DataMaking



Building Real-Time Sales Dashboard using Apache Spark, Apache Kafka and Python Django for Aspiring Data Engineers



About Us

DataMaking

A online technology education company that develops personalized online courses and software solutions on Data Engineering and Data Science field.

Contact Us

- Website: https://www.datamaking.com
- Email: support@datamaking.com | datamaking.elearning@gmail.com
- YouTube: https://www.youtube.com/c/datamaking
- LinkedIn: https://www.linkedin.com/in/parimargu



Prerequisites: Environment Setup

Programming/Runtime

- Python
- Java
- Scala

Big Data/Distributed System

- Apache Hadoop
- Apache Spark
- Apache Kafka

RDBMS

MySQL

Web/UI

- Python Django
- Chart.js



Data Source

Message Layer

Data Processing/Data Analytics

Data Storage

Visualisation/UI



Step No. 1

Simulated Online Orders Data(Sales) to Kafka Topic Step No. 2

Apache Kafka acts as Message Broker Step No. 3

Building Data Processing Pipeline using Apache Spark(Python and Scala) and Apache Hadoop

- Data Ingestion
- Data Processing

Step No. 4

Aggregated Data is stored in MySQL Database and HDFS(Distributed Storage) Step No. 5













Step 1 of Building Real-Time Sales Dashboard

Data Source

Simulated Online Orders Data(Sales) to Kafka Topic

Kafka Producer



Data Source

Message Layer

Data Processing/Data Analytics

Data Storage

Visualisation/UI

Step No. 1

Simulated Online Orders Data(Sales) to Kafka Topic



Step No. 2

Apache Kafka acts as Message Broker Step No. 3

Building Data Processing Pipeline using Apache Spark(Python and Scala) and Apache Hadoop

- Data Ingestion
- Data Processing

Step No. 4

Aggregated Data is stored in MySQL Database and HDFS(Distributed Storage) Step No. 5













Step 2 of Building Real-Time Sales Dashboard

- Single Node Apache Kafka Cluster
 - Kafka Broker/Server
 - Kafka Topic
 - Kafka Producer
 - Kafka Consumer



Data Source

Message Layer

Data Processing/Data Analytics

Data Storage

Visualisation/UI

Step No. 1

Simulated Online Orders Data(Sales) to Kafka Topic Step No. 2

Apache Kafka acts as Message Broker



Step No. 3

Building Data Processing Pipeline using Apache Spark(Python and Scala) and Apache Hadoop

- Data Ingestion
- Data Processing

Step No. 4

Aggregated Data is stored in MySQL Database and HDFS(Distributed Storage) Step No. 5













Step 3 of Building Real-Time Sales Dashboard

- Single Node Apache Hadoop/Spark Cluster
 - Apache Hadoop
 - Apache Spark
 - Data Ingestion
 - Data Processing
 - Data Pipeline



Data Source

Message Layer

Data Processing/Data Analytics

Data Storage

Visualisation/UI

Step No. 1

Simulated Online Orders Data(Sales) to Kafka Topic Step No. 2

Apache Kafka acts as Message Broker Step No. 3

Building Data Processing Pipeline using Apache Spark(Python and Scala) and Apache Hadoop

- Data Ingestion
- Data Processing



Step No. 4

Aggregated Data is stored in MySQL Database and HDFS(Distributed Storage) Step No. 5













Data Source

Message Layer

Data Processing/Data Analytics

Data Storage

Visualisation/UI

Step No. 1

Simulated Online Orders Data(Sales) to Kafka Topic Step No. 2

Apache Kafka acts as Message Broker Step No. 3

Building Data Processing Pipeline using Apache Spark(Python and Scala) and Apache Hadoop

- Data Ingestion
- Data Processing

Step No. 4

Aggregated Data is stored in MySQL Database and HDFS(Distributed Storage)



Step No. 5











DataMaking



Questions?



Reach out to Us

Telegram Group

Group Link: https://t.me/+ulSEPVLP64BIMTc1

Contact Us

- Email: <u>support@datamaking.com</u> | <u>datamaking.elearning@gmail.com</u>
- Website: https://www.datamaking.com
- WhatsApp: +91 6380745611
- YouTube: https://www.youtube.com/c/datamaking
- LinkedIn: https://www.linkedin.com/in/parimargu

DataMaking



Suggestions?





Thank you!