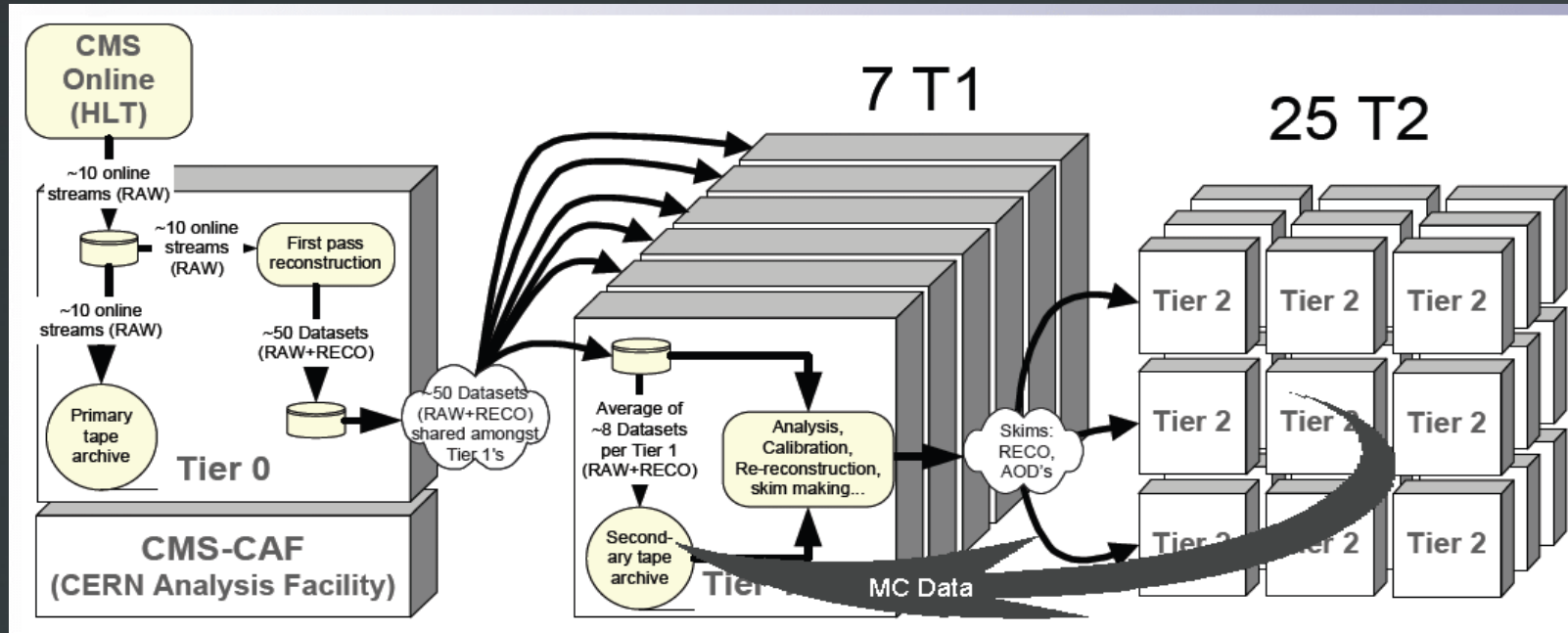


# Data distribution and LHC experiments

Giacinto DONVITO  
INFN-BARI

# CMS Tiers



## ■ Tier0:

- RAW Data Archive
- First Pass Reconstruction => "small" number of pre-defined datasets
- Send RAW+RECO to Tier1s

## ■ Tier1:

- Secondary Archive for RAW+RECO Data
- Calibration, Re-reconstruction, Skim Making
- Transfer RECO-AOD to Tier2
- MC Data Archive

## ■ Tier2:

- Chaotic Analysis
- MC Scheduled production
- Stage output from users analysis

# Datasets naming conventions

---

/ <physics process> / <main period> \_ <data taking conditions> \_  
<conditions data version> \_ <processing version> / <data format  
content> :

- QCDDiJetPt0to15, RelValZEE, TTbar-madgraph, ...
- CMSSW version, or Summer08, Fall08, Winter09, ... ⇒ not always trivial to relate to CMSSW version ☹
- IDEAL, STARTUP
- V9, V11, ... see <https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideFrontierConditions>
- v1, v2, ...
- GEN-SIM-RECO, GEN-SIM-RAW, AODSIM, ...

# CMS Data Tiers

- CMS Data is arranged into a hierarchy of data tiers. Each physics event is written into each data tier, where the tiers each contain different levels of information about the event. The different tiers each have different uses. The three main data tiers written in CMS are:
  1. RAW: full event information from the Tier-0 (i.e. from CERN), containing 'raw' detector information (detector element hits, etc) RAW is not used directly for analysis
  2. RECO ("RECOConstructed data"): the output from first-pass processing by the Tier-0. This layer contains reconstructed physics objects, but it still very detailed RECO can be used for analysis, but is too big for frequent or heavy use when CMS has collected a substantial data sample.
  3. AOD ("Analysis Object Data"): this is a "distilled" version of the RECO event information, and is expected to be used for most analyses AOD provides a trade-off between event size and complexity of the available information to optimize flexibility and speed for analyses

# CMS Dataset Bookkeeping System

A new data discovery service, [DAS](#), is released as beta-version, testers are needed.

DBS instances  [HELP](#)

## DBS discovery :: Adv. search :: Results

Physicist

Found 1 results. Show [all](#)

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

Sort by  | [asc](#)

### /Mu/Run2010B-Dec22ReReco\_v1/AOD

Created 22 Dec 2010 05:00:26 GMT, contains 33299679 events, 1540 files, 6 block(s), 3.6TB, located at 12 sites ([show](#), [hide](#)), LFNs: [cfl](#), [py](#), [plain](#), [L=N/A]  
[Release info](#), [Block info](#), [Run info](#), [Conf. files](#), [Parents](#), [Children](#), [Description](#), [PhEDEx](#), [Create ADS](#), [ADS](#), [crab.cfg](#)

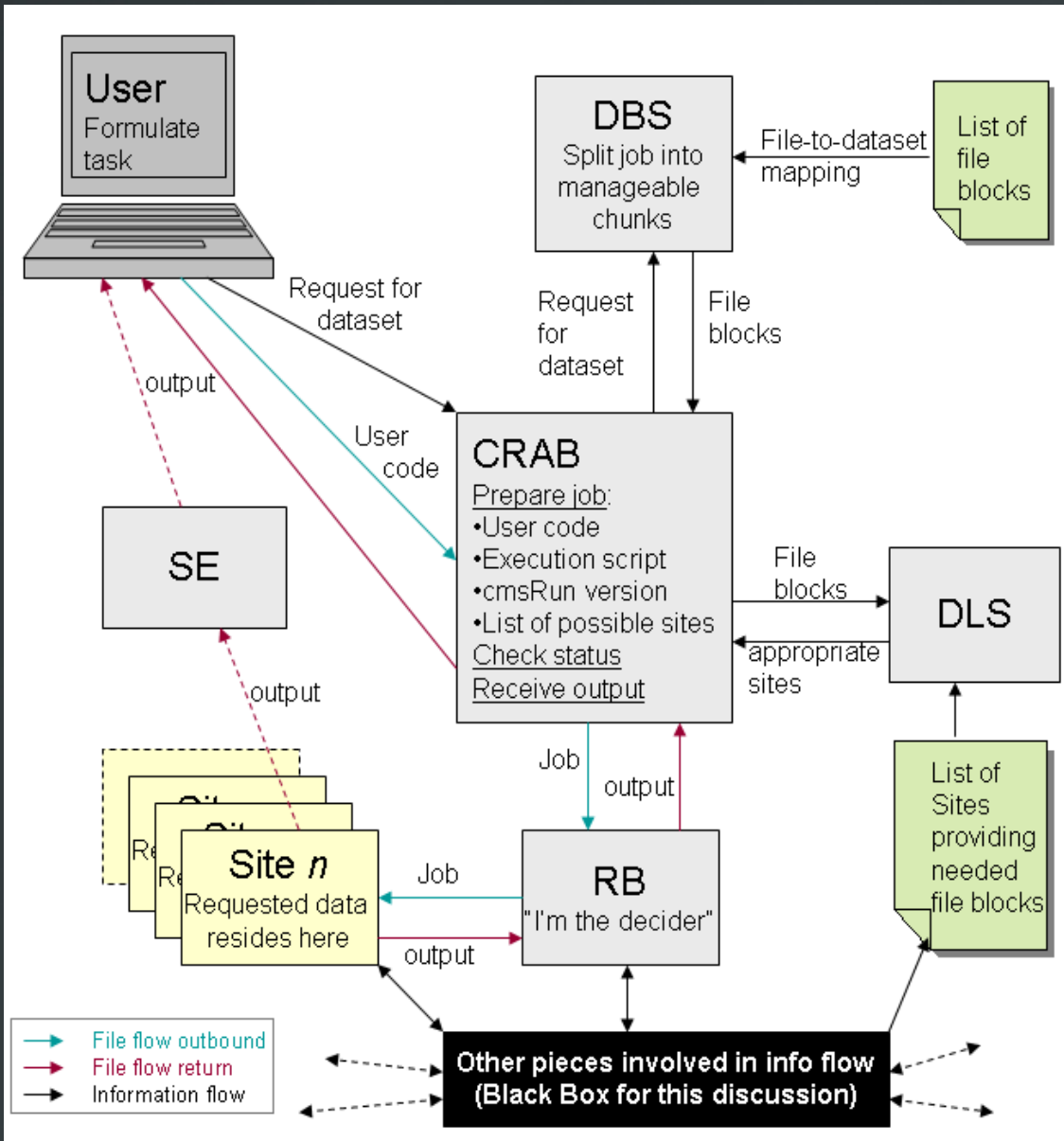
Location	Events	Files	size	LFNs
T2_ES_CIEMAT : srm.ciemat.es	33299679	1540	3.6TB	<a href="#">cfl</a> <a href="#">plain</a>
T2_ES_IFCA : storm.ifca.es	33299679	1540	3.6TB	<a href="#">cfl</a> <a href="#">plain</a>
T1_US_FNAL : cmssrm.fnal.gov	33299679	1540	3.6TB	<a href="#">cfl</a> <a href="#">plain</a>
T2_US_Wisconsin : cmssrm.hep.wisc.edu	33299679	1540	3.6TB	<a href="#">cfl</a> <a href="#">plain</a>
T1_IT_CNAF : storm-fe-cms.cr.cnaf.infn.it	33299679	1540	3.6TB	<a href="#">cfl</a> <a href="#">plain</a>
T3_FR_IPNL : lyogrid06.in2p3.fr	21348192	1040	2.3TB	<a href="#">cfl</a> <a href="#">plain</a>
T2_US_Purdue : srm-dcache.rcac.purdue.edu	33299679	1540	3.6TB	<a href="#">cfl</a> <a href="#">plain</a>
T2_BE_UCL : ingrid-se02.cism.ucl.ac.be	33299679	1540	3.6TB	<a href="#">cfl</a> <a href="#">plain</a>
T2_CN_Beijing : srm.ihep.ac.cn	33299679	1540	3.6TB	<a href="#">cfl</a> <a href="#">plain</a>
T2_IT_Legnaro : t2-srm-02.lnl.infn.it	33299679	1540	3.6TB	<a href="#">cfl</a> <a href="#">plain</a>
T2_DE_DESY : dcache-se-cms.desy.de	33299679	1540	3.6TB	<a href="#">cfl</a> <a href="#">plain</a>
T2_EE_Estonia : ganymede.hep.kbfi.ee	33299679	1540	3.6TB	<a href="#">cfl</a> <a href="#">plain</a>

Number of results per page

Result page:

/store/data/Run2010B/Mu/AOD/Dec22ReReco\_v1/0047/8AD8D267-F012-E011-95B9-E0CB4E1A114B.root  
/store/data/Run2010B/Mu/AOD/Dec22ReReco\_v1/0047/56F16FF0-CB13-E011-BD29-E0CB4E5536BB.root  
/store/data/Run2010B/Mu/AOD/Dec22ReReco\_v1/0031/DAAD8469-2D11-E011-9BCE-E0CB4E4408DE.root  
/store/data/Run2010B/Mu/AOD/Dec22ReReco\_v1/0030/E0482B52-0711-E011-8347-90E6BA19A212.root  
/store/data/Run2010B/Mu/AOD/Dec22ReReco\_v1/0029/E818A394-7A10-E011-AAA6-E0CB4E29C4F9.root

# How to submit an analysis job to the Grid – CMS



- CRAB takes care of interact with the grid and experiments services hiding all the details to the final user



# CMS Dataset Placement System

## PhEDEx transfer page

PhEDEx – CMS Data Transfers

DB Instance: Production  
Anne-Marie Magnan | [Sign out](#)  
Logged in via Certificate

[Overview](#) | [Create Request](#) | [View/Manage Requests](#)

### New Transfer Request

E-mail:

DBS:

Data Items:

/Primary/Processed/Tier  
or  
/Primary/Processed  
/Tier:Block  
(Use \* as wildcard)  
[More Help](#)

Destinations:

<input checked="" type="checkbox"/>	T2_UK_London_IC	<input type="checkbox"/>	TX_LOGBO_Buffer
<input type="checkbox"/>	T2_UK_London_RHUL	<input type="checkbox"/>	TX_TESTa_Buffer

Transfer Type:  [What's this?](#)

Subscription Type:  [What's this?](#)

Priority:  [What's this?](#)

Custodian:  [What's this?](#)

Group:  [What's this?](#)

Comment:

Select T2\_UK\_London\_IC  
Or  
T3\_UK\_London\_RHUL  
(ask Monica V.-A for more help as she manages data in the UK)