

Data Catalog is now a part of Dataplex.

# Dataplex

Break free from data silos with Dataplex's intelligent data fabric that enables organizations to centrally discover, manage, monitor, and govern their data across data lakes, data warehouses, and data marts with consistent controls, providing access to trusted data and powering analytics at scale.

- Single pane of glass for data management across data silos
- Centralized security and governance enabling distributed ownership of data with global control
- Unified search and data discovery, based on business context, across distributed data
- Built-in data intelligence to enable trust in data and accelerate time to insights
- An open platform with support for open source tools and a robust partner ecosystem

## BENEFITS

### **Freedom of choice**

Get the freedom to store data where you want for the best price and performance while choosing the best analytics tools, open source or cloud native, to accelerate the entire analytics lifecycle.

### **Intelligent automation**

Built-in data intelligence using Google's best-in-class AI/ML capabilities that automate data discovery, metadata harvesting, data lifecycle management, data quality and lineage to reduce management costs.

### **Unified governance**

Enable standardization and unification of metadata, security policies, governance, and data classification for consistency across distributed data.

# Key features

## **Centralized security and governance**

Enable central policy management, monitoring, and auditing for data authorization and classification, across data silos. Facilitate distributed data ownership based on business domains with global monitoring and governance.

## **Unified metadata management with Data Catalog**

Automate data discovery, classification, and metadata enrichment of structured, semi-structured, and unstructured data, stored in Google Cloud and beyond, with built-in data intelligence. Manage technical, operational, and business metadata, for all your data, in a unified, flexible, and powerful Data Catalog.

## **Simplified data discovery and exploration**

Easily search, find, and understand your data with built-in faceted-search interface using the same search technology as Gmail. Explore and analyze this data using a variety of analytics and data science tools.

## **Data quality and lineage**

Automate data quality across distributed data and enable access to data you can trust. Use automatically captured data lineage to better understand your data, trace dependencies, and effectively troubleshoot data issues.

## **Data organization and lifecycle management**

Logically organize your data that spans multiple storage services into business-specific domains using Dataplex lakes and data zones. Manage, curate, tier, and archive your data easily with one-click.

# What is Data Catalog?

Data Catalog is a fully managed, scalable metadata management service within [Dataplex](#).

Why do you need Data Catalog?

Most organizations today are dealing with a large and growing number of data assets.

Data stakeholders (consumers, producers, and administrators) within an organization face multiple challenges:

- **Searching for insightful data:**
  - Data consumers don't know the location and origin of data. They have to navigate data "swamps".

- Data consumers don't know what data to use to get insights because most data is not well documented and, even if documented, is not well maintained.
- Data can't be found and is often lost when it resides only in people's minds.
- **Understanding data:**
  - Is the data fresh, clean, validated, approved for use in production?
  - Which dataset out of several duplicate sets is relevant and up-to-date?
  - How does one dataset relate to another?
  - Who is using the data and who is the owner?
  - Who and what processes are transforming the data?
- **Making data useful:**
  - Data producers don't have an efficient way to put forward their data for consumers. If there's no self-service, consumers may overwhelm producers. Several data engineers can't manually provide data to thousands of data analysts.
  - Valuable time is lost if data consumers have to find out how to request data access, request it, wait without a defined response time, escalate, and wait again.

Without the right tools, the challenges become a major obstacle to the efficient use of data. Data Catalog provides a centralized place that lets organizations achieve the following:

- Gain a **unified view** to reduce the pain of searching for the right data.
- Support data-driven decision making and accelerate the insight time by enriching data with **technical and business metadata**.
- Improve **data management** to increase operational efficiency and productivity.
- Take **ownership** over the data to improve trust and confidence in it.

Data Catalog functions

Data Catalog provides three main functions:

- Searching for data entries for which you have access
- Tagging data entries with metadata
- Providing column-level security for BigQuery tables

In addition, Data Catalog can leverage the results of a [Cloud Data Loss Prevention](#) (DLP) scan to identify sensitive data directly within Data Catalog in the form of tag templates.

# What is Dataplex?

Dataplex is an intelligent data fabric that helps you unify distributed data and automate data management and governance across that data to power analytics at scale.

Additionally, you can discover and curate metadata across various silos using catalog capabilities in Dataplex. See [Data Catalog Overview](#).

For data stored in Cloud Storage and BigQuery, Dataplex enables you to:

- Build a domain-specific data mesh across data stored in multiple Google Cloud projects, without any data movement.
- Consistently govern and monitor data with a single set of permissions.
- Automate metadata discovery and make the data securely accessible and available for querying via BigQuery as external tables, and open source applications such as SparkSQL, Presto, and HiveQL.
- Run data quality and data lifecycle management tasks, including serverless Spark tasks.
- Explore data using fully-managed, serverless Spark environments with simple access to notebooks and SparkSQL queries.

## Why use Dataplex?

Enterprises have data distributed across data lakes, data warehouses, and data marts. Dataplex enables you to discover, curate, and unify this data without any data movement, organize it based on your business needs, and centrally manage, monitor, and govern this data. Dataplex enables standardization and unification of metadata, security policies, governance, classification, and data lifecycle management across this distributed data.

