
AWS
re:Invent

AIM404

Productionizing R workloads using Amazon SageMaker, featuring Siemens Energy

Michael Hsieh

Senior AI/ML Solutions Architect,

AWS

Edgar Bahilo Rodríguez

Lead Data Scientist,

Siemens Energy (Industrial Applications Division)

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

R on Amazon SageMaker

Demo

Customer Case Study : Siemens Energy and Digitalization, introducing SEER (Siemens Energy Eclectic Resource)



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

R on Amazon SageMaker



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker helps you build, train, and deploy models

SageMaker Ground Truth
Fully managed data labeling

SageMaker Processing
Built-in Python, BYO R/Spark

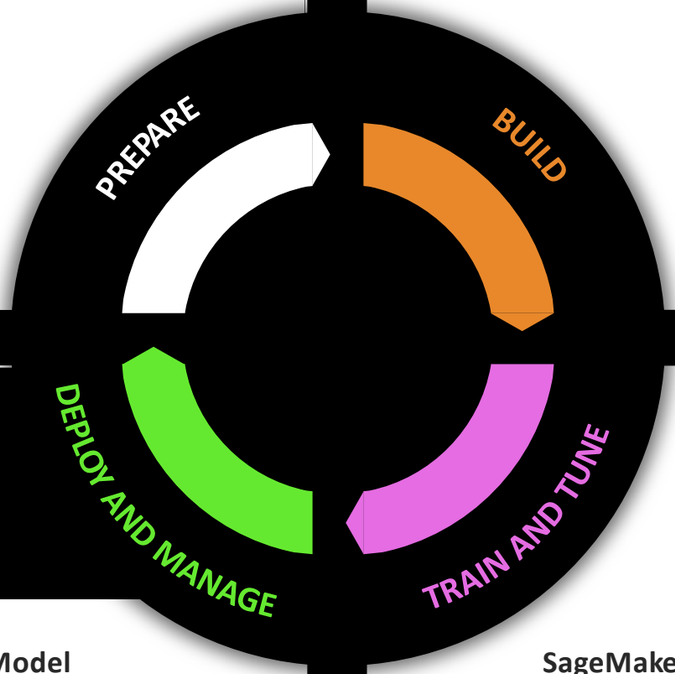
Inf1/Amazon Elastic Inference
High performance at lowest cost

Amazon Augmented AI
Add human review of model predictions

SageMaker Neo
Train once, deploy anywhere

One-click deployment
Supports real-time, batch, and multi-model

Model Monitor
Automatically detect concept drift



SageMaker Studio

SageMaker Studio Notebooks
One-click notebooks with elastic compute

Built-in and bring-your-own algorithms
Supervised and unsupervised algorithms

AWS Marketplace
Pre-built algorithms and models

SageMaker Autopilot
Automatically build and train models

One-click Training
Makes it easy to run training jobs

Automatic Model Tuning
One-click hyperparameter optimization

SageMaker Experiments
Capture, organize, and compare every step

SageMaker Debugger
Debug training runs

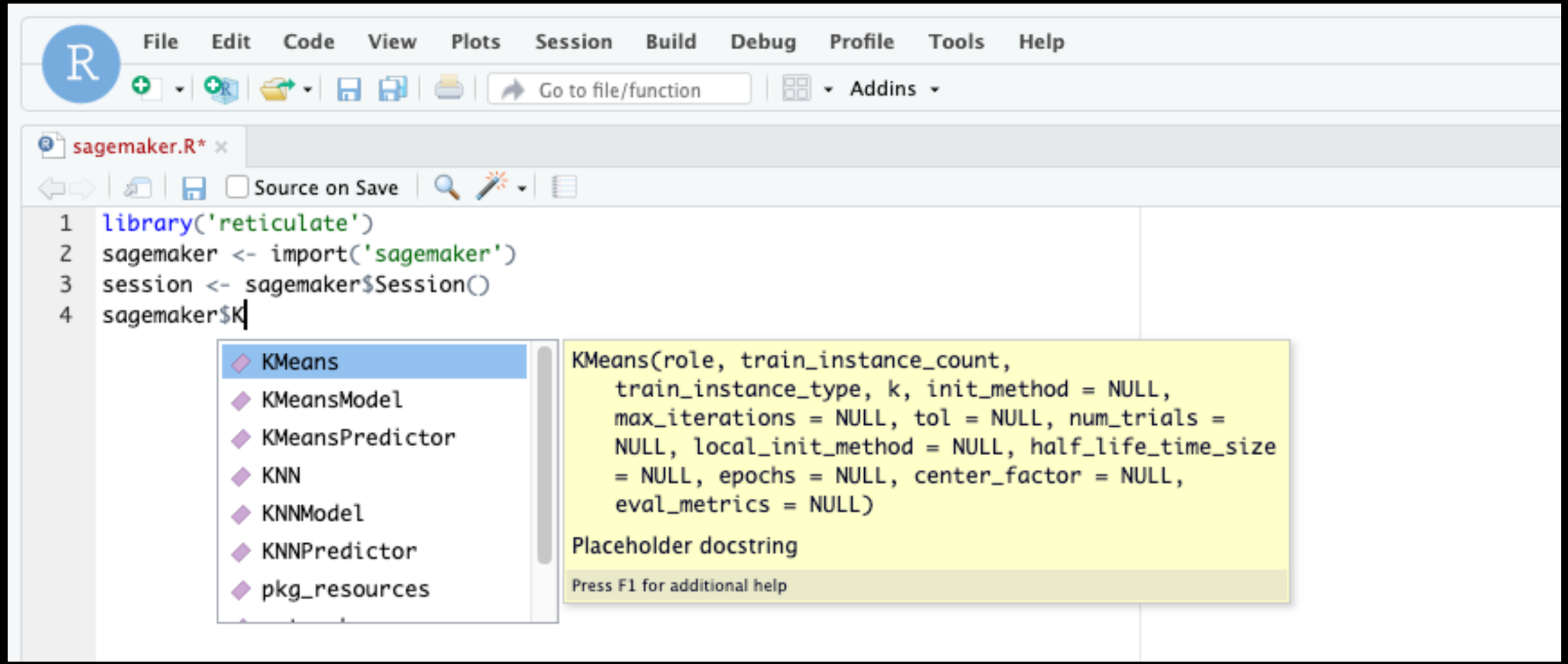
Managed spot training
Reduce training cost by 90%

Components outside of SageMaker Studio

Components included in SageMaker Studio

R on Amazon SageMaker

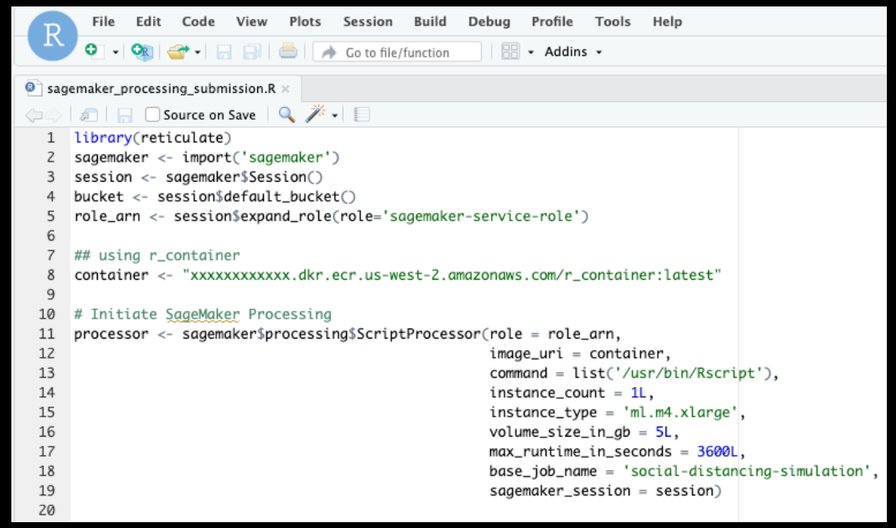
Data Scientists can use R in their preferred IDE, e.g. RStudio, to leverage Amazon's Machine Learning and Artificial Intelligence stack; full Amazon SageMaker capabilities can be accessed through *reticulate* library which is a R interface for the Python library, enabling machine learning practitioners to build, train, run, tune, and deploy R workloads and models.



3 ways of interacting with Amazon SageMaker

FROM YOUR IDE, NOTEBOOK, OR CLI

Your IDE, e.g. RStudio



SageMaker Notebook

```
In [ ]: s3_train_input <- sagemaker$s3_input(s3_data = s3_train,
                                             content_type = 'csv')
        s3_valid_input <- sagemaker$s3_input(s3_data = s3_valid,
                                             content_type = 'csv')
```

Training the model

Amazon SageMaker algorithm are available via a Docker container. To train an XGBoost model, specify the training containers in Amazon Elastic Container Registry (Amazon ECR) for the AWS Region.

```
In [ ]: registry <- sagemaker$amazon$amazon_estimator$registry(session$boto
        _region_name, algorithm='xgboost')
        container <- paste(registry, '/xgboost:latest', sep='')
        container
```

AWS CLI

```
# Fill in the values of these four variables
arn_role=<arn-of-your-notebook-role>
training_image=<training-image-for-region>
bucket=<name-of-your-s3-bucket>
region=<your-region>

prefix=/sagemaker/ml-forecasting-r
training_job_name=ml-forecasting-r-`date '+%Y-%m-%d-%H-%M-%S'`

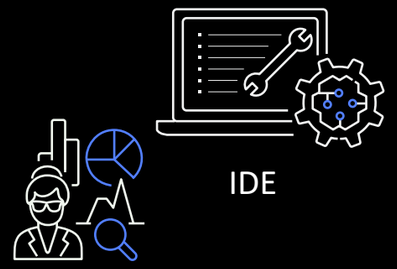
training_data=$bucket$prefix/train
eval_data=$bucket$prefix/validation
train_source={S3DataSource={S3DataType=S3Prefix,S3DataDistributionType=FullyReplicated,S3Uri=$training_data}}
eval_source={S3DataSource={S3DataType=S3Prefix,S3DataDistributionType=FullyReplicated,S3Uri=$eval_data}}

aws --region $region \
sagemaker create-training-job \
--role-arn $arn_role \
--training-job-name $training_job_name \
--algorithm-specification TrainingImage=$training_image,TrainingInputMode=File \
--resource-config InstanceCount=1,InstanceType=ml.c4.2xlarge,VolumeSizeInGB=10 \
--input-data-config ChannelName=train,DataSource=$train_source ChannelName=validation,DataSource=$eval_source \
--output-data-config S3OutputPath=$bucket$prefix \
--hyper-parameters learning_rate=0.0001,batch_size=128,num_round=100 \
--stopping-condition MaxRuntimeInSeconds=1800
```

Scale with Amazon SageMaker

WHILE PROTOTYPING FROM YOUR IDE

Machine learning



Data preprocessing



Amazon SageMaker Processing

Model training / tuning



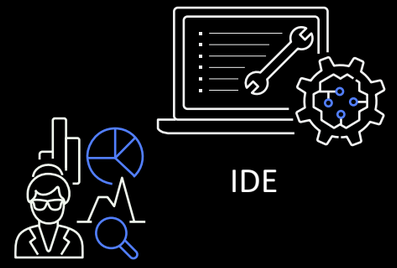
Amazon SageMaker Training / Tuning

Model hosting



Amazon SageMaker Endpoint / Batch Transform

Statistical analysis / simulation / optimization



Data preprocessing



Amazon SageMaker Processing

Analysis / simulation / optimization



Amazon SageMaker Processing

Analysis results



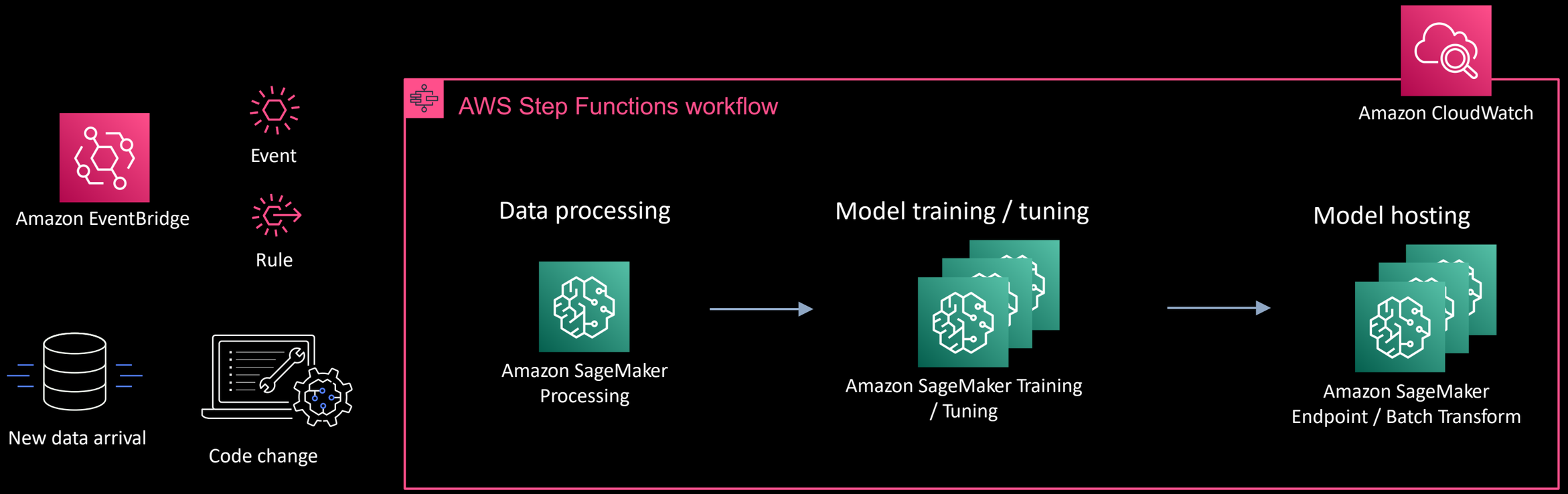
Amazon S3



Databases

Automation

AUTOMATE WITH A WORKFLOW THAT IS EVENT DRIVEN (DATA CHANGE, SCHEDULED EVENT)



RStudio Server | Amazon SageMaker

File Edit Code View Plots Session Build Debug Profile Tools Help

ubuntu

Go to file/function | Addins | Project: (None)

fable_demo.r | sagemaker_training.r | fable_sagemaker.r | build_and_push_docker.sh | Dockerfile

Source on Save | Run | Source

```
## Dependencies
## install.packages(c('readr', 'curl', 'ggplot2', 'dplyr', 'stringr', 'fable', 'tsibble', 'dplyr
library(fable)
library(tsibble)
library(dplyr)

## Setting parameters
city <- 'Melbourne'
ets_trend_method <- 'A' # additive
ic <- 'aic' # use aic as information criteria to select model

## Getting data
tourism_city <- tourism %>%
  filter(Region == city)
tourism

## Exploration
# Purpose: Business, Holiday, Visiting friends and family, or Others
tourism_city %>%
  autoplot(Trips)
```

3:1 (Top Level) R Script

Console | Terminal | Jobs

~/data-science-with-r-sagemaker.git/forecasting/

```
>
```

Environment | History | Connections | Tutorial

Import Dataset | List

Global Environment

Environment is empty

Files | Plots | Packages | Help | Viewer

Zoom | Export

Siemens Energy and digitalization



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Siemens Energy in a nutshell

OUR PORTFOLIO

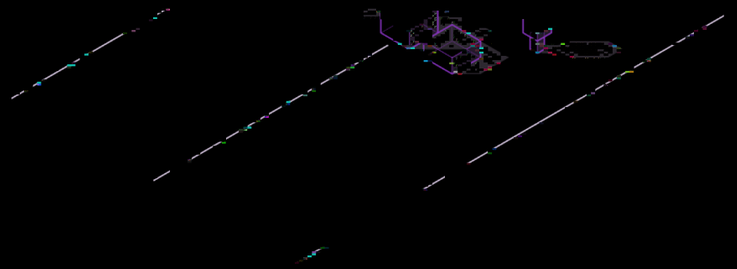
TRANSMISSION

RENEWABLES

GENERATION



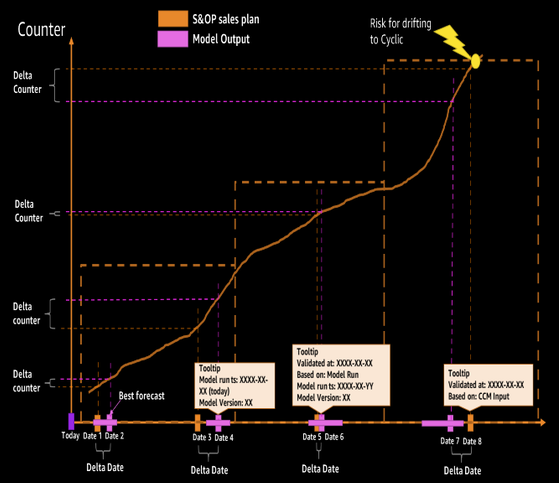
INDUSTRIAL APPLICATIONS



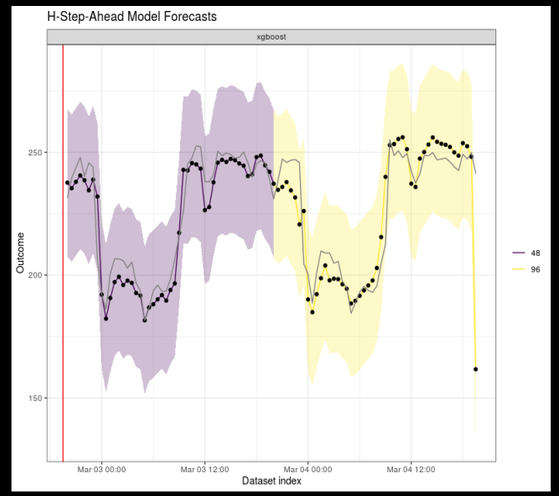
NEW ENERGY BUSINESS

Main data science and analytics use cases

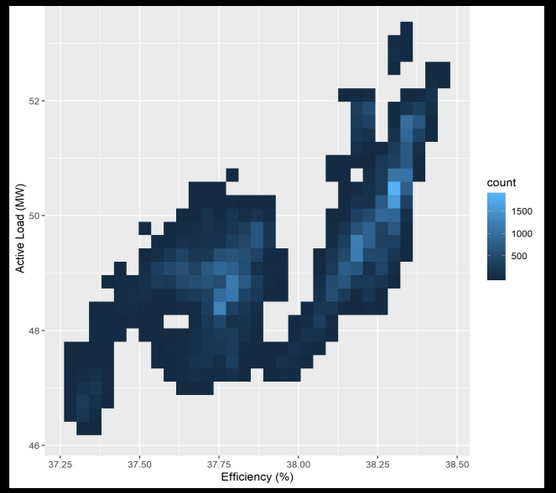
S&OP prognosis engine



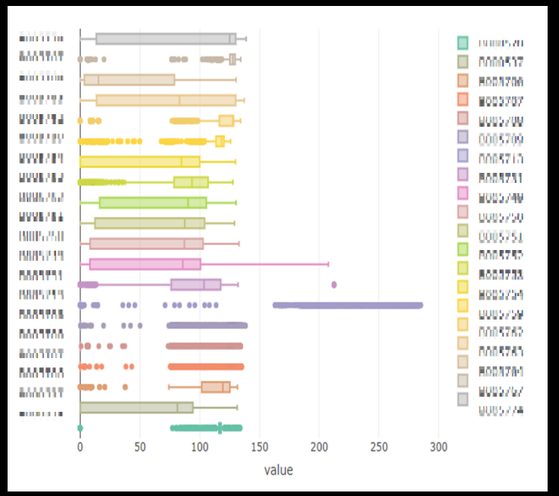
Dispatch optimizer



Analytics driven sales



Probabilistic design



Supporting the business cases with the right tools will accelerate delivery time and enable new use cases

Detailed use case : S&OP prognosis engine

Systematic data collection

Fleet counter data (EOH, EOC, Starts, and OH)

Different data sources (STA-RMS field services, customer portal)

Life calculation

Consumed life:

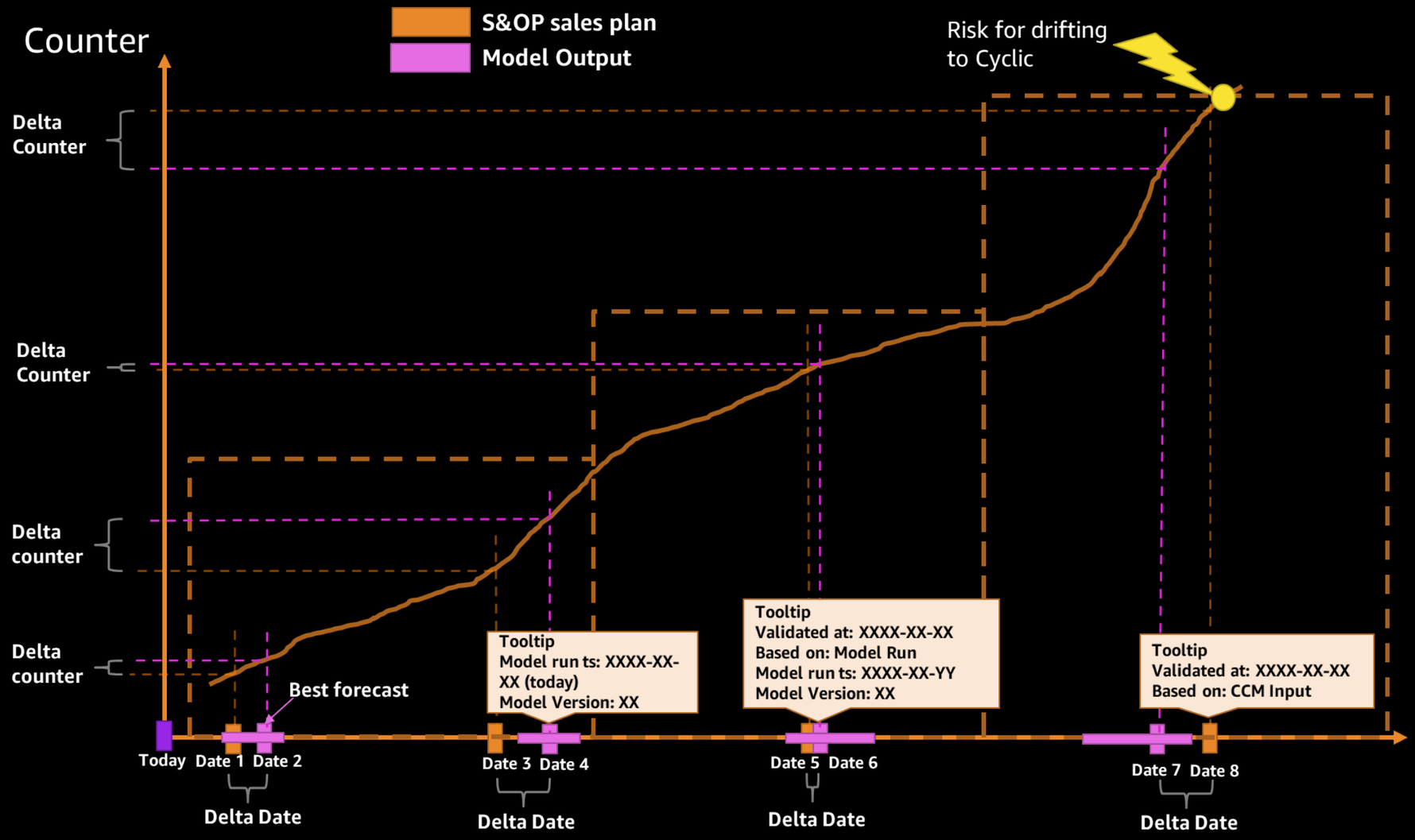
- Package level
- Assembly level (GG / PT / Core engine calculation)
- Component level calculation

Forecasting engine

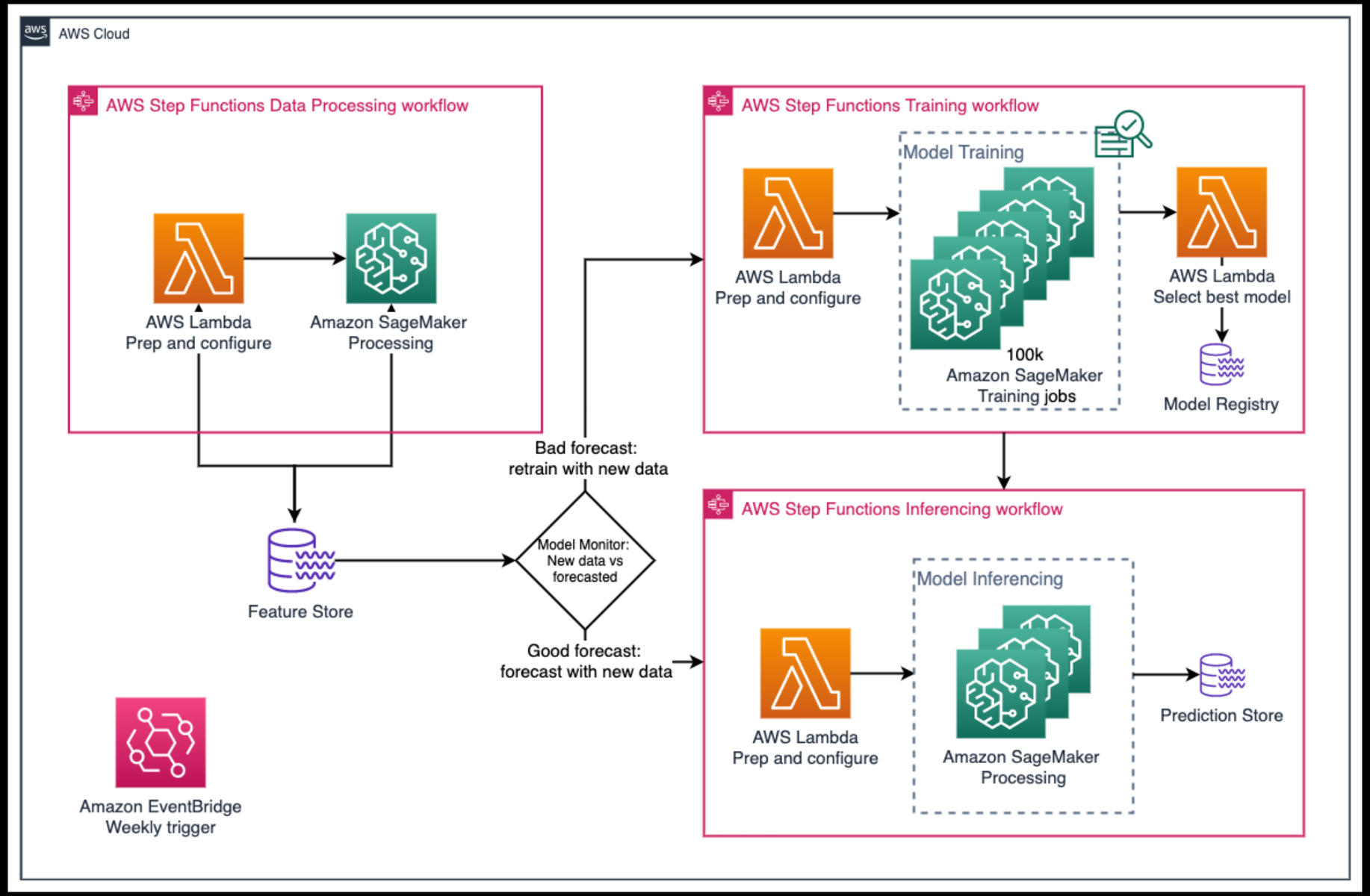
Processing 4,000+ time series

Training 22,000+ statistical and machine learning models

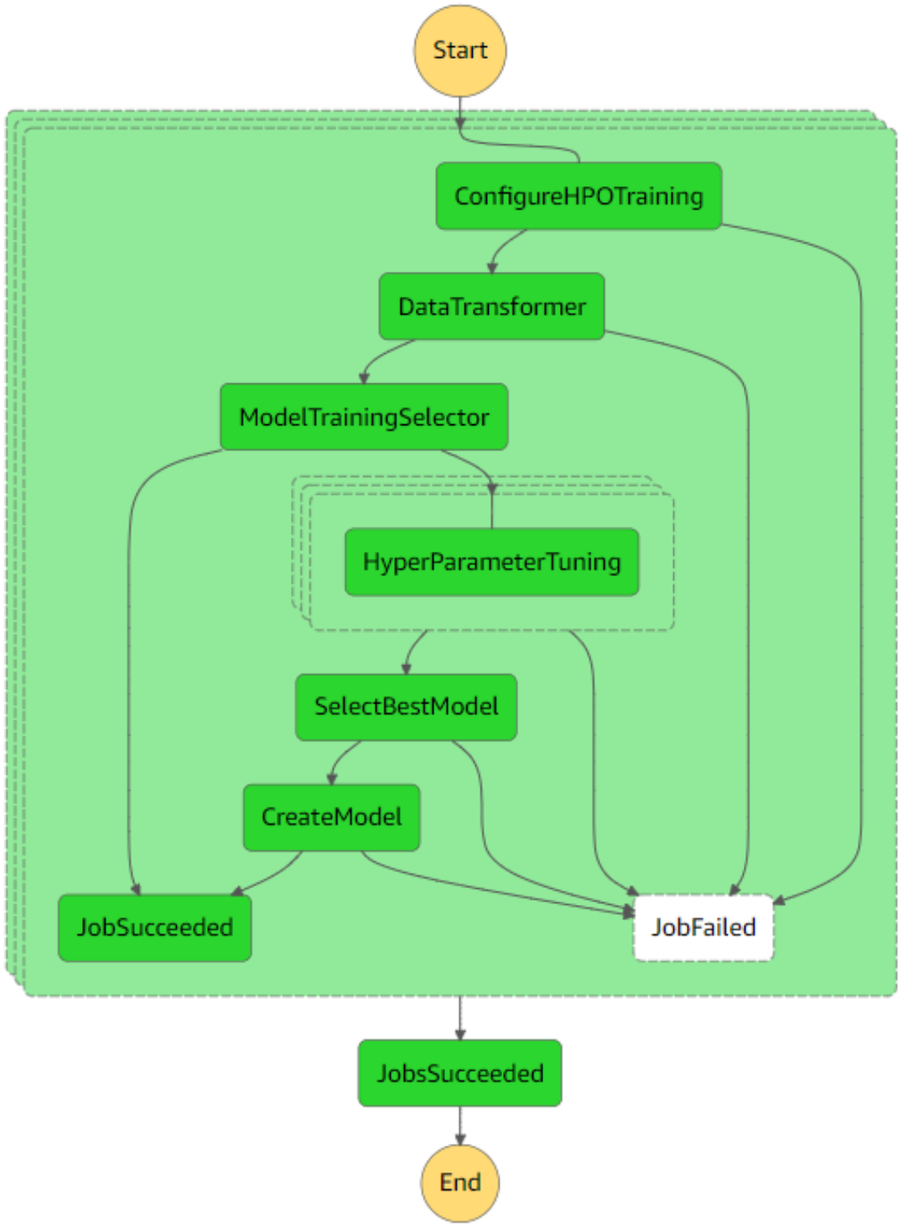
Predicting fleet counter data for different horizons (6M to 5Y)



Automating end-to-end ML workflows



Detailed training workflow



What to do when you have multiple algorithms

What to do when you have different hyperparameters

What to do when you are training multiple models for multiple physical assets

How to select the best of the best

AWS Step Functions

AWS Lambda

Amazon SageMaker

Integrating SageMaker in your training script

Setup

```
option_list = list(
  make_option(c("-i", "--input"),
    type="character",
    default=Sys.getenv("SM_CHANNEL_TRAINING"),
    help=”Input data dir",
    metavar="character"),
  make_option(c("-o", "--outputfolder"),
    type="character",
    default=Sys.getenv("SM_MODEL_DIR"),
    help=”Model binary output",
    metavar="character"),
  make_option(c("-hp", "--hyperparameters"),
    type="character",
    default=Sys.getenv("SM_HPS"),
    help="Hyperparameters in JSON format",
    metavar="character"),
  make_option(c("-c", "--cores"),
    type="integer",
    default=Sys.getenv("SM_NUM_CPUS"),
    help=”CPU count", metavar="number")
);
```

Enabling HPO

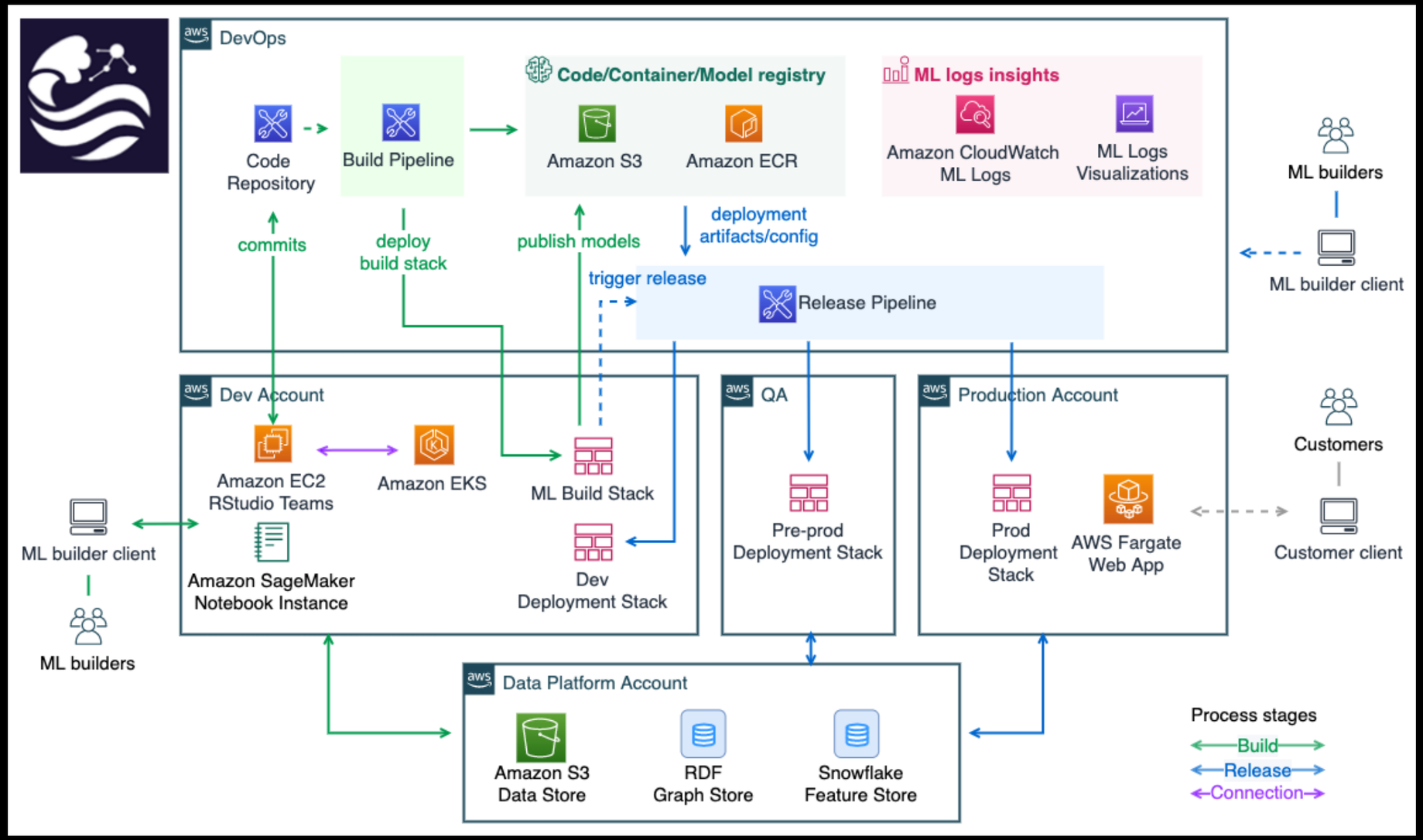
```
accuracy_explorer <- function(mode, models,... testing_data=NULL) {
  if(mode == "train") {
    ....
  } else if(mode=="test"){
    forecast_data <- forecast(models, testing_data)
    accuracy_metrics <- accuracy(forecast_data, testing_data) %>%
      select(..., RMSE, MAE, MAPE, MASE)) %>%
      mutate(horizon = max_forecast_horizon)
    print(paste0("MASE:", accuracy_metrics$MASE))
}
```

Model Training

```
trainModels <- function(input_data, model_type, ...) {
  ...
    if(model_type == "theta") {
      ts_models <- input_data %>%
                   model(theta_model = THETA(snsr_val_clean ~ season()))
      } else if(model_type == "ensemble") {
        ts_models <- input_data %>% model(ensemble =
          combination_model(ARIMA(snsr_val_clean),
          ARIMA(snsr_val_clean ~
                fourier(argument_parser$hyperparameters$season1,
                        K=argument_parser$hyperparameters$K)),
          fable.prophet::prophet(snsr_val_clean ~
                season(argument_parser$hyperparameters$season1,
                       K=argument_parser$hyperparameters$K,
                       type = argument_parser$hyperparameters$mode)
        ...
        }}
```

AI industrialization, introducing SEER

SIEMENS ENERGY ECLECTIC RESOURCE



Additional resources

PLEASE VISIT THESE SESSIONS AND PAGES FOR MORE ON AMAZON SAGEMAKER

AIM309 Secure and compliant machine learning for regulated industries

AIM314 Train and tune ML models to the highest accuracy using Amazon SageMaker

AIM406 Implementing MLOps practices with Amazon SageMaker

R User Guide to Amazon SageMaker
https://docs.aws.amazon.com/sagemaker/latest/dg/r-guide.html

Hands-on examples: R on Amazon SageMaker
https://github.com/aws/amazon-sagemaker-examples/tree/master/r_examples

Using R with Amazon SageMaker
https://aws.amazon.com/blogs/machine-learning/using-r-with-amazon-sagemaker/

Coding with R on Amazon SageMaker notebook instances
https://aws.amazon.com/blogs/machine-learning/coding-with-r-on-amazon-sagemaker-notebook-instances/

Creating a persistent custom R environment for Amazon SageMaker Notebooks
https://aws.amazon.com/blogs/machine-learning/creating-a-persistent-custom-r-environment-for-amazon-sagemaker/

Thank you!



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.