


# MIC-GDL-DATA-TRINO

2026

Emmanuel Perez Carrillo

 <p><b>tigo</b> <b>MILlicom</b> THE DIGITAL LIFESTYLE</p>	Millicom International Cellular S.A.	<b>Code:</b>	MIC-GDL-DATA
		<b>Version:</b>	1.0.0
		<b>Effective Date:</b>	14-FEB-2026
		<b>Next Review Date:</b>	14-FEB-2007
		<b>Classification:</b>	Public
		<b>Document Type:</b>	Guidelines
		<b>Document Owner:</b>	Emmanuel Perez Carrillo

<b>1</b>	<b>TABLA DE CONTENIDOS</b>	
2	PROPÓSITO .....	3
3	AUDIENCIA Y ALCANCE .....	3
4	DEFINICIONES .....	3
5	GUÍAS.....	4
5.1	Descripción general de Trino.....	4
5.2	Arquitectura a alto nivel.....	4
5.3	Casos de uso .....	5
5.4	Referencias .....	5
6	APÉNDICES.....	6
6.1	Revisión del Documento.....	6
6.2	Histórico de Revisiones .....	6

## 2 PROPÓSITO

El propósito de esta guía es brindar un panorama general del uso de este componente dentro de la Plataforma de Servicio de Datos o Datalake.

## 3 AUDIENCIA Y ALCANCE

Estas guías aplican para Millicom y sus subsidiarias (colectivamente “La compañía”), incluyendo a Socios, Terceras partes y Proveedores que realicen negocios en nombre de la Compañía. Las guías pueden actualizarse de manera periódica para reflejar cambios en las versiones de los productos o la manera de trabajo interna, la adopción de nuevas tecnologías, o de cualquier manera, para reflejar los requerimientos de la empresa.

Será muy útil para roles de datos en especial para Ingenieros DevOps, Ingenieros QA y equipos Soporte L2.

## 4 DEFINICIONES

Definiciones de conceptos de datos que facilitan el entendimiento del documento.

Término	Definición
CNCF	Acrónimo del inglés – Cloud Native Computing Foundation
Data Lake	Repositorio de datos de todo tipo de formatos para Big Data
DatalakeHouse	Repositorio de datos modelado para manejo de grandes volúmenes de información.
Modelo	Conjuntos de datos y catálogos que permiten analizar y describir un tema de interés desde varias perspectivas de negocio.
Capa Bronce	También conocida como Capa Raw, fría, Landing, Staging u ODS. Se distingue por ser una copia lo más cercano a la fuente de datos, donde tenemos los archivos en formato original, es decir, sin limpieza de datos ni reglas aplicadas, tal cual se recibieron de la fuente.
Capa Plata	También conocida como Capa Estándar, Intermedia, filtrada, aumentada o enriquecida. Se distingue por tener datos ya procesados combinados para facilitar su análisis.
Capa Oro	También conocida como Capa de analítica, de Consumo, Caliente o master. Se distingue por tener datos ya procesados que se unen fácilmente con otros catálogos para facilitar su análisis a través de identificadores estandarizados.

En nuestra plataforma se suele describir con el siguiente modelo conceptual:



## 5 GUÍAS

### 5.1 Descripción general de Trino

Trino es un motor de consultas SQL distribuido y de alto rendimiento para analítica de grandes volúmenes de datos. No es una base de datos en sí misma: no almacena datos, sino que se conecta a múltiples fuentes de datos (data lakes, bases relacionales, NoSQL, etc.) y permite consultarlas mediante ANSI SQL estándar.

Actuar como capa unificada de acceso y consulta de datos para la organización.

Permitir que analistas, científicos de datos y aplicaciones internas puedan:

- Ejecutar **consultas SQL interactivas** sobre grandes volúmenes de datos.
- Unir datos dispersos (data warehouse, data lake, bases OLTP, sistemas de logging).
- Estandarizar el acceso a datos mediante un único endpoint SQL.

Características clave:

- **Velocidad:** motor altamente paralelo y distribuido, orientado a baja latencia para analítica.
- **Escalabilidad:** utilizado para consultar datos a escala de petabytes/exabytes.
- **Simplicidad:** compatible con herramientas BI como Tableau, Power BI, Superset, R, entre otras.
- **Versatilidad:** sirve tanto para consultas interactivas de segundos como para procesos de tipo batch de varias horas.
- **Query federation:** permite hacer consultas federadas, uniendo datos de múltiples sistemas en una sola consulta SQL.
- **In-place analysis:** consulta datos directamente en S3, Hadoop, Cassandra, MySQL, PostgreSQL, etc., sin necesidad de mover o duplicar datos.
- **Open source:** proyecto gestionado por la Trino Software Foundation.

### 5.2 Arquitectura a alto nivel

A grandes rasgos, Trino se compone de:

- **Coordinador (Coordinator):**
  - Recibe las consultas SQL.
  - Las analiza, planifica y distribuye el trabajo entre los nodos worker.
  - Expone la interfaz HTTP para clientes (CLI, JDBC, herramientas BI).
- **Workers (nodos de trabajo):**

- Ejecutan las tareas de lectura, filtrado, agregación y join sobre las fuentes de datos.
- Se escalan horizontalmente según la carga y el volumen de datos.
- **Conectores (Connectors):**
  - Componentes plugin que permiten a Trino conectarse a diversas fuentes: Hive, Iceberg, Delta Lake, MySQL, PostgreSQL, Cassandra, Kafka, BigQuery, Snowflake, etc.
  - Cada conector se configura como un **catálogo** dentro de Trino.

### 5.3 Casos de uso

- Analítica interactiva: dashboards, exploración ad-hoc, A/B testing.
- Data Lake Analytics: análisis de datos en HDFS, S3 u otros object storage con formato Parquet/ORC/Avro.
- ETL/ELT batch: ejecutar procesos de transformación de datos usando SQL en vez de jobs específicos de cada sistema.
- Federación de datos: realizar consultas como:

```
SELECT o.order_id,
       o.order_date,
       c.customer_name
FROM   lakehouse.orders o    -- datos en data lake (ej. Iceberg/Hive)
JOIN   crm_mysql.customers c -- datos en MySQL
      ON o.customer_id = c.id;
```

- Capa de acceso centralizada: aplicaciones internas acceden a Trino por JDBC/HTTP, aislando los detalles de cada fuente.

### 5.4 Referencias

#### 1. Canal oficial de Trino en YouTube

- [Trino – YouTube](#)  
 Contiene:
  - Trino Community Broadcast.
  - Charlas técnicas y casos de uso.
  - Introducciones generales y sesiones de entrenamiento.

#### 2. Documentación oficial: <https://trino.io/docs/current/index.html>

3. Github oficial: <https://github.com/trinodb/trino>

## 6 APÉNDICES

### 6.1 Revisión del Documento

Revisado por	Rol	Fecha
Emmanuel Perez Carrillo	Millicom - Chief Data Officer	01-FEB-2026

### 6.2 Histórico de Revisiones

Versión	Descripción del cambio	Autor	Fecha
1.0.0	Versión inicial del documento	David Vasquez Fernandez	15-JAN-2026
1.0.0	Revisado por par – Luz Nino	Luz Nino	09-FEB-2026
1.0.0	Revisado por dueño – Emmanuel Perez	Emmanuel Perez	09-FEB-2026