

Management of Cost Effective Mass Storage Environments

Tuesday, 20 March 2018 11:00 (30 minutes)

The Scientific Data and Computing Center (SDCC) at Brookhaven National Laboratory (BNL) is a major scientific data storage site with more than 100 PB of archived experimental data, which amount is supposed to grow exponentially during the next several years. BNL is a Tier 1 data center for the ATLAS collaboration.

The challenge is to meet the data storage and processing requirements of the Relativistic Heavy Ion Collider (RHIC) experiments at BNL and especially the newly emerging enormous storage needs of ATLAS with limited budget.

In this past year SDCC has developed several tools to lower operating costs. One of the cost saving implementations is the introduction of JBOD in addition to existing RAID systems and LTO tape systems. The SDCC-developed software gathers statistical information to provide adaptive methods and techniques for device management and rigorous performance tuning and optimization of the entire storage environment. The software suite provides an intuitive user-friendly web-based interface enabling dynamic parameter visualization based on a priority-driven device-monitoring paradigm and addresses the following major technical issues:

- proprietary JBOD monitoring and management software enhancing the scalability of JBOD systems. It can monitor, configure and optimize the disk systems at 50% of the cost of commercial RAID systems.
- developed in-house data retrieval management and optimization application. ERADAT is SDCC developed software that manages and optimizes data retrievals from mass storage media, such as tape cartridges. This application will optimize data retrievals based on the data locations on the media, and manage the device allocations for multiuser concurrent operations.
- predict device failures before they actually happen and service the devices. This self-developed software periodically queries all mass storage devices for usage history and predicts failure before it actually happens. It alerts administrators when a device usage is nearing marginal limits and when a media error or device error is detected.
- detect and protect storage devices and media from unauthorized access. All media access and data retrievals are monitored and logged. The self-developed application will alert system administrators of any unusual data access/deletion patterns.
- encryption and duplication of data as requested by users.

Summary

The Scientific Data and Computing Center (SDCC) at Brookhaven National Laboratory (BNL) strives to meet the data storage and processing requirements while lowering the costs at the mean time.

Primary author: Mr CHOU, Tim (Brookhaven National Laboratory)

Presenter: Mr CHOU, Tim (Brookhaven National Laboratory)

Session Classification: Data Management & Big Data Session

Track Classification: Big Data & Data Management