BDEC meeting in Frankfurt, May 2016

Pathways to Convergence

Osamu Tatebe
University of Tsukuba
tatebe@cs.tsukuba.ac.jp

Convergence

Research and development of big data analysis in Cloud have been done in relatively poor network environment and a large number of commodity compute nodes. Major focus of the challenge is how to reduce network data transfer and how to tolerate the node and network failures. This challenge is also required in the exascale high-performance computing. Network speed becomes poor and poor compared with CPU performance improvement in each compute node since many cores or GPU acceleration boosts the CPU performance. Fault tolerance becomes really an issue in exascale computing since the number of parts increases. Exascale high-performance computing requires almost the same challenge as the big data analysis but in different performance order in network and CPU.

Regarding the storage architecture, local disks on compute nodes are mostly used for big data analysis in Cloud. This trend is also true in exascale high-performance computing to bridge the performance gap between CPU and storage. Non-volatile memory and NVMe-connected flash device will be in each compute node. Between the compute nodes and a parallel storage, there will be burst buffers. This hierarchical storage architecture is one of critical goals of the exascale high-performance computing, and the technological and architectural convergence of big data analysis and high-performance computing is quite natural for further progress of the performance of big data analysis and high-performance computing.

Pathways

The architecture of big data analysis and high-performance computing will be converged. The pathways to the convergence will be natural. Regarding the runtime system and the programming environment, it will be divergent for application requirements. Currently, MPI in HPC and MapReduce/Spark in big data analysis are major. The newer type of

applications appears, the more runtime system and domain specific programming language appear. On the other hand, the storage system will be converged. It will utilize a hierarchical storage system from non-volatile memory and flash device in each compute node, burst buffers, to a parallel storage system, and possibly an archive (cold storage) system. It also needs to provide required functionalities by MPI-IO, MapReduce/Spark, and newly appeared runtime systems. The storage system may not be a single system but several storage layers to cover functionalities of each storage software like burst buffer software and parallel file system.