

# DATAPLEX

Carlos Alejandro Baltazar Padilla

6B

28/02/2022

# ¿QUÉ ES DATAPLEX?

- Es una estructura de datos inteligente que permite a las organizaciones administrar, monitorear y gobernar de manera centralizada sus datos a lo largo de lagos, almacenes y bases de datos, en una sola experiencia integrada. La calidad de la data automatizada permite a los analistas y científicos de datos abordar la coherencia de los datos en las herramientas de su elección, para unificar y administrarlos sin moverlos o duplicarlos.





## ¿POR QUÉ USAR DATAPLEX?

- Las empresas tienen datos distribuidos en lagos de datos, almacenes de datos y mercados de datos. Dataplex le permite unificar estos datos sin ningún movimiento de datos, organizarlos según las necesidades de su negocio y administrar, monitorear y controlar estos datos de manera centralizada. Dataplex permite la estandarización y unificación de metadatos, políticas de seguridad, gobernanza, clasificación y gestión del ciclo de vida de los datos en todos estos datos distribuidos.

## ¿CÓMO FUNCIONA DATAPLEX?

- Dataplex administra los datos de una manera que no requiere movimiento o duplicación de datos. A medida que identifica nuevas fuentes de datos, Dataplex recolecta los metadatos para datos estructurados y no estructurados, utilizando verificaciones de calidad de datos integradas para mejorar la integridad. Dataplex registra automáticamente todos los metadatos en un almacén de metadatos unificado. También puede acceder a datos y metadatos a través de una variedad de servicios de Google Cloud, como BigQuery, Dataproc Metastore, Data Catalog y herramientas de código abierto, como Apache Spark y Presto.

## DATAPLEX ABSTRAE LOS SISTEMAS MEDIANTE EL USO DE LAS SIGUIENTES CONSTRUCCIONES:

Lago: una construcción lógica que representa un dominio de datos o una unidad de negocio. Por ejemplo, para organizar los datos según el uso del grupo, puede configurar un lago por departamento (por ejemplo, Minorista, Ventas, Finanzas).

Zona: un subdominio dentro de un lago, útil para categorizar los datos por etapa (por ejemplo, aterrizaje, sin procesar, `curated_data_analytics`, `curated_data_science`), uso (por ejemplo, contrato de datos) o restricciones (por ejemplo, controles de seguridad, niveles de acceso del usuario). Las zonas son de dos tipos, crudas y curadas.

Zona sin procesar: datos que están en su formato sin procesar y no sujetos a una verificación de tipos estricta.

Zona seleccionada: datos limpios, formateados y listos para análisis. Los datos son en columnas, divididos en Hive, en archivos Parquet, Avro, Orc o tablas de BigQuery. Los datos se someten a una verificación de tipo, por ejemplo, para prohibir el uso de archivos CSV porque no funcionan tan bien para el acceso SQL.

Activo: un activo se asigna a los datos almacenados en Cloud Storage o BigQuery. Puede asignar datos almacenados en proyectos de Google Cloud separados como activos en una sola zona.

Entidad: una entidad representa metadatos para datos estructurados y semiestructurados (tabla) y datos no estructurados (conjunto de archivos)

# CARACTERÍSTICAS GENERALES DE DATAPLEX



Administración unificada de datos impulsada por metadatos a través de un panel único



Seguridad y administración centralizadas que permiten la propiedad distribuida con control global



Inteligencia de datos integrada para unificar datos distribuidos sin movimiento de datos



Una plataforma abierta compatible con herramientas de código abierto y un ecosistema de socios sólido

# VENTAJAS

## Libertad de elección

Obtén la libertad de almacenar datos donde quieras obtener el mejor precio y con el mejor rendimiento, a la vez que eliges las mejores herramientas estadísticas, código abierto o nativo de la nube para acelerar todo el ciclo de vida de estadísticas.

## Administración unificada

Habilita la estandarización y la unificación de los metadatos, las políticas de seguridad, la administración y la clasificación de datos para la coherencia entre los datos distribuidos.

## Automatización inteligente

Inteligencia de datos integrada, con las mejores funciones de IA y AA de Google que automatizan el descubrimiento y la recopilación de datos, la administración del ciclo de vida de los datos y la calidad de los datos para reducir los costos de administración.

# CARACTERÍSTICAS CLAVE



## Seguridad y administración centralizadas

- Administración, supervisión y auditoría centrales de políticas para la autorización y clasificación de datos. Distribución de la propiedad de datos con supervisión y control globales entre los datos y artefactos relacionados, incluidos los modelos de aprendizaje automático.



## Administración de metadatos

- Automatiza el descubrimiento, la clasificación, la detección de esquemas y la recopilación de metadatos, y el registro de datos estructurados, semiestructurados o no estructurados en sistemas aislados de datos con inteligencia de datos integrada. Accede a estos datos con facilidad desde una variedad de herramientas de estadísticas y ciencia de datos.



## Calidad de los datos

- Automatiza la calidad de los datos en los datos distribuidos y habilita un acceso listo para usar a datos confiables.



## Administración del ciclo de vida de los datos

- Organiza tus datos de manera lógica que abarcan varios servicios de almacenamiento en dominios específicos de la empresa mediante data lakes y zonas de datos de Dataplex. Administra, selecciona, clasifica y archiva tus datos con facilidad con un solo clic.



## Creado para datos distribuidos

- Unifica los datos sin moverlos ni duplicarlos. Deja tus datos en los que se minimizan los costos y maximizan el rendimiento.



MEJORES  
PRACTICAS EN  
DATAPLEX

Elige un proyecto para tu lago  
(host)

Configuración y limitaciones  
del almacenamiento en la nube

Guía de seguridad y permisos

# CLIENTES

- Algunos clientes que usan DATAPLEX



## BotMaker

Botmaker ofrece una atención excepcional al cliente por medio de bots habilitados por voz y texto que responden dudas en tiempo real a través de inteligencia artificial y procesos automatizados.

[Leer el caso de éxito](#)

## PROPERATI

### Properati

Properati reúne la disponibilidad de propiedades inmobiliarias añadiendo datos con valor agregado para facilitar la búsqueda en línea de acuerdo al interés de las personas interesadas.

[Leer el caso de éxito](#)



## Minka

Minka es una plataforma de banca abierta construida en la nube con tecnología blockchain que permite a las fintechs e instituciones financieras tradicionales crear nuevos productos financieros en días en vez de años.

[Leer el caso de éxito](#)



## Compara Online

Compara Online es un comparador en línea de servicios financieros que despliega la mayor cantidad de productos de distintas aseguradoras para que el usuario elija de forma gratuita la mejor opción.

[Leer el caso de éxito](#)



## Enel Codensa

Codensa se enfrentó con un hallazgo arqueológico que le representó un desafío en términos de difundir información de interés público por una vía digital en un plazo de tiempo menor a un mes.

[Leer el caso de éxito](#)



## Magma Partners

Magma Partners implementó el uso de herramientas de Google obteniendo ventajas como la facilidad del trabajo colaborativo entre distintos equipos distribuidos en diferentes zonas geográficas, reducción de tiempos y centralización de información.

[Leer el caso de éxito](#)

# MÁS CLIENTES

 magazine luiza

## Magazine Luiza

Apigee, Firebase y GCP han ayudado a Magazine Luiza a convertirse en una de las empresas de comercio electrónico más exitosas de Brasil.

[Leer el caso de éxito](#)

 mercado libre

## Mercado Libre

Mercado Libre buscaba mejorar el trabajo colaborativo y la integración de los equipos distribuidos en todos los países donde tiene operaciones, así que las herramientas de Google Workspace contribuyeron a reconfigurar la dinámica de trabajo entre los equipos, creando una nueva cultura atractiva, sobre todo, para colaboradores millennials.

[Leer el caso de éxito](#)

 almundo

## Almundo

La agencia de viajes omnicanal, junto con las aplicaciones de Google Workspace, mejoró los procesos de trabajo para lograr un entorno más colaborativo y aumentó la productividad de sus equipos de trabajo para continuar su liderazgo en América Latina.

[Leer el caso de éxito](#)

 sbt

## SBT

En asociación con Ilegre, la empresa Sistema Brasileiro de Televisão (SBT) desarrolló un nuevo sitio web para incorporar el concepto de "mobile first", dinamizar el consumo del contenido y volverlo más atractivo para el público.

[Leer el caso de éxito](#)

 Farmacias del Ahorro  
Te queremos... bien.

## Farmacias del Ahorro

Farmacias del Ahorro es una cadena mexicana de venta de medicamentos y productos de cuidado personal que buscaba renovar sus herramientas de comunicación y colaboración para generar procesos más eficientes y productivos.

[Leer el caso de éxito](#)

 Grupo Amanus

## Grupo Amanus

Grupo Amanus mitigó las fallas en su sistema de correo electrónico y experimentó una nueva era de trabajo colaborativo entre sus equipos de trabajo gracias a Google Workspace.

[Leer el caso de éxito](#)

<https://cloud.google.com/customers>

# PRECIOS

- Los precios de Dataplex se basan en el uso de prepago, incluido lo siguiente:
  1. Procesamiento de Dataplex (medido por unidad de procesamiento de datos, DCU), que cubre las funciones de descubrimiento y exploración de datos en Dataplex. 1 DCU se compara con 1 CPU virtual y 4 GB de RAM.
  2. Dataplex Shuffle Storage (GB por mes), que cubre las necesidades de almacenamiento en disco para la exploración de datos en Dataplex.

# NIVEL GRATUITO DE DATAPLEX

- 100 DCU son 100 CPU virtuales y 400 GB de RAM. Aprox. (mensual)

Recurso	Límites mensuales de uso gratuito
Procesamiento de Dataplex	100 DCU-hora
Almacenamiento de Dataplex Shuffle	125 GB de almacenamiento de Dataplex Shuffle por DCU (o 17 GB por mes de almacenamiento de Dataplex Shuffle en total)

## ASOCIACIÓN CON LÍDERES DE LA INDUSTRIA (SOCIOS)


- Proveedores de estadísticas de datos líderes del sector para que Dataplex pueda integrarse con rapidez en tus inversiones existentes en estadísticas de datos.
- Aquí se pueden buscar los socios por filtros como región, especialización, iniciativas y productos que pueden administrar.

 accenture

 CONFLUENT

 Collibra

 HCL

 Informatica

 NVIDIA

 Starburst

 TRIFACTA

# FUENTES

- <https://cloud.google.com/dataplex#section-7>