



AWS Machine Learning Blog

Lowering total cost of ownership for machine learning and increasing productivity with Amazon SageMaker

by Kimberly Madia | on 11 FEB 2020 | in [Artificial Intelligence](#), [SageMaker](#) | [Permalink](#) | [Comments](#) | [Share](#)

You have many choices for building, training, and deploying machine learning (ML) models. Weighing the financial considerations of different cloud solutions requires detailed analysis. You must consider the infrastructure, operational, and security costs for each step of the ML workflow, as well as the size and expertise of your data science teams.

The Total Cost of Ownership (TCO) is often the financial metric that you use to estimate and compare ML costs. This post presents a TCO analysis for [Amazon SageMaker](#), a fully managed service to build, train, and deploy ML models. The findings show that the TCO over a three-year horizon is 54% lower compared to other cloud-based ML options such as self-managed [Amazon EC2](#) and AWS managed [Amazon EKS](#). The analysis ranged from small teams of five data scientists and extra-large teams with 250 data scientists and found that Amazon SageMaker provides a better TCO across teams of all sizes.

Analysis findings

The following table summarizes the findings. For the full TCO analysis, see [The total cost of ownership of Amazon SageMaker](#).

<u>OVERALL SUMMARY</u>		Amazon SageMaker 3-year TCO Savings	
		Compared to EC2	Compared to EKS
Small scenario	5 data scientists	-90%	-90%
Medium scenario	15 data scientists	-87%	-85%

Large scenario	50 data scientists	-79%	-65%
X-Large Scenario	250 data scientists	-77%	-54%

Typically, the TCO for Amazon SageMaker is lower in the first year compared to the EC2 or EKS options because you must spend more on building security and compliance, which come out-of-the-box in Amazon SageMaker. The TCO for Amazon SageMaker continues to remain significantly lower over time because Amazon SageMaker optimizes infrastructure usage automatically and doesn't require upkeep of security and compliance features.

The TCO analysis evaluated infrastructure (compute, storage, and network), operational, and security costs for each step of the ML workflow and company size (small, medium, large, extra-large.) During ML model building, you incur costs to explore and preprocess data and experiment with ML frameworks and algorithms. During training, you incur costs for training tools and processes as well as tuning ML model hyperparameters. Finally, during ML model deployment, you incur costs as your model makes inferences on unseen data. Across each step of the workflow, the analysis factors in the costs to employ engineers. It also evaluates security costs, which span all three phases of the ML workflow. Security includes the costs to secure ML workloads, achieve compliance with regulatory standards, and to maintain security and compliance on an ongoing basis.

ML costs can vary depending on the type of model you pick. This TCO analysis isn't based on one particular ML framework, algorithm, or model. Instead, it takes a common mix of both ML and deep learning models seen in production across AWS customers.

One reason Amazon SageMaker has a strong TCO is because it is a fully managed service. You don't need to build, manage, or maintain any infrastructure or tooling to support ML. Amazon SageMaker also runs your model on auto-scaling clusters that are spread across multiple Availability Zones to deliver both high performance and high availability. Because you pay for storage and network based on your usage, costs are controlled. In addition, Amazon SageMaker has built-in security and compliance for ML workloads, so you don't need to invest in additional security.

With self-managed ML with EC2, you take on the responsibility of provisioning and managing EC2 instances, including instance failure recovery, patching, automatic scaling, and building and maintaining required security and compliance. You can use the [AWS Deep Learning AMIs](#) with the ML frameworks and libraries pre-built, but you still need to optimize data access to get high throughput and also optimize your setup for scale and to enable distributed training. In addition, you need to build and maintain the required security and compliance features for your ML workloads.

With managed Kubernetes on AWS, services such as EKS make it easy to deploy, manage, and scale containerized workloads on EC2. However, you need to take on the additional cost overhead of managing your own cluster, tuning the performance and usage based on the memory, compute, and network requirements for your workloads. In addition, you need to build the right level of security, compliance, and availability for your ML workloads.

In addition to lower TCO, Amazon SageMaker's productivity features enable you to put ML ideas into production faster and improve data scientist productivity by up to 10 times. One of the most significant sources of

productivity gains is from Amazon SageMaker Studio. 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 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, which makes you much more productive. You can perform all ML development activities, including notebooks, experiment management, automatic model creation, debugging and profiling, and model drift detection, within the unified SageMaker Studio visual interface.

Testimonials

The following are some productivity gains from Amazon SageMaker customers.



Coinbase uses ML models on Amazon SageMaker to help with fraud prevention, identity verification, and large-scale compliance. **Using Amazon SageMaker, Coinbase reduced model training times from 20 hours to 10 minutes.**



Intuit developed ML models that can pull a year's worth of bank transactions to find deductible business expenses for customers. **Using Amazon SageMaker, Intuit reduced ML deployment time by 90%, from six months to one week.**



Using Amazon SageMaker, NuData Security prevents credit card fraud by analyzing anonymized user data to detect anomalous activity before a fraudulent transaction occurs. **With Amazon SageMaker, NuData reduced ML development time by 60%, simplified their ML architecture by 95%, and worked with a large bank to passively block nearly 100% of fraudulent attempt traffic within the bank's consumer friction tolerance.**



Using Amazon SageMaker, Voodoo can decide in real time which ad to show to their players and invoke their endpoint over 100 million times by over 30 million users daily, representing close to a billion predictions per day. **With AWS machine learning, Voodoo put an accurate model into production in less than a week, supported by a small team, and has built on top of it continuously as their team and business grow.**



Using TensorFlow on Amazon SageMaker, Siemens Financial Services developed an NLP model to extract critical information to accelerate investment due diligence, **reducing time to summarize diligence documents from 12 hours to 30 seconds.**

Celgene uses Apache MXNet on Amazon SageMaker for toxicology prediction to analyze biological impacts of potential drugs virtually, without putting patients at



risk. **A model that previously took two months to train can now be trained in four hours.**



ADP uses AWS ML, including Amazon SageMaker, to quickly identify workforce patterns and predict outcomes before they happen, such as employee turnover or the impact of an increase in compensation. **ADP reduced the time to deploy ML models from two weeks to just one day.**

Conclusion

Amazon SageMaker is a fully managed ML service that lets you build, train, tune, and deploy models at scale. The total cost of ownership of Amazon SageMaker over a three-year horizon is over 54% lower compared to other cloud options and developers can be up to 10 times more productive.

For the full TCO analysis, see [The total cost of ownership of Amazon SageMaker](#).

You can experience the benefits of Amazon SageMaker first hand! Get started by logging into the [Amazon SageMaker console](#).

About the Author



Kimberly Madia is a Principal Product Marketing Manager with AWS Machine Learning. Her goal is to make it easy for customers to build, train, and deploy machine learning models using Amazon SageMaker. For fun outside work, Kimberly likes to cook, read, and run on the San Francisco Bay Trail.

TAGS: [Amazon SageMaker](#)

Resources

[Getting Started](#)







[What's New](#)

[Top Posts](#)

[Official AWS Podcast](#)

[AWS Case Studies](#)

Follow

-  [Twitter](#)
-  [Facebook](#)
-  [LinkedIn](#)
-  [Twitch](#)
-  [RSS Feed](#)
-  [Email Updates](#)

Related Posts

[Creating a Seamless Handoff Between Amazon Pinpoint and Amazon Connect](#)

[Winners of AWS Machine Learning Research Awards announced](#)

[Machine Learning: Avoiding Garbage in, Garbage Out](#)

[175+ Customers Achieve Machine Learning Success with AWS's Machine Learning Solutions Lab](#)

[Mission: Technology-enabled, sustainable agriculture](#)

[Simplify Machine Learning Inference on Kubernetes with Amazon SageMaker Operators](#)

[AI Explorations of a Former Media CTO \(With No Prior Machine Learning Experience\)](#)

[Updating your addresses with Amazon Connect and Amazon Lex](#)