Fermilab

Managed by Fermi Research Alliance, LLC for the U.S. Department of Energy Office of Science

Space Monitoring as joint project

Natalia Ratnikova

WLCG experiment support meeting

CERN, March 3rd, 2016

Outline

- CMS data management overview
- Motivation and requirements for space monitoring
- System overview
- Deployment and usage status
- Areas of potential common interest
- Additional desired features
- Discussion

CMS storage resources by the beginning of LHC Run 2

- CMS Tier 1 and 2 storage space requirements* :

Year	2013	2014	2015	2016
Tier 1 Disk	26,000	26,000	26,000	33,000
Tier 1 Tape	50,000	55,000	74,000	100,000
Tier 2 Disk	26,000	27,000	29,000	38,000

- Increased pileup, higher HLT rate, data parking and scouting
- Volume will grow proportionally to LHC life time
- Phase 2 detector upgrade studies

➡ CMS expects severe resource constraints

* *Values are given in Tbytes according to WLCG-rebus pledges summary*

Evolution of the computing model

- Changed patterns in organized data processing
- Tier 1 disk and tape separation
- AAA xrootd driven data federations
- Dynamic data management
- New data types:
  - MiniAOD
  - phase 2 detector studies
  - parked data
- Diverse user analysis patterns
- Increased share of storage space for users and groups

Multiple data placement processes not necessarily aware of each other sharing the same storage resources

Space monitoring for distributed storage

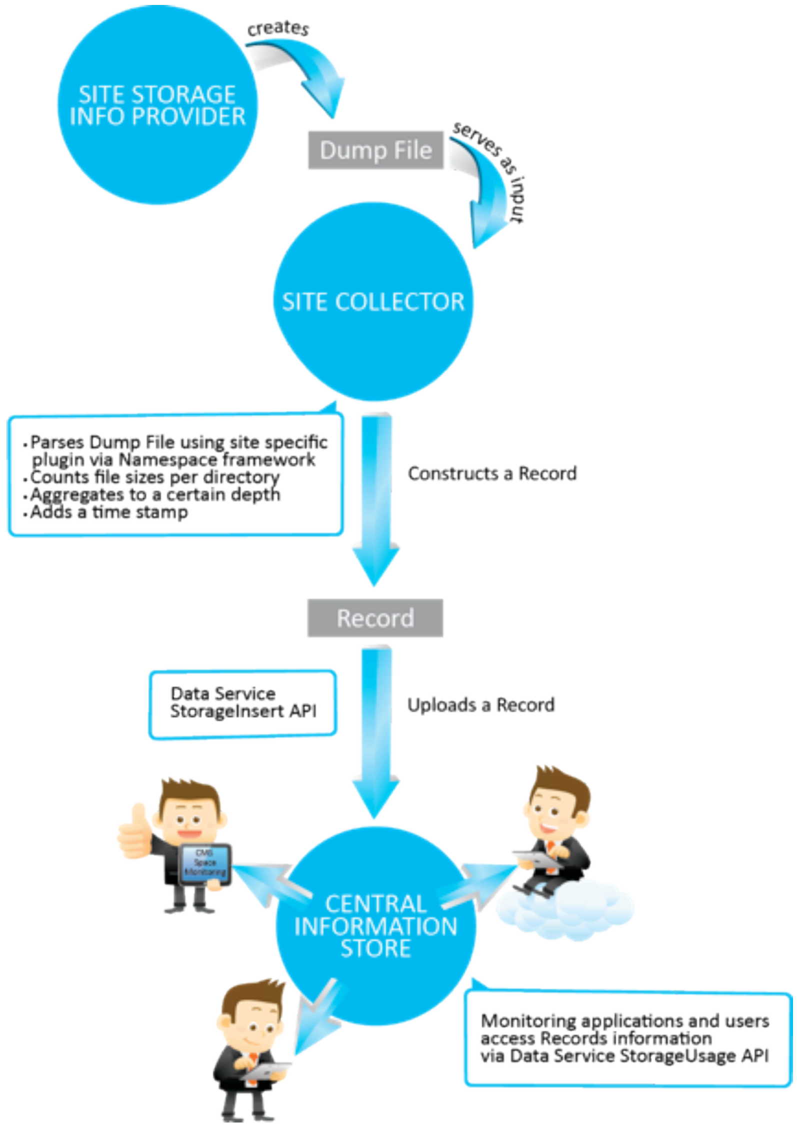
- CMS data live in a **global name space**, addressed by a logical file name (LFN), e.g.:

  /store/data, /store/mc, /store/user, /store/group, …
- Data are accessed by physical file names (PFNs) according to the LFN to PFN translation rules specified in the trivial file catalogs provided by the sites
- Space monitoring allows to track the space occupied by each level under /store across the sites.
- CMS central Transfer Management Database keeps track of data maintained by PhEDEx.
- Information on other files, users data, temporary production and test data, is only available from the direct storage dumps.

Related initiatives

- **CMS monitoring taskforce 2008** identified gaps in storage accounting
- **Storage dumps** to validate new Cache instance at GridKa
- **Syncat format** presented by Paul Millar at CHEP'09 – demonstrated abstraction from the storage technology
- Proposed the idea as a "**demonstrator**" at WLCG storage jamboree in Amsterdam 2009, supported by Tony & Daniele
- Elisa Lanciotti (WLCG ES group) work on identifying common formats and **tools for storage dumps** at Tier-1 sites
- **CMS monitoring workshop (2011)** data transfers team initial set of requirements for storage space monitoring
- **CMS Tier 2 data consistency campaign** launched in 2012
- Joint CMS, Atlas and LHCb presentation at **CHEP'12**

General architecture and workflow



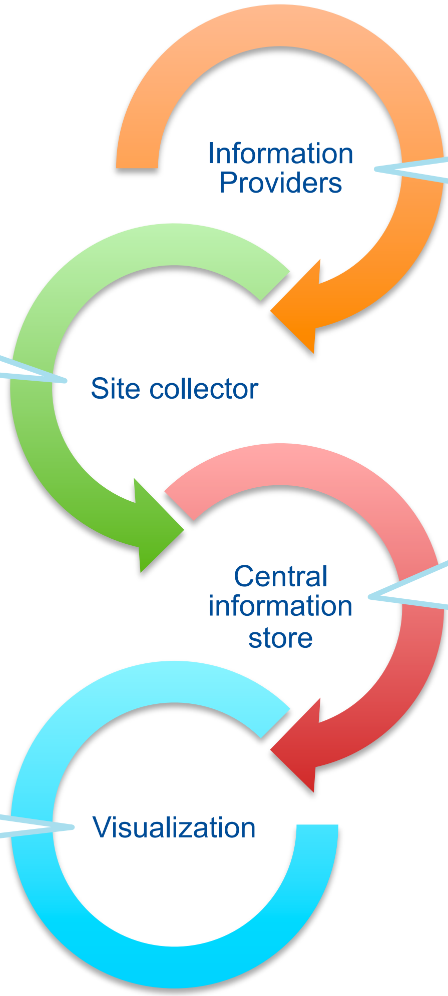
Components

- Site information providers (storage dump tools)
  – Storage technologies: Castor, dCache, DPM, EOS, Hadoop, LStore, Lustre, StoRM.
- Site collector (client tool)
  – Aggregates information and uploads to DMWMMON oracle database at CERN
- Central information store: web based data service
  – provides interfaces to upload and retrieve space usage records
  – APIs for authentication, list of sites, and troubleshooting access

System overview

Client side tool
Parses and aggregates storage dump and uploads to central information store

Tools for producing dumps of the site local storage catalogs



DMWMMON database and interfaces
Data service APIs to insert or retrieve space usage records

Interface to retrieve DMWMMON info and presentation layer in CMS dashboard

Deployment campaign

Issues encountered during this first stage of this deployment can be categorized into three groups:

1. Questions from sites about why they need to provide storage usage information and at what level of detail
2. Authentication problems uploading the information to the central data service
3. The long time it takes to take a dump for some storage systems.

Also some privacy and security concerns were raised by the sites.

Original requirements

- Collect storage space usage information at the sites
  - including data not managed in the central file catalog, such as user data, temp areas, legacy data, orphaned files etc
- Aggregate space usage information
  - must reflect the CMS data organization
  - scalability and privacy considerations
- Update information at weekly intervals
  - unlike PhEDEx, which uses a white board approach, keep old records to monitor storage usage evolution over time

Sources:

1. 2011 Computing and Offline monitoring workshop https://indico.cern.ch/event/137822/ Jakob's talk "Data operations" in Requirements session
2. Various discussions
3. Sites feedback during deployment campaign

Additional desired features

- Collect information about available space
  - storage technology dependent,
  - usually includes (file system) overhead
- Keep track of other metrics
  - number of files and directories
  - file access modification time (overlap with popularity?)
- Legacy data
- Aggregation of the file sizes on the level of the storage bookkeeping database (e.g. chimera)
  - possible and preferred for some storage technologies
  - replaces part of client functionality

Maintenance effort expected for code base and operations

- Site admins
  - set up and maintain process for producing storage dumps
  - Install/update spacemon client, eventually cvmfs client
  - maintain valid proxy for the uploads
- Site support team
  - monitor the sites uploads and help site admins as needed
- SpaceMon configuration
  - update default configuration distributed with a client following updates in DMWM Namespace
- Central infrastructure
  - maintain nodes in DMWMMON database and roles in SiteDB
  - dmwmmon data service hosted by CMSWEB, DB access
  - Visualization hosted by cms-dashboard

Near term development needs

- Summarized in :

https://twiki.cern.ch/twiki/bin/view/CMS/SpaceMonPlan2016

Some highlights:

- Data presentation & visualization
- Ensure consistent transactions on database level
- Add site validity interval
- Configuration for the aggregation parameters
- Test suite for new feature

Coordination points

- Coordination with WLCG Experiment support team
  - evaluating requirements for potential joint project involving ATLAS and LHCb
  - Support for tools for producing storage dumps
- Subprojects for summer students posted on CMS computing projects page
  - Advertised to Lithuanian students program at CERN
  - INFN students at Fermilab
- Contribution from the CMS site admins

Summary

- Proof of concept demonstrated with a quick prototype and successful deployment at [N] CMS sites
- Core of the new framework is ready
  - client code is fully decoupled from PhEDEx
  - central data service reuses PhEDEx core functions/utilities
- Use of CMS computing services & infrastructure
- Visualization/presentation layer is in design phase
- Many bells and whistles to be added
  - opportunities for summer students
- Feedbacks on requirements and the particular use cases are most crucial at this stage