AWS: REINVENT 2018

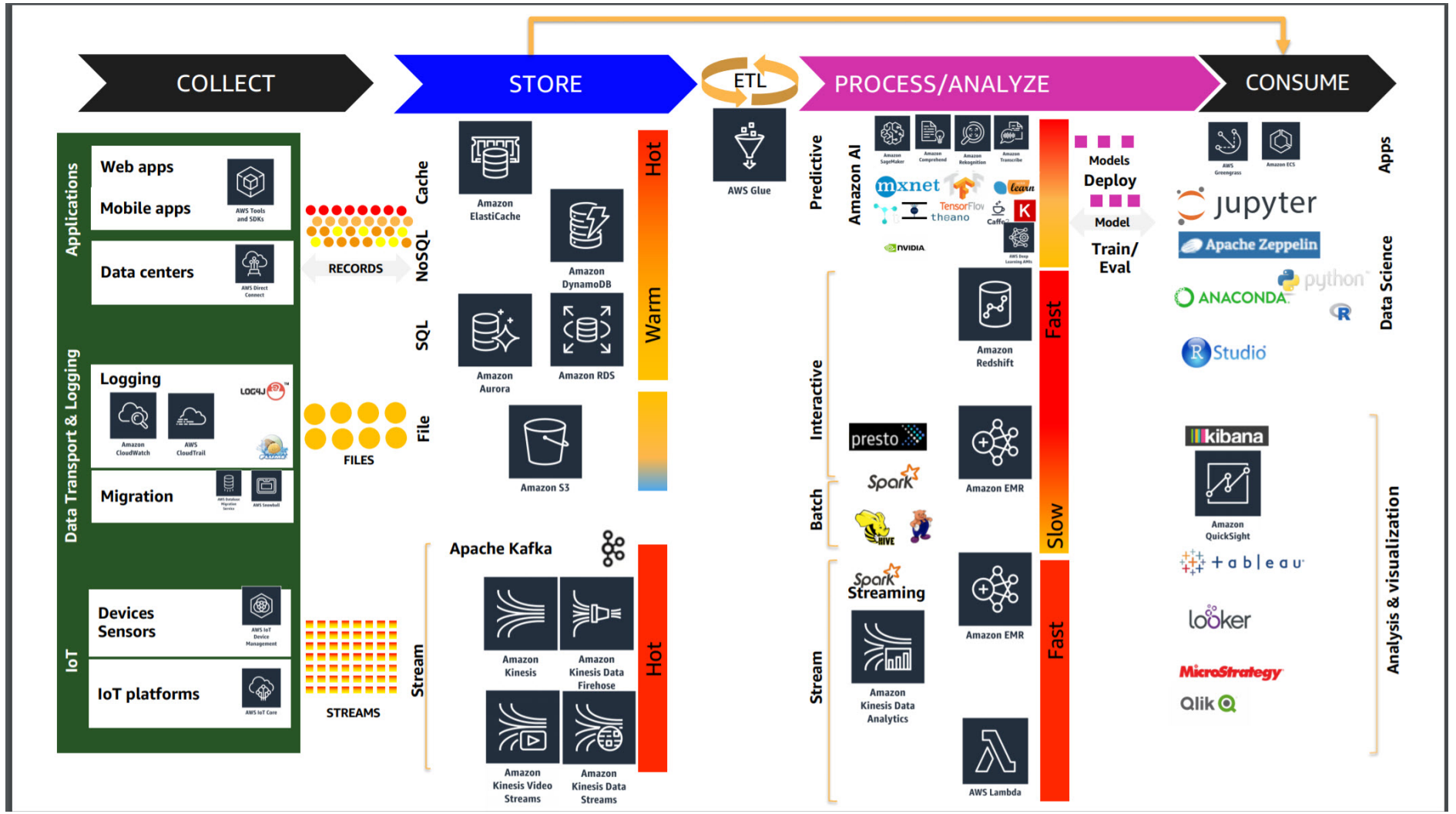
Mahesh KC

Note: Contents and images are copied from AWS Reinvent 2018 slideshare which solely belongs to AWS.

Agenda

- Data Pipeline
- Data Lake for Machine Learning
- Data Lake patterns for Voice, Vision, Advanced Analytics and ML using Serverless solutions
- Build, Deploy, and Server Machine-Learning Models on Streaming Data Using Amazon SageMaker, Apache Spark on Amazon EMR, and Amazon Kinesis
- Replicate and Manage Data Using Managed Databases and Serverless Technologies
- Capture Voice of Customer Insights with NLP & Analytics
- AWS Services

Data Pipeline



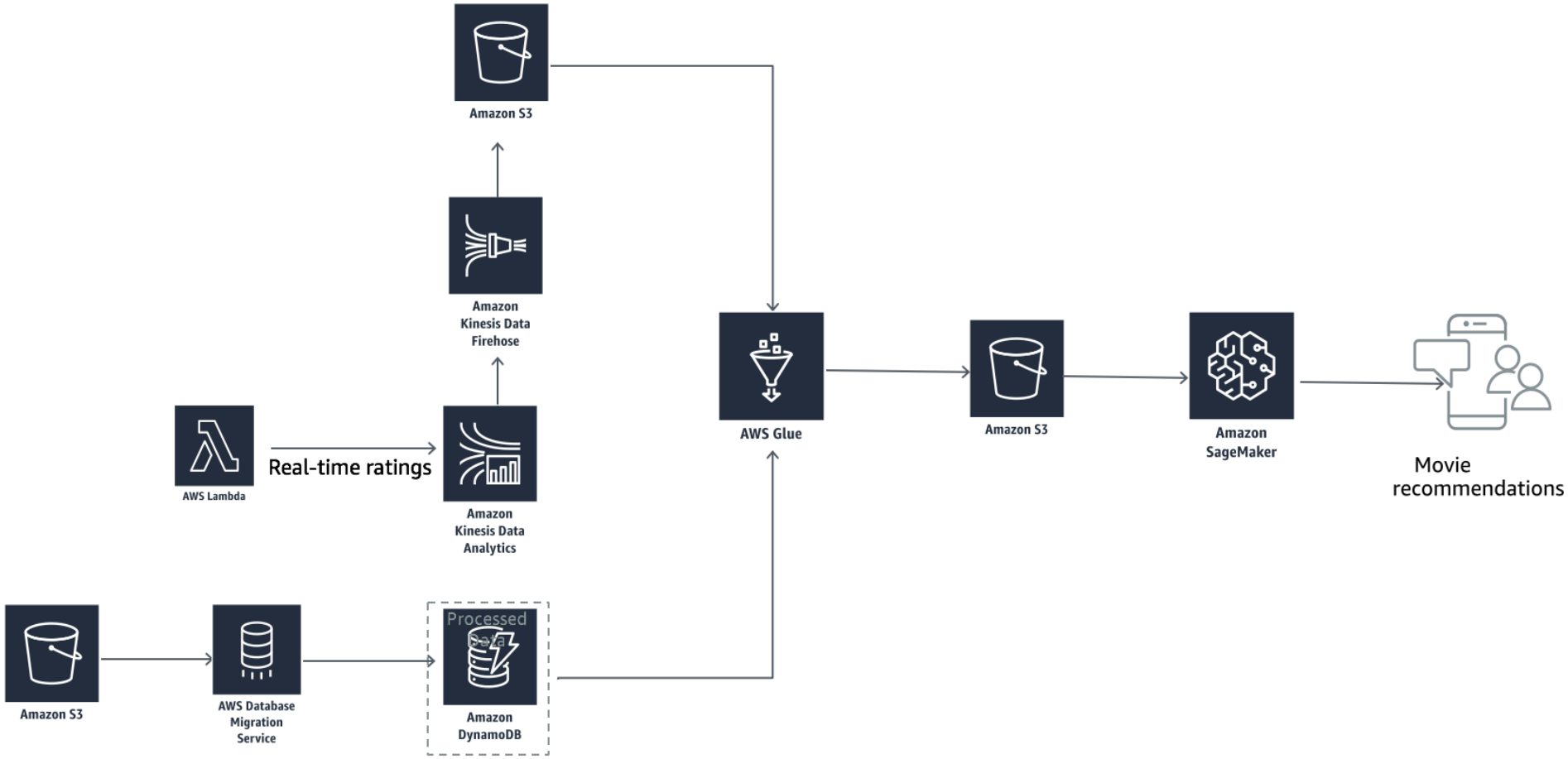
Data Lake for Machine Learning

Data lake characteristics

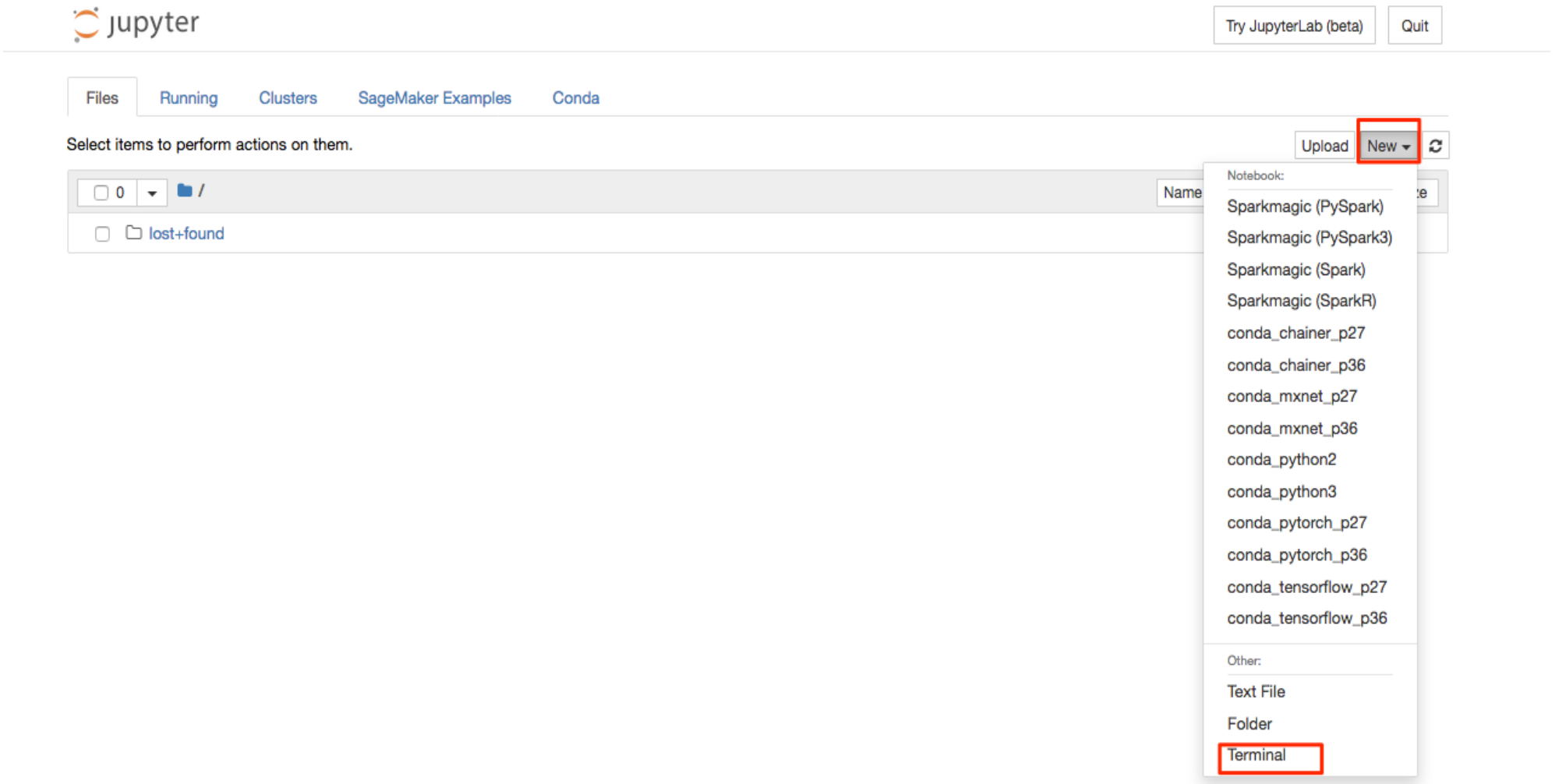
- Collect, store, process, consume, and analyze all organizational data
- Structured, semi-structured, and unstructured data
- Decoupled compute and storage
- Fast automated ingestion
- Schema on read
- AI/ML and BI/analytical use cases
- Complementary to data warehouses

Machine learning is more than building a model. It involves tasks that includes data sourcing, data ingestion, data transformation, pre-processing data for use in training, training a model and hosting the model.

Data Lake for Machine Learning Cont.



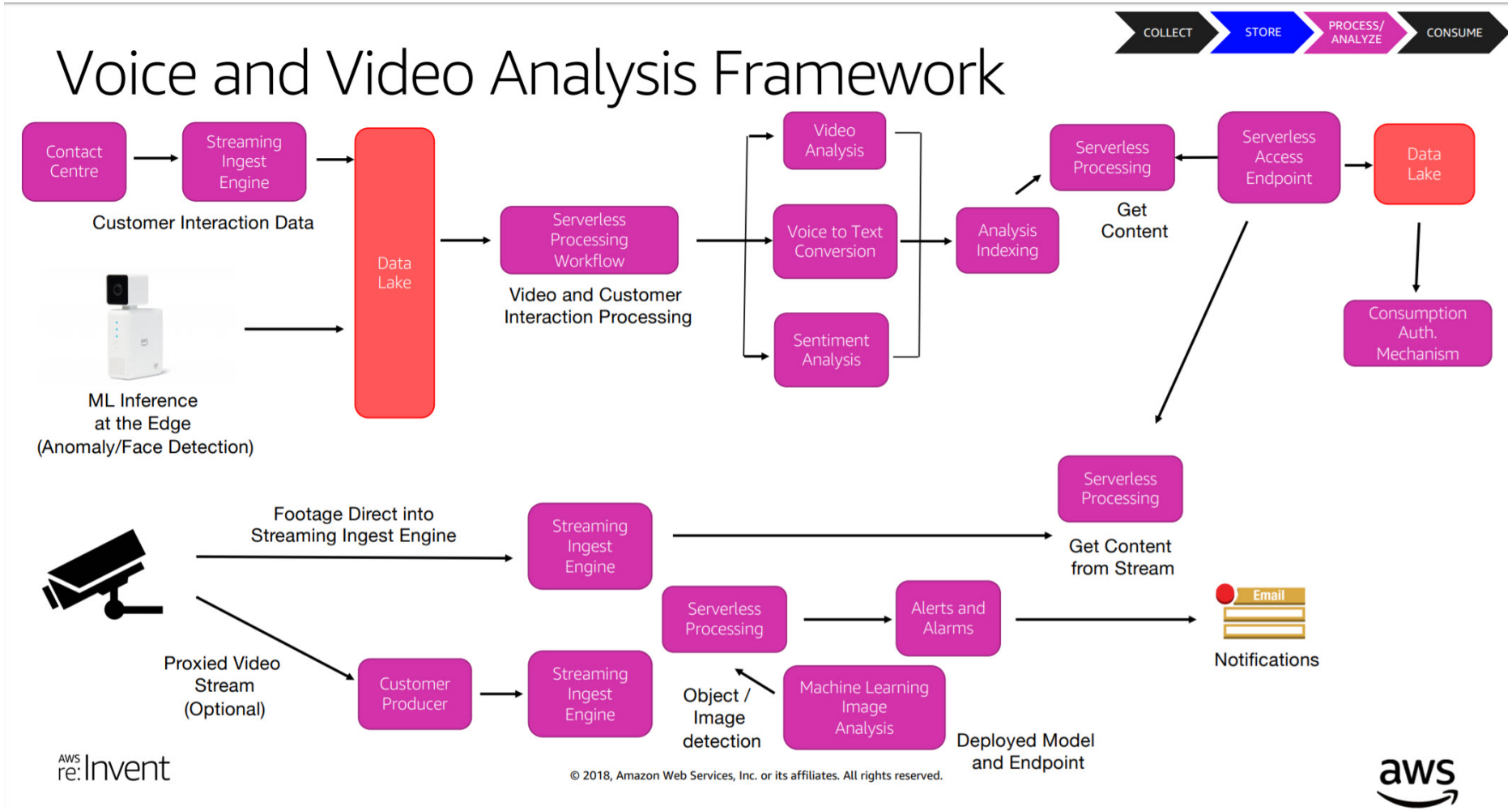
Data Lake for Machine Learning Cont.



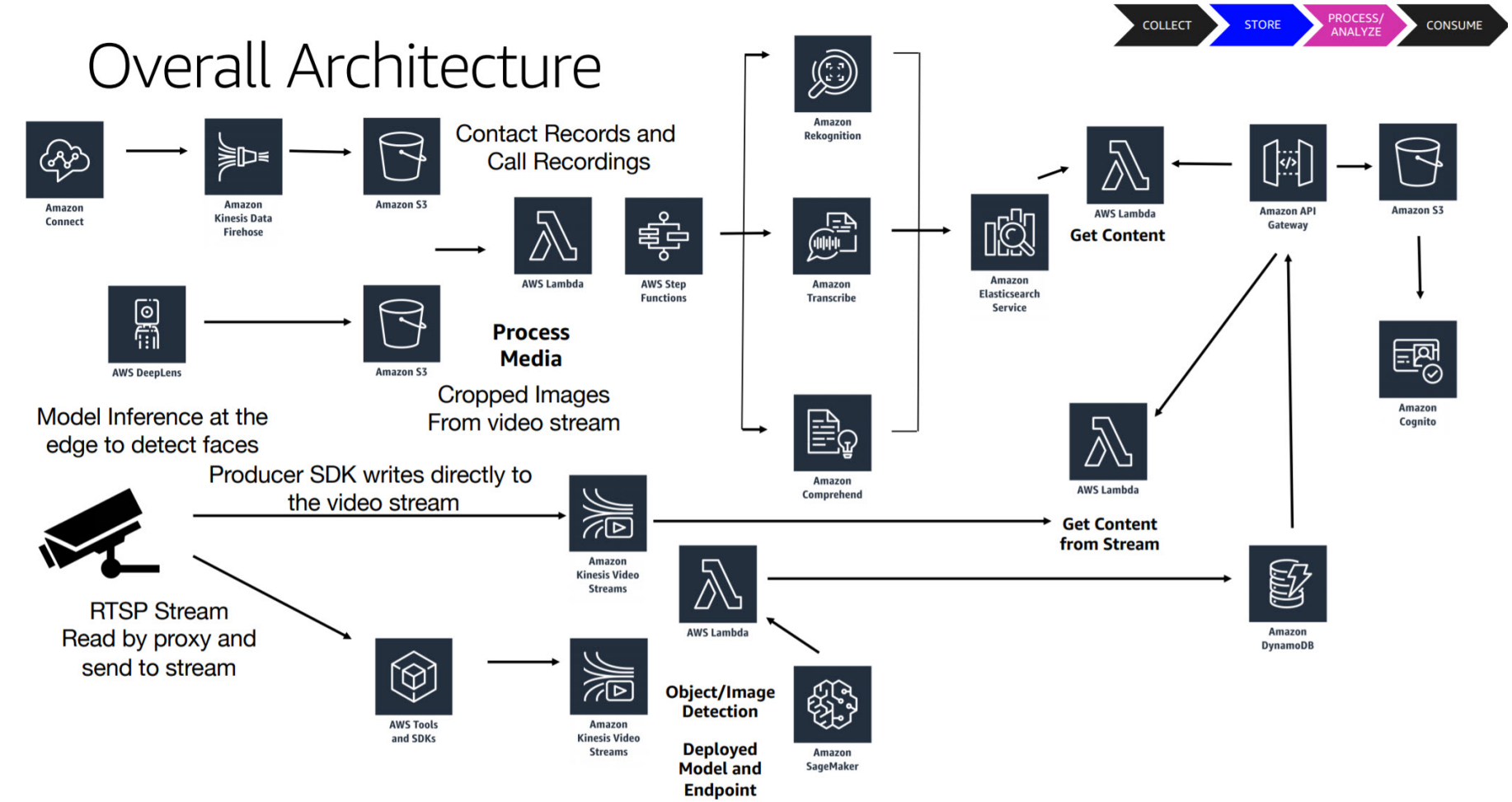
Data Lake patterns for Voice, Vision, Advanced Analytics and ML using Serverless solutions

- Collect video feed from camera (CCTV footage, production line camera)
- Collect voice fee using hotline (Automated hotline, Analyze contact center recordings)
- Analysis for sentiment/anomaly detection
- Use data sources for ML training and inference
- Real-time dashboard and alerting

Data Lake patterns for Voice, Vision, Advanced Analytics and ML using Serverless solutions Cont.



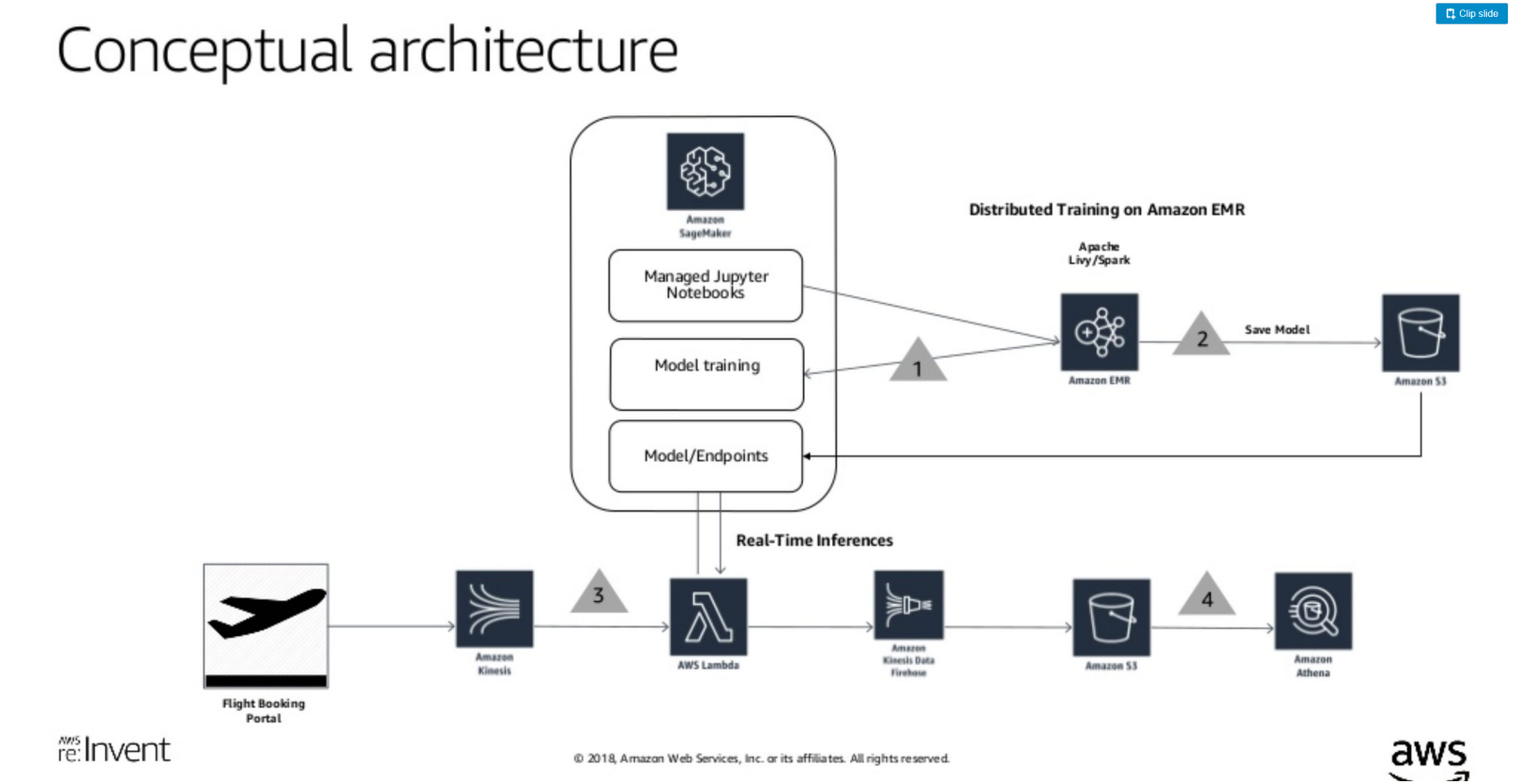
Data Lake patterns for Voice, Vision, Advanced Analytics and ML using Serverless solutions Cont.



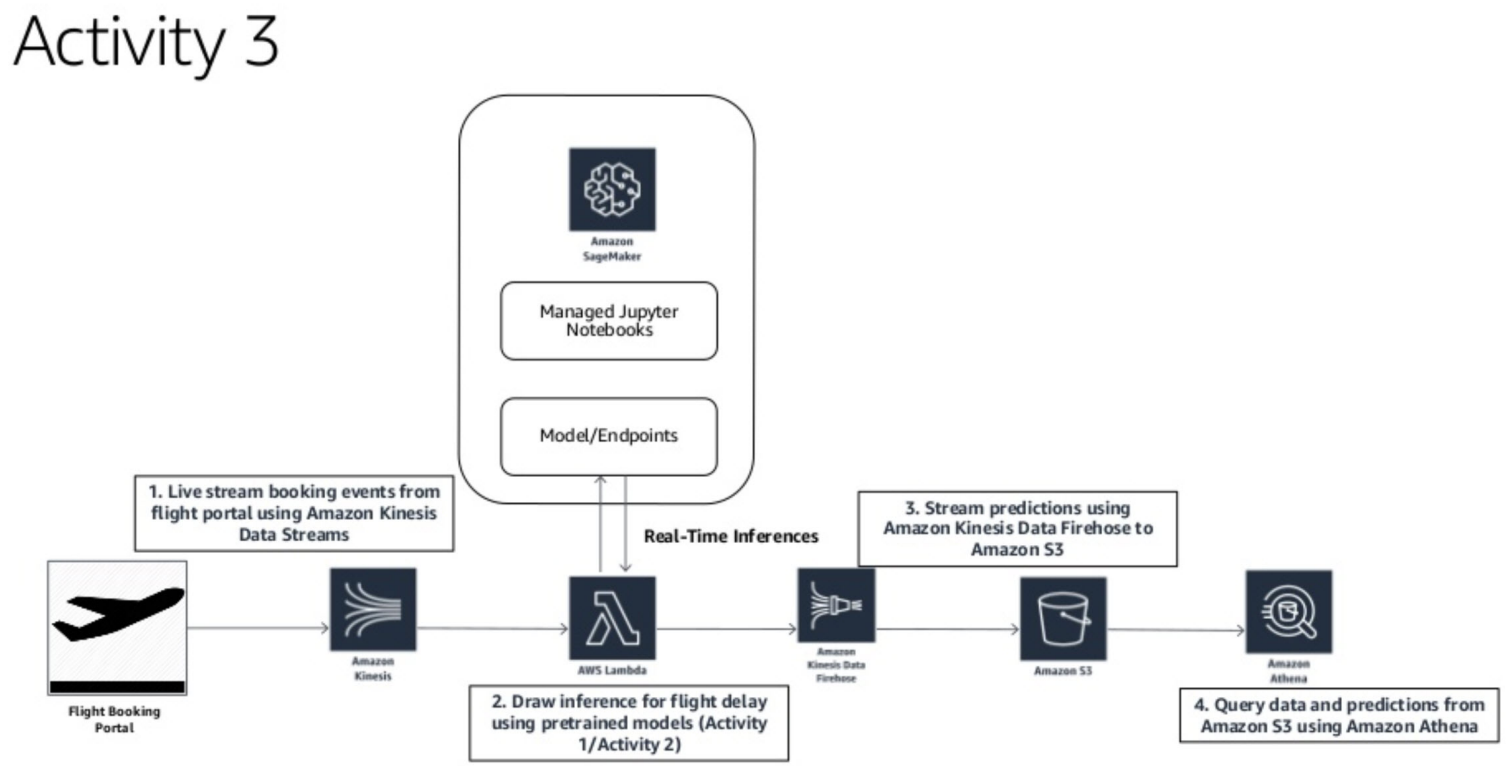
Build, Deploy, and Server Machine-Learning Models on Streaming Data Using Amazon SageMaker, Apache Spark on Amazon EMR, and Amazon Kinesis

- Explore data wrangling using Apache Spark, train and deploy ML models using Amazon SageMaker
- Train model using Apache Spark and deploy on Amazon Sagemaker
- Server ML model inferences for real-time data ingested through Amazon Kinesis Data Streams
- Query and compare inference results across deployed models

Build, Deploy, and Server Machine-Learning Models on Streaming Data Using Amazon SageMaker, Apache Spark on Amazon EMR, and Amazon Kinesis Cont.



Build, Deploy, and Server Machine-Learning Models on Streaming Data Using Amazon SageMaker, Apache Spark on Amazon EMR, and Amazon Kinesis Cont.

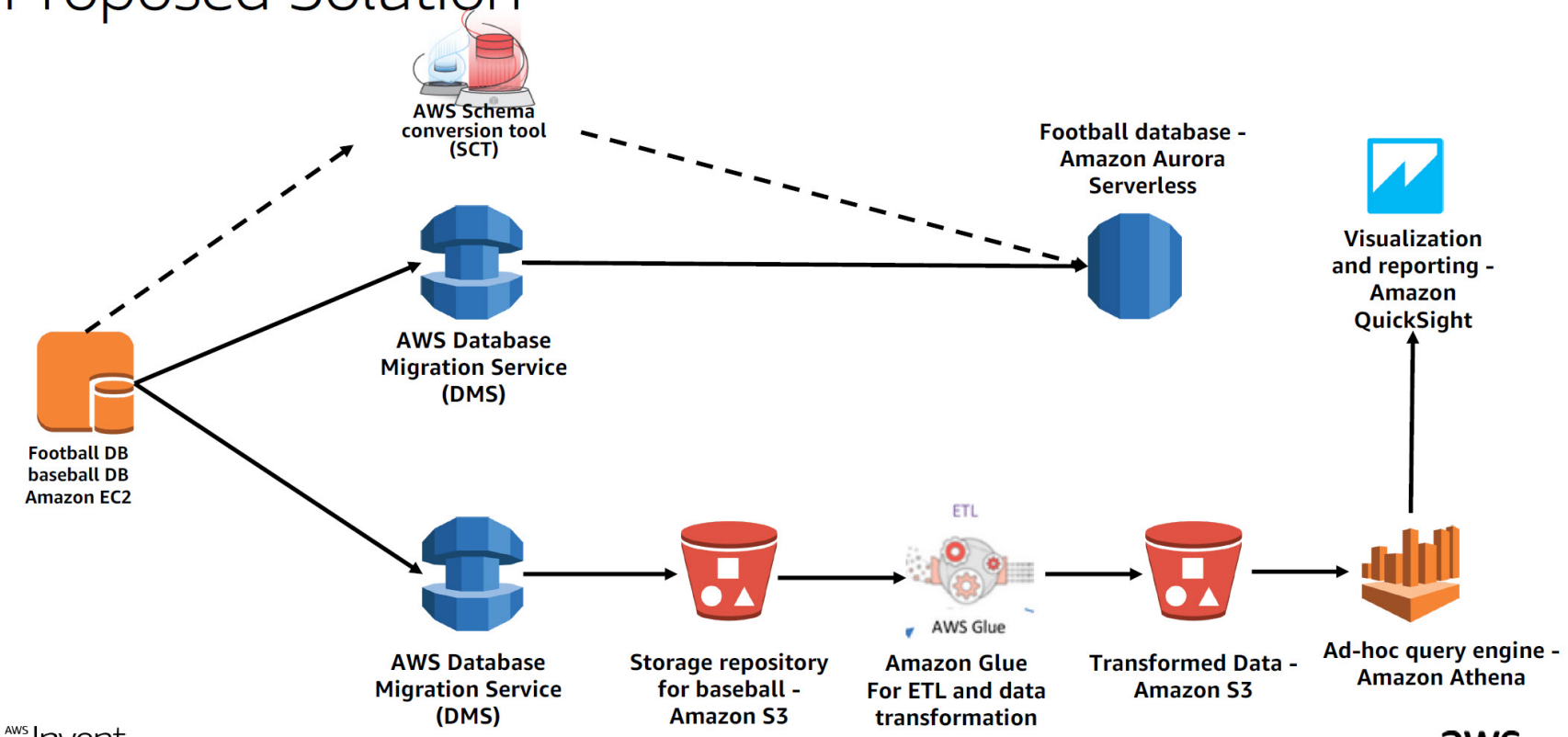


Replicate and Manage Data Using Managed Databases and Serverless Technologies

- Serverless technologies to replicate, analyze, and visualize data from relational and semi-structured datasets
- Leverage AWS SCT and AWS DMS to convert MS SQL server database to Amazon Aurora Serverless MySQL
- Use AWS Glue to build data catalog and transform data
- Use Amazon Athena to run ad-hoc SQL queries and interactively analyze data stored on S3 data store.
- Use Anazon QuickSight to visualize data and drive additional insights.

Replicate and Manage Data Using Managed Databases and Serverless Technologies Cont.

Proposed Solution



Replicate and Manage Data Using Managed Databases and Serverless Technologies Cont.

When to use AWS DMS and AWS SCT?

Modernize

Heterogeneous database migrations

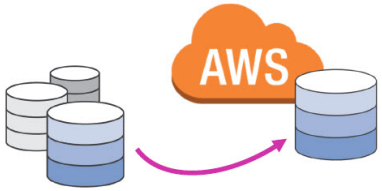
Modernize your database tier –

- Commercial to open-source
- Commercial to Amazon Aurora

Modernize your Data Warehouse –

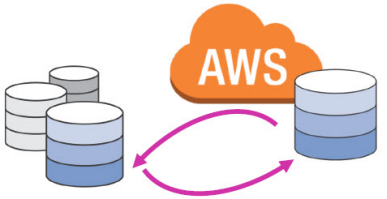
- Commercial to Amazon Redshift

Migrate



- Migrate business-critical applications
- Migrate from classic to VPC
- Migrate data warehouse to Amazon Redshift
- Upgrade to a minor version
- Consolidate shards into Aurora

Replicate



- Create cross-region read replicas
- Run your analytics in the cloud
- Keep your dev/test and production environment sync

Capture Voice of Customer Insights with NLP & Analytics

I haven't done this lab but instruction are on the link.

https://s3.amazonaws.com/serverless-analytics/reinvent-2018-comprehend-transcribe-connect/lab1.html

AWS Services