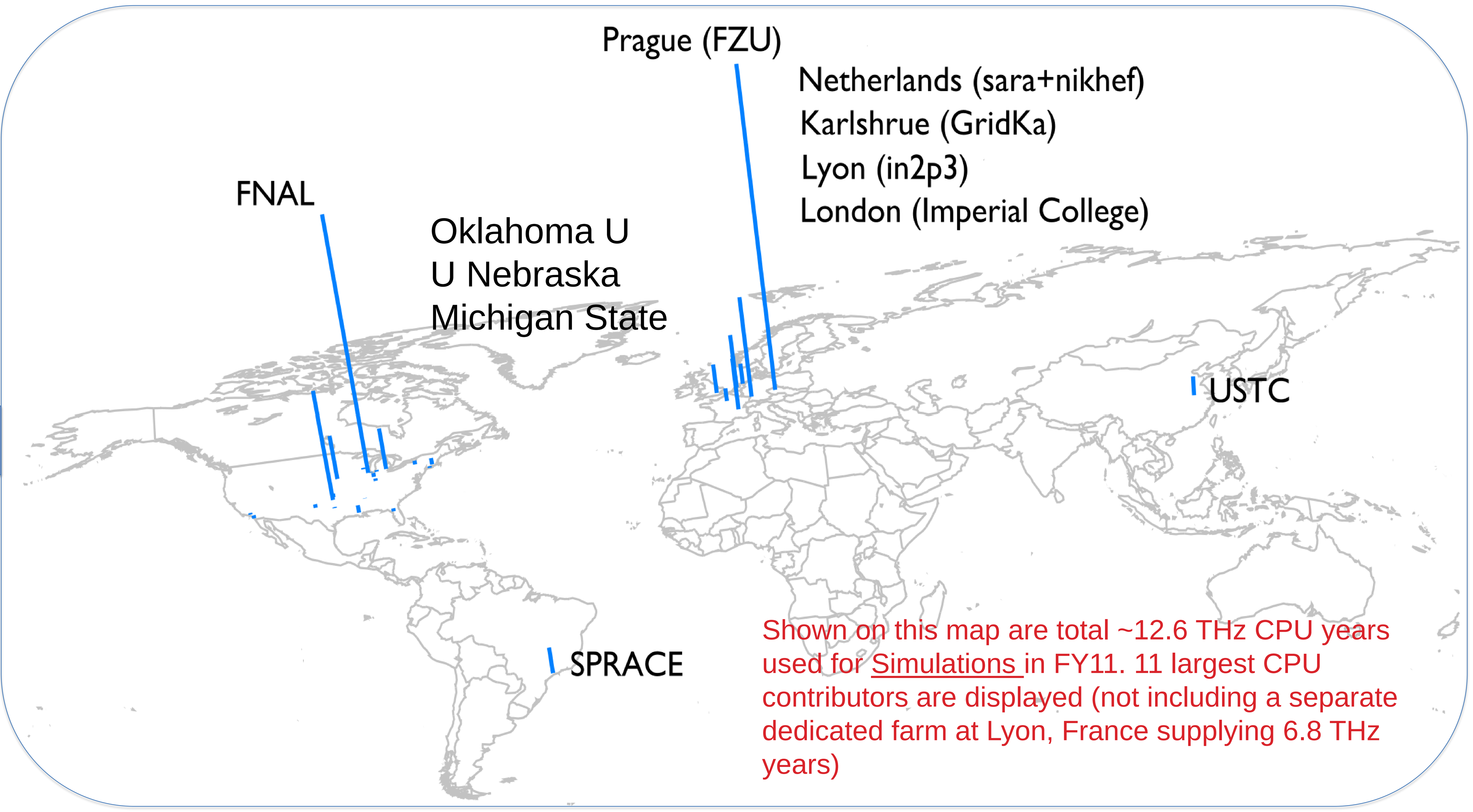
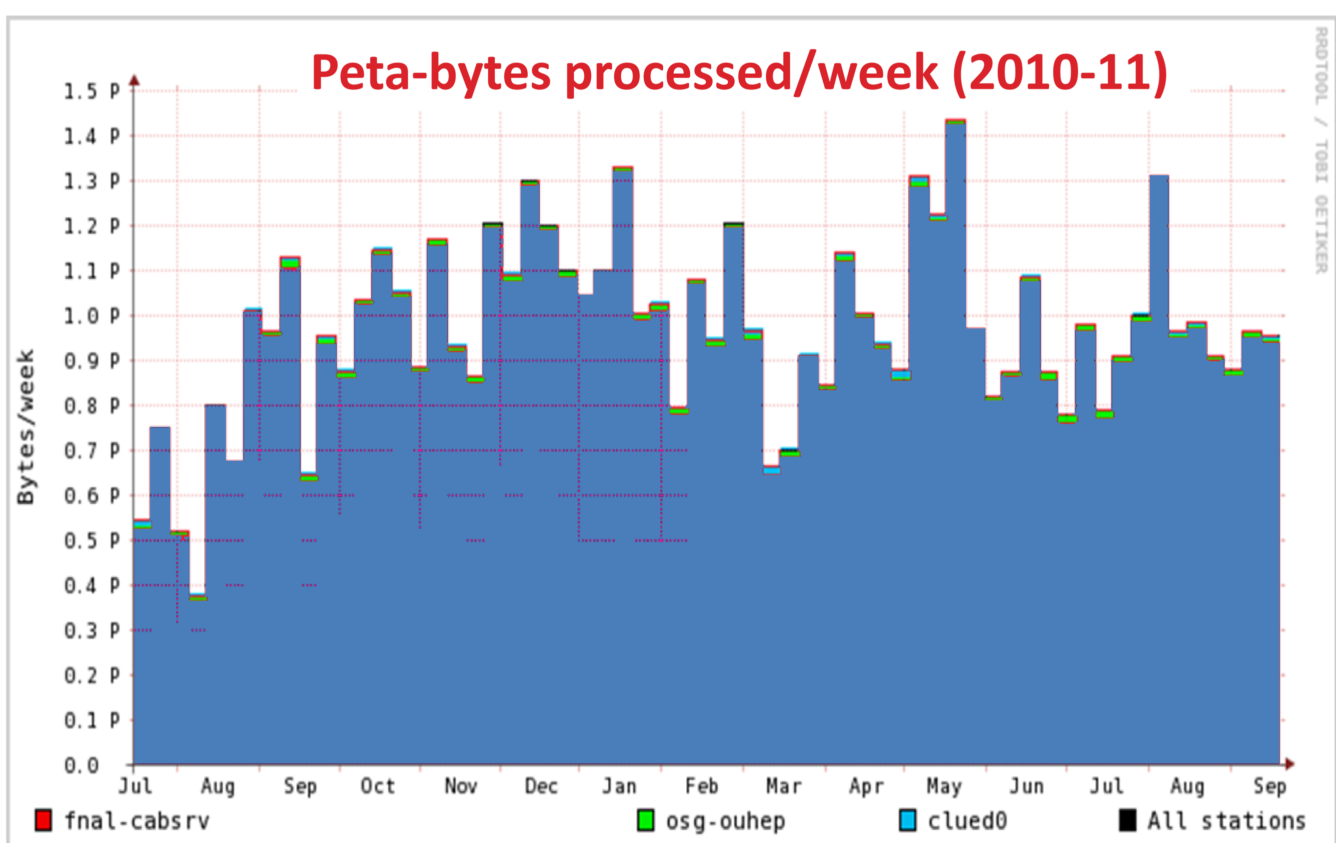
Data Handling and Cache Disks

CPU



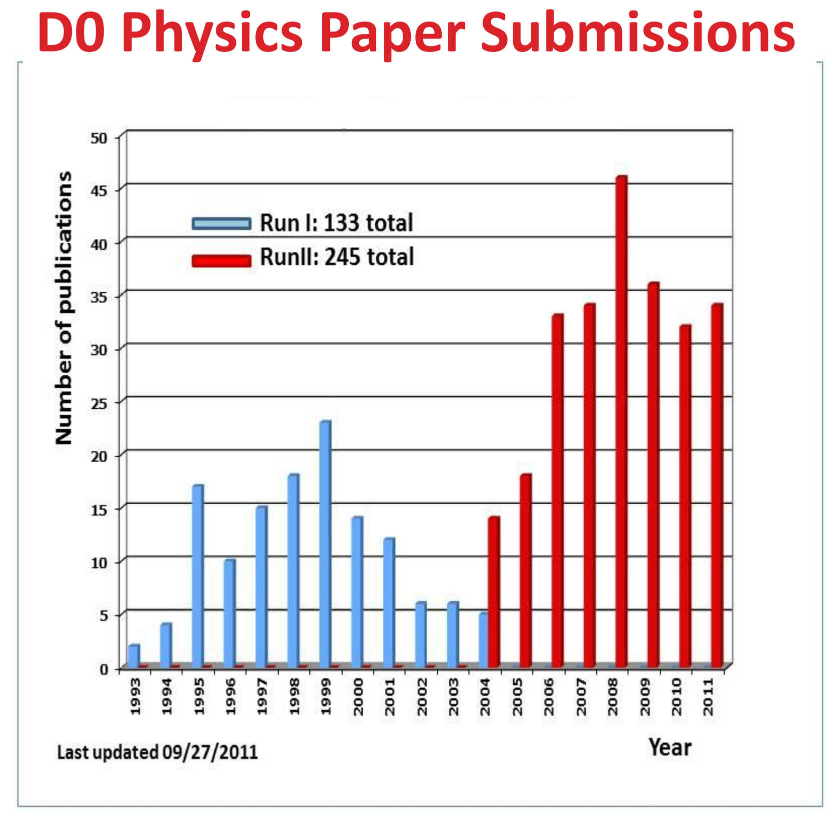
Shown on this map are total ~12.6 THz CPU years used for Simulations in FY11. 11 largest CPU contributors are displayed (not including a separate dedicated farm at Lyon, France supplying 6.8 THz years)

+

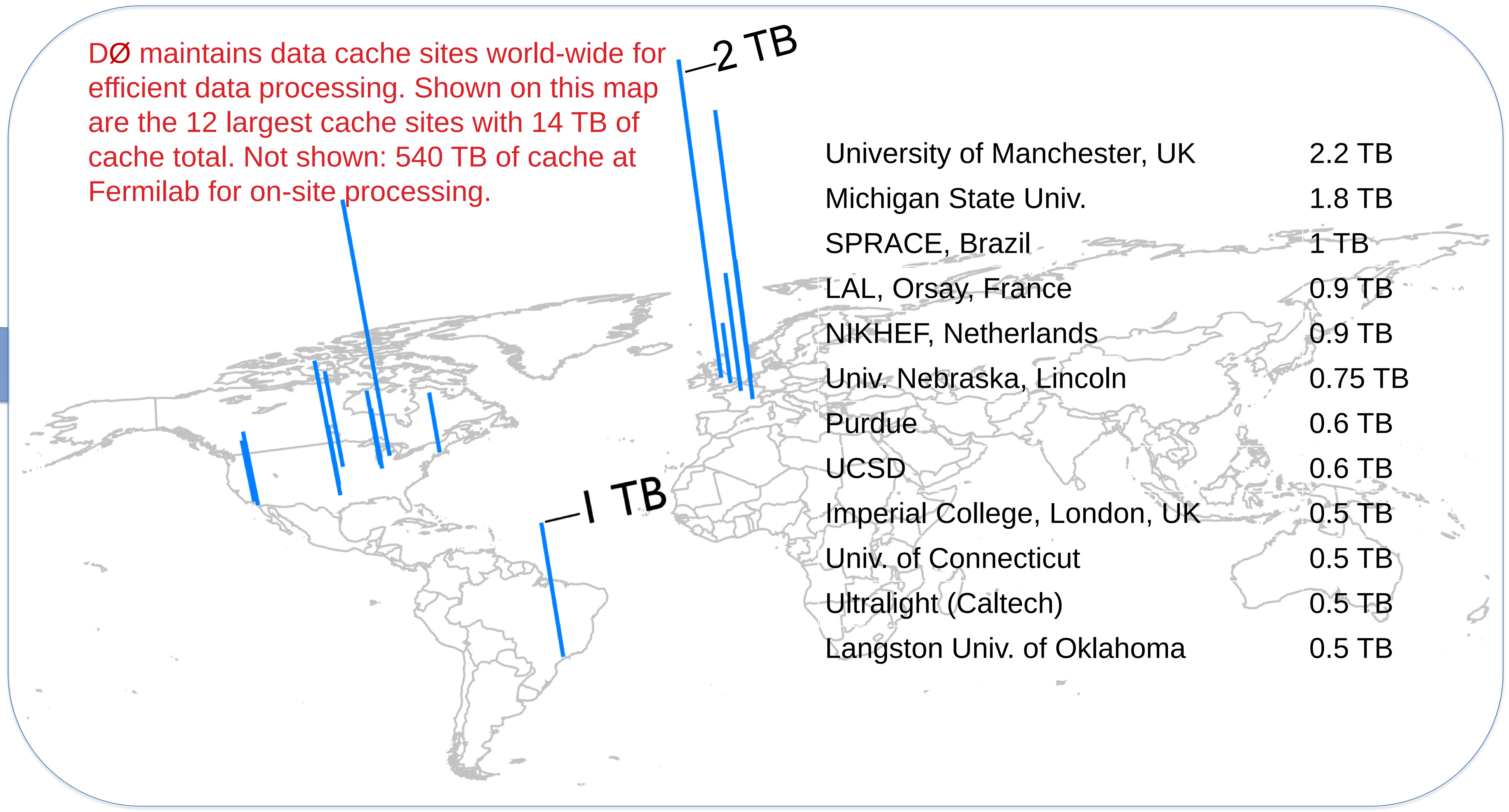


=

D0 Physics Paper Submissions



Data



DØ maintains data cache sites world-wide for efficient data processing. Shown on this map are the 12 largest cache sites with 14 TB of cache total. Not shown: 540 TB of cache at Fermilab for on-site processing.

University of Manchester, UK	2.2 TB
Michigan State Univ.	1.8 TB
SPRACE, Brazil	1 TB
LAL, Orsay, France	0.9 TB
NIKHEF, Netherlands	0.9 TB
Univ. Nebraska, Lincoln	0.75 TB
Purdue	0.6 TB
UCSD	0.6 TB
Imperial College, London, UK	0.5 TB
Univ. of Connecticut	0.5 TB
Ultralight (Caltech)	0.5 TB
Langston Univ. of Oklahoma	0.5 TB

The DØ Experiment used 33.1 Terahertz (THz) years of CPU last year (FY11) [which is equivalent to 33,100 one Gigahertz(GHz) CPUs running for a year] : 60% for Simulations, 40% for Reconstruction and Analysis

Physics Results

The DØ Experiment has recorded 7.7 Petabytes (PB) total of which about 1 PB are frequently accessed data for analysis and cached world-wide

Tevatron Run II Offline Computing

Experiments

The CDF and DØ experiments are particle physics detectors at the Fermilab Tevatron Collider. The experiments are huge, each 4-5 stories tall and weighing thousands of tons. Each experiment has approximately 500 physicist collaborators who design and build hardware, software, operate the experiment, and analyze the data.

These experiments have been running for over two decades and have produced results on a huge variety of particle physics phenomena including the discovery of the Top Quark in the 1990s.

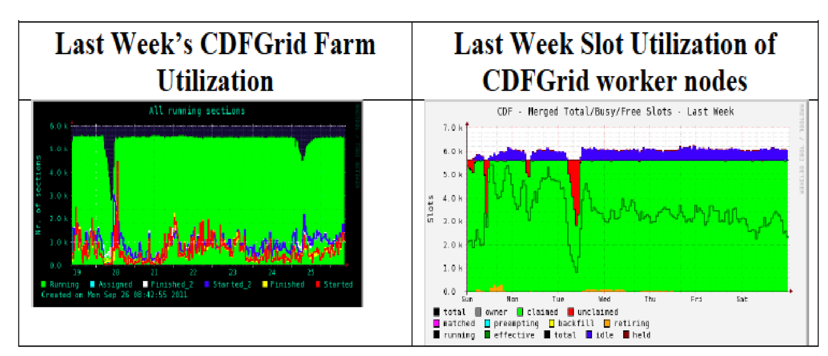
Raw data flow from the detectors into tape storage. Specialized reconstruction software is run to transform the raw data into analyzable data. The latter is processed many times with programs from individual physicists or small groups to study the physics. Furthermore, simulations of physical phenomena and the detectors are crucial for these analyses.

Reconstruction

The raw data set for the CDF experiment contains ~20 billion events and requires ~ 3 PB of storage.

The total reconstructed data set for CDF requires ~ 6 PB of storage.

Reprocessing requires 20-40% of the Fermilab Run II farms**.**



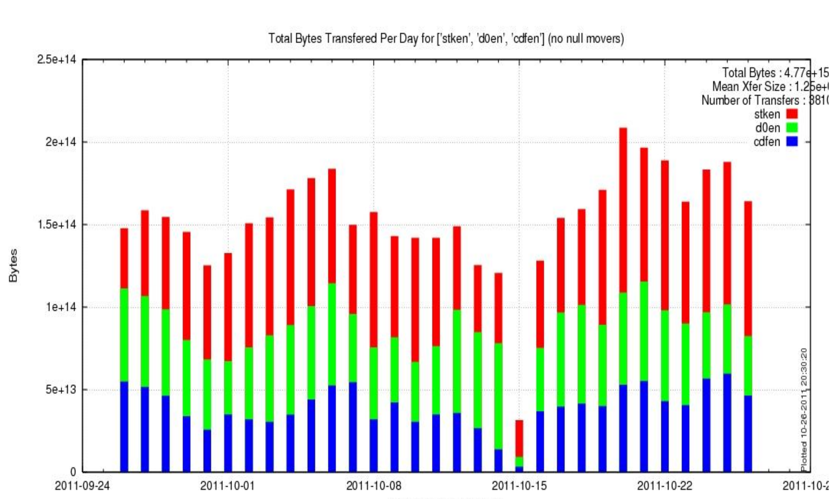
The CDF and DØ detectors do not record physics quantities of particles but instead raw information, like at what time a channel had a pulse of a certain height. A particle traversing the detector generates a large quantity of raw data. Reconstruction programs search for all raw information for a particle, reconstruct the path of the particle in the detector and derive the physics quantities of the particle.

The reconstructed data are used by researchers in the analysis of physics phenomena. Since all the raw data are inherently available at Fermilab and moving large data sets is cumbersome, the reconstruction effort has been done exclusively using Fermilab resources. While the Tevatron was running, the reconstruction processing had to keep up with the deluge of new data from the detectors. This activity consumed about 15 - 30% of the Run II compute farm, or 1500 - 4000 CPU cores/year.

Sometimes enhancements can be made in the reconstruction algorithms to increase sensitivity in various particle searches. Re-running reconstruction on any data set is very resource intensive and can take months. Now that the data set is complete, both experiments are reprocessing. This effort will complete over the course of the next several months.

Monte Carlo Simulation

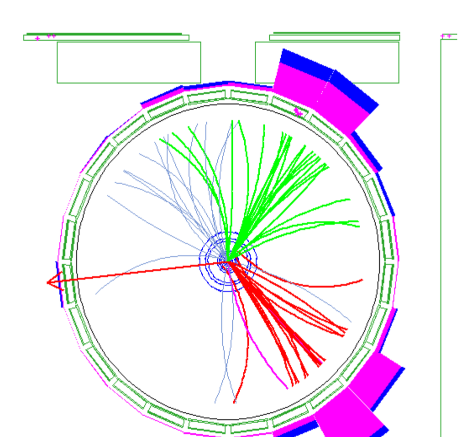
The DØ experiment produces (on average) 50 million simulated events a week, requiring 19.4 THz/year of CPU and generating 250 TB of data/year.



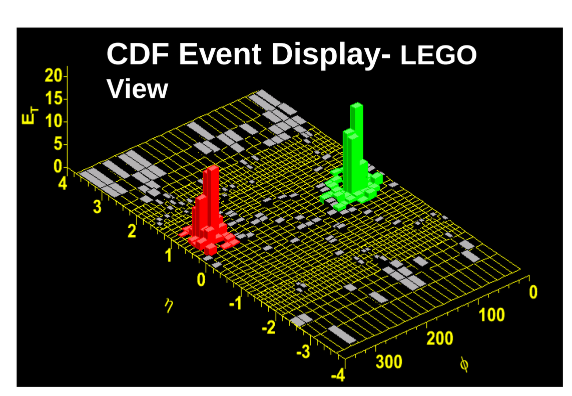
Monte Carlo simulations are used in generating events according to various physics processes and to simulate the detector response of the particles of those events as they traverse the detector. The method provides a statistical sampling over the many possible states and uncertainties.

Although not as I/O intensive as analysis jobs, the simulation computations represent from 1/3 to 1/2 of Run IIs computing needs. Because data access needs are smaller, it is often the Monte Carlo simulations that exploit the worldwide resources of the collaboration. More than 85% of simulations are done outside of Fermilab for the DØ experiment.

Physics Analysis



CDF Event Display -Transverse View



The particles reconstructed in the detector are analyzed to determine the most likely physics process that created the event. Events are then sorted by physics process and from statistical distributions physics quantities measured like the mass of the top quark.

Again because of the complexity and robustness of moving large data sets to remote sites, physics analysis is primarily completed on Run II compute farms. The nature of the analysis programs are highly specific to the particular scientist and can require a broad range of resources in terms of data handling, CPU power, parallelism, availability of specific algorithms, OS versions, etc. The batch system needs to be highly configurable in order to support these diverse needs and prioritize the analyses among the various research efforts as well as against the simulation and reconstruction needs.