

Data Without Limits

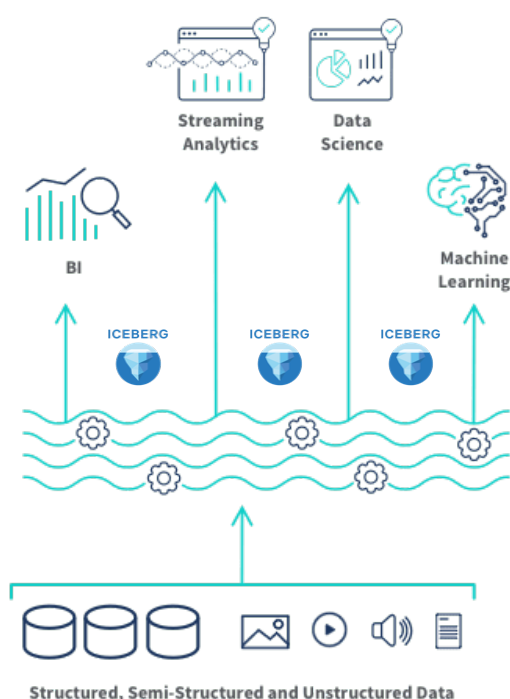
Powering Open Lakehouses with Apache Iceberg



Open lakehouses powered by Apache Iceberg offer a modern approach to data management, combining the scalability of data lakes with the reliability and performance of data warehouses — all while leveraging open standards.

Organizations are increasingly adopting open lakehouses with Apache Iceberg to reduce data silos, improve interoperability, and unlock significant cost savings in the cloud. Based on industry surveys, enterprises are expecting to save as much as **50% of their analytical costs** by moving to a lakehouse architecture.

With its support for multiple processing engines like Apache Spark, Snowflake, Databricks, Trino, and Presto, Apache Iceberg enables organizations to seamlessly access and analyze massive datasets, driving analytics, AI, machine learning, and real-time insights, all from a single repository.



What is an open lakehouse?

An open lakehouse combines the strengths of data warehouses and data lakes into a unified architecture, enabling seamless storage and analysis of structured, semi-structured, and unstructured data. Built on open standards, it ensures interoperability across various data processing engines, allowing organizations to:

- **Eliminate silos and reduce data movement** for more efficient workflows
- **Lower costs** with cloud-based object storage
- **Enhance governance** with consistent data quality and access control

Unlock the flexibility to support BI, machine learning, and data exploration — all in one platform.

Challenges abound in building performant lakehouses

However, this is still a rapidly evolving domain, and organizations have to navigate some murky waters to get maximum value out of lakehouses and to realize the true promise of Apache Iceberg.

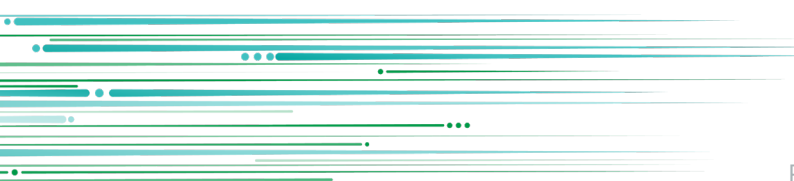
Some of the key challenges include:

- **Large scale ingestion into Iceberg:** Iceberg, is an open table format that, lacks native data ingestion capabilities. Organizations increasingly need data in real time or near real time for business-critical workloads. This means data teams need to look for external tools or APIs to write data from operational sources into Iceberg in real time or near real time easily and efficiently.
- **Iceberg table optimizations:** Iceberg optimization services require the user to tune and tweak optimization behavior by manually configuring each and every table to drive performance. This process is tedious, complicated, and simply not scalable. While customers can use open-source tools like Apache Spark to write and optimize Iceberg tables, they are difficult to scale beyond a few tables, require constant manual tuning and significant engineering resources to maintain.
- **Complex data pipeline management:** Customers must develop and maintain complex and error-prone code to ingest, optimize, and maintain data on their own. They have to account for high data volumes, managing sudden spikes in data volume (bursts), schema evolution, data quality, observability, and more.
- **Ensure data quality & trust:** Organizations want to ensure the data in their lakehouse is accurate, complete, timely, and trustworthy before it can be consumed for analytics and model training. More importantly, data engineers want to have an accurate view of the lineage, so they know where any specific piece of data came from, how it was split and merged with other data, and what transformations have been applied.

Introducing Qlik Open Lakehouse

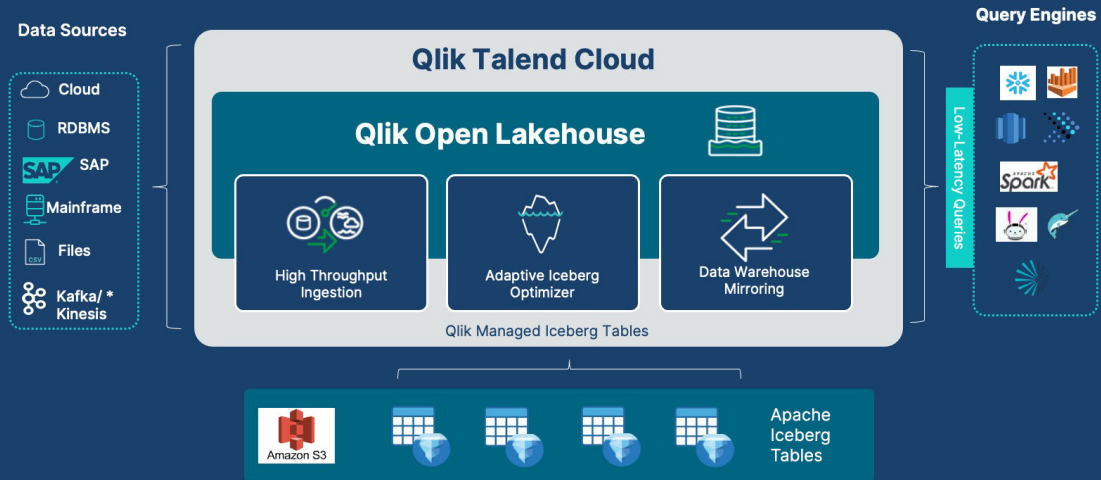
Qlik Open Lakehouse is a fully managed capability within Qlik Talend Cloud that makes it easy, effortless, and cost-effective for users to ingest, process, and optimize large amounts of data in Apache Iceberg-based lakehouses. With Qlik Open Lakehouse, customers can now set up a Lakehouse on their Amazon S3 environment and load data directly into Apache Iceberg tables with just a few clicks.

In addition, Qlik Open Lakehouse handles the hard parts by automatically mapping source to target data types, resolving type conflicts, partitioning, evolving schemas, updating and deleting rows, and optimizing the target Iceberg tables to reduce storage and accelerate queries. This enables Qlik Open Lakehouse to deliver **2.5x - 5x** better query performance, while reducing storage costs by up to **50%**.



Qlik Open Lakehouse

Fully managed and optimized open lakehouse powered by Apache Iceberg



Some of the key features include:

- **Real-time high throughput ingestion into Iceberg:** Users can quickly and easily ingest batch and real-time data from diverse sources directly into Iceberg tables, making data loading simple for lakehouses. Qlik Open Lakehouse's ingestion engine ensures frequently updating data, like those delivered using change data capture (CDC) from operational databases — including Type1/ Type 2 history — is written, merged, and optimized quickly and efficiently to meet customers' strict SLAs.
- **Adaptive Iceberg optimizations:** Qlik Open Lakehouse's industry-leading **Adaptive Iceberg Optimizer** technology continuously monitors tables and determines the ideal optimizations, compactions, and clean ups to execute based on each table's unique characteristics, delivering an unmatched performance boost (up to **2.5x - 5x improvement in performance**), and up to a 50% reduction in costs — all without writing a single line of code.
- **Data warehouse mirroring:** Automatically mirror data from Qlik Open Lakehouse Iceberg tables into your cloud data warehouse (such as Snowflake) to enable querying or additional downstream transformations, all **without duplicating or creating a copy of the data**. This ensures interoperability with your existing systems, minimizes data movement, and delivers further cost savings.
- **Unified, end-to-end offering for lakehouses:** Using Qlik Talend Cloud with Open Lakehouse, customers can now deploy a single **end-to-end solution** to ingest, transform, govern, optimize, and manage data pipelines for their Iceberg-based lakehouses, while ensuring data quality and trust — all without cobbling together multiple pieces of technology.
- **Open, industry-leading integrations:** Qlik Open Lakehouse **integrates with industry-leading catalogs** such as AWS Glue, Polaris (incubating), and Snowflake Open Catalog, enabling customers to deliver high-performance queries on petabytes of data using any iceberg-compatible platform or query engine such as Snowflake, Amazon Athena, Databricks, Apache Spark, Dremio, ClickHouse, Flink, Trino, Presto, and more.

High throughput ingestion into Iceberg

Ingest data in real time from hundreds of sources including databases, SaaS sources, and SAP directly into Iceberg.

Open and interoperable

Integrates with leading Iceberg catalogs, processing platforms, and query engines.

Continuous and adaptive Iceberg optimization

Optimize Iceberg tables with Adaptive Optimizer to drive **5x** query performance with zero manual effort.

Drive 50% cost savings with 5x faster queries

Drive faster queries and unlock cost savings of **>50%** through optimizations and storage cost reductions.

Mirror data to your data warehouses

Seamlessly mirror data to your Snowflake warehouse without the need to copy data.

Proven performance at petabyte scale

A battle-tested, proven solution for ingesting and managing data at PB scale.

Build on Iceberg with Qlik Open Lakehouse

Qlik Talend Cloud — with Open Lakehouse — offers an independent offering to ingest, optimize, process, and manage Iceberg tables on Amazon S3, while delivering comparable query performance to that of a data warehouse at 50% lower costs. It brings Qlik Talend Cloud's high throughput, low latency ingestion and efficient compute resources together with automated Iceberg optimization capabilities to deliver a 5x boost in query performance with 50% in cost savings.

Start building with Open Lakehouse on Qlik Talend Cloud!

For more details, go to:

Qlik Open Lakehouse



About Qlik

Qlik transforms complex data landscapes into actionable insights, driving strategic business outcomes. Serving over 40,000 global customers, our portfolio leverages advanced, enterprise-grade AI/ML and pervasive data quality. We excel in data integration and governance, offering comprehensive solutions that work with diverse data sources. Intuitive and real-time analytics from Qlik uncover hidden patterns, empowering teams to address complex challenges and seize new opportunities. Our AI/ML tools, both practical and scalable, lead to better decisions, faster. As strategic partners, our platform-agnostic technology and expertise make our customers more competitive.

qlik.com