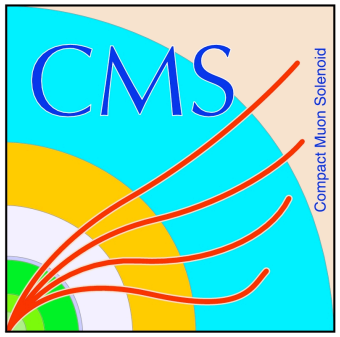
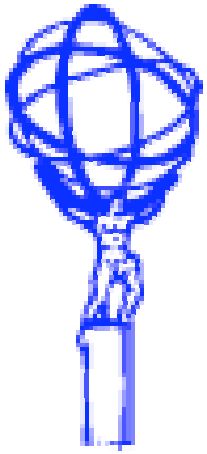
Experience with the WLCG Computing Grid

10 June 2010

Ian Fisk



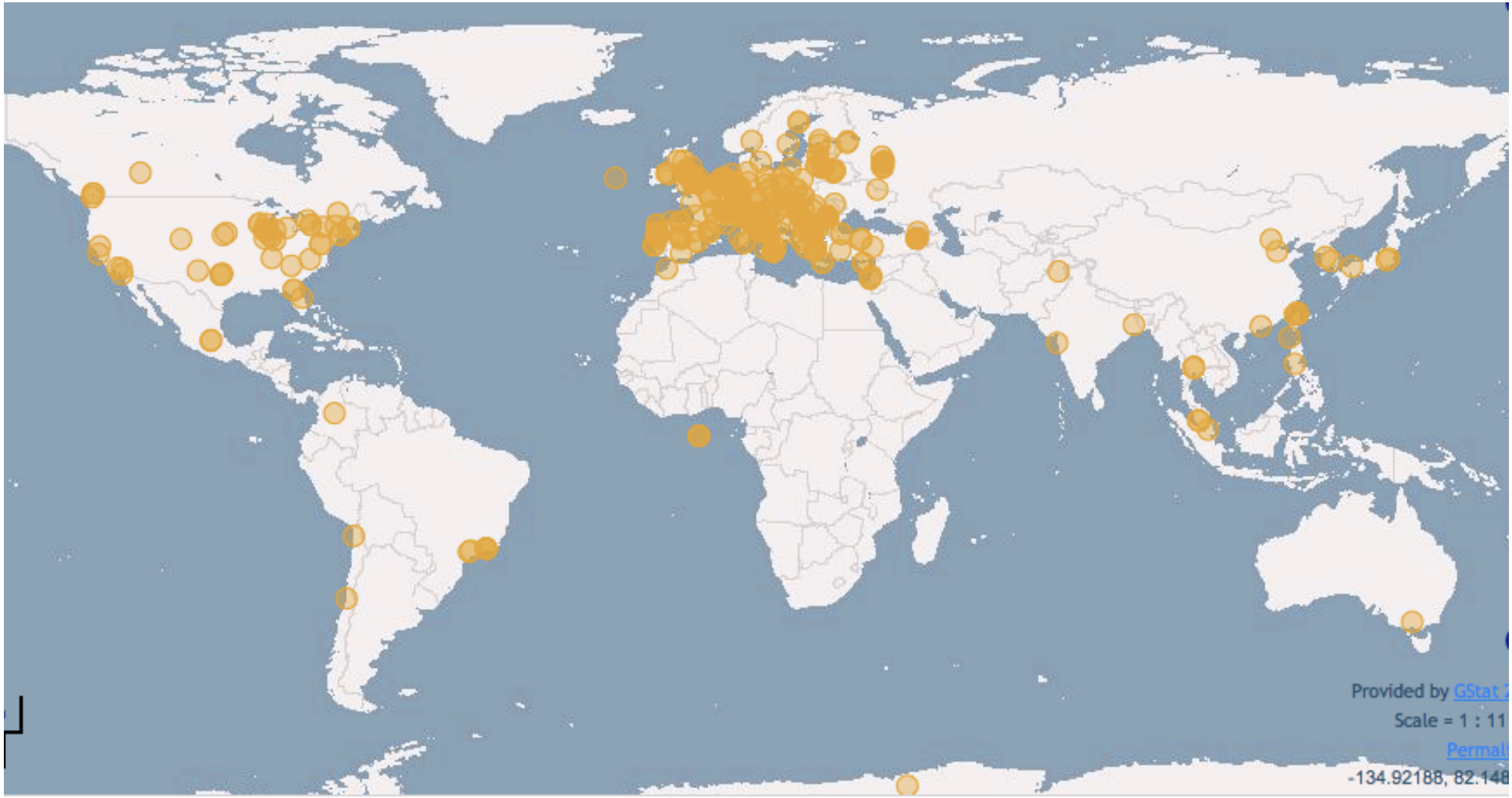


A Little History

- LHC Computing Grid was approved by CERN Council Sept. 20 2001
  - First Grid Deployment Board was Oct. 2002
  - LCG was built on services developed in Europe and the US.
  - LCG has collaborated with a number of Grid Projects
- It evolved into the Worldwide LCG (WLCG)
  - EGEE, NorduGrid, and Open Science Grid
  - Services Support the 4 LHC Experiments

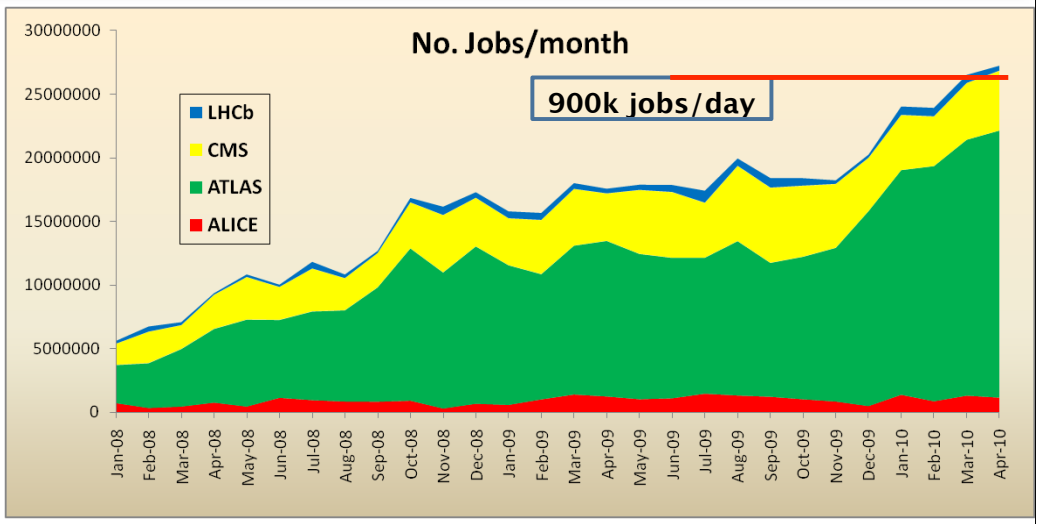
Today's WLCG

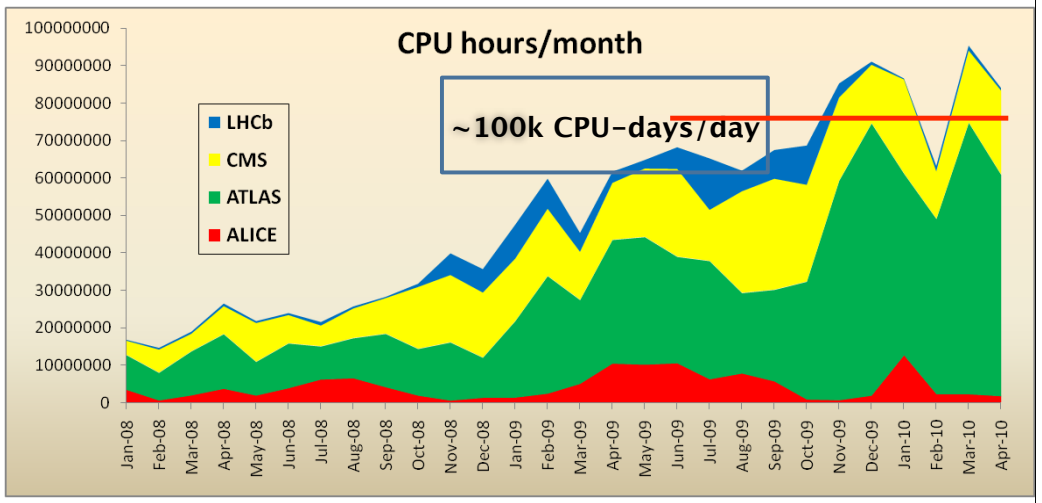
- More than 170 computing facilities in 34 countries
  - More than 100k Processing Cores
  - More than 50PB of disk



Today WLCG is

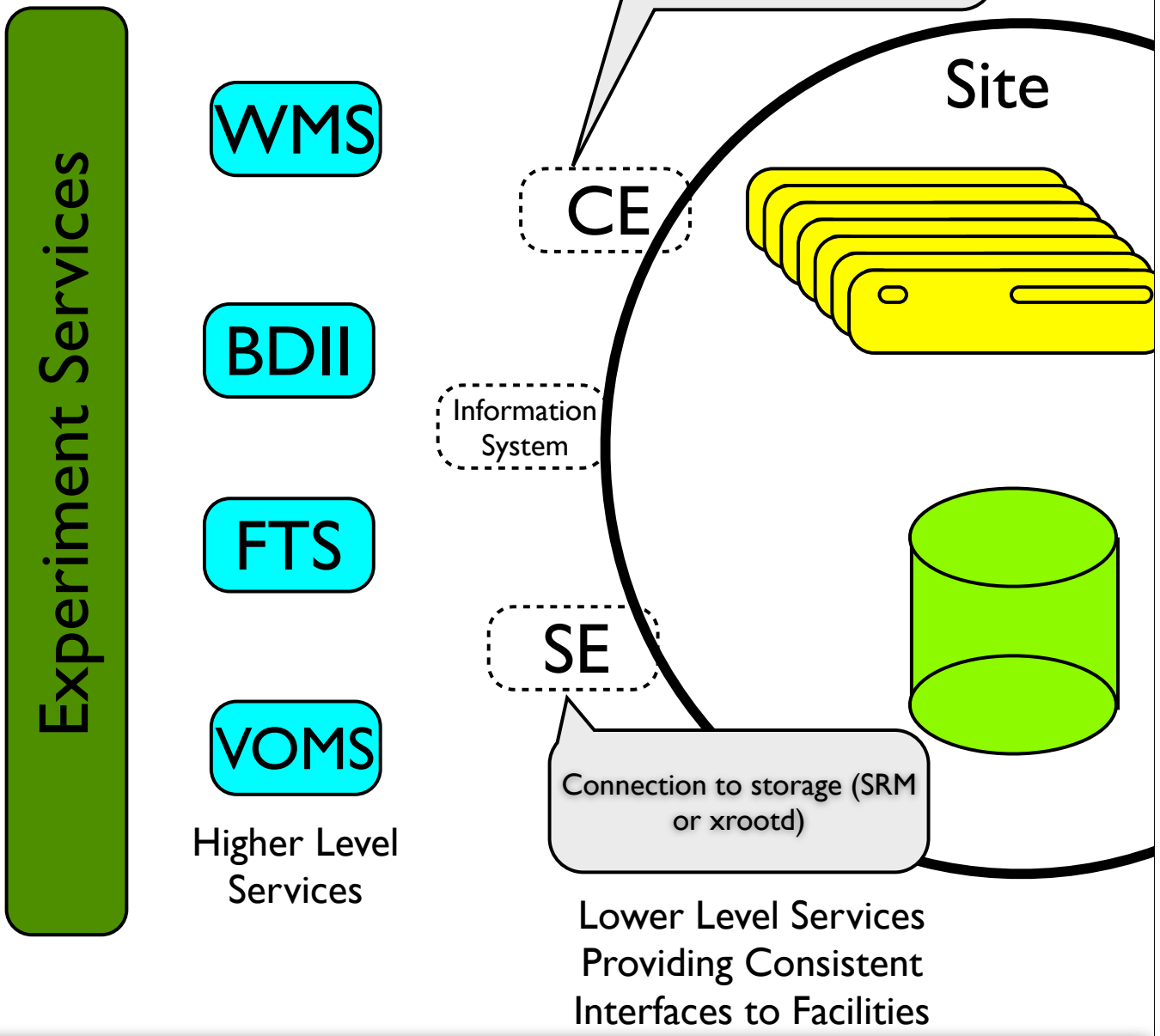
- Running increasingly high workloads:
  - Jobs in excess of 900k / day; Anticipate millions / day soon
  - CPU equiv. ~100k cores
- Workloads are
  - Real data processing !
  - Simulations
  - Analysis – more and more (new) users: several hundreds now
- Data transfers at unprecedented rates





Services

- Basic set of grid services at sites to provide access to processing and storage
- Tools to securely authenticate and manage the membership of the experiment
- Not all experiments use all services or use all services in the same way



Architectures

- To greater and lesser extents LHC Computing model are based on the MONARC model
  - Developed more than a decade ago
  - Foresaw Tiered Computing Facilities to meet the needs of the LHC Experiments

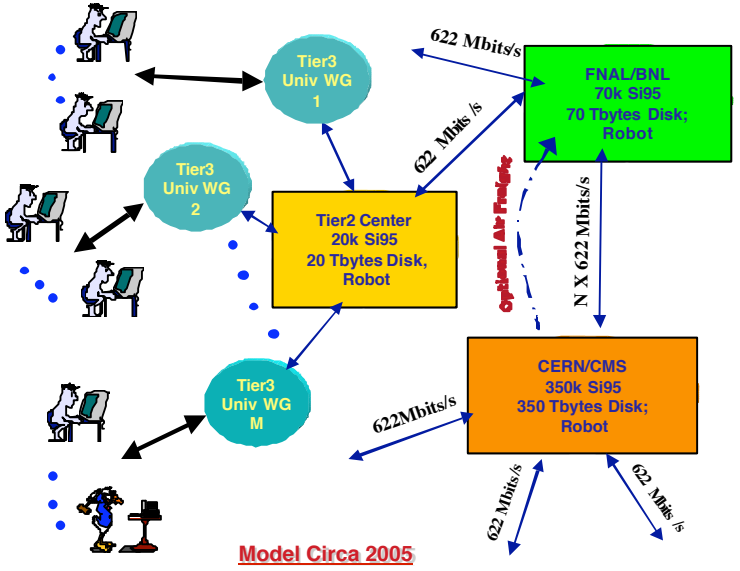
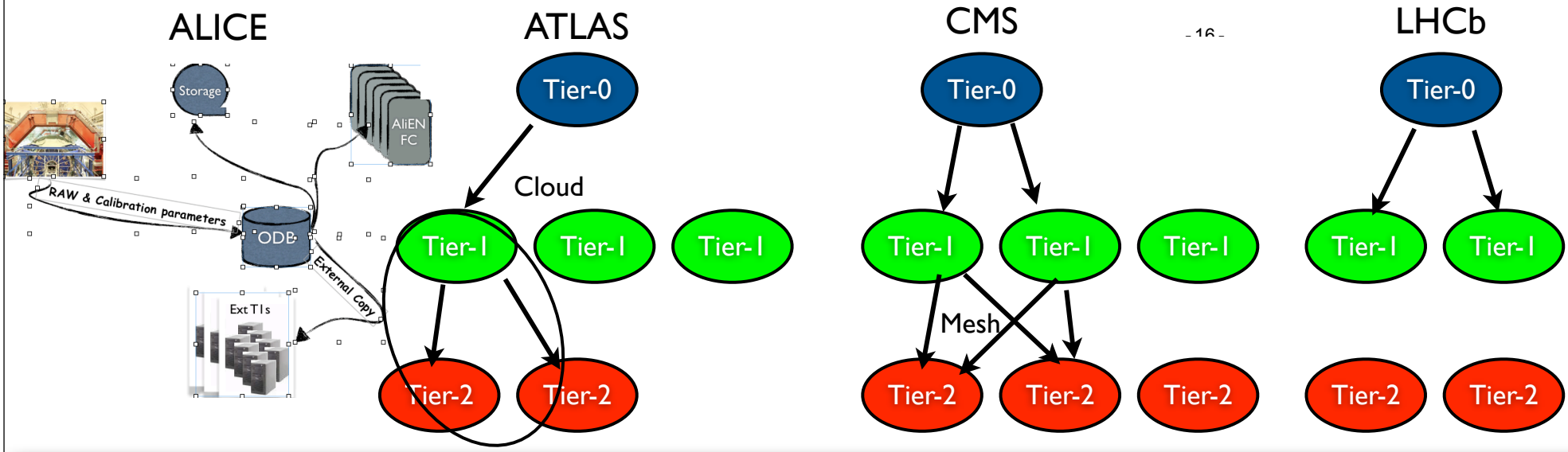
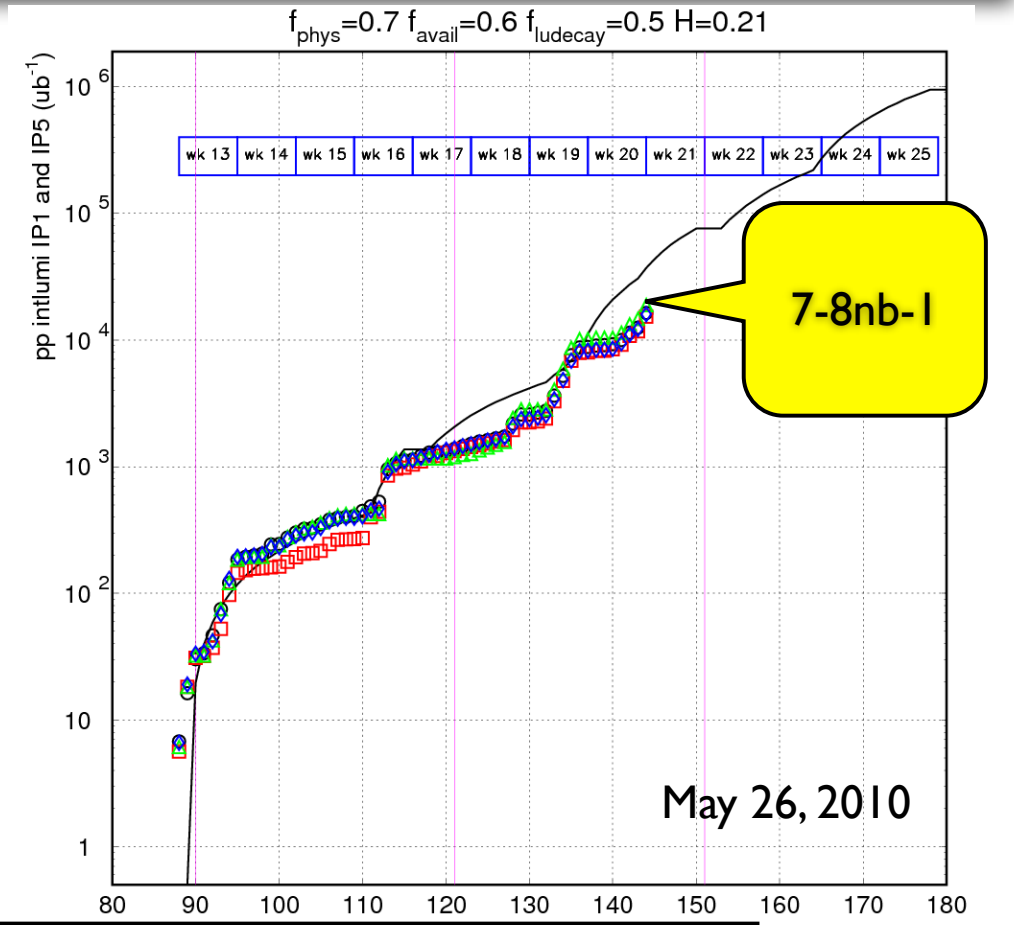


Fig. 4-1 Computing for an LHC Experiment Based on a Hierarchy of Computing Centers. Capacities for CPU and disk are representative and are provided to give an approximate scale).



Data Taking

- An extremely interesting region
  - Exponential Increase means a good weekend can double or triple the dataset
  - A significant failure or outage for a fill would be a big fraction of the total data

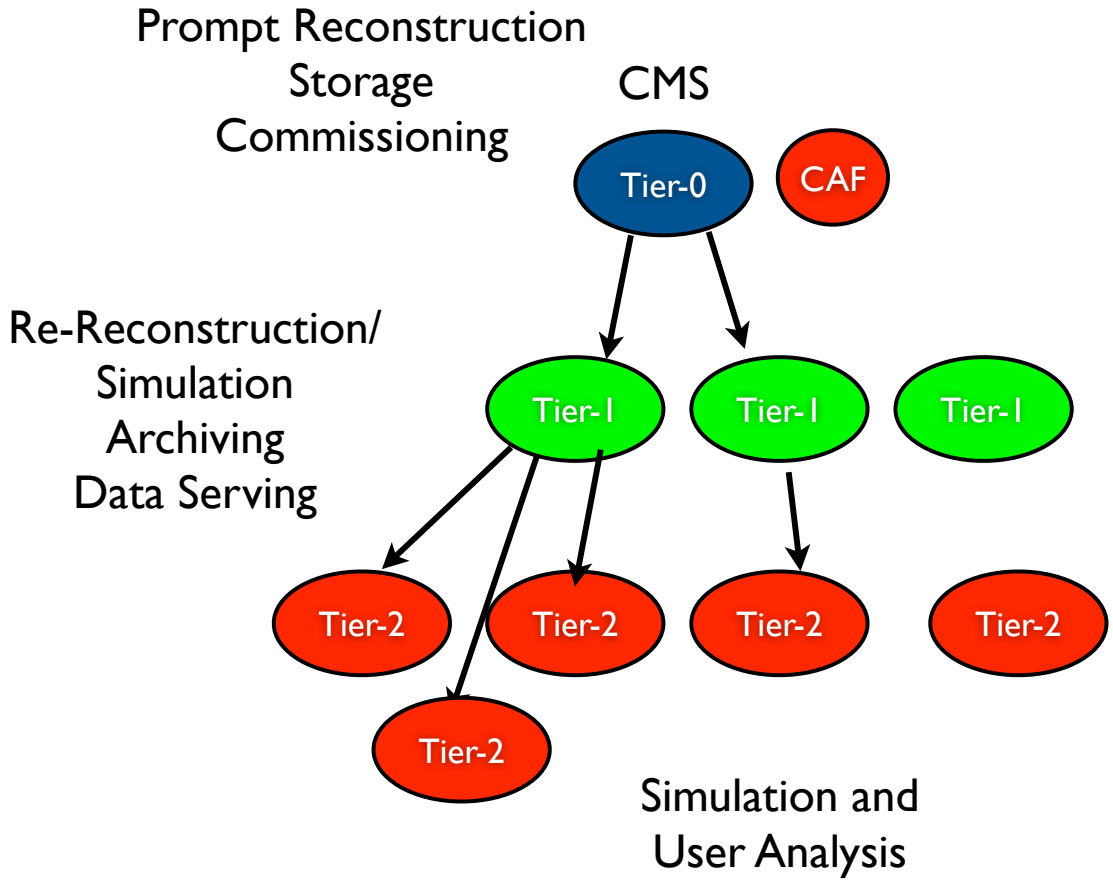


- Original planning for Computing in the first six months had higher data volumes (tens of inverse picobarn)
  - Total volumes of data are not stressing the resources
  - Slower ramp has allowed predicted activities to be performed more frequently

Experiment	Size
ALICE	~190TB
ATLAS	~120TB
CMS	~35TB
LHCb	~20TB

Activities

- Limited volume of data has allowed a higher frequency of workflows
  - Important during the commissioning phase
- All experiments report workflows executed are the type and location predicted in the computing model



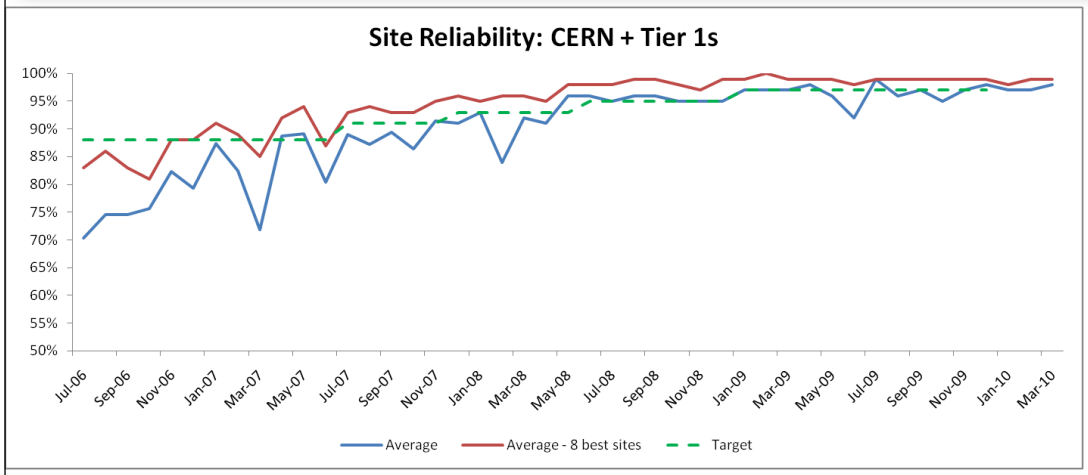
Reliability

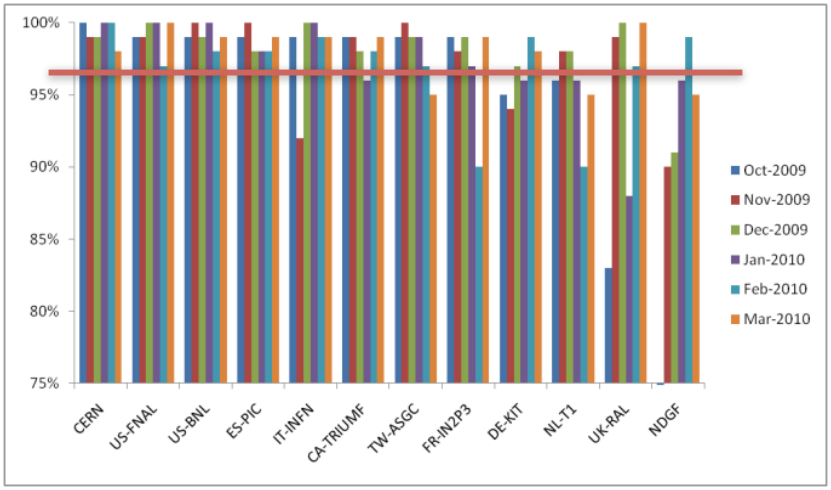
Distributed Computing is when a machine you've never heard of before, half a world away, goes down and stops you from working

"We've made the world's largest Tamagotchi" - Lassi Tuura

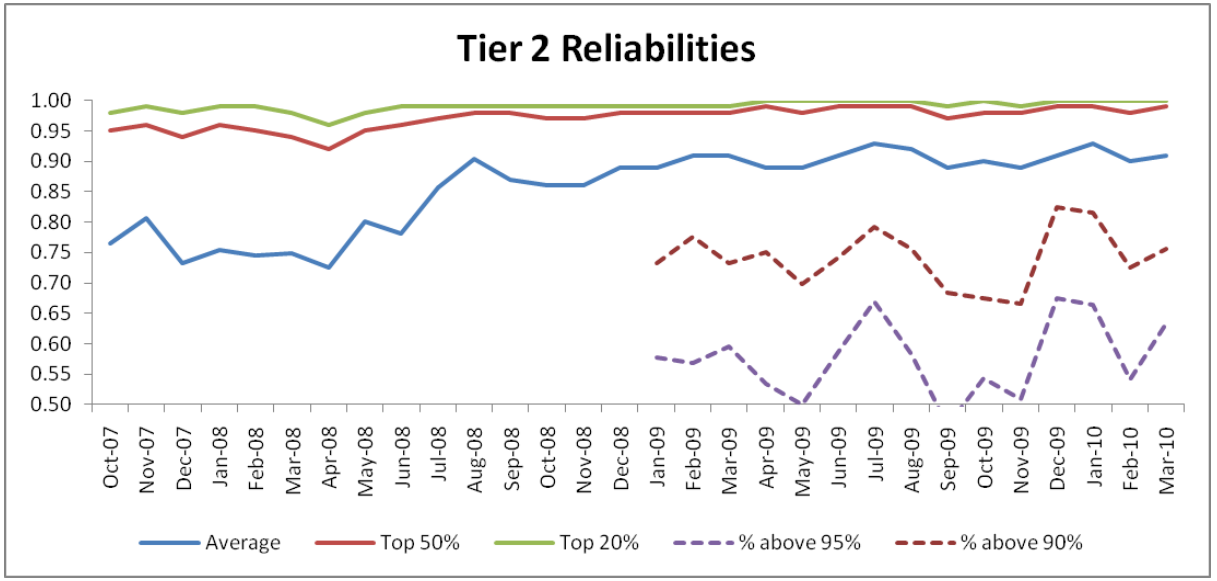
- The vast majority of the Computing Resources for the LHC are away from CERN.
  - WLCG Services are carefully monitored to ensure access to resources
  - Much of the effort in the last few years has been in improving operations

Reliability

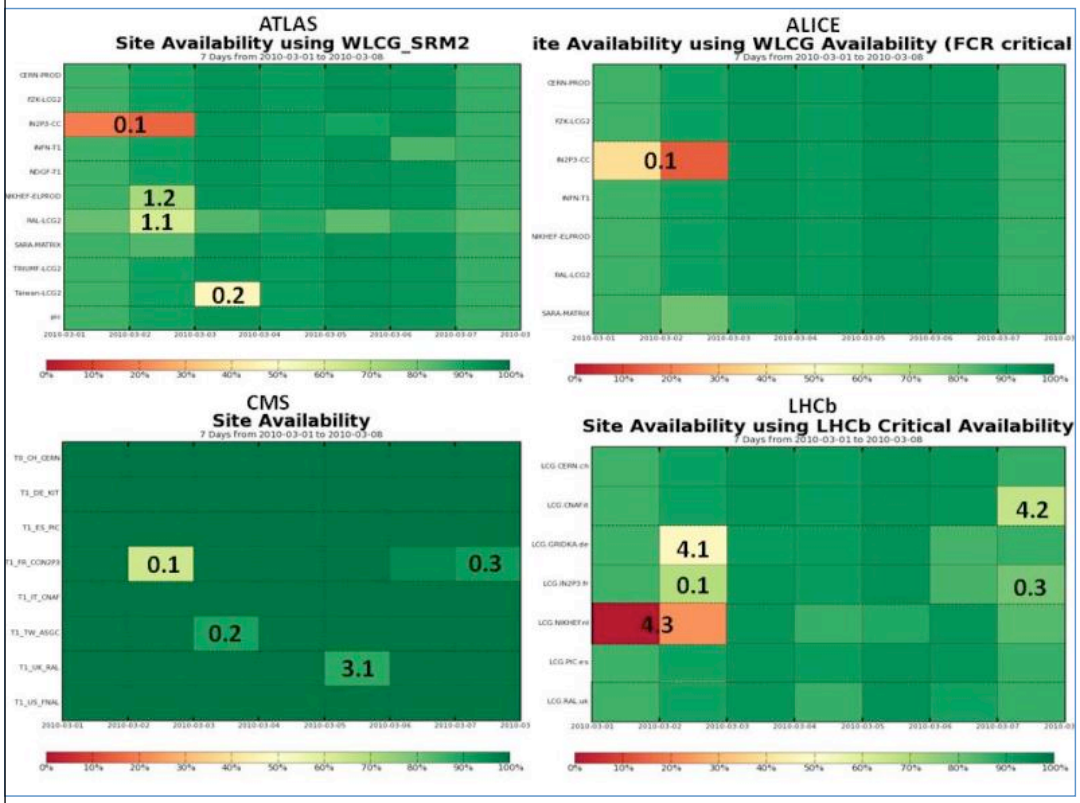


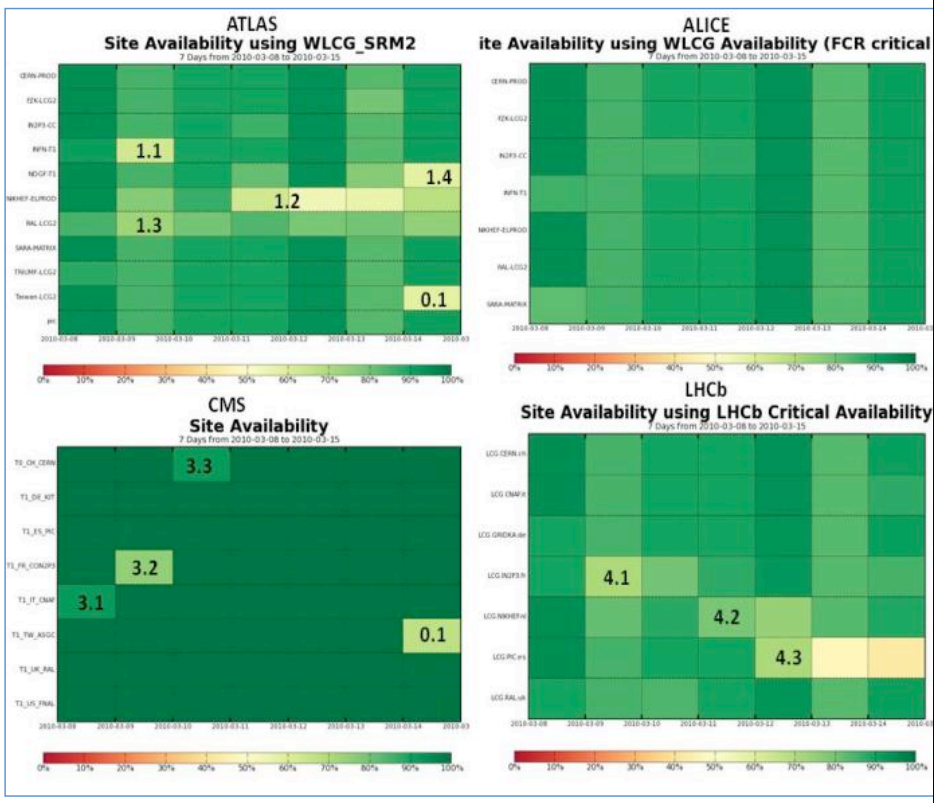


- Monitoring of basic WLCG Services
  - Clear improvement in preparation for data taking



Readiness as Seen by the Experiments

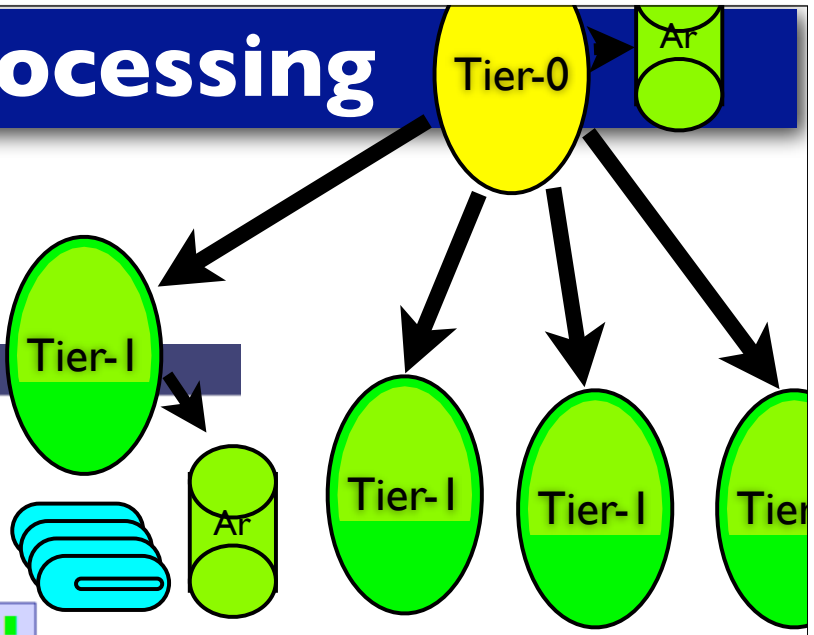




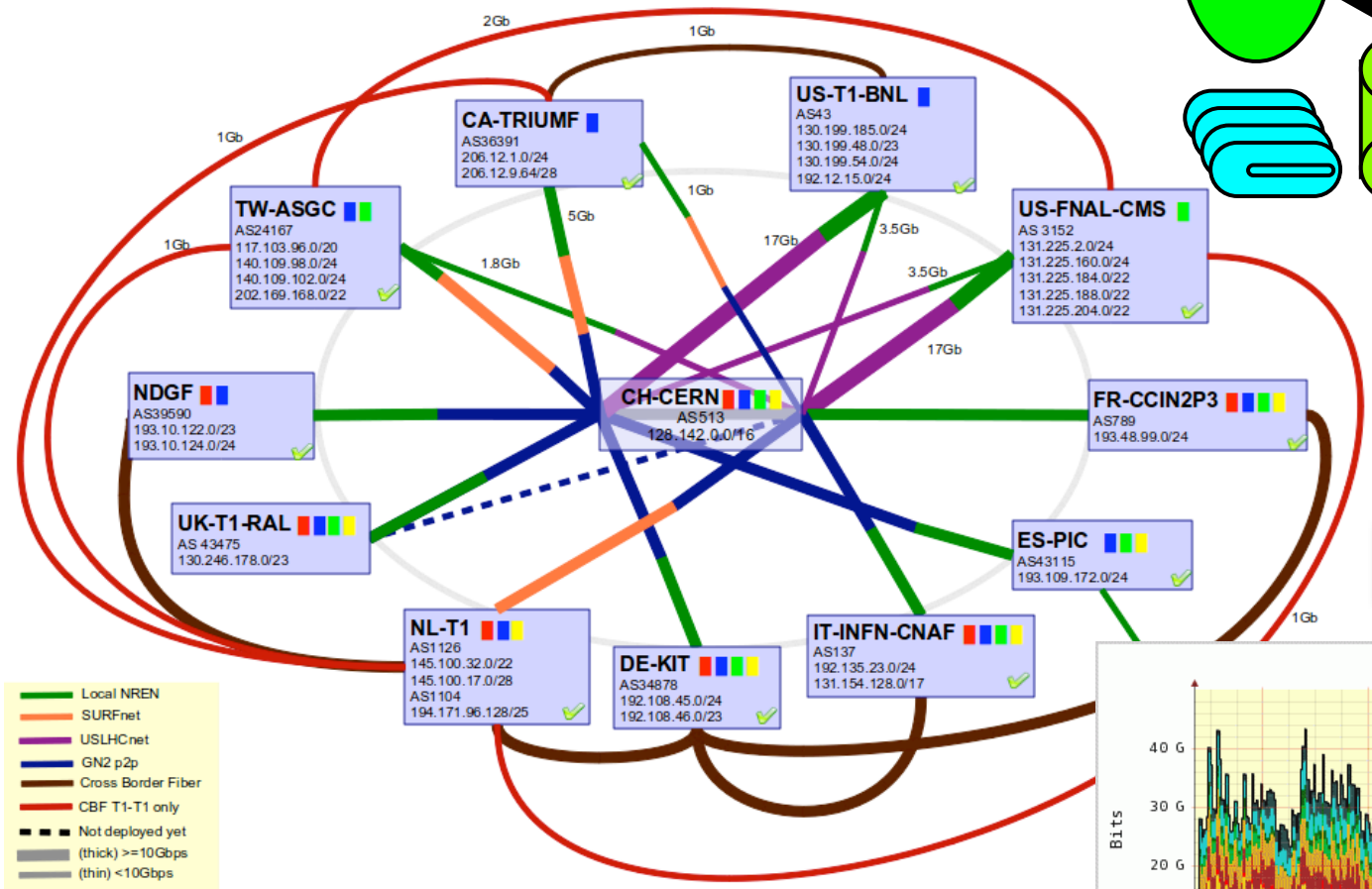
- Site readiness as seen by the experiments
  - LH week before data taking; RH 1st week of data
  - Experiment tests include specific workflows

Data Transfer and ReProcessing

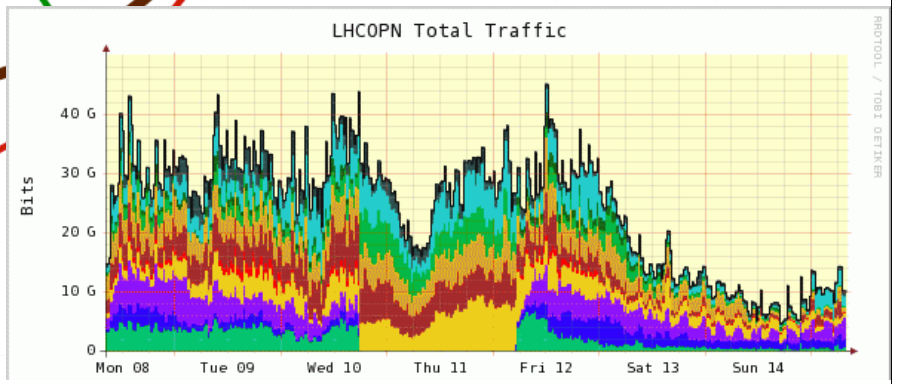
OPN links now fully redundant – last one – RAL – now in production



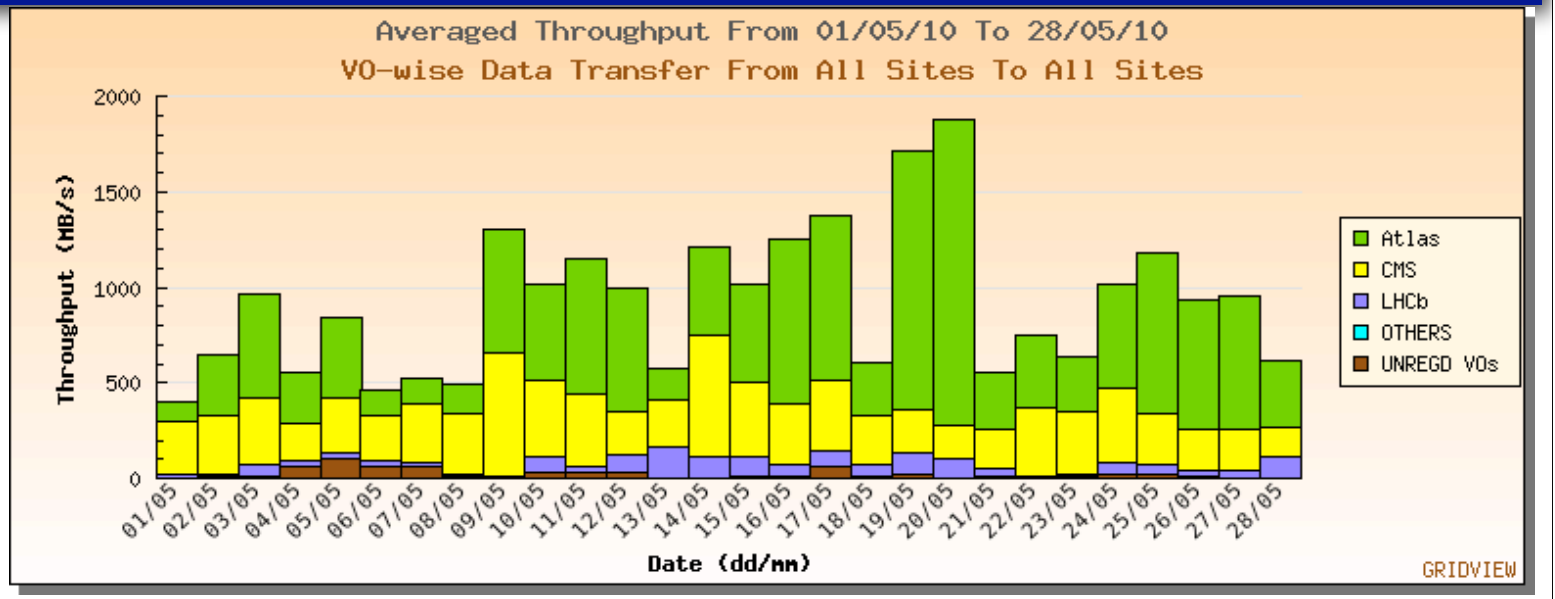
LHCOPN – current status



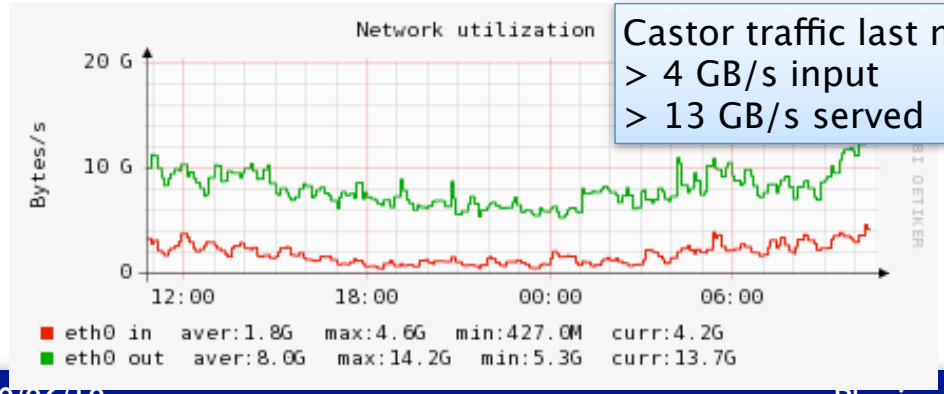
Fibre cut during STEP'09: Redundancy meant no interruption



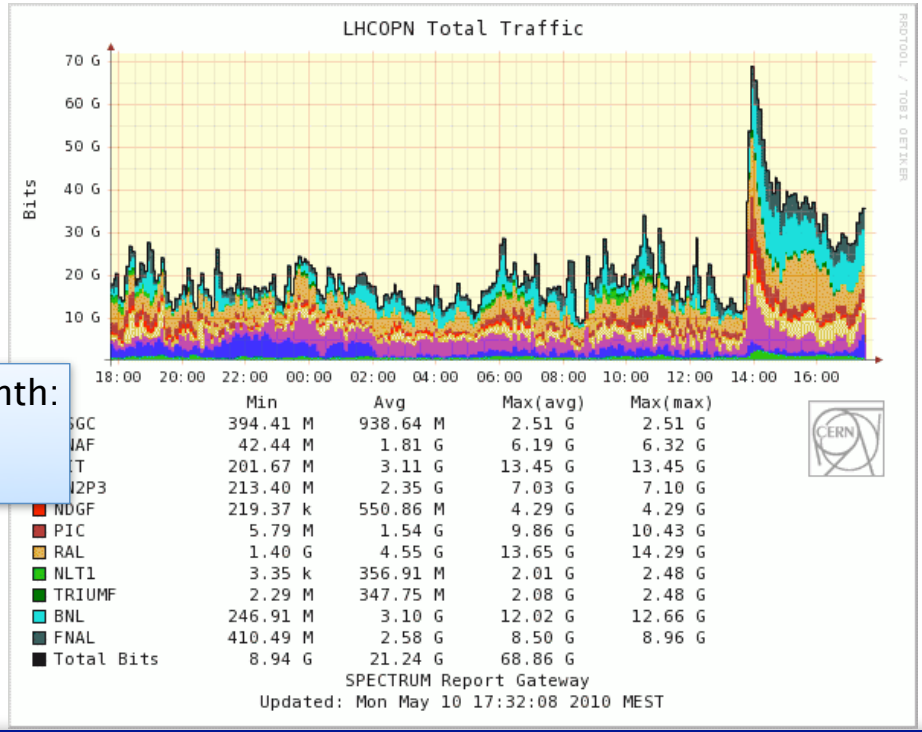
Transfers



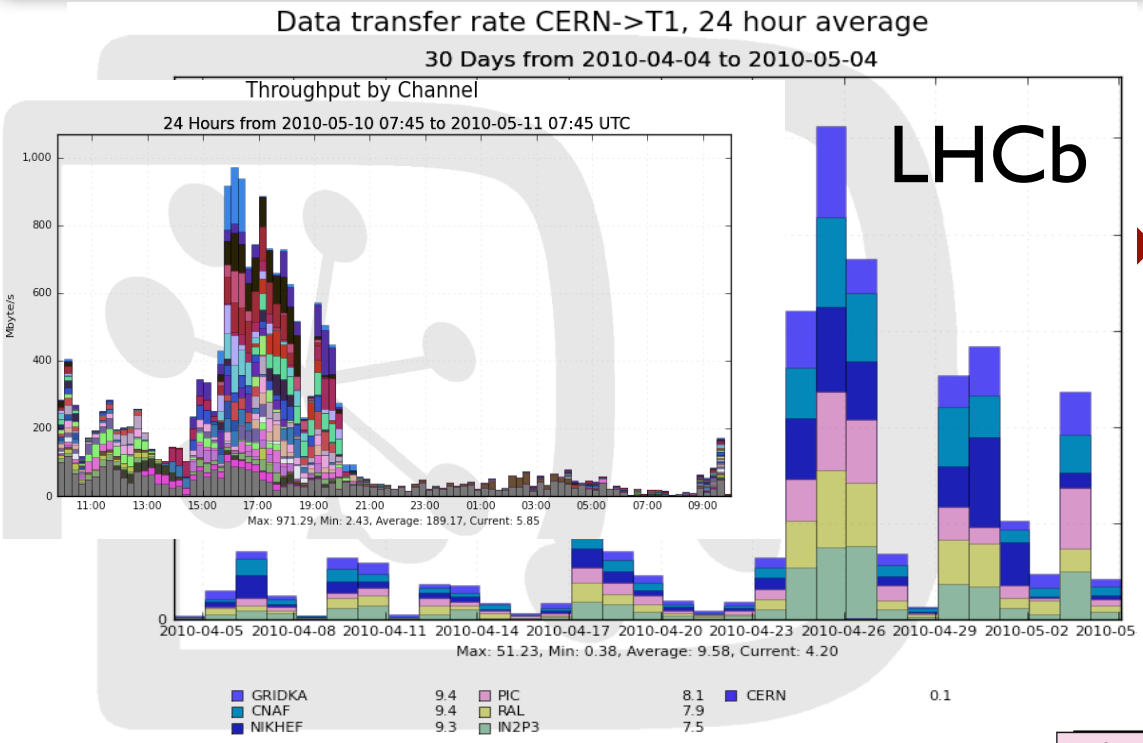
- Transfers from CERN to the Tier-1s are large, and driven by collision and cosmic running



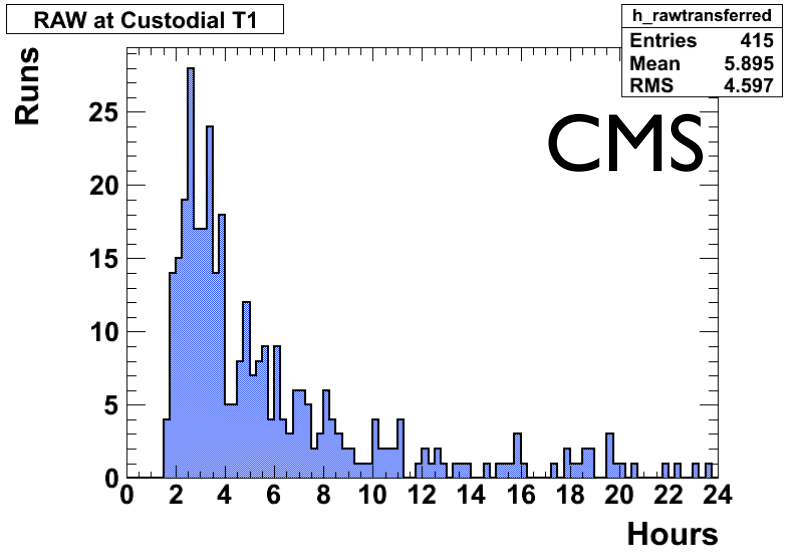
Castor traffic last month:
> 4 GB/s input
> 13 GB/s served

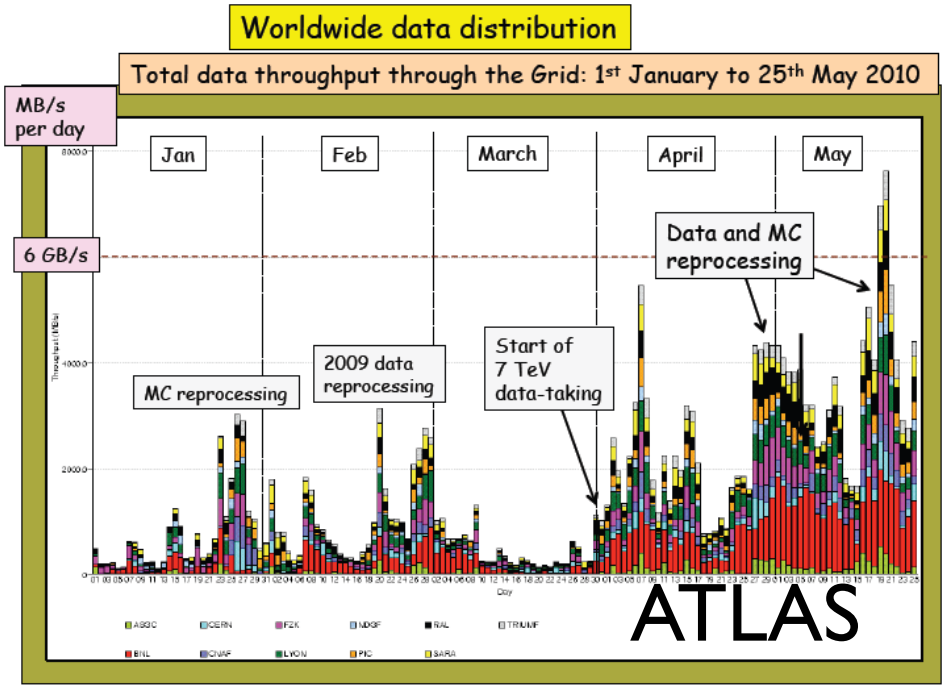


Transfer Peaks

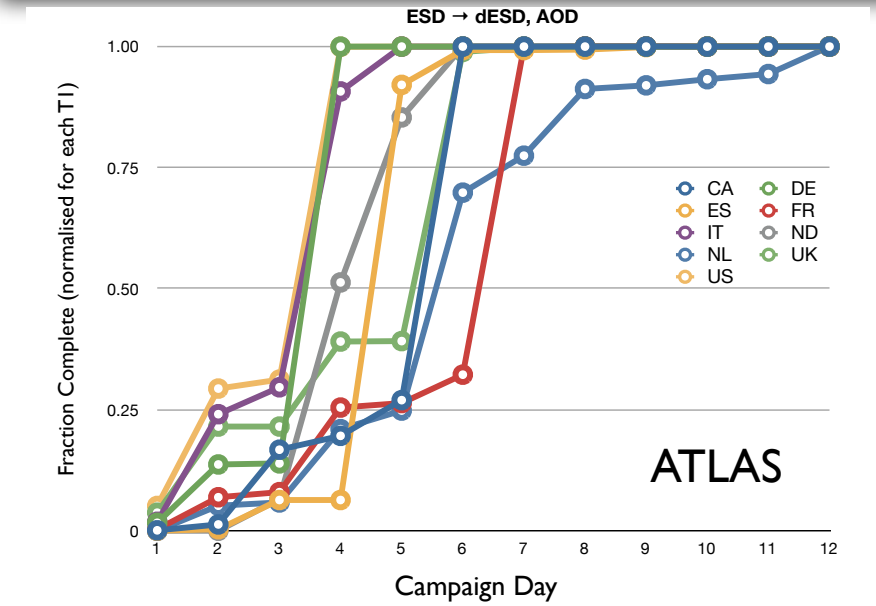


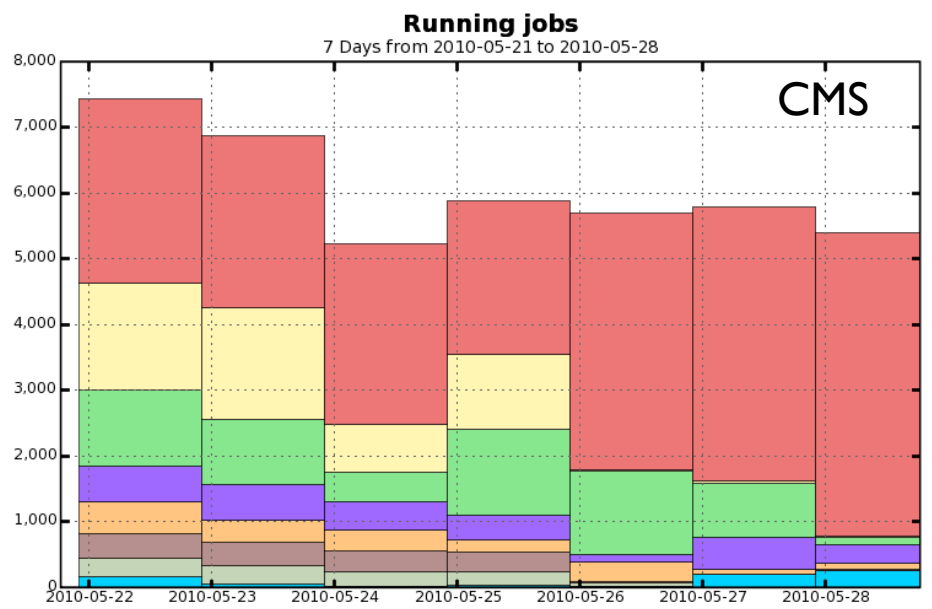
- Peak transfers are meeting the computing model expectations



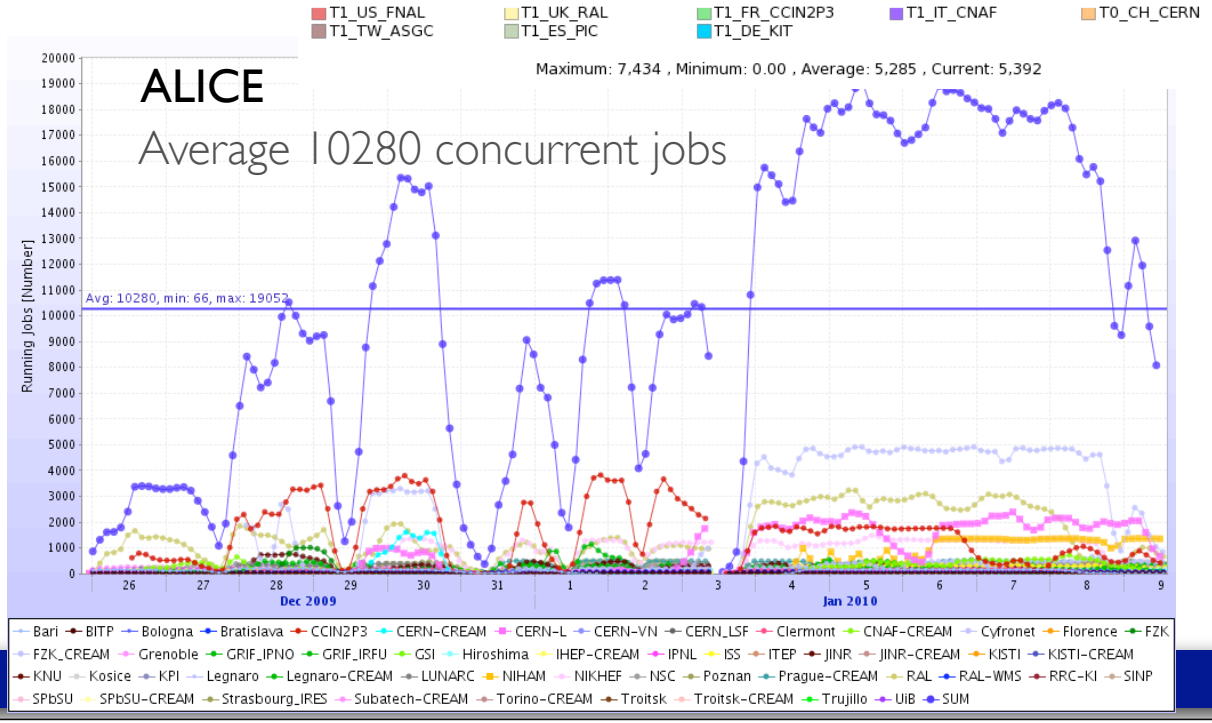


Data Reprocessing Profile





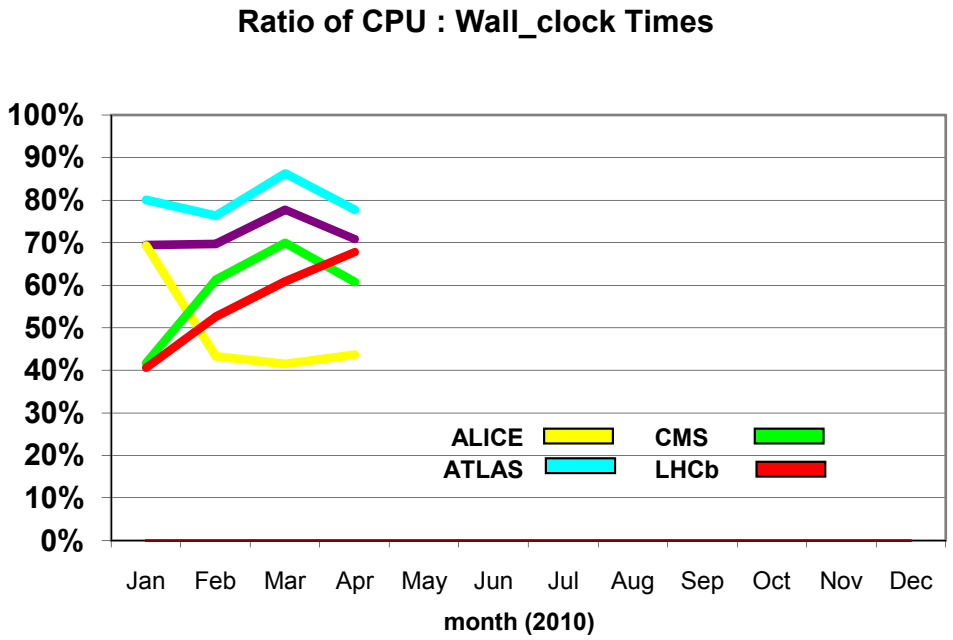
- Once data arrives at Tier-1 Centers is reprocessed
  - New calibrations, software and data formats



Data Processing

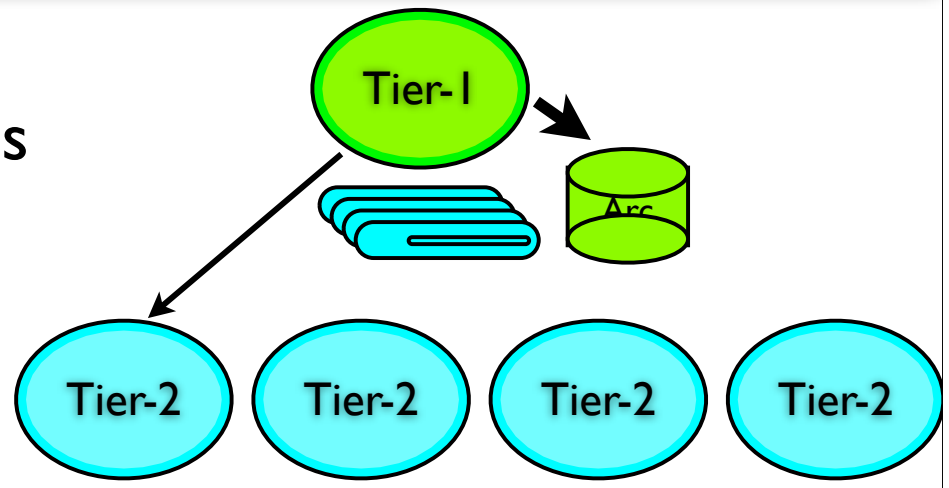
- LHC Experiments can currently reprocess the entire collected data in less than a week
  - These reprocessing passes will grow to several months as the data volumes increase
- Grid interfaces to the batch systems are scaling well
- Storage systems are ingesting and serving data

CPU Efficiency is reaching expectations for organized processing



Transfers to Tier-2s and Analysis

- Once data is onto the WLCG it is made accessible to analysis applications
  - Largest fraction of analysis computing is at Tier-2s



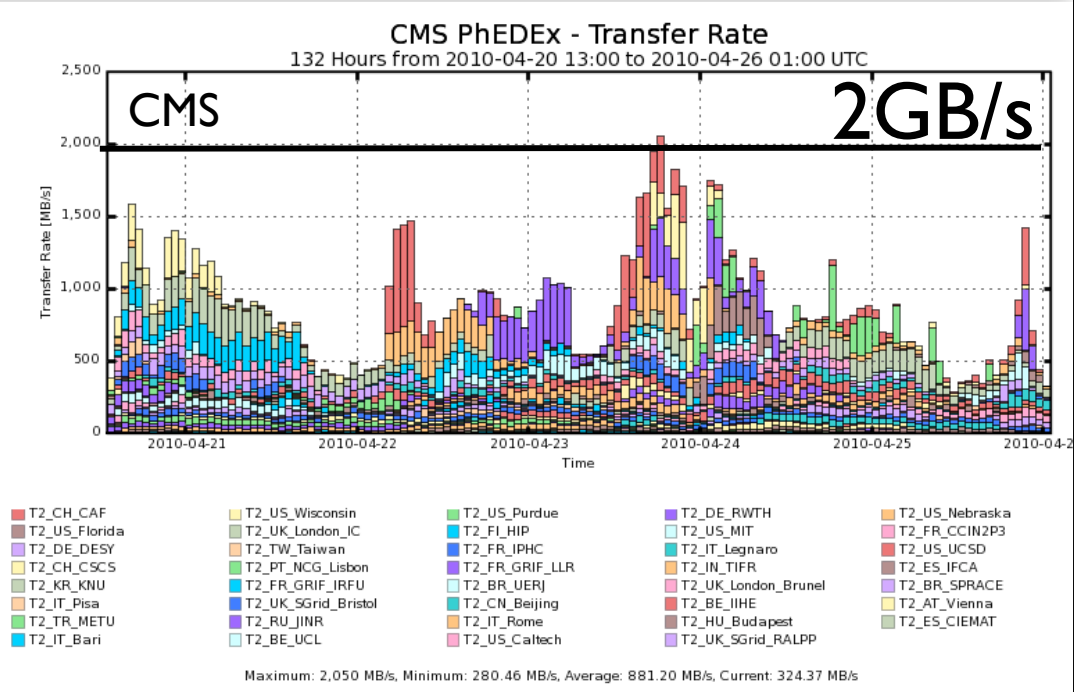
Example of Data Collection vs. Computing Exercise

	BEST 7 DAYS of STEP09		7 DAYS REPRO DISTRIB	
CLOUD	RATE	FILES	RATE	FILES
ASGC	155 MB/s	49,347	207 MB/s	285,051
BNL	640 MB/s	791,691	940 MB/s	1,113,927
CERN	208 MB/s	92,425	182 MB/s	584,935
CNAF	264 MB/s	83,217	237 MB/s	337,651
FZK	364 MB/s	184,151	568 MB/s	948,435
LYON	468 MB/s	227,413	388 MB/s	633,515
NDGF	152 MB/s	44,325	128 MB/s	209,555
PIC	339 MB/s	106,994	272 MB/s	314,142
RAL	405 MB/s	272,820	467 MB/s	637,770
SARA	320 MB/s	104,545	265 MB/s	364,706
TRIUMF	321 MB/s	120,007	264 MB/s	312,129

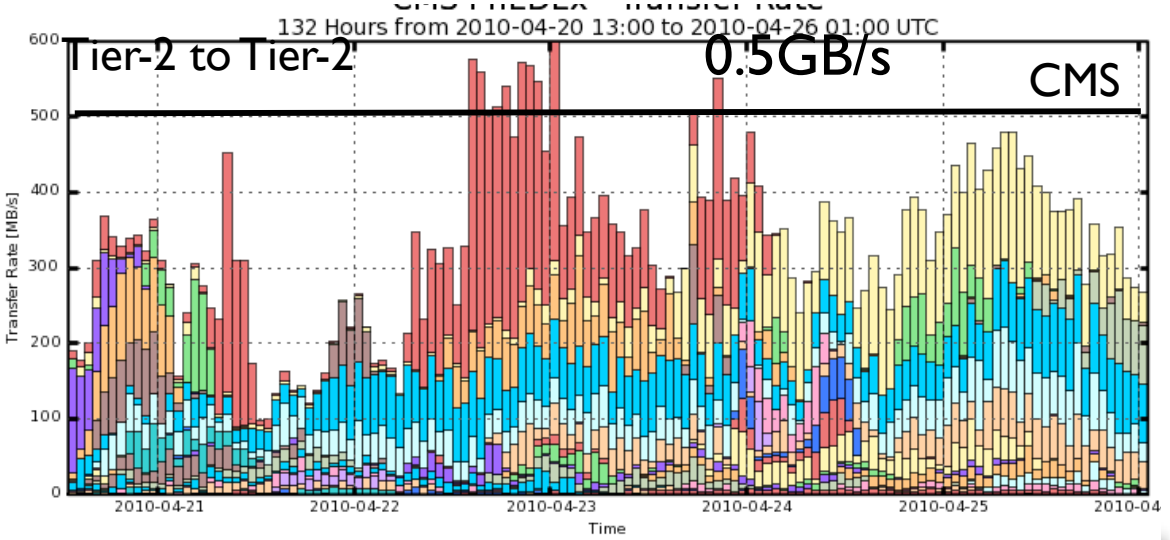
ATLAS

Data Moving to Tier-2s

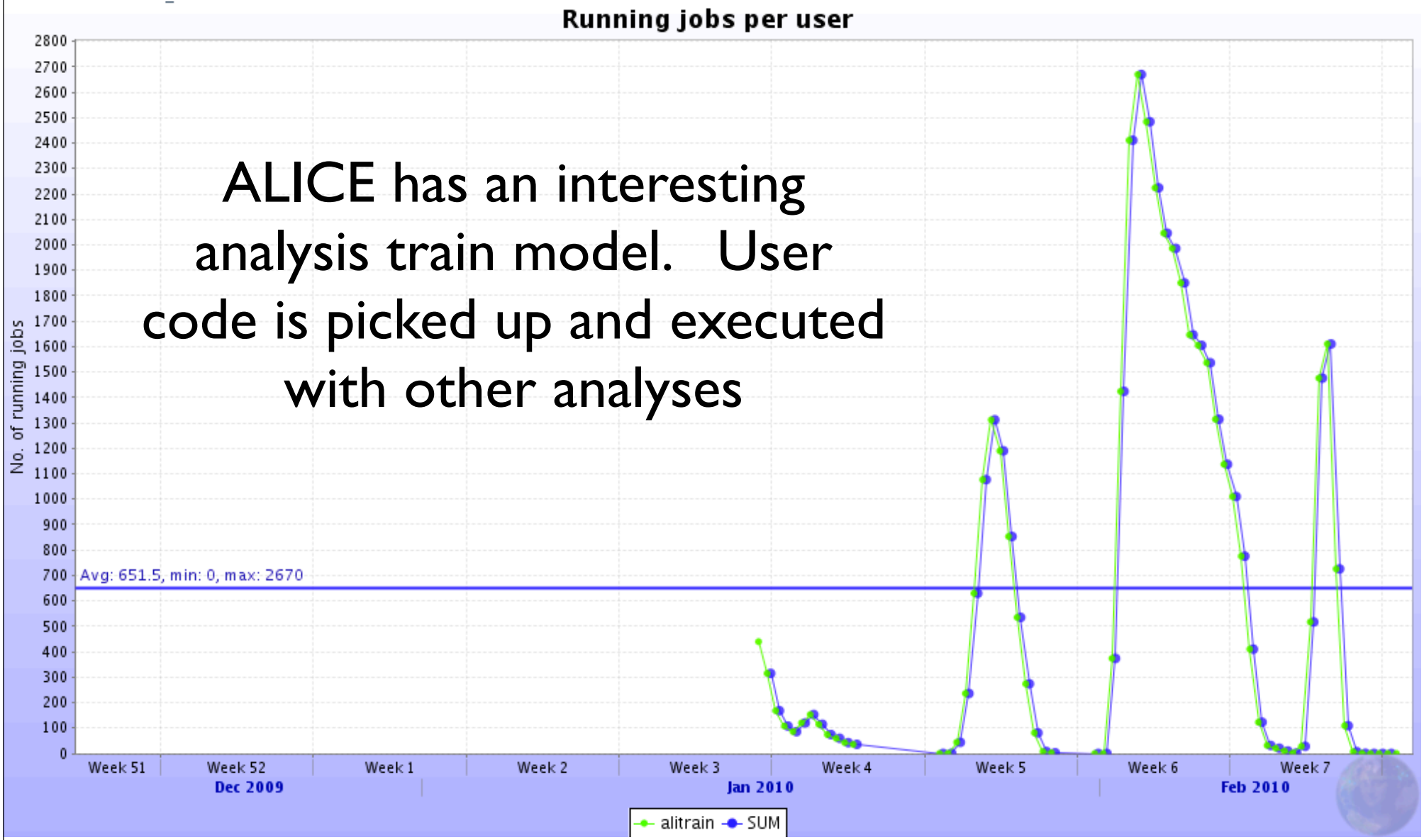
- Tier-1 to Tier-2 average is coming up
  - 49 Tier-2s sites have received data since the start of 7TeV Collisions



- After significant effort of the commissioning team we're now having serious Tier-2 to Tier-2 transfers
  - Good for group skims and replicating data

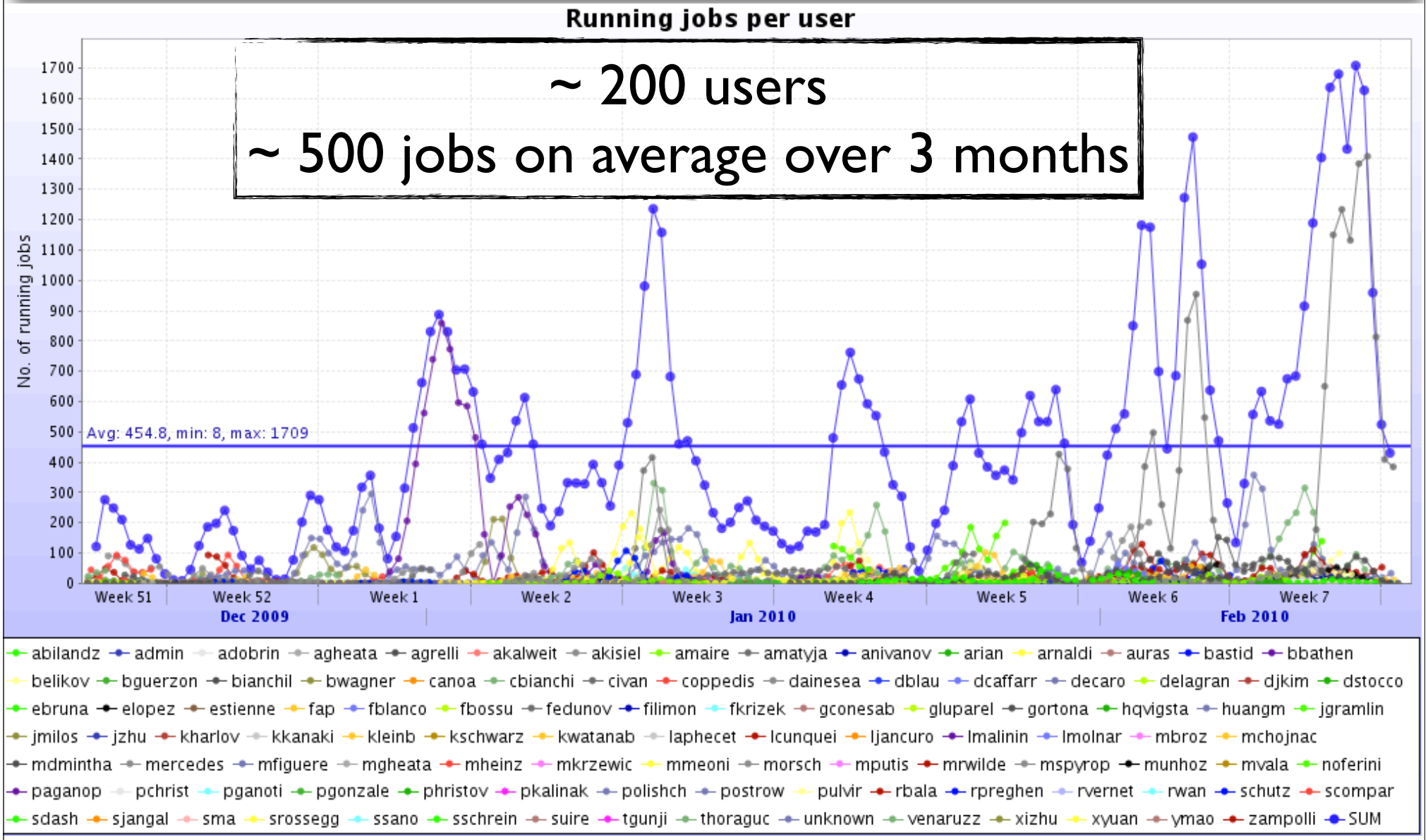


ALICE ANALYSIS train



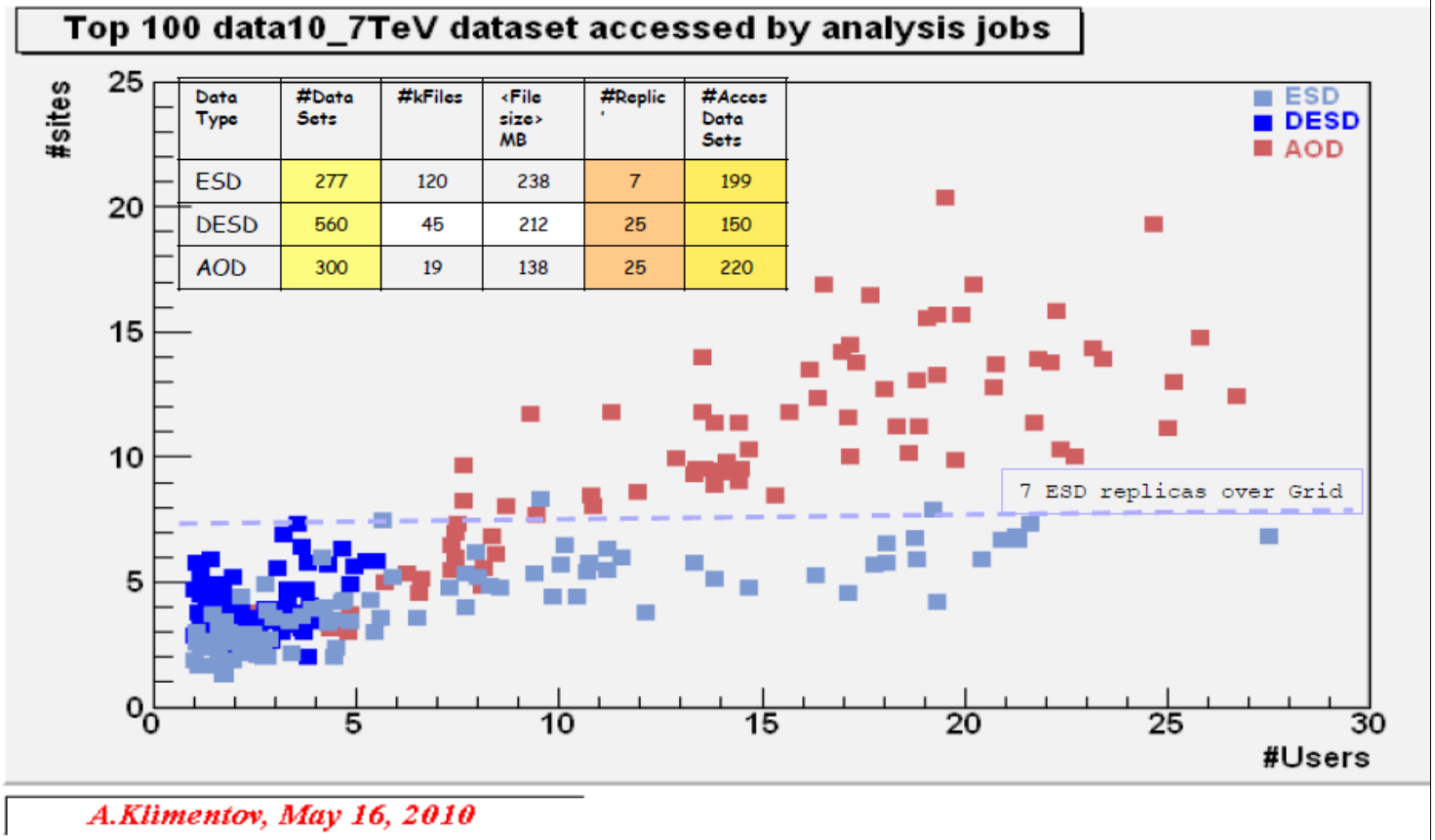
ALICE has an interesting analysis train model. User code is picked up and executed with other analyses

ALICE End user analysis



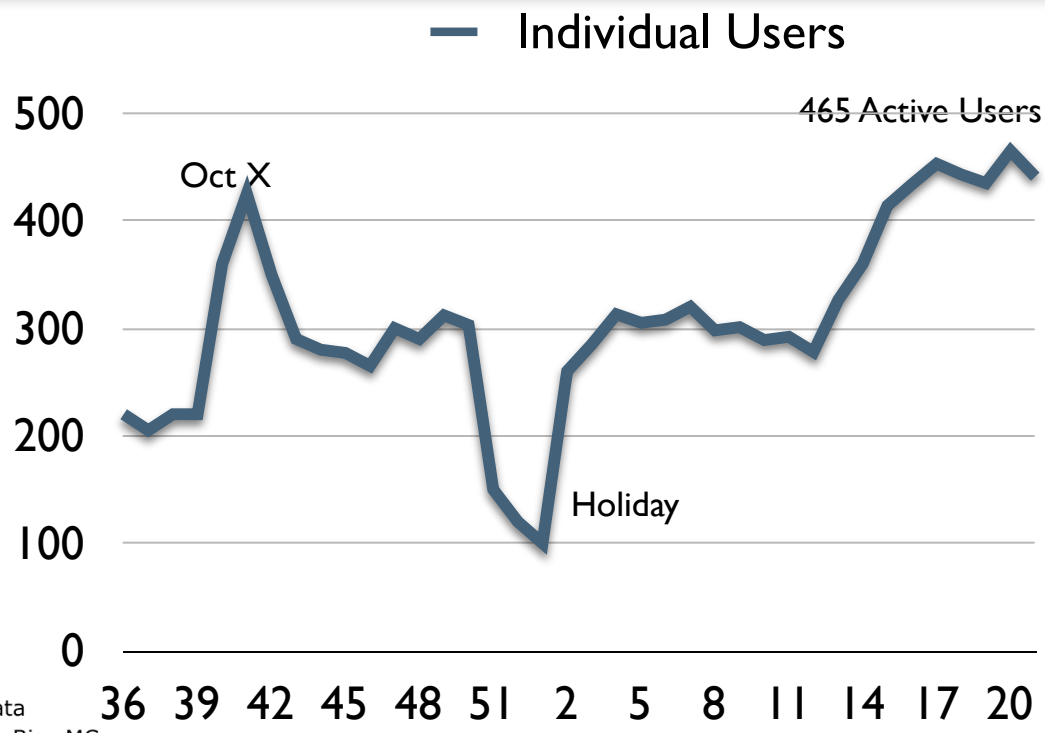
ATLAS Analysis data access

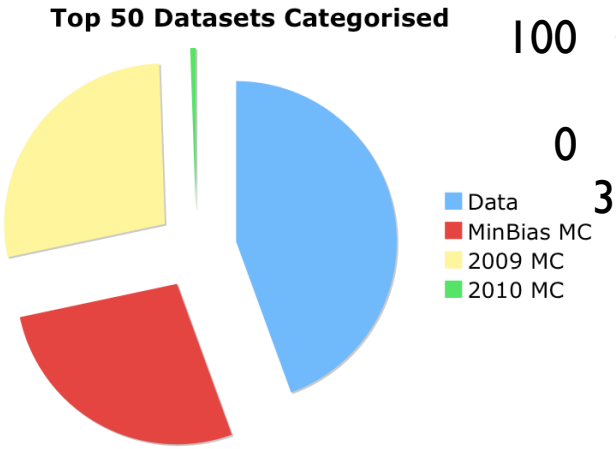
- Data were analyzed on the Grid already from the first days of data-taking
- Data are analyzed at many Tier-2, and some Tier-1, sites
  - See data access counts
- Many hundreds of users accessed data on the Grid



CMS Analysis Activity

- Number of people participating in analysis is increasing

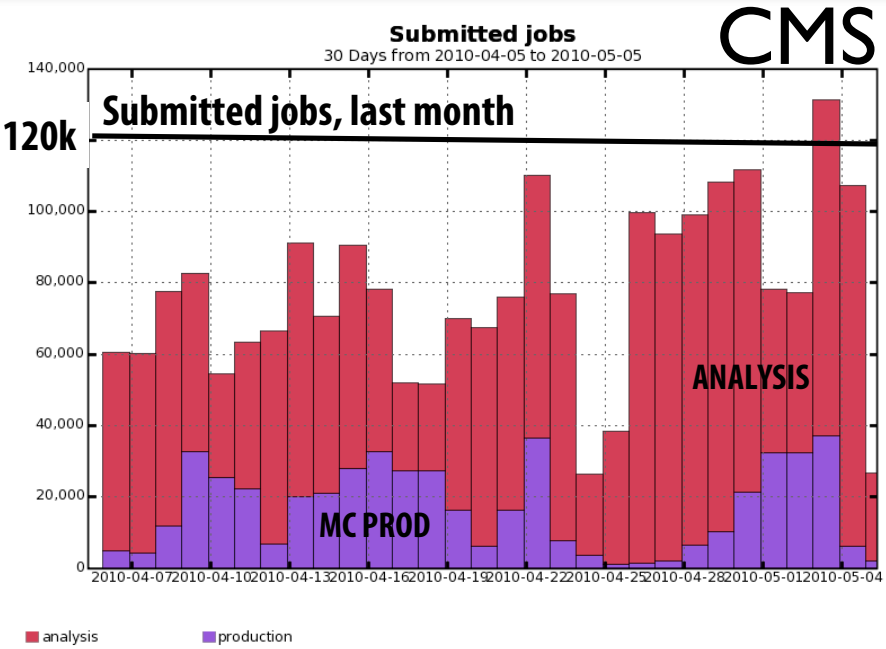




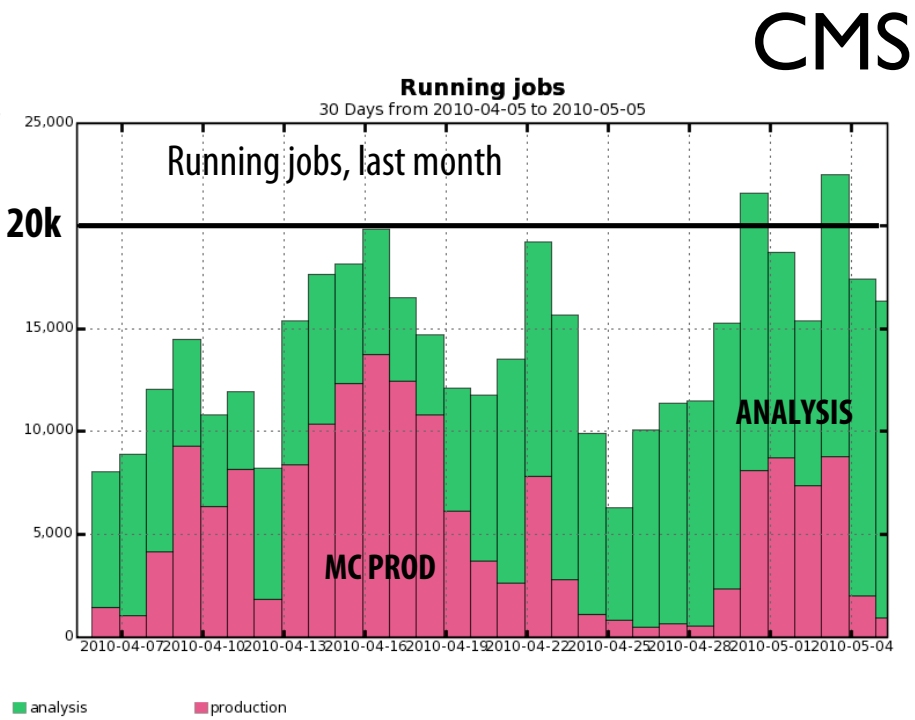
- Largest access to collision data and the corresponding simulation

- Total of 1715 datasets used within last month
- 50 most used datasets account for 2/3 of all jobs
- Among those, 3/4 are data and MinBias MC

CMS Analysis Ratio

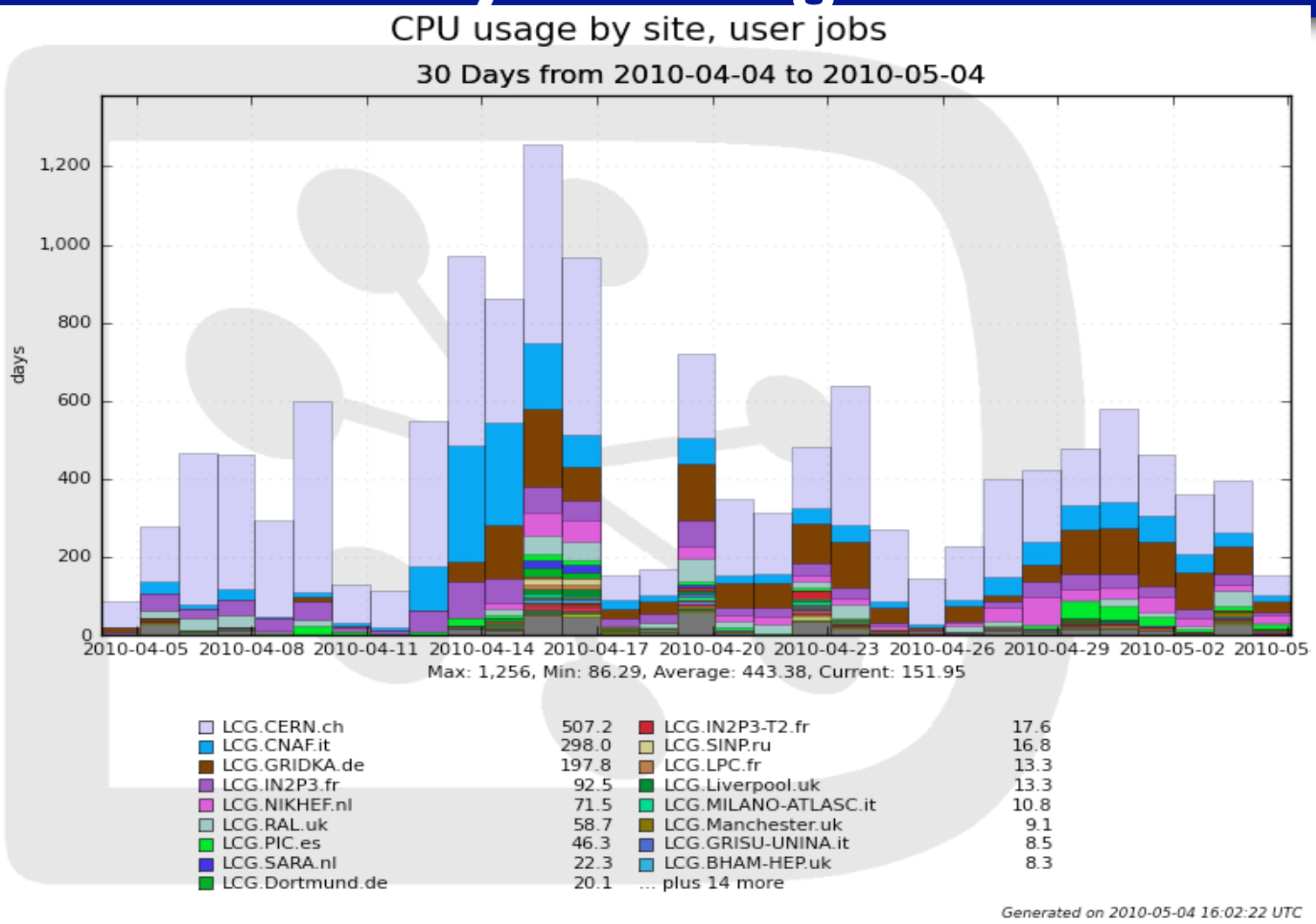


- Analysis users submit many more jobs



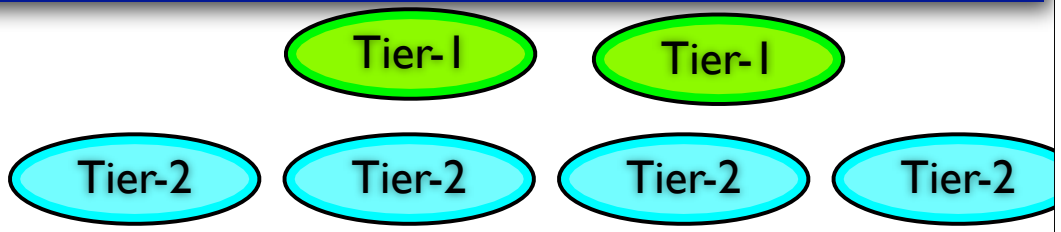
- Production jobs last longer so on average T2's are used 50%-50%

LHCb Analysis Usage



- Good balance of sites
  - Peak of 1200 CPU days delivered in a day

Simulated Event Production

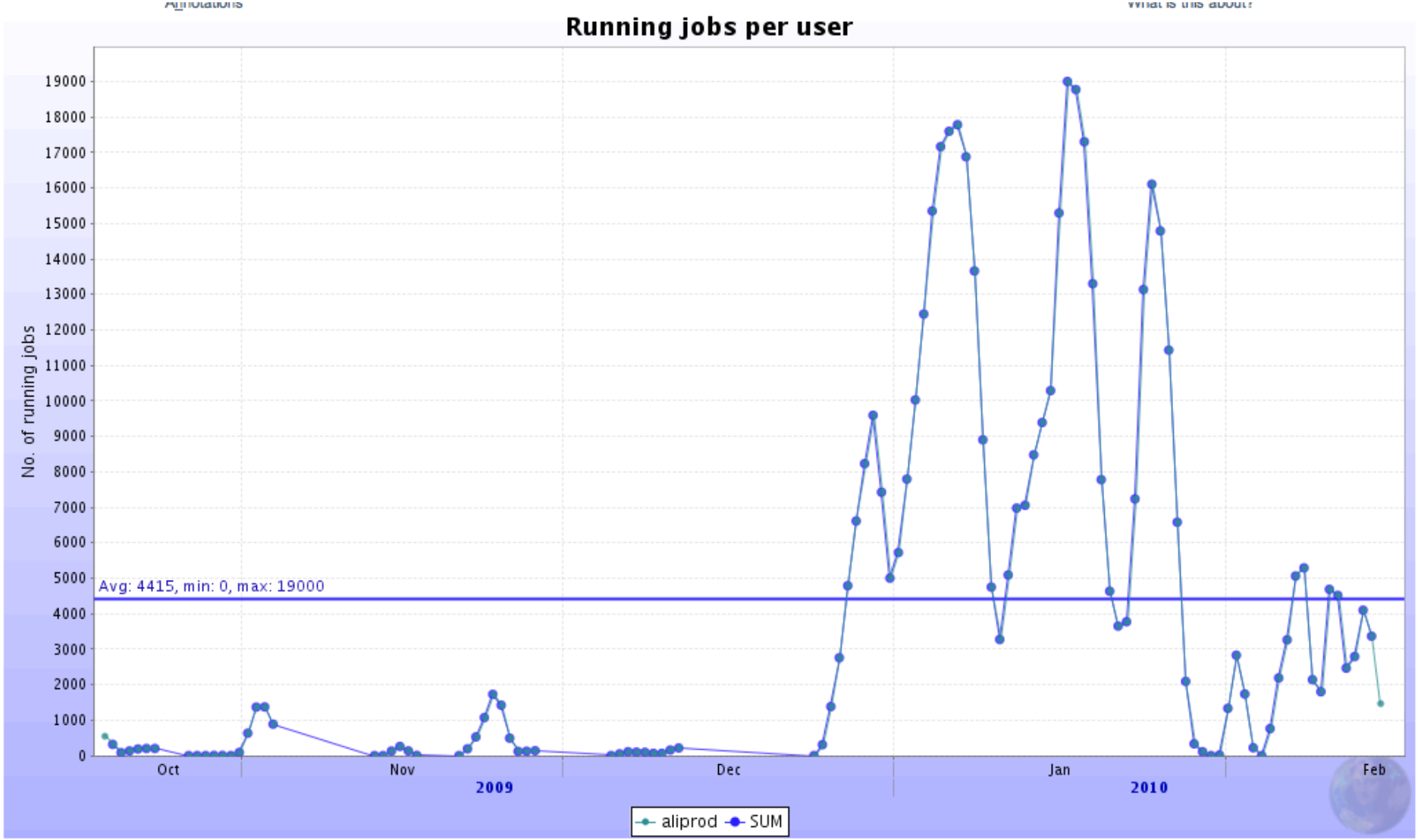


- Simulated Event Production is one of the earliest grid applications and very successful
  - Pile-up and realistic simulation making this a more interesting problem

ALICE MC production

MC production in all T1/T2 sites

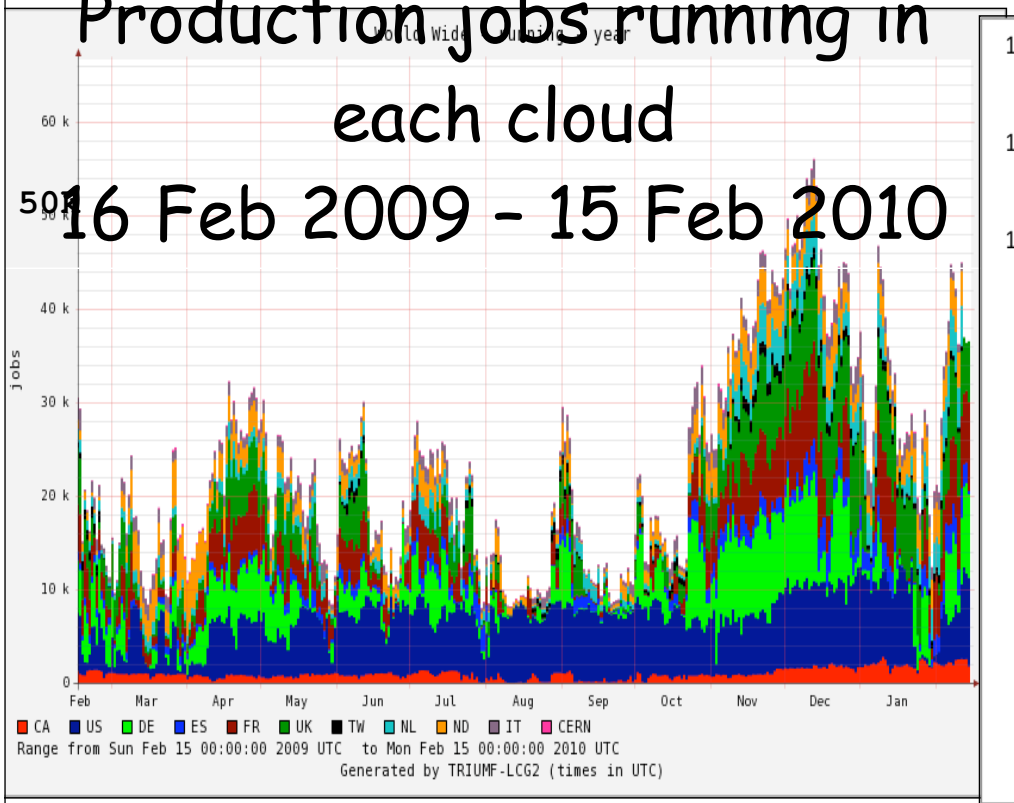
- 30 TB with replica



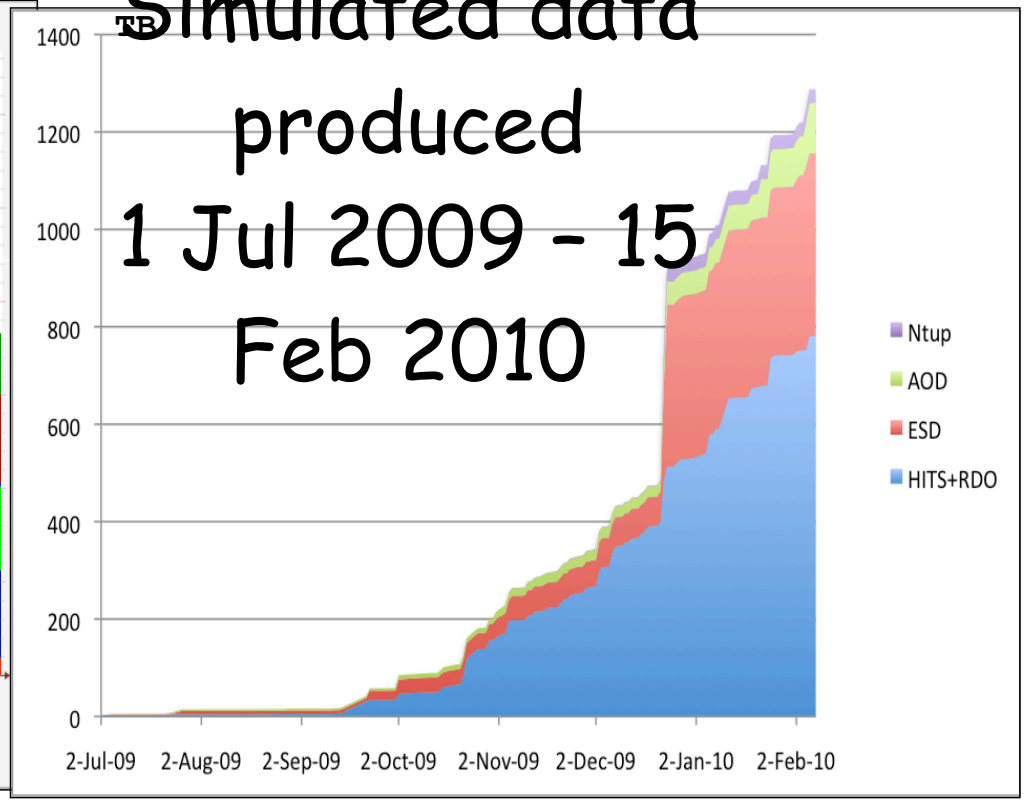
ATLAS 2009-10 simulation production

- Simulation production continues in the background all the time
  - Fluctuations caused by a range of causes, including release cycles, sites downtimes etc.
- Parallel effort underway for MC reconstruction and reprocessing
  - Including reprocessing of MC09 900 GeV and 2.36 TeV samples with AtlasTier0 15.5.4.10 reconstruction, same release as for data reprocessing, done in December
- More simulated ESDs produced since Dec'09 to match 2009 real data analysis

Production jobs running in each cloud
16 Feb 2009 - 15 Feb 2010

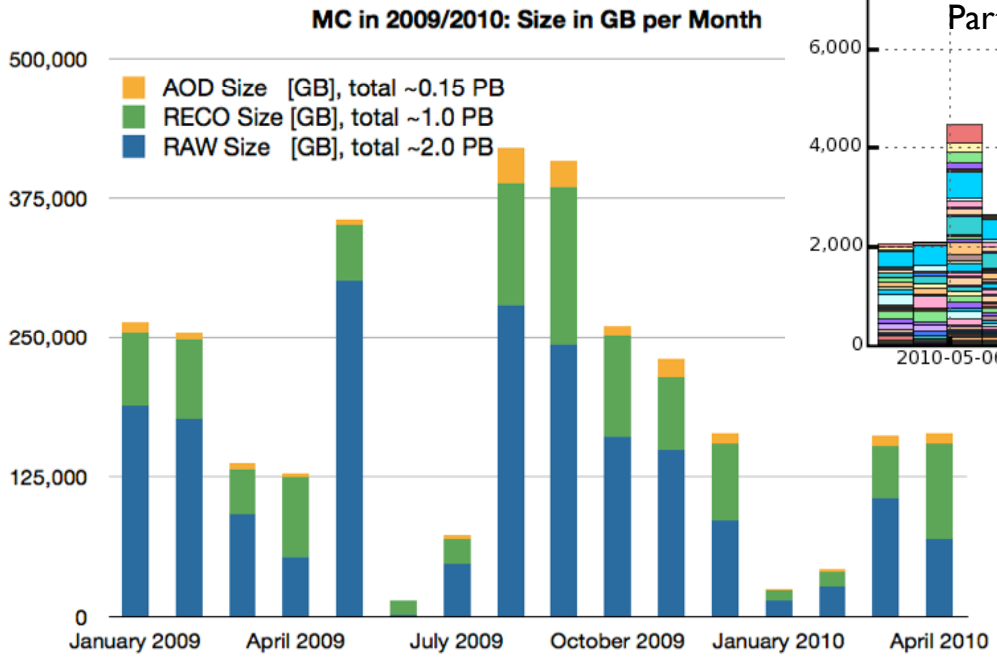


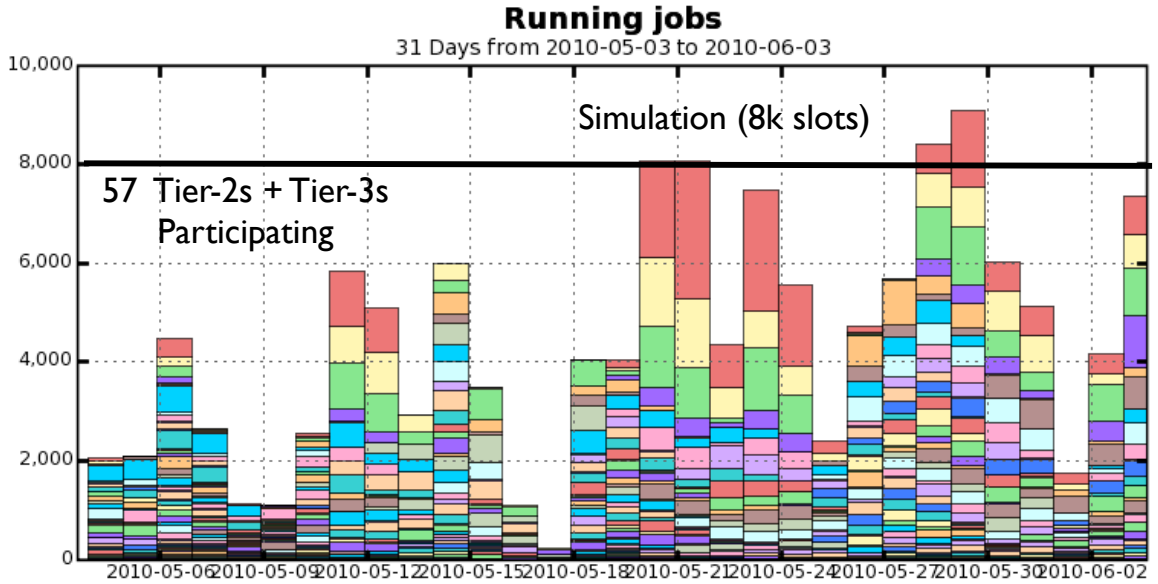
Simulated data produced
1 Jul 2009 - 15 Feb 2010



CMS MC Production by Month

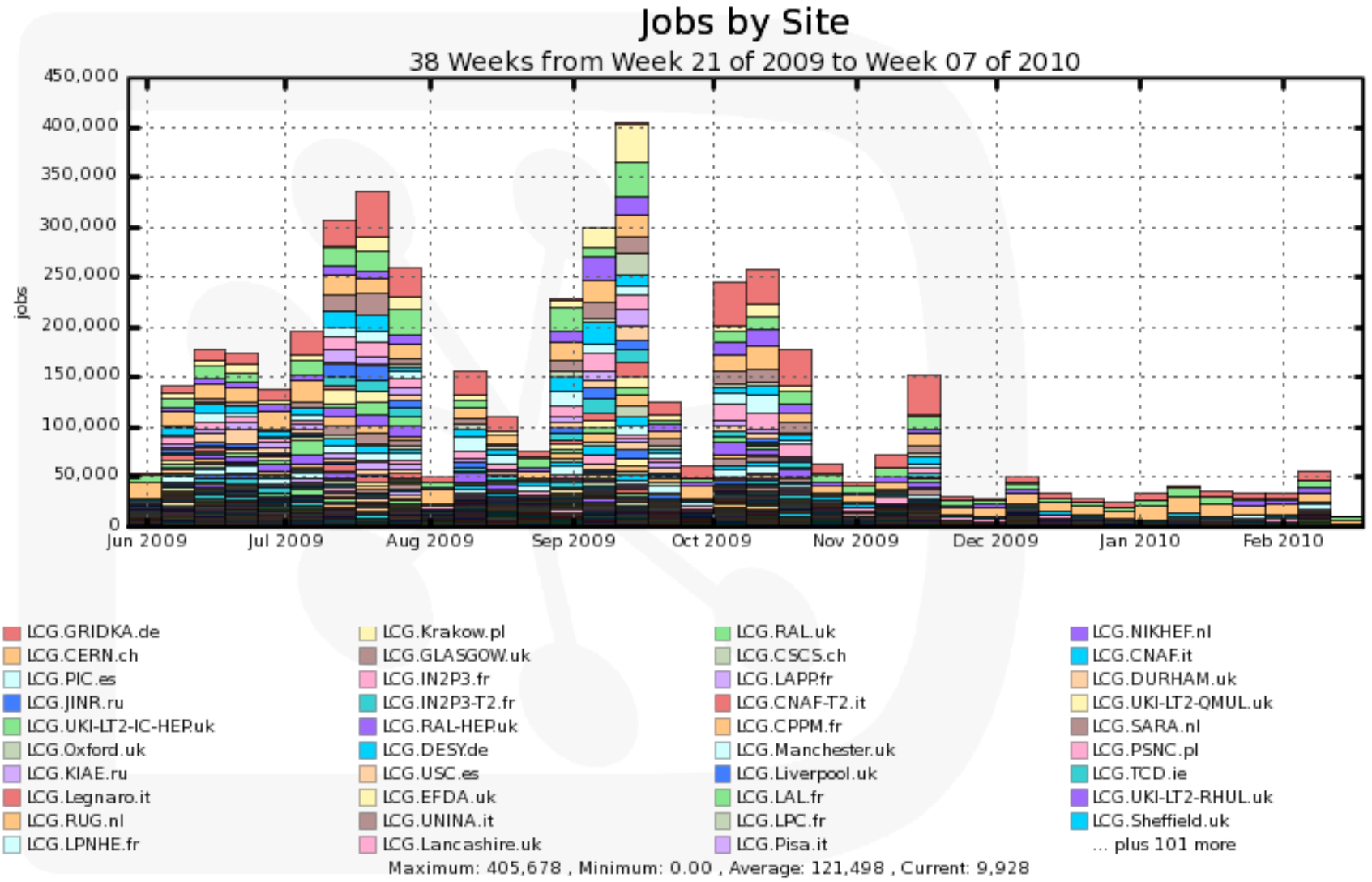
- Total number of events produced in 2009 is large
  - Big variations by month
  - Generally related to lack of work, not lack of capacity
- 2010 Sample growing





LHCb139 sites hit, 4.2 million jobs

- Start in June: start of MC09



Outlook

- The Grid Infrastructure is working for physics
  - Data is reprocessed
  - Simulated Events are Produced
  - Data is delivered to analysis users
- Integrated volume of data and livetime of the accelerator are still lower than the final planning
  - Not all resources are equally utilized
- Activity level is high
  - Still lots to learn with increasing datasets