



Management and Monitoring

Unit 13 - Summary

- **Management Overview:** This unit covers concepts about provisioning, managing, and monitoring AI infrastructure.
- **Cluster Management:** Cluster management includes infrastructure provisioning, resource management, and workload management.
- **Provisioning Tools:** Provisioning involves installing software and drivers onto hardware.
- **Monitoring Tools:** Monitoring involves checking the health and metrics for the AI infrastructure and software, services, and jobs running on the infrastructure.
- **Workload Management:** Workload management involves ensuring resources, scheduling jobs, and monitoring resource usage.

What is the main topic of this unit?

This unit is about Management and Monitoring of AI infrastructure, including Cluster Management, Infrastructure Provisioning, Resource Management and Monitoring, Workload Management and Monitoring, and Base Command Manager Overview.

What are the learning objectives of this unit?

By the end of this unit, you will be able to identify the general concepts about provisioning, managing, and monitoring AI infrastructure, describe the value of cluster management tools, describe the concepts for ongoing monitoring and maintenance, and identify tools that are used for provisioning, management, and monitoring.

What are some tools used for resource management and monitoring?

Some tools used for resource management and monitoring include DCGM Exporter, Prometheus, and Grafana.

What is Base Command Manager?

Base Command Manager is a tool that provides infrastructure provisioning, resource monitoring, and user access and workload management.