What is Amazon SageMaker?,"Amazon SageMaker is a fully managed service to prepare data and build, train, and deploy machine learning (ML) models for any use case with fully managed infrastructure, tools, and workflows."
"In which Regions is Amazon SageMaker available?
","For a list of the supported Amazon SageMaker AWS Regions, please visit the AWS Regional Services page. Also, for more information, see Regional endpoints in the AWS general reference guide."
"What is the service availability of Amazon SageMaker?
","Amazon SageMaker is designed for high availability. There are no maintenance windows or scheduled downtimes. SageMaker APIs run in Amazon’s proven, high-availability data centers, with service stack replication configured across three facilities in each AWS Region to provide fault tolerance in the event of a server failure or Availability Zone outage."
How does Amazon SageMaker secure my code?,"Amazon SageMaker stores code in ML storage volumes, secured by security groups and optionally encrypted at rest."
What security measures does Amazon SageMaker have?,"Amazon SageMaker ensures that ML model artifacts and other system artifacts are encrypted in transit and at rest. Requests to the SageMaker API and console are made over a secure (SSL) connection. You pass AWS Identity and Access Management roles to SageMaker to provide permissions to access resources on your behalf for training and deployment. You can use encrypted Amazon Simple Storage Service (Amazon S3) buckets for model artifacts and data, as well as pass an AWS Key Management Service (KMS) key to SageMaker notebooks, training jobs, and endpoints, to encrypt the attached ML storage volume. Amazon SageMaker also supports Amazon Virtual Private Cloud (VPC) and AWS PrivateLink support."
"Does Amazon SageMaker use or share models, training data, or algorithms?","Amazon SageMaker does not use or share customer models, training data, or algorithms. We know that customers care deeply about privacy and data security. That's why AWS gives you ownership and control over your content through simple, powerful tools that allow you to determine where your content will be stored, secure your content in transit and at rest, and manage your access to AWS services and resources for your users. We also implement responsible and sophisticated technical and physical controls that are designed to prevent unauthorized access to or disclosure of your content. As a customer, you maintain ownership of your content, and you select which AWS services can process, store, and host your content. We do not access your content for any purpose without your consent."
"How am I charged for Amazon SageMaker?
","You pay for ML compute, storage, and data processing resources you use for hosting the notebook, training the model, performing predictions, and logging the outputs. Amazon SageMaker allows you to select the number and type of instance used for the hosted notebook, training, and model hosting. You pay only for what you use, as you use it; there are no minimum fees and no upfront commitments. See

the Amazon SageMaker pricing page and the Amazon SageMaker Pricing calculator for details."
"How can I optimize my Amazon SageMaker costs, such as detecting and stopping idle resources in order to avoid unnecessary charges?","There are several best practices you can adopt to optimize your Amazon SageMaker resource utilization. Some approaches involve configuration optimizations; others involve programmatic solutions. A full guide on this concept, complete with visual tutorials and code samples, can be found in this blog post."
"What if I have my own notebook, training, or hosting environment?","Amazon SageMaker provides a full end-to-end workflow, but you can continue to use your existing tools with SageMaker. You can easily transfer the results of each stage in and out of SageMaker as your business requirements dictate."
Is R supported with Amazon SageMaker?,"Yes, R is supported with Amazon SageMaker. You can use R within SageMaker notebook instances, which include a preinstalled R kernel and the reticulate library. Reticulate offers an R interface for the Amazon SageMaker Python SDK, enabling ML practitioners to build, train, tune, and deploy R models.
"
How can I check for imbalances in my model?,"Amazon SageMaker Clarify helps improve model transparency by detecting statistical bias across the entire ML workflow. SageMaker Clarify checks for imbalances during data preparation, after training, and ongoing over time, and also includes tools to help explain ML models and their predictions. Findings can be shared through explainability reports."
What kind of bias does Amazon SageMaker Clarify detect?,"Measuring bias in ML models is a first step to mitigating bias. Bias may be measured before training and after training, as well as for inference for a deployed model. Each measure of bias corresponds to a different notion of fairness. Even considering simple notions of fairness leads to many different measures applicable in various contexts. You need to choose bias notions and metrics that are valid for the application and the situation under investigation. SageMaker currently supports the computation of different bias metrics for training data (as part of SageMaker data preparation), for the trained model (as part of Amazon SageMaker Experiments), and for inference for a deployed model (as part of Amazon SageMaker Model Monitor). For example, before training, we provide metrics for checking whether the training data is representative (that is, whether one group is underrepresented) and whether there are differences in the label distribution across groups. After training or during deployment, metrics can be helpful to measure whether (and by how much) the performance of the model differs across groups. For example, start by comparing the error rates (how likely a model's prediction is to differ from the true label) or break further down into precision (how likely a positive prediction is to be correct) and recall (how likely the model will correctly label a positive example)."
How does Amazon SageMaker Clarify improve model explainability?,Amazon SageMaker Clarify is integrated with Amazon SageMaker Experiments to

provide a feature importance graph detailing the importance of each input for your model’s overall decision-making process after the model has been trained. These details can help determine if a particular model input has more influence than it should on overall model behavior. SageMaker Clarify also makes explanations for individual predictions available via an API.
What is Amazon SageMaker Studio?,"Amazon SageMaker Studio provides a single, web-based visual interface where you can perform all ML development steps. SageMaker Studio gives you complete access, control, and visibility into each step required to prepare data and build, train, and deploy models. You can quickly 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 all in one place, making you much more productive. All ML development activities including notebooks, experiment management, automatic model creation, debugging and profiling, and model drift detection can be performed within the unified SageMaker Studio visual interface."
What is RStudio on Amazon SageMaker?,"RStudio on Amazon SageMaker is the first fully managed RStudio Workbench in the cloud. You can quickly launch the familiar RStudio integrated development environment (IDE) and dial up and down the underlying compute resources without interrupting your work, making it easy to build machine learning (ML) and analytics solutions in R at scale. You can seamlessly switch between the RStudio IDE and Amazon SageMaker Studio notebooks for R and Python development. All your work, including code, datasets, repositories, and other artifacts, is automatically synchronized between the two environments to reduce context switch and boost productivity."
"How does Amazon SageMaker Studio pricing work?
",There is no additional charge for using Amazon SageMaker Studio. You pay only for the underlying compute and storage charges on the services you use within Amazon SageMaker Studio.
In which Regions is Amazon SageMaker Studio supported?,You can find the Regions where Amazon SageMaker Studio is supported in the documentation here.
What ML governance tools does Amazon SageMaker provide?,"Amazon SageMaker provides purpose-built ML governance tools across the ML lifecycle. With SageMaker Role Manager, administrators can define minimum permissions in minutes. SageMaker Model Cards makes it easier to capture, retrieve, and share essential model information from conception to deployment, and SageMaker Model Dashboard keeps you informed on production model behavior, all in one place. View more details."
What does Amazon SageMaker Role Manager do?,"You can define minimum permissions in minutes with Amazon SageMaker Role Manager. SageMaker Role Manager provides a baseline set of permissions for ML activities and personas with a catalog of pre-built IAM policies. You can keep the baseline permissions, or customize them further based on your specific needs. With a few self-guided prompts, you can quickly input

common governance constructs such as network access boundaries and encryption keys. SageMaker Role Manager will then generate the IAM policy automatically. You can discover the generated role and associated policies through the AWS IAM console. To further tailor the permissions to your use case, attach your managed IAM policies to the IAM role that you create with SageMaker Role Manager. You can also add tags to help identify the role and organize across AWS services."
What does Amazon SageMaker Model Cards do?,Amazon SageMaker Model Cards helps you centralize and standardize model documentation throughout the ML lifecycle by creating a single source of truth for model information. SageMaker Model Cards auto-populates training details to accelerate the documentation process. You can also add details such as the purpose of the model and the performance goals. You can attach model evaluation results to your model card and provide visualizations to gain key insights into model performance. SageMaker Model Cards can easily be shared with others by exporting to a pdf format.
What does Amazon SageMaker Model Dashboard do? ,"Amazon SageMaker Model Dashboard gives you a comprehensive overview of deployed models and endpoints, letting you track resources and model behavior violations through one pane. It allows you to monitor model behavior in four dimensions, including data and model quality, and bias and feature attribution drift through its integration with Amazon SageMaker Model Monitor and Amazon SageMaker Clarify. SageMaker Model Dashboard also provides an integrated experience to set up and receive alerts for missing and inactive model monitoring jobs, and deviations in model behavior for model quality, data quality, bias drift, and feature attribution drift. You can further inspect individual models and analyze factors impacting model performance over time. Then, you can follow up with ML practitioners to take corrective measures."
What is Amazon SageMaker Autopilot?,"Amazon SageMaker Autopilot is the industry’s first automated machine learning capability that gives you complete control and visibility into your ML models. SageMaker Autopilot automatically inspects raw data, applies feature processors, picks the best set of algorithms, trains and tunes multiple models, tracks their performance, and then ranks the models based on performance, all with just a few clicks. The result is the best-performing model that you can deploy at a fraction of the time normally required to train the model. You get full visibility into how the model was created and what’s in it, and SageMaker Autopilot integrates with Amazon SageMaker Studio. You can explore up to 50 different models generated by SageMaker Autopilot inside SageMaker Studio so it’s easy to pick the best model for your use case. SageMaker Autopilot can be used by people without ML experience to easily produce a model, or it can be used by experienced developers to quickly develop a baseline model on which teams can further iterate."
What built-in algorithms are supported in Amazon SageMaker Autopilot?,Amazon SageMaker Autopilot supports 2 built-in algorithms: XGBoost and Linear Learner.
Can I stop an Amazon SageMaker Autopilot job manually?,"Yes. You can

stop a job at any time. When an Amazon SageMaker Autopilot job is stopped, all ongoing trials will be stopped and no new trial will be started."
How do I get started with Amazon SageMaker quickly?,"Amazon SageMaker JumpStart helps you quickly and easily get started with ML. SageMaker JumpStart provides a set of solutions for the most common use cases that can be deployed readily with just a few clicks. The solutions are fully customizable and showcase the use of AWS CloudFormation templates and reference architectures so you can accelerate your ML journey. SageMaker JumpStart also supports one-click deployment and fine-tuning of more than 150 popular open-source models such as transformer, object detection, and image classification models."
Which open-source models are supported with Amazon SageMaker JumpStart?,"Amazon SageMaker JumpStart includes 150+ pre-trained open-source models from PyTorch Hub and TensorFlow Hub. For vision tasks such as image classification and object detection, you can use models such as ResNet, MobileNet, and Single-Shot Detector (SSD). For text tasks such as sentence classification, text classification, and question answering, you can use models such as BERT, RoBERTa, and DistilBERT."
"What solutions come pre-built with Amazon SageMaker JumpStart?
","SageMaker JumpStart includes solutions that are preconfigured with all necessary AWS services to launch a solution into production. Solutions are fully customizable so you can easily modify them to fit your specific use case and dataset. You can use solutions for over 15 use cases including demand forecasting, fraud detection, and predictive maintenance, and readily deploy solutions with just a few clicks. For more information about all solutions available, visit the SageMaker getting started page. "
How can I share ML artifacts with others within my organization?,"With Amazon SageMaker JumpStart, data scientists and ML developers can easily share ML artifacts, including notebooks and models, within their organization. Administrators can set up a repository that is accessible by a defined set of users. All users with permission to access the repository can browse, search, and use models and notebooks as well as the public content inside of SageMaker JumpStart. Users can select artifacts to train models, deploy endpoints, and execute notebooks in SageMaker JumpStart."
Why should I use Amazon SageMaker JumpStart to share ML artifacts with others within my organization?,"With Amazon SageMaker JumpStart, you can accelerate time-to-market when building ML applications. Models and notebooks built by one team inside of your organization can be easily shared with other teams within your organization with just a few clicks. Internal knowledge sharing and asset reuse can significantly increase the productivity of your organization."
"How does Amazon SageMaker JumpStart pricing work?
","You are charged for the AWS services launched from Amazon SageMaker JumpStart, such as training jobs and endpoints, based on SageMaker pricing. There is no additional charge for using SageMaker JumpStart."
What is Amazon SageMaker Canvas?,"Amazon SageMaker Canvas is a no-code

service with an intuitive, point-and-click interface that lets you create highly accurate ML-based predictions from your data. SageMaker Canvas lets you access and combine data from a variety of sources using a drag-and-drop user interface, automatically cleaning and preparing data to minimize manual cleanup. SageMaker Canvas applies a variety of state-of-the-art ML algorithms to find highly accurate predictive models and provides an intuitive interface to make predictions. You can use SageMaker Canvas to make much more precise predictions in a variety of business applications and easily collaborate with data scientists and analysts in your enterprise by sharing your models, data, and reports. To learn more about SageMaker Canvas, please visit the SageMaker Canvas FAQ page."
"How does Amazon SageMaker Canvas pricing work?
","With Amazon SageMaker Canvas, you pay based on usage. SageMaker Canvas lets you interactively ingest, explore, and prepare your data from multiple sources, train highly accurate ML models with your data, and generate predictions. There are two components that determine your bill: session charges based on the number of hours for which SageMaker Canvas is used or logged into, and charges for training the model based on the size of the dataset used to build the model. For more information see the SageMaker Canvas pricing page."
How can I build a continuous integration and delivery (CI/CD) pipeline with Amazon SageMaker?,"Amazon SageMaker Pipelines helps you create fully automated ML workflows from data preparation through model deployment so you can scale to thousands of ML models in production. SageMaker Pipelines comes with a Python SDK that connects to Amazon SageMaker Studio so you can take advantage of a visual interface to build each step of the workflow. Then using a single API, you can connect each step to create an end-to-end workflow. SageMaker Pipelines takes care of managing data between steps, packaging the code recipes, and orchestrating their execution, reducing months of coding to a few hours. Every time a workflow executes, a complete record of the data processed and actions taken is kept so data scientists and ML developers can quickly debug problems."
How do I view all my trained models to choose the best model to move to production?,"Amazon SageMaker Pipelines provides a central repository of trained models called a model registry. You can discover models and access the model registry visually through SageMaker Studio or programmatically through the Python SDK, making it easy to choose your desired model to deploy into production."
What components of Amazon SageMaker can be added to Amazon SageMaker Pipelines?,"The components available through Amazon SageMaker Studio, including Amazon SageMaker Amazon Clarify, Amazon SageMaker Data Wrangler, Amazon SageMaker Feature Store, Amazon SageMaker Experiments, Amazon SageMaker Debugger, and Amazon SageMaker Model Monitor, can be added to SageMaker Pipelines."
How do I track my model components across the entire ML workflow?,"Amazon SageMaker Pipelines automatically keeps track of all model constituents and keeps an audit trail of all changes, thereby eliminating manual tracking, and can help you achieve compliance

goals. You can track data, code, trained models, and more with SageMaker Pipelines."
How does the pricing for Amazon SageMaker Pipelines work?,There is no additional charge for Amazon SageMaker Pipelines. You pay only for the underlying compute or any separate AWS services you use within SageMaker Pipelines.
Can I use Kubeflow with Amazon SageMaker?,"Yes. Amazon SageMaker Components for Kubeflow Pipelines are open-source plugins that allow you to use Kubeflow Pipelines to define your ML workflows and use SageMaker for the data labeling, training, and inference steps. Kubeflow Pipelines is an add-on to Kubeflow that lets you build and deploy portable and scalable end-to-end ML pipelines. However, when using Kubeflow Pipelines, ML ops teams need to manage a Kubernetes cluster with CPU and GPU instances and keep its utilization high at all times to reduce operational costs. Maximizing the utilization of a cluster across data science teams is challenging and adds additional operational overhead to the ML ops teams. As an alternative to an ML-optimized Kubernetes cluster, with SageMaker Components for Kubeflow Pipelines you can take advantage of powerful SageMaker features such as data labeling, fully managed large-scale hyperparameter tuning and distributed training jobs, one-click secure and scalable model deployment, and cost-effective training through Amazon EC2 Spot instances without needing to configure and manage Kubernetes clusters specifically to run the ML jobs."
How does Amazon SageMaker Components for Kubeflow Pipelines pricing work?,There is no additional charge for using Amazon SageMaker Components for Kubeflow Pipelines.
How can Amazon SageMaker prepare data for ML?,"Amazon SageMaker Data Wrangler reduces the time it takes to aggregate and prepare data for ML. From a single interface in Amazon SageMaker Studio, you can browse and import data from Amazon S3, Amazon Athena, Amazon Redshift, AWS Lake Formation, Amazon SageMaker Feature Store, and Snowflake in just a few clicks. You can also query and import data that is transferred from over 40 data sources and registered in AWS Glue Data Catalog by Amazon AppFlow. SageMaker Data Wrangler will automatically load, aggregate, and display the raw data. After importing your data into SageMaker Data Wrangler, you can see automatically generated column summaries and histograms. You can then dig deeper to understand your data and identify potential errors with the SageMaker Data Wrangler Data Quality and Insights report, which provides summary statistics and data quality warnings. You can also run bias analysis supported by Amazon SageMaker Clarify directly from SageMaker Data Wrangler to detect potential bias during data preparation. From there, you can use SageMaker Data Wrangler’s pre-built transformations to prepare your data. Once your data is prepared, you can build fully automated ML workflows with Amazon SageMaker Pipelines or import that data into Amazon SageMaker Feature Store."
How can I create model features with Amazon SageMaker Data Wrangler?,"Without writing a single line of code, Amazon SageMaker Data Wrangler can automatically transform your data into new features.

SageMaker Data Wrangler offers a selection of preconfigured data transforms, impute missing data, one-hot encoding, dimensionality reduction using principal components analysis (PCA), as well as time-series specific transformers. For example, you can convert a text field column into a numerical column with a single click. You can also author a code snippet from SageMaker Data Wrangler's library of snippets."
How can I visualize my data in Amazon SageMaker Data Wrangler?,"Amazon SageMaker Data Wrangler helps you understand your data and identify potential errors and extreme values with a set of robust pre-configured visualization templates. Histograms, scatter plots, and ML-specific visualizations, such as target leakage detection, are all available without writing a single line of code. You can also create and edit your own visualizations."
How does the pricing for Amazon SageMaker Data Wrangler work?,"You pay for all ML compute, storage, and data processing resources you use for Amazon SageMaker Data Wrangler. You can review all the details of SageMaker Data Wrangler pricing here. As part of the AWS Free Tier, you can also get started with SageMaker Data Wrangler for free."
How can I train machine learning models with data prepared in Amazon SageMaker Data Wrangler?,"Amazon SageMaker Data Wrangler provides a unified experience enabling you to prepare data and seamlessly train a machine learning model in Amazon SageMaker Autopilot. SageMaker Autopilot automatically builds, trains, and tunes the best ML models based on your data. With SageMaker Autopilot, you still maintain full control and visibility of your data and model. You can also use features prepared in SageMaker Data Wrangler with your existing models. You can configure Amazon SageMaker Data Wrangler processing jobs to run as part of your SageMaker training pipeline either by configuring the job in the user interface (UI) or exporting a notebook with the orchestration code."
How does Amazon SageMaker Data Wrangler handle new data when I have prepared my features on historical data?,"You can configure and launch Amazon SageMaker processing jobs directly from the SageMaker Data Wrangler UI, including scheduling your data processing job and parametrizing your data sources to easily transform new batches of data at scale."
How does Amazon SageMaker Data Wrangler work with my CI/CD processes?,"Once you have prepared your data, Amazon SageMaker Data Wrangler provides different options for promoting your SageMaker Data Wrangler flow to production and integrates seamlessly with MLOps and CI/CD capabilities. You can configure and launch SageMaker processing jobs directly from the SageMaker Data Wrangler UI, including scheduling your data processing job and parametrizing your data sources to easily transform new batches of data at scale. Alternatively, SageMaker Data Wrangler integrates seamlessly with SageMaker processing and the SageMaker Spark container, allowing you to easily use SageMaker SDKs to integrate SageMaker Data Wrangler into your production workflow."
"What model does Amazon SageMaker Data Wrangler Quick Model use?

","In a few clicks of a button, Amazon SageMaker Data Wrangler splits and trains an XGBoost model with default hyperparameters. Based on the problem type, SageMaker Data Wrangler provides a model summary, feature summary, and confusion matrix to quickly give you insight so you can iterate on your data preparation flows. "
"What size data does Amazon SageMaker Data Wrangler support?
","Amazon SageMaker Data Wrangler supports various sampling techniques—such as top-K, random, and stratified sampling for importing data—so that you can quickly transform your data using SageMaker Data Wrangler's UI. If you are using large or wide datasets, you can increase the SageMaker Data Wrangler instance size to improve performance. Once you have created your flow, you can process your full dataset using SageMaker Data Wrangler processing jobs."
Does Amazon SageMaker Data Wrangler work with Amazon SageMaker Feature Store?,You can configure Amazon SageMaker Feature Store as a destination for your features prepared in Amazon SageMaker Data Wrangler. This can be done directly in the UI or you can export a notebook generated specifically for processing data with SageMaker Feature Store as the destination.
How do I store features for my ML models?,"Amazon SageMaker Feature Store provides a central repository for data features with low latency (milliseconds) reads and writes. Features can be stored, retrieved, discovered, and shared through SageMaker Feature Store for easy reuse across models and teams with secure access and control. SageMaker Feature Store supports both online and offline features generated via batch or streaming pipelines. It supports backfilling the features and provides both online and offline stores to maintain parity between features used in model training and inference."
How do I maintain consistency between online and offline features?,Amazon SageMaker Feature Store automatically maintains consistency between online and offline features without additional management or code. SageMaker Feature Store is fully managed and maintains consistency across training and inference environments.
How can I reproduce a feature from a given moment in time?,Amazon SageMaker Feature Store maintains time stamps for all features at every instance of time. This helps you retrieve features at any period of time for business or compliance requirements. You can easily explain model features and their values from when they were first created to the present time by reproducing the model from a given moment in time.
What are offline features?,"Offline features are used for training because you need access to very large volumes over a long period of time. These features are served from a high-throughput, high-bandwidth repository."
What are online features?,Online features are used in applications required to make real-time predictions. Online features are served from a high-throughput repository with single-digit millisecond latency for fast predictions.
How does pricing work for Amazon SageMaker Feature Store?,"You can get started with Amazon SageMaker Feature Store for free, as part of

the AWS Free Tier. With SageMaker Feature Store, you pay for writing into the feature store, and reading and storage from the online feature store. For pricing details, see the SageMaker Pricing Page."
What does Amazon SageMaker offer for data labeling?,"Amazon SageMaker provides two data labeling offerings, Amazon SageMaker Ground Truth Plus and Amazon SageMaker Ground Truth. Both options allow you to identify raw data, such as images, text files, and videos, and add informative labels to create high-quality training datasets for your ML models. To learn more, visit the SageMaker Data Labeling webpage."
What is geospatial data? ,"Geospatial data represents features or objects on the Earth’s surface. The first type of geospatial data is vector data which uses two-dimensional geometries such as, points, lines, or polygons to represent objects like roads and land boundaries. The second type of geospatial data is raster data such as imagery captured by satellite, aerial platforms, or remote sensing data. This data type uses a matrix of pixels to define where features are located. You can use raster formats for storing data that varies. A third type of geospatial data is geo-tagged location data. It includes points of interest—for example, the Eiffel Tower—location tagged social media posts, latitude and longitude coordinates, or different styles and formats of street addresses. "
What are Amazon SageMaker geospatial capabilities? ,"Amazon SageMaker geospatial capabilities make it easier for data scientists and machine learning (ML) engineers to build, train, and deploy ML models for making predictions using geospatial data. You can bring your own data, for example, Planet Labs satellite data from Amazon S3, or acquire data from Open Data on AWS, Amazon Location Service, and other Amazon SageMaker geospatial data sources. "
Why should I use geospatial ML on Amazon SageMaker?,"You can use Amazon SageMaker geospatial capabilities to make predictions on geospatial data faster than do-it-yourself solutions. Amazon SageMaker geospatial capabilities make it easier to access geospatial data from your existing customer data lakes, open-source datasets, and other Amazon SageMaker geospatial data sources. Amazon SageMaker geospatial capabilities minimize the need for building custom infrastructure and data pre-processing functions by offering purpose-built algorithms for efficient data preparation, model training, and inference. You can also create and share custom visualizations and data with your organization from Amazon SageMaker Studio. Amazon SageMaker geospatial capabilities include pre-trained models for common uses in agriculture, real estate, insurance, and financial services."
What are Amazon SageMaker Studio notebooks?,"Amazon SageMaker Studio notebooks are quick start, collaborative, managed Jupyter notebooks. Amazon SageMaker Studio notebooks integrate with purpose-built ML tools in SageMaker and other AWS services for end-to-end ML development in Amazon SageMaker Studio, the fully integrated development environment (IDE) for ML."
How are Amazon SageMaker Studio notebooks different from the instance-based notebooks offering?,"SageMaker Studio notebooks offer a few important features that differentiate them from the instance-based

notebooks. With the Studio notebooks, you can quickly launch notebooks without needing to manually provision an instance and waiting for it to be operational. The startup time of launching the UI to read and execute a notebook is faster than the instance-based notebooks. You also have the flexibility to choose from a large collection of instance types from within the UI at any time. You do not need to go to the AWS Management Console to start new instances and port over your notebooks. Each user has an isolated home directory independent of a particular instance. This directory is automatically mounted into all notebook servers and kernels as they're started, so you can access your notebooks and other files even when you switch instances to view and run your notebooks. SageMaker Studio notebooks are integrated with AWS IAM Identity Center (successor to AWS SSO), making it easy to use your organizational credentials to access the notebooks. Notebook sharing is an integrated feature in SageMaker Studio notebooks. You can share your notebooks with your peers using a single click or even co-edit a single notebook at the same time."
How do Amazon SageMaker Studio notebooks work?,"Amazon SageMaker Studio notebooks are one-click Jupyter notebooks that can be spun quickly. The underlying compute resources are fully elastic, so you can easily dial up or down the available resources and the changes take place automatically in the background without interrupting your work. SageMaker also enables one-click sharing of notebooks. You can easily share notebooks with others and they'll get the exact same notebook, saved in the same place. With SageMaker Studio notebooks you can sign in with your corporate credentials using AWS IAM Identity Center (successor to AWS SSO). Sharing notebooks within and across teams is easy, since the dependencies needed to run a notebook are automatically tracked in work images that are encapsulated with the notebook as it is shared."
What are the shared spaces in Amazon SageMaker?,"Machine learning practitioners can create a shared workspace where teammates can read and edit Amazon SageMaker Studio notebooks together. By using the shared paces, teammates can coedit the same notebook file, run notebook code simultaneously, and review the results together to eliminate back and forth and streamline collaboration. In the shared spaces, ML teams will have built-in support for services like BitBucket and AWS CodeCommit, so they can easily manage different versions of their notebook and compare changes over time. Any resources created from within the notebooks, such as experiments and ML models, are automatically saved and associated with the specific workspace where they were created so teams can more easily stay organized and accelerate ML model development."
How do Amazon SageMaker Studio notebooks work with other AWS services?,"Amazon SageMaker Studio notebooks give you access to all SageMaker features, such as distributed training, batch transform, hosting, and experiment management. You can access other services such as datasets in Amazon S3, Amazon Redshift, AWS Glue, Amazon EMR, or AWS Lake Formation from SageMaker notebooks."
How does Amazon SageMaker Studio notebooks pricing work?,"You pay for

both compute and storage when you use SageMaker Studio notebooks. See Amazon SageMaker Pricing for charges by compute instance type. Your notebooks and associated artifacts such as data files and scripts are persisted on Amazon EFS. See Amazon EFS Pricing for storage charges. As part of the AWS Free Tier, you can get started with Amazon SageMaker Studio notebooks for free."
Do I get charged separately for each notebook created and run in SageMaker Studio?,"No. You can create and run multiple notebooks on the same compute instance. You pay only for the compute that you use, not for individual items. You can read more about this in our metering guide. In addition to the notebooks, you can also start and run terminals and interactive shells in SageMaker Studio, all on the same compute instance. Each application runs within a container or image. SageMaker Studio provides several built-in images purpose-built and preconfigured for data science and ML. You can read more about the SageMaker Studio developer environment in the guide for using SageMaker Studio notebooks."
How do I monitor and shut down the resources used by my notebooks?,You can monitor and shut down the resources used by your SageMaker Studio notebooks through both SageMaker Studio visual interface and the AWS Management Console. See the documentation for more details.
"I’m running a SageMaker Studio notebook. Will I still be charged if I close my browser, close the notebook tab, or just leave the browser open?","Yes, you will continue to be charged for the compute. This is similar to starting Amazon EC2 instances in the AWS Management Console and then closing the browser. The Amazon EC2 instances are still running and you still incur charges unless you explicitly shut down the instance."
Do I get charged for creating and setting up an Amazon SageMaker Studio domain?,"No, you don’t get charged for creating or configuring an Amazon SageMaker Studio domain, including adding, updating, and deleting user profiles."
How do I see the itemized charges for Amazon SageMaker Studio notebooks or other Amazon SageMaker services?,"As an admin, you can view the list of itemized charges for Amazon SageMaker, including SageMaker Studio, in the AWS Billing console. From the AWS Management Console for SageMaker, choose Services on the top menu, type ""billing"" in the search box and select Billing from the dropdown, then select Bills on the left panel. In the Details section, you can click on SageMaker to expand the list of Regions and drill down to the itemized charges."
What is Amazon SageMaker Studio Lab?,"Amazon SageMaker Studio Lab is a free ML development environment that provides the compute, storage (up to 15 GB), and security—all at no cost—for anyone to learn and experiment with ML. All you need to get started is a valid email ID; you don’t need to configure infrastructure or manage identity and access or even sign up for an AWS account. SageMaker Studio Lab accelerates model building through GitHub integration, and it comes preconfigured with the most popular ML tools, frameworks, and libraries to get you started immediately. SageMaker Studio Lab

automatically saves your work so you don’t need to restart between sessions. It’s as easy as closing your laptop and coming back later."
Why should I use Amazon SageMaker Studio Lab?,"Amazon SageMaker Studio Lab is for students, researchers, and data scientists who need a free notebook development environment with no setup required for their ML classes and experiments. SageMaker Studio Lab is ideal for users who do not need a production environment but still want a subset of the SageMaker functionality to improve their ML skills. SageMaker sessions are automatically saved, enabling users to pick up where they left off for each user session."
How does Amazon SageMaker Studio Lab work with other AWS services?,"Amazon SageMaker Studio Lab is a service built on AWS and uses many of the same core services as Amazon SageMaker Studio, such as Amazon S3 and Amazon EC2. Unlike the other services, customers will not need an AWS account. Instead, they will create an Amazon SageMaker Studio Lab specific account with an email address. This will give the user access to a limited environment (15 GB of storage, and 12 hour sessions) for them to run ML notebooks."
What is Amazon SageMaker Canvas?,"Amazon SageMaker Canvas is a visual drag-and-drop service that allows business analysts to build ML models and generate accurate predictions without writing any code or requiring ML expertise. SageMaker Canvas makes it easy to access and combine data from a variety of sources, automatically clean data and apply a variety of data adjustments, and build ML models to generate accurate predictions with a single click. You can also easily publish results, explain and interpret models, and share models with others within your organization to review."
"What data sources does Amazon SageMaker Canvas support?
","Amazon SageMaker Canvas enables you to seamlessly discover AWS data sources that your account has access to, including Amazon S3 and Amazon Redshift. You can browse and import data using the SageMaker Canvas visual drag-and-drop interface. Additionally, you can drag and drop files from your local disk, and use pre-built connectors to import data from third-party sources such as Snowflake."
How do I build an ML model to generate accurate predictions in Amazon SageMaker Canvas?,"Once you have connected sources, selected a dataset, and prepared your data, you can select the target column that you want to predict to initiate a model creation job. Amazon SageMaker Canvas will automatically identify the problem type, generate new relevant features, test a comprehensive set of prediction models using ML techniques such as linear regression, logistic regression, deep learning, time-series forecasting, and gradient boosting, and build the model that makes accurate predictions based on your dataset."
"How long does it take to build a model in Amazon SageMaker Canvas? How can I monitor progress during model creation?
","The time it takes to build a model depends on the size of your dataset. Small datasets can take less than 30 minutes, and large datasets can take a few hours. As the model creation job progresses, Amazon SageMaker Canvas provides detailed visual updates, including percent job complete and the amount of time left for job completion."

What is Amazon SageMaker Experiments?,"Amazon SageMaker Experiments helps you organize and track iterations to ML models. SageMaker Experiments helps you manage iterations by automatically capturing the input parameters, configurations, and results, and storing them as ""experiments"". You can work within the visual interface of Amazon SageMaker Studio, where you can browse active experiments, search for previous experiments by their characteristics, review previous experiments with their results, and compare experiment results visually."
What is Amazon SageMaker Debugger?,"Amazon SageMaker Debugger automatically captures real-time metrics during training, such as confusion matrices and learning gradients, to help improve model accuracy. The metrics from SageMaker Debugger can be visualized in Amazon SageMaker Studio for easy understanding. SageMaker Debugger can also generate warnings and remediation advice when common training problems are detected. SageMaker Debugger also automatically monitors and profiles system resources such as CPUs, GPUs, network, and memory in real time, and provides recommendations on re-allocation of these resources. This enables you to use your resources efficiently during training and helps reduce costs and resources."
Does Amazon SageMaker support distributed training?,"Yes. Amazon SageMaker can automatically distribute deep learning models and large training sets across AWS GPU instances in a fraction of the time it takes to build and optimize these distribution strategies manually. The two distributed training techniques that SageMaker applies are data parallelism and model parallelism. Data parallelism is applied to improve training speeds by dividing the data equally across multiple GPU instances, allowing each instance to train concurrently. Model parallelism is useful for models too large to be stored on a single GPU and require the model to be partitioned into smaller parts before distributing across multiple GPUs. With only a few lines of additional code in your PyTorch and TensorFlow training scripts, SageMaker will automatically apply data parallelism or model parallelism for you, allowing you to develop and deploy your models faster. SageMaker will determine the best approach to split your model by using graph partitioning algorithms to balance the computation of each GPU while minimizing the communication between GPU instances. SageMaker also optimizes your distributed training jobs through algorithms that fully utilize the AWS compute and network in order to achieve near-linear scaling efficiency, which allows you to complete training faster than manual open-source implementations."
What is Amazon SageMaker Training Compiler?,"Amazon SageMaker Training Compiler is a deep learning (DL) compiler that accelerates DL model training by up to 50 percent through graph- and kernel-level optimizations to use GPUs more efficiently. SageMaker Training Compiler is integrated with versions of TensorFlow and PyTorch in SageMaker, so you can speed up training in these popular frameworks with minimal code changes."
How does Amazon SageMaker Training Compiler work?,"Amazon SageMaker Training Compiler accelerates training jobs by converting DL models

from their high-level language representation to hardware-optimized instructions that train faster than jobs with the native frameworks. More specifically, SageMaker Training Compiler uses graph-level optimization (operator fusion, memory planning, and algebraic simplification), data flow-level optimizations (layout transformation, common sub-expression elimination), and backend optimizations (memory latency hiding, loop oriented optimizations) to produce an optimized model training job that more efficiently uses hardware resources and, as a result, trains faster."
How can I use Amazon SageMaker Training Compiler?,"Amazon SageMaker Training Compiler is built into the SageMaker Python SDK and SageMaker Hugging Face Deep Learning Containers. You don’t need to change your workflows to access its speedup benefits. You can run training jobs in the same way as you already do, using any of the SageMaker interfaces: Amazon SageMaker notebook instances, Amazon SageMaker Studio, AWS SDK for Python (Boto3), and AWS Command Line Interface. You can enable SageMaker Training Compiler by adding a TrainingCompilerConfig class as a parameter when you create a framework estimator object. Practically, this means a couple of lines of code added to your existing training job script for a single GPU instance. Most up-to-date detailed documentation, sample notebooks, and examples are available in the documentation."
What is the pricing of Amazon SageMaker Training Compiler? ,Training Compiler is a SageMaker Training feature and is provided at no additional charge exclusively to SageMaker customers. Customers can actually reduce their costs with Training Compiler as training times are reduced.
What is Managed Spot Training?,"Managed Spot Training with Amazon SageMaker lets you train your ML models using Amazon EC2 Spot instances, while reducing the cost of training your models by up to 90%."
How do I use Managed Spot Training?,"You enable the Managed Spot Training option when submitting your training jobs and you also specify how long you want to wait for Spot capacity. Amazon SageMaker will then use Amazon EC2 Spot instances to run your job and manages the Spot capacity. You have full visibility into the status of your training jobs, both while they are running and while they are waiting for capacity."
When should I use Managed Spot Training?,"Managed Spot Training is ideal when you have flexibility with your training runs and when you want to minimize the cost of your training jobs. With Managed Spot Training, you can reduce the cost of training your ML models by up to 90%."
How does Managed Spot Training work?,"Managed Spot Training uses Amazon EC2 Spot instances for training, and these instances can be pre-empted when AWS needs capacity. As a result, Managed Spot Training jobs can run in small increments as and when capacity becomes available. The training jobs need not be restarted from scratch when there is an interruption, as Amazon SageMaker can resume the training jobs using the latest model checkpoint. The built-in frameworks and

the built-in computer vision algorithms with SageMaker enable periodic checkpoints, and you can enable checkpoints with custom models."
Do I need to periodically checkpoint with Managed Spot Training?,"We recommend periodic checkpoints as a general best practice for long-running training jobs. This prevents your Managed Spot Training jobs from restarting if capacity is pre-empted. When you enable checkpoints, Amazon SageMaker resumes your Managed Spot Training jobs from the last checkpoint."
How do you calculate the cost savings with Managed Spot Training jobs?,"Once a Managed Spot Training job is completed, you can see the savings in the AWS Management Console and also calculate the cost savings as the percentage difference between the duration for which the training job ran and the duration for which you were billed. Regardless of how many times your Managed Spot Training jobs are interrupted, you are charged only once for the duration for which the data was downloaded."
Which instances can I use with Managed Spot Training?,"Managed Spot Training can be used with all instances supported in Amazon SageMaker.
"
Which AWS Regions are supported with Managed Spot Training?,"Managed Spot Training is supported in all AWS Regions where Amazon SageMaker is currently available.
"
Are there limits to the size of the dataset I can use for training?,"There are no fixed limits to the size of the dataset you can use for training models with Amazon SageMaker.
"
What algorithms does Amazon SageMaker use to generate models?,"Amazon SageMaker includes built-in algorithms for linear regression, logistic regression, k-means clustering, principal component analysis, factorization machines, neural topic modeling, latent dirichlet allocation, gradient boosted trees, sequence2sequence, time-series forecasting, word2vec, and image classification. SageMaker also provides optimized Apache MXNet, Tensorflow, Chainer, PyTorch, Gluon, Keras, Horovod, Scikit-learn, and Deep Graph Library containers. In addition, Amazon SageMaker supports your custom training algorithms provided through a Docker image adhering to the documented specification."
"What is Automatic Model Tuning?
",Most ML algorithms expose a variety of parameters that control how the underlying algorithm operates. Those parameters are generally referred to as hyperparameters and their values affect the quality of the trained models. Automatic model tuning is the process of finding a set of hyperparameters for an algorithm that can yield an optimal model.
What models can be tuned with Automatic Model Tuning?,"You can run automatic model tuning in Amazon SageMaker on top of any algorithm as long as it’s scientifically feasible, including built-in SageMaker algorithms, deep neural networks, or arbitrary algorithms you bring to SageMaker in the form of Docker images."

Can I use Automatic Model Tuning outside of Amazon SageMaker?,Not at this time. The best model tuning performance and experience is within Amazon SageMaker.
What is the underlying tuning algorithm for Automatic Model Tuning?,"Currently, the algorithm for tuning hyperparameters is a customized implementation of Bayesian Optimization. It aims to optimize a customer-specified objective metric throughout the tuning process. Specifically, it checks the object metric of completed training jobs, and uses the knowledge to infer the hyperparameter combination for the next training job."
"Does Automatic Model Tuning recommend specific hyperparameters for tuning?
","No. How certain hyperparameters impact the model performance depends on various factors, and it is hard to definitively say one hyperparameter is more important than the others and thus needs to be tuned. For built-in algorithms within Amazon SageMaker, we do call out whether or not a hyperparameter is tunable."
"How long does a hyperparameter tuning job take?
","The length of time for a hyperparameter tuning job depends on multiple factors, including the size of the data, the underlying algorithm, and the values of the hyperparameters. Additionally, customers can choose the number of simultaneous training jobs and total number of training jobs. All these choices affect how long a hyperparameter tuning job can last."
"Can I optimize multiple objectives simultaneously, such as optimizing a model to be both fast and accurate?
","Not at this time. Currently, you need to specify a single objective metric to optimize or change your algorithm code to emit a new metric, which is a weighted average between two or more useful metrics, and have the tuning process optimize towards that objective metric."
"How much does Automatic Model Tuning cost?
","There is no charge for a hyperparameter tuning job itself. You will be charged by the training jobs that are launched by the hyperparameter tuning job, based on model training pricing."
"How do I decide to use Amazon SageMaker Autopilot or Automatic Model Tuning?
","Amazon SageMaker Autopilot automates everything in a typical ML workflow, including feature preprocessing, algorithm selection, and hyperparameter tuning, while specifically focusing on classification and regression use cases. Automatic Model Tuning, on the other hand, is designed to tune any model, no matter whether it is based on built-in algorithms, deep learning frameworks, or custom containers. In exchange for the flexibility, you have to manually pick the specific algorithm, hyperparameters to tune, and corresponding search ranges."
What is reinforcement learning?,Reinforcement learning is a ML technique that enables an agent to learn in an interactive environment by trial and error using feedback from its own actions and experiences.
"Can I train reinforcement learning models in Amazon SageMaker?
","Yes, you can train reinforcement learning models in Amazon

SageMaker in addition to supervised and unsupervised learning models."
How is reinforcement learning different from supervised learning?,"Though both supervised and reinforcement learning use mapping between input and output, unlike supervised learning where the feedback provided to the agent is the correct set of actions for performing a task, reinforcement learning uses a delayed feedback where reward signals are optimized to ensure a long-term goal through a sequence of actions."
"When should I use reinforcement learning?
","While the goal of supervised learning techniques is to find the right answer based on the patterns in the training data, the goal of unsupervised learning techniques is to find similarities and differences between data points. In contrast, the goal of reinforcement learning (RL) techniques is to learn how to achieve a desired outcome even when it is not clear how to accomplish that outcome. As a result, RL is more suited to enabling intelligent applications where an agent can make autonomous decisions such as robotics, autonomous vehicles, HVAC, industrial control, and more."
"What type of environments can I use for training RL models?
","Amazon SageMaker RL supports a number of different environments for training RL models. You can use AWS services such as AWS RoboMaker, open-source environments or custom environments developed using Open AI Gym interfaces, or commercial simulation environments such as MATLAB and SimuLink."
"Do I need to write my own RL agent algorithms to train RL models?
","No, Amazon SageMaker RL includes RL toolkits such as Coach and Ray RLLib that offer implementations of RL agent algorithms such as DQN, PPO, A3C, and many more."
Can I bring my own RL libraries and algorithm implementation and run them in Amazon SageMaker RL?,"Yes, you can bring your own RL libraries and algorithm implementations in Docker Containers and run those in Amazon SageMaker RL."
"Can I do distributed rollouts using Amazon SageMaker RL?
",Yes. You can even select a heterogeneous cluster where the training can run on a GPU instance and the simulations can run on multiple CPU instances.
What deployment options does Amazon SageMaker provide? ,"After you build and train models, Amazon SageMaker provides three options to deploy them so you can start making predictions. Real-time inference is suitable for workloads with millisecond latency requirements, payload sizes up to 6 MB, and processing times of up to 60 seconds. Batch transform is ideal for offline predictions on large batches of data that are available up front. Asynchronous inference is designed for workloads that do not have sub-second latency requirements, payload sizes up to 1 GB, and processing times of up to 15 minutes. "
What is Amazon SageMaker Asynchronous Inference?,"Amazon SageMaker Asynchronous Inference queues incoming requests and processes them asynchronously. This option is ideal for requests with large payload sizes and/or long processing times that need to be processed as they arrive. Optionally, you can configure auto-scaling settings to scale

down the instance count to zero when not actively processing requests to save on costs. "
How do I configure auto-scaling settings to scale down the instance count to zero when not actively processing requests?,"You can scale down the Amazon SageMaker Asynchronous Inference endpoint instance count to zero in order to save on costs when you are not actively processing requests. You need to define a scaling policy that scales on the ""ApproximateBacklogPerInstance"" custom metric and set the ""MinCapacity"" value to zero. For step-by-step instructions, please visit the autoscale an asynchronous endpoint section of the developer guide. "
What is Amazon SageMaker Serverless Inference?,"Amazon SageMaker Serverless Inference is a purpose-built serverless model serving option that makes it easy to deploy and scale ML models. SageMaker Serverless Inference endpoints automatically start the compute resources and scale them in and out depending on traffic, eliminating the need for you to choose instance type, run provisioned capacity, or manage scaling. You can optionally specify the memory requirements for your serverless inference endpoint. You pay only for the duration of running the inference code and the amount of data processed, not for idle periods."
Why should I use Amazon SageMaker Serverless Inference?,"Amazon SageMaker Serverless Inference simplifies the developer experience by eliminating the need to provision capacity up front and manage scaling policies. SageMaker Serverless Inference can scale instantly from tens to thousands of inferences within seconds based on the usage patterns, making it ideal for ML applications with intermittent or unpredictable traffic. For example, a chatbot service used by a payroll processing company experiences an increase in inquiries at the end of the month while for rest of the month traffic is intermittent. Provisioning instances for the entire month in such scenarios is not cost-effective, as you end up paying for idle periods. SageMaker Serverless Inference helps address these types of use cases by providing you automatic and fast scaling out of the box without the need for you to forecast traffic up front or manage scaling policies. Additionally, you pay only for the compute time to run your inference code (billed in milliseconds) and for data processing, making it a cost-effective option for workloads with intermittent traffic."
What is Amazon SageMaker shadow testing?,"SageMaker helps you run shadow tests to evaluate a new ML model before production release by testing its performance against the currently deployed model. SageMaker deploys the new model in shadow mode alongside the current production model and mirrors a user-specified portion of the production traffic to the new model. It optionally logs the model inferences for offline comparison. It also provides a live dashboard with a comparison of key performance metrics, such as latency and error rate, between the production and shadow models to help you decide whether to promote the new model to production."
Why should I use SageMaker for shadow testing?,"SageMaker simplifies the process of setting up and monitoring shadow variants so you can

evaluate the performance of the new ML model on live production traffic. SageMaker eliminates the need for you to orchestrate infrastructure for shadow testing. It lets you control testing parameters such as the percentage of traffic mirrored to the shadow variant and the duration of the test. As a result, you can start small and increase the inference requests to the new model after you gain confidence in model performance. SageMaker creates a live dashboard displaying performance differences across key metrics, so you can easily compare model performance to evaluate how the new model differs from the production model."
What is Amazon SageMaker Inference Recommender?,"Amazon SageMaker Inference Recommender is a new capability of Amazon SageMaker that reduces the time required to get ML models in production by automating performance benchmarking and tuning model performance across SageMaker ML instances. You can now use SageMaker Inference Recommender to deploy your model to an endpoint that delivers the best performance and minimizes cost. You can get started with SageMaker Inference Recommender in minutes while selecting an instance type and get recommendations for optimal endpoint configurations within hours, eliminating weeks of manual testing and tuning time. With SageMaker Inference Recommender, you pay only for the SageMaker ML instances used during load testing, and there are no additional charges."
Why should I use Amazon SageMaker Inference Recommender?,"You should use SageMaker Inference Recommender if you need recommendations for the right endpoint configuration to improve performance and reduce costs. Previously, data scientists who wanted to deploy their models had to run manual benchmarks to select the right endpoint configuration. They had to first select the right ML instance type out of the 70+ available instance types based on the resource requirements of their models and sample payloads, and then optimize the model to account for differing hardware. Then, they had to conduct extensive load tests to validate that latency and throughput requirements are met and that the costs are low. SageMaker Inference Recommender eliminates this complexity by making it easy for you to: 1) get started in minutes with an instance recommendation; 2) conduct load tests across instance types to get recommendations on your endpoint configuration within hours; and 3) automatically tune container and model server parameters as well as perform model optimizations for a given instance type."
How does Amazon SageMaker Inference Recommender work with other AWS services?,"Data scientists can access Amazon SageMaker Inference Recommender from SageMaker Studio, AWS SDK for Python (Boto3), or AWS CLI. They can get deployment recommendations within SageMaker Studio in the SageMaker model registry for registered model versions. Data scientists can search and filter the recommendations through SageMaker Studio, AWS SDK, or AWS CLI."
"Can Amazon SageMaker Inference Recommender support multi-model endpoints or multi-container endpoints?
","No, we currently support only a single model per endpoint."
What type of endpoints does SageMaker Inference Recommender

support?,Currently we support only real-time endpoints.
Can I use SageMaker Inference Recommender in one Region and benchmark in different Regions?,"At launch, we will support all Regions supported by Amazon SageMaker, except the AWS China Regions."
Does Amazon SageMaker Inference Recommender support Amazon EC2 Inf1 instances?,"Yes, we support all types of containers. Amazon EC2 Inf1, based on the AWS Inferentia chip, requires a compiled model artifact using either the Neuron compiler or Amazon SageMaker Neo. Once you have a compiled model for an Inferentia target and the associated container image URI, you can use Amazon SageMaker Inference Recommender to benchmark different Inferentia instance types."
What is Amazon SageMaker Model Monitor?,"Amazon SageMaker Model Monitor allows developers to detect and remediate concept drift. SageMaker Model Monitor automatically detects concept drift in deployed models and provides detailed alerts that help identify the source of the problem. All models trained in SageMaker automatically emit key metrics that can be collected and viewed in Amazon SageMaker Studio. From inside SageMaker Studio, you can configure data to be collected, how to view it, and when to receive alerts."
Can I access the infrastructure that Amazon SageMaker runs on?,"No. Amazon SageMaker operates the compute infrastructure on your behalf, allowing it to perform health checks, apply security patches, and do other routine maintenance. You can also deploy the model artifacts from training with custom inference code in your own hosting environment."
"How do I scale the size and performance of an Amazon SageMaker model once in production?
","Amazon SageMaker hosting automatically scales to the performance needed for your application using Application Auto Scaling. In addition, you can manually change the instance number and type without incurring downtime by modifying the endpoint configuration."
How do I monitor my Amazon SageMaker production environment?,"Amazon SageMaker emits performance metrics to Amazon CloudWatch Metrics so you can track metrics, set alarms, and automatically react to changes in production traffic. In addition, Amazon SageMaker writes logs to Amazon CloudWatch Logs to let you monitor and troubleshoot your production environment."
What kinds of models can be hosted with Amazon SageMaker?,Amazon SageMaker can host any model that adheres to the documented specification for inference Docker images. This includes models created from Amazon SageMaker model artifacts and inference code.
How many concurrent real-time API requests does Amazon SageMaker support?,Amazon SageMaker is designed to scale to a large number of transactions per second. The precise number varies based on the deployed model and the number and type of instances to which the model is deployed.
What is Batch Transform?,"Batch Transform enables you to run predictions on large or small batch data. There is no need to break down the dataset into multiple chunks or manage real-time endpoints. With a simple API, you can request predictions for a large number of

data records and transform the data quickly and easily."
What is Amazon SageMaker Edge Manager?,"Amazon SageMaker Edge Manager makes it easier to optimize, secure, monitor, and maintain ML models on fleets of edge devices such as smart cameras, robots, personal computers, and mobile devices. SageMaker Edge Manager helps ML developers operate ML models on a variety of edge devices at scale."
How do I get started with Amazon SageMaker Edge Manager?,"To get started with Amazon SageMaker Edge Manager, you need to compile and package your trained ML models in the cloud, register your devices, and prepare your devices with the SageMaker Edge Manager SDK. To prepare your model for deployment, SageMaker Edge Manager uses SageMaker Neo to compile your model for your target edge hardware. Once a model is compiled, SageMaker Edge Manager signs the model with an AWS generated key, then packages the model with its runtime and your necessary credentials to get it ready for deployment. On the device side, you register your device with SageMaker Edge Manager, download the SageMaker Edge Manager SDK, and then follow the instructions to install the SageMaker Edge Manager agent on your devices. The tutorial notebook provides a step-by-step example of how you can prepare the models and connect your models on edge devices with SageMaker Edge Manager."
What devices are supported by Amazon SageMaker Edge Manager?,"Amazon SageMaker Edge Manager supports common CPU (ARM, x86) and GPU (ARM, Nvidia) based devices with Linux and Windows operating systems. Over time, SageMaker Edge Manager will expand to support more embedded processors and mobile platforms that are also supported by SageMaker Neo."
Do I need to use Amazon SageMaker to train my model in order to use Amazon SageMaker Edge Manager?,"No, you do not. You can train your models elsewhere or use a pre-trained model from open source or from your model vendor."
Do I need to use Amazon SageMaker Neo to compile my model in order to use Amazon SageMaker Edge Manager?,"Yes, you do. Amazon SageMaker Neo converts and compiles your models into an executable that you can then package and deploy on your edge devices. Once the model package is deployed, the Amazon SageMaker Edge Manager agent will unpack the model package and run the model on the device."
How do I deploy models to the edge devices?,Amazon SageMaker Edge Manager stores the model package in your specified Amazon S3 bucket. You can use the over-the-air (OTA) deployment feature provided by AWS IoT Greengrass or any other deployment mechanism of your choice to deploy the model package from your S3 bucket to the devices.
How is Amazon SageMaker Edge Manager SDK different from the SageMaker Neo runtime (dlr)?,"Neo dlr is an open-source runtime that only runs models compiled by the Amazon SageMaker Neo service. Compared to the open source dlr, the SageMaker Edge Manager SDK includes an enterprise grade on-device agent with additional security, model management, and model serving features. The SageMaker Edge Manager SDK is suitable for production deployment at scale."
How is Amazon SageMaker Edge Manager related to AWS IoT

Greengrass?,"Amazon SageMaker Edge Manager and AWS IoT Greengrass can work together in your IoT solution. Once your ML model is packaged with SageMaker Edge Manager, you can use AWS IoT Greengrass's OTA update feature to deploy the model package to your device. AWS IoT Greengrass allows you to monitor your IoT devices remotely, while SageMaker Edge Manager helps you monitor and maintain the ML models on the devices."
How is Amazon SageMaker Edge Manager related to AWS Panorama? When should I use Amazon SageMaker Edge Manager versus AWS Panorama?,"AWS offers the most breadth and depth of capabilities for running models on edge devices. We have services to support a wide range of use cases, including computer vision, voice recognition, and predictive maintenance.For companies looking to run computer vision on edge devices such as cameras and appliances, you can use AWS Panorama. Panorama offers ready-to-deploy computer vision applications for edge devices. It's easy to get started with AWS Panorama by logging into the cloud console, specifying the model you would like to use in Amazon S3 or in SageMaker, and then writing business logic as a Python script. AWS Panorama compiles the model for the target device and creates an application package so it can be deployed to your devices with just a few clicks. In addition, independent software vendors who want to build their own custom applications can use the AWS Panorama SDK, and device manufacturers can use the Device SDK to certify their devices for AWS Panorama. Customers who want to build their own models and have more granular control over model features can use Amazon SageMaker Edge Manager. SageMaker Edge Manager is a managed service to prepare, run, monitor, and update ML models across fleets of edge devices such as smart cameras, smart speakers, and robots for any use case such as natural langue processing, fraud detection, and predictive maintenance. SageMaker Edge Manager is for ML edge developers who want control over their model, including engineering different model features and monitoring models for drift. Any ML edge developer can use SageMaker Edge Manager through the SageMaker console and the SageMaker APIs. SageMaker Edge Manager brings the capabilities of SageMaker to build, train, and deploy models in the cloud to edge devices."
In which AWS Regions is Amazon SageMaker Edge Manager available?,"Amazon SageMaker Edge Manager is available in six AWS Regions: US East (N. Virginia), US East (Ohio), US West (Oregon), EU (Ireland), EU (Frankfurt), and Asia Pacific (Tokyo). For details, see the AWS Regional Services list."
What is Amazon SageMaker Neo?,Amazon SageMaker Neo enables ML models to train once and run anywhere in the cloud and at the edge. SageMaker Neo automatically optimizes models built with popular deep learning frameworks that can be used to deploy on multiple hardware platforms. Optimized models run up to 25 times faster and consume less than a tenth of the resources of typical ML models.
How do I get started with Amazon SageMaker Neo?,"To get started with Amazon SageMaker Neo, log into the Amazon SageMaker console, choose a trained model, follow the example to compile models, and deploy the

resulting model onto your target hardware platform."
What are the major components of Amazon SageMaker Neo?,"Amazon SageMaker Neo contains two major components: a compiler and a runtime. First, the Neo compiler reads models exported by different frameworks. It then converts the framework-specific functions and operations into a framework-agnostic intermediate representation. Next, it performs a series of optimizations. Then, the compiler generates binary code for the optimized operations and writes them to a shared object library. The compiler also saves the model definition and parameters into separate files. During execution, the Neo runtime loads the artifacts generated by the compiler—model definition, parameters, and the shared object library to run the model."
Do I need to use Amazon SageMaker to train my model in order to use Amazon SageMaker Neo to convert the model?,No. You can train models elsewhere and use Neo to optimize them for Amazon SageMaker ML instances or AWS IoT Greengrass supported devices.
Which models does Amazon SageMaker Neo support?,"Currently, Amazon SageMaker Neo supports the most popular deep learning models that power computer vision applications and the most popular decision tree models used in Amazon SageMaker today. Neo optimizes the performance of AlexNet, ResNet, VGG, Inception, MobileNet, SqueezeNet, and DenseNet models trained in MXNet and TensorFlow, and classification and random cut forest models trained in XGBoost."
Which hardware platforms does Amazon SageMaker Neo support?,"You can find the lists of supported cloud instances, edge devices, and framework versions in the Amazon SageMaker Neo documentation."
In which AWS Regions is Amazon SageMaker Neo available?,"To see a list of supported Regions, view the AWS Regional Services list."
What are Amazon SageMaker Savings Plans?,"Amazon SageMaker Savings Plans offer a flexible usage-based pricing model for Amazon SageMaker in exchange for a commitment to a consistent amount of usage (measured in $/hour) for a one- or three-year term. Amazon SageMaker Savings Plans provide the most flexibility and help to reduce your costs by up to 64%. These plans automatically apply to eligible SageMaker ML instance usages, including SageMaker Studio notebooks, SageMaker On-Demand notebooks, SageMaker Processing, SageMaker Data Wrangler, SageMaker Training, SageMaker Real-Time Inference, and SageMaker Batch Transform regardless of instance family, size, or Region. For example, you can change usage from a CPU instance ml.c5.xlarge running in US East (Ohio) to an ml.Inf1 instance in US West (Oregon) for inference workloads at any time and automatically continue to pay the Savings Plans price."
Why should I use Amazon SageMaker Savings Plans?,"If you have a consistent amount of Amazon SageMaker instance usage (measured in $/hour) and use multiple SageMaker components or expect your technology configuration (such as instance family, or Region) to change over time, SageMaker Savings Plans make it simpler to maximize your savings while providing flexibility to change the underlying technology configuration based on application needs or new innovation. The Savings Plans rate applies automatically to all eligible ML instance

usage with no manual modifications required."
How can I get started with Amazon SageMaker Savings Plans?,"You can get started with Savings Plans from AWS Cost Explorer in the AWS Management Console or by using the API/CLI. You can easily make a commitment to Savings Plans by using the recommendations provided in AWS Cost Explorer to realize the biggest savings. The recommended hourly commitment is based on your historical On-Demand usage and your choice of plan type, term length, and payment option. Once you sign up for a Savings Plan, your compute usage will automatically be charged at the discounted Savings Plans prices and any usage beyond your commitment will be charged at regular On-Demand rates."
How are Savings Plans for Amazon SageMaker different from Compute Savings Plans for Amazon EC2?,The difference between Savings Plans for Amazon SageMaker and Savings Plans for EC2 is in the services they include. SageMaker Savings Plans apply only to SageMaker ML Instance usage.
How do Savings Plans work with AWS Organizations/Consolidated Billing?,"Savings Plans can be purchased in any account within an AWS Organization/Consolidated Billing family. By default, the benefit provided by Savings Plans is applicable to usage across all accounts within an AWS Organization/Consolidated Billing family. However, you can also choose to restrict the benefit of Savings Plans to only the account that purchased them."