Merging Big Data and HPC for Large-scale Analysis/Analytics at the NASA Advanced Supercomputing (NAS) Division

Piyush Mehrotra, Jeff Becker
{piyush.mehrotra, jeffrey.c.becker}@nasa.gov
NASA Advanced Supercomputing (NAS) Division, NASA Ames, Moffett Field CA 94035

NASA is the curator and host for large quantities of observation and simulation data. By some estimates, NASA's approximately 100 current active missions, observing Earth, the solar system and deep space, produce more than 50 PBs of data annually. Similarly, large-scale simulations run on its supercomputers are starting to produce terabytes and sometimes petabytes of data from each run. As NASA's flagship site for computational science and engineering at scale, the NASA Advanced Supercomputing (NAS) Division, supports a user base that is at the forefront of data intensive and data driven science. Our users' codes utilize and generate very large datasets and analyzing these datasets to extract information and knowledge is a fundamental part of their workflows.

Recently we at NAS reached out to many of our users to better understand the challenges they face in their analysis efforts. Guided by the results of the survey, as well as many studies that have addressed the convergence of big data and big compute, we have designed an environment called, Data Intensive Supercomputing Environment (DISE), to meet all the challenges faced by scientists performing data analytics and analysis of NASA's large datasets. These challenges include discovery of data and tools, access to and movement of data, storage and management of data, algorithms/tools for performing the analysis/analytics workflow execution and management and finally dissemination of the results. The main focus is to co-locate the big data with big compute so as to reduce the transfer time and also support analytics across multiple large data sources.

NAS offers an integrated environment that supports analysis and analytics at scale from prototyping through production. Much of our experience in this arena has been gained through the NASA Earth Exchange (NEX) a collaborative platform developed to engage and enable the Earth science community. Using NEX as a model, we propose the architecture shown in Figure 1 for supporting Big Data Science and HPC at NAS. As depicted in the figure, DISE aims to provide a data-centric environment with the data resources being globally available across all platforms through a high-bandwidth low-latency network. Users will access the environment and its resources through a domain-specific web portal that can be used to browse and search relevant datasets, and copy files of interest to reside locally (if they're not already mirrored at the NAS). They will also be able to collaborate with other scientists and share data and information.

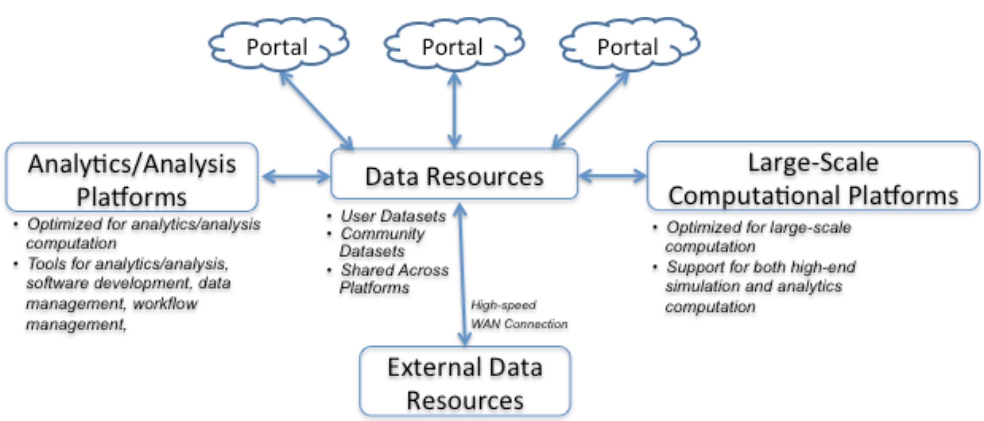

Figure 1: Notional Architecture for Data Intensive Supercomputing Environment at NAS

Given that a data analysis pipeline is generally complex and requires different types of processing, we envision that the infrastructure will support a heterogeneous set of computer resources so that the user can effectively utilize the appropriate hardware for the various subtasks including preprocessing and large-scale data analytics and analysis. This would include specialized hardware for analytics such as GPGPU nodes for machine learning and visualization, Hadoop clusters for map-reduce style computation along with more traditional large-scale shared-memory and distributed memory systems all accessing data via a high-performance global file system such as Lustre. The environments will provide access to computational tools for analysis and analytics. These include statistical packages such as R, and machine learning tool kits such as Google's tensorflow along with software for data management and workflow management and capture.

Current Status: The NAS facility currently provides support for traditional large-scale computation including Pleiades a 6.4 PF distributed-memory cluster with more than 11K nodes with a few NVIDIA GPGPU and the Intel Xeon Phi based nodes, and a large shared-memory system, Endeavour, consisting of two SGI UV nodes with 1024 cores- 4 TB and 512 cores – 2 TB of memory respectively. A 128-node GPU based cluster, hyperwall, supports visualization of large-scale data sets. A small Hadoop cluster has been set up to experiment with data mining techniques.

We have initiated several projects that are focusing on the different aspects of Big Data analysis pipeline. Given that such applications require fast I/O. we have been studying the various methods of using of NVMe SSDs in our high end compute platforms. Each node of our visualization cluster has an Intel NVMe 2TB SSD and we are currently testing system level caching using Excellero kernel software, which provides RDMA access to the SSD's over InfiniBand We have also equipped the object storage servers of one of our Lustre file systems with two Intel NVMe 1.6TB SSDs and are testing Intel Cache Acceleration software to use the SSD's for block device caching.

On the software side, we are investigating a framework for semi-automatically tagging of data both for security and data discovery. The focus is to use the resultant meta-data to properly protect the data from a security perspective while remaining easily discovered by those who require the data. For discovery, we have been working jointly with NASA Langley on the Ontology Driven Interactive Search Environment for Earth Sciences (ODISEES) project, a semantic-based framework for high-level discovery and access of Earth science data. The NEX team is developing a VisTrails-based workflow management system to specify and manage Earth science data-processing pipelines in an HPC environment. The effort also includes semi-automatic generation of provenance data and workflow capture to not only re-validate the processes but also to use them across different architectures and environment. The team is also testing machine-learning algorithms for object classification, segmentation, and feature extraction from very high-resolution imagery datasets. These techniques are being applied to the quarter million image scenes from the 1-m multispectral National Agriculture Imagery Program (NAIP) dataset to estimate tree cover for the continental United States given the large complexities and heterogeneity in land cover types. Our security group has initiated a project to utilize data mining techniques to analyze network flows to characterize and identify malicious activities.

Traditionally, HPC centers have focused on code performance providing infrastructure for executing large-scale simulations efficiently. Increasingly, these simulations are producing enormous amounts of data, which needs to be managed and analyzed. At the NAS facility, we have been focusing on extending the HPC infrastructure to handle big data analytics not only by enhancing the compute and I/O hardware but also providing services targeted to data discovery, access, and management along with analysis and analytics of such datasets. Such an environment will allow NASA scientists to do Big Science utilizing large-scale computing and data in a single environment.