## PFLOCK Report

Andres Calderon

University of California, Riverside

January 10, 2020

# Task analysis...

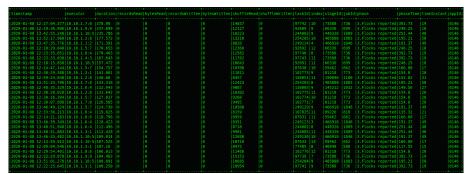
- ▶ Apache Spark divides the work into a number of Stages, each one is also divide into a number of Tasks.
- ► Each Task evaluates the data from a particular Partition and it is sent to an Executor for evaluation.
- ► Tracking the Tasks will give us a notion of the Partitions and Executors performance...

## Task analysis...

- SparkListener class allows us to monitor Taks metrics. We can capture info about:
  - executor: The id of the executor where the Task was evaluated.
  - duration: Execution time of the task.
  - recordsRead/Written: Total number of records read or written.
  - bytesRead/Written: Total number of bytes read or written.
  - ▶ shuffleRead/Written: Total number of records read or written from the shuffle by this task.

### Task analysis...

- ▶ Running experiments with Epsilon=30, Mu=3 and Delta=3.
- ▶ Collecting Task metrics. Focus on Time by time implementation...



### Stage analysis...

- ➤ SparkListener also provide info about the whole stage. Some interesting metrics are:
  - ▶ numTasks: The number of tasks in which the stage is divided.
  - name: The Spark function which invoke the stage and its line of code.
  - details: The stack trace when the stage was called.

