Amazon SageMaker pricing

Pricing overview

Collaborate and build faster using familiar AWS tools for model development, generative AI application development, data processing, and SQL analytics, powered by a single data and AI platform. Train and deploy models cost-effectively at scale with infrastructure, tools, and fully managed workflows built with governance built for a team your enterprise can easily work. When using SageMaker Unified Studio, you pay for the tools and compute you use, and the data you store. You pay only for each capability of SageMaker in consumed below.

Amazon SageMaker Unified Studio

SageMaker Unified Studio is a single data and AI development environment that provides an integrated experience to use all your data and tools for analytics and AI. When you create a project in SageMaker Unified Studio, it sets up the Amazon Simple Storage Service (Amazon S3) bucket, AWS Glue Data Catalog, and fine-grained access controls for you. You can then add Amazon Athena query engine, Amazon Redshift compute, and other resources to your project. There is no additional charge for using SageMaker Unified Studio. You pay only for the resources you use, allowing for cost-effective scaling of your data and AI operations. To learn more, see Amazon SageMaker Unified Studio.

In addition, each AWS service that you use through the SageMaker Unified Studio has its own pricing structure. You pay for the compute and storage resources consumed in the AWS Free Tier period applies. To learn more, see the pricing details below.

To accurately estimate your costs, review the individual pricing for the various AWS services available in SageMaker Unified Studio below. For detailed pricing examples, however, consult the official AWS pricing pages for each AWS service you plan to use.

Quick setup option

AWS offers a quick setup option to help you get started creating an AWS IAM Identity Center (IDC)-based domain. There is an additional charge for any interacting resources that AWS sets up on your behalf if you choose the quick setup option. For example, if you choose quick setup, AWS automatically creates a default VPC with public subnets in each Availability Zone, an Amazon S3 bucket, and an AWS IAM role. To learn more about quick setup, see the Amazon SageMaker Unified Studio documentation.

Other AWS Free Tier benefits

AWS Lambda and Athena also use a shared AWS Free Tier allocations as per-per-use agreements (PFAs) for AWS services that you run through SageMaker Unified Studio. Amazon Bedrock also has a free tier for certain models. To learn more about the AWS Free Tier for SageMaker Unified Studio, navigate to the Free Tier page in the AWS Billing and Cost Management console. For Amazon SageMaker workloads, AWS offers a Free Tier usage for the first 2 months. Free Tier usage is calculated monthly and is available for resources that exist AWS Free Tier benefits are associated with your AWS account, and any usage through service consoles or other interfaces also consumes your available Free Tier benefits. For example, if you launch an Amazon EC2 instance through SageMaker Unified Studio or other AWS interfaces, it will be the Amazon EC2 consumed in the AWS Free Tier period applies. To learn more, see the pricing details below.

SageMaker Unified Studio Free Tier

The SageMaker Unified Studio Free Tier helps you quickly get started developing with data and AI and at no cost by offering a selection of data, machine learning, and generative AI tools and resources. The Free Tier is available for AWS accounts that you use through the SageMaker Unified Studio.

Always-free features

Core requests in the SageMaker Unified Studio, including AWS setup through manual creation of IAM Identity Center, project management, and collaboration, are always free. You can also use the Amazon S3 storage for up to 5 GB and access your SageMaker Unified Studio environment. The first 10 GB of Amazon Redshift compute are free at the beginning of the month.

Other AWS Free Tier benefits

AWS Lambda and Athena also use a shared AWS Free Tier allocations as per-per-use agreements (PFAs) for AWS services that you run through SageMaker Unified Studio. Amazon Bedrock also has a free tier for certain models. To learn more about the AWS Free Tier for SageMaker Unified Studio, navigate to the Free Tier page in the AWS Billing and Cost Management console. For Amazon SageMaker workloads, AWS offers a Free Tier usage for the first 2 months. Free Tier usage is calculated monthly and is available for resources that exist AWS Free Tier benefits are associated with your AWS account, and any usage through service consoles or other interfaces also consumes your available Free Tier benefits. For example, if you launch an Amazon EC2 instance through SageMaker Unified Studio or other AWS interfaces, it will be the Amazon EC2 consumed in the AWS Free Tier period applies. To learn more, see the pricing details below.

Amazon SageMaker Notebooks

Amazon SageMaker notebooks offer data and AI teams a high-performance, serverless programming environment for analytics and machine learning (ML) workflows. Amazon SageMaker notebooks can be launched quickly, integrate with Apache Spark for distributed data processing, and come with built-in data connectors and ML algorithms. You can also configure notebooks to run on powerful compute resources with optional GPU acceleration and share notebooks with others using SageMaker Unified Studio.

You are charged for the instance type and storage you choose, based on the duration of use.

Amazon SageMaker Unified Studio Notebooks

Location Type: AWS Region

Region: Europe (London)

Select Compute Type for the Primary Node: Standard Instances

50 available instances

Instance Name	vCPU	Memory	price
ml.t3.medium	2	4 GiB	$0.067
ml.t3.large	2	8 GiB	$0.113
ml.t3.xlarge	4	16 GiB	$0.227
ml.t3.2xlarge	8	32 GiB	$0.453
ml.m7.large	2	8 GiB	$0.140
ml.m7.xlarge	4	16 GiB	$0.280
ml.m7.2xlarge	8	32 GiB	$0.559
ml.m7.4xlarge	16	64 GiB	$1.119
ml.m7.8xlarge	32	128 GiB	$2.238
ml.m7.12xlarge	48	192 GiB	$3.357
ml.m7.16xlarge	64	256 GiB	$4.476
ml.m7.24xlarge	96	384 GiB	$6.713
ml.m7.48xlarge	192	768 GiB	$13.427
ml.r8.large	2	8 GiB	$0.152
ml.r8.xlarge	4	16 GiB	$0.304
ml.r8.2xlarge	8	32 GiB	$0.607
ml.r8.4xlarge	16	64 GiB	$1.214
ml.r8.8xlarge	32	128 GiB	$2.427
ml.r8.12xlarge	48	192 GiB	$3.641
ml.r8.16xlarge	64	256 GiB	$4.854

Notebook storage

	Price
General Purpose SSD Storage	$0.1296 per GB-month of provisioned storage

Amazon SageMaker Data Agent

SageMaker Data Agent is an AI agent within SageMaker notebooks that accelerates data querying, exploratory data analysis, and machine learning model development. SageMaker Data Agent integrates with your data catalog, understands your business context, and helps you build ML models. It responds to requests to generate code that correctly references your tables, and columns. SageMaker Data Agent follows a pay-as-you-go pricing model. You are charged for the underlying compute and storage resources.

It builds a set of unit AI agents to assist you in designs. Simple prompts such as "Add a stage function or a graph that is quick algorithms, consume data from Amazon S3 and Amazon Redshift data warehouse, to help you with implementation pipeline or building an ML model. Querying data becomes more than 1 credit. Credits are metered for the second decimal points, so the minimum billable unit is 0.01 credits. Each AWS Free Tier account is provided 25 credits per month to use for the first 30 days.

For example, you can generate an entire data transformation pipeline in 4 to 5 minutes. SageMaker Data Agent is available in SageMaker notebooks and is available in the catalog. Note that it will automatically expand hours of manual coding. debugging incorrect table references, and tweaking an pipeline logic can be accomplished in minutes, or even seconds.

Data Agent Credit

Pricing dimension	Price
Data Agent Credit	$0.04 per credit

Lakehouse Architecture

The next generation of Amazon SageMaker is built on an open lakehouse architecture that is compatible with Apache Iceberg. It unifies data in Amazon S3 data lakes and Amazon Redshift data warehouses, helping you build powerful analytics and AI/ML applications on a single copy of data. You only pay for the resources you use, allowing for cost-effective scaling of your data and AI operations. To learn more, see Amazon SageMaker Lakehouse. Amazon SageMaker also integrates with the Data Zero-ETL and Amazon S3 tables. Depend on your choice of Amazon S1 or Amazon Redshift Managed Storage (RMS) for storage. Amazon Redshift Managed Storage (RMS) is available in the catalog. Note that it will automatically expand hours of manual coding. Debugging incorrect table references, and tweaking an pipeline logic can be accomplished in minutes, or even seconds.

Amazon SageMaker AI

SageMaker AI follows a pay-as-you-go pricing model with no upfront commitments or minimum fees. The key pricing dimensions for SageMaker AI include compute instances, storage volumes, data processed in and out, models inferences, Amazon SageMaker notebooks, Amazon Elastic Inference (EI) accelerators, and SageMaker AI Studio. You are charged for model deployment, and ML/Ops (Amazon SageMaker Pipelines and Model Monitor) and machine learning governance (Amazon SageMaker Model Cards and Model Dashboard) features are billing examples. The actual cost you pay based on factors such as AWS Regions, instance types, storage volumes, and inference options. To learn more access and detailed pricing information, consult SageMaker AI pricing.

Amazon Bedrock

Amazon Bedrock offers a flexible, consumption-based pricing model. You pay only for the Amazon Bedrock resources you use, without any minimum fees or upfront commitments. Pricing is based on the number of inputs and outputs that you process, the customization of your models, and the optional features, such as the Agents, Flows, Knowledge Bases, Guardrails, and Evaluations features and the models available through Amazon Bedrock. For the most accurate and detailed pricing information, consult Amazon Bedrock pricing.

SQL Analytics

SQL with Amazon Redshift offers a flexible, consumption-based pricing model. You pay only for the compute resources you use, without minimum fees or upfront commitments. Pricing is based on the number of nodes and instance types in your Amazon Redshift cluster. To learn more about Amazon Redshift and the variety of compute components, including compute and storage options, see Amazon Redshift pricing. Amazon Redshift Serverless charges for compute capacity used during workloads on a per-second basis. You pay only for the compute capacity you use when you are connected to your data warehouse. For the most accurate and detailed pricing information, consult Amazon Redshift pricing.

Amazon SageMaker Data Processing

SageMaker Data Processing, which brings together capabilities from Amazon Athena, Amazon EMR, AWS Glue, and Amazon Managed Workflows for Apache Airflow (Amazon MWAA), offers a flexible, pay-as-you-go pricing model without upfront commitments or minimum fees. You are charged for the number of data processing units (DPUs) and the duration of your workloads or jobs. DPU is a relative measure of processing power that consists of CPU, memory, and network resources, each of which can vary depending on factors such as AWS Regions, instance types, and usage patterns. To learn more about SageMaker Data Processing, see the Athena, Amazon EMR, AWS Glue, and Amazon MWAA pricing pages.

Amazon Q Developer

Amazon Q Developer is available within SageMaker Unified Studio to support development tasks like coding, testing, security scanning, resource optimization, and troubleshooting. Amazon Q Developer offers a free tier and Pro tier. For the most accurate and detailed pricing information, consult Amazon Q Developer pricing.

Amazon SageMaker Data and AI Governance

The next generation of SageMaker simplifies discovery, governance, and collaboration for data and AI across your lakehouse data, models, and applications. With SageMaker Catalog, built on Amazon DataZone, you can consistently define and enforce access policies using a simple permissions model with fine-grained controls.

SageMaker Catalog helps you govern your data and AI assets centrally and consistently across data. Pricing is based on four key dimensions: Catalog requests, Metadata storage, Governed data assets, and Data quality. Billing occurs monthly. Each account receives some free usage per billing month, including a set amount of metadata storage, API requests, and compute units. Additional usage is billed at standard rates. For example, if you store 2 GB of metadata, the first 1 GB may be free and 1 GB is billed at $0.30 per GB-month. The rates for each of the dimensions and discount tiers can be found below.

SageMaker Catalog free usage

SageMaker Catalog, built on Amazon DataZone, provides free usage as a limit so that you can explore its capabilities without having to worry about incurring charges. Each account where you create or access SageMaker Catalog resources gets the following free tier: 1,000 API requests, and 0.1 free compute units. You will only be charged for SageMaker Catalog resources if you go over any of these. If your account exceeds these free usage limits, you will be billed at the standard rates. For example, if you store 2 GB of metadata in SageMaker Catalog, and 1 month are provided free of charge and do not count towards the 10,000 free API requests.

Pricing dimension	Price	Free usage
Requests	$10 per 100,000 requests	10,000 free requests to Catalog-month per account. In addition to an API call that can provide free tier.
Metadata storage	$0.30 per GB	20 MB of free storage to Catalog-month per account
Compute	$ 1.176 per compute unit	0.1 free compute units in a Catalog-month per account
Recommendations	$0.01 per 1,000 input tokens/ $0.03 per 1,000 output tokens	No free usage

Pricing Dimensions Details and Examples

Metadata storage

You will be charged at standard rate of $0.3 per GB in a billing month for metadata storage if you exceed the free limit of 20 MB. Metadata storage includes the size of your business, dynamic structure and forms, metadata forms, projects, environments, users, and data assets. The table below provides a few scenarios to explain how metadata storage-based pricing works.

Metadata storage	Charges incurred (billing month)	Price	Description
20 MB	$0	$0.00 per GB, after 20 MB of free usage	20 MB of storage is free in a billing month
100 MB	$0.0117	$0.30 per GB, after 20 MB of free usage	Charges incurred = (100 MB - 20 MB) / (0.00 TB/GB * 1024 GB/TB)
1.00 (1024 MB)	$0.29	$0.30 per GB, after 20 MB of free usage	Charges incurred = (1024 MB - 20 MB) / (0.00 TB/GB * 1024

Requests

SageMaker Catalog provides certain core APIs such as CreateDomain, CreateProject, and CreateAsset are never charged. All the other APIs will be charged a standard rate of $10 per 100,000 requests. For example, if you make 1 million requests in a billing month, you will be billed for 900,000 request in an interaction with SageMaker Catalog, whether through the SageMaker Unified Studio, AWS Management Console, or directly via the Amazon DataZone API or SDK, an SageMaker Catalog API request is made. Actions performed in the SageMaker Unified Studio or Amazon DataZone portal, viewing data asset details, or creating a project.

Number of requests excluding free APIs (per account in a billing month)	Charges incurred (billing month)	Price	Description
6,000	$0	$10 per 100,000 requests included after 10,000 free usage	First 6,000 requests in a billing month are free
16,000	$0.07	$10 per 100,000 requests included after 10,000 free usage	Charges incurred = (16,000 - 10,000) / (0.00 TB/GB * $10
100,000	$0.63	$10 per 100,000 requests included after 10,000 free usage	Charges incurred = (100,000 - 10,000) / (0.00 TB/GB * $10

Free requests that do not incur any charges

The following requests are provided free of charge when used from SageMaker Unified Studio. See the full list of SageMaker Catalog APIs.

Request	Description	Free APIs
Domain management	These include API requests to create and manage Amazon DataZone domains.	CreateDomain, GetDomain, UpdateDomain, ListDomains, TagResource, UntagResource, GetTagResource, GetServiceAuthorization
Environment and Blueprints	These include all the API requests for working with environments and environment blueprints in Amazon DataZone.	CreateEnvironment, GetEnvironment, UpdateEnvironment, DeleteEnvironment, ListEnvironments, CreateEnvironmentBlueprint, GetEnvironmentBlueprint, UpdateEnvironmentBlueprint, DeleteEnvironmentProfile, GetEnvironmentProfile, ListEnvironmentProfiles, CreateEnvironmentBlueprintConfiguration, GetEnvironmentBlueprintConfiguration, UpdateEnvironmentBlueprintConfiguration, DeleteEnvironmentBlueprintConfiguration, ListEnvironmentBlueprintConfigurations
Project Management	These include all the API requests for working with projects and projects profiles in Amazon DataZone.	CreateProject, GetProject, UpdateProject, DeleteProject, ListProjects, CreateProjectMembership, DeleteProjectMembership, ListProjectMemberships, CreateProjectProfile, GetProjectProfile, UpdateProjectProfile, DeleteProjectProfile, ListProjectProfiles
User and Group Management	These include API requests for managing users and groups within your Amazon DataZone domains.	CreateUserProfile, GetUserProfile, UpdateUserProfile, SearchUserProfiles, CreateGroupProfile, GetGroupProfile, SearchGroupProfiles, SearchInventory, SearchTypes, SearchGroupProfiles
Policy and authorization	These include the API requests for creating and managing	AddPolicyGrant, RemovePolicyGrant, ListPolicyGrants

Compute

You will be charged at standard rate of $1.176 compute unit if you exceed the included 0.1 compute units in a billing month. Compute units are used to measure the compute capacity consumed when you run various functions from your data sources, including AWS Glue databases and Amazon Redshift. They are also used when you opt in for automatic synchronizations from 3rd to SageMaker Catalog. Compute units are billed for both data ingestion and publishing and synchronizations. The table below provides a few scenarios to explain how compute-based pricing works.

Compute units (per account in a billing month)	Charges incurred (billing month)	Price	Description
0.2	$0	$1.176 per compute unit, after 0.2 compute units of free usage	0.2 compute units free in a billing month
1	$1.42	$1.176 per compute unit, after 0.2 compute units of free usage	Charges incurred = (1 - 0.2 MB) / (1.176 * 0 sec
10	$11.48	$1.176 per compute unit, after 0.2 compute units of free usage	Charges incurred = (10 - 0.2 MB) + (1.176 * $1.176

Recommendations

SageMaker Catalog provides a generative AI-based capability to improve data discovery, data understanding, and data usage by enriching the business data catalog. In a single step, data producers can generate comprehensive business descriptions for data assets and their schemas. Data consumers get AI recommendations in an end-to-end user. You will be charged for input and output tokens for generating asset business names, column descriptions, asset descriptions, and inferring terms based on the number of tokens processed. Tokens are a basic unit that represent a few characters of text. Input tokens are the number of tokens provided to the prompt and used to generate a response. Output tokens are the number of tokens generated in response. To calculate the recommendation cost that SageMaker Catalog provides has to you and are charged at $0.001 per 1,000 tokens.

Next Steps

© 2026, Amazon Web Services, Inc. or its Affiliates. All rights reserved.