



Cloud Platforms for AI

Jade, Alfredo, Monique,
Morgan, Emmanuel
08-July-2025

What is a Cloud Platform for AI?

Scalable & On-Demand Resources

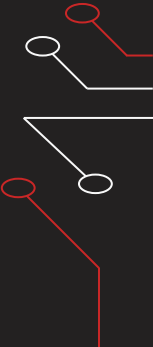
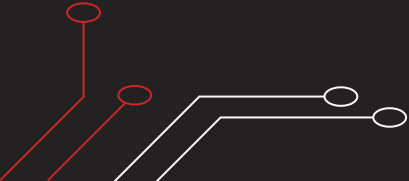
Cloud platforms provide flexible computing power that scales with AI workload demands.

No Local Infrastructure Needed

Train, deploy, and manage AI models without relying on on-premise hardware.

Popular Cloud AI Services:

- AWS SageMaker
- Google Cloud AI Platform
- Microsoft Azure AI





Key Components of AI Cloud Platforms

Compute Resources

- Intel processors, TPUs, GPUs, and CPUs
- Used for training and running inference on AI models

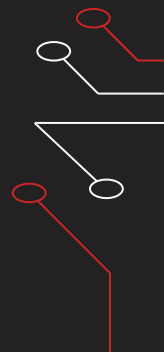
Storage

- Data lakes and model artifacts stored in object stores
- Examples: Amazon S3, Google Cloud Storage (GCS)

Data Pipelines

- Tools like Dataflow and ETL processes using Apache Beam
- Enable data transformation and movement at scale

Orchestration

- Workflow management tools to coordinate tasks
 - Examples: Apache Airflow, Vertex AI Pipeline
- 

Techniques & Approaches

Machine Learning & Deep Learning

- Prebuilt and custom model training (e.g., TensorFlow, PyTorch)
- AutoML for low-code/no-code model development
- Transfer learning for faster, efficient training

MLOps (Machine Learning Operations)

- CI/CD pipelines for model deployment
- Model versioning and rollback
- Monitoring for drift and performance

Serverless & Scalable Compute

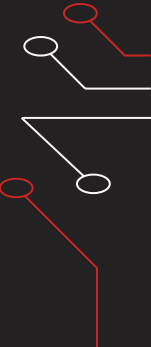
- On-demand compute (e.g., AWS Lambda, Google Cloud Functions)
- Elastic scaling for training and inference
- Cost-efficient resource management

Data Engineering & Pipelines

- ETL/ELT tools (e.g., AWS Glue, Azure Data Factory)
- Real-time data streaming (e.g., Kafka, Pub/Sub)
- Data validation and preprocessing automation

Security & Governance

- Role-based access control (RBAC)
- Data encryption at rest and in transit
- Compliance with GDPR, HIPAA, etc.



Popular Platforms

Cloud Provider	AI & Data Services
Amazon Web Services (AWS)	- SageMaker (ML platform)
	- EC2 (compute)
	- S3 (storage)
	- Glue (data integration)
Google Cloud Platform (GCP)	- Vertex AI (ML lifecycle)
	- BigQuery (data warehouse)
	- AutoML (custom model training)
Microsoft Azure	- Azure ML (model development)
	- Data Factory (ETL pipelines)
	- Cognitive Services (prebuilt AI APIs)
IBM Cloud	- Watson AI (AI services)
	- DataStage (data integration)



Use Cases & Applications

Industry	AI Use Case	Cloud Platforms
Retail	Product recommendations	AWS, GCP
Healthcare	Medical image analysis	Azure, IBM Watson
Finance	Fraud detection models	AWS, GCP
Manufacturing	Predictive maintenance	Azure, IBM

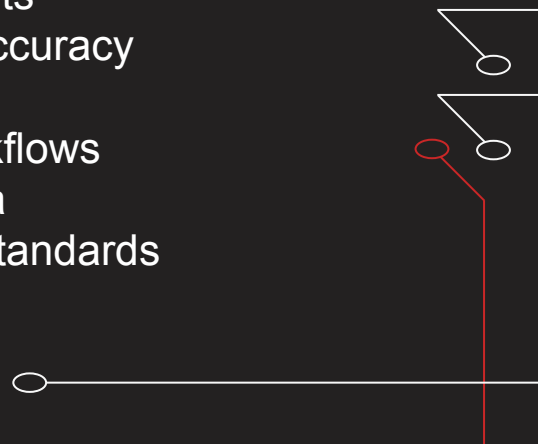
Challenges & Best Practices

Limiting Factors

- Cost Control – Managing compute and storage expenses
- Model Drift – Ensuring models stay accurate over time
- Data Privacy – Protecting sensitive and personal information
- Integration – Seamlessly connecting AI with existing systems

Best Practices

- Version Control – Track changes in models and datasets
- Performance Monitoring – Regularly evaluate model accuracy and efficiency
- Pipeline Automation – Streamline data and model workflows
- Data Quality & Fairness – Ensure clean, unbiased data
- Regulatory Compliance – Adhere to legal and ethical standards



Emerging Trends

MLOps & Serverless AI Platforms

- Tools like Google Vertex AI streamline model deployment and lifecycle management
- Serverless architecture reduces infrastructure overhead

Hybrid Cloud Solutions

- Seamless integration between on-premises systems and cloud environments
- Offers flexibility, compliance, and performance optimization
- Combined Data-to-Model Pipelines
- Unified workflows from data ingestion to model deployment
- Enhances automation, traceability, and scalability

AI-as-a-Service (AlaaS)

- Rapid growth in prebuilt, cloud-hosted AI solutions
- Lowers entry barriers for businesses adopting AI

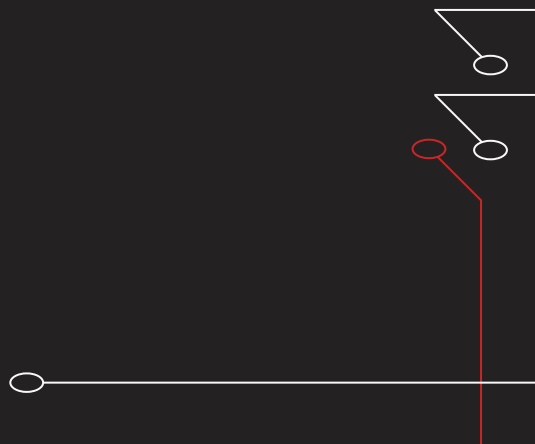


Demo

AI and the Future of Cloud – Google Cloud Tech (15 min)

A thoughtful discussion on how AI is transforming cloud infrastructure, featuring insights from industry expert Janakiram MSV. Topics include:

- The role of cloud in AI development
- AI in DevOps and automation
- Future applications of AI in cloud environments



Summary

Accelerated Deployment & Scalability

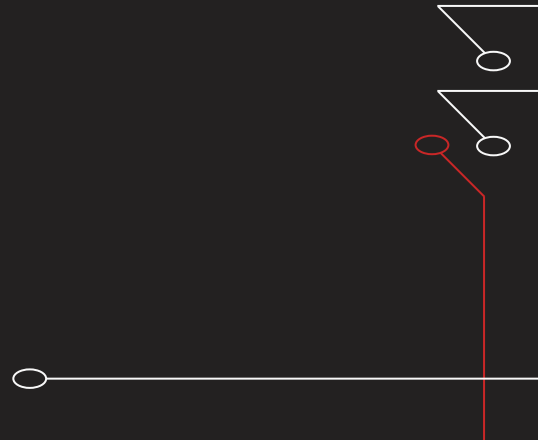
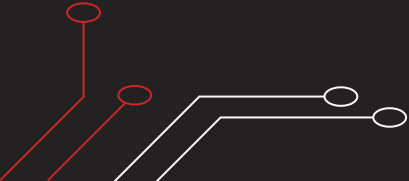
- Cloud platforms enable rapid rollout of AI solutions
- Easily scale resources up or down based on demand

Integrated Ecosystems

- Bundled tools for data management, compute power, and workflow orchestration
- Streamlined development and operations

Platform Selection Criteria

- Depends on:
 - AI workload requirements
 - Cost considerations
 - Specific services and capabilities needed



References

- Amazon Web Services. (n.d.). Amazon SageMaker. AWS.
<https://aws.amazon.com/sagemaker/>
- Google Cloud. (n.d.). Vertex AI documentation.
<https://cloud.google.com/vertex-ai>
- Microsoft. (n.d.). Azure Machine Learning.
<https://azure.microsoft.com/en-us/products/machine-learning>
- IBM. (n.d.). Watson AI. <https://www.ibm.com/watson>

