



Amazon SageMaker e MLOps

Renato Barbosa
AI/ML - Business Development Manager
Amazon Web Services

Qual é a maior dificuldade de machine learning (ML)?

“A parte mais difícil de ML não é o ML, mas a massiva quantidade de esforços para manter sistemas de ML. É fácil de ficar dependente e difícil de sustentar.

- Anthony Penta, Sr. Manager & Principal Scientist, Amazon Consumer Payments

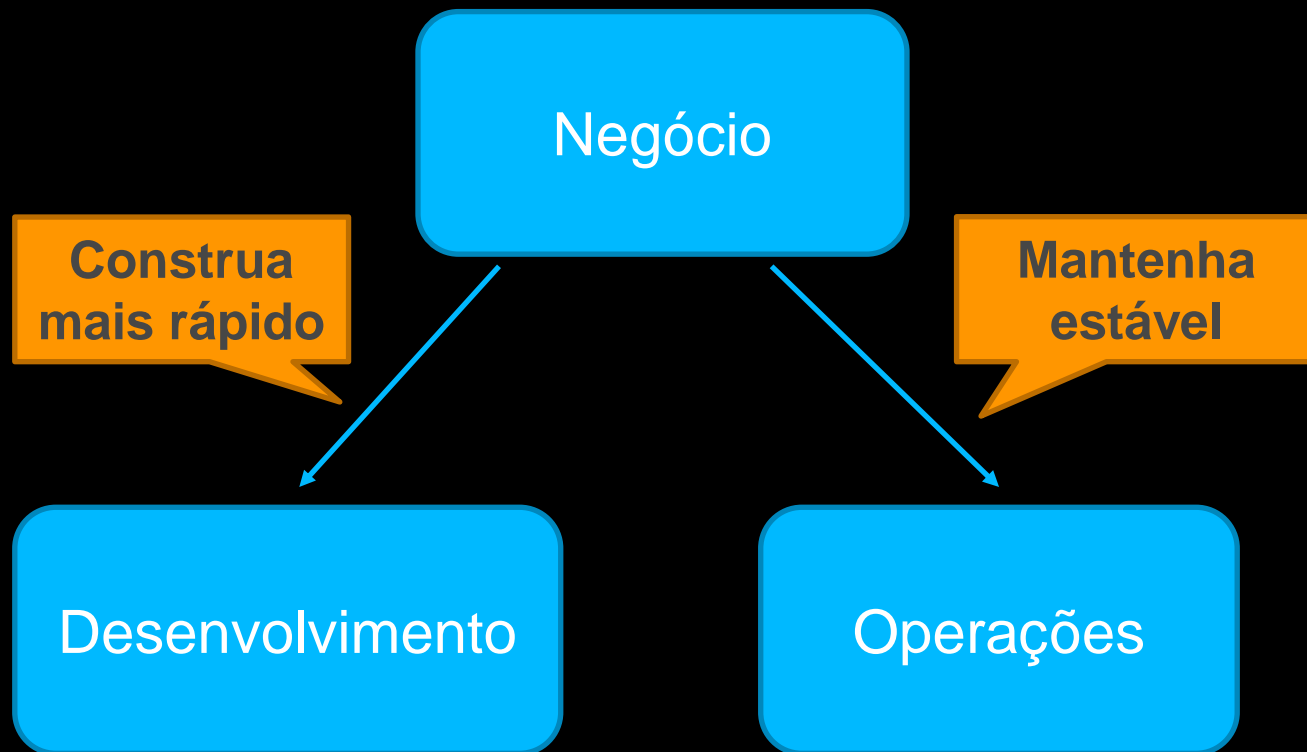
Hidden Technical Debt in Machine Learning Systems

<https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf>

Por que estamos aqui hoje?

Forças concorrentes:

A mudança é a causa raiz da maioria das interrupções

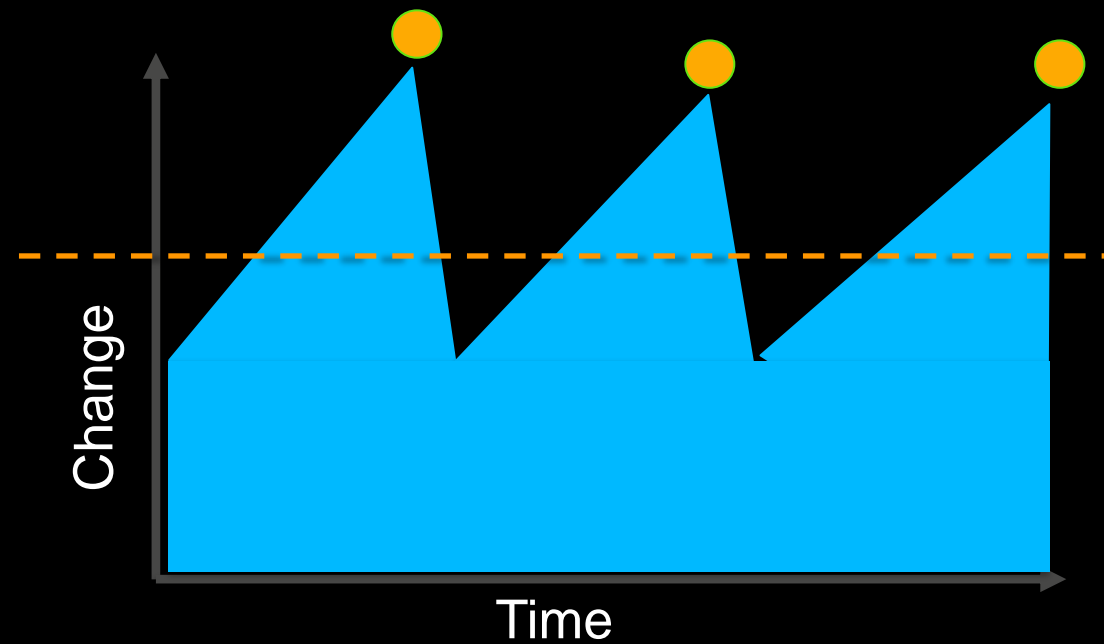


Duas opções:

1. Faça disso um grande negócio
2. Não-eventos iterativos pequenos

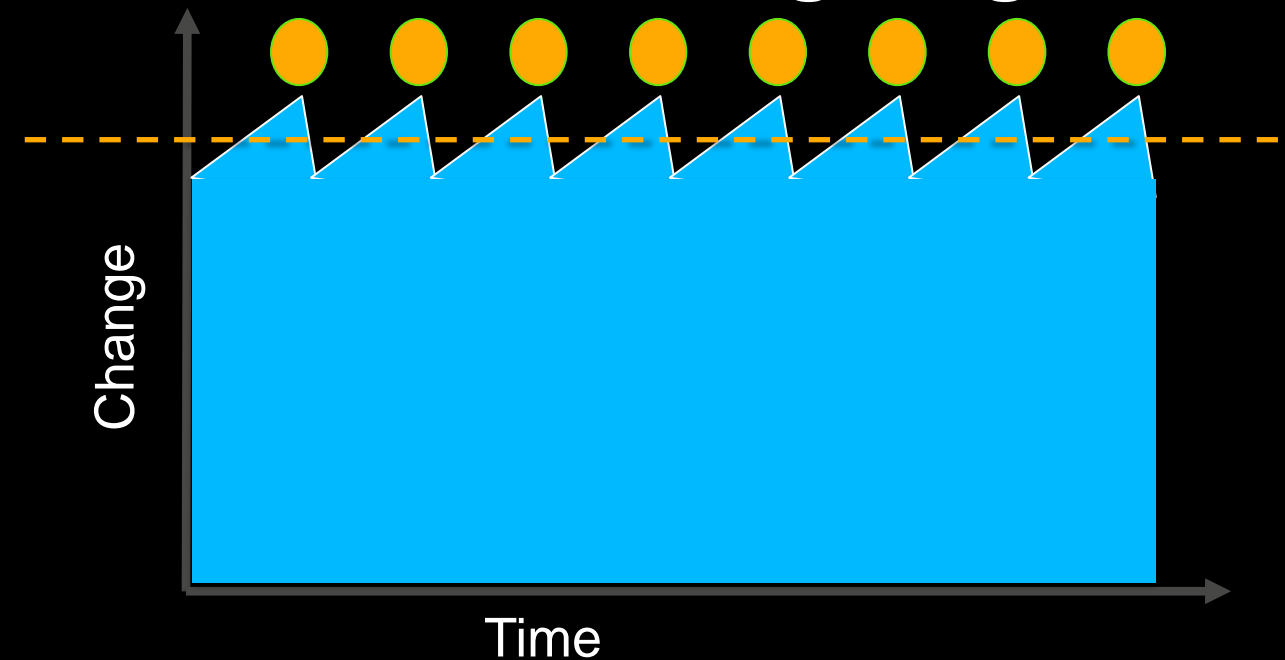
A implantação com mais frequência reduz o risco

Eventos de lançamento raros:
“Metodologia em cascata”



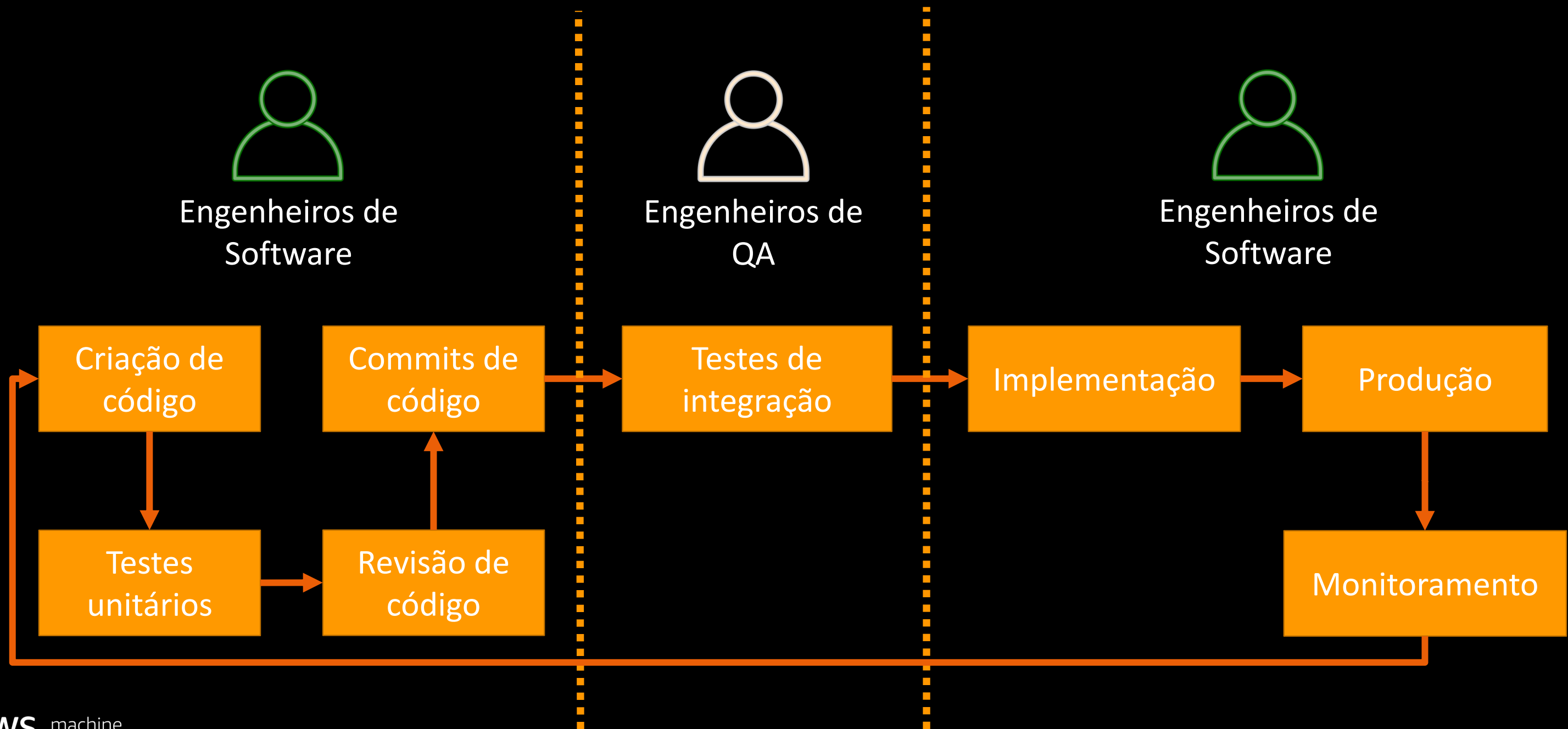
Esforço maior
"Risco aumentado"

Eventos de lançamento frequentes:
“Metodologia ágil”

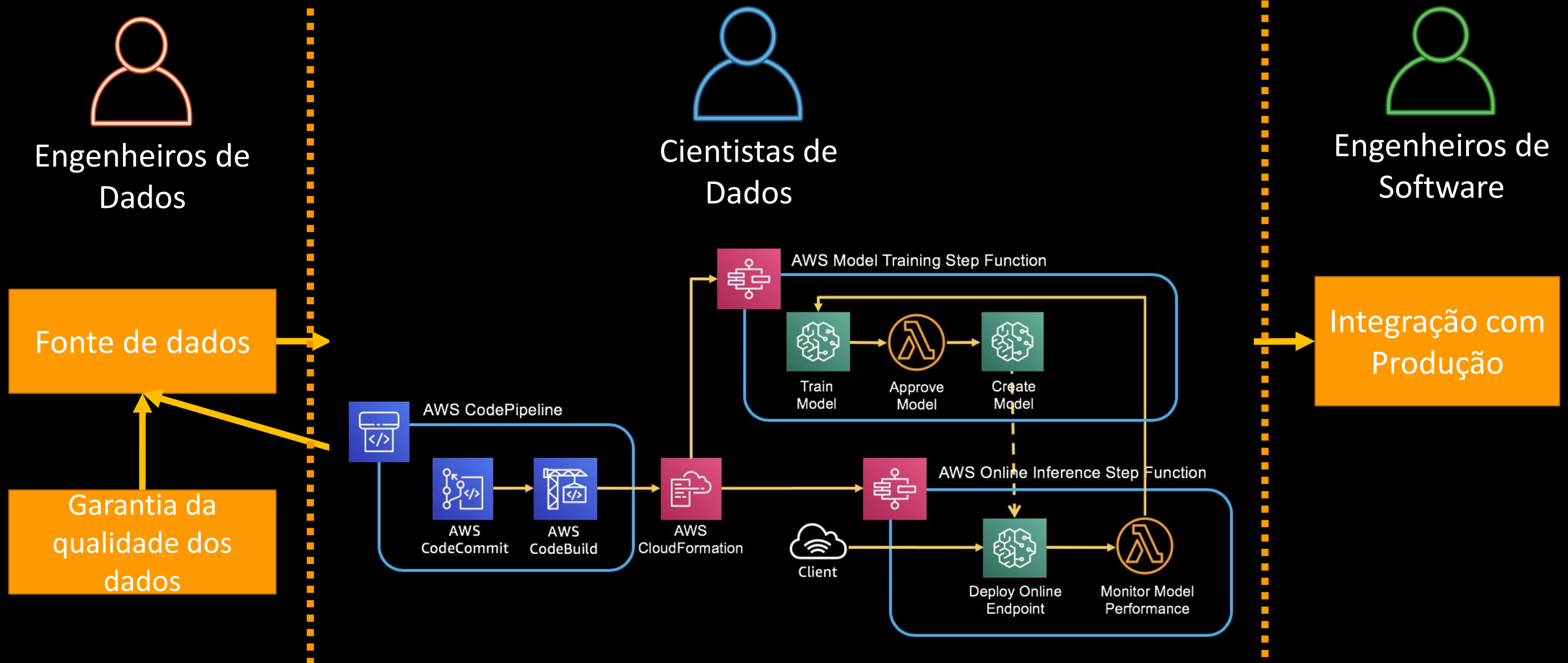


Esforço menor
"Risco minimizado"

Ciclo de Vida de Desenvolvimento de Software (SDLC)



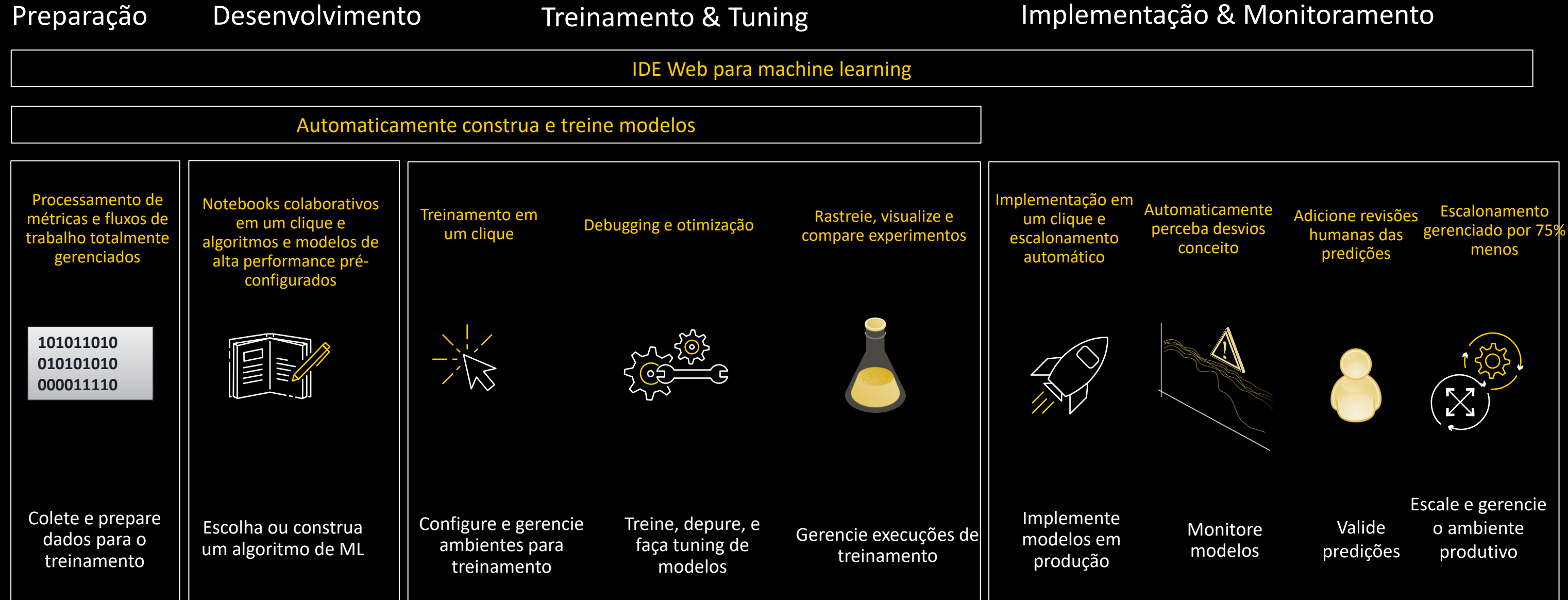
Ciclo de Vida de Desenvolvimento de Modelos (MDLC)



Amazon SageMaker

Amazon SageMaker:

Construa, treine, implante e opere modelos de ML em escala



Use o Amazon SageMaker Studio para atualizar modelos e ver o impacto na qualidade do modelo imediatamente

The screenshot displays the Amazon SageMaker Studio interface. On the left, a Jupyter notebook titled 'xgboost_customer_churn.ipynb' is open, showing Python code for data preprocessing and splitting. The code includes comments in Portuguese and instructions on how to convert categorical features to numeric and split the data into training, validation, and test sets. The right side of the interface features two panels: 'Trial Component Chart' and 'Trial Component List'.

Trial Component Chart

The chart shows 'trainloss_last' on the y-axis (ranging from 0.0 to 0.4) against 'period' on the x-axis (ranging from 0 to 6). Multiple colored lines represent different trials, showing a general downward trend in loss over time.

Trial Component List

The list displays 10 rows of trial components. The first four rows are highlighted, showing a status of 'Completed' for training jobs.

Status	Experiment	Type	Trial	Trial c
✓ Completed	customer-churn-predi...	Training job	Trial-3	Tra
✓ Completed	customer-churn-predi...	Training job	Trial-2	Tra
✓ Completed	customer-churn-predi...	Training job	Trial-1	Tra
✓ Completed	customer-churn-predi...	Training job	Trial-0	Tra

At the bottom of the interface, the status bar indicates 'Mode: Command', 'Ln 1, Col 1', and the file name 'xgboost_customer_churn.ipynb'.

Use os Notebooks do Amazon SageMaker para compartilhar facilmente seu trabalho com colegas e rapidamente provisione recursos computacionais

The screenshot displays the Amazon SageMaker Studio interface. On the left, a file explorer shows a directory structure with files like 'build', 'mnist', 'sagemaker-experiments', 'sagemaker-python-sdk-staging', 'src', '12345.ipynb', and 'AutoML Interactive Execution Not...'. The main area shows a code editor with Python code for linear algebra and data processing. A 'Select instance' dialog box is open, showing the current instance 'ml.t3.medium' and options to select a new instance from 'Standard (3/14)', 'Compute optimized (1/8)', or 'GPU instances (1/9)' categories. The 'Standard (3/14)' category is selected, showing options like '2 vCPU + 1 GiB (system default)', '2 vCPU + 4 GiB (ml.t3.medium)', and '2 vCPU + 8 GiB + 8 ECU (ml.m5.large)'. The 'Compute optimized (1/8)' category shows '4 vCPU + 8 GiB + 17 ECU (ml.c5.xlarge)'. The 'GPU instances (1/9)' category shows '4 vCPU + 16 GiB + 125 GB NVMe SSD (ml.g4dn.xlarge)'. A 'Create shareable snapshot' dialog box is also open, showing options to share the snapshot as 'my_notebook.ipynb' and to include Git repo information and notebook output.

Select instance

Current instance - ml.t3.medium Fast startup

vCPU	Memory	ECU	Instance Storage	Network Performance
2	4 GiB	Variable	EBS only	Low-Moderate

Select new instance

☒ Show fast startup only

Standard (3/14)

- 2 vCPU + 1 GiB (system default) Fast startup
- 2 vCPU + 4 GiB (ml.t3.medium) Fast startup ✓
- 2 vCPU + 8 GiB + 8 ECU (ml.m5.large) Fast startup

Compute optimized (1/8)

- 4 vCPU + 8 GiB + 17 ECU (ml.c5.xlarge) Fast startup

GPU instances (1/9)

- 4 vCPU + 16 GiB + 125 GB NVMe SSD (ml.g4dn.xlarge) Fast startup

Create shareable snapshot

Share as

my_notebook.ipynb

☒ Include Git repo information

☐ Include notebook output

Cancel Create snapshot

Use algoritmos pré-configurados do Amazon SageMaker ou traga seus próprios

Classificação

- Linear Learner
- XGBoost
- KNN

Visão Computacional

- Classificação de Imagens
- Detecção de objetos
- Segmentação semântica

Modelagem de tópicos

- LDA
- NTM

Texto

- BlazingText
- Supervisionado
- Não-supervisionado

Recomendação

- Factorization Machines

Predição

- DeepAR

Tradução de Sequências

- Seq2Seq

Detecção de anomalias

- Random Cut Forests
- IP Insights

Clustering

- KMeans

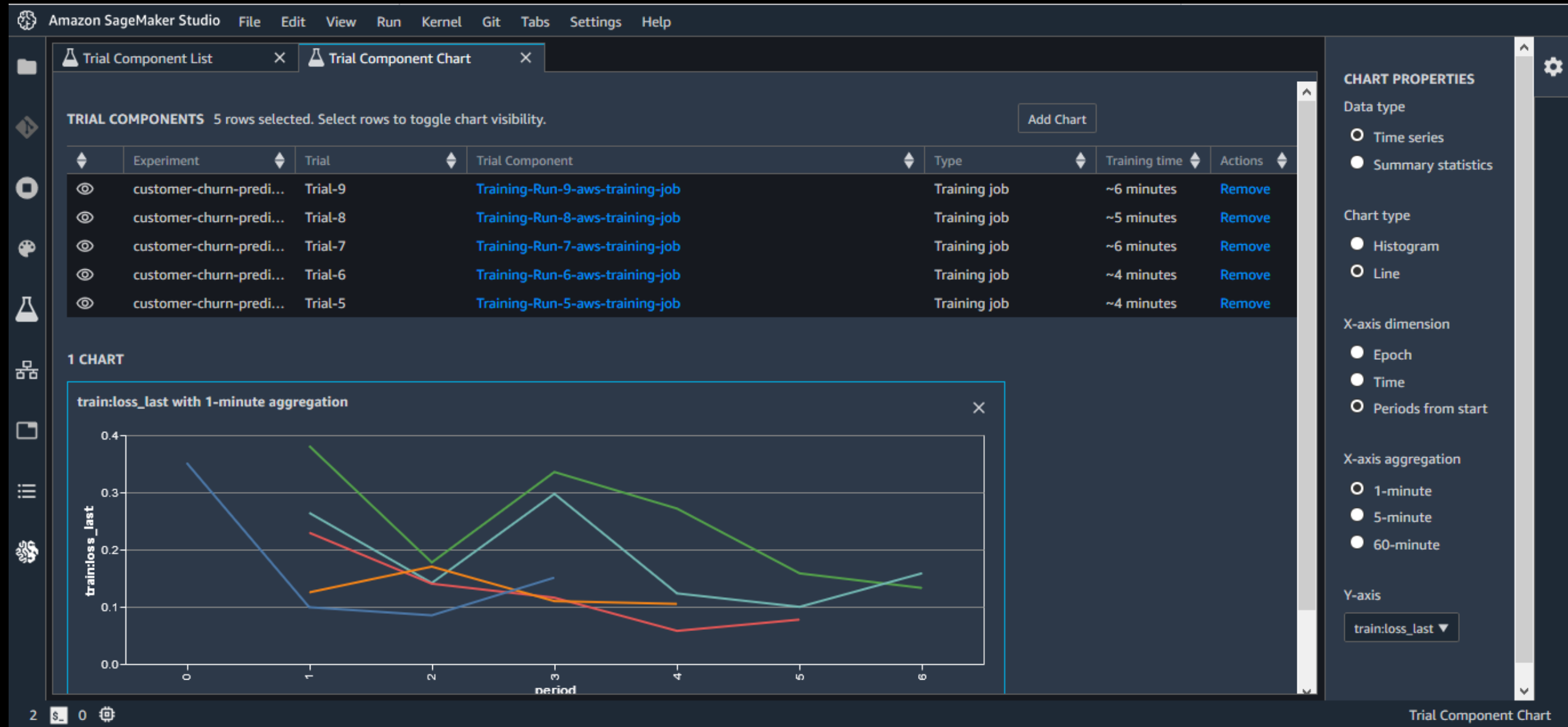
Regressão

- Linear Learner
- XGBoost
- KNN

Redução de dimensionalidade

- PCA
- Object2Vec

Use o Amazon SageMaker Experiments para rastrear e gerenciar milhares de experimentos



Use o Amazon SageMaker Debugger para identificar problemas como vanishing gradients

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

SMDebugger-CloudWatch-Lo

conda_tensorflow_p36

Using SageMaker Rules

In this example we'll demonstrate how to use SageMaker rules to be evaluated against your training. You can find the list of SageMaker rules and the configurations best suited for using them here.

We specify a few rules that check for overfitting, decrease in loss across epochs and for saturated activations.

```
[8]: estimator = TensorFlow(
    role=sagemaker.get_execution_role(),
    base_job_name='mnist-tensorflow-example',
    train_instance_count=1,
    train_instance_type='ml.p3.2xlarge',
    image_name=cpu_training_image,
    entry_point=entrypoint_script,
    framework_version='1.15',
    py_version='py3',
    train_max_run=3600,
    script_mode=True,
    sagemaker_session=sess,
    ## New parameter
    rules = [ Rule.sagemaker(rule_configs.vanishing_gradient()),
              Rule.custom(name='Overfitting', # used to identify the rule
                          image_uri='759209512951.dkr.ecr.us-west-2.amazonaws.com',
                          instance_type='ml.c4.xlarge', # instance type to run the rule
                          source='my_custom_rule.py', # path to the rule source file
                          rule_to_invoke='CustomGradientRule', # name of the class
                          volume_size_in_gb=400, # EBS volume size required to be attached to the instance
                          collections_to_save=[CollectionConfig(name='losses')], # name of the collection
                          rule_parameters={
                              "threshold": "20.0" # this will be used to initialize the rule
                          },
                          hyperparameters = {'num_epochs' : 100 }
                        ) ]
    )
```

Note that Sagemaker-Debugger is only supported for py_version='py3' currently.

Let's start the training by calling `fit()` on the MXNet estimator

```
[9]: # After calling fit, SageMaker will spin off 1 training job and 1 rule job for you
# The rule evaluation status(es) will be visible in the training logs
# at regular intervals

estimator.fit(wait=False)
```

Result

Describe Trial Component

Experiment: Unassigned
Trial: Unassigned

Trial stages

Charts Metrics Parameters Artifacts AWS Settings Debugger

mnist-tensorflow-example-2019-12-02-09-52-13-126-aws-training-job

Created 15 minutes ago

Status Issues Found

Last modified 4 minutes ago

Rule name VanishingGradient

Job ARN arn:aws:sagemaker:us-west-2:3

Status Issues Found

Last modified 4 minutes ago

Rule name Overfitting

Job ARN arn:aws:sagemaker:us-west-2:3

Trial Component Chart

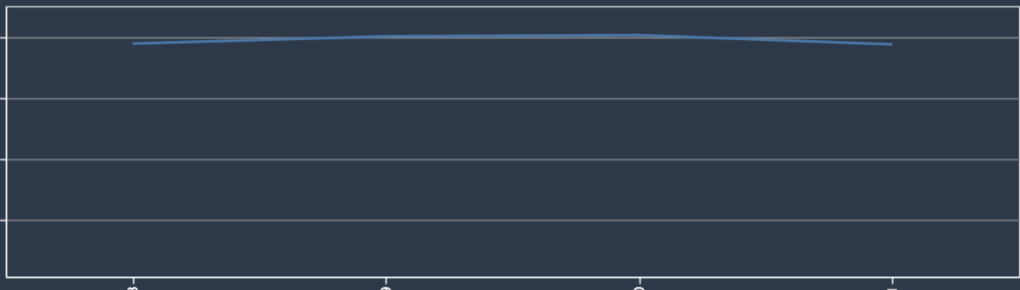
TRIAL COMPONENTS 1 rows selected. Select rows to toggle chart visibility.

Experiment Trial Trial Component Type Train


N/A N/A mnist-tensorflow-example-2019-12-02-09-52-13-126-aws-trainin... Training job ~10

2 CHARTS

sparse_softmax_cross_entropy_loss/value:0_avg with 1-minute aggregation

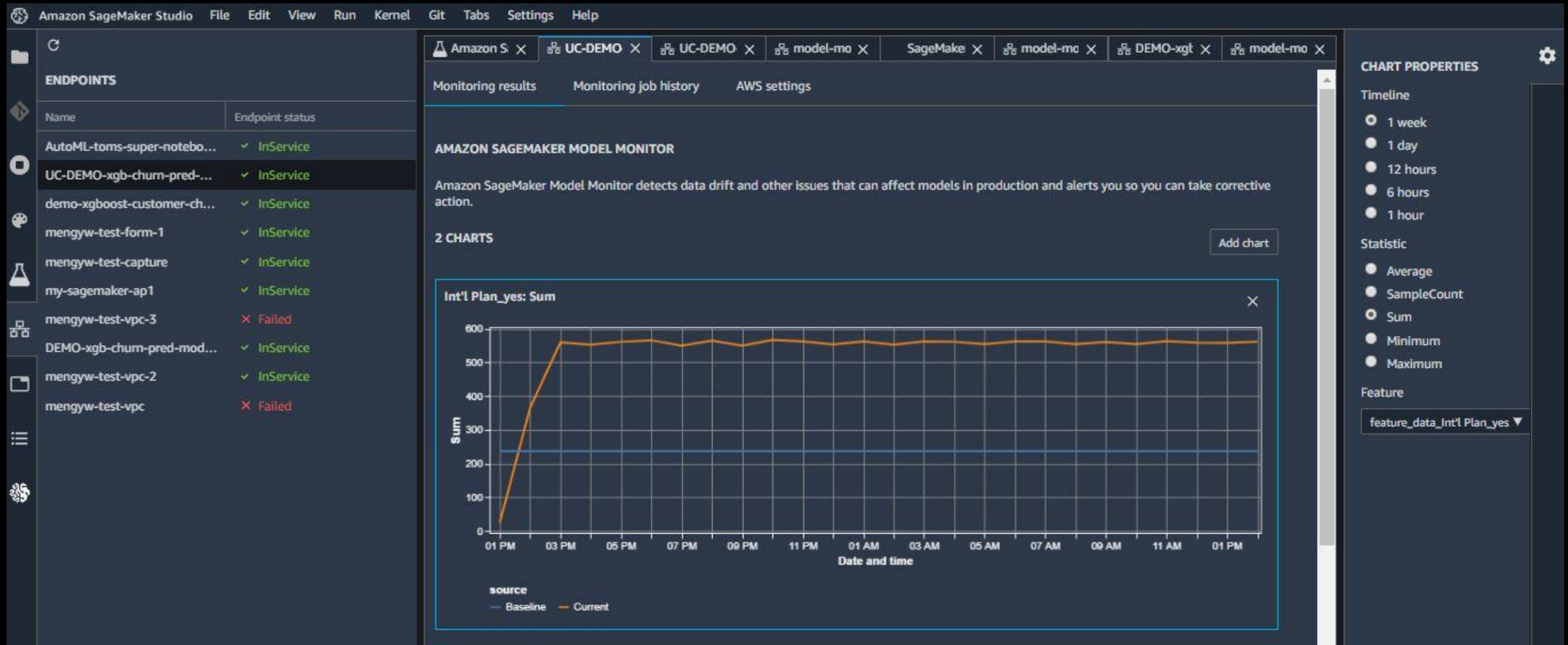


trialComponentName
— mnist-tensorflow-example-2019-1...

 machine learning

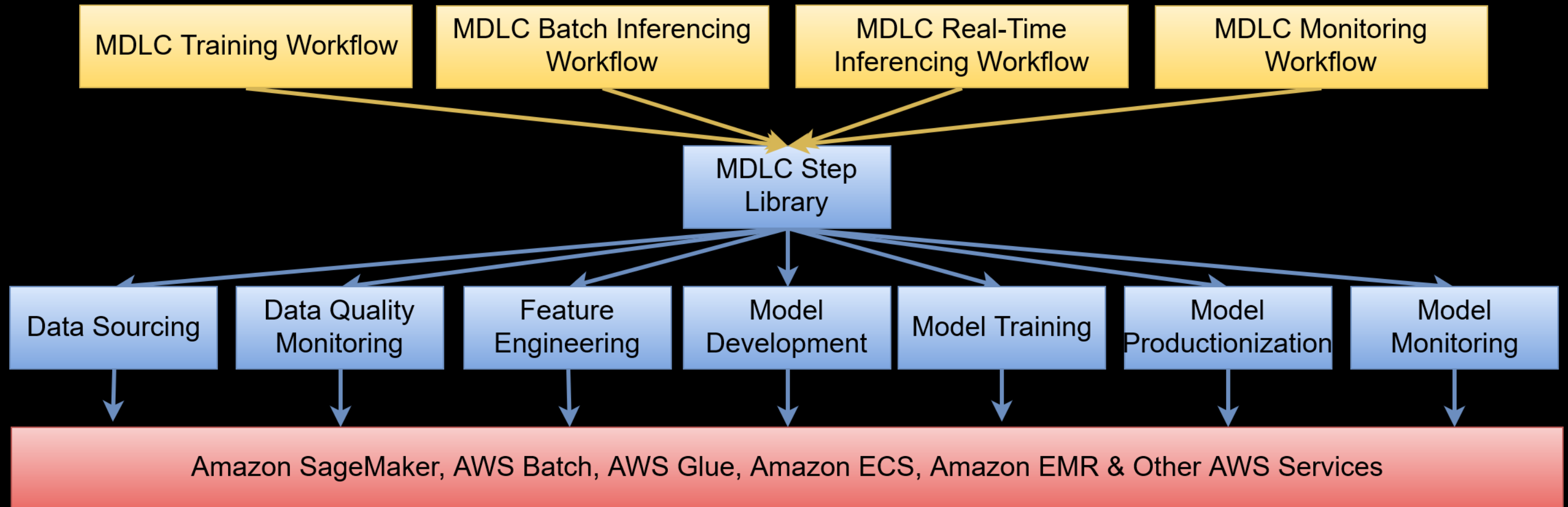
© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Use o Amazon SageMaker Model Monitor para identificar desvios no modelo e tome ações

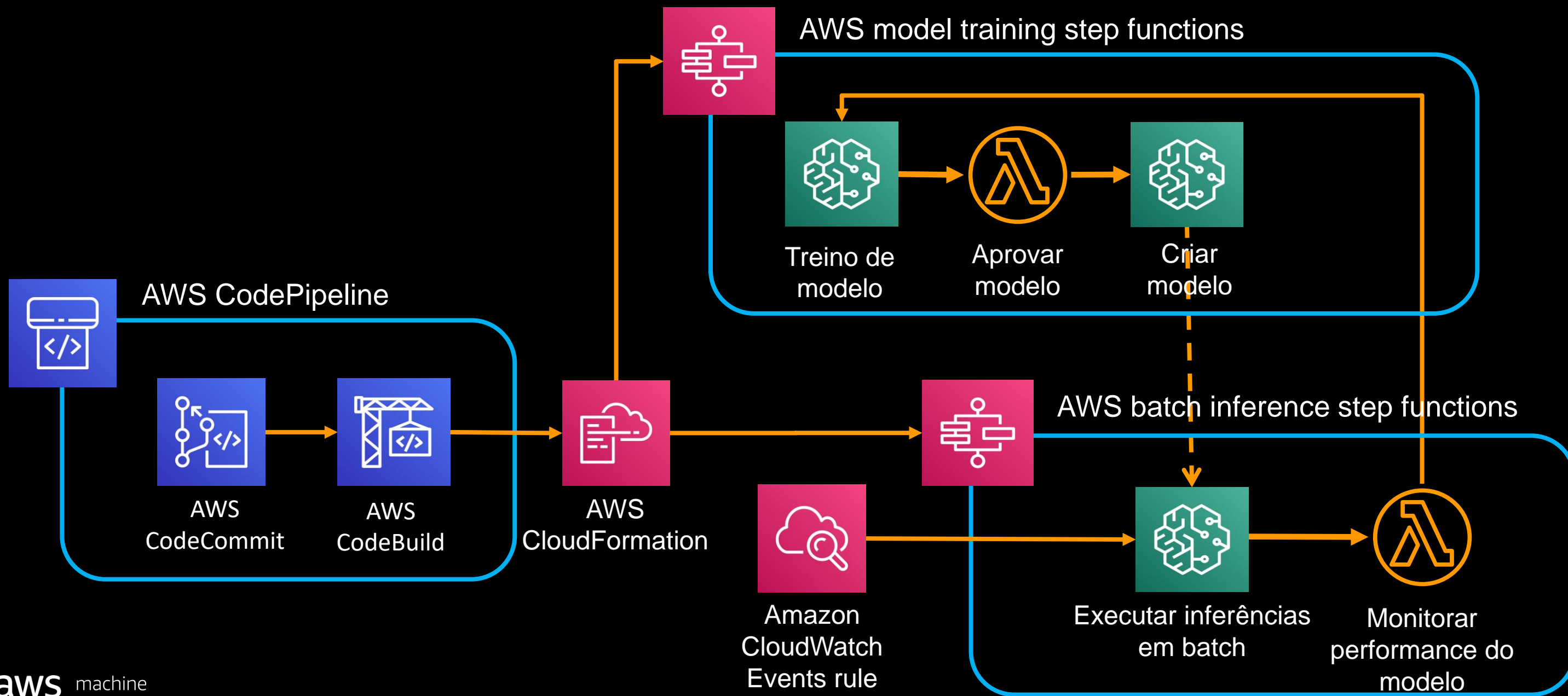


Caso - Amazon.com

Fluxo de trabalho com o MDLC : Componentes

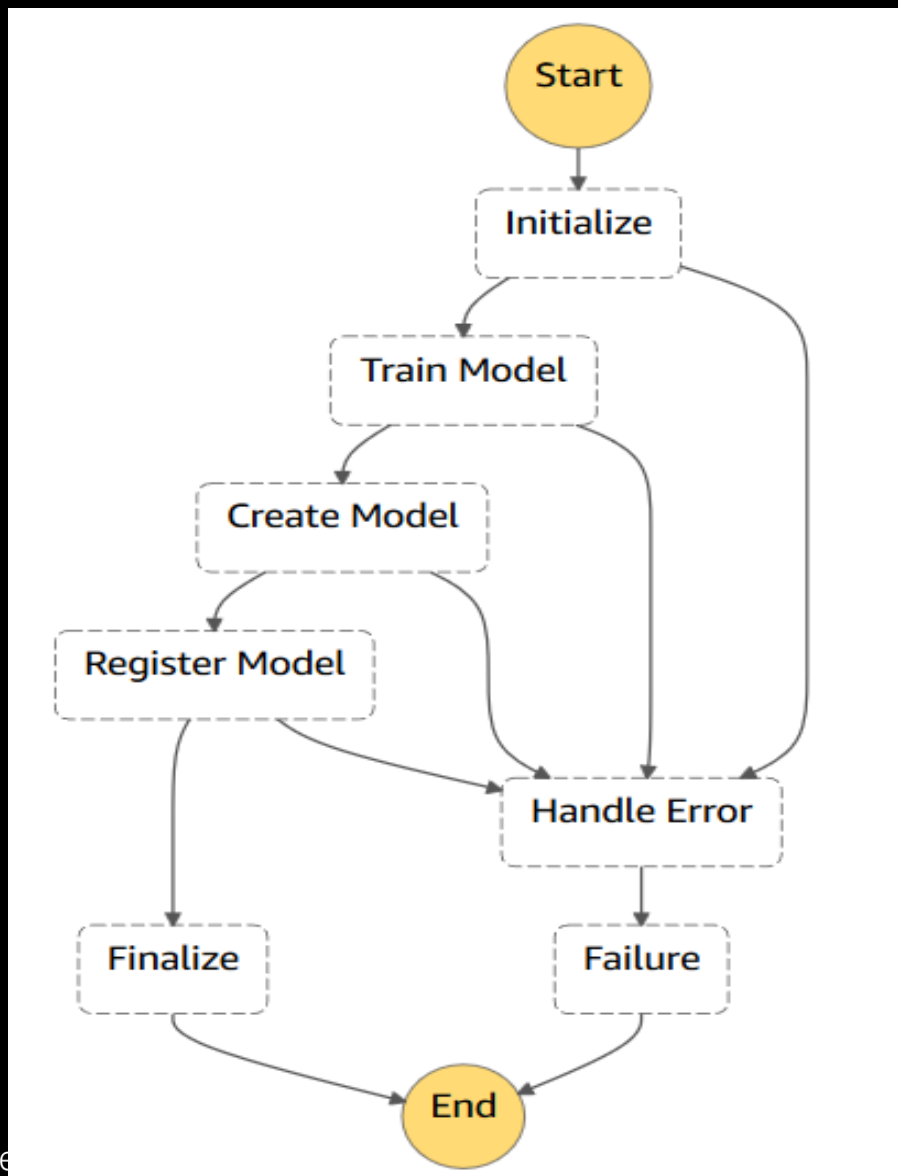


Fluxo de trabalho MDLC de inferência em lote (offline)

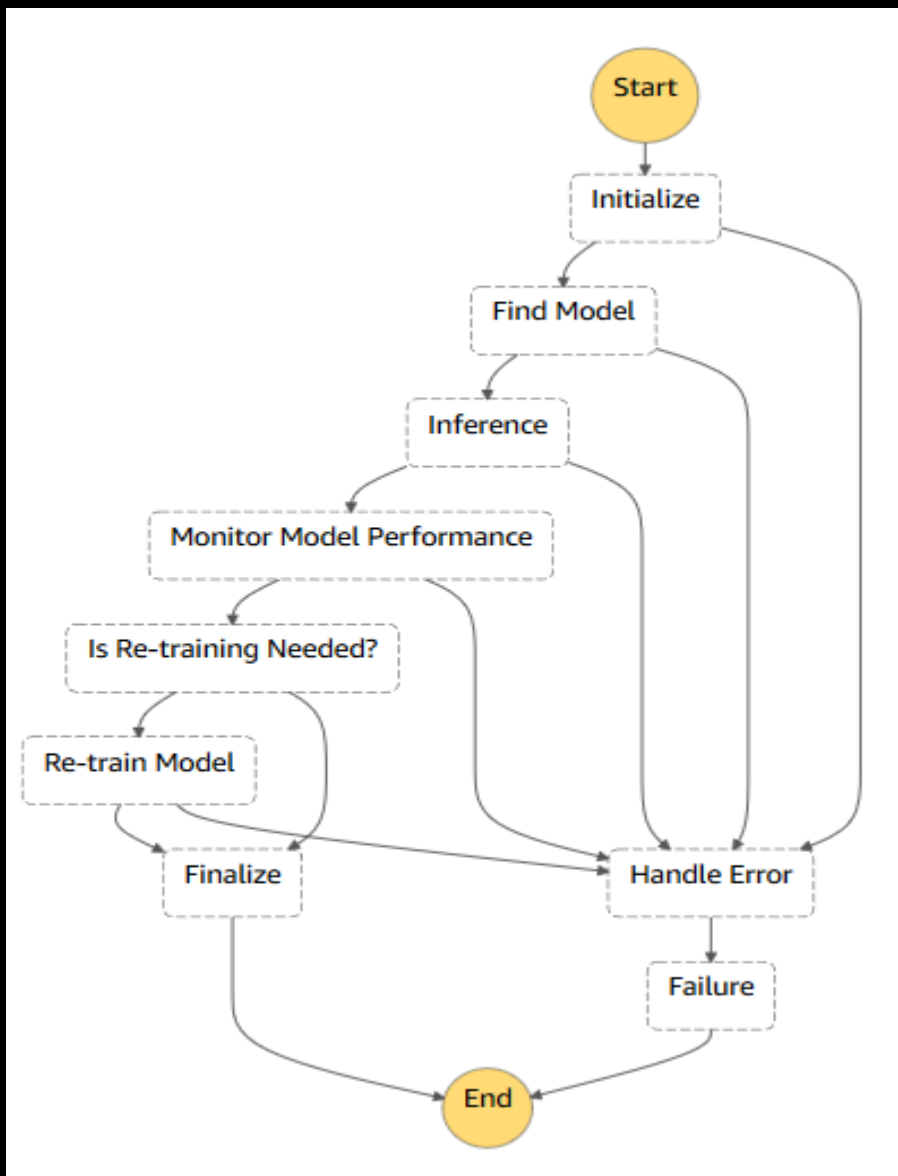


Fluxo de trabalho MDLC: AWS Step Functions

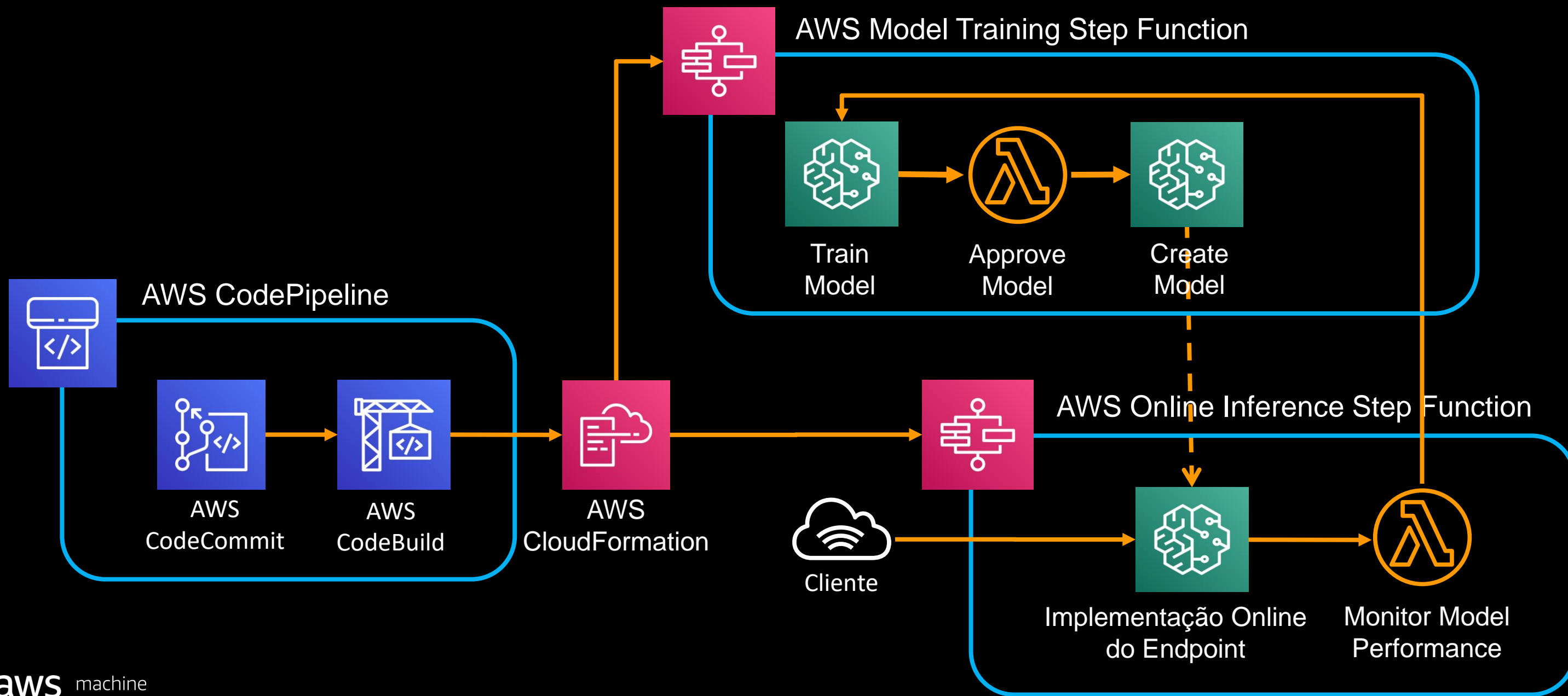
Máquina de estados para treinamento



Máquina de estados para inferência

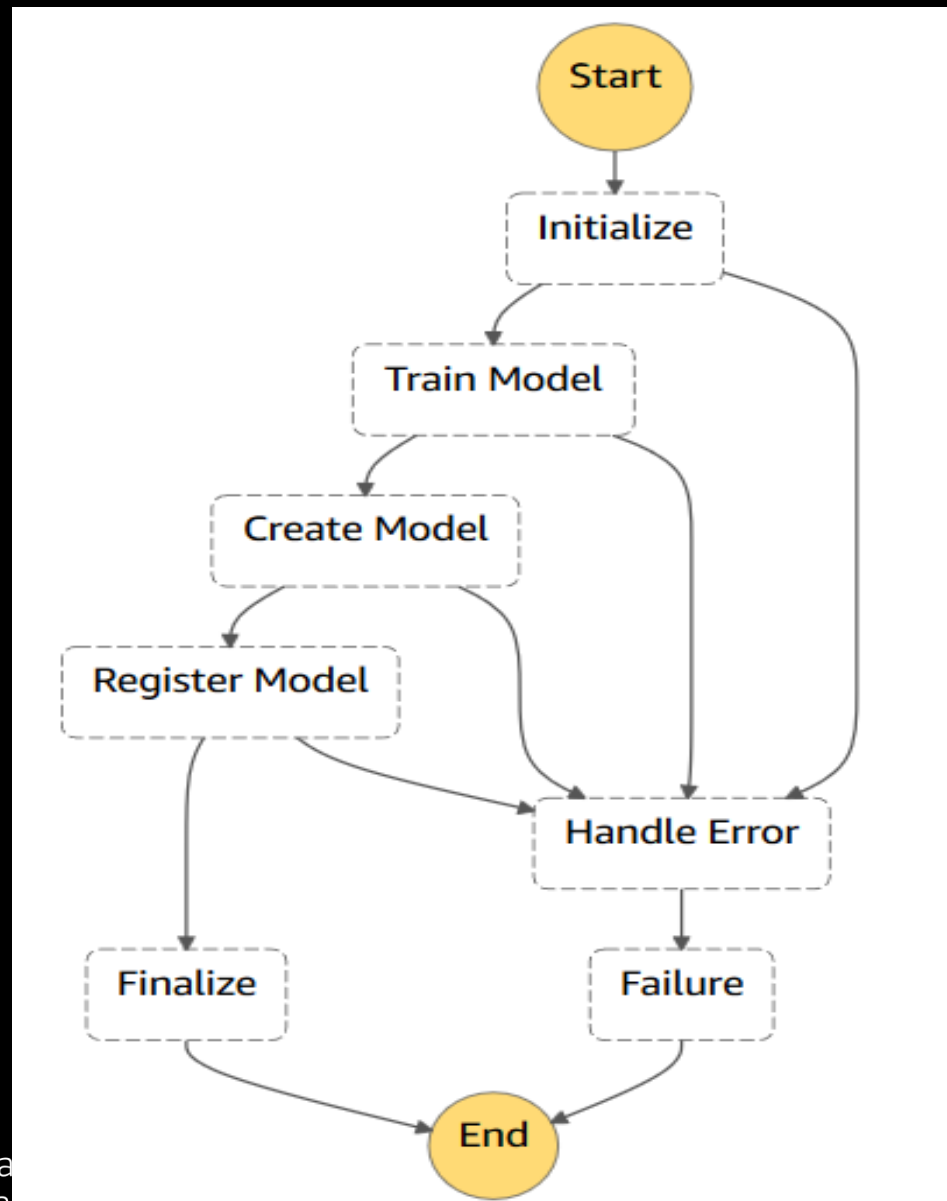


Fluxo de trabalho MDLC de inferência em tempo real (online)

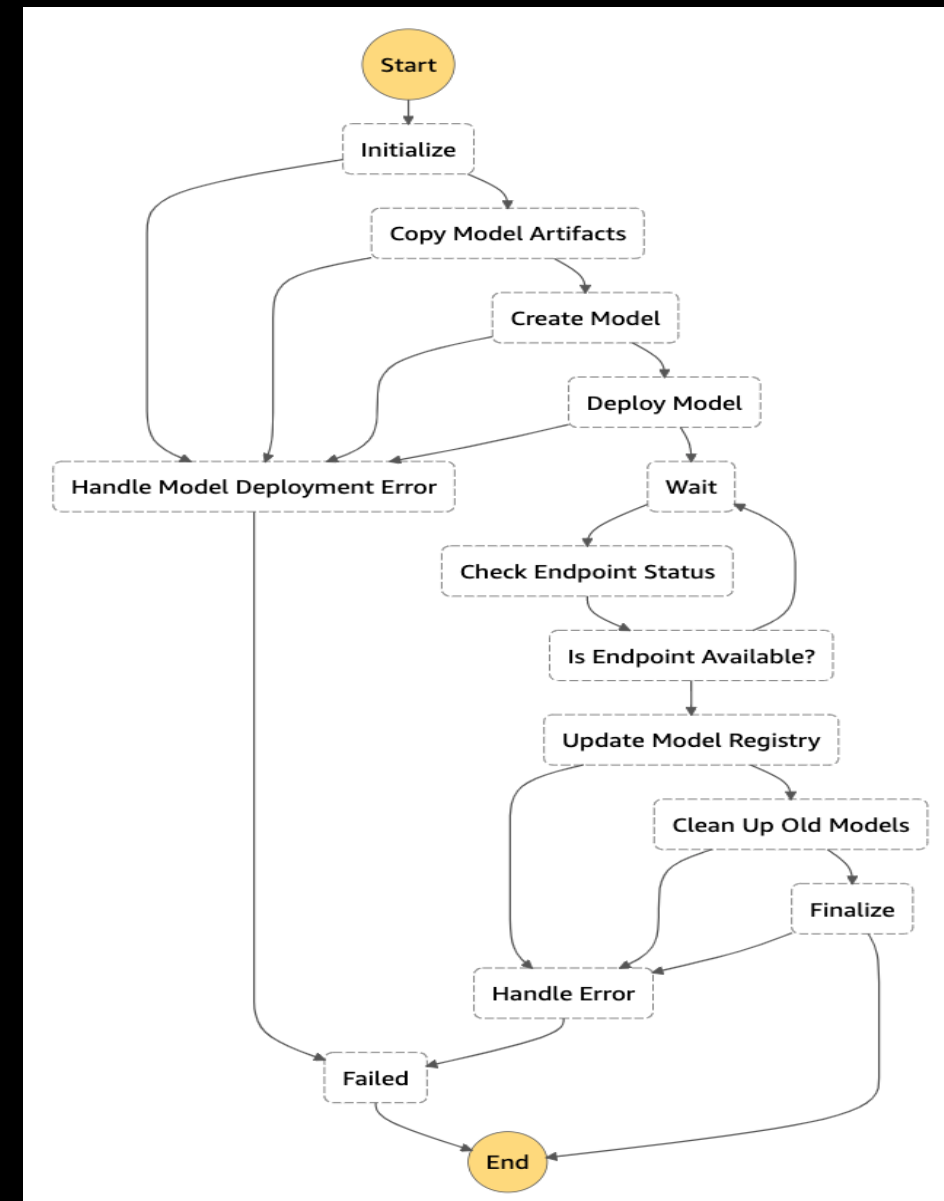


MDLC workflow: AWS Step Functions

Máquina de estado para treinamento



Máquina de estado para inferência



Demo

Por onde começar?

Materiais extras

Apresentação re:Invent 2019:

“Amazon.com automating machine learning deployments at scale”

<https://bit.ly/2WGDsJN>

Amazon Deequ

<https://github.com/awslabs/deequ>

Workshop MLOps na AWS

<https://bit.ly/2MxZBqj>

Notebook de exemplo SageMaker Studio, Experiments, Debugger, Model Monitor

<https://bit.ly/2UxMd6a>

Desenvolvendo os skills de seu time

Aprendizado na prática

Treinamento + Certificações



UDACITY

coursera

edX

AWS DeepRacer
Reinforcement Learning

AWS DeepComposer
Generative AI

AWS DeepLens
Deep Learning

AWS ML Training
&
Certificações

Parceiros
com plataformas de
Ensino online

Available AWS Certifications

Professional

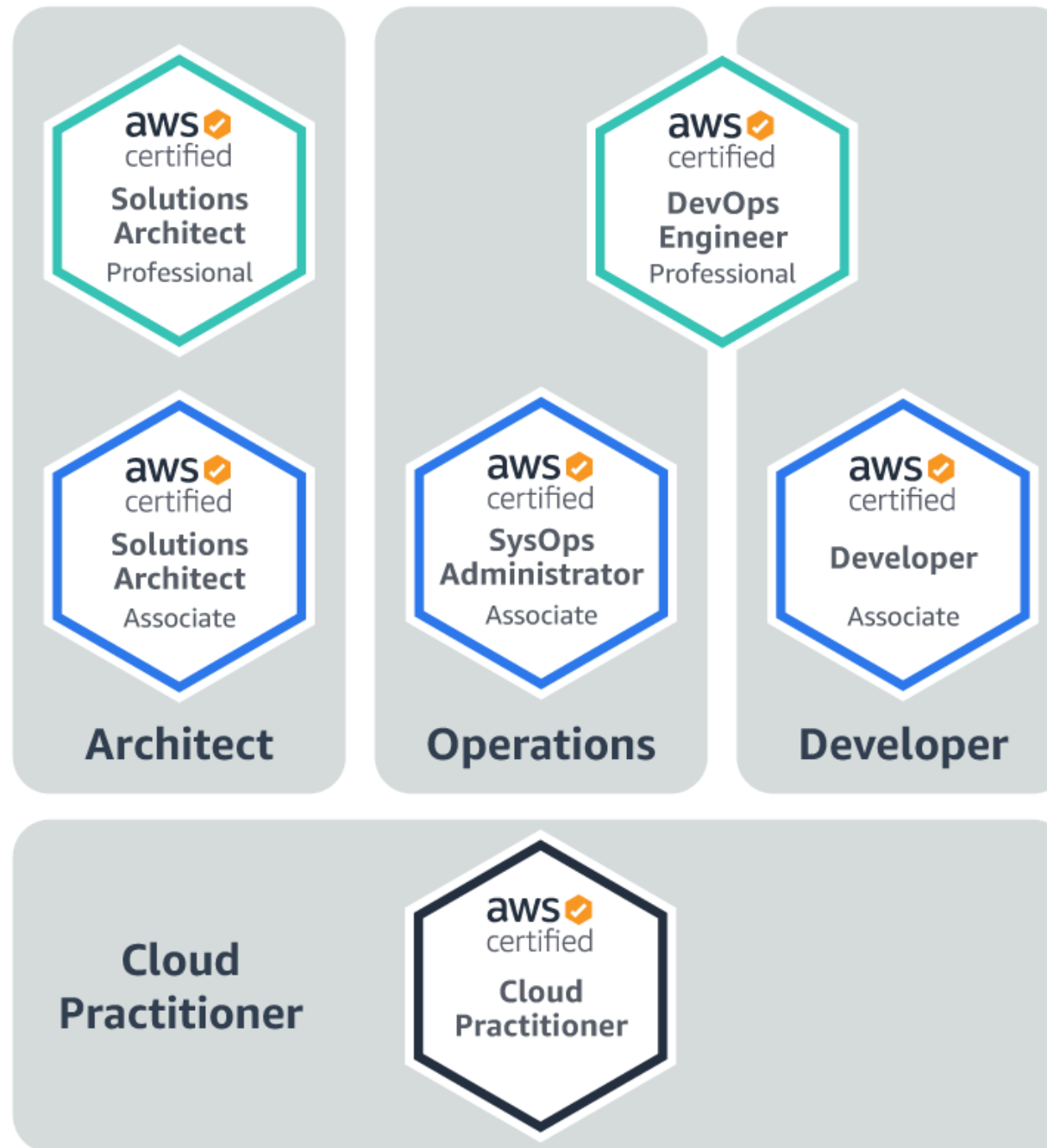
Two years of comprehensive experience designing, operating, and troubleshooting solutions using the AWS Cloud

Associate

One year of experience solving problems and implementing solutions using the AWS Cloud

Foundational

Six months of fundamental AWS Cloud and industry knowledge



Specialty

Technical AWS Cloud experience in the Specialty domain as specified in the **exam guide**



Obrigado!

Renato Barbosa
rbbarbos@amazon.com

