Catalogs in Apache Iceberg: Choosing the Right One

hadoop

REST

HIVE

Nessie

JDBC

snowflake®

DynamoDB

APACHE GRAVITINO

APACHE POLARIS™

Lakekeeper

AWS Glue

Never miss an update
Follow for more!

RisingWave

What is an Iceberg Catalog?

An **Iceberg Catalog** helps you locate and manage tables by storing and organizing their metadata.

It acts as a lookup system—when a system needs to interact with a table, the catalog tells it **where to find the table and how to access it** .

Catalogs are essentially the gateway to your data. They play a critical role in data accessibility. The file format wars are behind us—now, it’s about building actionable insights. But we still face a complex frontier: catalogs.

So, what exactly is a catalog? Is it a schema registry? A metadata store? A business discovery tool? It can be ambiguous. However, in this context, we're referring to **Iceberg catalogs** .

A Brief History: Why Catalogs Matter

As internet users and data volumes grew, vertically scalable databases (Oracle, MySQL) gave way to massively parallel systems (Greenplum, Teradata). Then came Hadoop, which democratized big data by combining:

- Distributed storage (HDFS)
- A reverse-engineered MapReduce
- A resource orchestrator

Hive added a SQL layer on Hadoop, followed by improvements such as Avro (enhanced schema evolution) and Parquet (improved columnar read performance).

Now in the **Lakehouse era** , data resides in object stores (like S3), stored in open formats (like Iceberg), and queryable by multiple engines.

But even with all these pieces—data, format, engines—you still need a **catalog** .