

# AWS LakeFormation Build A Secure Data Lake



# Speaker



**Sanchit Jain**

Lead Architect - AWS at Quantiphi  
AWS APN Ambassador

**FOLLOW ME**



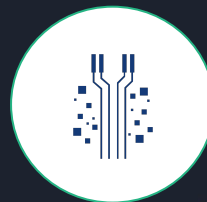
# Agenda



Challenges in  
Traditional Data Lake



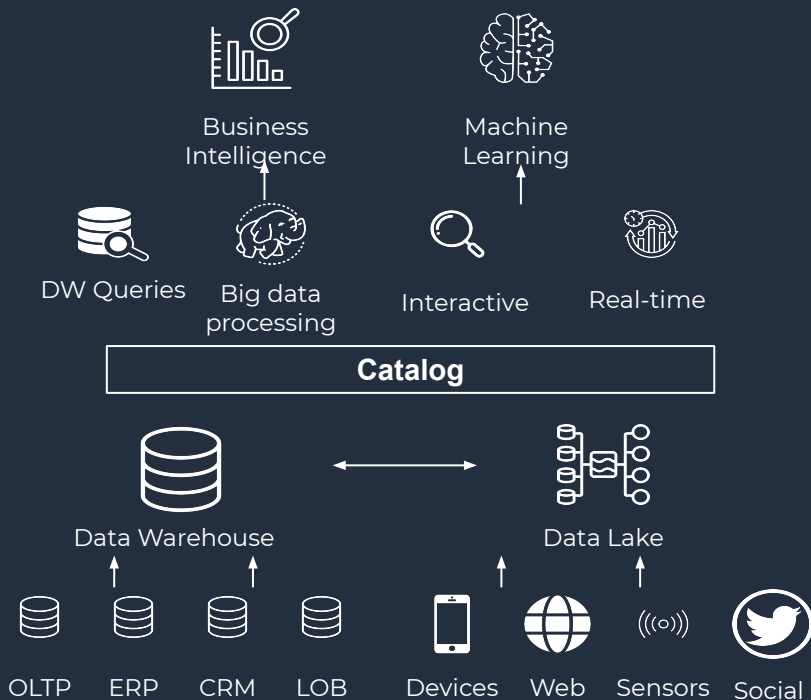
Introduction to  
Lake formation



Features of  
Lakeformation

# What is a data lake?

A data lake is a **centralized repository** that allows you to store all your **unstructured and structured data** at any scale



Data lakes provide:

- Relational and non relational data
- Scale-out to TBs
- Diverse set of analytics and machine learning tools
- Work on data without any data movement
- Designed for low cost storage and analytics

# Data Lake Challenges

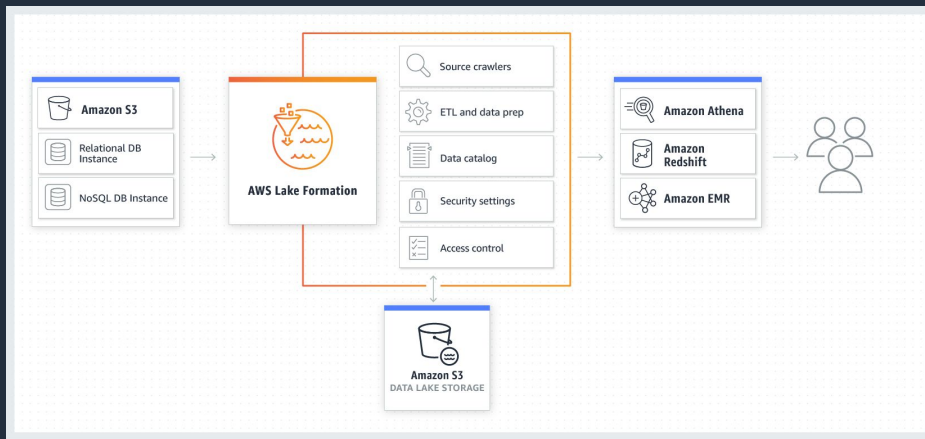
- Maintaining a data catalog/enabling self-service access
- Configuring and managing access controls/Data governance
- Audit logging

- Building data lakes can still take months

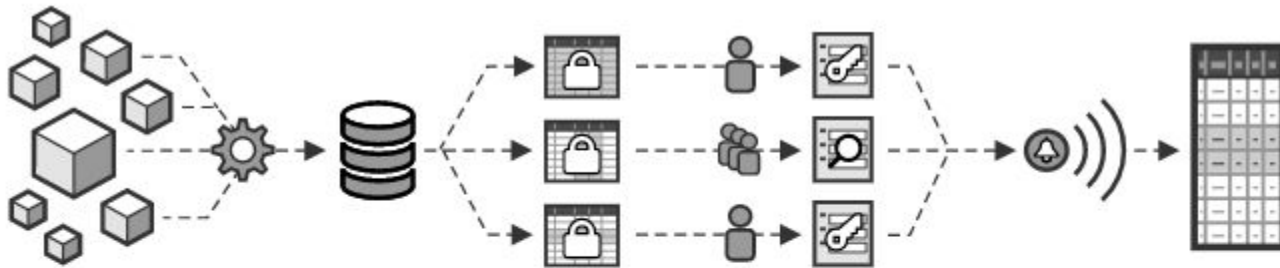


# Introduction to Lake Formation

- Builds and secures data in data lakes easily
- Integrated with AWS Glue
  - Facilitates setting up data catalogs
  - Builds workflows for loading data to S3
- Acts like gatekeeper to all the data in the data lake
- Integrated with Athena, Quicksight and Redshift Spectrum



# How it works



## Ingest & Organize

Automatically ingest, clean, encrypt, and register existing Amazon S3 bucket content, including log data from CloudTrail, CloudFront, and Amazon ELB.

## Secure & Control

Define access control that provides the right data to the right users, groups, and roles. Flexible database, table, and column permissions enable granular security.

## Collaborate & Use

Search and discover using catalog metadata. All access is checked against policy, so your data is protected even if tools change or new data arrives.

## Monitor & Audit

Be alerted of access requests and policy exceptions. Review activity history with detailed change logs and data lineage.

# Key components

- **Blueprints/ Data importers** - templates for ETL, metadata (schema) and partition management
- **Enhanced data catalog** - enable users to record more metadata and tag data catalog objects (i.i. Databases, tables, columns)
- **ML Transformations** - ML algorithms that customers can use to create their own ML transforms (i.e. record deduplication)
- **Enhanced security and governance** - security and governance layer at the data catalog level



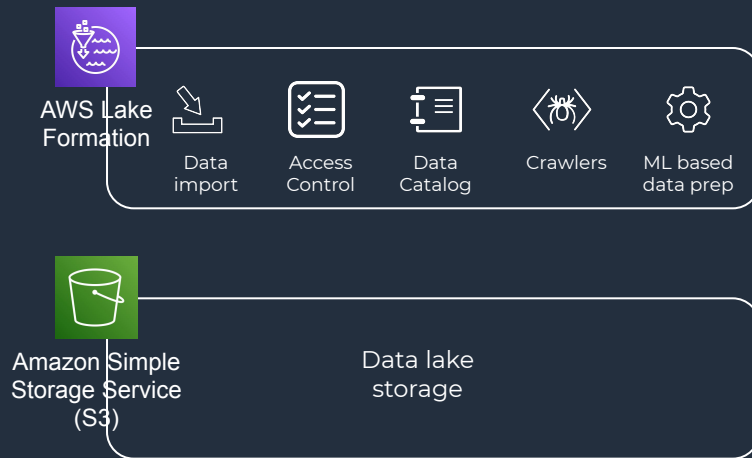
# Register data

Amazon S3 forms the storage layer for Lake Formation

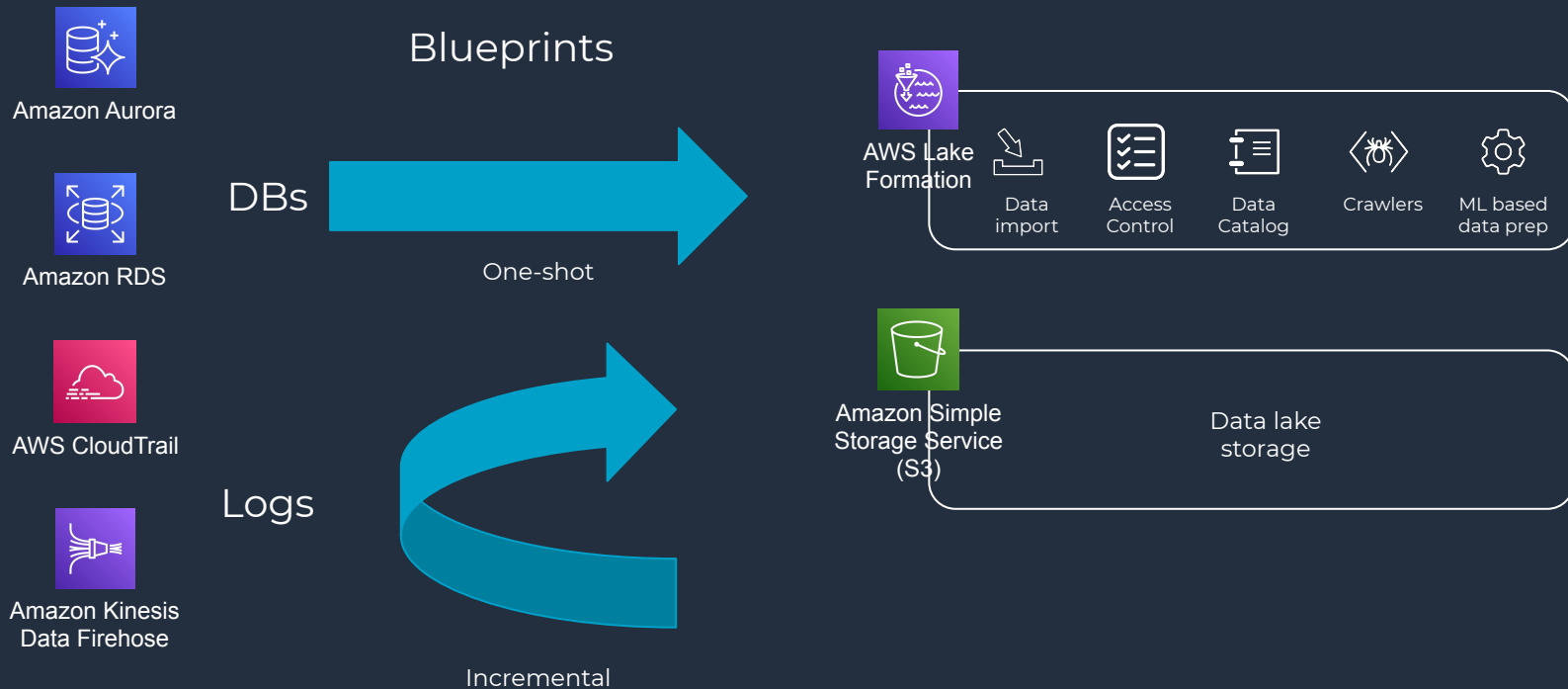
Register existing S3 buckets that contain your data

Ask lake formation to create required S3 buckets and import data into them

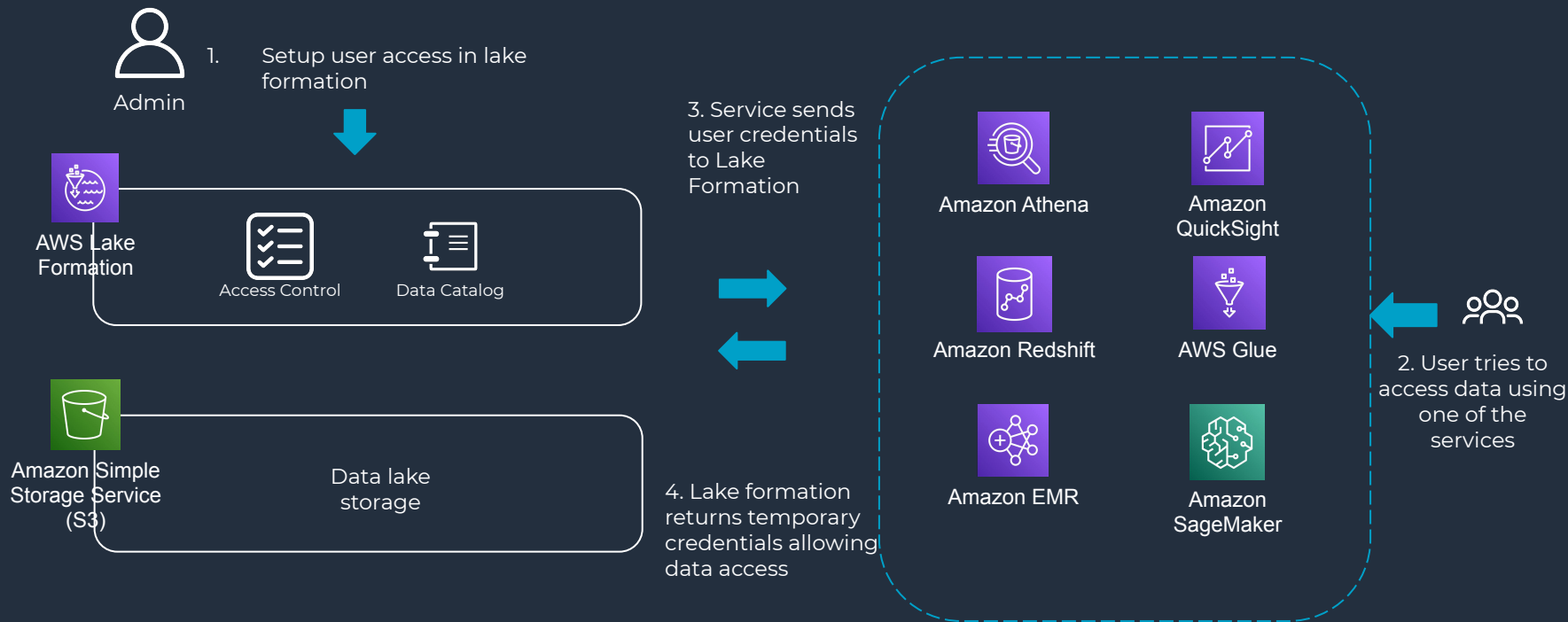
Data is stored in your account. You have **direct access** to it. No lock-in.



# Easy Ingestion



# Secure once, access in multiple ways



# AWS Lake Formation Pricing

**No Additional Charges** - Only pay for the underlying services used.

# THANK YOU

