Data Storage Accounting and Verification at LHC Experiments

E. Lanciotti[1], N. Magini[1], N. Ratnikova[2], A. Sanchez-Hernandez[3], C. Serfon[4], T. Wildish[5], X. Zhang[6]

Common Objectives

All major experiments at Large Hadron Collider (LHC) need to measure real storage usage at Grid sites. This information is equally important for efficient resource management, planning, and operations.

Common Technologies

To verify consistency between central data catalogs and local storage experiments are asking the grid sites to provide a full list of files on the storage element, including file size, checksum, and other attributes. Storage dumps (SD) produced at regular time intervals give a realistic view of the storage use. Tools to produce SD are:

- driven by specific storage technology:
  Castor, dCache, DPM, EOS, hadoop/bestman, Lustre, LStore, StoRM/GPFS, xrootd, ...
- agnostic to experiment's data handling model
- successfully re-used by the experiments.

Scheduling Considerations

Producing Storage dumps and running consistency checks may impose additional load on the system. The optimal schedule depends on the storage technology used, local site infrastructure, and other activities sharing the same resources. Analysis of the use cases and the operational experience to date show that weekly produced reports based on storage dumps satisfy most accounting needs.

Typical Use Cases

- Data placement:
  - how much data are stored at the sites by data type, dataset, data owner, etc
  - How much space at the site is occupied by custodial data alone
  - how much space would be saved by the removal of a data replica from a site
- Data production monitoring:
  - how many files have been produced, grouped by production type (MC, reco), data type
- Check availability of free pledged storage resources
- Disk usage:
  - how much disk space is used by pre-staged/pinned/produced data
- Monitor storage usage by users and groups.
- Detect files not known to the central data catalogs, legacy data
- Verify consistency of the data catalogs

Conclusions

Regular monitoring of the space usage and data verification help to improve the integrity and performance of the data handling system.

Both the importance and the complexity of these tasks increase with the constant growth of the total data volumes during the active data taking period at the LHC. Experiments are constantly working on improvement of storage monitoring and accounting tools. Adopting common solutions helps to reduce development efforts and maintenance costs at the sites.
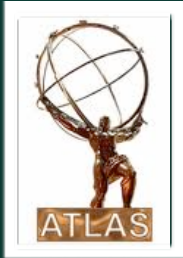
Central ATLAS data catalog keeps accounting of the used space according to DDM (ATLAS data management system) and SRM, and total allocated space from SRM, for each space token.
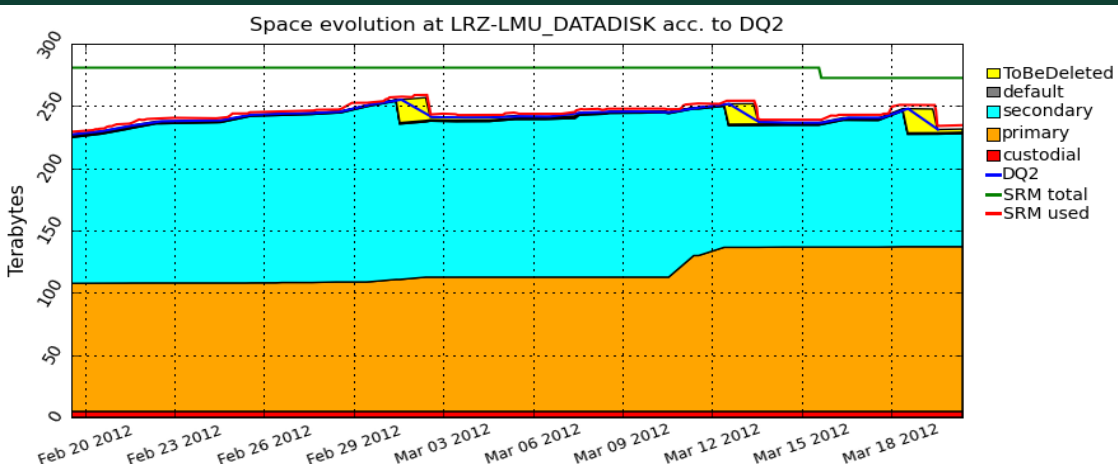
Plots show historical space usage and also spot possible discrepancy between SRM and file catalogue. If used space gets close to the total available, automatic site cleaning systems identify secondary replicas to be removed.
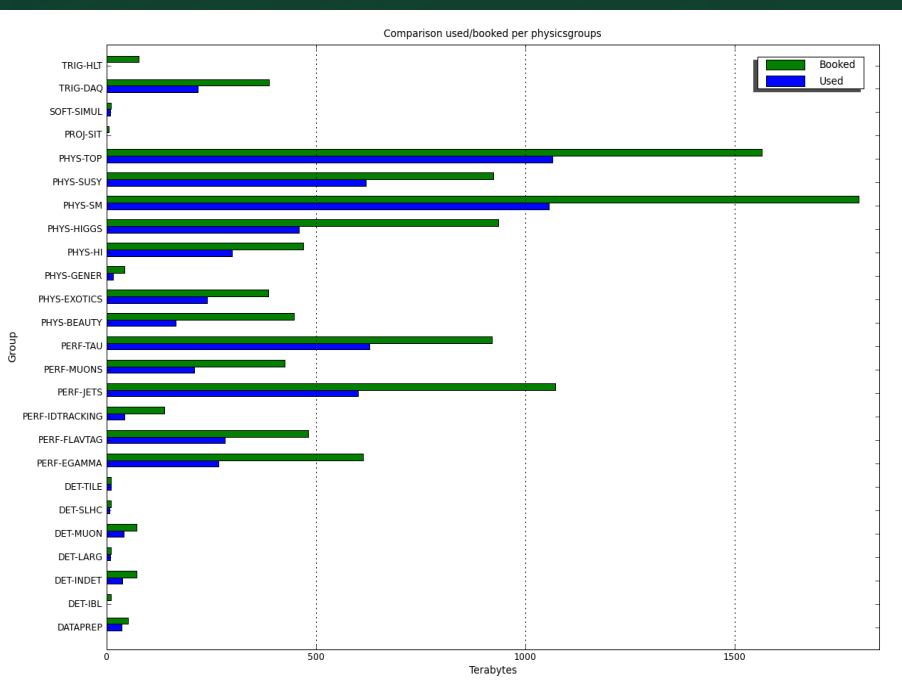

For monitoring physics groups space usage: plots of allocated space and used space. Also break-down by site.

Development ongoing to completely change the backend, allowing more powerful and performing data accounting by project, data type and other relevant meta-data.

CMS stores data information in two databases: DBS (Data Bookkeeping System) keeps records of produced data and production conditions. TMDB (Transfer Management Database) keeps information about the replicas stored at the sites.
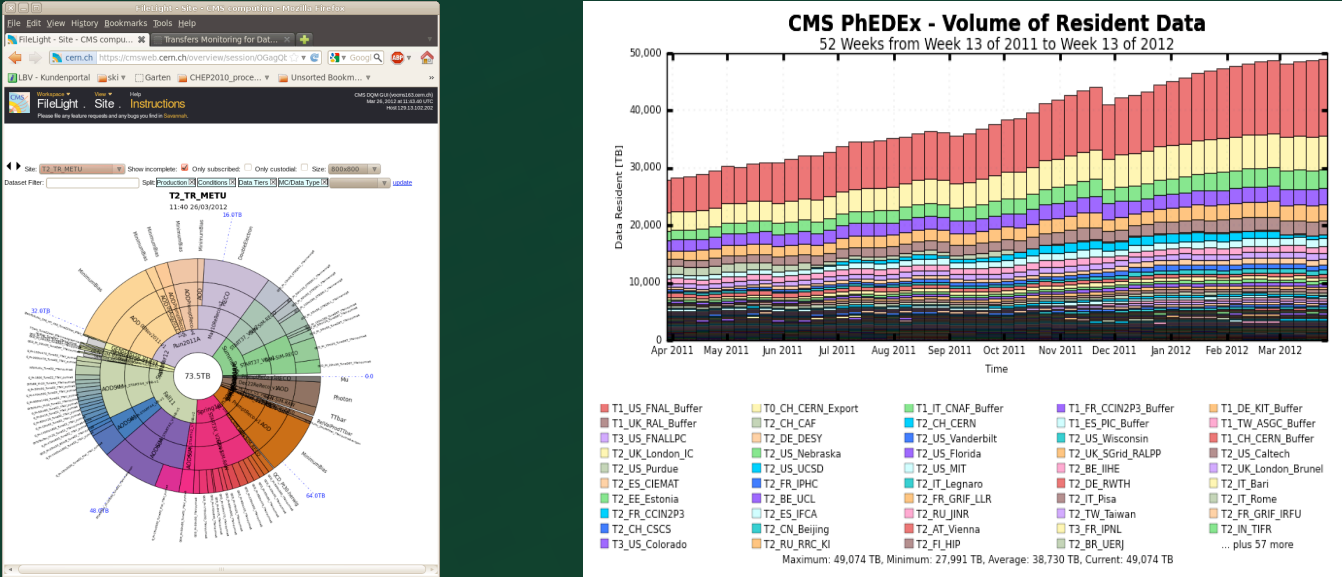

Overview visualization tool provides stack charts and pie diagrams of the used space groupped by data type, groups or by sites.

New Space Monitoring tool based on storage dumps provides:

- accounting for all data on the SE
- historical view of storage usage
- auto-detection of orphaned files
- possibility to interactively explore aggregated storage information

Information is aggregated at the sites and stored persistently in a dedicated central database.

New system can provide historical view and a possibility to explore the contents interactively, visualized via Treemap. This tool is not CMS specific in any way.
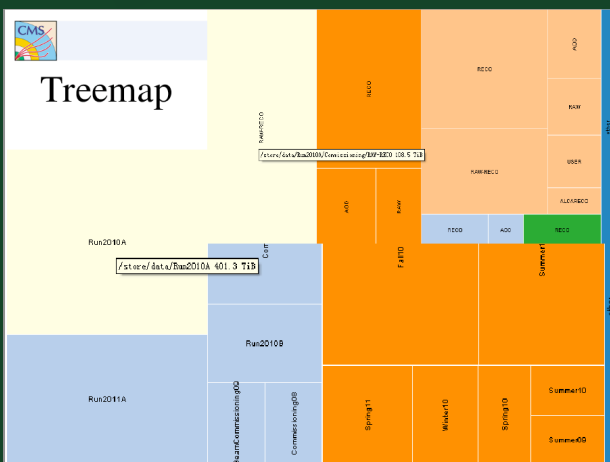
Ongoing Storage Dumps deployment campaign at Tier-1 and Tier-2 sites


LHCb data are managed centrally through the DMS (Data Management System) of the LHCbDirac framework. Existing accounting for storage use provided by SLS is based on information reported by SRM. New accounting is based on two catalogs:

- LFC replica catalog
- LHCb Bookkeeping meta-data catalog for production files
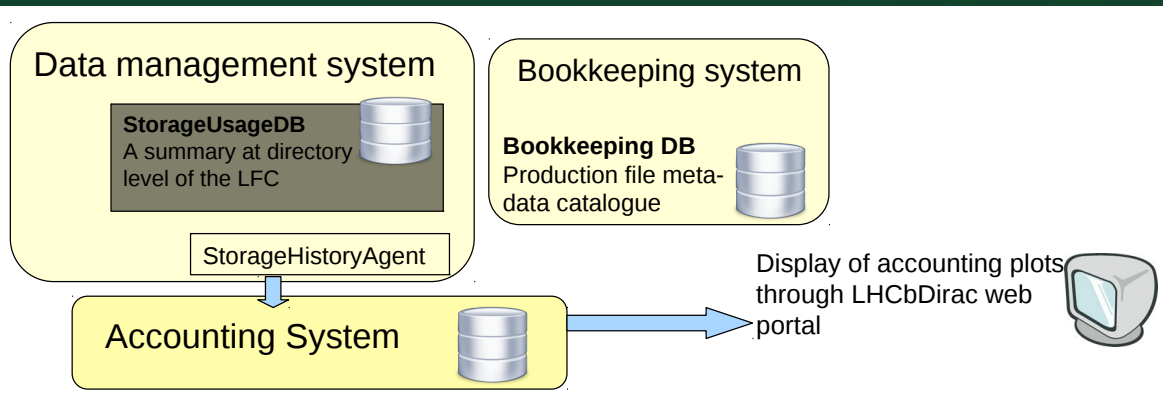
Storage accounting implementation in LHCbDirac framework.

Example of monitoring the progress of data reprocessing. The plot shows how many files have been produced for a given reprocessing, grouping by file type.
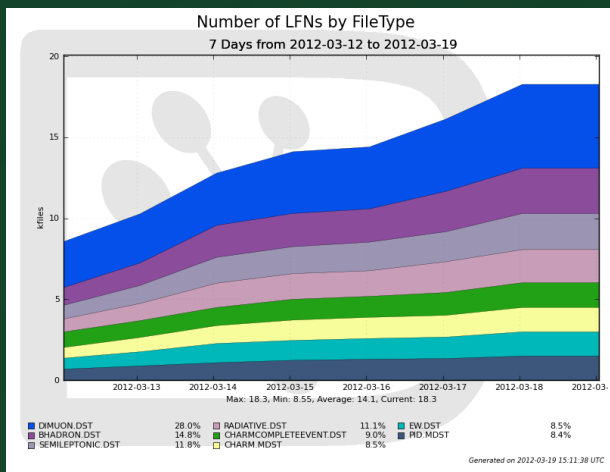
Example of monitoring the progress of the datasets replication. The plot shows how much space is used at the sites where the processed output data are replicated.
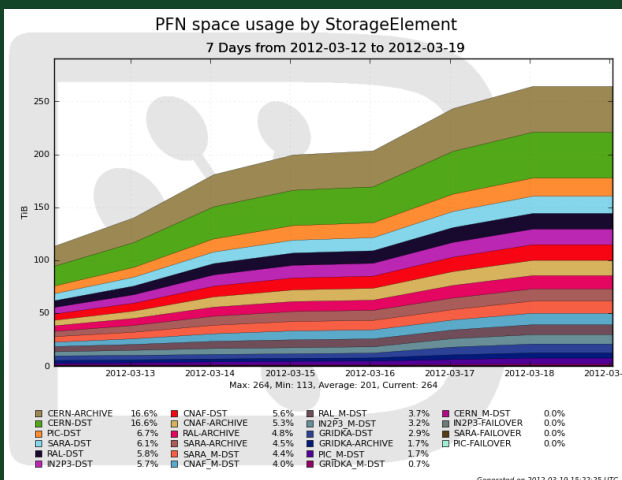
Caveat: the consistency between file catalogues and storage element has to be ensured.

---

[1] CERN (CH), [2] Karlsruhe Institute of Technology (DE) and ITEP (RU), [3] Cinestav (MX), [4] Ludwig-Maximilians-Univ. München (DE), [5] Princeton University (US), [6] IHEP, Beijing (CN)