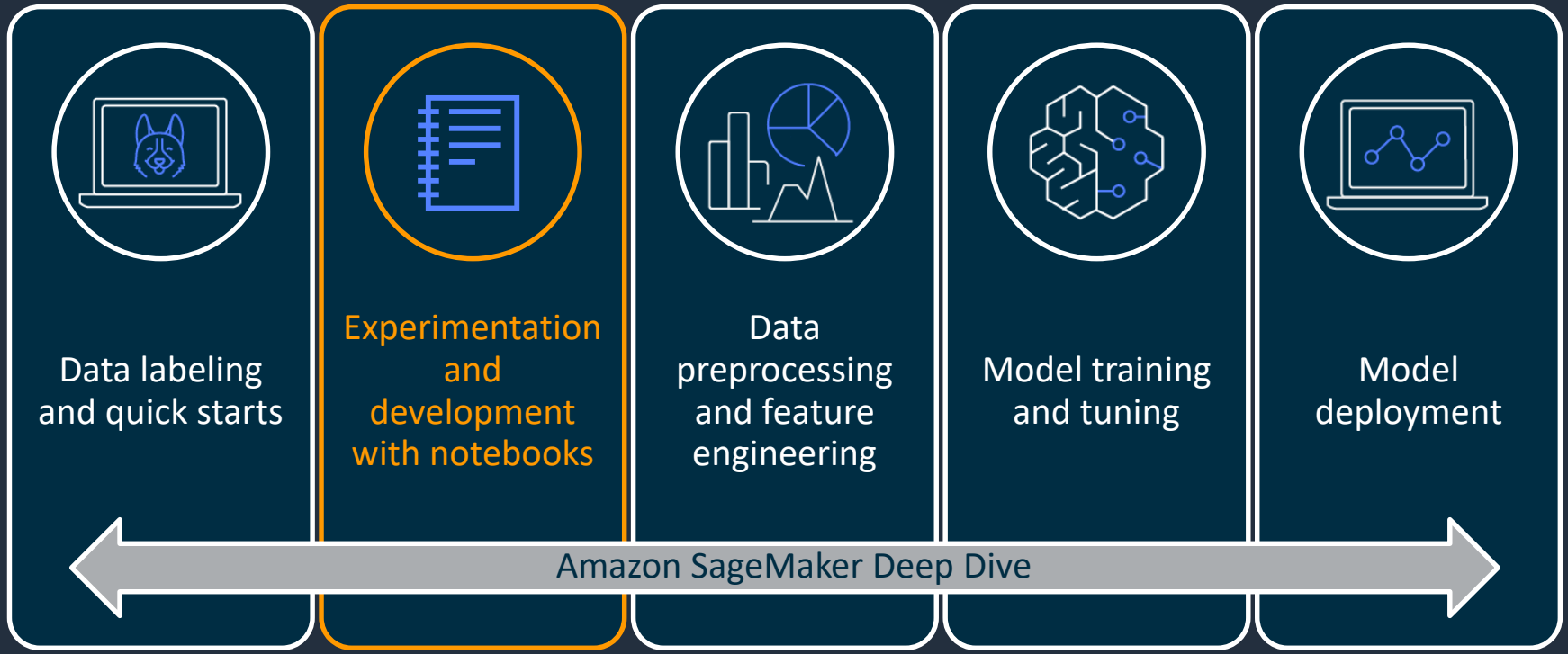
aws partnercast

Experimentation and development with notebooks

Amazon SageMaker Deep Dive Series

Amazon SageMaker Deep Dive Series



Amazon SageMaker key benefits

Most complete, end-to-end ML service

Accelerate ML development

20+ tools covering the entire ML development lifecycle

Boost data scientist productivity

The world's first integrated development environment (IDE)

Reduce cost

Eliminate costs of writing custom integration code with integrated functionality optimized for ML

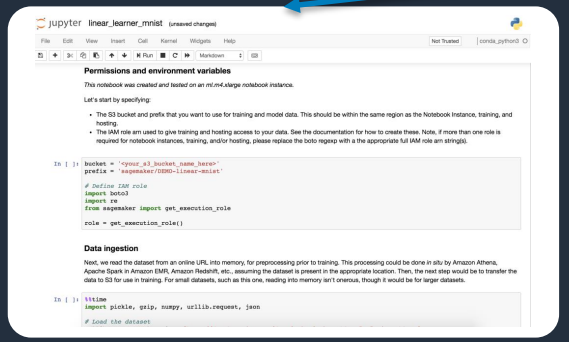
Amazon SageMaker overview

Features covered in this session

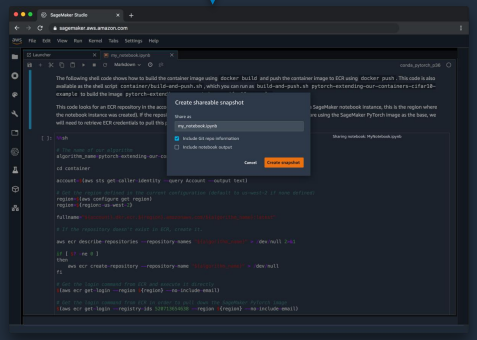
Development with SageMaker IDE's

Familiar integrated development environments with managed infrastructure

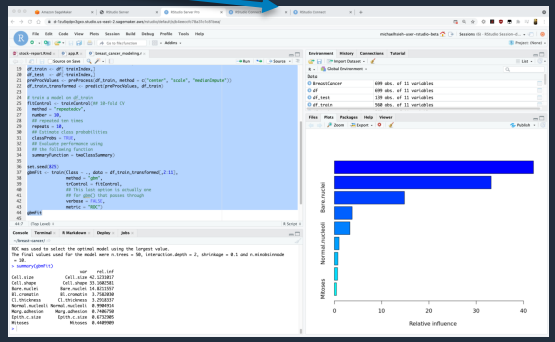
Amazon SageMaker



Classic Jupyter notebooks



Studio Jupyter notebooks



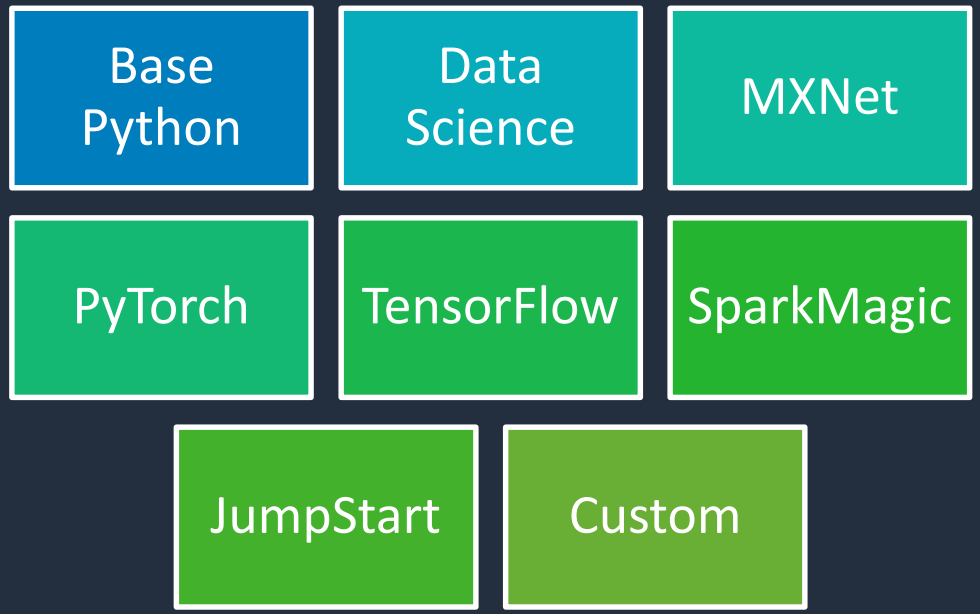
Rstudio notebooks

aws training and certification

ML instance types

Kernels

Managed environments with pre-installed packages for data science.

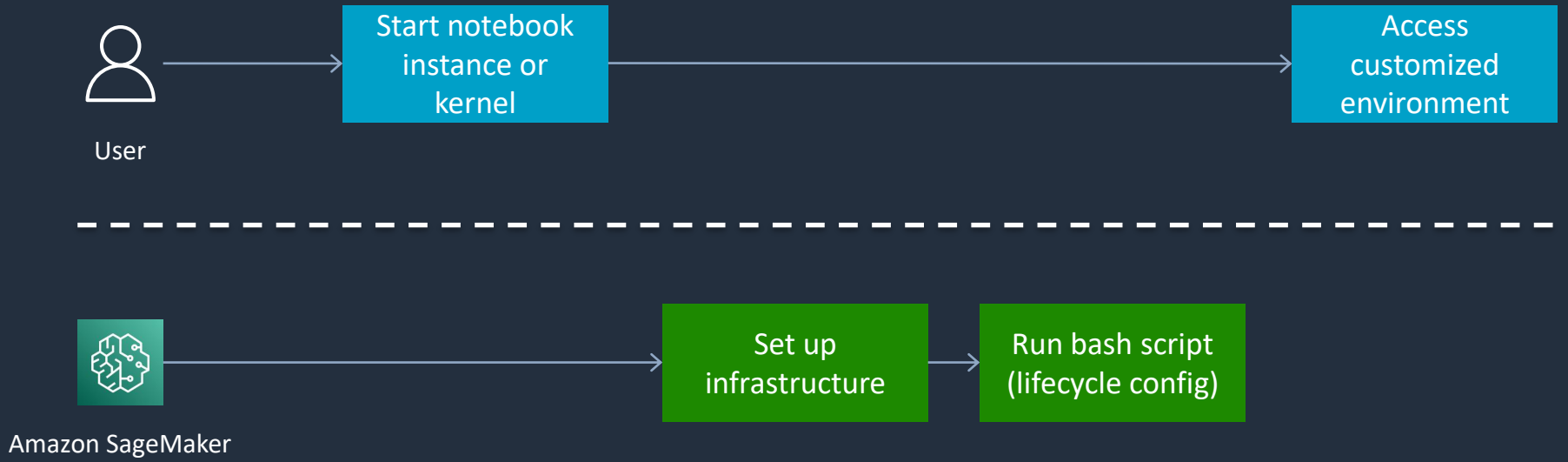


Packages in Data Science kernel:

- bokeh
- boto3
- matplotlib
- nltk
- numpy
- pandas
- plotly
- scipy
- seaborn
- sklearn
- ...

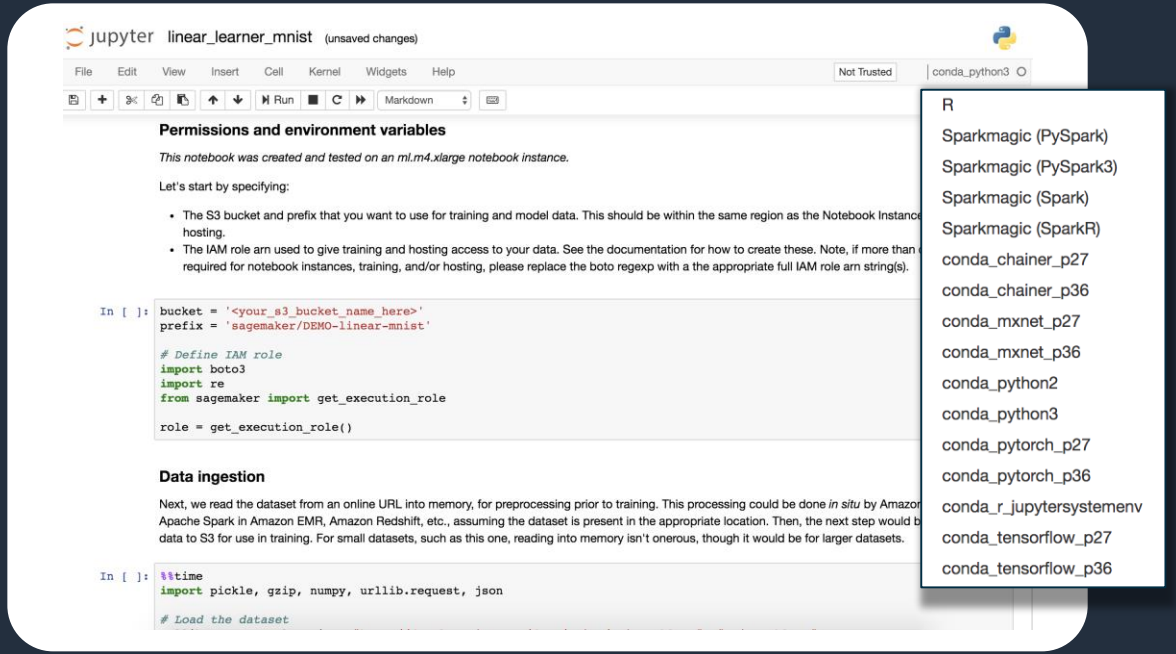
Lifecycle configurations

Automate customization of notebook environments with bash scripts. Use to install libraries, extensions, mount file systems, connect to other services (EMR, code repository), and more...



Classic Jupyter notebooks

- Managed Jupyter notebooks
- Supports Jupyter Lab
- Multiple built-in kernels
- Install external dependencies
- Integrate with Git
- Sample notebooks
- VPC connectivity
- Automation with Lifecyle Configuration



Under the hood

CONDA®

15+ pre-built kernels

Demo

Classic Jupyter notebooks

SageMaker Studio notebooks

Easy access with Single Sign-On

Access your notebooks in seconds

Fully managed and secure

Administrators manage access and permissions

Fast setup

Start your notebooks without spinning up compute resources

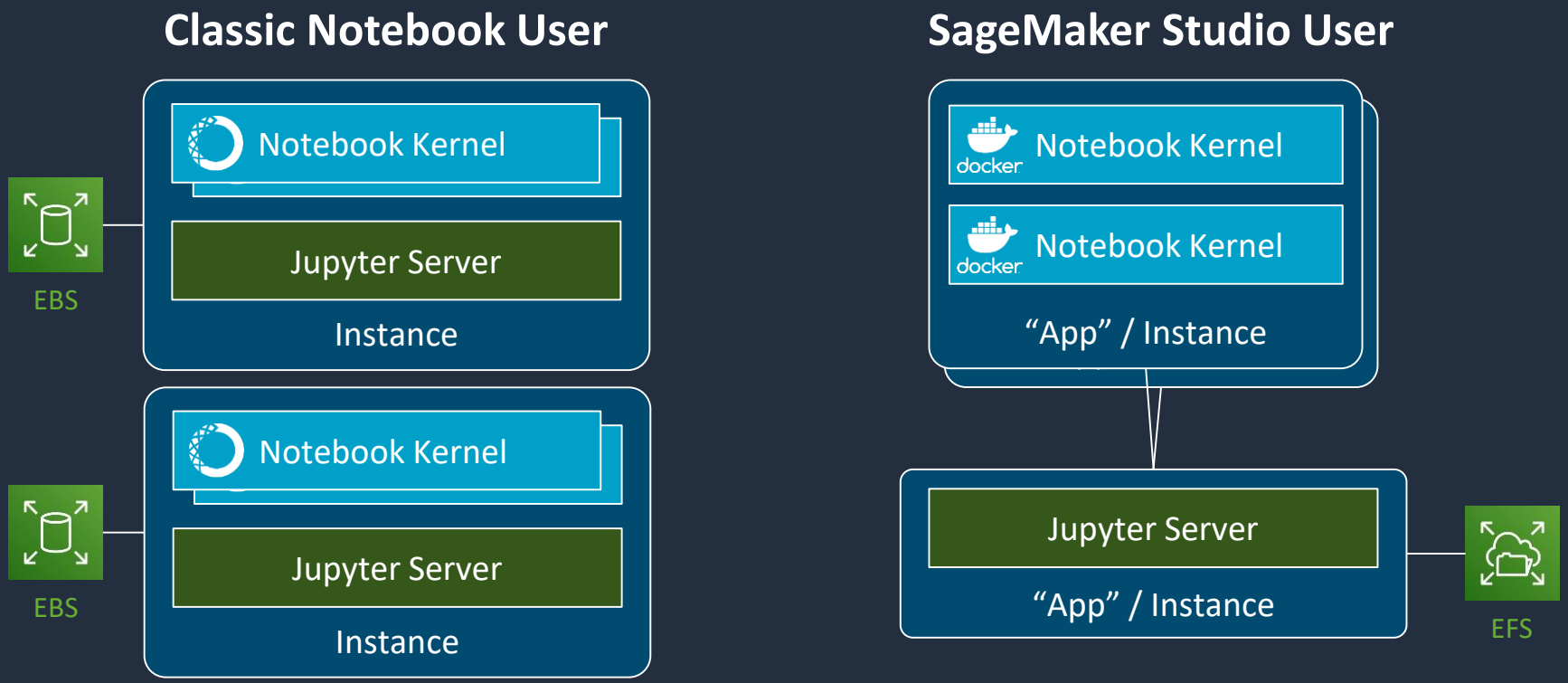
Easy collaboration

Share notebooks with a single click

Flexible

Dial up or down compute resources (coming soon)

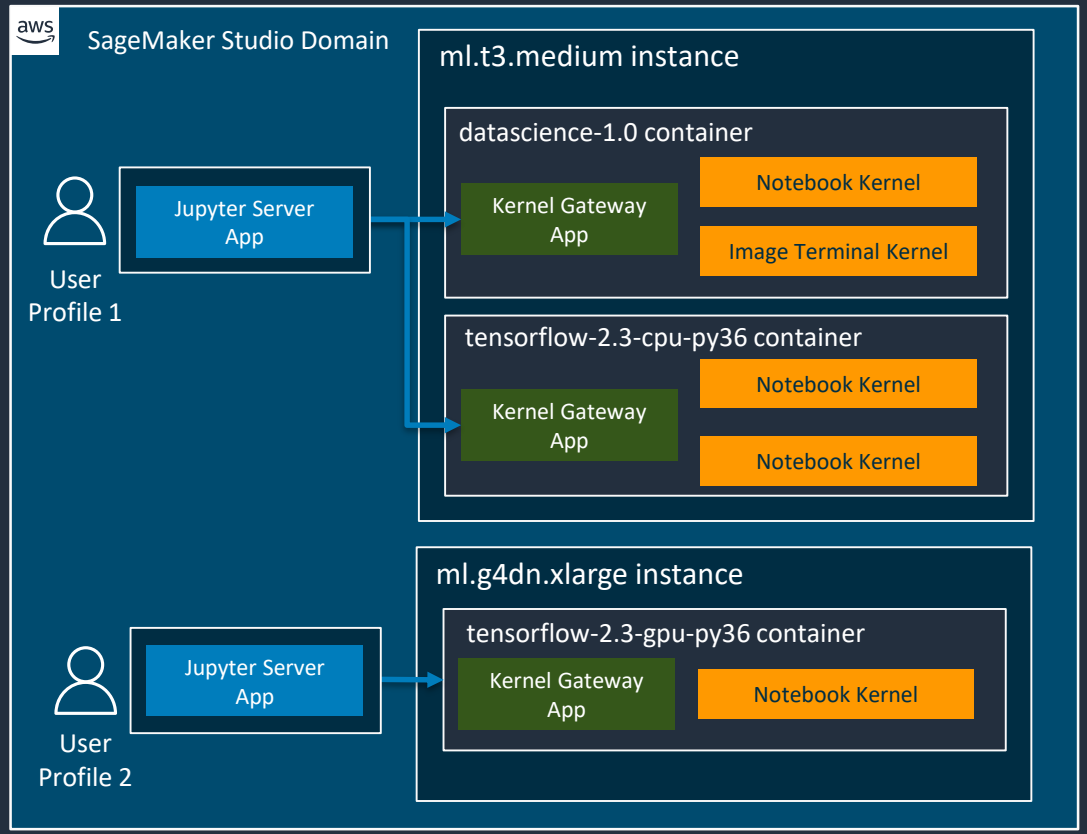
Differences under the hood



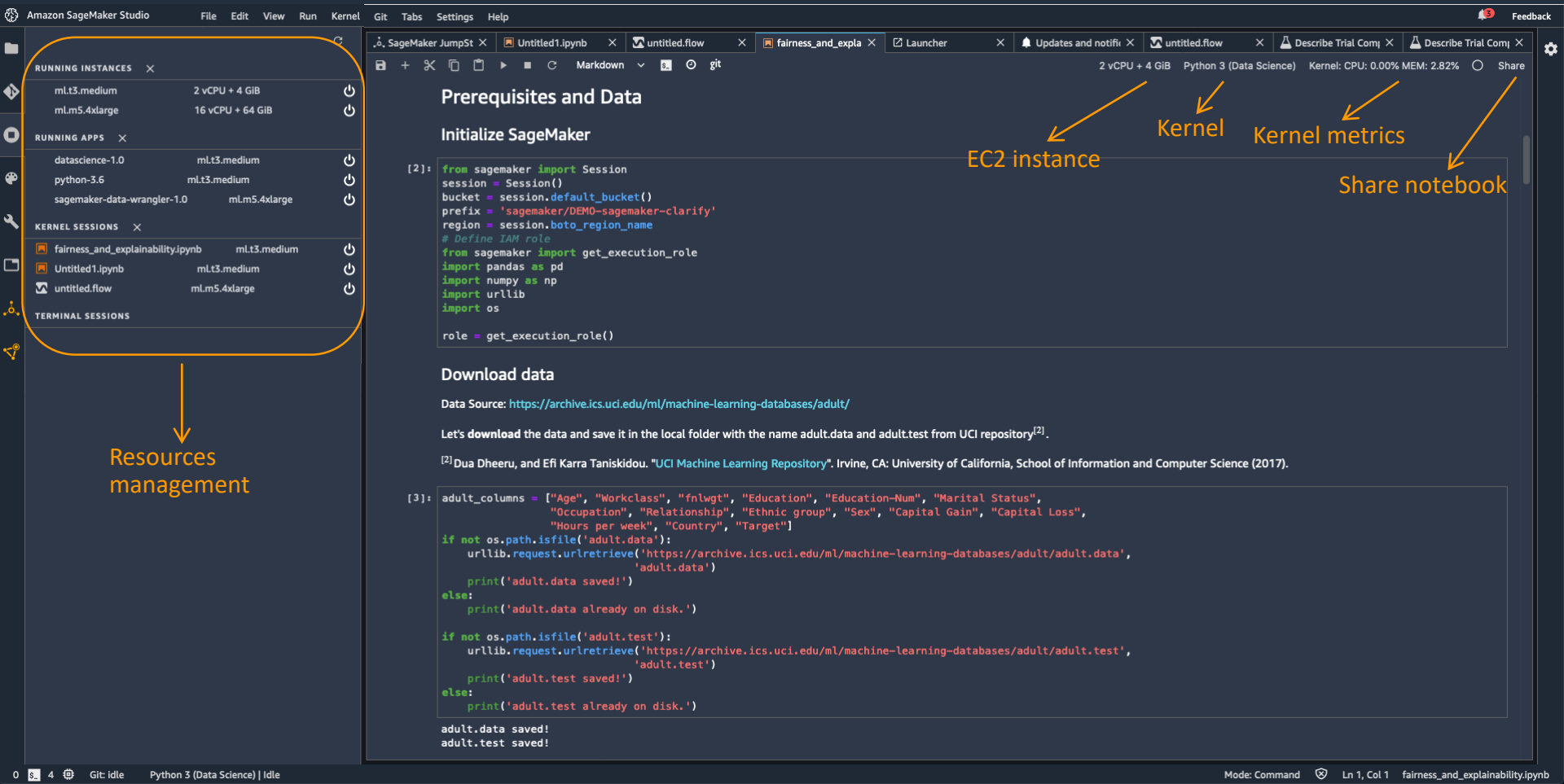
Right sizing notebook compute

SageMaker Studio is based on a Kernel Gateway architecture

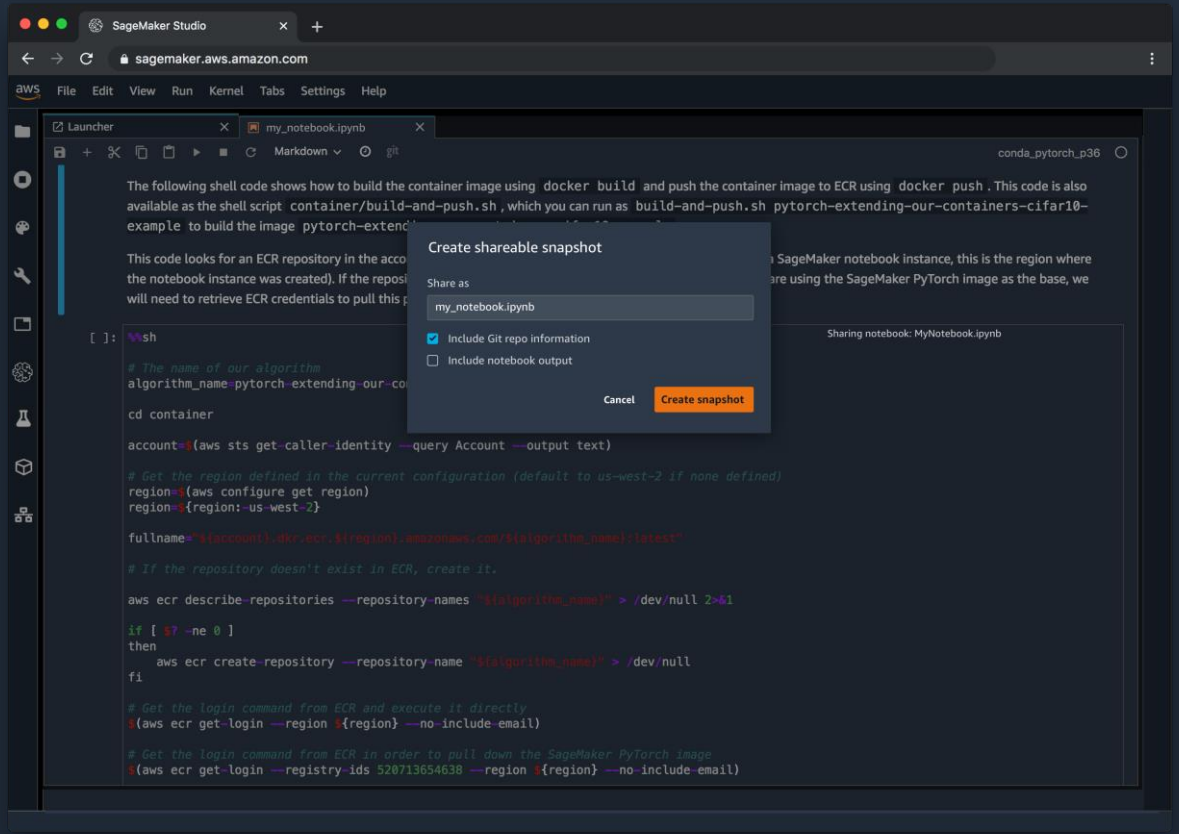
- JupyterServer runs the Jupyter server. Enables access to EFS and notebooks without selecting a kernel.
- Kernel Gateway runs a SageMaker image container (an environment). Can run up to 4 Kernel Apps (if CPU/RAM/GPU requirements are met).
- Kernel App runs individual notebooks, terminals, or other components.



SageMaker Studio UI



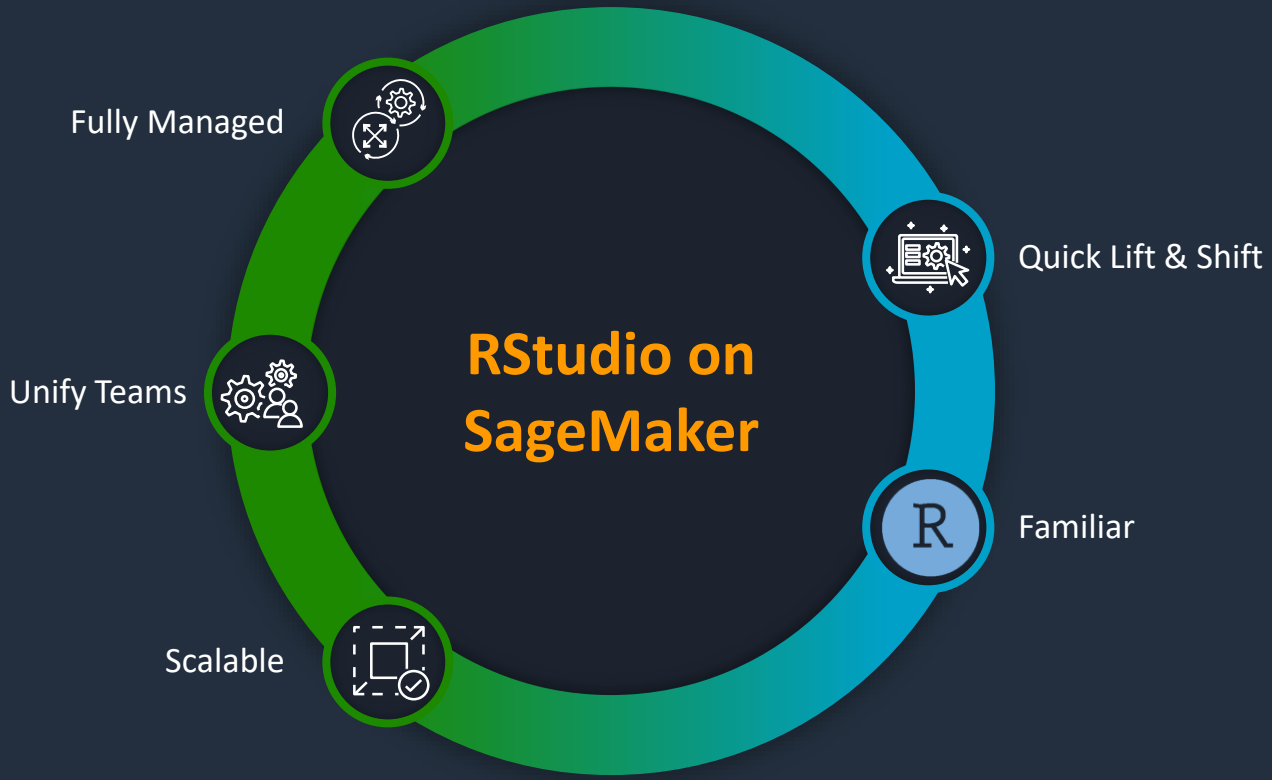
Shareable notebooks



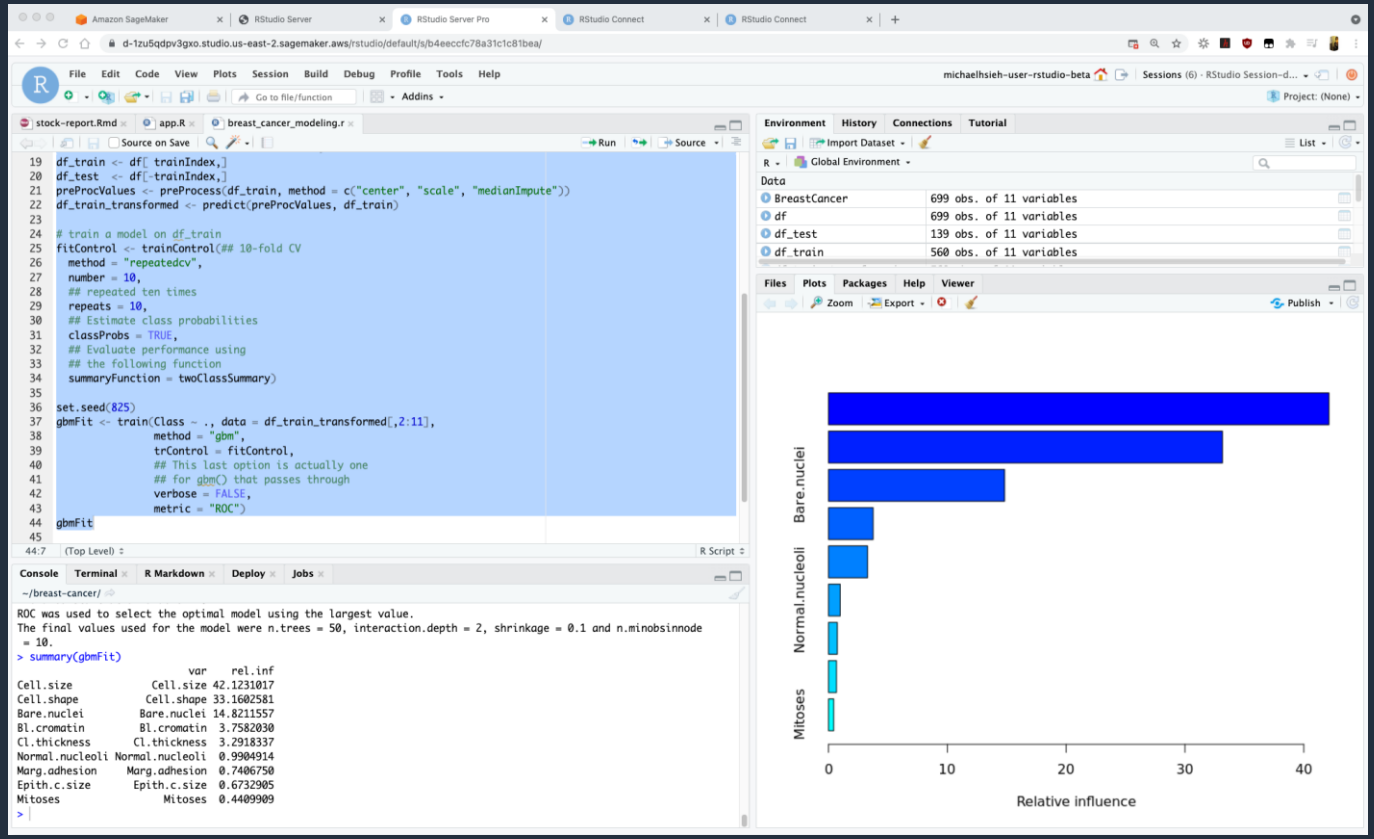
- Read-only copy (snapshot) of through a secure URL
- All dependences included.
- Snapshots are independent
- Changes in original notebook are not reflected in older snapshots
- Changes in snapshot, are also not reflected in original notebook

Demo

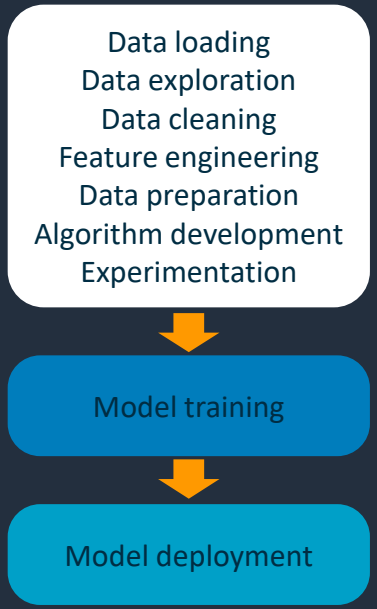
Studio notebooks



RStudio on Amazon SageMaker



ML stages have different resource requirements



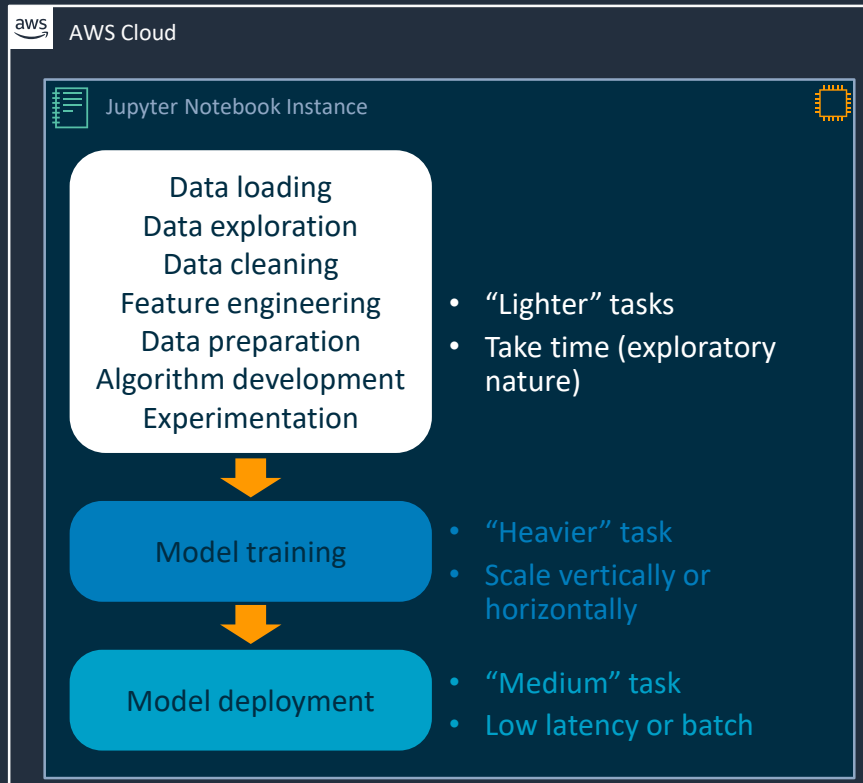
- "Lighter" tasks
- Take time (exploratory nature)

- "Heavier" task
- Scale vertically or horizontally

- "Medium" task
- Low latency

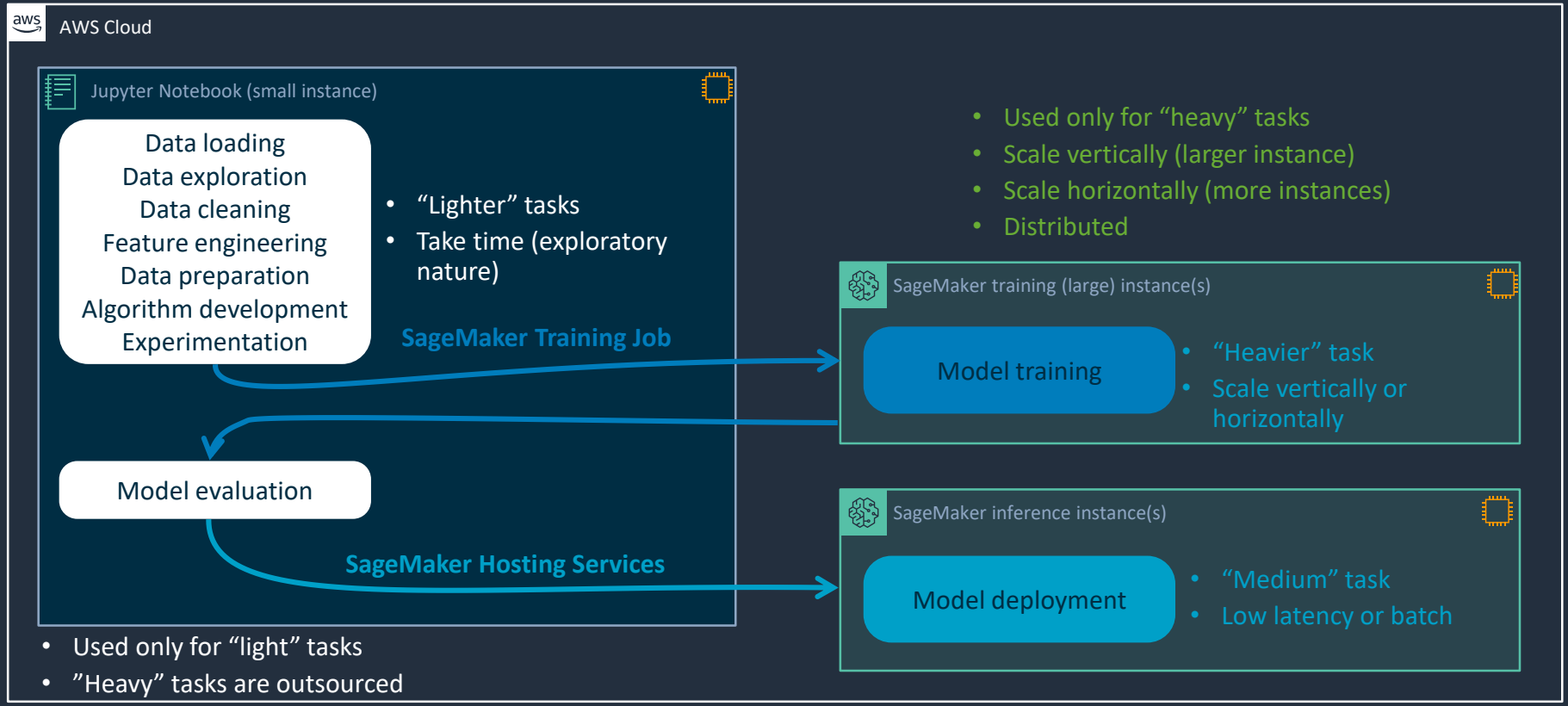
How to choose the right infrastructure?

Option 1: Run all in one Jupyter notebook instance



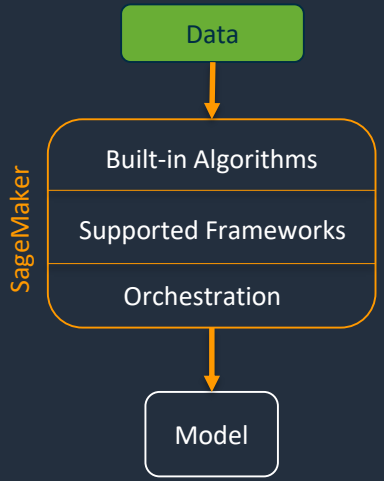
- Used both for "light" and "heavy" tasks
- Difficult to right-size, requires manual effort
- Easy to forget to downsize

Option 2: Decouple “heavy” from “light” tasks



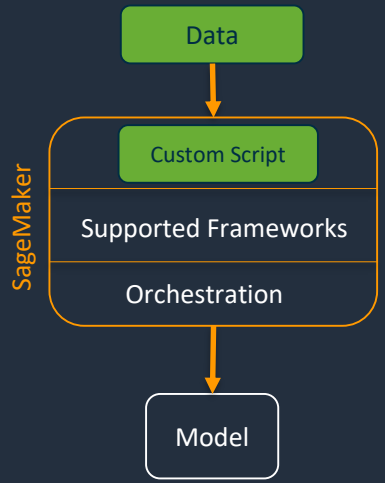
Algorithms in SageMaker

SageMaker
Built-in algorithms



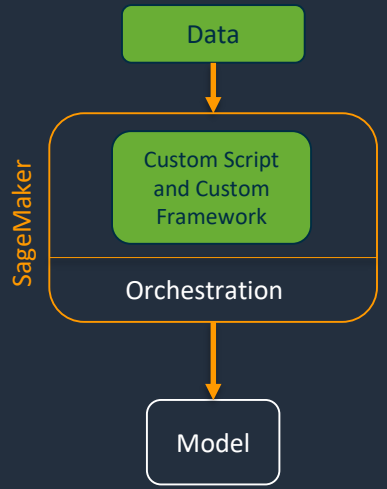
Built-in high performance algorithms

Custom script on
supported framework



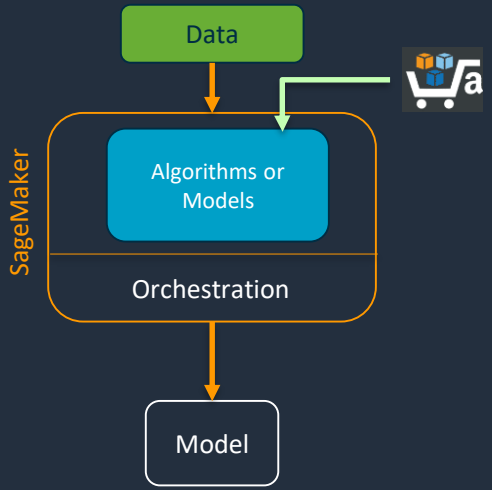
Supported frameworks include Apache MXNet, TensorFlow , Scikit-learn, PyTorch, Spark

BYO algorithm and
framework



Docker containers with your own algorithms and frameworks

AWS Marketplace
algorithms



3rd party algorithms and models

Amazon SageMaker has built-in algorithms or bring your own

Computer vision

Image classification | Object detection | Semantic segmentation

Topic modeling

LDA | NTM

Classification and regression

Linear Learner | XGBoost | KNN

Recommendations

Factorization machines

Forecasting

DeepAR

Working with text

BlazingText

Embeddings

Object2Vec

Clustering

KMeans

Sequence translation

Seq2Seq

Anomaly detection

Random cut forests | IP Insight

Feature reduction

PCA

Amazon SageMaker interfaces

1. SageMaker Console UI

2. AWS Command Line Interface (CLI):

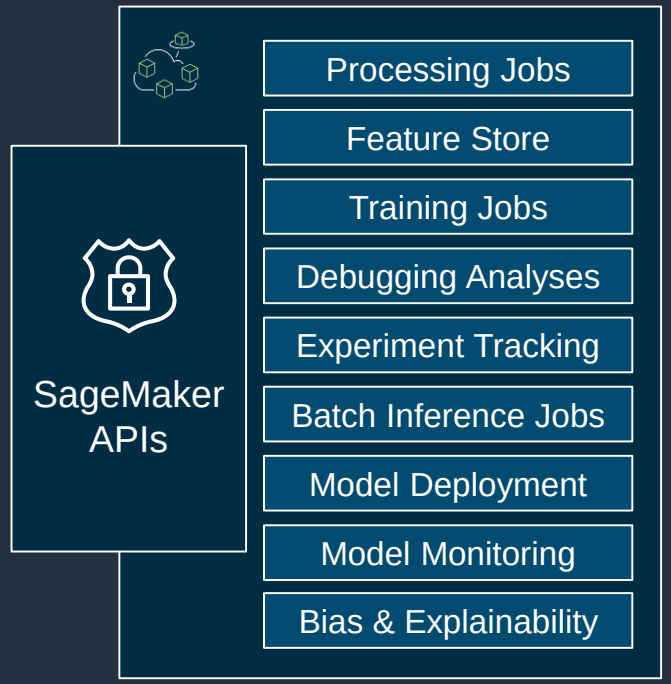
```
aws sagemaker create-endpoint
    --endpoint-name <value>
    --endpoint-config-name <value>
```

3. **sagemaker**: High-level Python SDK

```
model.deploy(initial_instance_count=1,
             instance_type='ml.m4.xlarge')
```

4. **boto3**: General AWS Python SDK

```
client.create_endpoint(EndpointName='string',
                       EndpointConfigName='string')
```

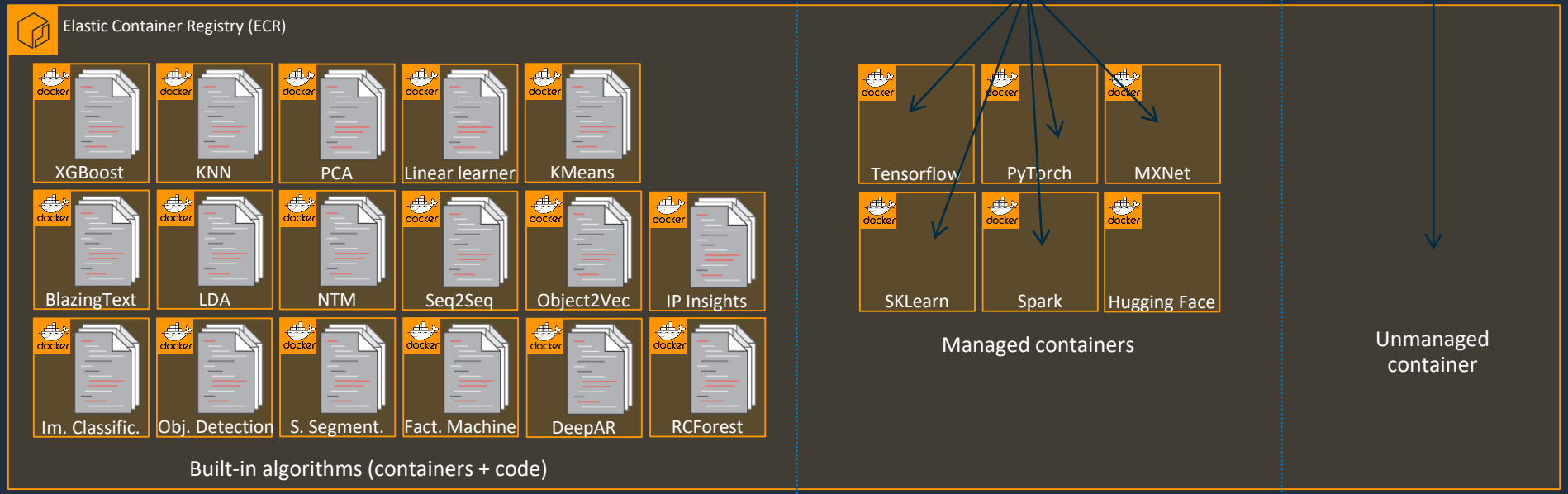


Under the hood

Select one of the built-in containers

Add your training code to one of the supported framework managed containers

Bring your own container



Build you own container

Build custom Docker containers for SageMaker with your preferred runtime, libraries, and security configurations.

1

Extend existing, open-source, SageMaker framework containers.

OR

2

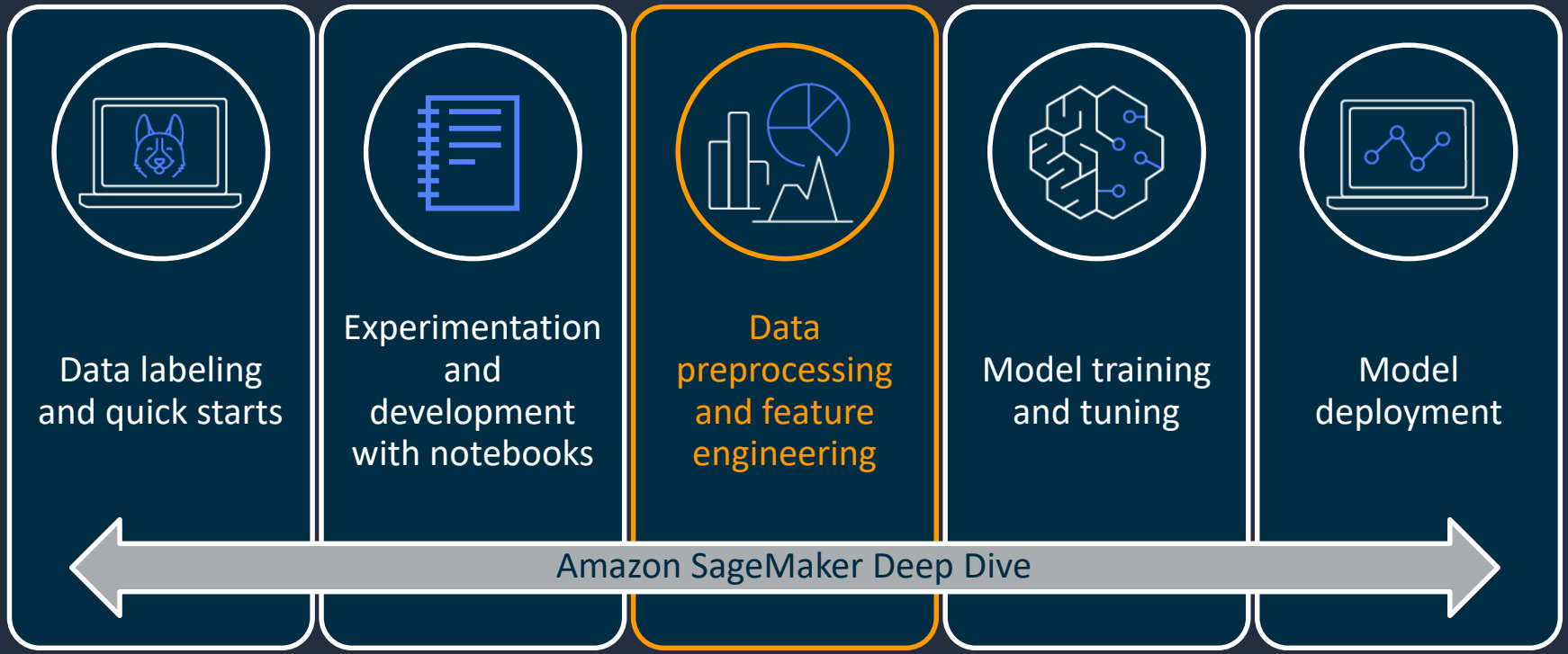
Create a container using the SageMaker Training Toolkit and SageMaker Inference Toolkit.

Features covered in this session

Resources

[Documentation] Amazon SageMaker Notebooks
[Documentation] Amazon SageMaker Studio Notebooks
[Documentation] RStudio on Amazon SageMaker
[Documentation] Algorithms in Amazon SageMaker
[Documentation] Amazon SageMaker API Reference Guide
[Documentation] Amazon SageMaker Python SDK
[Documentation] SageMaker boto3 API
[Code Samples] Amazon SageMaker Examples
[Workshop] Amazon SageMaker Immersion Day
[Workshop] Amazon SageMaker 101

Join us in the next session



aws partnercast

Thank you!

Australia & New Zealand
Brad Ryan
AI/ML Partner Development Specialist
brdryn@amazon.com

Sara van de Moosdijk
AI/ML Partner Solutions Architect
sarmoosd@amazon.com

ASEAN
Ling Chang
AI/ML Partner Development Specialist
linglych@amazon.com

Vasileios Vonikakis
AI/ML Partner Solutions Architect
vonikakv@amazon.com

India
Ankit Kandoi
Data Analytics & ML Partner Development Specialist
akkandoi@amazon.com

aws training and certification