**Section 1**

* What is a Data Lake?
  + A Data Lake is a centralized repository that lets an individual/organization store unstructured and structured data at scale. Data can be stored as-is without having to first perform any sought of transformation on it.
  + Benefits of a data lake can include:
    - Data lake offers the ability to ingest data from multiple sources in less time without defining a schema.
    - Data lake can handle large volumes of big data and requires low-cost hardware.
    - Data lake allows for more complex and advanced analytics such as machine learning and predictive modeling using raw data.
  + How it differs from a data warehouse
    - I like to use an analogy to describe the difference:
      * A data lake is like water lake where many water streams flow into it and it’s up to everyone to get water the way he/she wants from the lake while a dataware house is like a producer of water where you are handed bottled water in a particular size and shape of the bottle.
* Data ingestion to Data lake is based on ELT (Extract, Load and Transform) process while Data ingestion to a Data warehouse is based on ETL(Extract, Transform and Load).
* Schema for a data lake is not defined essential schema-on-read while schema for a data warehouse is known before ingestion (schema-on-write)
* Serverless Architecture:
  + Serverless Architecture is an approach to building and running applications and services without having to manage the underlying server infrastructure essentially, a cloud provider like AWS allocates machine resources and takes care of servers based on customers demands.
  + Benefits of serverless?
    - Serverless frees up administrators from architecture-related responsibilities.
    - Serverless is easy to deploy as developers are not bug down with responsibility of designing, developing, and deploying serverless infrastructure but focus on implementing code logic for the business and application release.
    - Inherent Scalability – cloud providers are responsible for scaling capacity on demand so developers do not spend time setting up, tuning and auto-scaling as the function instances of serverless computing are created or removed automatically.
    - Boost in productivity of developers – developers who use serverless architecture can simply deploy their code without being concerned with maintenance and planning issues that come with servers hence this accelerates product delivery cycles and rapidly scale operations.
  + ETL Pipeline Diagram
    - A diagram of a server

      Description automatically generated
  + Modern MLOPs is the practice of efficiently developing, testing, deploying and maintaining machine learning applications in production. MLOps automates, monitors the entire machine learning lifecycle, and enables seamless collaboration across teams, resulting in faster time to production and reproducible results.
    - How organizations should be approaching management from a tool and system perspective
      * Firstly, the organization must understand it specific needs by accessing workflows processes and pain points that could benefit from tooling and system improvements.
      * Evaluate Exiting Tools – Access the tools and systems currently in use and determine their effectiveness and limitations.
      * Research and choose tools wisely – Implement tools that foster collaboration and effective communication among team members. Consider factors like ease of integration, scalability, cost and support.
      * Automate Repetitive tasks- identify repetitive tasks that consume valuable time and resources then implement automation tools or scripts to streamline these processes, freeing up employees to focus on more strategic tasks.
      * Scalability and Flexibility- Choose tools and systems that can scale with the organization’s growth and adapt to changing needs.