



Contribution ID: 453

Type: **Poster**

Featherweight Communication Between The Tasks For Spark

Tuesday 5 November 2019 16:15 (15 minutes)

ABSTRACT

Apache Spark is a splendid framework for big data analysis nowadays. A Spark application can be divided into some jobs which are triggered by an action of RDD, then the jobs will be divided into stages by the DAGScheduler, after these processes, we will get the task which is a unit of work within a stage, corresponding to one RDD partition.

Task is the smallest unit when Spark executes the application. However there is no communication between tasks in the current Spark framework. This article discusses the reasons why we need to extend Spark by compiling an API which can offer featherweight communication between tasks. This API won't break current communication mode and can be portable to standard Spark installations. At last we give some examples to explain how to address the specified situation in the high energy physics field.

Keywords

Spark, featherweight communication, task, high energy physics

Consider for promotion

No

Author: Mr WANG, YI

Presenter: Mr WANG, YI

Session Classification: Posters

Track Classification: Track 5 –Software Development