



DATA GOVERNANCE & SECURITY

Analytics in 15

Centralize Data Lake
Security Management with AWS

Ramkumar Nottath

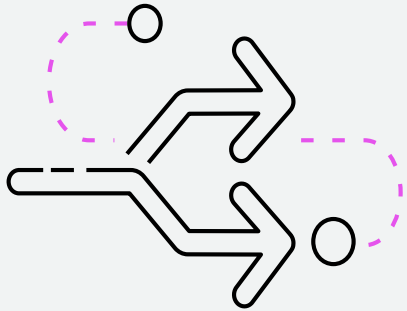
Sr. Solutions Architect, Data & Analytics
Amazon Web Services

Agenda

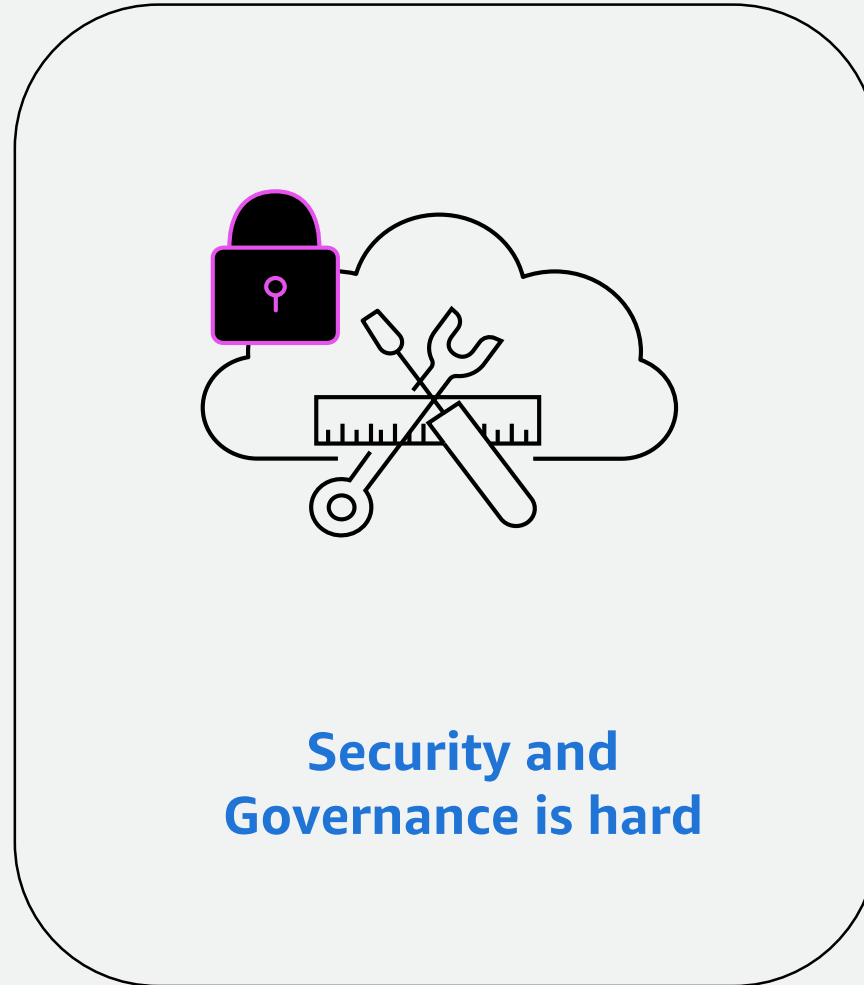
- Customer challenges in building data lakes
- Centralize Data Lake Security Management with AWS
- How can you get started?

Customer challenges

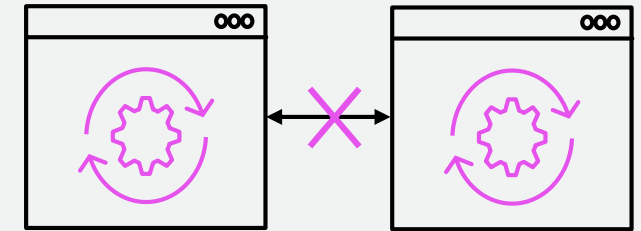
Challenges in building data lakes



Data Ingestion & Data Management Challenges



Security and Governance is hard

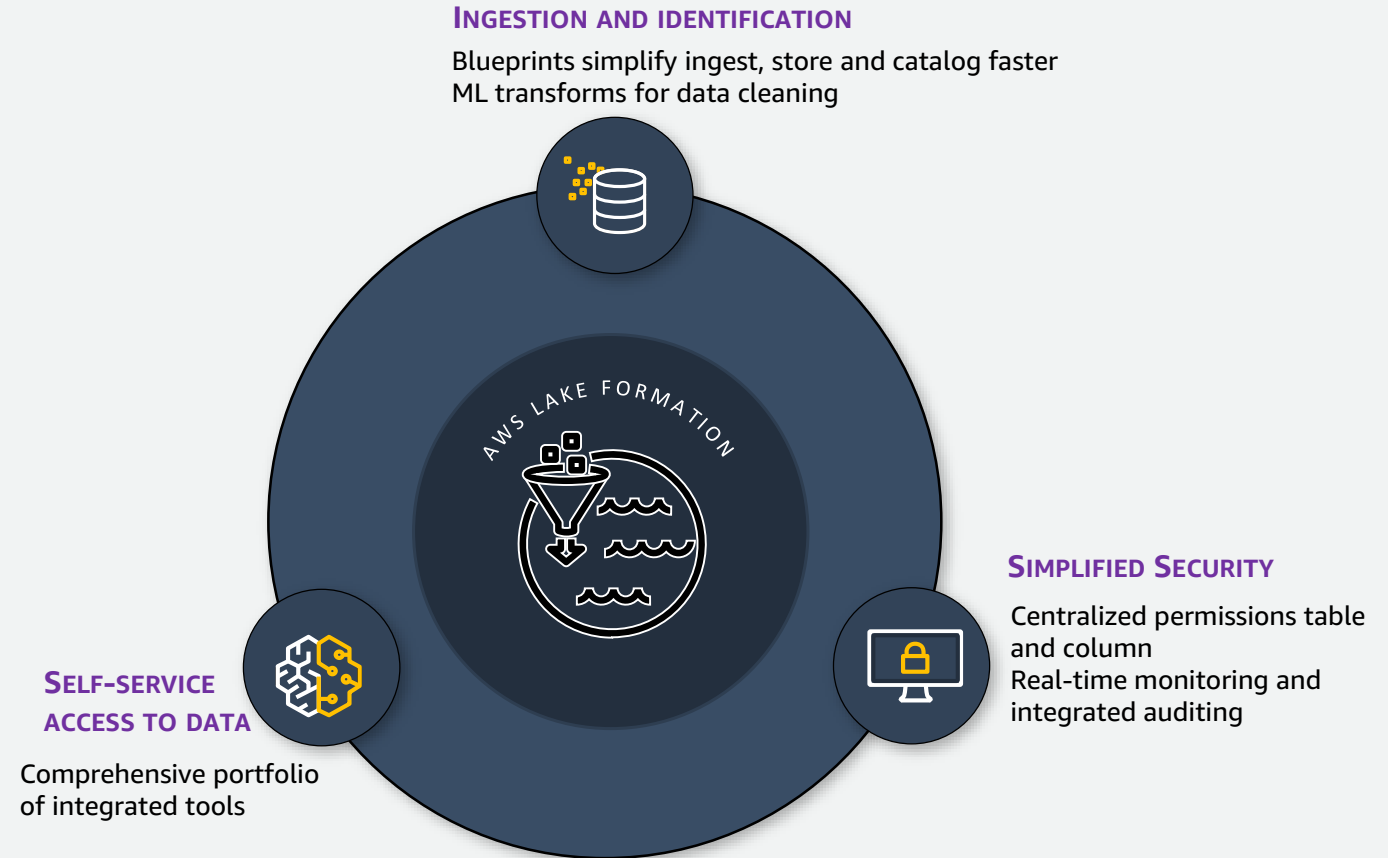


Complex Self-Service Analytics, Data Sharing, External Integrations

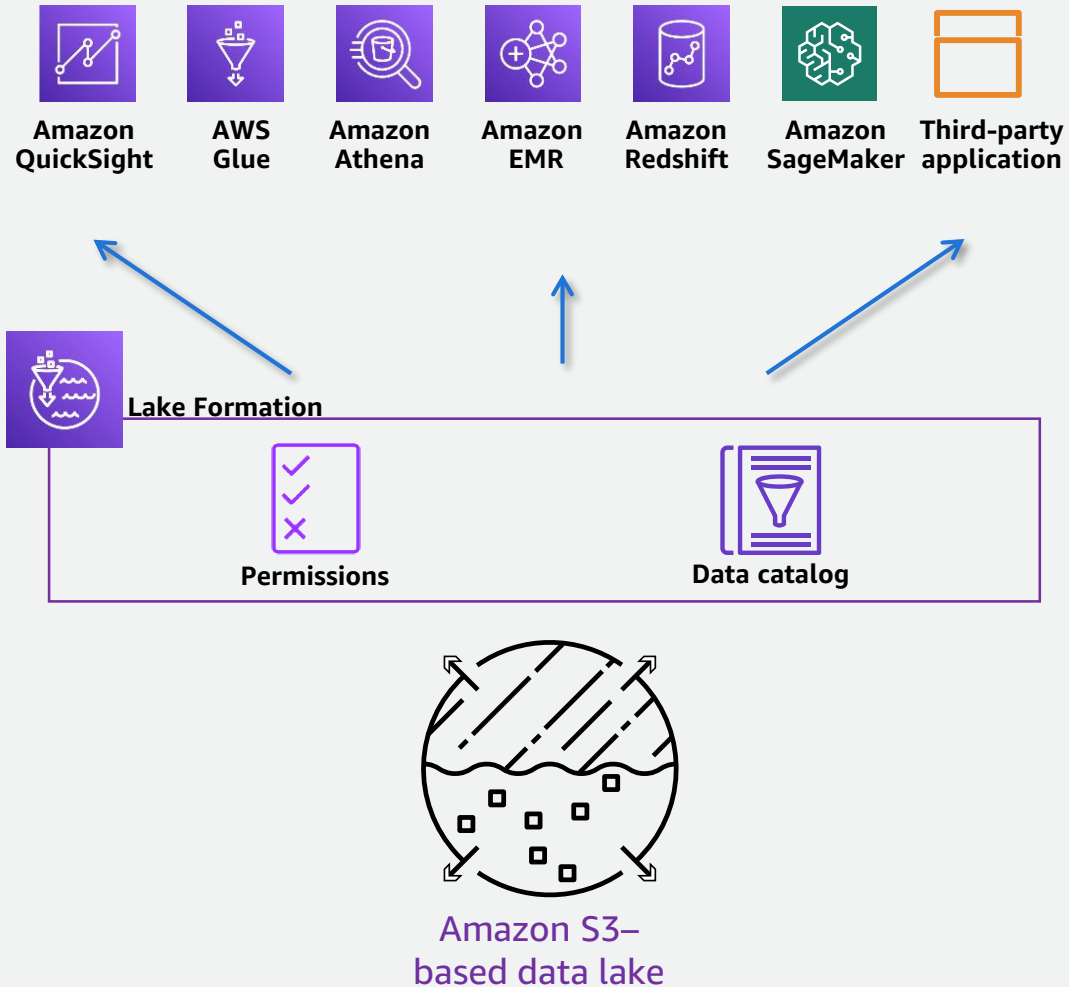
Centralize Data Lake Security Management with AWS



AWS Lake Formation is a fully managed serverless service that allows you to build and secure data lakes in days



Lake Formation secures your data lake



DB-style fine-grained permissions on resources

Scale permissions management Lake Formation
Tag-Based Access Control (LF-TBAC)

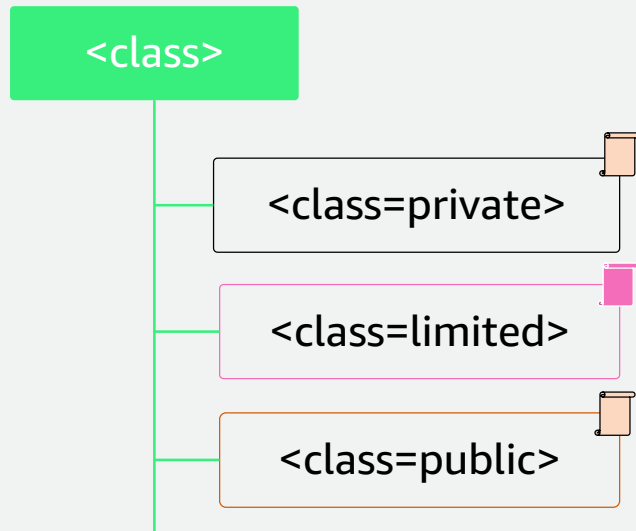
Unified Amazon S3 permissions

Integrated with services and tools

Easy to audit permissions and access

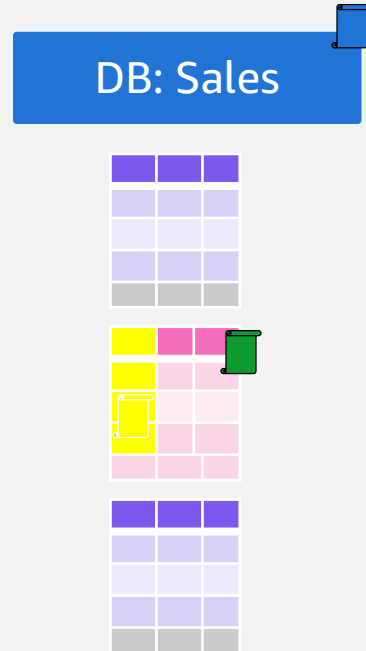
Scale permissions management with LF-TBAC

Define LF-Tags



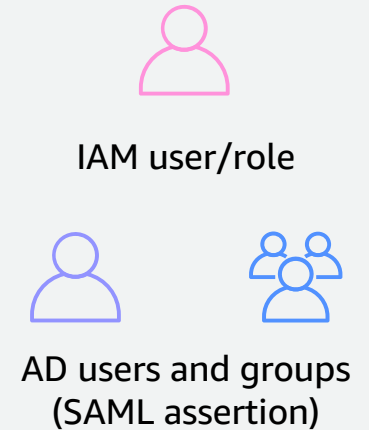
- Specify who can assign LF-Tags and values

Assign LF-Tags to resources



- Tag databases, tables, columns
- LF-Tags are hierarchical and may be overridden

Create policies on LF-Tags



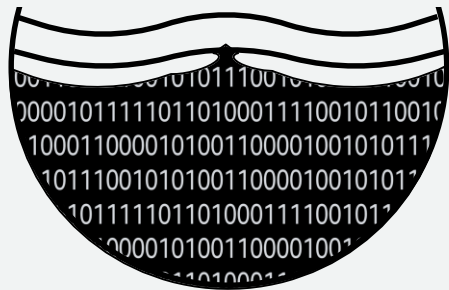
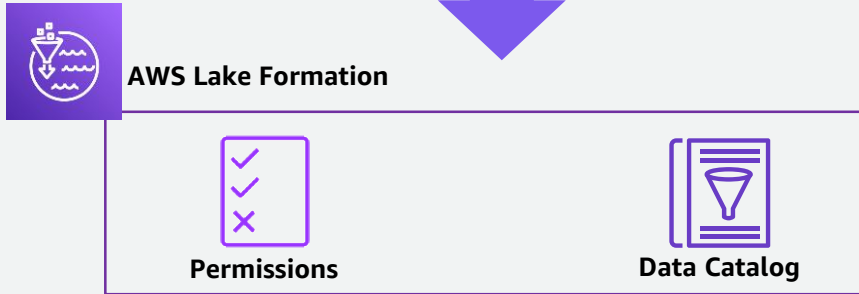
- Scale by applying permission on LF-Tags

AWS Lake Formation centralized authorization

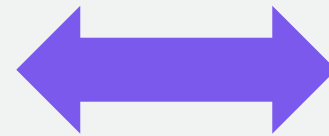


DATA SECURITY OFFICER

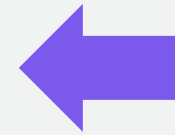
1. Administrator sets up user permissions on lake resources: databases, tables, columns, rows, or cells.



Amazon S3 data lake storage



3. Lake Formation **authorizes** access to catalog resources & Amazon S3 objects



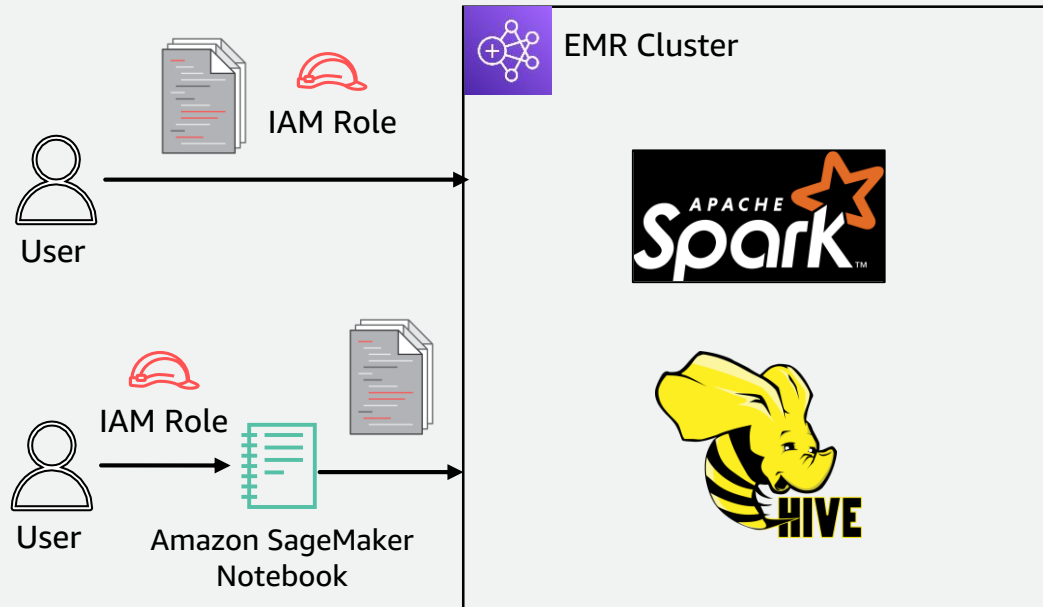
DATA ANALYST

2. User access data via integrated services

New Integrations For Centralized Data Lake Security Management

Expanded support for Lake Formation permissions – Amazon EMR Runtime Roles

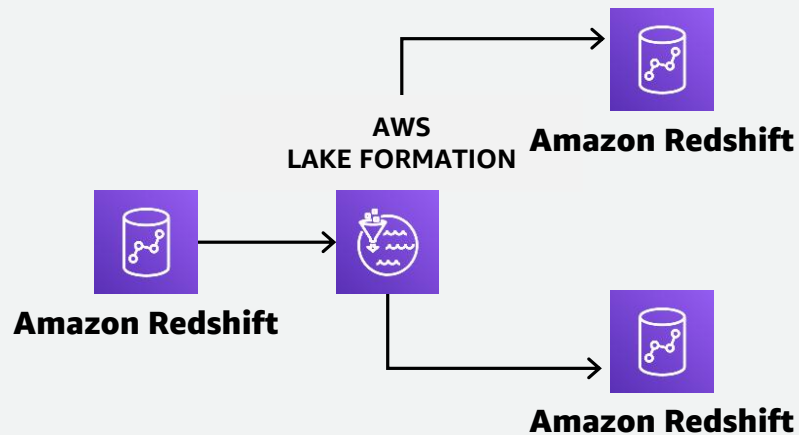
EMR Runtime Roles



- Jobs submitted using a Runtime Role enabled cluster will use Lake Formation for enforcement
- Jobs can be submitted through EMR Steps, Amazon SageMaker Studio Notebooks, or directly to the cluster using JDBC for Hive or Livy for Spark

Data sharing access control with AWS Lake Formation

Centrally manage data sharing with AWS Lake Formation



Centrally manage granular access to data across all consuming data services

Improve security and governance with row level and column level granular permissions on data sharing

No manual scripting or complex querying

Define policies once and enforce those consistently for multiple consumers

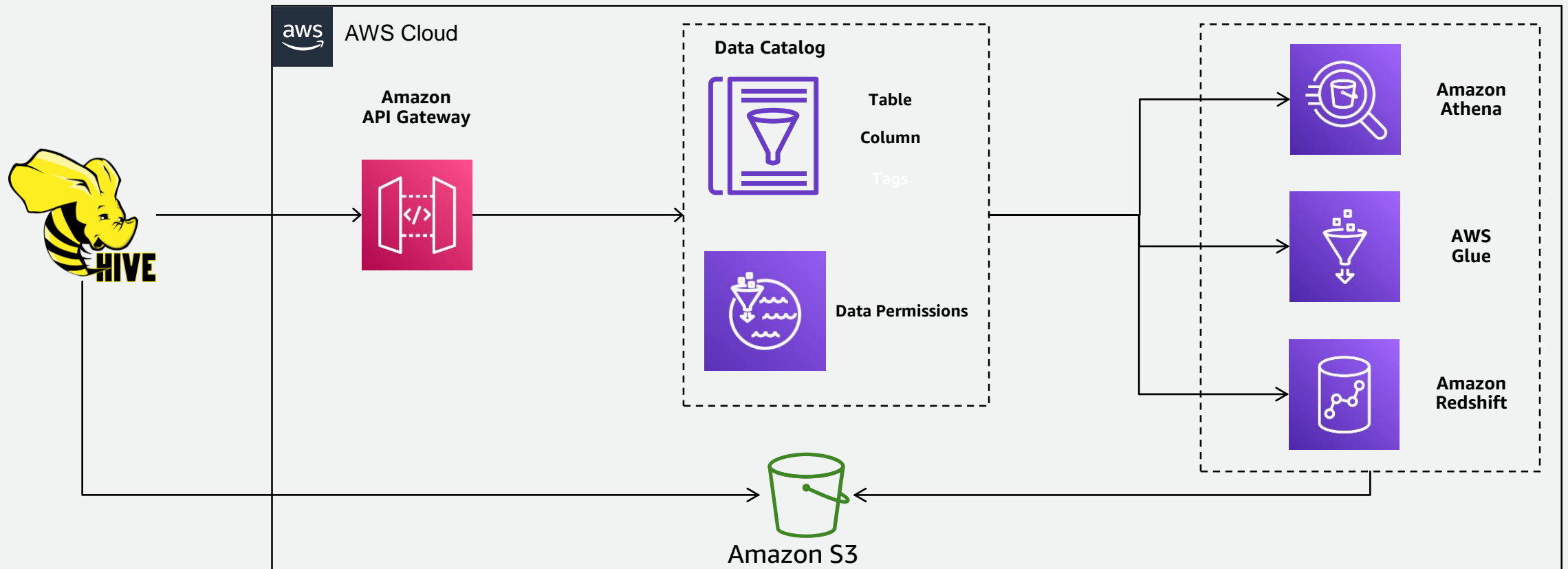
Hive Metastore Federation

ACCESS HIVE METASTORE DATA WITH ATHENA, REDSHIFT AND ETL

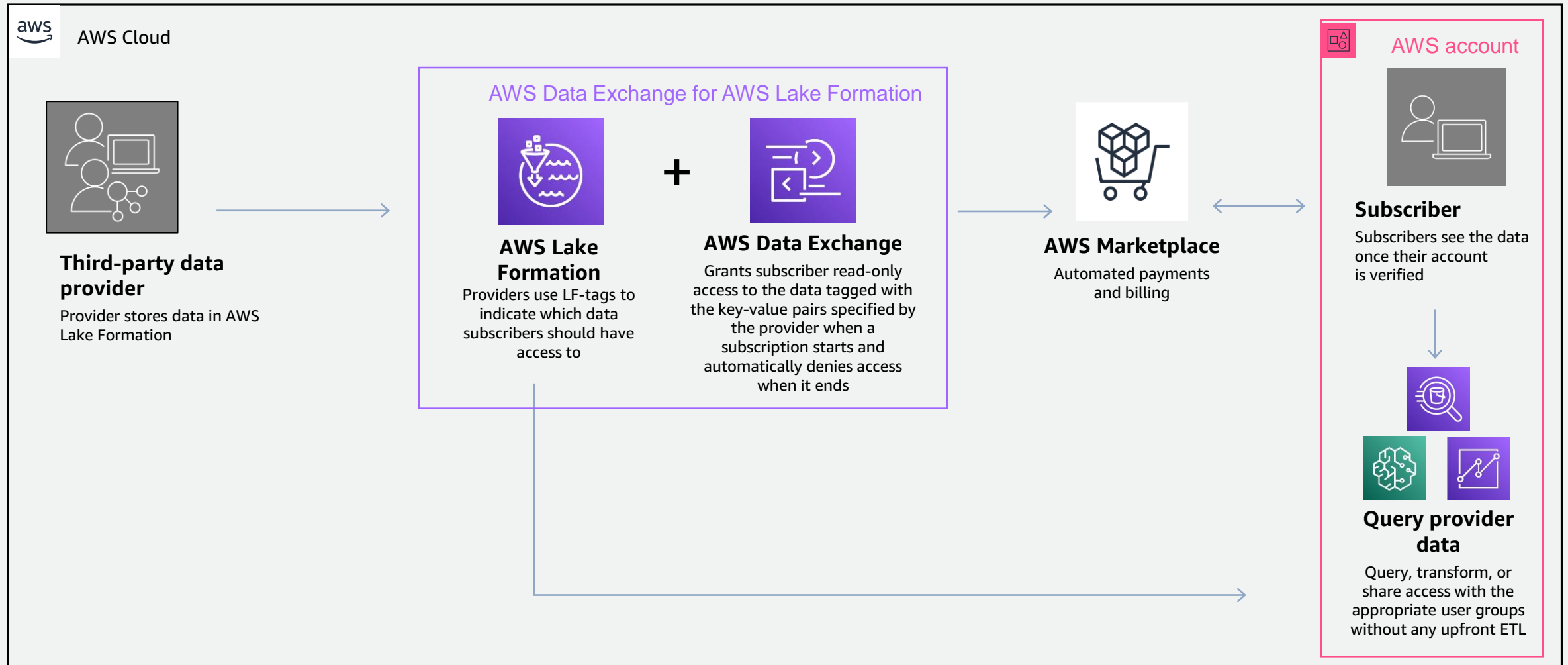
1. Connect Hive Metastore to Lake Formation

2. Provide Access to Users

3. Use with Athena, Redshift, Glue and other services



AWS Data Exchange for AWS Lake Formation



How can you get started?

Call to action: Next steps

THINK BIG

Discovery Workshop

Data-Driven Everything

START SMALL

Data Labs

AWS ProServe POC

SCALE FAST

AWS ProServe

Partners



Thank you!

Ramkumar Nottath

rnottath@amazon.com

www.linkedin.com/in/ramnottath