Data Streaming Application in kafka and Spark Processing

This is a standalone application for stream data processing from source, process the data in pyspark and load the data as JSON in the local directory.

The data pipeline is given below,

Process Flow Diagram

| Data from source system's to KAFKA | → | Subscribe the topics from Kafka and dump in a file. | → | PySpark process the data and dump the final JSON | → | BAU Monitor process loads the dashboard | → |