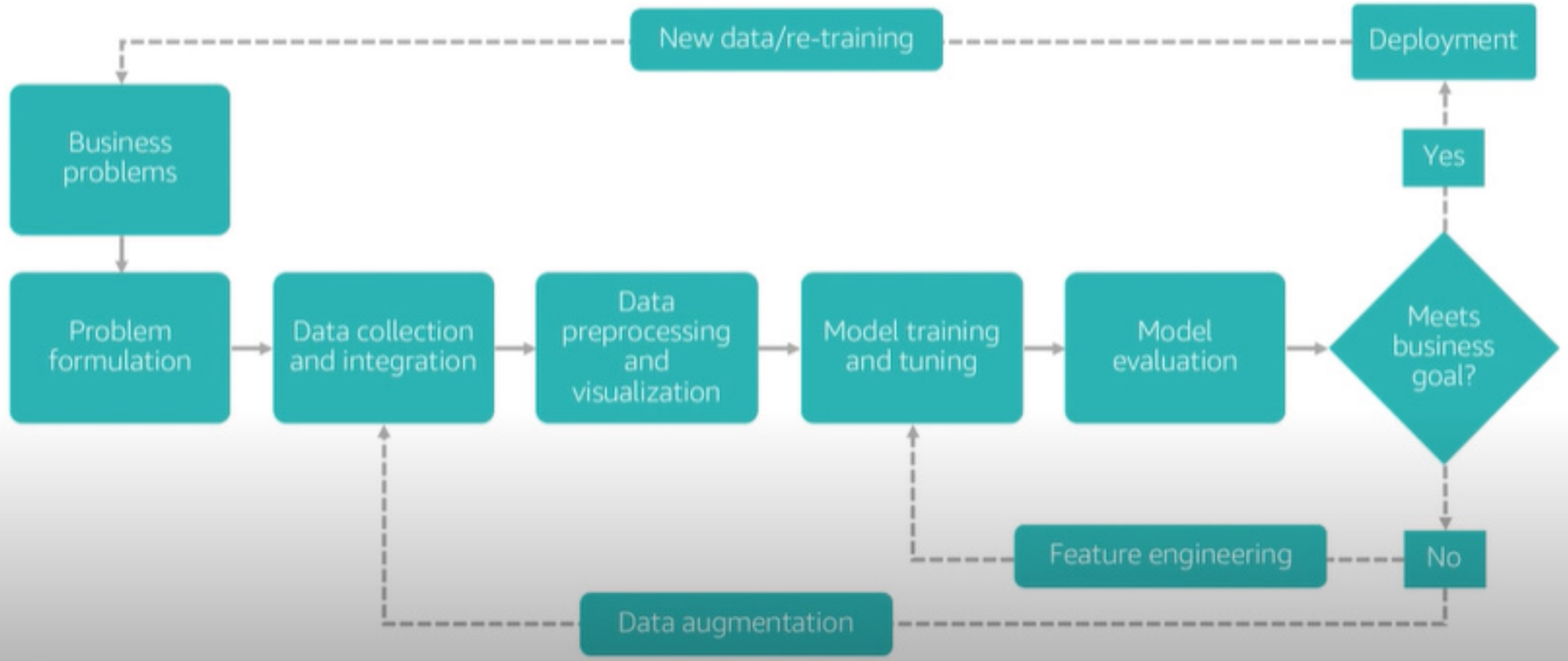
The Machine Learning Pipeline

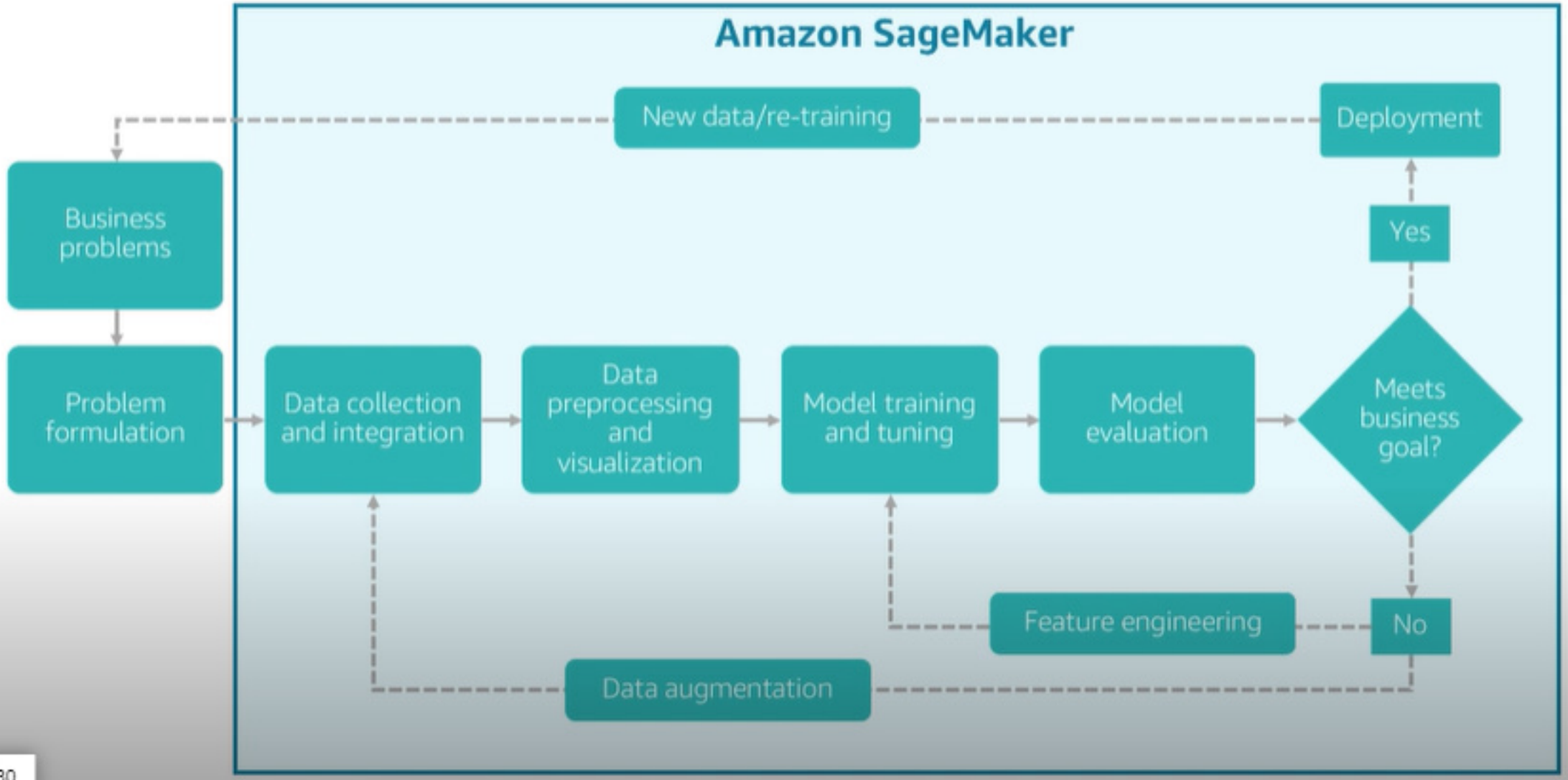


00:30

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

Where does Amazon SageMaker fit?

aws training and certification

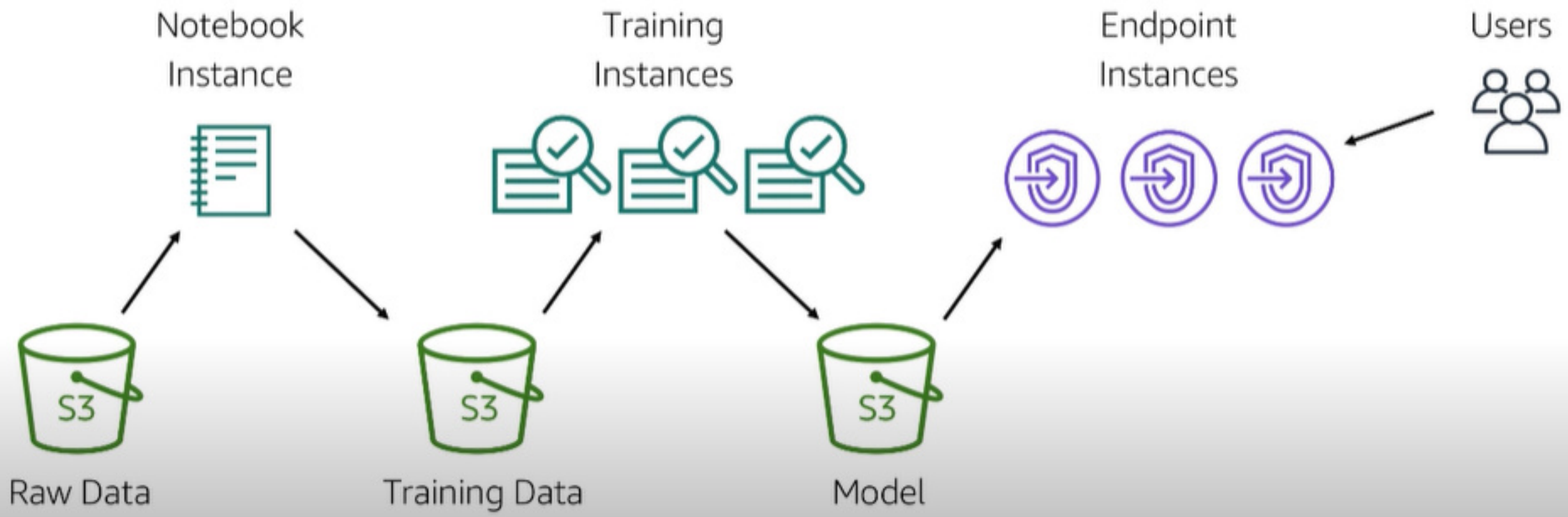


00:30

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

Amazon SageMaker Workflow

aws training and certification



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

Amazon SageMaker - Instance types

aws training and certification

	Ideal for notebooks	Ideal for traditional machine learning training				Ideal for deep learning	Ideal for inference
Instance Family	t family	m family	r family	c family	p family	g family	Elastic Inference
Workload Type	Short jobs/ notebooks	Standard CPU to memory ratio	Memory optimized	Compute optimized	Accelerated computing-training and inference	Accelerated inference, smaller training jobs	Cost-effective Inference Accelerators

For lower demand workloads → For higher demand workloads

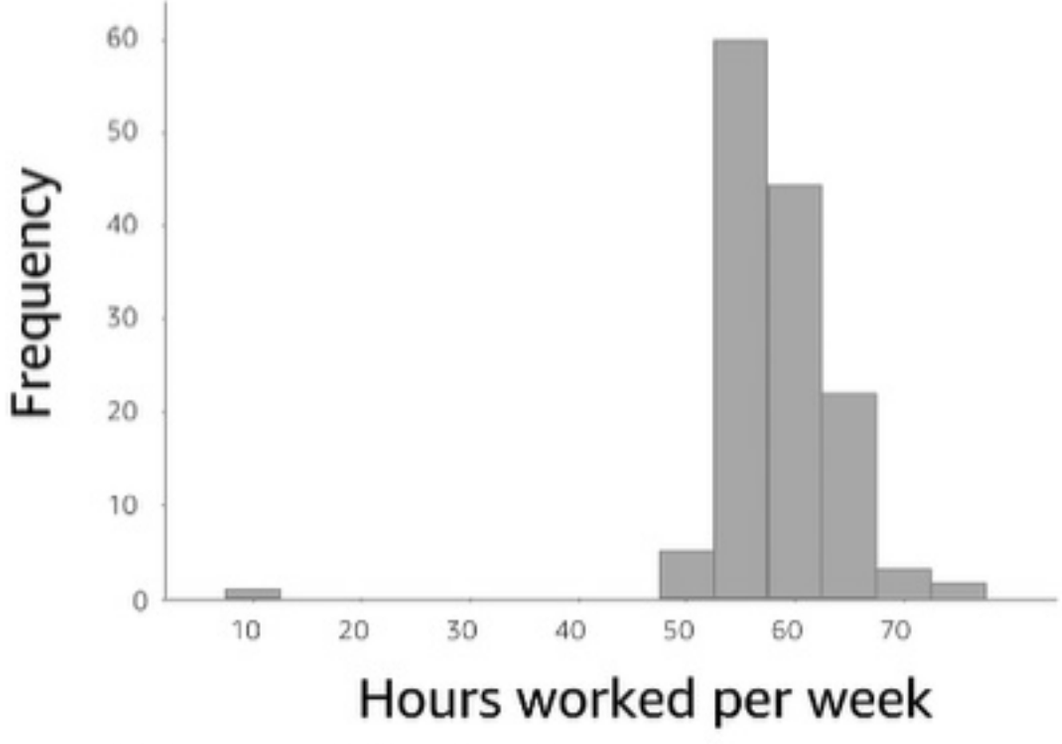
https://aws.amazon.com/sagemaker/pricing/instance-types/

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

SageMaker > Data Exploration

aws training and certification

Large amounts of data are easier for the brain to process visually.

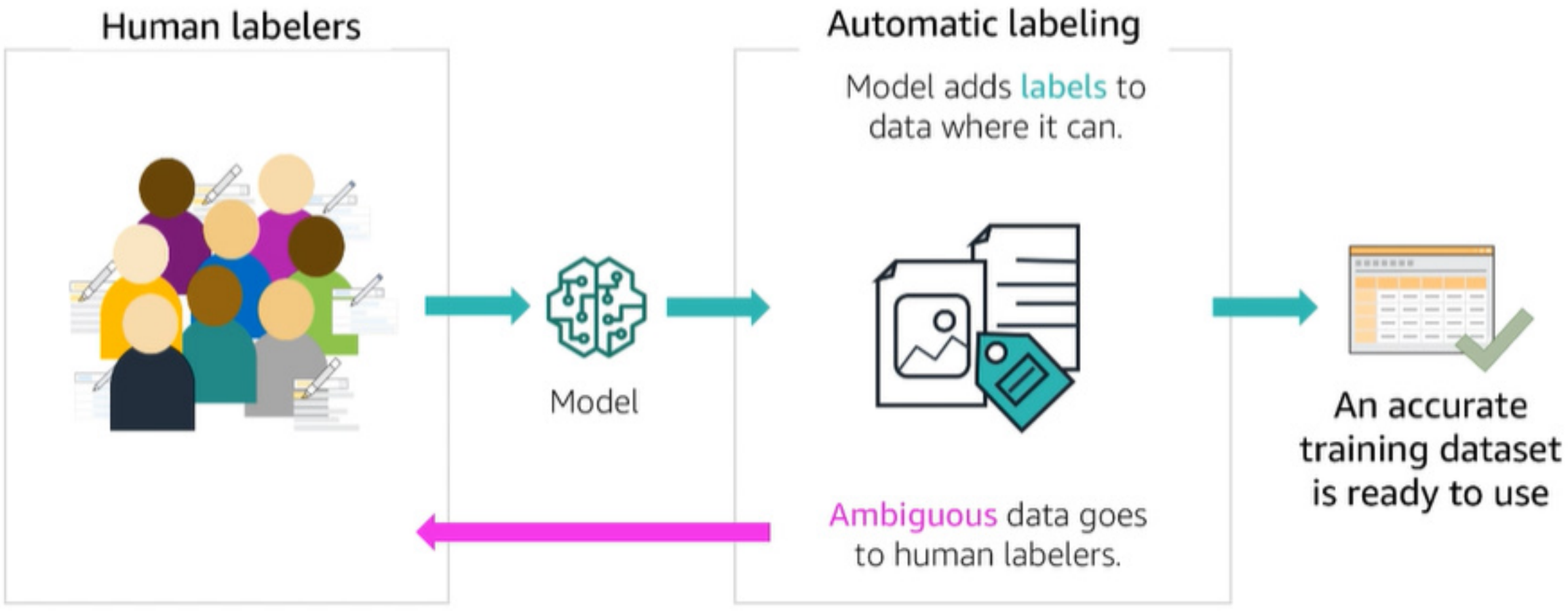


Positive correlation

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

SageMaker Ground Truth > Dataset Labels

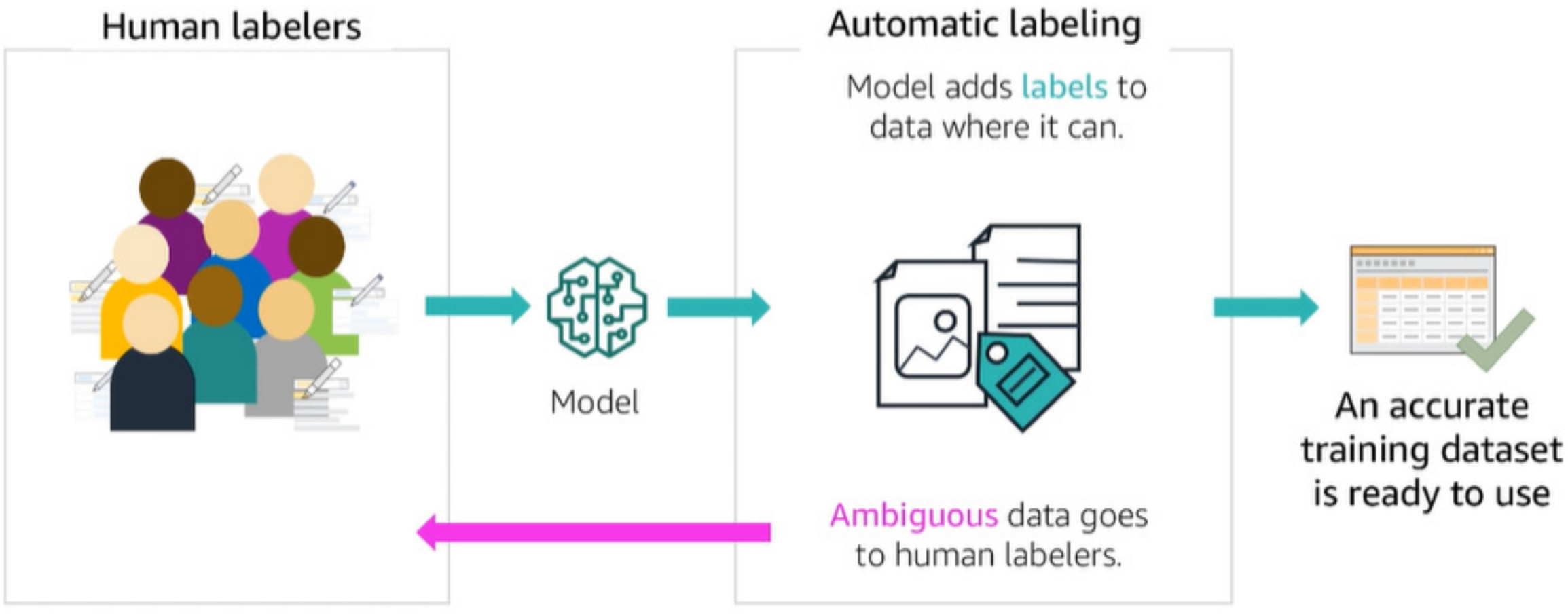
aws training and certification



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

SageMaker Ground Truth > Dataset Labels

aws training and certification



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

SageMaker > Model Selection

aws training and certification

Algorithm options

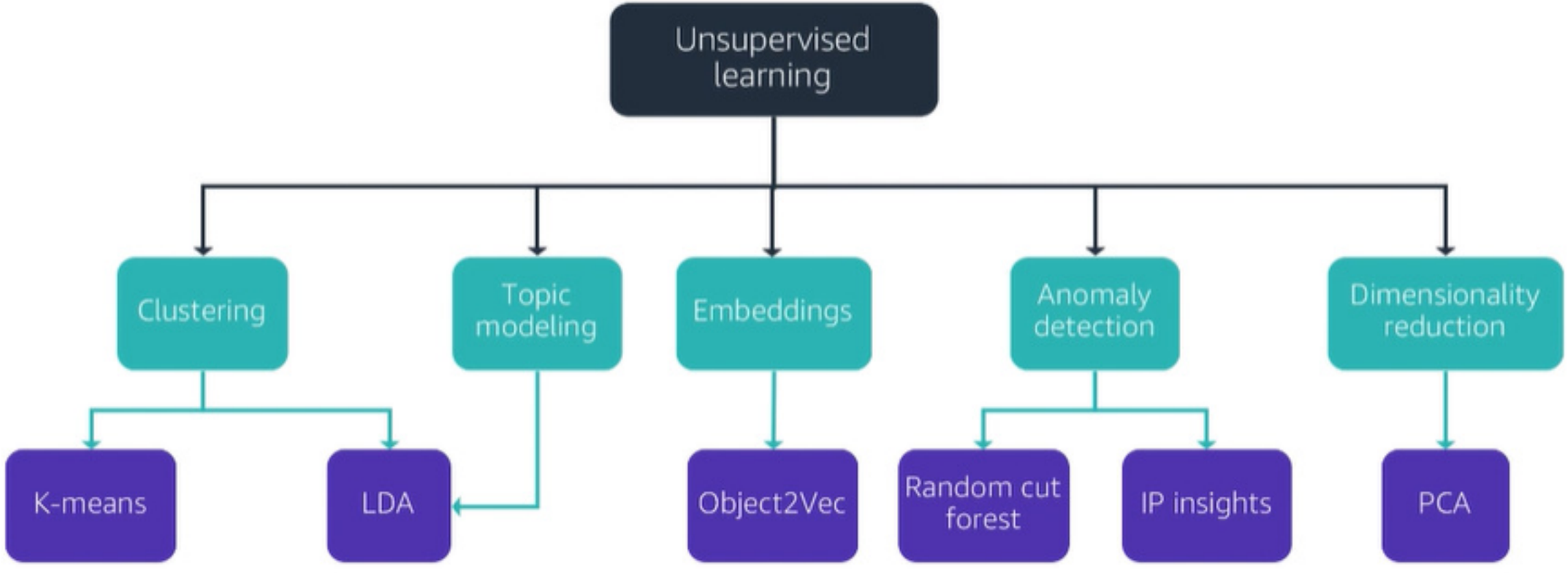
- Use a SageMaker built-in algorithm
- Write a script in your framework
- Get an algorithm from AWS Marketplace
- Bring your own algorithm

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

Choosing a SageMaker built-in algorithm

aws training and certification

What problem do you need to solve?

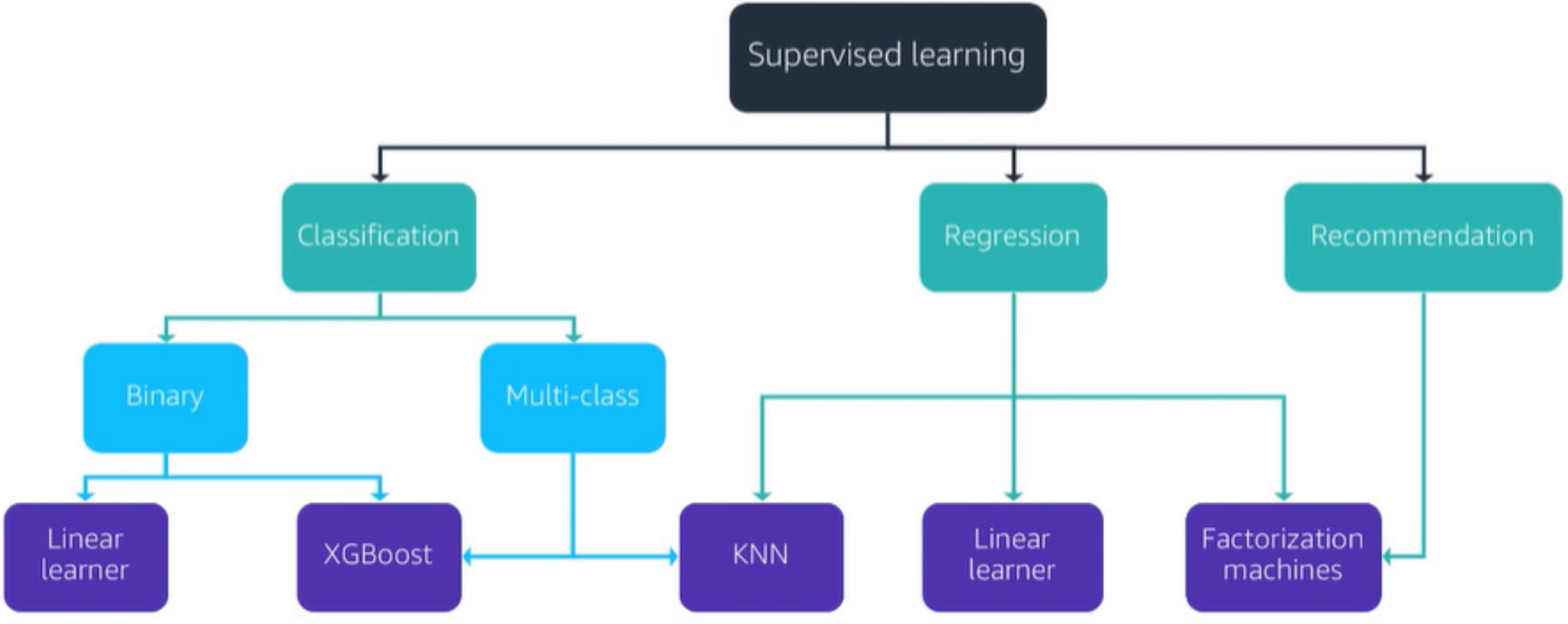


© 2020 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Choosing a SageMaker built-in algorithm

aws training and certification

What problem do you need to solve?

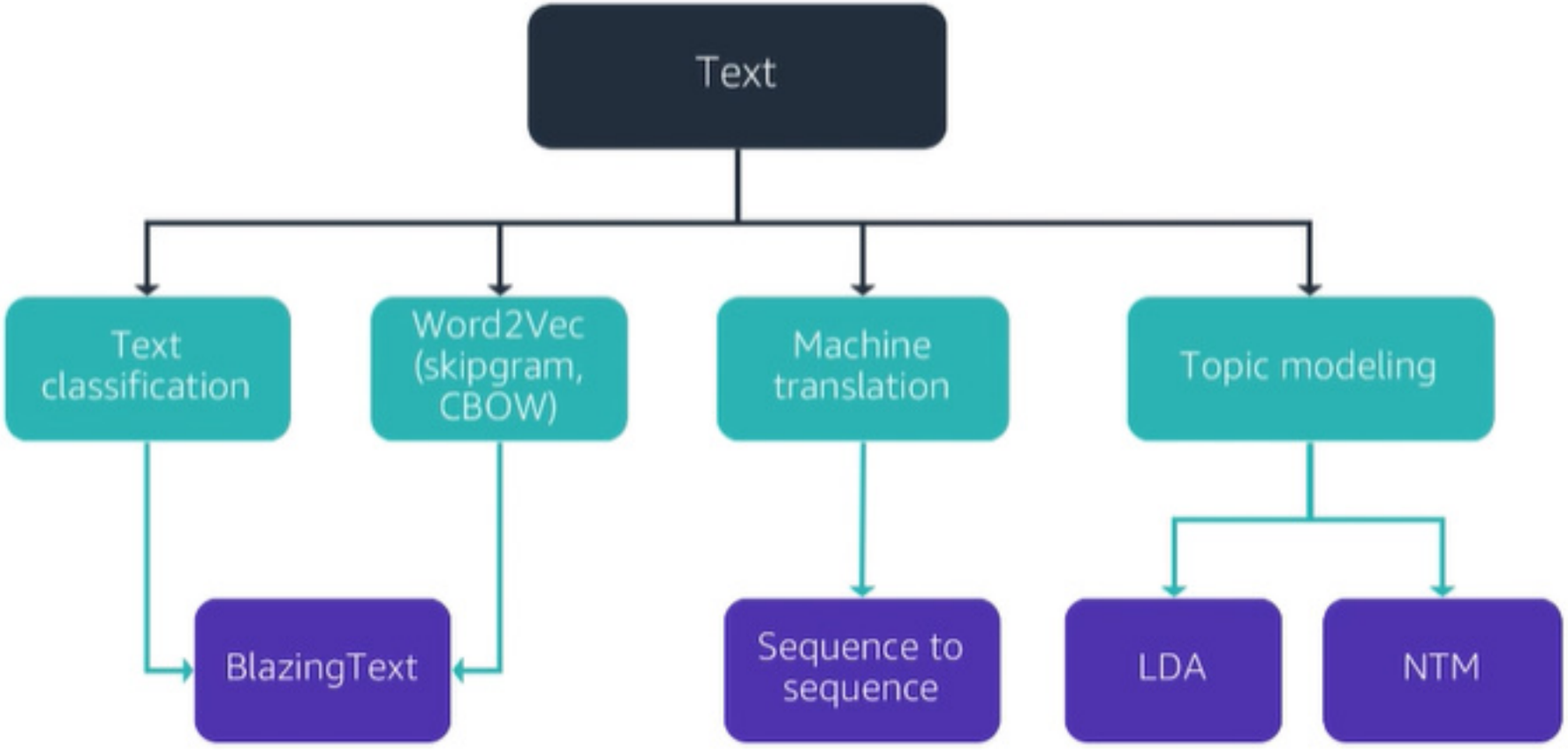


© 2020 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Choosing a SageMaker built-in algorithm

aws training and certification

What kind of data are you training with?

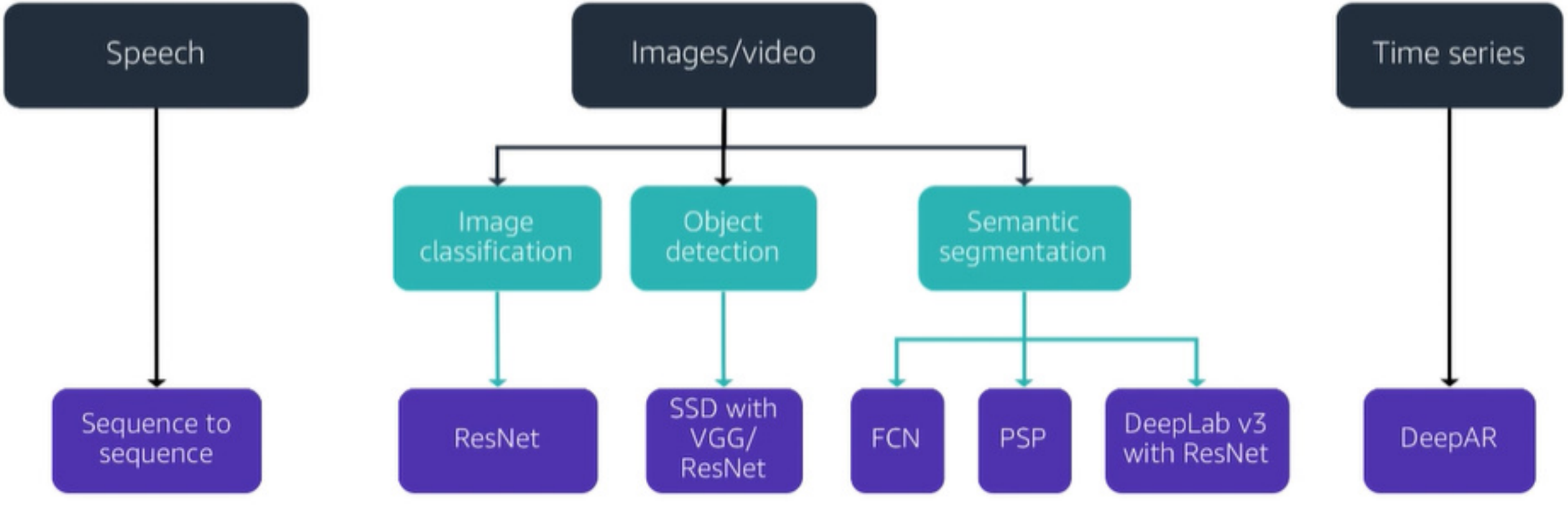


© 2020 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Choosing a SageMaker built-in algorithm

aws training and certification

What kind of data are you training with?



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

Formatting the data for an algorithm

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

Amazon SageMaker data formats

aws training and certification

Most commonly supported for built-in algorithms

Comma separated values

Record-IO protobuf

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

CSV formatting: Label on the left

aws training and certification

Participant ID	Age	Income	Education	State	Flu shot	Outbreak zone?
123456	39	45,000/ year	4-yr degree	New York	Yes	No
123457	23	1,500/ month	Baccalauréat	Minnesota	No	Yes
123458	78		Masters/PhD		Yes	Yes
123459	20	3,000,000 /year	HS diploma	California	No	No
123460	154	53,000 /year		Masters/PhD		
...	...	...	...	...	...	...

No	123456	39	45,000/ year	4-yr degree	New York	Yes
Yes	123457	23	1,500/ month	Baccalauréat	Minnesota	No
Yes	123458	78		Masters/PhD		Yes
No	123459	20	3,000,00 0/year	HS diploma	California	No
	123460	154	53,000 /year		Masters/PhD	
...	...	...	...	...	...	...

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

RecordIO-protobuf formatting

aws training and certification

Amazon SageMaker automatically performs some transformations on your recordIO-protobuf formatted data.

- Python: We recommend using those transformations
- Other languages: We recommend using the protobuf definition file provided in the AWS documentation:
  - https://docs.aws.amazon.com/sagemaker/latest/dg/cdf-training.html

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

Running training jobs in Amazon SageMaker

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

Training with Amazon SageMaker

aws training and certification

Start with an SDK:

Amazon SageMaker Python SDK

or

AWS SDK for Python (Boto 3)

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

Options for scripts, algorithms, and models

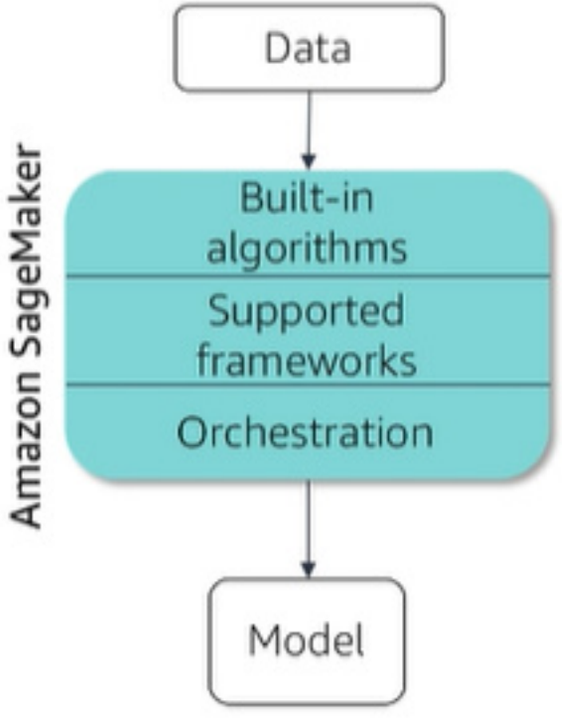
- Use a built-in algorithm *(previously discussed)*
- Use pre-built container images
- Extend a pre-built container image
- Build your own custom container image

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

Use a built-in algorithm

aws training and certification

Amazon SageMaker built-in algorithms



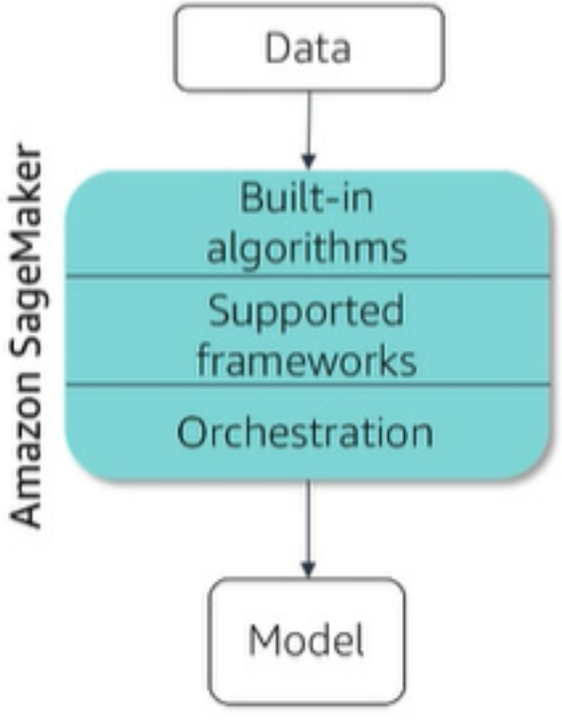
17 built-in high-performance algorithms

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

Build your own custom container image

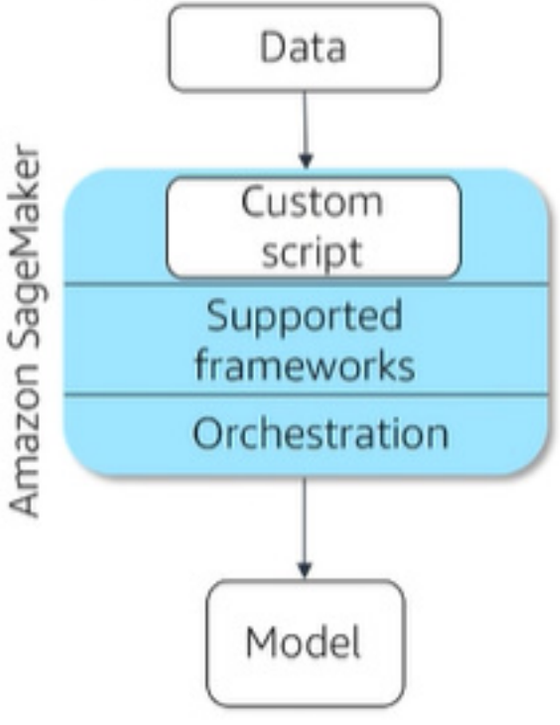
aws training and certification

Amazon SageMaker built-in algorithms



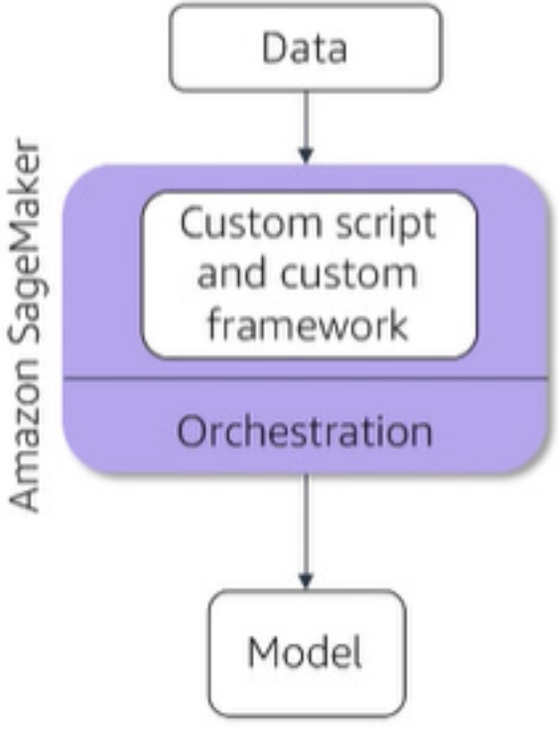
17 built-in high-performance algorithms

Custom script on supported framework



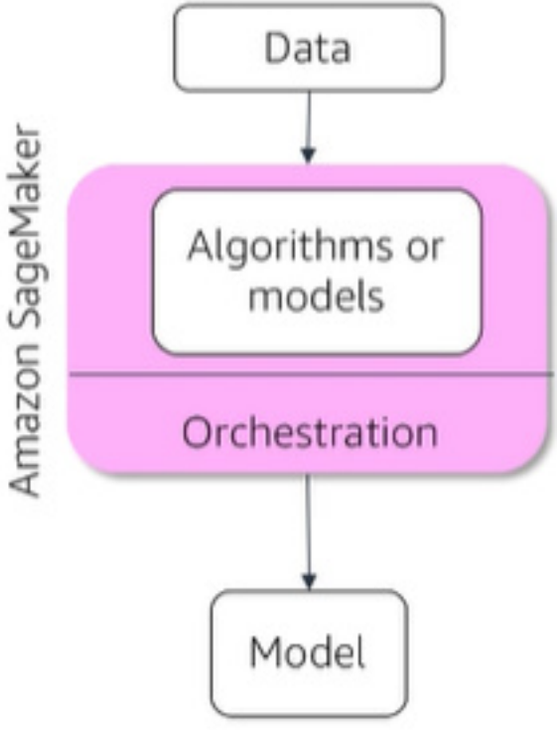
Supported frameworks: Apache MXNet, TensorFlow, Scikit-learn, PyTorch, Chainer

BYO algorithm and framework



Docker containers with your own algorithms and frameworks

AWS Marketplace algorithms



Third-party algorithms and models

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

AWS Marketplace for machine learning

aws training and certification

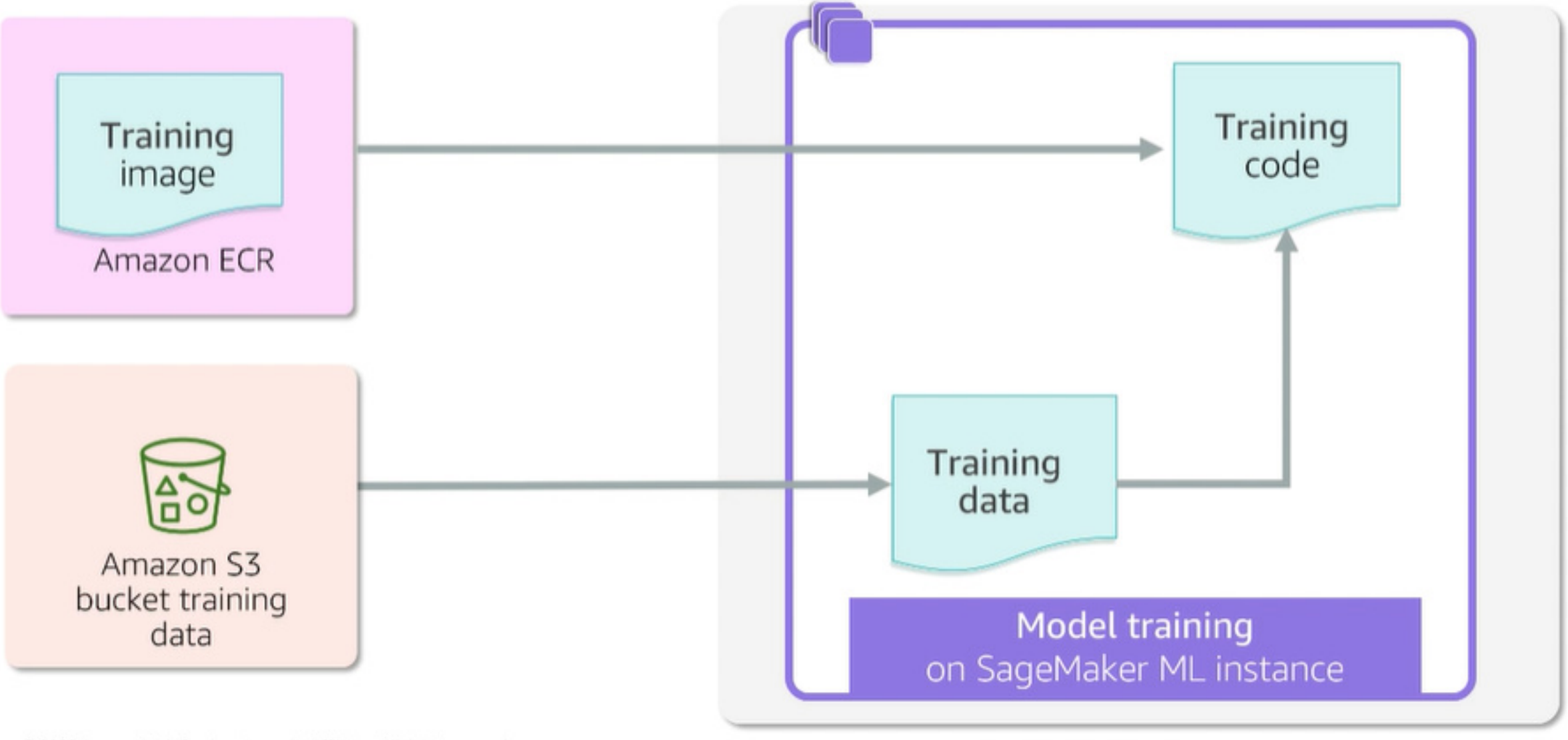
Curated catalog of **250+ ML model packages** and **algorithms**

- **14+** industry segments (computer vision, natural language processing, speech recognition, fraud detection, etc.)
- Found through Amazon SageMaker console or AWS Marketplace
- Deploy through SageMaker console, SageMaker SDK, or AWS CLI
- Accessed through RESTful endpoints
- Consolidated into your AWS billing
- Pricing: Free, Free Trial, and Paid Subscriptions

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

Creating a training job in Amazon SageMaker

aws training and certification



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

Example training deployment code

aws training and certification

Import →

Hardware →

Parameters →

Start Training →

```
import boto3
import sagemaker

sess = sagemaker.Session()

pca = sagemaker.estimator.Estimator(containers[boto3.Session().region_name],
                                    role,
                                    train_instance_count=1,
                                    train_instance_type='ml.c4.xlarge',
                                    output_path=output_location,
                                    sagemaker_session=sess)
pca.set_hyperparameters(feature_dim=50000,
                        num_components=10,
                        subtract_mean=True,
                        algorithm_mode='randomized',
                        mini_batch_size=200)

pca.fit({'train': s3_train_data})
```

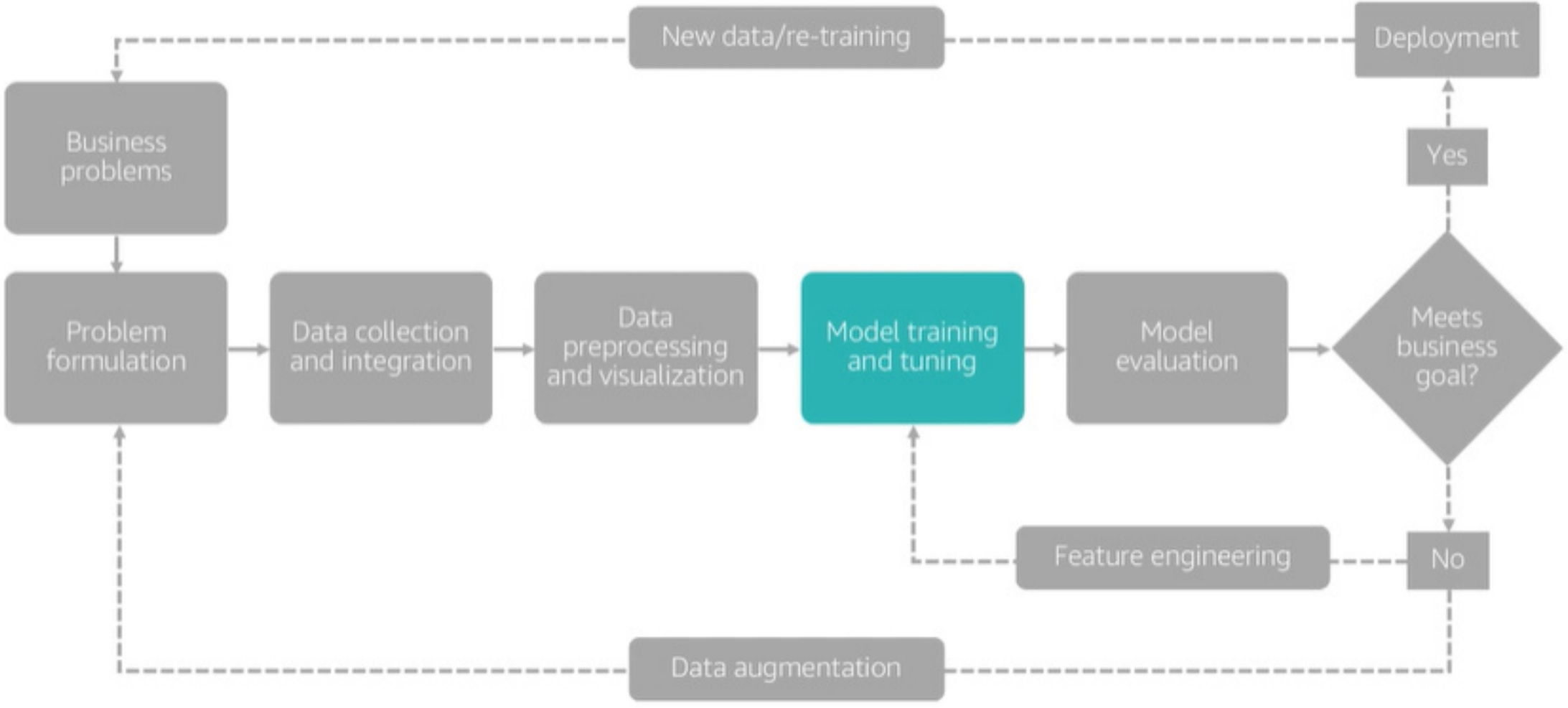
© 2020 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Hyperparameter tuning

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

The ML pipeline

aws training and certification



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

Hyperparameter categories

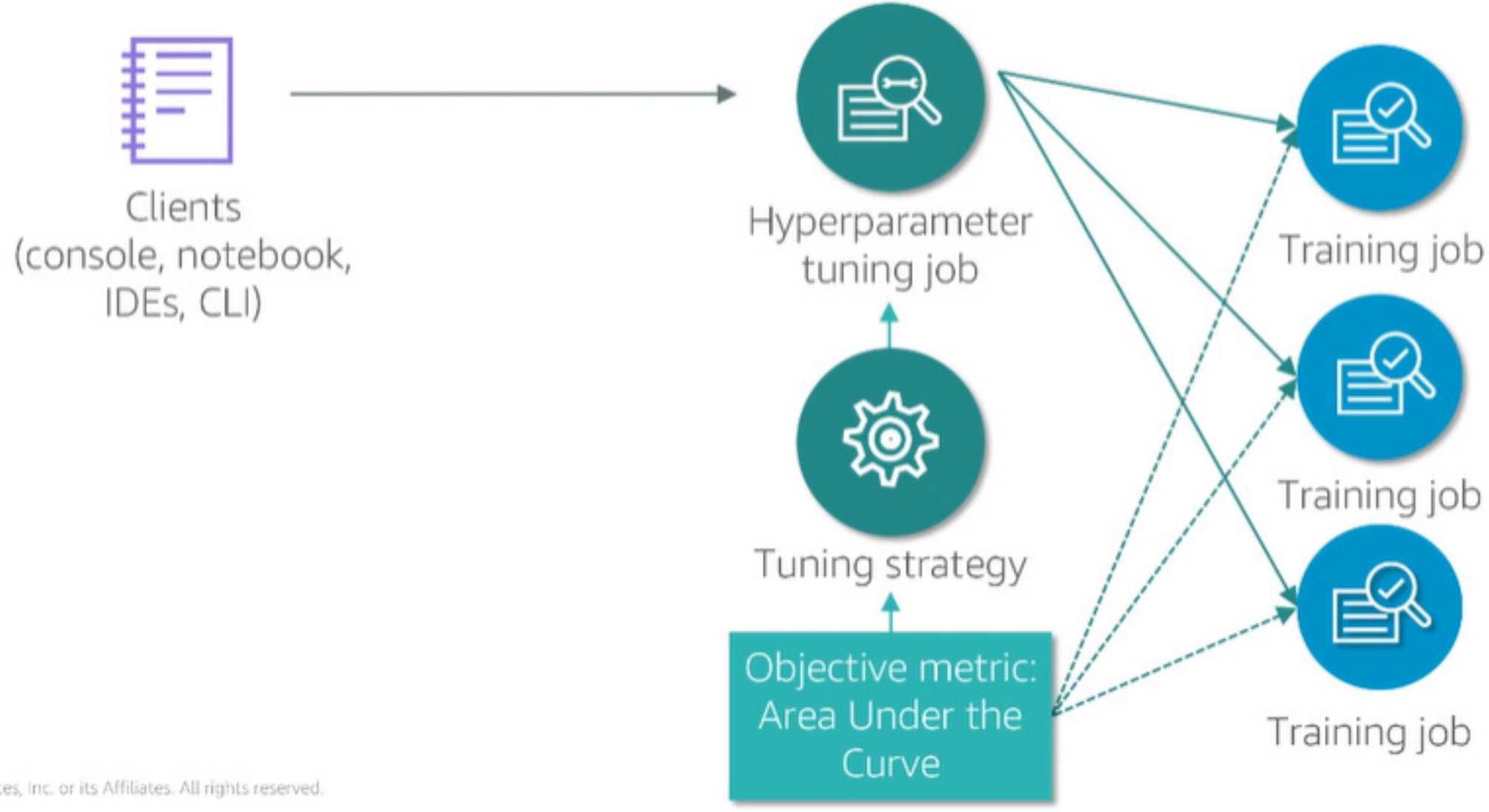
aws training and certification

Model	Optimizer	Data
Help define the model	How the model learns patterns on data	Define attributes of the data itself
Filter size, pooling, stride, padding	Gradient descent, stochastic gradient descent	Useful for small/homogenous datasets

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

Amazon SageMaker offers automated hyperparameter tuning

aws training and certification



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

Amazon SageMaker offers automated hyperparameter tuning

aws training and certification

Clients
(console, notebook,
IDEs, CLI)

Hyperparameter tuning job

Tuning strategy

Objective metric: Area Under the Curve

Training job

Training job

Training job

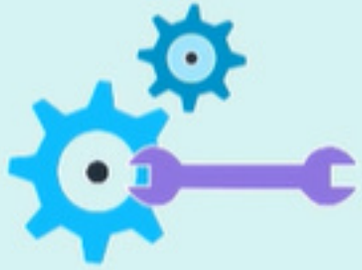
Model name	Objective metric	Eta	Max_depth	…
Model3	0.8	0.7	6	…
Model1	0.75	0.09	5	…
…	…	…	…	…

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

Your model is ready…

aws training and certification

Trained Model



Tuned Model

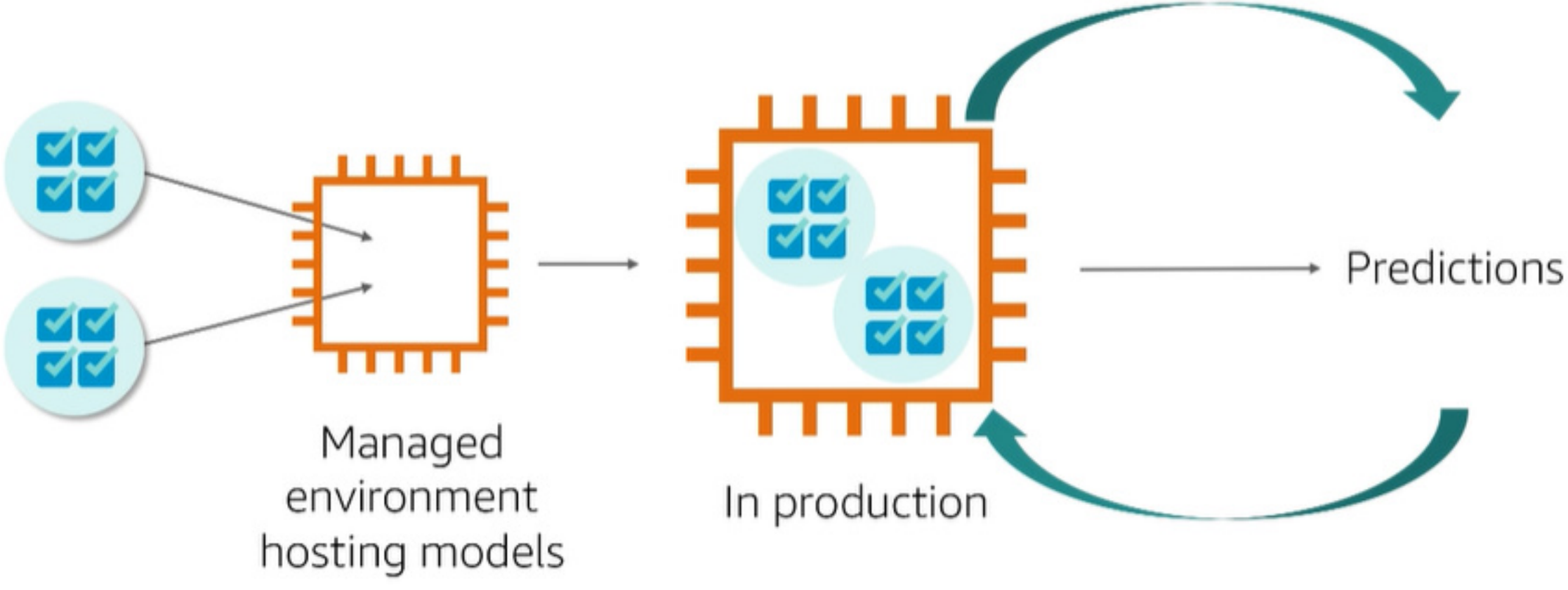
Tested/Ready Model

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Deployment

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

The goal of deployment



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon SageMaker for Deployment

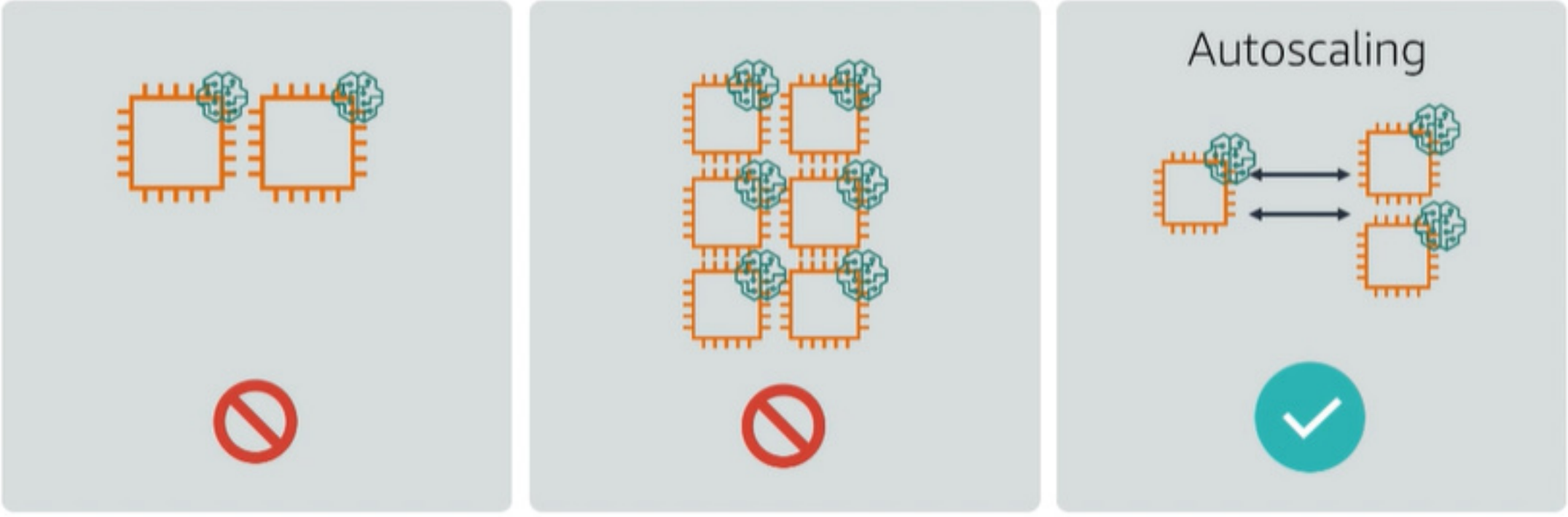
aws training and certification

Amazon Sagemaker

- Has a variety of deployment options
- Provides model hosting services
- Provides an HTTPS endpoint

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Enable autoscaling



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Consider GPU

Deep learning models require high power GPU instances for training, but inferences against the deep learning models do not typically need the full power of a GPU.

- **Amazon Elastic Inference** enables you to attach low-cost, GPU-powered acceleration to Amazon EC2 and Amazon SageMaker instances to reduce the cost of running deep learning inferences

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Creating a good baseline

Without SageMaker, you are responsible for:

- Creating the AMI (Amazon Machine Image) containing your model artifact
- Launching one or more EC2 instances with this AMI
- Configuring the autoscaling options necessary to scale

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Batch vs. Real time inferences

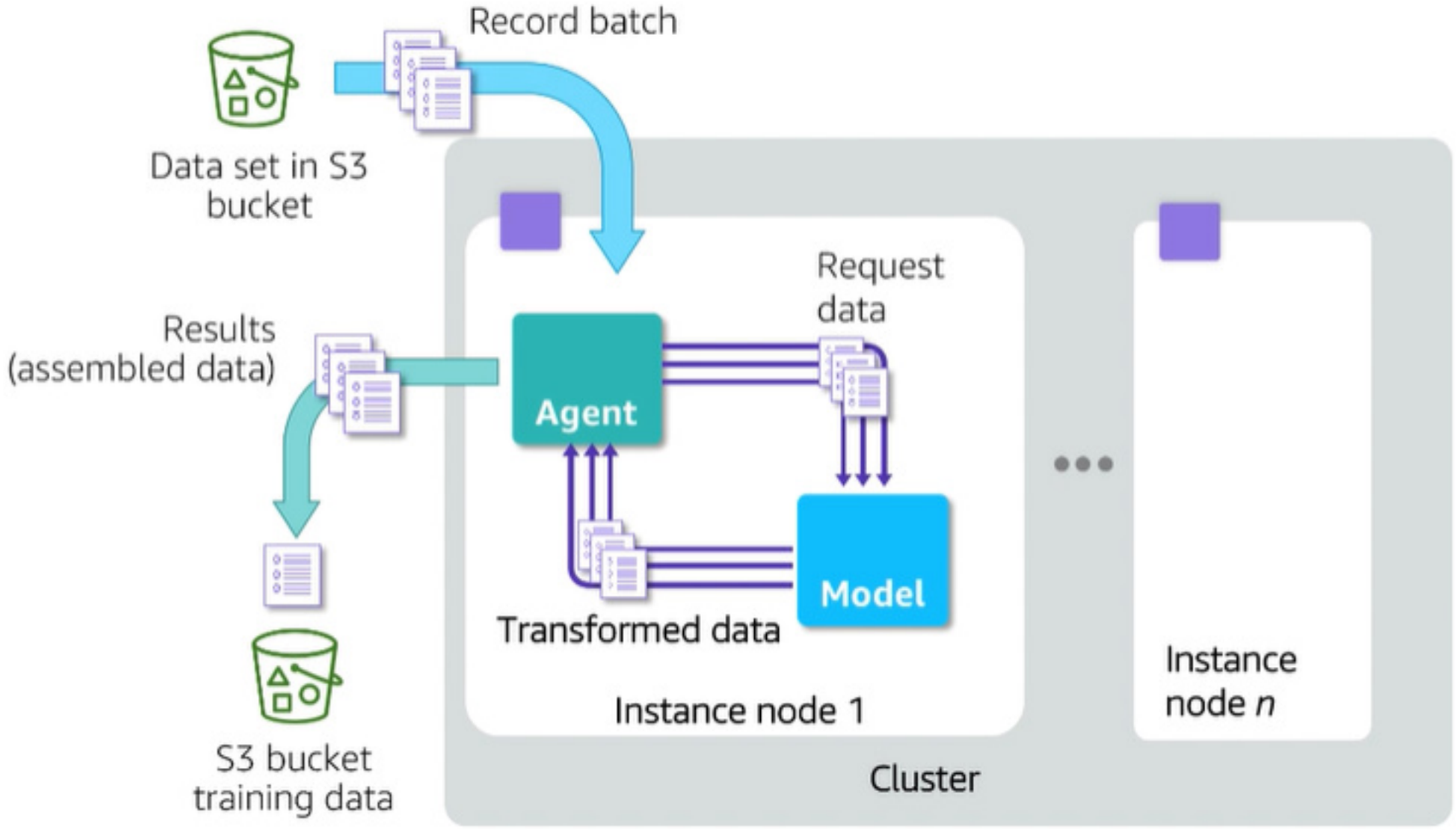
aws training and certification

On-demand as a batch	Endpoints available 24/7
• Model available periodically • Predictions are asynchronous • Dataset has multiple rows • Results available on a scheduled basis when the job's done	• Model available all the time • Predictions are synchronous • Single observation of data • Results available as users interact in real time

Vs.

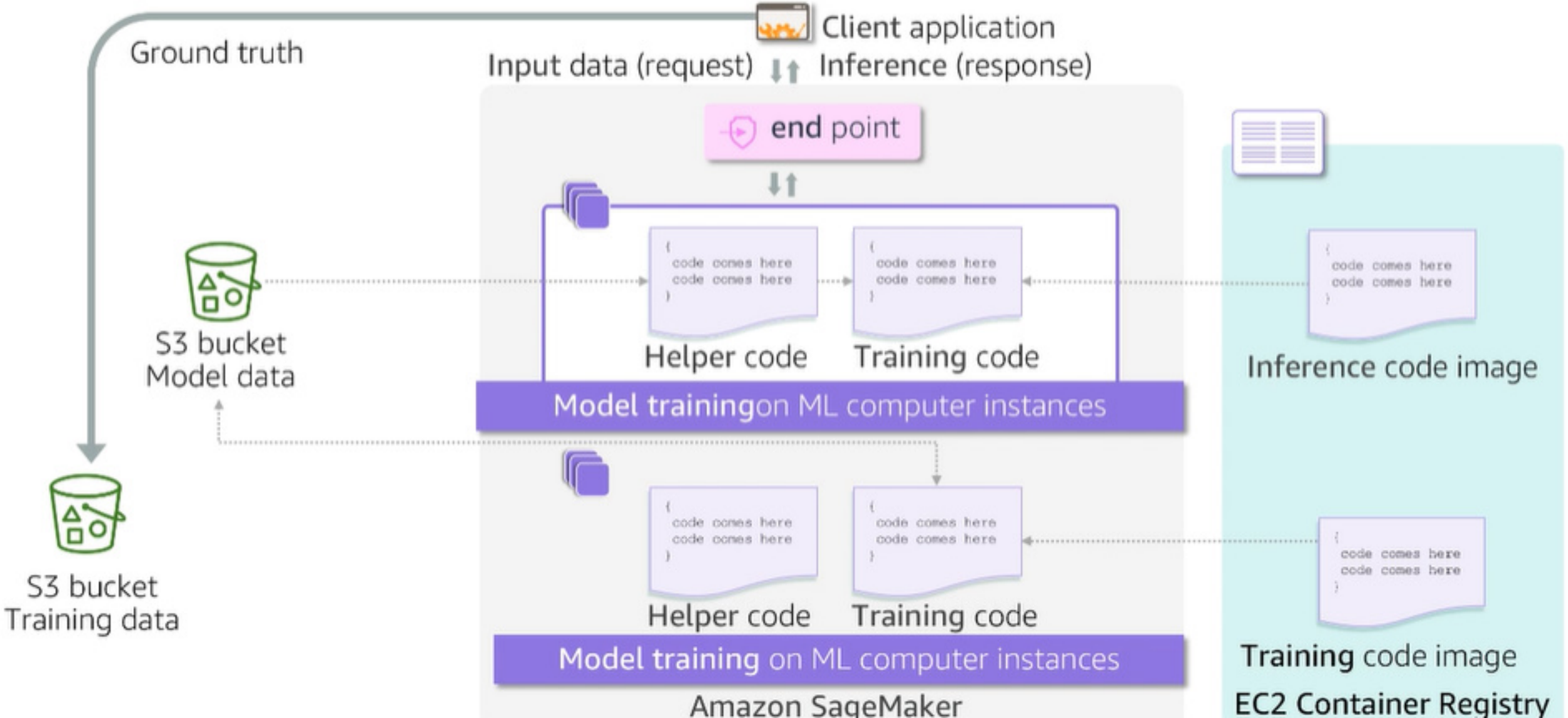
© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Getting predictions on a scheduled basis



Getting predictions in real time
Create an HTTPS endpoint

aws training and certification



When to use batch transform

aws training and certification

Use batch transform when you:

- Want to get inferences for an entire dataset and index them to serve inferences in real time
- Don't need a persistent endpoint that applications (for example, web or mobile apps) can call to get inferences
- Don't need the subsecond latency that Amazon SageMaker hosted endpoints provide
- Additionally, you can also use batch transform to preprocess your data before using it to train a new model or generate inferences.

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon SageMaker inference pipelines

- Pass raw input data
- On batch or real-time inference requests
- Use with built-in or custom algorithms

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

What are my next steps?

aws training and certification

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

Learn at your own pace

Expand your cloud security skills with our self-paced digital course, Amazon SageMaker: Build an Object Detection Model Using Images Labeled with Ground Truth.

Learn from AWS experts

Build your cloud security skills with our classroom courses, *Practical Data Science with Amazon SageMaker*.

Ramp-Up Guides

Our Machine Learning *Ramp-Up Guide* offers many resources to help build your knowledge of Machine Learning in the AWS Cloud.

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

Thank You

© 2020 Amazon Web Services, Inc. or its affiliates. All rights reserved. This work may not be reproduced or redistributed, in whole or in part, without prior written permission from Amazon Web Services, Inc. Commercial copying, lending, or selling is prohibited. Corrections or feedback on the course, please email us at: aws-course-feedback@amazon.com. For all other questions, contact us at: https://aws.amazon.com/contact-us/aws-training/. All trademarks are the property of their owners.

aws training and certification