I have used the Nifi , Kafka to ingest data from the dataset & tried to understand the sales data in real time.

The E-commerce data set was used for this purpose.

**Operations / Transformations :**

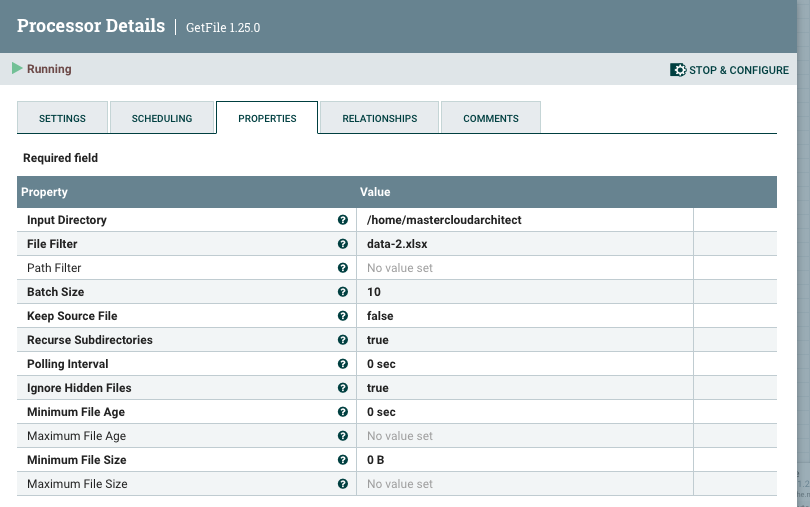
Two workflows were created - Publishing and Consuming workflow.

This is because in Apache NiFi, it is not possible to directly connect a PublishKafka processor to a ConsumeKafka processor.

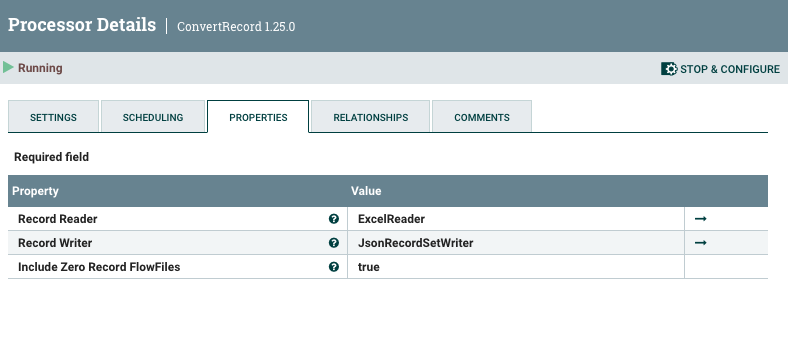
NiFi is designed to handle data flows where data is consumed, processed, and then output to various destinations. Therefore the data is published to Kafka and then a separate consumer flow is used to consume and process the data.

**Dataflow / Processors to Publish Data to Kafka**:

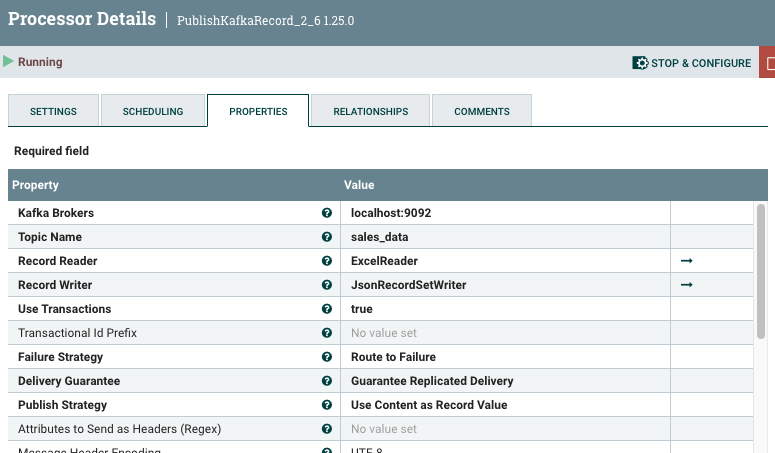
1. **GetFile Processor**: Reads data from a file.
   * **Input Directory**: /home/mastercloudarchitect
   * **File Filter**: data-2.xlsx



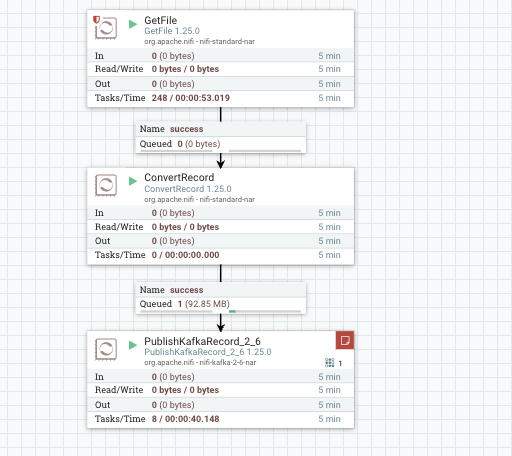
1. **ConvertRecord Processor**: Converts data format.
   * **Record Reader**: ExcelReader
   * **Record Writer**: JsonRecordSetWriter



1. **PublishKafka\_2\_6 Processor**: Publishes data to Kafka.
   * **Kafka Brokers**: localhost:9092
   * **Topic Name**: sales\_data

****

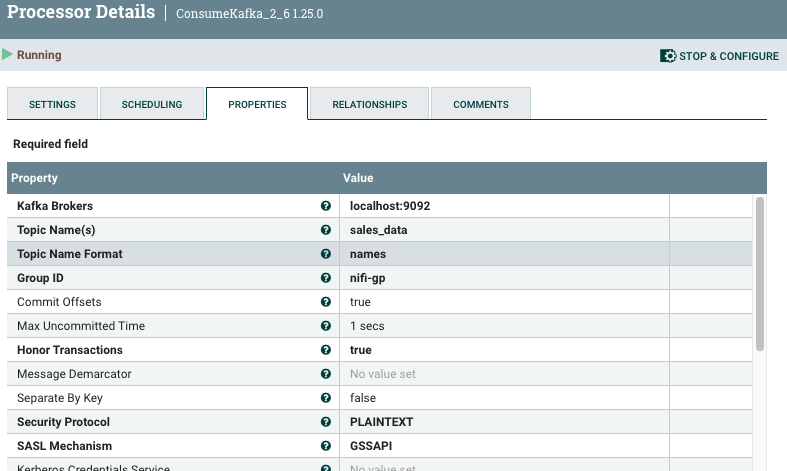
**GetFile** → **ConvertRecord** → **PublishKafka\_2\_6**



 **Dataflow / Processors to Consume Data from Kafka**:

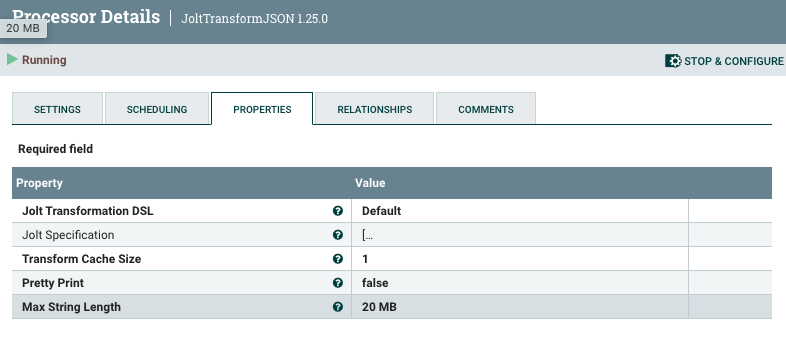
1. **ConsumeKafka\_2\_6 Processor**: Consumes data from Kafka.

* **Kafka Brokers**: localhost:9092
* **Topic Name**: sales\_data
* **Group ID**: nifi-group



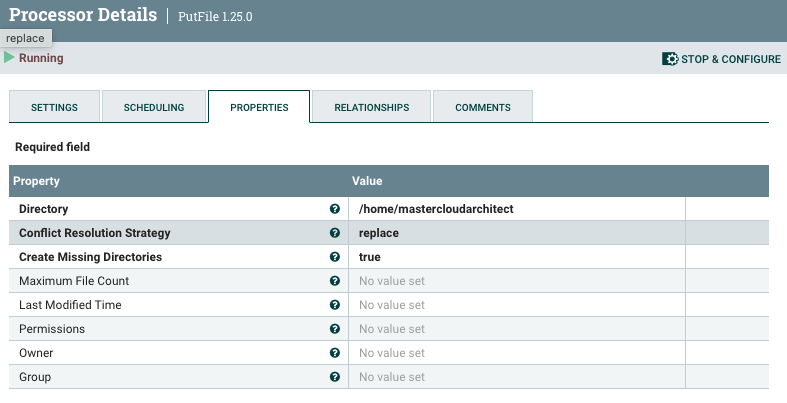
1. **ConvertRecord Processor**: Converts data format (if needed).

* **Record Reader**: JsonTreeReader
* **Record Writer**: ExcelRecordSetWriter (or any other format you prefer)

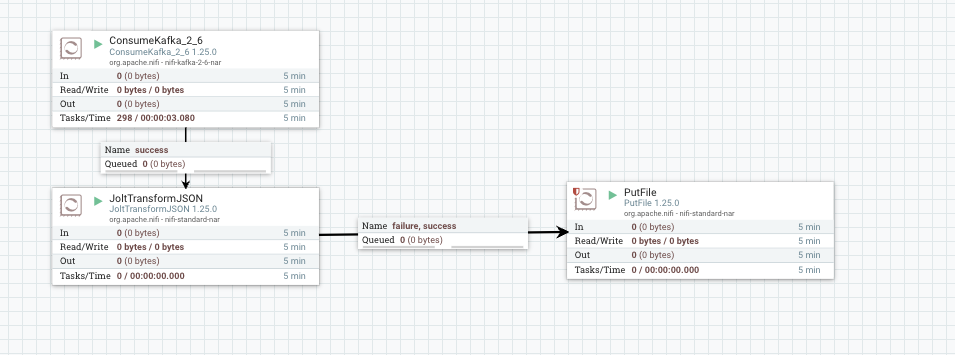


1. **PutFile Processor**: Writes data to a file.

* **Directory**: /home/mastercloudarchitect
* **Conflict Resolution Strategy**: replace



**ConsumeKafka\_2\_6** → **ConvertRecord** → **PutFile**

****