

BNL Service Challenge 3

7 – June – 2005

Experience from Service Challenge 2:

We used four file transfer servers with 1 Gigabit WAN network connection to CERN. The performance/throughput challenges (70~80MB/second disk to disk) was met. We enabled data transfer between dCache/SRM and CERN SRM at Openlab. We designed our own script to control SRM data transfer; we also enabled data transfer between BNL GridFTP servers and CERN Openlab GridFTP servers controlled by Radiant software.

Service Challenge 2 Lessons learned:

Many components need to be tuned. The performance was not stable 7 x 24. The long 250 ms network RRT and high packet dropping rate prevented us from simply utilizing the 1Gbps BNL WAN connection. We had to use multiple TCP streams and multiple file transfers to fill up network pipe. We found that our data transfer servers had sluggish parallel file I/O with EXT2/EXT3 and lot of processes with "D" state. When we increased the number of file streams, the performance on file system became worse. We switched to XFS file system and there was an improvement. Still need to tune file system parameter and hard drivers to optimize the overall data transfer performance.

BNL Service Challenge 3 Throughput Phase - Goals

1. Efficiently use our available resource, such as Network, disk, and tape service
 - Sufficient network bandwidth: 2Gbit/sec
 - BNL large scale of dCache/SRM
 - Quality of service: performance: 150Mbyte/sec to disk, up to 60 Mbytes/second to tape, done with efficiency and effectiveness.
2. Functionality/Services, high reliability, data integration, high performance
 - Robust file transfer service
 - Storage Servers
 - File Transfer Software (FTS)
 - Data Management software (SRM, dCache)
 - Archiving service: tape servers, tape robots, tapes, tape drives,
3. Provide Grid middleware support for ATLAS Experiment Applications and facilitate ATLAS software interfacing with the Grid services challenged in SC3.
4. Enable Tier 2 to participation.

BNL Service Challenge 3 Throughput Phase - Evaluation Criteria:

1. Were goals defined above met?
2. Was data transfer sustained 7 x 24?
3. Many Grid service are monitored so many events will be reflected in the logs for problem diagnoses: what, when, why and how the fixes improve.
4. Number of human intervention required? Number of FTE needed to support SC3?

BNL Service Challenge 3 Throughput Phase - Status and Plan:

- **1: Network Status and Plan:** BNL campus LAN was upgraded into 10 Gbps on May/24/2005. A new USATLAS Accessing switch is connected to BNL campus

network via 2 x 1 gigaE Channel on May/25/2005. BNL campus LAN was connected to PoP at ES-net with 2 x 1 gigaE Channel. Our network connectivity (2.0Gbps full duplex) satisfies the 150 MB/second target data rate from CERN to BNL and the remaining BNL network activities). There won't be an upgrade of BNL WAN connectivity to 10Gbps in 2005.

- **2: BNL FTS Installation Plan:**

- May 25~June 3: Install FTS (Xin Zhao) and its backend Oracle database (Grace Tsai). Completed by Xin and Grace.
- June 13: Install FTS client software on BNL AFS area for local users and Tier 2 users (Xin).

- **3: USATLAS Tier 1 SRM/Storage Element Status and Plan:**

Start from existing BNL USATLAS production dCache system for service challenge. Existing dCache/SRM provides the SRM 1.1 interface, required by SC3 and total disk space of 50 TB. It has one core server hosting SRM door, GridFtp door and dCap door, two dCache external write pool nodes with fast Ethernet connectivity.

Before SC3 starts the following will occur:

- June 11~June 15: Add eight dedicated write pool nodes, each with 1GB memory, dual gigabit Ethernet and 3 x 160GB SCSI drivers. Four write pools will be XFS file system based to compare performance with standard file system. (**Zhenping: pool deployment, Jason: OS system and XFS file system configurations**)
- May 25~June 11: Prepare the core dCache servers: host certificate, firewall conduit requests. (**Zhenping/Jason, Jason** is responsible for keep track of certificate)
- June 8: Start three days data transfer test to use FTS data transfer with the existing BNL dCache. Push data from CERN to BNL with CERN FTS and pull data from CERN to BNL with BNL FTS instance. (**Xin**)
- June 20: Provide ATLAS "agent" box for SC3 service phase. (Xin)
- June 13~June 17: Add 230 nodes with 120TB of disk into dCache system (most not of use for SC3), add one dedicate SRM door, 2~4 GridFtp/dCap door and related PNFS database upgrade. (**Zhenping/Xin**)
- June 17~June 23: BNL USATLAS production dCache stability test and tuning. (**Zhenping, Xin**)
- June 23~June 29: Repeat FTS data transfer with the upgraded dCache system. (**Xin**)

- **4: USATLAS Tier 1 Tape Status and Plan:** dCache will be used as the front end of BNL HPSS system. Data written in dCache will be staged into tape system. USATLAS owns only two tapes drives so borrowing from RHIC will be negotiated.

- June 17~June 23: Test data transfer from CERN disk to BNL tape systems with FTS. (**Xin**, with help from Zhenping if needed)

- July 15~July 29: Borrow four-six tape drives from RHIC for 24 hours to accomplish 60MB/second data transfer from CERN to BNL tape system. (to be discussed)
- **5: USATLAS Tier 2 Status and Plan:** Both University of Chicago and Boston University confirmed their participation into SC3. Use the existing ganglia system to monitor data traffic between Tier 2 and Tier 1. During the Tier 2 exercise we should try to exercise both the download (in principle of data for analysis) to the Tier 2's of a fraction, 50% to each of the data being received from CERN and the upload (in principle of simulate data generated at the Tier 2) from Tier 2's of a small data flow, a few files per hours:
 - July: University of Chicago will have dCache ready. Ideally UC needs to have dCache ready before July 18. Rob Gardner will confirm the exact date. (**Xin** will be the main contact for understanding tier 2's needs)
 - July: Boston University will not have dCache ready until the late summer before SC3 service phase. It is probably too late for SC3 throughput phase. Saul confirmed his participation into SC3 and he would provide vanilla GridFtp server(s) for the throughput phase. FTS developer confirmed that FTS supports GridFtp server. (**Xin** will be the main contact point here too).
 - July 20: Tier 2 test of FTS client software. (**Xin** will provide help)
- **6: LFC Status and Plan:** ATLAS Experiments would test LFC with POOL catalog and use it during service phase. "ATLAS service will use the LFC as a file catalog at several sites (Gilbert Poulard)". We would deploy LFC right after SC3 planning meeting (June 13~June 15). A dedicated file catalog host was allocated. USATLAS tier2 site will share LFC server at BNL.
 - June 25, LFC client and server will be up and running at BNL. (**Wensheng**)
 - July 1, LFC clients are going to be deployed in AFS directory and Tier 2 sites (UC and Boston University). (**Wensheng**).