

CSEE5590-0001/490-0003: Big Data Programming

Lesson Plan # 11

<https://forms.gle/xMAmr3zATrtMG5cX7>

Lesson Title: *Apache Spark Streaming*

Lesson Description: *Apache Spark Streaming*

Lesson Overview:

Spark Streaming is an extension of the core Spark API that enables scalable, high-throughput, fault-tolerant stream processing of live data streams. ... Spark Streaming provides a high-level abstraction called discretized stream or ***DStream***, which represents a continuous stream of data.

In class exercise:

1. Spark Streaming using Log File Generator:

Spark Streaming using log file generator. Use the instructions in the slides

2. Spark Streaming for TCP Socket:

Write a spark word count program of Spark Streaming received from a data server listening on a TCP socket.

Hint:

For Netcat utility in Windows

<https://github.com/rsanchez-wsu/jfiles/wiki/Windows-10-Telnet-&-NetCat>

3. Spark Streaming for Character Frequency using TCP Socket:

Output Example

Character Count	=>	List of Words (Distinct)
2	=>	[to, in, go]
3	=>	[the, law, and, can, how]
4	=>	[java, data, file]

Note: Ignore the case-sensitivity. Assume all the words are in lower case.

ICP Submission Guidelines:

1. ICP Submission is individual however, it can be completed as a Team during session.
2. If completed, should be presented to TA or Instructor before the completion of the class
3. Submission after the deadline is considered as late submission. (Check the late submission policy in the syllabus)
4. ICP Code with brief explanation should be pushed to GitHub.
5. Submit your screenshots as well to GitHub and documentation. The screenshot should have both the code and the output.
6. Submit a demo video 2-3 min showing your assignment with a voice over explaining your work if you are unable to complete ICP within the deadline due to genuine reason.
7. Provide the video submission link through the GitHub and submission form <https://forms.gle/xMAmr3zATrtMG5cX7>

Cheating, plagiarism, disruptive behavior and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy. See detailed description of university policy at the following URL:
<https://catalog.umkc.edu/special-notice/academic-honesty/>