## HU Extension Assignment 09 E63 Big Data Analytics

### Handed out: 10/28/2017 Due by 4:00 PM EST on Saturday, 11/04/2017

**Problem 1.** Install Kafka on you Linux VM**.** If on your own VM with CentOS7.4 you should be able to install Kafka using yum:

$ sudo yum install kafka

Kafka code is most probably installed in the directory /usr/hdp/current/kafka-broker. Create an environmental variable KAFKA\_HOME pointing to that directory. Place the directory /usr/hdp/current/kafka-broker/bin in the PATH variable in the .bash\_profile file in your home directory. Source .bash\_profile (e.i. issue command $ source .bash\_profile ), so that you can invoke Kafka scripts from any directory. Make sure that Zookeeper server is started. Kafka configuration files reside in the directories: $KAFKA\_HOME/config. Create a topic**.** Demonstrate that provided scripting clientkafka-console-producer.sh receives and displays messages produced by kafka-console-consumer.sh client.

**Problem 2**. Make supplied python script kafka\_consumer.py receive messages produced by supplied python script kafka\_producer.py. Modify kafka\_producer.py so that you can pass server name and the port of the Kafka broker and the name of Kafka topic on the command line. Also, modify that script so that it continuously reads your terminal inputs and sends every line to Kafka consumer. Demonstrate that kafka\_consumer.py can read and display messages of modified kafka\_producer.py. Provide working code of modified kafka\_producer.py. Describe to us the process of installing Python packages, if any, you needed for this problem.

**Problem 3.**  Rather than using splitAndSend.sh bash script to generate traffic towards Spark Streaming engine, write a Kafka Producer which will read orders.txt file and send 1,000 orders to a Kafka topic every second. Write a Kafka consumer that will deliver those batches of orders to Spark Streaming engine. Base your Kafka consumer on provided direct\_word\_count.py script. Let Spark streaming engine count the number of orders different stocks where bought in each batch. Display for us a section of results in your solution. Describe to us the process of installing and invoking Python packages, if any, you needed for this problem.

**Problem 4.** Install Cassandra server on your VM. Use Cassandra SQL Client, cqlsh, to create and populate table person. Let every person by described by his or her first and last name, and city where he or she lives. Let every person possess up to three cell phones. Populate your table with three individuals using cqlsh client. Demonstrate that you can select the content of your table person including individuals’ cell phones. Write a simple client in a language of your choice that will populate 3 rows in Casandra’s table person, subsequently update one of those rows, for example change the city where a person lives, and finally retrieve that modify row from Cassandra and write its content to the console. Describe to us the process of installing and invoking Java, Scala or Python packages, if any, you needed for this problem.

Please, describe every step of your work and present all intermediate and final results in a Word document. **Please, copy and past text version of all essential command and snippets of results into the Word document with explanations of the purpose of those commands. We cannot retype text that is in JPG images**. Please, always submit a separate copy of the original, working scripts and/or class files you used. Sometimes we need to run your code and retyping is too costly. Please include in your MS Word document only relevant portions of the console output or output files. Sometime either console output or the result file is too long and including it into the MS Word document makes that document too hard to read. PLEASE DO NOT EMBED files into your MS Word document. For issues and comments visit the class Discussion Board. If you use some other language other than Python in your daily work with NLP, please be free to use that language and a framework of your choice to do this assignment.