Enterprises are struggling to make high quality data easily discoverable and accessible for analytics, across multiple silos, to a growing number of people and tools within their organization. They are often forced to make tradeoffs—to move and duplicate data across silos to enable diverse analytics use cases or leave their data distributed but limit the agility of decisions.

Today we are excited to announce [Dataplex](http://cloud.google.com/dataplex), an intelligent data fabric that provides a way to centrally manage, monitor, and govern your data across data lakes, data warehouses and data marts, and make this data securely accessible to a variety of analytics and data science tools.

Dataplex provides an integrated analytics experience, bringing together the best of Google Cloud and open source tools, so you can rapidly curate, secure, integrate, and analyze data at scale. With built-in data intelligence using Google Artificial Intelligence (AI) and machine learning (ML) capabilities and a flexible consumption model, you can now spend less time wrestling with infrastructure and more time focused on driving business outcomes.

[Graphical user interface, application, Teams

Description automatically generated](https://storage.googleapis.com/gweb-cloudblog-publish/images/integrated_analytics_experience.max-2800x2800.jpg)

Dataplex enables you to:

* Achieve **freedom of choice** to store data wherever you want for the right price/performance and choose the best analytics tools for the job, including Google Cloud and open source analytics technologies such as Apache Spark and Presto.
* Enforce **consistent controls** across your data to ensure unified security and governance
* Take advantage of the **built-in data intelligence** using Google’s best in class AI/ML capabilities to automate much of the manual toil around data management and get access to higher quality data.

Early customers like Equifax, Loblaw, and ANZ are excited about using Dataplex to address data management complexity.

*“Dataplex will greatly simplify the existing analytics workflows within Equifax with its unified data fabric and single interface for policy management and governance across all our analytics data. Its built-in data discovery and data quality features will ensure that our data scientists and analysts always have access to high quality data that they can trust. Dataplex aligns well with our enterprise data strategy and we are excited to partner with Google Cloud on this.”  
—****Kumar Menon, SVP, Data Fabric & Decision Science Technology, Equifax.***

*“Loblaw is Canada's food and pharmacy leader, and we are excited to be an early adopter of Dataplex. We could significantly benefit from Dataplex as it provides a single pane of glass for end-to-end data management and governance. We are particularly interested in improving platform resilience and data quality by detecting anomalies as early as possible in the data pipeline with the help of Dataplex.”  
—****Elton Martins, Senior Director of Data Insights & Analytics, Loblaw***

*“We are undergoing a major data transformation at ANZ, bringing together our various data assets and building a cohesive data ecosystem for customer benefit. 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. We are excited to partner with GCP on Dataplex and test the product in private preview.”  
—****Ashish Shekhar, Head of Technology – Enterprise Analytics & Applied AI, ANZ***

Dataplex is built for distributed data. We are starting with data stored in [Google Cloud Storage](https://cloud.google.com/storage) and [BigQuery](https://cloud.google.com/bigquery), with support for other data sources coming soon. It provides a workflow-driven experience helping you build an open data platform and make data easily accessible to your end users while ensuring your policies and best practices are consistently enforced.

Organizing and curating your data

One of the core tenets of Dataplex is letting you organize and manage your data in a way that makes sense for your business, without data movement or duplication. For that, we are providing logical constructs like lakes, data zones and assets. Those constructs enable you to abstract away the underlying storage systems and become the foundation for setting policies around data access, security, lifecycle management, and so on.

For example, you can create a lake per department within your organization (Retail, Sales, Finance, etc.) and create data zones that map to data readiness and usage (landing, raw, curated\_data\_analytics, curated\_data\_science, etc.).

Once you have your lakes and zones setup, you can now attach data to these zones as assets. You can add data from different types of storage (e.g. GCS Bucket and BigQuery dataset) under the same zone. You can also attach data across multiple projects under the same zone.

[Graphical user interface, diagram

Description automatically generated](https://storage.googleapis.com/gweb-cloudblog-publish/images/Organizing_and_curating_your_data.max-2800x2800.jpg)

You can ingest data into your lakes and zones using the tools of your choice including services such as [Dataflow](https://cloud.google.com/dataflow), [Data Fusion](https://cloud.google.com/data-fusion), [Dataproc](https://cloud.google.com/dataproc), [Pub/Sub](https://cloud.google.com/pubsub)or choose from one of our partner products. Dataplex provides built-in 1-click templates for common data management tasks.

Securing your data

Dataplex enables you to define and enforce consistent policies across your data, irrespective of where it physically resides. Data owners can easily set up policies for specific data domains based on business needs, without thinking about where the data is stored while data stewards get global visibility into governance policies and permissions across their data.

You can apply security and governance policies for your entire lake, a specific zone, or an asset. Dataplex maps the policies to the underlying storage and pushes down permissions to the storage layer to provide end-to-end secure data access. Additionally, you can secure not just data but also related artifacts like notebooks, scripts, and models using the same set of access policies.

Making high quality data available for analytics and data science

One of the biggest differentiators for Dataplex is our **data intelligence** capabilities using Google’s best in class AI/ML technologies.  As you bring the data under management, Dataplex will automatically harvest the metadata for both structured and unstructured data, with built-in data quality checks. All of the metadata is automatically registered in a unified metastore, and made available for search and discovery. It is also published to [Bigquery](https://cloud.google.com/bigquery/?utm_source=google&utm_medium=cpc&utm_campaign=na-US-all-en-dr-skws-all-all-trial-e-dr-1009892&utm_content=text-ad-none-any-DEV_c-CRE_509008811228-ADGP_Desk%20%7C%20SKWS%20-%20EXA%20%7C%20Txt%20~%20Data%20Analytics%20~%20BigQuery_Big%20Query-KWID_43700060008413254-kwd-47616965283&utm_term=KW_bigquery-ST_bigquery&gclid=Cj0KCQjws-OEBhCkARIsAPhOkIYMg8VQuFSgDl2Nmbjy4Min9roNvvlEhAyKcmJ5zqcb6zfkk4-gDksaAueOEALw_wcB&gclsrc=aw.ds), [Dataproc Metastore](https://cloud.google.com/dataproc-metastore/docs/overview), and [Data Catalog](https://cloud.google.com/data-catalog) - ensuring that you have the same consistent data context and access across your tools.

For example, when you write parquet files to a Google Cloud Storage bucket - Dataplex will automatically extract metadata of these files, detect a tabular schema, including hive-style partitions, run data quality checks, and make this data queryable in BigQuery as an external table and from any open source or partner application - with the same consistent security and access policies you defined at the logical data layer.

Your data scientists and analysts now have secured access to this data that meets your quality bar and governance rules via the tools of their choice, without needing any additional processing.

[Graphical user interface

Description automatically generated](https://storage.googleapis.com/gweb-cloudblog-publish/images/dataplex.max-2800x2800.jpg)

One-click access to collaborative analytics

Dataplex provides fully managed, one-click analytics environments enabling you to use the power of Apache Spark and BigQuery with support for other engines coming in the future.

As data administrators, you now have the flexibility to pre-configure these environments with the right cost and financial governance measures without taking on the overhead of managing and maintaining the infrastructure required to power these environments. You can easily configure different environments for different types of workloads and share it with multiple users using their IAM credentials. Dataplex manages the provisioning, monitoring, scaling, and shutdown of these environments.

As data scientists, analysts, and engineers, you now have a turn-key experience to run your analysis using notebooks and a SQL workbench. You can search for notebooks and scripts alongside data, save and share your work with other users, and schedule your notebooks or scripts for recurring workloads - all using the same integrated experience within Dataplex.

Graphical user interface, application, Word

Description automatically generated

Building an open platform with industry leaders

We are partnering with industry leaders such as [Accenture](https://www.accenture.com/us-en/services/software-platforms/accenture-google-cloud-business-group), [Collibra](https://www.collibra.com/google), [Confluent](https://www.confluent.io/partner/google-cloud-platform/), [Informatica](https://www.informatica.com/solutions/explore-ecosystems/google-cloud-platform.html), [HCL](https://www.hcltech.com/cloud/google-cloud), [Starburst](https://www.starburst.io/platform/deployment-options/starburst-on-google-cloud/), [NVIDIA](https://cloud.google.com/nvidia), [Trifacta](https://www.trifacta.com/solutions/google/), and others to build an open platform to power analytics at scale. Our partners are excited about the capabilities that Dataplex will provide:

*“Collibra is excited to partner with Dataplex to provide data governance and data quality for consistent controls across distributed data. Pairing Collibra's multi-cloud and hybrid solution with Dataplex allows enterprises to securely open up access to more, higher quality data for users and analytics using a single unified view.”****—Jim Cushman, Chief Product Officer, Collibra***

*"Dataplex builds on Google Cloud’s commitment to open source by integrating with Apache Kafka®, a leading open source platform for event streaming. We at Confluent, the platform for data in motion that completes Apache Kafka® to be enterprise-ready, are excited to partner with Dataplex to enable customers to bring together distributed, real-time data and build a unified data fabric for end to end analytics."****—Paul Mac Farland, VP, Head of Customer Solutions and Innovation, Confluent***

*"We are excited to partner with Google Cloud's Dataplex team as we look to provide our joint customers an integrated and open data fabric for analytics at scale. Extending Dataplex's data management and data quality capabilities with Starburst Enterprise will accelerate time to value for enterprises looking to connect distributed data without having to move data."****—Justin Borgman, CEO and Co-founder, Starburst***

Next Steps

Dataplex is now available in preview for a select number of customers. For more information, visit our [website](http://cloud.google.com/dataplex) or watch the [recording](https://cloudonair.withgoogle.com/events/summit-data-cloud?talk=breakout_track2_intelligent_data_analytics). If you would like to sign up, please click [here](https://docs.google.com/forms/d/e/1FAIpQLSfQ-jxrRbd4A2iwKuVMSvefdsGEd3HTeojluZNmiYfdEFRYjg/viewform).

[RELATED ARTICLE](https://cloud.google.com/blog/products/data-analytics/transform-your-business-with-the-data-cloud)

[Transforming your business with the data cloud](https://cloud.google.com/blog/products/data-analytics/transform-your-business-with-the-data-cloud)

[Accelerate your business transformation with the data cloud.](https://cloud.google.com/blog/products/data-analytics/transform-your-business-with-the-data-cloud)