**Title : Real Time Smart Bank Data Streaming Capture**

**GROUP - 03**

Sindhu Nagesha (017419987)

Prayag Nikul Purani (017416737)

Syed Faraaz Ahmed (017428619)

Sai Vivek Chunduri (017435301)

Department of Applied Data Science

San Jose State University

DATA 228: Big Data Tech and Applications

Submitted to: Vishnu S Pendyala

March 29, 2024

### **1.Progress towards the goal achieved so far**

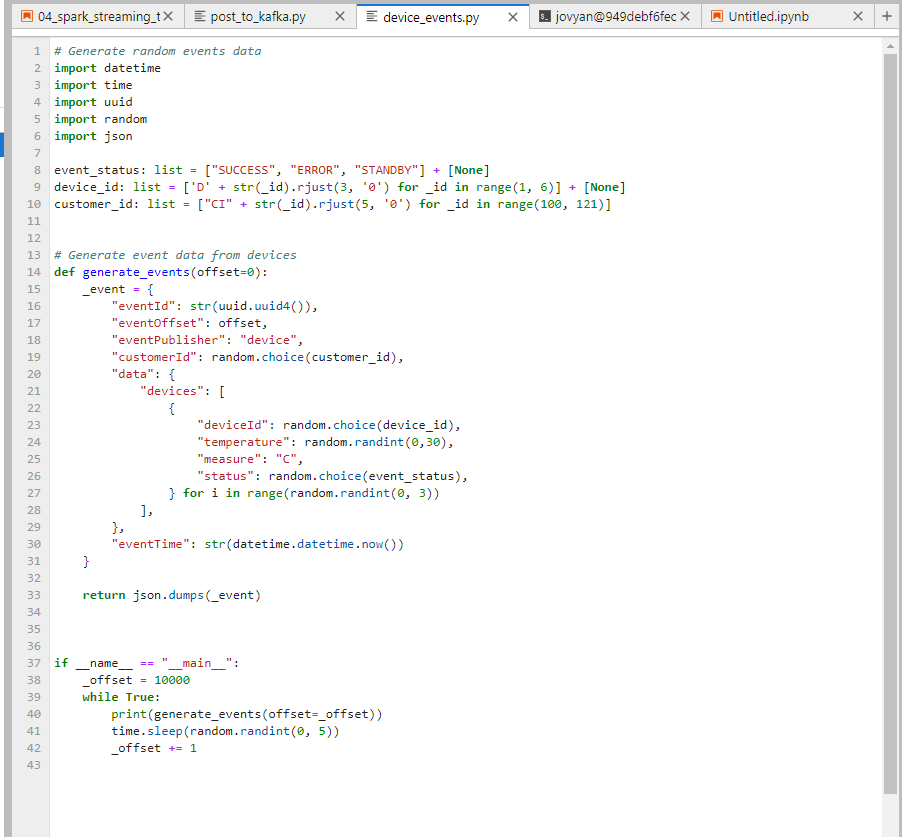
### 

So, we are ready with your integration with kafka and the data is produced from the real-time application and is ready to be written to kafka clusters and the data is efficiently read by the python in jupyter notebook.

* Integration with Kafka: We have successfully integrated our system with Kafka, a distributed streaming platform. This integration allows us to produce and consume real-time data efficiently.



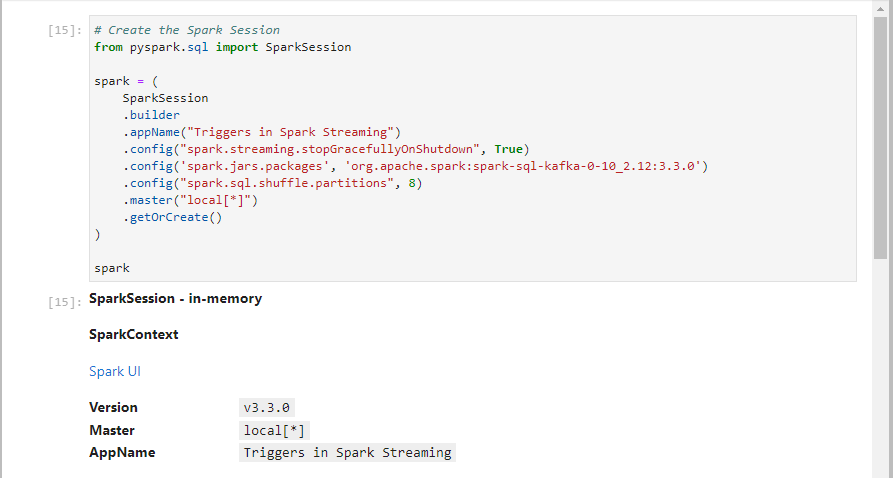
* Data Production: The real-time application responsible for generating bank data is up and running smoothly. It produces data streams that are ready to be written to Kafka clusters.



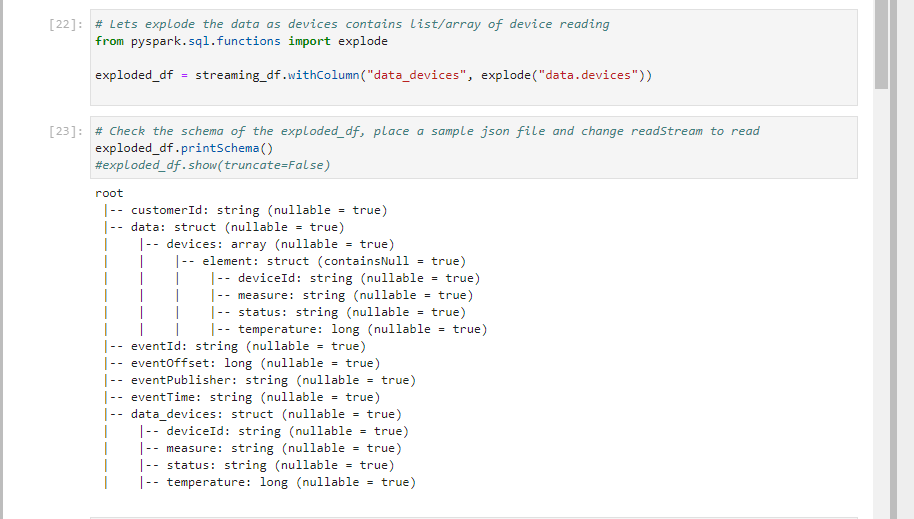
* Kafka Cluster Setup: Our Kafka clusters are configured and operational. They are capable of handling the incoming data streams effectively, ensuring high availability and fault tolerance.

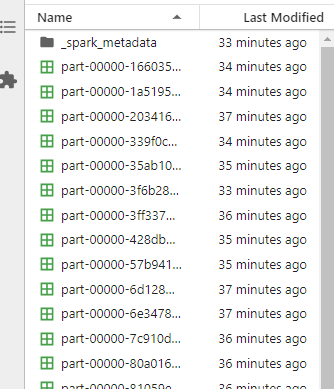
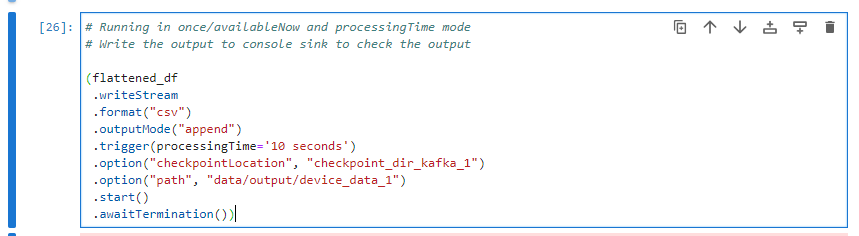


* Data Consumption: Python scripts running within a Jupyter notebook environment are set up to efficiently consume data from Kafka. These scripts are capable of processing the streaming data in real-time.









* Testing and Validation: We have conducted extensive testing to ensure the end-to-end functionality of our system. This includes testing data production, Kafka integration, data consumption, and data processing within the Jupyter notebook environment.

**2. Findings / results so far**

Integration with Kafka:

Our team has successfully integrated our system with Kafka, enabling seamless data streaming from our real-time application to Kafka clusters. This integration ensures efficient and reliable transfer of data, laying the foundation for further analysis and processing.

Data Production and Consumption:

Data from our real-time application is being consistently produced and written to Kafka clusters. Python scripts within Jupyter Notebook environments are effectively consuming this data, enabling real-time analysis and insights.

**3. Difficulties being encountered and how we planned to resolve them**

The main difficulties that have arisen are the connection of the kafka and the API for the data and the solution for this which we came up with was the docker image containing all the things like kafka, jupyter Note-book, zookeeper and the after this in future we will add all the other to it and integration parts to the docker only so there will not be a version issue and the integration of the program will be a easier task.

### **4. Remaining tasks**

### Data Processing and Analysis: Develop Python scripts within the Jupyter notebook to analyze and process the incoming bank data. This could involve tasks such as real-time fraud detection, customer behavior analysis, or financial trend prediction.

### Visualization: Implement visualizations within the Jupyter notebook to present the analyzed data in a clear and insightful manner. This could include interactive charts, graphs, and dashboards.

### Scalability and Optimization: Continuously monitor and optimize the performance of our system to handle increasing data loads efficiently. This may involve scaling Kafka clusters, optimizing data processing algorithms, and improving resource utilization.

### Security Measures: Implement robust security measures to protect sensitive bank data throughout the entire data streaming pipeline. This includes encryption, access control, and compliance with relevant data privacy regulations.

### Deployment and Monitoring: Deploy the system into a production environment and establish monitoring mechanisms to ensure its smooth operation. This involves setting up alerts, logging, and performance monitoring tools to detect and address any issues promptly.

* Sensor data monitoring: Develop visualization dashboards or interfaces to display the sensor data in a user-friendly and intuitive manner. This could include real-time charts, graphs, maps, or custom visualizations tailored to the specific sensor data and monitoring objectives.
* Visualization: The last and most important step in the project is to visualize the data and to get some insight from the data and to plot some real-time graphs.

### **5. Any others that we think is relevant**

Personalized Financial Insights: Utilizing artificial intelligence and data analytics to provide personalized financial insights to customers in real-time. This could include tailored budgeting suggestions, investment recommendations, or alerts about upcoming bills or expenses based on the customer's spending patterns and financial goals.

Automated Financial Health Monitoring: Developing systems that continuously monitor customers' financial health by analyzing their income, expenses, savings, and investment activities in real-time. These systems can provide proactive alerts and recommendations to help individuals manage their finances better and achieve their financial goals.