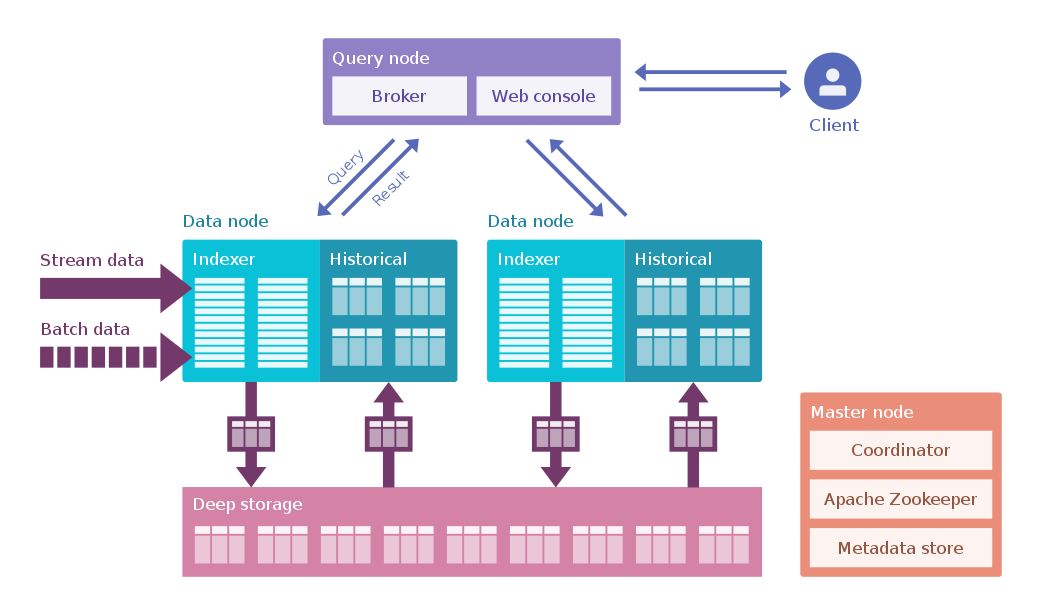
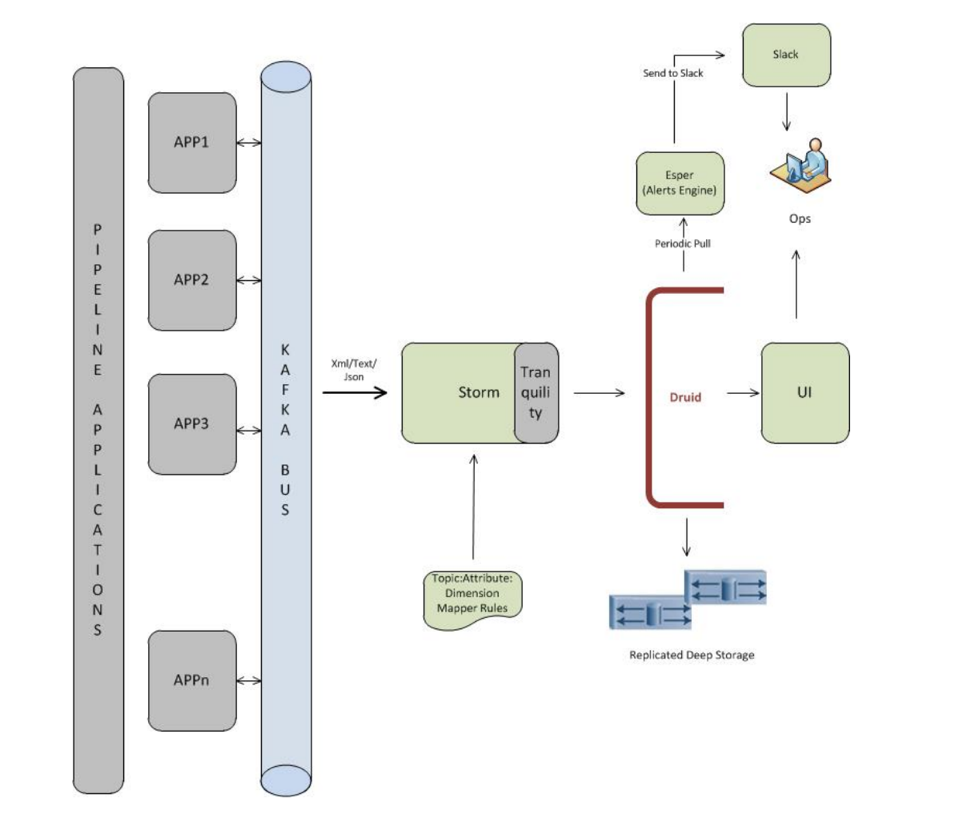
10/31 Big Data Challenge Round 2 – Druid

* Apache Druid is a column-oriented, open-source, distributed data store written in Java. Druid is designed to quickly ingest massive quantities of event data, and provide low-latency queries on top of the data
* Language: Java
* Operating system: Cross-platform
* Druid is commonly used in business intelligence/**OLAP** applications to analyze high volumes of real-time and historical data
* Fully deployed, Druid runs as a cluster of specialized processes (called nodes in Druid) to support a fault-tolerant architecture where data is stored redundantly, and there is no single point of failure
* The cluster includes external dependencies for coordination (Apache ZooKeeper), metadata storage (e.g. MySQL, PostgreSQL, or Derby), and a deep storage facility (e.g. HDFS, or Amazon S3) for permanent data backup
* Features
  + Low latency (streaming) data ingestion
    - Support streaming data ingestion and offers insight on events immediately after they occur
  + Arbitrary slice and dice data exploration
  + Sub-second analytic queries
    - Support multi-dimensional filtering, aggregation, and is able to target the very data to do query
  + Approximate and exact computations
  + Scalable
    - Able to deal with trillions of events for total, millions events for each second
  + Integrates natively with Kafka and Storm



Resources Referred:

<https://medium.com/walmartglobaltech/event-stream-analytics-at-walmart-with-druid-dcf1a37ceda7>

https://en.m.wikipedia.org/wiki/Apache\_Druid