

AI and Cloud Computing - How the Cloud is accelerating AI

Suresh Babu Rajasekaran

Submitted: 02/11/2022

Revised: 21/12/2022

Accepted: 06/01/2023

Abstract – Enterprises worldwide are rapidly adopting cloud-native architectures and applications. Enterprises also increasingly infuse their products and services with AI capabilities to improve business efficiencies, achieve cost savings, boost sales, develop new use cases, and enhance customer experiences. But enterprises worldwide often struggle to understand how to seamlessly combine Cloud Computing and AI technologies and how these two technologies complement each other. This paper discusses how combining Cloud computing and AI, the two new-age technologies will propel enterprises to unleash their full potential in serving their customers and stay ahead of the competition in the world of digital transformation.

Keywords: artificial intelligence, cloud computing, AI in cloud computing, cyber security, digital transformation, enterprises, phishing.

1. INTRODUCTION

Artificial Intelligence (AI) is the most disruptive technology innovation of our lifetime and is finally meeting with the expansion of cloud computing. Although organizations understand the performance, reliability, scalability and simplicity of cloud computing and the importance and potential of AI, they often struggle to understand how to seamlessly adopt these two technologies and how these two technologies complement each other. In this paper we will see the challenges faced by organizations, how those challenges are addressed and how cloud computing has evolved into an excellent platform for enterprises in developing and distributing AI technologies at scale.

2. WHAT IS CLOUD COMPUTING?

Cloud computing is the on-demand delivery of IT resources over the Internet with pay-as-you-go (PAYG) pricing, meaning you only pay for what you have consumed. Instead of buying, owning, and maintaining physical data centers and servers, you can access technology services, such as computing power, storage, and databases, on an as-needed basis from cloud service providers (CSPs), some of the popular CSPs include

1. Amazon Web Services (AWS)
2. Google Cloud Platform (GCP)
3. Microsoft Azure
4. Oracle Cloud Infrastructure (OCI)
5. Salesforce
6. IBM
7. NVIDIA

3. WHY CLOUD COMPUTING?

Cloud computing is a paradigm shift from the

(srajasekaran.pm@gmail.com)

traditional way businesses think about IT resources and offers numerous benefits to businesses of all sizes, with cost savings, global scalability, performance, security, speed, and reliability being the primary advantages.

3.1 COST

Cloud computing offers cost savings to enterprises by minimizing huge capex that is required to setup a private data center i.e., initial hardware and software purchase costs, energy, and real estate expenses to run and maintain a private datacenter, in addition to labor, and physical security costs. CSPs leverage to economy of scale to provide these services at a lower cost to enterprises.

3.2 GLOBAL SCALE

A malware attack is a type of cyber-attack where the attacker uses malicious software (known as malware) to damage or disable computer systems. Malware is typically designed to cause harm to computer systems, such as by deleting files, stealing sensitive information, or disrupting the normal operation of a system. There are many different types of malwares, including viruses, worms, and ransomware.

3.3 PERFORMANCE

Cloud computing is user-friendly, highly manageable,

highly scalable, and highly reliable, guaranteeing an impressive 99.9% uptime, this enables businesses especially that requires low network latency to provide uninterrupted services to their customers which increases customer satisfaction and improves brand perception. CSPs also provide excellent monitoring and observability capabilities that enables enterprises to debug their applications in case of an outage.

3.4 SECURITY

Cloud computing can significantly improve enterprises security posture due to its advanced and robust security features, such as data encryption, multi-factor authentication, automatic maintenance, centralized management, and ample storage for backing up large numbers of files. These features ensure that enterprises data is protected and secure and adheres to any data policies like GDPR or PIPL.

3.5 SPEED

Cloud computing services offer businesses the flexibility to provision computing resources in minutes. With a few mouse clicks, businesses can quickly and easily scale resources without needing to purchase additional hardware. This makes planning computing capacity much easier and gives businesses the ability to launch applications quickly and at scale.

3.6 RELIABILITY

Cloud computing enables businesses to implement data backup, disaster recovery, and business continuity strategies with ease. By leveraging the cloud network provider's multiple redundant sites, organizations can save costs compared to setting up redundant on-premises infrastructure.

4. TYPES OF CLOUD COMPUTING?

Cloud computing models are based on the same principle of virtualization and differ in terms of location, accessibility, and storage capacity. Organizations can choose from three main types of cloud computing: Public Cloud, Private Cloud, or Hybrid Cloud.

4.1 PUBLIC CLOUD

Public Cloud is a vast array of readily available computing resources owned and managed by third-party cloud vendors like Google Cloud Platform (GCP), Amazon Web Services (AWS), and Microsoft Azure.

4.2 PRIVATE CLOUD

Private Cloud is a cloud computing resource exclusively for a single organization or entity that is physically located in the company's on-site data center or hosted by a third-party provider.

4.3 HYBRID CLOUD

Hybrid Cloud combines public and private clouds, allowing data and applications to be shared between them.

5. TYPES OF CLOUD SERVICES

Cloud computing services can be grouped into four broad categories: infrastructure as a service (IaaS), platform as a service (PaaS), serverless, and software as a service (SaaS).

5.1 IAAS

IaaS provides businesses with an efficient and cost-effective way to access and manage IT infrastructure without needing to purchase, install and maintain their own hardware. Companies can use IaaS to quickly develop and deploy applications, scale on demand, and access resources from anywhere in the world. Furthermore, IaaS offers a high level of flexibility, allowing businesses to pay only for the resources they use and to quickly adapt to changing needs and demands.

5.2 PAAS

Platform as a Service (PaaS) provides an environment for developers to create, test, deploy, and manage software applications on demand.

5.3 SERVERLESS COMPUTING

Serverless Computing enables developers to create apps and services without the need to manage the underlying infrastructure. This allows developers to focus on building their applications and services rather than spending time managing servers and infrastructure. With Serverless Computing, applications can automatically scale as needed and offer cost savings on infrastructure and server maintenance.

5.4 SAAS

Software as a Service (SaaS) is an on-demand, usually subscription-based, method of delivering software applications over the Internet. A user connects over the internet, typically using her web browser on the device, and the cloud provider hosts and manages the software application and underlying infrastructure.

6. WHAT IS AI?

Simply put, AI is a combination of different technologies and frameworks that allows machines to rationalize and take actions that have the best chance of achieving a specific goal. At its foundation are machine learning and its more complex sibling, deep learning. They bring to life various AI applications such as computer vision, natural language processing, and the ability to harness huge volumes of data to make accurate

predictions and discover hidden insights.

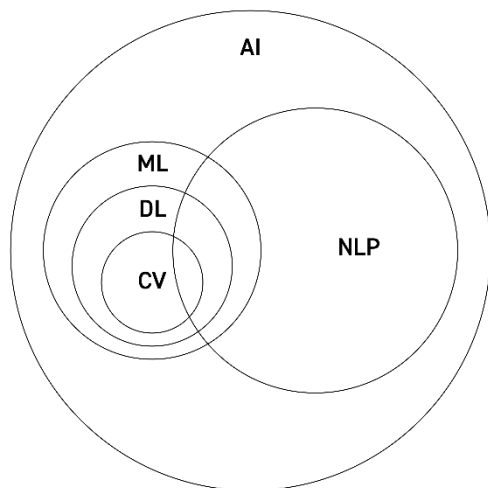


Fig. 1 Key pillars of AI

Some key pillars that constitute AI are:

6.1 MACHINE LEARNING (ML)

Machine Learning, as the name implies, gives a machine (a computer) the ability to learn. This learning process involves analyzing massive amounts of data, identifying hidden trends or patterns, and predicting future outcomes.

6.2 DEEP LEARNING

Deep learning is a branch of ML that trains the computer to think like humans. It runs on artificial neural networks that try to emulate the human brain through different types of data, weights, and biases. Deep learning accurately recognizes, classifies, and describes the patterns hidden inside the data using a combination of these elements.

This subset of AI helps analyze rich, complex, and multi-dimensional data, including speech, images, and video. Ultimately, it allows businesses to automate intelligent tasks, identify trends, make predictions, recommend products and services to customers, and improve efficiency.

6.3 NATURAL LANGUAGE PROCESSING (NLP)

Natural Language Processing, or NLP, is a branch of AI that allows computers to understand text and spoken words like humans do. NLP combines computational linguistics, ML, statistical, and deep learning models. This process extracts or generates meaning and intent from the text in a natural, readable, and grammatically correct form.

Enterprises can use this technology to analyze sentiment and automate speech-to-text conversion tasks. NLP also enables businesses to develop virtual assistants and chatbots to assist customers.

6.4 COMPUTER VISION

Computer vision is an interdisciplinary scientific field that deals with how computers can extract meaning and intent (a higher-level understanding) from visual elements like digital images or videos. From an engineering perspective, computer vision aims to perceive and automate tasks as human visual systems do. Some of the important use cases of computer vision include facial recognition, Optical Character Recognition (OCR), self-driving cars, factory automation, and medical image analysis.

7. AI USE CASES

With rapid and constant innovations happening in AI, new applications and use cases are being discovered every day in almost all domains and industries. AI is now widely adopted in a variety of industries like Agriculture, Banking and Finance, Government & Public Sector, Cybersecurity, Autonomous Vehicles, IoT, Healthcare, Defense, Retail, Manufacturing, and Smart Cities to drive efficiency, eliminate waste and generate revenue.

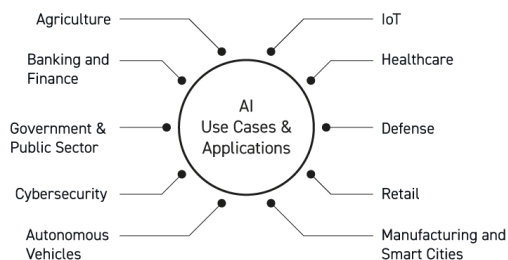


Fig. 2 Popular use cases of AI

8. THE CHALLENGES WITH AI ADOPTION

Some of the major inhibitors that enterprises face today in developing and distributing AI solutions at scale include:

8.1 COSTS

AI is still too expensive to be accessible to every company worldwide. ML and deep learning technologies require AI experts, access to large data sets, and powerful hardware accelerators like Graphics Processing Units (GPUs). All these elements require significant investments in money, talent, and time, which is quite challenging to achieve for most businesses.

8.2 LACK OF TALENT

There is a severe shortage of data scientists, researchers, and AI engineers in the job market. Most companies worldwide require more domain experts and resources to take full advantage of AI.

8.3 LACK OF TOOLS AND TECHNOLOGIES FOR ML OPERATIONS

The workflow and frameworks in developing AI applications are evolving and oftentimes are proprietary and lack common standards and interoperability.

8.4 LACK OF SUFFICIENT VOLUME AND QUALITY OF DATA

The quality of the data determines the effectiveness of an AI application. The lower data quality is one of the most critical challenges to AI adoption. Many enterprises collect large volumes of inconsistent and redundant data, which results in data decay. Fortunately, streamlining the data collection process, data cleansing, warehousing, and labeling can improve the data quality.

8.5 AI ETHICS

Ethics is becoming a major concern as more enterprises integrate AI with their business operations. A larger portion of data fed into the AI/ML models is prejudiced and biased. These models are inclined towards human biases and amplify them, making their decisions or conclusions primarily unreliable.

9. HOW THE CLOUD HELPS IN AI ADOPTION

Cloud democratizes AI making it accessible for enterprises of all sizes. We have briefly seen what AI means, various use cases and challenges that enterprises face in developing and distributing AI based solutions. Let's explore the combined power and advantages of AI and Cloud.

Training ML models using large data sets requires a large array of powerful and expensive CPUs and GPUs. A combination of public, private, or hybrid cloud setup can provide with such huge amounts of raw compute power in a cost-effective way. In addition, the storage, network, frameworks, data pipelines and orchestration mechanism, and other resources required to train, build, and deploy huge ML models at scale can be achieved relatively quicker and cheaper in a cloud environment. ML scientists and developers can now focus on their core work on improving the model rather than worrying about infrastructure by migrating their workloads to a cloud setup rather than running on premises or in a private data center.

Enterprises today rely heavily on High-Performance Computing, or HPC, to solve complex business problems at lightning speeds ranging from detecting a credit card fraud in real-time to a groundbreaking medical discovery to rendering a beautiful scene in a movie. With the advancement in cloud computing, organizations are moving their HPC workloads to the cloud. This transition transforms the dynamics of research and development of HPC based applications as it enables faster iteration and

fewer prototypes. It also accelerates and reduces the time and cost involved in bringing a product to the market.

Some of the common AI applications that can be accelerated in the cloud include:

9.1 IOT

AI powered IoT devices can seamlessly connect to cloud to store, retrieve and process any real time data that are generated by them without any network latency.

9.2 BUSINESS INTELLIGENCE

Business intelligence (BI) involving artificial intelligence can gain insight into the market, target audience, and competitors of customers. The cloud makes it more efficient to store and transfer data across BI systems.

9.3 CHATBOTS

Chatbots, powered by artificial intelligence (AI) and natural language processing (NLP), facilitate communication between users and applications. Cloud services can be leveraged here to store and process the data retrieved by the chatbot, as well as to connect it with other applications for further processing.

9.4 AI AS A SERVICE (AIAAS)

Companies can take advantage of AI outsourcing services provided by public cloud service providers, allowing them to test their software and ML algorithms without disrupting their main infrastructure. This cost-effective solution enables businesses to deploy AI-powered applications with ease.

9.5 COGNITIVE CLOUD COMPUTING

Companies like NVIDIA, IBM, and Google have created cognitive cloud platforms to enable businesses to access cognitive insights and apply Artificial Intelligence in their operations. These platforms are being used in a variety of industries, including finance, retail, and healthcare.

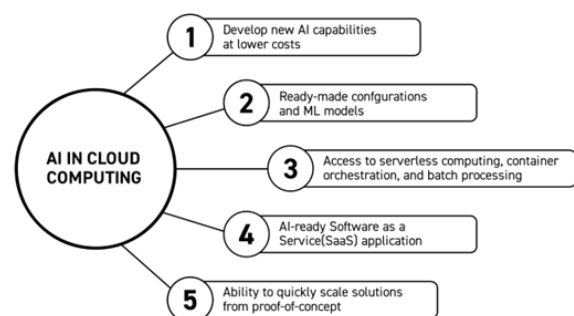


Fig. 3 AI in Cloud Computing

10. THE BENEFITS OF DEPLOYING AI IN THE CLOUD

There are numerous benefits and advantages in deploying all your AI workload in the cloud versus running them on-premises or in a private data center, some of them are

10.1 COST SAVINGS

With the innovations in public and private cloud virtualization - building, testing, and deploying ML-based models is significantly cost-effective, allowing small and medium businesses to remain competitive.

10.2 PRODUCTIVITY & AUTOMATION

Developing AI algorithms is complex and involves numerous repetitive tasks or provisioning and deprovisioning compute, network and storage instances which takes up a large amount of time for IT admins, Cloud helps IT admins to focus more on value-driven tasks by eliminating the repetitive tasks by automating it without the need to manually manage software and hardware provisioning.

10.3 ANALYTICS

Cloud powered AI models can mine data stored in the cloud in real-time, offering valuable insights to end users and enabling them to make more informed decisions.

10.4 DATA MANAGEMENT

AI powered business uses cases in customer support, HR, marketing, finance, operations, and supply chain management when deployed in cloud enables stakeholders to make decisions in real time thereby avoiding the lag time and ensures business continuity.

10.5 BETTER SAAS TOOLS

AI-based algorithms can be integrated with popular SaaS apps to optimize customer service, marketing, ERP, and supply chain management, as well as provide enhanced value to end users. Einstein is Salesforce's Cloud CRM powered by AI, is one such example that offers actionable insights to customers from the massive amount of data.

11. AI IN THE CLOUD: THE CHALLENGES

Using AI in the cloud is a challenging endeavor and is not without its challenges, two primary challenges being – data privacy and connectivity

11.1 DATA PRIVACY

Data privacy is a concern because cloud allows companies to collect customer, supplier, and marketing data through their cloud AI offerings with little

understanding of the security risks involved. Sensitive company and customer data could be the prime target for hackers if the processes around data while designing and developing AI algorithms are not clearly defined.

11.2 CONNECTIVITY

Connectivity is another issue, as any AI algorithm or processing system functioning on the cloud requires a stable internet connection to function properly. Without a reliable connection, ML processes can be hindered, and real-time events and analytics can get compromised.

12. AI, CLOUD COMPUTING AND GEOPOLITICS

The politicization of AI technology is inevitable, opening possibilities for new forms of conflict. Technology has always shaped geopolitical forces. Geopolitical actors have always used technology to achieve their goals. In some countries, AI can be used at scale using cloud platforms to manipulate public opinion by controlling the information people see and using surveillance to limit freedom of expression.

Moreover, AI-powered automation will transform employment patterns in ways that benefit some economies more than others. One way to mitigate the geopolitical risks of AI and ensure the comprehensive oversight it needs is by having open standards and transparent discussion about AI's benefits, limitations, and complexities.

13. THE FUTURE OF AI AND CLOUD COMPUTING

The merger of AI and cloud computing will bring tremendous shifts in the IT industry. According to Gartner, nearly 90% of businesses worldwide will embrace a cloud-first principle by 2025, and one in every three enterprises will be using cloud-based AI by 2024. The AI and Cloud Computing combination will accelerate the digital transformation of many businesses worldwide. The deeper integration of AI into cloud computing will lead to improved automation and help discover new use cases.

Enterprises worldwide have cloud budget and have pre-committed spend agreement with major CSPs, CSP Marketplaces now offer unprecedented access to software to various teams within an enterprise, and they don't need to go through different POs (Purchase Orders) for each software and services as all billing and transactions will be done within the CSP realm, this also accelerates cloud and AI adoption across businesses.

Technology giants like Amazon, Microsoft, Google, NVIDIA, and IBM are at the forefront of the Cloud Computing and AI/ML innovation and are closely followed by numerous AI start-ups who have raised over

\$43 billion in funds to further the AI software and hardware research and development. There are numerous incentives and free trials along with education, tutorials, and know-hows provided by these technology innovators to make it easier for enterprises to adopt and scale their AI offerings in the cloud. There is also increase in the IT spend by the enterprises to upgrade their IT infrastructure and we are in the middle of an important upgrade cycle. We are just getting started.

REFERENCES

- [1]. The Critical Role of Data in Scaling AI and ML Initiatives by Ritu Jyoti, February 2022, Snowflake, Inc.
- [2]. Artificial Intelligence in Cloud Computing by Michael Kriech, April 2022 Datacenters.com
- [3]. The Role of Cloud Computing in Artificial Intelligence by Kivanc Uslu, April 2021, Towards Data Science
- [4]. The Role of AI in Cloud Computing, By Michael Brenner January 2022, Nutanix, Inc.
- [5]. IBM Global AI Adoption Index 2022, IBM Corporation

DISCLAIMER

This disclaimer informs readers that the views, thoughts, and opinions expressed in this paper belong solely to the authors, and not necessarily to the author's employer, institution or other groups or individuals.

Suresh Babu Rajasekaran is an experienced product leader with over sixteen years of industry experience and is currently the Group Product Manager at NVIDIA. In his current role, he is focused on helping enterprises to adopt, accelerate and scale AI solutions in their businesses. Suresh worked at companies like Samsung, Adobe, and Autodesk before joining NVIDIA. Suresh is passionate about Artificial Intelligence and its profound impact on Humans. Suresh holds a Masters in Software Engineering from Carnegie Mellon University and an MBA from The University of Chicago Booth school of Business. He currently lives in the San Francisco Bay Area with his wife and two kids. The opinions shared in this paper are the author's own and are not affiliated to any institute or organization.