CASE STUDY IN CLOUD COMPUTING

Amazon Web Services

Amazon Web Services, Inc. (AWS) is a subsidiary of Amazon that provides on-demand cloud computing platforms and APIs to individuals, companies, and governments, on a metered pay-as-you-go basis. These cloud computing web services provide distributed computing processing capacity and software tools via AWS server farms. One of these services is Amazon Elastic Compute Cloud (EC2), which allows users to have at their disposal a virtual cluster of computers, available all the time, through the Internet. AWS's virtual computers emulate most of the attributes of a real computer, including hardware central processing units (CPUs) and graphics processing units (GPUs) for processing; local/RAM memory; hard-disk/SSD storage; a choice of operating systems; networking; and pre-loaded application software such as web servers, databases, and customer relationship management (CRM).

Amazon Personalize

Amazon Personalize is an ML service that makes it easy for developers to create individualized recommendations for customers using their applications.

ML is increasingly used to improve customer engagement by powering personalized product and content recommendations, tailored search results, and targeted marketing promotions. However, developing the ML capabilities necessary to produce these sophisticated recommendation systems has been beyond the reach of most organizations today due to the complexity of developing ML functionality. Amazon Personalize allows developers with no prior ML experience to easily build sophisticated personalization capabilities into their applications, using ML technology perfected from years of use on Amazon.com.

With Amazon Personalize, you provide an activity stream from your application – page views, signups, purchases, and so forth – as well as an inventory of the items you want to recommend, such as articles, products, videos, or music. You can also choose to provide Amazon Personalize with additional demographic information from your users such as age, or geographic location. Amazon Personalize processes and examines the data, identifies what is meaningful, selects the right algorithms, and trains and optimizes a personalization model that is customized for your data.

Amazon Personalize offers optimized recommenders for retail and media and entertainment that make it faster and easier to deliver high-performing personalized user experiences. Amazon Personalize also offers intelligent user segmentation so you can run more effective prospecting

campaigns through your marketing channels. With our two new recipes, you can automatically segment your users based on their interest in different product categories, brands, and more.

All data analyzed by Amazon Personalize is kept private and secure, and only used for your customized recommendations. You can start serving your personalized predictions via a simple API call from inside the virtual private cloud that the service maintains. You pay only for what you use, and there are no minimum fees and no upfront commitments.

Amazon Personalize is like having your own Amazon.com ML personalization team at your disposal, 24 hours a day.

MACHINE LEARNING SERVICES

1. Amazon Rekognition

Amazon Rekognition makes it easy to add image and video analysis to your applications using proven, highly scalable, deep learning technology that requires no ML expertise to use. With Amazon Rekognition, you can identify objects, people, text, scenes, and activities in images and videos, as well as detect any inappropriate content. Amazon Rekognition also provides highly accurate facial analysis and facial search capabilities that you can use to detect, analyze, and compare faces for a wide variety of user verification, people counting, and public safety use cases.

With Amazon Rekognition Custom Labels, you can identify the objects and scenes in images that are specific to your business needs. For example, you can build a model to classify specific machine parts on your assembly line or to detect unhealthy plants. Amazon Rekognition Custom Labels takes care of the heavy lifting of model development for you, so no ML experience is required. You simply need to supply images of objects or scenes you want to identify, and the service handles the rest.

2. Amazon SageMaker

With Amazon SageMaker, you can build, train, and deploy ML models for any use case with fully managed infrastructure, tools, and workflows. SageMaker removes the heavy lifting from each step of the ML process to make it easier to develop high-quality models. SageMaker provides all of the components used for ML in a single toolset so models get to production faster with much less effort and at lower cost.

3. Amazon SageMaker Canvas

Amazon SageMaker Canvas expands access to ML by providing business analysts with a visual point-and-click interface that allows them to generate accurate ML predictions on

their own — without requiring any ML experience or having to write a single line of code.

4. Amazon SageMaker Studio Lab

Amazon SageMaker Studio Lab is a free ML development environment that provides the compute, storage (up to 15GB), and security—all at no cost—for anyone to learn and experiment with ML. All you need to get started is a valid email address—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 in between sessions. It's as easy as closing your laptop and coming back later.

5. Amazon SageMaker Data Labeling

Amazon SageMaker provides data labeling offerings 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.

AZURE

Microsoft Azure, formerly known as Windows Azure, is Microsoft's public cloud computing platform. It **provides a broad range of cloud services, including compute, analytics, storage and networking.**

The Azure platform aims to help businesses manage challenges and meet their organizational goals. It offers tools that support all industries -- including e-commerce, finance and a variety of Fortune 500 companies -- and is compatible with open source technologies. This gives users the flexibility to use their preferred tools and technologies. In addition, Azure offers four different forms of cloud computing: infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS) and serverless functions.

Microsoft charges for Azure on a pay-as-you-go (PAYG) basis, meaning subscribers receive a bill each month that only charges them for the specific resources and services they have used.

How does Microsoft Azure work?

Once customers subscribe to Azure, they have access to all the services included in the Azure portal. Subscribers can use these services to create cloud-based resources, such as VMs and databases. Azure resources and services can then be assembled into running environments used to host workloads and store data.

Key service capabilities for the full machine learning lifecycle

Data labeling

Create, manage, and monitor labeling projects, and automate iterative tasks with machine learning–assisted labeling.

Data preparation

Quickly iterate on data preparation at scale on Apache Spark clusters within Azure Machine Learning, interoperable with Azure Synapse Analytics.

Collaborative notebooks

Maximize productivity with IntelliSense, easy compute and kernel switching, and offline notebook editing. Launch your notebook in Visual Studio Code for a rich development experience, including secure debugging and support for Git source control.

Automated machine learning

Rapidly create accurate models for classification, regression, time-series forecasting, natural language processing tasks, and computer vision tasks. Use model interpretability to understand how the model was built.

Drag-and-drop machine learning

Use machine learning tools like designer for data transformation, model training, and evaluation, or to easily create and publish machine learning pipelines.

Reinforcement learning

Scale reinforcement learning to powerful compute clusters, support multiple-agent scenarios, and access open-source reinforcement-learning algorithms, frameworks, and environments.

Responsible building

Get model transparency at training and inferencing with interpretability capabilities. Assess model fairness through disparity metrics and mitigate unfairness. Improve model reliability and identify and diagnose model errors with the error analysis toolkit. Help protect data with differential privacy.

Experimentation

Manage and monitor runs or compare multiple runs for training and experimentation. Create custom dashboards and share them with your team.

Registries

Use organization-wide repositories to store and share models, pipelines, components, and datasets across multiple workspaces. Automatically capture lineage and governance data using the audit trail feature.

Git and GitHub

Use Git integration to track work and GitHub Actions support to implement machine learning workflows.

Managed endpoints

Use managed endpoints to operationalize model deployment and scoring, log metrics, and perform safe model rollouts.

Autoscaling compute

Use managed compute to distribute training and to rapidly test, validate, and deploy models. Share CPU and GPU clusters across a workspace and automatically scale to meet your machine learning needs.

Interoperability with other Azure services

Accelerate productivity with Microsoft Power BI and services such as Azure Synapse Analytics, Azure Cognitive Search, Azure Data Factory, Azure Data Lake, Azure Arc, Azure Security Center, and Azure Databricks.

Hybrid and multicloud support

Run machine learning on existing Kubernetes clusters on premises, in multicloud environments, and at the edge with Azure Arc. Use the simple machine learning agent to start training models more securely, wherever your data lives.

Enterprise-grade security

Build and deploy models more securely with network isolation and end-to-end private IP capabilities, role-based access control for resources and actions, custom roles, and managed identity for compute resources.

Cost management

Reduce IT costs and better manage resource allocations for compute instances, with workspace and resource-level quota limits and automatic shutdown.

Google Cloud Platform (GCP)

offered by Google, is a suite of cloud computing services that runs on the same infrastructure that Google uses internally for its end-user products, such as Google Search, Gmail, Google Drive, and YouTube

Alongside a set of management tools, it provides a series of modular cloud services including computing, data storage, data analytics and machine learning.[3] Registration requires a credit card or bank account details.

Google Cloud Platform provides infrastructure as a service, platform as a service, and serverless computing environments.

In April 2008, Google announced App Engine, a platform for developing and hosting web applications in Google-managed data centers, which was the first cloud computing service from the company. The service became generally available in November 2011. Since the announcement of App Engine, Google added multiple cloud services to the platform.

Ranging from healthcare to finance to automation there is no such industry where Artificial Intelligence (AI) and Machine Learning (ML) is not transforming the world and giving ease of life to every individual.

But we can not imagine that AI and Cloud Computing are totally unrelated for a second. In **2020**, more than 90% of companies are using cloud services for large data storage, big data analysis, managing data lakes, or data streaming, which are the initial core steps for any Machine Learning model.

Pay-as-you-go pricing model in cloud computing makes it easier for all of us to access services like GPU based computing servers and data lakes to manage large datasets from any cloud

provider like Amazon Web Services (AWS), Google Cloud Platform (GCP) or Microsoft Azure. AWS is undoubtedly leading the market with more than 60% market share public services but GCP is having a slight edge if we specifically talk about AI and ML cloud services.

In this blog, we will talk about a few out-of-the-box AI and ML services we can use in Google Cloud Platform (GCP).

Cloud AutoML

Cloud AutoML enables developers with limited machine learning and programming expertise to train high-quality models specific to their business needs. This service leverages Google's proprietary research technology and also relies on Google's state-of-the-art transfer learning and neural architecture search technology to achieve faster performance and more accurate predictions. People can train, evaluate, improve, and deploy custom machine learning models to solve problems for Vision, Translation, and Natural Language by using a simple graphical user interface within a few minutes.

Text-to-Speech-to-Text

Google provides very easy to use APIs to convert Text-to-Speech and Speech-to-Text with a few lines of code. These APIs are backed by many great key features including 90+ Wavenet voices, Voice and language selection, Global vocabulary, Pitch tuning, Noise robustness, Domain-specific models, Content filtering, Auto-detect language (beta), Automatic punctuation (beta), Text and SSML support, etc. This service is very useful for some different use cases like text bots, voice bots, transcribe multimedia content, customer service, voice generation, etc. These API's are priced based on the number of characters sent to the service to be synthesized into audio or the amount of audio successfully processed respectively each month.

Dialogflow

Google Dialogflow is a development suite for creating chatbots and conversational IVR for websites, mobile applications, popular messaging platforms, and IoT devices. This tool is powered by Google's machine learning and natural language processing algorithms to recognize a user's intent, understand user sentiment, and extract prebuilt entities such as time, date, and numbers. Dialogflow supports 20+ languages and one-click integration with 10+ different platforms. Apart from this, it has some more key features such as automatic spelling correction, built-in analytics, and expand your bots to voice interactions

AI Platform

Google AI Platform is a one-stop solution for machine learning developers and data scientists to take their ML projects from experiment to production and deployment. AI Platform integrated with several easy-to-use tools like BigQuery and Data Labeling Service to help you build and run your own machine learning applications quickly. You can store and manage the large amount of data with BigQuery and then prepare or label this data for model training using Data Labeling Service.

AI Hub

This is one of my favorites AI services provided by Google. AI Hub is Google Cloud's hosted repository of plug-and-play AI components, end-to-end AI pipelines, and out-of-the-box ML algorithms. You can discover best AI content, pre-trained models, and a wide range of open datasets to further modify them for your custom needs. You can share your ML pipelines, notebooks, models, and other AI content via AI Hub. AI Hub also provides enterprise-grade sharing capabilities for organizations to privately host their AI content for users internally.

Tensorflow Enterprise

In the last but not least, Tensorflow Enterprise provides enterprise-grade support, performance, and managed services for your ML & AI workloads directly by the Tensorflow creators.

TensorFlow Enterprise delivers so many out-of-the-box features like prioritized patches and bug fixes, fine-tuned TensorFlow containers, automatic provisioning, and scaling of resources like CPUs, GPUs, and TPUs, and managed services like AI Platform and Kubernetes Engine. And the best thing is it is available at no additional cost.

Why ML Capabilities Of GCP Is Way Ahead Of

AWS & Azure

“All vendors offer strong ML services and functionalities, but this is where GCP stands out as their years of search engine expertise, and research come into play,” said Diwakar Chittora, Founder & CEO, IntelliPaat.

Here is how GCP offers more benefits than AWS, Azure.

Tensor Processing Unit (TPU)
TPUs are Google’s custom-developed application-specific integrated circuits (ASICs) to accelerate ML workloads. A big advantage for GCP is Google’s strong commitment to AI and ML. “The models that used to take weeks to train on GPU or any other hardware can put out in hours with TPU. AWS and Azure do have AI services, but to date, AWS and Azure have nothing to match the performance of the Google TPU,” said Jeevan Pandey, CTO, TelioLabs.

Benefits of TPU:

- Can be leveraged by Google Cloud Machine Learning Engine to run complex ML models
- Increased performance of linear algebra computation
- Reduced time-to-accuracy when training complex neural network models
- Scalable across different machines

Vertex AI
Vertex AI brings together the Google Cloud services for building ML under one, unified UI and API. It has pre-trained APIs for vision, video, natural language, etc. Vertex AI integrates with widely used open-source frameworks such as TensorFlow, PyTorch, and scikit-learn, along with supporting all ML frameworks via custom containers for training and prediction.

Benefits of Vertex AI:

- Easily train and compare models using AutoML
- Faster movement of models from experimentation to production
- Access to AI toolkit used internally to power Google, including computer vision, language and conversation data.

- Smoother end-to-end ML workflow that removes the complexity of self-service model maintenance and repeatability with MLOps tools

Open-source

Google cloud's open-source contributions, especially in tools like Kubernetes –a portable, extensible, open-source platform for managing containerized workloads and services, facilitating declarative configuration and automation– have worked to their advantage.

"Kubernetes helps in the AI and ML workflow as it supports and leverages the speed of today's cloud GPUs," said Chittora.

Benefits of Kubernetes:

- Works on any container runtime or any infrastructure, including public cloud, private cloud and on-premises server
- Can host workloads on a single cloud or across many clouds.
- 100% open-source project offering more flexibility

Speech and translate APIs

Google cloud's speech and translate APIs are much more widely used than their counterparts. According to Gartner's 2021 Magic Quadrant, Google cloud has been named the leader for Cloud AI services. Pre-trained ML models can be instantly used to classify objects in an image into millions of predefined categories. Additionally, one of the top ML services from Google cloud is Vision AI, powered by AutoML.

Benefits of speech and translate APIs:

- Can translate many languages
- Can detect source text language
- Speech API recognises more than 80 languages
- Affordable pricing
- Products are highly scalable, easy-to-use and accurate

AutoML

AutoML enables developers with limited machine learning expertise to build custom ML models in minutes. It is a cloud-based ML platform and uses a No-Code approach with a set of prebuilt models via a set of APIs. It is tightly integrated with all Google's services and stores data in the cloud. Trained models can be deployed via the REST API interface.

"Data Analysts can combine a custom model and pretrained models in a single product, and this feature keeps GCP above its competitors," said Saket Saurabh, Head, Cloud and Analytics at STEMROBO.

Benefits of AutoML:

- A no-code way to build models
- Customers can train their own neural networks
- Can apply data and integrate predictions whenever you need
- Not limiting like single type of ML model offered by others

Other benefits

Abishek Chiffon, Data Scientist at Ideas2IT, outlined ML and AI services that put GCP ahead of AWS and Azure:

- Offers AI Hub, a repository of models in a format that can be deployed in Kubeflow, Deep Learning VMs in GPU or TPU.
- AI Platform Classic tool helps create ML training jobs with TensorFlow, Keras, PyTorch, Scikit-learn, and XGBoost and allows training at scale with user containers and custom frameworks.
- Google has TensorFlow to build and launch self-contained Deep Learning models.
- Its AI Platform Notebooks are VM instances that come pre-integrated with TensorFlow and PyTorch instances, Deep Learning packages, and Jupyter notebook.
- Deep Learning VM images come pre-installed with all software, deep learning and ML frameworks on a Google Compute Engine instance.

“AWS and Azure have different tools for different services, and also the training algorithm of GCP has matured over time through its Search and AI platforms used worldwide. GCP is turning out to be a popular choice for experienced and new data scientists and is ready to trump AWS and Azure soon,” said Saurabh of STEMROBO.