Update on recent facilities upgrades in the Stern Center for Research Computing

There were a number of recent facility and service upgrades that impact research computing users. A list of all of the services and facilities can be found at http://scrc.stern.nyu.edu

1) **Increased/upgraded Storage** - We upgraded our storage network with the addition of another 96 Terabytes of storage to handle the rapidly increasing size and number of research data sets. We now have over 200 TB of disk storage. Our primary storage device was upgraded with solid state disks, to dramatically increase it's performance.

2) **Network upgrades** - In order to support the rapidly increasing size of research data sets, we have now moved all of our major servers and most of our storage systems onto a 10 gb network backbone, increasing network capacity by a factor of 10. We hope to complete upgrading our remaining storage in the next few months. All of our vmware cloud servers and most of our storage devices are now running at 10 gigabits / second.

3) **NYU Big Data Cluster** -The university HPC group has just deployed a very large hadoop (big data) cluster running the latest version of Cloudera Enterprise, which includes hadoop, hive, pig, spark, impala, oozie, tez, hue, ... The cluster (dumbo) has 44 128gb nodes, 704 cores, and over 1 petabyte of disk storage. Early benchmarks indicate it is very fast. (Sorts 1 TB of data in about 5 minutes). We are working with the university to provide a high speed link (10gb) between the Stern Research cluster and the NYU/ITS cluster so that researchers can easily move data back and forth between the two facilities. Dumbo also shares very high speed storage with the large HPC cluster (mercer). This will allow researchers to use both clusters to analyze the same data, for instance, running large simulations on the mercer HPC cluster and then using dumbo, the big data cluster to store and analyze/visualize the results.

4) **NYU data center network upgrade** -Just announced yesterday, the NYU South Data Center, where our equipment is located, is upgrading it's network backbone to 10 gigabits/sec from 1 gb/sec. We are already in talks about how to use this increased bandwidth to be able to move data between Stern Research Computing storage devices and the NYU/ITS HPC and big data clusters.

Coming soon...

We are actively working with the university HPC team to be able to connect SAS, Tableau, and other systems that support a relational data base interface to the new big data cluster. Hive (an sql like system) will be used to allow remote applications to treat the big data cluster as a large, extremely fast relational data base store. We expect to be able to connect to the new big data cluster both from within the Stern cluster, as well as from the NYU HPC cluster (which now can run SAS jobs). This should bring exciting new capabilities to both Stern and the university.

Posted 14th October 2015 by Norman White

Labels: 10gb big data cloud computing cloudera hive HPCresearch computing sas tableau vmware