Contribution ID: **702** Contribution code: **contribution ID 702** Type: **Poster**

Blaze: High performance Big Data Computing System for High Energy Physics

High energy physics (HEP) is moving towards extremely high statistical experiments and super-large-scale simulation of theory such as Standard Model. In order to handle the challenge of rapidly increase of data volumes, distributed computing and storage frameworks in Big Data area like Hadoop and Spark make computations easily to scale out. While in- memory RDD based programming model assumes workload perform local computation and occasionally exchange messages, it's inefficient at HEP use cases, because the scientific computations, such as partial wave analysis (PWA) and lattice quantum chromodynamics (LQCD), are based on numerical linear algebra and iterative algorithms that rely on message passing between tasks. In this paper, we present a computing system by modifying Spark to support OpenMPI, and it performs as a unified system that integrated MPI in DAG and task scheduling strategy. Therefore, the insufficient of expressiveness in Spark model are supplemented by inter-task message passing, it also empowers MPI with the ability of data-locality computing and provides a solution of fault tolerant.

Significance

This presentation proposes a computing system based on Spark and OpenMPI for HEP use case. The experiments shows that matrix multiplication and conjugate gradient (CG) algorithm are at least 50% faster than Spark applications. In addition, we also present a LQCD end-to-end workflow, including numerical simulation, data management and analysis stages.

References

Speaker time zone

Compatible with Asia

Primary authors: XIA, Libin (IHEP); SUN, Wei (IHEP); SUN, Gongxing (INSTITUE OF HIGH ENERGY PHYSICS); LIU, Xiaoyu

Presenter: XIA, Libin (IHEP)

Session Classification: Posters: Raspberry

Track Classification: Track 1: Computing Technology for Physics Research