LHC Computing


Ian Fisk
CHEP Conference
Victoria, Canada
September 3, 2007

LICE, ATLAS, CMS and LHCB are
ed by LHC Computing

Data expected in late July of 2008

Active preparations for computing for
5-6 years

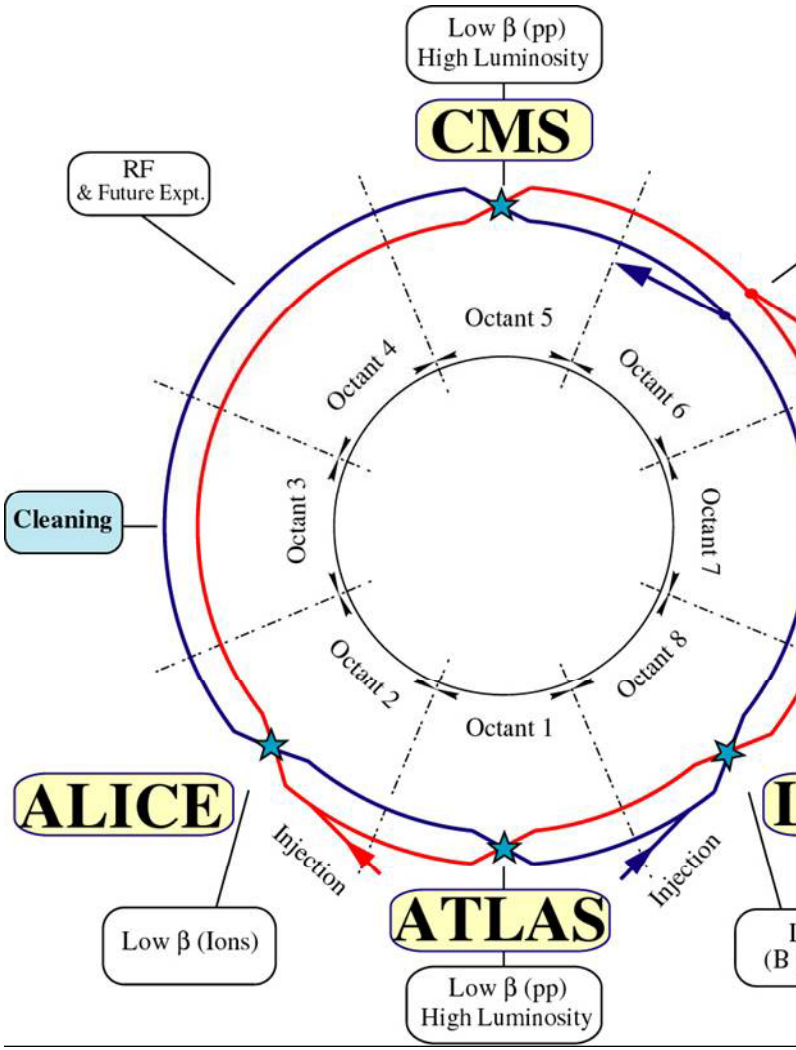
Big increase the proposed scale of

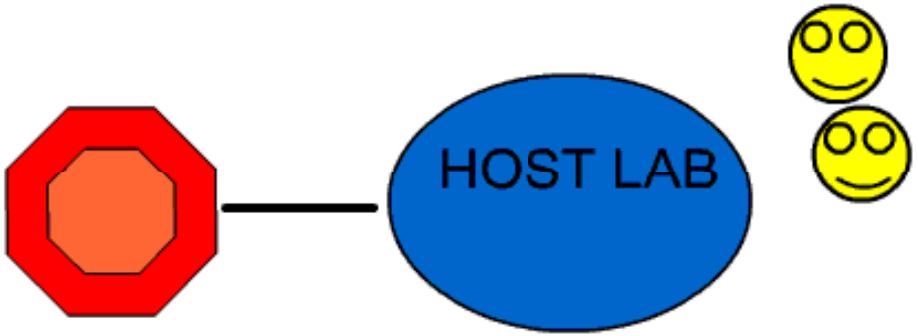
Distribution

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

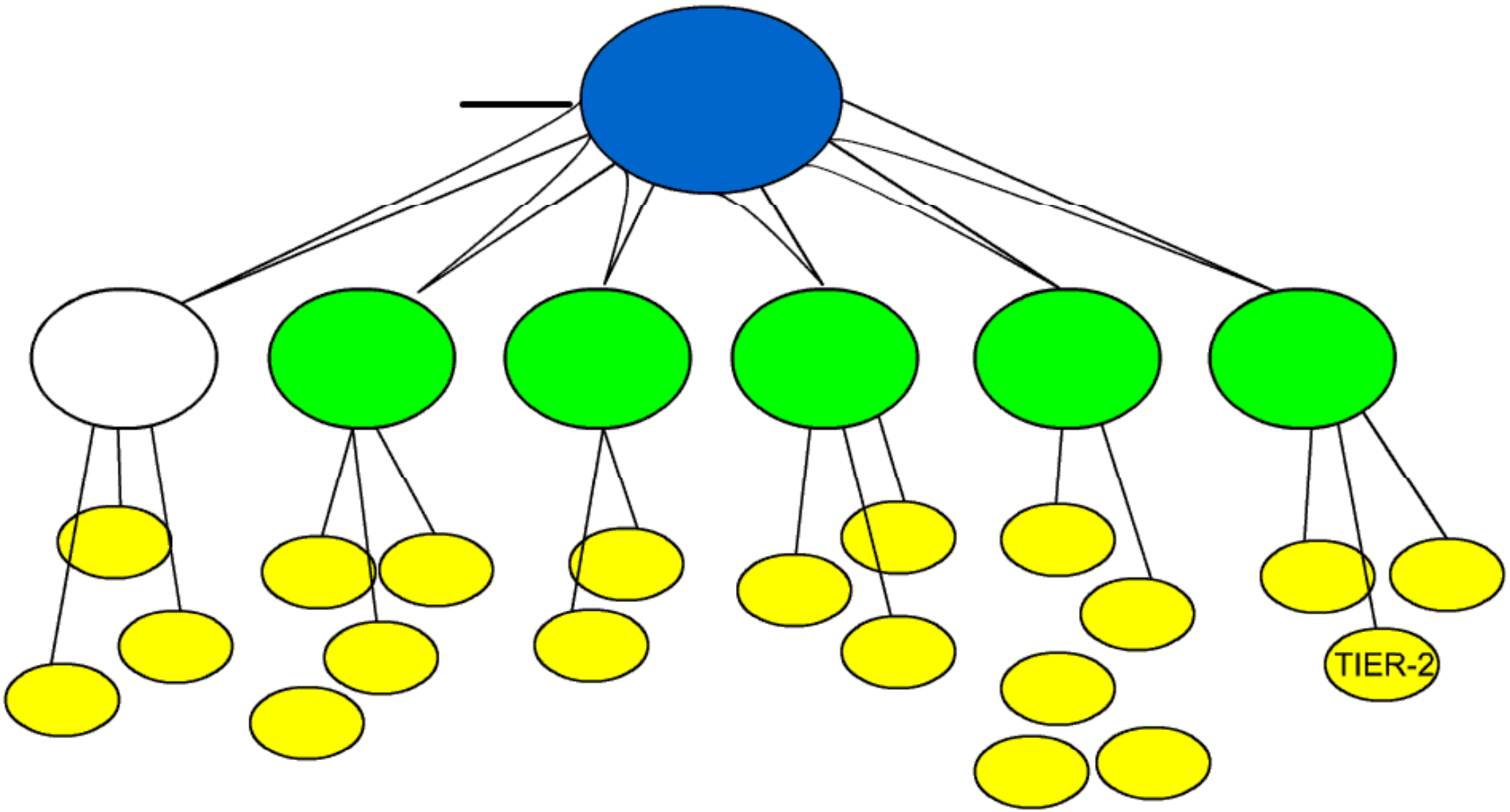


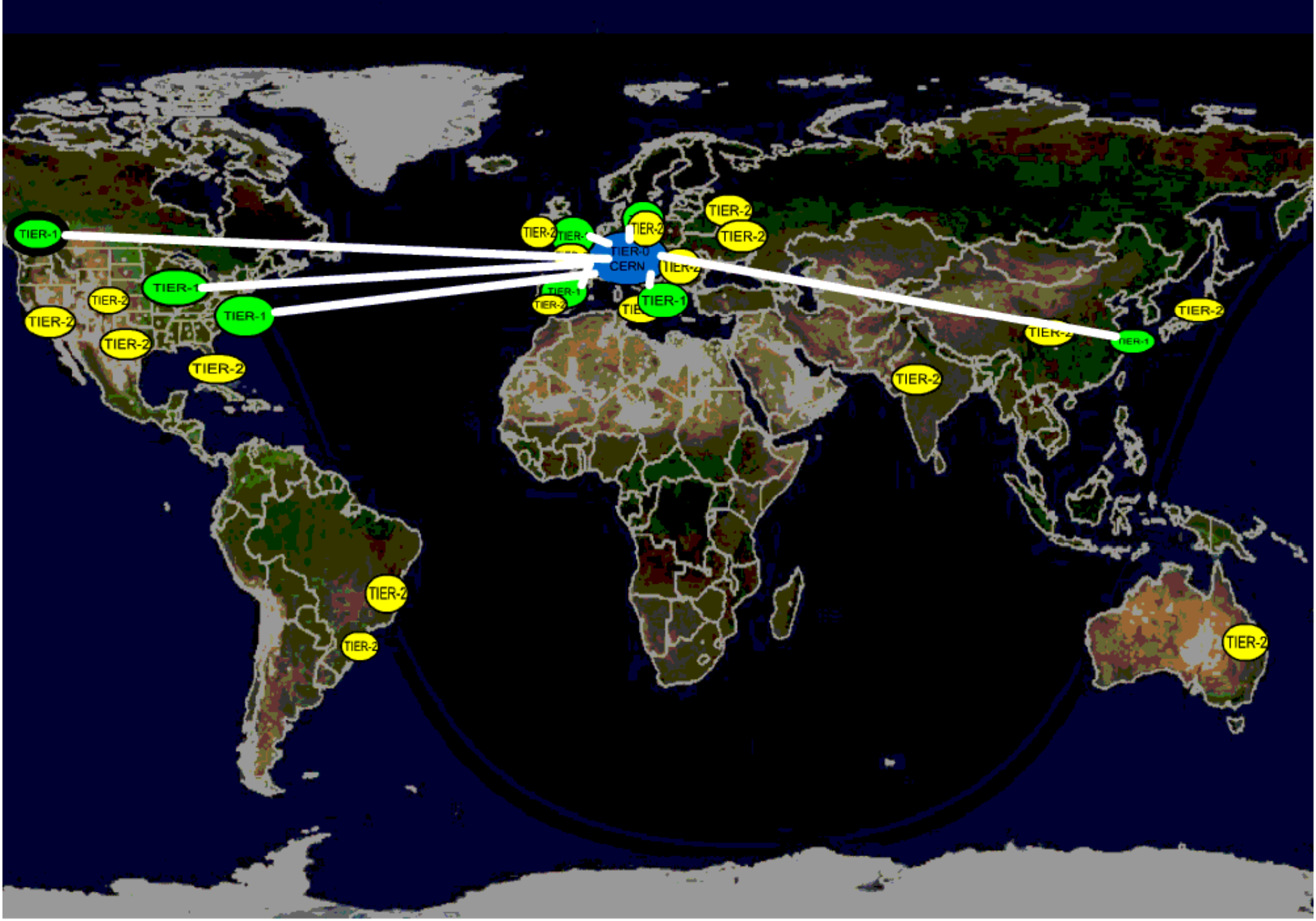


beginning the computing was centralized

eriments began to develop distributed computing models

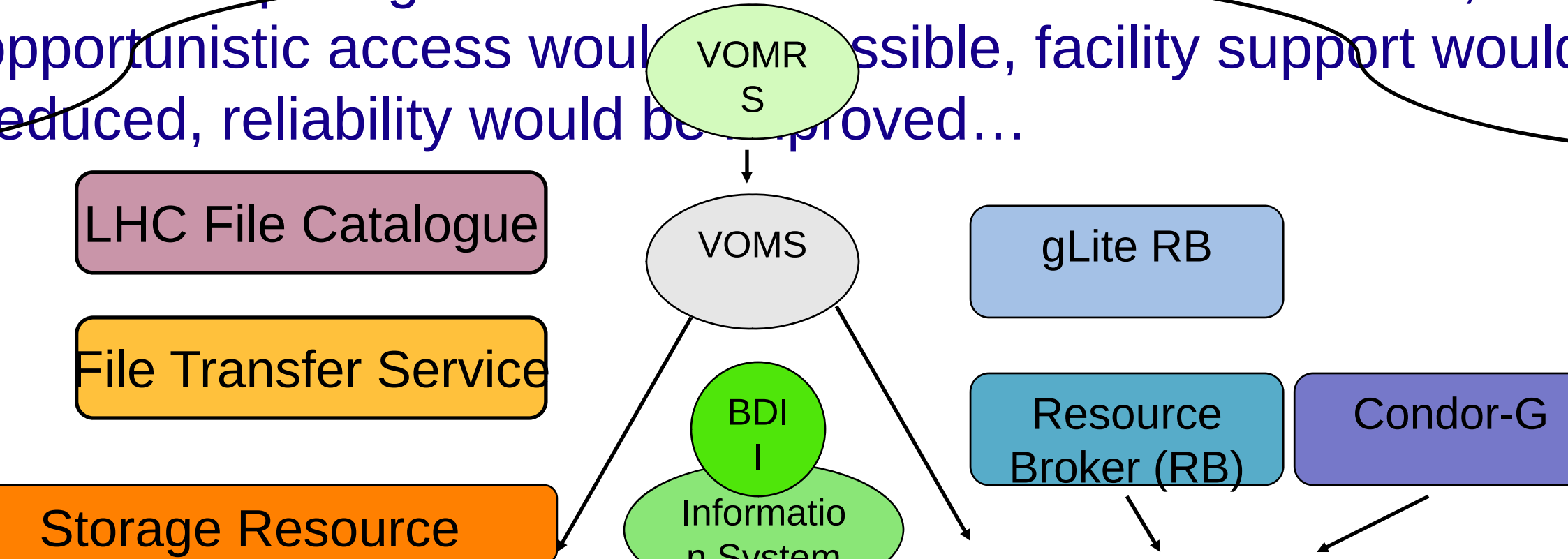
Two examples: Babar had Tier-As that users could connect t
access to the data and resources. CDF had distributed ana
centers


TIER-2



l Private Network (OPN) connects CERN and Tier-1s. Othe

original CDF distributed analysis facility the services, opera
m, storage solution, authentication method, etc. were develo
PhEDEx
DQ2
he services were unique and the e were se
CRAB/ ProdAgent
Dira
Alien
Dirac
with nt of grid ser nterfaces
Ganga/ProdSys
Panda
Alie
protocols was chosen
Entire computing facilities could be shared between VOs,
opportunistic access woul ssible, facility support would
educed, reliability would be proved…



evel of distribution and the number of services requires an
nced system to check the health of the globally distributed
em

WLCG has developed a series of Site Availability Monitors (S
ests

Series of automatically submitted and tracked tests

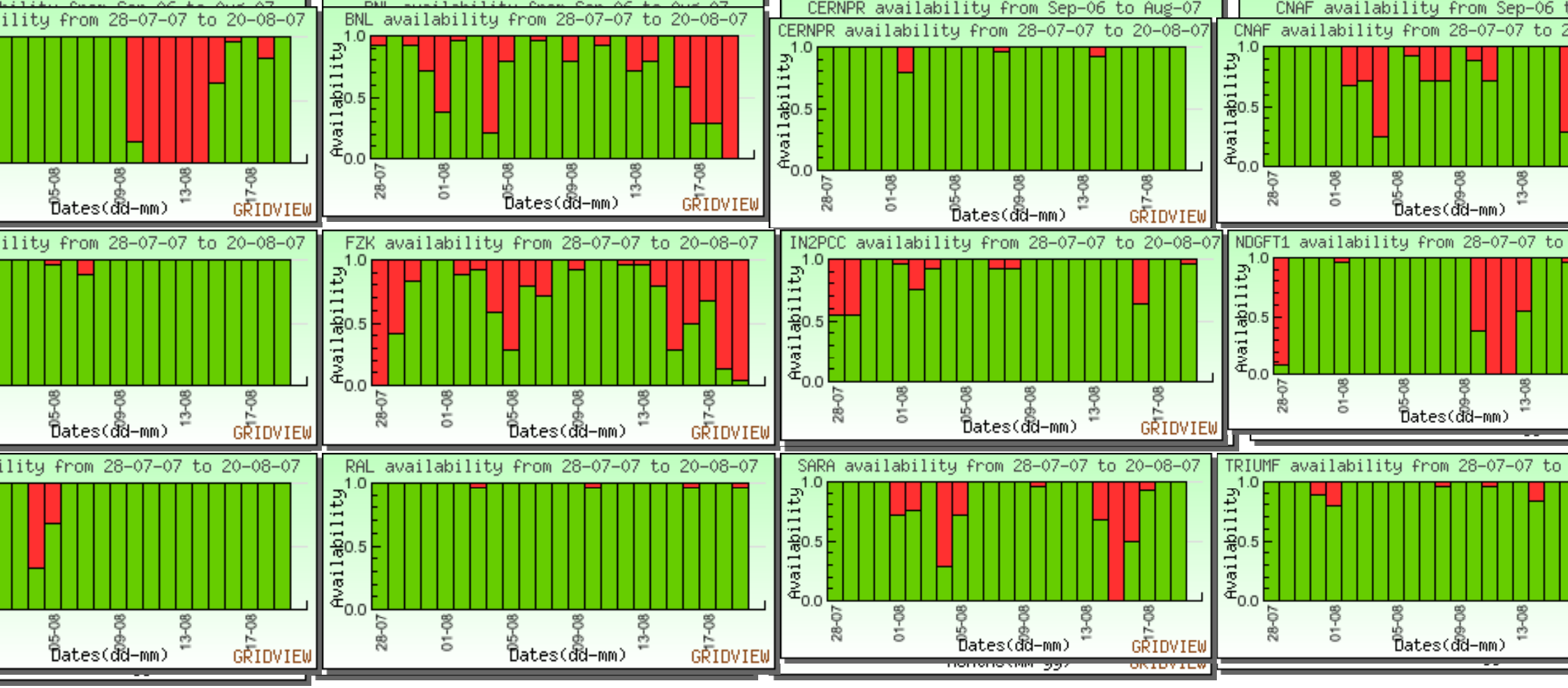
Validate the processing services all the way down to worker
nodes

Validate storage services

nformation systems

Tests run every few hours and results are tracked and publis

vear VOs have begun to introduce their own tests



learly 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.

arities in the data management functionality for the 4 experim

All experiments have services that sit on top of the grid servic
o define the mappings between events, files, and eventually
latasets

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

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 w
s used in Babar, where the storage element namespace is u
to resolve logical file names to physical files names without a
central service

Experiment data management systems drive the replication o
lata

ools to define datasets tend to be experiment specific becau
unctionality is driven physics requirements and choices for w
e supported

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

- Datasets are more dynamic

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

- Simplifies the definitions and access

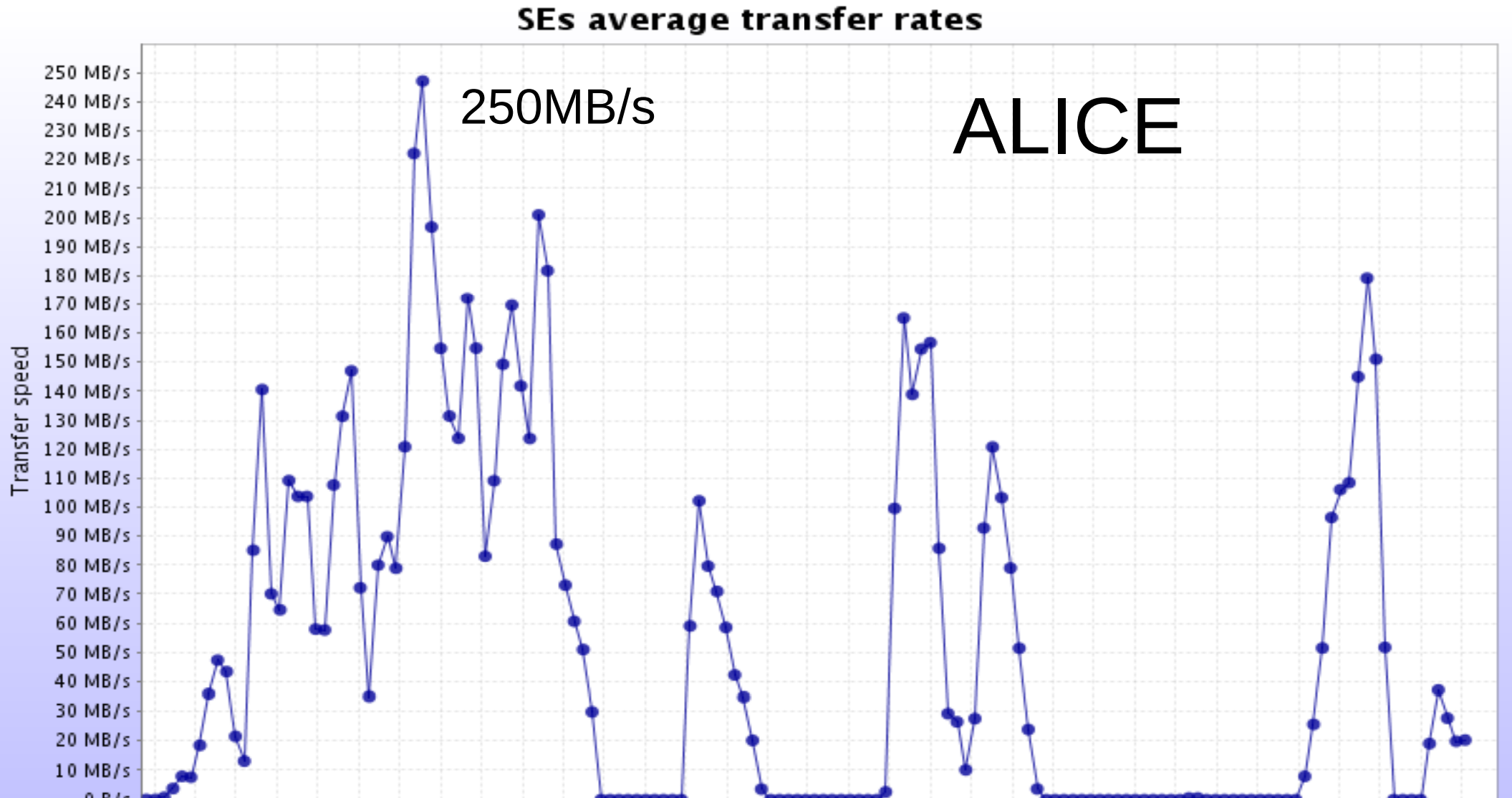
CMS is in-between with datasets being defined and stored in
central bookkeeping service, but operations and users can d
new datasets as needed

ATLAS has a system that allows querying datasets from the

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-1centers, successful demonstration
SRM and FTS



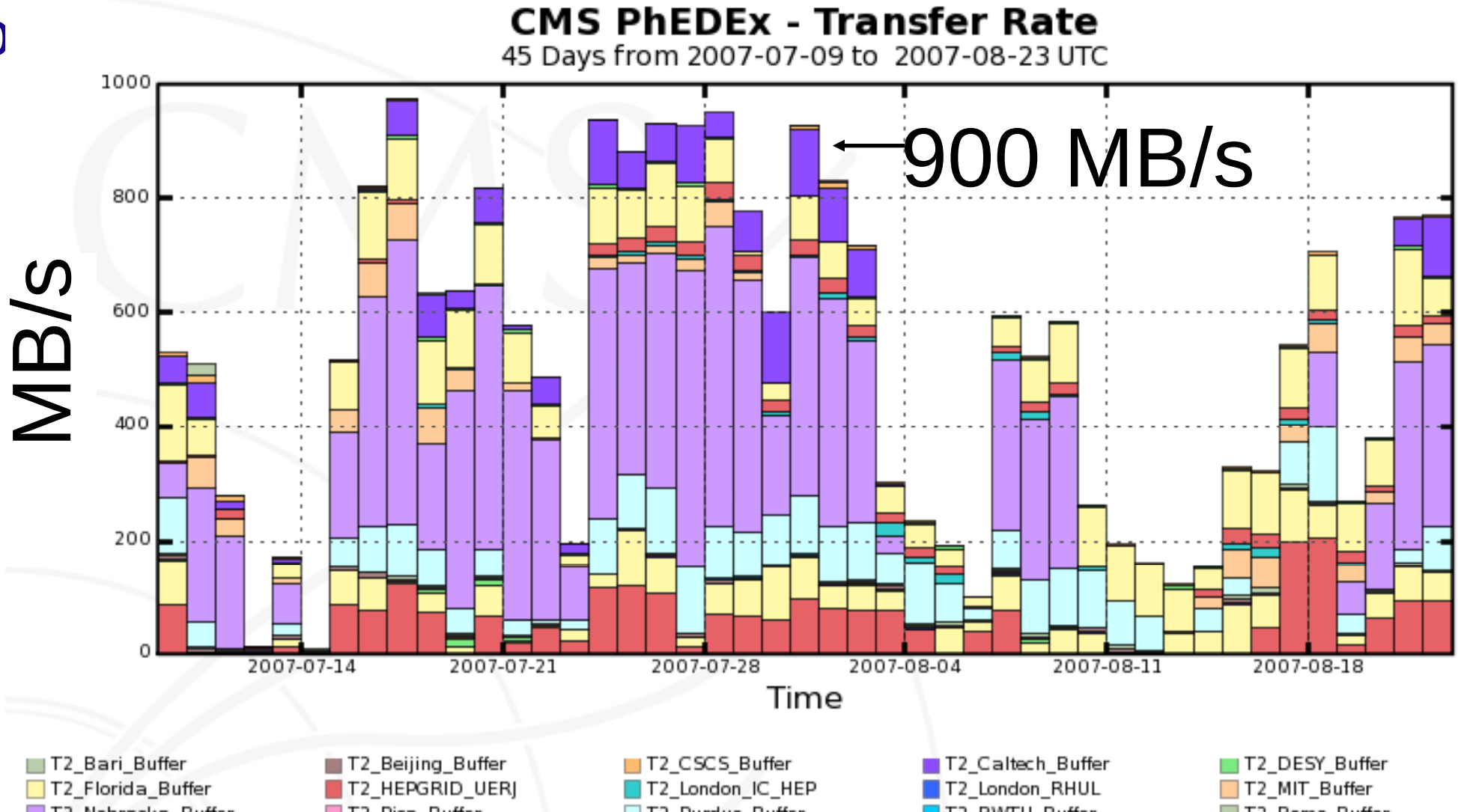
expects Tier-2 storage to be treated like a dynamic cache

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

n 2008 data rates are expected in bursts of 50MB/s-500MB/s

nk

Plo



Access to applications has been a difficult area for LHC
uting 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

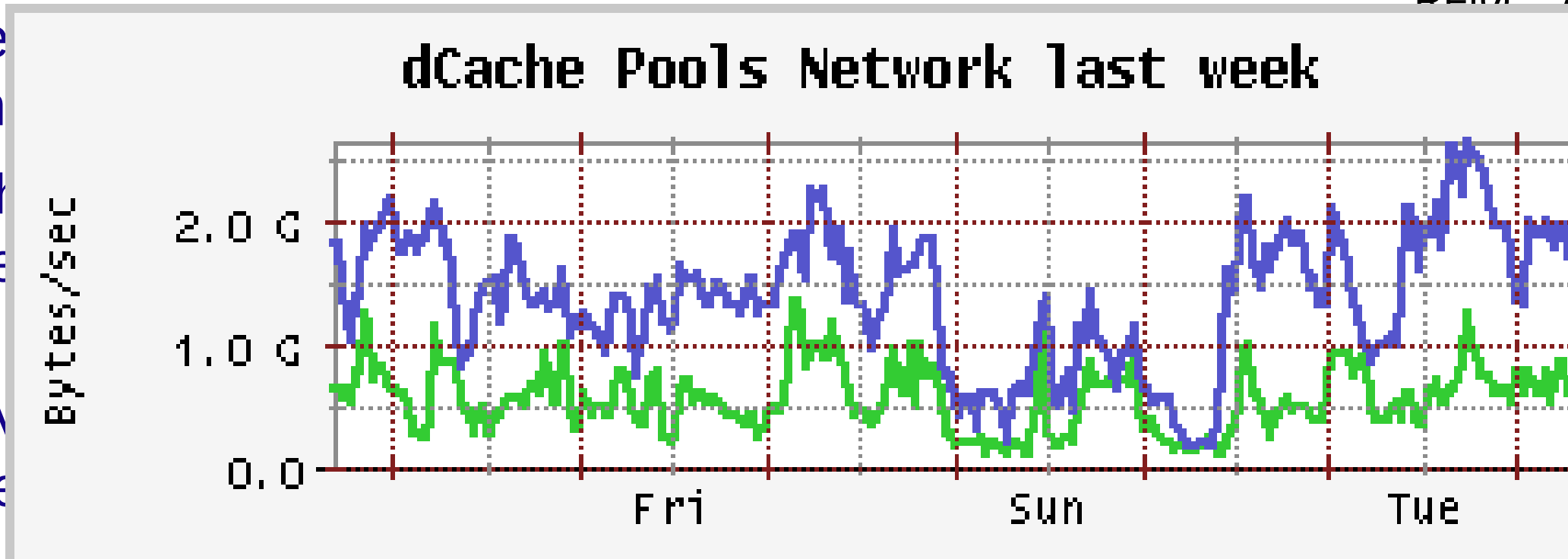
simplest solution, the mass storage system handles data
ng and serving to applications using an efficient local
col (rfio, dcap, xrootd)

dcap
RFIO

n orde
espon

LHCb h
elease
isk

Other V
torage



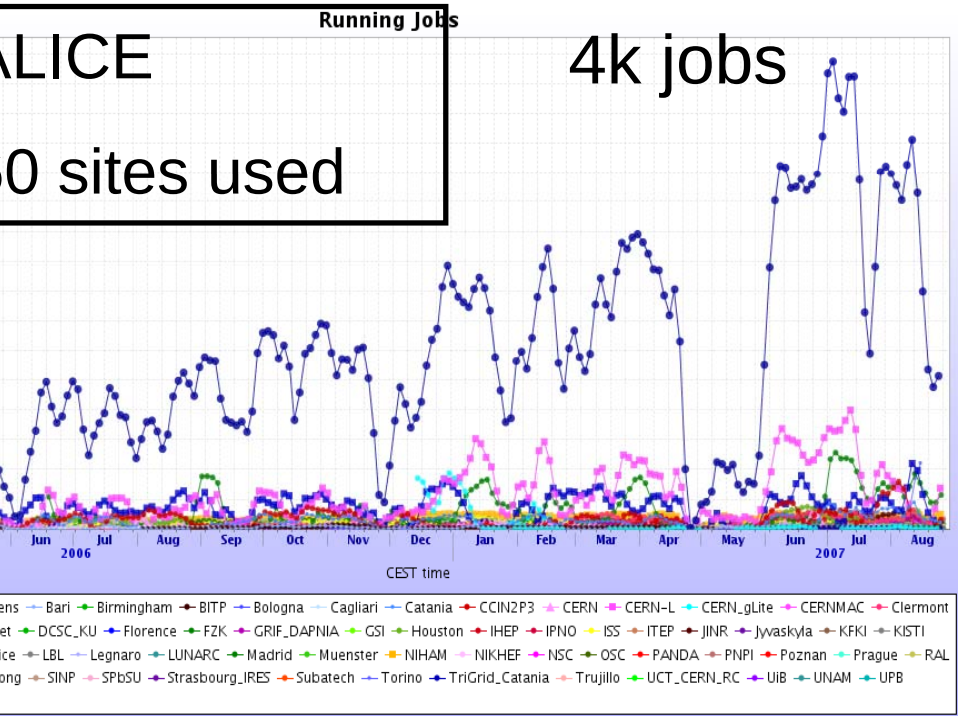
roduction is an ideal candidate for distributed processing

Large output and CPU requirements but small input and predictable applications. All four experiments are succeeding

ALICE

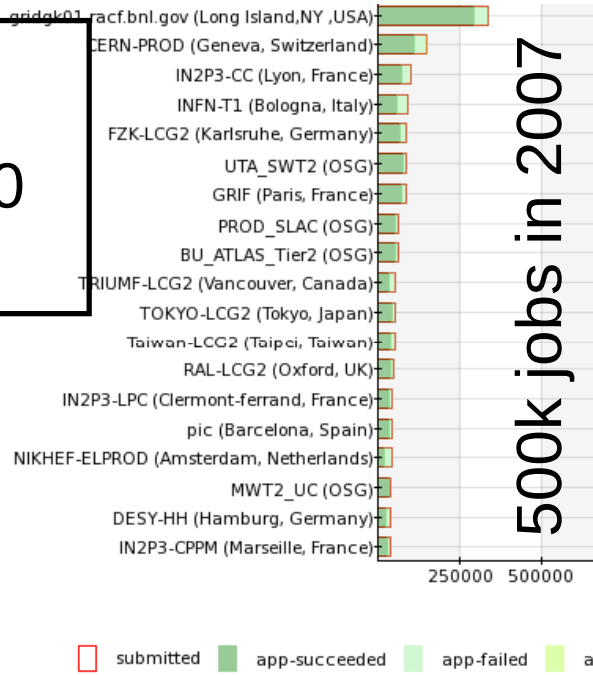
60 sites used

4k jobs



ATLAS

More than 100 sites used

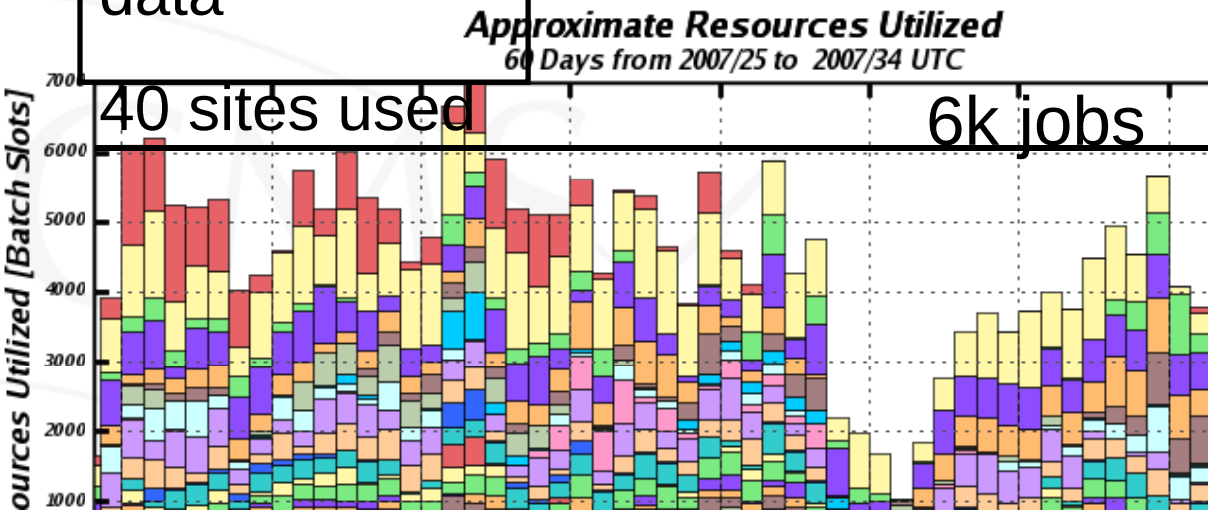


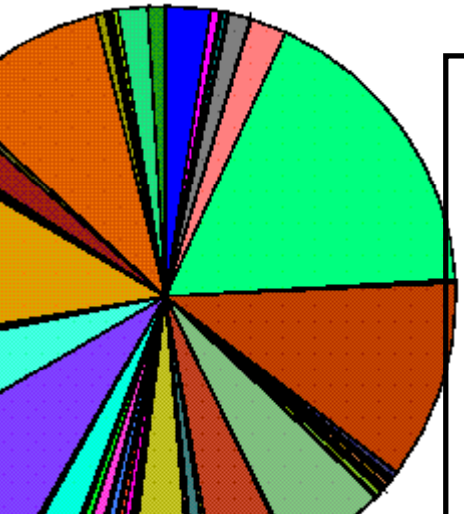
CMS

296M Events

~180TB of data

40 sites used





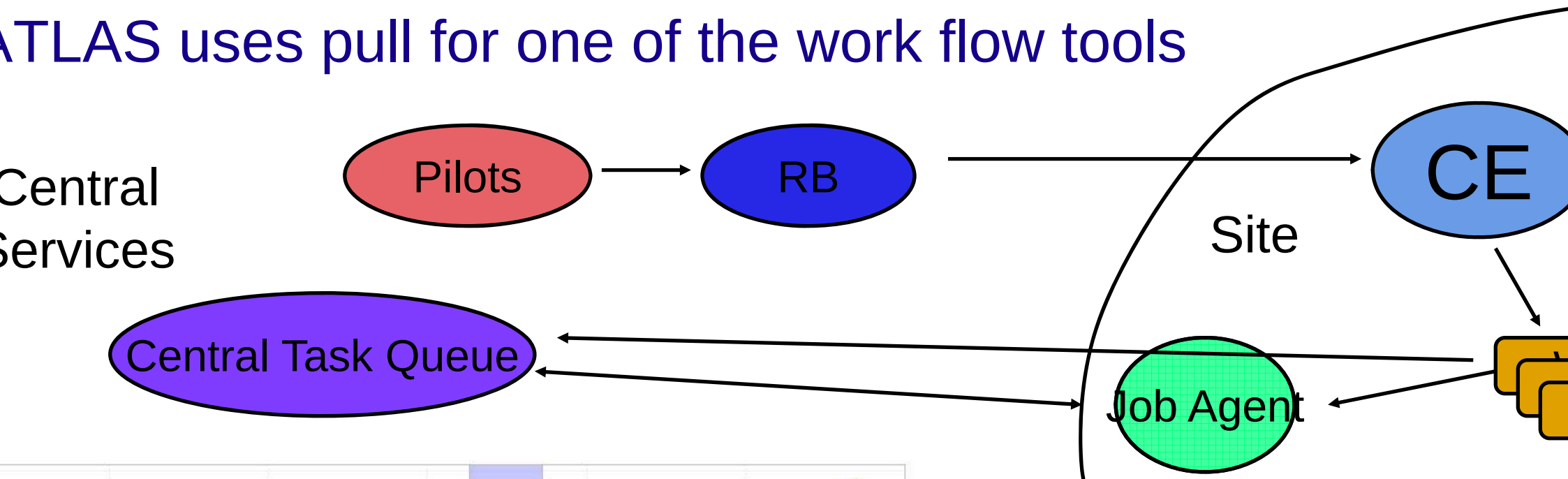
LHCb

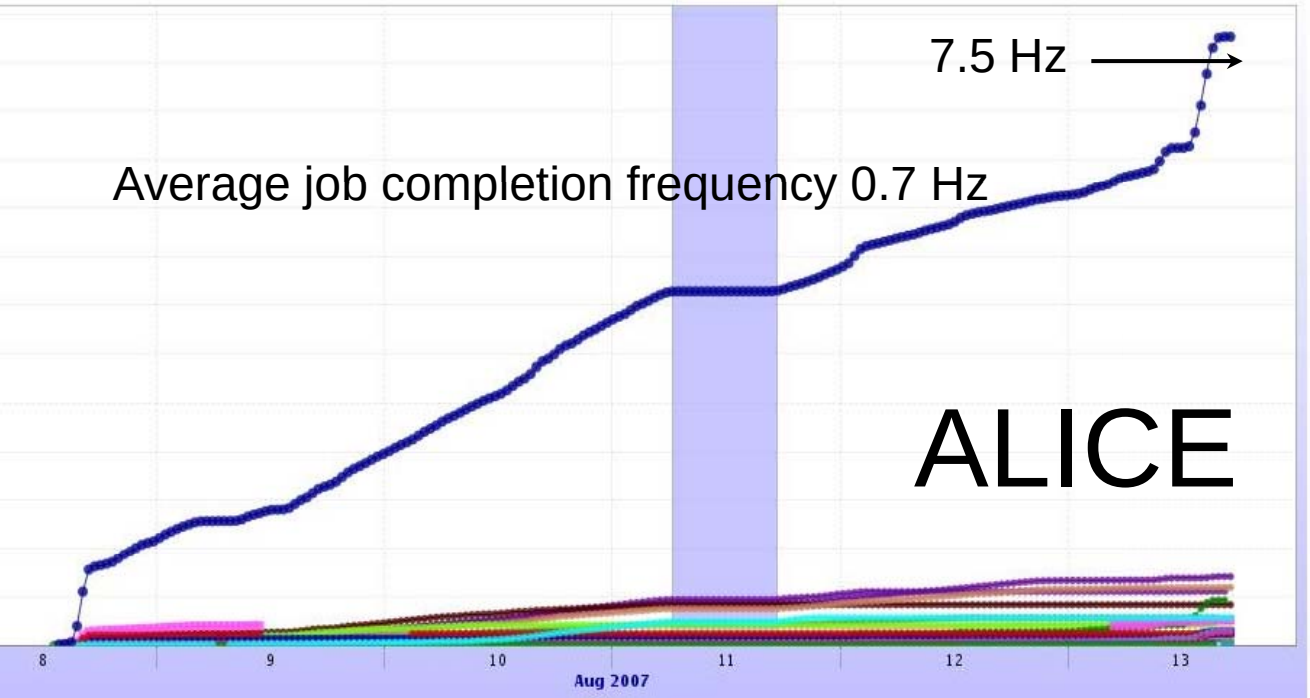
375M Events

~90TB of data produced

ALICE and LHCb have developed pull based job submissior
ems for both Production and Analysis

ATLAS uses pull for one of the work flow tools



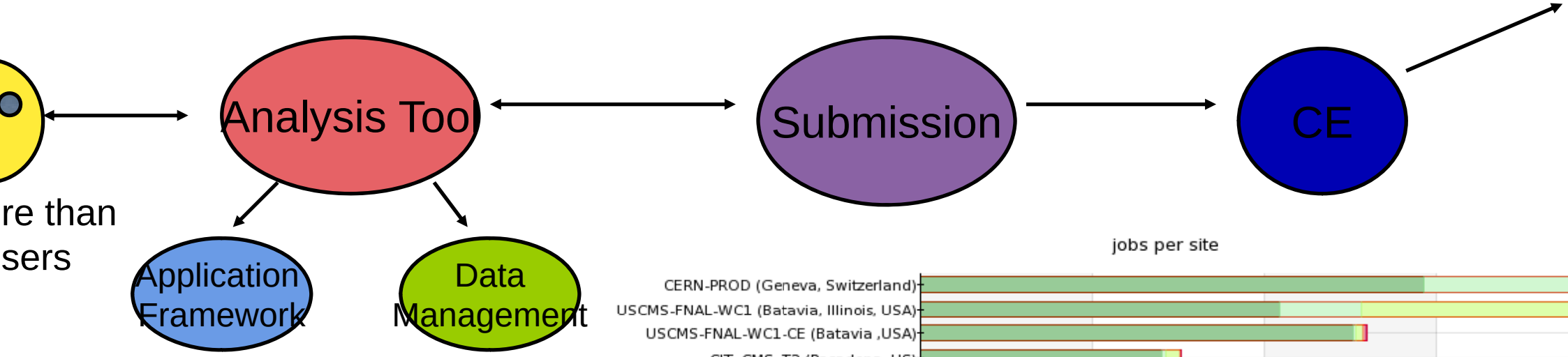


Average rate is about 3
The final rate expected
➔Central queue sca
Central task queue is L
Dirac system reached
Processes

sis processing is more interesting need to match processing resource
arge quantities of data.

Systems used in ATLAS and CMS are similar in the steps

Ganga and Panda in ATLAS and CRAB is CMS



re than
sers

ntly CMS is averaging

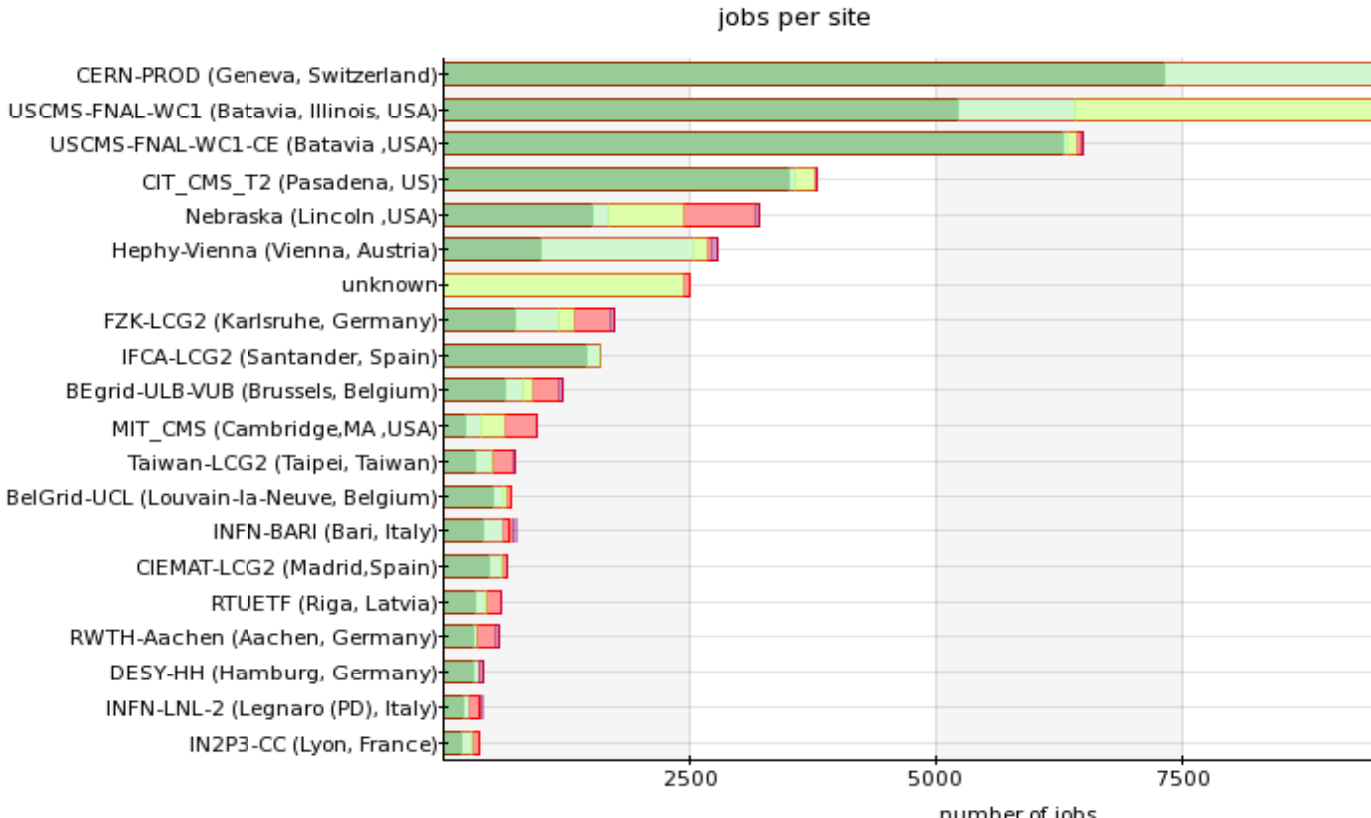
bs per day

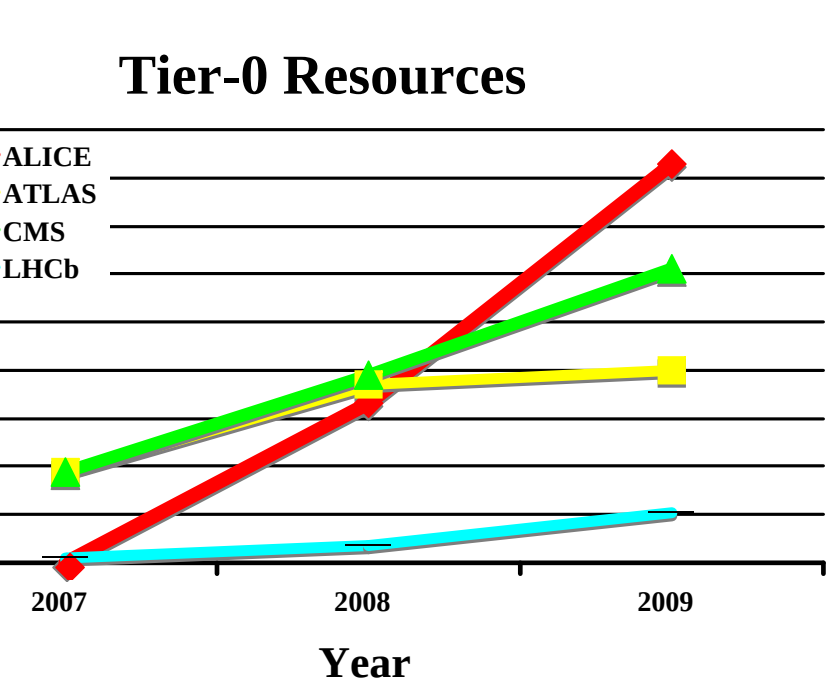
Estimate in 08 is > 100k

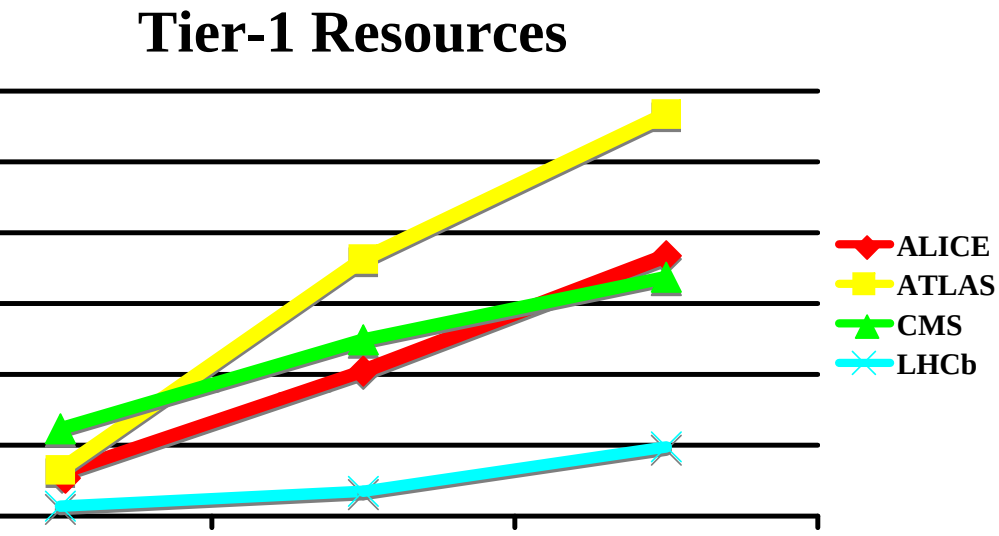
S dashboard reports

r numbers

User scripts outside of





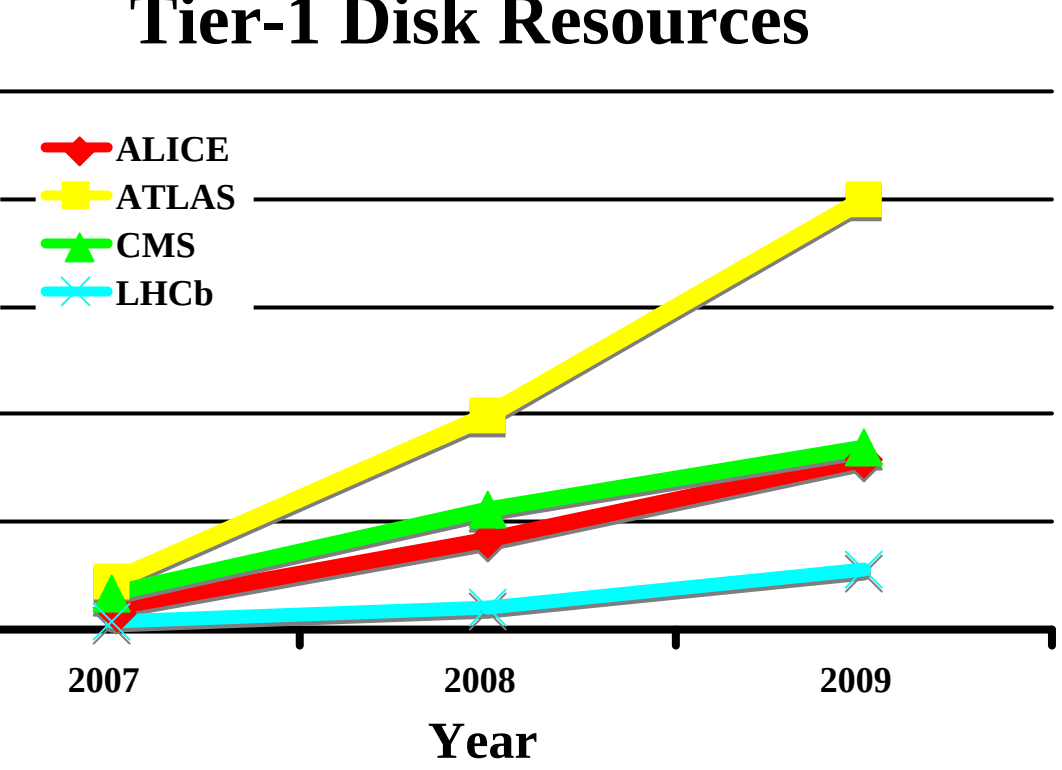


The total quantity of computing resources needs to more than d all the experiments over 2 years

- Some of this can be accoun by Moore's law improvemen
- While large there is experien running farms this large

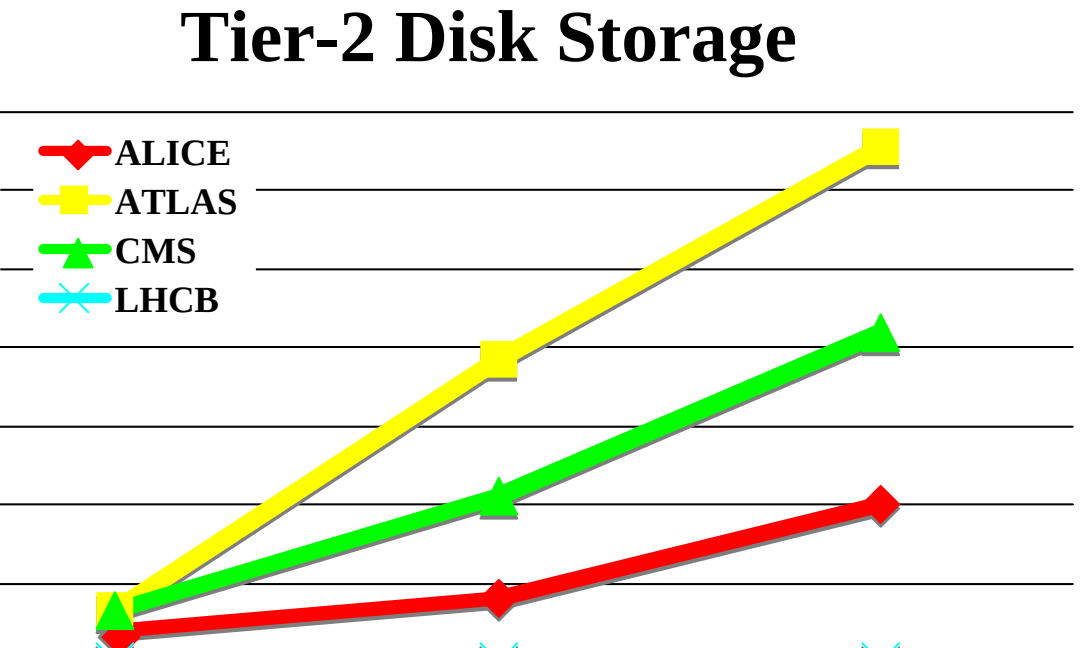
In order to reach the scale requir lot of processes

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



Disk ramp is a little more concerni

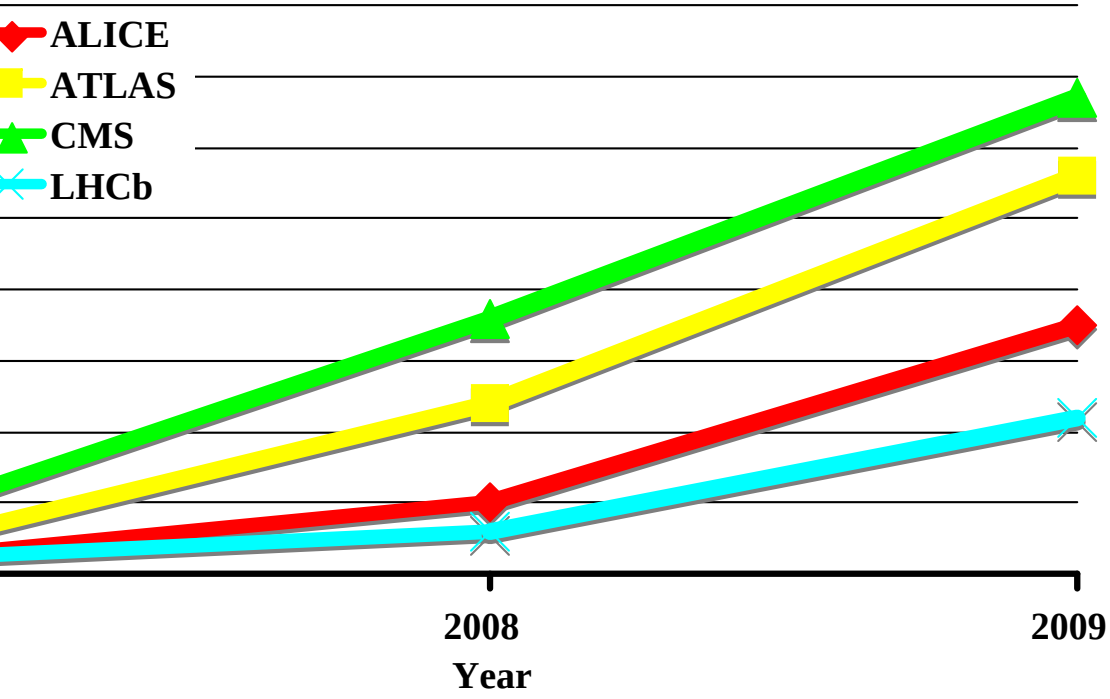
- The required increase canno accommodated by technolog improvements alone.
  - There are a limited number examples of multi-peta byte installations
  - Issues of facility operations scalability of storage name s



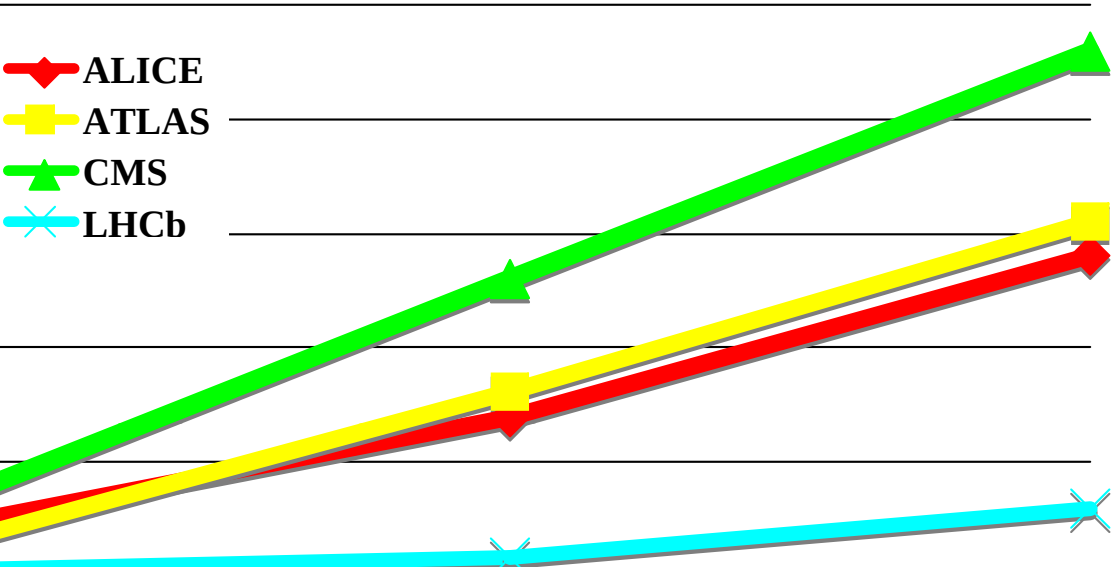
Performance and stability of mass storage is dependent on how it is

- Rely on experiment for reasc file sizes and access rates

Tier-0 Tape Resources



Tier-1 Tape Resources



Tape resources are some of most scalable

- Robotic storage is desig to handle large quantitie data

Also one of the services that requires the longest operatio experience of operate reliab

- Not all Tier-1s are equal experienced

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

- Standard operating mod

experiments have begun demonstrating computing infrastruc
e scale expected to be seen in running conditions

Transfers from CERN

Resources utilized for simulated event production

of work left in the final year of preparation

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

User analysis access needs to ramp up

complicated computing environment and we are still learning
e and operate it