Instructor Notes: DI-1500 v9.1

There have been a few minor changes to update the course for v9.0.

Please follow the instructions outlined below to ensure you can complete the required labs:

If you need to reset the course files:

1. In a terminal:

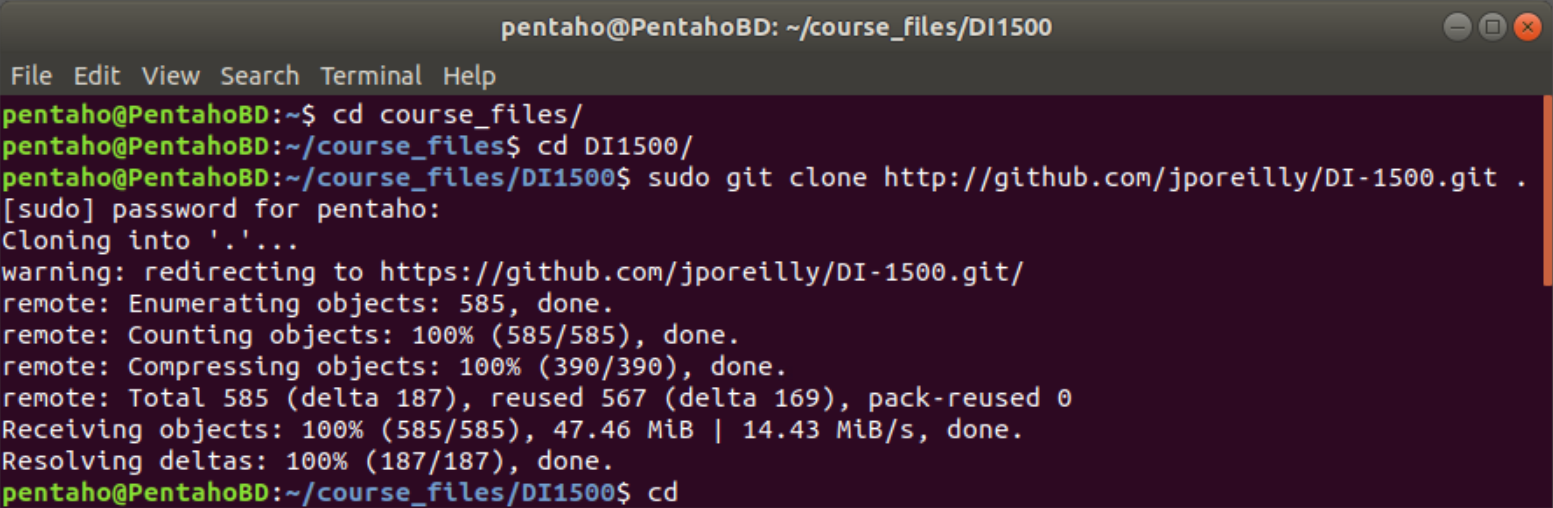
$ cd course-files

1. Delete the existing DI1500 folder:

course-files$ rm -rf DI1500

1. Clone GitHub repo:

$ sudo git clone <http://github.com/jporeilly/DI-1500.git> DI1500



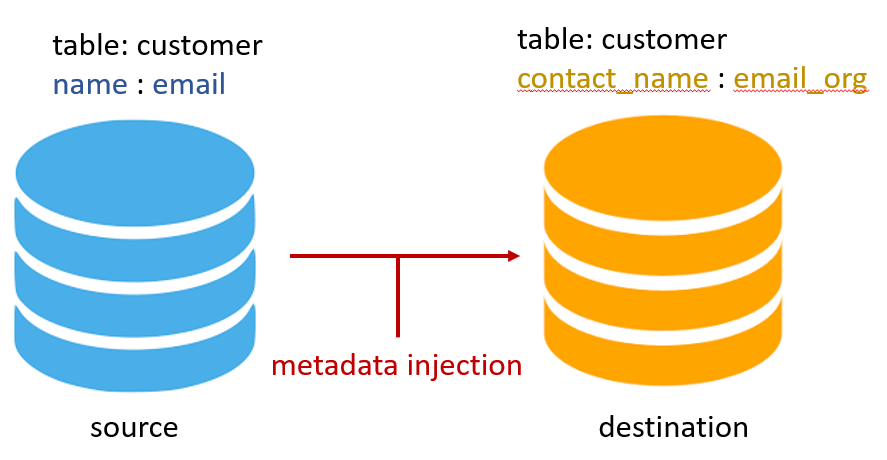
This will clonethe latest course files. The window will automatically close once completed.

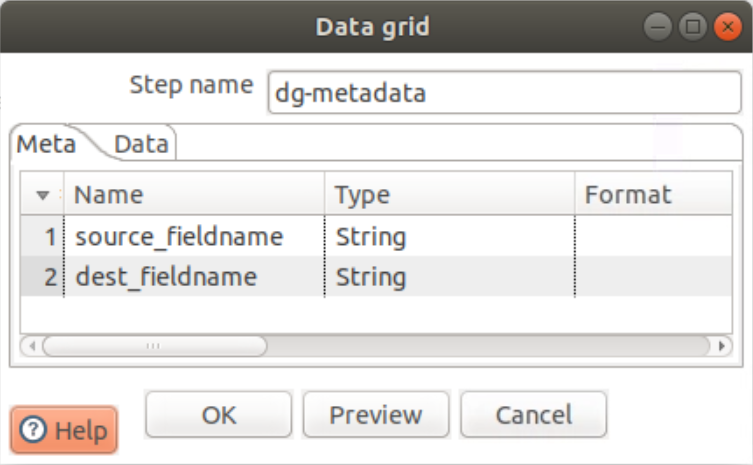
1. Check that the course files have been updated.

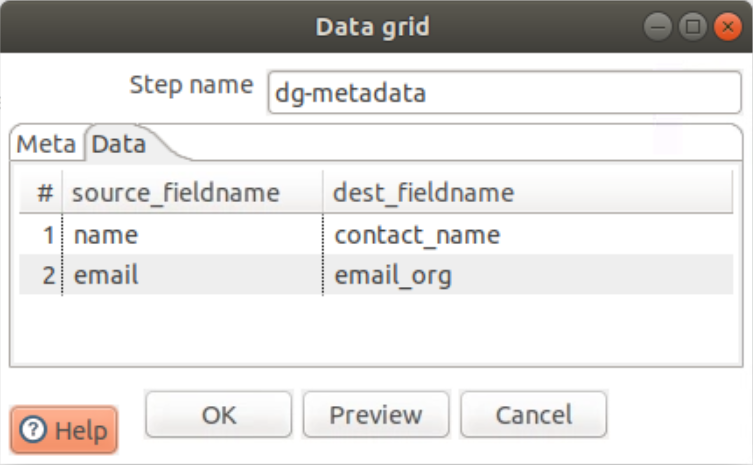
Module 1 – Metadata Injection

The data has been changed, so that you can introduce a database scenario. In moving data from one staging area to another the table properties may need to be changed (column names, varchar, etc..).

In these series of Labs you will be onboarding some data where the column names need to changed to match the destination.







Module 2 – PDI as a Data Source

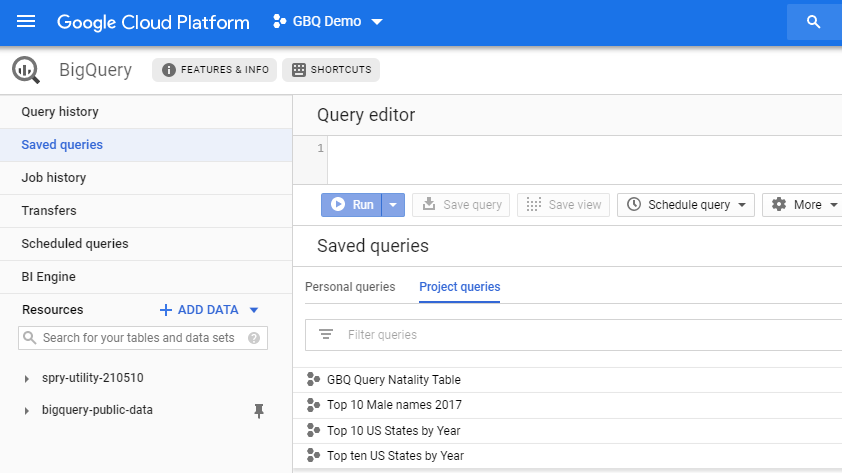
CDA

From October 14th APIXU relaunches as WeatherStack. The tr\_weather\_cda.ktr has been updated to reflect the changes in api.

Google BigQuery

The lab should only be demonstrated by the Pentaho Instructor. Due to 2 factor authentication process, you may have to show pptx deck of GCP

The following file will help execute example queries. If you switch to the new UI you will find the queries saved.



The Transformation has been updated to include the required account authentication email and path to key.

Account details:

Module 3 – Streaming Data

**MQTT – Mosquitto**

The Mosquitto Broker service now starts automatically.

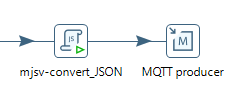
Added: Use Case – Logistics Sensor Data

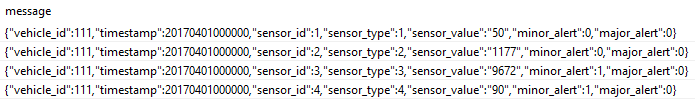
Note: This transformation will only be added if the reset script is RUN. The output can be used as a Pentaho Data Service to populate a C-Tools dashboard.

**Logistics Sensor Data**

tr\_mqtt\_producer.ktr

the sensordata stream is streamed in JSON format



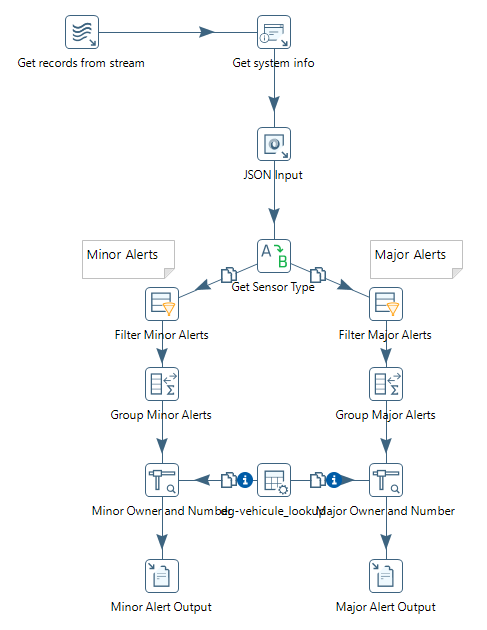


tr\_mqtt\_consumer.ktr

By default the consumer is processing:

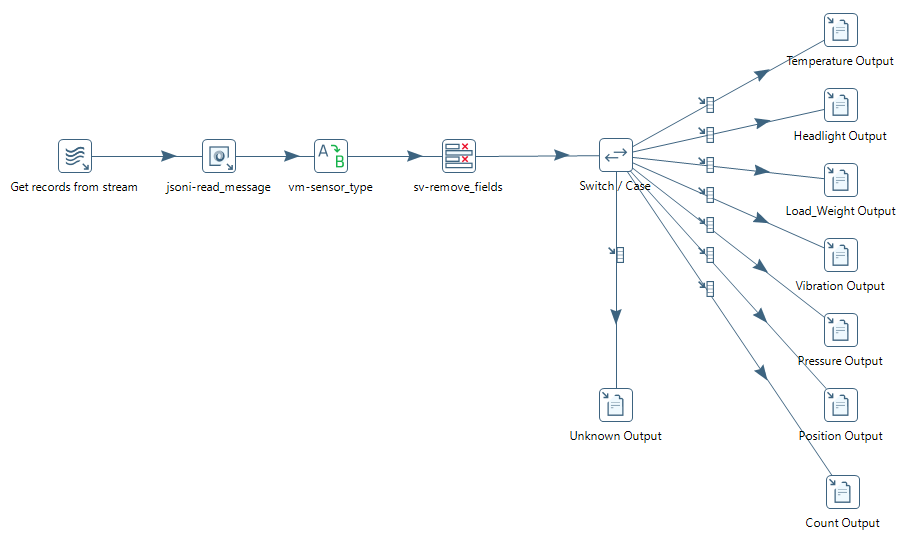


tr\_process\_alert\_data.ktr



tr\_process\_sensor\_data.ktr

Or point the consumer step to..



**Apache Kafka**

Confluent distro has been replaced with the docker image: Apache Kafka + KafDrop UI. Docker will start automatically.

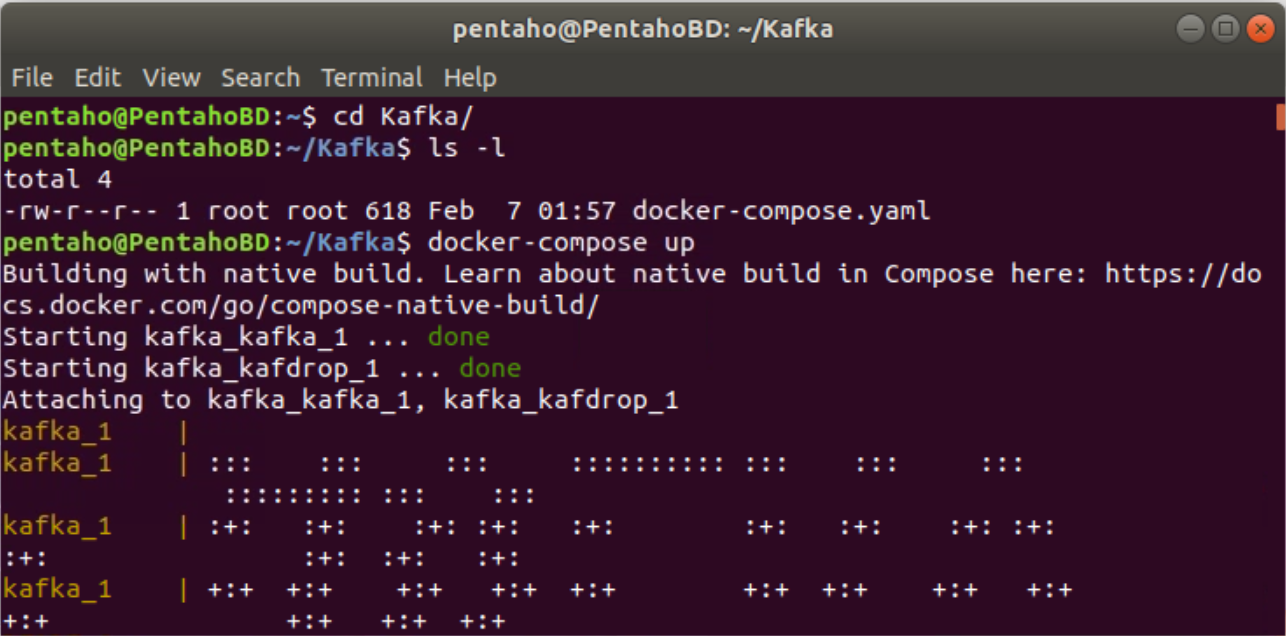
Check Docker is up and running:

$ systemctl status docker



To load image:

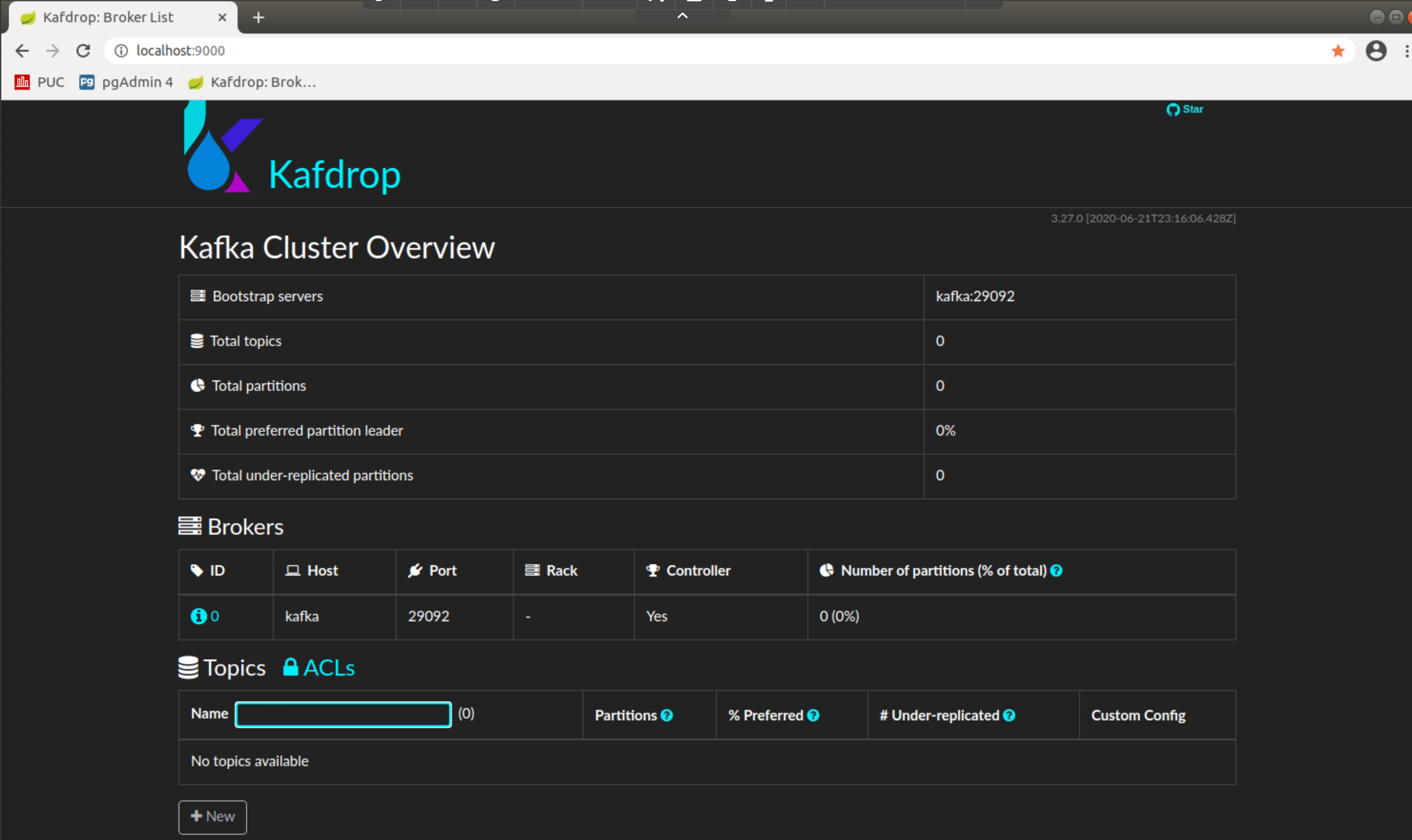
$ cd Kafka

$ docker-compose up

In your Chrome browser:

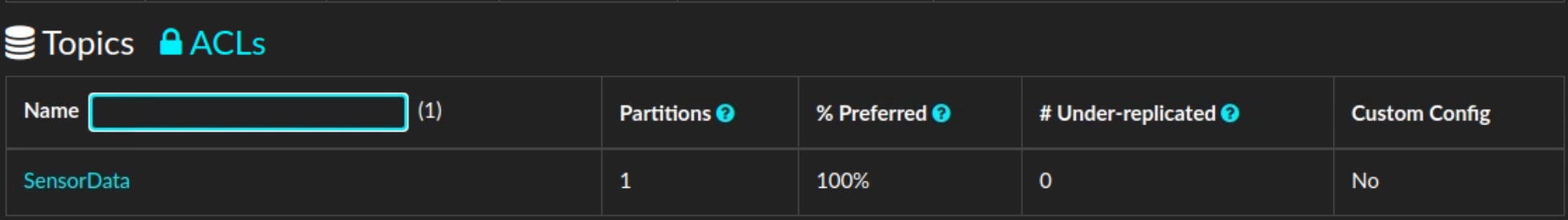
<http://localhost:9000>

The tool displays information such as brokers, topics, partitions, consumers (including lag) and lets you view messages.

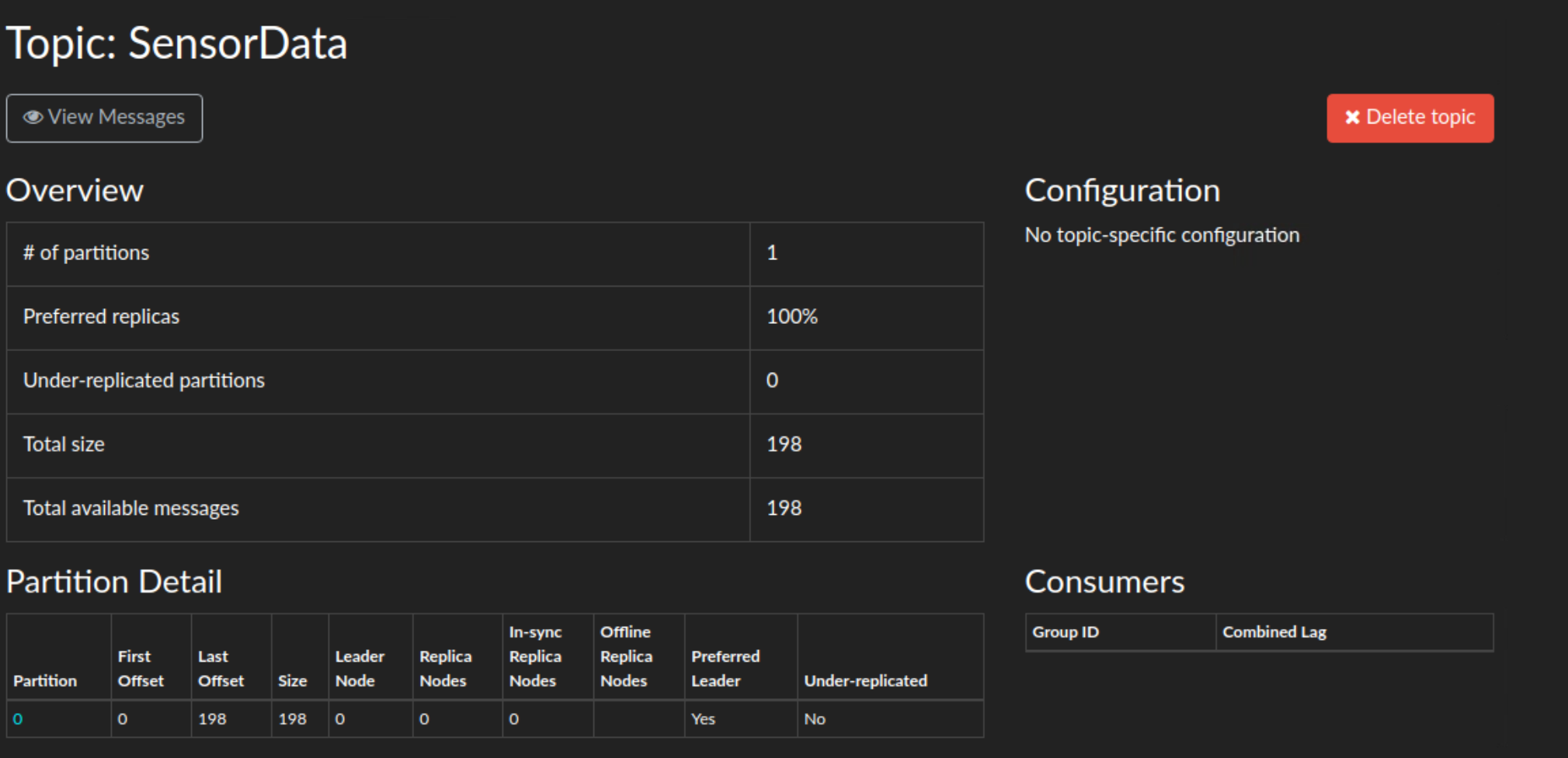


Run transformation

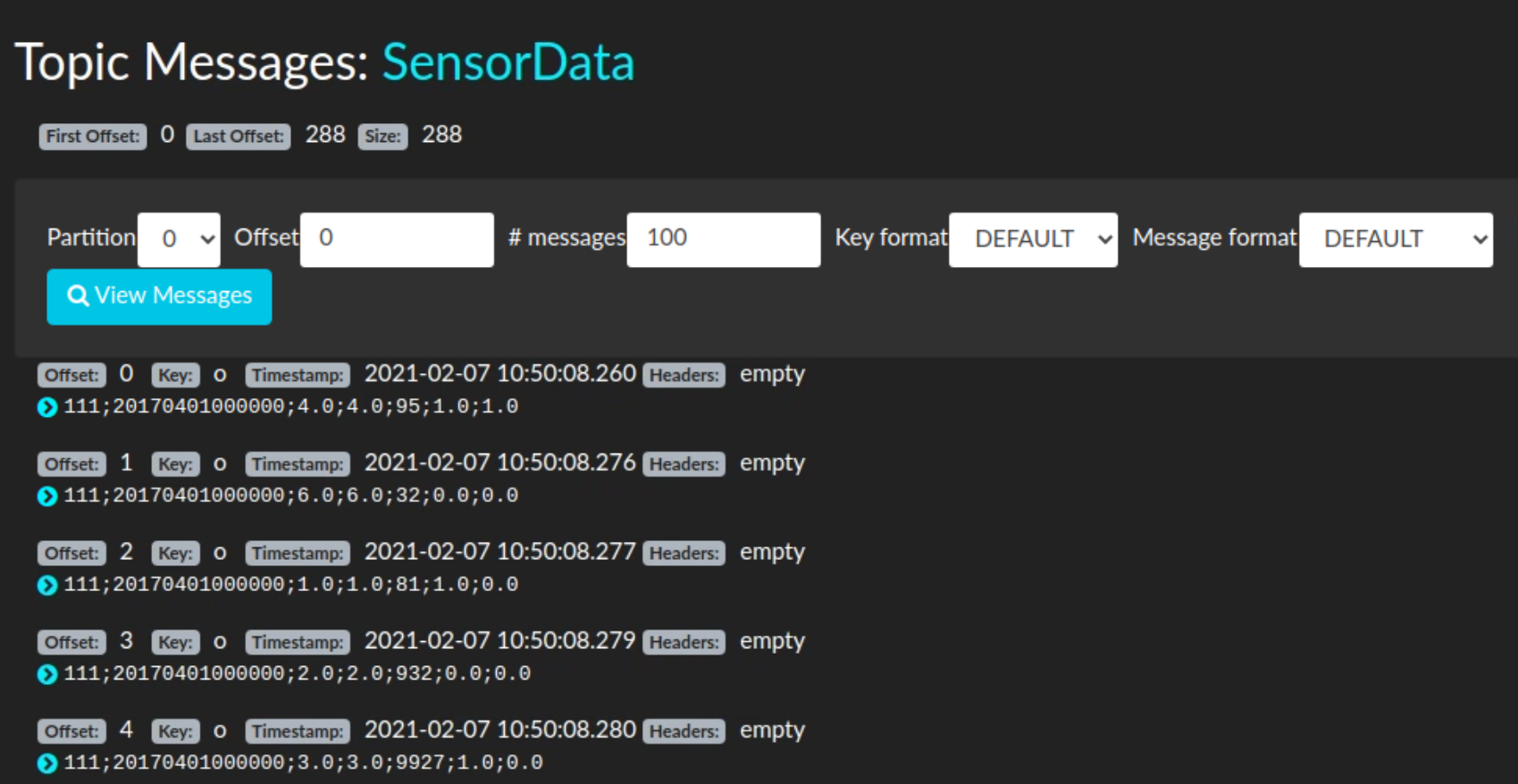
Click on Topic: SensorData



Click on View Messages



Check messages



Module 4 – Scalability

Lab: ‘Round Robin or Copy’ has been removed from the course files but kept in the pptx.

Lab: ‘Transactional Transformations’ has been removed from the course files but kept in the pptx.

kettle.properties file has been modified to include checkpoint logging