

UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS



**ANÁLISIS DE CUMPLIMIENTO DE METAS DE TRADING EN INTERBANK
UTILIZANDO CLOUD**

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, 17 de noviembre de 2025

JUSTIFICACIÓN DEL USO DE LA NUBE

Se eligió Google Cloud Platform (GCP) porque ofrece una base moderna, fácil de usar y alineada con las necesidades del área de Trading. BigQuery permite analizar grandes volúmenes de información con gran velocidad, mientras que Dataproc brinda flexibilidad para procesamientos más complejos, especialmente considerando que la mayoría de nuestros datos se manejan en batch. Además, GCP ofrece un modelo de costos claro y predecible. En conjunto, esta plataforma permite que el equipo de BI responda con rapidez ante la incertidumbre del mercado y genere estrategias oportunas para cumplir las metas de utilidad.

Comparativa

Característica	Google Cloud Platform (GCP) (Elegida)	Amazon Web Services (AWS)	Microsoft Azure
1. Escalabilidad (DW)	Automática y Sin Límites (BigQuery). No nos preocupamos de la infraestructura. El sistema escala solo para manejar cualquier cantidad de datos de Trading. Ventaja Técnica: Máxima agilidad. No hay tiempos muertos esperando que se ajusten los servidores.	Requiere Gestión (Redshift). Tienes que prever el tamaño del servidor. Si sube la carga de trabajo, se debe ajustar manualmente. Nota: Ahora existe también serverless	Requiere Planificación (Synapse). Necesitas definir la capacidad por adelantado.
2. Procesamiento de los CSV (ETL)	Máximo Control (Cloud Dataproc). Permite usar Apache Spark/Hive con total flexibilidad de código para limpiar y transformar los CSV. Ventaja: Ideal cuando se necesita optimizar al máximo el rendimiento de los jobs de transformación pesada (miles de registros).	Máxima Simplicidad (AWS Glue). Es un servicio Serverless que gestiona automáticamente toda la infraestructura de transformación. Ventaja: El equipo se centra solo en la lógica de transformación, sin gestionar servidores. Desventaja: Menor control para optimizaciones manuales de cluster o para usar librerías de Spark muy personalizadas..	Máxima Integración (Data Factory). ADF es el Orquestador Principal que centraliza y gestiona todo el flujo de datos. La transformación pesada se delega a servicios conectados como Azure Synapse o Databricks.
3. Inteligencia Predictiva (ML)	Integración Nativa (Vertex AI). Es la nube mejor integrada para construir modelos de Machine Learning (ML) sobre BigQuery. Ventaja	Plataforma Sólida (SageMaker). Excelente, pero requiere más pasos para mover los datos	Buen Ecosistema (ML Studio). Ideal para equipos que usan herramientas

	Técnica: Mejores estrategias. Permite crear modelos que predigan si un mes será bueno o malo para la utilidad.	entre el DW (Redshift) y la plataforma de ML.	de desarrollo de Microsoft.
4. Modelo de Cobro (DW)	Solo por lo que Consultas. BigQuery cobra por el volumen de datos que escaneas al hacer una pregunta (consulta). Ventaja Financiera: Incentiva la eficiencia. Si el equipo diseña bien las tablas, el costo de análisis es muy bajo. Se paga solo por el análisis real, no por un servidor encendido.	Pago por Servidor Encendido. Se paga por hora o minuto que el servidor de Redshift esté activo, sin importar si lo estás usando o no. Riesgo Financiero: Requiere gestión constante para apagarlo.	Pago por Uso de Recursos. Similar a AWS, basado en el tiempo de computación.
5. Previsibilidad de Costos	Muy Alta. BigQuery te dice cuánto costará una consulta antes de ejecutarla. Ventaja Financiera: Cero sorpresas. El equipo puede operar dentro de su presupuesto sin miedo a consultas accidentales y costosas.	Media. El costo está ligado a la complejidad del <i>cluster</i> y su uso, lo que hace más difícil predecir el gasto.	Media. Similar a AWS, el costo es más fácil de predecir con cargas de trabajo constantes.
6. Seguridad de Datos Sensibles	Robusta. Seguridad de nivel global. Ventaja Técnica/Financiera: La seguridad es de las más altas del mercado, esencial para proteger los datos transaccionales y de utilidad del área de Trading.	Líder en Cumplimiento. Muy fuerte en seguridad y cumplimiento normativo.	Excelente para Empresas. Muy fuerte en el sector empresarial, especialmente con Azure AD.

SELECCIÓN DE SERVICIOS CLOUD Y ARQUITECTURA

El proyecto de BI se dividirá en cuatro fases principales, donde cada servicio juega un rol esencial:

1. El proyecto de BI se dividirá en cuatro fases principales, donde cada servicio juega un rol esencial:

En esta fase, los datos de origen (los 5 archivos CSV que se tiene) son cargados y guardados en su formato original antes de ser procesados.

Servicio a usar:

Google Cloud Storage:

Almacén de Datos Fuente (Data Lake). Es el destino inicial para los cinco archivos CSV de Trading, saldos y demás. También se usa para almacenar los scripts de transformación de Dataproc.

2. Fase de Transformación y Procesamiento (ETL)

Esta es la fase crítica donde los datos brutos se limpian, validan, enriquecen y se estructuran en modelos listos para el análisis.

Servicio a usar:

Dataproc:

Motor de Transformación (ELT). Ejecuta jobs de Apache Spark para leer los CSV desde Cloud Storage, aplicar la lógica de negocio (limpiar valores nulos, estandarizar formatos de fechas, calcular la utilidad, etc.), y luego cargar el resultado en BigQuery.

3. Fase de Orquestación y Data Warehousing

En esta fase, se define cuándo y cómo se ejecutan los procesos (Orquestación) y se almacena el resultado final para la consulta de BI (Data Warehousing).

Servicios a usar:

Composer:

Orquestador del Flujo de Trabajo (Scheduler). Define la secuencia y frecuencia de las tareas: Primero, carga los CSV a Storage. Luego, ejecuta el job de Dataproc. Finalmente, verifica la carga en BigQuery.

BigQuery:

Almacén de Datos (Data Warehouse). Es el destino final de los datos limpios y transformados por Dataproc. Aquí se crean las tablas finales (hechos y dimensiones) que serán usadas por el área de Trading.

4. Fase de Visualización y Consumo de Datos (BI)

La fase final, donde los usuarios de negocio acceden a los datos transformados para la toma de decisiones.

Servicio a usar:

Looker Studio:

Herramienta de Visualización. Se conecta directamente a las tablas de BigQuery para crear los dashboards y reportes de BI que mostrarán la utilidad, la performance de las comunicaciones, y los factores influyentes.

ARQUITECTURA:

