboración del dashboard se utilizaron medidas creadas en DAX y diferentes visualizaciones en Power BI, integrando información proveniente de las tablas de hechos y dimensiones. Se aplicaron funciones de agregación, filtros y formato condicional para representar los indicadores clave de desempeño, permitiendo un análisis interactivo y visualmente claro del comportamiento de la utilidad y la retención de clientes.



Vistas acumulativas con filtros

La curva ascendente refleja un crecimiento sostenido de la utilidad a lo largo de los meses, alcanzando un punto máximo hacia el final del periodo analizado. Esto evidencia un comportamiento positivo en los resultados del área de trading, señalando una mejora en el desempeño financiero mes a mes.



Vistas de comparación por campos relevantes

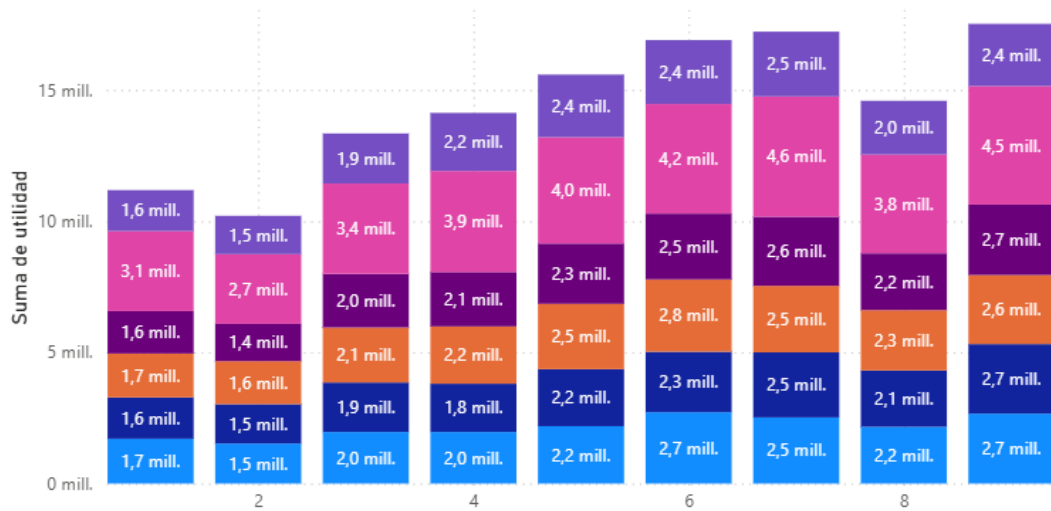
El gráfico muestra la productividad individual de los ejecutivos del área de trading. Se observa que Juan Gutiérrez lidera en generación de utilidad promedio, mientras que otros ejecutivos mantienen un desempeño más moderado. Esto permite identificar buenas prácticas y posibles brechas de rendimiento entre el personal.

Utilidad Promedio generada por Ejecutivo



Suma de utilidad por PERIODOS y segmento_fx

segmento_fx ● CHURN ● COMEX ● ESTRELLA ● FINTECH ● NUEVOS ● RECURRENTE



KPI destacado con semáforo

El indicador revela que el 51 % de los clientes recurrentes ha dejado de operar con la institución. Este valor, resaltado en rojo, indica un nivel crítico de deserción, lo que sugiere la necesidad de implementar estrategias de fidelización, comunicación proactiva y análisis de causas de fuga para este segmento clave.

51.00

Tasa de Fuga Clientes
Recurrentes (%)

segmento_fx, tipo_cuenta

- ✓ ☐ CHURN
- ✓ ☐ COMEX
- ✓ ☐ ESTRELLA
- ✓ ☐ FINTECH
- ✓ ☐ NUEVOS
- ✓ ☒ RECURRENTE

14. Limitaciones identificadas en el proceso de desarrollo de la solución

- Algunas limitaciones que identificamos fue el hecho de tener que solicitar la información y que nos la aprueben, estos temas de seguridad generaron que no analizar indicadores más a profundidad que nos habrían ayudado a obtener más insights.
- Nos aprobaron una cantidad limitada de datos, es decir nos aprobaron tener información a partir del periodo 202501, sin embargo, se solicitará mayor información.
- La versión de SQL que maneja Ambari Sandbox es algo desfasada, tuvimos que corregir varias queries que habíamos optimizado por temas de versiones

15. Propuesta de mejora o trabajo futuro en la solución elaborada.

- Implementar el cálculo de la Utilidad por Comunicación Enviada para identificar qué segmentos o canales de marketing son más eficientes.

- Medir el tiempo promedio que transcurre entre el envío de una comunicación y la realización de una transacción para optimizar el timing de las campañas.
- Crear un segmento de Alto Valor que identifique a los clientes más rentables del período anterior para enfocar los esfuerzos de retención.
- Reemplazar la estricta prohibición de rotación de segmentos por un modelo de probabilidad de cambio para mejorar el realismo de la simulación del ciclo de vida del cliente.
- Crear un análisis de predicción de clientes Churn para poder evitar su fuga

CONCLUSIONES

1. La identificación de KPIs estratégicos como el cumplimiento de utilidades, la tasa de fuga de clientes y la productividad por ejecutivo permite un seguimiento más preciso del desempeño y orienta la gestión hacia los objetivos planteados.
2. El modelo conceptual preliminar elaborado ofrece una visión integral de los factores que influyen en los resultados, facilitando el análisis predictivo y la detección de patrones relevantes en el negocio.
3. El inventario de fuentes OLTP asegura la consistencia y disponibilidad de la información, lo cual es fundamental para garantizar la confiabilidad de los análisis desarrollados en el sistema de inteligencia de negocios.
4. La adopción de Hortonworks como plataforma tecnológica brinda escalabilidad y capacidad de integración, fortaleciendo la infraestructura analítica y asegurando que la solución propuesta pueda evolucionar conforme a las necesidades de la organización.
5. En conclusión, el caso de Interbank refleja cómo incluso una institución sólida y con un portafolio diversificado puede enfrentar problemáticas específicas en unidades clave como el área de trading. La identificación de los principales dolores operativos y estratégicos —desde la demora en comunicaciones hasta la fuga de clientes o la limitada motivación de los ejecutivos— constituye un primer paso fundamental para plantear soluciones basadas en información confiable y herramientas tecnológicas. La implementación de arquitecturas de datos modernas, como Hortonworks con sus componentes de ingesta, almacenamiento y análisis, ofrece una base robusta para la toma de decisiones estratégicas, tácticas y operativas. Con un enfoque en la eficiencia, la innovación y la orientación al cliente, Interbank puede no solo superar estas dificultades, sino también transformar los retos en oportunidades para afianzar su liderazgo en el sector financiero peruano.