Startup guide to data cost optimization

How startups can scale and save using data services on AWS

Table of contents

Introduction to data cost optimization

Cost optimization is a top-of-mind consideration for startups. Meaningful savings can be achieved with a wide variety of techniques, but how your startup tackles cost optimization is dependent on the stage of its growth. Many startups are laser-focused on product development, which can mean a choice between time spent building extra functionality or time spent managing costs. Should you reorganize your account structures or build cost-analytics pipelines? Either way, prioritizing low effort, high-impact architectural changes can help maintain your momentum.

Amazon Web Services (AWS) helps business decision-makers control their startup costs by continuously optimizing spend with a scalable data strategy. We meet startups where they are on their data journey and help them start anywhere with an end-to-end data strategy.

The breadth of AWS services and pricing options provides you with the flexibility to manage cost while delivering the capacity your business requires. With an end-to-end data strategy on AWS, startups can rapidly build scalable data lakes, use a broad and deep collection of purpose-built data services, ensure compliance via unified data access, security, and governance, and scale their systems at a low cost without compromising performance—all to inform better decisions with speed and agility at scale.

In this eBook, we will share easy-to-implement cost-optimization techniques that can help you achieve the best data strategy while controlling spend. Get back to developing features that drive real value for your customers. Leverage the following insights to extend your runway with a long-term strategy for cost management.

Embracing the Well-Architected Framework

As a startup founder, your days are filled with competing priorities: You want to focus your time on innovation and product-market fit—not worry about underlying infrastructure or unexpected costs.

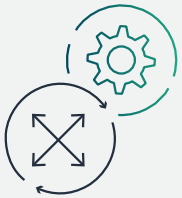
To alleviate distractions and gain the freedom and control necessary to drive innovation, it is important that you lay the right foundational bricks for your startup's architecture framework—especially your data analytics application workloads. Not having a Well-Architected Framework can lead to unruly costs and delays in product development that can negatively impact your startup's bottom line and growth.

Understanding how to design, deploy, and architect analytics application workloads in the AWS Cloud can help optimize cost and scale your startup to new heights. With a Well-Architected Framework for data analytics workloads, you can extend your runway and create a long-term strategy for cost management of your data infrastructure.

The cost-optimization pillar within the Well-Architected Framework includes the continual process of refinement and improvement of a system over its entire lifecycle to optimize cost, from the initial design of your first proof of concept to the ongoing operation of production workloads. It's a yearslong, continual process.

Choose the right solution and pricing model. Build cost-aware systems that allow you to achieve business outcomes and minimize costs. To perform cost optimization over time, you need to identify data, infrastructure resources, and analytics jobs that can be removed or downsized.

Embracing the Well-Architected Framework (cont'd)



Operational excellence

Run, manage, and monitor production workloads to deliver business value and continuous improvements on supporting processes and events.

Security

Protect information, systems, and assets, while delivering value through risk assessments and mitigation strategies.

Reliability

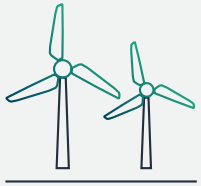
Auto-recover workloads from infrastructure, power, or system failures with dynamic resource management to meet operational thresholds.

Performance efficiency

Use computing resources efficiently to support on-demand changes for delivering workloads with maximum performance to meet SLAs.

Cost optimization

Avoid and eliminate unneeded costs or replace resources with cost-effective resources without impacting best practices and business needs.



Sustainability

Evaluate the design, architecture, and implementation of workloads to reduce their energy consumption and improve their efficiency to support a more sustainable future.

Learn more about the Well Architected Framework ›

Cost optimization as an ongoing process

Cost optimization is the continual process of refinement and improvement of a system over its entire lifecycle—from the initial design of your first proof of concept to the ongoing operation of production workloads. It's a yearslong, continual process. It's why choosing the right solution and pricing model for your startup is so important. You and your team need to build cost-aware systems that allow your startup to achieve its business objectives while minimizing costs. To perform cost optimization over time, you need to identify data, infrastructure resources, and analytics jobs that can be removed or downsized.

There is great value in determining analytics workflow costs at each individual data processing step or individual pipeline branch. The benefit of understanding analytics workflow costs at this granular level will help you decide where to focus engineering resources for development. It can also help you perform return-on-investment (ROI) calculations for the analytics portfolio as a whole.

To ensure that you always have the most cost-efficient workloads, periodically review your workloads to discover opportunities to implement new services, features, and components. It is common for analytics workloads to have an ever-growing number of users and an exponential growth of data volume. You should implement a standardized process across your organization to identify and remove unused resources, such as unused data, infrastructure, and ETL (extract, transform, and load) jobs.

Six ways to optimize the cost of your data strategy

How do you select the right compute and storage solution for your data analytics workload for your startup? Your initial design choice could have a significant cost impact. Understand the resource requirements of your workload, including its steady state and spikiness, and then select the solution and tools that meet your requirements. Avoid over-provisioning to allow more cost-optimization opportunities.

With those objectives in mind, here are six best practices that can help you optimize the costs of your data strategy:

1. **Optimize the cost of information infrastructure at scale ›**
2. **Decouple storage data from compute data ›**
3. **Plan and provision capacity for predictable workload usage ›**
4. **Use On-Demand Instances capacity for unpredictable workload usage ›**
5. **Avoid data duplication ›**
6. **Build your data strategy leveraging AWS Activate credits ›**

1: Optimize the cost of information infrastructure at scale

A startup's ability to scale quickly means that it can make decisions faster, stay agile with rapidly changing business needs, and keep up with customer demand. To keep up with the growing volume and velocity of data and make data-driven decisions in real time, you need to store and query your data with the most scalable, highest-performing data services. AWS offers the most scalable, highest-performing services for storing, querying, and acting on data. And we're continuously optimizing to give you the best price performance. AWS provides databases that are designed for the best price performance for your use case at scale.

Amazon Redshift is the fastest and most economical data warehouse in the cloud, with up to three times better price performance than other cloud data warehouses. AWS infrastructure innovation benefits you across data services. Many of our data services enable you to run workloads on **AWS Graviton** instances with little or no code change to get better price performance. These services include **Amazon Relational Database Service** (Amazon RDS) with up to 50 percent better price performance, **Amazon EMR** with up to 30 percent lower cost, up to 15 percent improved performance for Spark, and up to 20 percent performance improvement for **Amazon Aurora**. With the **AWS Nitro System**, **Amazon Elastic Compute Cloud** (Amazon EC2) instances can deliver a more than 15 percent higher throughput performance on some workloads as compared to other major cloud providers running the same CPU, so applications accessing your databases can run even faster, improving time to insight for analytics.

up to 50%
better price performance with Amazon RDS

up to 30%
lower cost with Amazon EMR

up to 20%
performance improvement for Amazon Aurora

2: Decouple storage data from compute data

It's common for data assets to grow exponentially year over year. However, your compute needs for your startup might not grow at the same rate. Decoupling storage data from compute data allows you to manage the cost of storage and compute separately. That frees you to implement different cost-optimization features to minimize cost. In order to do this, you must:

Compress data when applicable

Most analytics services can access compressed data. Compressing your data and storing it in columnar formats can improve query speed and reduce the cost of query transactions.

Consider data temperature when choosing data store and storage class

Tiered storage transitions datasets to different storage tiers based on their usage patterns to optimize storage cost. **Amazon Simple Storage Service** (Amazon S3) offers a variety of storage classes that help you to optimize costs over time. Consider a cost-saving lifecycle plan that migrates data from S3 Standard to S3 Infrequent Access (or, in some cases, Amazon S3 One Zone-Infrequent Access, *S3 One Zone-IA*) to Amazon Simple Storage Service Glacier (Amazon S3 Glacier).

Use low-cost compute resources, such as Spot Instances, when applicable

Decoupling storage data from compute data allows stateless logic implementation, but consider using **Amazon EC2** Spot Instances for additional cost savings. Spot Instances are also an excellent option for reducing the cost of development or quality assurance (QA) applications.

Deploy compute close to data to reduce data transfer costs

In AWS, data transfer charges apply to data transferred out of AWS to the internet or over AWS Direct Connect, data transferred between AWS Regions, and data transferred between Availability Zones. Data transfer into AWS and data transfer within an Availability Zone are at no cost.

Use Amazon S3 Select and Amazon S3 Glacier Select to reduce data retrieval

Amazon S3 Select and Amazon S3 Glacier Select allow applications to retrieve only a subset of data from an object by using simple SQL expressions.

3: Plan and provision capacity for predictable workload usage

For well-defined data workloads, planning capacity in advance (based on average usage patterns) helps improve resource utilization and eliminates over-provisioning. For a spiky workload, set up automatic scaling to meet user and workload demand. This can be done by:

Choosing the right instance type based on workload pattern and growth ratio

Consider resource needs, such as CPU, memory, and networking, that will meet the performance requirements of your workload. Choose the right instance type to avoid over-provisioning. An optimized Amazon EC2 instance runs your workloads with optimal performance and infrastructure cost. For example, choose the smaller instance if your growth ratio is low, as this allows more granular incremental change.

Deploying rightsizing based on average or medium workload usage

Rightsizing is the process of matching instance types and sizes to your workload performance and capacity requirements at the lowest possible cost. It's also the process of looking at deployed instances and identifying opportunities to downsize without compromising capacity or other requirements, which will result in lower costs.

Using automatic scaling capabilities to meet peak demand

Many analytics system components can scale dynamically to meet demand and reduce over-provisioning. When possible, architect your analytics system to use horizontal scaling instead of vertical scaling.

4: Access capacity on demand for unpredictable workloads

Serverless solutions are either charged by the amount of data processed or the amount of compute resources used—but only when there is a workload actively running. Compare this approach to using long-running infrastructure but where you don't need to pay for idle resources.

Use Amazon Athena for ad hoc SQL workloads

Amazon Athena is a serverless query service that makes it easy to analyze data directly in Amazon S3 using standard SQL. With Amazon Athena, you only pay for the queries that you run. You are charged based on the amount of data scanned per query.

Use AWS Glue instead of Amazon EMR for infrequent ETL jobs

AWS Glue is a fully managed ETL service that makes it simple and cost-effective to categorize your data, clean it, enrich it, and move it reliably between various data stores and data streams. With AWS Glue, you pay only for the resources used during the ETL process. In contrast, **Amazon EMR** is typically utilized for frequently running jobs requiring semi-persistent data storage.

Use on-demand resources for transient workloads or short-term development and testing needs

Choose on-demand resources such as Amazon EMR on EC2 Spot Instances to perform experimental or test data processing with low cost. During the initial planning and testing phases of analytics workloads, using on-demand resources provides flexibility to iterate through many alternative architectures until all aspects of a well-architected analytics project are optimized for the job. This includes selecting the right instance type or using a serverless architecture to meet the required time duration of analytics pipeline runs.

5: Avoid data duplication with a centralized storage layer

Cost-conscious teams want to share data at many levels to allow for broad and deep insights but also to minimize complexity and expense. For example, data needs to be shared from a central data warehouse that loads and transforms constant streams of updates with business intelligence (BI) and analytics clusters that serve a variety of workloads, such as dashboard applications, ad hoc queries, and data science. However, unloading data from one system (the producer) and copying it into another (the consumer) can be expensive and introduce delays, especially as the number of consumers grows. This approach requires building and maintaining separate ETL jobs to provide relevant subsets of the data for each consumer. The complexity increases with the security practices that must be followed to monitor business-critical data usage regularly and ensure compliance. Whenever possible, analytics workloads should adopt technical solutions that allow data sharing without data duplication. Remember to do the following when making this shift for your startup:

Implement a central storage layer to share data among tenants

- A centralized storage layer, such as a data lake architecture, can make the same data resources available to different teams.
- Security isolations can be enforced so that only relevant subsets of data are made available to teams that must have access to those portions of the data lake.
- Use a market-proven solution, such as AWS Lake Formation, to build a secure data lake. Data lakes on AWS help you break down data silos to maximize end-to-end data insights—all with the best price performance. Hundreds of thousands of data lakes run on AWS.

6: Leverage up to $100,000 in AWS Activate credits*

As a startup founder, you can join the AWS Activate program and enjoy its benefits anytime. Depending on your startup stage, you may qualify for up to $100,000 in AWS Activate credits*, as well as Activate exclusive offers including reduced cost or free access to third-party services, products, and memberships, valued up to $800,000.

Plus, when you're approved for credits, you will also gain access to more saving opportunities, as outlined below.

- **Exclusive offers:** If you have AWS Activate credits, you also unlock access to exclusive offers from trusted organizations such as **Notion**, **Miro**, **HubSpot**, and **Stripe**. The discounts, free trial periods, credits, and other startup-friendly perks are all designed to help your startup grow and succeed.
- **Cost-optimization tools:** If you have AWS Activate credits, your AWS Activate Console will help you optimize spend with a Cost and Credits Summary. The summary gives you real-time visibility into your balances, alerts you of pending credit expiration, and offers tailored cost-optimization tips to help you make the most of your spend. Don't rush this step; we want you to build at your own pace. When you are ready to work on your data architecture, you can apply for the credits that can help offset that cost.

Learn more about AWS Activate ›

* All AWS credits are in USD and subject to the AWS Promotional Credit Terms & Conditions.

Meet the startups that are successfully optimizing data spend

Discover how startups at every stage reduced their spend by leveraging the AWS cost-optimization pillar with an end-to-end data strategy

TextNow saves 93% on data transfer fees with AWS PrivateLink

The challenge

Founded in 2009 by two university students, mobile app TextNow is on a mission to connect customers through communications technology for free or as close to free as possible. When the COVID-19 pandemic hit in March 2020, companies began to cut digital advertising spending. “Our advertising business is our main revenue driver,” says Andres Ivanov, software technical manager at TextNow. “We needed a plan that would help us regain profit.”

The AWS solution

The company began exploring how it could save money on AWS and found an opportunity to reduce data transfer fees by using **AWS PrivateLink** Service Ready Partner **Datadog**. AWS PrivateLink enables businesses to establish private connectivity between virtual private clouds and services hosted on AWS or on premises without exposing data to the public internet.

The results

By securing traffic over a private network, TextNow was able reduce data transfer fees significantly while increasing data security. In the US East AWS Regions, for example, each gigabyte of data it transfers now costs $0.01 instead of $0.09.

Read the full story ›

textnow

“Using AWS PrivateLink, we reduced our data transfer fees by 93%. That’s amazing for us and a key driver for what we’re doing with this new solution.”

Andres Ivanov, Software Technical Manager, TextNow

DCI reduces cloud costs by 27%, enables long-term growth on AWS

The challenge

Digital Commerce Intelligence (DCI) provides market and competitive insights, allowing its customers to build data-driven strategies. As DCI grew, its cloud services provider lacked the flexibility it needed, resulting in unpredictable compute and database costs. An investor suggested the company migrate to AWS to avoid the kind of billing issues it had with its previous cloud provider.

The AWS solution

The company migrated to AWS in just six months using the AWS Startup Program. DCI's participation in **AWS Activate**—which offers free tools and resources to startups—helped the business move fast. Guided by its account team and AWS support engineers, DCI was able to migrate its data collection tools, SQL Server, messaging queue, Kubernetes clusters, image registry, and compute to AWS.

The results

So far, the migration to AWS has reduced monthly IT costs by 27 percent. Those savings matter because—to run its algorithms and deliver results to its customers—DCI needs to ingest and process a lot of data. These results give DCI customers the market insights they need to run their businesses more intelligently.

Read the full story ›

DIGITAL COMMERCE INTELLIGENCE

"Lower costs mean we can spend more on people and on product development—things that make the business more competitive."

Konstantinos Kitsaras, CTO, Digital Commerce Intelligence

SkyTouch Technology gains agility and cost benefits by migrating its business-critical apps to the AWS Cloud

The challenge

A division of Choice Hotels, SkyTouch Technology offers the industry's most widely used cloud-based hotel property management system. The hotel industry was hit hard by the COVID-19 pandemic. With occupancy rates dropping more than 30 percent, one of SkyTouch's biggest challenges was how to save on operating costs.

The AWS solution

SkyTouch migrated its critical business applications to AWS, which was a major undertaking. Not only did SkyTouch need to minimize the impact of downtime for its customers, but it also had to plan for numerous integrations between systems, customers, and third parties. The **AWS Migration Acceleration Program** (MAP) for Windows helped SkyTouch reach migration goals even faster with best practices, tools, and incentives.

The results

Since partnering with AWS in 2018, downtime for the SkyTouch property management system was reduced by 77 percent year over year. The company has also successfully optimized infrastructure costs by implementing AWS Auto Scaling and taking advantage of AWS Savings plans and AWS Reserved Instances. Those efforts have reduced data costs by 35 percent when compared to on-demand pricing.

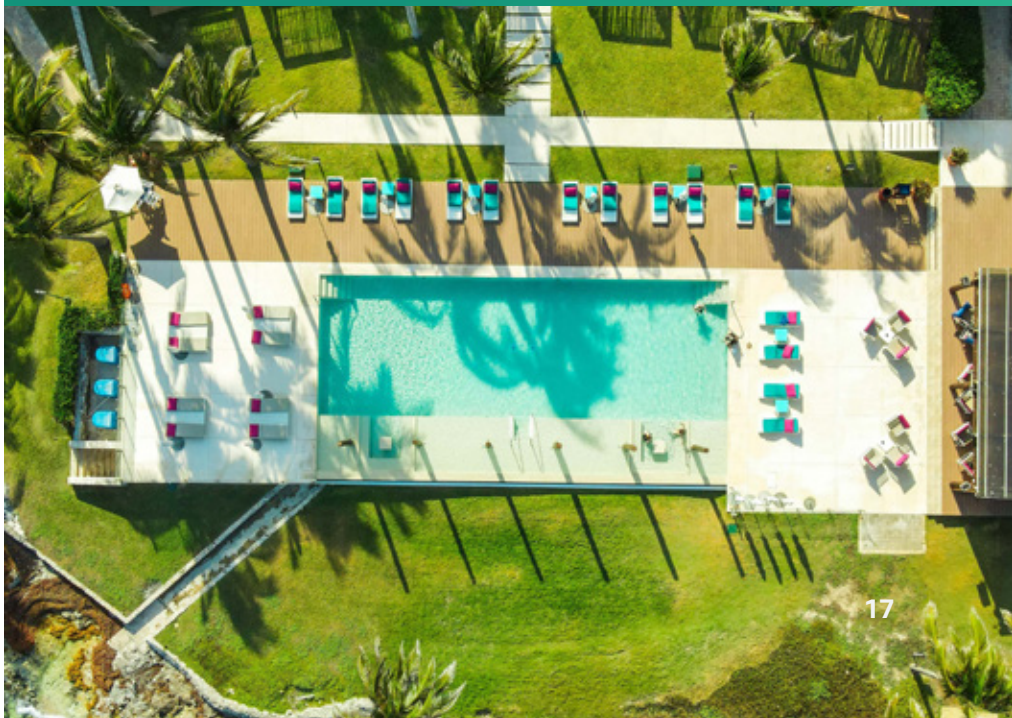
Read the full story ›

aws

SKYTOUCH TECHNOLOGY

"Our demand is highly cyclical. Provisioning infrastructure that can easily scale up and down allows us to match capacity to demand. This flexibility also significantly reduces the overhead of right-sizing infrastructure resources."

Steve Garvey, Director of Cloud Platform Engineering, SkyTouch Technology



Majelan uses Amazon Athena to understand user behavior and control costs

The challenge

Launched in 2019, Majelan provides free access to more than 18 million podcasts from 50 countries. The intelligent application collects granular data on how content is listened to, with the aim of making recommendations and improving programming. At first, Majelan used Amazon Elasticsearch Service and its own servers to perform the analysis. Growth in the number of users and an increase in the use of the platform meant it was hitting the limits of these services in terms of data storage and analysis.

The AWS solution

Majelan needed a high-performance solution capable of analyzing a large volume of data. After sharing its technical needs with both AWS and Google Cloud Platform (GCP), the startup finally chose AWS and **Amazon Athena**.

The results

Amazon Athena makes it possible to run SQL queries on large volumes of data and reduce query time from 10–15 minutes to a few seconds. The solution has also provided Majelan with vastly improved control of costs. "If we don't do calculations and data processing, we only pay Athena for the relatively low cost of storage," notes Thomas Fillon, data director at Majelan.

Read the full story ›

majelan

"Using indexing robots that crawl through S3 files and provide Athena with a structured view means the service does not have to read 200 million events with each request. This is important because Athena's cost is linked to the amount of data read to perform the query. Limiting costs can be done by making intelligent queries that don't use all the data, for example, by partitioning files by date, user, or type of use."

Thomas Fillon, Data Director, Majelan

Apify powers web insights, cuts cloud costs by 25% with AWS

The challenge

Apify, a Prague-based startup founded in 2015, has developed a web-scraping and automation platform and set of open-source tools that help businesses improve their operations by collecting and analyzing large volumes of web data and automating web processes. As a startup, Apify needed a reliable, cost-effective infrastructure that could easily scale as customer demand grew.

The AWS solution

Apify began using AWS through **AWS Activate**, a program that offers startups free tools and resources. Apify's approach means that it needs to process vast amounts of information fast so customers can reliably access and analyze web data. It processes 20 million web automation jobs monthly using **Amazon EC2** and **Amazon Elastic Kubernetes Service** (Amazon EKS), a managed container service to run and scale Kubernetes applications in the cloud or on premises.

The results

By building its offering on AWS, the company has grown rapidly and now provides services to 1,000 organizations in 179 countries. It has also scaled to process 1,000 terabytes of data a month and reduced its total cloud costs by 25 percent using Amazon EC2 Spot Instances, which run fault-tolerant workloads at a discount of up to 90 percent.

Read the full story ›

APIFY

"The support we received through AWS Activate for credits, training, and cost control was a key reason we chose AWS. It undoubtedly contributed to our early success."

Marek Trunkat, CTO, Apify

Next steps

Optimizing the cost of your data strategy should be the number-one priority for your startup so that you and your team can continue investing time where your business needs it the most.

AWS enables you to take control of costs and continuously optimize your data spend while building modern, scalable applications to meet your needs. AWS provides a breadth of services and pricing options that offer the flexibility to effectively manage your costs and maintain the performance and capacity you require so your startup can achieve its highest saving potential.

To get started on AWS, visit AWS Data for Startups ›