Contribution ID: **36** Type: **Poster**

Spark and Dask Performance Analysis Based on ARL Image Library

Significant scientific discoveries are currently being driven by the analysis of large volumes of image data, and the SKA-square kilometer array telescope is one of them in astronomy. At present, there are many computation and transmission frameworks supporting such tasks, but the specific performance of the frameworks for astronomical science data processing (SDP) remains to be verified. In this paper, we evaluate two popular frameworks, Spark and Dask, using a standard image processing pipeline of SKA SDP. The evaluation is carried out from multiple angles such as total cores, data size and the number of threads per process. And then we find that the task scheduling models can be further improved by genetic algorithm, which leads to a local optimal solution. More contributions of this paper consist of some basic ideas of the coordination between computation topology model, data transmission model of processors and physical machines, and also the routing model.

Suggested duration

Primary authors: Mr FU, Kaiyu (Shanghai Jiao Tong University); Dr LI, Qiuhong (Fudan University); Ms FAN, Siyu (Shanghai Jiao Tong University); Ms LI, Ting (Shanghai Jiao Tong University); Dr HUANG, Tian (University of Cambridge); Dr LUO, Yuan (Shanghai Jiao Tong University, China)

Presenter: Mr FU, Kaiyu (Shanghai Jiao Tong University)

Track Classification: SKA Regional Centres