

Steps to Run SparkAB

Step 1: Launch Kafka

Launch Kafka by running the script `start-kafka.sh`. You can do this by opening a terminal, navigating to the directory containing the script, and then entering the command:

Unset

```
./start-kafka.sh
```

Step 2: Run the Program

In a new terminal window, navigate to the root folder of your project and run PySpark. You can do this by entering the following command:

Unset

```
pyspark
```

Steps 3: Run DataShow.ipynb

Open the Jupyter notebook titled `DataShow.ipynb` in the root folder of your project. Execute all the cells in the notebook. To do this, you can click the 'Run' button in the notebook's toolbar.

Step 4: Run Spark_Consumer

Navigate to the root folder of your project and open `Spark_Consumer.ipynb`. Run all the cells in this notebook just as you did in Step 3.

Step 5: Run Kafka_Producer

Next, open `Kafka_Producer.ipynb` in the same manner as the previous steps and execute all the cells in the notebook.

Step 6: Observe the Output

Return to the `DataShow.ipynb` notebook. You should now see new information displayed every 2 minutes.