Build Your Open Data Lakehouse on Apache Iceberg with Dremio and Fivetran

At a Glance

- ✓ Fivetran automates the most time-consuming parts of the ELT process so it's easy to get data into the cloud data lake.

- ✓ Apache Iceberg tables deliver data lakehouse management capabilities for large data volumes while reducing the risk of vendor lock-in.

- ✓ The Dremio Open Data Lakehouse provides all of the functionality of the data warehouse directly on data lake storage, enabling high-performance self-service analytics for technical and non-technical data consumers.

The Data Explosion

Businesses today are facing a data explosion. They're collecting data from a wide range of sources and in a variety of formats. The growth of data volumes is forcing necessary shifts in data architectures. In the past, when businesses were concerned with structured data from business systems in the data center, data warehouses were the standard solution for Business Intelligence (BI) and reporting workloads.

Today, businesses are focused on the newest and fastest-growing sources of customer and operational data, which is often semistructured and unstructured data from sources outside of the data center. Data teams use cloud data lakes as cheap and efficient storage for a wide variety of data types. However, data lakes lack the data management and data governance capabilities of data warehouses.

Apache Iceberg: An Analytic Table Format for Data Lakes

Apache Iceberg is an open table format built specifically for high performance analytics on large volumes of data. It is quickly growing in popularity relative to other table formats, including Hive, Delta Lake, and Hudi. Apache Iceberg delivers the simplicity of SQL to the data lake.

Built for Big Data: Apache Iceberg is an analytic table format designed to deliver analytics on the largest datasets in the world. It features data optimization features like compaction and hidden partitioning ensuring high performance queries at massive scale.

Engine Freedom: Most processing engines are designed for specific use cases, but are not designed for every job. With Apache Iceberg, data teams can use the best engine for each specific workload, including Dremio Sonar, Spark, Flink, Trino, Presto, and more.

Open Data, No Lock-In: With Apache Iceberg, data remains in the data lake in an open table format. Data teams can use best-in-breed tools available today and remain agile enough to adopt the next wave of innovation.

Use Fivetran to Write Data into Iceberg Tables in Amazon S3

Fivetran is the automated data movement platform that makes it easy for data teams to move data across cloud data platforms. Data teams can connect data sources with more than 300 pre-built, no-code source connectors, select their destination, and easily move the data. Fivetran automates many of the time-consuming aspects of the Extract, Load & Transform (ELT) process, including data normalization, updates, and transformations. Fivetran protects data in flight and at rest with automated security and governance.

dremio | Fivetran

Getting Insights from the Data Lake is Easier than Ever with Fivetran and Dremio

Dremio and Fivetran make it easy to get from data to insights:

- ✔ Fivetran automates the ELT process so data teams can quickly populate their AWS S3 data lakes

- ✔ Apache Iceberg tables ensure high-performance analytics based on all of the data in the data lake

- ✔ The Dremio Open Data Lakehouse simplifies data management and data governance while providing broad and easy access to data for technical and non-technical data consumers.

Dremio is the Easy and open Data Lakehouse

Dremio delivers all of the functionality of the data warehouse directly on data lake storage. Dremio Cloud on AWS S3 features two products:

Dremio Arctic ◆

Dremio Sonar •))

An open data lakehouse management service built on Apache Iceberg that gives data teams the tools to easily and efficiently manage data in AWS S3. Dremio Arctic features Data as Code functionality so teams can use branches, tags, and commits to easily make changes to data without impacting production workloads, and automatic table maintenance to ensure high performance analytics.

A SQL query engine, query accelerator, and semantic layer based on Apache Arrow. Dremio Sonar is purpose-built for sub-second query performance on massive volumes of data, and enables self-service access to data for technical and non-technical data consumers.

Try Dremio and Fivetran today!

Dremio Cloud is available today for AWS customers. Try it for free today at
www.dremio.com/get-started
All you pay for is AWS resources.

Fivetran is free to try for 14 days. Get started at
https://fivetran.com/signup