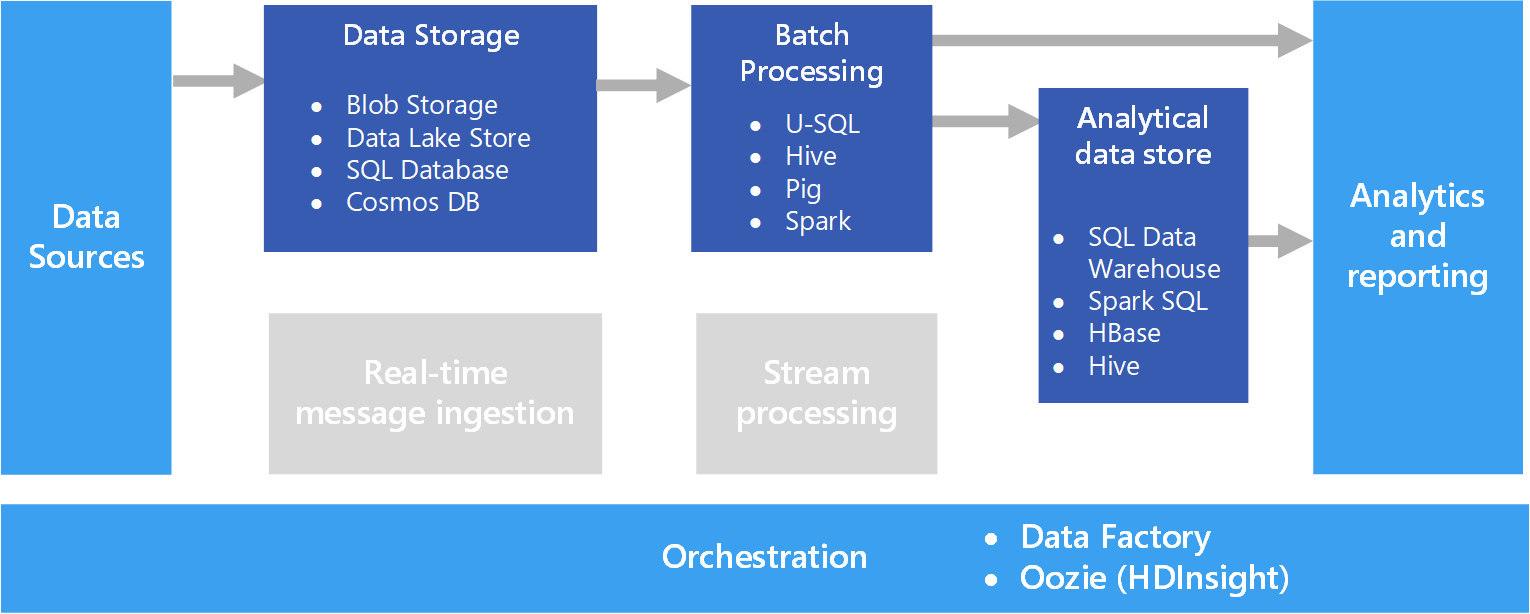
S555058

Batch Processing:

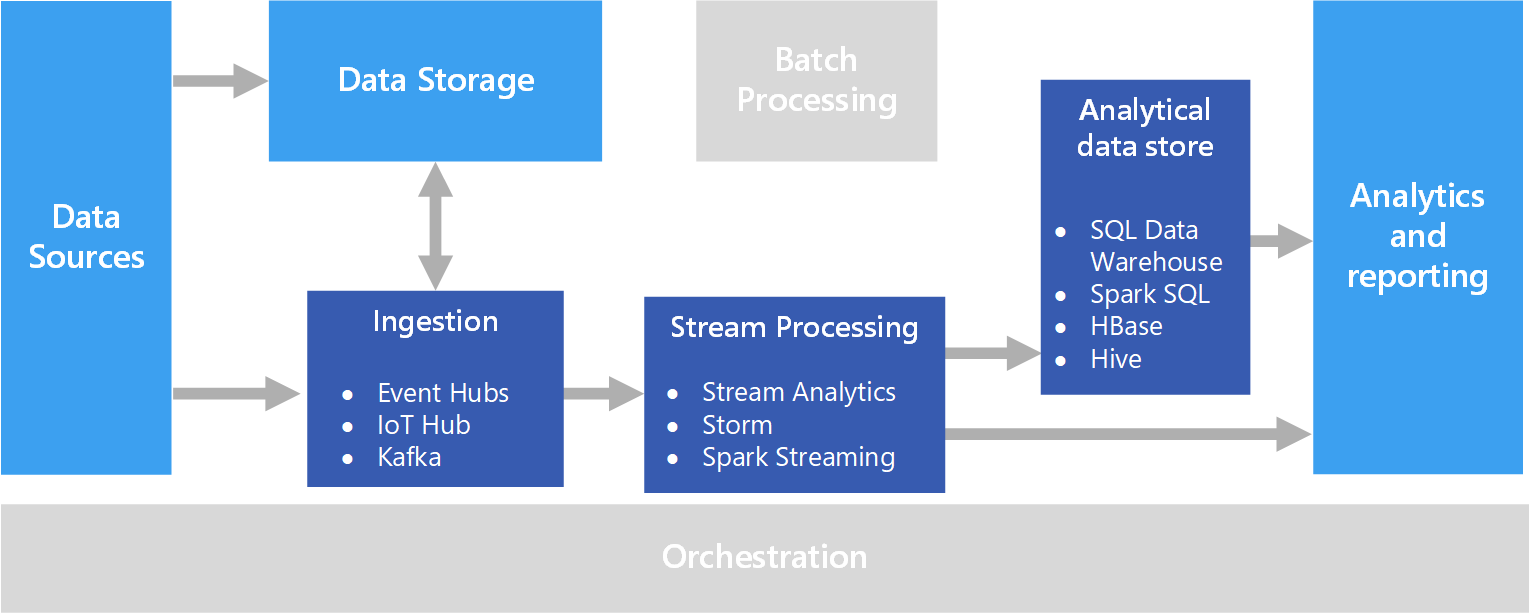
batch processing is transforming a large set of flat, semi-structured CSV or JSON files into a schematized and structured format that is ready for further querying. Typically the data is converted from the raw formats used for ingestion (such as CSV) into binary formats that are more performant for querying



Real-time Processing:

Real time processing deals with streams of data that are captured in real-time.

Example: real-time traffic monitoring solution might use sensor data to detect high traffic volumes. This data could be used to dynamically update a map to show congestion, or automatically initiate high-occupancy lanes or other traffic management systems.



**Apache Kafka:**

Apache Kafka is an open source distributed streaming platform that can be used to build real-time streaming data pipelines and applications. Kafka also provides message broker functionality.

Kafka is used to build real-time streaming data pipelines and real-time streaming applications. A data pipeline reliably processes and moves data from one system to another, and a streaming application is an application that consumes streams of data.

**Kafka Architecture:**

