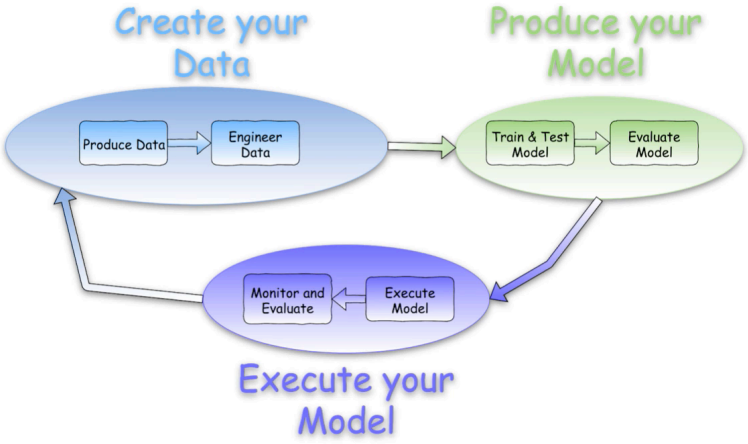
Whizlabs - ML Specialty Exam Course - Implementation and Operations

https://www.whizlabs.com/learn/course/aws-mls-practice-tests/video/3510

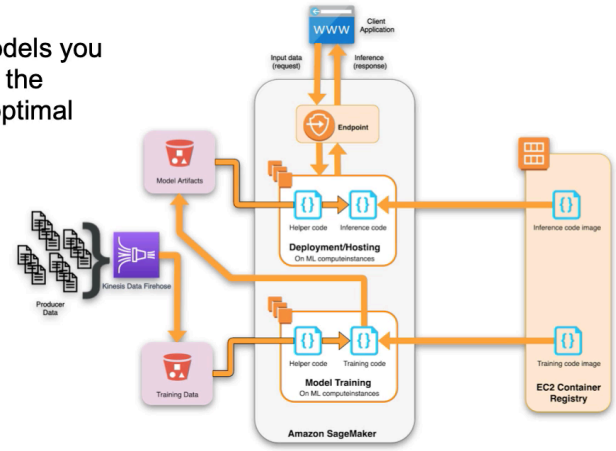
AWS Machine Learning - Introduction - Machine Learning Implementation and Operations

The Machine Learning Cycle - Implementing and Operating the Model



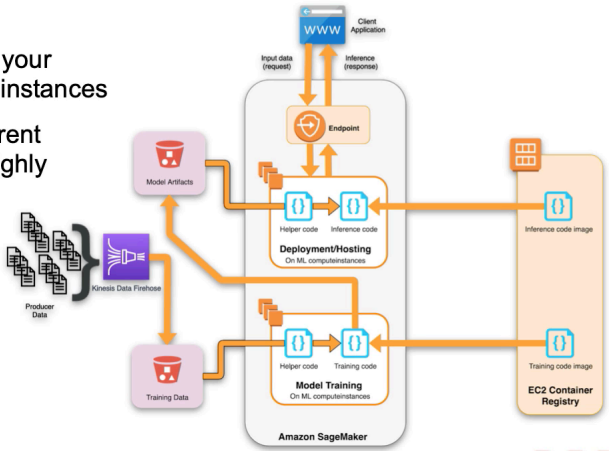
Machine Learning Implementation and Operation - Performance

- To optimize the performance of your models you can use Automatic Model Tuning to find the hyperparameters that will give you the optimal performance
- You can use SageMaker hosting to automatically scale your model to the performance needed for your model



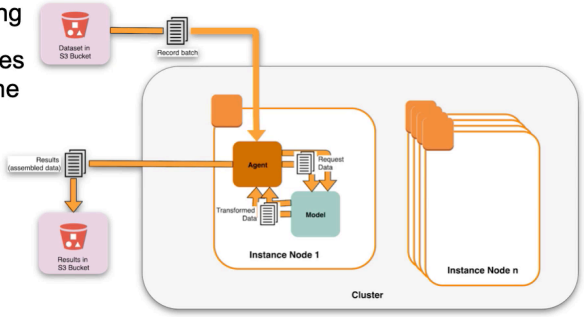
Machine Learning Implementation and Operation - Availability

- SageMaker hosting allows you to configure your endpoints to elastically scale your endpoint instances
- Use two or more endpoint instances in different availability zones to make your endpoints highly available



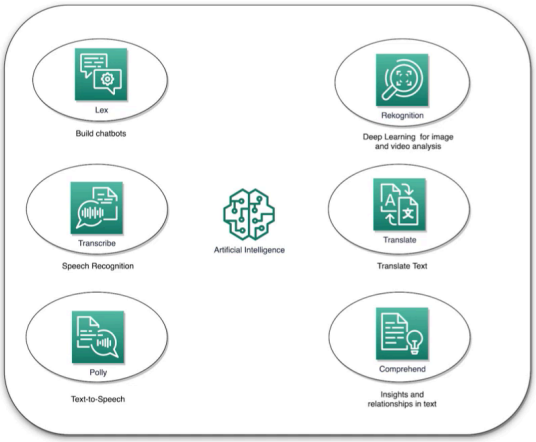
Machine Learning Implementation and Operation - Scalability

- SageMaker hosting automatically scales your endpoint instances to the performance needed for your application through Application Auto Scaling
- You can manually adjust the number of instances and the instance type without suffering downtime by modifying your endpoint configuration
- SageMaker scales to large numbers of transactions per second. The actual scale depends on your deployed model and the number and type of instances your model endpoints run on



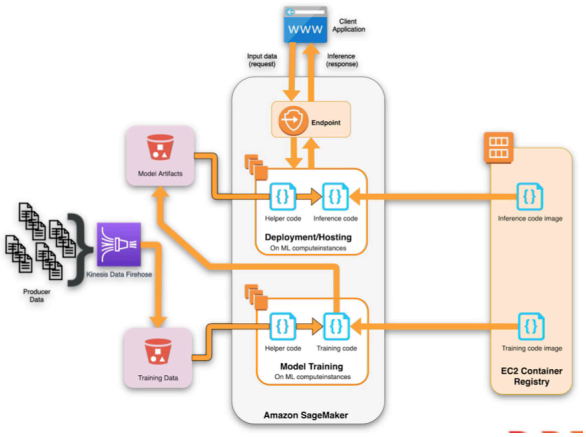
Machine Learning Implementation and Operation - Recommend Appropriate ML Services

- Select from several AWS Services to implement the appropriate Machine Learning solution for your business problem



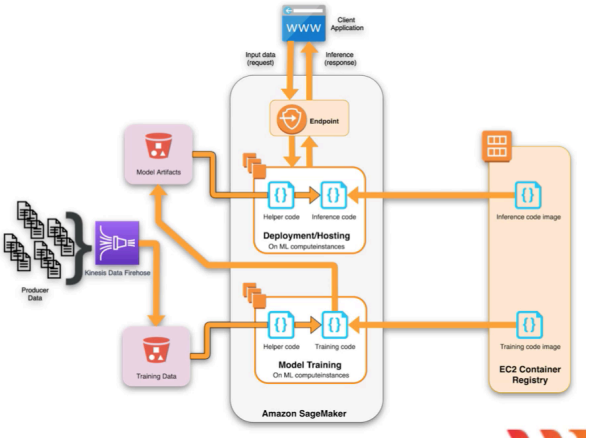
Machine Learning Implementation and Operation - Security

- SageMaker model artifacts and other system artifacts are encrypted in transit and at rest
- Requests to your endpoint API and can be made over a secure (SSL) connection
- Assign IAM roles to your model instance to provide permission to access resources on your behalf for deployment
- Use encrypted S3 buckets for model artifacts and data



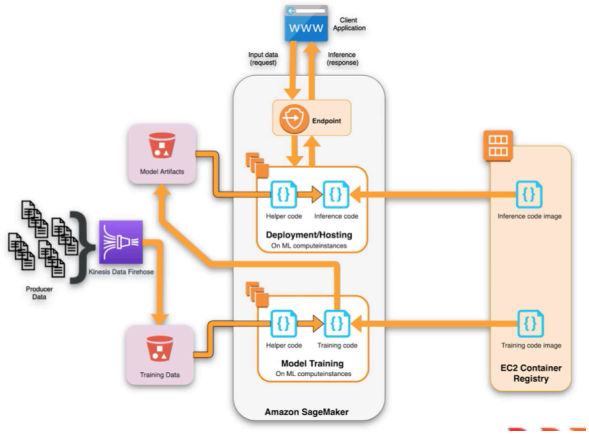
Machine Learning Implementation and Operation - Security

- SageMaker model artifacts and other system artifacts are encrypted in transit and at rest
- Requests to your endpoint API and can be made over a secure (SSL) connection
- Assign IAM roles to your model instance to provide permission to access resources on your behalf for deployment
- Use encrypted S3 buckets for model artifacts and data
- Pass a KMS key to your endpoints to encrypt their attached ML storage volume



Machine Learning Implementation and Operation - Deploy/Operationalize the Model

- Clients send HTTPS requests to your endpoint to obtain inferences
- Can deploy multiple variants of a model to the same HTTPS endpoint (test variations of your model in production)
- Can also configure a production variant to achieve auto scaling
- You can modify your endpoints without suffering downtime by modifying your endpoint configuration

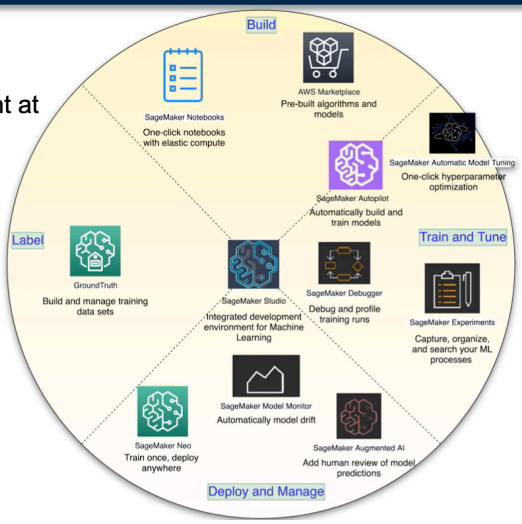


Implementing and Operating the model

AWS Machine Learning - Build machine learning solutions for performance, availability, scalability, resiliency, and fault tolerance

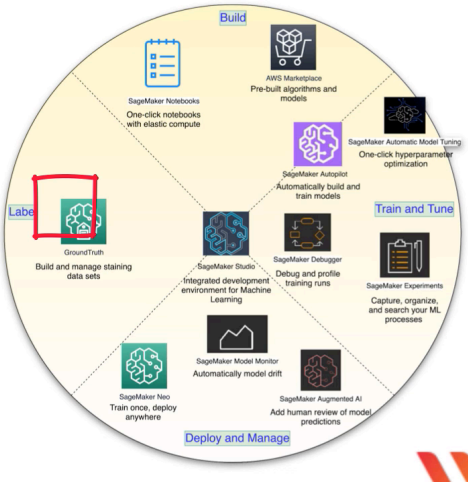
Machine Learning Implementation and Operation

- SageMaker makes it simple to develop high quality models by enabling quick deployment at scale
- SageMaker provides the tools needed for machine learning in a single toolset that allows you to get models to production faster with much less effort and at lower cost than with the traditional tool suite



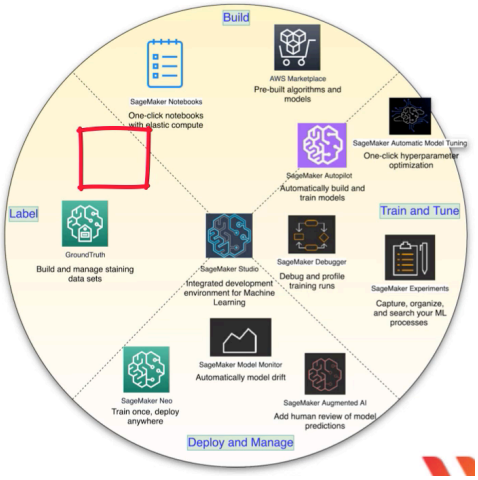
Machine Learning Implementation and Operation - Build with SageMaker Studio

- Single, web-based visual interface where you can perform all your machine learning development steps
  - Build, train, and deploy models
  - Upload data, create new notebooks, train and tune models, move back and forth between steps to adjust experiments, compare results, and deploy models to production



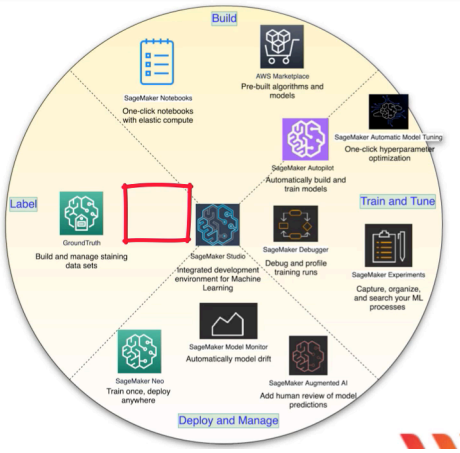
Machine Learning Implementation and Operation - Performance with SageMaker Autopilot

- Build, train, and tune models with complete visibility and control
  - Inspects input data, applies feature engineering, picks optimum set of algorithms, trains and tunes multiple models, tracks model's performance, and ranks your models based on their performance
- Gives you the best model for your problem at hand while saving time
- Can be used by developers without extensive machine learning experience



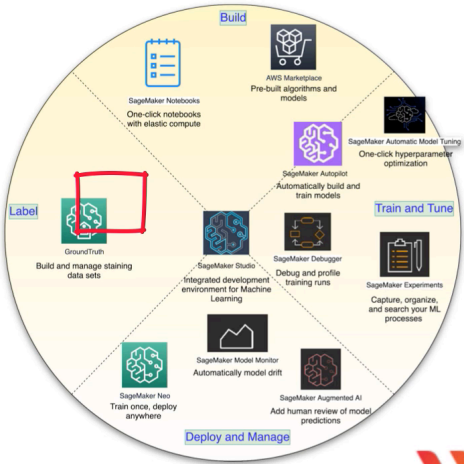
Machine Learning Implementation and Operation - Evaluate with SageMaker Experiments

- Organize your artifacts, track metrics, and evaluate training runs using SageMaker Experiments
  - Manage your model iterations by capturing the input parameters, configurations, and results, and saving them as experiments



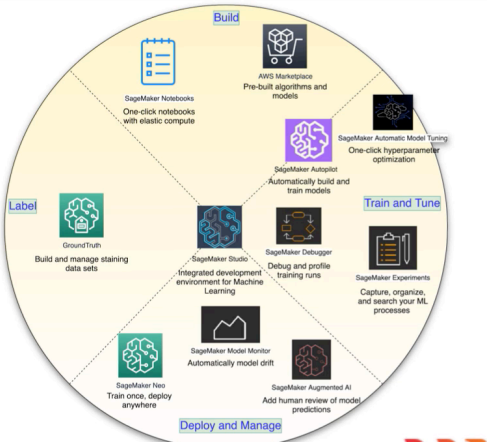
Machine Learning Implementation and Operation - Fault Tolerance with SageMaker Debugger

- Analyze, identify, and alert problems for machine learning
  - Track real-time metrics when training such as validation, confusion matrices, and learning gradients to increase model accuracy
  - Generate warnings and corrective action advice when you experience training issues



Machine Learning Implementation and Operation - Resiliency with SageMaker Monitor

- Detect and remediate concept drift where the patterns used to train the model have changed over time
  - Identifies concept drift in deployed models and gives detailed notifications that help identify the source of the drift

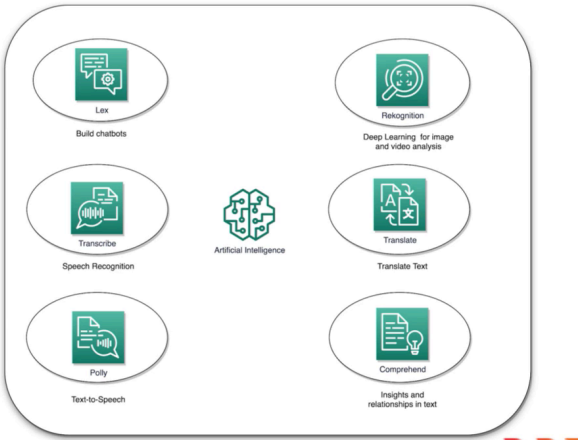


ML Services

AWS Machine Learning - Recommend and implement the appropriate machine learning services and features for a given problem

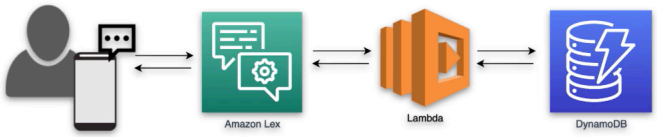
Machine Learning Services - Select the appropriate service for your problem

- Select from several AWS Services to implement the appropriate Machine Learning solution for your business problem



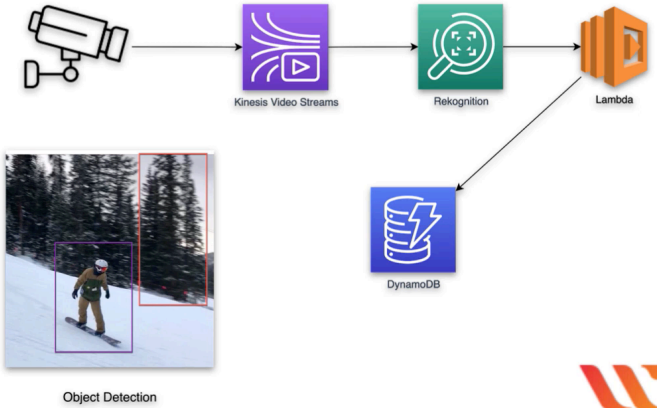
Machine Learning Services - Lex

- Build conversational interfaces into your applications using voice and text
- ML using deep learning algorithms of automatic speech recognition (ASR) for converting speech to text, and natural language understanding (NLU) to recognize the intent of the text
- Use the same deep learning algorithms used in Alexa to build natural language conversational chatbots
- Fully managed service that scales automatically



Machine Learning Services - Rekognition

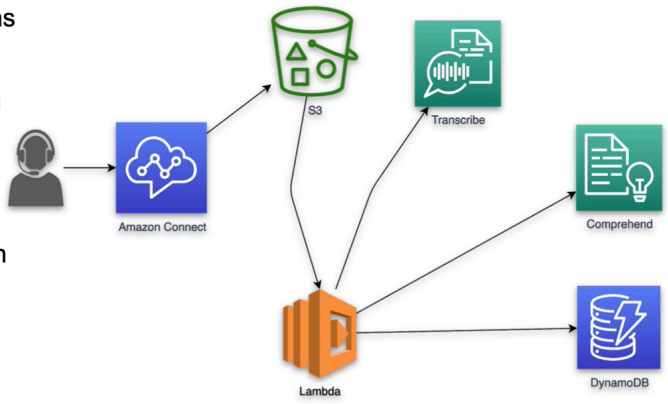
- Add video and image analysis to your applications
- ML using deep learning algorithms to identify objects, people, text, scenes, activities, and inappropriate content
- Facial recognition and facial search for verification, people counting, celebrity identification, public safety
- Rekognition Custom Labels feature automatically identifies and labels in your video or images



Object Detection

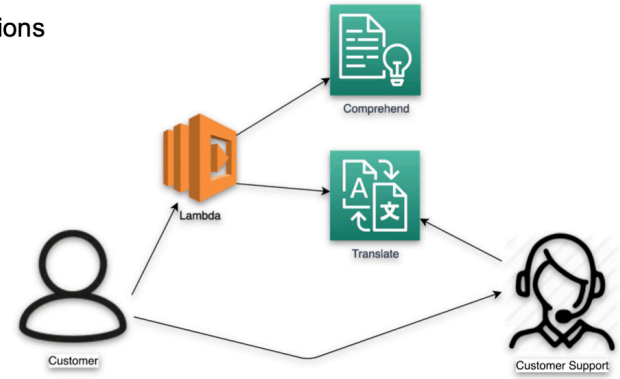
Machine Learning Services - Transcribe

- Add speech-to-text to your applications
- ML using deep learning algorithms such as automatic speech recognition (ASR) to convert speech to text
- Can process speech in batch or near real time in a streaming application
- Can automatically redact content such as personally identifiable information (PII)



Machine Learning Services - Translate

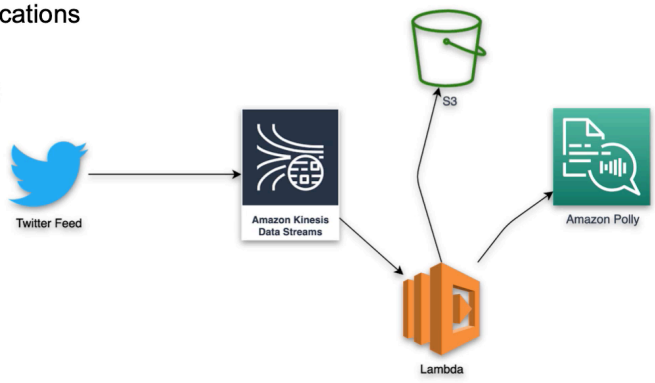
- Add language translation to your applications
- Neural machine translation service for language translation that uses deep learning models to produce accurate, natural sounding translation
- Localize content
- On-demand translation
- Continually learning from expanding datasets



Comprehend will detect the language, then the lambda can call translate

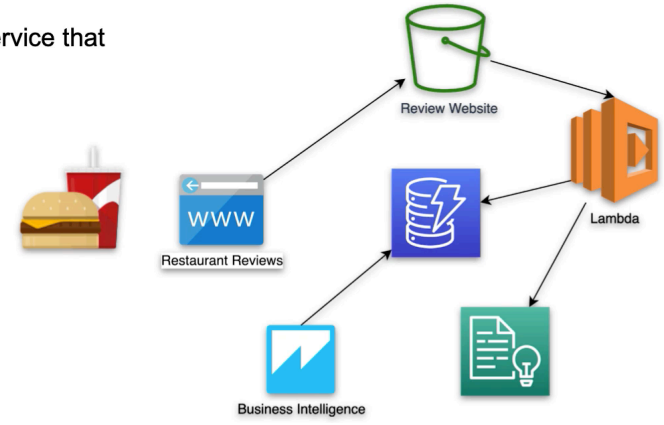
Machine Learning Services - Polly

- Add text-to-speech (TTS) to your applications
- Uses deep learning algorithms to synthesize human speech that sounds natural
- Has many voices/languages from which to choose
- Also has Neural Text-to-Speech (NTTS) voices which support newscaster and conversational styles
- Can create a brand voice for your organization, a customized NTTS voice



Machine Learning Services - Comprehend

- Natural language processing (NLP) service that finds insights and relationships in text
- Identifies language, extracts key phrases, places, people, events
- Understands positive or negative sentiment of the text
- Tokenizes text and parts of speech and organizes text files by topic
- Can extend to recognize your own organization's vocabulary



AWS Security Practices
SM provides EC2 based instances - run them as dedicated EC2 instances for our use. They are NOT shared.
We can map them to our VPC environment, allowing us to have control over the access to our resources
- network level control such as VPC Endpointor Security Group => control access to notebooks, training jobs, hosted ML models
The way it is done, we create an ENI (Elastic Network Interface) in our VPC - we attach it to the SM Service account where the instance runs.

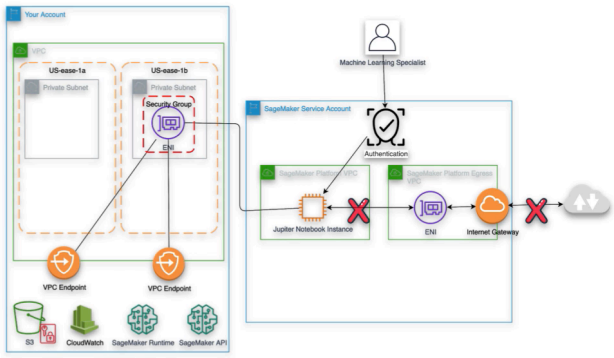
=> For Jupiter notebooks, this means that access to everything is controlled by us and our network config
No internet access to our notebooks

We can only access the SM resources from within our VPC, using our VPC Endpoint => Private connectivity

AWS Machine Learning - Apply Basic AWS Security Practices to Machine Learning Solutions

Machine Learning Services - Secure your SageMaker Instances

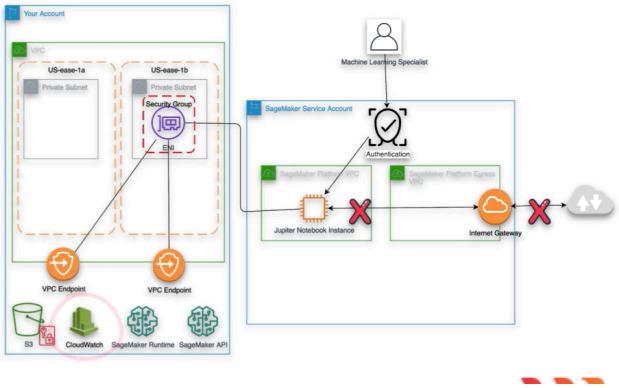
- Secure your jupyter notebook instances
  - Your SageMaker infrastructure uses EC2 instances dedicated for your use
  - Can map your SageMaker resources to VPC so you can use your network controls
  - Control access to your jupyter notebooks and your hosted models through IAM
  - Can only access your SageMaker resources from within your VPC using your VPC Endpoints (Private Connectivity)



We can also encrypt our data at rest and in transit
- STS encryption from S3

Machine Learning Services - Secure your SageMaker Instances

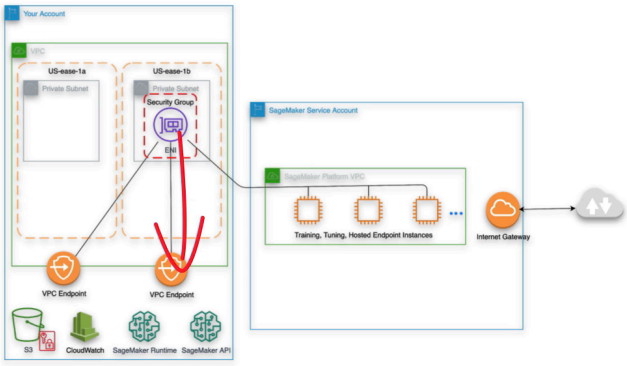
- Secure your jupyter notebook instances
  - Encrypt your data at rest and in flight from your datasets on S3 to your notebooks and through to your hosted endpoints
  - You can use lifecycle configurations to harden the OS of your SageMaker EC2 instances or install agents
  - SageMaker is integrated with CloudWatch and CloudTrail for logging training job and hosted model activity as well as API calls



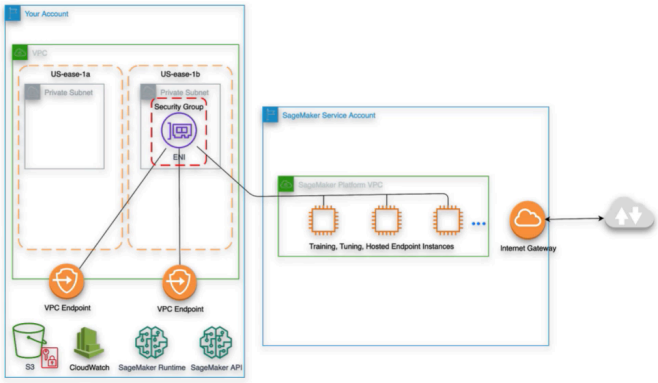
We can apply the same approach to our training/tuning jobs and hosted model instances.
They don't have access to internet
Access all resources they need through our => ENI => VPC Endpoint, to finally access the S3 bucket where the data is or our models get stored

Machine Learning Services - Secure your SageMaker Instances

- Secure your training job, tuning job, and your hosted model endpoint instances

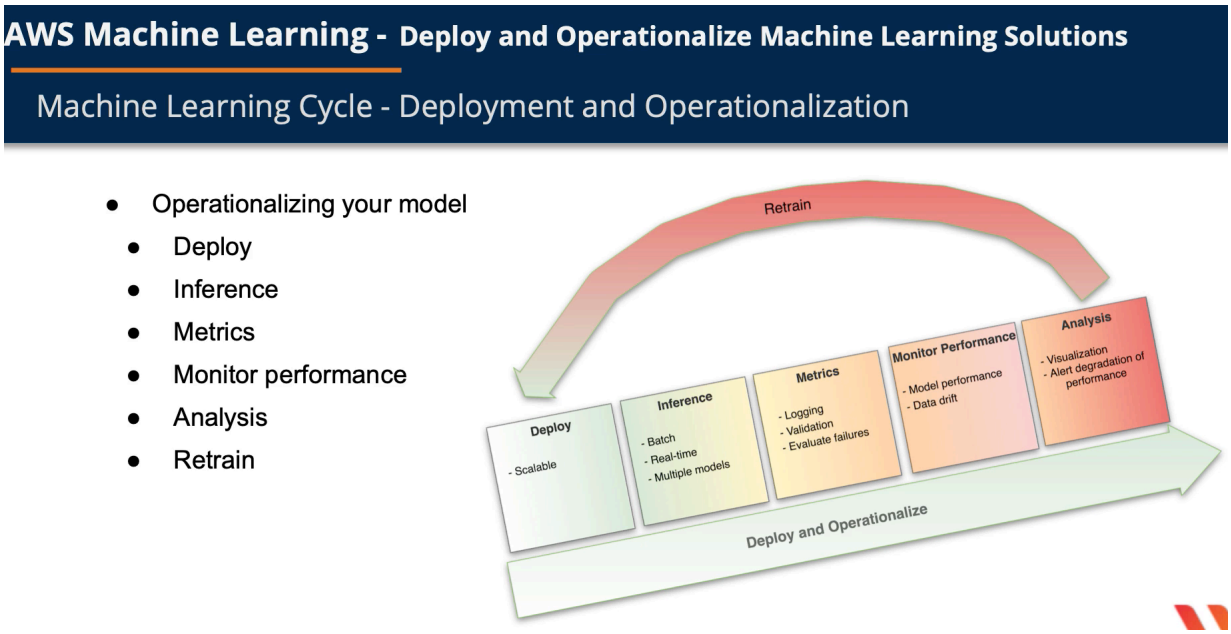


- Secure your training job, tuning job, and your hosted model endpoint instances
  - Can map your training, tuning, and hosted model endpoint instances to VPC so you can use your network controls
  - Can restrict your training, tuning, and endpoint instances to resources within your VPC using your VPC Endpoints (Private Connectivity)



Deploy and Operationalize ML solutions

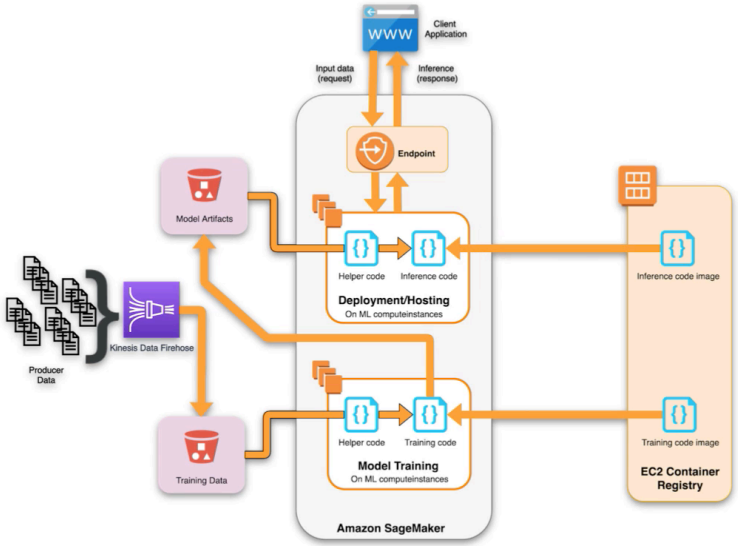
5 steps, and re-train over and over again



- Operationalizing your model
  - Deploy
  - Inference
  - Metrics
  - Monitor performance
  - Analysis
  - Retrain

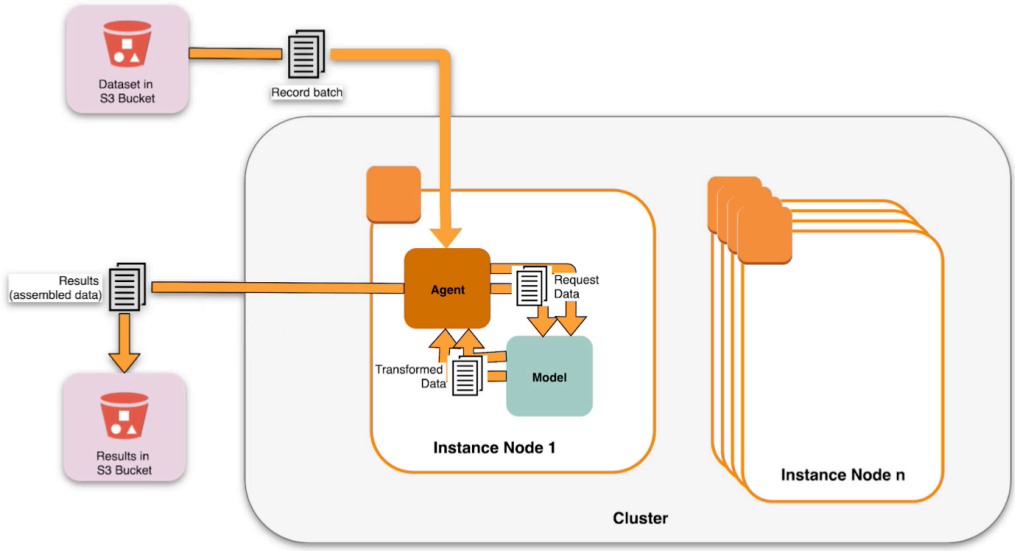
Real Time Endpoint

Machine Learning Cycle - Deployment and Operationalization - Real-Time



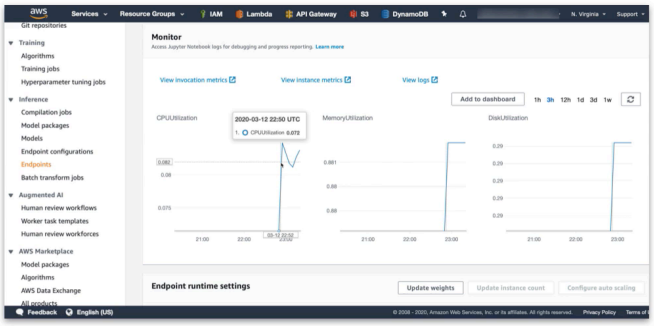
Batch Transform

Machine Learning Cycle - Deployment and Operationalization - Batch

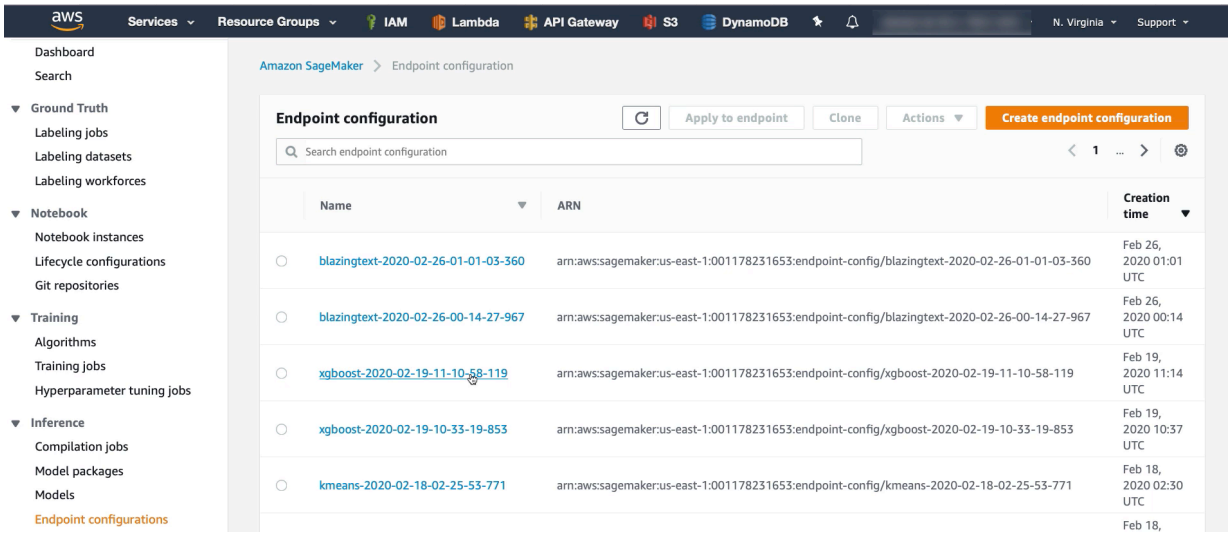


Machine Learning Cycle - Model Monitor

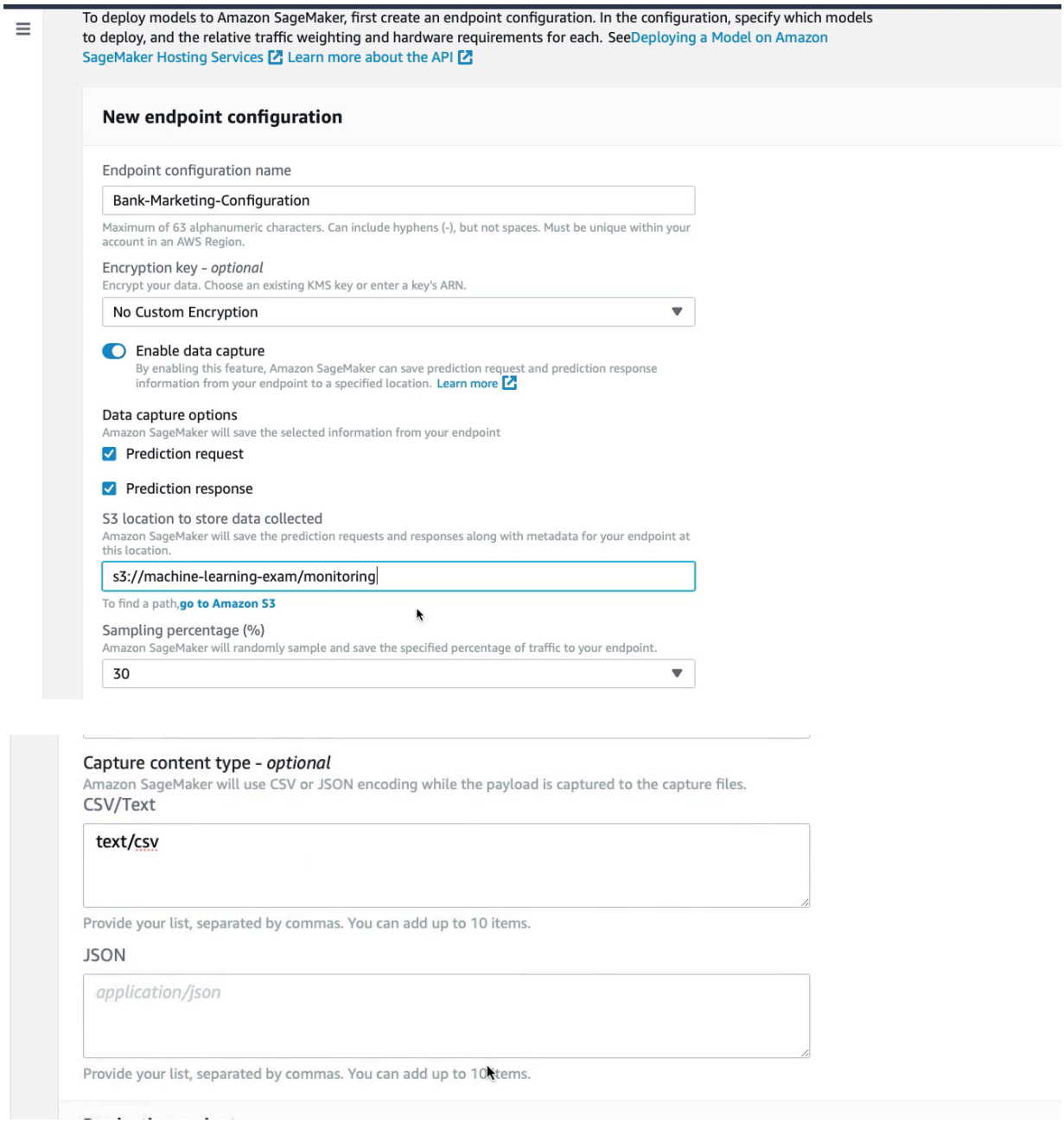
- Monitor your production model to detect deviations in data quality compared to a baseline dataset
  - Monitor your endpoint using Model Monitor (csv or flat-json)
    1. Baseline
    2. SageMaker suggests baseline constraints
    3. Create baselining job
    4. Create a continuous monitoring schedule
    5. Start continuous monitoring
- Analyze/Retrain



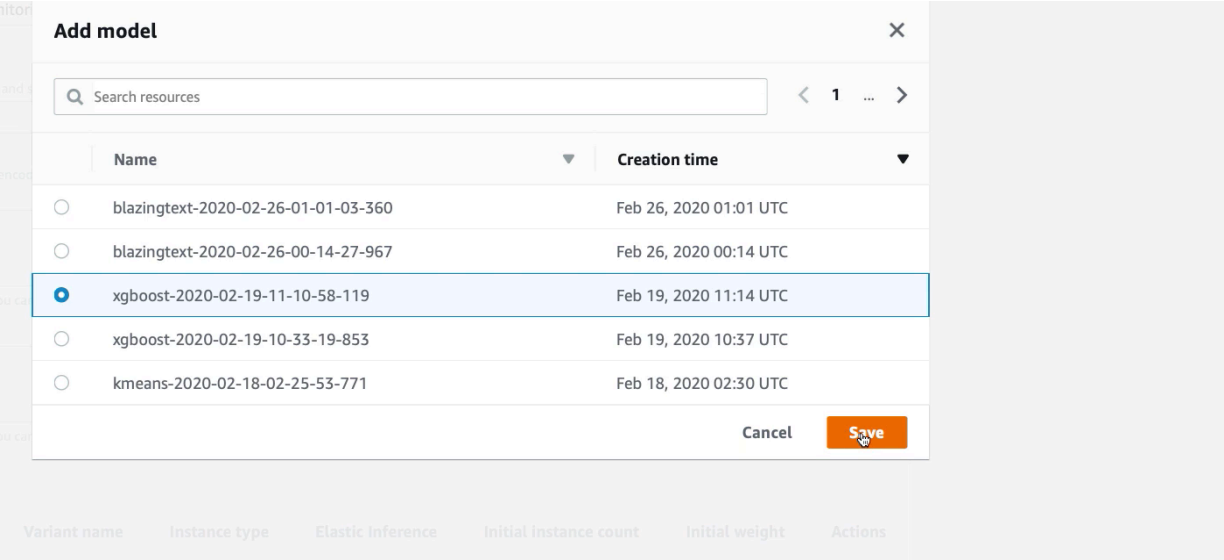
Let’s create an end point configuration to monitor it

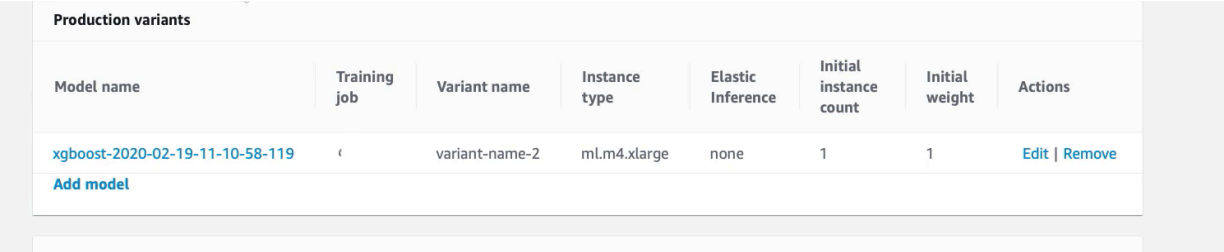


Enable data capture

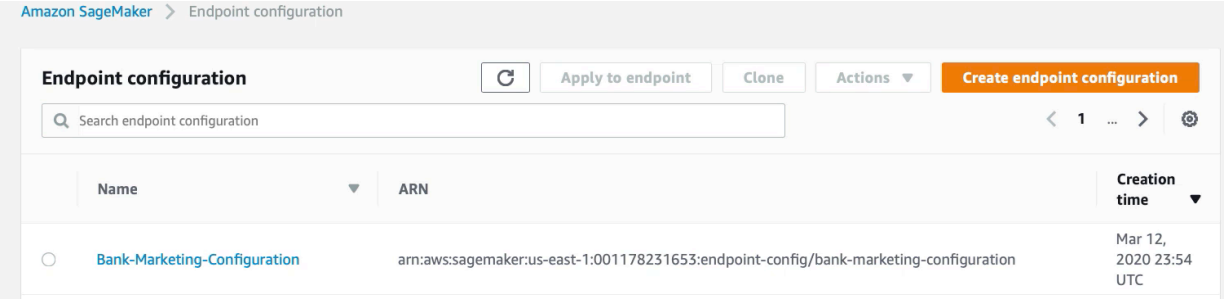


Production Variants - add a 2nd model

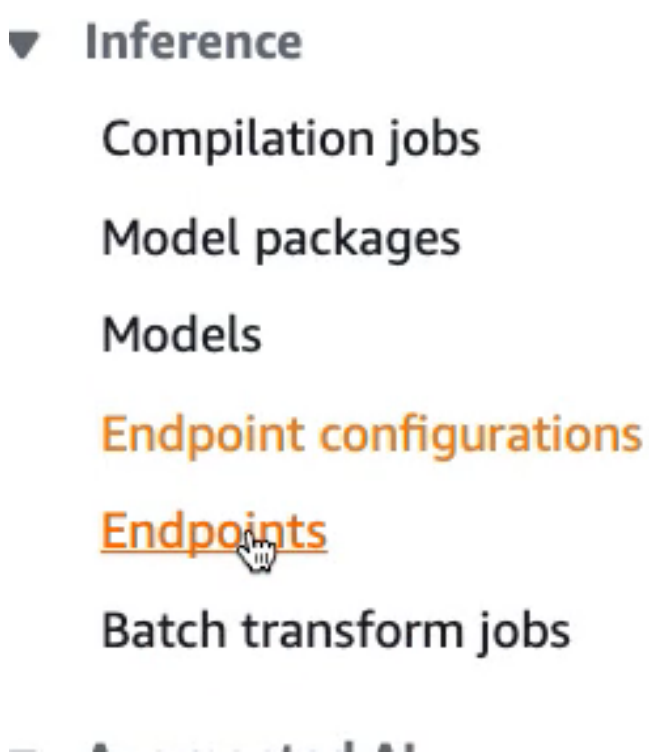


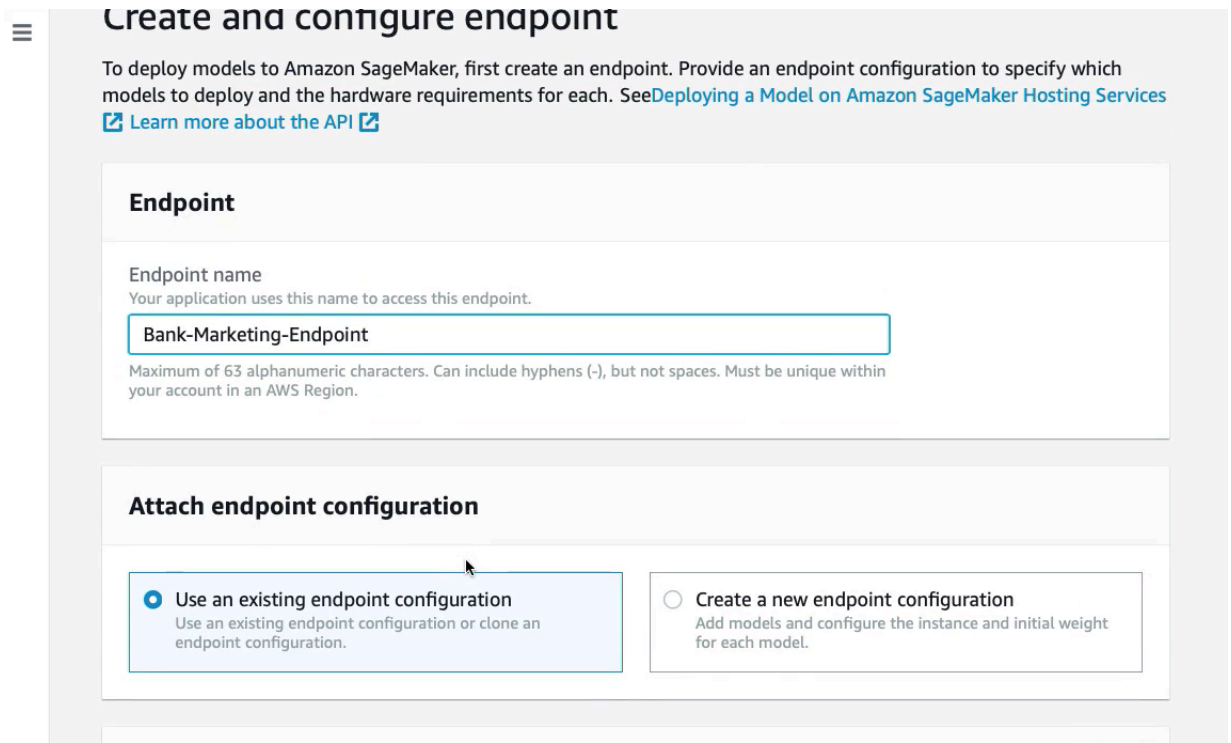


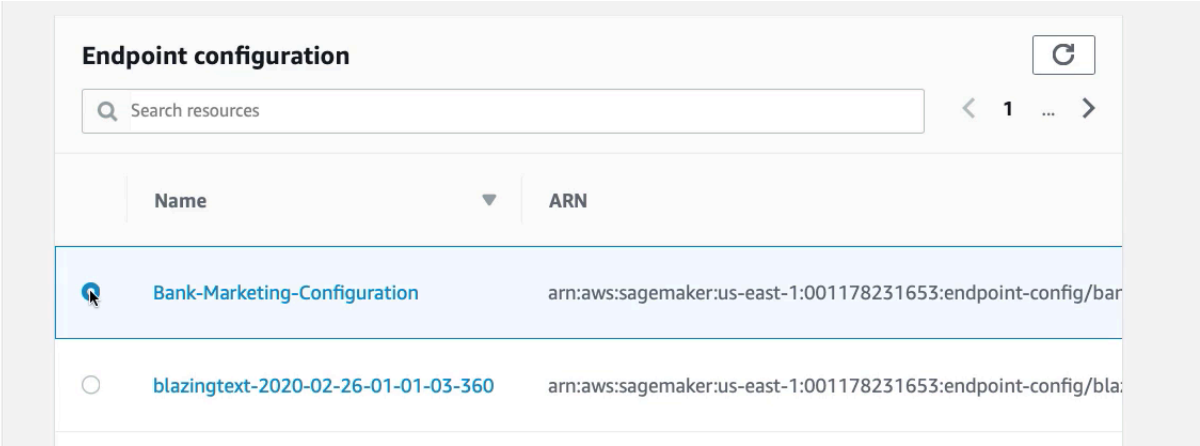
It’s now created:



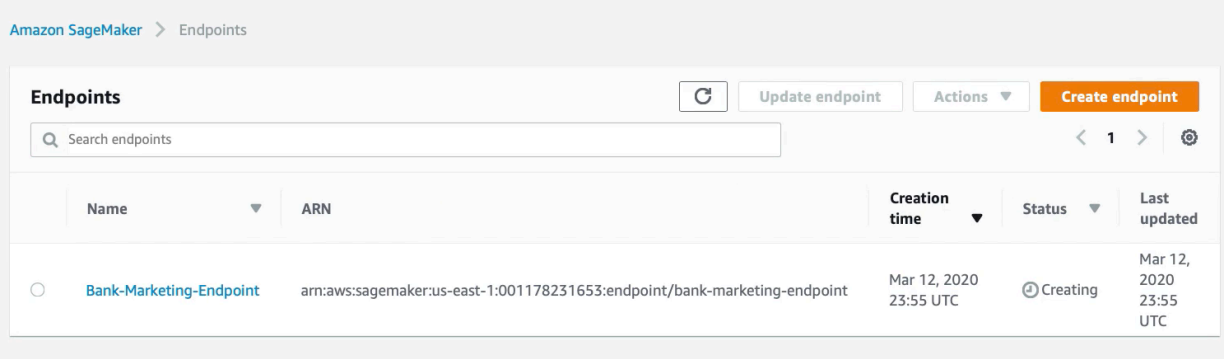
Create an Endpoint now

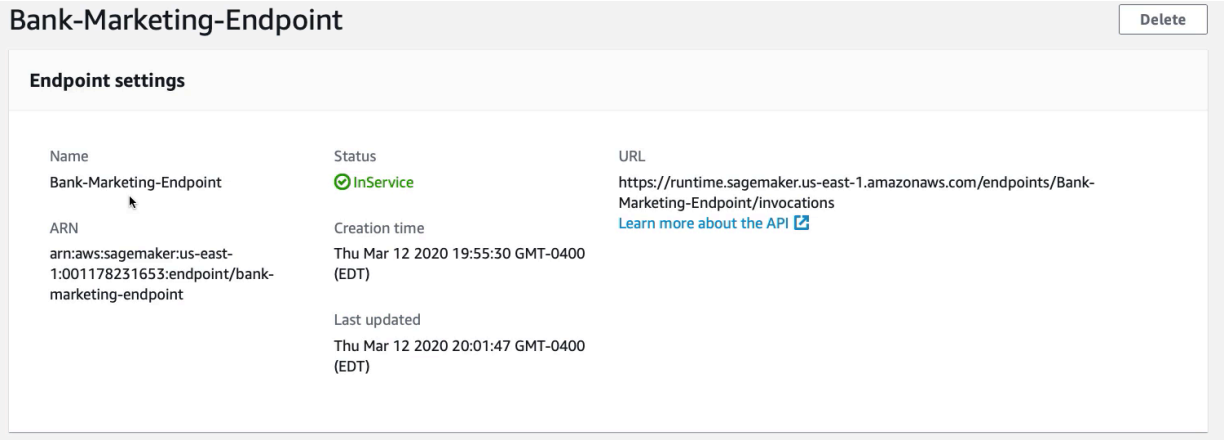

Inference
Compilation jobs
Model packages
Models
Endpoint configurations
Endpoints
Batch transform jobs


Create and configure endpoint
To deploy models to Amazon SageMaker, first create an endpoint. Provide an endpoint configuration to specify which models to deploy and the hardware requirements for each. SeeDeploying a Model on Amazon SageMaker Hosting Services
Learn more about the API
Endpoint
Endpoint name
Your application uses this name to access this endpoint.
Bank-Marketing-Endpoint
Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.
Attach endpoint configuration
Use an existing endpoint configuration
Use an existing endpoint configuration or clone an endpoint configuration.
Create a new endpoint configuration
Add models and configure the instance and initial weight for each model.



Endpoint has been created





To invoke the endpoint, use POST

Request Syntax

```
POST /endpoints/EndpointName/invocations HTTP/1.1
Content-Type: ContentType
Accept: Accept
X-Amzn-SageMaker-Custom-Attributes: CustomAttributes
X-Amzn-SageMaker-Target-Model: TargetModel

Body
```

Data capture settings

Enable data capture	Current sampling percentage (%)	S3 location to store data collected
Yes	30	s3://machine-learning-exam/monitoring

Data capture status
Started

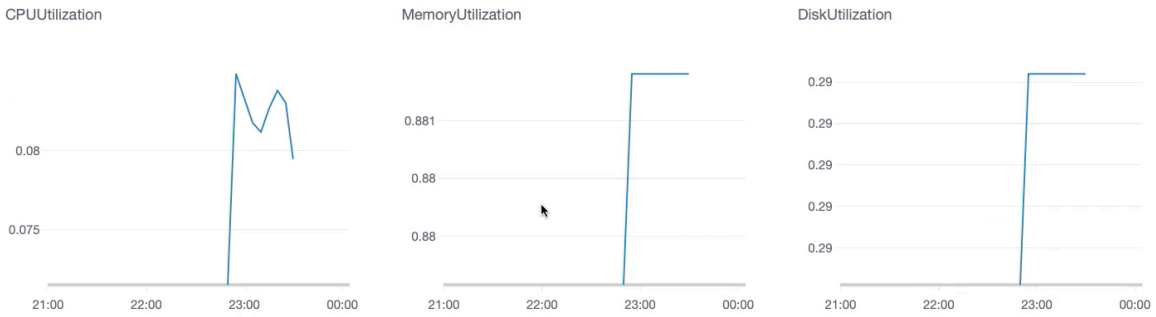
Monitoring metrics

Monitor

Access Jupyter Notebook logs for debugging and progress reporting. Learn more

View invocation metrics | View instance metrics | View logs

Add to dashboard | 1h 3h 12h 1d 3d 1w

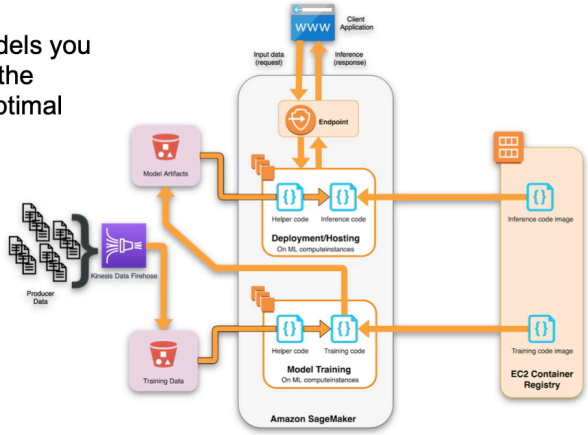


Exam tips

AWS Machine Learning - Exam Tips - Machine Learning Implementation and Operations

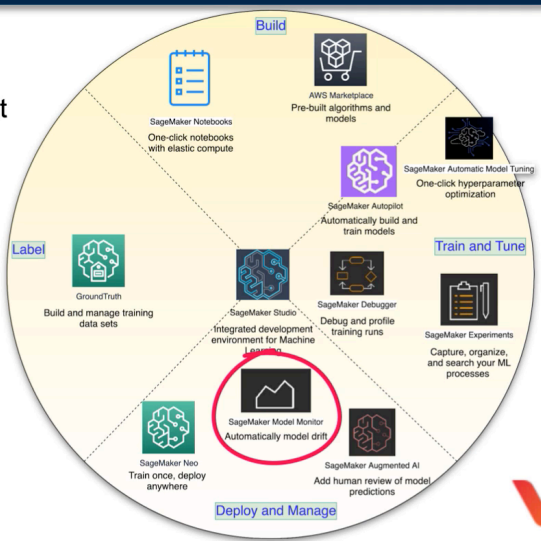
Machine Learning Implementation and Operation - Performance

- To optimize the performance of your models you can use Automatic Model Tuning to find the hyperparameters that will give you the optimal performance
- You can use SageMaker hosting to automatically scale your model to the performance needed for your model



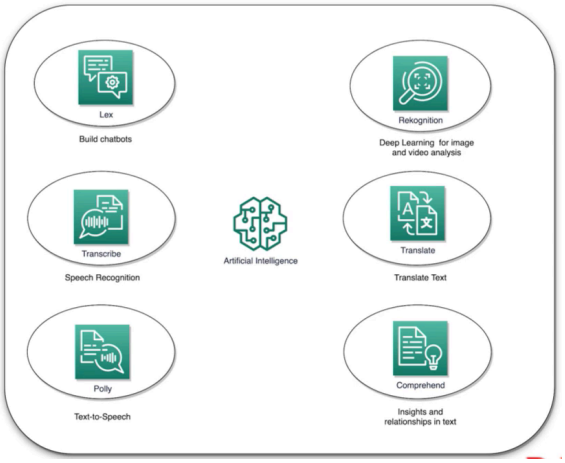
Machine Learning Implementation and Operation

- SageMaker makes it simple to develop high quality models by enabling quick deployment at scale
- SageMaker provides the tools needed for machine learning in a single toolset
  - SageMaker Studio
  - SageMaker Autopilot
  - SageMaker Experiments
  - SageMaker Debugger
  - SageMaker Model Monitor



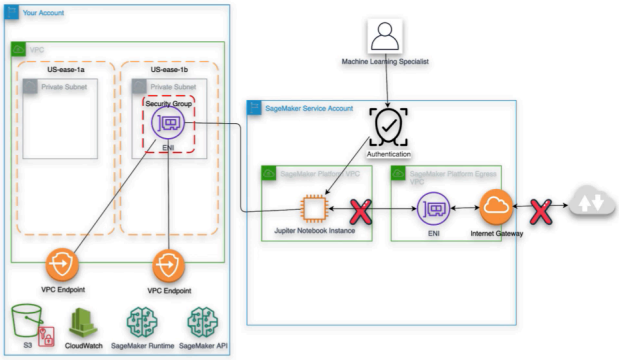
Machine Learning Services - Select the appropriate service for your problem

- Select from several AWS Services to implement the appropriate Machine Learning solution for your business problem



Machine Learning Services - Secure your SageMaker Instances

- Secure your jupyter notebook instances
  - Your SageMaker infrastructure uses EC2 instances dedicated for your use
  - Can map your SageMaker resources to VPC so you can use your network controls
  - Control access to your jupyter notebooks and your hosted models through IAM
  - Can only access your SageMaker resources from within your VPC using your VPC Endpoints (Private Connectivity)



Exam Day tips

AWS Machine Learning - Exam Day Tips

Arrive prepared and relaxed

- Use an outline of all of the material covered in this course
  - Study the material continuously the week leading up to your exam date
- Take practice exams
- Arrive at the exam location at least 15 minutes before your start time
- Leave your phone, watch and any other electronics in your car
- Make sure you are relaxed

Only One Step Away!