

Manage and govern distributed data with Google Cloud Dataplex

Data Engineering Spotlight
July 2022

Irina Farooq
Director of Product Management

Prajakta Damle
Group Product Manager



Agenda

01

Distributed Data
Challenges

02

Dataplex

03

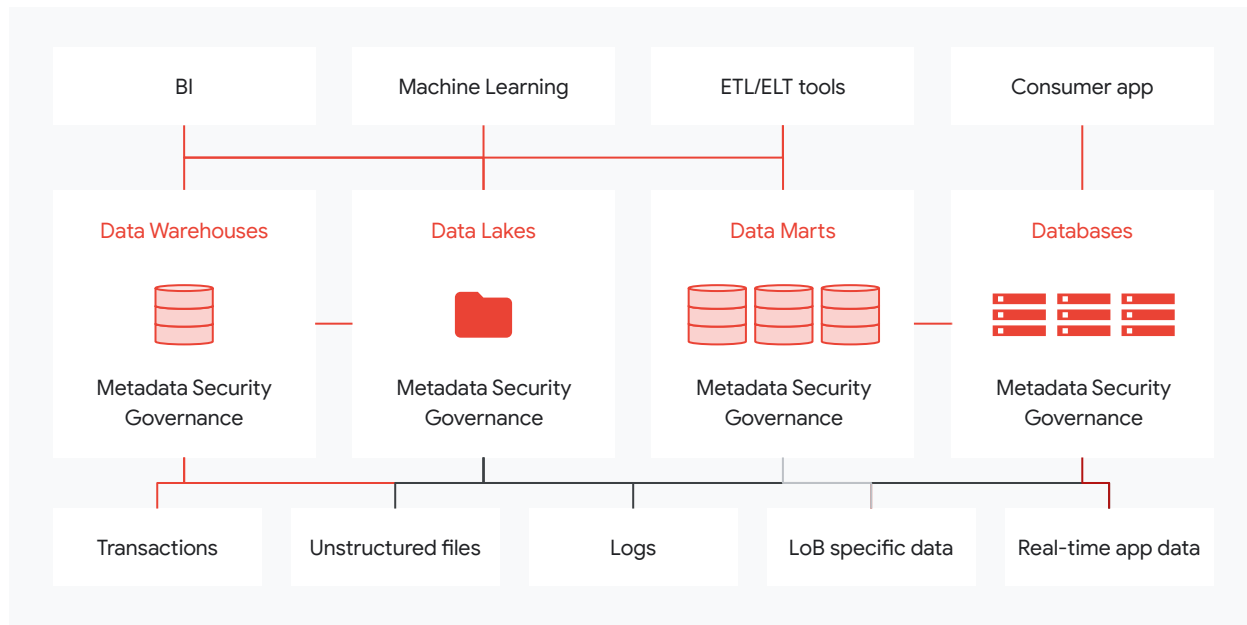
Introducing
enhancements

04

Demo

Enterprise data is distributed

Most organisations have grown organically and so has complexity.
Data landscape also exists outside of IT span of control.



Data silos



Security



Connecting tools



Standardization



Financial governance

Dataplex

Intelligent data fabric that unifies your distributed data to help automate data management and power analytics at scale.

Unified metadata across distributed data

Automatic data discovery, and metadata harvesting, enriched with business context. Logically unify and organize your data without any data movement.

Centralized security & governance

Central policy management, monitoring and auditing for data authorization, retention, and classification.

Intelligent data management

Built-in AI-driven intelligence with data classification, data quality, data lineage, and lifecycle management.



Unified data management and governance

metadata | quality | lifecycle | classification | security

Data Warehouses



Data Lakes



Data Marts





Dataplex enables us to deliver a **business domain specific, self-service data platform** across distributed data, with de-centralized data ownership but centralized governance and visibility. We are very excited to adopt Dataplex as a **central component for building a unified data mesh** across our analytics data.

Saral Jain,
Director of Engineering



We are excited to use Dataplex to enable centralized governance for our distributed data. Dataplex formalizes our data mesh vision and gives us the right set of controls for cross-domain data organization, data security, and data quality.”

Balaji Maragalla,
Director Big Data Platform



Having a first party, cloud-native, **product to architect a Data Mesh in GCP** is crucial for effective **data sharing and data quality** amongst teams. Dataplex streamlines productivity, allowing teams to build data domains and orchestrate data curation across the enterprise.”

Alexandre Cote,
Product Leader



Dataplex’s vision and capabilities align well with our current data strategy to build a **unified data fabric for all our analytics and AI/ML use cases.”**

Ashish Shekhar, Head of Technology
Enterprise Analytics & Applied AI



Dataplex enables us to effectively manage and **standardize data quality, data security, and data privacy for data across these domains.** We are looking forward to building trust in our data with Google Cloud’s Dataplex”

Iwao-san,
Director of Data Analytics



Sources:

[Introducing google cloud dataplex](https://cloud.google.com/blog/products/data-analytics/build-a-data-mesh-on-google-cloud-with-dataplex-now-generally-available)
<https://cloud.google.com/blog/products/data-analytics/build-a-data-mesh-on-google-cloud-with-dataplex-now-generally-available>

Announced! the unification between Dataplex and Data Catalog



Dataplex

+



Data Catalog

=

Analytics

BigQuery

Dataproc

Dataflow

AI platform

Data studio

Dataplex

Data catalog (discover, organize, enrich)

Unified security and governance (permissions management)

Data intelligence (data quality, data lineage, classification, lifecycle mgmt.)

Logical data organization

Storage

Structured

Semi-structured

Unstructured

Streaming data*

GCP

Multi-cloud*

On-premises*

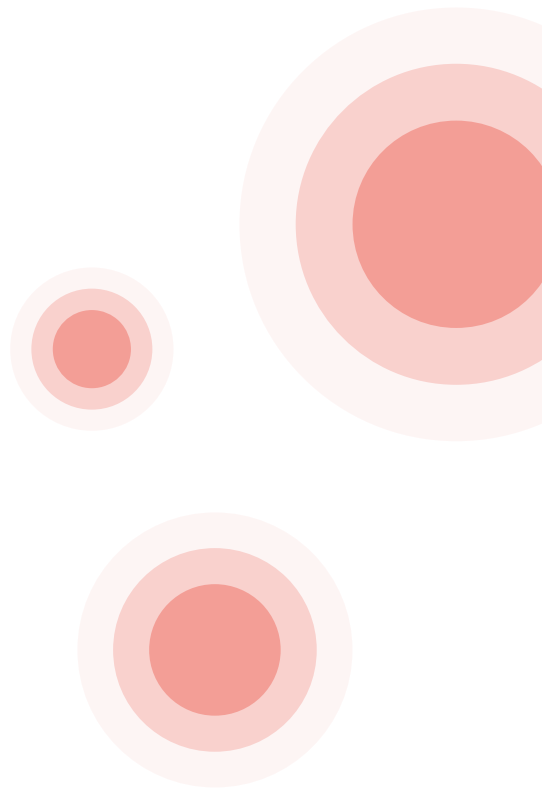
**future capabilities*



With Data Catalog now being part of Dataplex, we get a unified, simplified, and streamlined experience to effectively discover and govern our data, which enables team productivity and analytics agility for our organization. **We can now use a single experience to search and discover data with relevant business context**, organize and govern this data based on business domains, and enable access to trusted data for analytics and data science - all within the same platform."

Elton Martins

Senior Director of Data Engineering at Loblaw Companies Limited



What Dataplex does for you

Self-select what is most relevant to you



Organize

Organize data
into data domains



Discover

Auto-discover
data & metadata



Enrich

Business context



Trust

With lineage and quality



Curation

Per org policies



Secure

Active metadata driven
security, privacy



Monitor

Monitor & audit changes

Discover data with business context

Automate cataloging of metadata

Automatically harvest up-to-date metadata for structured and unstructured data (docs, pdfs, imgs) from GCS, BQ and Pub/Sub

Enrich with business context

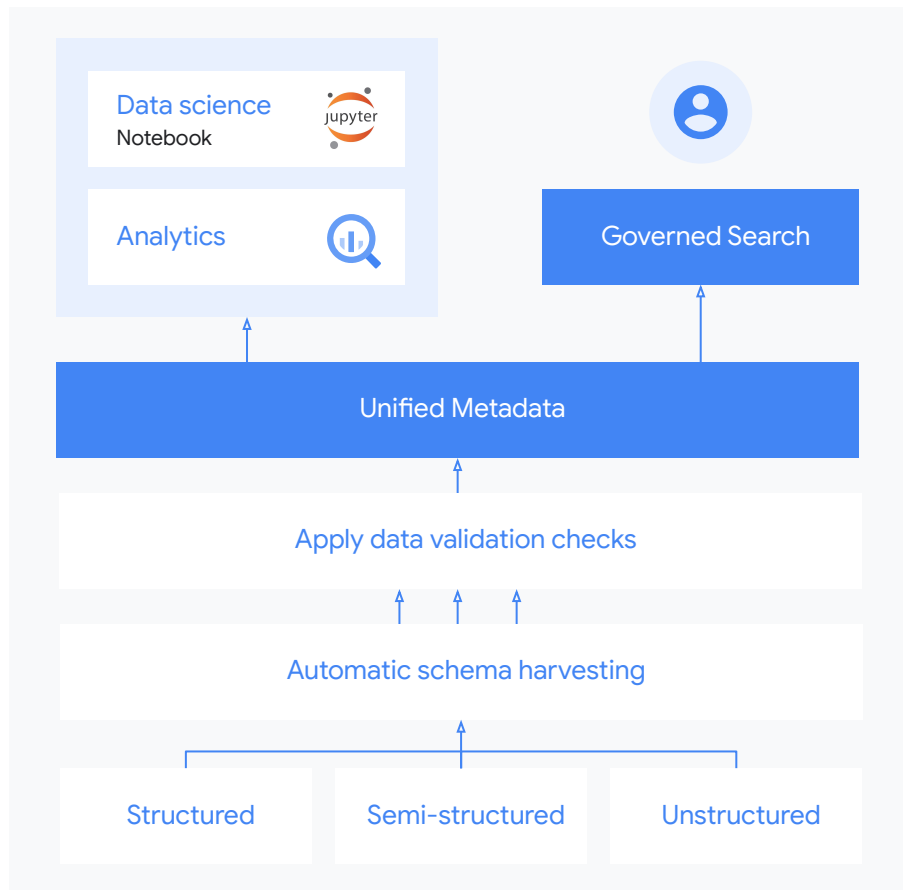
Augment technical metadata with business context (tags, data stewards, etc) to enable self-service workflows

Enable global governed search

Allow users to quickly identify the data they need with powerful faceted search that complies with ACLs

Power data exploration

Provide a one-click access to notebooks and SQL workbench within Dataplex to quickly explore identified data



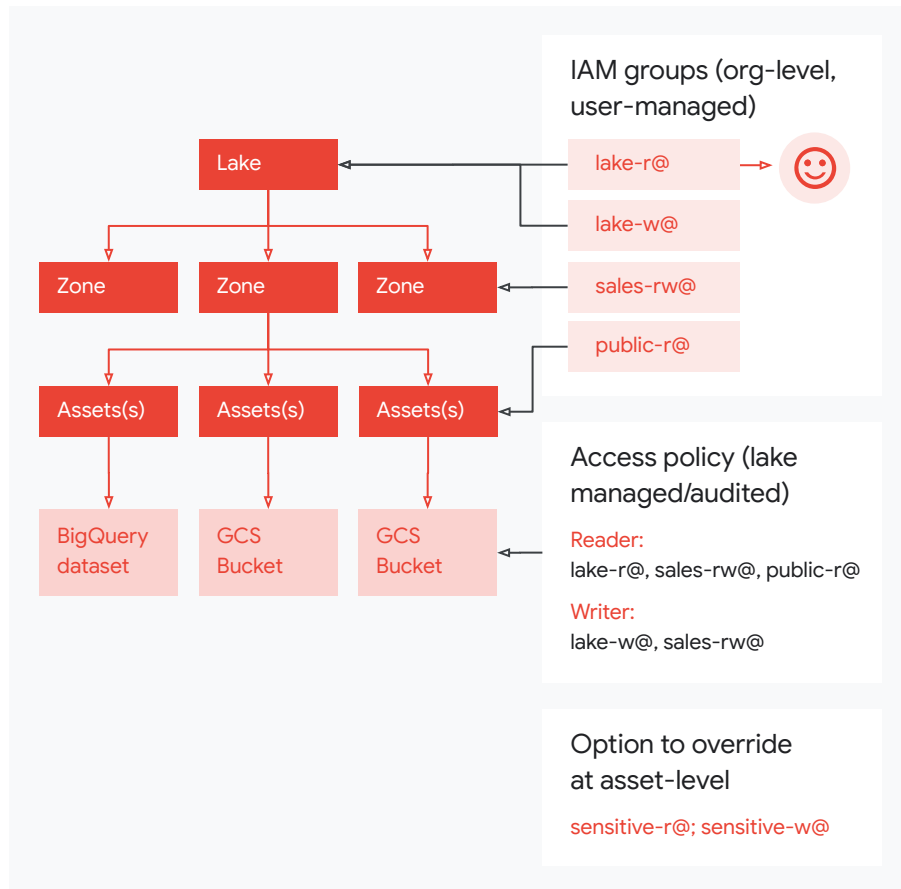
Centrally secure data access

Distributed data ownership with global visibility

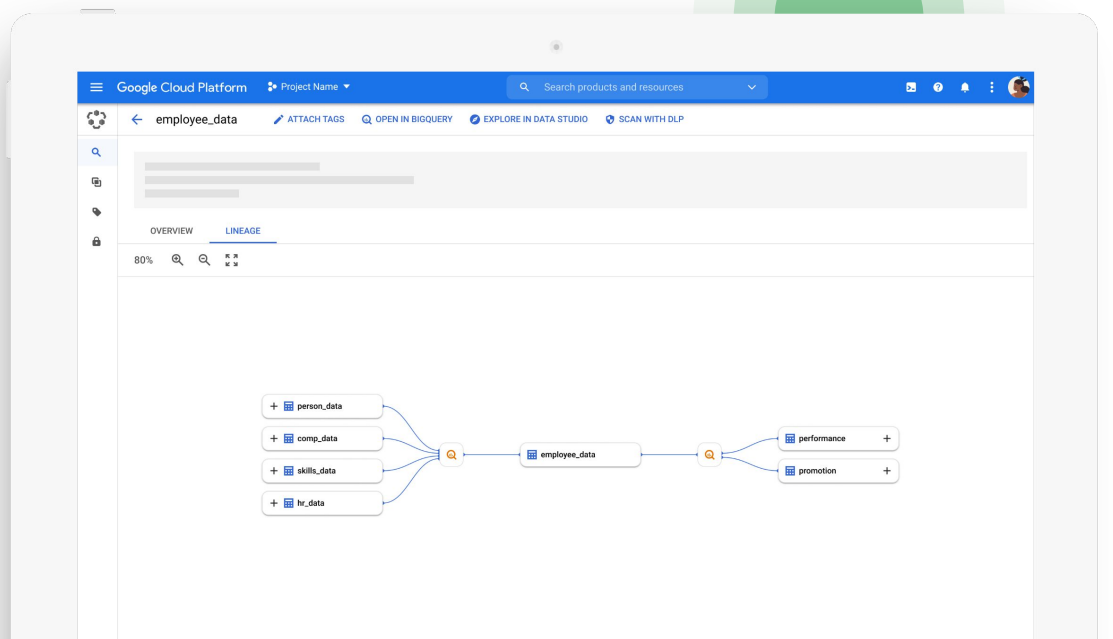
- Use existing IAM groups
- Local owners can grant permissions at the Lake, Zone, or Asset-levels
- Permissions inherited by default with an option to override
- Data stewards can globally monitor and audit

Consistent and unified policy management

- Centrally manage data access permissions
- Apply consistent permissions based on logical domains to data stored across projects and products
- Dataplex automatically applies policies to underlying storage in GCS and BigQuery.



Enhance trust in data with data lineage and data quality



Understand & trust

Ability to understand where your data comes from and the transformations it goes through to trust the insights generated.

Impact & root cause analysis

Trace downstream impact and upstream causes as soon as data issues are captured with data quality checks

Compliance

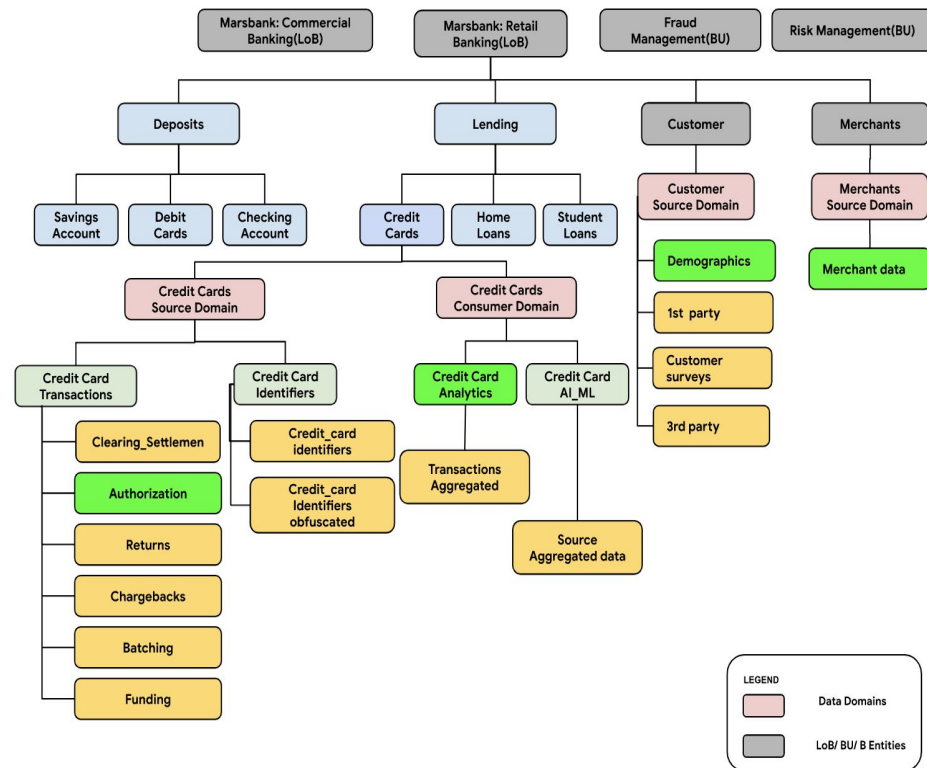
Understand how various data types flow through your data estate to ensure compliance to regulations

Demo

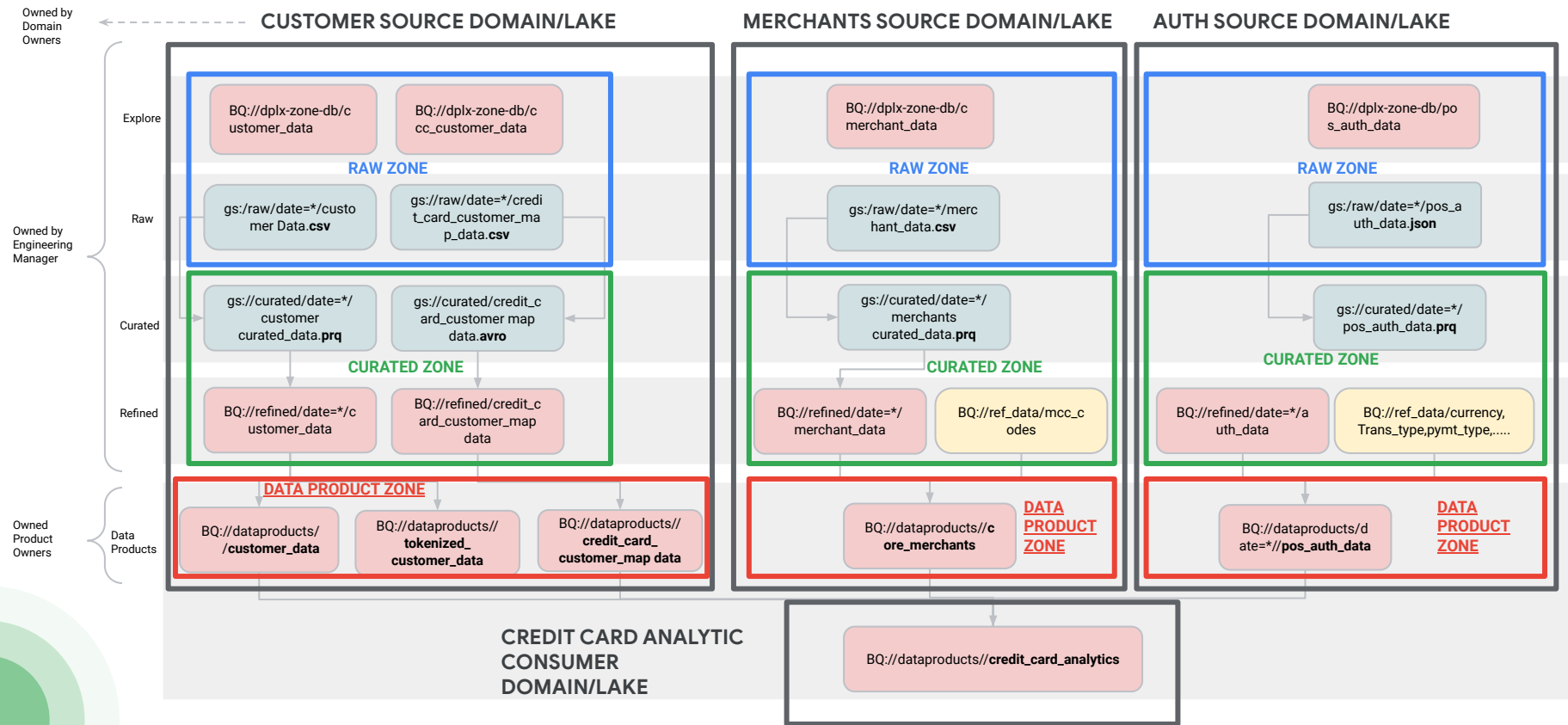
The scenario

Let's take a look at the data landscape of “**Mars Bank**”, which provides financial services to individual consumers and businesses.

The Bank currently stores they data in Google Cloud Storage and BigQuery



Domain based architecture



Additional Resources

Google Cloud White Paper

<https://services.google.com/fh/files/misc/build-a-modern-distributed-datamesh-with-google-cloud-whitepaper.pdf>

DB Insights White paper

<https://cloud.google.com/resources/db-insight-dataplex-analyst-report>

Dataplex Product Detail Page

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