Open Table Formats for VantageCloud Lake and AI Unlimited

Data management challenges

Organizations with petabytes of data that must be stored for extended periods of time and consistently updated need ways to reduce costs and complexity. Moving, replicating, and transforming data across systems is time-consuming and often results in unsynchronized data stored in multiple silos.

Open table formats (OTFs) provide a standardized storage abstraction layer that defines the structure and rules for how analytical engines interact with data in object storage. This architecture enables you to maintain a single data store for use by multiple analytical engines and tools without needing to host, replicate, and manage multiple copies of large datasets.

While OTFs accelerate open access to the storage layer for better cost optimization, they do require external management of the storage layer using proprietary catalogs. This includes placing, managing, landing, and securing data, while catalogs act as central repositories and queries to find and interact with data.

Teradata's unique approach

Teradata takes a unique, open, and connected approach to supporting OTFs, enabling cross-read, cross-write, and cross-query of data stored in a variety of OTFs and open catalogs in multi-cloud and multi-data-lake environments. This future-ready approach allows enterprises to employ a truly modern data strategy, with unmatched agility and flexibility to deliver Trusted AI at scale—all without the need to move, replicate, or transform data.

Teradata offers freedom of choice with first-party support for Apache Iceberg and Linux Foundation Delta Lake OTFs, residing in object stores like AWS S3, Azure Data Lake Storage, and Google Cloud Storage. These data stores utilize open file formats like Parquet, Avro, ORC for Iceberg OTF, and Parquet for Delta Lake OTF.

Teradata's implementation of OTFs supports simultaneous connections to multiple metadata catalogs compatible with OTFs, such as Apache Iceberg and Delta Lake. Use your preferred catalog, such as Hive Metastore or AWS Glue, with your preferred OTF storage layer. While many data platforms lock users into a vendor-preferred catalog and limit them to cross-reading if they use different or multiple catalogs,

Teradata provides an open ecosystem for OTFs, catalogs, and cloud service providers (CSPs), allowing Teradata analytics engines to run workloads for business intelligence, data engineering, data science, and machine learning.
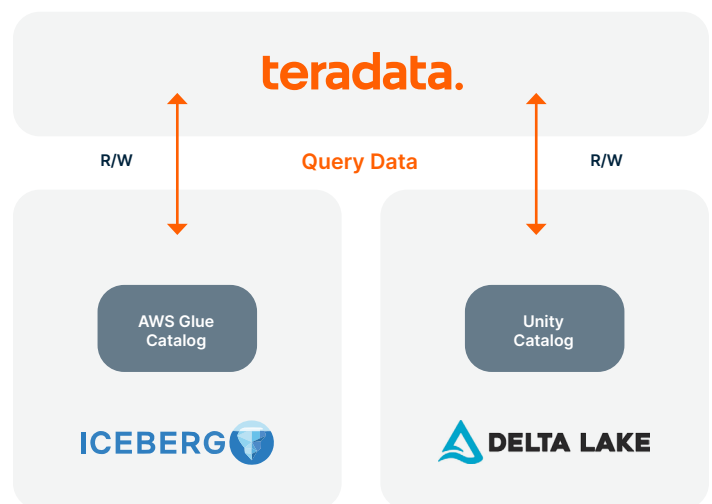
OTFs: Key features

- **Atomicity, consistency, isolation, and durability (ACID) transactions.** Ensure transactional consistency for multiple, concurrent read and write operations.

- **Schema evolution.** Modify schemas (column definition, order, and data types) or partitions without re-creating new tables or partitions, saving time and data movement costs.

- **Time travel.** Return to previous table states for historical analysis across time periods.

- **Rollback.** Revert to prior versions to quickly correct issues and return tables to a known good state.

teradata.

The best engines for your use cases: VantageCloud Lake and AI Unlimited

Leverage Teradata VantageCloud Lake and Teradata AI Unlimited against your own datasets, in your own cloud object stores—exploring AI use cases, accelerating innovation, and maximizing value with an agile approach.

As a cloud-native analytics and data platform, VantageCloud Lake can run independent elastic workloads for the entire enterprise. It supports multiple tiers of storage for different workload needs, including:

- Block File System for operational/tactical queries
- Object File System for performant storage for clusters
- New: OTFs for easily accessing data to cross-read, cross-write, and cross-query across CSPs

AI Unlimited delivers the power of an on-demand AI/ML engine, with an enterprise-grade AI service that maximizes exploration and discovery in a low-cost environment. It offers unlimited flexibility and scale to make AI innovation faster, easier, and more cost-effective. With AI Unlimited, you can get started quickly, analyze data where it lives in Iceberg or Delta Lake tables, and eliminate fear of cost overruns, thanks to on-demand compute with pay-as-you-go pricing.

Both VantageCloud Lake and AI Unlimited include Teradata's industry-leading ClearScape Analytics™ capabilities. With ClearScape Analytics, you can quickly and effectively scale AI/ML to solve your most complex challenges, reduce costs and friction, and accelerate time to value throughout your organization.

As a first-class citizen, Teradata follows all data management practices and read/write features in OTF open-source specifications—and doesn't deviate from upstream OTF or catalog specifications. With this approach, we easily adapt to continued innovations in Iceberg and Delta Lake OTF standards and enable support for additional OTF standards and compatible catalogs. With OTFs, VantageCloud Lake and AI Unlimited engines can efficiently perform compute operations—such as data engineering, exploration, and analysis—on external, customer-managed, shared data stores, all while upholding data consistency and integrity.
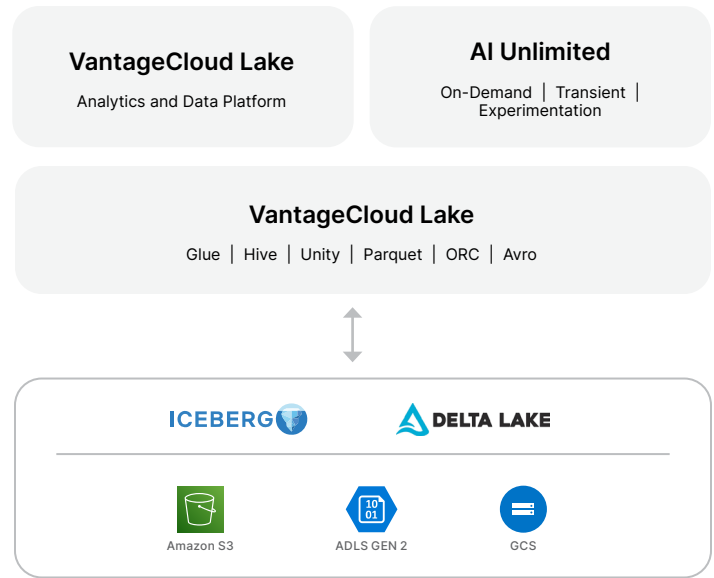
Optimize performance for parallel processing of shared data

Teradata's OTF implementation delivers differentiated, parallel query performance by leveraging world-class features, such as:

- AMP parallelism, which breaks down requests into smaller components that are executed in parallel
- Optimizer, which determines the most cost-effective query plan for data in all Teradata storage tiers
- Teradata Active System Management (TASM) workload management, which allows prioritization of workloads, performance tuning, and workload monitoring and management

About Teradata

At Teradata, we believe that people thrive when empowered with trusted information. We offer the most complete cloud analytics and data platform for AI. By delivering harmonized data and Trusted AI, we enable more confident decision-making, unlock faster innovation, and drive the impactful business results organizations need most.
See how at Teradata.com.