

Amazon SageMaker

End-to-End ML pipeline

1. Data acquisition
2. Data preparation
3. Data Analysis
4. Modeling
5. Deploy/Productionization
6. Monitoring



→ Problem-specific
Sub-stages
& variations

End-to-End ML pipeline

1. Data acquisition
2. Data preparation
3. Data Analysis
4. Modeling
5. Deploy/Productionization
6. Monitoring

[On-premise]

DB/DW/Log servers

compute / cluster servers

Pros & Cons of on-premise setup:

- (-) up front high hardware costs
- (-) less flexible to changes in workload
- ~~(-)~~ Administrative & ops overhead → install, update, h/w setup
- (-) security overhead
- (+) full-control of h/w, s/w & security

Cloud

[AWS, Azure, GCP]

→ renting hardware, software, Ops
as you need.

Popular Open-source tools/libraries

- Python , Matplotlib, Numpy / Scipy, Seaborn, Notebooks, Pandas
- SKlearn, Xgboost, Light GBM, CatBoost, . . .
- TF, Keras, PyTorch, Caffe
- Spark (MLlib), Dask ,
- Kubeflow, MLFlow,

	On-premise + open source tools	Cloud + open source tools	Cloud ML platforms
Hardware	x	✓	✓
Software	x	x	✓
Security	x	(partial)	✓
Migration	easy	easy	hard

Cloud-based ML platforms

Amazon Sagemaker

Azure ML

GCP Vertex AI + Other services



Hardware

Open-source software

Ops

[well-packaged]

Pros & Cons of cloud ML platforms

- (-) lock-in ; harder to move to others
- (+) no H/w & S/w & ops issues ; focus on core -ML
- (+) best practices & tools available by default
- (+) can Speed-up, ML : idea to product , by a few times-

e.g.: why we love Google Colab?

→ (Notebook Server)

No installation headaches

Accessible anywhere

Pro for heavier workloads.

Makes life simpler & easier.

When to use Amazon SageMaker?

- data & team on AWS
- No need to migrate in foreseeable future.
- No internal compliance issues/conflicts -

NOT
when to use Amazon SageMaker?

- full control of everything
- large team (Admin, ops, Engineers ...)
- Internal compliance & security

SageMaker

Key Features →

Prepare

Data Wrangler

Feature Store

Ground Truth

Clarify

Build

Studio

Autopilot

JumpStart

Train & Tune

Debugger

Distributed Training

Deploy & Manage

Pipelines

Model Monitor

Kubernetes Integration

Edge Manager

Neo

<https://aws.amazon.com/sagemaker/>

Simple end-end system

- UCI credit card Default
- Xgboost model
- Focus on The Sagemaker specific details
- we will dive deeper later

Very
easy
to
follow

Refer : <https://aws.amazon.com/getting-started/hands-on/build-train-deploy-monitor-machine-learning-model-sagemaker-studio/>

AWS

- create an account
- login
- needs a payment method.

Sagemaker - home page

The screenshot shows the Amazon SageMaker homepage. On the left, a sidebar menu is open under the heading "Amazon SageMaker". A red box highlights the "Amazon SageMaker Studio" option, which is also the target of a red arrow pointing from the top-left of the image. Other menu items include "Dashboard", "Search", "Images", and several sections starting with "▶".

The main content area features the title "Amazon SageMaker" and the subtitle "Build, train, and deploy machine learning models at scale". Below this, a sub-subtitle reads "The quickest and easiest way to get ML models from idea to production." To the right, there's a "Get started" section with a "SageMaker Studio" button, a "Pricing (US)" section, and a "Related services" section listing "AWS Glue", "Amazon EC2", and "Amazon Elastic Block Store (EBS)".

How it works

Label	Build	Train
		
Set up and manage labeling jobs for highly accurate training datasets within Amazon SageMaker, using active learning and human labeling.	Connect to other AWS services and transform data in Amazon SageMaker notebooks.	Use Amazon SageMaker's algorithms and frameworks, or bring your own, for distributed training.

SETUP

Quickstart vs Standard Setup

SageMaker Studio

Get started

[Learn more about getting started with SageMaker Studio](#)

Quick start (radio button selected)

Let Amazon SageMaker handle configuring account and setting the permissions that you or a team in your organization need to use SageMaker Studio. Choosing this option uses standard encryption, which you can't change. If you need more control over configuration, choose Standard setup.

User name: tutorial-appliedroots

The user name can have up to 63 characters. Valid characters: A-Z, a-z, 0-9, and - (hyphen)

Execution role: Create a new role

Please provide a valid Arn.

SageMaker Projects and JumpStart New: Enable access and provisioning of AWS Service Catalog Portfolio of products in Amazon SageMaker Studio for Amazon SageMaker Projects and JumpStart. [Learn more](#)

Enable Amazon SageMaker project templates and JumpStart for this account and Studio users: If enabled, the administrator can view the Amazon SageMaker provided project templates and JumpStart solutions published in AWS Service Catalog and users who are configured to use the domain execution are allowed to create projects using those templates and solutions with JumpStart. A launch constraint role and a project use role are automatically generated in IAM for your account.

Standard setup (radio button)

Control all aspects of account configuration, including permissions and encryption. Choose this option if you are an administrator setting up SageMaker Studio for your organization.

[Cancel](#) [Submit](#)

Execution role:

Create an IAM role

Passing an IAM role gives Amazon SageMaker permission to perform actions in other AWS services on your behalf. Creating a role here will grant permissions described by the [AmazonSageMakerFullAccess](#) IAM policy to the role you create.

The IAM role you create will provide access to:

S3 buckets you specify - *optional*

Any S3 bucket
Allow users that have access to your notebook instance access to any bucket and its contents in your account.

Specific S3 buckets
Example: bucket-name-1, bucket
Comma delimited. ARNs, "*" and "/" are not supported.

None

Any S3 bucket with "sagemaker" in the name

Any S3 object with "sagemaker" in the name

Any S3 object with the tag "sagemaker" and value "true" [See Object tagging](#)

S3 bucket with a Bucket Policy allowing access to SageMaker [See S3 bucket policies](#)

[Cancel](#) [Create role](#)

S3 : simple storage service

Get started

[Learn more about getting started with SageMaker Studio](#)

Quick start

Let Amazon SageMaker handle configuring account and setting the permissions that you or a team in your organization need to use SageMaker Studio. Choosing this option uses standard encryption, which you can't change. If you need more control over configuration, choose Standard setup.

User name

tutorial-appliedroots

The user name can have up to 63 characters. Valid characters: A-Z, a-z, 0-9, and - (hyphen)

Execution role

SageMaker Studio requires permissions to access other AWS services, such as Amazon SageMaker and Amazon S3. The execution role must have the [AmazonSageMakerFullAccess policy](#) attached. If you don't have a role with this policy attached, we can create one for you.

AmazonSageMaker-ExecutionRole-20210815T163173



Success! You created an IAM role.

[AmazonSageMaker-ExecutionRole-20210815T163173](#)



SageMaker Projects and JumpStart [New](#)

Enable access and provisioning of AWS Service Catalog Portfolio of products in Amazon SageMaker Studio for Amazon SageMaker Projects and JumpStart. [Learn more](#)

Enable Amazon SageMaker project templates and JumpStart for this account and Studio users

If enabled, the administrator can view the Amazon SageMaker provided project templates and JumpStart solutions published in AWS Service Catalog and users who are configured to use the domain execution are allowed to create projects using those templates and solutions with JumpStart. A launch constraint role and a project use role are automatically generated in IAM for your account.

Standard setup

Control all aspects of account configuration, including permissions and encryption. Choose this option if you are an administrator setting up SageMaker Studio for your organization.

Cancel

Submit

Data to S3

after few mins

SageMaker Studio is ready
Choose your user name, then choose Open Studio to get started.

Amazon SageMaker > SageMaker Studio > Control Panel

SageMaker Studio Control Panel

Choose your user name, then choose Open Studio to get started

User name	Last modified	Created	
tutorial-appliedroots	Aug 15, 2021 11:11 UTC	Aug 15, 2021 11:11 UTC	<input checked="" type="checkbox"/> Open Studio

Studio Summary

Status Ready	Studio ID d-bmgtl9yg27en	Execution role arn:aws:iam::047419640029:role/service-role/AmazonSageMaker-ExecutionRole-20210815T163173	Authentication method AWS Identity and Access Management (IAM)
------------------------	-----------------------------	---	---

Use the Studio ID for troubleshooting and tracking usage.
The status shown is for the SageMaker Studio service, and is not the status of compute resources such as EC2 instances to execute notebooks.

Projects

- Amazon SageMaker project templates enabled for this account
 - Launch constraint role: arn:aws:iam::047419640029:role/service-role/AmazonSageMakerServiceCatalogProductsLaunchRole
 - Product use role: arn:aws:iam::047419640029:role/service-role/AmazonSageMakerServiceCatalogProductsUserRole
- Amazon SageMaker project templates enabled for Studio users
 - Execution role: arn:aws:iam::047419640029:role/service-role/AmazonSageMaker-ExecutionRole-20210815T163173

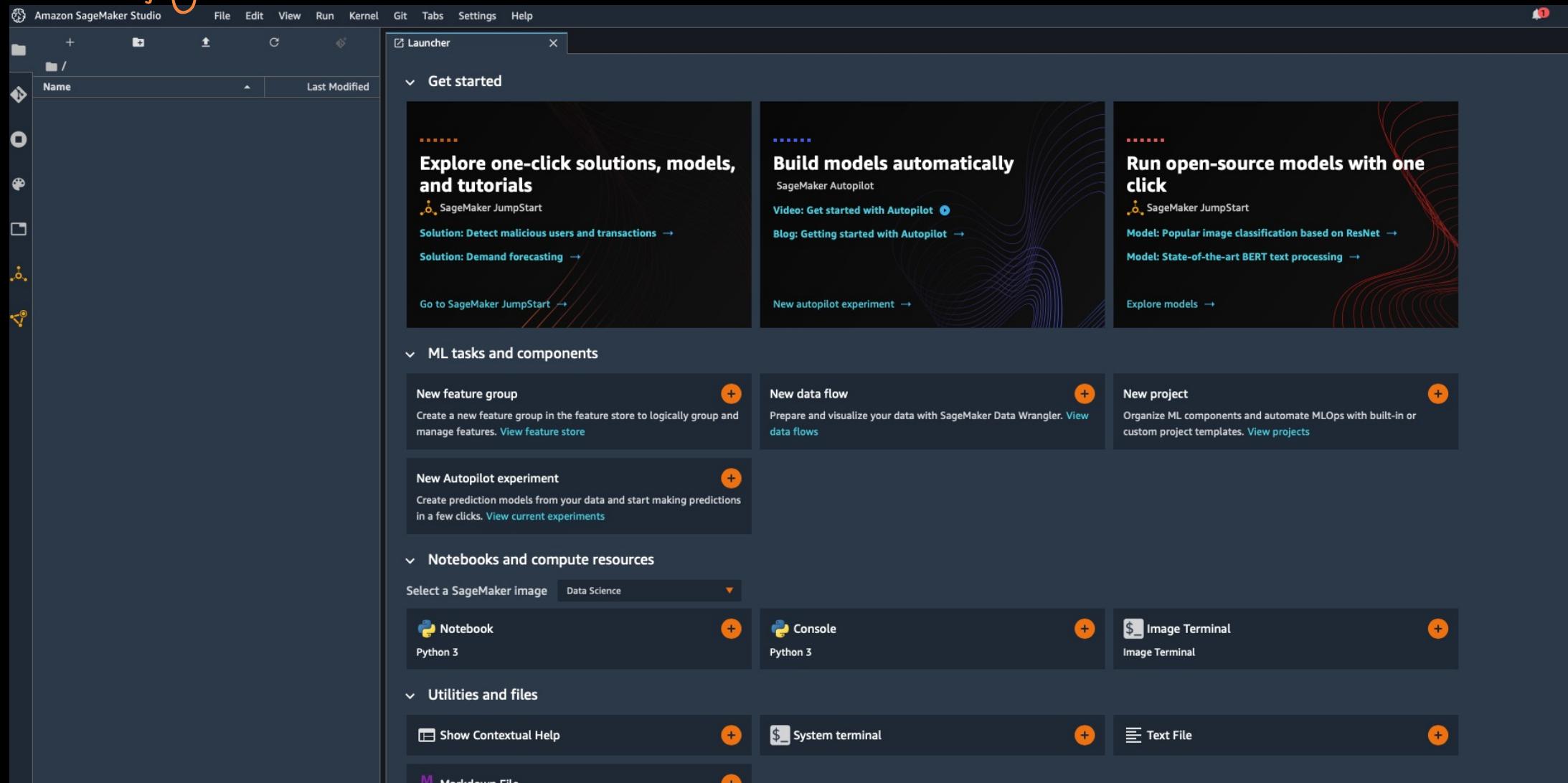
[How to delete Studio](#) [Delete Studio](#) [Edit Settings](#)

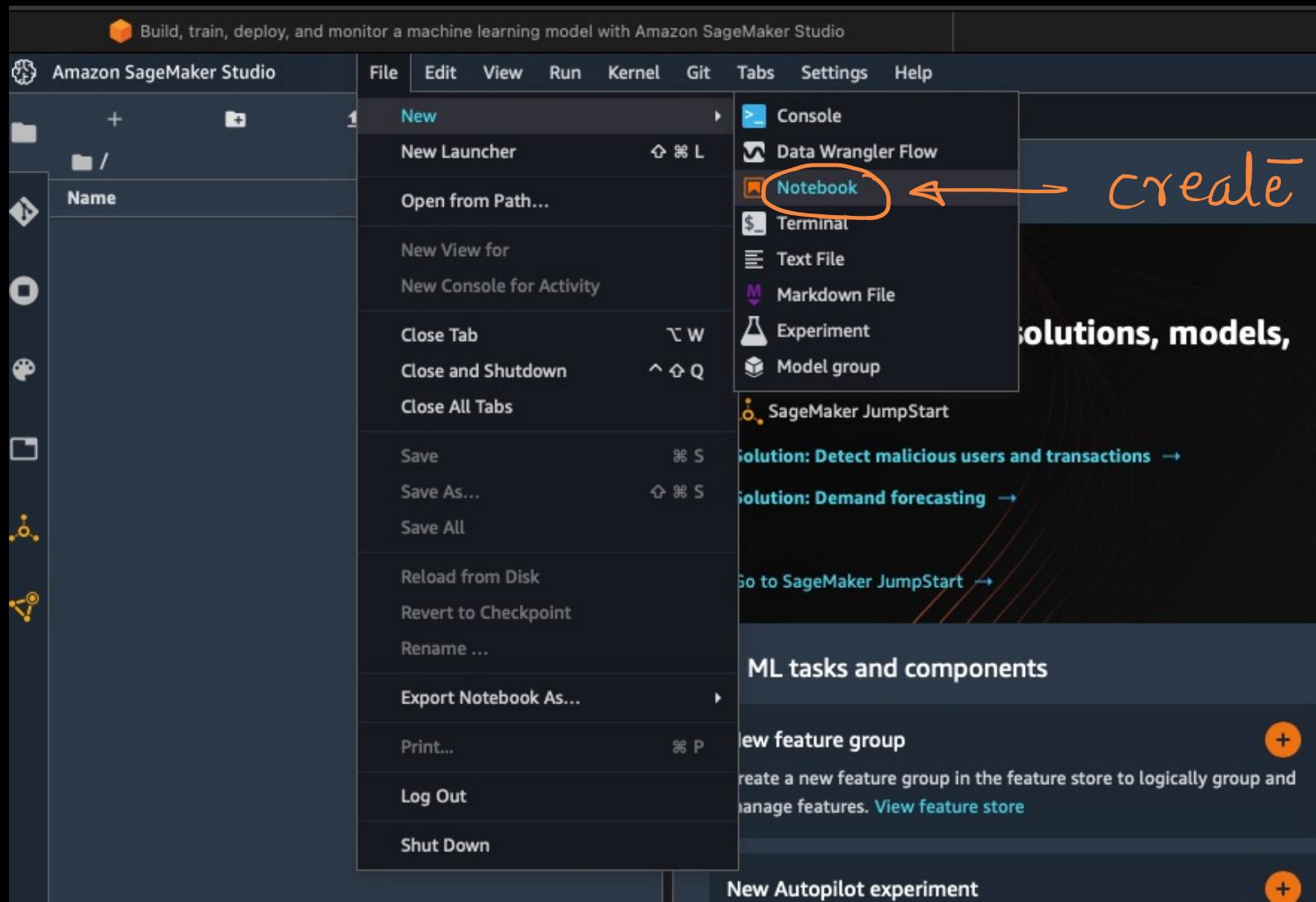


Amazon SageMaker Studio

Creating the JupyterServer application default...

Jupyter-like





create a notebook

solutions, models,

Solution: Detect malicious users and transactions →

Solution: Demand forecasting →

Go to SageMaker JumpStart →

ML tasks and components

New feature group

Create a new feature group in the feature store to logically group and manage features. [View feature store](#)

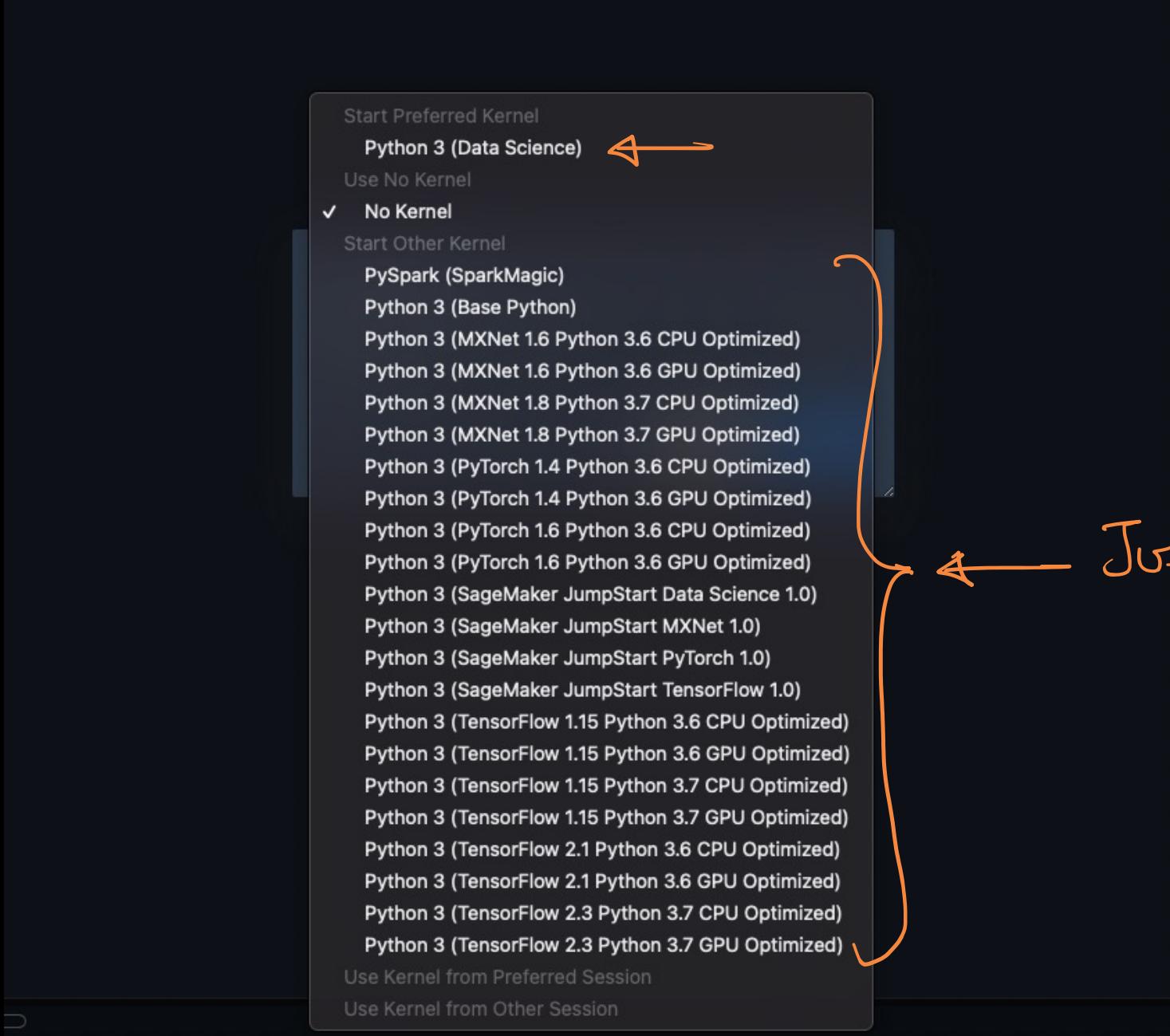
Select Kernel

Select kernel for: "Untitled.ipynb"

No Kernel

Cancel

Select



Jupyter Notebook
Kernels

name

default EC2

kernel

The screenshot shows a Jupyter Notebook interface with the following details:

- Launcher** tab is selected.
- The notebook file is **Untitled.ipynb**.
- The kernel configuration is set to **2 vCPU + 4 GiB Python 3 (Data Science)**.
- The code cell [1] contains Python code to handle SageMaker version compatibility and role setup:

```
[1]: import boto3
import sagemaker
from sagemaker import get_execution_role
import sys
import IPython

if int(sagemaker.__version__.split('.')[0]) == 2:
    print("Installing previous SageMaker Version and restarting the kernel")
    !{sys.executable} -m pip install sagemaker==1.72.0
    IPython.Application.instance().kernel.do_shutdown(True)

else:
    print("Version is good")

role = get_execution_role()
sess = sagemaker.Session()
region = boto3.session.Session().region_name
print("Region = {}".format(region))
sm = boto3.Session().client('sagemaker')
```

SageMaker Version

```
[1]: import boto3
      import sagemaker
      from sagemaker import get_execution_role
      import sys
      import IPython

      if int(sagemaker.__version__.split('.')[0]) == 2:
          print("Installing previous SageMaker Version and restarting the kernel")
          !{sys.executable} -m pip install sagemaker==1.72.0
          IPython.Application.instance().kernel.do_shutdown(True)

      else:
          print("Version is good")

      role = get_execution_role()
      sess = sagemaker.Session()
      region = boto3.session.Session().region_name
      print("Region = {}".format(region))
      sm = boto3.Session().client('sagemaker')
```

For AWS admin

→ Show output on Notebook

```
[1]: # import libraries
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import os
from time import sleep, gmtime, strftime
import json
import time

[2]: # Sagemaker-Experiments to track various training experiments
!pip install sagemaker-experiments
from sagemaker.analytics import ExperimentAnalytics
from smexperiments.experiment import Experiment
from smexperiments.trial import Trial
from smexperiments.trial_component import TrialComponent
from smexperiments.tracker import Tracker
```

S3-bucket & folders

```
# S3 bucket & folders to store all project files
rawbucket= sess.default_bucket() # Alternatively you can use our custom bucket here.

prefix = 'sagemaker-modelmonitor' # use this prefix to store all files pertaining to this workshop.

dataprefix = prefix + '/data'
traindataprefix = prefix + '/train_data'
testdataprefix = prefix + '/test_data'
testdatanolabelprefix = prefix + '/test_data_no_label'
trainheaderprefix = prefix + '/train_headers'
```

INFO:sagemaker:Created S3 bucket: sagemaker-us-west-2-047419640029

AWS homepage

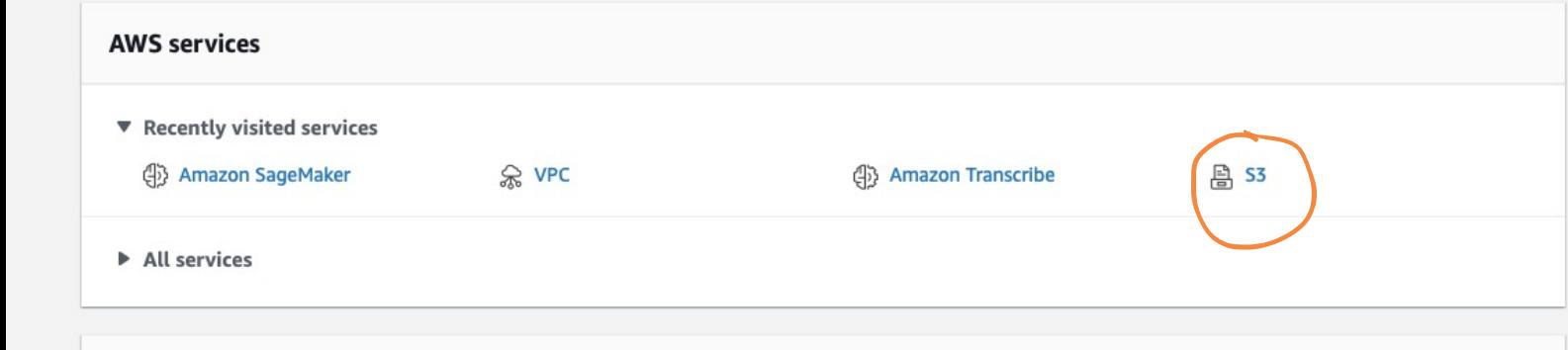
AWS Management Console

AWS services

▼ Recently visited services

- Amazon SageMaker
- VPC
- Amazon Transcribe
- S3

► All services



Amazon S3

► Account snapshot

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

[View Storage Lens dashboard](#)

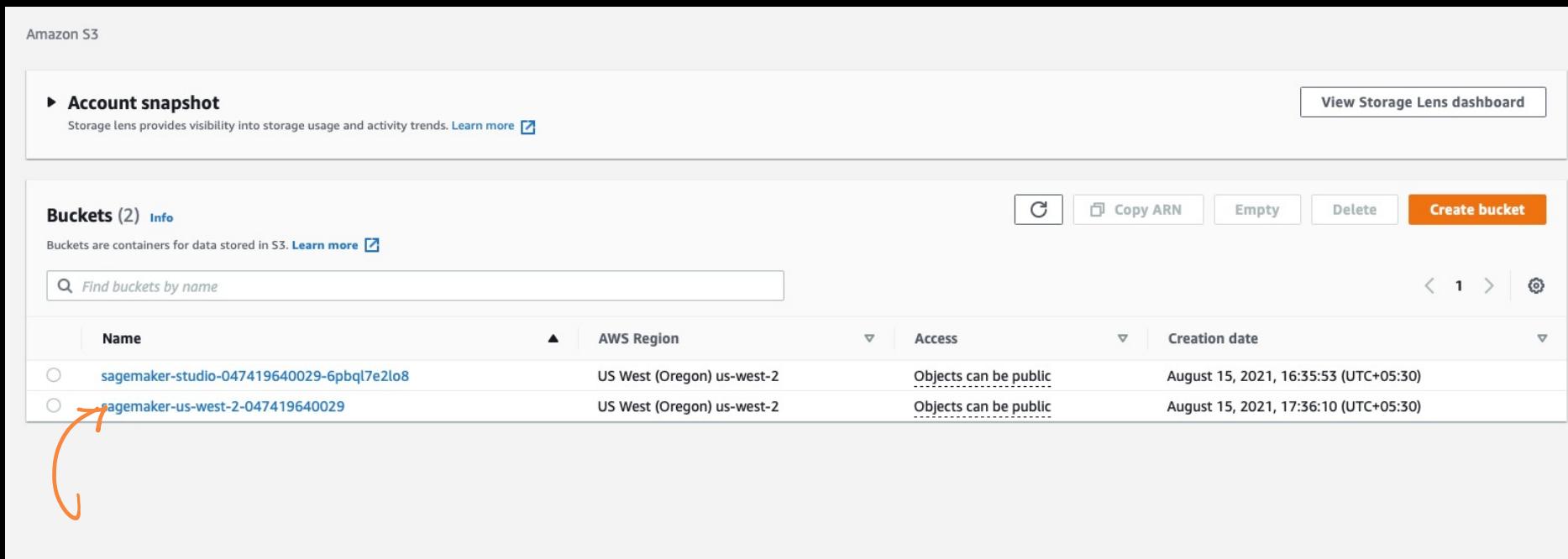
Buckets (2) [Info](#)

Buckets are containers for data stored in S3. [Learn more](#)

Find buckets by name

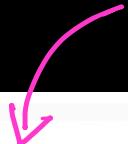
[Create bucket](#)

Name	AWS Region	Access	Creation date
sagemaker-studio-047419640029-6pbql7e2lo8	US West (Oregon) us-west-2	Objects can be public	August 15, 2021, 16:35:53 (UTC+05:30)
sagemaker-us-west-2-047419640029	US West (Oregon) us-west-2	Objects can be public	August 15, 2021, 17:36:10 (UTC+05:30)



Part - II

AWS Free-tier (SageMaker)



Product	Description	Free Tier Offer Details	Product Pricing
Amazon SageMaker Machine Learning	Build, train, and deploy machine learning models fast.	Free 2 month trial 250 hours per month of t2.medium notebook usages 50 hours per month of m4.xlarge for training 125 hours per month of m4.xlarge for hosting	Amazon SageMaker Pricing

What payment methods does AWS accept?

Last updated: 2020-06-12

What payment methods can I use to pay my AWS bill?

Resolution

AWS accepts the following payment methods:

- Most major credit and debit cards. If your account address is in India or Europe, then you can use Visa, MasterCard, or American Express.
- [ACH direct debit payments](#) (for qualifying United States customers).
- [SEPA direct debit payments](#) (for qualifying European customers).

Note: Your seller of record information might impact the payment methods available for your account.

Download data from web:

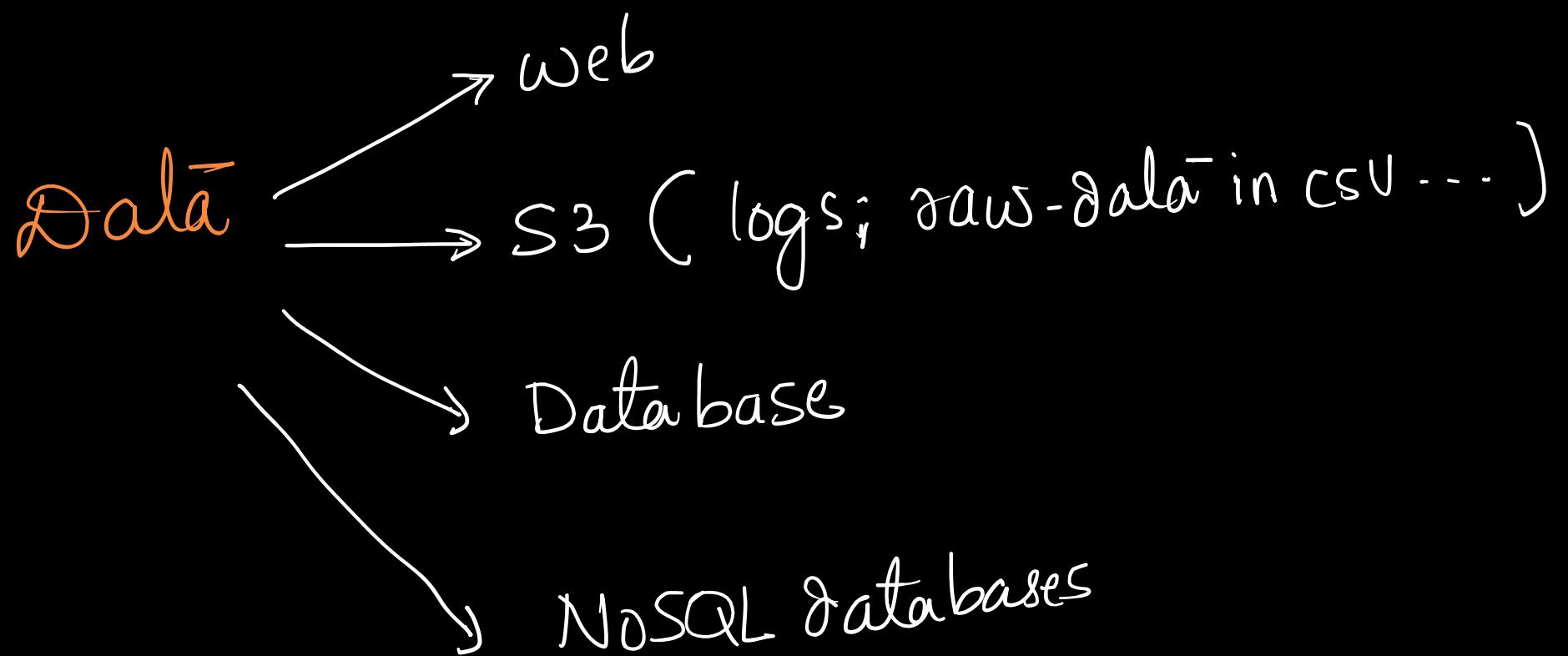
```
# download data using wget
! wget https://archive.ics.uci.edu/ml/machine-learning-databases/00350/default%20of%20credit%20card%20clients.xls
data = pd.read_excel('default of credit card clients.xls', header=1)
data = data.drop(columns = ['ID'])
data.head()
```

```
--2021-08-15 12:09:38-- https://archive.ics.uci.edu/ml/machine-learning-databases/00350/default%20of%20credit%20card%20clients.xls
Resolving archive.ics.uci.edu (archive.ics.uci.edu)... 128.195.10.252
Connecting to archive.ics.uci.edu (archive.ics.uci.edu)|128.195.10.252|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 5539328 (5.3M) [application/x-httdp-php]
Saving to: 'default of credit card clients.xls'

default of credit c 100%[=====>] 5.28M 20.4MB/s in 0.3s

2021-08-15 12:09:39 (20.4 MB/s) - 'default of credit card clients.xls' saved [5539328/5539328]
```

	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	PAY_5	...	BILL_AMT4	BILL_AMT5	BILL_AMT6	PAY_AMT1	PAY_AMT2
0	20000	2	2	1	24	2	2	-1	-1	-2	...	0	0	0	0	0
1	120000	2	2	2	26	-1	2	0	0	0	...	3272	3455	3261	0	1
2	90000	2	2	2	34	0	0	0	0	0	...	14331	14948	15549	1518	1
3	50000	2	2	1	37	0	0	0	0	0	...	28314	28959	29547	2000	2
4	50000	1	2	1	57	-1	0	-1	0	0	...	20940	19146	19131	2000	36



Refer:

<https://docs.aws.amazon.com/sagemaker/latest/dg/data-wrangler-import.html>

:

[Data Wrangler]

```
[9]: # rename last col as "Label" and move it to the first column for Xgboost formatting  
data.rename(columns={"default payment next month": "Label"}, inplace=True)  
lbl = data.Label  
data = pd.concat([lbl, data.drop(columns=['Label'])], axis = 1)  
data.head()
```

```
[9]:
```

	Label	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	...	BILL_AMT3	BILL_AMT4	BILL_AMT5	BILL_AMT6	PAY_-
0	1	20000	2	2	1	24	2	2	-1	-1	...	689	0	0	0	0
1	1	120000	2	2	2	26	-1	2	0	0	...	2682	3272	3455	3261	
2	0	90000	2	2	2	34	0	0	0	0	...	13559	14331	14948	15549	
3	0	50000	2	2	1	37	0	0	0	0	...	49291	28314	28959	29547	
4	0	50000	1	2	1	57	-1	0	-1	0	...	35835	20940	19146	19131	

5 rows × 24 columns

default next month or not

```
# data to S3
if not os.path.exists('rawdata/rawdata.csv'):
    mkdir rawdata
    data.to_csv('rawdata/rawdata.csv', index=None)
else:
    pass

# Upload the raw dataset
raw_data_location = sess.upload_data('rawdata', bucket=rawbucket, key_prefix=dataprefix)
print(raw_data_location)

s3://sagemaker-us-west-2-047419640029/sagemaker-modelmonitor/data
```

} → disk on EC2 instance

to S3 ↗

Data Preprocessing

<https://sagemaker.readthedocs.io/en/stable/frameworks/sklearn/sagemaker.sklearn.html#scikit-learn-processor>

```
# Data Preprocessing using SkLearn
# Use scikit-learn processing container (Docker) for more popular a EC2 instance
# INsatnce types and pricing: https://aws.amazon.com/sagemaker/pricing/
from sagemaker.sklearn.preprocessing import SKLearnProcessor
sklearn_processor = SKLearnProcessor(framework_version='0.20.0',
                                      role=role,
                                      instance_type='ml.c5.xlarge',
                                      instance_count=1)
```

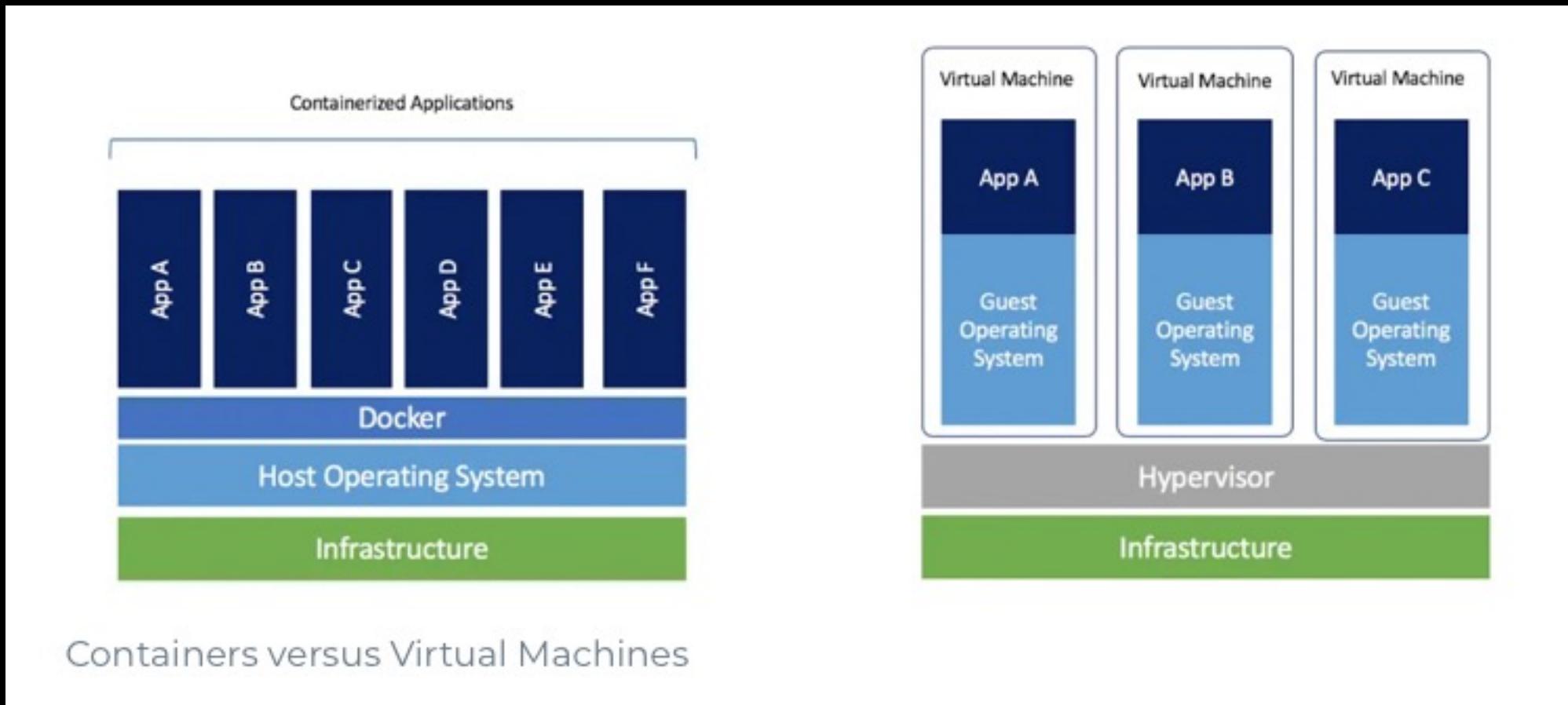


EC2 - instance type

<https://aws.amazon.com/sagemaker/pricing/>

Compute-optimized: 16 vCPU, 32 GB RAM

Docker:



Source : Docker.com

iPython Magic Commands

[<https://ipython.readthedocs.io/en/stable/interactive/magics.html>]

```
%%writefile preprocessing.py

import argparse
import os
import warnings

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.exceptions import DataConversionWarning
from sklearn.compose import make_column_transformer

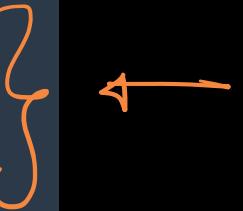
warnings.filterwarnings(action='ignore', category=DataConversionWarning)

if __name__ == '__main__':
    parser = argparse.ArgumentParser()
    parser.add_argument('--train-test-split-ratio', type=float, default=0.3)
    parser.add_argument('--random-split', type=int, default=0)
    args, _ = parser.parse_known_args()
```

EC2 local-disk

[15]: ! ls -l

```
total 5448
-rw-r--r-- 1 root root  25476 Aug 15 12:27 Untitled.ipynb
-rw-r--r-- 1 root root 5539328 Jan 26  2016 'default of credit card clients.xls'
-rw-r--r-- 1 root root   3035 Aug 15 12:27 preprocessing.py
drwxr-xr-x 2 root root  6144 Aug 15 12:18 rawdata
```



Must-do : [WHY?]

```
# Copy the preprocessing code over to the s3 bucket
codeprefix = prefix + '/code'
codeupload = sess.upload_data('preprocessing.py', bucket=rawbucket, key_prefix=codeprefix)
print(codeupload)

s3://sagemaker-us-west-2-047419640029/sagemaker-modelmonitor/code/preprocessing.py
```

```
: # S3 paths to store train and test data after processing
train_data_location = rawbucket + '/' + traindataprefix
test_data_location = rawbucket + '/' + testdataprefix
print("Training data location = {}".format(train_data_location))
print("Test data location = {}".format(test_data_location))
```

```
Training data location = sagemaker-us-west-2-047419640029/sagemaker-modelmonitor/train_data
Test data location = sagemaker-us-west-2-047419640029/sagemaker-modelmonitor/test_data
```

Start sklearn Processing Job:

```
from sagemaker.processing import ProcessingInput, ProcessingOutput

sklearn_processor.run(code=codeupload, # preprocessing.py location
                      inputs=[ProcessingInput(
                        source=raw_data_location, # S3 location
                        destination='/opt/ml/processing/input')], # local disk location as per preprocessing.py
                      outputs=[ProcessingOutput(output_name='train_data',
                                                source='/opt/ml/processing/train', # local disk
                                                destination='s3://'+ train_data_location), #S3 location|
                               ProcessingOutput(output_name='test_data',
                                                source='/opt/ml/processing/test',
                                                destination="s3://" + test_data_location),
                               ProcessingOutput(output_name='train_data_headers',
                                                source='/opt/ml/processing/train_headers',
                                                destination="s3://" + rawbucket + '/' + prefix + '/train_headers')],
                      arguments=['--train-test-split-ratio', '0.2'] # command line arguments to preprocessing.py
)

preprocessing_job_description = sklearn_processor.jobs[-1].describe()

output_config = preprocessing_job_description['ProcessingOutputConfig']
for output in output_config['Outputs']:
    if output['OutputName'] == 'train_data':
        preprocessed_training_data = output['S3Output']['S3Uri']
    if output['OutputName'] == 'test_data':
        preprocessed_test_data = output['S3Output']['S3Uri']
```

How to
understand
this
code?

```
INFO:sagemaker:Creating processing-job with name sagemaker-scikit-learn-2021-08-15-12-34-15-337
```

```
Job Name: sagemaker-scikit-learn-2021-08-15-12-34-15-337
Inputs: [{"InputName": "input-1", "S3Input": {"S3Uri": "s3://sagemaker-us-west-2-047419640029/sagemaker-modelmonitor/data", "LocalPath": "/opt/ml/processing/input", "S3DataType": "S3Prefix", "S3InputMode": "File", "S3DataDistributionType": "FullyReplicated", "S3CompressionType": "None"}}, {"InputName": "code", "S3Input": {"S3Uri": "s3://sagemaker-us-west-2-047419640029/sagemaker-modelmonitor/code/preprocessing.py", "LocalPath": "/opt/ml/processing/input/code", "S3DataType": "S3Prefix", "S3InputMode": "File", "S3DataDistributionType": "FullyReplicated", "S3CompressionType": "None"}}]
Outputs: [{"OutputName": "train_data", "S3Output": {"S3Uri": "s3://sagemaker-us-west-2-047419640029/sagemaker-modelmonitor/train_data", "LocalPath": "/opt/ml/processing/train", "S3UploadMode": "EndOfJob"}}, {"OutputName": "test_data", "S3Output": {"S3Uri": "s3://sagemaker-us-west-2-047419640029/sagemaker-modelmonitor/test_data", "LocalPath": "/opt/ml/processing/test", "S3UploadMode": "EndOfJob"}}, {"OutputName": "train_data_headers", "S3Output": {"S3Uri": "s3://sagemaker-us-west-2-047419640029/sagemaker-modelmonitor/train_headers", "LocalPath": "/opt/ml/processing/train_headers", "S3UploadMode": "EndOfJob"}}]
.....
```

AWS Services ▾ Search for services, features, marketplace products, and docs [Option+S] Srikant Varma Chekuri ▾ Oregon ▾ Support ▾

Amazon SageMaker X

Amazon SageMaker Studio

Dashboard

Search

Images

▶ Ground Truth

▼ Notebook

Notebook instances

Lifecycle configurations

Git repositories

▼ Processing

Processing jobs

▶ Training

▶ Inference

▶ Edge Manager

▶ Augmented AI

▶ AWS Marketplace

Amazon SageMaker > Processing jobs

Processing jobs

Search processing jobs

Name ARN Creation time Duration Status

Name	ARN	Creation time	Duration	Status
sagemaker-scikit-learn-2021-08-15-12-34-15-337	arn:aws:sagemaker:us-west-2:047419640029:processing-job/sagemaker-scikit-learn-2021-08-15-12-34-15-337	Aug 15, 2021 12:34 UTC	4 minutes	Completed

A red arrow points from the "Processing jobs" link in the left sidebar to the "Completed" status of the job in the main table.

Sagemaker Homepage

(OR)

The screenshot shows the Amazon SageMaker homepage with a sidebar and a main content area. The sidebar on the left lists various services under 'Amazon SageMaker Studio' such as Dashboard, Search, Images, Ground Truth, Notebook, Processing, Training, Inference, Edge Manager, Augmented AI, and AWS Marketplace. The main content area has four main sections: Prepare, Build, Train & Tune, and Deploy & Manage. Under 'Recent activity', there are five categories: Ground Truth, Notebook, Training, Inference, and Processing. The Processing section shows one job running and one created, highlighted with a green and grey background respectively. A red arrow points from the text 'Recent activity' to this section.

Recent activity

Recent activity within the Last 7 days

Ground Truth	Notebook	Training	Inference	Processing
Labeling jobs	Notebook instances	Training jobs	Models	Processing jobs
No recent activity.	No recent activity.	No recent activity.	No recent activity.	1 Running
		Hyperparameter tuning jobs	Endpoints	1 Created
		No recent activity.	No recent activity.	
			Batch transform jobs	
			No recent activity.	

Feedback English (US) © 2008 - 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use Cookie preferences

Processing jobs



Actions ▾

Create processing job

 Search processing jobs

< 1 > ⚙

Name	ARN	Creation time	Duration	Status
sagemaker-scikit-learn-2021-08-15-12-34-15-337	arn:aws:sagemaker:us-west-2:047419640029:processing-job/sagemaker-scikit-learn-2021-08-15-12-34-15-337	Aug 15, 2021 12:34 UTC	3 minutes	InProgress

[Amazon SageMaker](#) > [Processing jobs](#) > sagemaker-scikit-learn-2021-08-15-12-34-15-337

sagemaker-scikit-learn-2021-08-15-12-34-15-337

Job settings

Job name	Status	IAM role ARN	Processing time (seconds)
sagemaker-scikit-learn-2021-08-15-12-34-15-337	InProgress	arn:aws:iam::047419640029:role/service-role/AmazonSageMaker-ExecutionRole-20210815T163173	-
ARN	Creation time		
arn:aws:sagemaker:us-west-2:047419640029:processing-job/sagemaker-scikit-learn-2021-08-15-12-34-15-337	Aug 15, 2021 12:34 UTC		
	Last modified time		
	Aug 15, 2021 12:34 UTC		

App specification

Processing image	Instance count	Additional volume size (GB)	Volume encryption key
246618743249.dkr.ecr.us-west-2.amazonaws.com/sagemaker-scikit-learn:0.20.0-cpu-py3	1	30	-
Instance type		Maximum runtime (seconds)	
ml.c5.xlarge		86400	
Entrypoint			
python3			

Notebook output

```
.....  
/miniconda3/lib/python3.7/site-packages/sklearn/externals/joblib/externals/cloudpickle/cloudpickle.py:47: DeprecationWarning: the imp module is deprecated in favour of importlib; see the module's documentation for alternative uses  
    import imp  
Received arguments Namespace(random_split=0, train_test_split_ratio=0.2)  
Reading input data from /opt/ml/processing/input/rawdata.csv  
Running preprocessing and feature engineering transformations  
Train data shape after preprocessing: (24000, 23)  
Test data shape after preprocessing: (6000, 23)  
Saving training features to /opt/ml/processing/train/train_data.csv  
Complete  
Save training data with headers to /opt/ml/processing/train_headers/train_data_with_headers.csv  
Saving test features to /opt/ml/processing/test/test_data.csv  
Complete
```



Amazon S3 > **sagemaker-us-west-2-047419640029** > sagemaker-modelmonitor/

sagemaker-modelmonitor/

Objects (5)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in this bucket.

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last
<input type="checkbox"/>	code/	Folder	-
<input type="checkbox"/>	data/	Folder	-
<input type="checkbox"/>	test_data/	Folder	-
<input type="checkbox"/>	train_data/	Folder	-
<input type="checkbox"/>	train_headers/	Folder	-

Experiments

↳ collection of processing & training jobs related
to a project

↓
data, params, models

```
# Create a SageMaker Experiment
cc_experiment = Experiment.create(
    experiment_name=f"Build-train-deploy-{int(time.time())}",
    description="Predict credit card default from payments data",
    sagemaker_boto_client=sm)
print(cc_experiment)

Experiment(sagemaker_boto_client=<botocore.client.SageMaker object at 0x7f8765fce990>,experiment_name='Build-train-deploy-1630234090',description='Predict credit card default from payments data',tags=None,experiment_arn='arn:aws:sagemaker:us-west-2:047419640029:experiment/build-train-deploy-1630234090',response_metadata={'RequestId': '62ce51cd-c850-475e-ade6-e826eb899ac1', 'HTTPStatusCode': 200, 'HTTPHeaders': {'x-amzn-requestid': '62ce51cd-c850-475e-ade6-e826eb899ac1', 'content-type': 'application/x-amz-json-1.1', 'content-length': '101', 'date': 'Sun, 29 Aug 2021 10:48:10 GMT'}, 'RetryAttempts': 0})
```

Expt



multiple Trials (one single training end-end)

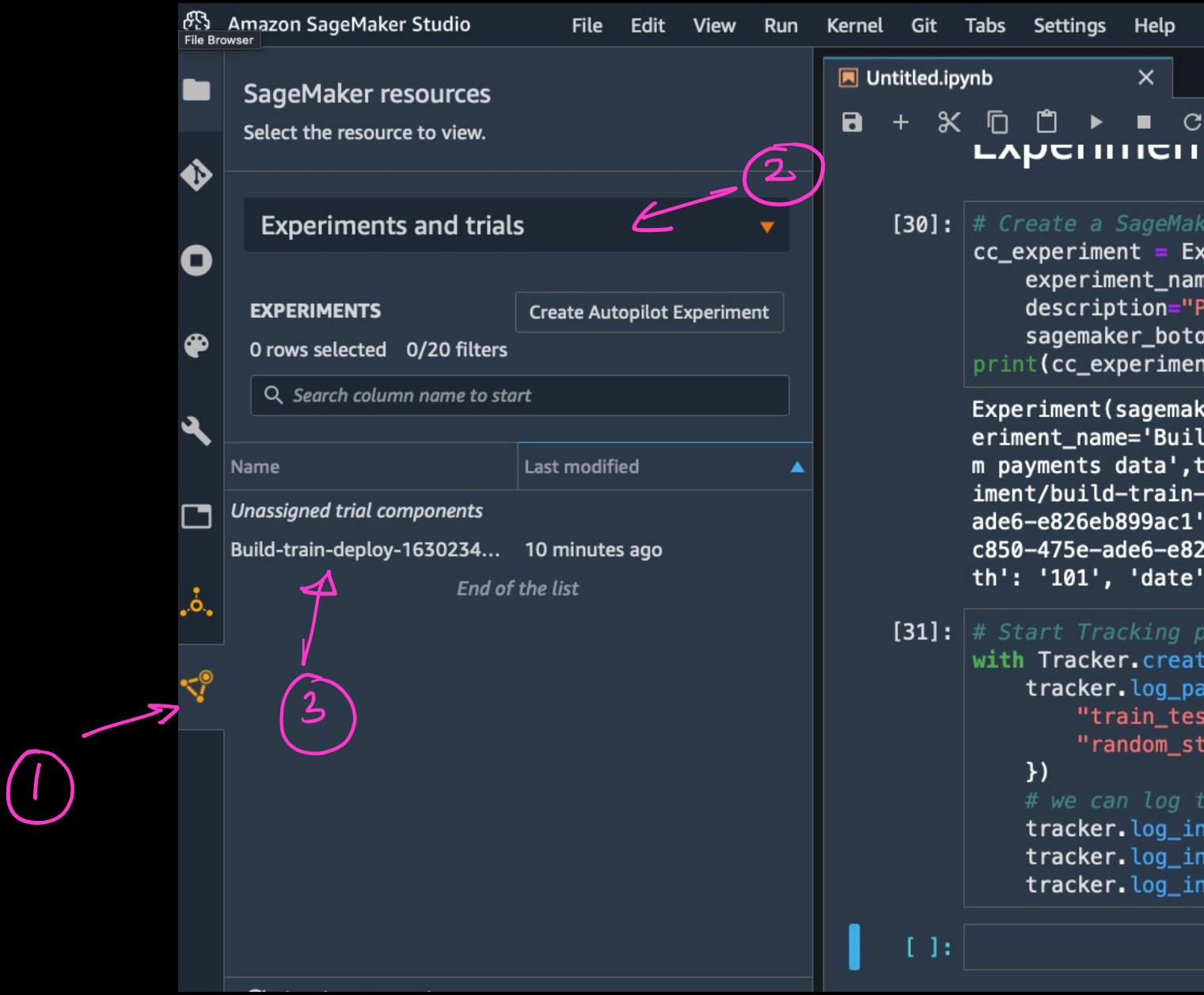
Tracking everything related in
an expt is key

→ reproducibility

```
# Start Tracking parameters used in the Pre-processing pipeline.  
with Tracker.create(display_name="Preprocessing", sagemaker_boto_client=sm) as tracker:  
    tracker.log_parameters({  
        "train_test_split_ratio": 0.2,  
        "random_state": 0  
    })  
    # we can log the s3 uri to the dataset we just uploaded  
    tracker.log_input(name="ccdefault-raw-dataset", media_type="s3/uri", value=raw_data_location)  
    tracker.log_input(name="ccdefault-train-dataset", media_type="s3/uri", value=train_data_location)  
    tracker.log_input(name="ccdefault-test-dataset", media_type="s3/uri", value=test_data_location)
```

Train - test split & random-state

dataset used



Untitled.ipynb X

LAPTOP-11111111

```
[30]: # Create a SageMaker experiment
cc_experiment = Experiment(
    experiment_name='Build-train-deploy',
    description='Payments data',
    sagemaker_boto3_session=boto3.Session(),
)
print(cc_experiment)
```

```
[31]: # Start Tracking parameters
with Tracker.create_tracker() as tracker:
    tracker.log_parameters(
        "train_test_random_state"
    )
    # we can log things
    tracker.log_inference_endpoints()
    tracker.log_inference_endpoints()
    tracker.log_inference_endpoints()
```

```
# create a Trial with Xgboost
from sagemaker.amazon.amazon_estimator import get_image_uri
container = get_image_uri(boto3.Session().region_name, 'xgboost', '1.0-1') # Docker container to train Xgboost model
s3_input_train = sagemaker.s3_input(s3_data='s3://' + train_data_location, content_type='csv')
preprocessing_trial_component = tracker.trial_component

# create trial
trial_name = f"cc-default-training-job-{int(time.time())}"
cc_trial = Trial.create(
    trial_name=trial_name,
    experiment_name=cc_experiment.experiment_name,
    sagemaker_boto_client=sm
)

cc_trial.add_trial_component(preprocessing_trial_component)
cc_training_job_name = "cc-training-job-{}".format(int(time.time()))

xgb = sagemaker.estimator.Estimator(container,
                                      role,
                                      train_instance_count=1,
                                      train_instance_type='ml.m5.xlarge',
                                      train_max_run=86400,
                                      output_path='s3://{}{}'.format(rawbucket, prefix),
                                      sagemaker_session=sess) # set to true for distributed training
```

•

•

•

```
xgb.set_hyperparameters(max_depth=5,
                        eta=0.2,
                        gamma=4,
                        min_child_weight=6,
                        subsample=0.8,
                        verbosity=0,
                        objective='binary:logistic',
                        num_round=100)

xgb.fit(inputs = {'train':s3_input_train},
        job_name=cc_training_job_name,
        experiment_config={
            "TrialName": cc_trial.trial_name, #log training job in Trials for lineage
            "TrialComponentDisplayName": "Training",
        },
        wait=True,
    )
time.sleep(2)
```

Amazon SageMaker X

Amazon SageMaker Studio

Dashboard

Search

Images

▶ Ground Truth

▶ Notebook

▶ Processing

▼ Training

- Algorithms
- Training jobs**
- Hyperparameter tuning jobs

▶ Inference

▶ Edge Manager

▶ Augmented AI

▶ AWS Marketplace

Amazon SageMaker > Training jobs

Training jobs

Search training jobs

Name	Creation time	Duration	Status
cc-training-job-1630235216	Aug 29, 2021 11:06 UTC	-	InProgress

A red arrow points from the "Training jobs" link in the sidebar to the "Training jobs" section in the main content area.

```
WARNING:sagemaker.amazon.amazon_estimator:'get_image_uri' method will be deprecated in favor of 'ImageURIProvider' class in SageMaker Python SDK v2.  
WARNING:sagemaker:'s3_input' class will be renamed to 'TrainingInput' in SageMaker Python SDK v2.  
WARNING:sagemaker.estimator:Parameter image_name will be renamed to image_uri in SageMaker Python SDK v2.  
INFO:sagemaker:Creating training-job with name: cc-training-job-1630235216  
2021-08-29 11:06:56 Starting - Starting the training job...  
2021-08-29 11:06:58 Starting - Launching requested ML instances...  
2021-08-29 11:07:52 Starting - Preparing the instances for training.....  
2021-08-29 11:08:56 Downloading - Downloading input data...  
2021-08-29 11:09:07 Training - Downloading the training image...  
2021-08-29 11:09:55 Uploading - Uploading generated training model  
2021-08-29 11:09:55 Completed - Training job completed  
INFO:sagemaker-containers:Imported framework sagemaker_xgboost_container.training  
INFO:sagemaker-containers:Failed to parse hyperparameter objective value binary:logistic to Json.  
Returning the value itself  
INFO:sagemaker-containers>No GPUs detected (normal if no gpus installed)  
INFO:sagemaker_xgboost_container.training:Running XGBoost Sagemaker in algorithm mode  
INFO:root:Determined delimiter of CSV input is ','  
INFO:root:Determined delimiter of CSV input is ','  
[11:09:44] 24000x23 matrix with 552000 entries loaded from /opt/ml/input/data/train?format=csv&label_column=0&delimiter=,  
INFO:root:Single node training.  
INFO:root:Train matrix has 24000 rows
```

```
INFO:root:Determined delimiter of CSV input is ',  
[11:09:44] 24000x23 matrix with 552000 entries lo  
t=csv&label_column=0&delimiter=,  
INFO:root:Single node training.  
INFO:root:Train matrix has 24000 rows  
[0]#011train-error:0.17854  
[1]#011train-error:0.17746  
[2]#011train-error:0.17704  
[3]#011train-error:0.17717  
[4]#011train-error:0.17621  
[5]#011train-error:0.17608  
[6]#011train-error:0.17504  
[7]#011train-error:0.17467  
[8]#011train-error:0.17513  
[9]#011train-error:0.17475  
[10]#011train-error:0.17408  
[11]#011train-error:0.17354  
[12]#011train-error:0.17363  
[13]#011train-error:0.17313  
[14]#011train-error:0.17325  
[15]#011train-error:0.17296  
[16]#011train-error:0.17292
```

```
[90]#011train-error:0.16017  
[91]#011train-error:0.16000  
[92]#011train-error:0.16029  
[93]#011train-error:0.16012  
[94]#011train-error:0.16025  
[95]#011train-error:0.15975  
[96]#011train-error:0.15983  
[97]#011train-error:0.15967  
[98]#011train-error:0.15937  
[99]#011train-error:0.15887  
Training seconds: 59  
Billable seconds: 59 ↗
```

for ml.m5.xlarge
box

Amazon SageMaker Studio

File Edit View Run

SageMaker resources
Select the resource to view.

Experiments and trials

Build-train-deploy-1630234090

TRIALS
0 rows selected 0/20 filters

Search column name to start

Name	Last modified
cc-default-training-job-1630... 3 minutes ago	

End of the list

1

Amazon SageMaker Studio

File Edit View Run

SageMaker resources
Select the resource to view.

Experiments and trials

Build-train-deploy-1630234090 / cc-default-training-job-1630235216

TRIAL COMPONENTS
0 rows selected 0/20 filters

Search column name to start

Name	Last modified
Training 4 minutes ago	
Preprocessing 22 minutes ago	

End of the list

2

Untitled.ipynb X  Describe Trial Component X

C half a minute ago Stop training job

Experiment: Build-train-deploy-1630234090

Trial: cc-default-training-job-1630235216

Trial Component Created: 8 minutes ago

Trial Component Status: Completed

Training job detailed status: Completed

Trial components: Training ▾

Charts Metrics Parameters Artifacts AWS settings Debugger Model explainability Bias report Trial Mappings

Name	Minimum	Maximum	Standard Deviation	Final value
train:error	0.15887	0.17854	0.005174460853748282	0.15887

Offline batch Inference :

```
# Copy test-data from S3 to local box for batch inference
test_data_path = 's3://' + test_data_location + '/test_data.csv'
! aws s3 cp $test_data_path .

download: s3://sagemaker-us-west-2-047419640029/sagemaker-modelmonitor/test_data/test_data.csv to ./test_data.csv
```

```
test_full = pd.read_csv('test_data.csv', names = [str(x) for x in range(len(data.columns))]) # column-names are 0,1,2,3....
test_full.head()

   0      1      2      3      4      5      6      7      8      9     ...     14     15     16     17     18     19     20     21     22     23
0 0 -0.341476  0.201175  20000.0  1.0  1.0  2.0  33.0  1.0  2.0  ...  17399.0  19057.0  18453.0  19755.0  19288.0  2260.0  0.0  1600.0  0.0  644.0
1 0 -0.136859  0.199594  20000.0  2.0  2.0  2.0  35.0  0.0  0.0  ...  19347.0  18600.0  19000.0  19000.0  20000.0  0.0  1000.0  0.0  1000.0  0.0
2 0 -0.284364  0.185736  230000.0  2.0  1.0  1.0  44.0  1.0 -1.0  ...  949.0   2864.0   933.0    0.0    0.0  2873.0  933.0  0.0  0.0  0.0
3 0 -0.040569  0.289360  100000.0  1.0  2.0  1.0  42.0  0.0  0.0  ...  99998.0  16138.0  17758.0  18774.0  20272.0  2000.0  2000.0  2000.0  2000.0  2000.0
4 0  0.079132  0.186502  150000.0  1.0  1.0  2.0  29.0 -2.0 -2.0  ...  6917.0   831.0   6469.0   5138.0   7810.0   833.0  6488.0  5153.0  7833.0  7130.0

5 rows × 24 columns

label = test_full['0']
```

AWS Services ▾

Search for services, features, marketplace products, and docs [Option+S]

Srikanth Varma Chekuri ▾ Oregon ▾ Support ▾

Amazon SageMaker X

Amazon SageMaker Studio

Dashboard

Search

Images

▶ Ground Truth

▶ Notebook

▶ Processing

▼ Training

- Algorithms
- Training jobs
- Hyperparameter tuning jobs

▼ Inference

- Compilation jobs
- Model packages
- Models
- Endpoint configurations
- Endpoints
- Batch transform jobs**

▶ Edge Manager

▶ Augmented AI

▶ AWS Marketplace

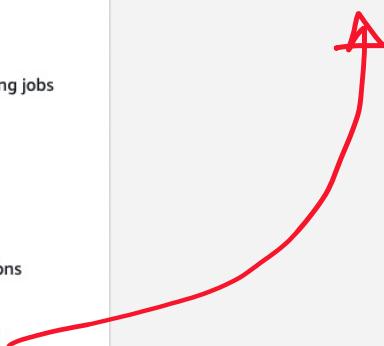
Amazon SageMaker > Batch transform jobs

Batch transform jobs

Search batch transform jobs

Name	Status	Duration	Creation time
sagemaker-xgboost-2021-08-29-11-32-49-876	InProgress	4 minutes	Aug 29, 2021 11:32 UTC
sagemaker-xgboost-2021-08-29-11-28-07-035	Completed	4 minutes	Aug 29, 2021 11:28 UTC
sagemaker-xgboost-2021-08-29-11-27-16-674	Completed	4 minutes	Aug 29, 2021 11:27 UTC

< 1 > ⚙



```
[46]: import json
import io
from urllib.parse import urlparse

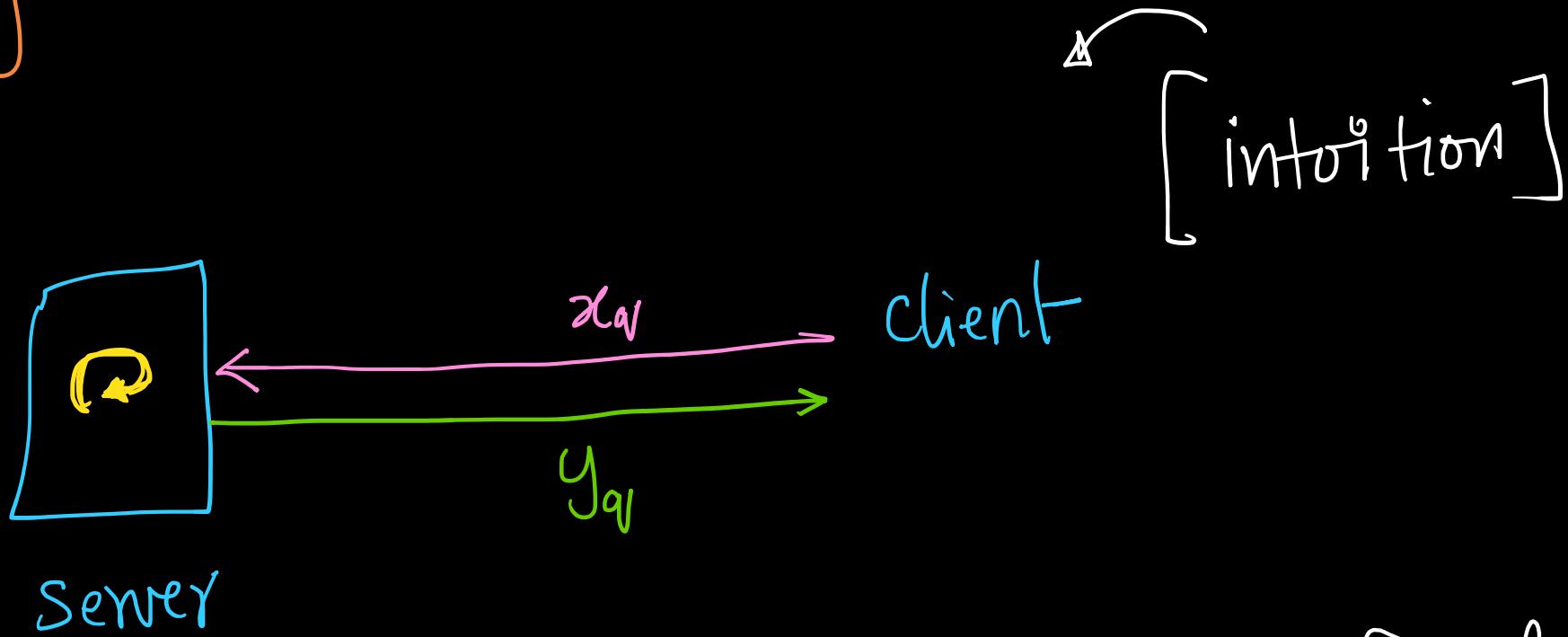
def get_csv_output_from_s3(s3uri, file_name):
    parsed_url = urlparse(s3uri)
    bucket_name = parsed_url.netloc
    prefix = parsed_url.path[1:]
    s3 = boto3.resource('s3')
    obj = s3.Object(bucket_name, '{}/{}'.format(prefix, file_name))
    return obj.get()["Body"].read().decode('utf-8')
output = get_csv_output_from_s3(sm_transformer.output_path, 'test_data.csv.out')
output_df = pd.read_csv(io.StringIO(output), sep=",", header=None)
output_df.head(8)
output_df['Predicted']=np.round(output_df.values)
output_df['Label'] = label
from sklearn.metrics import confusion_matrix, accuracy_score
confusion_matrix = pd.crosstab(output_df['Predicted'], output_df['Label'], rownames=['Actual'], colnames=['Predicted'], margins = True)
confusion_matrix
```

	Predicted	0	1	All
Actual				
0.0	4460	815	5275	
1.0	243	482	725	
All	4703	1297	6000	

} Not a great model

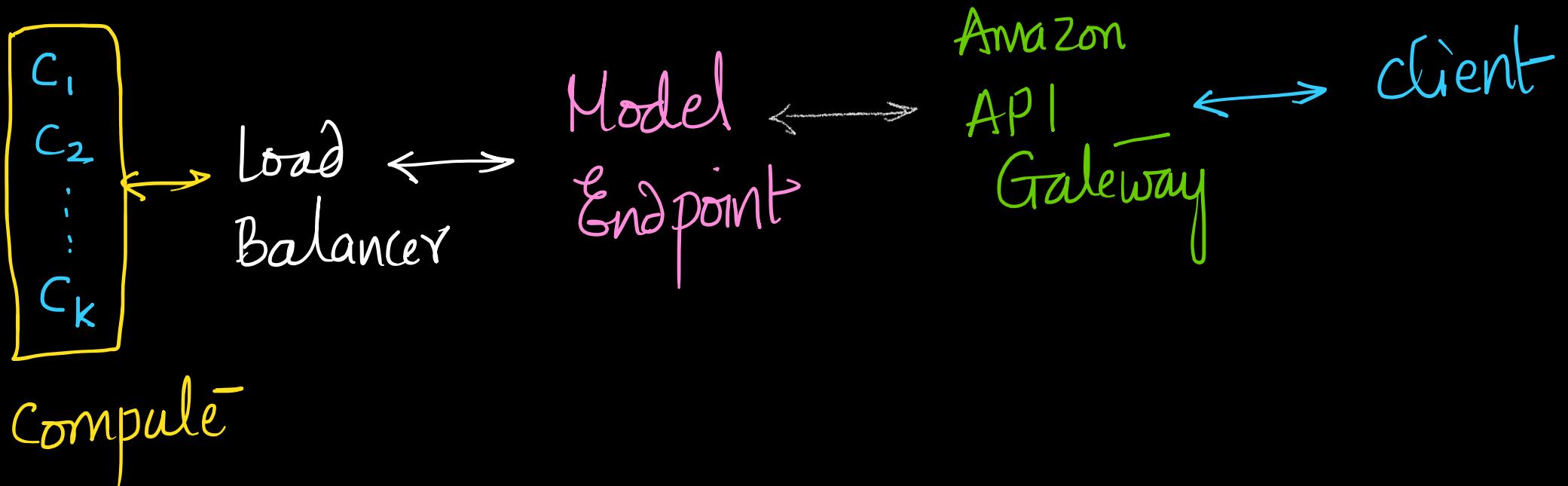
Hyper-param tuning using Bayesian opt
(later)

Deploy as an API (over HTTPS)



- calling a remote function

AWS - Model Endpoints



Code, links & references
in the Notebook itself

Distributed Training

<https://docs.aws.amazon.com/sagemaker/latest/dg/distributed-training.html>

- ↳ inbuilt library of popular Models (Xgboost , by SM)
- TensorFlow/ PyTorch/ MxNet (Most - popular use - case)
- Spark ML

Each algorithm solves a type of prediction problem

Classification

- Linear Learner *
- XGBoost *
- KNN
- Factorization Machines

Working with Text

- Blazing Text
 - Supervised
 - Unsupervised *

Sequence Translation

- Seq2Seq *

Computer Vision

- Image Classification <>
- Object Detection <>
- Semantic Segmentation

Recommendation

- Factorization Machines * (+ KNN)

Anomaly Detection

- Random Cut Forests *
- IP Insights *

Regression

- Linear Learner
- XGBoost
- KNN

Topic Modeling

- LDA
- NTM

Forecasting

- DeepAR *

Clustering

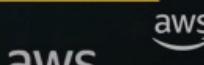
- Kmeans *
- KNN

Feature Reduction

- PCA
- Object2Vec

* = distributed training

<> = incremental training



<https://youtu.be/CDg55-Gklm4>

Xgboost

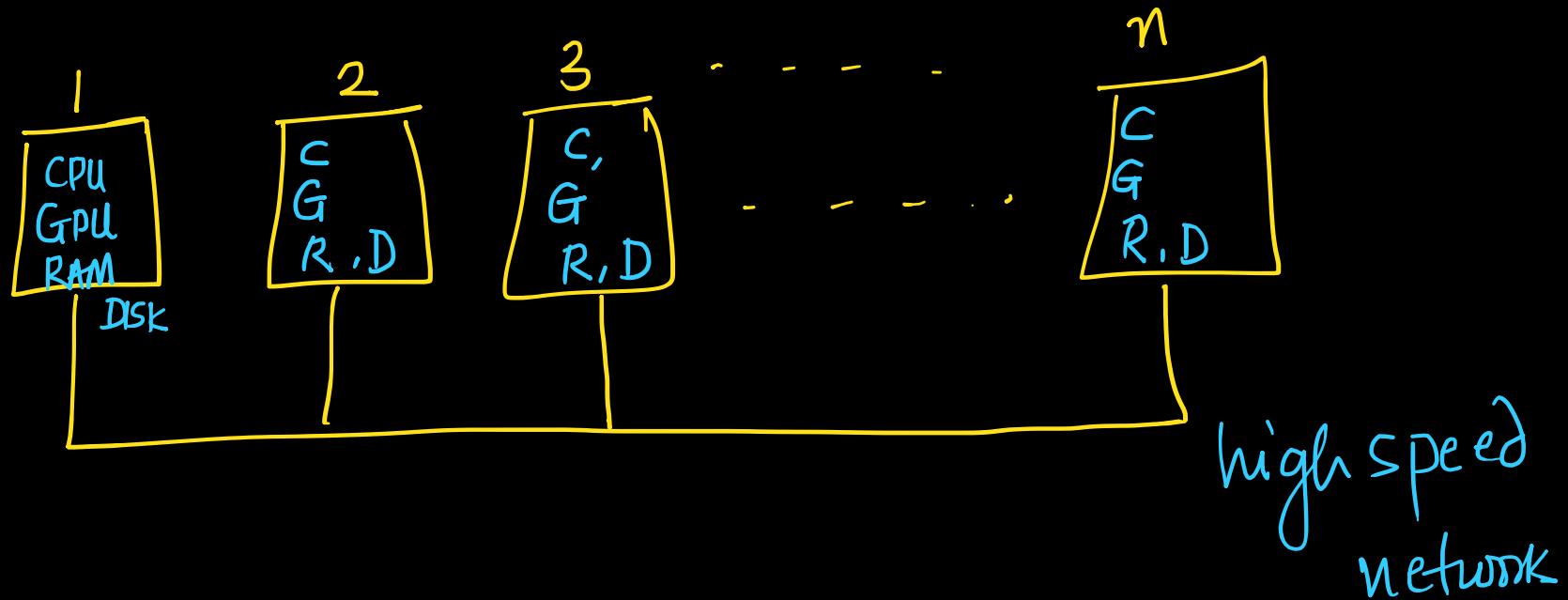
Tensorflow 2

PyTorch

SparkML



→ implicitly distributed / parallel

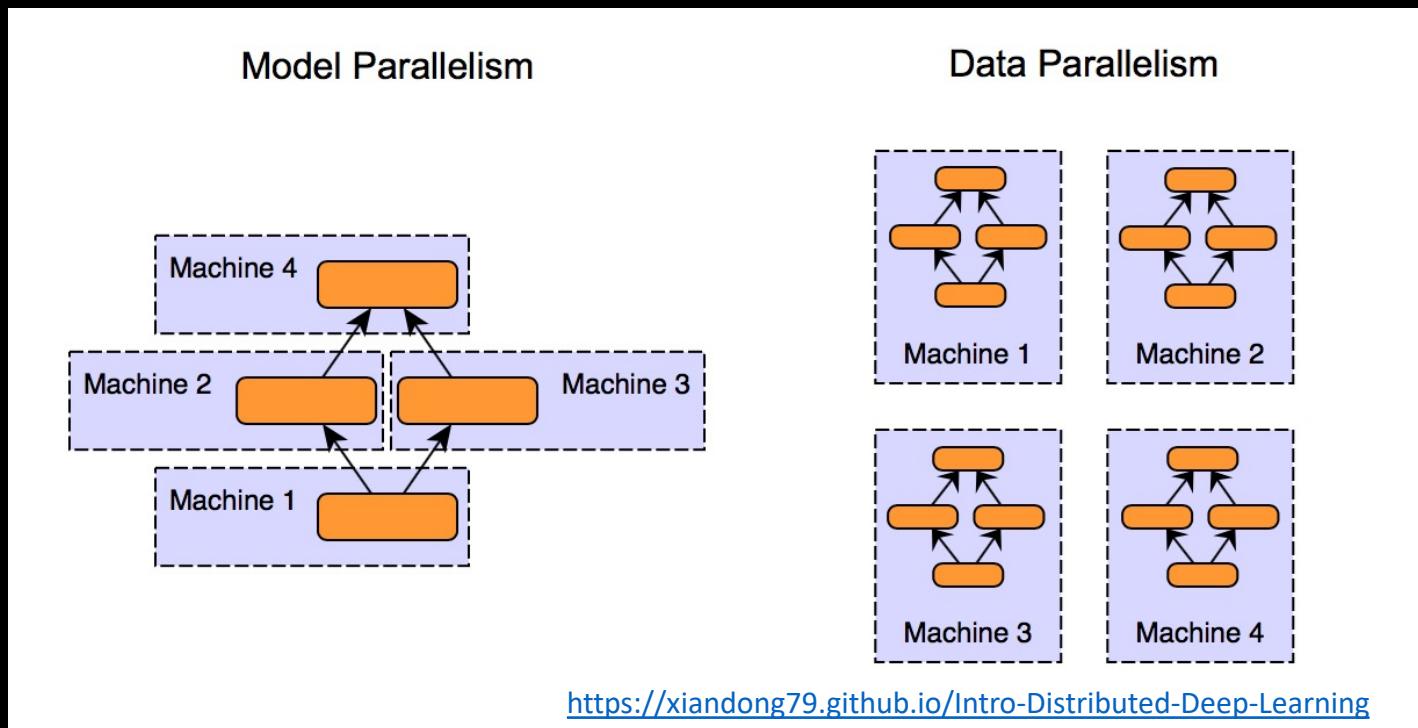


Deep-Learning (recap)

→ epoch vs iteration

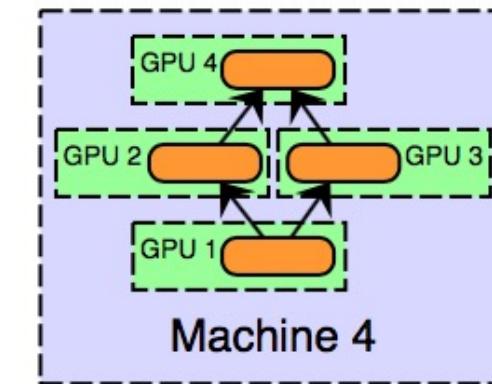
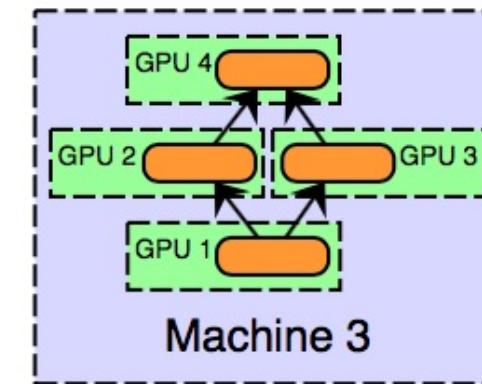
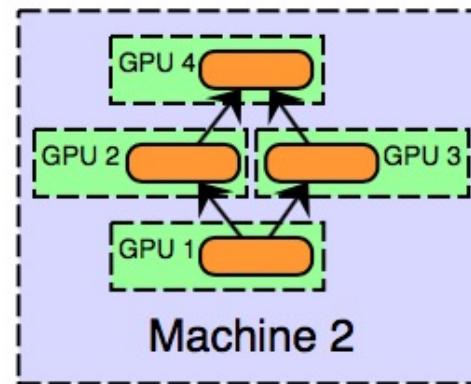
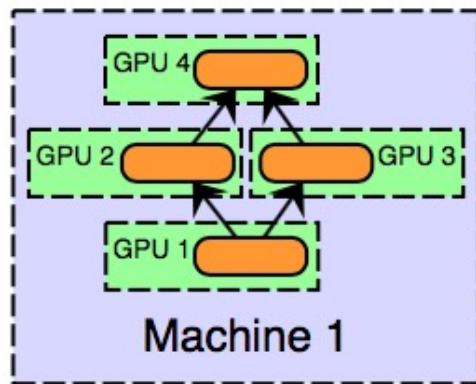
→ batch-size (aka global batch size)

Data Parallelism vs Model Parallelism



<https://www.amazon.science/latest-news/the-science-of-amazon-sagemakers-distributed-training-engines>

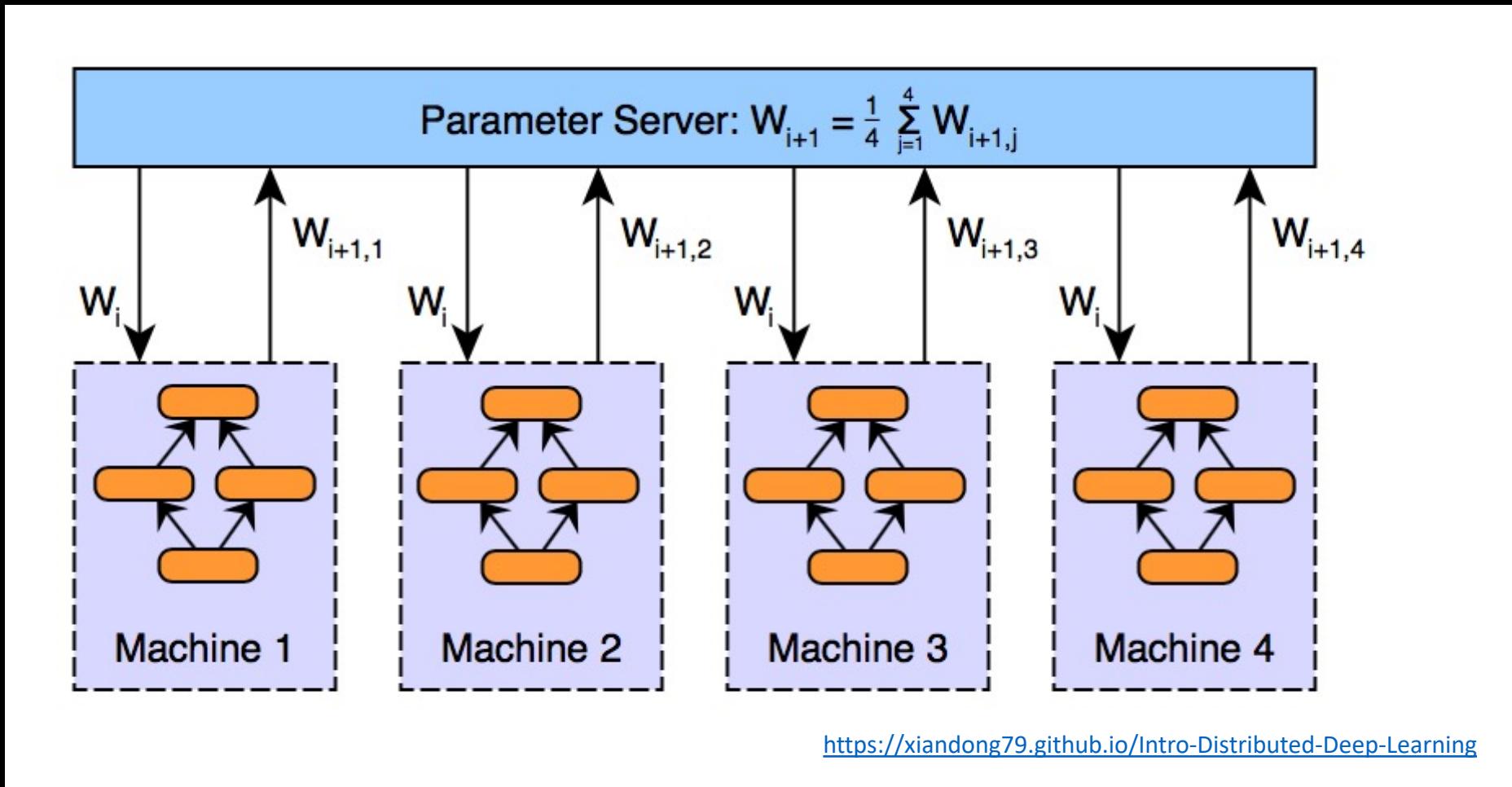
Model and Data Parallelism



<https://xiandong79.github.io/Intro-Distributed-Deep-Learning>

Multi-GPU box

Data Parallelism



All-reduce : [Dala Parallelism]

<https://docs.aws.amazon.com/sagemaker/latest/dg/data-parallel-intro.html>

SageMaker Data Parallel + TensorFlow 2 (SDP)

https://sagemaker-examples.readthedocs.io/en/latest/training/distributed_training/tensorflow/data_parallel/mnist/tensorflow2_smdataparallel_mnist_demo.html

SageMaker Model Parallelism (SMP)

- Automated Model Splitting (using CPU + RAM)
- Pipelined Execution (Interleaved)
- TensorFlow DAG

<https://docs.aws.amazon.com/sagemaker/latest/dg/model-parallel-core-features.html>

SageMaker Model Parallel + TensorFlow 2 (SMP)

https://sagemaker-examples.readthedocs.io/en/latest/training/distributed_training/tensorflow/model_parallel/mnist/tensorflow_sm_modelparallel_mnist.html

<https://docs.aws.amazon.com/sagemaker/latest/dg/model-parallel-customize-training-script-tf.html>

Distributed Xgboost on SageMaker :

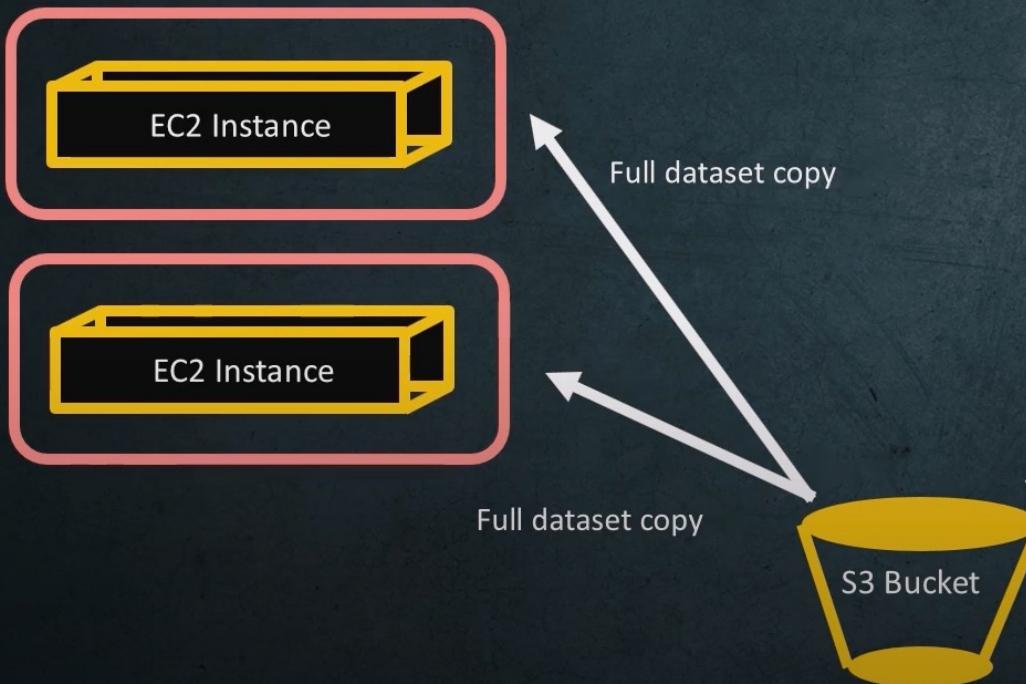
<https://engineering.zalando.com/posts/2020/06/distributed-xgb-sagemaker.html>

- num_instances
- tree_method
- distribution='ShardedByS3Key'

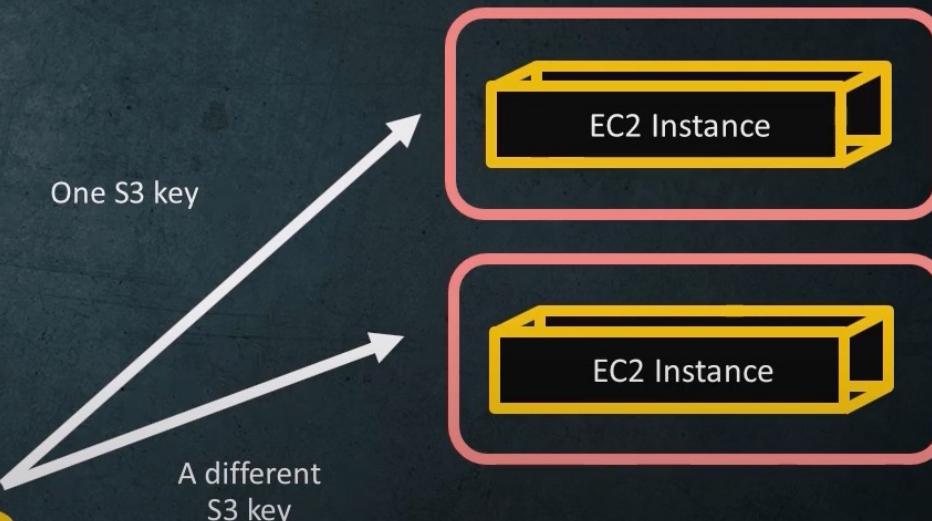
<https://xgboost.readthedocs.io/en/latest/treemethod.html>

(use 'approx'
or 'hist')

Fully Replicated



Sharded by S3 Key



aws



<https://youtu.be/CDg55-GkIm4>

SageMaker + Spark ML

<https://sagemaker-examples.readthedocs.io/en/latest/sagemaker-spark/index.html>

<https://sagemaker-pyspark.readthedocs.io/en/latest/api.html#algorithms>

