

Tools used in big data

Apache Hadoop
 Apache Spark
 NoSQL Database
 Apache Kafka
 Elasticsearch and Kibana

V's of Big Data

- **Volume**
- **Velocity**
- **Variety**
- **Veracity (Validity)**
- Value
- Variability
- Visualization
- Virality

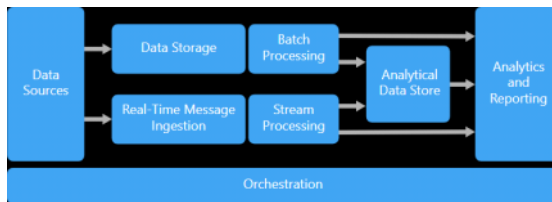
Challenges of Big Data

- Storage and management**
 - Hadoop distributed file system (HDFS)
 - Distributed and scale storage
 - Store data in files instead of rows and cols
- Processing power**
- Data integration**
- Data quality**
- Privacy and security**
- Analytics and insights**

Types of data

- Structured data - relational, tabulated - excel
- Un structured data - files, (eg. Notepad) - social media, image, Video
- Semi structured - data from iot devices - xml, json

Big Data Architecture Style



Data Lakes

A centralized repo that allows to store manage and analyze vast amt of data of all types.
 Technology for building datalakes

- Azure HDInsight
- Azure datalakes store
- Azure datalakes analytics

Tools and services for big data analysis

