Data Types

- **Structured** - Data that can be represented using tables with very strict schema. Each row must follow defined schema. Some tables have defined relationships between them. Typically used in relational databases.
- **Semi-structured** - Data that can be represented using tables but without strict defined schema. Rows must only have unique key identifier.
- **Unstructured** - Any files in any format. Like binary files, application files, images, movies, etc.

Cosmos DB

- Globally distributed NoSQL (semi-structured data) Database service
- Schema-less
- Multiple APIs (SQL, MongoDB, Cassandra, Gremlin, Table Storage)
- Designed for
    - Highly responsive (real time) applications with super low latency responses <10ms
    - Multi-regional applications

SQL Database

- **Relational database** service in the cloud (PaaS) (DBaaS - Database as a Service)
- **Structured data service** defined using schema and relationships
- **Rich Query Capabilitie**s (SQL)
- **High-performance**, reliable, fully managed and secure database for building - applications

Azure SQL product family

- Azure **SQL Database** – Reliable relational database based on SQL Server
- Azure **Database for MySQL** – Azure SQL version for MySQL database engine
- Azure **Database for PostgreSQL** – Azure SQL version for PostgreSQL database engine

- Azure **SQL Managed Instance** – Fully fledged SQL Server managed by cloud provider
- Azure **SQL on VM** – Fully fledged SQL Server on IaaS
- Azure **SQL DW (Synapse)** – Massively Parallel Processing (MPP) version of SQL Server

What is Big Data?

Big Data is a field of technology that helps with the **extraction**, **processing** and **analysis** of information that is **too large or complex** to be dealt with by traditional software.

The three V's rule

Big data typically has one of the following characteristics

- **Velocity** - how fast the data is coming in or how fast we are processing it
  - Batch
  - Periodic
  - Near Real Time
  - Real Time
- **Volume** - how much data we are processing
  - Megabytes
  - Gigabyte
  - Terabytes
  - Petabytes
- **Variety** - how structured/complex the data is
  - Tables
  - Databases
  - Photo, Audio
  - Video, Social Media

Azure Synapse Analytics

- Big data analytics platform (PaaS)
- Multiple components
  - Spark
  - Synapse SQL
    - SQL pools (dedicated – pay for provisioned performance)
    - SQL on-demand (ad-hoc – pay for TB processed)

- Synapse Pipelines (Data Factory – ETL)
- Studio (unified experience)

Azure HDInsight

- Flexible multi-purpose big data platform (PaaS)
- Multiple technologies supported (Hadoop, Spark, Kafka, HBase, Hive, Storm, Machine Learning)

Azure Databricks

- Big data collaboration platform (PaaS)
- Unified workspace for notebook, cluster, data, access management and collaboration
- Based on Apache Spark
- Integrates very well with common Azure data services

Azure Marketplace

- Think of it like an "Azure Shop" where you purchase services and solutions for the Azure platform
- Each product is a template which contains one or multiple services
- Products are delivered by first and third-party vendors
- Solutions can leverage all service categories like IaaS, PaaS and SaaS