# Extra Notes-Big Data

7 Aug 2023

#### Data-at-rest

SQL, NoSQL databases
Folders and files
Logs and crash reports
Dump files
Extracts, backups

Either data transfer or migration



Data-in-motion

Games, videos, CCTVs, IoT, sensors/actuators

Router (HA), buffers

Data-at-rest (batch)

SQL, NoSQL databases
Folders and files
Logs and crash reports
Dump files
Extracts, backups

Azure/AWS/GCP

SQL migration tools

Azure- DataFactory GCP- DataFlow



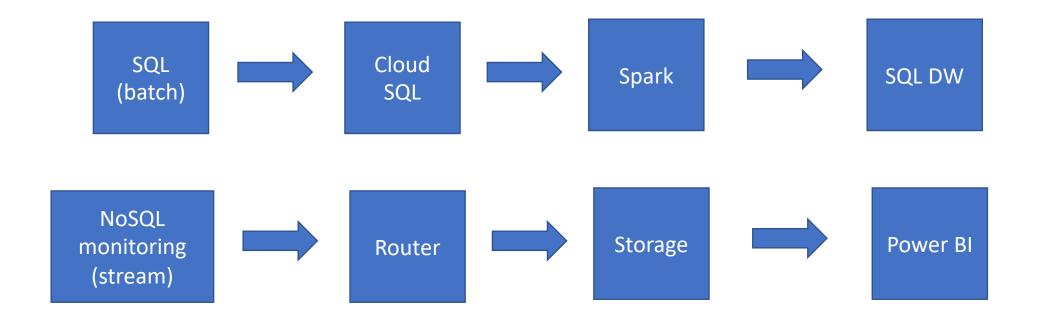
Data-in-motion (stream)

Games, videos, CCTVs, IoT, sensors/actuators

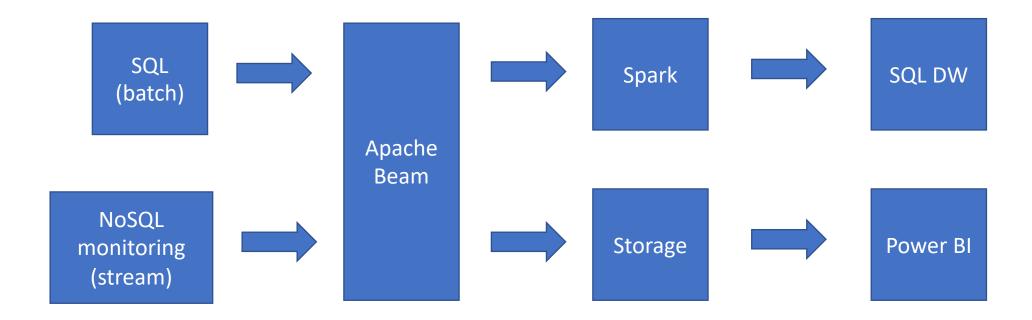
Router (HA), buffers

on prem- Kafka, Storm Azure- IoT Hub or events hub GCP- Pub/Sub and DataFlow AWS- SQS, SNS, EventBridge

### Lambda

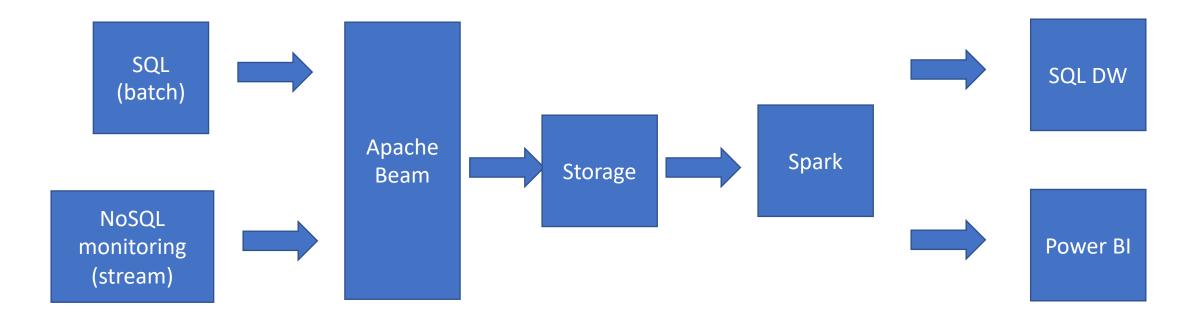


## Kappa



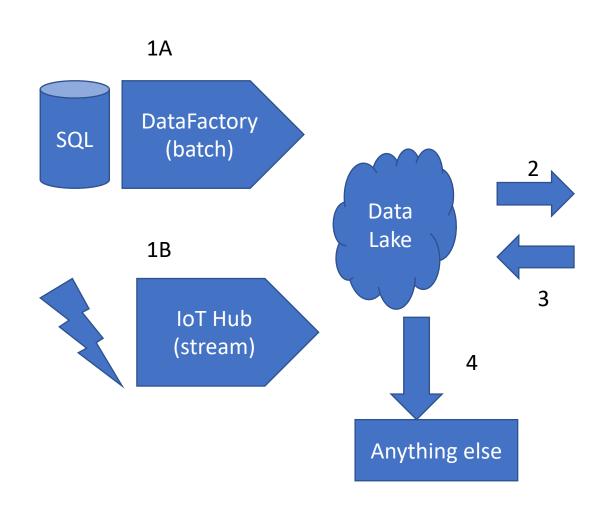
Beam = Batch + stEAM

### Kappa



Beam = Batch + stEAM

#### Option 1



#### Transformation

Native
Spark/Hadoop/Kafka/Hive
/Hbase/Sqoop/Zookeeper
- everything APACHE OSS
(HDInsight)

Only cloud-enabled Spark without anything else (Databricks)

Data warehousing SQL / Spark (Synapse Analytics)

No-code transformations (Data Factory)

