

CMS Trivial File Catalog and Site Local Configuration

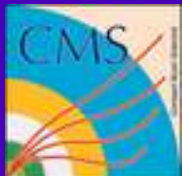
Peter Elmer
Princeton University
WLCG Asian Tier-2 Workshop
03 Dec, 2006



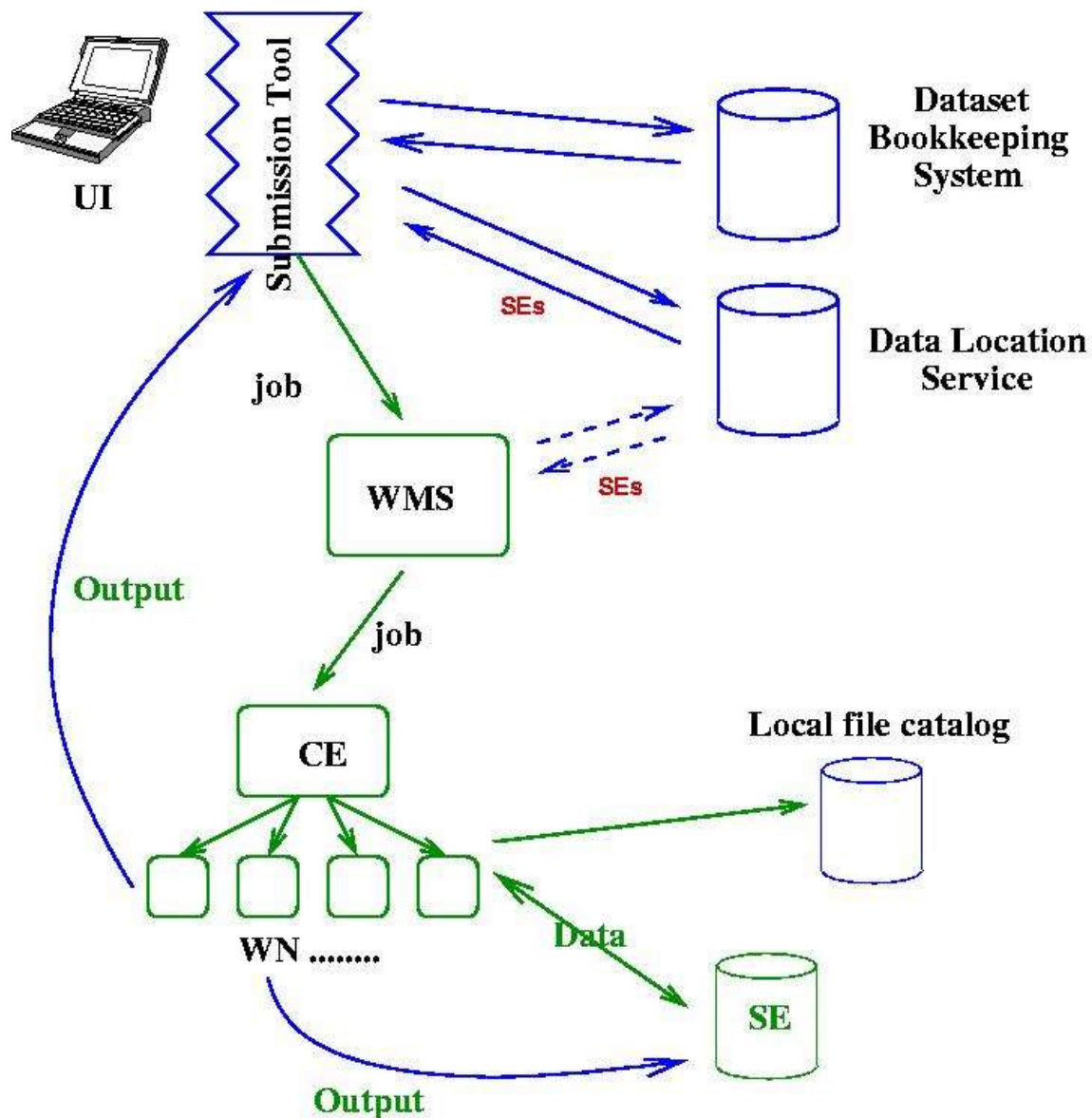
DM Overview



- The CMS DM system consists of 4 CMS components:
 - Dataset Bookkeeping System (DBS)
 - Global instance is installed at CERN
 - Data Location Service (DLS)
 - Global instance is installed at CERN
 - Phedex (previous presentation)
 - **Must be installed for site**
 - Trivial File Catalog (TFC) and Site Local Configuration
 - **Must be installed at site**



DM and Jobs




- User on UI:
 - Dataset bookkeeping system
- Either User on UI or RB:
 - Data location service
- Job on worker node:
 - Data access/storage
 - Local file catalog (either “trivial” or standard)
- Output management – no connections from WN outwards, apart from via output sandbox or asynchronous management by Phedex



Data Discovery Page

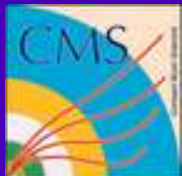


<http://cmsdbs.cern.ch/discovery/>

 DBS/DLS DATA DISCOVERY PAGE Home page: [users](#) [experts](#)

<p>Navigator</p> <p>Keyword search</p> <p>History</p> <p>Help</p> <p>Hide panel</p>	<p>Navigator menu</p> <p>Tier sites: All</p> <p>Application: /CMSSW_1_1_1/Skimming/cmsRun</p> <p>Primary dataset: mc-physval-111-UDSUDSbar</p> <p>Data tier: All</p> <p><input type="button" value="Find"/></p> <p>DBS glossary</p>
--	--

This is the user data discovery tool (webpages), but it illustrates also what PRODAGENT/CRAB are doing in a more automated way. (Lee described more details about its use for data discovery.)



Datasets, blocks and LFN's



/CSA06-084-os-HLTSoup/DIGI/CMSSW_0_8_4-GEN-SIM-DIGI-HLT-1157592494-merged

contains 5167619 events, 3185 files, 6.0TB.

Show: **Blocks** Summary Both

row	Location ↓	Events	Files	status	size	LFN list	Block name
02_1	castor.sc.grid.sinica.edu.tw	399583	241	OPEN	471.6GB	cff, plain	/CSA06-084-os-HLTSoup/CMSSW_0_8_4-GEN-SIM-DIGI-HLT-1157592494-merged#7238a753-a3d9-4dfd-9bf9-bf6b94c58382
13_1	cithep59.ultralight.org	4782	3	OPEN	5.8GB	cff, plain	/CSA06-084-os-HLTSoup/CMSSW_0_8_4-GEN-SIM-DIGI-HLT-1157592494-merged#f76e1f36-a9e9-485f-a86c-99261051f6f0
05_1	cmssrm.fnal.gov	114256	71	OPEN	135.9GB	cff, plain	/CSA06-084-os-HLTSoup/CMSSW_0_8_4-GEN-SIM-DIGI-HLT-1157592494-merged#162ec163-93f4-463f-9058-a40238c7ccc6
05_2	cmssrm.fnal.gov	859368	531	OPEN	1019.9GB	cff, plain	/CSA06-084-os-HLTSoup/CMSSW_0_8_4-GEN-SIM-DIGI-HLT-1157592494-merged#5e5a810e-2d7d-4b64-8689-eed28cba8588
14_1	dcache.rcac.purdue.edu	138779	86	OPEN	164.1GB	cff, plain	/CSA06-084-os-HLTSoup/CMSSW_0_8_4-GEN-SIM-DIGI-HLT-1157592494-merged#18485b8a-0d6f-4c43-9f9f-8cc823b3a6c3

An example LFN:

/store/CSA06/2006/7/27/CSA06-081-os-minbias/00AF79E0-5627-DB11-B5F8-00304875AA19.root

DBS/DLS just provides the SE name where the data is located (so we know where to submit the job) and an LFN. But once the job starts, it needs to know how to open the file (rfio, dcap, pnfs, root, etc. and other such “physical” details, all specific to the site, i.e. the PFN. This is something like:

rfio:/castor/cern.ch/cms/store/CSA06/2006/7/...../00AF79E0-5627-DB11-B5F8-00304875AA19.root



Trivial File Catalog

- Rather than have a true catalog to map “logical file names” (LFN's) to “physical filenames” (PFN's), e.g. implemented in:
 - LFC
 - MySQL
 - xml catalogs (as in the old CMS EDM)

we have chosen something simpler, the so called trivial file catalog. This is basically just a set of algorithmic rules for mapping LFN's to PFN's (as strings), in practice by prepending something. This is much simpler than maintaining a full, catalog at sites.

Phedex also uses the trivial file catalog for transfers.



Other site local configuration



- I've talked about accessing files, but there are other potential site local configurations for running CMSSW. The additional example we have at the moment is how to access the local frontier squid proxy.
- The model for this is effectively the same: the job arrives on the worker node with no specific information for that site, and must find the information at the site itself.



How to make a TFC and site local config



Rather than duplicate information to this presentation, I'll now switch to the Twiki page where this is documented:

<https://twiki.cern.ch/twiki/bin/view/CMS/SWIntTrivial>