

Creating a Data Streaming

1. Create a kafka topic via cloudera@quickstart terminal:
\$ kafka-topics --create --zookeeper quickstart.cloudera:2181 --replication-factor 1 --partitions 1 --topic topic0302
2. Check if the topic was created via cloudera@quickstart terminal
\$ kafka-topics --list --zookeeper localhost:2181
3. Create a kafka-producer and kafka-consumer using jupyter-notebook in the my_root environment
4. Consumer script: consumer%2Bpredictive%2Brunning-version-0402-lmplot-all-test+final.ipynb
5. Producer script: producer-running-0302+fulldataset.ipynb
6. Run the producer script (sometimes it does not start, so try running the consumer first and then producer)
7. Monitoring the streaming via the terminal:
\$ kafka-console-consumer --zookeeper localhost:2181 --topic topic0302 --from-beginning