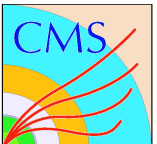


Data Operations, Where is the Data? How do I access it? Is it OK? DBS.

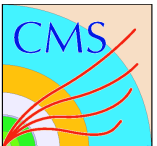
David A. Mason, Fermilab

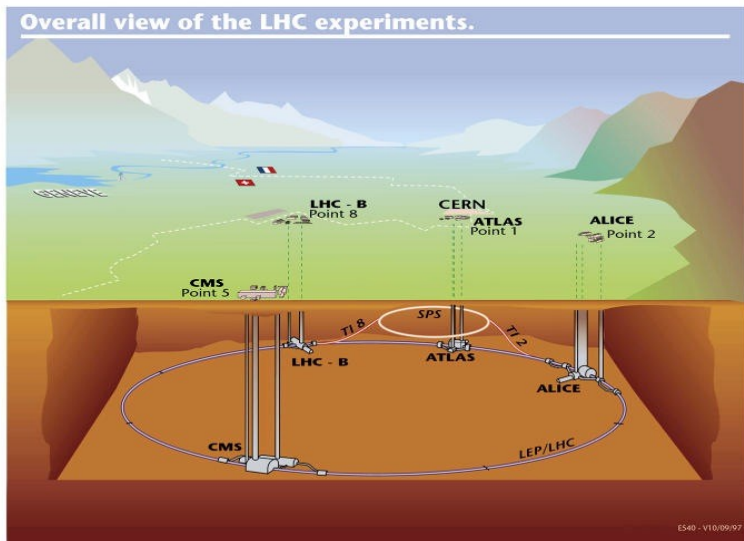


Outline

- Data Operations
- Where is the Data?
- How do I access it?
- Is it OK?
- DBS.

(but probably will talk a bit about DBS earlier)

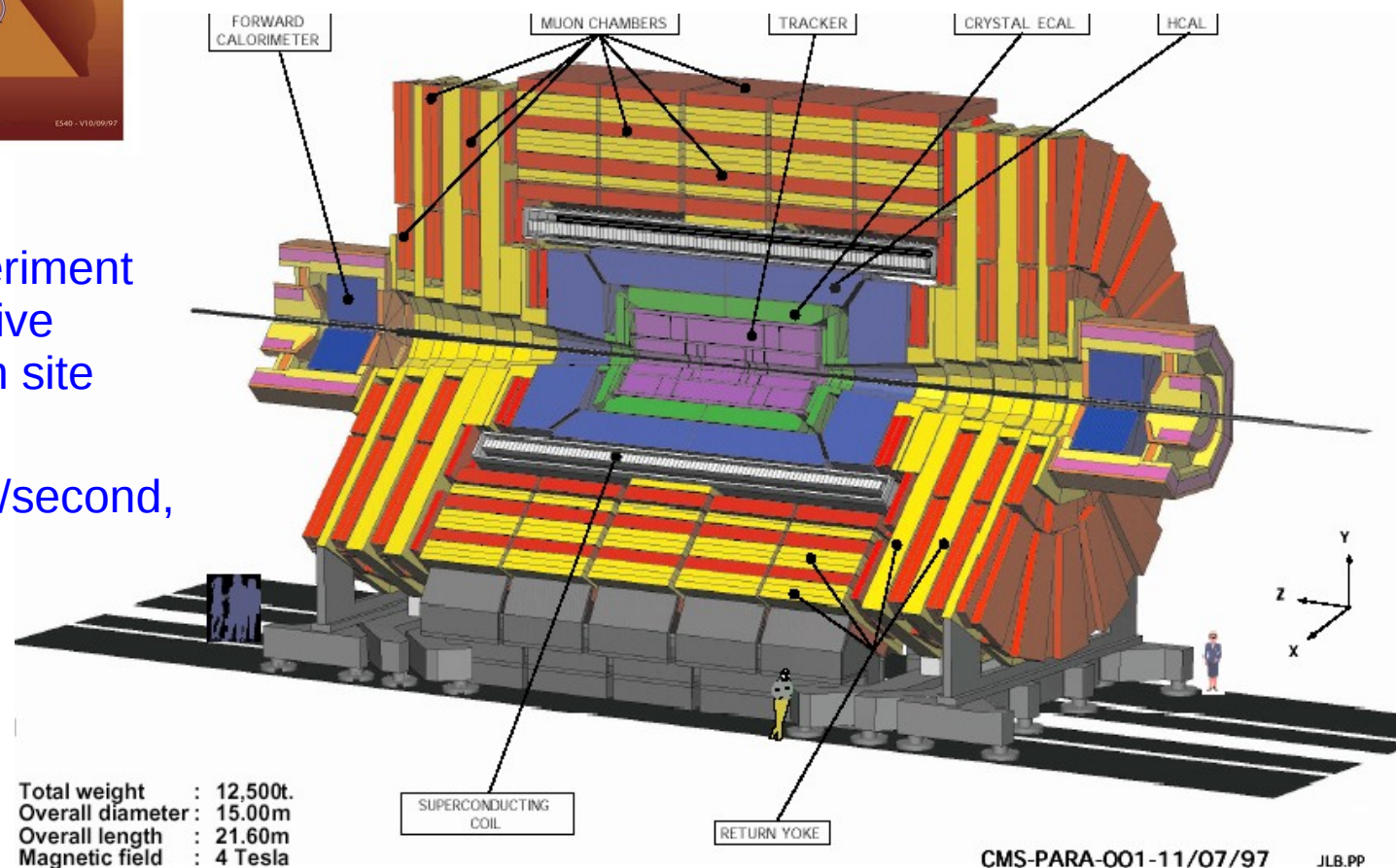




The Data Starts Here

CMS: LHC p-p collider experiment located about a half hour drive from the main CERN Meyrin site

Nominal datarate of 450 MB/second, with bursts of 1 GB/second



CMS-PARA-001-11/07/97 JLB.PP

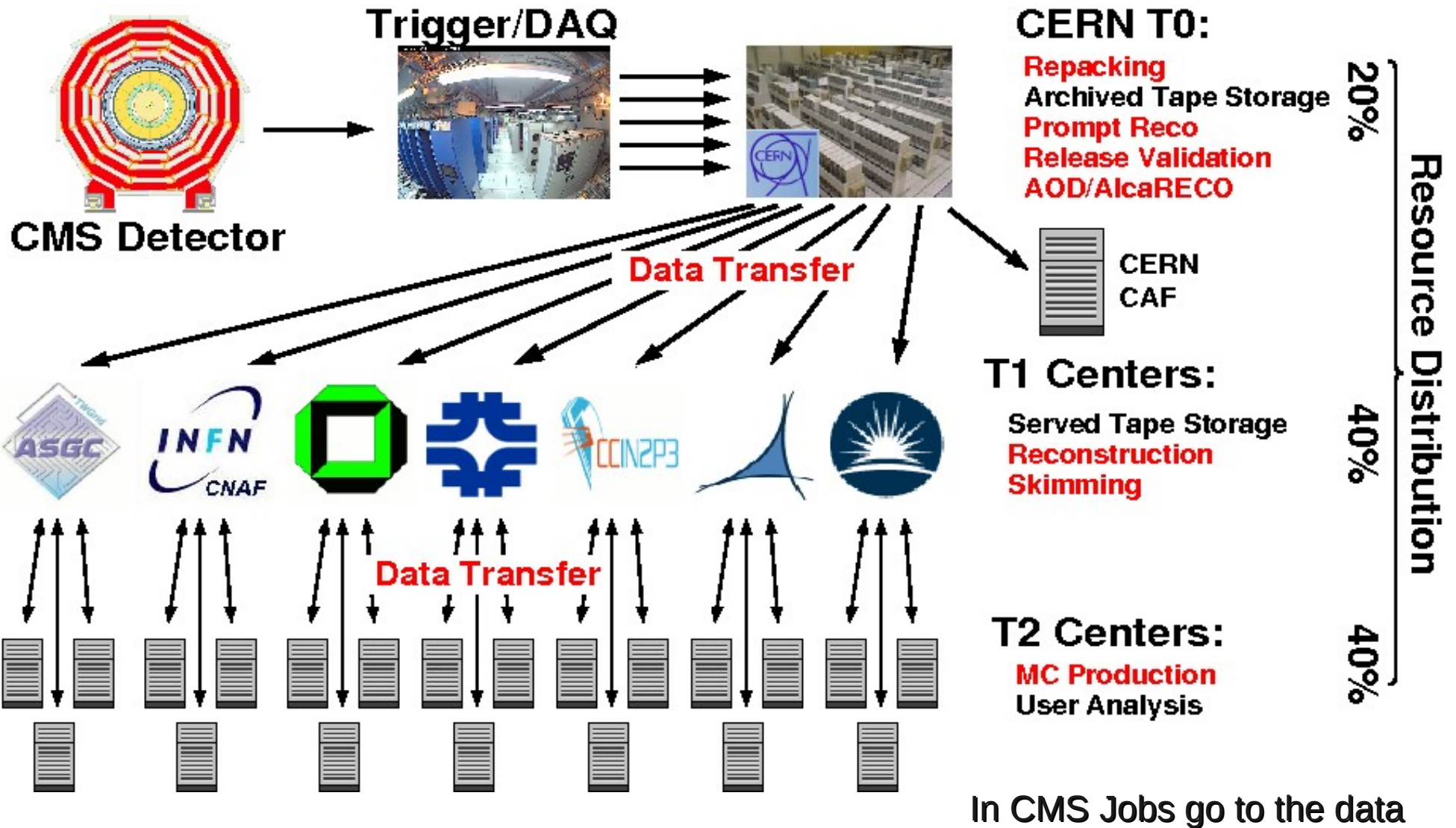


A-Term 2009

Slide 3



CMS Computing Model



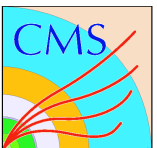
In CMS Jobs go to the data

Tasks in red are performed by central data operations teams



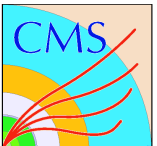
A-Term 2009

Slide 4



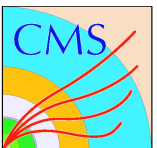
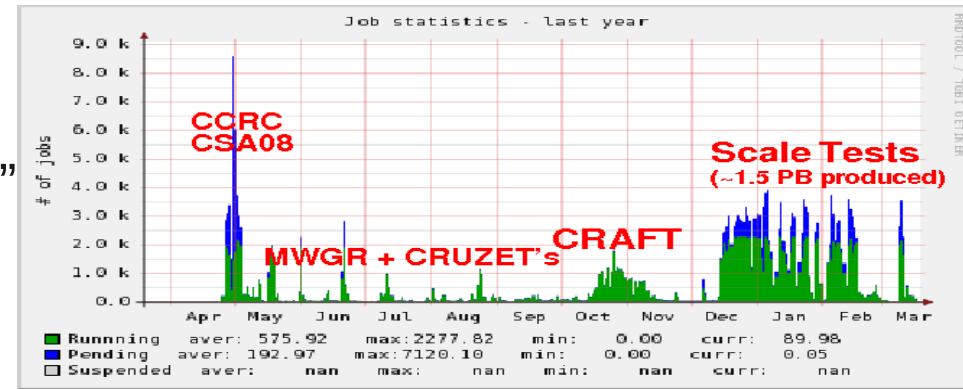
CMS Data Operations

- Everything in **red** in the previous slide is in the hands of CMS Data Operations.
- CMS employs teams of operators around the world to manage data processing, data transfer and Monte Carlo production.
 - Teams take full advantage of grid based tools to manage CMS workflows independent of physical location
 - Multiple timezones around the world increases operational coverage
- **CERN CMS Centre** and **FNAL Remote Operations Center (ROC)** serve as central “home bases” for the data operations.
- From there (as well as wherever we happen to be planted) we've been working to commission CMS computing systems in concert with detector commissioning activities over the past two years.



CMS Commissioning Datataking

- CMS is now in its second round of ramping up to take collision data.
- During these commissioning periods datataking starts off with “midweek global runs” -- weekly two day “tools down” datataking periods where the goal is to test out new things and checkpoint how well (or if) we can still take data.
- After a while we move to ~week long “CRUZET” -- Cosmic Running at Zero Tesla. Goal is more stable, extended length running, with hopefully analysis quality zero field cosmic ray data.
- Eventually when detector commissioning hardware work has been complete, we have moved to more extended periods (month or longer) of Cosmic Running At Four Tesla (CRAFT). We did this once last year in October/November, and now are taking data in CRAFT09.
- Data Ops T0 team has been (necessarily) actively involved in all these commissioning activities, with the T1 teams being involved in reprocessing this data at intervals since.



Before answering “where” you need to know what its called...

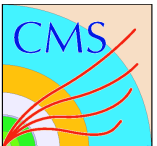
- We have a convention for naming datasets:
`/PrimaryDataset/AcquisitionEra-ProcessingVersion/DataTier`
 - Primary Dataset – defined by a collection of trigger bits.
 - Acquisition Era – defines a period of datataking, e.g. “CRAFT09”
 - Processing Version – usually something like “v1”, “v2”, ... , typically used to separate datasets when the format of data has to change.
 - Data Tier – an indicator of what kinds of objects are in the dataset -- “RAW”, “RECO”, etc.
- Some relevant examples:

`/Cosmics/CRAFT09-v1/RAW`

`/Cosmics/CRAFT09-PromptReco-v1/RECO`

`/Calo/CRAFT09-v1/RAW`

`/Cosmics/Commissioning09-PromptReco-v6/RECO`



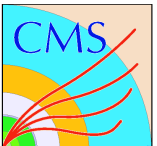
Current 2009 Cosmic datasets

- <https://twiki.cern.ch/twiki/bin/view/CMS/CRAFT09AnalysisInfo>
- 3 primary datasets exist for cosmics data which are being reconstructed: Cosmics, Calo, MinimumBias (MinimumBiasNoCalo)
- Started off with acquisition era of Commissioning09 – ranging from detector in open configuration, few subdetectors available to running all detectors that were able to join (few weeks ago). Generally this data is where things have been tested, broken, tested again, moved around, turned sideways, etc.
 - Broken 3 times for RAW, 6 times for RECO!
- In Spring took 1 week of data with CRUZET09 as the acquisition era
 - Because of the not quite stable nature of CRUZET2 it actually falls in the Commissioning09 acquisition era
- Currently taking data with an acquisition era of CRAFT09
- Can find these guys using the DBS Discovery Page
aha now we learn about DBS:



DBS

- Data Bookkeeping Service – A database should contain every file of official data produced and processed, including:
 - Dataset it belongs to
 - CMSSW version that made it
 - Configuration that made it
 - When files were entered into database
 - From what dataset/file(s) processed data came from
 - How many files are in a dataset and how big they are
- DBS does not tell you where the files are!
- This is the domain of PhEDEx, CMS's transfer infrastructure.
- The DBS Discovery Page pulls in information from DBS and PhEDEx to tell you where data is
- http://cmsweb.cern.ch/dbs_discovery – some examples.



Commissioning09 data:

DBS data discovery page - Mozilla Firefox

File Edit View History Bookmarks Tools Help

https://cmsweb.cern.ch/dbs_discovery/aSearch?caseSensitive=on&userMode=user&sortOrder=d

Members PhEEx WebMail Opera Software - ... http://zebu.uoreg... Connections BizJournal SmartUpdate Mktplace Shopping Cart

Dashboard DBS Discovery DataTransfer SiteDB CondDB Support Login

Home - aSearch - Navigator - RSS - Status - Runs - Admin - Tools - Help - Contact View

DBS instances cms_dbs_prod_global **HELP**

find dataset where dataset like /Cosmics/Commissioning09-v*/RAW and dataset.status=VALID Search Reset

DBS discovery :: Adv. search :: Results Physicist

Found 3 results. Show [all](#) View results: [grid](#) | [list](#) mode Sort by DATASET [desc](#) | [asc](#)

DATASET	CREATED	CREATOR	SIZE	BLOCKS	FILES	EVENTS	SITES	CRAB	JL	LINKS
/Cosmics/Commissioning09-v1/RAW	04/03/09	vocms39.cern.ch	7.4TB	365	cff , py , txt , (3577)	276908001	7	cfg	N/A	More info v
/Cosmics/Commissioning09-v2/RAW	17/06/09	vocms39.cern.ch	3.8TB	129	cff , py , txt , (1326)	85279498	4	cfg	N/A	More info v
/Cosmics/Commissioning09-v3/RAW	15/07/09	vocms39.cern.ch	1.9TB	61	cff , py , txt , (774)	20044174	5	cfg	N/A	More info v

Number of results per page [10](#) Result page: [1](#)



Where is the data:

DBS data discovery page - Mozilla Firefox

File Edit View History Bookmarks Tools Help

https://cmsweb.cern.ch/db_s_discovery/aSearch?caseSensitive=on&userMode=user&sortOrder=d

Members PhEDEx WebMail Opera Software - ... http://zebu.uoreg... Connections BizJournal SmartUpdate Mktplace Shopping Cart

Dashboard DBS Discovery DataTransfer SiteDB CondDB Support Login

Home - aSearch - Navigator - RSS - Status - Runs - Admin - Tools - Help - Contact View

DBS instances cms_dbs_prod_global HELP

find dataset where dataset like /Cosmics/Commissioning09-v*/RAW and dataset.status=VALID Search Reset

DBS discovery :: Adv. search :: Results Physicist

Found 3 results. Show [all](#) View results: [grid](#) | [list](#) mode Sort by DATASET desc | asc

DATASET	CREATED	CREATOR	SIZE	BLOCKS	FILES	EVENTS	SITES	CRAB	fL	LINKS
/Cosmics/Commissioning09-v1/RAW	04/03/09	vocms39.cern.ch	7.4TB	365	cff , py , txt , (3577)	276908001	7	cfg	N/A	More info
/Cosmics/Commissioning09-v2/RAW	17/06/09	vocms39.cern.ch	3.8TB	129	cff , py , txt , (1326)	852794				
/Cosmics/Commissioning09-v3/RAW	15/07/09	vocms39.cern.ch	1.9TB	61	cff , py , txt , (774)	20044				

Number of results per page 10

NO NAME : castorsrm.cr.cnaf.infn.it
T2_US_Caltech : cit-se.ultraviolet.org
T1_US_FNAL : cmssrm.fnal.gov
T1_CH_CERN : srm-cms.cern.ch
T2_CH_CAF : caf.cern.ch
NO NAME : bonner-cms3.rice.edu
T2_PL_Warsaw : se.polgrid.pl



Commissioning09 RECO data:

DBS instances [HELP](#)

find dataset where dataset like /Cosmics/Commissioning09-PromptReco-v*/RECO and dataset.status=

DBS discovery :: Adv. search :: Results Physicist

Found 6 results. Show [all](#) View results: [grid](#) | [list](#) mode Sort by DATASET | [asc](#)

DATASET	CREATED	CREATOR	SIZE	BLOCKS	FILES	EVENTS	SITES	CRAB	JL	LINKS
/Cosmics/Commissioning09-PromptReco-v1/RECO	04/03/09	vocms39.cern.ch	11.3TB	350	cfl , py , txt , (3144)	255830953	4	cfl	N/A	<input type="button" value="More info"/>
/Cosmics/Commissioning09-PromptReco-v2/RECO	10/06/09	vocms39.cern.ch	990.0GB	42	cfl , py , txt , (275)	14493257	4	cfl	N/A	<input type="button" value="More info"/>
/Cosmics/Commissioning09-PromptReco-v3/RECO	17/06/09	vocms39.cern.ch	3.6TB	107	cfl , py , txt , (916)	57206264	4	cfl	N/A	<input type="button" value="More info"/>
/Cosmics/Commissioning09-PromptReco-v4/RECO	02/07/09	vocms39.cern.ch	2.8TB	38	cfl , py , txt , (394)	27324103	4	cfl	N/A	<input type="button" value="More info"/>
/Cosmics/Commissioning09-PromptReco-v5/RECO	15/07/09	vocms39.cern.ch	341.7GB	21	cfl , py , txt , (157)	4017675	4	cfl	N/A	<input type="button" value="More info"/>
/Cosmics/Commissioning09-PromptReco-v6/RECO	23/07/09	vocms39.cern.ch	803.3GB	27	cfl , py , txt , (325)	10641212	4	cfl	N/A	<input type="button" value="More info"/>

Number of results per page Result page:



CRAFT09 RECO data:

DBS instances [HELP](#)

find dataset where dataset like /*/CRAFT09-* and (tier=RAW or tier = RECO) and dataset.status=VALID

DBS discovery :: Adv. search :: Results

Physicist

Found 13 results. Show [all](#)

View results: [grid](#) | [list](#) mode

Sort by DATASET | [asc](#)

DATASET	CREATED	CREATOR	SIZE	BLOCKS	FILES	EVENTS	SITES	CRAB	JL	LINKS
/AICaP0/CRAFT09-v1/RAW	25/07/09	vocms39.cern.ch	7.3GB	70	cff , py , txt , (127)	6669990	2	cfg	N/A	More info <input type="button" value="v"/>
/AICaPhiSymEcal/CRAFT09-v1/RAW	24/07/09	vocms39.cern.ch	1.8TB	78	cff , py , txt , (857)	56294880	2	cfg	N/A	More info <input type="button" value="v"/>
/AICaPhiSymHcal/CRAFT09-v1/RAW	25/07/09	vocms39.cern.ch	1.1GB	72	cff , py , txt , (72)	13878	2	cfg	N/A	More info <input type="button" value="v"/>
/Calo/CRAFT09-PromptReco-v1/RECO	25/07/09	vocms39.cern.ch	12.2TB	206	cff , py , txt , (5033)	167095099	4	cfg	N/A	More info <input type="button" value="v"/>
/Calo/CRAFT09-v1/RAW	24/07/09	vocms39.cern.ch	20.8TB	293	cff , py , txt , (5427)	172353534	4	cfg	N/A	More info <input type="button" value="v"/>
/Cosmics/CRAFT09-PromptReco-v1/RECO	25/07/09	vocms39.cern.ch	14.2TB	223	cff , py , txt , (5252)	159173600	6	cfg	N/A	More info <input type="button" value="v"/>
/Cosmics/CRAFT09-v1/RAW	24/07/09	vocms39.cern.ch	19.8TB	283	cff , py , txt , (5609)	165899321	5	cfg	N/A	More info <input type="button" value="v"/>
/HcalHPDNoise/CRAFT09-v1/RAW	24/07/09	vocms39.cern.ch	10.9MB	97	cff , py , txt , (97)	0	3	cfg	N/A	More info <input type="button" value="v"/>
/MinimumBiasNoCalo/CRAFT09-PromptReco-v1/RECO	25/07/09	vocms39.cern.ch	266.6MB	93	cff , py , txt , (93)	0	4	cfg	N/A	More info <input type="button" value="v"/>
/MinimumBiasNoCalo/CRAFT09-v1/RAW	24/07/09	vocms39.cern.ch	10.9MB	97	cff , py , txt , (97)	0	4	cfg	N/A	More info <input type="button" value="v"/>
/RPCMonitor/CRAFT09-v1/RAW	24/07/09	vocms39.cern.ch	347.0GB	92	cff , py , txt , (1216)	106740485	2	cfg	N/A	More info <input type="button" value="v"/>
/RandomTriggers/CRAFT09-v1/RAW	24/07/09	vocms39.cern.ch	54.6GB	97	cff , py , txt , (108)	737967	2	cfg	N/A	More info <input type="button" value="v"/>
/TestEnables/CRAFT09-v1/RAW	24/07/09	vocms39.cern.ch	5.6TB	131	cff , py , txt , (2263)	53060460	2	cfg	N/A	More info <input type="button" value="v"/>



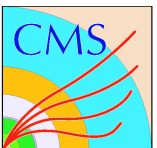
A-Term 2009

Slide 13



OK – How do I access this stuff?

- CRAB -- CMS Remote Analysis Builder
 - To first order, creates and submits CMSSW jobs to where the dataset you're interested in lives in.
 - You need to know the dataset name you're going to run on, and should make sure its living at a Tier 2/3 center to run the jobs at. (i.e. LPC CAF)
 - You can request data be moved to your favorite T2 (subject to that admin's approval) using PhEDEx
 - <https://cmsweb.cern.ch/phedex/prod/Request::Create>
 - Grid cert required and you need to enter your DN name (printed out when you get a grid proxy)
 - <https://cmsweb.cern.ch/sitedb>
 - More info on what CRAB can do in Eric Vaandering's talk tomorrow



Is the Data OK?

ADVANCED KEYWORD SEARCH

DBS instances

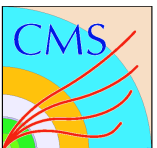
```
Exception(error)
Exception: _____ API Invoked countQuery_____ Unexpected execution exceptionDBS returns:
code   = 4000,
msg    = _____ API Invoked countQuery_____ Unexpected execution exception,
detail = Invalid Token everything on line 1 at column 43
QUERY  find dataset where dataset like I want and everything = wonderful and dataset.status=VALID
POSITION                                     ^
```

- To be honest, this is your job!
- There are tools however to help this along.
- Run summary page tells you what detectors were active
<https://cmswbm.web.cern.ch/cmswbm/cmsdb/servlet/RunSummary>
- DBS currently has quality flags:
 - Query looks like: find run,sum(file.numevents) where dataset = /X/Y/Z and dq=JetMET=GOOD
- Data quality will actually be moved out of DBS into its own database
 - Though hopefully you won't need to know that because a Data Aggregation Service (DAS) will sit in front to put the two together for you



A-Term 2009

Slide 15



Summary

- Data Ops manages all central data processing and production for CMS
- CMS has been taking cosmic data with varying levels of stability in 2008 and '09.
- Commissioning09 = less stable, CRAFT09 is more stable
- DBS discovery page shows how big datasets are and accesses PhEDEx's database to show you where it is
- Once the data is at a Tier 2 you can run your analysis on it using CRAB
- Although there is readily available info indicating what detectors are in a run, whether a physics group says it OK, “Is it OK” is ultimately for you to answer!
- Go start looking at real data and learn about your detector!
Data is actually being taken now!

