Contribution ID: **9**   Type: **Poster presentation**

The Repack Challenge

Monday 14 October 2013 15:00 (45 minutes)

Physics data stored in CERN tapes is quickly reaching the 100 PB milestone. Tape is an ever-changing technology that is still following Moore's law in terms of capacity. This means we can store every year more and more data in the same amount of tapes. However this doesn't come for free: the first obvious cost is the new higher capacity media. The second less known cost is related to moving the data from the old tapes to the new ones. This activity is what we call repack. Repack is vital for any large tape user: without it, one would have to buy more tape libraries and more floor space and, eventually, data on old non supported tapes would become unreadable and be lost forever. The challenge is not an easy one. First, to make sure we won't need any more tape slots in the near future, we will have to repack 120 PB from 2014 to 2015, this in turn means that we will have to be able to cope with peaks of 3.5 GB/s smoothly. Secondly, all the repack activities will have to run concurrently and in harmony with the existing experiment tape activities. Making sure that this works out seamlessly implies careful planning of the resources and the various policies for sharing them fairly and conveniently. Our previous setup allowed for an average repack performance of only 360 MB/s. Our needs demand this figure increase tenfold by 2013. To tackle this problem we needed to fully exploit the speed and throughput of our modern tape drives. This involved careful dimensioning and configuration of the disk arrays (the middle step between an old source tape and a new higher capacity destination tape) and all the links between them and the tape servers (the machines responsible for managing the tape drives). We also planned a precise schedule and provided a visual monitoring tool to check the progress over time. The new repack setup we deployed brought an average 80% increase in the throughput of tape drives, allowing them to perform closer to their design specifications. This improvement in turn meant a 40% decrease in the number of drives needed to achieve the 3.5 GB/s goal. CERN is facing its largest data migration challenge yet. By restructuring the repack infrastructure we allowed the vital repack and LHC experiments activities to coexist without the need for new expensive tape drives.

Primary author: KRUSE, Daniele Francesco (CERN)

Presenter: KRUSE, Daniele Francesco (CERN)

Session Classification: Poster presentations

Track Classification: Data Stores, Data Bases, and Storage Systems