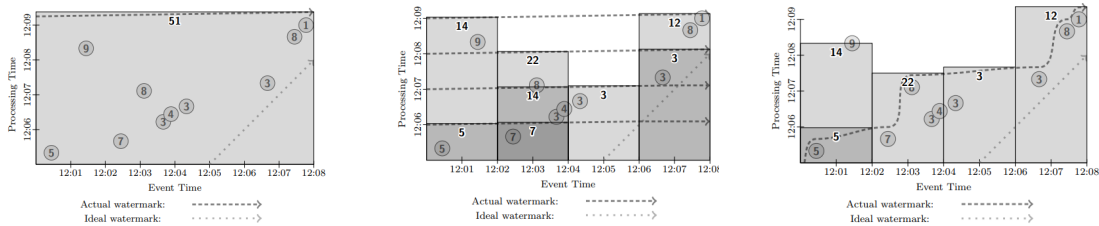


Data Intensive Computing - Review Questions 4

Deadline: Oct. 1, 2022

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. Through the following pictures, explain how Google Cloud Dataflow supports batch, mini-batch, and streaming processing.



4. Explain briefly how the command `pregel` works in GraphX?
5. Assume we have a graph and each vertex in the graph stores an integer value. Write three pseudo-codes, in Pregel, GraphLab, and PowerGraph to find the minimum value in the graph.