# **ITCS 3190 Assignment 5**

Total 100 points

*Due: May 4, 2020. No late submission accepted*

**Please refer to the slides and relative videos to finish this assignment**

1. (25 pts) What is Spark Streaming?

Large scale stream processing framework

1. (25 pts) What are the requirements from Spark Streaming platform? (5 aspects)
2. Scalable to large clusters
3. Second-scale latencies
4. Simple programming model
5. Integrated with batch and interactive processing
6. Efficient fault tolerance in stateful computations

1. (25 pts) How does Spark Streaming implement Discretized Streaming Processing?

By running a streaming computation as a series of small deterministic batch jobs

1. (25 pts) State the fault-tolerance strategy adopted by Spark Streaming?

RDDs remember the sequence of operations to get from the original fault-tolerant input so it can be recreated

Input data is replicated in memory of worker notes making it fault-tolerant