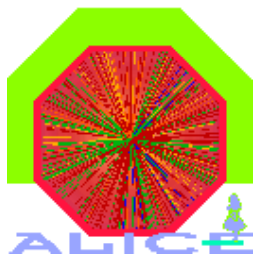


LHC Computing



Ian Fisk

CHEP Conference
Victoria, Canada
September 3, 2007

ALICE, ATLAS, CMS and LHCb are
led by LHC Computing

Data expected in late July of 2008

Active preparations for computing for
5-6 years

Big increase the proposed scale of

Distribution

Data Transfer

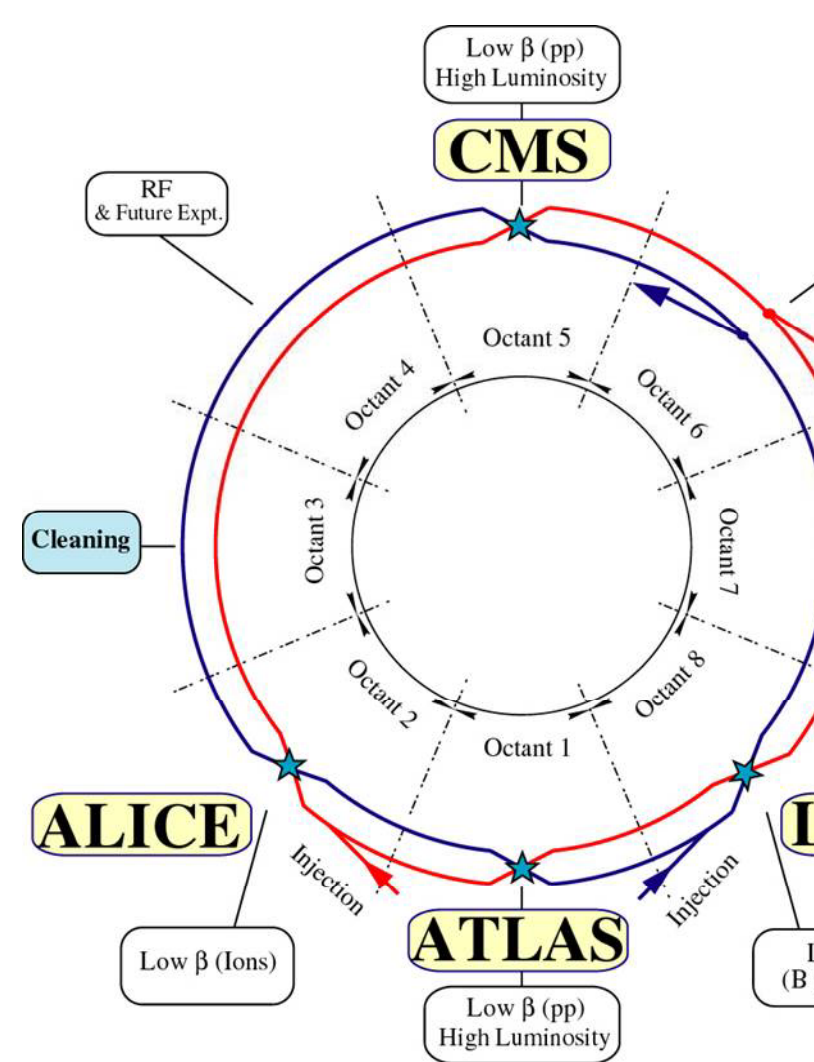
Data Access and Analysis

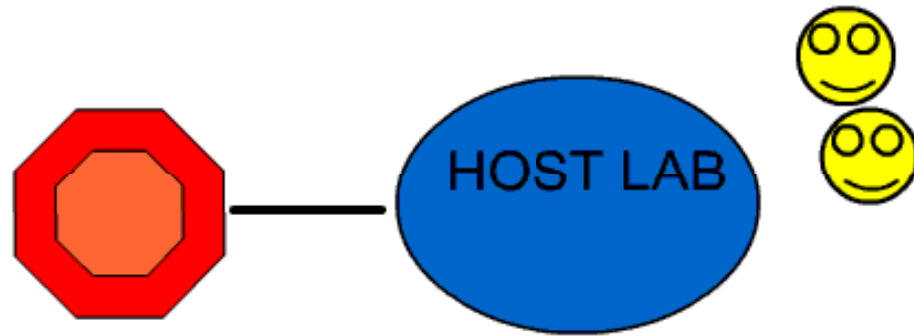
LHC experiments have enjoyed an
unprecedented level of support from

grid projects, national funding

agencies, national labs, and

universities



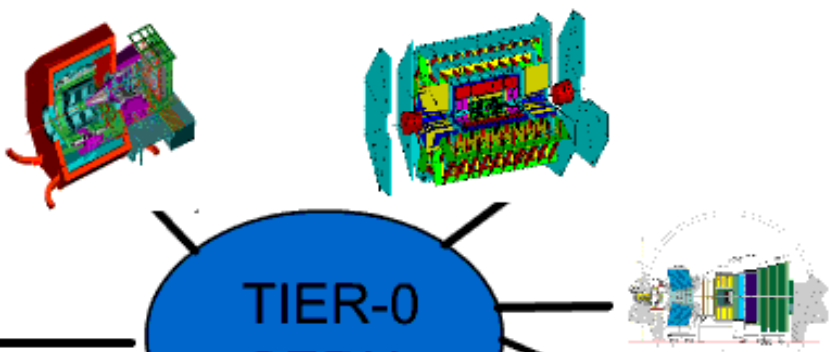


beginning the computing was centralized

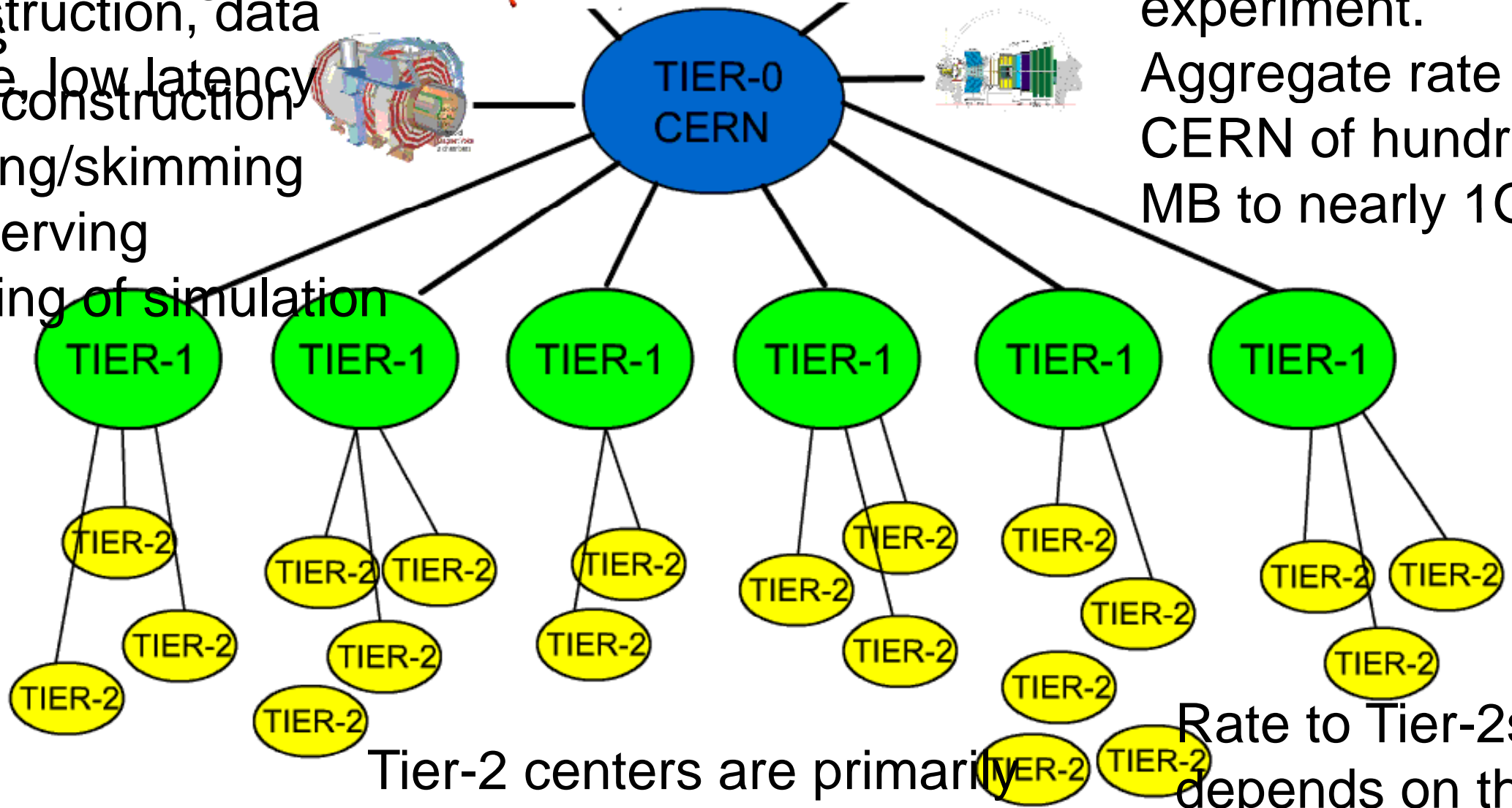
periments began to develop distributed computing models

Two examples: Babar had Tier-As that users could connect to for access to the data and resources. CDF had distributed analysis centers

Tier-0 center at CERN
 centers are primarily at
 ed for prompt
 al labs or large
 construction, data
 sites
 chive, low latency
 -Reconstruction
 k
 lipping/skimming
 ta serving
 chiving of simulation



by event size and
 trigger rate for ea
 experiment.
 Aggregate rate fr
 CERN of hundred
 MB to nearly 1GB

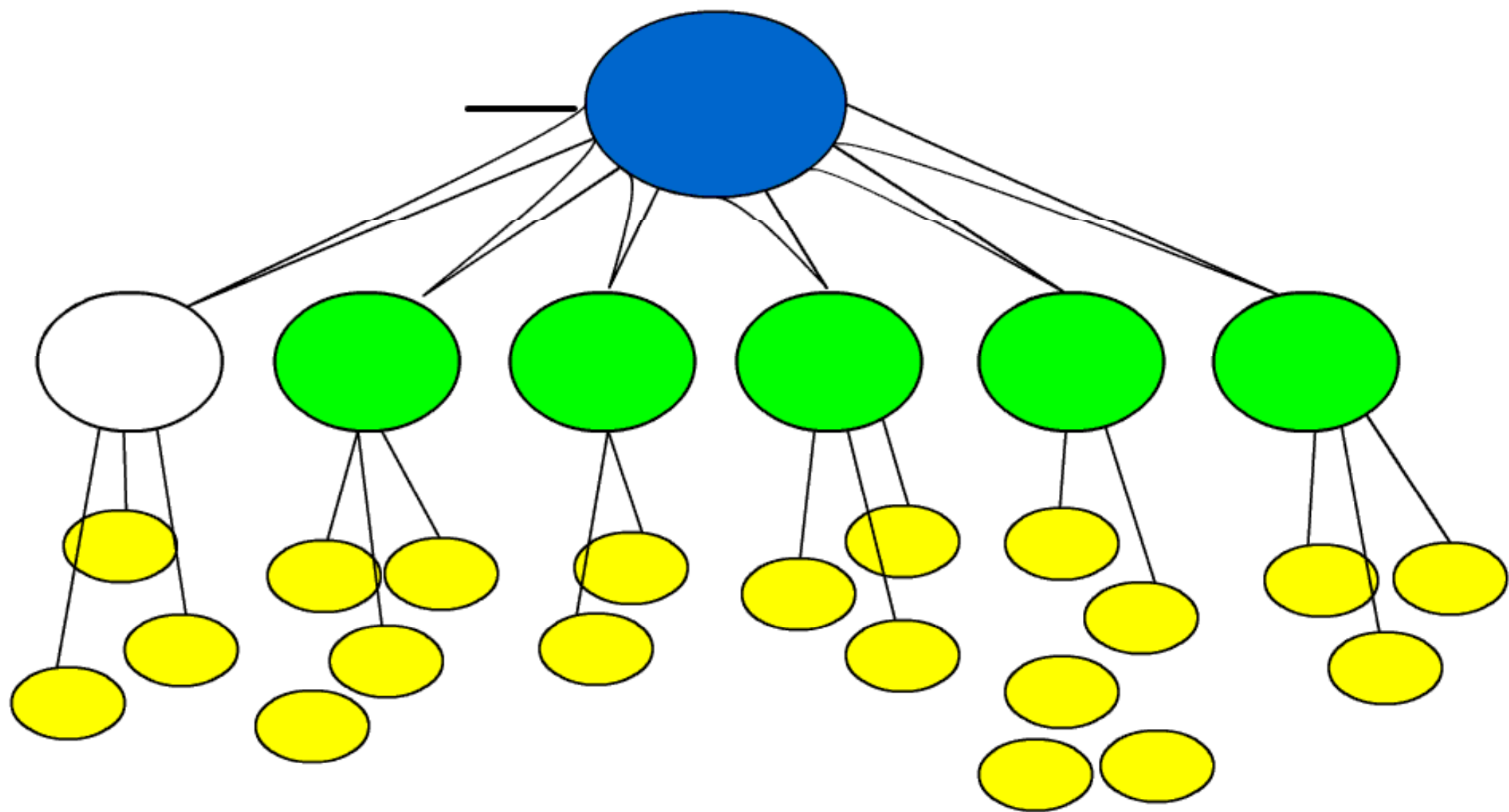


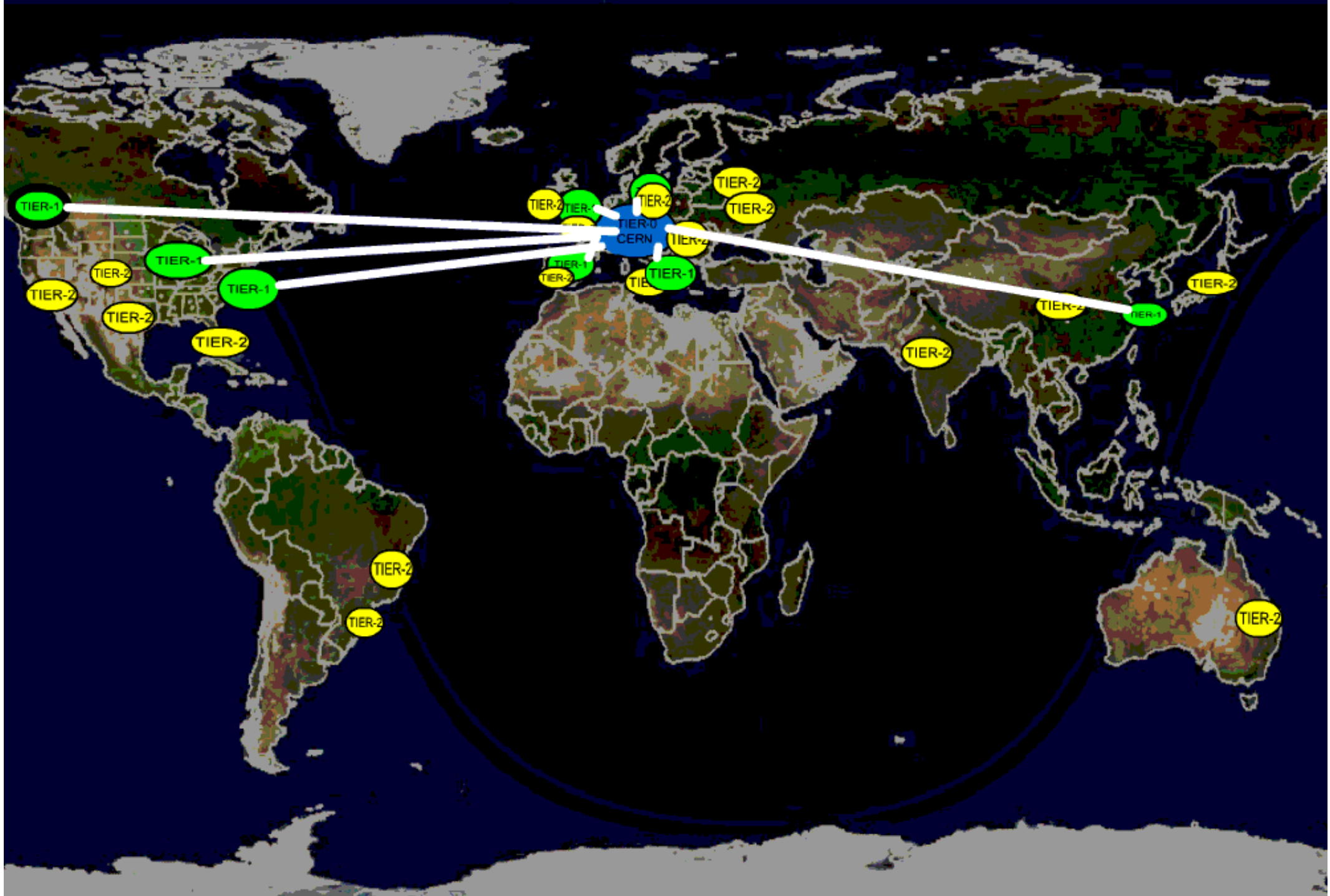
Tier-2 centers are primarily
 at universities

Rate to Tier-2s
 depends on the
 experiment and

IARC Tiered computing model came in the late 90's
 Level of distribution motivated by the desire to empower and
 • Simulation
 • User Analysis (ALICE,

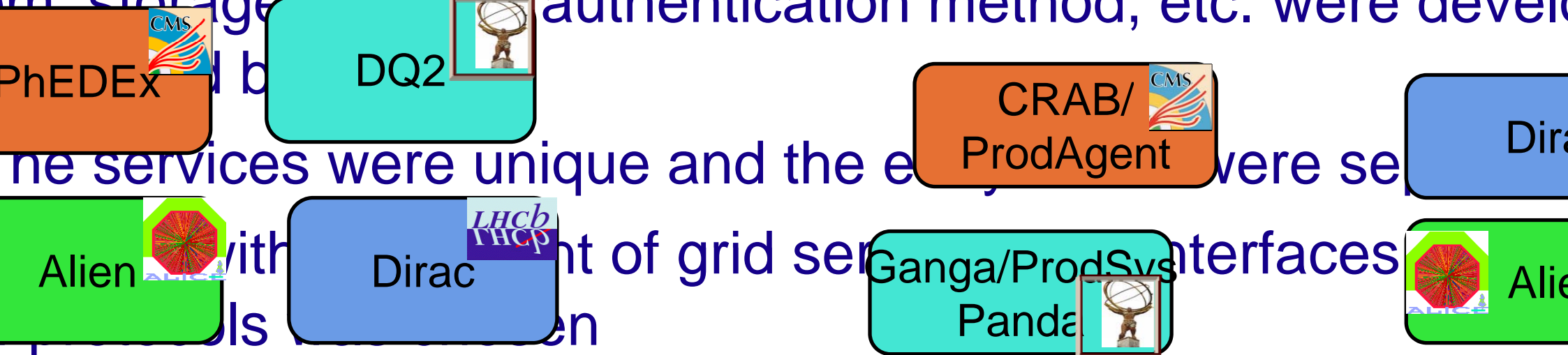
expectations for
 updating storage



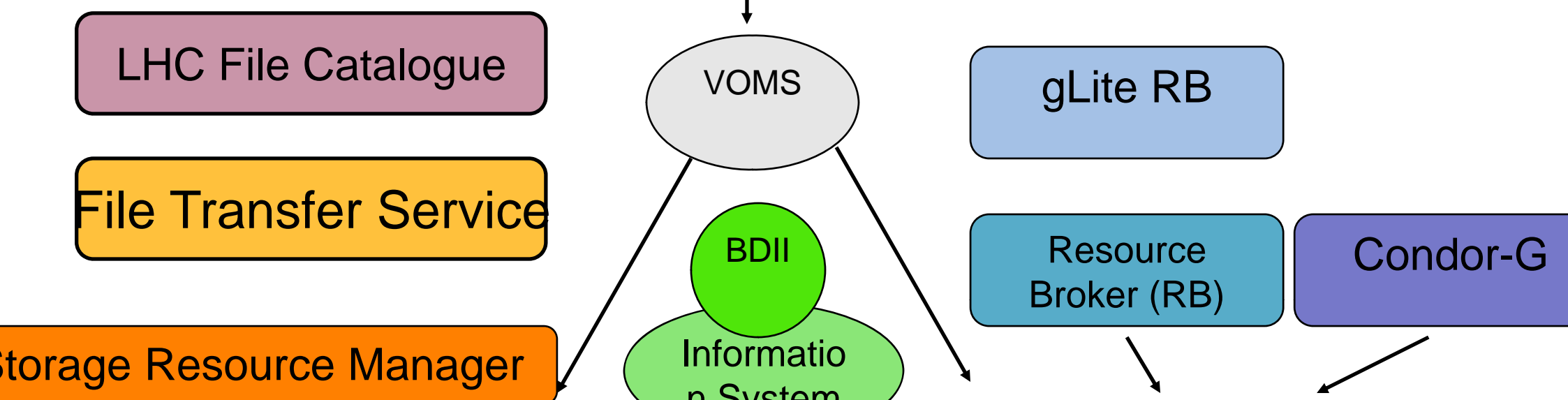


Open Private Network (OPN) connects CERN and Tier 1. Other

original CDF distributed analysis facility the services, operat
m storage solution authentication method, etc. were develo



Entire computing facilities could be shared between VOs,
opportunistic access would be possible, facility support would
be reduced, reliability would be improved...



level of distribution and the number of services requires an advanced system to check the health of the globally distributed system

VLCG has developed a series of Site Availability Monitors (SAMs) tests

Series of automatically submitted and tracked tests

Validate the processing services all the way down to worker nodes

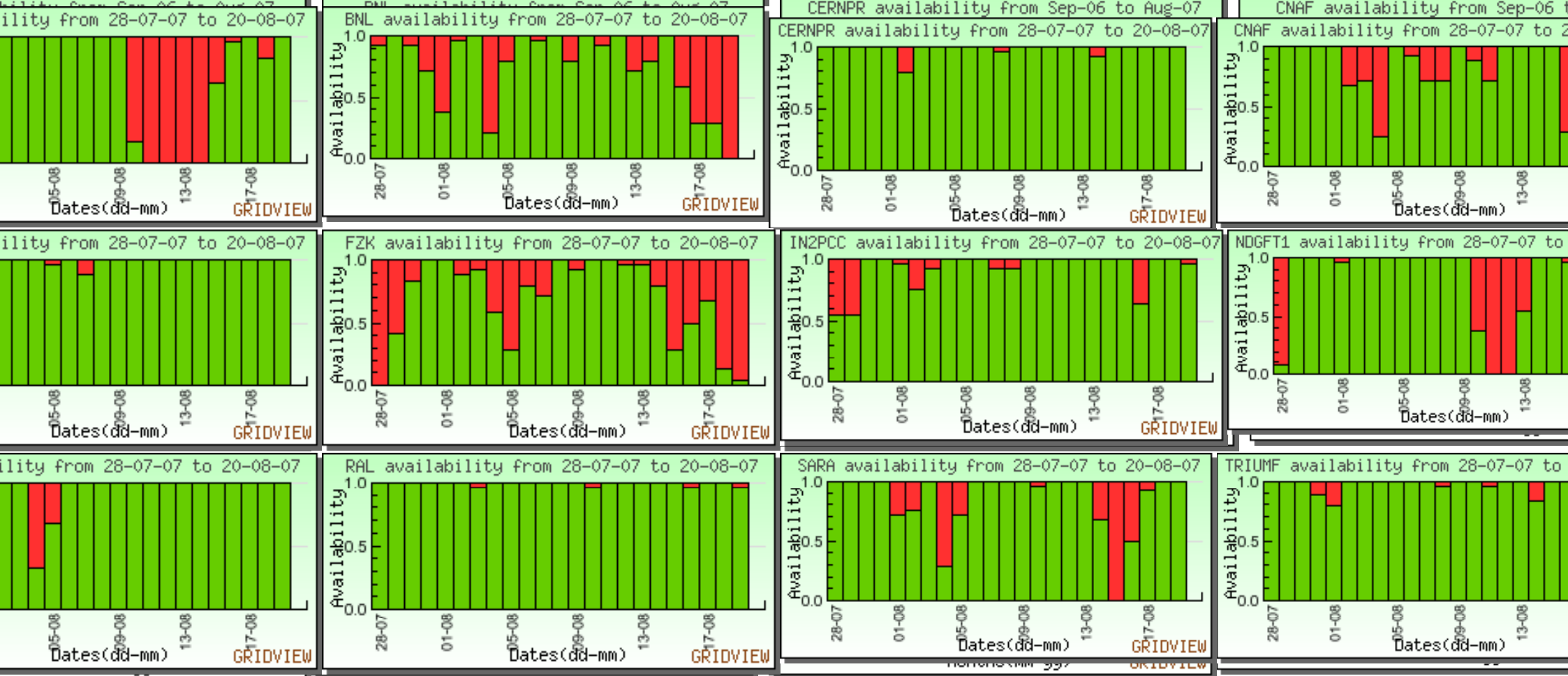
Validate storage services

Information systems

Tests run every few hours and results are tracked and published

near VOs have begun to introduce their own tests

Verify the experiment workflows within the SAM framework



Clearly areas for improvement

Underlying services need to end up in the much higher 90s

Experiments have worked on retries and failovers in both workflows and transfers to improve the efficiency.

varieties in the data management functionality for the 4 experiments

All experiments have services that sit on top of the grid services to define the mappings between events, files, and eventually datasets

A dataset is typically defined as a collection of logical file names

The files are immutable and can be replicated between sites

ATLAS and LHCb both use the LHC File Catalog (LFC) in production

CMS uses a TFC (Trivial File Catalog) technique similar to what was used in Babar, where the storage element namespace is used to resolve logical file names to physical file names without a central service

Experiment data management systems drive the replication of data

Tools to define datasets tend to be experiment specific because functionality is driven physics requirements and choices for what to be supported

Can be very flexible like ALICE's Event TAG service that allows users to place cuts and receive a new list of files for that particular query

- Datasets are more dynamic

In LHCb the specialized data sample lends itself to a predefined set of stripped datasets that are centrally produced

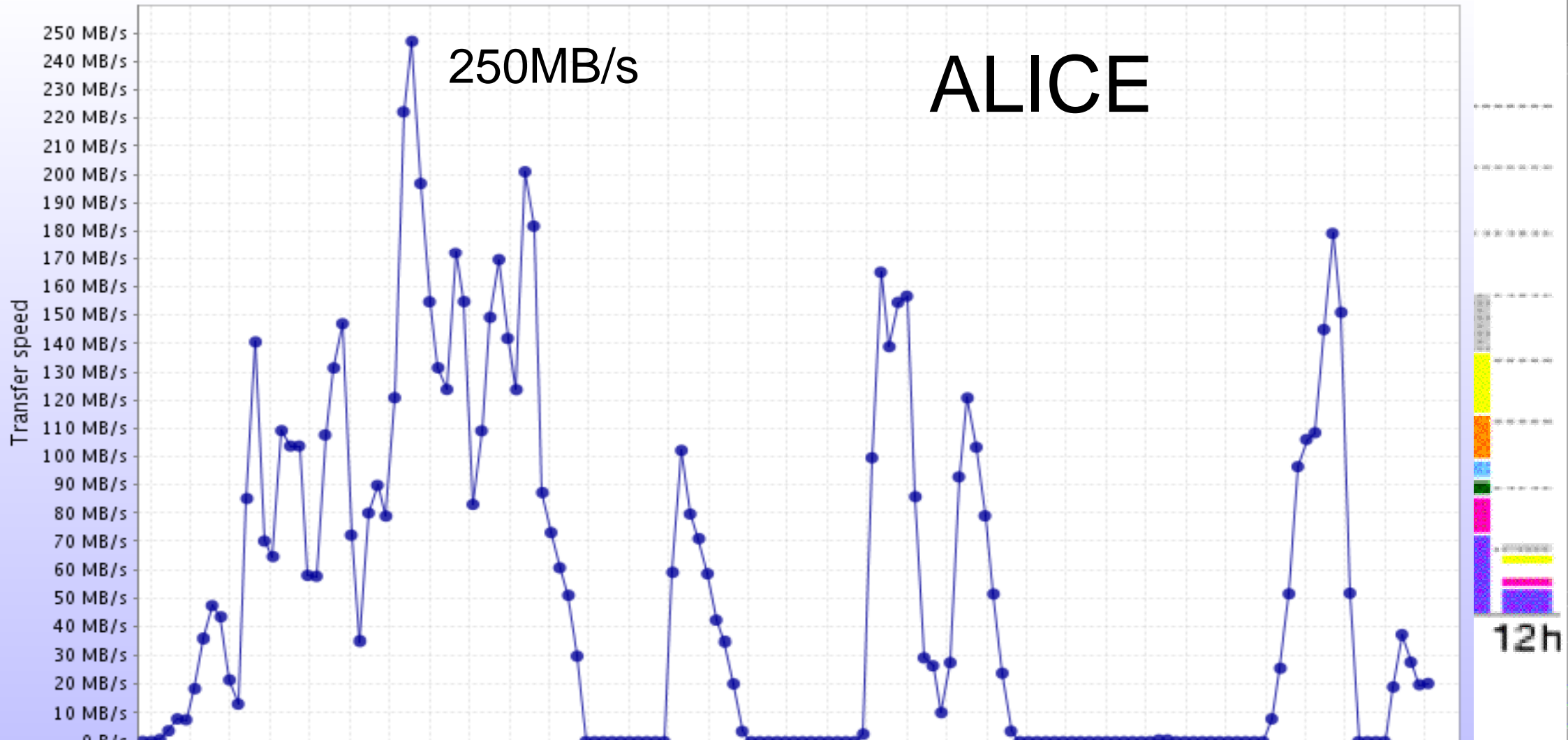
- Simplifies the definitions and access

CMS is in-between with datasets being defined and stored in a central bookkeeping service, but operations and users can define new datasets as needed

ATLAS has a system that allows querying datasets from the command line and bringing down a few files

S has the largest nominal CERN to Tier-1 transfer rate is AT
Tests this spring reached ~75% of the eventual target
Successful use of 11 Tier-1 centers, successful demonstration
SRM and FTS

SEs average transfer rates

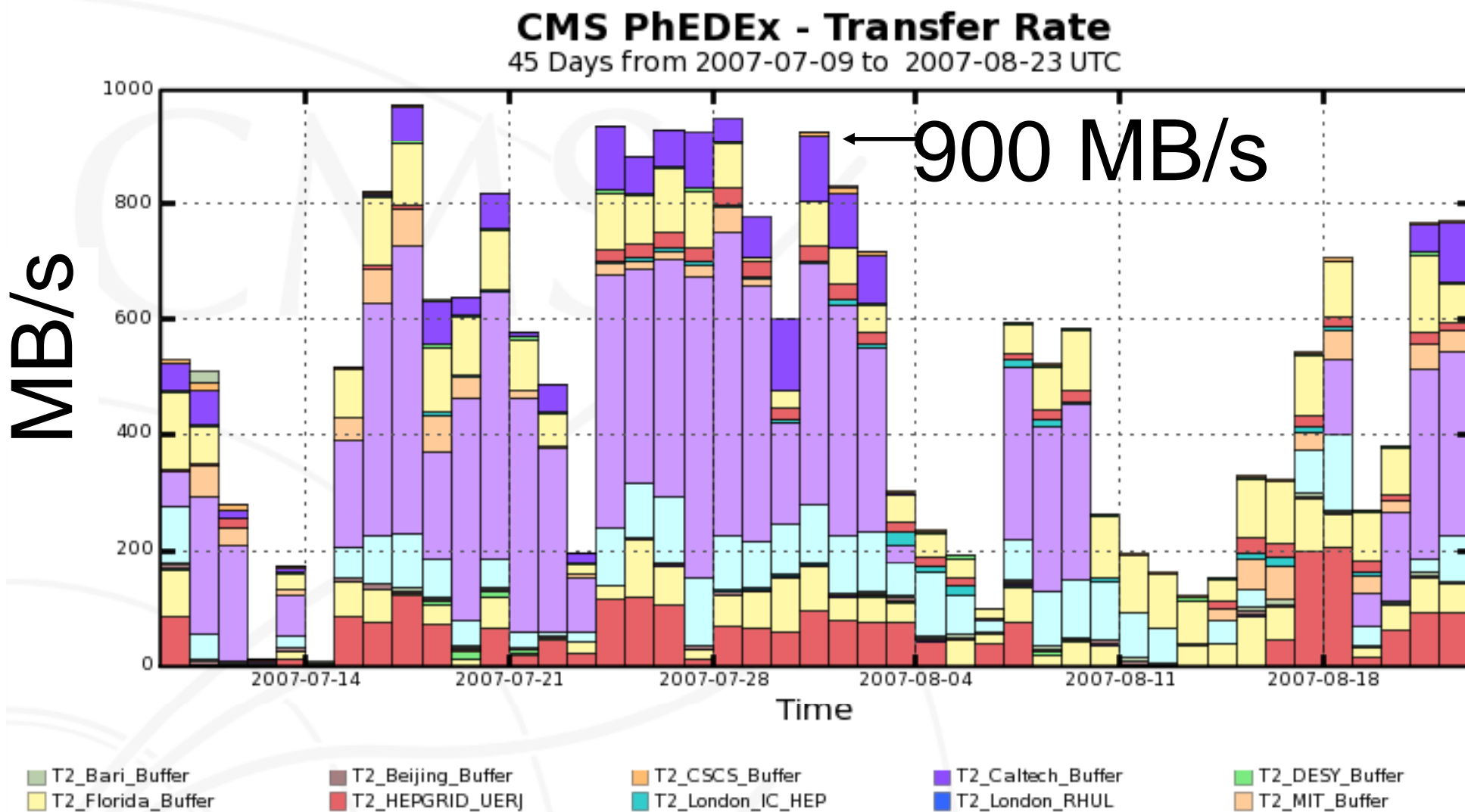


expects Tier-2 storage to be treated like a dynamic cache

Tier-2s can be updated with data from any Tier-1.

In 2008 data rates are expected in bursts of 50MB/s-500MB/s per link

Plot below of data exported from FNAL to 21 Tier-2s



Access to applications has been a difficult area for LHC
during commissioning

large number of sites, CPUs, and large volume of data

hierarchical mass storage

Need to be mindful of file size and rates of opening files

simplest solution, the mass storage system handles data

ing and serving to applications using an efficient local

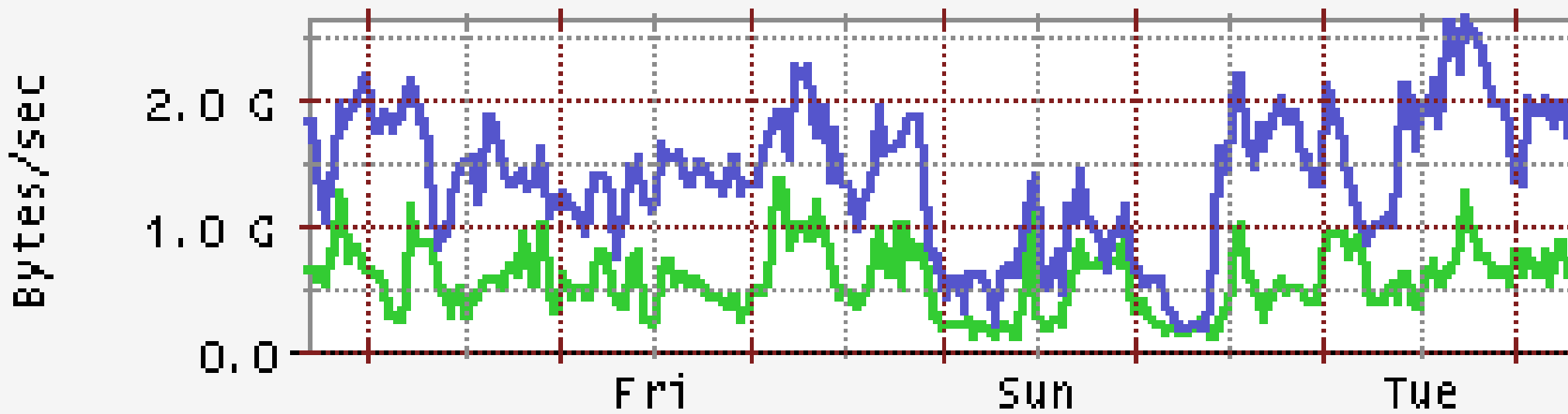
col (rfio, dcap, xrootd)



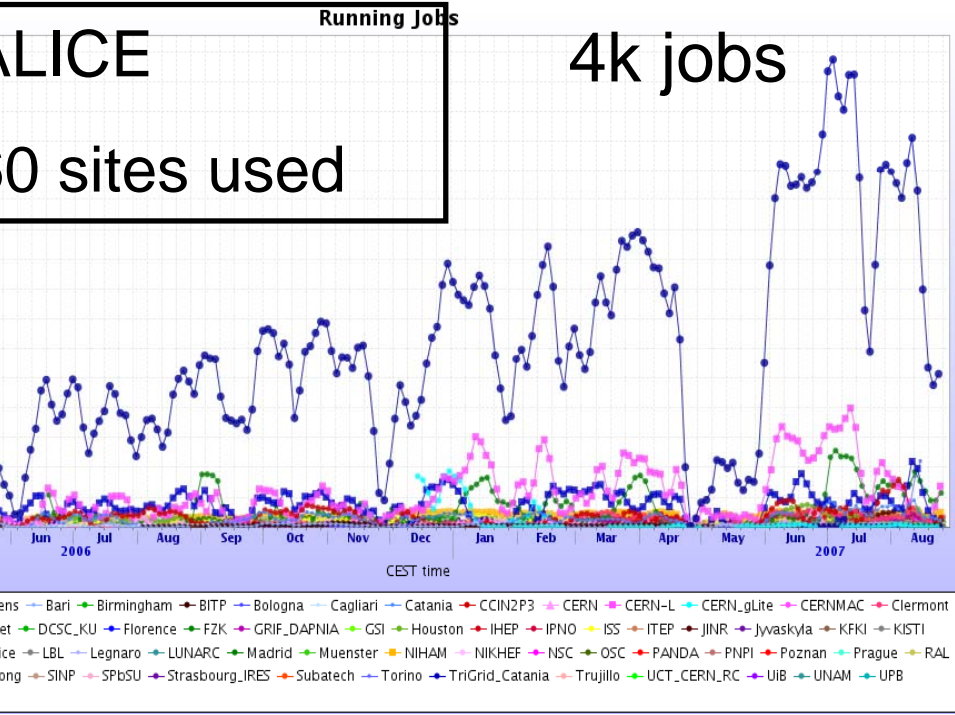
dcap/
RFIO

n orde
espon
HCb
elease
isk
Other V
storage

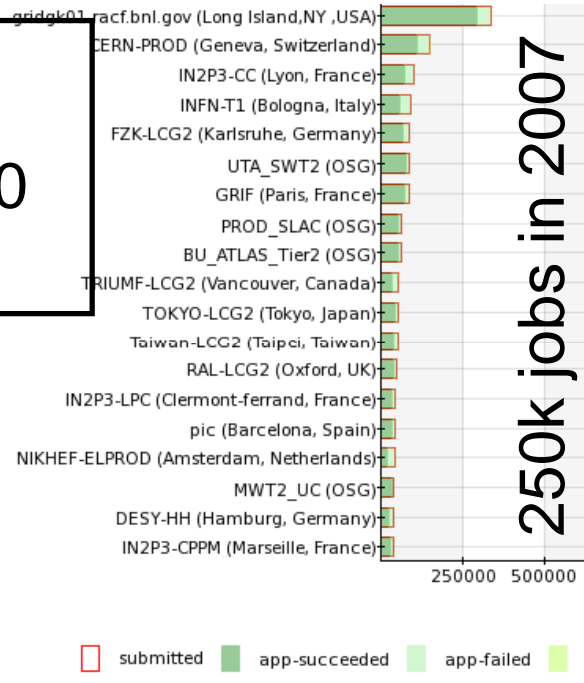
dCache Pools Network Last week



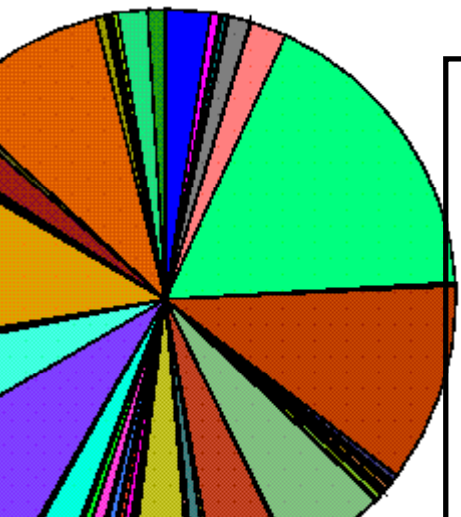
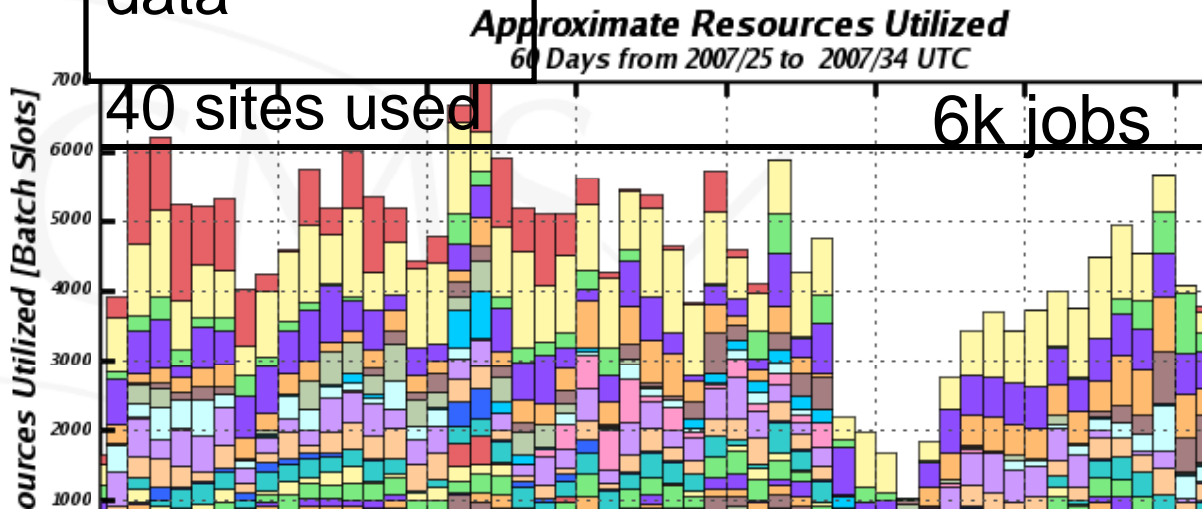
Production is an ideal candidate for distributed processing
 Large output and CPU requirements but small input and
 predictable applications. All four experiments are succeeding



ATLAS
 More than 100 sites used



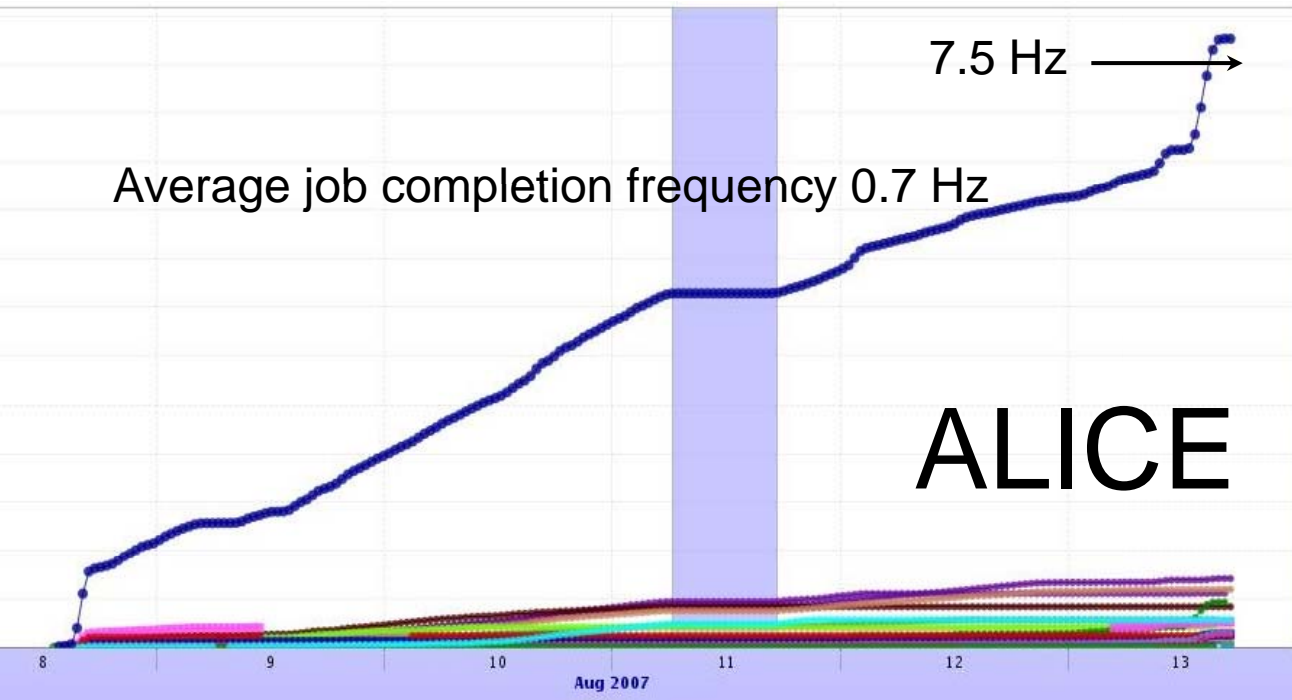
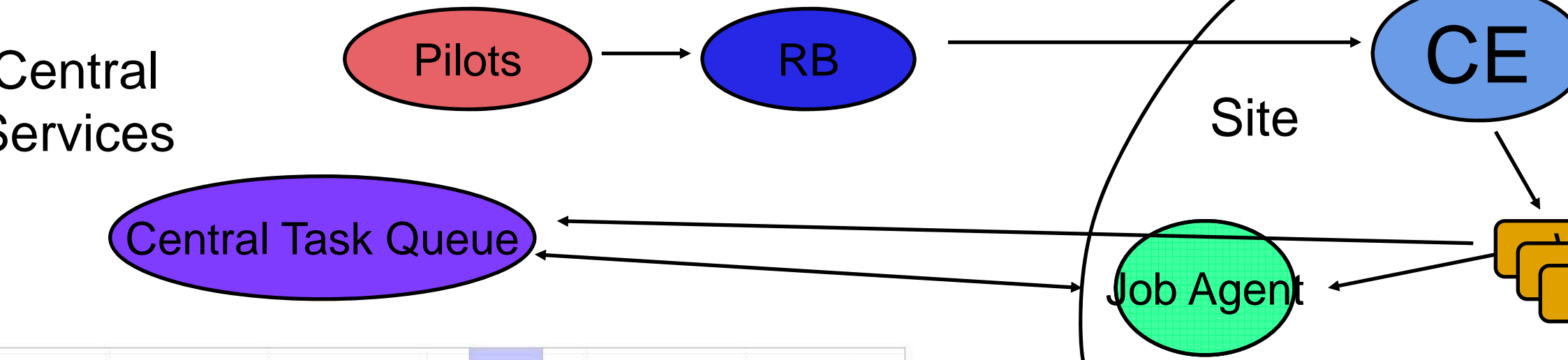
CMS
 170M Events
 ~180TB of data



LHCb
 375M Events
 ~90TB of data produced

ALICE and LHCb have developed pull based job submission systems for both Production and Analysis

ATLAS uses pull for one of the work flow tools

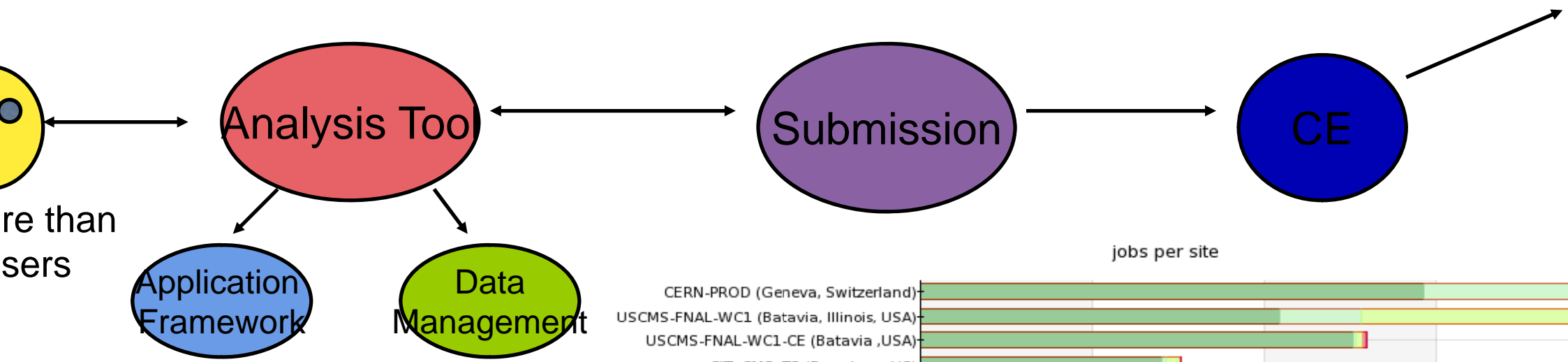


Average rate is about 3
The final rate expected
→ Central queue sca
Central task queue is L
Dirac system reached
Processes

Analysis processing is more interesting need to match processing resources to large quantities of data.

Systems used in ATLAS and CMS are similar in the steps

Ganga and Panda in ATLAS and CRAB is CMS



more than users

Currently CMS is averaging

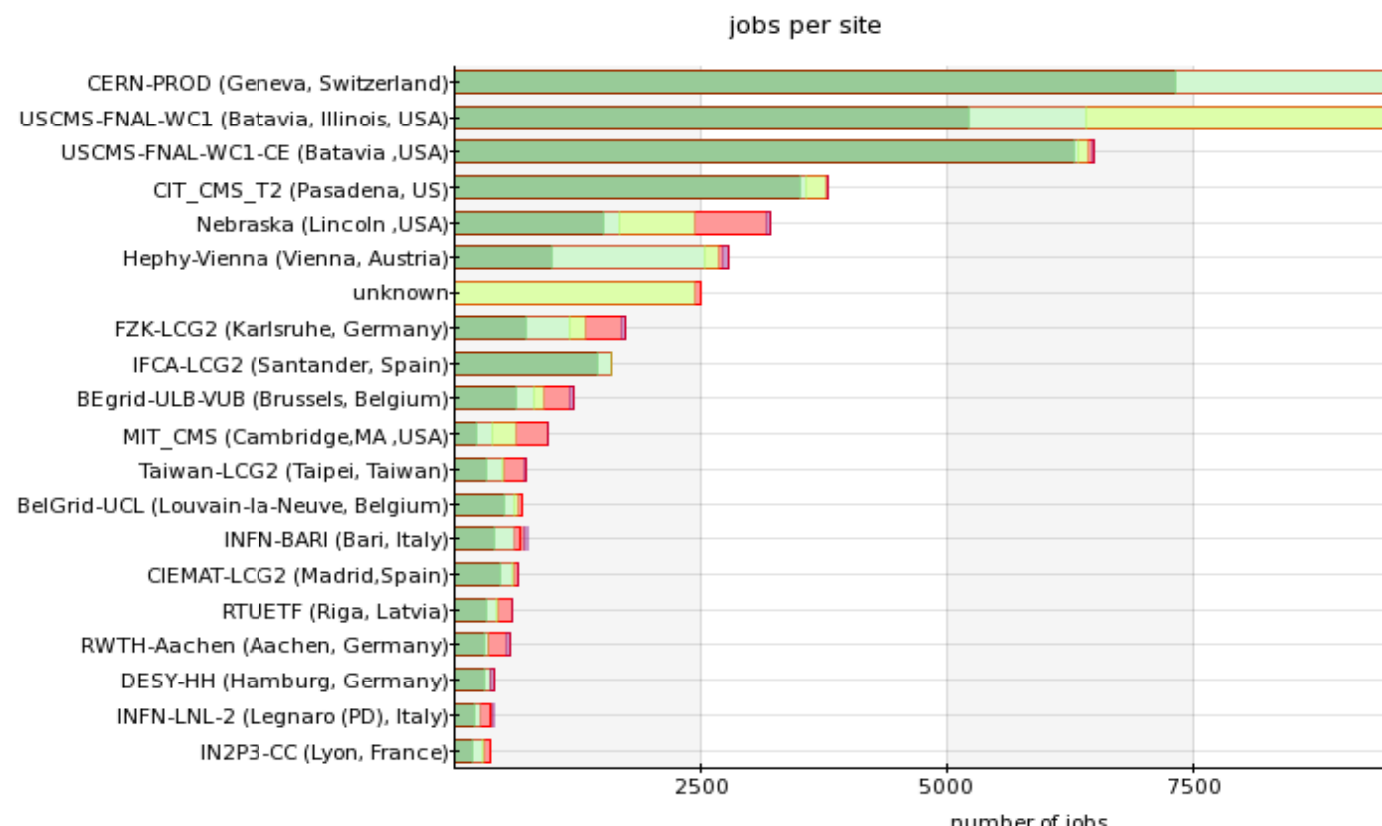
jobs per day

Estimate in 08 is > 100k

Dashboard reports

for numbers

User scripts outside of



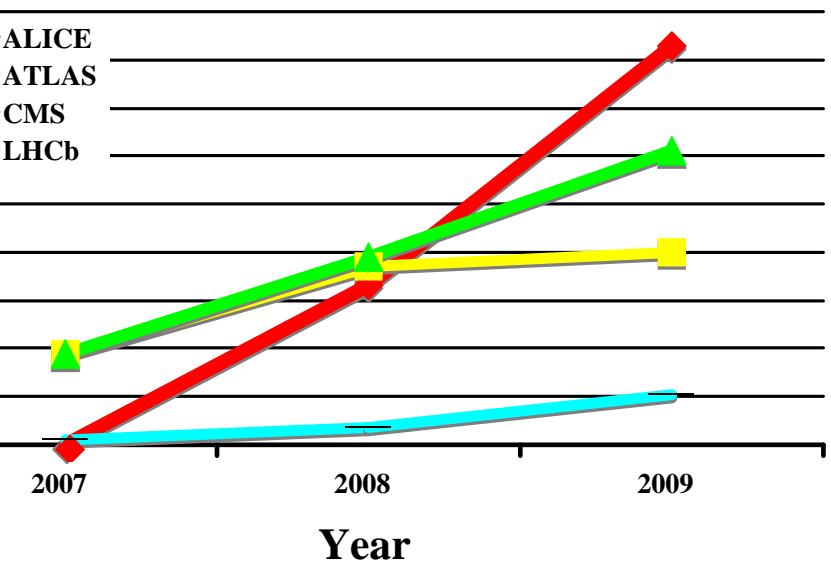
The total quantity of computing resources needs to more than double all the experiments over 2 years

- Some of this can be accounted for by Moore's law improvements
- While large there is experience running farms this large

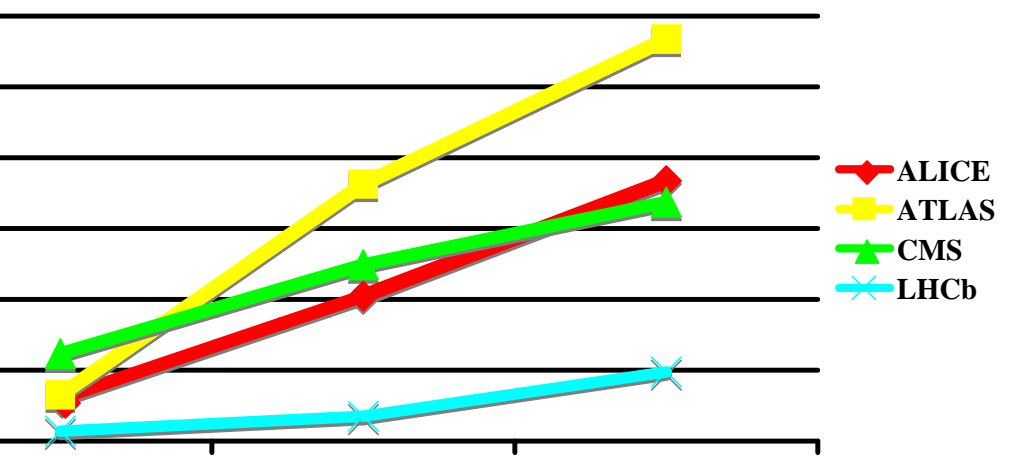
In order to reach the scale required a lot of processes

- Node purchased in 2003 had 2kSI2k and needed 2-3 processes to utilize them
- Node purchases in 2007 had 15kSI2k but requires 8-10 processes to utilize them
- Impacts the scale of required

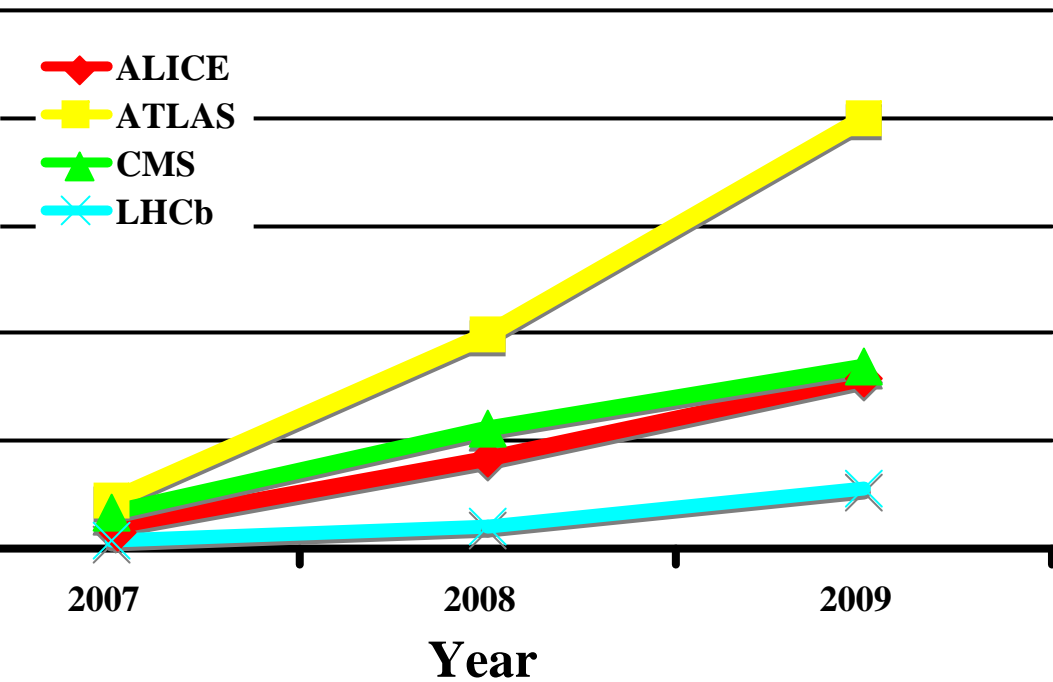
Tier-0 Resources



Tier-1 Resources



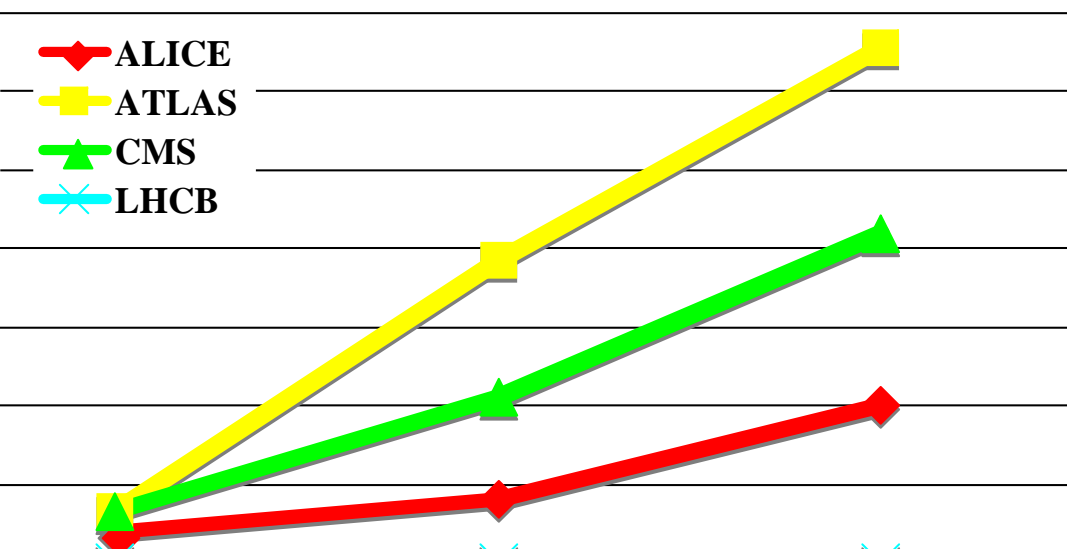
Tier-1 Disk Resources



Disk ramp is a little more concerning

- The required increase cannot be accommodated by technological improvements alone.
- There are a limited number of examples of multi-peta byte installations
- Issues of facility operations and scalability of storage name spaces

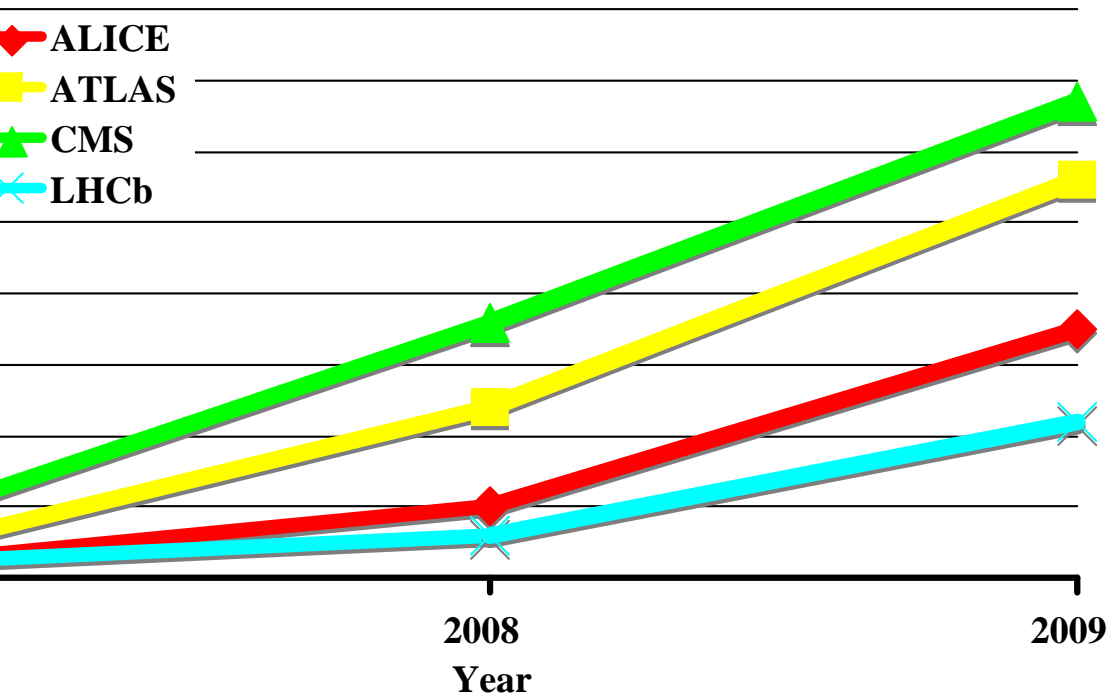
Tier-2 Disk Storage



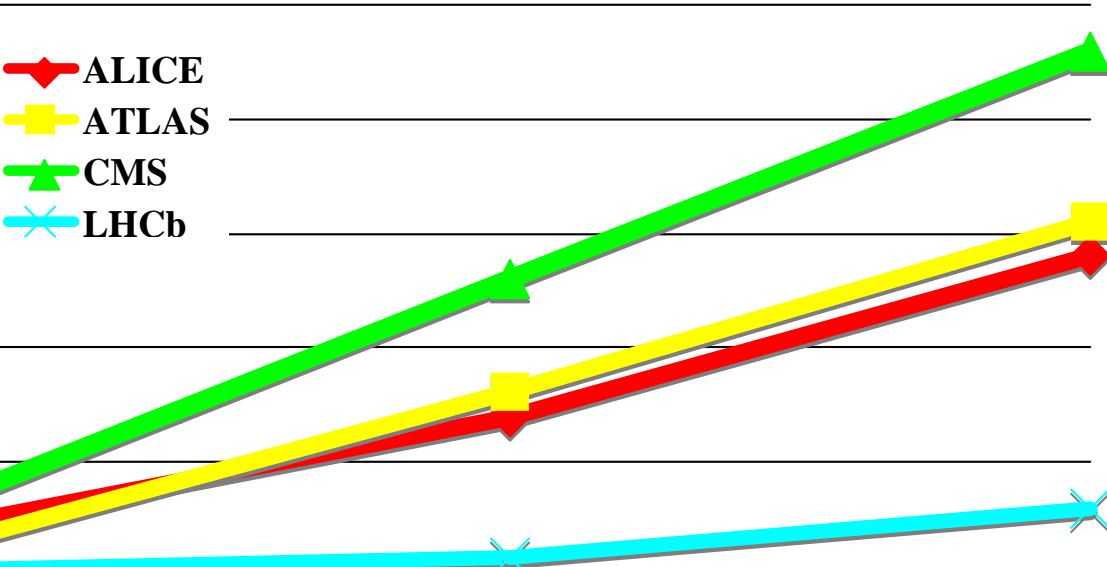
Performance and stability of mass storage is dependent on how it is used

- Rely on experiment for reasonable file sizes and access rates

Tier-0 Tape Resources



Tier-1 Tape Resources



Tape resources are some of the most scalable

- Robotic storage is designed to handle large quantities of data

Also one of the services that requires the longest operational experience of operation reliability

- Not all Tier-1s are equally experienced

Most of the LHC experiments plan to operate in the write once read many times regime

- Standard operating mode

experiments have begun demonstrating computing infrastructure
at the scale expected to be seen in running conditions

Transfers from CERN

Resources utilized for simulated event production

Amount of work left in the final year of preparation

A big increase in scale needed in facility infrastructure and the
ability to use it routinely

User analysis access needs to ramp up

Very complicated computing environment and we are still learning
how to build and operate it