Cloud Computing Concepts

CS 3132

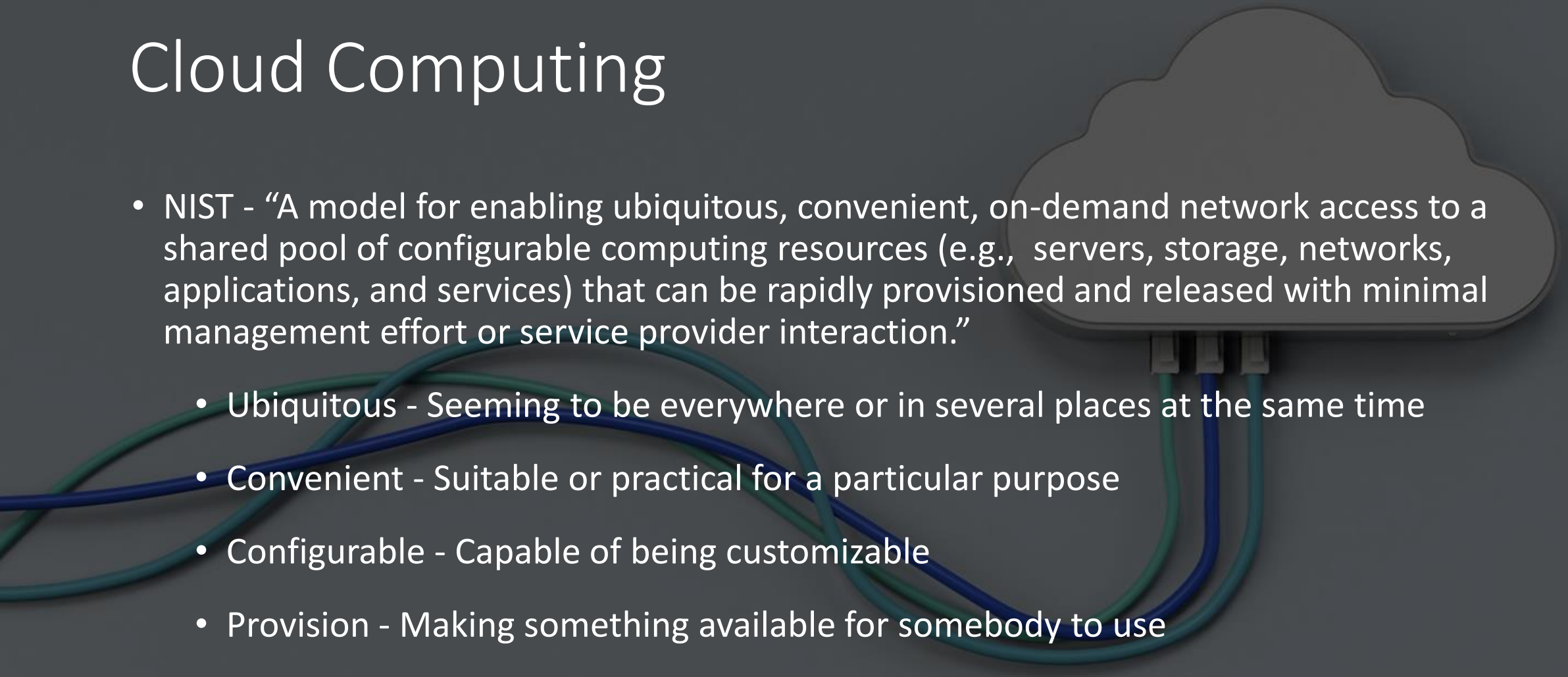
Dr. Anand Kumar Mishra

NIIT University

Cloud Computing

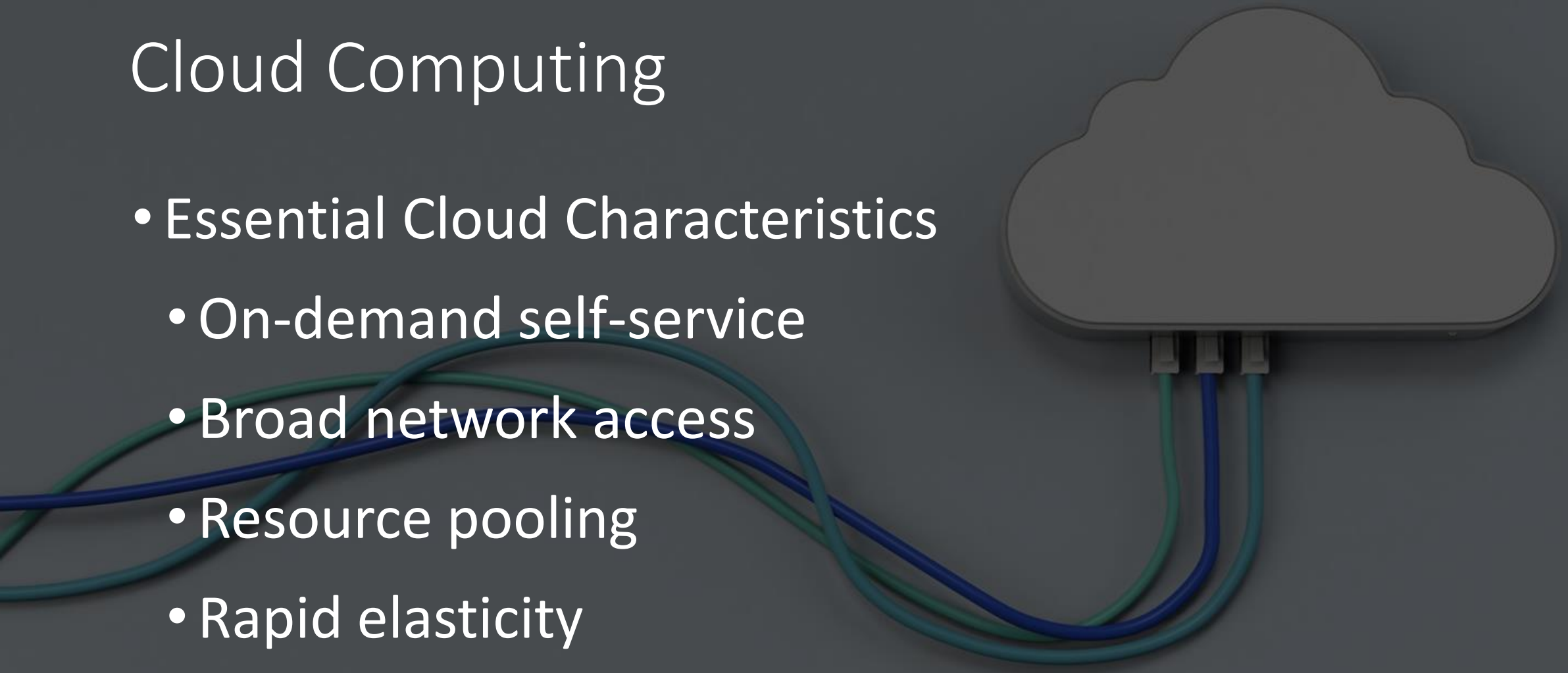
- NIST - “A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., servers, storage, networks, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

Cloud Computing



- NIST - “A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., servers, storage, networks, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”
  - Ubiquitous - Seeming to be everywhere or in several places at the same time
  - Convenient - Suitable or practical for a particular purpose
  - Configurable - Capable of being customizable
  - Provision - Making something available for somebody to use

Cloud Computing



- Essential Cloud Characteristics
  - On-demand self-service
  - Broad network access
  - Resource pooling
  - Rapid elasticity
  - Measured service

Cloud Computing - On-Demand Self-Service

- Enables consumers to get computing resources as and when required, without any human intervention
- Facilitates consumer to leverage “ready to use” services or, enables to choose required services from the service catalog
- Allows provisioning of resources using self-service interface
  - Self-service interface should be user-friendly

Cloud Computing - Broad Network Access

Cloud services are accessed via the network, usually the internet, from a broad range of client platforms such as:

- Desktop computer
- Laptop
- Mobile phone
- Thin Client

Eliminates the need for accessing a particular client platform to access the services

Enables accessing the services from anywhere across the globe

Cloud Computing - Resource Pooling

IT resources (compute, storage, network) are pooled to serve multiple consumers

- Based on multi-tenant model

Consumer has no knowledge about the exact location of the resources provided

Resources are dynamically assigned and reassigned based on the consumer demand

Cloud Computing - Rapid Elasticity

Ability to scale IT resources rapidly, as required, to fulfill the changing needs without interruption of service

- Resources can be both scaled up and scaled down dynamically

To the consumer, the Cloud appears to be infinite

- Consumers can start with minimal computing power and can expand their environment to any size

Cloud Computing - Measured Service

Consumers are billed based on the metered usage of Cloud resources

- Cost incurred on a pay-per-use basis
- Pricing/billing model is tied up with the required service levels
- Resource usage is monitored and reported, which provides transparency for chargeback to both Cloud service provider and consumer about the utilized service

Cloud systems automatically control and optimize resource use

- By leveraging a metering capability at some level of abstraction appropriate to the type of service
  - storage, processing, bandwidth, and active user accounts

Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service