Agenda

Spark Streaming with Kafka

[CSCIE63 2016, by Marina Popova]

1. Kafka 0.8.2.2 setup
   a. MacOS
   b. **CentOS6.7 VM + JDK1.8 (not Cloudera Quickstart VM)**
2. Kafka administration demo
   a. Topic creation/status
   b. Cluster state inspection
   c. Log/offset management and inspection
3. Kafka console consumer and producer demo
4. Kafka-python lib:
   a. Installation
   b. Producer script demo - send Apache logs to Kafka
5. Spark Streaming apps with Kafka:
   a. Eclipse setup with Maven
   b. WordCount version - reading data from Kafka
   c. Windowed aggregation example: aggregate Apache logs by client Ips (data is produced by the Kafka-python app - into Kafka)
   d. Running locally (MacOS) and on a VM