**DP-203: Data Engineering on Microsoft Azure**

Evolving World of Data

While extracting data from data pool and migrating it to data repository data is transformed as both source and destination might have different schemas. This process is called ETL(Extract Transform Load).

But this process has a disadvantage that the transformation stage sometimes take longer time which can potentially tie up source resources.

To overcome this issue we have an alternative approach called Extract, Load and Transform(ELT) where extracted data is immediately stored on a large repository system such as Azure Cosmos DB or Azure Data Lake Storage which reduces the contention (disagreement) on Source resources.

Cloud Use-cases

**Web:** Azure Cosmos DB multimaster replication model to create a data architecture that supports web and mobile applications.

**Healthcare :** Azure Databricks to accelerate big-data analytics and AI solutions.

**IoT solutions :** Azure IoT Hub, to design a data solution architecture that captures information from IoT devices.

* Unstructured data schema can be defined at query time.

# Understand data storage in Azure Storage

Azure Storage offers four configuration options:

* **Azure Blob**: A scalable object store for text and binary data
* **Azure Files**: Managed file shares for cloud or on-premises deployments
* **Azure Queue**: A messaging store for reliable messaging between application components
* **Azure Table**: A NoSQL store for no-schema storage of structured data