

## **Used Tech Stack:**

**Druid**

**Imply**

**Kafka**

**Zookeeper**

## **Reason to choose Druid and Kafka:**

First thing, we need to store the data on every minute. So we are going to store and process a huge amount of data. And also we need to process ~16k ops/second. So we need to collect data on real time on very high speed. And as kafka is able to process that huge amount of data in real time so I think kafka is a good choice. One other reason to use kafka is that we can ingest data from kafka to druid.

Druid is designed to store and analyse huge amount of streaming data. Also, we can also integrate Druid with other application. So it's easy to query druid with other application and can be used for fast analyses. There are lots of different query available like groupBy, topN, time boundry, timeserie etc.

## **Using Imply:**

Imply is high performance analytics solution to store, query and analyse data. And also it is built around druid, so when it come to designing and visualising data on real time IMPLY is the best solution. In imply, you can do lot of thing like creating data dashboard, datacube, running sql query on druid etc. So, this makes all thing very easy.

## **Reference:**

1. <http://druid.io/>
2. <https://imply.io/>
3. <https://blog.netsil.com/a-comparison-of-time-series-databases-and-netsils-use-of-druid-d-b805d471206>