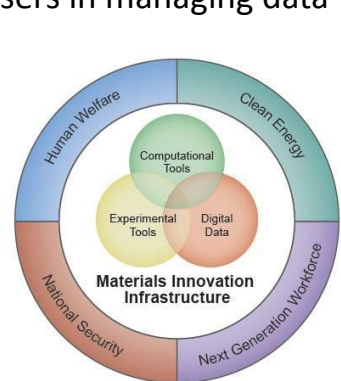


Development of a Cube.js Dashboard for Visualization of PARADIM Instrument Streaming Data

By: Avery Lenihan; Mentor: David Elbert; PI: Tyrel McQueen

Introduction

The Materials Genome Initiative(MGI)¹ centers on the production of shared data to accelerate the materials design loop. PARADIM is created a platform-wide streaming data ecosystem to automate production of FAIR data for users. This project investigated visualization of data flow in PARADIM infrastructure to aid staff and users in managing data



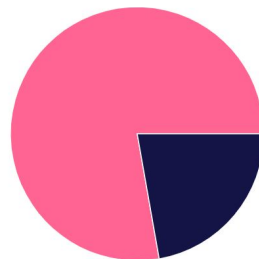
Development

- A dashboard was identified as adding value for users
- Multiple options were evaluated including Kowl, Confluent Cloud, and DataDog
- Cube.js was selected due to flexibility, power, and established feature set
- Kafka Eagle is leveraged to gather JMX metrics from the PARADIM brokers into a MySQL Database
- Version controlled schema development augmented automated Cube.js schema features

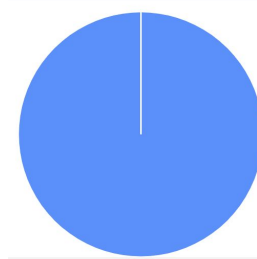
Conclusions

- Data tools provide API thst allow ready linking for new functionality such as a dashboard
- Manual schema production is critical to produce function.
- Cube.js is a flexible, powerful, and easy to implement solution
- Further development can produce a dashboard that is readily deployable and of value to PARDIM users and staff

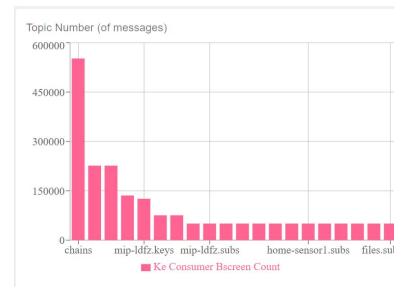
Cube.js Visualization Examples:



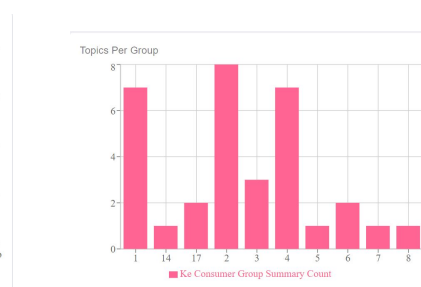
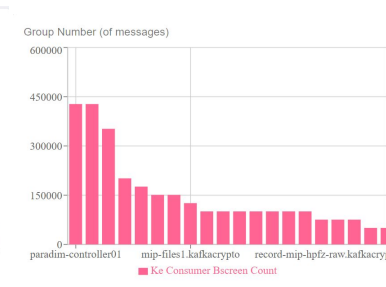
Distribution of Kafka and Zookeeper messages



Message distribution filtering out messages of value 0



Distribution of messages in topics & groups



Number of topics per group.