The CMS Dataset Bookkeeping Service

Lee Lueking
CMS Offline Software and Computing

Sept. 3, 2007

CHEP 2007: Software Components, Tools, and Databases

Outline

- Motivation and overview
- Terminology and relationships
- Features and architecture
- Deployment and Operational experience

Ackowledgements: The DBS Team

- FNAL: Anzar Afaq, Yuyi Guo, Sergey Kosyakov, Vijay Sekhri
- Cornell:, Valentin Kuznetsov , Dan Riley, Drew Dolgert, Chris Jones

https://twiki.cern.ch/twiki/bin/view/CMS/DBS-TDR

Be sure to see Valentin Kuznetsov CHEP poster #224

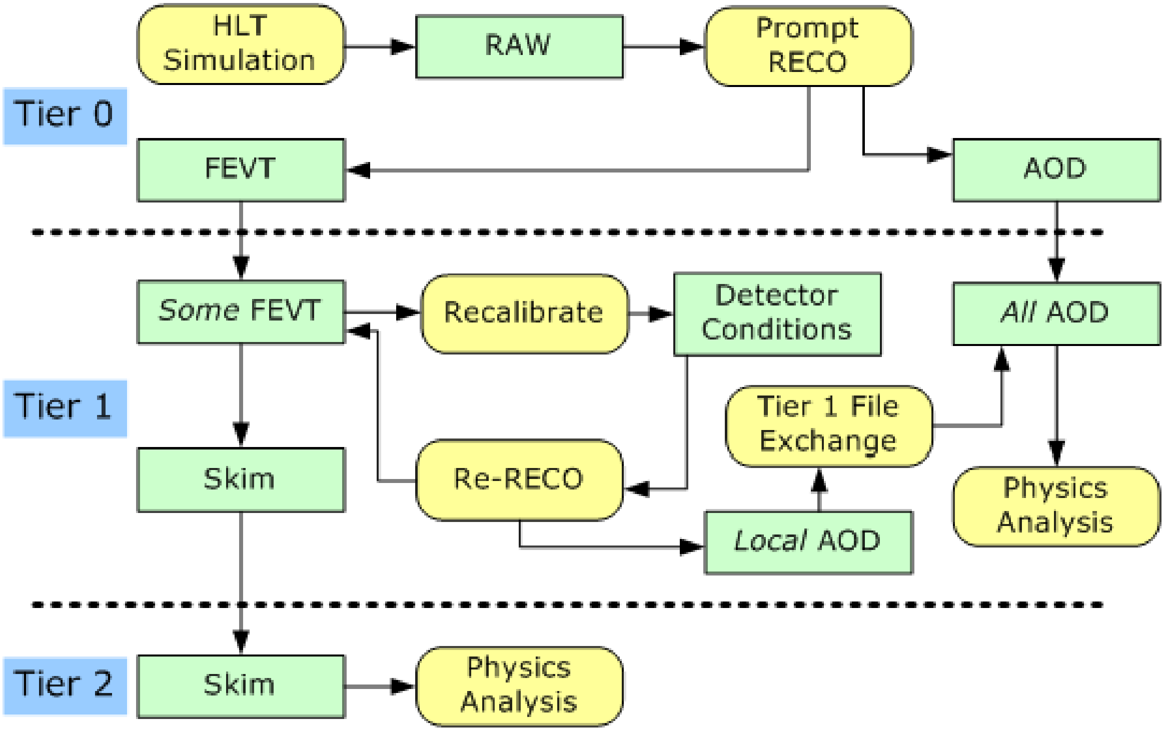
What is DBS?

Define, Discover and Use CMS event data

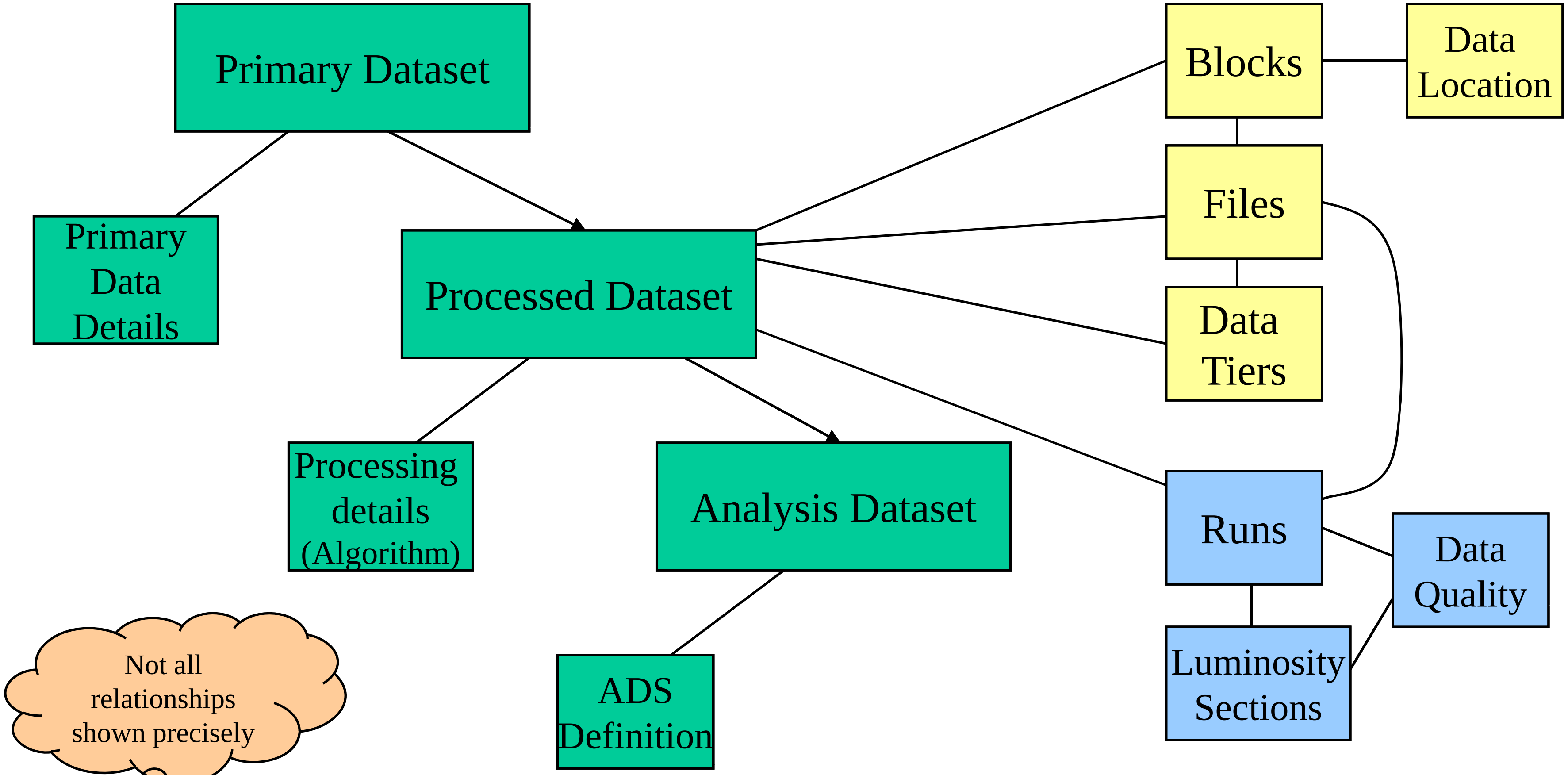
- Data definition:
  - Dataset specs: runs, lumi sections, algorithms, root branches,...
  - Track data parentage
- Data discovery:
  - What data exists
  - Dataset organization in terms of files/fileblocks
  - Site datablock location information
- Use:
  - WEB, CLI and API interfaces
  - Distributed analysis tool (CRAB)
  - Production data processing (ProdAgent)
  - Data distribution tool (PhEDEx)
  - End User Data Discovery

DBS Use Cases

- Typical workflows in data processing
- **Tier 0,1,2** refer to processing tiers (computing model)
- **RAW, FEVT, AOD** refer to data tiers (data model).
- Important steps include:
  - Adding new data
  - Processing existing data
  - Merging data (a.k.a combining, concatenating)
  - Skimming data (a.k.a. filtering or streaming)
- In all cases, the data provenance (history, parentage) is recorded.



Terms and Relationships



Not all relationships shown precisely

Data PATH=/PrimaryDS/ProcessedDS/Tier[/ADS-definition]

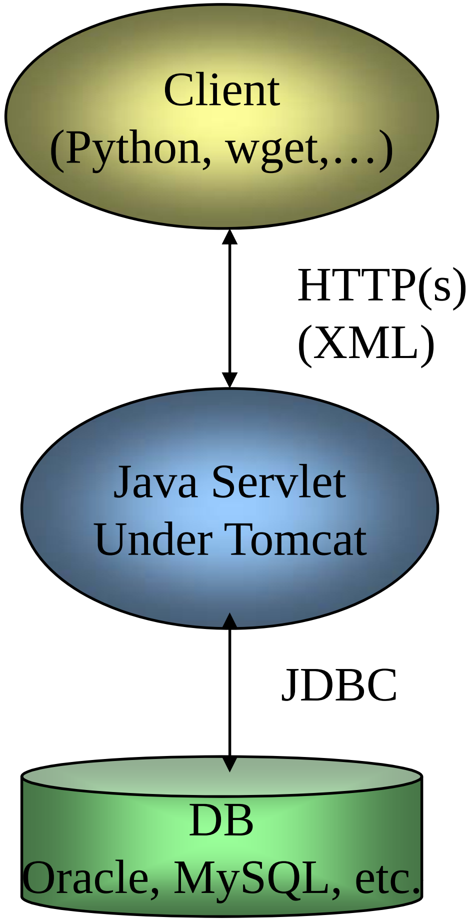
Schema Concepts (1)

- **Dataset**
  - **Primary Dataset:** determined by High Level Trigger (HLT) trigger classification or MC production parameters
  - **Processed Dataset:** a slice of a primary dataset with a consistent processing history. Note: *May include multiple copies of some events with slight differences in processing.*
  - **Analysis Dataset:** a snapshot of a subset of processed dataset representing a coherent sample for physics analysis
- **Luminosity Section**
  - Sub-section of a run during which time the instantaneous Luminosity is unchanging. (2^{20} orbits = 93 Seconds)
  - Unit of accounting for integrated luminosity
  - Production data **files** will contain one or many whole luminosity sections
- **Run:** Period of data-taking over which conditions are stable
- **Data Quality**: DQ flags set for Run or specific Lumi Sections of a Run

Schema Concepts (2)

- **Data Tier**
  - A set of objects grouped together for each event
  - Defined by the software release configuration files
- **Files**
  - Parentage relationships between files is recorded in addition to Dataset lineage
  - Files can be marked as “unavailable” if they are lost or corrupted.
- **Files Blocks**
  - Files grouped into blocks of reasonable size or logical content.
  - Tracking many files grouped into blocks has advantages in data transfers
  - Physical storage locations is recorded at the block level
- **Block Location**
  - Location of File Blocks at Site Storage Elements (SE)

DBS Architecture



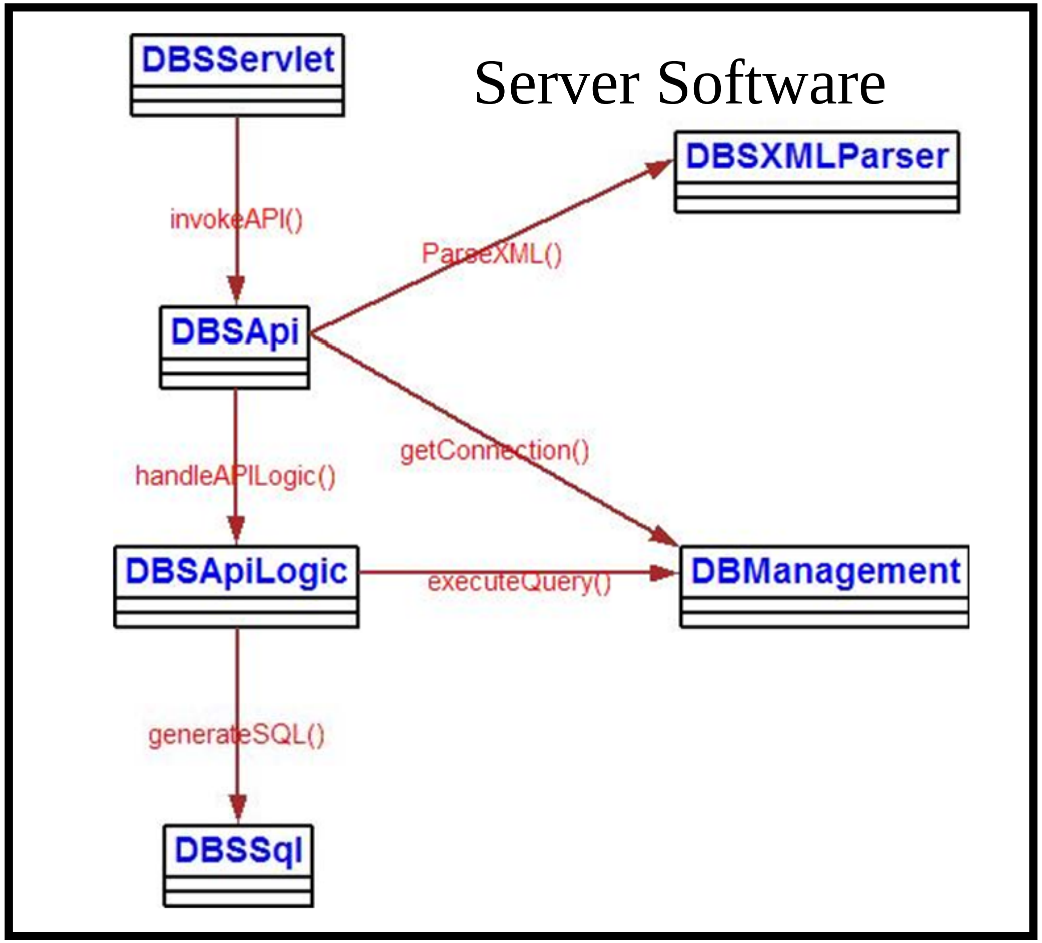
Client
(Python, wget,…)

HTTP(s)
(XML)

Java Servlet
Under Tomcat

JDBC

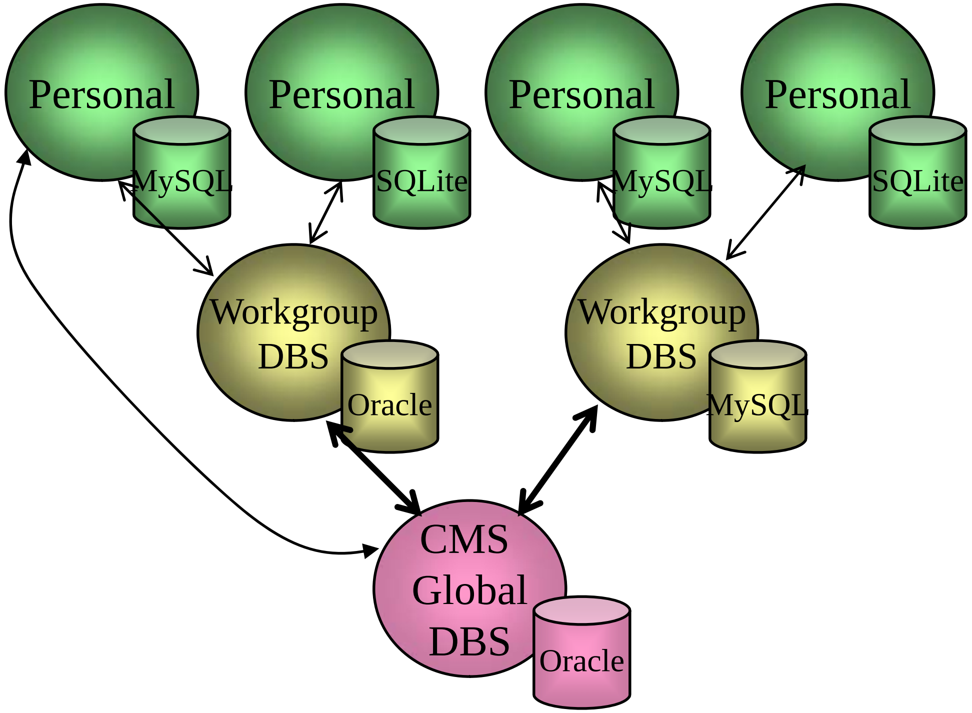
DB
Oracle, MySQL, etc.



Server Software

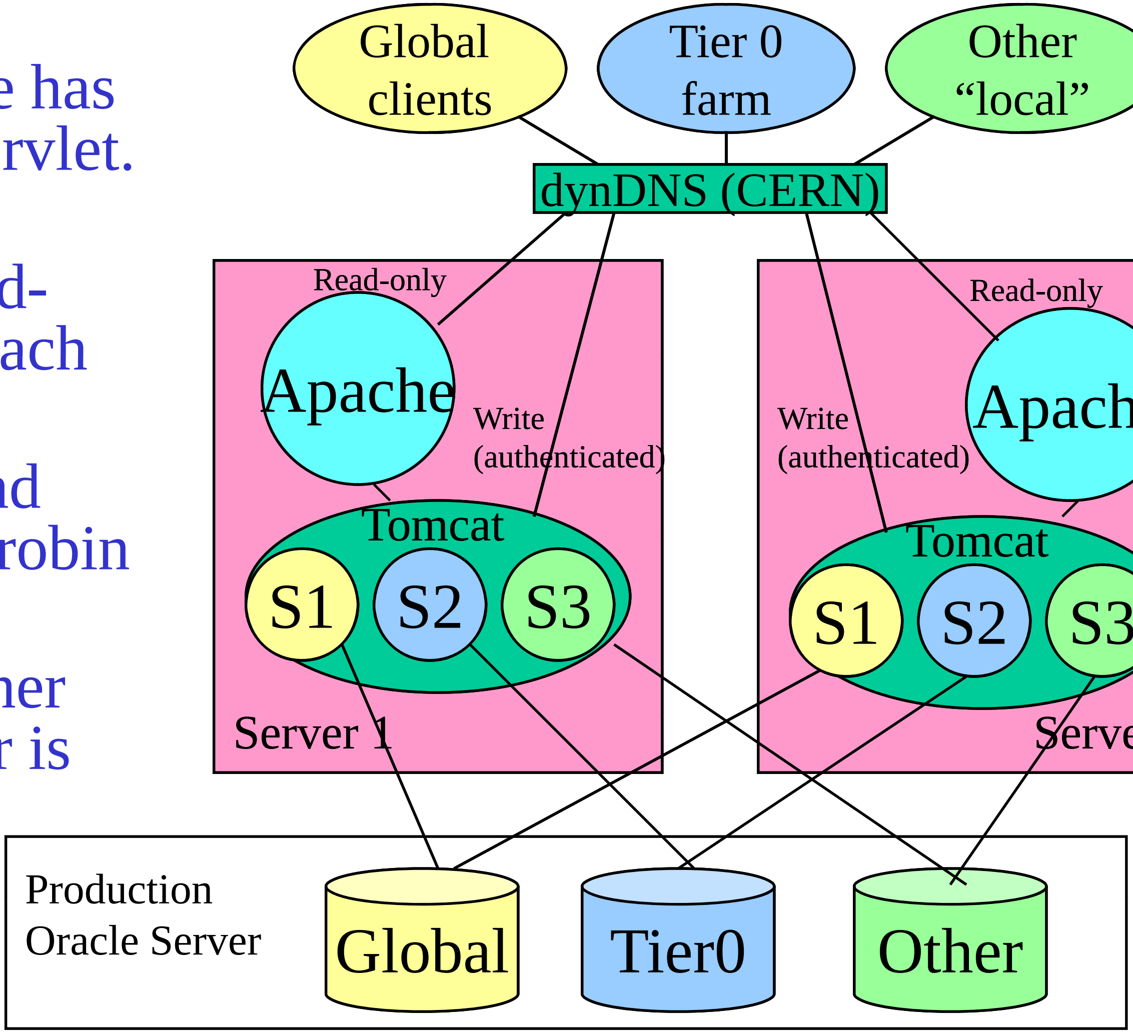
Scalability

- DBS is designed to support a hierarchical deployment model
- This enables the scalability needed to meet the needs of CMS
  - A single **Global** instance is the official repository for all CMS data
  - **Workgroup** instances are used for production processing and analysis groups
  - **Personal** instances can be used for private work. Still in evaluation.
- Instances can have Oracle, MySQL, or SQLite DB stores.
- Datasets are migrated from instance to instance. Certain rules apply to maintain consistency and streamline.

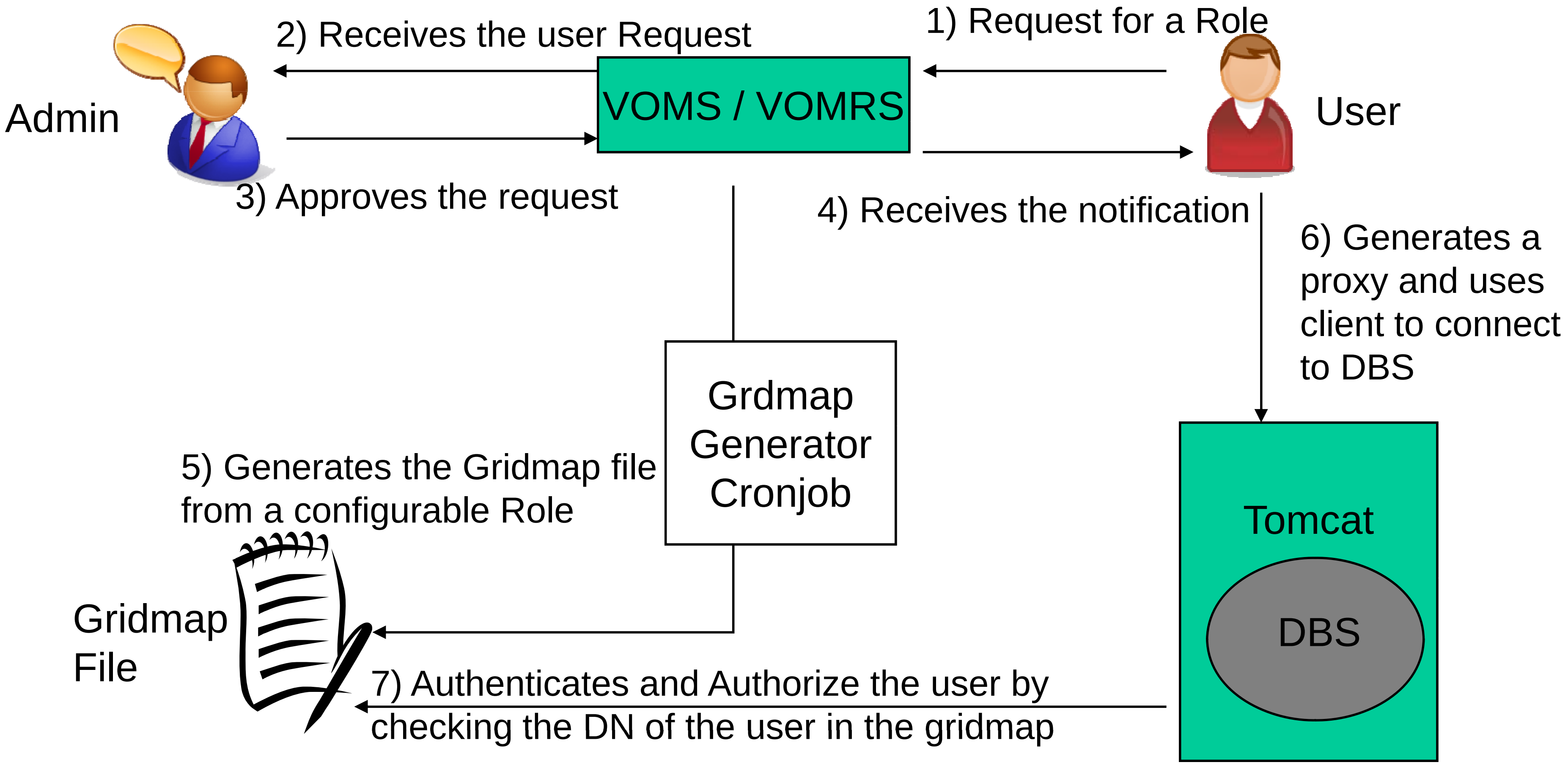


CERN Deployment

- Each DBS instance has DB account and servlet.
- Read-only and Authenticated Read-write servlets for each instance.
- Load balancing and failover via round robin DNS.
- Client retries to other server if one server is down.

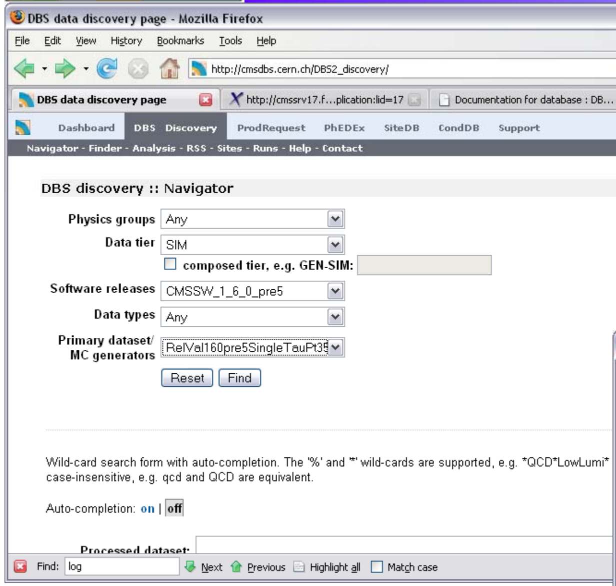


Security Architecture



1) Request for a Role
2) Receives the user Request
Admin
VOMS / VOMRS
User
3) Approves the request
4) Receives the notification
6) Generates a proxy and uses client to connect to DBS
Grdmap Generator Cronjob
5) Generates the Gridmap file from a configurable Role
Tomcat
DBS
Gridmap File
7) Authenticates and Authorize the user by checking the DN of the user in the gridmap

DBS Discovery Page


DBS data discovery page - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://cmsdbs.cern.ch/DBS2_discovery/
gnuplot log plots
DBS data discovery page
http://cmssrv17.f...plication:lid=17
Documentation for database : DB...
Leave Req. for Eric Wicklund, 16-...
Bracing for a Double Dip
Dashboard DBS Discovery ProdRequest PhEDEx SiteDB CondDB Support
You are not logged in, click here to login.
Navigator - Finder - Analysis - RSS - Sites - Runs - Help - Contact
View
Reset
Find
Find: log Next Previous Highlight all Match case

DBS discovery :: Navigator — Physicist

Physics groups: Any

Data tier: SIM

□ composed tier, e.g. GEN-SIM:

Software releases: CMSSW_1_6_0_pre5

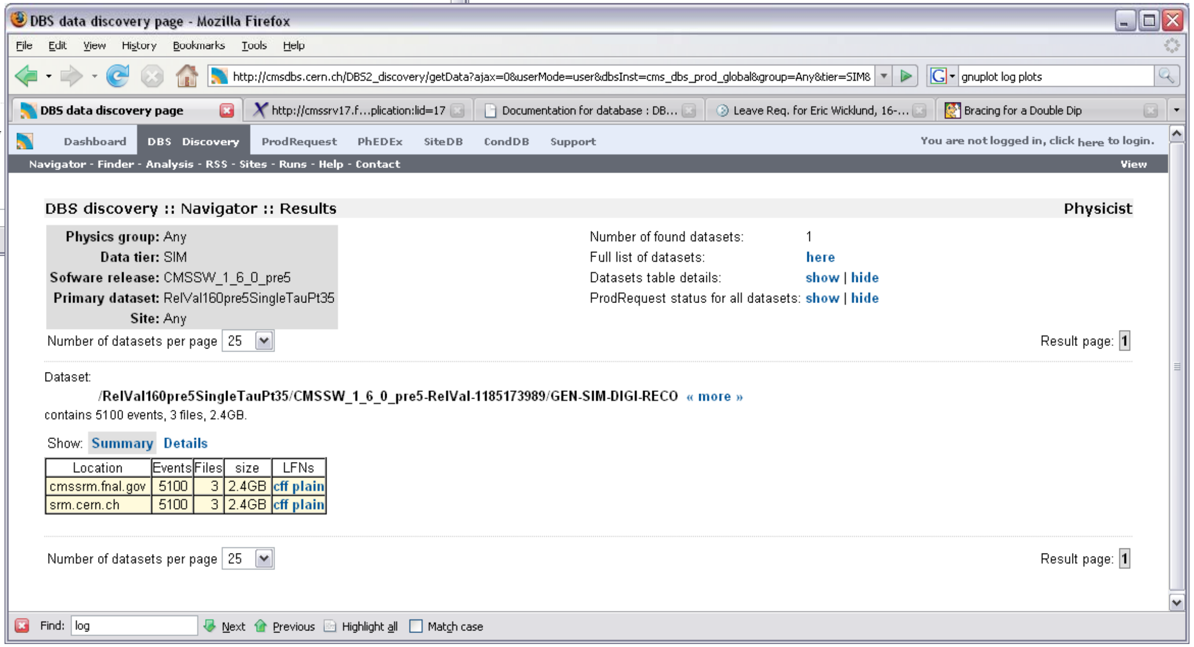
Data types: Any

Primary dataset/ MC generators: RelVal160pre5SingleTauPt35

Wild-card search form with auto-completion. The '%' and '*' wild-cards are supported, e.g. *QCD*LowLumi* case-insensitive, e.g. qcd and QCD are equivalent.

Auto-completion: on | off

Processed dataset:


DBS data discovery page - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://cmsdbs.cern.ch/DBS2_discovery/getData?ajax=0&userMode=user&dbsInst=cms_dbs_prod_global&group=Any&tier=SIM&
gnuplot log plots
DBS data discovery page
http://cmssrv17.f...plication:lid=17
Documentation for database : DB...
Leave Req. for Eric Wicklund, 16-...
Bracing for a Double Dip
Dashboard DBS Discovery ProdRequest PhEDEx SiteDB CondDB Support
You are not logged in, click here to login.
Navigator - Finder - Analysis - RSS - Sites - Runs - Help - Contact
View
Find: log Next Previous Highlight all Match case

DBS discovery :: Navigator :: Results — Physicist

Physics group: Any
Data tier: SIM
Sofware release: CMSSW_1_6_0_pre5
Primary dataset: RelVal160pre5SingleTauPt35
Site: Any

Number of found datasets: 1
Full list of datasets: here
Datasets table details: show | hide
ProdRequest status for all datasets: show | hide

Number of datasets per page 25 — Result page: 1

Dataset:
/RelVal160pre5SingleTauPt35/CMSSW_1_6_0_pre5-RelVal-1185173989/GEN-SIM-DIGI-RECO « more »
contains 5100 events, 3 files, 2.4GB.

Show: Summary Details

Location	Events	Files	size	LFNs
cmssrm.fnal.gov	5100	3	2.4GB	cff plain
srm.cern.ch	5100	3	2.4GB	cff plain

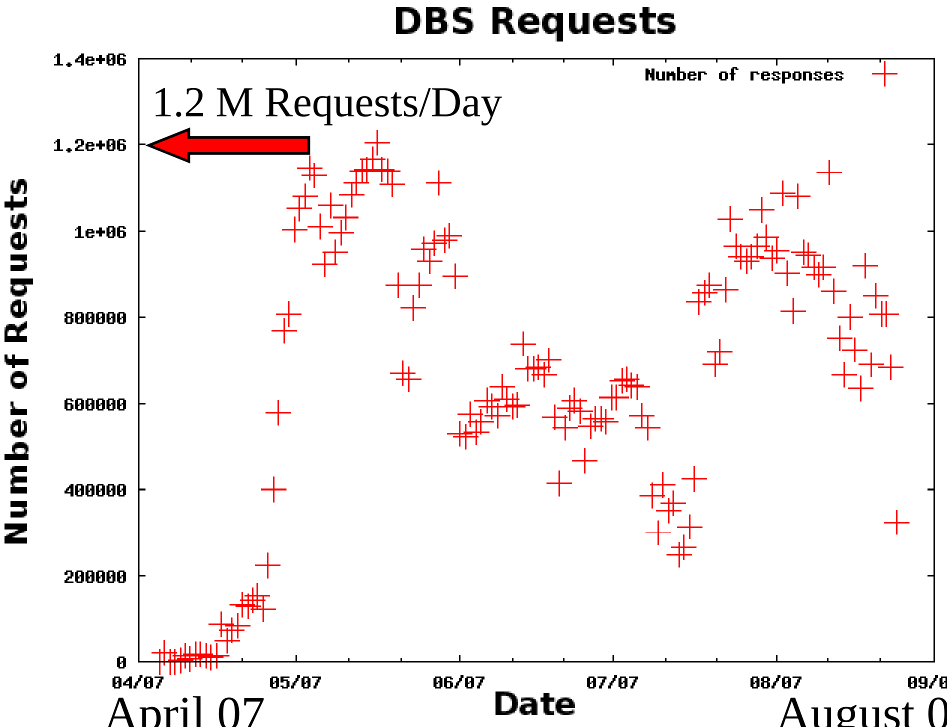
Number of datasets per page 25 — Result page: 1

- Convenient web-based interface to DBS
- See Valentin Kuznetsov’s presentation. (CHEP #223)

Operational Experience

- DBS-1 was a prototype in operation from the summer of 2006 to April 2007.
- DBS-2, described in this talk, has been in operation since April 2007 and used to record all MC production and analysis steps. All data was migrated from DBS-1 to DBS-2
- Deployment (on DBS CERN servers):
  - One Global DBS instance is the authoritative source of data information for CMS at large.
  - Six local instances are used for MC production.
  - One Tier0 instance and several test instances.
- Highly reliable and stable
- STATS (Global instance):
  - Events : 740 M, Files : 321834, Blocks : 12063
  - Primary DS : 1431, Processed DS : 2638

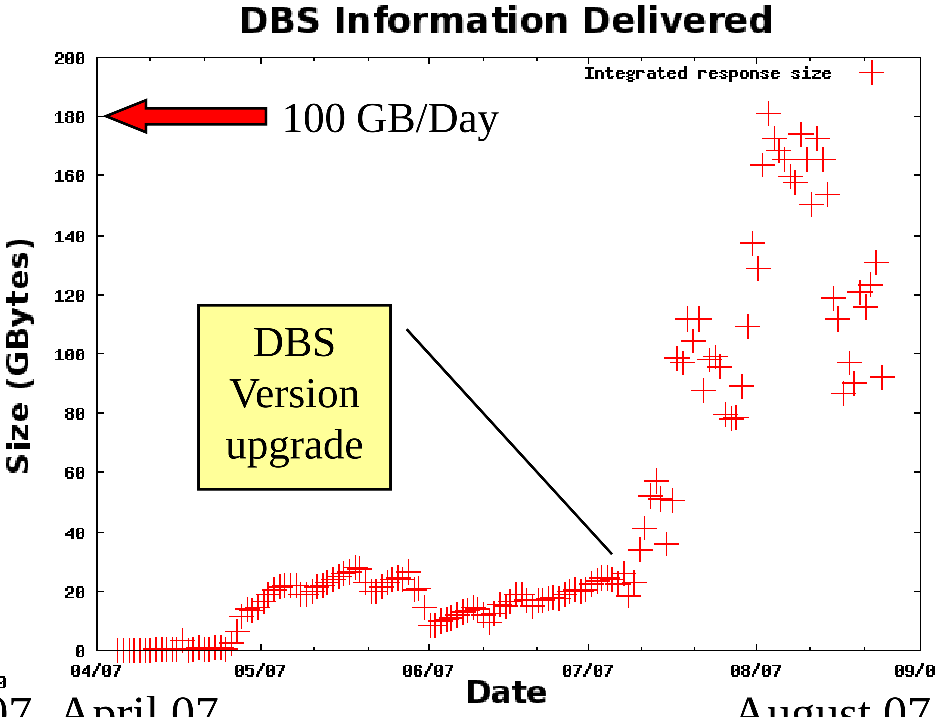
Operational Experience



1.2 M Requests/Day

April 07 August 07

- Requests to DBS clients per day
- Activity tied to MC processing



100 GB/Day

DBS Version upgrade

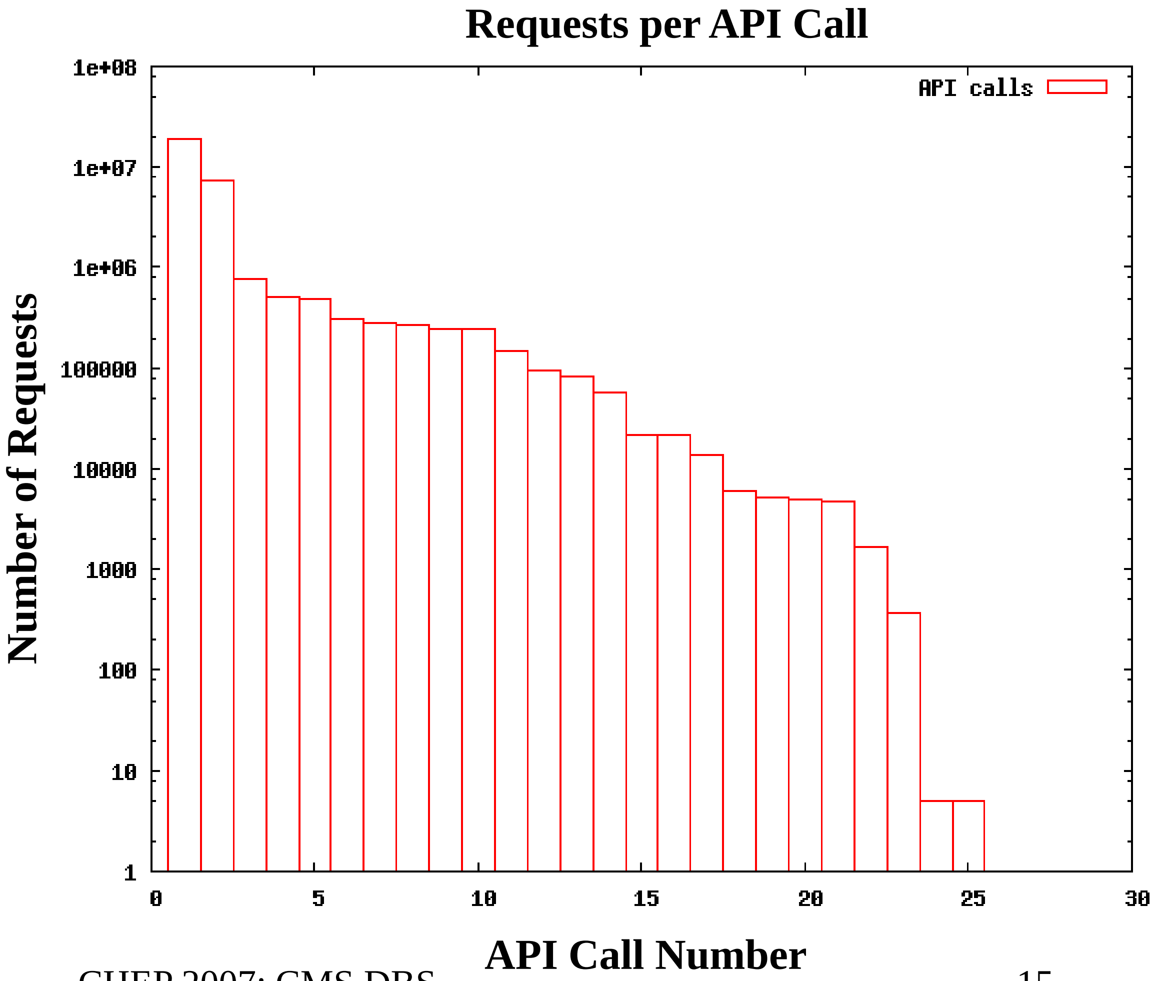
April 07 August 07

- Total information delivered by server to DBS clients per day
- Improvements planned to streamline response payloads

API Call Frequencies

Top Ten List

1 listBlocks
2 listFiles
3 insertFiles
4 insertStorageElement
5 listFileLumis
6 insertAlgorithm
7 insertProcessedDataset
8 insertPrimaryDataset
9 insertRun
10 insertLumiSection



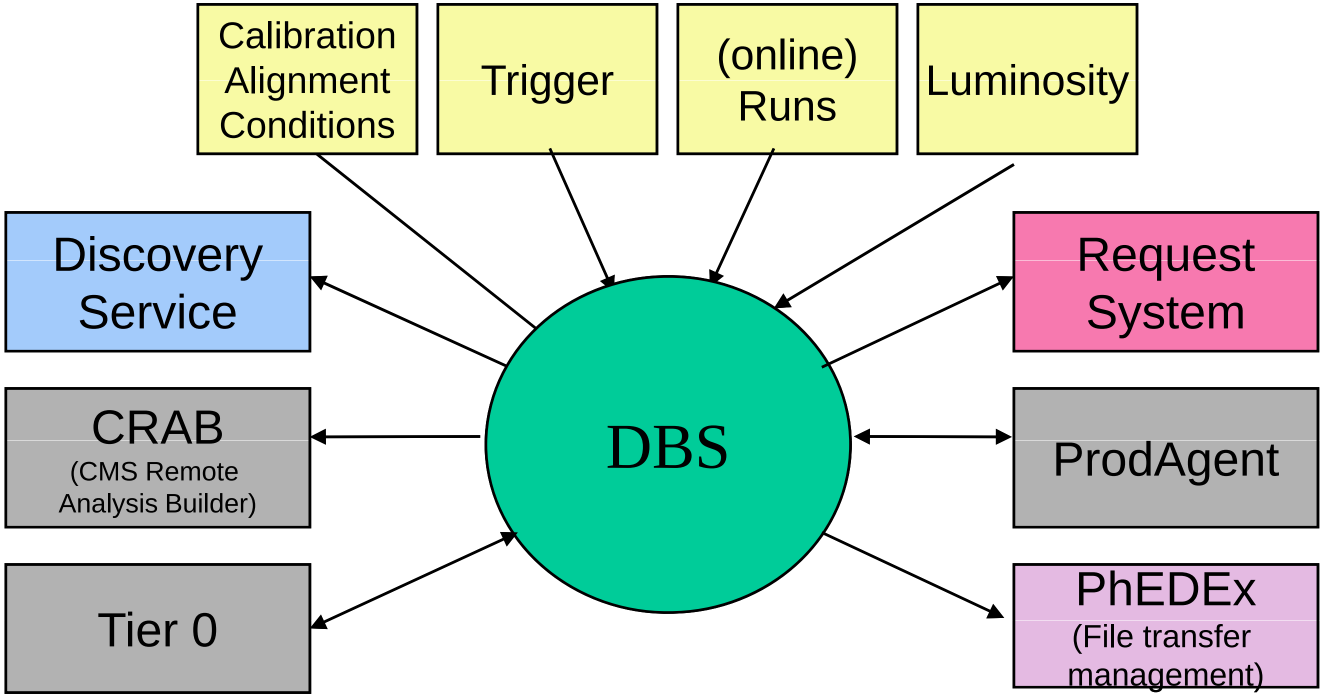
Summary

- DBS is used by CMS to record and track the history of all event data.
- It is designed to accommodate the data processing and event data models of CMS, and integrate with the workflow tools.
- The current deployment has demonstrated good functionality, scalability, stability and performance.
- Examination of usage patterns, performance metrics and additional use cases are being used to plan future enhancements.

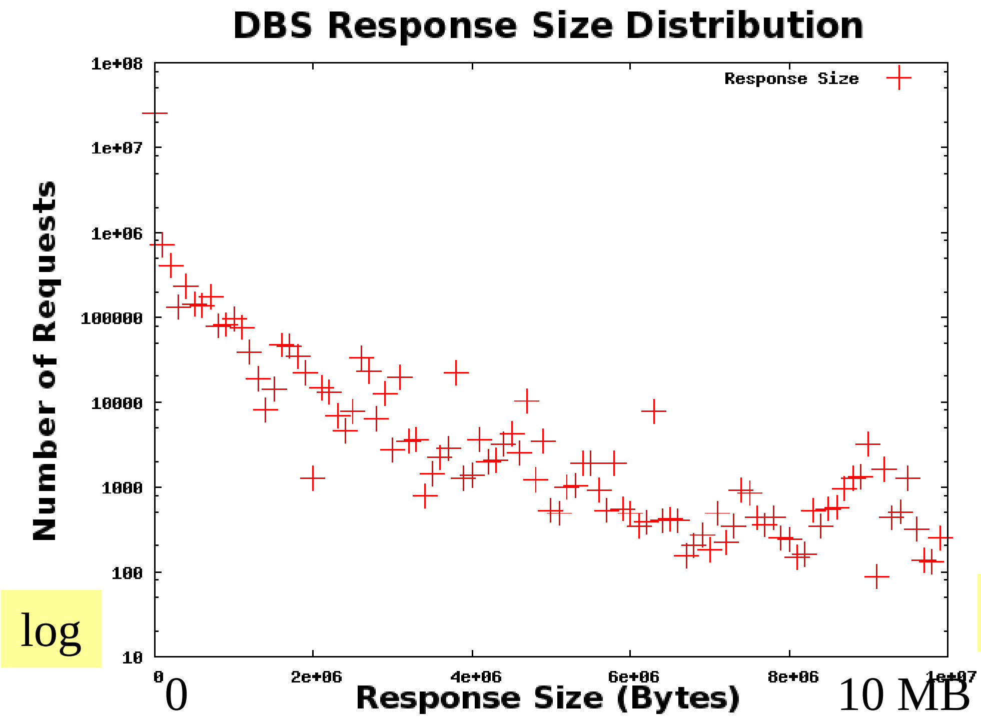
Finish

DBS in CMS

- DBS has many connections to CMS information systems
- DBS tracks event data from its initial sources through all processing steps.
- DBS is used by CMS Workflow Management tools



Response Size and Time

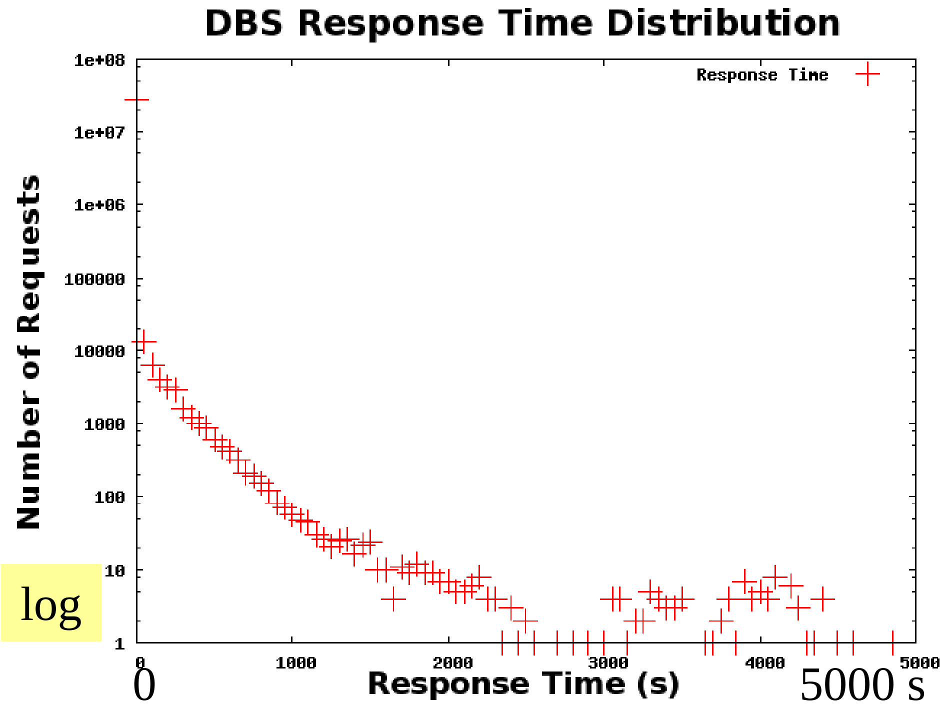


log

0

10 MB

- Response sizes range up to a few Mbytes
- Large responses are lists of files.
- Plans to reduce some overheads



log

0

5000 s

- Response times can range up to a few minutes and correspond to payload size.
- Typical “insert” is very fast.