Module 1: Introduction to Big Data and Apache Kafka

Case Study

Case Study

CMS (Catalog Management System)

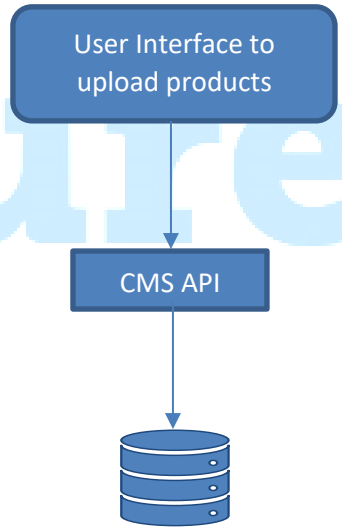
We have completed module 1 and now you must have a basic understanding of Kafka and its applications. So, it's high time to spend some time solving a real-time use-case, so that you have a fair amount of idea about why and how to leverage Kafka as a part of a solution strategy.

Statement:

John lives in a village and he submitted an idea to "Government Start-up Program".
He wants to build an eCommerce portal like Amazon, Flipkart or EBay where he will ask sellers/local brands to upload their products on portal from where users/buyers can visit and purchase them.

The government approved the idea and funded his project. Now, John doesn't have much knowledge about the system, so he hired some freelancers and they built basic version where anyone can upload products.

As the number of sellers increased the number of products also increased exponentially. This has left a negative impact on John's system, wherein sometimes his current system stops responding.



Existing System

This problem has also led to a decrease in the number of sellers uploading their products.

Luckily, John visits LinkedIn and finds you as a perfect fit to solve this problem.

You have gone through the current system catalog and collected products attributes which are:

- Product Name
- Product Description
- Product Code
- Quantity

- Color [Multiple]
- Size [Multiple]
- Brand [Multiple]
- Price
- MRP
- Category [Clothing, Footwear, Electronics, etc.]
- Sub Category [Shirts, Jeans etc.]
- Price
- Inventory
- Seller Code
- Seller Rating
- Image URL

You can add more attributes if you want. John has chosen some database to store all product data. Products volume can be large like 1 lakh products per minute or maybe more. We don't want some sellers/brands to wait while they upload products and products should also get stored in our system.

You need to redesign existing system architecture which takes care of

- Near Real time uploading of Products
- System failure in case of high load
- Handling of bulk upload functionality

While redesigning the above system you need to answer below mentioned questions

- How can you solve above problem using Kafka?
- If you think Kafka is not, then Why not?
- If you think it is perfect for the solution, then why?
- How will you handle such large catalog using Kafka?
- If products are continuously changing, how will you maintain latest state?

Considering all these facts mentioned above, you have to submit a high level design including flow diagram.