

Data Intensive Computing - Review Questions 4

Deadline: September 29, 2019

1. What is `DStream` data structure, and explain how a stateless operator, such as `map`, works on `DStream`?
2. Explain briefly how `mapWithState` works, and how it differs from `updateStateByKey`.
3. Explain how Flink provides Fault Tolerance?
4. Through the following pictures, explain how Google Cloud Dataflow supports batch, mini-batch and streaming processing.

