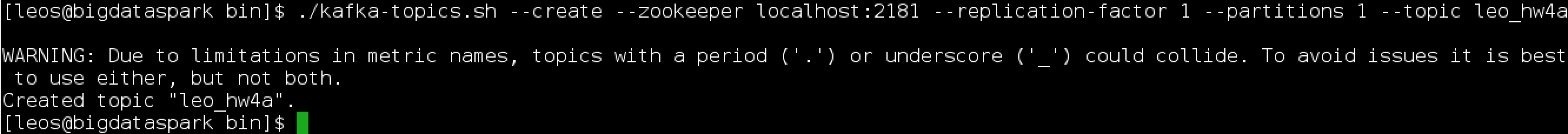
BIGDATA 220B HOMEWORK 4

Kafka Revisited

|  |  |  |
| --- | --- | --- |
| Leo Salemann | 8/23/17 | [Leo.salemann@me.com](mailto:Leo.salemann@me.com)  Leos@uw.edu |

# Create the Kafka Topic



# Modify the Consumer (station\_counter.py, pattern=re.compile('Piccadilly Circus')

# 

Start the Consumer



# Start the Producer



# After re-running the consumer, re-routing output to an external file ...



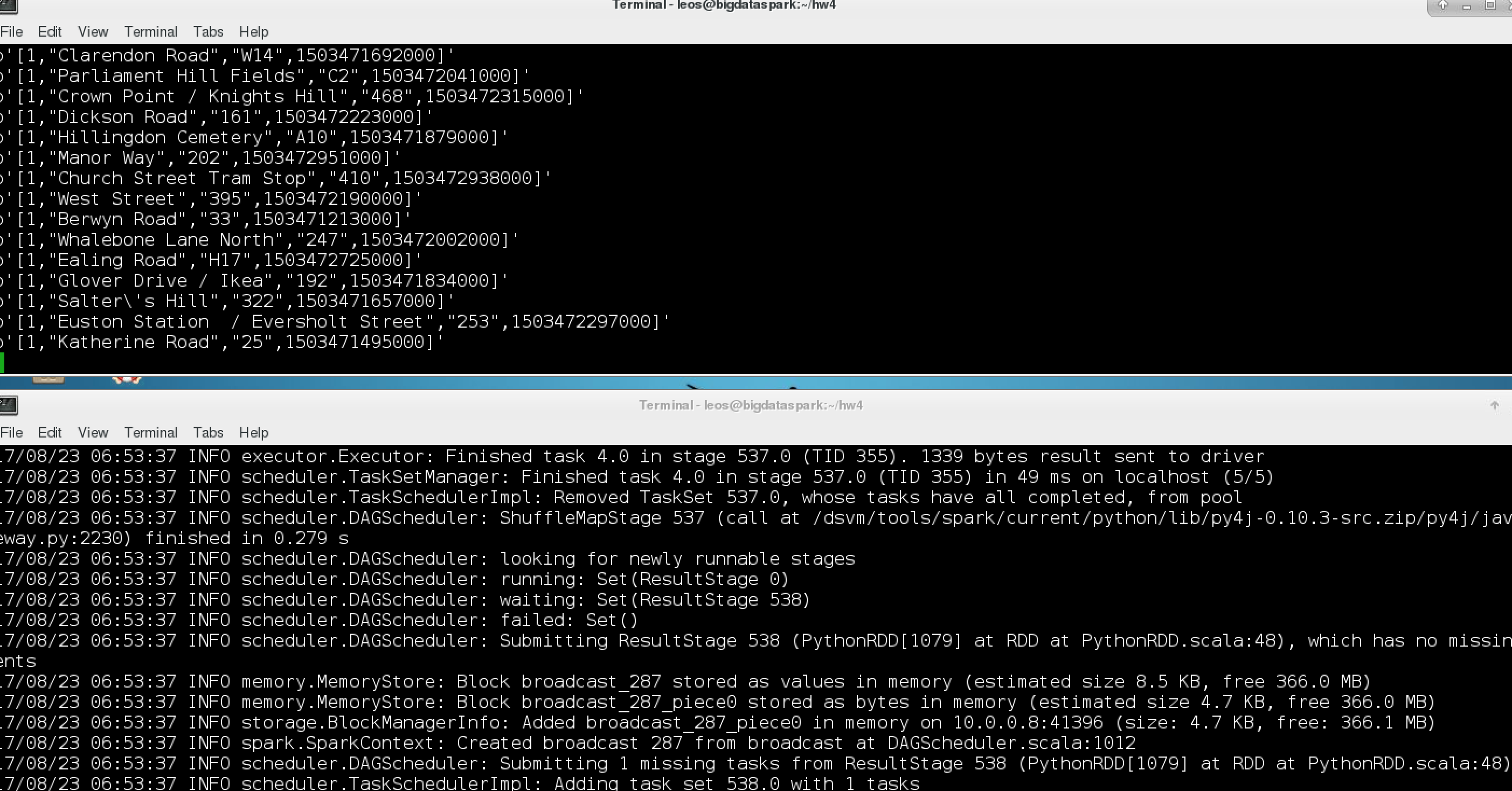
# … I finally got some output.



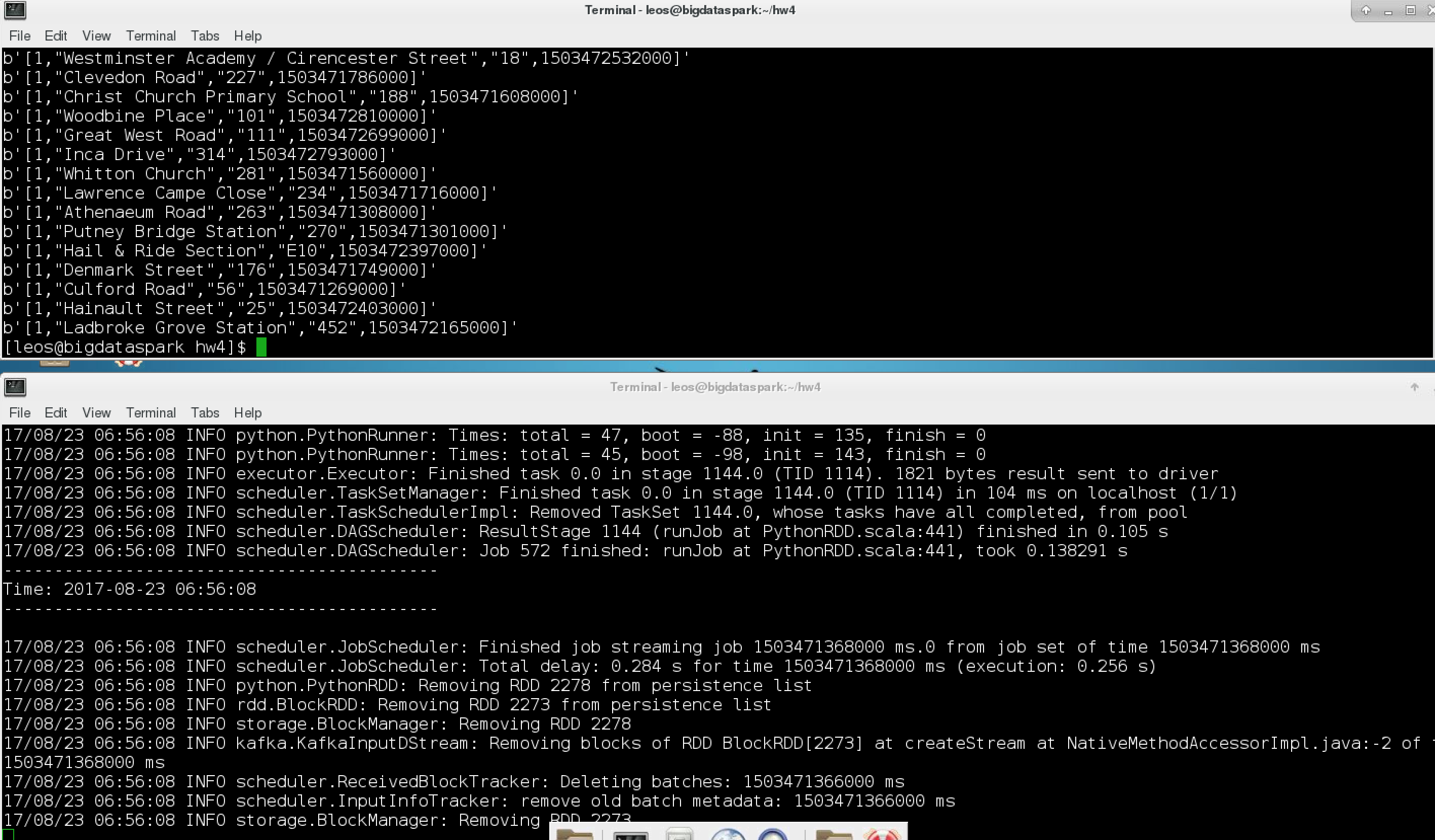
# … But most of the time, this happened.

After this, the screen is awash in diagnostic data, with no hope of a discernable result.

Shortly after running ...



Producer finished, consumer/spark job keeps going ..



This is the closest thing that bore any resemblance to an output

