Contribution ID: **13**   Type: **Short talk**

DASK enabled external tasks for in transit analytics with DEISA

Thursday 23 March 2023 14:00 (10 minutes)

In situ paradigm represents a relevant alternative to classical post hoc workflows as it allows bypassing disk accesses, thus reducing the IO bottleneck. However, as most in situ data analytics tools are built on MPI, and they are complicated to set up and use, especially to parallelize irregular algorithms. In a previous work, we provided a tool that couples MPI simulations with in situ task-based analytics written in Dask Distributed called Deisa[1]. In our old prototype, data and metadata were exchanged synchronously at each timestep (that overloads the scheduler), and a new task graph was submitted to process that step every time (time dependencies need to be managed manually to write algorithms).

In this work, we have addressed these limitations and improved our design by introducing asynchronicity and reducing the traffic to the scheduler. In addition, we avoid metadata fetch and allow submitting time-independent task graphs thanks to three main concept: "deisa virtual arrays" data structure, "contracts" to make selections of only needed data, and "external tasks" in Dask distributed to support getting data from external running programs (a running MPI simulation in our case).

We have implemented these improvements on top of the work presented in [1]. We have added a new Deisa plugin in the PDI Data interface and included our "external tasks" contribution into a forked version Dask Distributed repository. We have tested our approach using a heat equation mini-app with several analytics, such as temporal derivative and incremental PCA, and working on production use cases.

Deisa[1], an in situ analytics tool, [1] A. Gueroudji, J. Bigot and B. Raffin, "DEISA: Dask-Enabled In Situ Analytics," 2021 IEEE 28th International Conference on High Performance Computing, Data, and Analytics (HiPC), 2021, pp. 11-20, doi: 10.1109/HiPC53243.2021.00015.

JLESC topic

Primary authors: Mrs GUEROUDJI, Amal (French Alternative Energies and Atomic Energy Commission (CEA Sacaly)); Dr RAFFIN, Bruno (Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, 38000 Grenoble, France); Dr BIGOT, Julien (French Alternative Energies and Atomic Energy Commission (CEA Sacaly))

Presenter: Mrs GUEROUDJI, Amal (French Alternative Energies and Atomic Energy Commission (CEA Sacaly))

Session Classification: Short Talks on Tasking

Track Classification: I/O, storage and in-situ processing