Apache Iceberg Crash Course

The Role of Apache Iceberg Catalogs

Curriculum

July 11: What is a Data Lakehouse and What is a Table Format?

July 16: The Architecture of Apache Iceberg, Apache Hudi and Delta Lake

July 23: The Read and Write Process for Apache Iceberg Tables

Aug 13: Understanding Apache Iceberg's Partitioning Features

Aug 27: Optimizing Apache Iceberg Tables

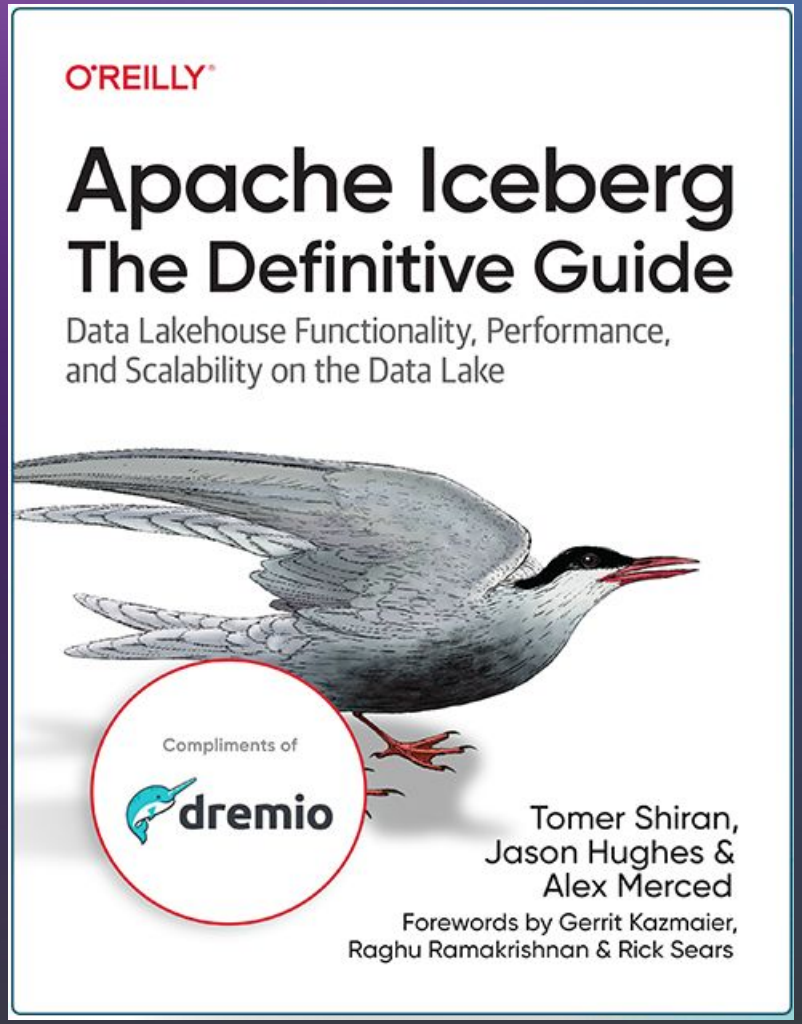
Sep 3: Streaming with Apache Iceberg

Sep 17: The Role of Apache Iceberg Catalogs

Oct 1: Versioning with Apache Iceberg

Oct 15: Ingesting Data into Apache Iceberg with Apache Spark

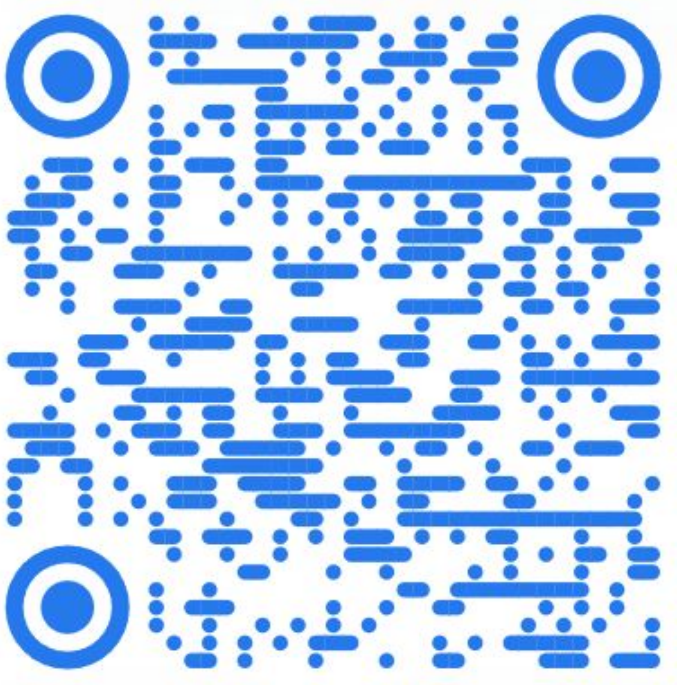
Oct 29: Ingesting Data into Apache Iceberg with Dremio


O'REILLY
Apache Iceberg
The Definitive Guide
Data Lakehouse Functionality, Performance, and Scalability on the Data Lake
Compliments of
dremio
Tomer Shiran, Jason Hughes & Alex Merced
Forewords by Gerrit Kazmaier, Raghu Ramakrishnan & Rick Sears

Podcast

dremio.com/gnarly-data-waves
Youtube | Spotify| iTunes

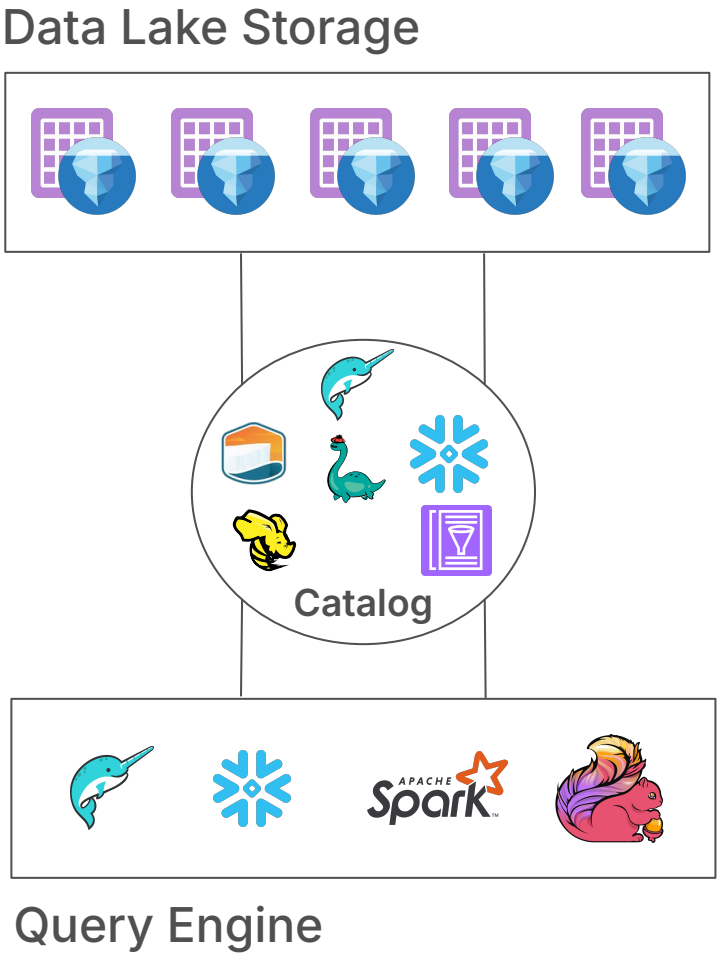
Questions



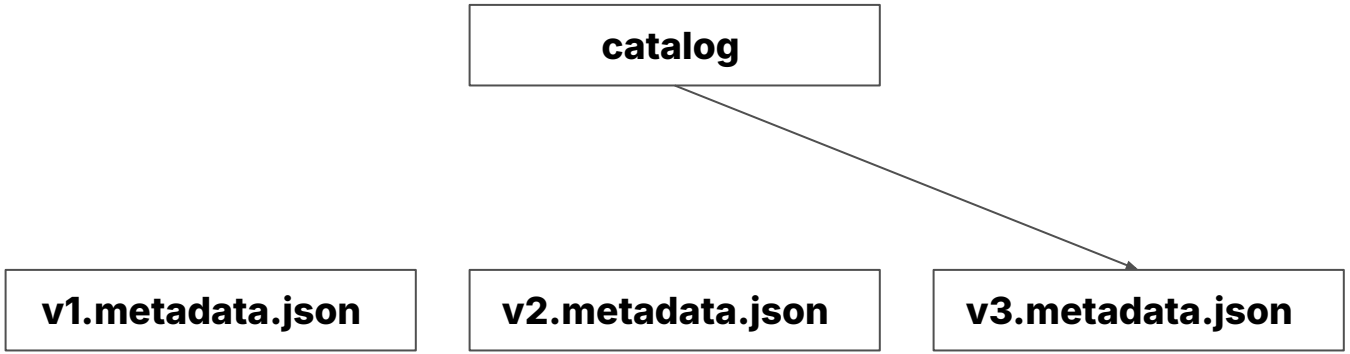
community.dremio.com
Apache Iceberg Category

Apache Iceberg Catalogs

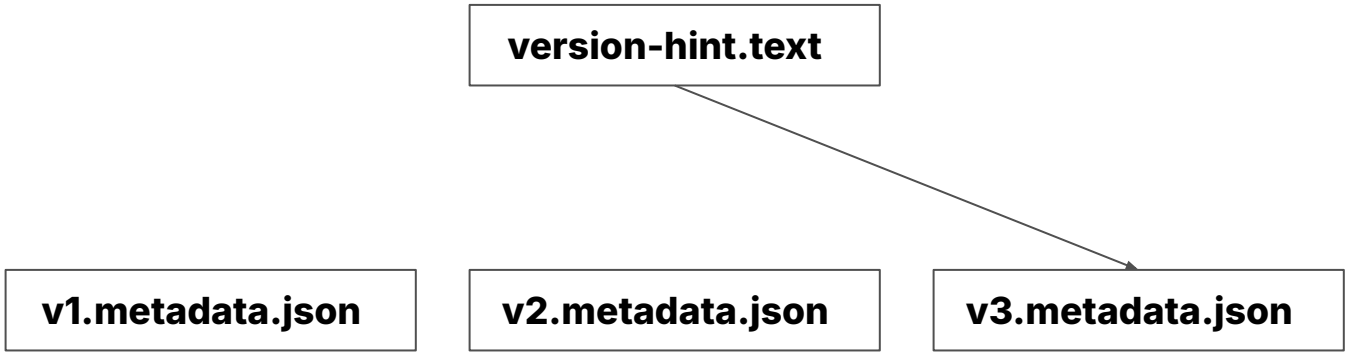
Role of the Iceberg Catalog



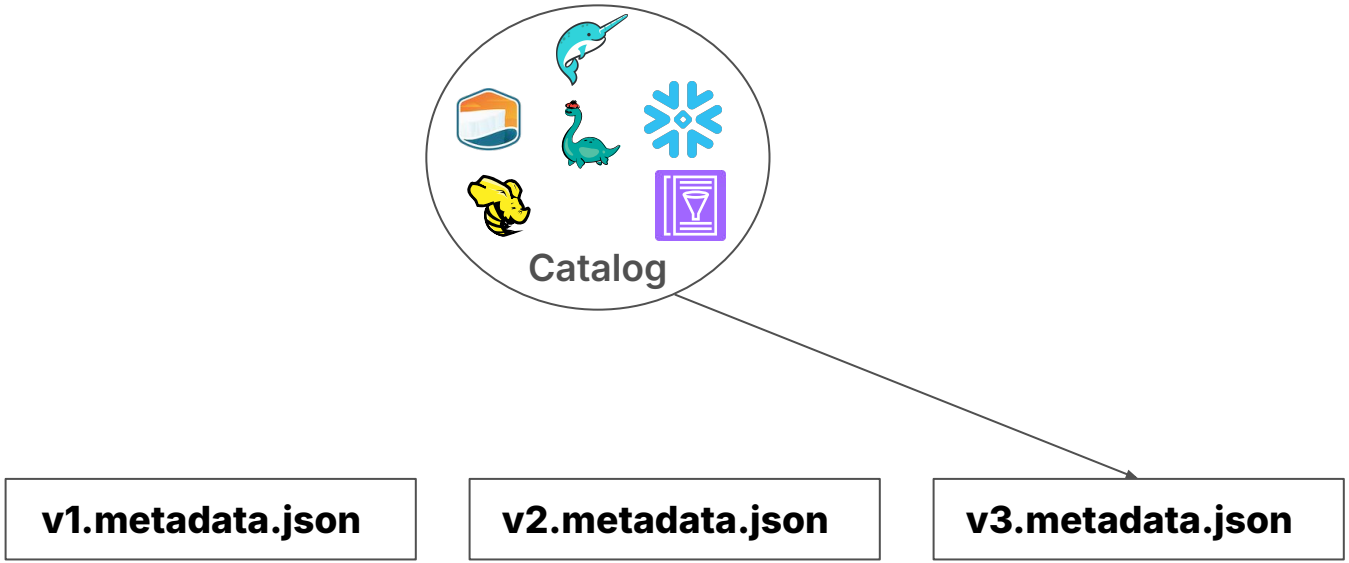
The Catalog Table Reference



The Catalog Table Reference - File System Catalog (Hadoop)



The Catalog Table Reference - Service Catalog



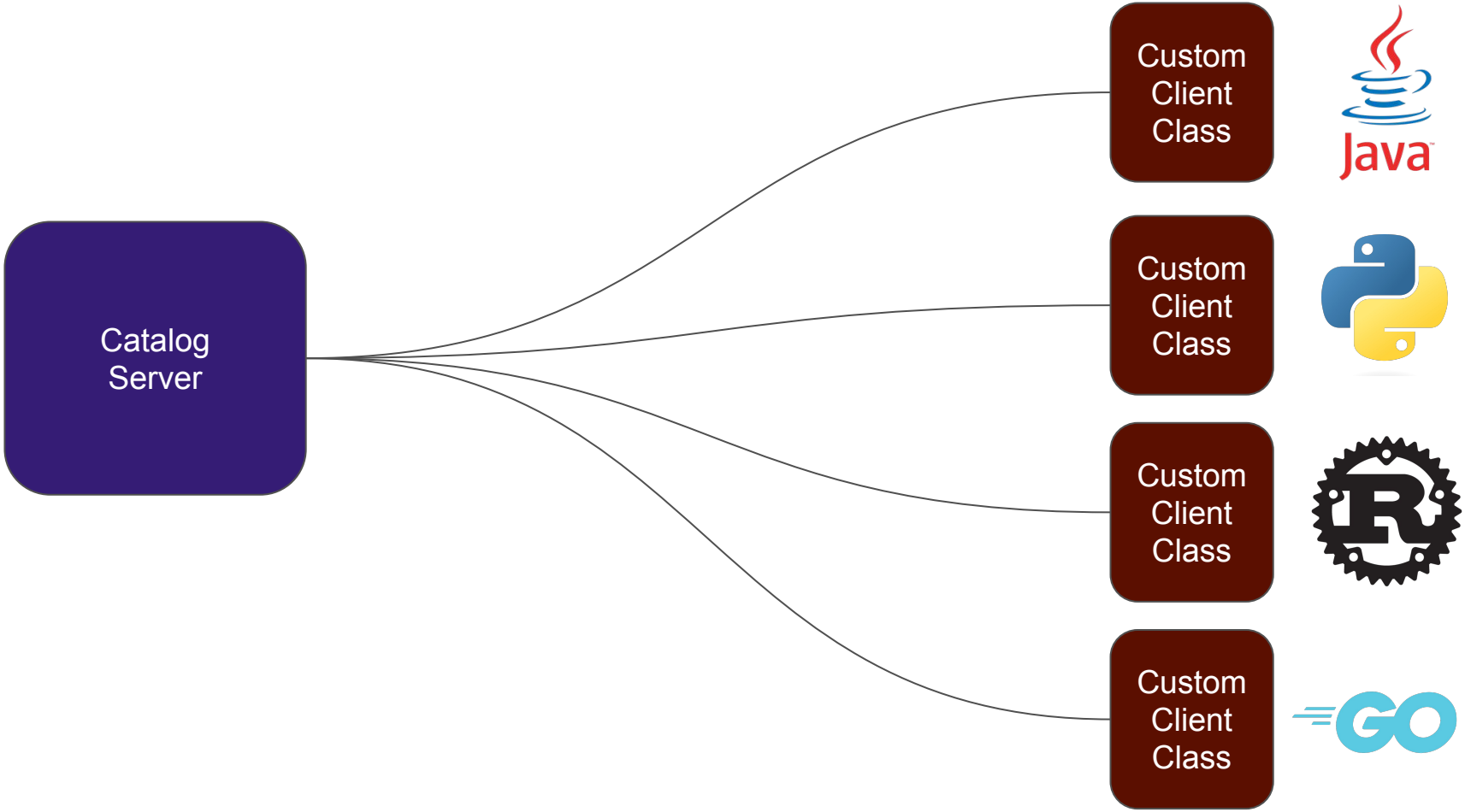
The Apache Iceberg REST Catalog

Problems

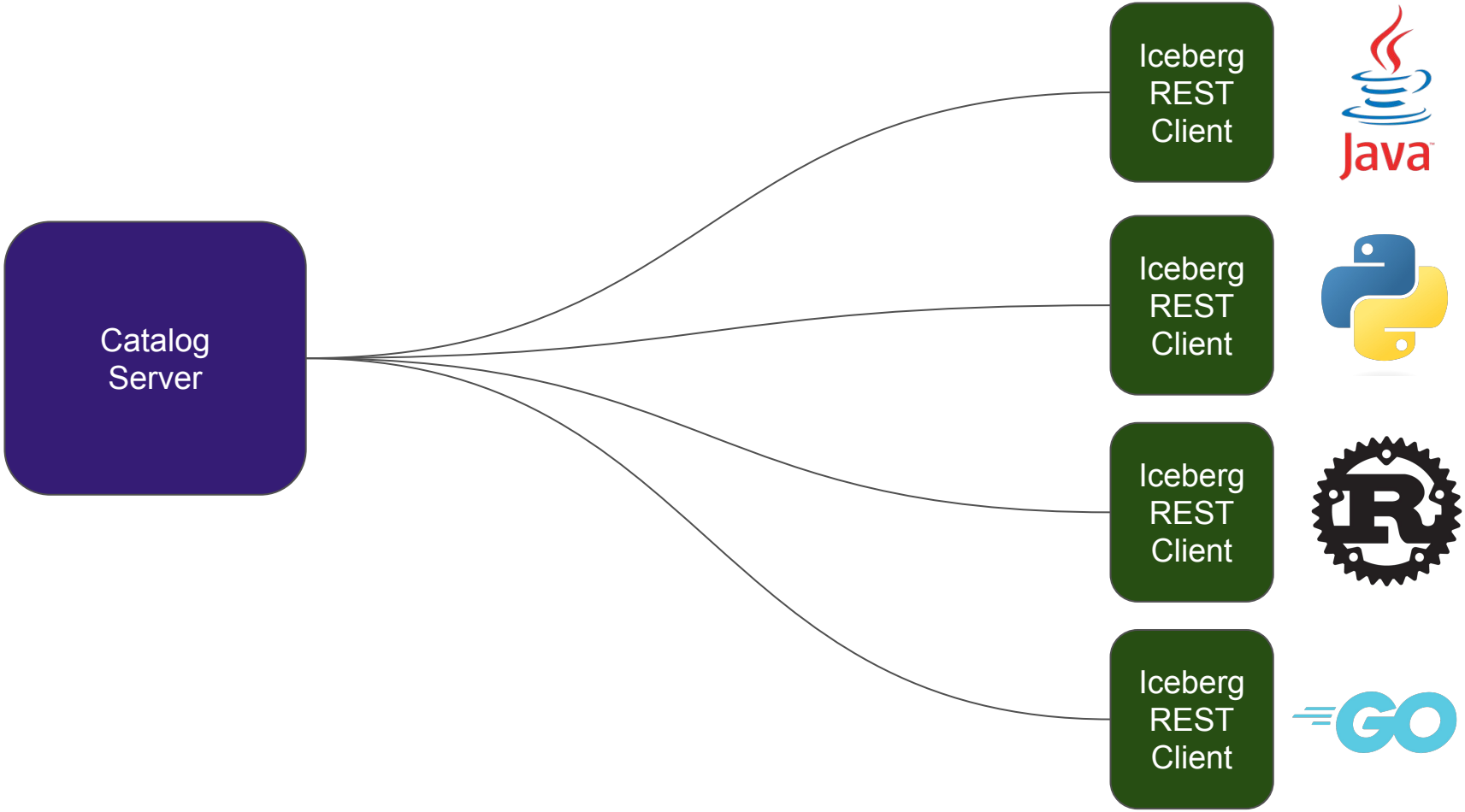
- Have to be re-implemented per language
- Engines would have to implement support for each catalog
- No control whether end-users are using most up to date version

Confidential - Do Not Share or Distribute

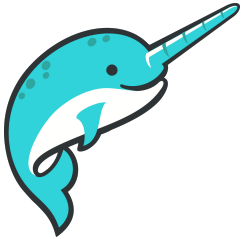
The Catalog Status Quo



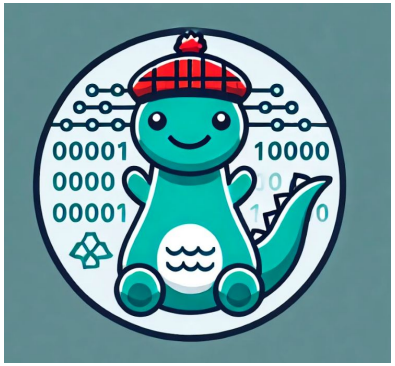
The Apache Iceberg REST Specification



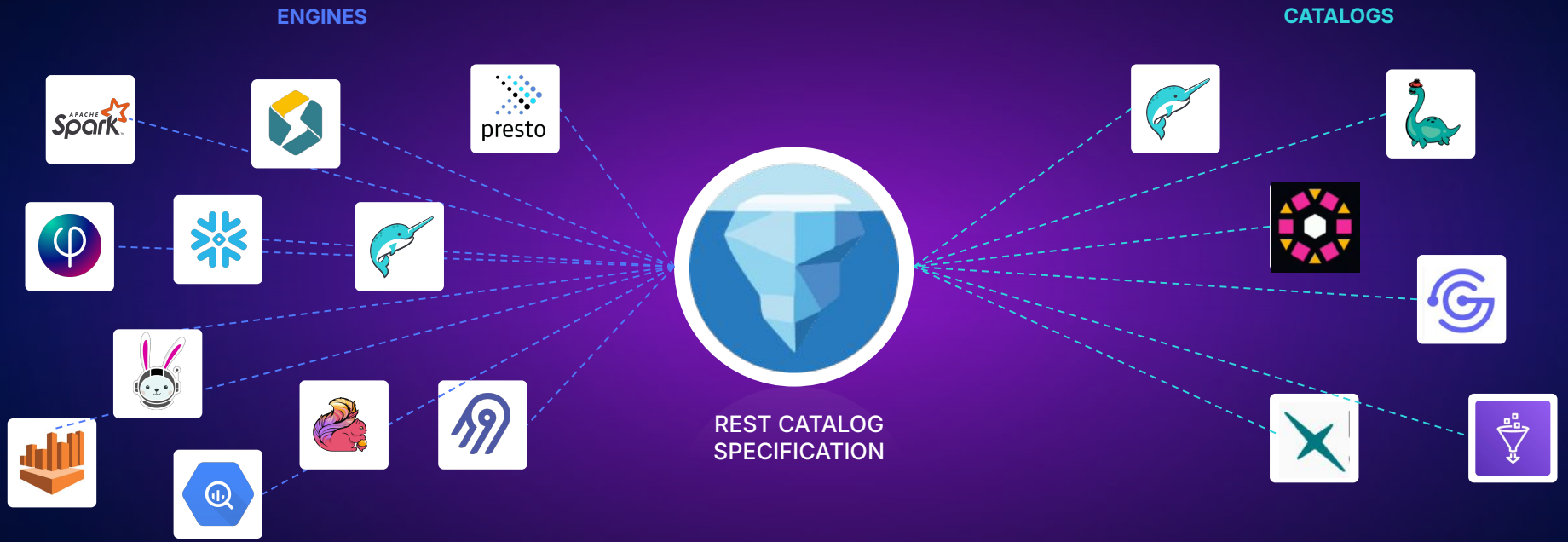
REST Catalog Client Support



Open Source Catalogs using REST CATALOG Spec



Announcing Support for the Iceberg REST Catalog Specification



Any Engine That Supports REST Catalog Specification Can Connect to Any Catalog that Supports REST Catalog