

21st International Conference on Computing in High Energy and Nuclear Physics (CHEP2015)



Contribution ID: 259

Type: **oral presentation**

Deep Storage for Big Scientific Data

Tuesday, April 14, 2015 5:30 PM (15 minutes)

Brookhaven National Lab (BNL)'s RHIC and Atlas Computing Facility (RACF), is supporting science experiments such as RHIC as its Tier-0 center and the U.S. ATLAS/LHC as a Tier-1 center. Scientific data is still growing exponentially after each upgrade. The RACF currently manages over 50 petabytes of data on robotic tape libraries, and we expect a 50% increase in data next year. Not only do we have to address the issue of efficiently archiving high bandwidth data to our tapes, but we also have to face the problem of randomly restoring files from tapes. In addition, we have to manage tape resource usage and technology migration, which is moving data from low-capacity media to newer, high-capacity tape media, in order to free space within a tape library. BNL's mass storage system is managed by a software called IBM HPSS. To restore files from HPSS, we have developed a file retrieval scheduling software, called TSX. TSX provides dynamic HPSS resource management, schedules jobs efficiently, and enhances visibility of real-time staging activities and advanced error handling to maximize the tape staging performance.

Primary author: YU, David (BNL)

Co-authors: Mrs CHE, Guangwei (BNL); Mr NOVAKOV, Ognian (BNL); Mr CHOU, Tim (BNL)

Presenter: YU, David (BNL)

Session Classification: Track 3 Session

Track Classification: Track3: Data store and access