CMS Database Status Report

Lee Lueking

CD CMS Activities Meeting

Feb 21, 2006

Contents

- Overview
- CMS “Conditions” Data
- Online: HCAL + PIXEL
- Offline: POOL + Frontier
- Conclusions

CMS Non-event Data Model

Today's Report Will Focus on Conditions Data

- Conditions data: (Calibration, Alignment, Geometry, Configuration, Monitoring)
  - Online DB: Schemas designed specific to sub-system; Oracle DB server at P5
  - Offline DB: POOL-ORA repositories
    - HLT Farm: Oracle DB server at P5
    - Tier-0:Oracle DB server at IT
  - Online to offline: Transform online format to POOL-ORA payloads. Transfer POOL-ORA payloads to offline via Oracle Streams.
  - Offline (Tier-0) to Tier-N:
    - Plan A: POOL-ORA repository ⇨ FroNTier "pass through" server ⇨SQUID proxy/caching servers ⇨ CMS Client POOL-RAL .
    - Plan B: Oracle replication to Tier-1 sites, if Plan A is insufficient or fails.
- Event Data Management System (DMS)
  - Tier-0: Dataset Bookkeeping Service (DBS), Dataset Location Service (DLS)
  - Tier-1: Local DBS and DLS for internal bookkeeping. Possible use of Oracle replication w/ Tier-0 if needed for performance/availability.
  - Tier-2: Local DBS and DLS for internal bookkeeping (not Oracle).

Conditions Overview(1)

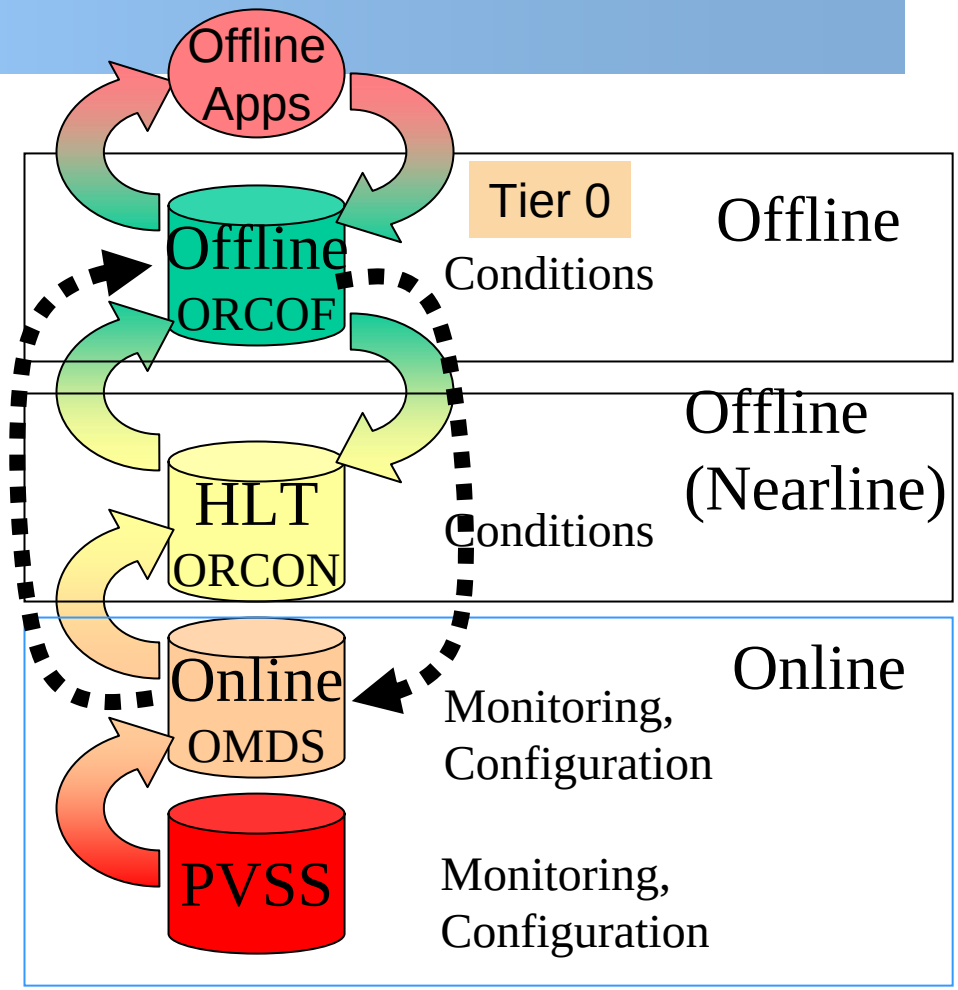
- Generically referred to as Detector non-event-data (NED)
- Three main areas of work:
  - Calibration: Conditions information such as detector slow controls, monitoring, and NED such as pedestals, and gains.
  - Alignment: Detector specific and global alignment including the Laser Alignment System.
  - Infrastructure: Database and offline framework software for maintaining and accessing NED. Tools transferring NED from online to offline, and offline to computing Grid.
- There are 8 sub-detectors, and several other sub-systems (i.e. trigger, luminosity) that will need management of their NED.
- Activity is coordinated by Lee Lueking (FNAL) and Oliver Buchmüller (CERN). (LL replaced by Luca Malgeri (CERN) eff. Feb. 2006)

Conditions Overview(2)

- Last summer CMS assembled the infrastructure components for the calibration, alignment and geometry service.
- Starting in November, we began integration and testing of the existing systems.
- Sub-detectors included in the testing were ECAL, HCAL, Si Strip Tracker, and Drift Tubes.
- The work included large scale testing, to understand how the online to offline, and Tier-0 to Tier-n delivery systems will perform under load and perturbation.
- The system will be used for the Cosmic Challenge the end of March.

On and Off line Databases

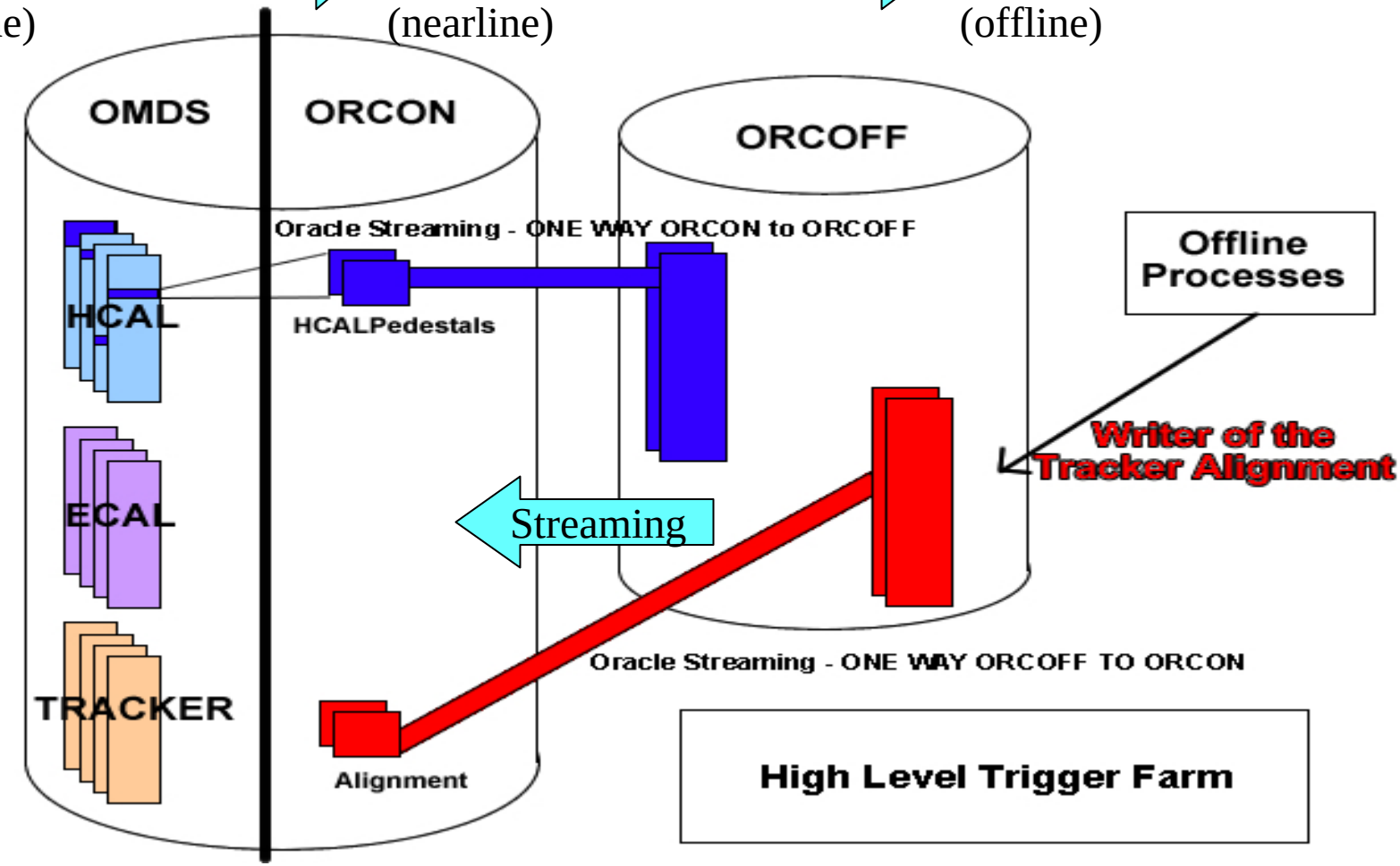
- The offline (ORCOF) DB will run on a production server in CERN/IT.
- The other servers will be located at P5 (current thinking).
- Assumed flow of data is shown by colored arrows. Additional flows are possible (dashed arrows).
- Currently testing mechanism(s) for managing these flows in a fail-safe way.



Conditions=calibration,alignment, slow controls data

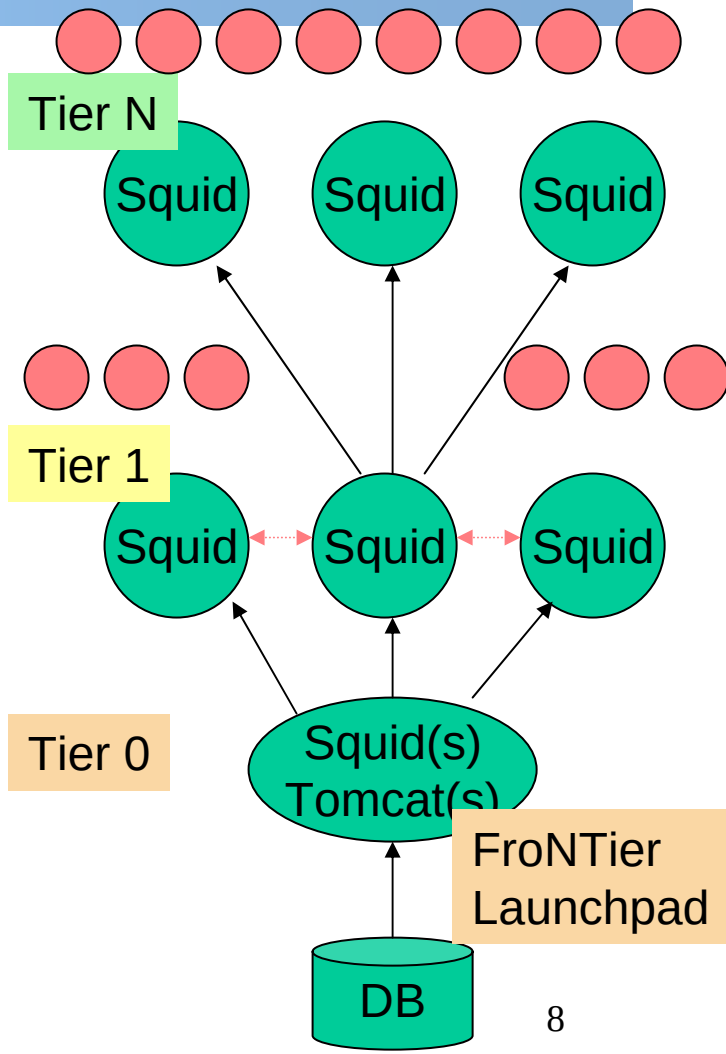
Online to Offline Infrastructure

RDBMS (online) → Transform → POOL-ORA (nearline) → Streaming → POOL-ORA (offline)



Offline Deployment Strategy

- Tier-0 (CERN):
  - Central offline calibration and alignment DB.
  - Application servers (specifically Apache Tomcat) deliver data from DB to Clients via HTTP.
  - Caching servers (specifically Squids) cache and deliver data to clients.
- Tier-1: (FNAL for example)
  - Caching servers (Squid) cache and deliver data to clients, and Tier-N sites.
- Tier-N:
  - Caching servers (Squid) cache and deliver data to clients.



USCMS Manpower

Person	Contribution
Michael Case (UC Davis)	Detector Geomety infrastructure, alignment software support
Yuyi Guo (FNAL)	HCAL and Si Pixel Database and infrastructure for data loading and maintenance.
Saima Iqal (FNAL)	Online to offline testing, online and offline database software support.
Zhen Xie (Princeton)	Offline calibration and alignment “conditions service” within CMSSW framework. Includes needed CMS POOL infrastructure.
Lee Lueking (FNAL)	Calib and Align coordination. Offline DB access infrastructure and coordination with LCG and CERN-IT.
Gennadiy Lukhan (FNAL/PPD)	HCAL and Si Pixel DB Schema. Online DB applications and interfaces.

Calibration Status Summary

Area\Detector	Tracker Strip	Tracker PIXEL	Muon DT	Muon CSC	Muon RPC	ECAL	HCAL	Pre Shower
Calib DB (Online)	☺	☺	☺	☺		☺	☺	👍
Offline Object Definition(s)	☺	😐	☺	☺		☺	☺	👍
O2O Xfer/Xform	☺		☺	☺		☺	☺	
Unified Event Setup + IOV	☺		☺	☺		☺	☺	
Apply cal in FW	☺		😐	😐		☺	☺	
Perf & Stress test	👍					☺	☺	

Not Participating in Cosmic Challenge

Only Cable Mapping Needed for CC

Not Participating in Cosmic Challenge

☺ = Near final, 😐 = Working example, 👍 =Starting, blank/? = none/unknown

HCAL & PIXEL Online

Generalized Schema

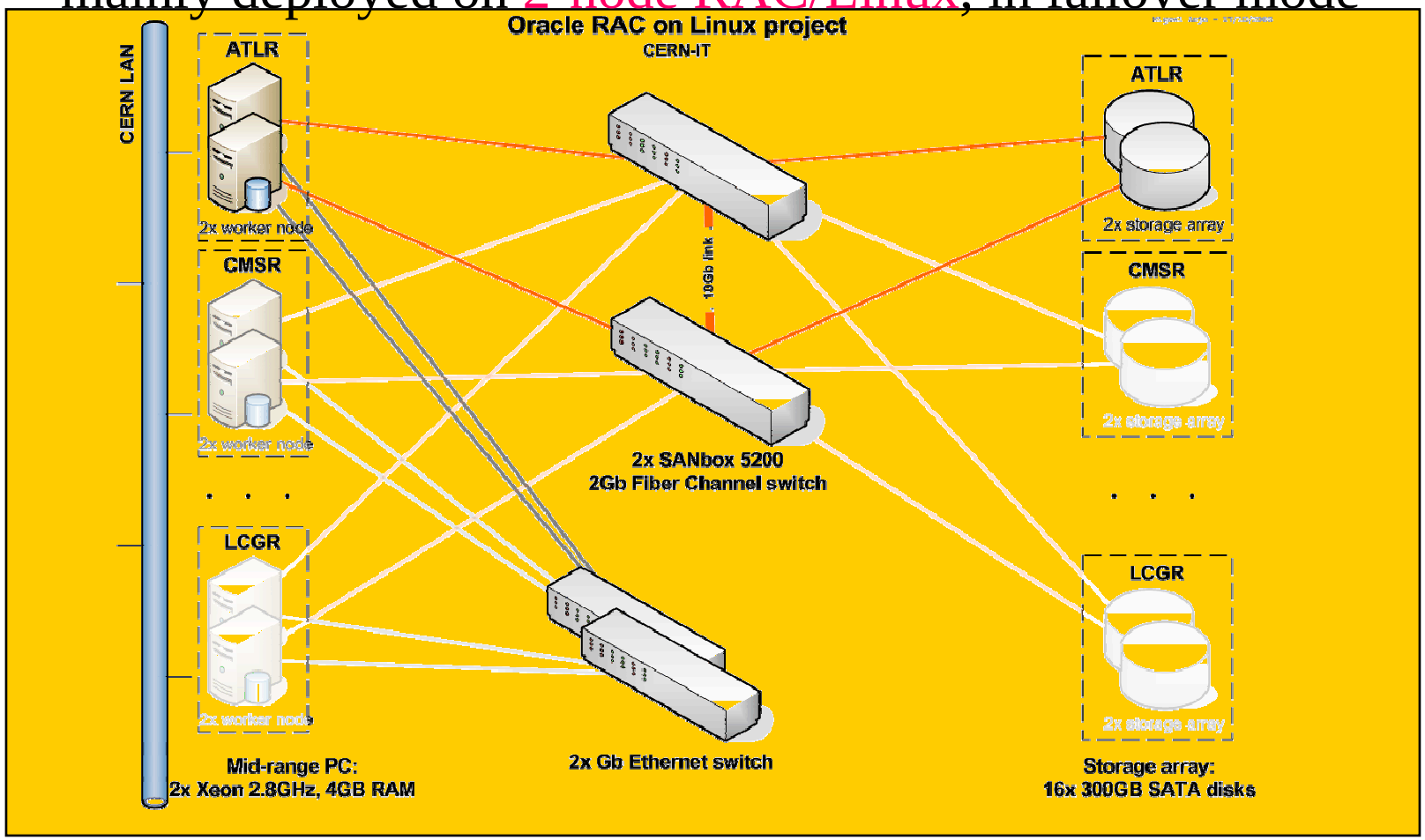
- A Generalized approach was developed enabling construction, configuration, and conditions data to use the same schema.
- Tools have been developed to help automate the generation of tables, load the data, and an API is available for access.
- Both HCAL and PIXEL are using this schema.
  - HCAL has loaded construction info, and calibration data for the Magnet Test/Cosmic Challenge (MTCC) & 2006 testbeam.
  - PIXEL is loading detailed testing and construction info and tracking their progress building the barrel and forward pixel detectors. This data will be used for configuration when the pixel detector is commissioned in 2008.
- Yuyi Guo and Gennadiy (PPD/CMS) are the major contributors.

Infrastructure Support

- FNAL/CD/DSG provides infrastructure support for these efforts:
  - CMSCALD – Development server
  - CMSCALIR – Integration server
- CERN/IT/DB provides a CMS production server on which HCAL has, and PIXEL will soon have, a schema.
- CERN/IT/DB provided a validation server for CMS testing in November-January (described later).

CERN Service Architecture

- The Physics Database Production and Validation services are mainly deployed on 2-node RAC/Linux, in failover mode



Hardware evolution for 2006

- Ramping up of the hardware resources in 2006-2008

Current State							
ALICE	ATLAS	CMS	LHCb	Grid	3D	Non-LHC	Validation
-	2-node offline	2-node	2-node	2-node	-	-	2x2-node
	2-node online test						
Proposed structure in 2006 (end of Q106)							
2-node	n-node or nx2-node	4-node or 2x2-node	n-node or nx2-node	n-node or nx2-node	2-node	2-node (PDB replacement)	
	2-node valid/test	2-node valid/test	2-node valid/test	2-node pilot		Compass??	
	Online?						

Integration & Testing

November Test Goals

- Understand the Performance of POOL-ORA
- Demonstrate the feasibility for deployment of Oracle Streams (online to offline) and Frontier (offline to Tier-N) infrastructure.
- Explore the maintenance and operation issues of the model.
- Test for the following:
  - Functionality
  - Performance
  - Reliability
  - Scalability
- Compare Frontier caching approach to direct Oracle Access.

Database Configuration

- Four types of accounts:
  - Owner: Owns tables
  - Writer: Read, Insert
  - Reader: Read
  - Admin: Read,Insert,Update,Delete
- Each Sub detector has its's own schema. A general "user" has read role for all schemas.
- Loaded with estimated 6 Months simulated data for HCAL, ECAL.
- SiStrip Tracker and Muon Drift Tubes (DT) have small amounts of data loaded.

Database Storage Conclusions

- The space summary of for the users on the validation server is shown below.

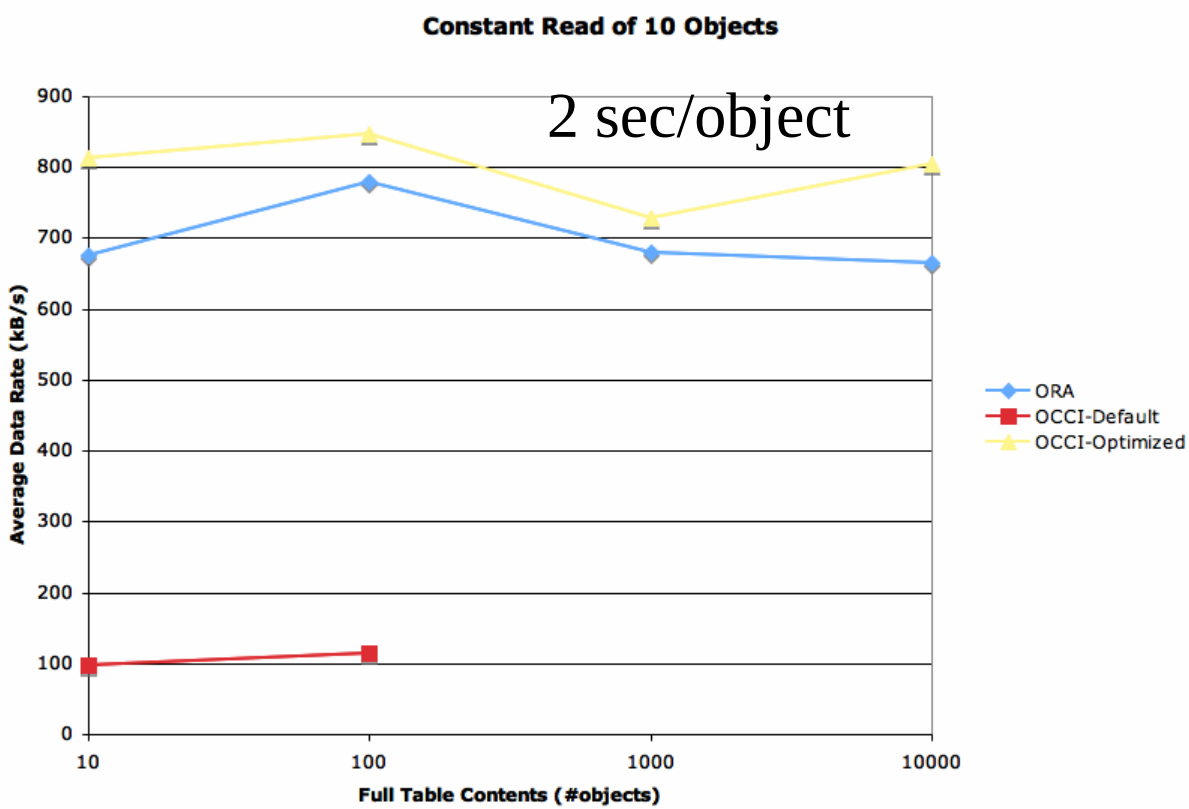
```
OWNER SIZE_IN_MEGABYTES
------------------------------ --------------------
CMS_VAL_DT_POOL_OWNER      54
CMS_VAL_ECAL_POOL_ADMIN  0.0625
CMS_VAL_ECAL_POOL_OWNER  44941
CMS_VAL_GENERAL_POOL_OWNER 0.5
CMS_VAL_HCAL_POOL_ADMIN  2.625
CMS_VAL_HCAL_POOL_OWNER 12093
CMS_VAL_STRIP_POOL_OWNER 1.75
```

- Analyzing 10,000 Ecal Pedestals in the DB, they are about 1.6 MB each considering payload data alone a total of 16000 MB of data.
- The size used in the database is 2.8 x that stored, and this seems like a reasonable database overhead, especially since we are not considering additional indexes and the IOVs.

ECAL Oracle Write and Read Tests

Ricky Egeland

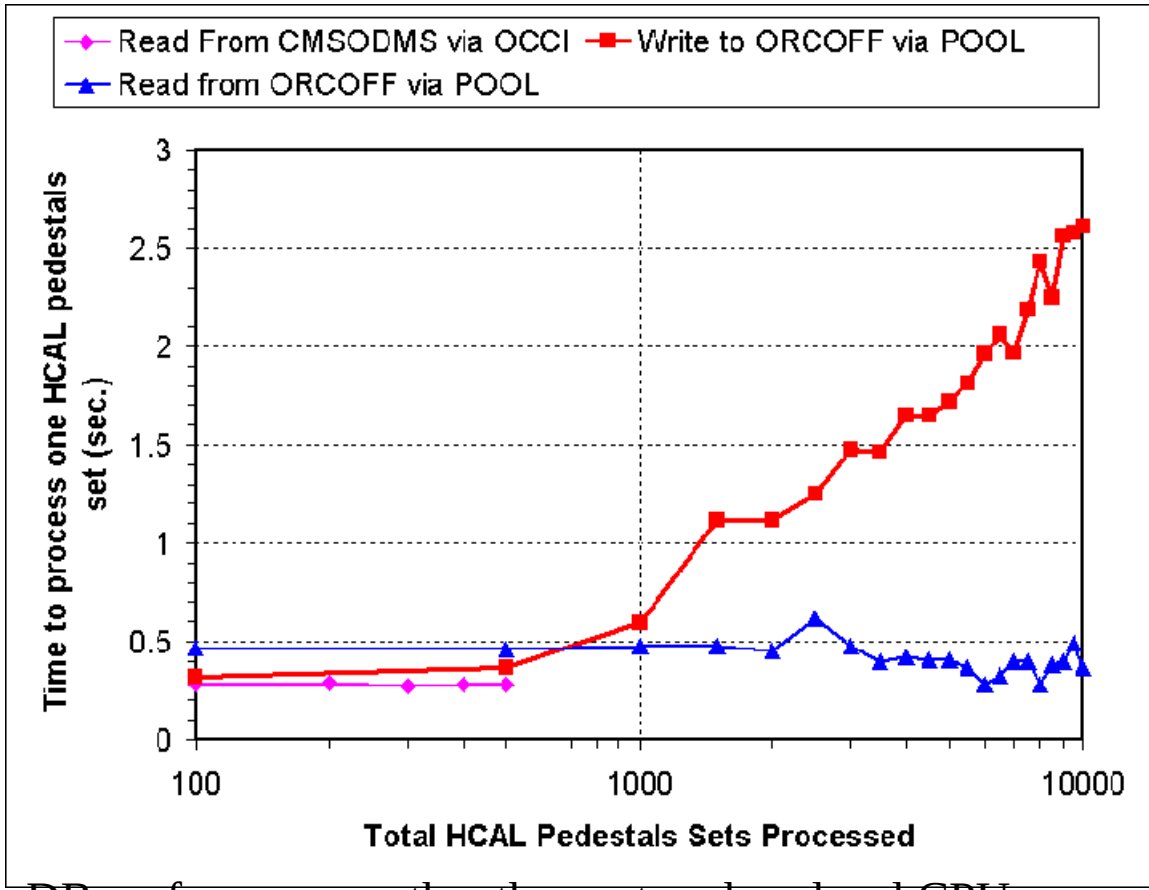
- Write and read ECAL data to/from validation test server. Client @ CERN
- Comparison of access rates for 10 1.63MB ECAL objects
- Three modes:
  - POOL-ORA (blue),
  - OCCI basic (red),
  - OCCI optimized (yellow)
- Time of 2 sec per object is consistent with previous measurement.



HCAL Direct Oracle Test

Fedor Ratnikov

- Test:
  - Read from online (CMS-OMDS),
  - Write to offline (ORCOF) DB.
  - Read from the Offline DB w/POOL-ORA.
- HCAL pedestal objects are ~200 kByte each.
- Client on lxcmsd2 not heavily used a the time of test.
- Results: POOL-ORA access is consistent with (somewhat better) other measurements. Read time is not dependent on number of objects in DB.



- Bottleneck for calibrations IO is DB performance, rather than network or local CPU.

Using Frontier w/ POOL

- The Object to Relational mapping provided in the original Frontier product (ala CDF) is not used.
- A query pass-through feature is added to the Frontier servlet.
- A POOL(CORAL)/Oracle-Frontier plugin is provided that uses a special Frontier client library.
- Frontier features employed are: 1) HTTP is used for the transport protocol from the Frontier server to client, 2) Caching is provided by Squid proxy/caching servers.

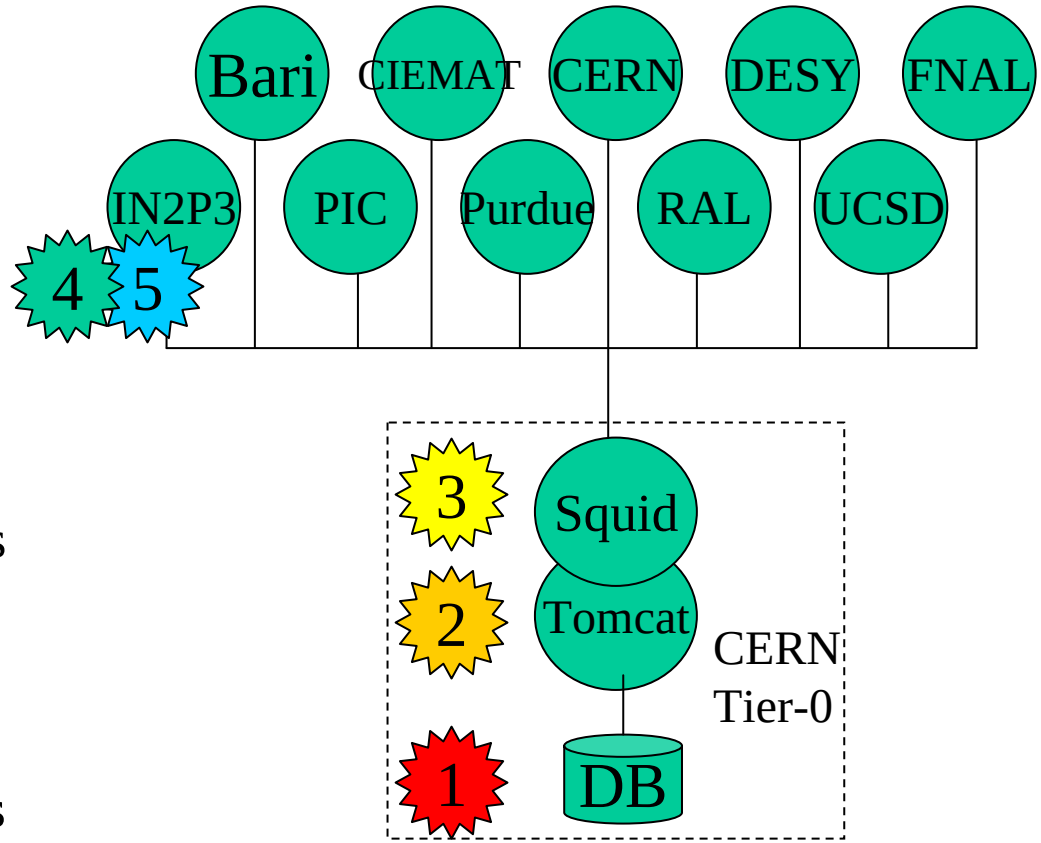
Participating Sites

Site Name	Grid Tier	Location	Contact Person	Squid Installed	Squid Works	Python Test	CMSSW Test
Bari	2	Bari, Italy	Giacinto Donvito	Y	N	Y	N
CERN	0	Geneva, Switzerland	Sinisa Veseli	Y	Y	Y	N
CIEMAT	2	Madrid, Spain	Jose Hernandez	Y	Y	Y	Y
DESY	2	Hamburg, Germany	Michael Ernst	Y	Y	Y	Y
FNAL	1	Batavia IL, USA	Gary Stiehr	Y	Y	Y	Y
IN2P3	1	Lyon, France	Artem Trunov	Y	?	N	N
PIC	1	Barcelona, Spain	Jose Hernandez	Y	Y	Y	Y
Purdue	2	West Lafayette, IN, USA	Christopher A. Baumbauer	Y	Y	Y	N
RAL	1	Didcot, UK	Catalin Condurache	Y	Y	Y	Y
UCSD	2	San Diego, CA, USA	Terrence Martin	Y	N	Y	N

- Tests were submitted LCG and OSG Grid sites
- Setup Conclusions:
  - Squid installation completed within ~1 week very few problems.
  - Squid worked “out-of-the-box” at the 80% level
  - Simple test w/ Python client quickly confirmed operation
  - CMS SW framework release required more debugging

The Tier-0 to Tier-N Test w/ Frontier and Oracle

1. Direct Oracle access with the POOL Oracle plugin.
2. Direct Frontier server with the POOL Frontier plugin.
3. CERN Squid access with the POOL Frontier plugin.
4. First Local Squid access with the POOL Frontier plugin.
5. Second Local Squid access with the POOL Frontier plugin. (assures cache loaded)



Clients: 1. Python (No direct Oracle) , 2. CMSSW Client

CMSSW Client: Single & Parallel Access Performance

HCAL (200 kB)

Site	Oracle@CERN (1)	Frontier@CERN (2)	Squid @ CERN (3)	Squid @ site (5)
CIEMAT	0.5 s	0.46 s	0.36s (0.37s)	0.18s (0.18s)
FNAL	1.6 s	0.68 s	0.87s (0.68s)	0.15s (0.15s)
RAL	0.38 s	0.38 s	0.29s	0.21s
CERN	0.8s	6.0	1.7	Squid@CERN

ECAL (1.6 MB)

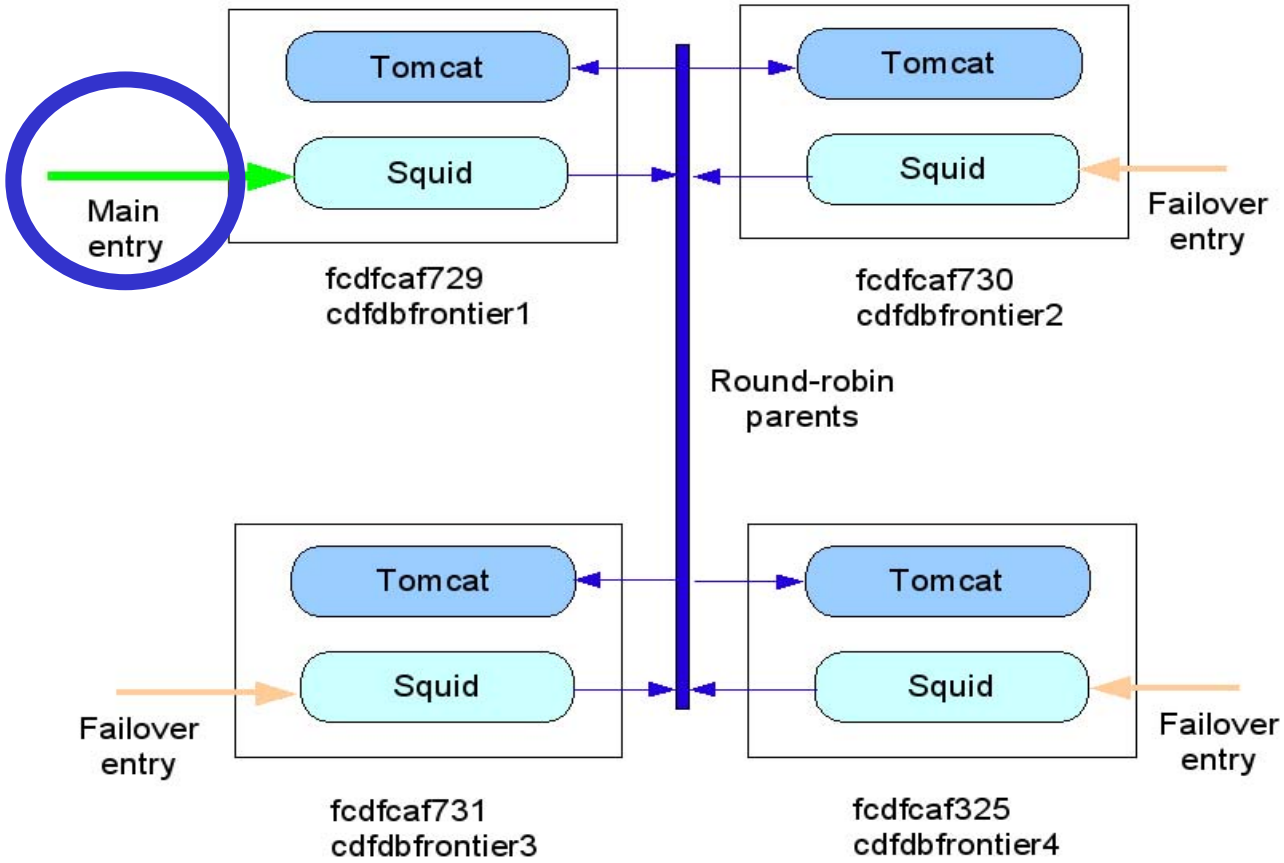
Site	Oracle@CERN	Frontier@CERN	Squid @ CERN	Squid @ site
CIEMAT	39.4 s	26.5 s	23.3s (24.8s)	12.7s (12.9s)
FNAL	78.8s	36.4 s	38.9s (31.5s)	8.51s (8.8s)
RAL	25.1 s	25.9 s	20.1s	14.6s
CERN	2 (2.2)	21	14	Squid@CERN

Numbers in (blue) are for 20 simultaneous clients.

Test Conclusions

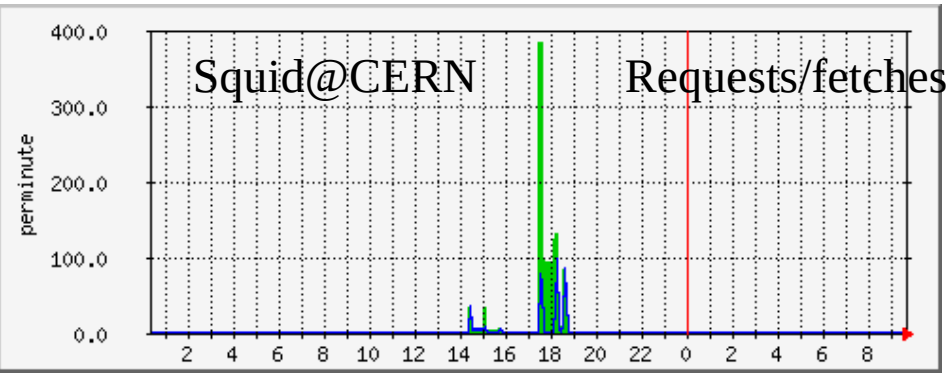
- Squid installation at 10 participating sites went smoothly.
- Squid and Tomcat stable at CERN. Frontier servlet stable.
- POOL/Oracle-Frontier plugin debugged. Additional work identified.
- Frontier object “packing” has excessive “bloat” (6.7 x for ECAL). Possible fix (uses zip) in progress.
- Frontier local cache for CMSSW clients reduced access times by as much as 10x (FNAL) over direct Oracle access form CERN.
- Oracle access at CERN is comparable or better than local Frontier Squid.
- ECAL and HCAL tests completed.
- SiStrip starting and important due to object sizes (~10 MB); PIXEL (~100MB). Initial results indicates serious optimization is needed.

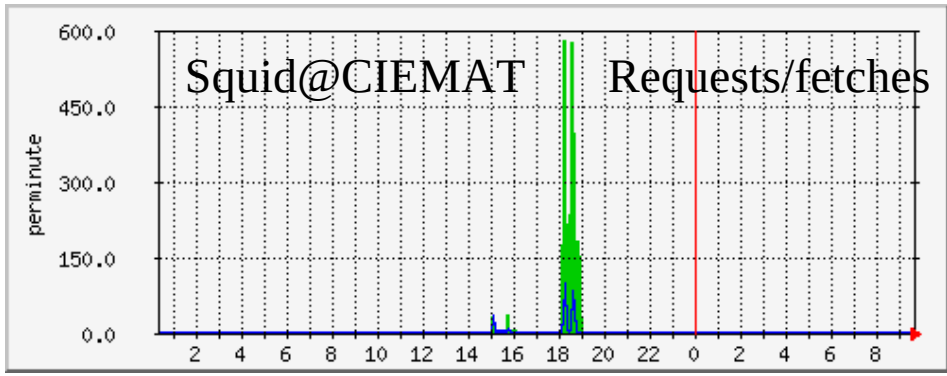
Frontier Launchpad

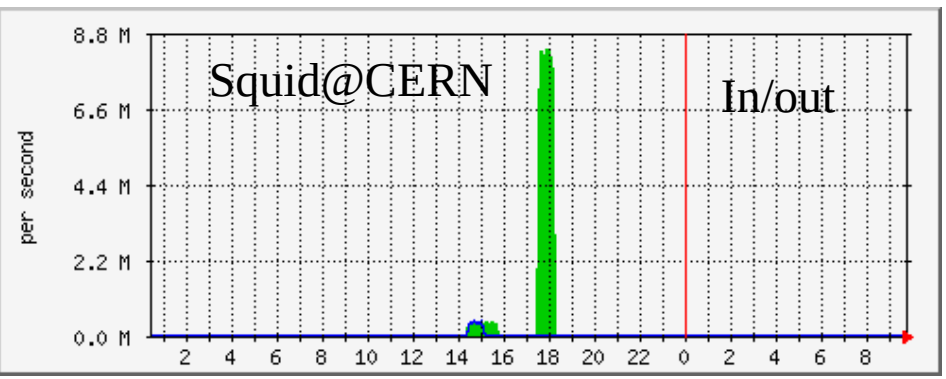


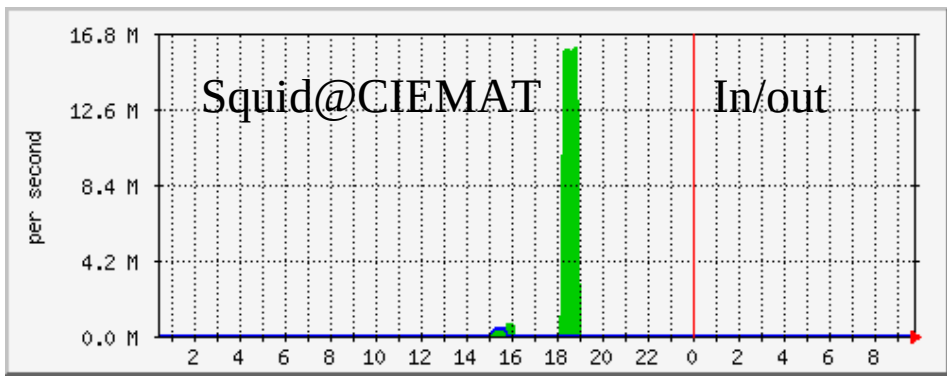
- Four new nodes in place at CERN (3 production, 1 test)
- Will configure in fail-over/load balancing schem similar to CDF experience.
- Main entry squid uses tomcats in round robin fashion

Monitoring Squids W/ SNMP interface (MRTG plots shown)









Test: 20 Parallel CMSSW Clients @ CIEMAT

Conclusion

- Excellent progress has been made over the last year to provide a comprehensive calibration and alignment system for CMS
- Good progress has been made with all sub-detectors to establish the needed infrastructure for the upcoming Magnet Test/Cosmic Challenge.
- Fermilab's responsibility for HCAL and PIXEL online databases is on track and on schedule.
- A detailed plan for deployment is in place and the integration and testing that started last November has demonstrated its feasibility.
- A significant part of this success is due to the contributions made by USCMS and Fermilab in particular.

Finish