**Project** : **Streaming data pipeline and persisting data to Nosql db**

This program reads psv and JSON format data and process it and persist into postgres(Nosql) database.

I have created a streaming data pipeline using kafka and spark structured streaming. FileReader reads data from two different file formats one is psv and other is json data and process it based on the data type , applying filters on data and writes data to postgres jdbc sink for persistence.

Below are the steps and screenshots of outputs:

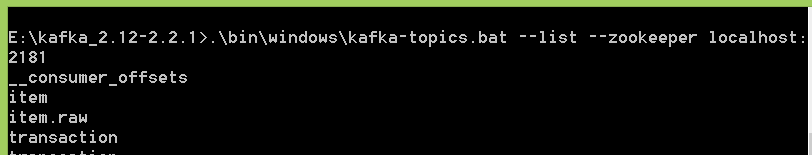
* Started zooker server using below command

**.\bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties**

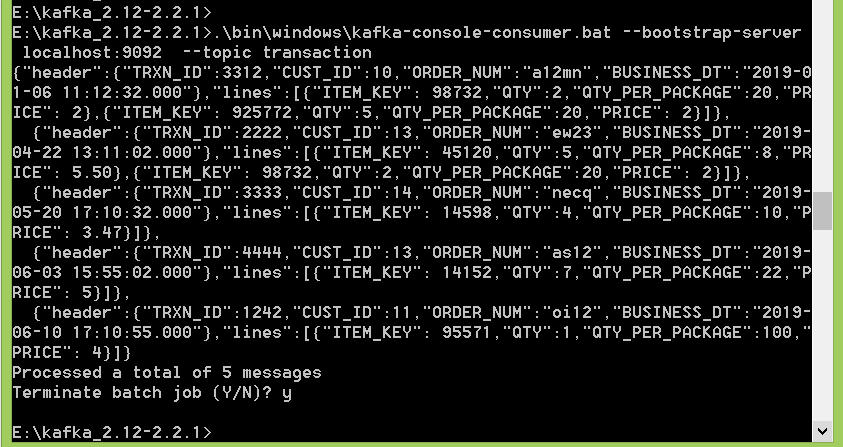
* Started kafka server

**.\bin\windows\kafka-server-start.bat .\config\server.properties**

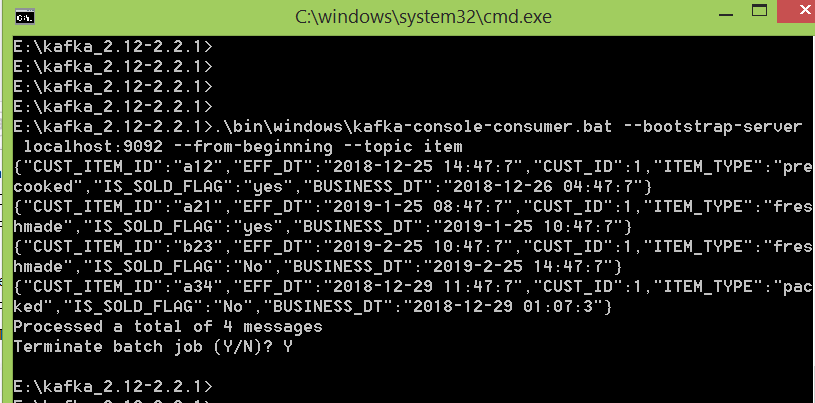
* intsall postgres DB into your loacl system
* Re-Build the project
* Execute FileReadertest and FileWritertest classes
* Kafka topics: item and transaction topic are created manually and executed my unittest classes , loaded data to kafka topics as shown below :



transaction topic data: json format



item topic : psv data:



Filewriter class created tables and ingested as shown below screenshots:

item, header tables got created with columns as motioned schema in Dataschema class , line will have one additional column from header to join them for further process .

