BigLake

Built on years of investment in BigQuery, BigLake is a storage engine that allows organizations to unify data warehouses and lakes, and enable them to perform uniform fine-grained access control, and accelerate query performance across multi-cloud storage and open formats.

- Store a single copy of data with uniform features across data warehouses & lakes.
- Fine-grained access control and multi-cloud governance over distributed data.
- Seamless integration with open source analytics tools and open data formats.

BENEFITS

Freedom of choice

Unlock analytics on distributed data regardless where and how it's stored, while choosing the best analytics tools, open source or cloud native over a single copy of data.

Secure and performant data lakes

Fine-grained access control across open source engines like Apache Spark, Presto and Trino, and open formats such as Parquet. Performant queries over data lakes powered by BigQuery.

Unified governance & management at scale

Integrates with [Dataplex](#) to provide management at scale, including logical data organization, centralized policy & metadata management, quality and lifecycle management for consistency across distributed data.

KEY FEATURES

Key features

Fine grained security controls

BigLake eliminates the need to grant file level access to end users. Apply table, row, column level security policies on object store tables similar to existing BigQuery tables.

Multi-compute analytics

Maintain a single copy of data and make it uniformly accessible across Google Cloud and open-source engines, including [BigQuery](#), [Vertex AI](#), [Dataflow](#), Spark, Presto, Trino, and Hive using BigLake connectors. Centrally manage security policies in one place, and have it consistently enforced across the query engines by the API interface built into the connectors.

Multi-cloud governance

Discover all BigLake tables, including those defined over Amazon S3, Azure data lake Gen 2 in [Data Catalog](#). Configure fine grained access control and have it enforced across clouds when querying with [BigQuery Omni](#).

Performance acceleration

Achieve industry leading performance over data lake tables on Google Cloud, AWS and Azure, powered by proven BigQuery infrastructure.

Built on open formats

Gain access to the most popular open data formats including Parquet, Avro, ORC, CSV, JSON. The API serves multiple compute engines through Apache Arrow.