# Question 1: Cloud Computing for Deep Learning

1. **Elasticity and Scalability in Cloud Computing for Deep Learning (10 points)**
   * **Elasticity** refers to the ability of a cloud computing system to automatically adjust resources (such as compute power, storage, and networking) based on demand. For deep learning, this means dynamically scaling up GPUs/TPUs during model training and scaling down when the demand decreases, optimizing cost efficiency.
   * **Scalability** is the capacity of a system to handle an increasing workload by adding more resources, either **vertically** (upgrading existing hardware) or **horizontally** (adding more machines). In deep learning, scalability ensures that models can be trained on massive datasets efficiently by distributing workloads across multiple cloud instances.

# Comparison of AWS Sage Maker, Google Vertex AI, and Microsoft Azure Machine Learning Studio

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **AWS Sage Maker** | **Google Vertex AI** | **Microsoft Azure ML Studio** |
| **Ease of Use** | Provides pre-built Jupyter notebooks and automated ML capabilities. | Simplifies end-to-end AI workflows with AutoML and pipelines. | Offers a no-code UI along with support for advanced ML pipelines. |
| **Compute Support** | Supports CPU, GPU, and AWS Inferentia chips for cost-effective inference. | Offers TPUs, GPUs, and CPUs for AI model training and deployment. | Provides flexible compute options, including NVIDIA GPUs and FPGAs. |
| **AutoML Capabilities** | Includes built-in AutoML for training without deep expertise. | Strong AutoML support with hyperparameter tuning. | Advanced AutoML with drag-and-drop features. |
| **Model Deployment** | One-click deployment with endpoints for real- time inference. | Seamless model deployment via AI pipelines. | Supports real-time and batch inference with MLOps integration. |
| **Integration & Ecosystem** | Deep integration with AWS services (S3, Lambda, EC2, etc.). | Native support for Google Cloud services (BigQuery, Dataflow, etc.). | Well-integrated with Azure services (Power BI, Cognitive Services, etc.). |
| **Pricing** | Pay-as-you-go pricing with spot instance discounts. | Flexible pricing based on usage; discounts for sustained use. | Consumption-based pricing with enterprise discounts. |

Each platform has strengths depending on specific needs:

* + **AWS SageMaker**: Best for users already in the AWS ecosystem. It provides end-to-end machine learning services, including built-in algorithms, automated model tuning, and easy AWS storage and compute integration.
  + **Google Vertex AI**: Optimized for large-scale AI workloads and deep learning applications, especially with TPUs (Tensor Processing Units). It offers seamless integration with Google Cloud services and a unified AI platform for model training, deployment, and monitoring.
  + **Azure Machine Learning Studio**: Well-suited for enterprises using Microsoft services. It provides robust MLOps capabilities, AutoML features, and strong integration with tools like Power BI and Azure DevOps.