



Data Center and Cloud Computing

Unit 6 - Summary

- **AI Infrastructure:** AI infrastructure includes compute, storage, and networking components, with accelerated systems providing computational power for AI training and inferencing.
- **Energy Efficiency:** Balancing speed and sustainability is important in AI infrastructure, with GPUs leading in energy efficiency.
- **Cloud Computing:** The power of accelerated computing can be extended to the cloud, minimizing OPEX and maximizing ROI.
- **Infrastructure Management:** AI infrastructure management includes provisioning, monitoring, and workload management for informed decision-making.
- **Orchestration and Scheduling:** Orchestration and job scheduling are important for automating provisioning and deployment and managing workloads.

What are the hosting environments for AI workloads?

The hosting environments for AI workloads include data centers and the cloud. Compute, storage, and management nodes are networked together to form a data center. The power of accelerated computing can also be extended to the cloud.

What are the components constituting AI data centers?

The components constituting AI data centers include compute platforms, networking, and storage for AI. An accelerated system is the next phase in the evolution of computers, providing computational power for AI training and inferencing. The network serves as the backbone of the AI infrastructure, and high-speed storage access is required for AI workloads.

What are the requisites and methods for managing and monitoring AI data centers?

The requisites and methods for managing and monitoring AI data centers include infrastructure provisioning, resource monitoring, and workload management. Orchestration and job scheduling are also important for automating provisioning and deployment, and managing workloads.