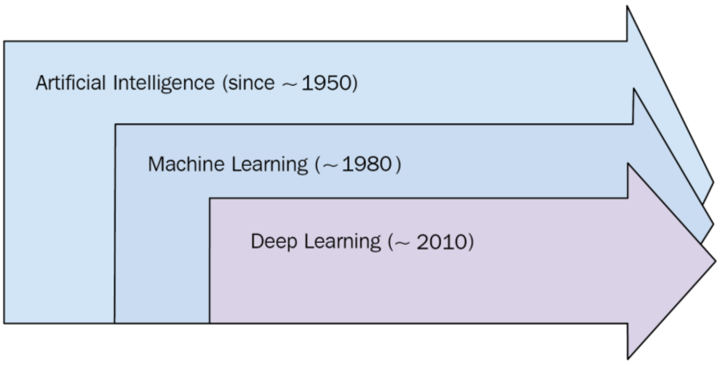
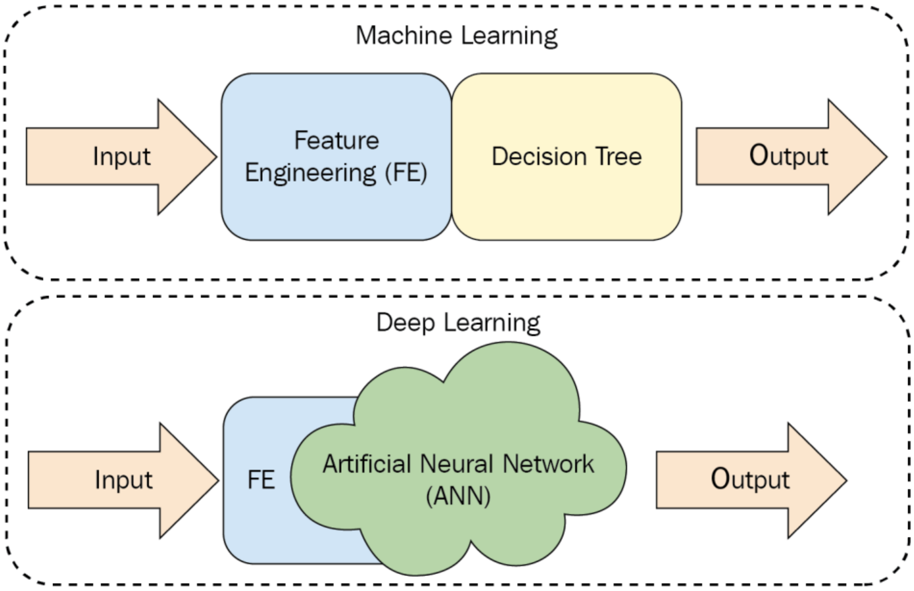
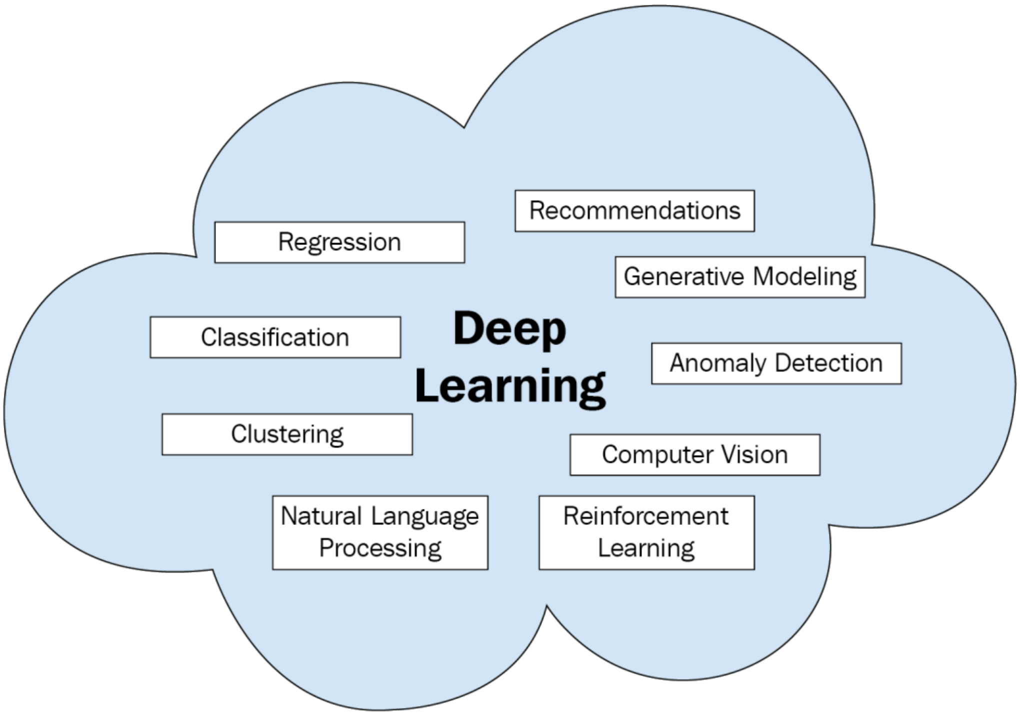
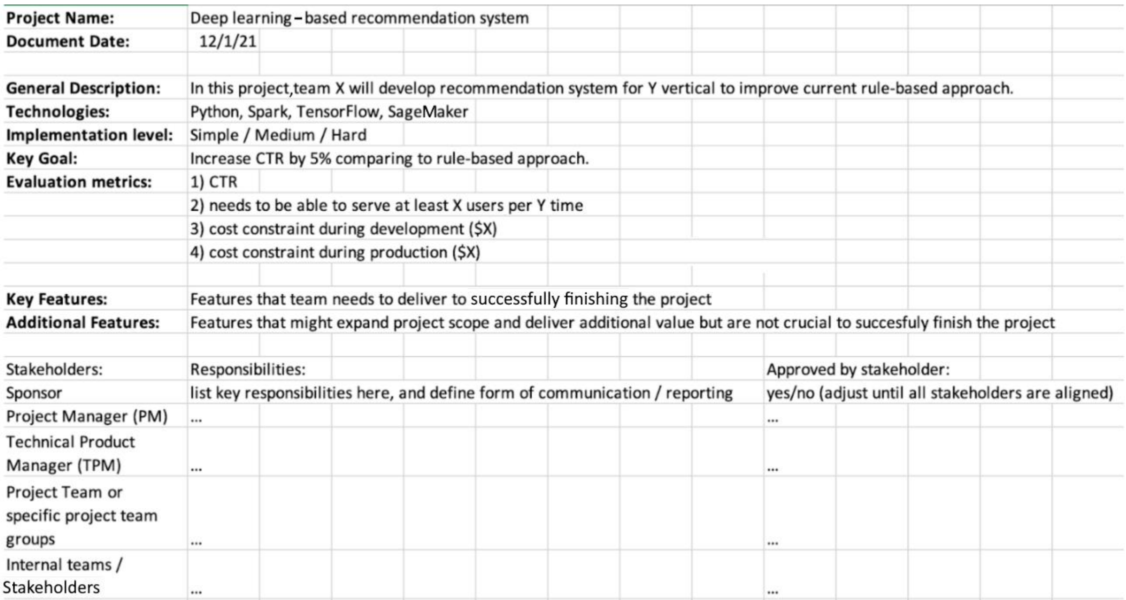
Chapter 1: Effective Planning of Deep Learning-Driven Projects



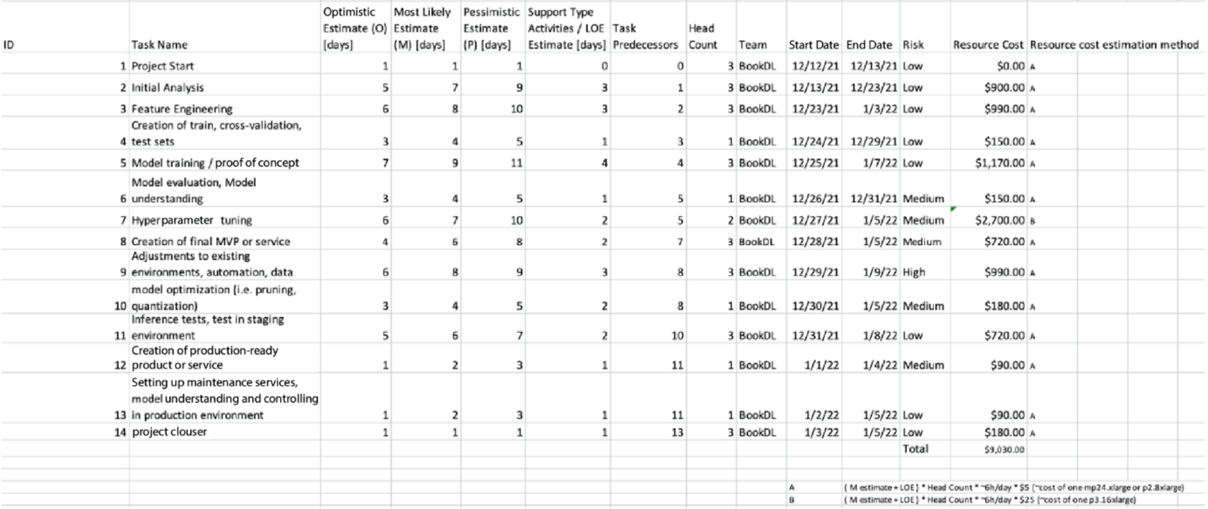




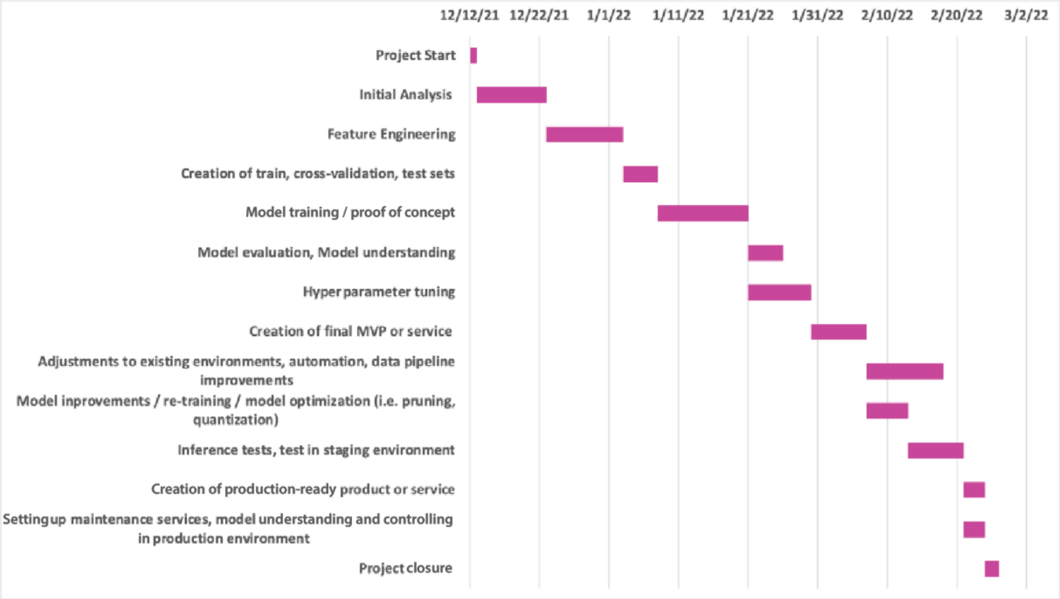


Project Name:	Deep learning – based recommendation system	
Document Date:	12/1/21	
General Description:	In this project,team X will develop recommendation system for Y vertical to improve current rule-based approach.	
Technologies:	Python, Spark, TensorFlow, SageMaker	
Implementation level:	Simple / Medium / Hard	
Key Goal:	Increase CTR by 5% comparing to rule-based approach.	
Evaluation metrics:	1) CTR	
	2) needs to be able to serve at least X users per Y time	
	3) cost constraint during development ($X)	
	4) cost constraint during production ($X)	
Key Features:	Features that team needs to deliver to successfully finishing the project	
Additional Features:	Features that might expand project scope and deliver additional value but are not crucial to succesfuly finish the project	
Stakeholders:	Responsibilities:	Approved by stakeholder:
Sponsor	list key responsibilities here, and define form of communication / reporting	yes/no (adjust until all stakeholders are aligned)
Project Manager (PM)	...	...
Technical Product Manager (TPM)	...	...
Project Team or specific project team groups	...	...
Internal teams / Stakeholders	...	...

ID	Task Name
1	Project Start
2	Initial Analysis
3	Feature Engineering
4	Creation of train, cross-validation, test sets
5	Model training / proof of concept
6	Model evaluation, Model understanding
7	Hyperparameter tuning
8	Creation of final MVP or service
9	Adjustments to existing environments, automation, data
10	model optimization (i.e. pruning, quantization)
11	Inference tests, test in staging environment
12	Creation of production-ready product or service
13	Setting up maintenance services, model understanding and controlling in production environment
14	Project Closure



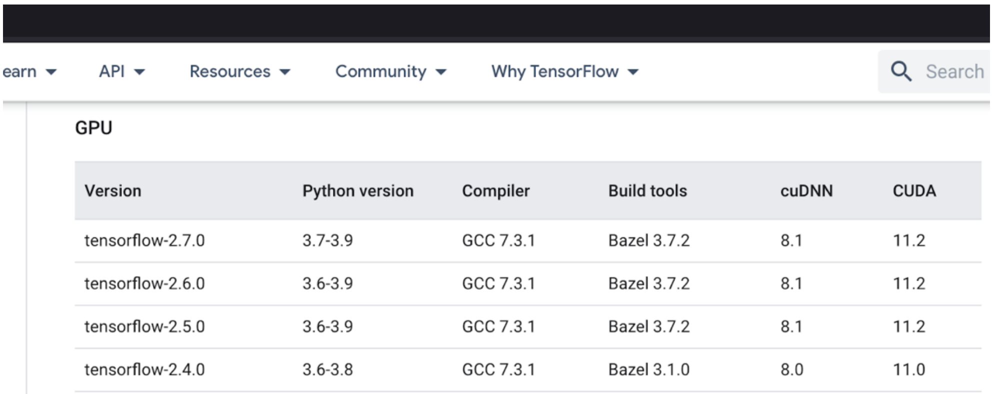
ID	Task Name	Optimistic Estimate (O) [days]	Most Likely Estimate (M) [days]	Pessimistic Estimate (P) [days]	Support Type Activities / LOE Estimate [days]	Task Predecessors	Head Count	Team	Start Date	End Date	Risk	Resource Cost	Resource cost estimation method
1	Project Start	1	1	1	0	0	3	BookDL	12/12/21	12/13/21	Low	$0.00	A
2	Initial Analysis	5	7	9	3	1	3	BookDL	12/13/21	12/23/21	Low	$900.00	A
3	Feature Engineering	6	8	10	3	2	3	BookDL	12/23/21	1/3/22	Low	$990.00	A
4	Creation of train, cross-validation, test sets	3	4	5	1	3	1	BookDL	12/24/21	12/29/21	Low	$150.00	A
5	Model training / proof of concept	7	9	11	4	4	3	BookDL	12/25/21	1/7/22	Low	$1,170.00	A
6	Model evaluation, Model understanding	3	4	5	1	5	1	BookDL	12/26/21	12/31/21	Medium	$150.00	A
7	Hyperparameter tuning	6	7	10	2	5	2	BookDL	12/27/21	1/5/22	Medium	$2,700.00	B
8	Creation of final MVP or service	4	6	8	2	7	3	BookDL	12/28/21	1/5/22	Medium	$720.00	A
9	Adjustments to existing environments, automation, data	6	8	9	3	8	3	BookDL	12/29/21	1/9/22	High	$990.00	A
10	model optimization (i.e. pruning, quantization)	3	4	5	2	8	1	BookDL	12/30/21	1/5/22	Medium	$180.00	A
11	Inference tests, test in staging environment	5	6	7	2	10	3	BookDL	12/31/21	1/8/22	Low	$720.00	A
12	Creation of production-ready product or service	1	2	3	1	11	1	BookDL	1/1/22	1/4/22	Medium	$90.00	A
13	Setting up maintenance services, model understanding and controlling in production environment	1	2	3	1	11	1	BookDL	1/2/22	1/5/22	Low	$90.00	A
14	project clouser	1	1	1	1	13	3	BookDL	1/3/22	1/5/22	Low	$180.00	A
											Total	$9,030.00	
									A	(M estimate + LOE) * Head Count * ~6h/day * $5 (~cost of one mp24.xlarge or p2.8xlarge)			
									B	(M estimate + LOE) * Head Count * ~6h/day * $25 (~cost of one p3.16xlarge)			



Stakeholder	Role
Sponsor	- Initiating the project - Defining a business justification for the project - Canceling the project when it is no longer needed
Project lead	- Motivating team members for the success of the project - Interacting with external stakeholders to make sure that the project is not delayed unexpectedly
Project manager	- Planning, monitoring, and ensuring the stable execution of the project - Analyzing risks - Making sure the project is on schedule
Data engineers	- Preprocessing the necessary data into a form that data scientists can use
Data scientists	- Analyzing the data and developing a model for the project
DevOps	- Migrating the model and data preprocessing logics to the cloud - Supporting software engineers with the deployment of the deliverable
Software engineers	- Developing the necessary tools for the project - Building the deliverable - Deploying the deliverable to the target users

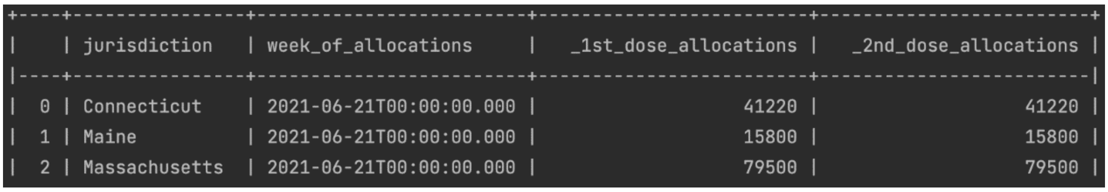
Stakeholder	Role
Data collector	Collecting the raw data that the project depends on
Labeling company	Labeling the raw data for model training
User	Interacting with the deliverable and providing feedback
C-suite executives	Allocating resources to the project

Chapter 2: Data Preparation for Deep Learning Projects

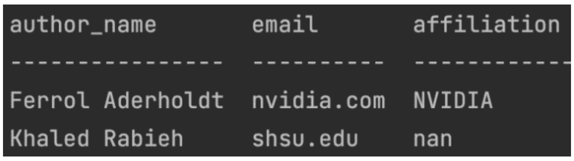


GPU

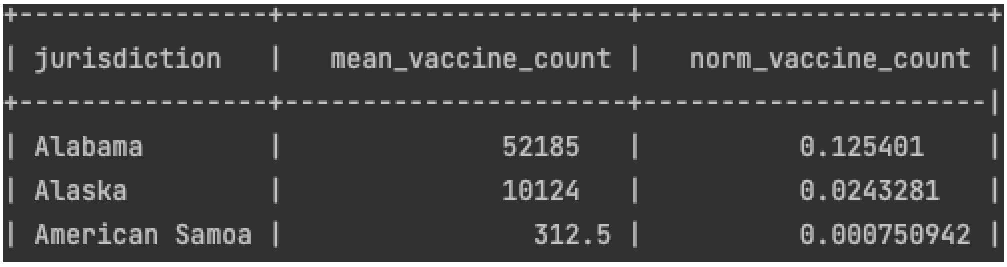
Version	Python version	Compiler	Build tools	cuDNN	CUDA
tensorflow-2.7.0	3.7-3.9	GCC 7.3.1	Bazel 3.7.2	8.1	11.2
tensorflow-2.6.0	3.6-3.9	GCC 7.3.1	Bazel 3.7.2	8.1	11.2
tensorflow-2.5.0	3.6-3.9	GCC 7.3.1	Bazel 3.7.2	8.1	11.2
tensorflow-2.4.0	3.6-3.8	GCC 7.3.1	Bazel 3.1.0	8.0	11.0



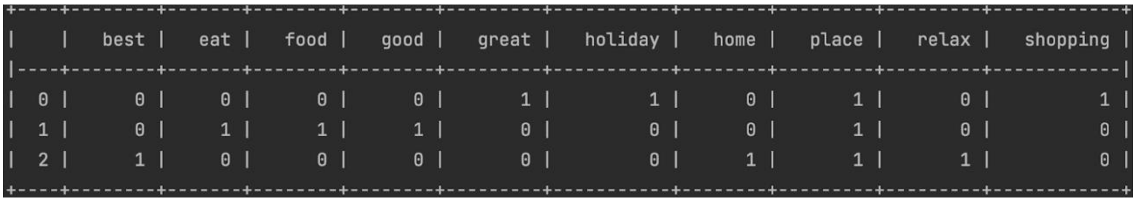
```
+----+---------------+-----------------------+-----------------------+-----------------------+
|    | jurisdiction  | week_of_allocations   |   _1st_dose_allocations |   _2nd_dose_allocations |
|----+---------------+-----------------------+-----------------------+-----------------------|
|  0 | Connecticut   | 2021-06-21T00:00:00.000 |                 41220 |                 41220 |
|  1 | Maine         | 2021-06-21T00:00:00.000 |                 15800 |                 15800 |
|  2 | Massachusetts | 2021-06-21T00:00:00.000 |                 79500 |                 79500 |
```



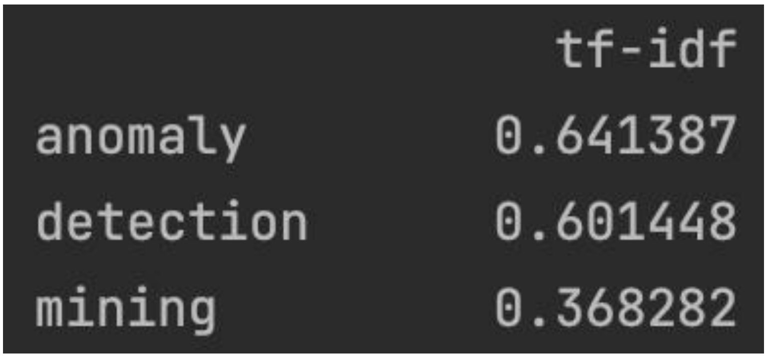
```
author_name       email       affiliation
----------------  ----------  -----------
Ferrol Aderholdt  nvidia.com  NVIDIA
Khaled Rabieh     shsu.edu    nan
```



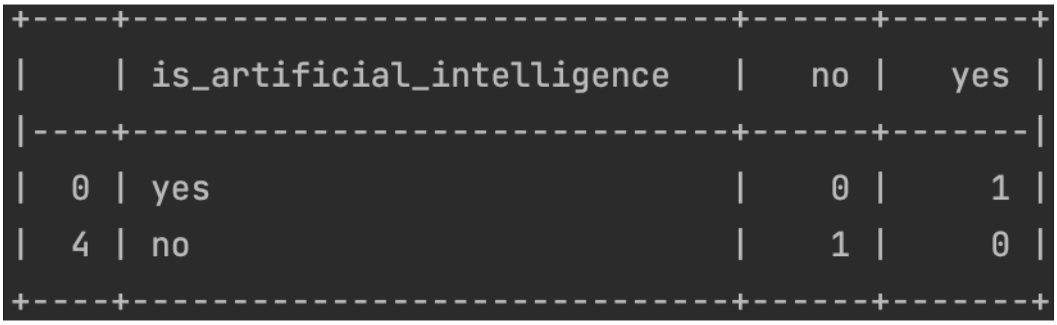
```
+----------------+--------------------+--------------------+
| jurisdiction   |   mean_vaccine_count |   norm_vaccine_count |
+----------------+--------------------+--------------------|
| Alabama        |              52185   |          0.125401    |
| Alaska         |              10124   |          0.0243281   |
| American Samoa |                312.5 |          0.000750942 |
```



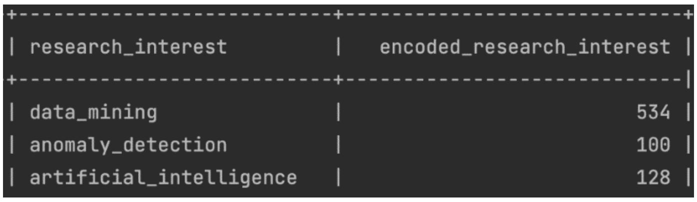
```
+----+--------+-------+--------+--------+---------+-----------+--------+---------+--------+------------+
|    |   best |   eat |   food |   good |   great |   holiday |   home |   place |   relax |   shopping |
|----+--------+-------+--------+--------+---------+-----------+--------+---------+--------+------------|
|  0 |      0 |     0 |      0 |      0 |       1 |         1 |      0 |       1 |       0 |          1 |
|  1 |      0 |     1 |      1 |      1 |       0 |         0 |      0 |       1 |       0 |          0 |
|  2 |      1 |     0 |      0 |      0 |       0 |         0 |      1 |       1 |       1 |          0 |
+----+--------+-------+--------+--------+---------+-----------+--------+---------+--------+------------+
```



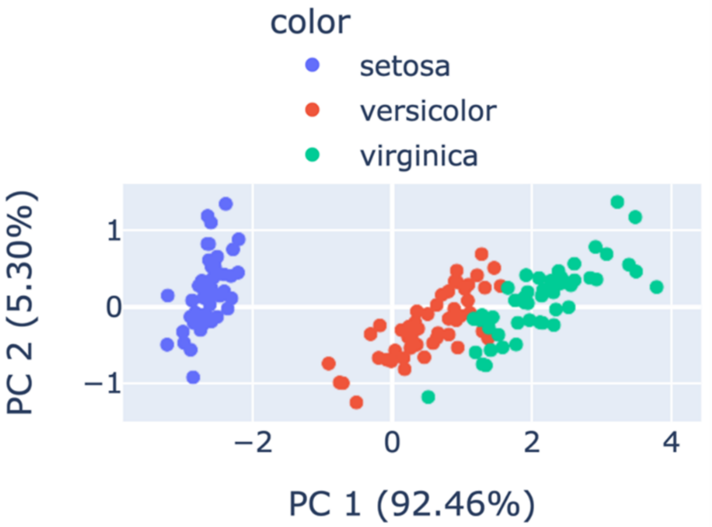
```
              tf-idf
anomaly     0.641387
detection   0.601448
mining      0.368282
```

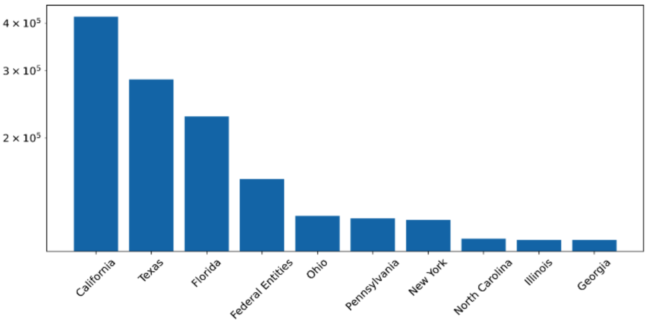


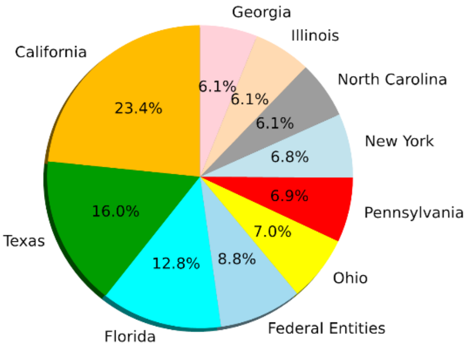
```
+----+---------------------------+-----+-------+
|    | is_artificial_intelligence  |  no |   yes |
|----+---------------------------+-----+-------|
|  0 | yes                         |   0 |     1 |
|  4 | no                          |   1 |     0 |
+----+---------------------------+-----+-------+
```

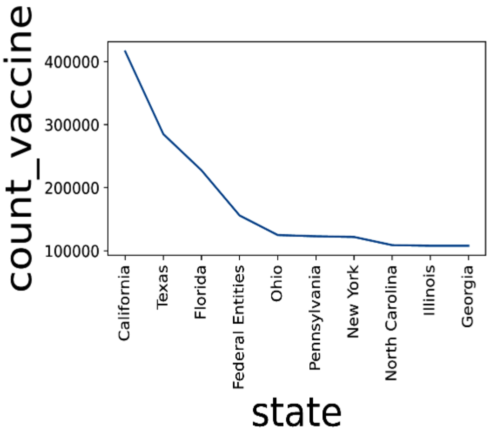


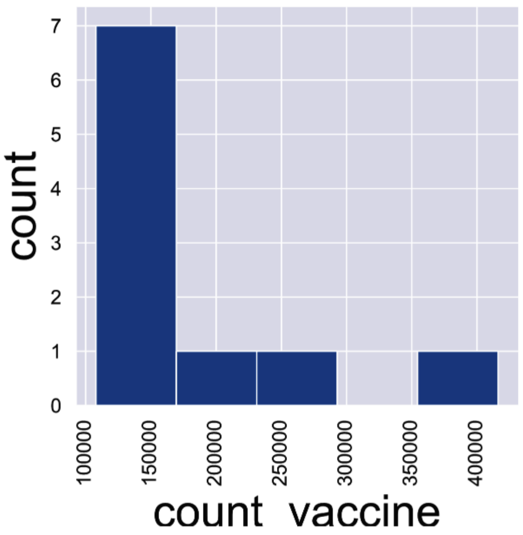
```
+-------------------------+-------------------------+
| research_interest       |  encoded_research_interest |
+-------------------------+-------------------------|
| data_mining             |                     534 |
| anomaly_detection       |                     100 |
| artificial_intelligence |                     128 |
```

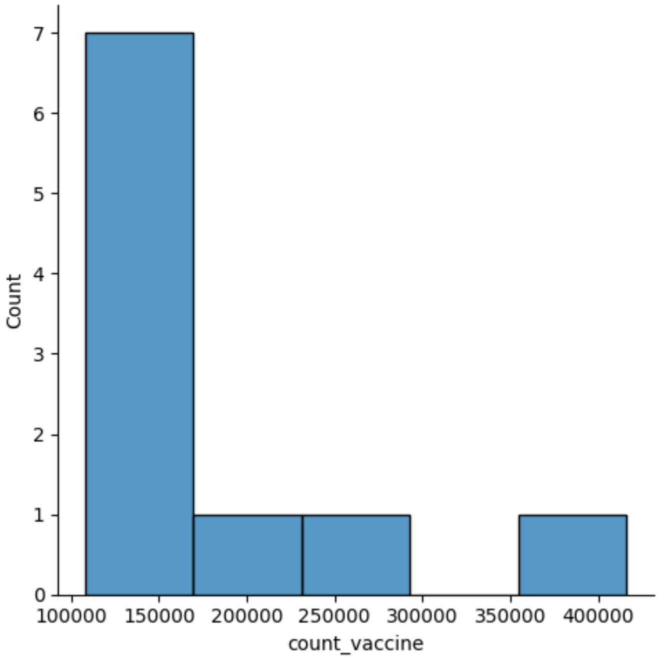




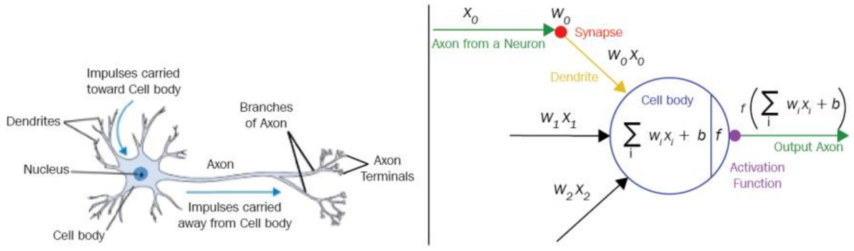



count_vaccine
400000
300000
200000
100000
California
Texas
Florida
Federal Entities
Ohio
Pennsylvania
New York
North Carolina
Illinois
Georgia
state

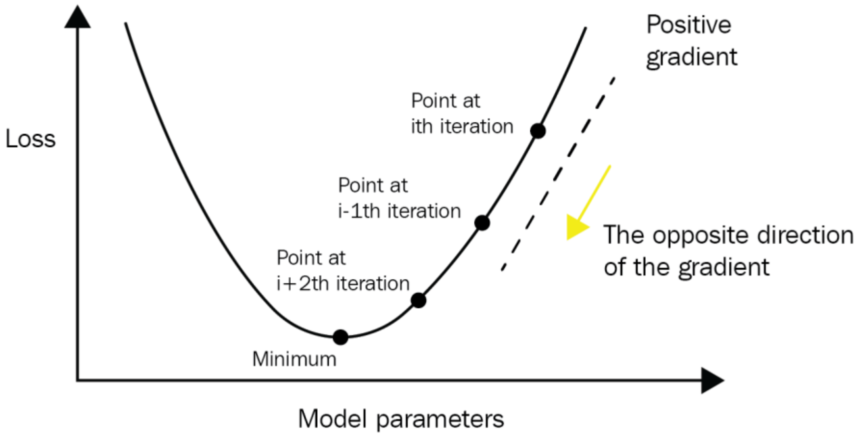

count
7
6
5
4
3
2
1
0
100000
150000
200000
250000
300000
350000
400000
count_vaccine

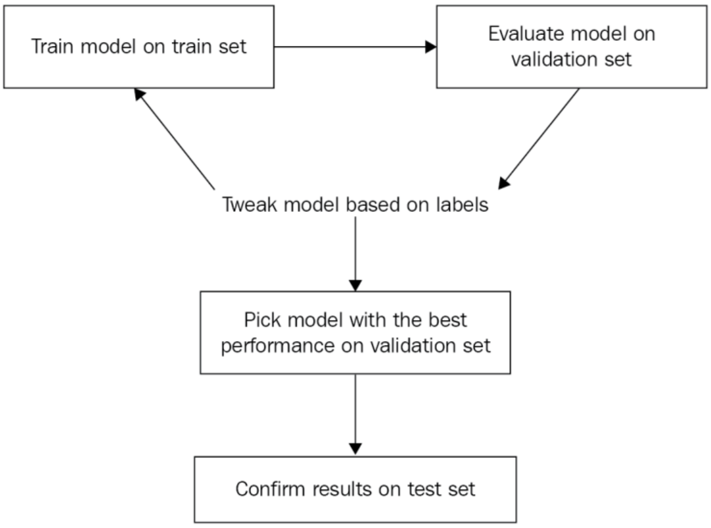

Count
7
6
5
4
3
2
1
0
100000
150000
200000
250000
300000
350000
400000
count_vaccine

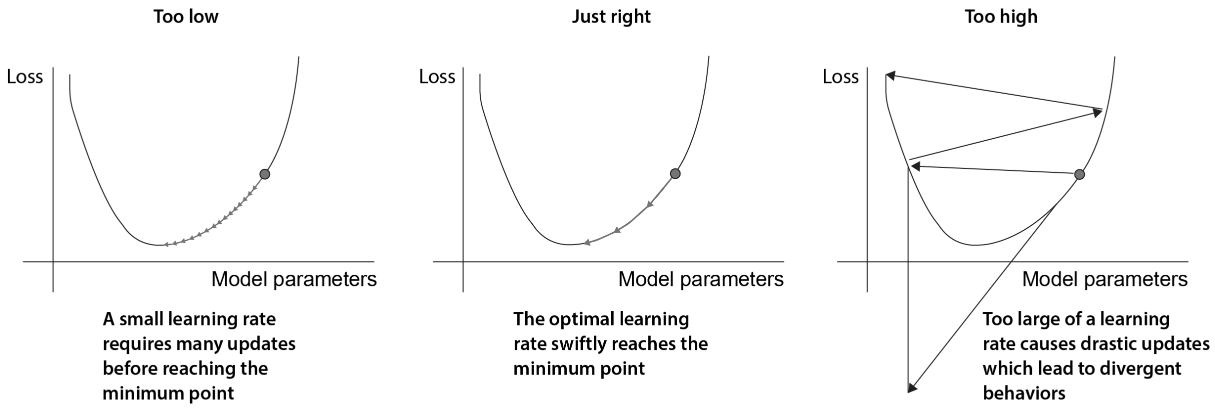
Chapter 3: Developing a Powerful Deep Learning Model

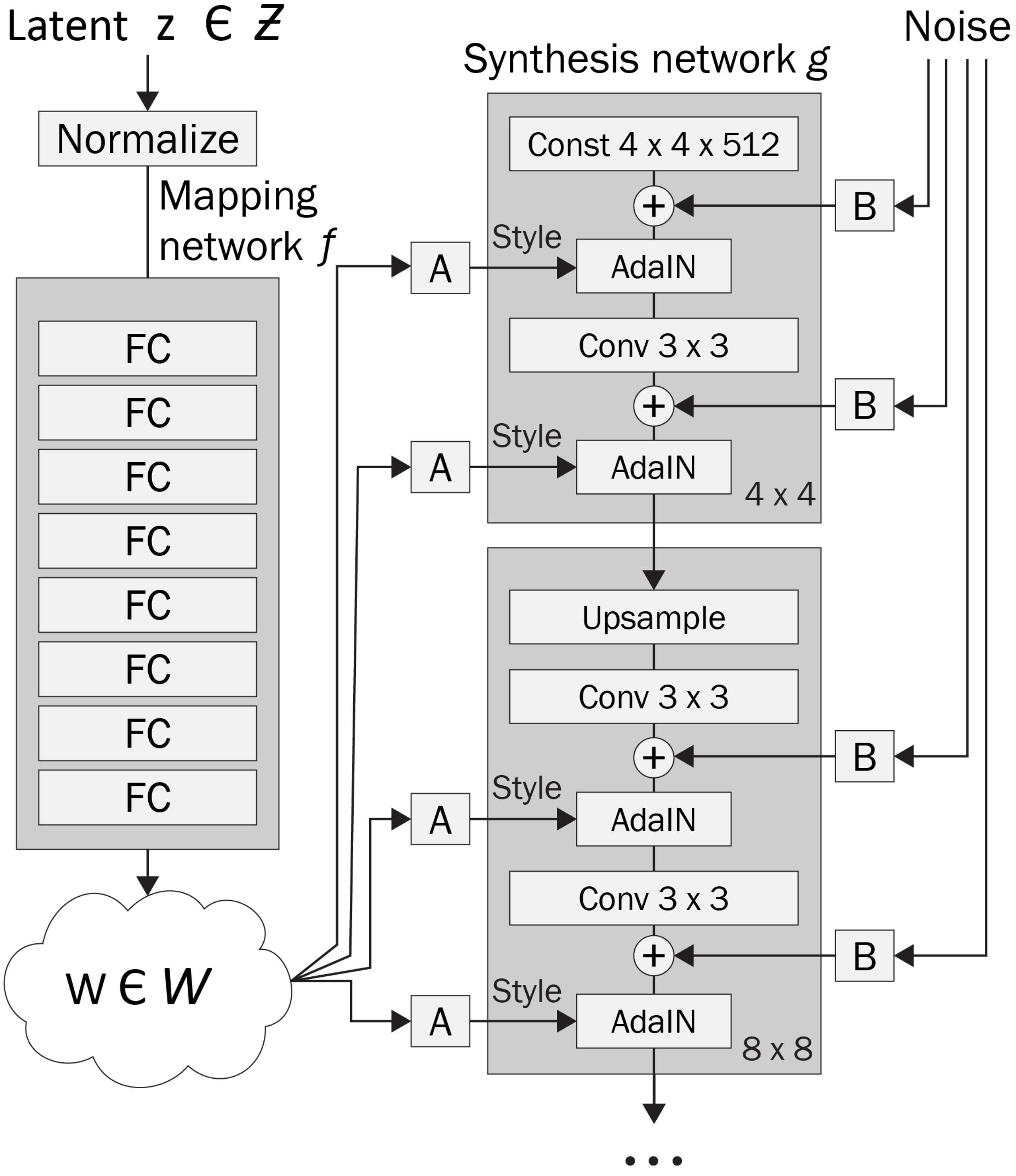


Drawing of a Biological Neuron (Left) and its Mathematical Model (Right).




Train model on train set
Evaluate model on validation set
Tweak model based on labels
Pick model with the best performance on validation set
Confirm results on test set

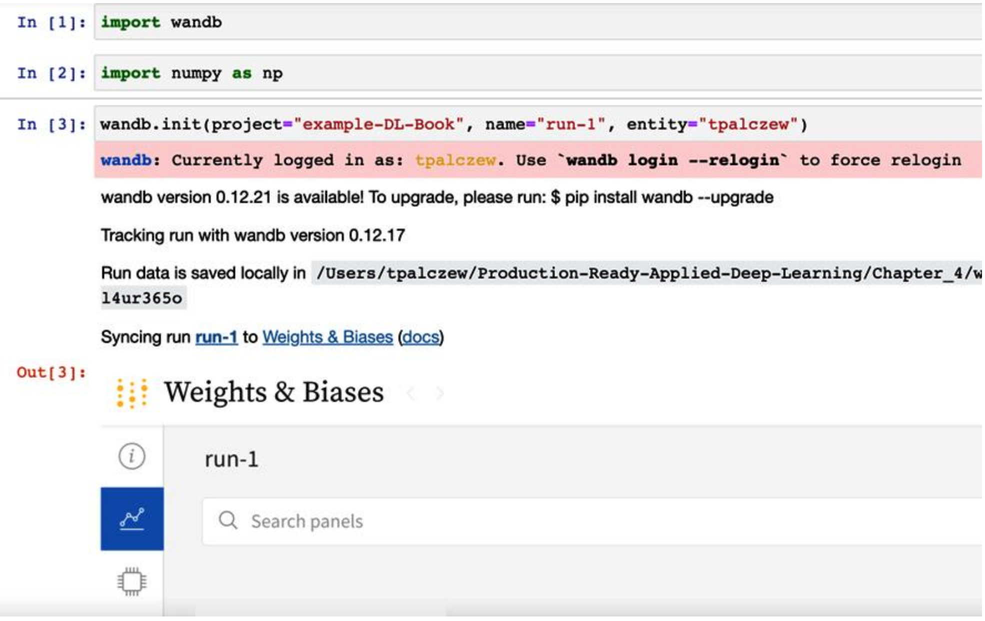

Too low
Loss
Model parameters
A small learning rate requires many updates before reaching the minimum point
Just right
Loss
Model parameters
The optimal learning rate swiftly reaches the minimum point
Too high
Loss
Model parameters
Too large of a learning rate causes drastic updates which lead to divergent behaviors


Latent z ∈ Z
Normalize
Mapping network f
FC
FC
FC
FC
FC
FC
FC
FC
w ∈ W
Synthesis network g
Noise
Const 4 x 4 x 512
+
B
Style
A
AdaIN
Conv 3 x 3
+
B
Style
A
AdaIN
4 x 4
Upsample
Conv 3 x 3
+
B
Style
A
AdaIN
Conv 3 x 3
+
B
Style
A
AdaIN
8 x 8
• • •

Discriminator	Activation	Output Shape	Params
Input image	–	3 x 1024 x 1024	–
Conv 1 x 1	LReLU	16 x 1024 x 1024	64
Conv 3 x 3	LReLU	16 x 1024 x 1024	2.3k
Conv 3 x 3	LReLU	32 x 1024 x 1024	4.6k
Downsample	–	32 x 512 x 512	–
Conv 3 x 3	LReLU	32 x 512 x 512	9.2k
Conv 3 x 3	LReLU	64 x 512 x 512	18k
Downsample	–	64 x 256 x 256	–
Conv 3 x 3	LReLU	64 x 256 x 256	37k
Conv 3 x 3	LReLU	128 x 256 x 256	74k
Downsample	–	128 x 128 x 128	–
Conv 3 x 3	LReLU	128 x 128 x 128	148k
Conv 3 x 3	LReLU	256 x 128 x 128	295k
Downsample	–	256 x 64 x 64	–
Conv 3 x 3	LReLU	256 x 64 x 64	590k
Conv 3 x 3	LReLU	512 x 64 x 64	1.2M
Downsample	–	512 x 32 x 32	–
Conv 3 x 3	LReLU	512 x 32 x 32	2.4M
Conv 3 x 3	LReLU	512 x 32 x 32	2.4M
Downsample	–	512 x 16 x 16	–
Conv 3 x 3	LReLU	512 x 16 x 16	2.4M
Conv 3 x 3	LReLU	512 x 16 x 16	2.4M
Downsample	–	512 x 8 x 8	–
Conv 3 x 3	LReLU	512 x 8 x 8	2.4M
Conv 3 x 3	LReLU	512 x 8 x 8	2.4M
Downsample	–	512 x 4 x 4	–
Minibatch stddev	–	513 x 4 x 4	–
Conv 3 x 3	LReLU	512 x 4 x 4	2.4M
Conv 4 x 4	LReLU	512 x 1 x 1	4.2M
Fully-connected	linear	1 x 1 x 1	513
Total trainable parameters			23.1M

GPUs	1024x1024	512x512
1	41 days 4 hours	24 days 21 hours
2	21 days 22 hours	13 days 7 hours
4	11 days 8 hours	7 days 0 hours
8	6 days 14 hours	4 days 10 hours

Chapter 4: Experiment Tracking, Model Management, and Dataset Versioning



```
In [1]: import wandb

In [2]: import numpy as np

In [3]: wandb.init(project="example-DL-Book", name="run-1", entity="tpalczew")
        wandb: Currently logged in as: tpalczew. Use `wandb login --relogin` to force relogin
```

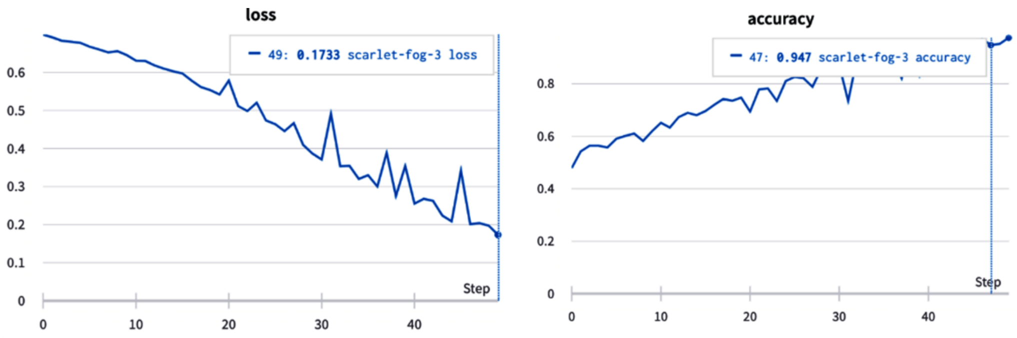
wandb version 0.12.21 is available! To upgrade, please run: $ pip install wandb --upgrade

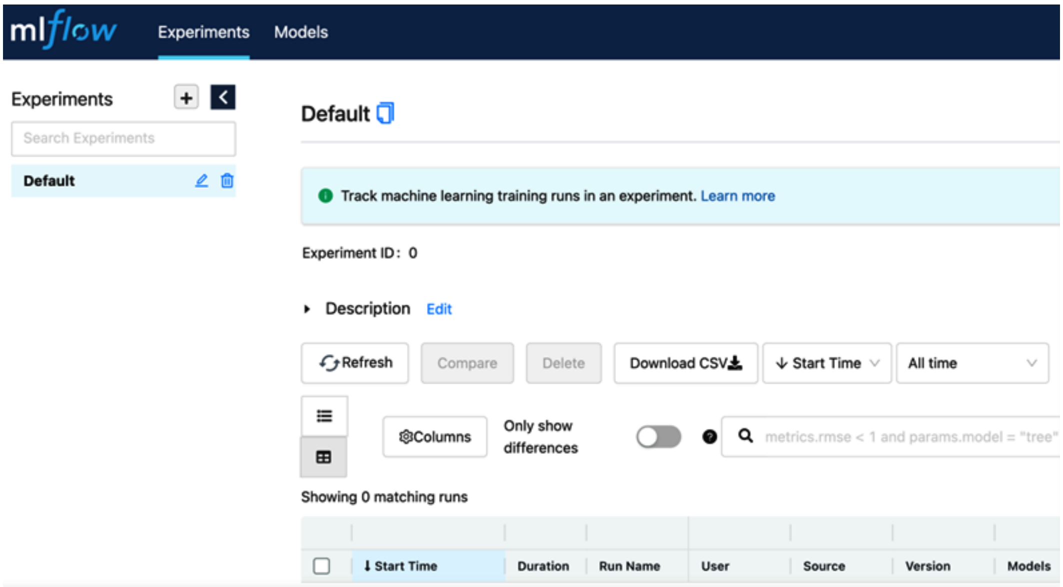
Tracking run with wandb version 0.12.17

Run data is saved locally in `/Users/tpalczew/Production-Ready-Applied-Deep-Learning/Chapter_4/w` `14ur365o`

Syncing run run-1 to Weights & Biases (docs)

Out[3]:





Default

Track machine learning training runs in an experiment. Learn more

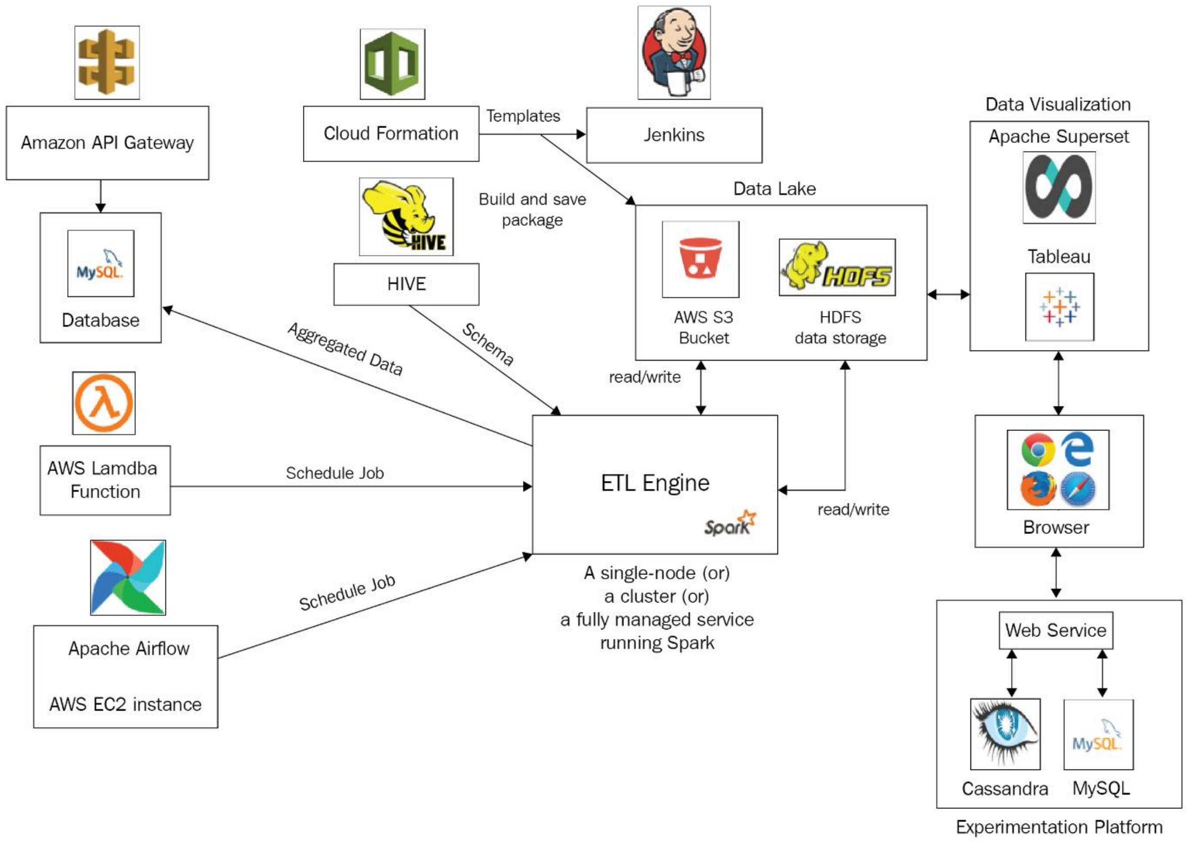
Experiment ID: 0

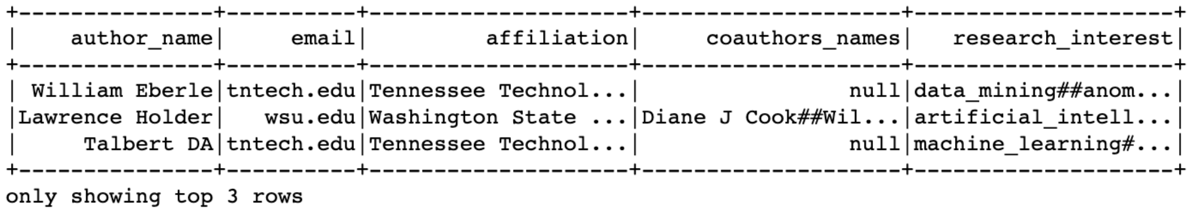
Description Edit

Showing 0 matching runs

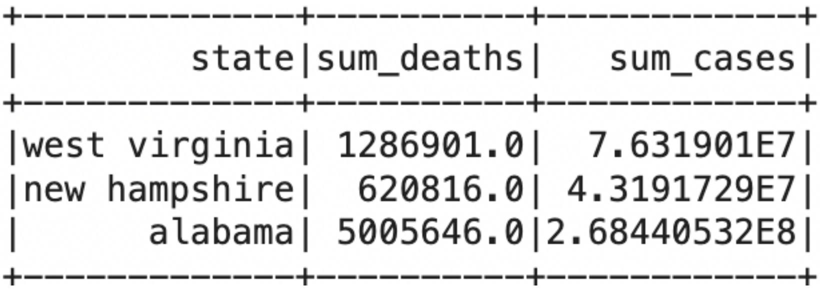
□	Start Time	Duration	Run Name	User	Source	Version	Models

Chapter 5: Data Preparation in the Cloud

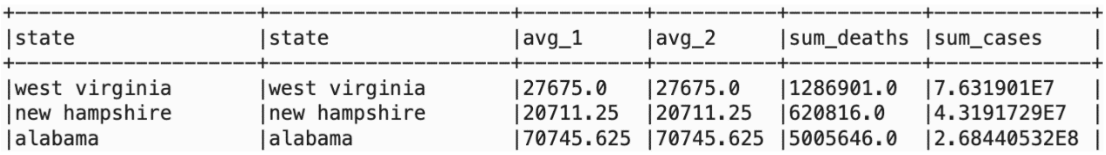




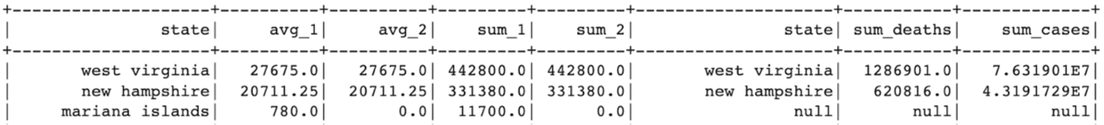
```
+---------------+----------+--------------------+--------------------+--------------------+
|    author_name|     email|         affiliation|     coauthors_names|   research_interest|
+---------------+----------+--------------------+--------------------+--------------------+
| William Eberle|tntech.edu|Tennessee Technol...|                null|data_mining##anom...|
|Lawrence Holder|   wsu.edu|Washington State ...|Diane J Cook##Wil...|artificial_intell...|
|     Talbert DA|tntech.edu|Tennessee Technol...|                null|machine_learning#...|
+---------------+----------+--------------------+--------------------+--------------------+
only showing top 3 rows
```



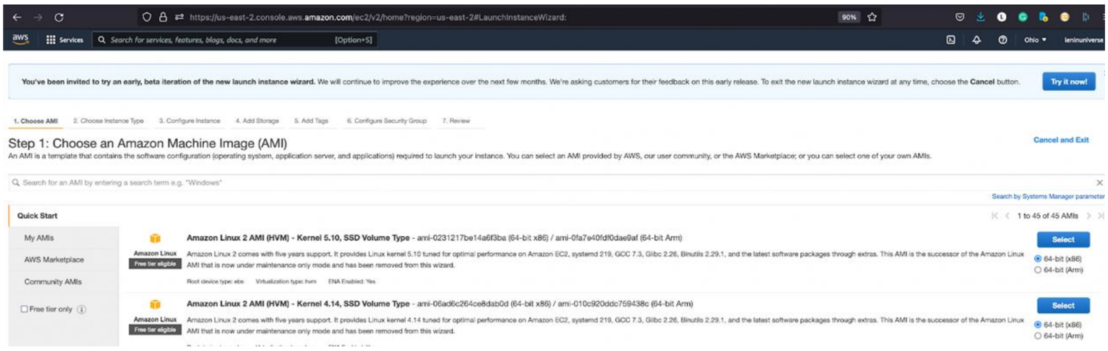
```
+-------------+----------+------------+
|        state|sum_deaths|   sum_cases|
+-------------+----------+------------+
|west virginia| 1286901.0|  7.631901E7|
|new hampshire|  620816.0| 4.3191729E7|
|      alabama| 5005646.0|2.68440532E8|
+-------------+----------+------------+
```



```
+------------------------------+------------------------------+---------------+---------------+-----------------+-----------------+
|state                         |state                         |avg_1          |avg_2          |sum_deaths       |sum_cases        |
+------------------------------+------------------------------+---------------+---------------+-----------------+-----------------+
|west virginia                 |west virginia                 |27675.0        |27675.0        |1286901.0        |7.631901E7       |
|new hampshire                 |new hampshire                 |20711.25       |20711.25       |620816.0         |4.3191729E7      |
|alabama                       |alabama                       |70745.625      |70745.625      |5005646.0        |2.68440532E8     |
```



```
+--------------------+----------+----------+---------+---------+--------------------+-----------+-------------+
|               state|     avg_1|     avg_2|    sum_1|    sum_2|               state| sum_deaths|    sum_cases|
+--------------------+----------+----------+---------+---------+--------------------+-----------+-------------+
|       west virginia|   27675.0|   27675.0| 442800.0| 442800.0|       west virginia|  1286901.0|   7.631901E7|
|       new hampshire|  20711.25|  20711.25| 331380.0| 331380.0|       new hampshire|   620816.0|  4.3191729E7|
|     mariana islands|     780.0|       0.0|  11700.0|      0.0|                null|       null|         null|
```



https://us-east-2.console.aws.amazon.com/ec2/v2/home?region=us-east-2#LaunchInstanceWizard:

90%
Services
Search for services, features, blogs, docs, and more [Option+S]
Ohio
leninuniverse

You've been invited to try an early, beta iteration of the new launch instance wizard. We will continue to improve the experience over the next few months. We're asking customers for their feedback on this early release. To exit the new launch instance wizard at any time, choose the **Cancel** button.

Step 1: Choose an Amazon Machine Image (AMI)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs.

Search for an AMI by entering a search term e.g. "Windows"
Search by Systems Manager parameter
Quick Start
My AMIs
AWS Marketplace
Community AMIs
□ Free tier only
1 to 45 of 45 AMIs

Amazon Linux 2 AMI (HVM) - Kernel 5.10, SSD Volume Type - ami-0231217be14a6f3ba (64-bit x86) / ami-0fa7e40fdf0dae9af (64-bit Arm)

Amazon Linux — Free tier eligible

Amazon Linux 2 comes with five years support. It provides Linux kernel 5.10 tuned for optimal performance on Amazon EC2, systemd 219, GCC 7.3, Glibc 2.26, Binutils 2.29.1, and the latest software packages through extras. This AMI is the successor of the Amazon Linux AMI that is now under maintenance only mode and has been removed from this wizard.

Root device type: ebs Virtualization type: hvm ENA Enabled: Yes

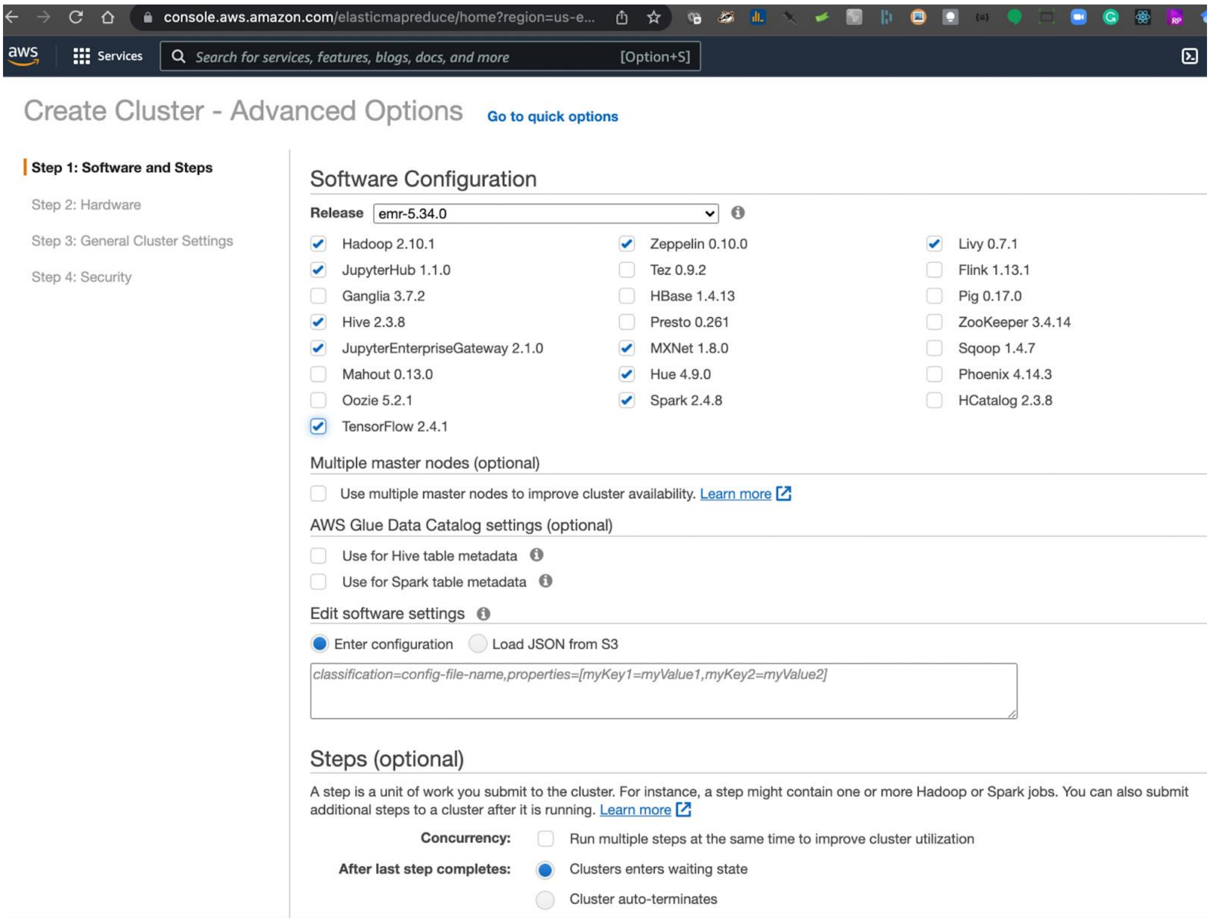
Select ● 64-bit (x86) ○ 64-bit (Arm)

Amazon Linux 2 AMI (HVM) - Kernel 4.14, SSD Volume Type - ami-06ad6c264ce8dab0d (64-bit x86) / ami-010c920ddc759438c (64-bit Arm)

Amazon Linux — Free tier eligible

Amazon Linux 2 comes with five years support. It provides Linux kernel 4.14 tuned for optimal performance on Amazon EC2, systemd 219, GCC 7.3, Glibc 2.26, Binutils 2.29.1, and the latest software packages through extras. This AMI is the successor of the Amazon Linux AMI that is now under maintenance only mode and has been removed from this wizard.

Select ● 64-bit (x86) ○ 64-bit (Arm)



console.aws.amazon.com/elasticmapreduce/home?region=us-e...

aws
Services
Search for services, features, blogs, docs, and more [Option+S]

Create Cluster - Advanced Options

Go to quick options
Step 1: Software and Steps
Step 2: Hardware
Step 3: General Cluster Settings
Step 4: Security

Software Configuration

Release: emr-5.34.0

☑ Hadoop 2.10.1
☑ JupyterHub 1.1.0
□ Ganglia 3.7.2
☑ Hive 2.3.8
☑ JupyterEnterpriseGateway 2.1.0
□ Mahout 0.13.0
□ Oozie 5.2.1
☑ TensorFlow 2.4.1
☑ Zeppelin 0.10.0
□ Tez 0.9.2
□ HBase 1.4.13
□ Presto 0.261
☑ MXNet 1.8.0
☑ Hue 4.9.0
☑ Spark 2.4.8
☑ Livy 0.7.1
□ Flink 1.13.1
□ Pig 0.17.0
□ ZooKeeper 3.4.14
□ Sqoop 1.4.7
□ Phoenix 4.14.3
□ HCatalog 2.3.8

Multiple master nodes (optional)

□ Use multiple master nodes to improve cluster availability. Learn more

AWS Glue Data Catalog settings (optional)

□ Use for Hive table metadata
□ Use for Spark table metadata

Edit software settings

● Enter configuration ○ Load JSON from S3

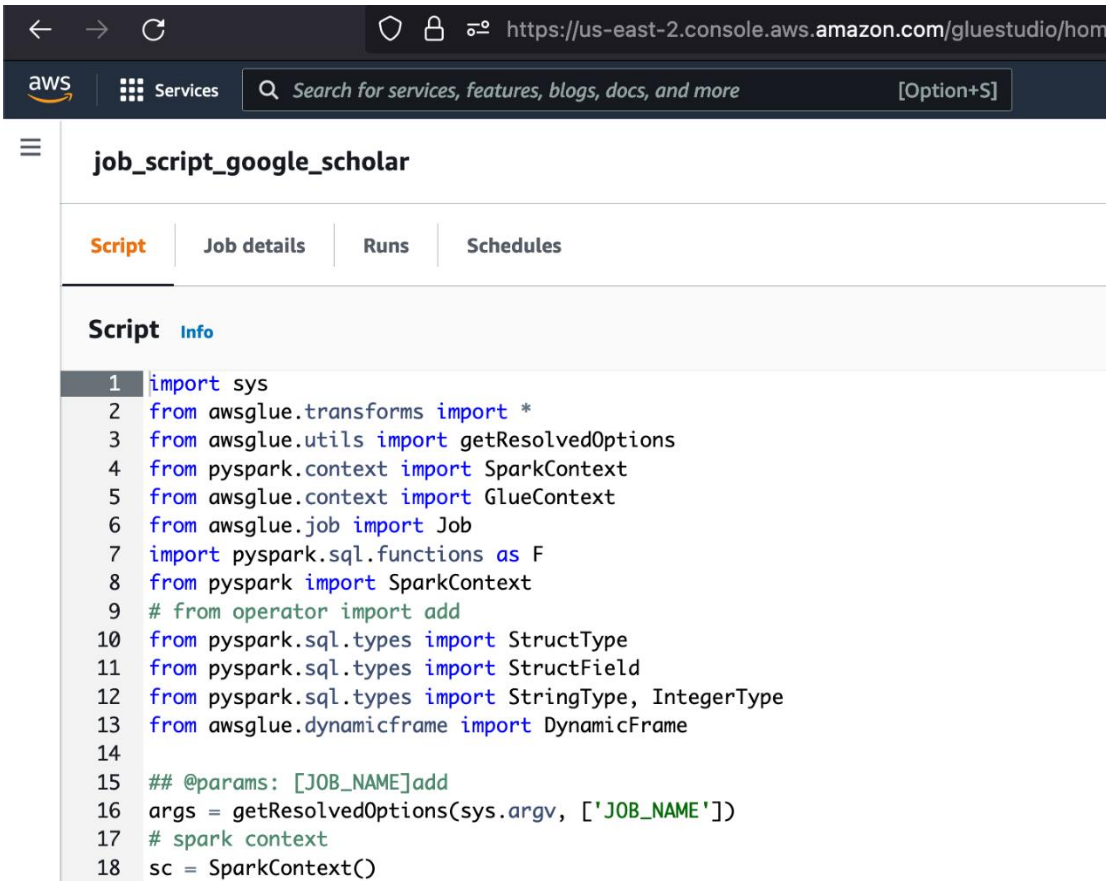
classification=config-file-name,properties=[myKey1=myValue1,myKey2=myValue2]

Steps (optional)

A step is a unit of work you submit to the cluster. For instance, a step might contain one or more Hadoop or Spark jobs. You can also submit additional steps to a cluster after it is running. Learn more

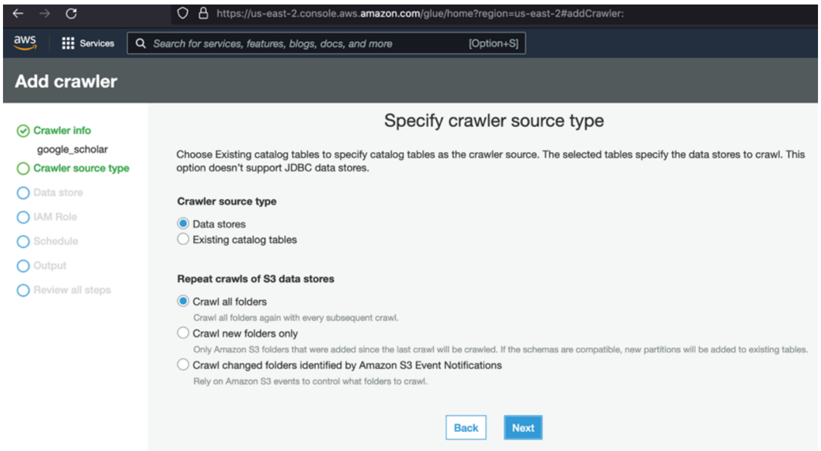
Concurrency: □ Run multiple steps at the same time to improve cluster utilization

After last step completes: ● Clusters enters waiting state
○ Cluster auto-terminates


https://us-east-2.console.aws.amazon.com/gluestudio/hom
aws
Services
Search for services, features, blogs, docs, and more
[Option+S]
job_script_google_scholar
Script
Job details
Runs
Schedules
Script
Info

```python
import sys
from awsglue.transforms import *
from awsglue.utils import getResolvedOptions
from pyspark.context import SparkContext
from awsglue.context import GlueContext
from awsglue.job import Job
import pyspark.sql.functions as F
from pyspark import SparkContext
# from operator import add
from pyspark.sql.types import StructType
from pyspark.sql.types import StructField
from pyspark.sql.types import StringType, IntegerType
from awsglue.dynamicframe import DynamicFrame

## @params: [JOB_NAME]add
args = getResolvedOptions(sys.argv, ['JOB_NAME'])
# spark context
sc = SparkContext()
```


https://us-east-2.console.aws.amazon.com/glue/home?region=us-east-2#addCrawler:
aws
Services
Search for services, features, blogs, docs, and more
[Option+S]
Add crawler
Crawler info
google_scholar
Crawler source type
Data store
IAM Role
Schedule
Output
Review all steps

Specify crawler source type

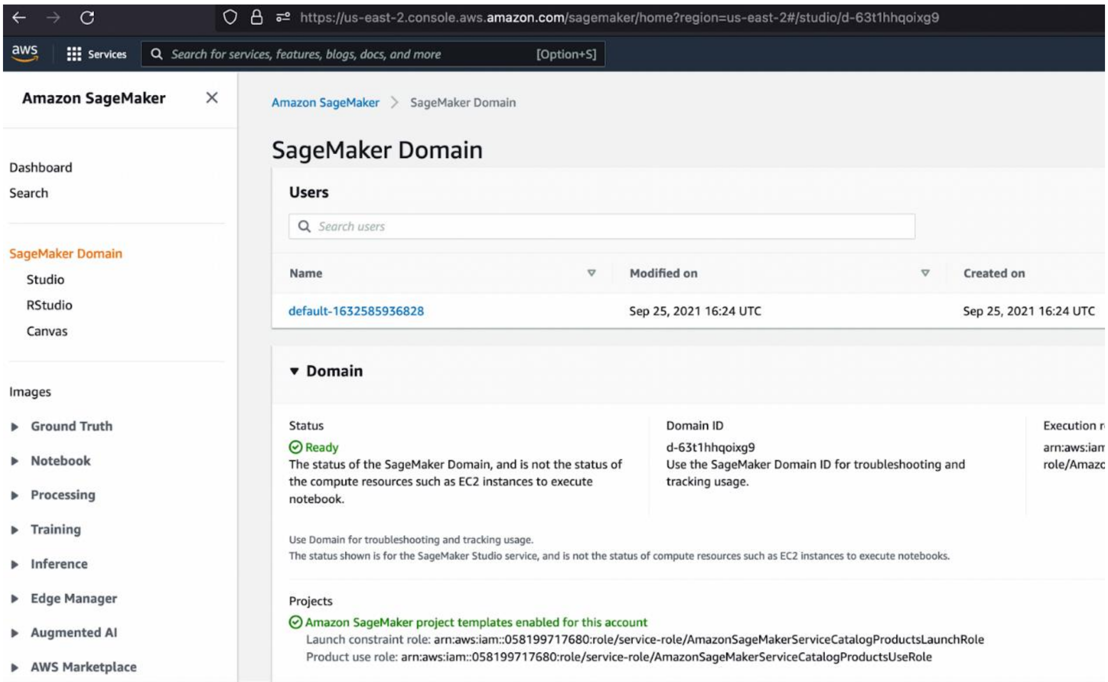
Choose Existing catalog tables to specify catalog tables as the crawler source. The selected tables specify the data stores to crawl. This option doesn't support JDBC data stores.

Crawler source type

● Data stores
○ Existing catalog tables

Repeat crawls of S3 data stores

● Crawl all folders
Crawl all folders again with every subsequent crawl.
○ Crawl new folders only
Only Amazon S3 folders that were added since the last crawl will be crawled. If the schemas are compatible, new partitions will be added to existing tables.
○ Crawl changed folders identified by Amazon S3 Event Notifications
Rely on Amazon S3 events to control what folders to crawl.


https://us-east-2.console.aws.amazon.com/sagemaker/home?region=us-east-2#/studio/d-63t1hhqoixg9
aws
Services
Search for services, features, blogs, docs, and more
[Option+S]
Amazon SageMaker
Dashboard
Search
SageMaker Domain
Studio
RStudio
Canvas
Images
Ground Truth
Notebook
Processing
Training
Inference
Edge Manager
Augmented AI
AWS Marketplace
Amazon SageMaker > SageMaker Domain

SageMaker Domain

Users

Search users

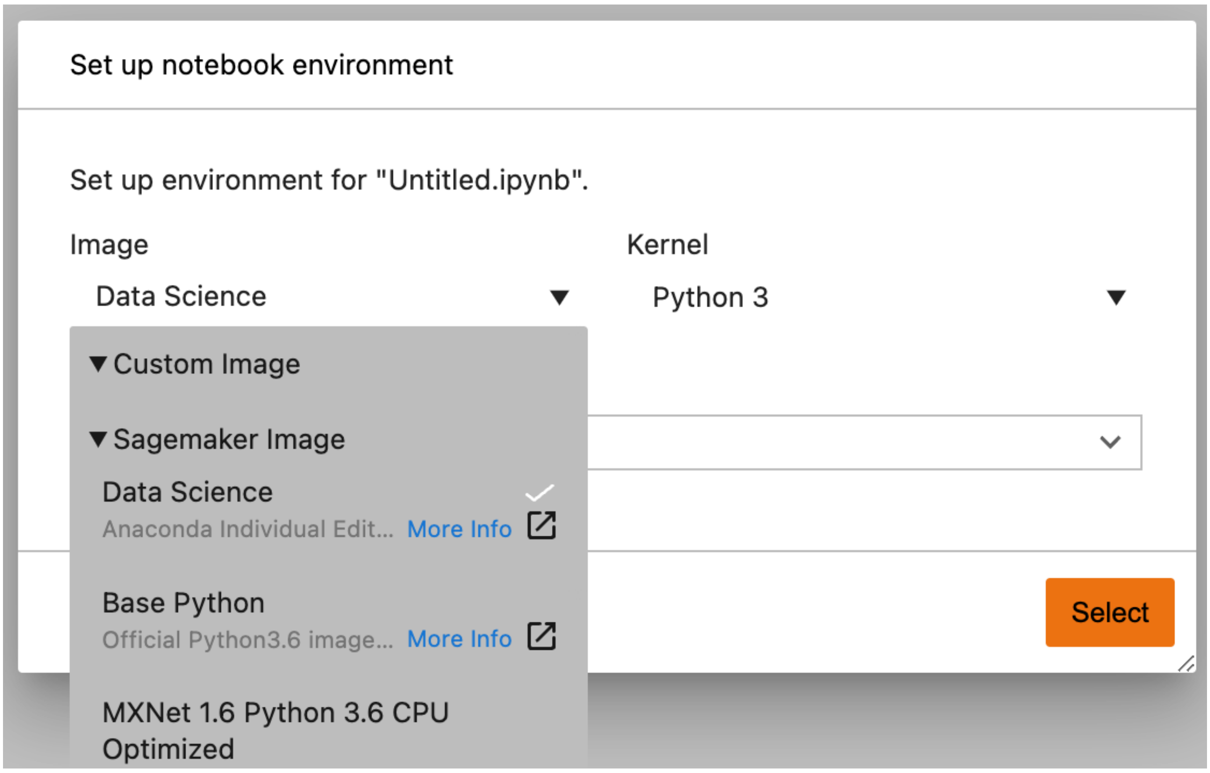
Name	Modified on	Created on
default-1632585936828	Sep 25, 2021 16:24 UTC	Sep 25, 2021 16:24 UTC

▼ Domain

Status	Domain ID	Execution r
✓ Ready The status of the SageMaker Domain, and is not the status of the compute resources such as EC2 instances to execute notebook.	d-63t1hhqoixg9 Use the SageMaker Domain ID for troubleshooting and tracking usage.	arn:aws:iar role/Amazc

Use Domain 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::058199717680:role/service-role/AmazonSageMakerServiceCatalogProductsLaunchRole
Product use role: arn:aws:iam::058199717680:role/service-role/AmazonSageMakerServiceCatalogProductsUseRole



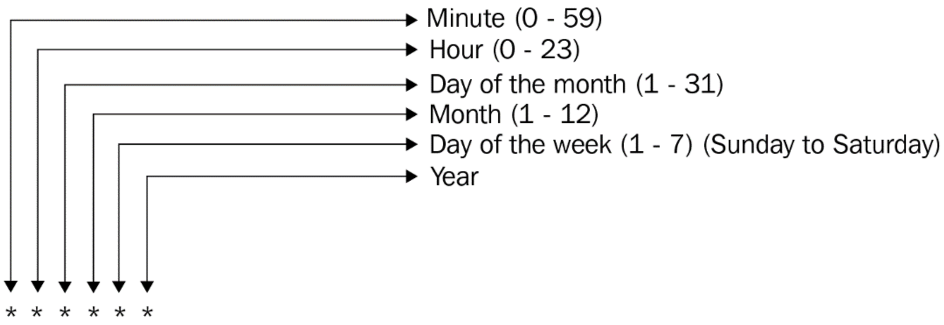
Set up notebook environment

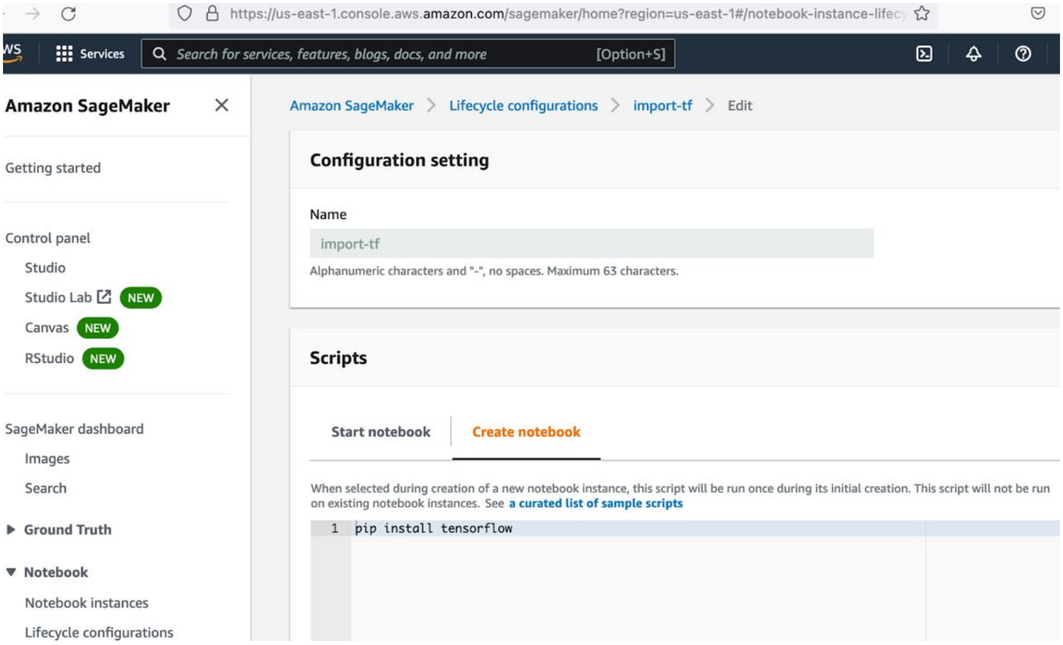
Set up environment for "Untitled.ipynb".

Image: Data Science

Kernel: Python 3

▼ Custom Image
▼ Sagemaker Image
Data Science ✓
Anaconda Individual Edit... More Info
Base Python
Official Python3.6 image... More Info
MXNet 1.6 Python 3.6 CPU Optimized
Select


Minute (0 - 59)
Hour (0 - 23)
Day of the month (1 - 31)
Month (1 - 12)
Day of the week (1 - 7) (Sunday to Saturday)
Year
* * * * * *


https://us-east-1.console.aws.amazon.com/sagemaker/home?region=us-east-1#/notebook-instance-lifec
Services
Search for services, features, blogs, docs, and more
[Option+S]
Amazon SageMaker
Getting started
Control panel
Studio
Studio Lab
NEW
Canvas
NEW
RStudio
NEW
SageMaker dashboard
Images
Search
Ground Truth
Notebook
Notebook instances
Lifecycle configurations
Amazon SageMaker > Lifecycle configurations > import-tf > Edit

Configuration setting

Name: import-tf

Alphanumeric characters and "-", no spaces. Maximum 63 characters.

Scripts

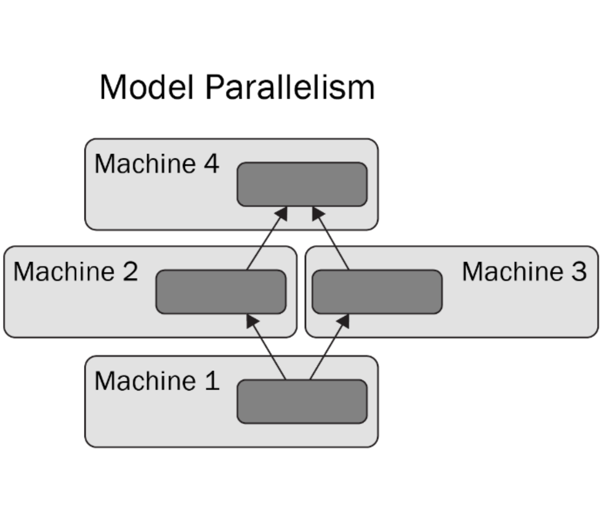
Start notebook | Create notebook

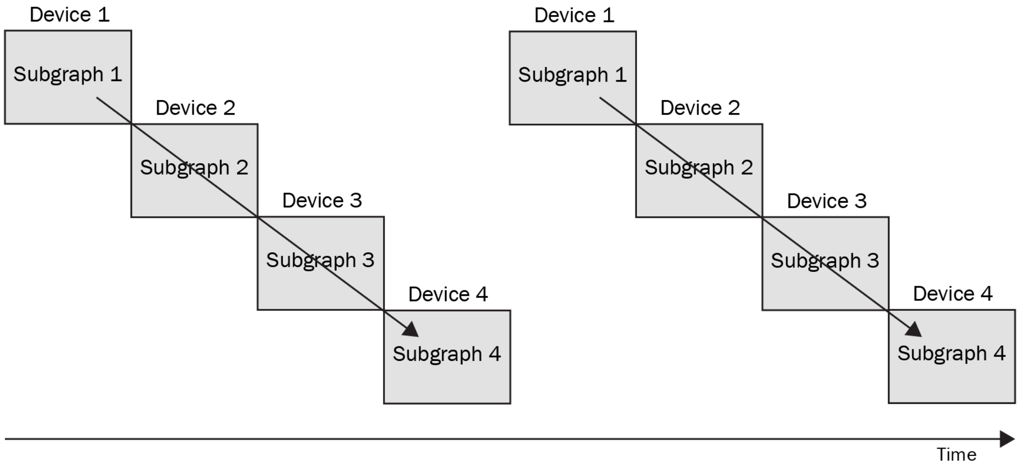
When selected during creation of a new notebook instance, this script will be run once during its initial creation. This script will not be run on existing notebook instances. See a curated list of sample scripts

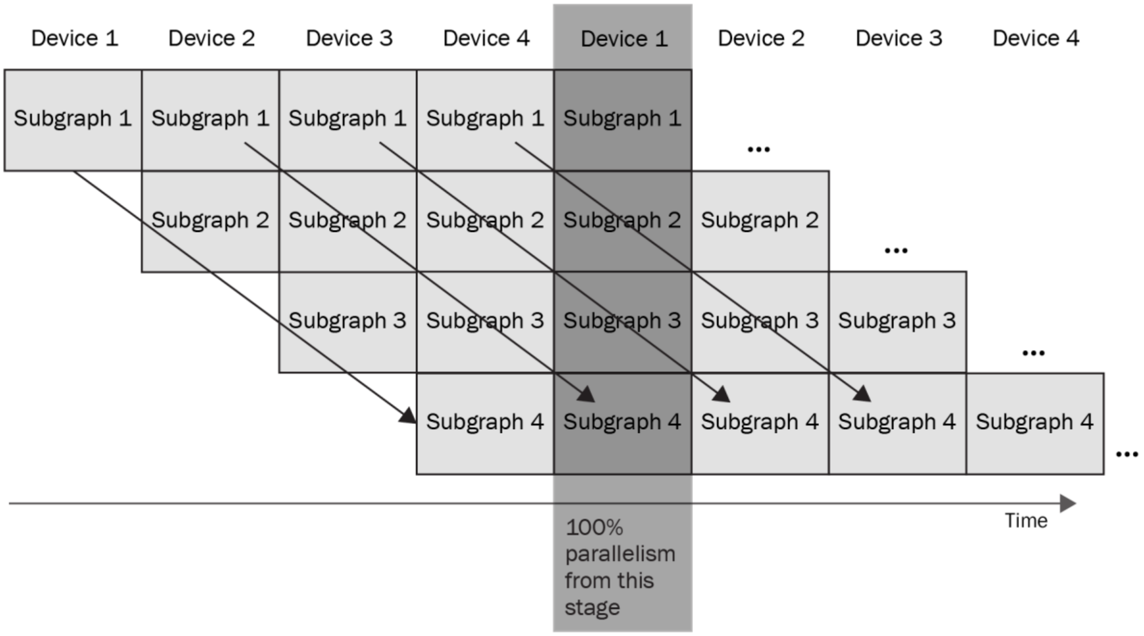
```
pip install tensorflow
```

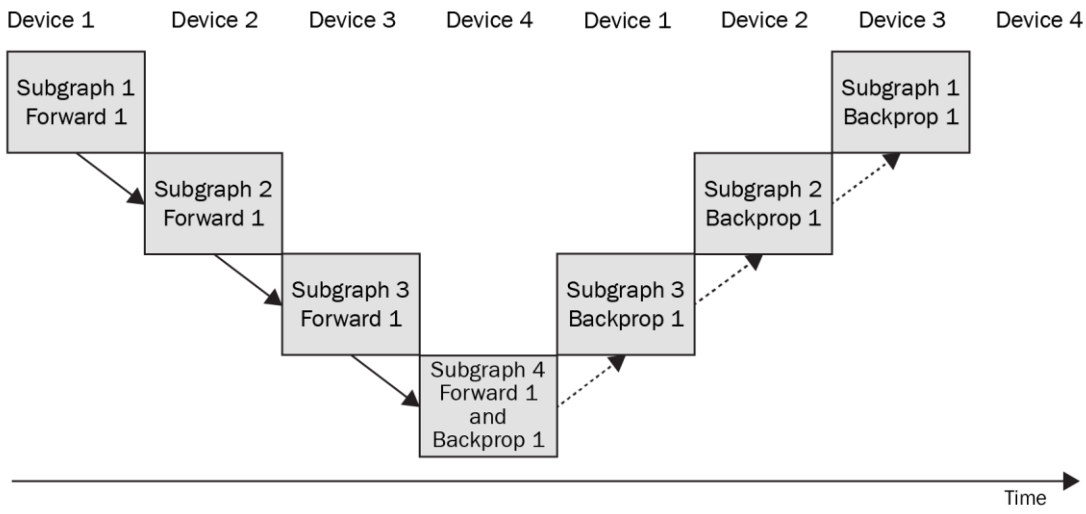
Supports	Single-Node EC2 Instance	Glue	EMR	SageMaker
Support for serverless architecture	No	Yes	No	No
Availability of a built-in workspace for collaboration among developers	No	No	Yes	No
Variety of EC2 instance types	More	Less	More	More
Availability of a built-in scheduler	No	Yes	No	Yes
Availability of a built-in job monitoring UI	No	Yes	No	Yes
Availability of a built-in model monitoring	No	No	No	Yes
Support for a fully managed service from model development to deployment	No	No	No	Yes
Availability of a built-in visualizer for analyzing the processed data	No	No	No	Yes
Availability of a predefined environment for ETL logic development	Yes	No	Yes	Yes

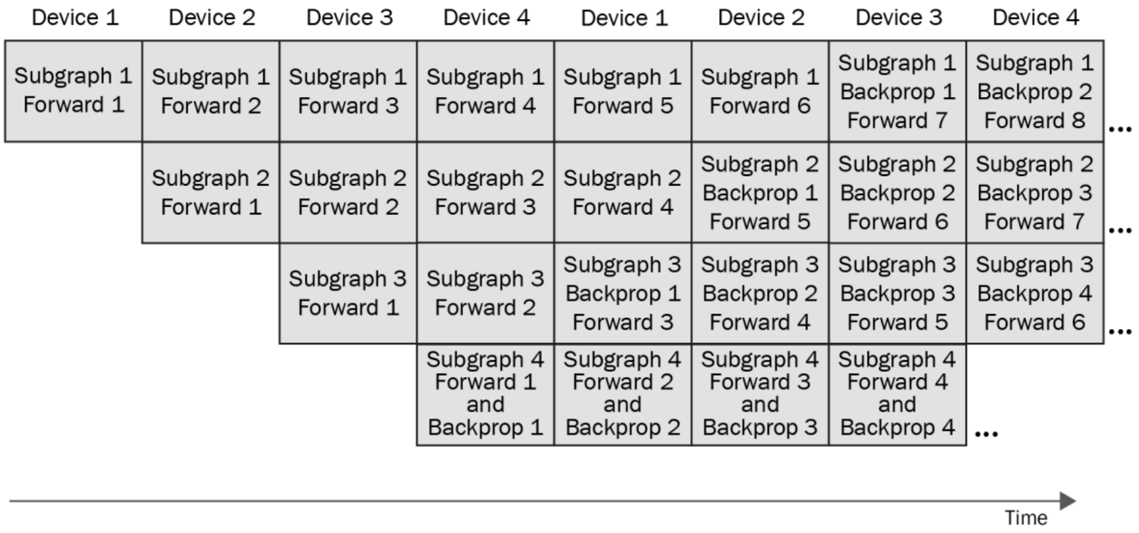
Chapter 6: Efficient Model Training



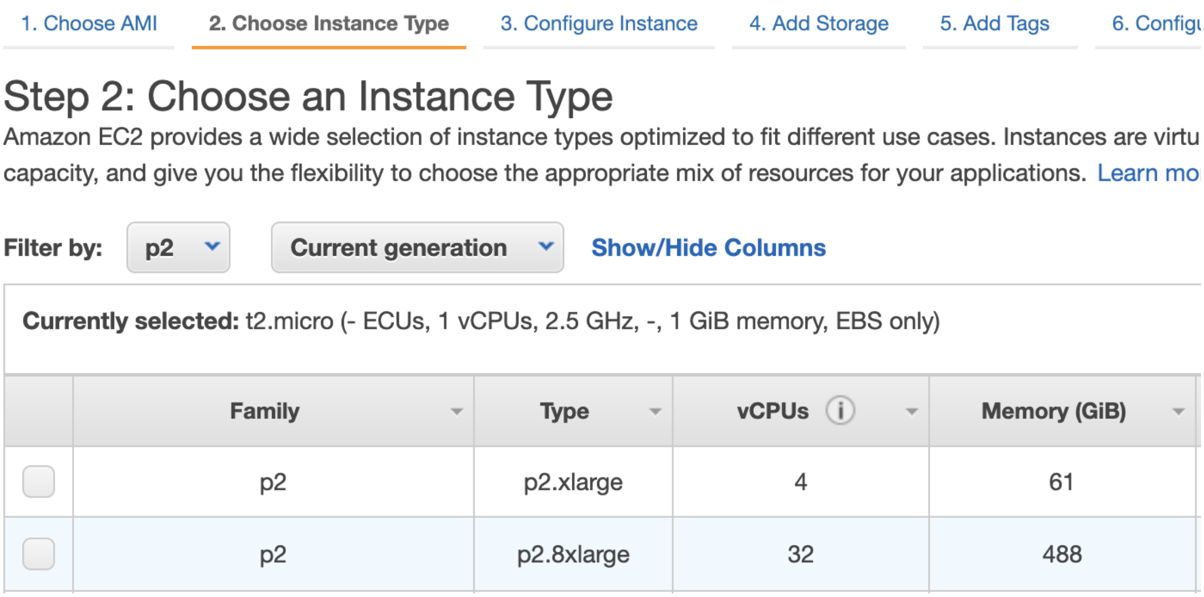








Device 1	Device 2	Device 3	Device 4	Device 1	Device 2	Device 3	Device 4	
Subgraph 1 Forward 1	Subgraph 1 Forward 2	Subgraph 1 Forward 3	Subgraph 1 Forward 4	Subgraph 1 Forward 5	Subgraph 1 Forward 6	Subgraph 1 Backprop 1 Forward 7	Subgraph 1 Backprop 2 Forward 8	...
	Subgraph 2 Forward 1	Subgraph 2 Forward 2	Subgraph 2 Forward 3	Subgraph 2 Forward 4	Subgraph 2 Backprop 1 Forward 5	Subgraph 2 Backprop 2 Forward 6	Subgraph 2 Backprop 3 Forward 7	...
		Subgraph 3 Forward 1	Subgraph 3 Forward 2	Subgraph 3 Backprop 1 Forward 3	Subgraph 3 Backprop 2 Forward 4	Subgraph 3 Backprop 3 Forward 5	Subgraph 3 Backprop 4 Forward 6	...
			Subgraph 4 Forward 1 and Backprop 1	Subgraph 4 Forward 2 and Backprop 2	Subgraph 4 Forward 3 and Backprop 3	Subgraph 4 Forward 4 and Backprop 4	...	



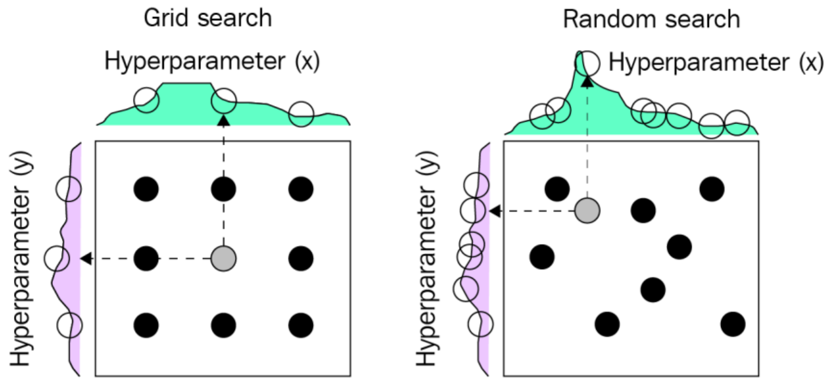
Step 2: Choose an Instance Type

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtu capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. Learn mo

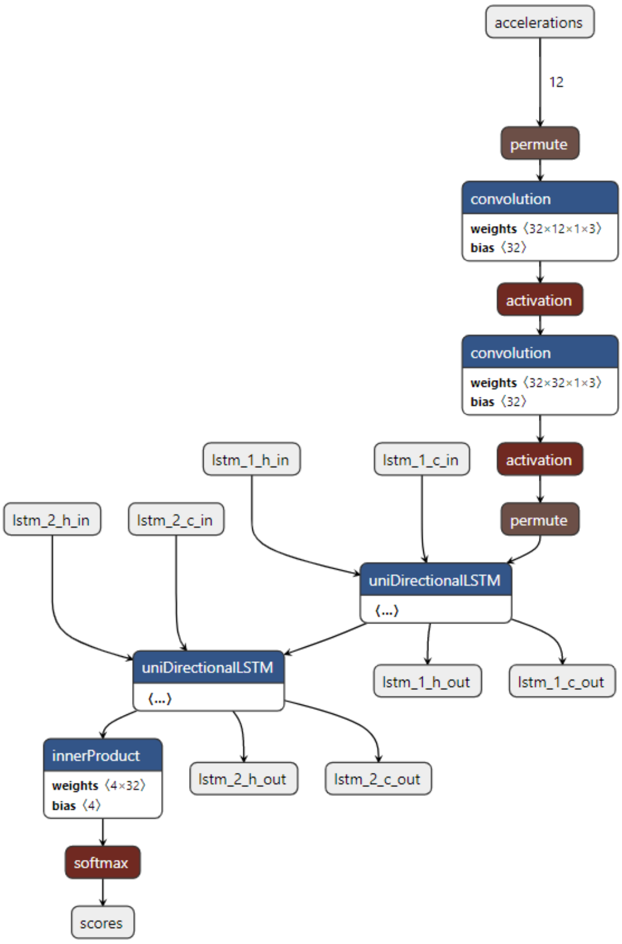
Filter by: p2 | Current generation | **Show/Hide Columns**

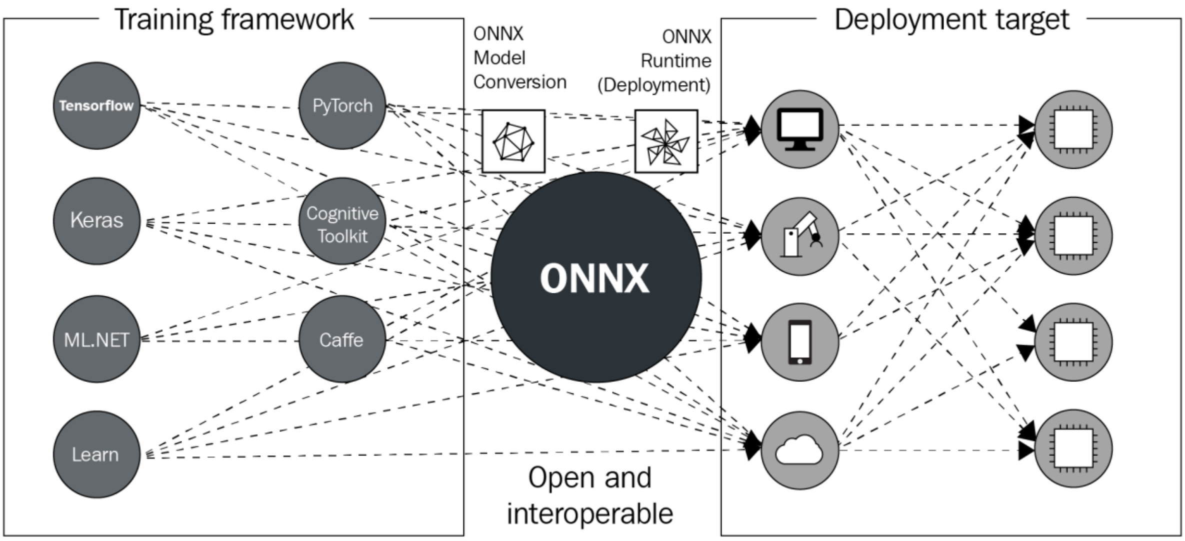
Currently selected: t2.micro (- ECUs, 1 vCPUs, 2.5 GHz, -, 1 GiB memory, EBS only)

	Family	Type	vCPUs	Memory (GiB)
□	p2	p2.xlarge	4	61
□	p2	p2.8xlarge	32	488

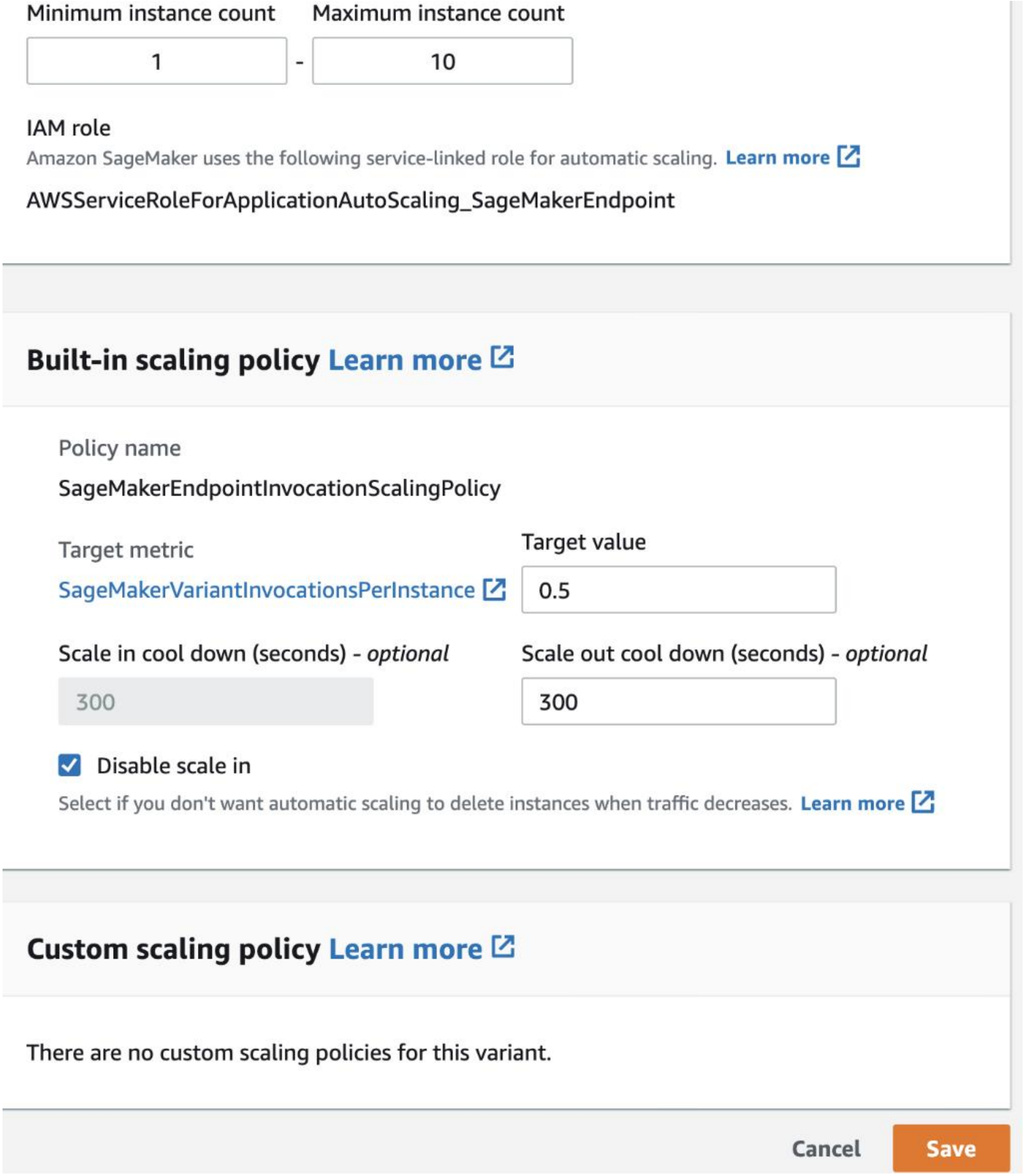

Grid search
Hyperparameter (x)
Hyperparameter (y)
Random search
Hyperparameter (x)
Hyperparameter (y)

Chapter 8: Simplifying Deep Learning Model Deployment





Chapter 9: Scaling a Deep Learning Pipeline



Minimum instance count: 1 - Maximum instance count: 10

IAM role

Amazon SageMaker uses the following service-linked role for automatic scaling. Learn more

AWSServiceRoleForApplicationAutoScaling_SageMakerEndpoint

Built-in scaling policy Learn more

Policy name

SageMakerEndpointInvocationScalingPolicy

Target metric

SageMakerVariantInvocationsPerInstance

Target value: 0.5

Scale in cool down (seconds) - *optional*: 300

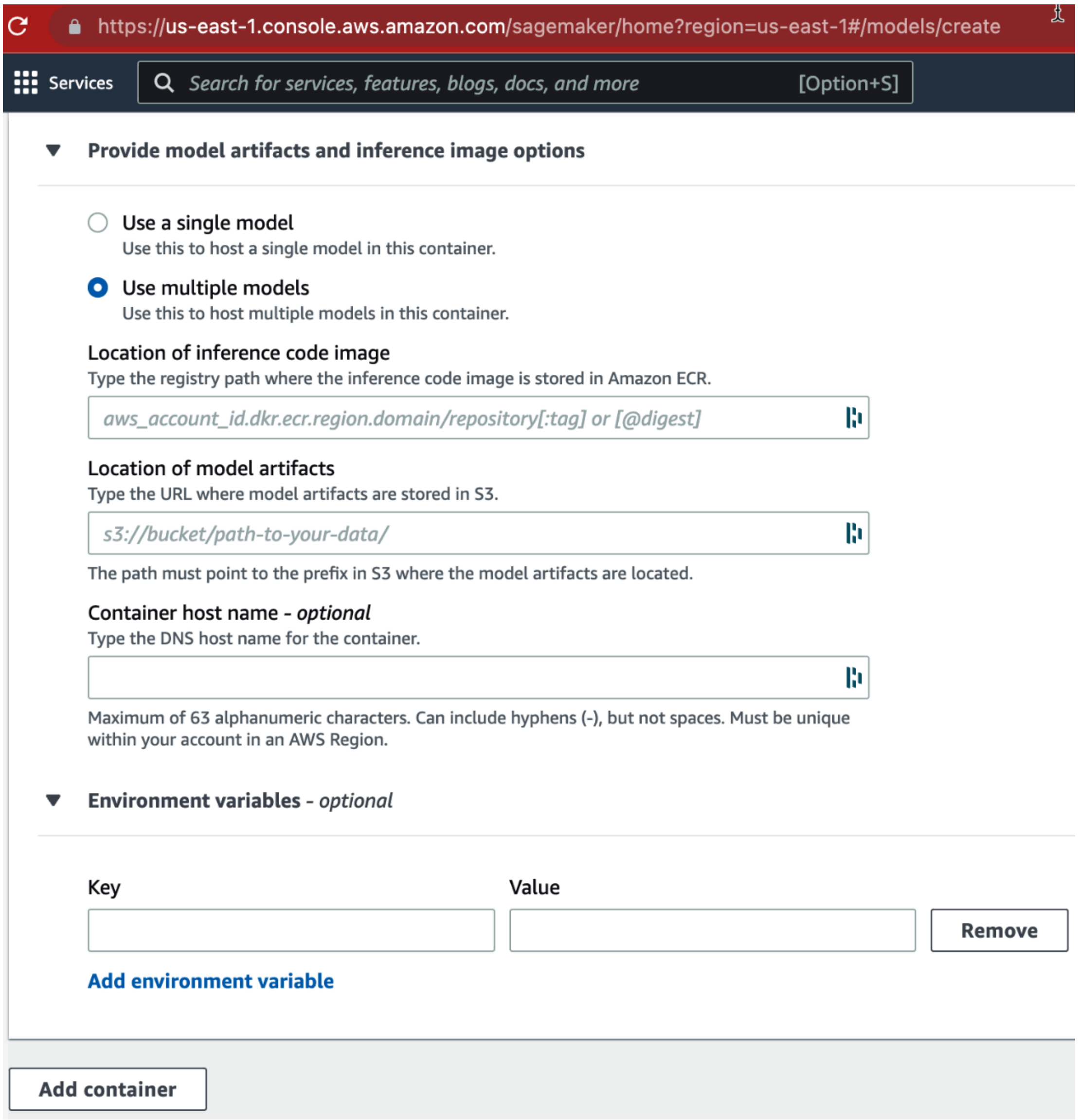
Scale out cool down (seconds) - *optional*: 300

☑ Disable scale in

Select if you don't want automatic scaling to delete instances when traffic decreases. Learn more

Custom scaling policy Learn more

There are no custom scaling policies for this variant.



https://us-east-1.console.aws.amazon.com/sagemaker/home?region=us-east-1#/models/create

Search for services, features, blogs, docs, and more [Option+S]

▼ **Provide model artifacts and inference image options**

○ Use a single model
Use this to host a single model in this container.

● Use multiple models
Use this to host multiple models in this container.

Location of inference code image
Type the registry path where the inference code image is stored in Amazon ECR.
aws_account_id.dkr.ecr.region.domain/repository[:tag] or [@digest]

Location of model artifacts
Type the URL where model artifacts are stored in S3.
s3://bucket/path-to-your-data/
The path must point to the prefix in S3 where the model artifacts are located.

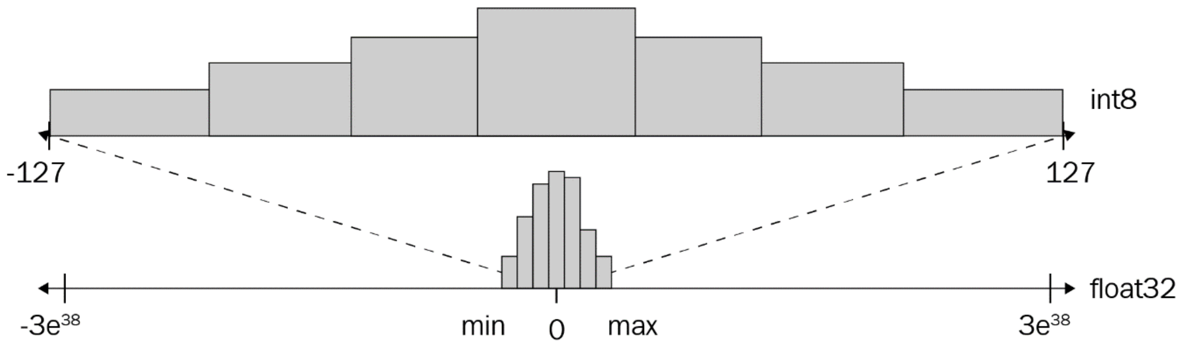
Container host name - *optional*
Type the DNS host name for the container.

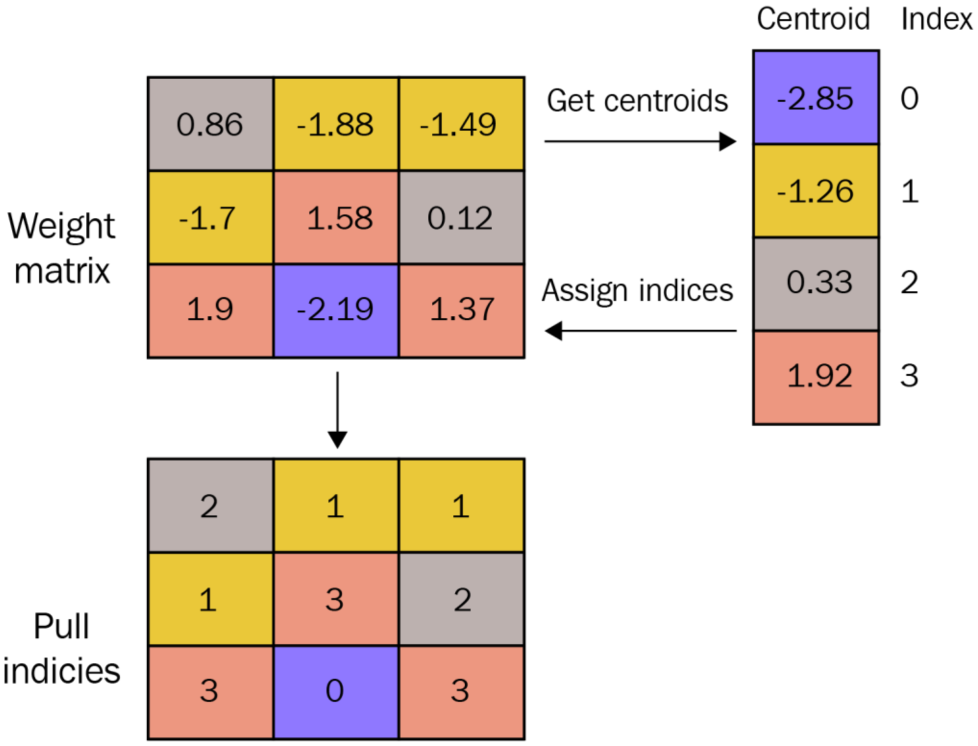
Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.

▼ **Environment variables** - *optional*

Key	Value	
		Remove

Chapter 10: Improving Inference Efficiency

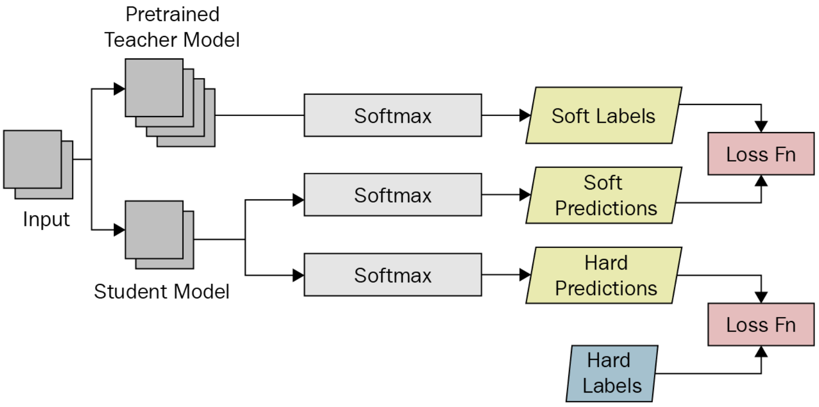




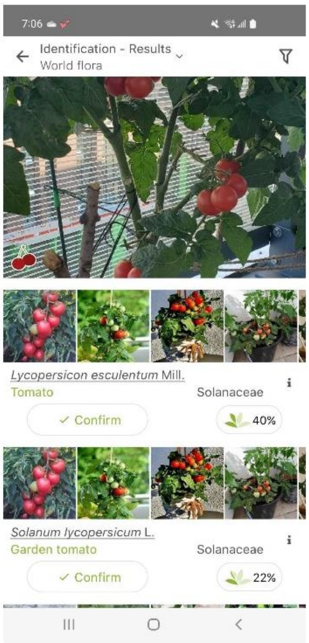
0.86	-1.88	-1.49
-1.7	1.58	0.12
1.9	-2.19	1.37

Centroid	Index
-2.85	0
-1.26	1
0.33	2
1.92	3

2	1	1
1	3	2
3	0	3

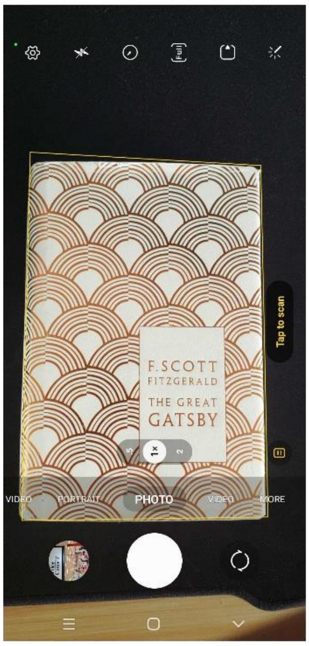


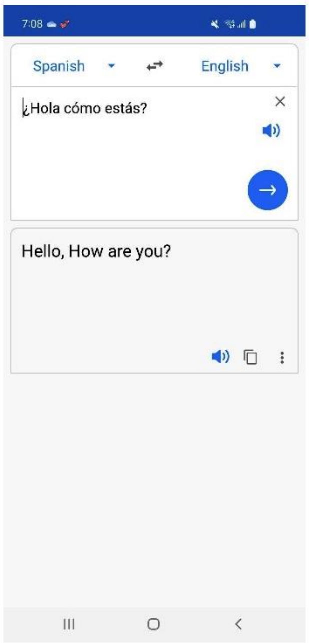
Chapter 11: Deep Learning on Mobile Device



Lycopersicon esculentum Mill.
Tomato
Solanaceae
✓ Confirm
40%

Solanum lycopersicum L.
Garden tomato
Solanaceae
✓ Confirm
22%





¿Hola cómo estás?

Hello, How are you?