Data Intensive Computing - Review Questions 4

Deadline: September 25, 2020

1. **1 point.** What is `DStream` data structure, and explain how a stateless operator, such as `map`, works on `DStream`?

   _____

2. **1 point.** Explain briefly how `mapWithState` works, and how it differs from `updataStateByKey`.

   _____

3. **1 point.** Explain how Spark streaming and MillWheel provide fault tolerance?

   _____

4. **2 points.** Throgh the following pictures, explain how Google Cloud Dataflow supports batch, mini-batch, and streaming processing.