



Advanced Analytics on AWS Jumpstart

Dylan Tong, AI/Machine Learning Partner Solutions Architect

Jumpstart Workshop Menu

- Predictive Churn Analytics
 - Focus on batch, pipelining, and database access.
- TBD...

Advanced Analytics Workshop

Advanced Analytics using AI Services

INTELLIGENT BUILDING BLOCKS

DATA
SOURCES



Amazon
Forecast



Amazon
Comprehend

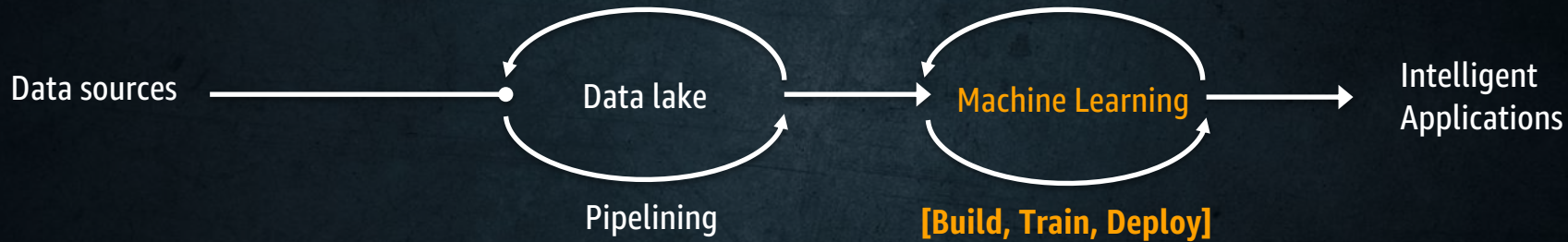


Amazon
Personalize

ETL



Build Strategy



Data Lake Core

AWS
Lake Formation

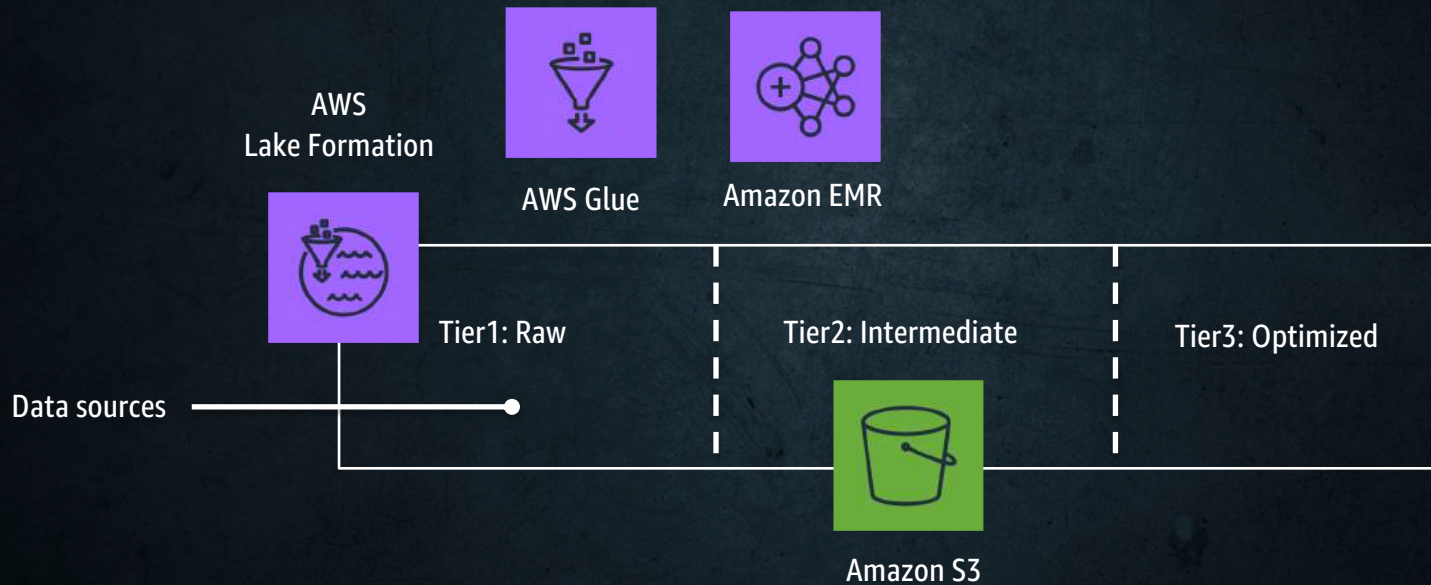


- Catalog
- Centralized Security
- Storage

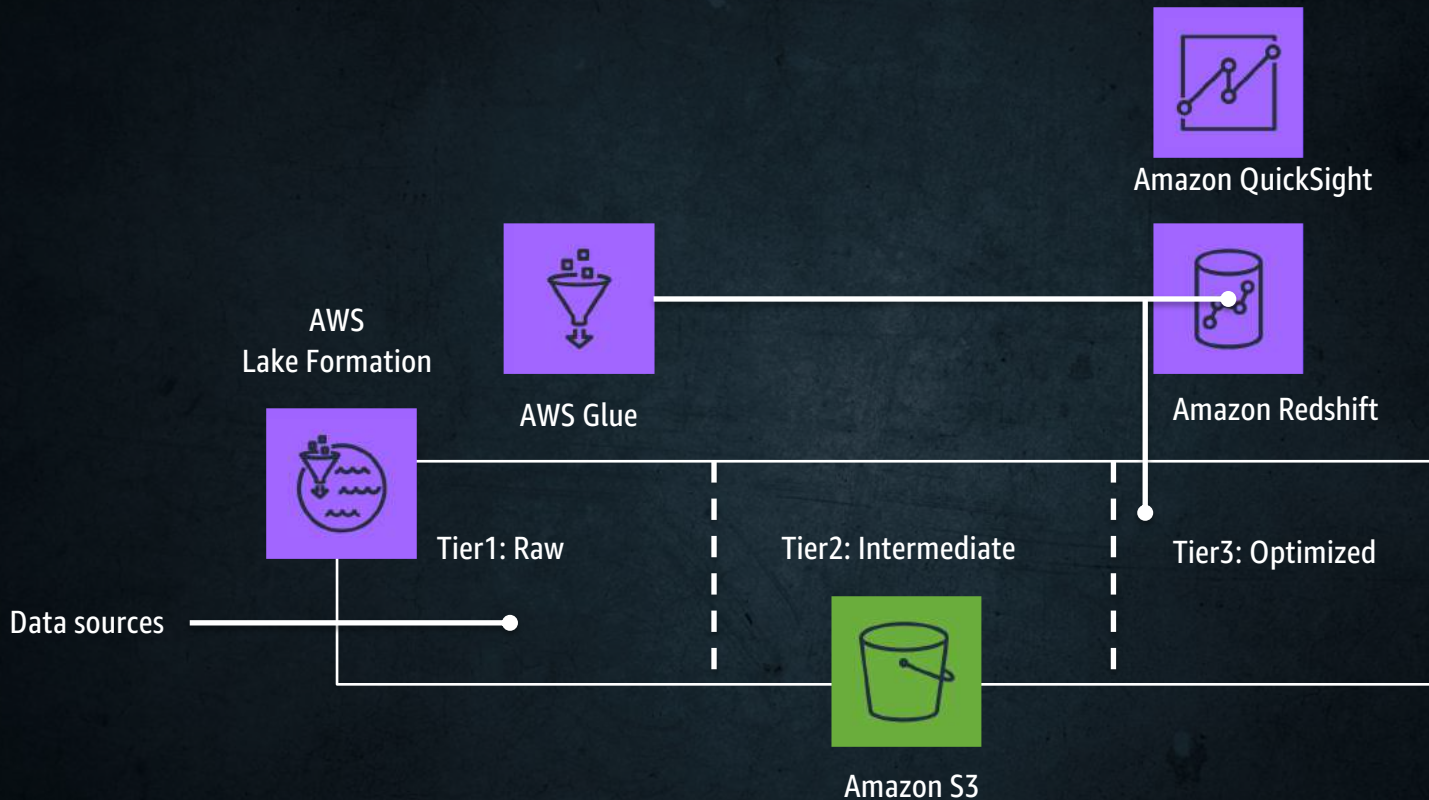


Amazon S3

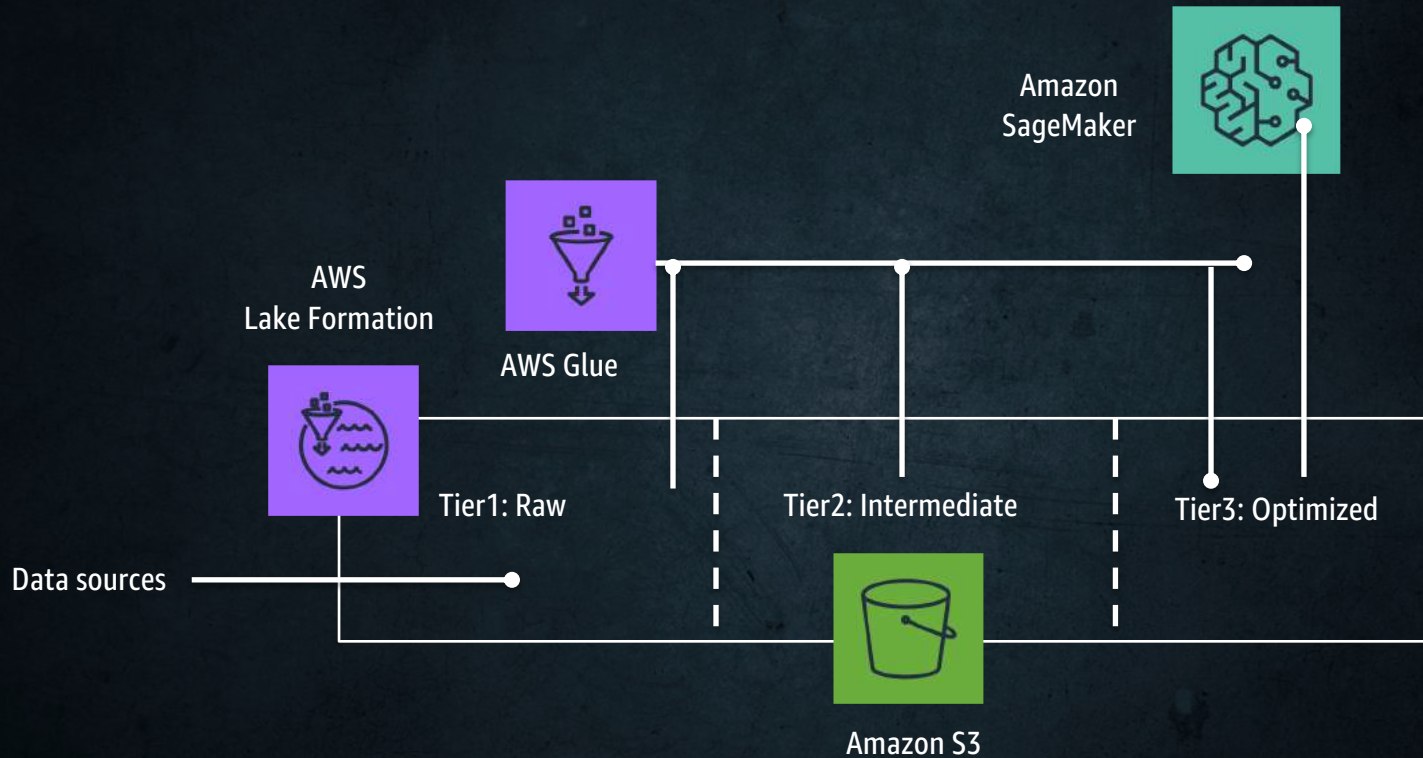
Data Processing



Traditional BI

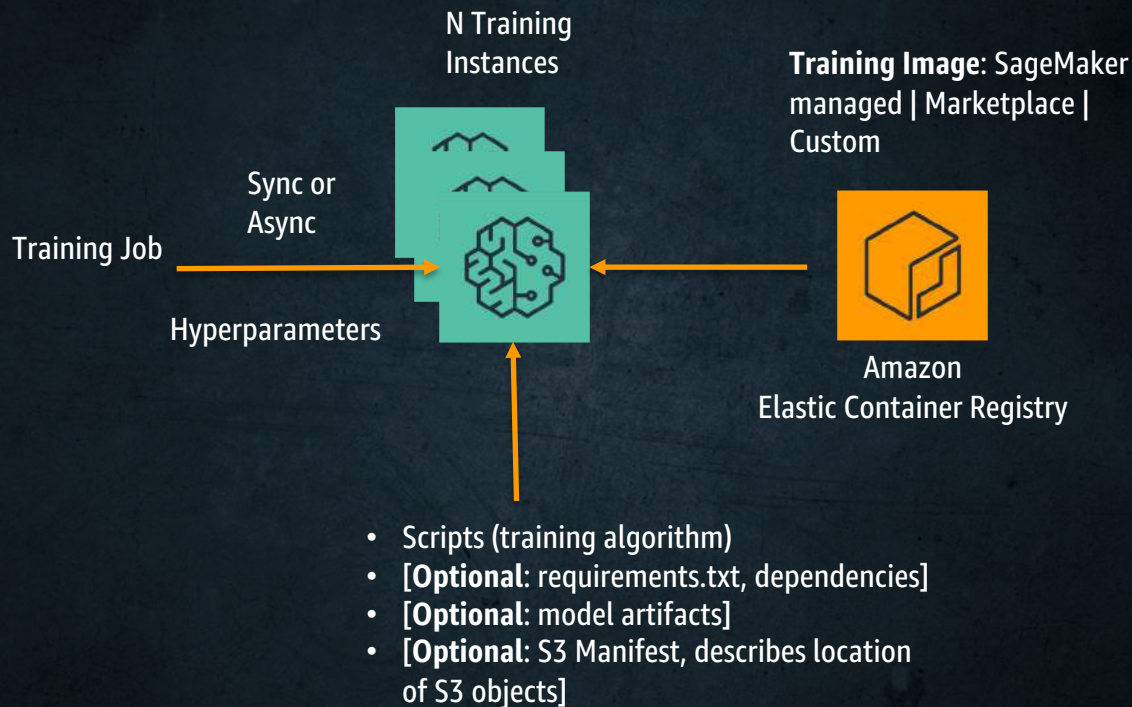


Machine Learning: Build and Training



Training Mechanics

SageMaker Training



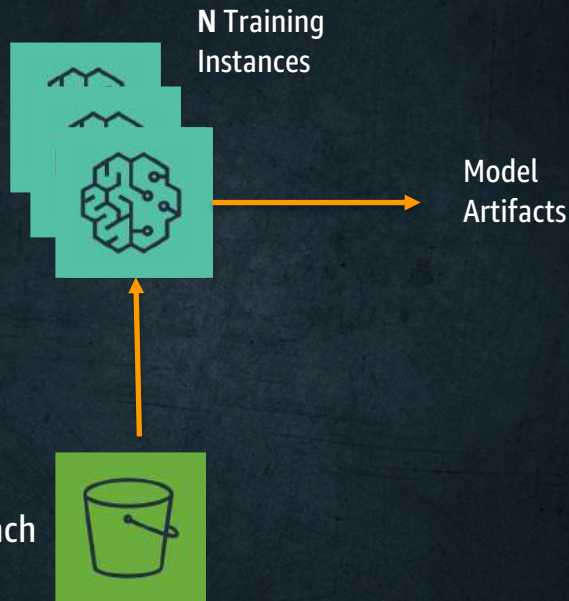
SageMaker Training

Modes:

- File: copy full dataset to local storage
- Pipe: stream directly from S3

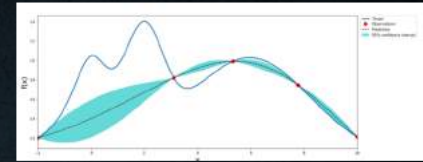
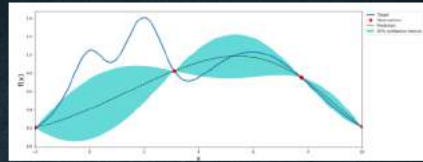
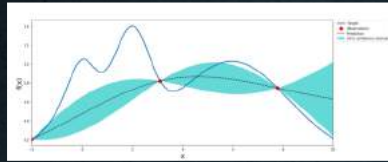
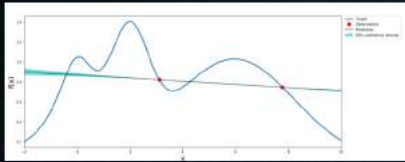
Distribution Type:

- FullyReplicated (full data set sent to each instance)
- ShardbyS3Key (1/N objects sent to each instance)



Hyperparameter Optimization

Bayesian Optimization

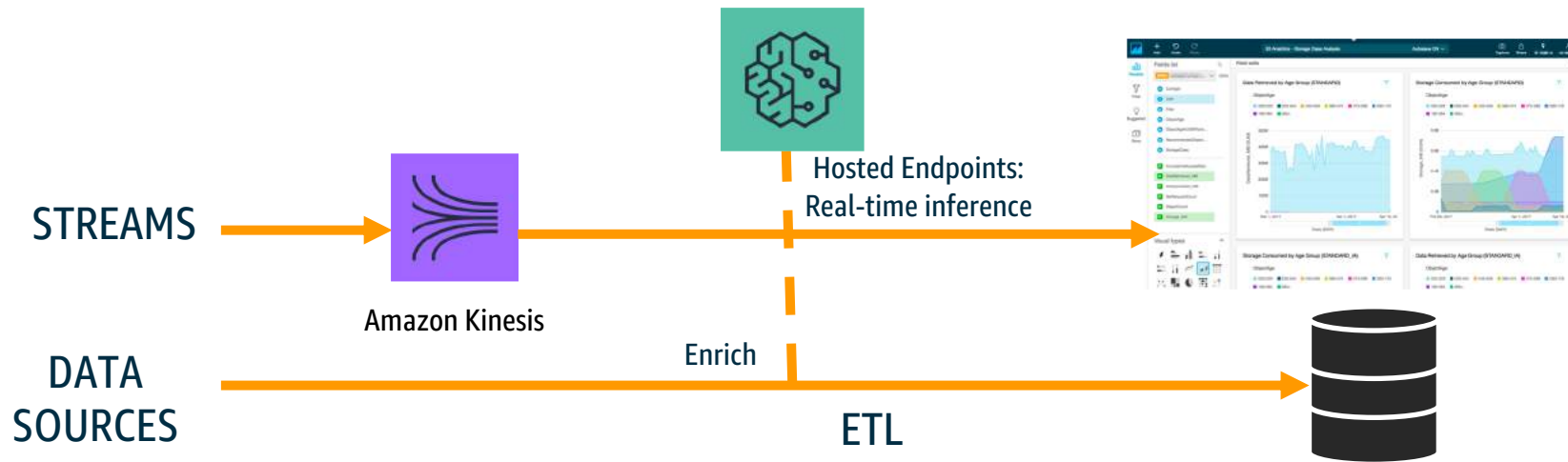


...

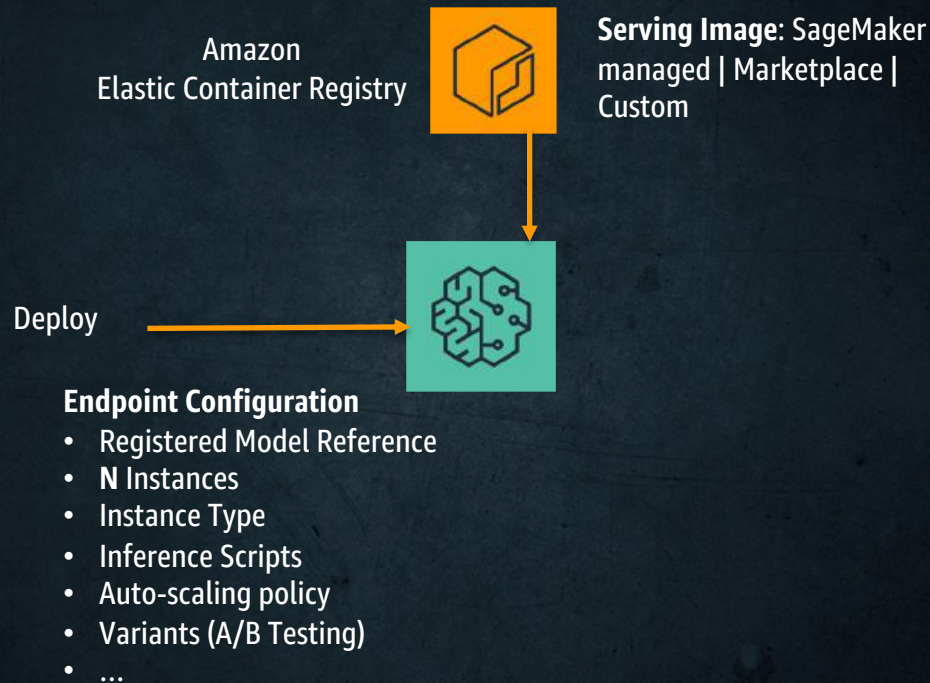


Inference Mechanics

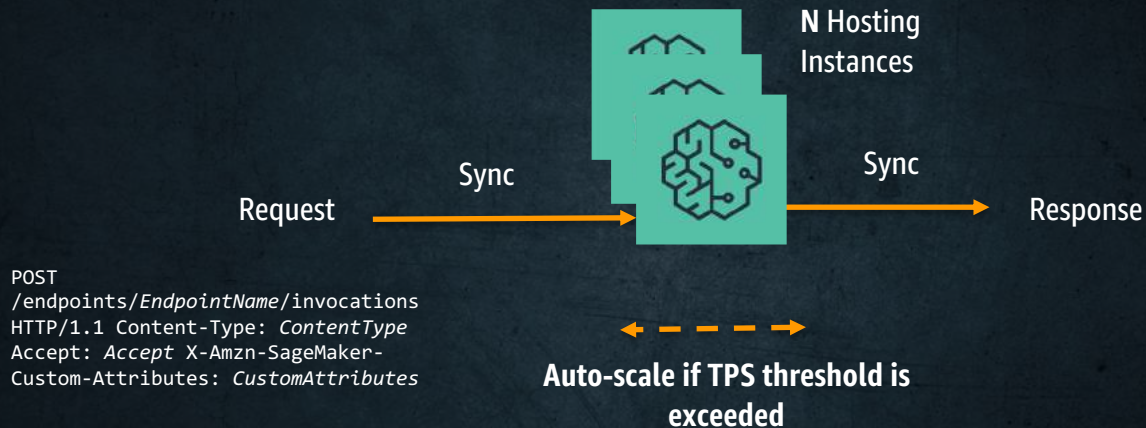
Advanced Analytics on AWS



Real-time Inference



Real-time Inference



Advanced Analytics on AWS



AWS Step Functions



DATA
SOURCES

ETL

DATA LAKE

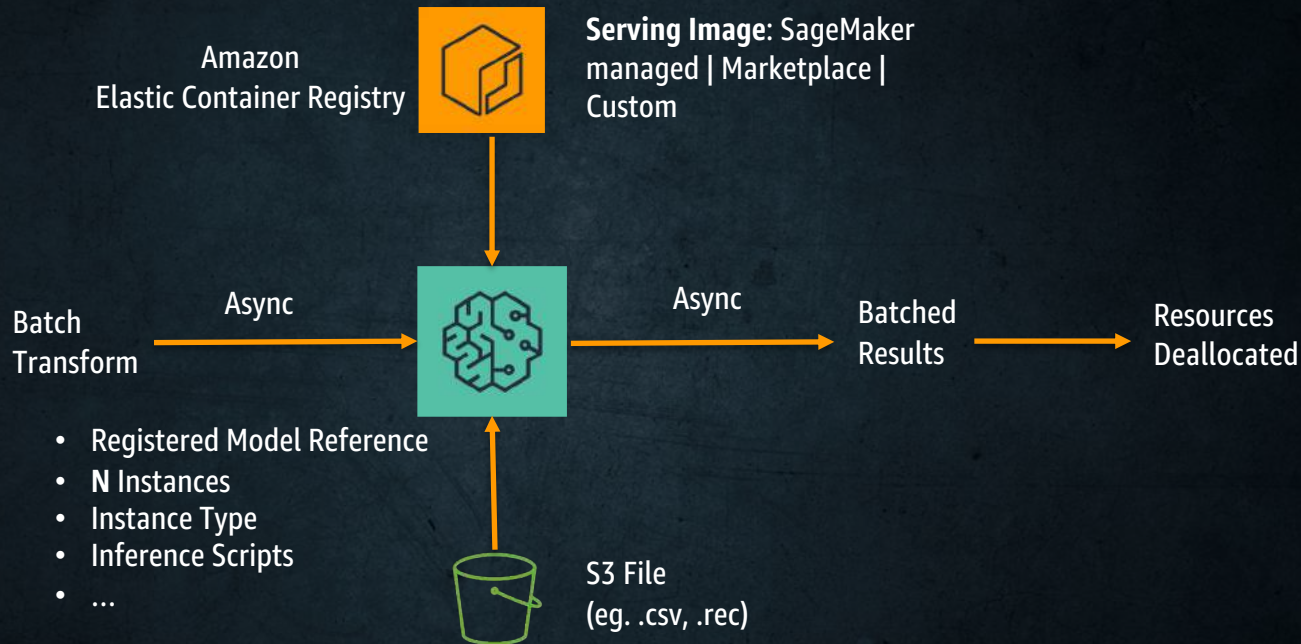


Batch Transform

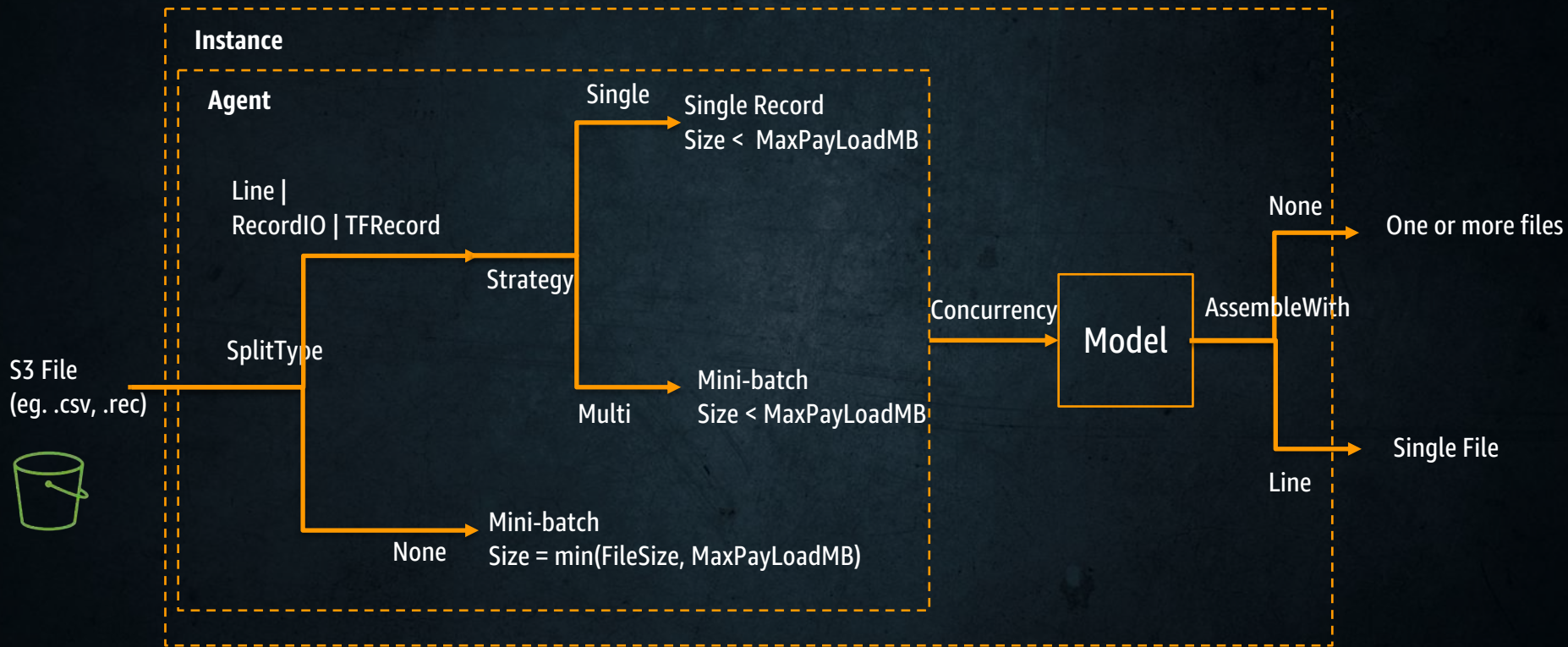
DATA LAKE



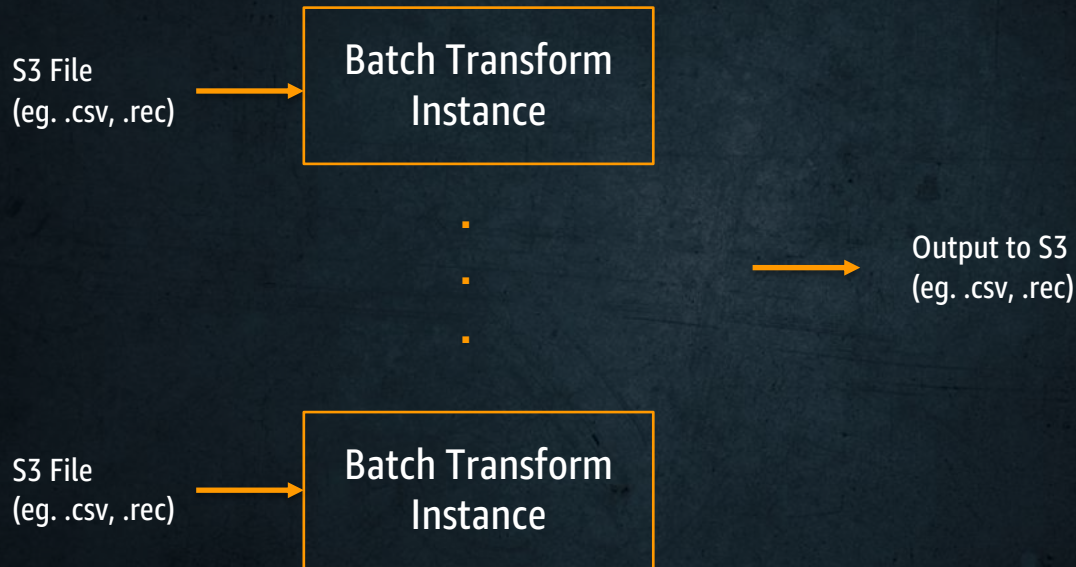
Batch Transform



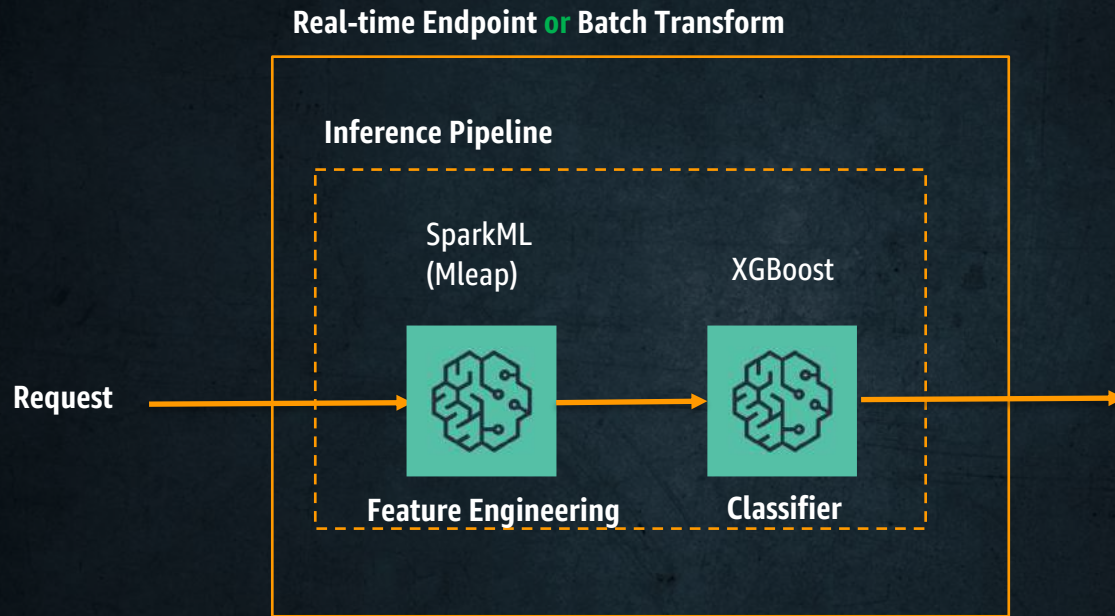
Batch Processing



Batch Processing



Inference Pipeline



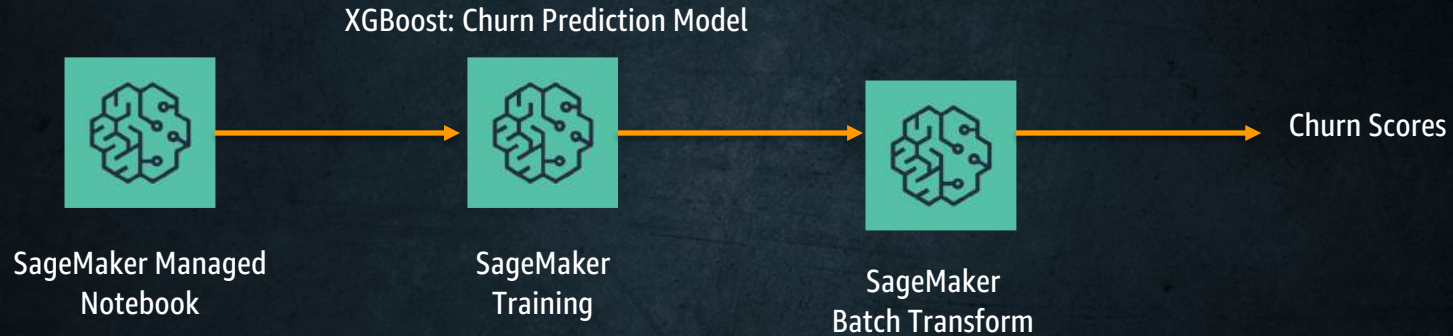
Example inference Pipeline: SparkML for feature engineering and XGBoost for predictions

Workshop Objectives and Agenda

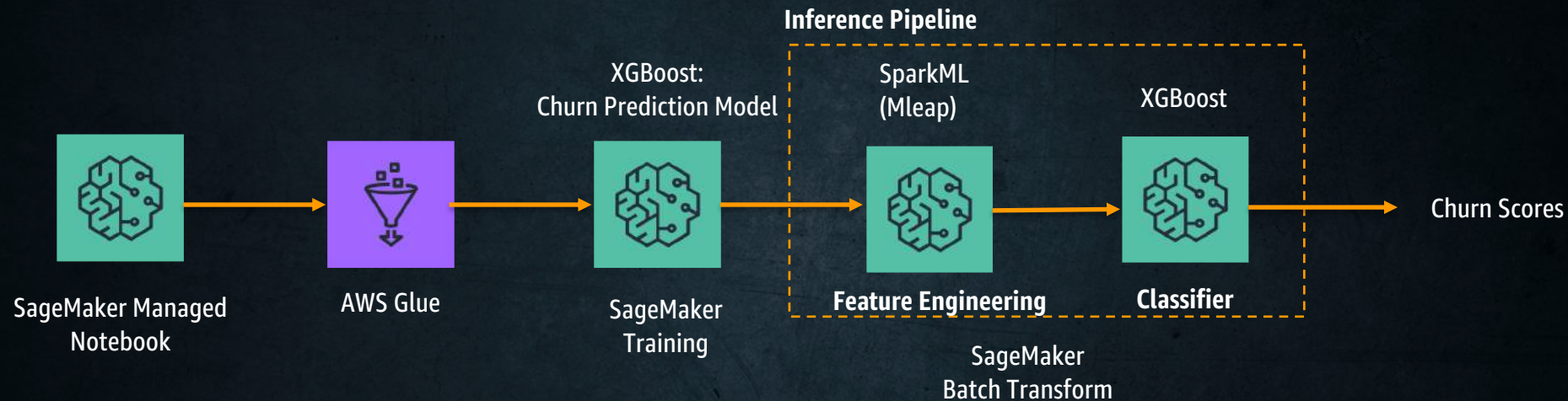
Explore Databases from SageMaker



Train (XGBoost) and Batch Inference



Scaling the Solution



<https://github.com/dylan-tong-aws/aws-advanced-analytics-jumpstarter>

Appendix

Future labs

- Send feedback: dylatong@amazon.com