Contribution ID: **19** Type: **Oral presentation**

Improving the I/O scalability for the next generation of Earth system models: OpenIFS-Cassandra integration as a case study

Thursday, 26 September 2019 14:30 (30 minutes)

Earth system models (ESMs) have increased the spatial resolution to achieve more accurate solutions, producing an enormous amount of data. However, some ESMs use inefficient sequential I/O schemes that do not scale well when many parallel resources are used. This is the case of the OpenIFS model.

OpenIFS is a free and simplified version of the Integrated Forecasting System (IFS), available under a license from the European Centre for Medium-Range Weather Forecasts (ECMWF). IFS is a global data assimilation and forecasting system which includes the modelling of the atmospheric composition.

The output in OpenIFS is done through an inefficient serial I/O scheme where the master process gathers all the data and sequentially writes into GRIB files.

When OpenIFS is used for climate modelling, there is a series of necessary post-processing operations to be performed, such as file format conversion to netCDF or regridding. In addition, if users want to query only a subset of the simulated data, they have to post-process entire files, and then filter the data. It is a slow, inefficient process which is difficult to parallelize.

To mitigate these issues, we modified OpenIFS to perform distributed I/O on all processes. Then, adopted Apache Cassandra, a Key-Value distributed database, to store all the information. Also, we tried Intel's modified Cassandra based on persistent memory whose higher throughput and lower latencies promise faster simulations and analysis. As a result, I/O intensive simulations are more efficient while data is managed transparently.

By using a Key-Value distributed database, scientists can query forecasts without post-processing entire files. Information is indexed at run-time, enabling users to access particular subsets without reading large amounts of data. Besides, our proposed system supports sampling to speedup anomaly and pattern detections, which simplifies data exploration greatly.

In this work, it became apparent that Cassandra and its connectors are not designed for HPC environments. They are based on thread-pools, which ultimately compete with MPI for CPU-time, reducing the potential gains of distributed I/O. In future research, we would like to explore a better integration with HPC software and persistent-memory to overcome these limitations.

Keywords

OpenIFS, Cassandra, persistent-memory

Primary authors: YEPES ARBÓS, Xavier (Barcelona Supercomputing Center); SANTAMARIA, Pol (Barcelona Supercomputing Center); SERRADELL MARONDA, Kim (Barcelona Supercomputing Center); Dr BECERRA, Yolanda (Barcelona Supercomputing Center); Prof. LABARTA, Jesus (Barcelona Supercomputing Center); Dr SERVAT, Harald (Intel Corporation)

Presenter: YEPES ARBÓS, Xavier (Barcelona Supercomputing Center)

Track Classification: NEXTGenIO Workshop on applications of NVRAM storage to exascale I/O