



AI in the Cloud

Unit 12 - Summary

- **AI in Cloud:** AI in the cloud is the future of computing and will transform every industry.
- **AI Infrastructure:** Enterprise AI requires time, expertise, and the right approach to architecture.
- **AI & Cloud Computing:** The combination of AI and cloud computing unleashes intelligent solutions through cloud-powered innovation.
- **Cloud Consumption Models:** Choosing the right consumption model for your business needs is important.
- **NVIDIA Solutions:** NVIDIA offers a robust ecosystem for AI deployment in the cloud.

What is AI in the cloud?

AI in the cloud refers to the use of cloud computing to enhance AI deployments, allowing for intelligent solutions through cloud-powered innovation.

What are some AI use cases in the cloud?

Some AI use cases in the cloud include large language models, speech AI, recommenders, cybersecurity, sentiment analysis, supply chain optimization, and predictive maintenance.

What are some key considerations for deploying AI in the cloud?

Key considerations for deploying AI in the cloud include assessing your organization's AI capabilities and readiness for growth, choosing between on-premises and cloud solutions, and leveraging a hybrid approach for maximum potential.

What are some supported cloud service providers and consumption models?

NVIDIA offers a robust ecosystem for AI deployment in the cloud, with support for multiple cloud service providers and consumption models. Customers can choose between managing their own cloud infrastructure or having the cloud provider manage it for them.



What are some NVIDIA solutions in the cloud?

NVIDIA offers a range of solutions in the cloud, including GPU-accelerated VMs/instances, AI Enterprise software, and the NVIDIA AI platform. These solutions provide an end-to-end open platform for production AI for everyone, everywhere.