**30 AUG 2023**

Big data

* Data that has terabyte of size and more.
* 4V’s
* Volume, Variety, Velocity, Veracity
* Batch processing – processing in the batch
* Stream processing – live processing
* Stream analytics helps in understanding 360 degrees of customer.
* Parallel processing
* Distributed computing

Data Warehouse

* Save data as object.
* Structured data
* Dimension modeling
* Acid properties
* Fixed schema

Data Lake

* Structured, semi structured and unstructured.
* Flexible schema
* Acid properties are not available.
* Metadata

Lake House

* Combination of both

Cloud Computing

* Delivery of computing resources over internet
* Private, Public, Hybrid
* Capex model – long term expenses
* Opex model - day to day expenses
* IaaS – only provides the infrastructure, like storage networking.
* PaaS - provides environment for building testing.
* SaaS – only we can use the software.

Core Azure Architecture Components

* Region pairs – at least 300 miles of separation, automatic replication of some services, updates are rollout sequentially to minimize the downtime.
* Availability zones – provide protection against downtime in the same region.
* Azure Resources – components like storage, VM’s, and network
* Azure Subscriptions – provides with the authenticated and authorized access to azure services.
* Azure Resource Manager – enables us to create, update and delete resources.
* Resource Group – container to manage all the resources in a single unit.