



# Intro to CMS Computing Model & Some of its components

Frank Würthwein
UCSD

# High Level Requirements for user analysis computing

- Code Development Environment
  - Compile, run, debug with fast turn around
  - Very agile & reasonably interactive
  - Complete data access at modest IO
- Large scale processing environment
  - Large scale parallelization
  - Large CPU & IO
  - Perfect bookkeeping that's trivial to use
  - Latencies commensurate with resource consumption

# How large is “large scale”

- ~200MBytes/sec = HLT output rate
  - $10^7$secs per year => ~2 Petabytes/year
  - Add MC and multiple releases in year 1.

***~10 Petabyte/year of CMS primary data***

| Name | Description | Value | Units |
|---|---|---|---|
| L2Rate | pp Rate to Tape | 150 | Hz |
| HIRate | Weighted mean HI event Rate | 50 | Hz |
| LHCYear | Days of pp Running/year | 10000000 | sec |
| HIYear | Seconds of HI Running/year | 1.E+06 | sec |
| NRawEvts | Number of pp Raw Events/year | 1.5E+09 | (derived) |
| NHIEvts | Number of HI Events/year | 5.0E+07 | (derived) |
| | | | |
| RawSize | Raw Data Event Size | 1.5 | MB |
| SimSize | Simulated Event Size | 2 | MB |
| RecSimSize | Reconstructed Sim Event Size | 0.4 | MB |
| RECOSize | Reco Size | 0.25 | MB |
| AODSize | AOD Size | 0.05 | MB |
| TAGSize | Tag and DPD Size | 0.01 | MB |
| HIRawSize | Weighted Mean Heavy Ion Raw Event | 7 | MB |
| HIRecoSize | Weighted Mean Heavy Ion Reco Size | 1 | MB |
| HIAODSize | Weighted Mean Heavy Ion AOD Size | 0.2 | MB |
| | | | |
| NSimEvt | Number of Simulated Events | 1.5.E+09 | Evts/Year |
| FracSimT1 | Fraction of NSimEvts done at T1 | 0% | |
| NSimPrivate | Number of Private Sim at T2s | 8.E+08 | Evts/Year |

# Computing Model in a Nutshell

- Tier-0 @ CERN
  - First pass processing & archival storage
- 7x Tier-1 Centers
  - Archival storage of 1/7th data each
  - All reprocessing of the data
  - All skimming of the data
- “50”x Tier-2 Centers
  - All MC production
  - All user analysis
    - Except CERN CAF, maybe -> probably mostly calib et al.
    - Except LPC CAF, maybe for US only (?)

# CMS Global Data Grid

- 2500 physicists, 40 countries
- 10s of Petabytes/yr by 2008
- 1000 Petabytes in < 10 yrs?

# Computing Model by numbers

T0

| | | |
|---|---|---|
| CPU scheduled | 4588 | kSI2K |
| Disk | 380 | Tbytes |
| Active tape | 3775 | Tbytes |
| Tape I/O | 400 | MB/s |

*2.3kSi2k/core for 2.33GHz Xeon in our new 8ways.*

Each T1

| | | |
|---|---|---|
| CPU scheduled | 1399 | kSI2K |
| CPU analysis | 1101 | kSI2K |
| Disk | 1304 | Tbytes |
| Active tape | 2090 | Tbytes |
| Data Serving I/O Rate | 800 | MB/s |

Each T2

| | | |
|---|---|---|
| CPU scheduled | 257 | kSI2K |
| CPU analysis | 579 | kSI2K |
| Disk | 418 | Tbytes |

***Note: Actual T1's and T2's are very different from this!***
***E.g.: US T2 = 1000 kSi2k plus 200TB useable disk by 2008.***

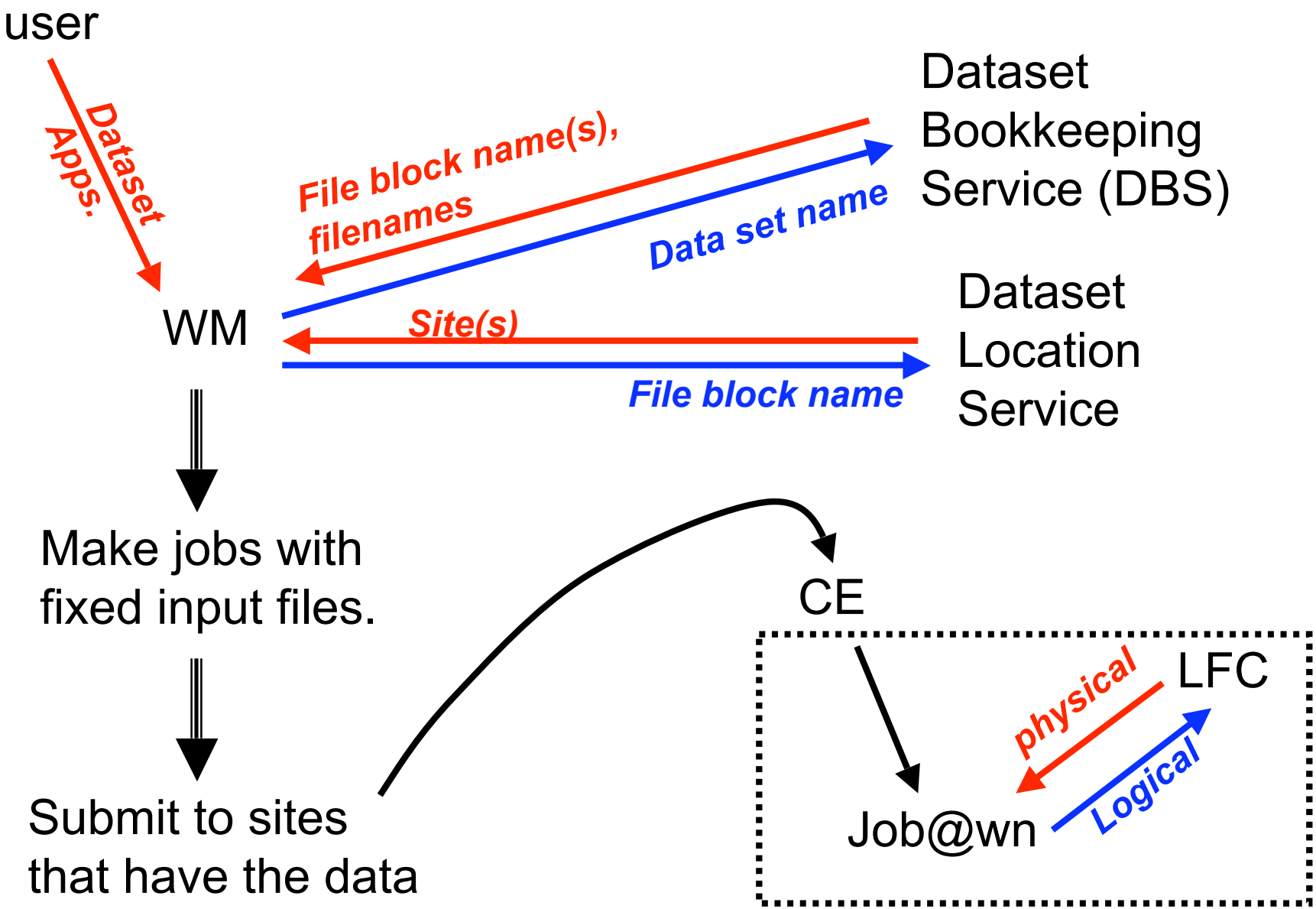
# Pledges for 2008 as presented at CRB today

2008

| | Plg. CPU | Plg. Disk | Plg. Tape |
|---|---|---|---|
| | kSI2k | TB | TB |
| | | | |
| FZK | 1200 | 650 | 900 |
| IN2P3 | 1490 | 780 | 1180 |
| PIC | 760 | 350 | 835 |
| CNAF | 1925 | 875 | 735 |
| ASGC | 1530 | 675 | 585 |
| RAL | 1330 | 620 | 1280 |
| FNAL | 4256 | 1986 | 4700 |
| CERN | | | |
| | | | |
| Total | 12491 | 5936 | 10215 |

Comp.Model: 17500 9000 14000

# CMS Baseline Services



# Some features of CMS Baseline Services

- Blocks of files are statically placed at sites.
  - Each T2 has a data manager:
    - Decides what data to move to and delete at T2.
    - Still to be negotiated how much of space is controlled centrally!
- Location service is at “block of files” level:
  - 1 block = O(1%) of Tier-2 disk space
  - 1 block = O(1‰) of CMS data volume (AOD)
- Input files for job fixed at submission time.
- Logical to physical translation is strictly local.

# Explaining some buzzwords

- PhEDEx = system that moves data to sites.
- DBS = dataset bookkeeping service
- DLS = data location service
- LFC/TFC = namespace mapping at site
- CRAB = job submission system
- OSG CE = generalized api for site batch system.
- dCache = site storage system
- FroNTier = mechanism to get calibrations etc.
- What else?