

AWS: REINVENT 2018

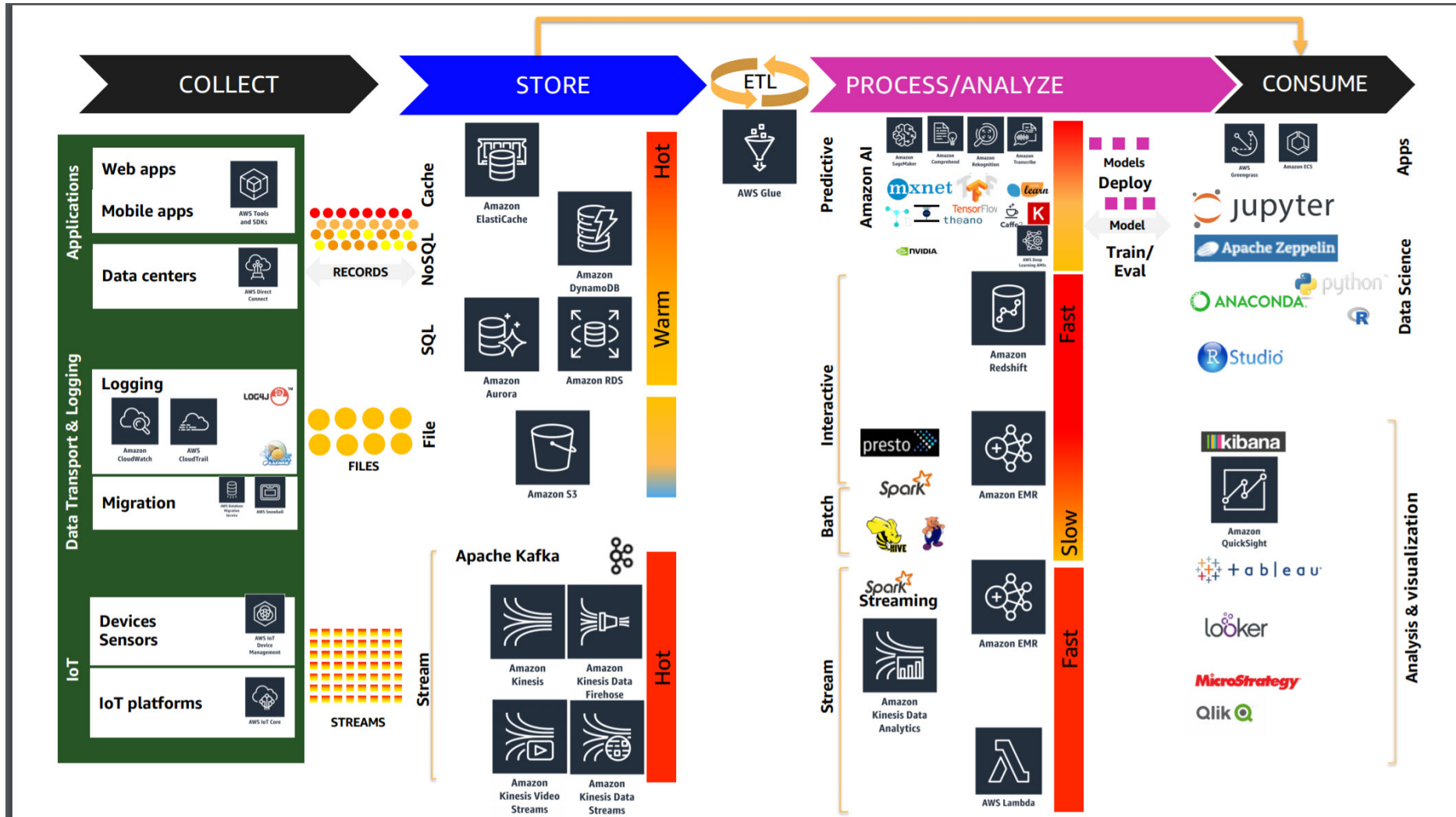
Mahesh KC

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



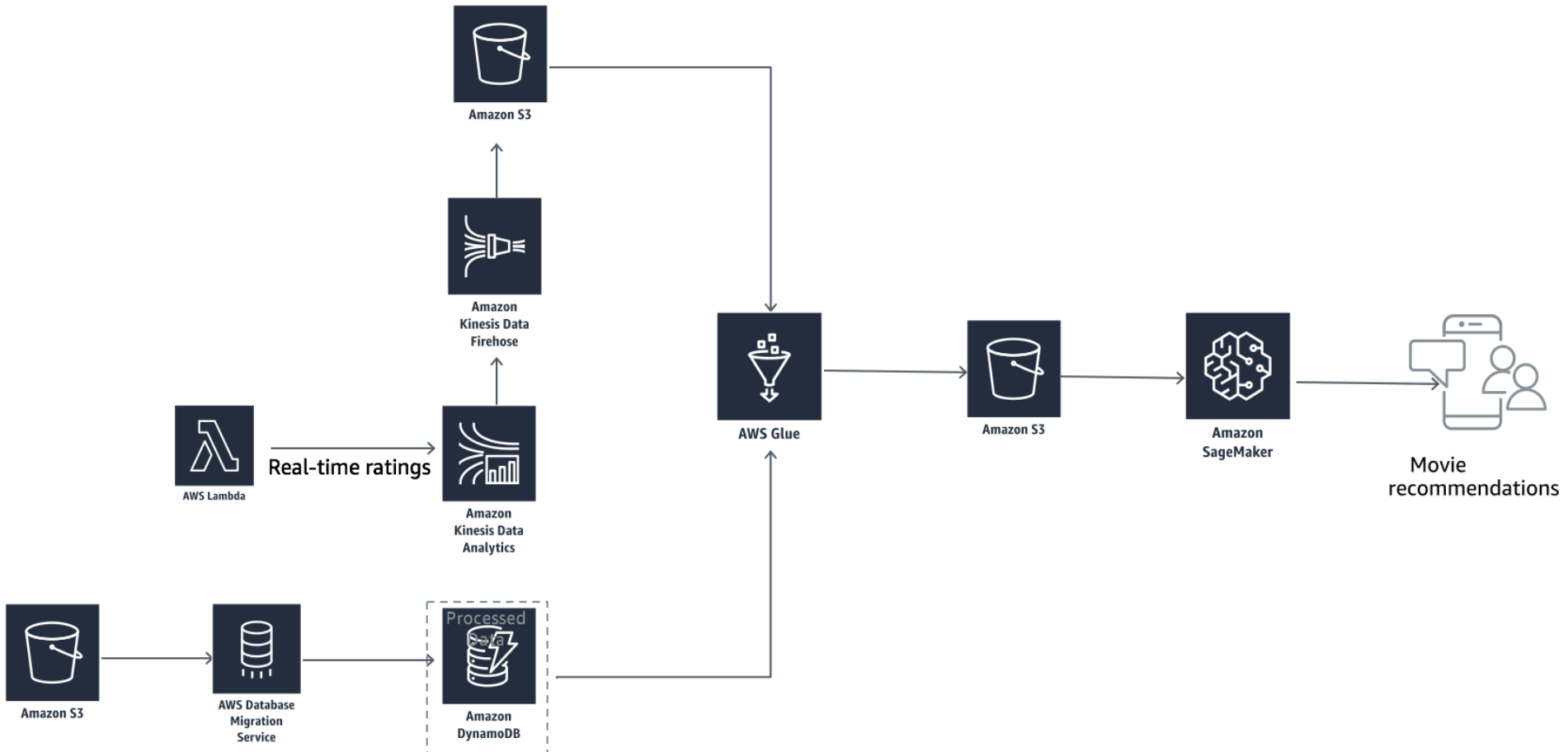
Data Lake for Machine Learning

Data lake characteristics

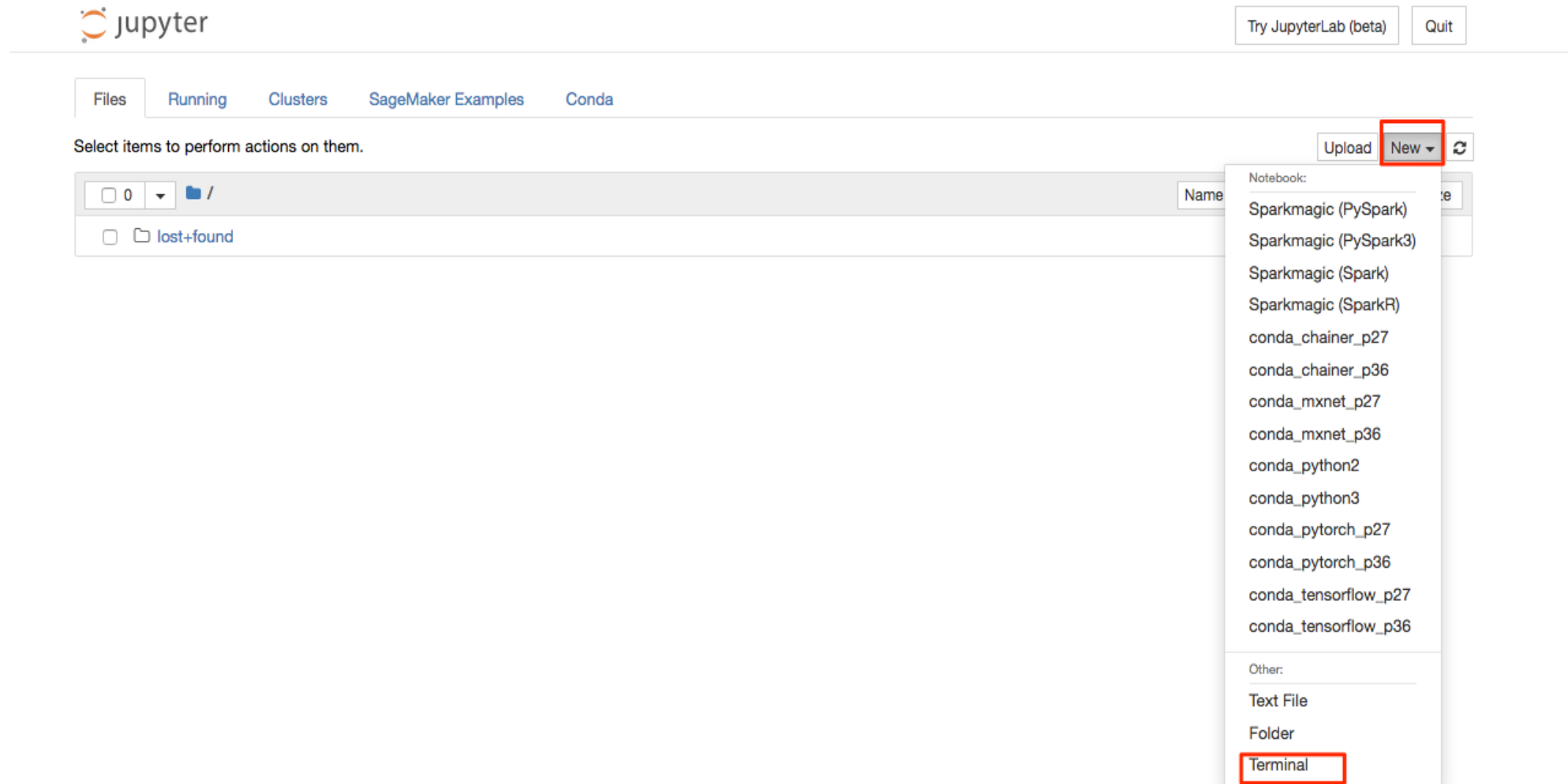
- Collect, store, process, consume, and analyze all organizational data
- Structured, semi-structured, and unstructured data
- Decoupled compute and storage
- Fast automated ingestion
- Schema on read
- AI/ML and BI/analytical use cases
- Complementary to data warehouses

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.



The screenshot displays the JupyterLab web interface. At the top left is the Jupyter logo. On the top right, there are buttons for 'Try JupyterLab (beta)' and 'Quit'. Below these is a navigation bar with tabs for 'Files', 'Running', 'Clusters', 'SageMaker Examples', and 'Conda'. The 'Files' tab is active, showing a file browser with a message 'Select items to perform actions on them.' and a list of files including 'lost+found'. To the right of the file list are buttons for 'Upload', 'New', and a refresh icon. The 'New' button is highlighted with a red box, and its dropdown menu is open. The menu is divided into two sections: 'Notebook:' and 'Other:'. The 'Notebook:' section lists various notebook templates such as 'Sparkmagic (PySpark)', 'Sparkmagic (PySpark3)', 'Sparkmagic (Spark)', 'Sparkmagic (SparkR)', and several 'conda' environments for different frameworks like 'chainer', 'mxnet', 'python', 'pytorch', and 'tensorflow'. The 'Other:' section includes 'Text File', 'Folder', and 'Terminal', with the 'Terminal' option highlighted by a red box.

jupyter

Try JupyterLab (beta) Quit

Files Running Clusters SageMaker Examples Conda

Select items to perform actions on them.

Upload New

Notebook:

- Sparkmagic (PySpark)
- Sparkmagic (PySpark3)
- Sparkmagic (Spark)
- Sparkmagic (SparkR)
- conda_chainer_p27
- conda_chainer_p36
- conda_mxnet_p27
- conda_mxnet_p36
- conda_python2
- conda_python3
- conda_pytorch_p27
- conda_pytorch_p36
- conda_tensorflow_p27
- conda_tensorflow_p36

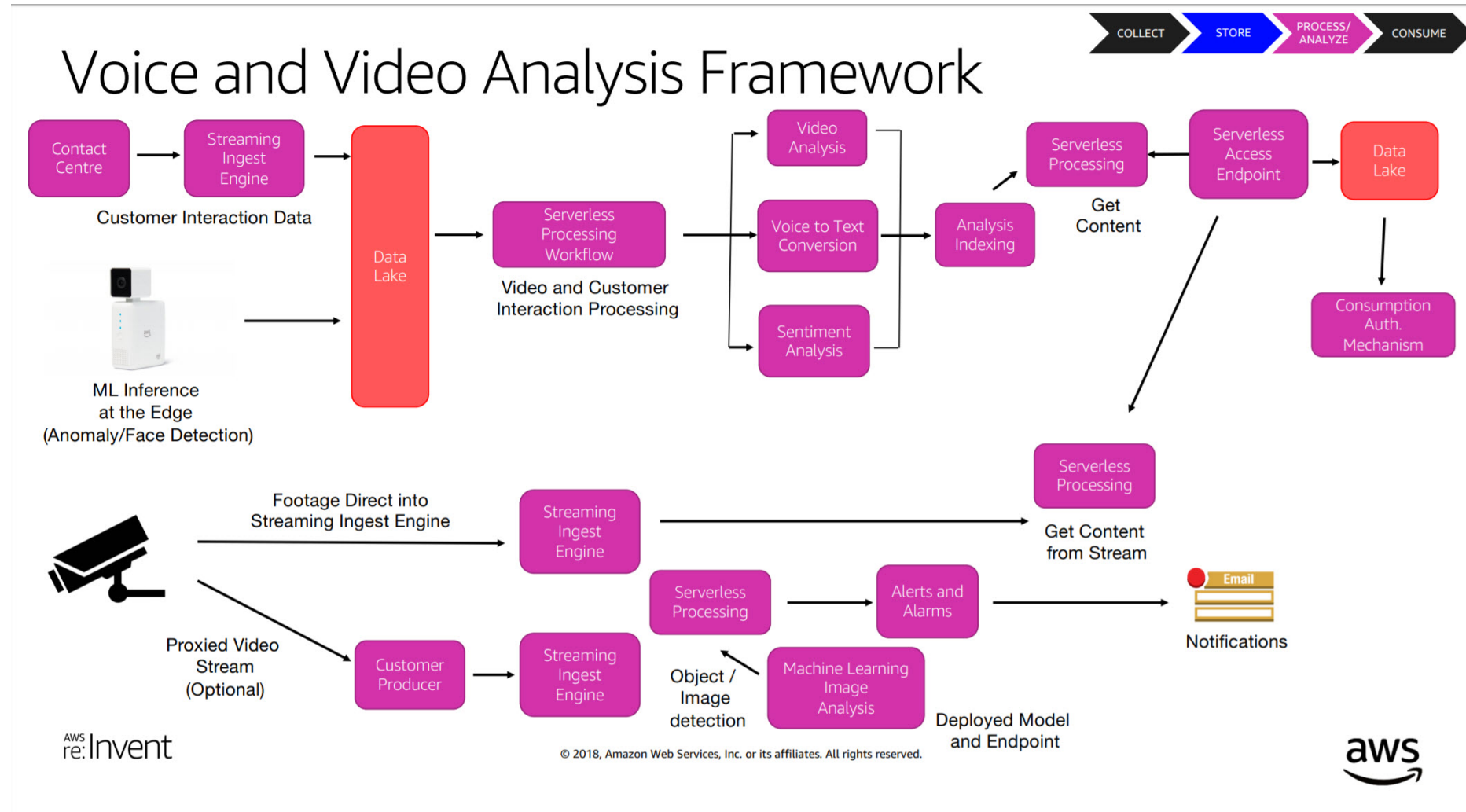
Other:

- Text File
- Folder
- Terminal

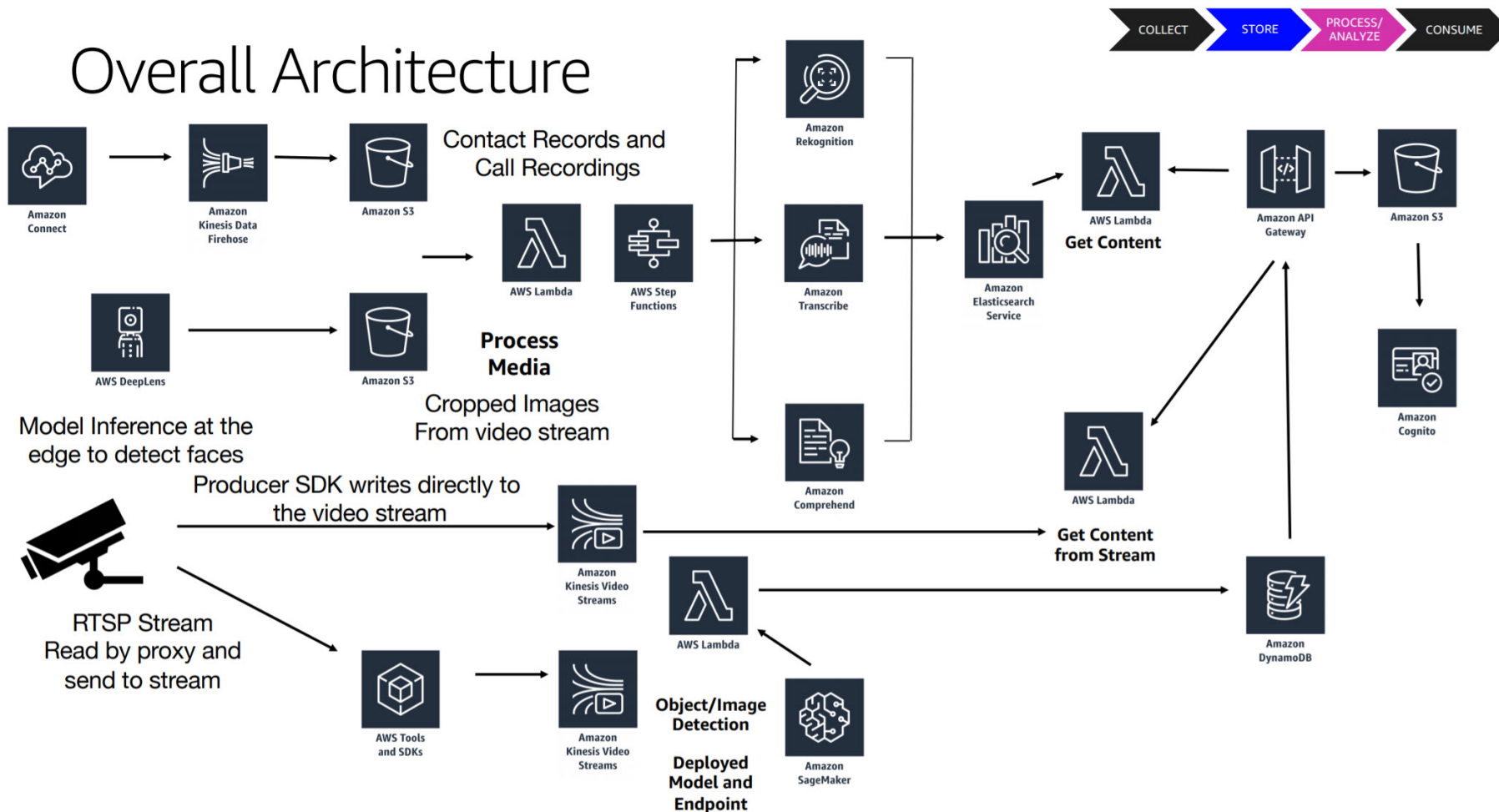
Data Lake patterns for Voice, Vision, Advanced Analytics and ML using Serverless solutions

- 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.



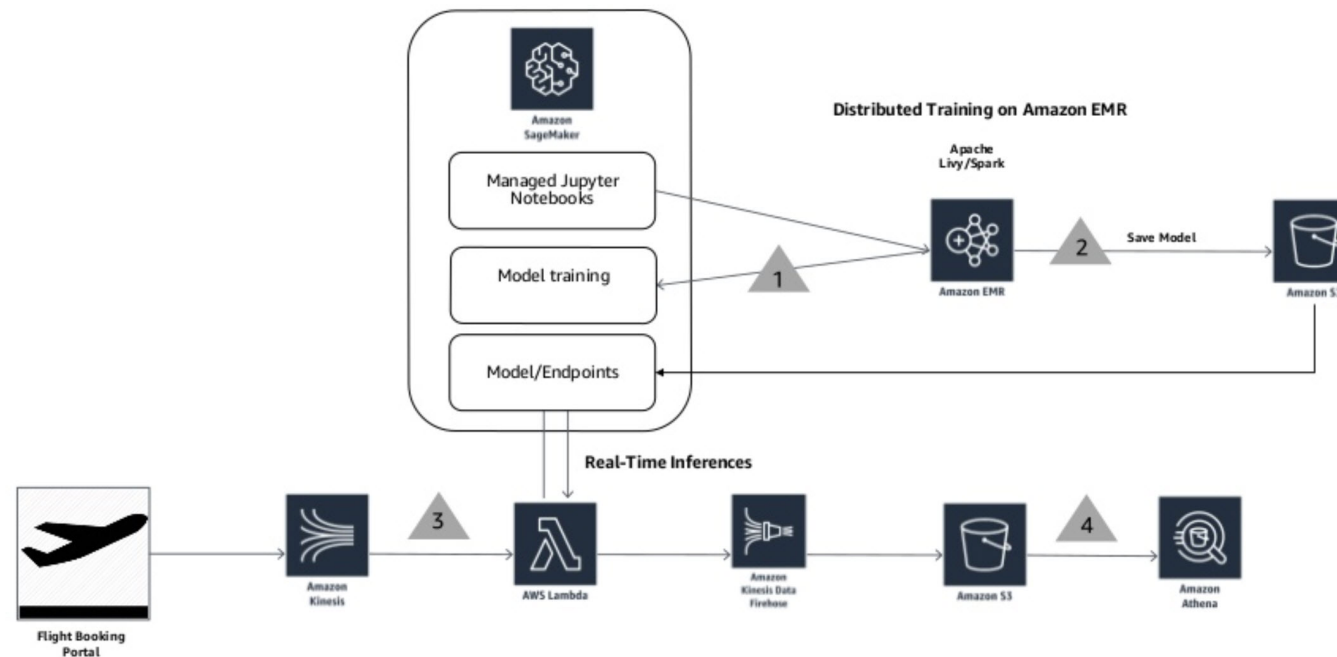
Build, Deploy, and Server Machine-Learning Models on Streaming Data Using Amazon SageMaker, Apache Spark on Amazon EMR, and Amazon Kinesis

- 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.

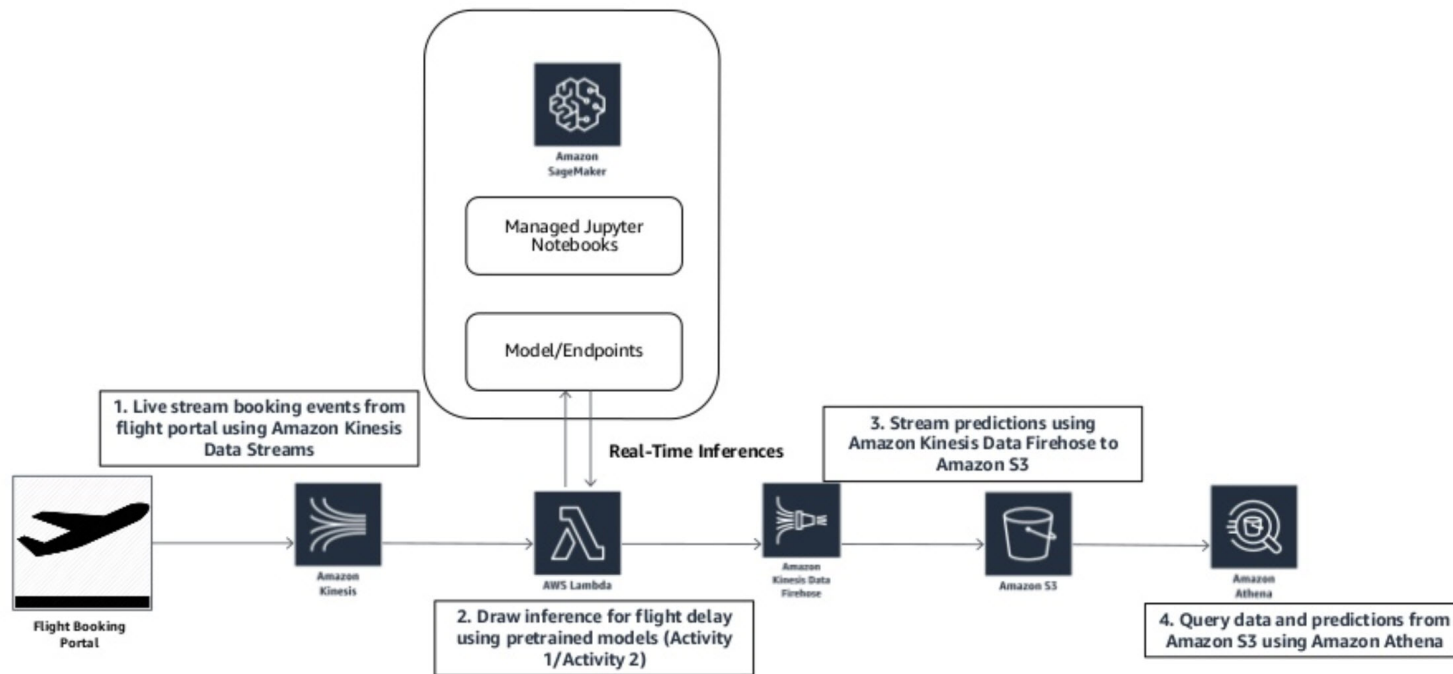
Conceptual architecture

Clip slide



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

Activity 3

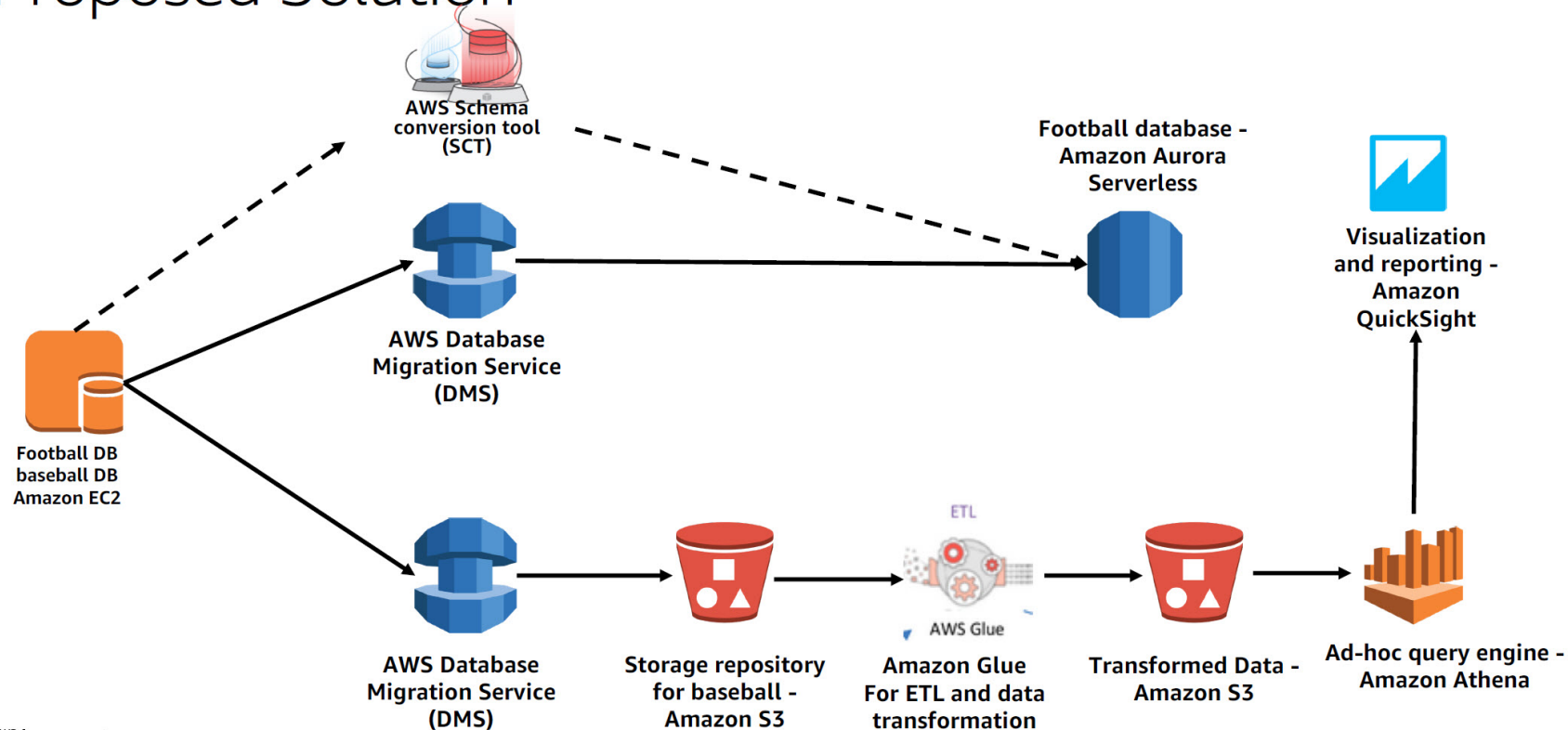


Replicate and Manage Data Using Managed Databases and Serverless Technologies

- 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 Amazon 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

AWS
re:Invent

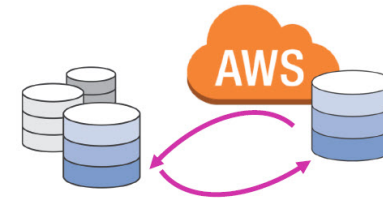
Migrate



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

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.

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

▼ All services

Compute

EC2
Lightsail 
ECS
EKS
Lambda
Batch
Elastic Beanstalk
ECR

Storage

S3
EFS
FSx
S3 Glacier
Storage Gateway


Database

RDS
DynamoDB
ElastiCache
Neptune
Amazon Redshift

Migration & Transfer

AWS Migration Hub
Application Discovery Service
Database Migration Service
Server Migration Service
AWS Transfer for SFTP
Snowball
DataSync

Networking & Content Delivery

VPC
CloudFront
Route 53
API Gateway
Direct Connect
AWS Cloud Map
Global Accelerator 

Developer Tools

CodeStar
CodeCommit
CodeBuild
CodeDeploy
CodePipeline
Cloud9
X-Ray

Robotics

AWS RoboMaker

Blockchain

Amazon Managed Blockchain

Satellite

Ground Station

Management & Governance

CloudWatch
AWS Auto Scaling
CloudFormation
CloudTrail
Config
OpsWorks
Service Catalog
Systems Manager
Trusted Advisor
Managed Services
Control Tower
AWS License Manager
AWS Well-Architected Tool


Media Services

Elastic Transcoder
Kinesis Video Streams
MediaConnect
MediaConvert
MediaLive
MediaPackage
MediaStore
MediaTailor


Machine Learning

Amazon SageMaker
Amazon Comprehend
AWS DeepLens
Amazon Lex
Machine Learning
Amazon Polly
Rekognition
Amazon Transcribe
Amazon Translate
Amazon Personalize
Amazon Forecast

Analytics

Athena
EMR
CloudSearch
Elasticsearch Service
Kinesis
QuickSight 
Data Pipeline
AWS Glue
MSK

Security, Identity, & Compliance

IAM
Resource Access Manager
Cognito
Secrets Manager
GuardDuty
Inspector
Amazon Macie 
AWS Organizations
AWS Single Sign-On
Certificate Manager
Key Management Service
CloudHSM
Directory Service
WAF & Shield
Artifact
Security Hub

AWS Cost Management

AWS Cost Explorer
AWS Budgets
AWS Marketplace Subscriptions

Mobile

AWS Amplify
Mobile Hub
AWS AppSync
Device Farm

AR & VR

Amazon Sumerian


Application Integration

Step Functions
Amazon MQ
Simple Notification Service
Simple Queue Service
SWF

Customer Engagement

Amazon Connect
Pinpoint
Simple Email Service

Business Applications

Alexa for Business
Amazon Chime 
WorkDocs
WorkMail

Desktop & App Streaming

WorkSpaces
AppStream 2.0

Internet of Things

IoT Core
Amazon FreeRTOS
IoT 1-Click
IoT Analytics
IoT Device Defender
IoT Device Management
IoT Events
IoT Greengrass
IoT SiteWise
IoT Things Graph

Game Development

Amazon GameLift