



AWS re:Invent

GPSTEC322

Modernizing your data platform using AWS

Suresh Veeragoni

Sr. Partner Solutions Architect
Amazon Web Services

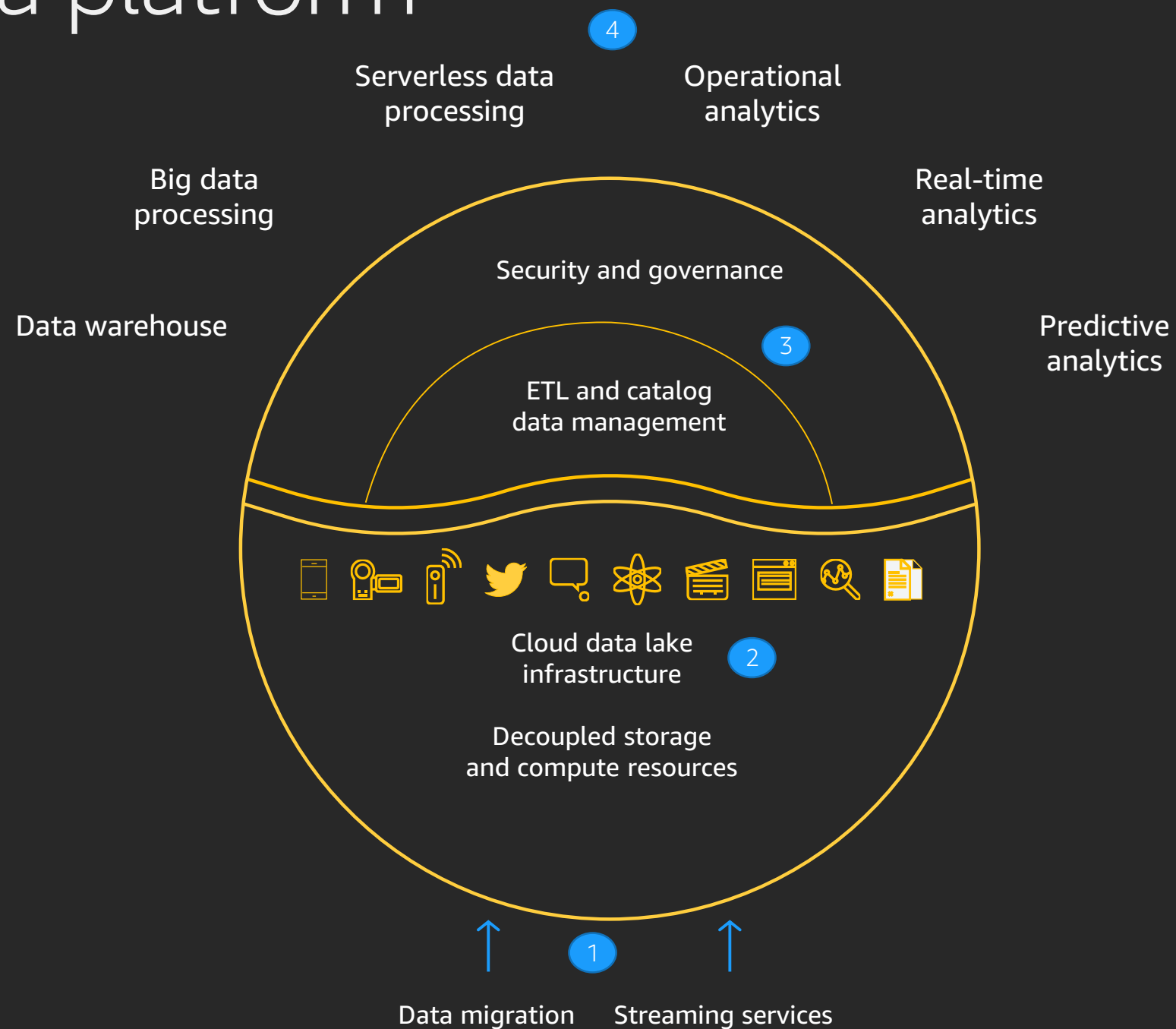
Sanjay Doiphode

Sr. Partner Solutions Architect
Amazon Web Services

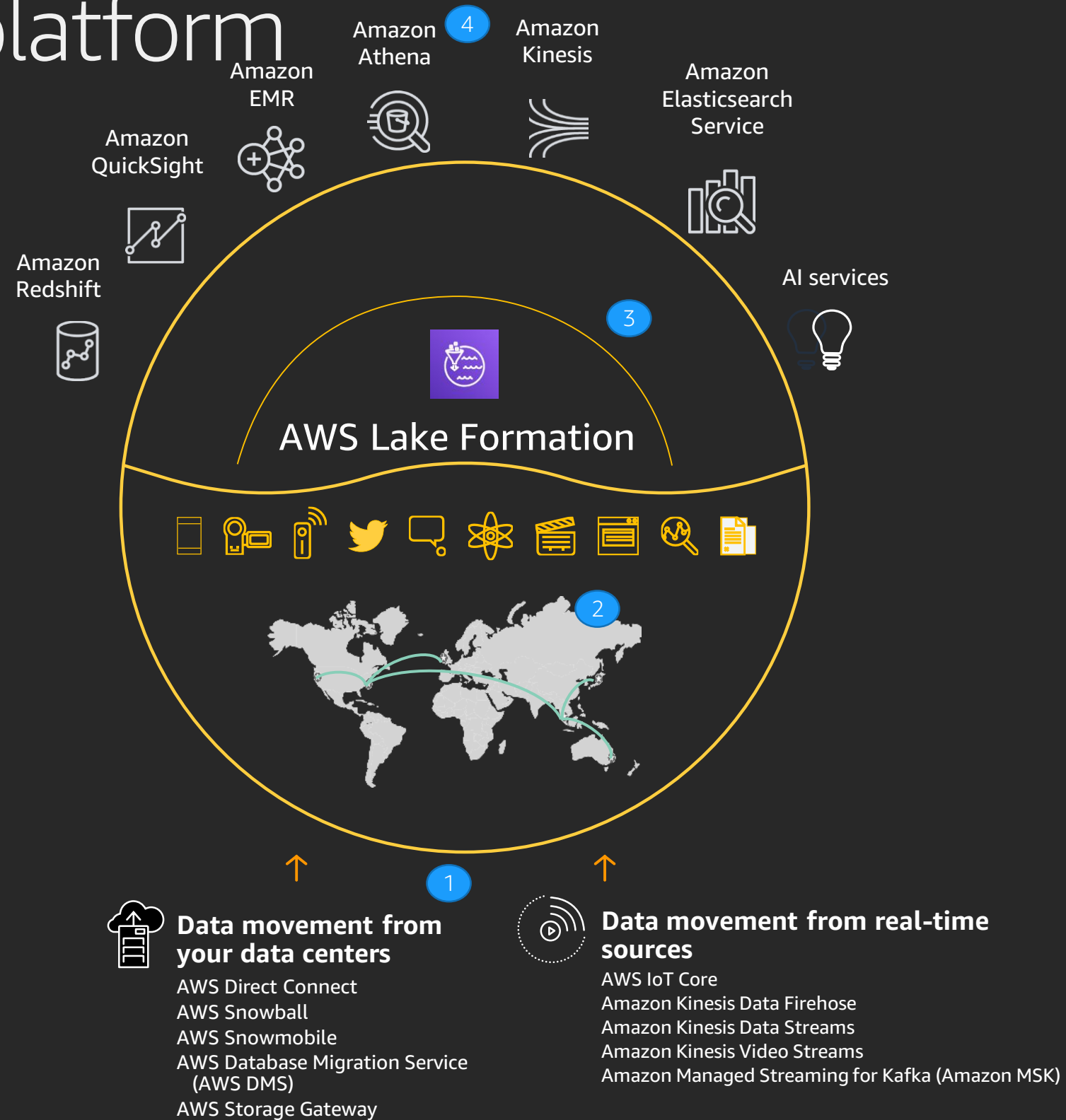
Challenges

- Ingesting semi-structured and unstructured data
- Ingesting and analyze real-time data
- Managing cost
- Decoupling storage and compute
- Scaling the data platform
- Managing infrastructure
- Data discovery
- Centralized security
- Data silos

Modern data platform



Modern data platform



AWS databases and analytics

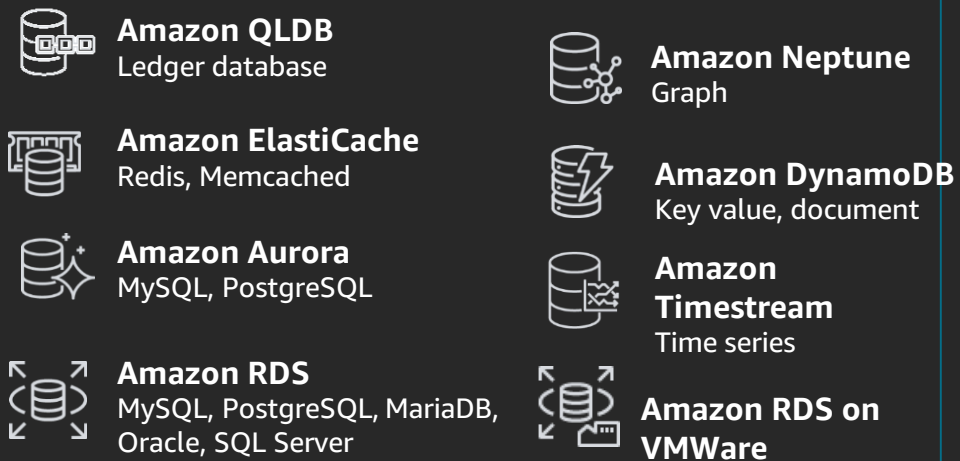
Broad and deep portfolio, built for builders

Business intelligence and machine learning

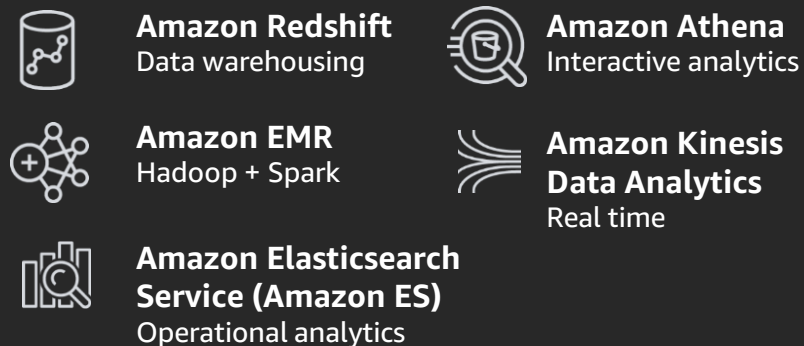


AWS Marketplace
250+ solutions

Databases



Analytics



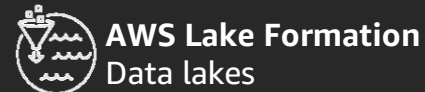
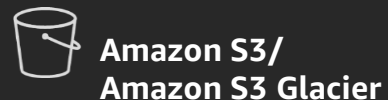
Blockchain



730+ database solutions

600+ analytics solutions

25+ blockchain solutions



Data lake



20+ data lake solutions

Data movement

AWS DMS | AWS Snowball | AWS Snowmobile | Amazon Kinesis Data Firehose | Amazon Kinesis Data Streams | AWS Data Pipeline | AWS Direct Connect

30+ solutions

AWS databases and analytics

Broad and deep portfolio, built for builders

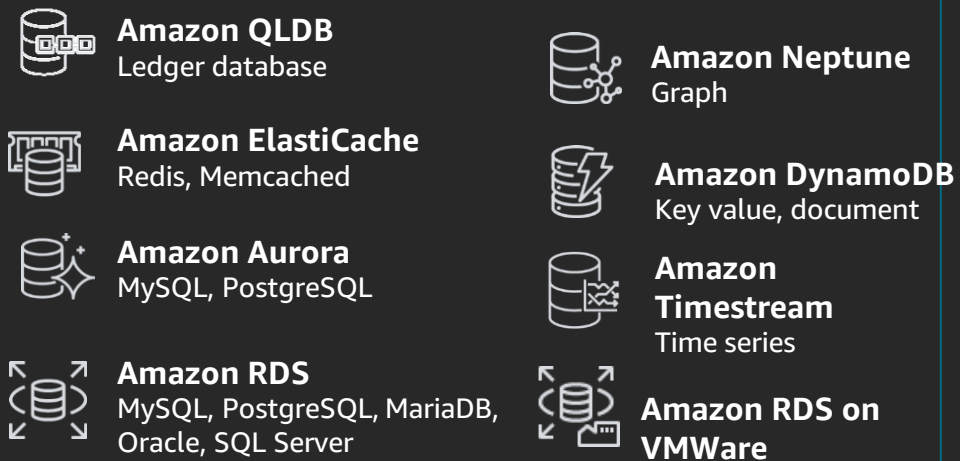
Business intelligence and machine learning



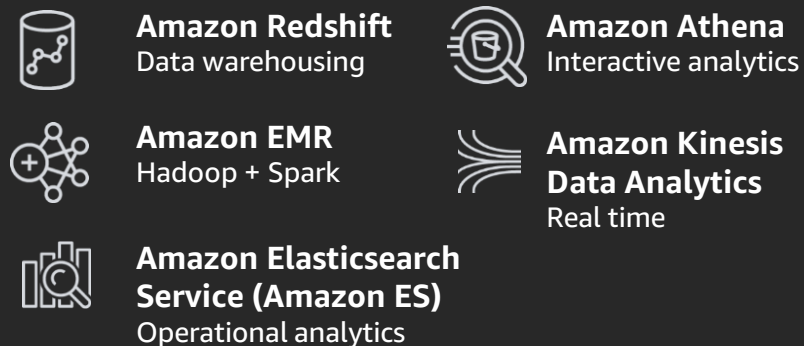
AWS Marketplace

250+ solutions

Databases



Analytics



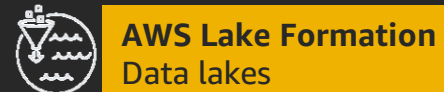
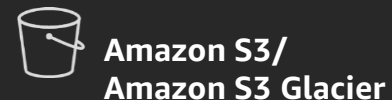
Blockchain



730+ database solutions

600+ analytics solutions

25+ blockchain solutions



Data lake



20+ data lake solutions

Data movement

AWS DMS | AWS Snowball | AWS Snowmobile | Amazon Kinesis Data Firehose | Amazon Kinesis Data Streams | AWS Data Pipeline | AWS Direct Connect

30+ solutions

Many partners to complement AWS offerings

accenture

TRIFACTA

CSC

talend

Qlik

47 Lining
A Hitachi Vantara Company

NorthBay

DXC.technology

mongoDB

DATAGUISE

Booz | Allen | Hamilton
strategy and technology consultants

Cloudwick

PARIVEDA
SOLUTIONS

DATASTAX

MATILLION

Paxata

New Relic

slalom

Capgemini
CONSULTING. TECHNOLOGY. OUTSOURCING

AQUILENT

TERADATA

Informatica

DOMO

DATA
REPUBLIC

CISCO

sumologic

Adobe

clearscale

credibiliT

CloudMas

ScienceLogic

druva

SAP

splunk

ZALONI
THE DATA LAKE COMPANY

redhat

CORPINFO

matrix
International Financial Services

MEGAZONE
Cloud Innovator

SUSE

blazeclan
Cloud IT Better

CITRIX

intel

looker

SAS

Cognizant

Dedalus

NetApp

REAN
CLOUD

Reply

tableau
SOFTWARE

MicroStrategy

ATTUNITY

Qubo

CLOUDNEXA

emind
YOUR CLOUD EXPERTS

BULLETPROOF

Data modernization tooling

Data lifecycle



Discover

AWS Glue
(crawlers)



Ingest

AWS Glue
Kinesis Data
Streams and
Kinesis Data
Firehose
AWS DMS
AWS Schema
Conversion Tool
(AWS SCT)
AWS Snow family
DX
Amazon MSK
AWS Storage
Gateway
AWS DataSync



Store

Amazon S3
Amazon S3
Glacier
Amazon RDS
Aurora
DynamoDB



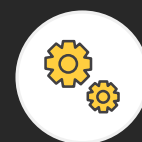
Secure

AWS Identity and
Access
Management
(IAM)
AWS Key
Management
Service (AWS
KMS)
Amazon Macie
AWS CloudTrail
Amazon
CloudWatch



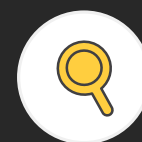
Catalog

AWS Glue (Data
Catalog)
Amazon EMR
(Hive metastore)



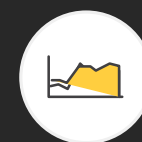
Prepare

AWS Glue ETL
(Serverless
Apache Spark)
Amazon EMR
(Apache Spark
and Hadoop)



Analyze

Amazon
SageMaker
Amazon Redshift
Athena
Amazon Kinesis
Data Analytics
Amazon EMR
(Apache Spark &
Hadoop)
Amazon ES
AI services



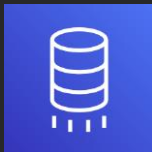
Visualize

Amazon
QuickSight
Amazon EMR
notebooks

Modernize your database and data warehouse

Modernize your database tier using AWS DMS

- Commercial to open source
- Commercial to Amazon Aurora

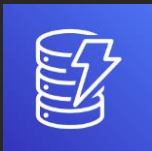


AWS DMS

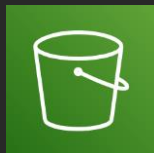
ORACLE



Aurora



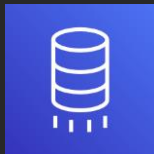
DynamoDB



Amazon S3

Modernize your warehouse using AWS SCT

- Commercial to Amazon Redshift



AWS SCT

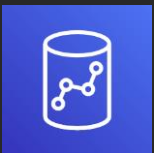
Vertica



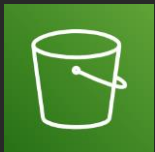
Netezza

Greenplum

ORACLE



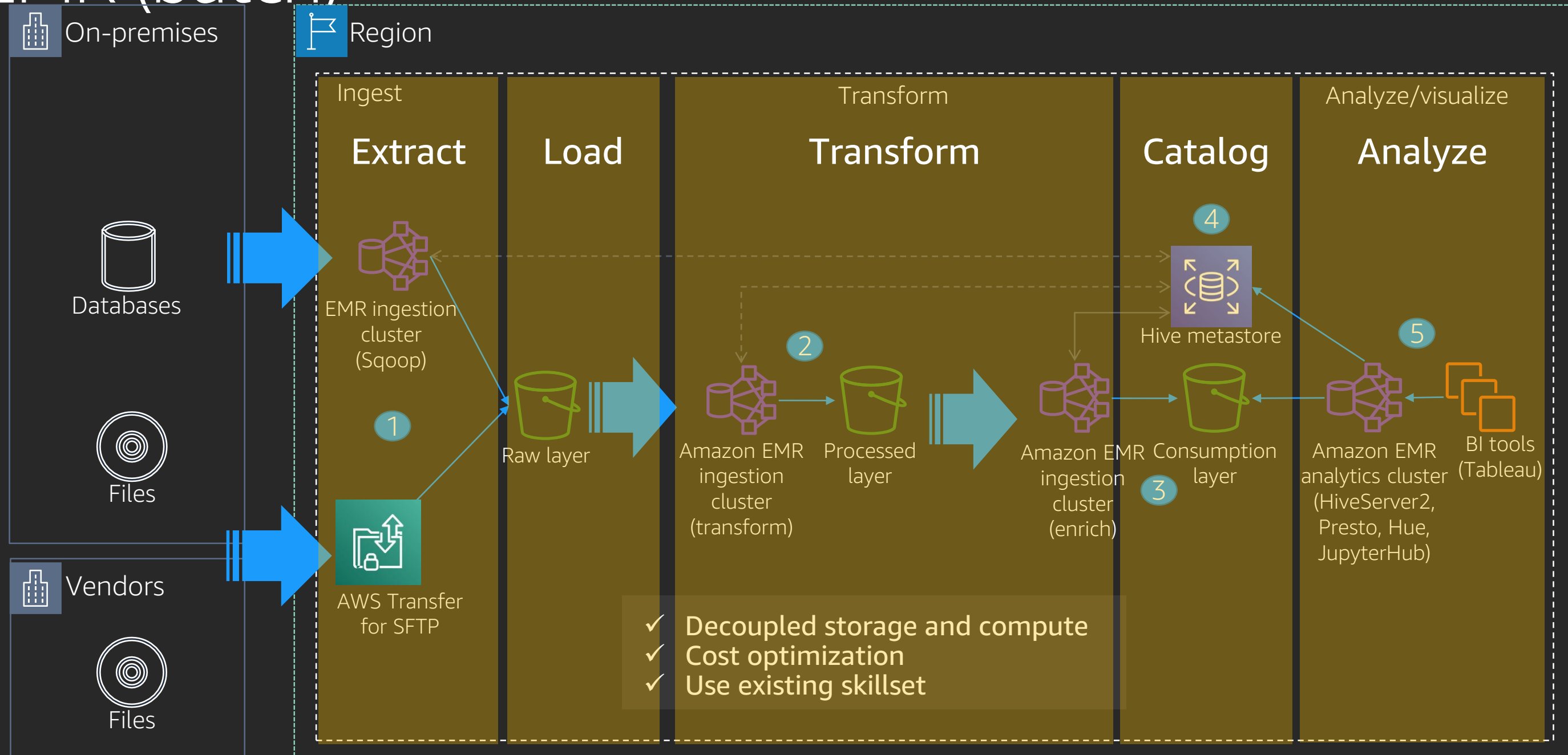
Amazon Redshift



Amazon S3

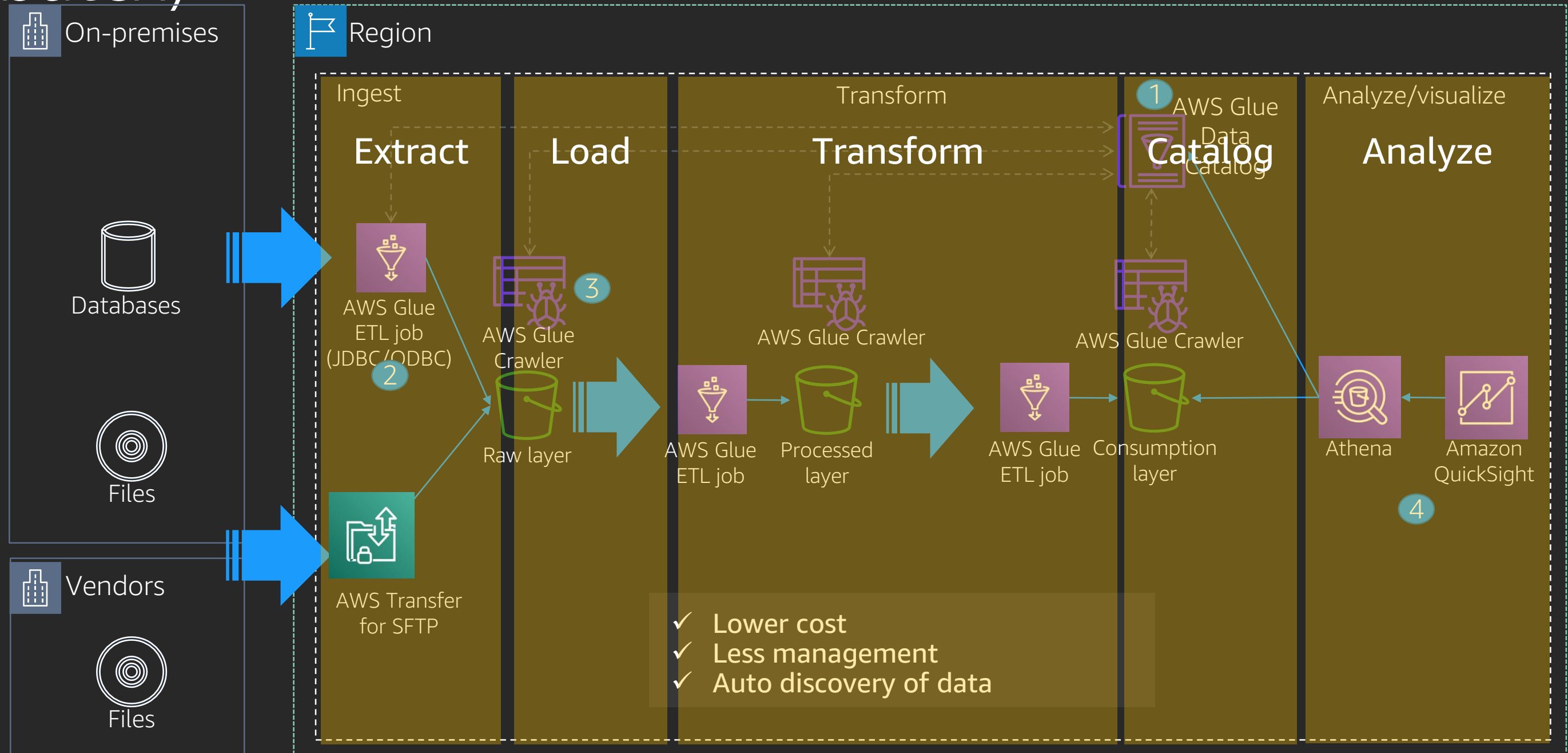
ETL/ELT reference architectures

Server-based data lake ingest pipeline with Amazon EMR (batch)

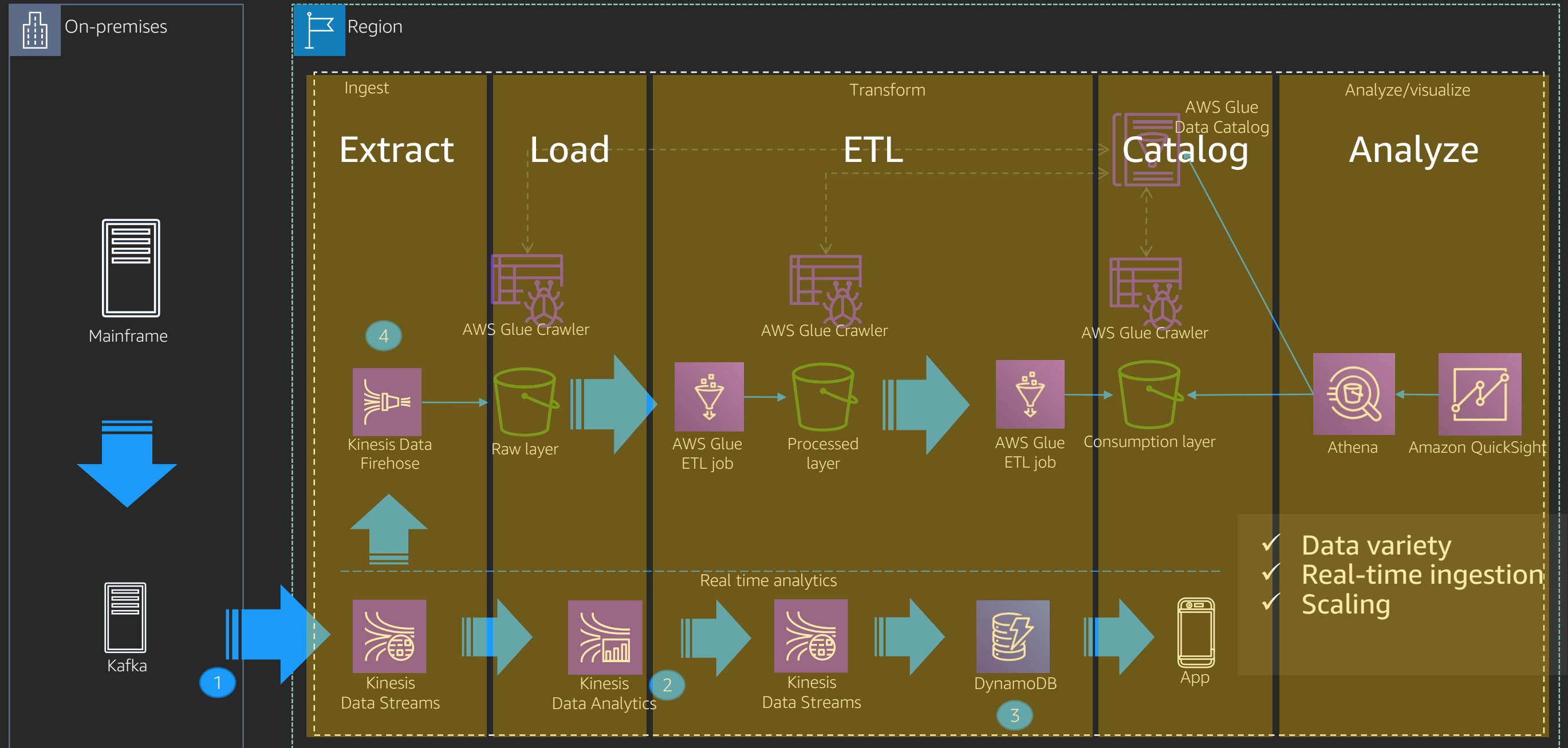


Serverless data lake ingest pipeline with AWS Glue

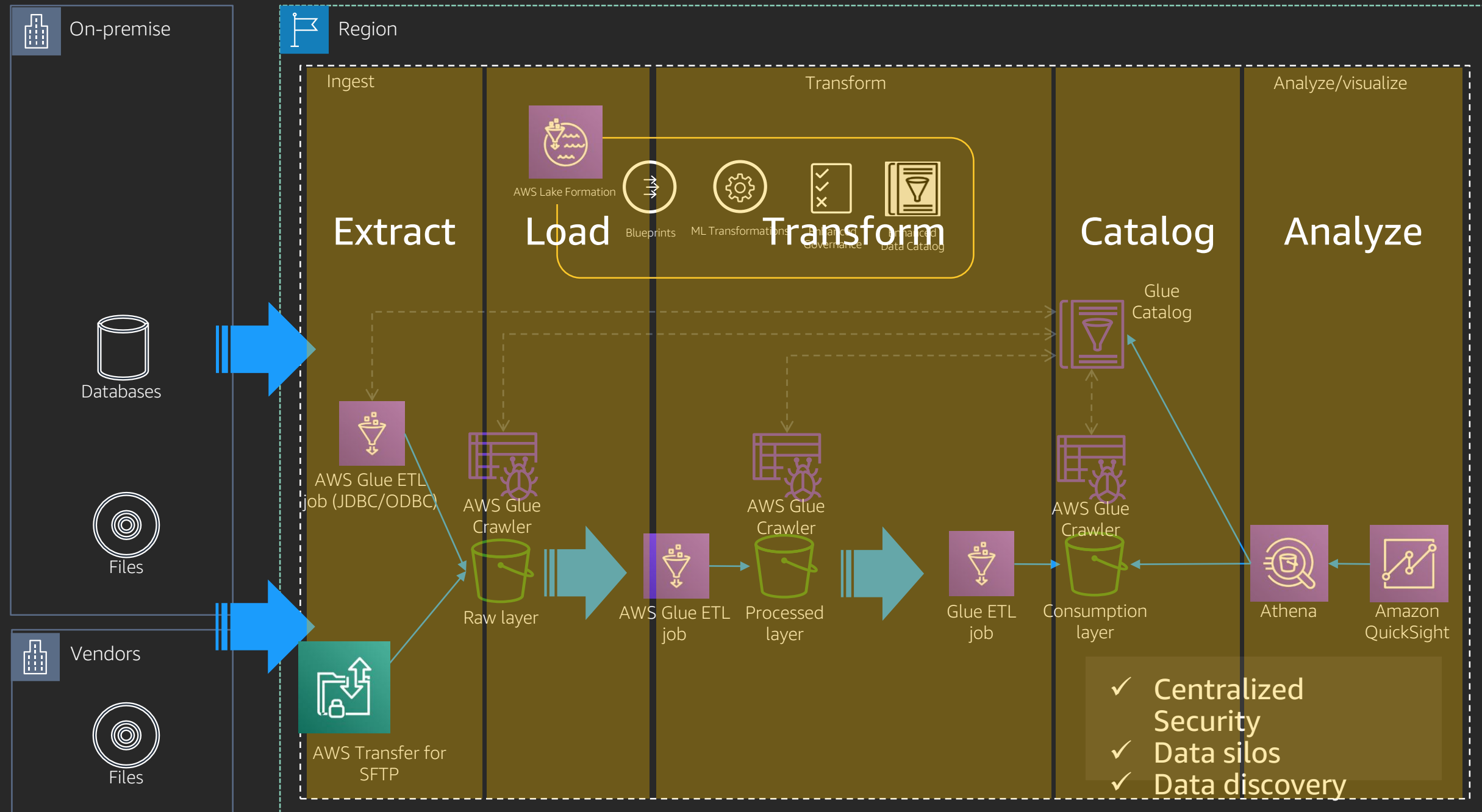
(batch)



Serverless data lake ingest pipeline with AWS Glue and Amazon Kinesis (streaming)



Serverless data lake ingest pipeline with AWS Lake Formation



No ETL

- Amazon Redshift Federated Query (Preview)
- Amazon Athena Federated Query (Preview)
- AQUA (Advanced Query Accelerator) for Amazon Redshift

Amazon Redshift Federated Query

PREVIEW

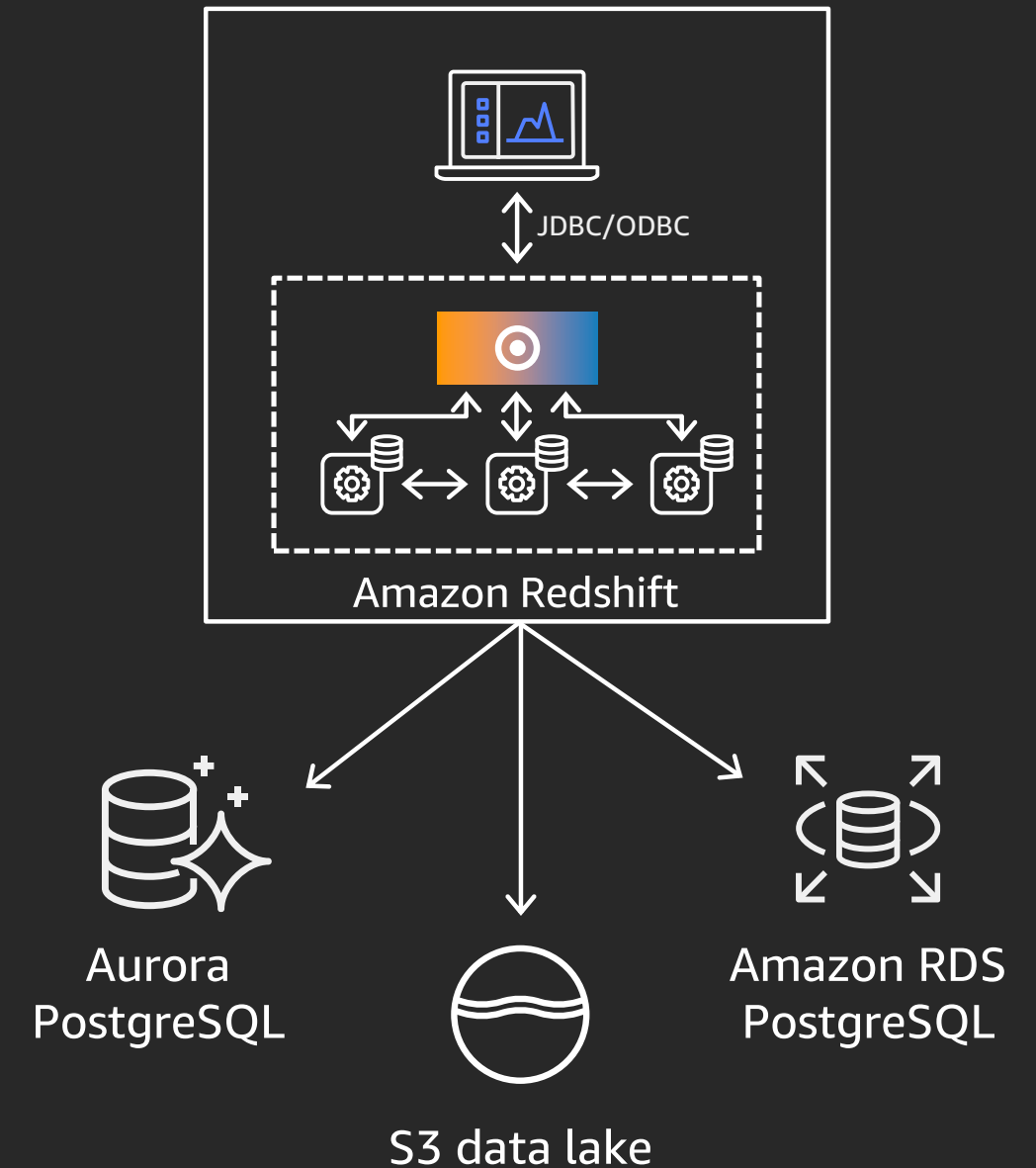
Queries on RDS and Aurora PostgreSQL databases

Analytics on live data without data movement

Unified analytics across data warehouse, data lake, and operational databases

Flexible and easy way to ingest data

Performant and secure access to data



Amazon Athena Federated Query

PREVIEW

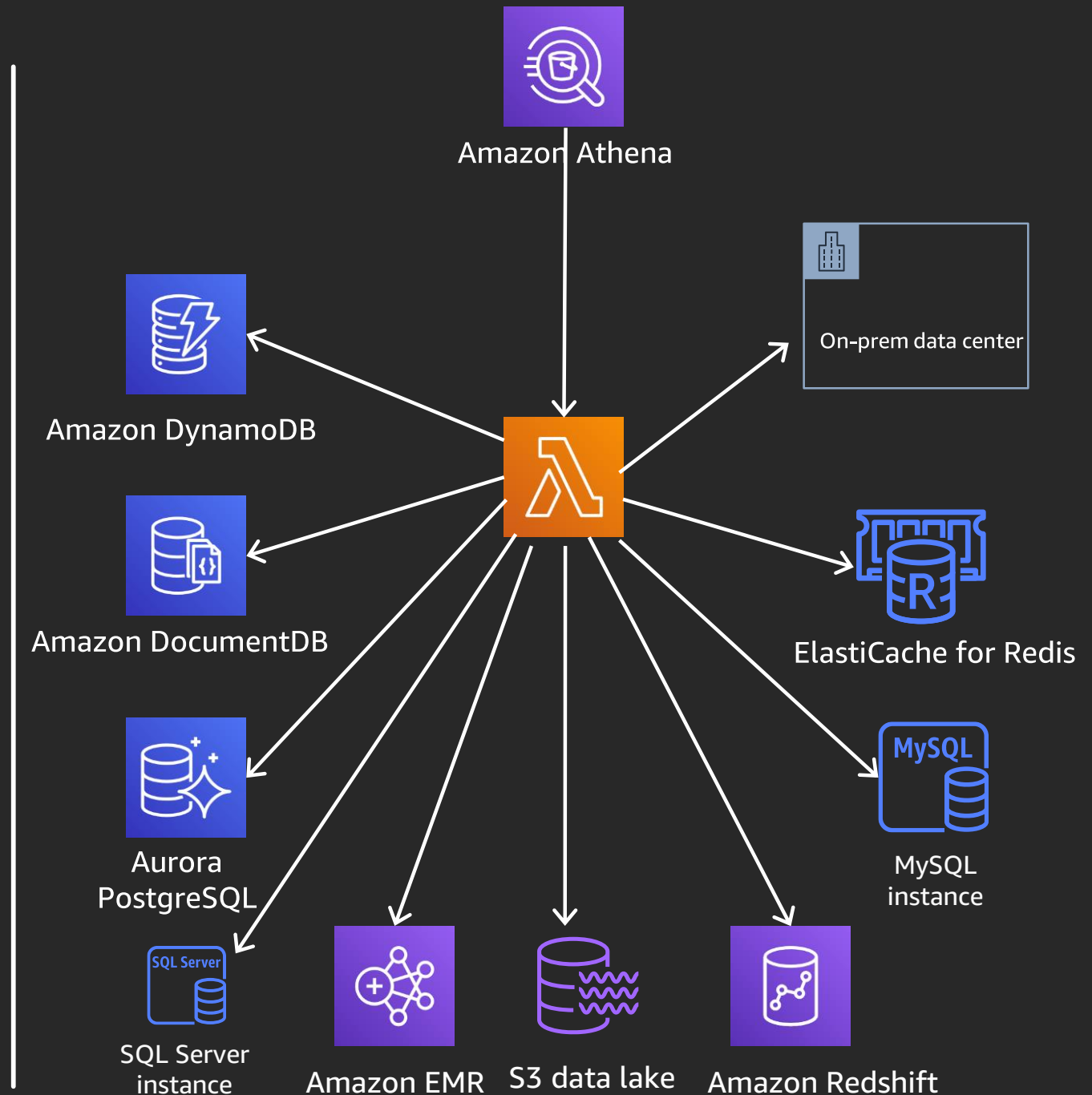
10 new sources available today

Support for your own custom sources

No ETLs, no servers, no problems

100% open source. SDK and connectors are on GitHub

One-click deployment

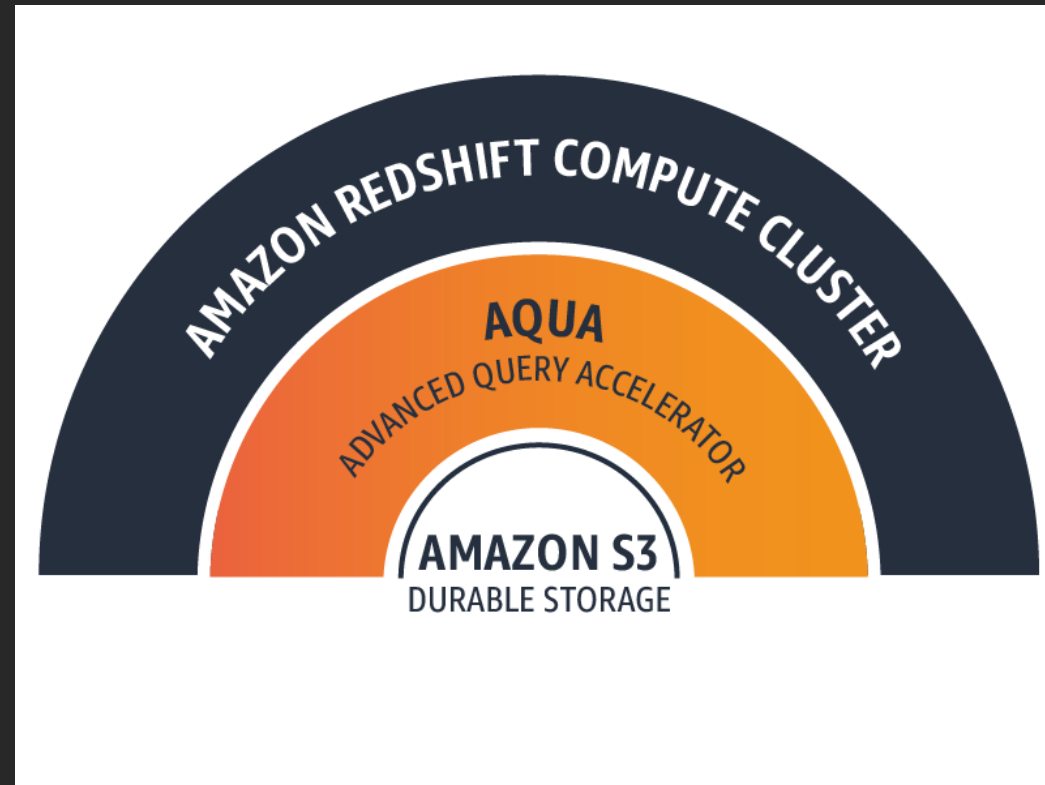


PREVIEW

COMING IN 2020

AQUA – Advanced Query Accelerator

Amazon Redshift runs **10x faster** than any other cloud data warehouse without increasing cost



AQUA brings compute to the storage layer so data doesn't have to move back and forth

High-speed cache on top of Amazon S3 scales out to process data in parallel across many nodes

AWS designed processors accelerate data compression, encryption, and data processing

100% compatible with the current version of Amazon Redshift

Solutions summary

- Ingesting semi-structured and unstructured data
- Ingesting and analyze real-time data
- Managing cost
- Decoupling storage and compute
- Scaling the data platform
- Managing infrastructure
- Data discovery
- Centralized security
- Data silos

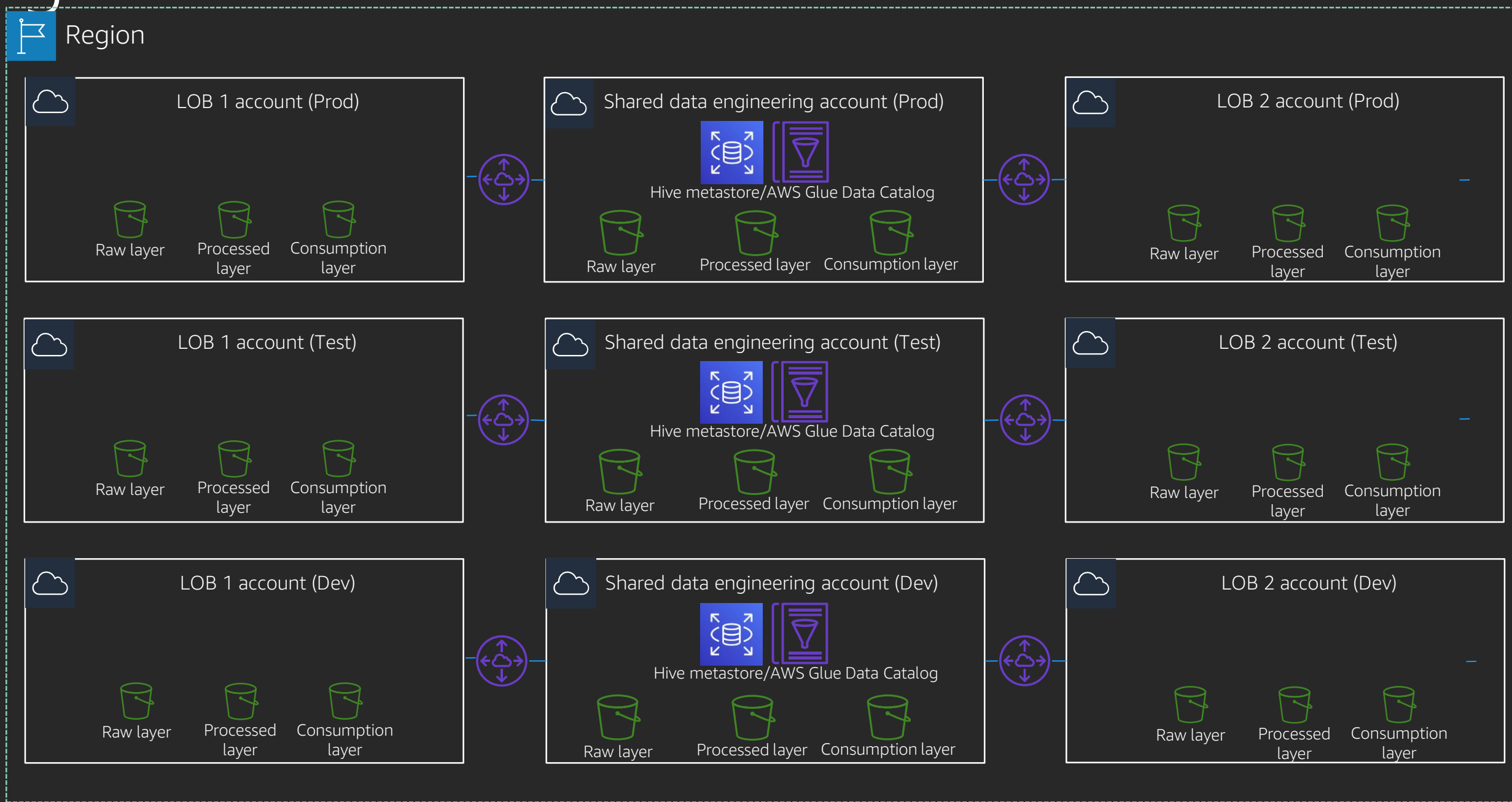
Open discussion

Thank you!



Please complete the session
survey in the mobile app.

Recommended pattern for catalog and bucket organization



Centralized metastore/catalog (ingest/transform)

