AWS re:Invent

WPS305

Building your geospatial data lake

Ryan Chung
Solutions Architect Mgr.
Amazon Web Services

Catherine Hall
Solutions Architect
Amazon Web Services

Luke Wells
Solutions Architect
Amazon Web Services

AWS re:Invent

aws

There is **more data** than people think

Data	Data platforms need to	
grows	live for	scale
>10x	**15**	**1,000x**
every 5 years	years	

Source: IDC, Data Age 20215: The Evolution of Data to Life-Critical Don't Focus on Big Data, Focus on the Data That's Big, April 2017.

Customers want more value from their data

Growing exponentially

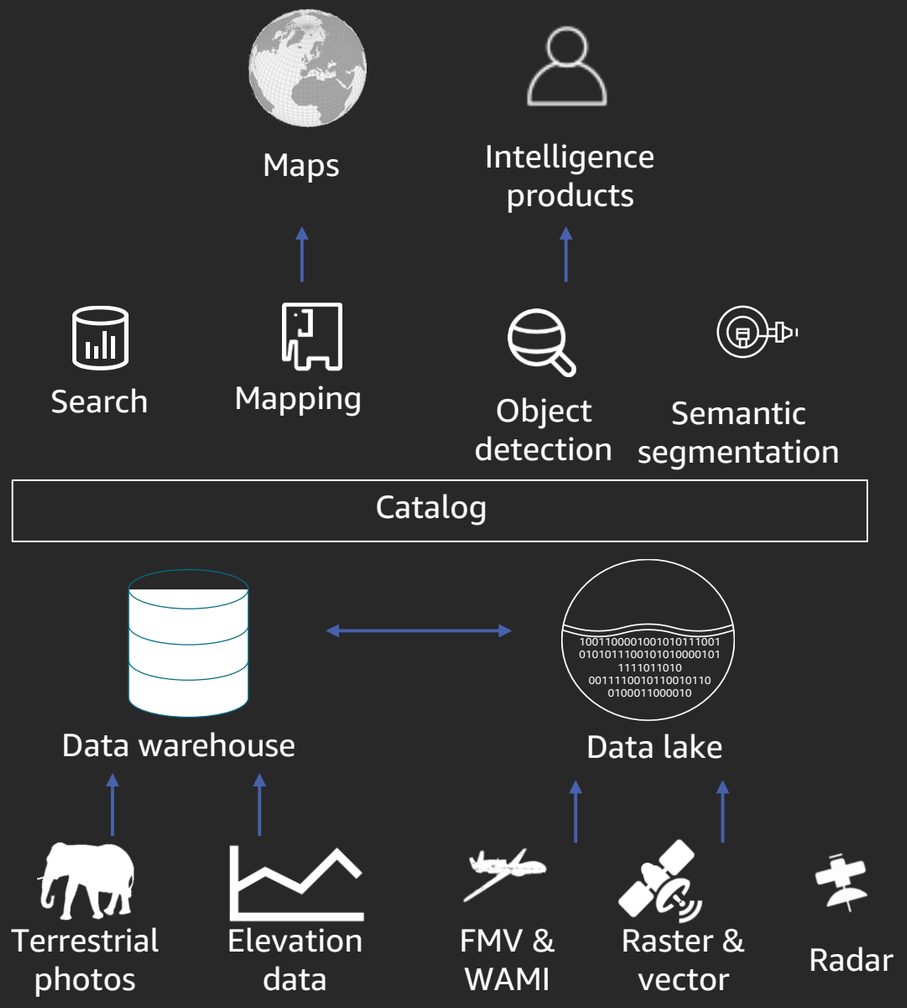
From new sources

Increasingly diverse

Used by many people

Analyzed by many applications

Why geospatial data lakes?



Data lakes provide:

Relational and non-relational data

Scale-out to exabytes

Diverse set of analytics and machine learning tools

Work on data without any data movement

Designed for low-cost storage and analytics

Characteristics of a data lake

Collect anything

Dive in anywhere

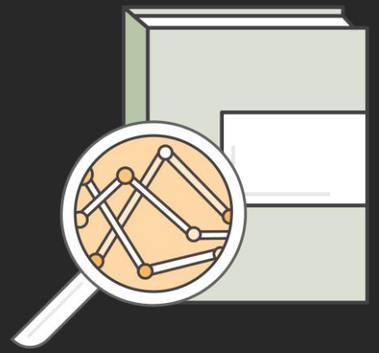
Flexible access

Future proof

Important components of a data lake

Access &
user interface

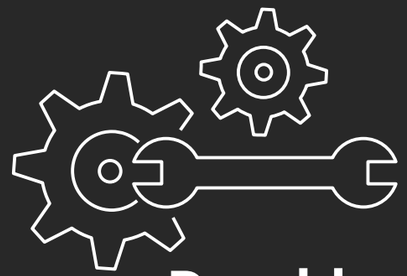
Ingest & store



Catalog
& search

Protect
& secure

Why Amazon S3 for data lakes?

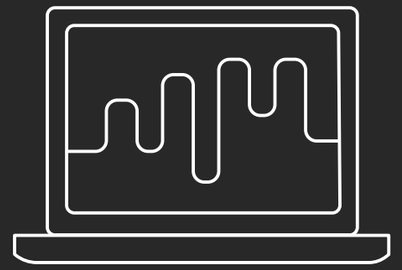


Durable

Designed for 11 9s of durability

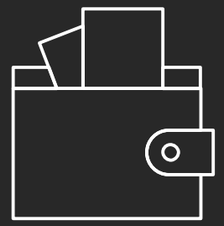
Available

Designed for **99.99**% availability



High performance

- Multiple upload
- Range GET



Easy to use

- Simple REST API
- AWS SDKs
- Read-after-create consistency
- Event notification
- Lifecycle policies

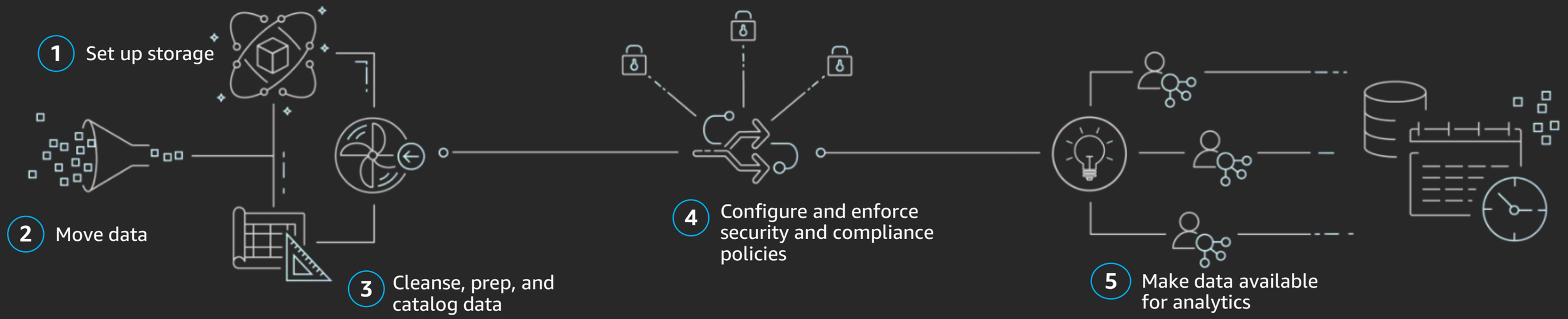
Scalable

- Store as much as you need
- Scale storage and compute independently
- No minimum usage commitments

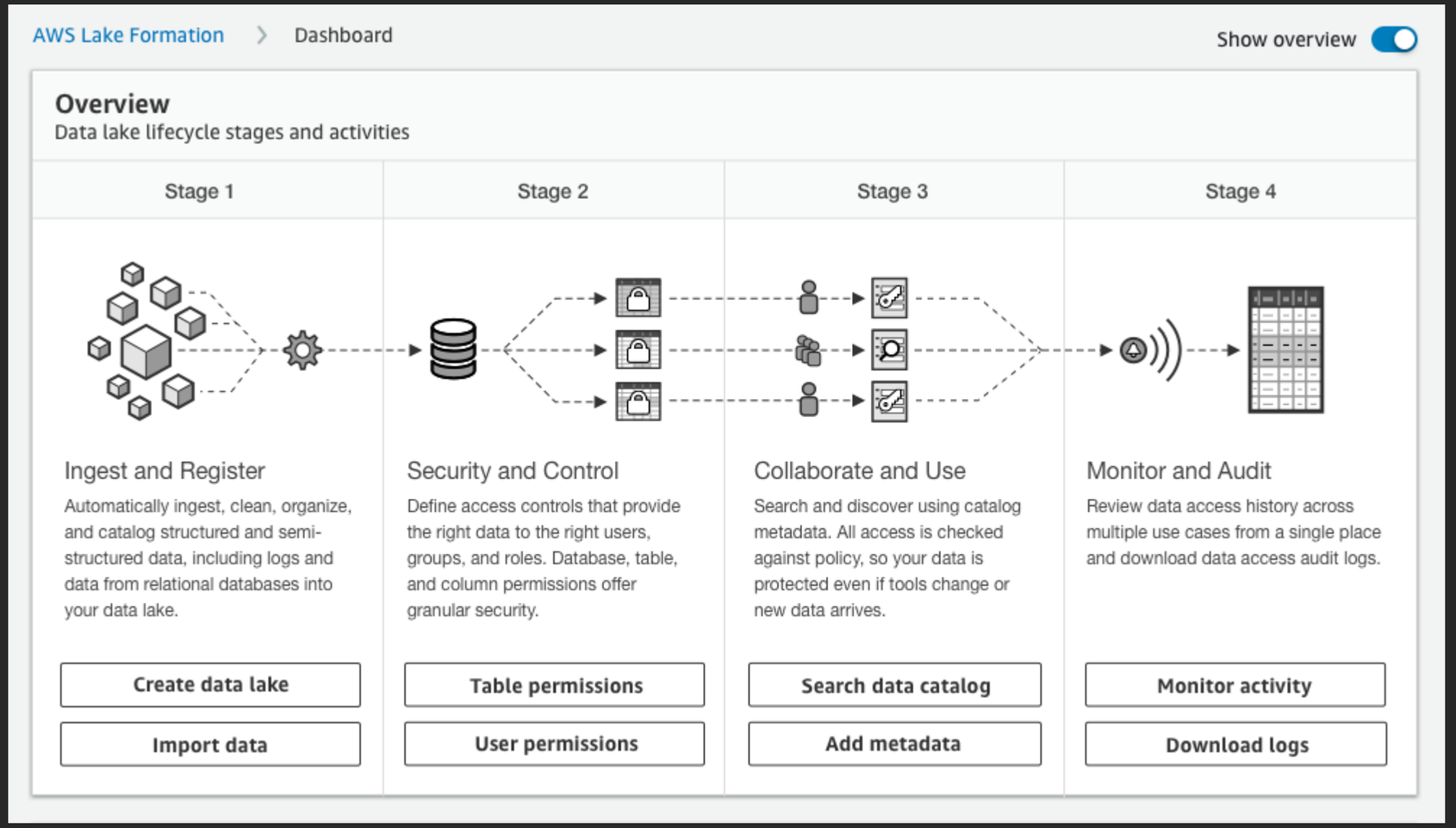
Integrated

- Amazon Redshift/Spectrum
- Amazon EMR
- Amazon Athena
- Amazon DynamoDB

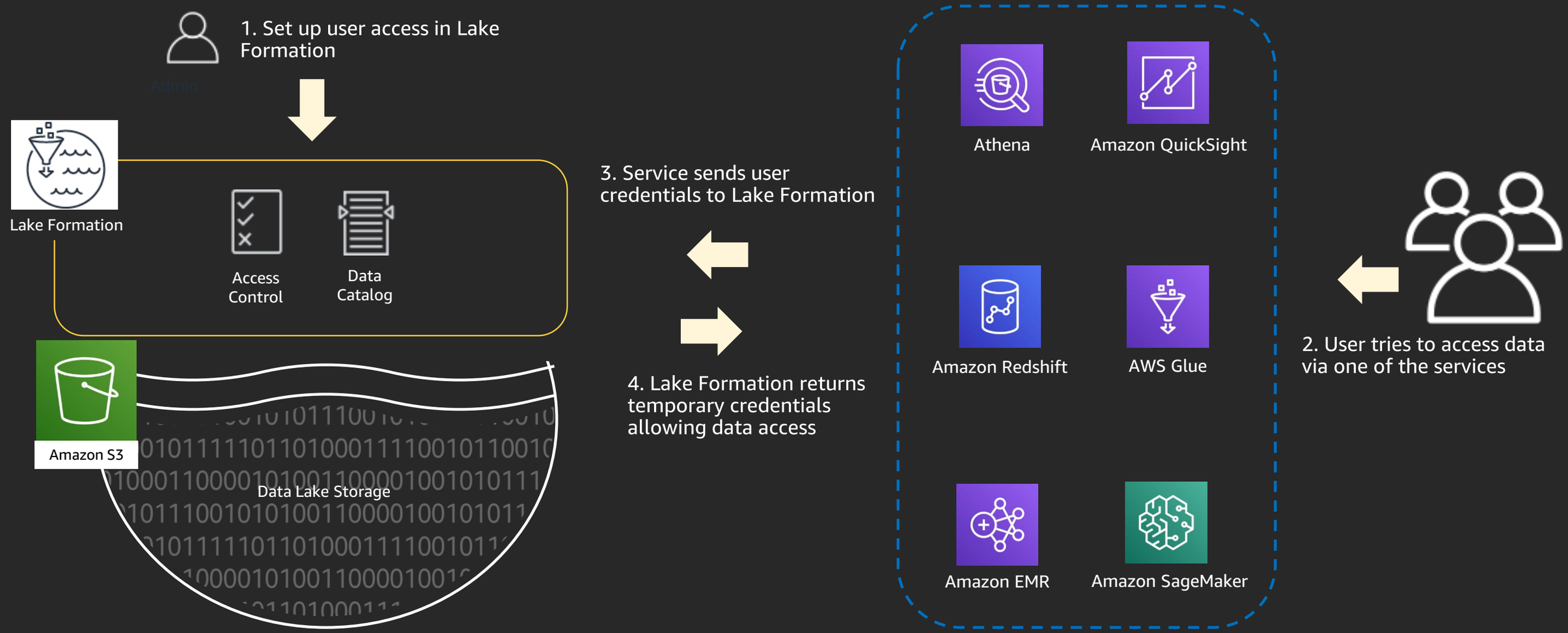
Typical steps of building a data lake



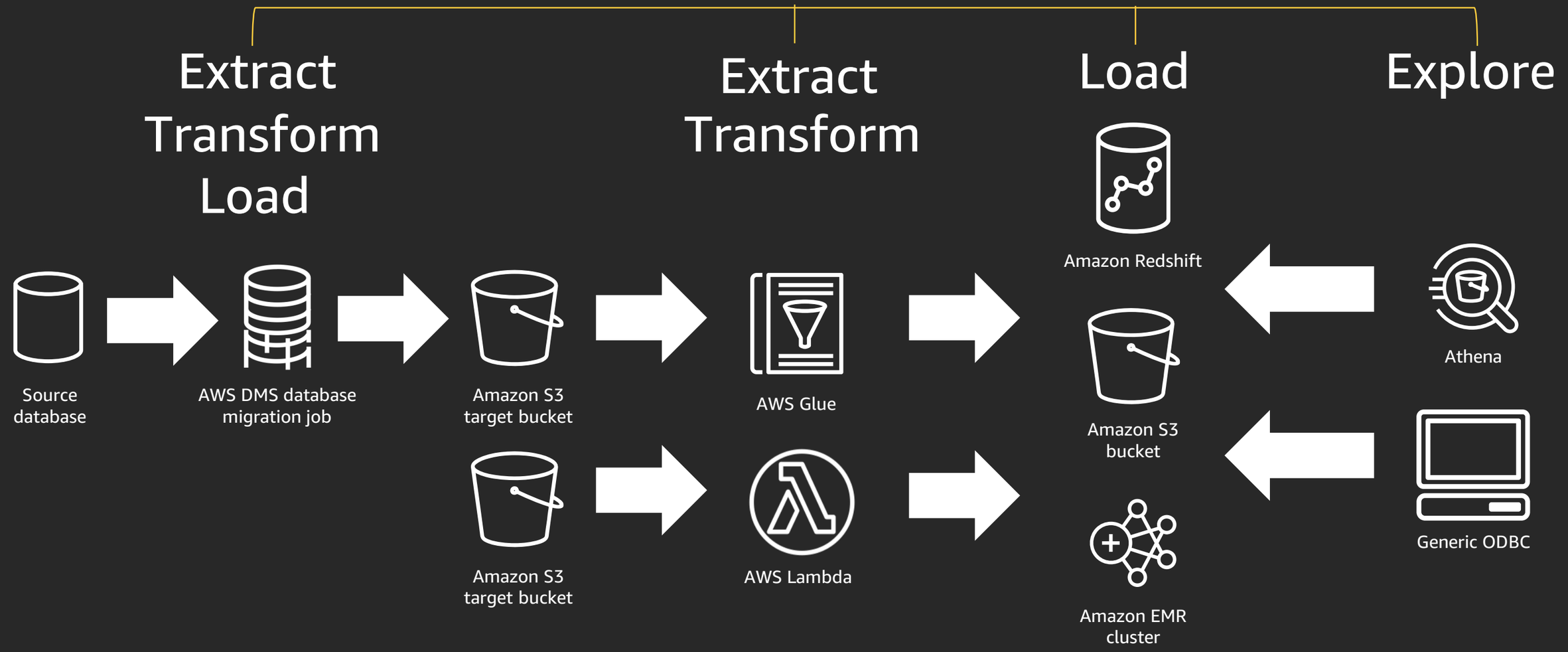
How AWS Lake Formation works



Secure once, access in multiple ways



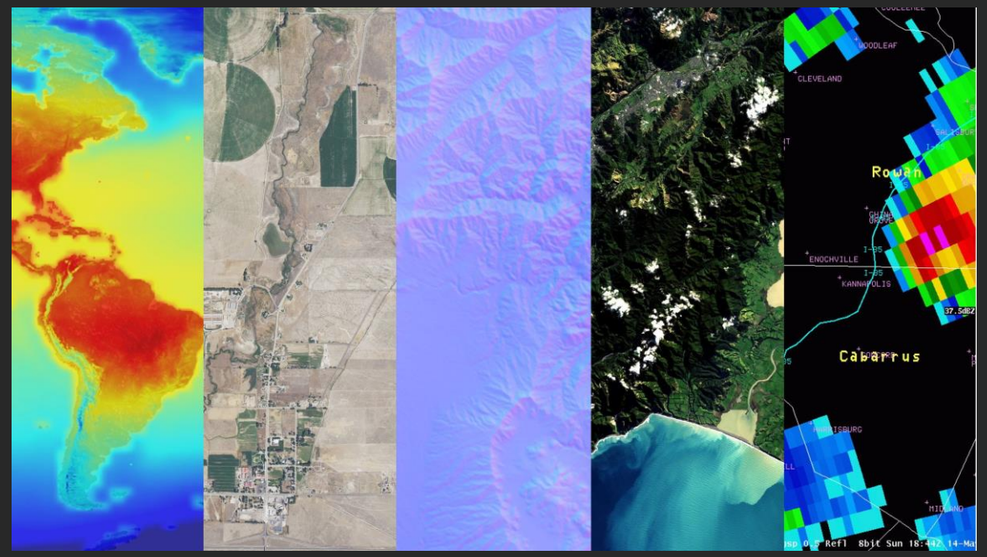
Make your geospatial data discoverable



Getting started

AWS Public Datasets

https://registry.opendata.aws



Earth on AWS

https://aws.amazon.com/earth

AWS Free Tier

https://aws.amazon.com/free/

Questions?

Related sessions

AIM366-R – SpaceNet: ML to solve mapping challenges

WPS307 – Building a serverless GIS/geo-positioning and alerting solution

WPS319-R – Best practices for working with large-scale geospatial data

STG401 – Manage objects and optimize cost with Amazon S3 and Amazon S3 Glacier

Thank you!

Ryan Chung
chungr@amazon.com

Catherine Hall
hallcat@amazon.com

Luke Wells
lawells@amazon.com

AWS re:Invent

aws

Please complete the session survey in the mobile app.