Wyatt Blair

EMT-678-WS: Big Data Technologies

10/18/2024

**Class 6 Streams and Monk – How Yelp is Approaching Kafka in 2020**

This paper details how the application Yelp handles data-streams. Specifically, it details the interactions between Kafka and Monk in their system. They also indicate the fact that not all data-streams have equivalent needs. In general, they have two policies which data-steams are treated with: “ACK” and “FIRE\_AND\_FORGET”. Under the “ACK” policy, Yelp opts to show their user an error in the event of Kafka being unavailable. In other instances, if Kafka is unavailable the system may want to store that event and then eventually flush it to Kafka. For instance, if a user is checking in somewhere and Kafka is down, there is not much the system can do other than retrying. This distinction in data-streams seems to implicitly take the CAP theorem into account; with “ACK” streams valuing consistency over availability and “FIRE\_AND\_FORGET” streams availability over consistency.

The paper discusses a number of other useful data-stream management techniques such as assigning priority to particular Kafka clusters. Another technique involves the use of a Monk buffer which helps ease the flow of data being pushed to Kafka clusters at any given moment.