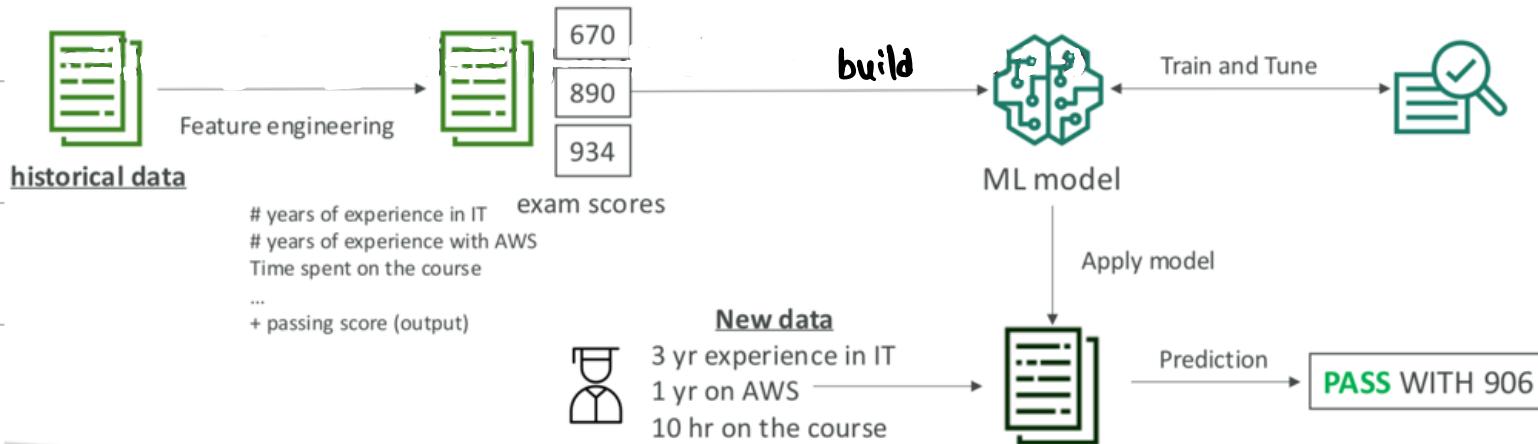


## Amazon SageMaker

- fully managed service for developers / data scientists to build ML models.
- difficult to do all the processes in one place + provision servers.



1. Collect and prepare data

2. Build and train machine learning models

3. Deploy the model and monitor the performance of the predictions.

- Built-in algorithms:

↳ Supervised learning (linear regression, classification, KNN)

↳ Unsupervised learning (PCA, K-means, Anomaly Detection)

↳ Textual Algorithms (NLP, Summarization ...)

↳ Image Processing (classification, detection)

- Automatic Model Tuning (AMT): Once the objective metric is defined, AMT will choose hyperparameter ranges, search strategy, maximum runtime of a tuning job and early stop condition. It saves both money & time

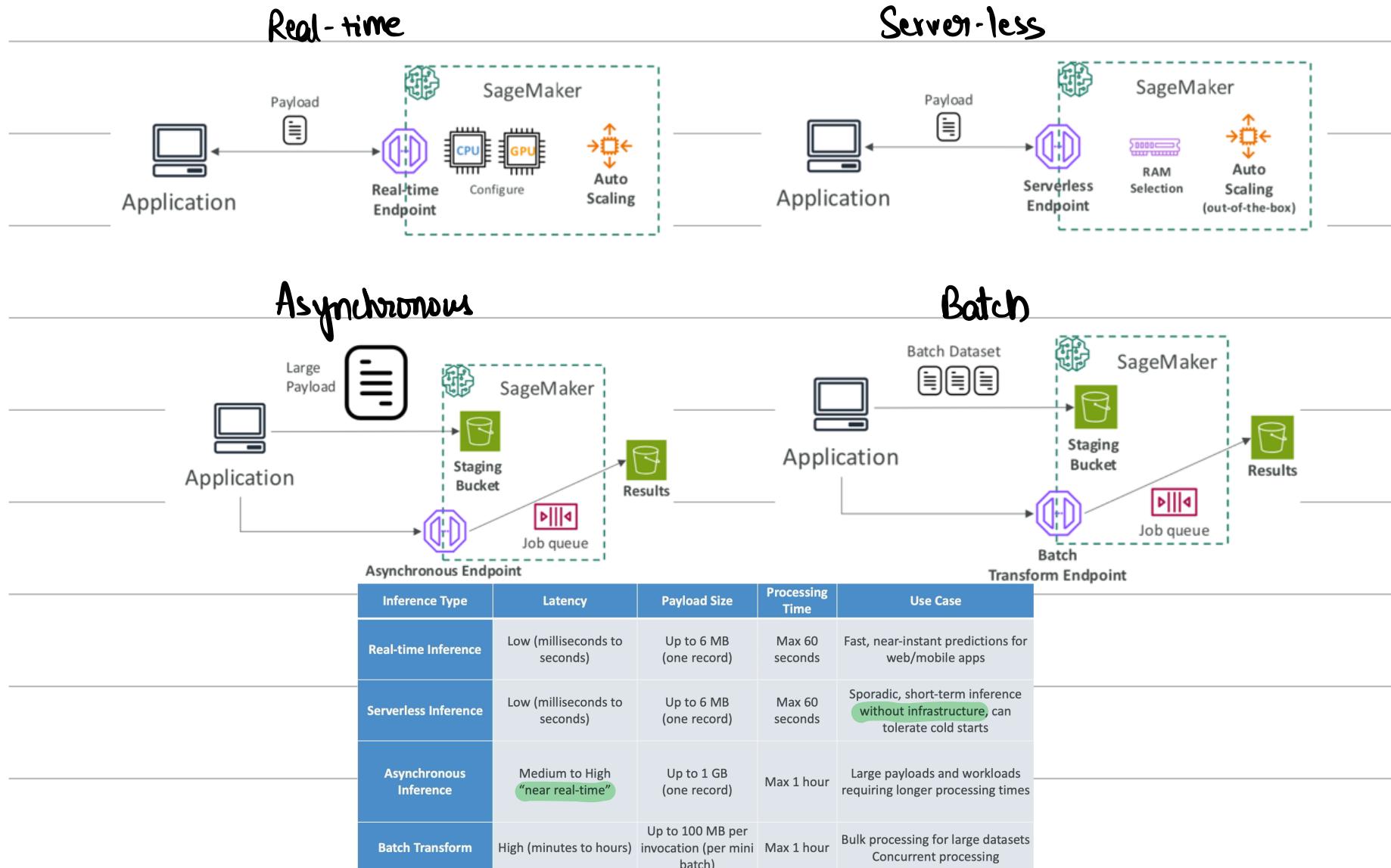
- Deploy with one click and have automatic scaling.

1. Real-time: one prediction at a time.

2. Serverless: idle period b/w traffic spikes - cold starts

3. Asynchronous: large payloads & long processing times. Requested Resources are in S3

4. Batch: Used to predict for an entire dataset & uses S3



- You perform all the work using a unified interface called SageMaker Studio.

- Sage Maker - Data Wrangler:

↳ Used to prepare image data & tabular data for machine learning.

↳ platform to perform data preparation, transformation and feature engineering.

↳ Acts as a data quality tool and also supports SQL.

- Sage Maker - Feature Store:

↳ features are inputs to ML models used during training and used for inference.

↳ It is important to have high quality features across your datasets in your company for reuse.

↳ It ingests features from a variety of sources.

↳ Transform data to feature with feature store (or) publish directly from Data Wrangler.

↳ You can access these features within SageMaker Studio too.

## - Sage Maker Clarity:

- ↳ Evaluate foundation models on the human-factors (friendliness) humor)
- ↳ Your own employees or AWS employees ; your own dataset or use built-in.
- ↳ It is a part of Sage Maker Studio.

## i. Model Explainability:

- ↳ It explains how an ml model is making predictions
- ↳ Understand the model characteristics and debug predictions.

Ex: Why is the model predicting a -ve value?

## ii. Detect Bias(human):

- ↳ Ability to detect and explain biases in your datasets and models.
- ↳ When you specify input feature, bias is measured using statistical metrics.

## - Sage Maker Ground Truth:

- ↳ RLHF: Reinforcement Learning from human feedback is included in 'reward' function so that the model is more aligned to the human preferences.
- ↳ We can also use human feedback while creating & evaluating models & to create labels.
- ↳ Sage Maker Ground Truth Plus: Label Data.

## - Sage Maker - ML Governance:

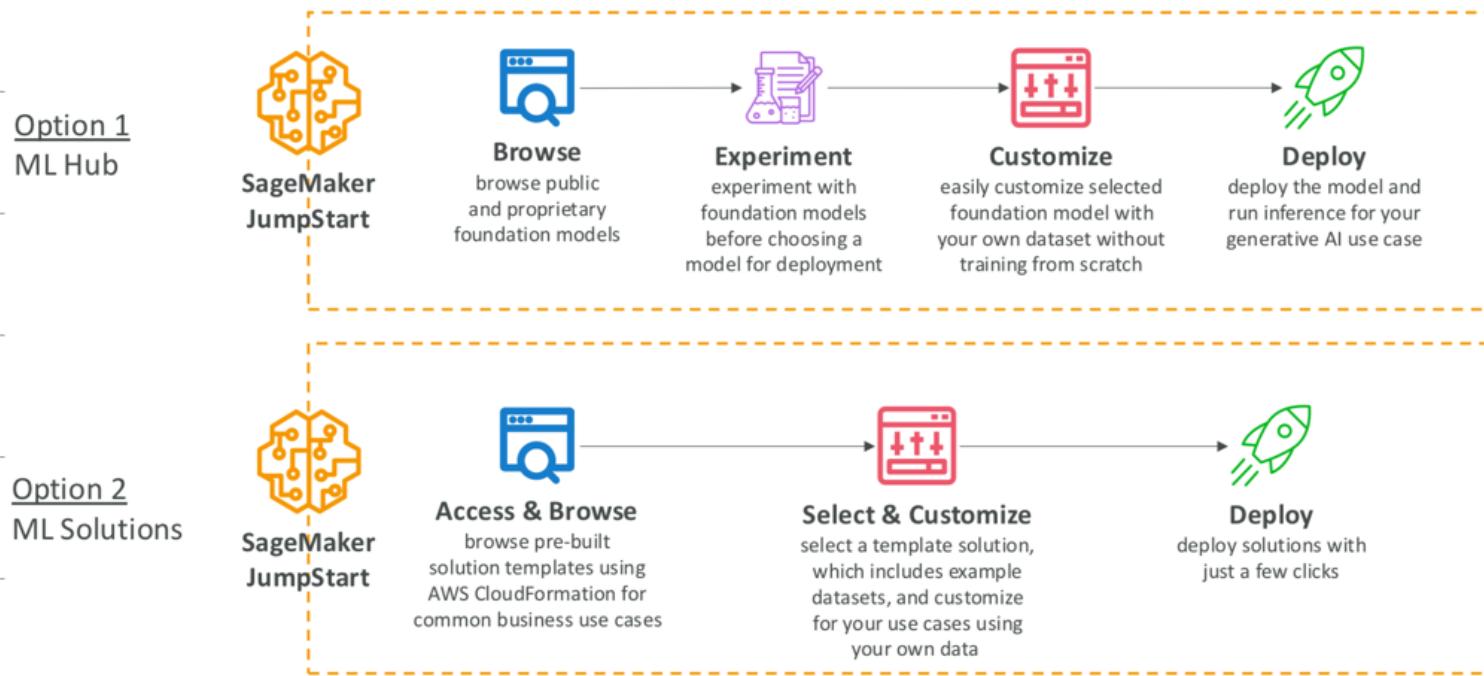
- ↳ Sage Maker Model Cards contains all the essential model information.
- ↳ Sage Maker Model Dashboard is a centralized repository about the information and insights for all the models.
- ↳ Sage Maker Role manager gives permissions and define roles for persons like the data scientists or MLops engineers.

U

- ↳ Sage Maker Model Monitor monitors the quality of your model in production either continuously or on-schedule and gives alerts for deviations.
- ↳ Sage Maker Model Registry is a centralized repository to track, manage and version ML models and access associated metadata. You can also manage approval status of a model, automate model deployment and share the models.
- ↳ SageMaker Pipelines is a workflow that automates the process of building, training and deploying a ML model like CI/CD. The supported Step types are:
  1. Processing
  2. Training
  3. Tuning
  4. AutoML (automatically train a model)
  5. Model (create/register SM model)
  6. ClarifyCheck (data bias, model bias, explainability)
  7. QualityCheck (Data Quality, Model Quality).

## Sage Maker JumpStart:

- ↳ ML hub to find pre-trained FM, CV models or NLP models.
- ↳ They can be customized for our data/usecase and deploy on SM directly.
- ↳ You also have prebuilt models for demand forecast, credit rate prediction, fraud detection and computer vision.



## - SageMaker Canvas:

- ↳ Build ML models using a visual interface or access ready-to-use models.
- ↳ They are powered by SM Autopilot which uses AutoML
- ↳ Leverage Data Wrangler for data preparation.

## - MLflow on Amazon SageMaker:

- ↳ open-source tool that helps to manage entire ML lifecycle.
- ↳ It has tracking servers for runs and experiments.

## - Summary:

- **SageMaker:** end-to-end ML service
- **SageMaker Automatic Model Tuning:** tune hyperparameters
- **SageMaker Deployment & Inference:** real-time, serverless, batch, async

- **SageMaker Studio:** unified interface for SageMaker
  - **SageMaker Data Wrangler:** explore and prepare datasets, create features
  - **SageMaker Feature Store:** store features metadata in a central place
  - **SageMaker Clarify:** compare models, explain model outputs, detect bias
  - **SageMaker Ground Truth:** RLHF, humans for model grading and data labeling
- 
- **SageMaker Model Cards:** ML model documentation
  - **SageMaker Model Dashboard:** view all your models in one place
  - **SageMaker Model Monitor:** monitoring and alerts for your model
  - **SageMaker Model Registry:** centralized repository to manage ML model versions
  - **SageMaker Pipelines:** CICD for Machine Learning
  - **SageMaker Role Manager:** access control
  - **SageMaker JumpStart:** ML model hub & pre-built ML solutions
  - **SageMaker Canvas:** no-code interface for SageMaker
  - **MLFlow on SageMaker:** use MLFlow tracking servers on AWS