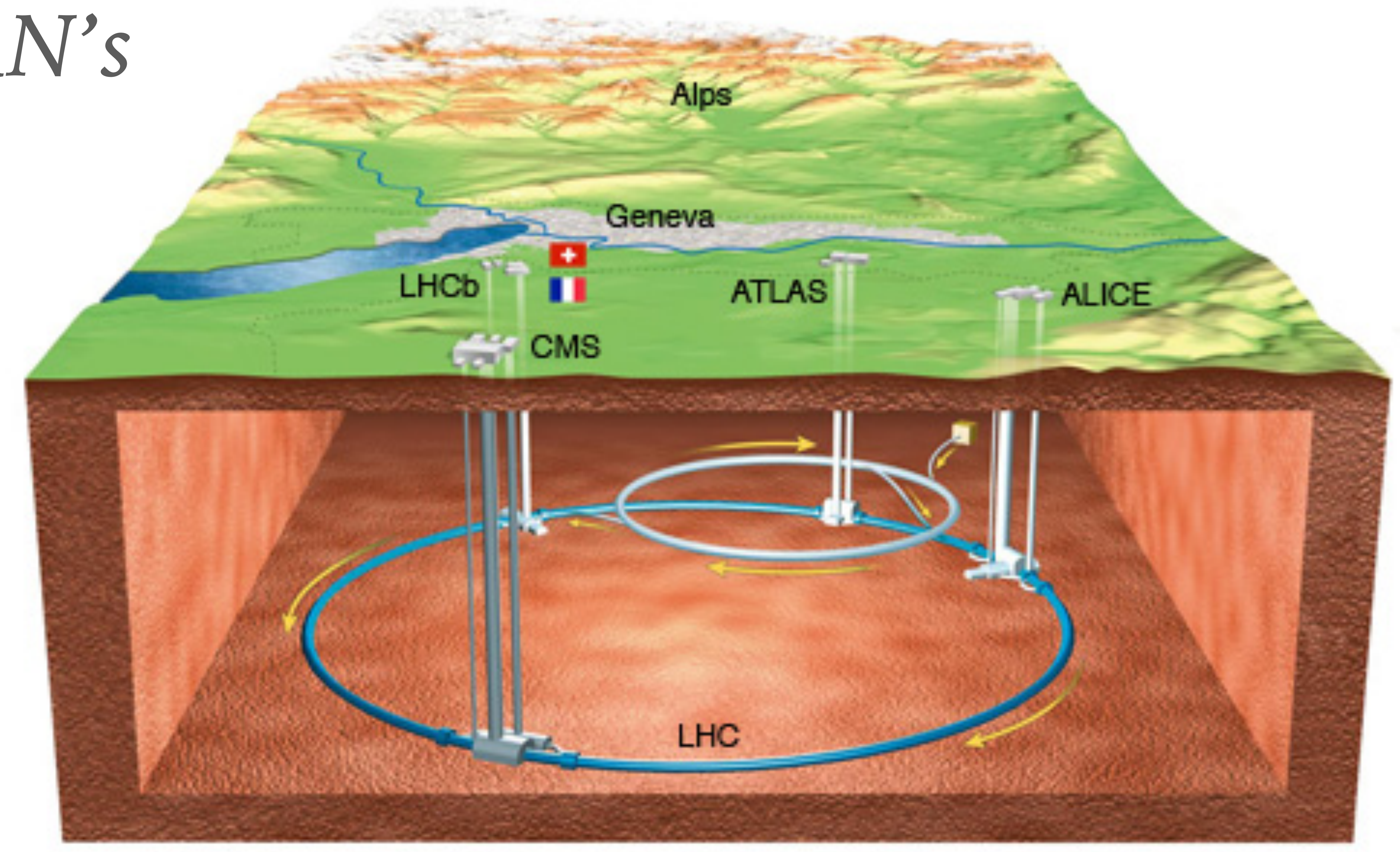
Integration Of PanDA Workload Management System With Supercomputers for ATLAS

Danila Oleynik, *Alexei Klimentov, Jack Wells, Kaushik De, Torre Wenaus, Tadashi Maeno, Paul Nilsson, Fernando Barreiro, Wen Guan, Sergey Panitkin, Shantenu Jha*

THE ATLAS EXPERIMENT AT THE LHC

The Large Hadron Collider (LHC) is the world's largest and most powerful particle accelerator. It first started up on 10 September 2008, and remains the latest addition to CERN's accelerator complex.

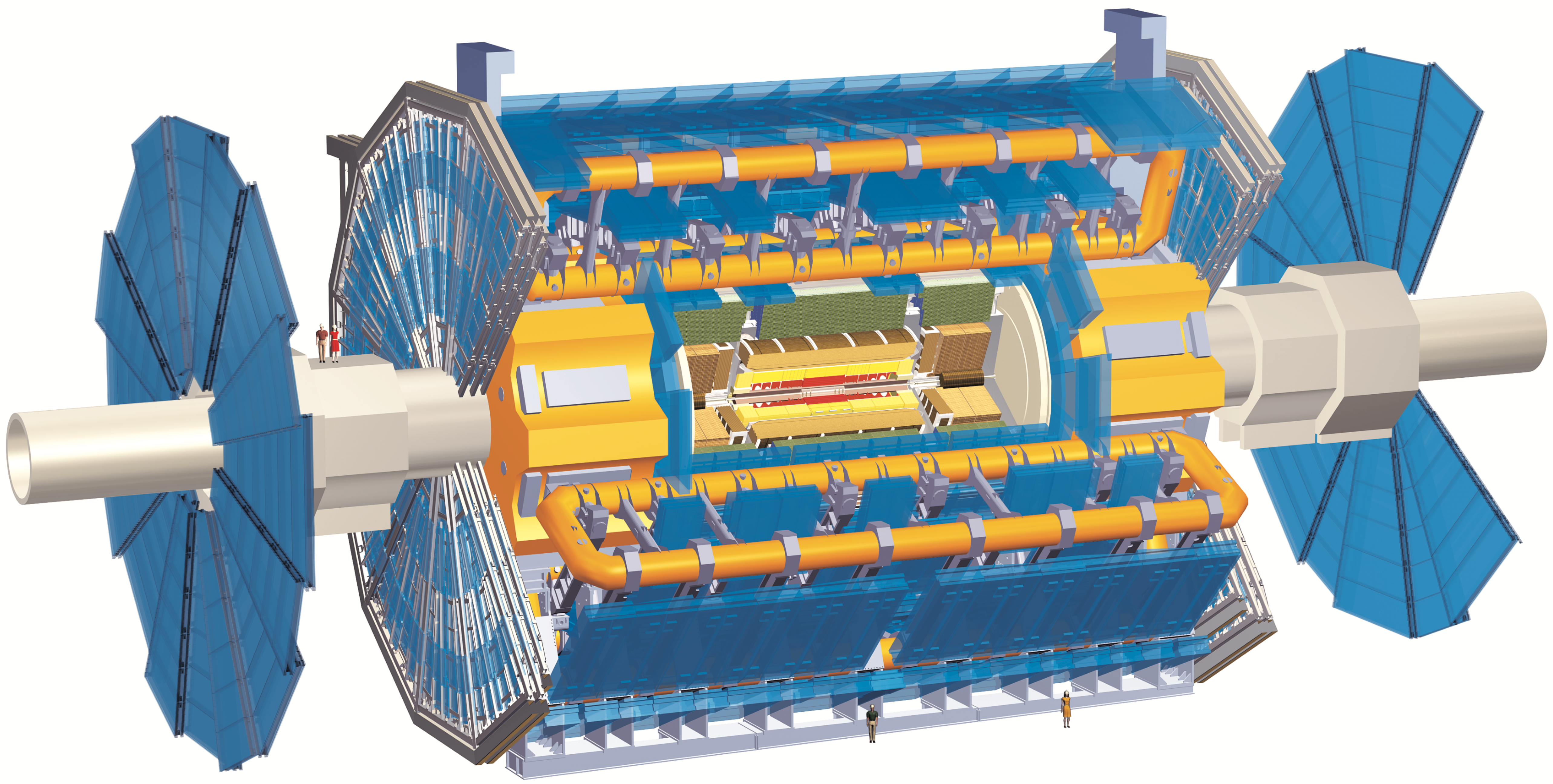


THE ATLAS EXPERIMENT AT THE LHC

3000 scientists

38 countries

THE ATLAS EXPERIMENT AT THE LHC



The ATLAS Detector	
Diameter:	25 m
Length:	46 m
Barrel toroid	26 m
Overall weight:	7000 tonnes
~ 100 million electronic channels	
~ 3 000 km of cables	

THE ATLAS EXPERIMENT AT THE LHC

The Nobel Prize in Physics 2013

Photo: A. Mahmoud

François Englert

Prize share: 1/2

Photo: A. Mahmoud

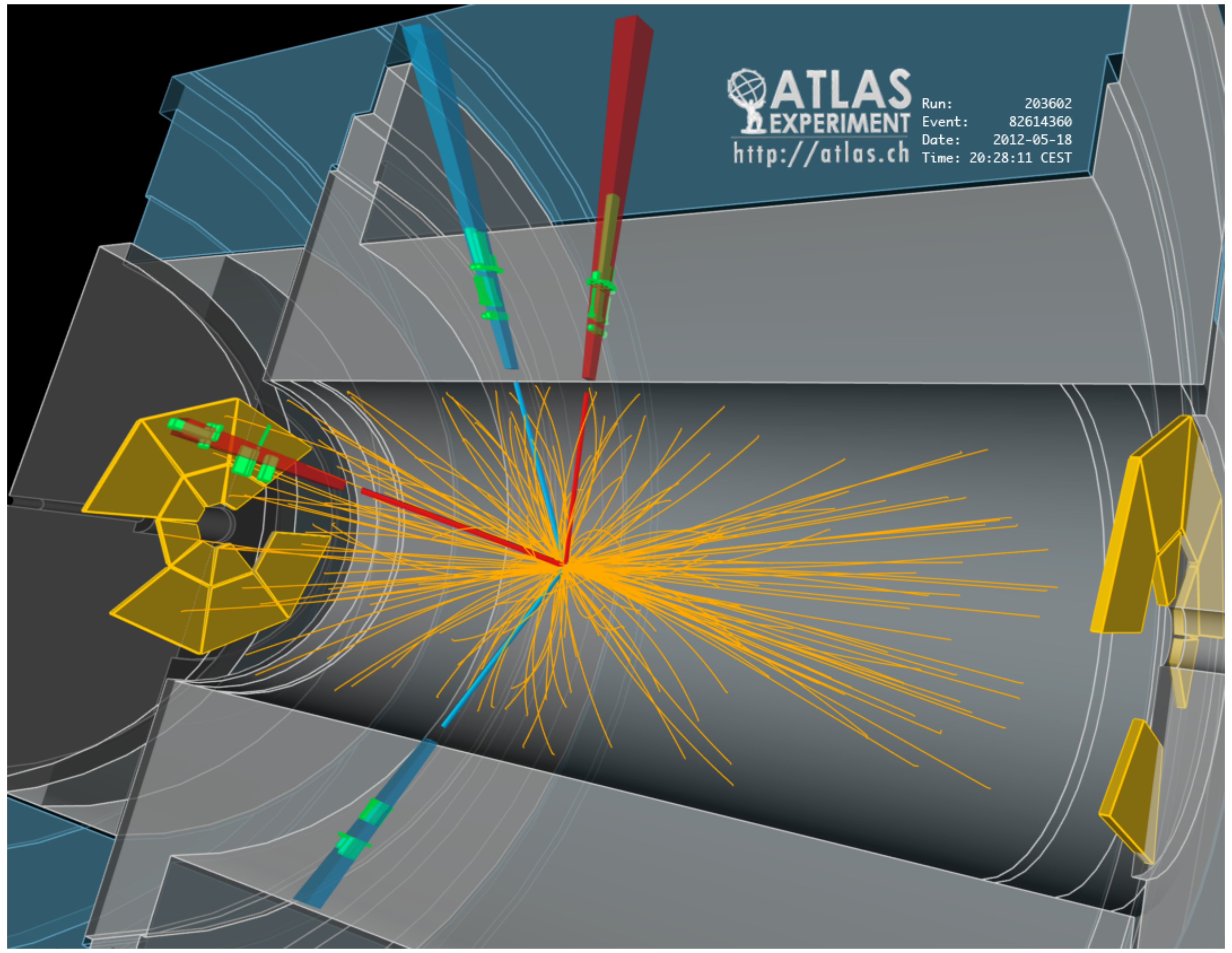
Peter W. Higgs

Prize share: 1/2

The Nobel Prize in Physics 2013 was awarded jointly to François Englert and Peter W. Higgs *"for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider"*

BIG DATA: OFTEN JUST A BUZZ WORD, BUT NOT WHEN IT COMES TO ATLAS…

- Bunches of protons cross 40 million times a second.
- Each bunch contains 10^{11} protons.
- Number of proton-proton collisions in the detector: 1 billion per second.
- When any of the protons collide, the process is called an “event”.

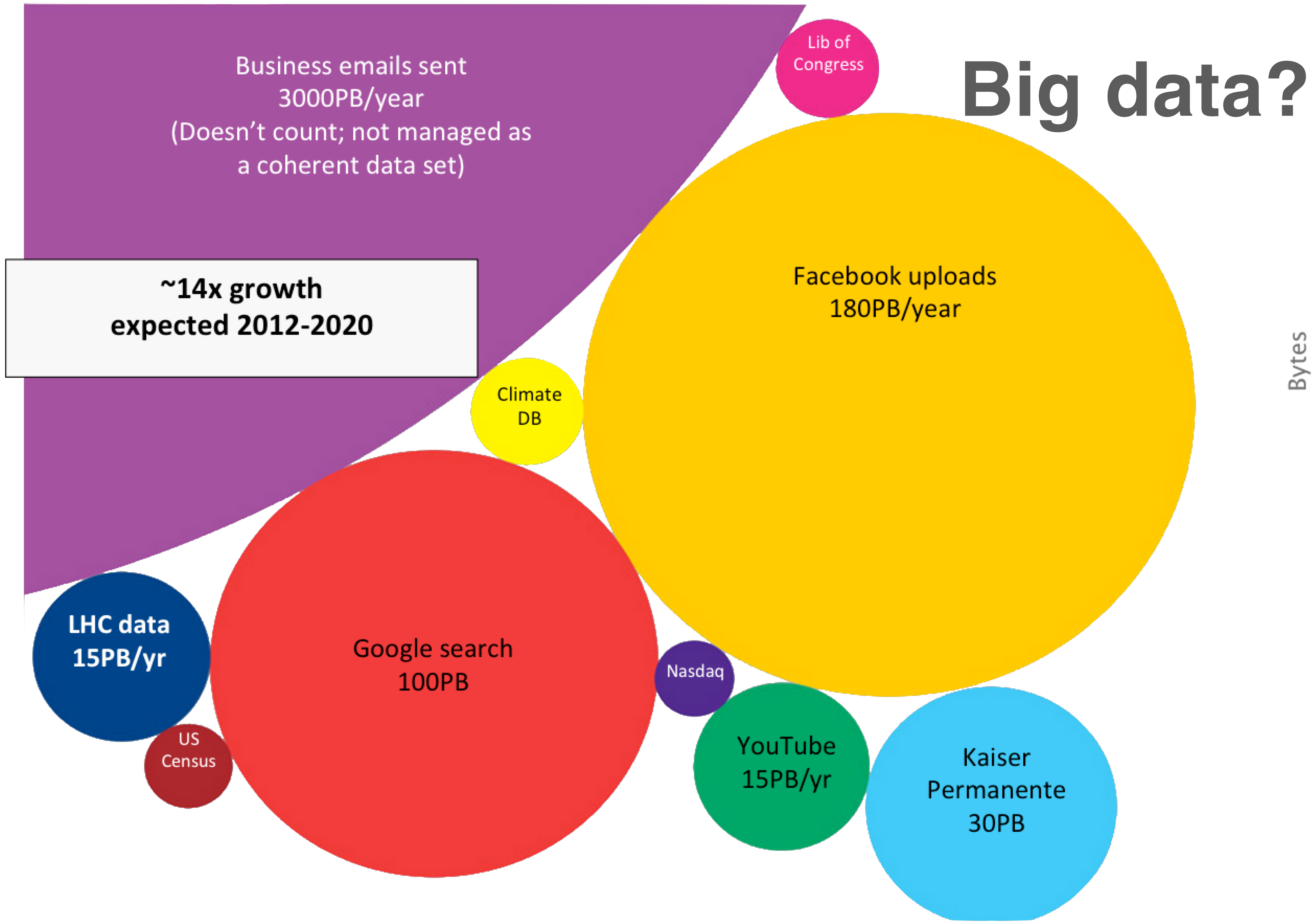


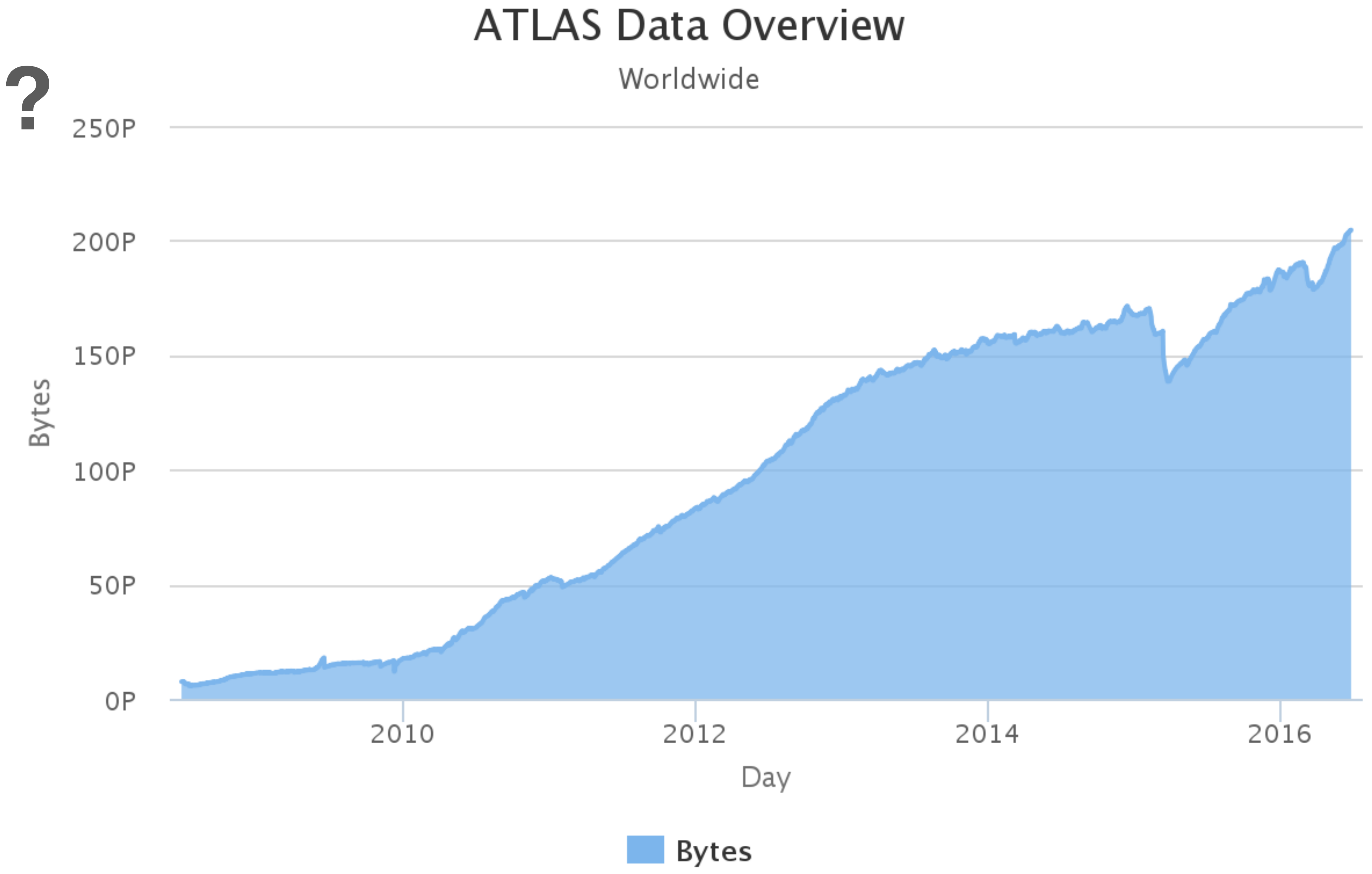
If all data would be recorded, this would fill 100 000 CDs per second. This would create a stack of CDs 150 m (450 ft) high every second, which could reach to the moon and back twice each year.

ATLAS detector has a 3 level trigger system (reduction in three steps). Total event reduction factor by the trigger system: 200 000.

Initially trigger and data acquisition records 320 Mbytes per second

BIG DATA: OFTEN JUST A BUZZ WORD, BUT NOT WHEN IT COMES TO ATLAS…





Current ATLAS data set, considering all data products and replicas

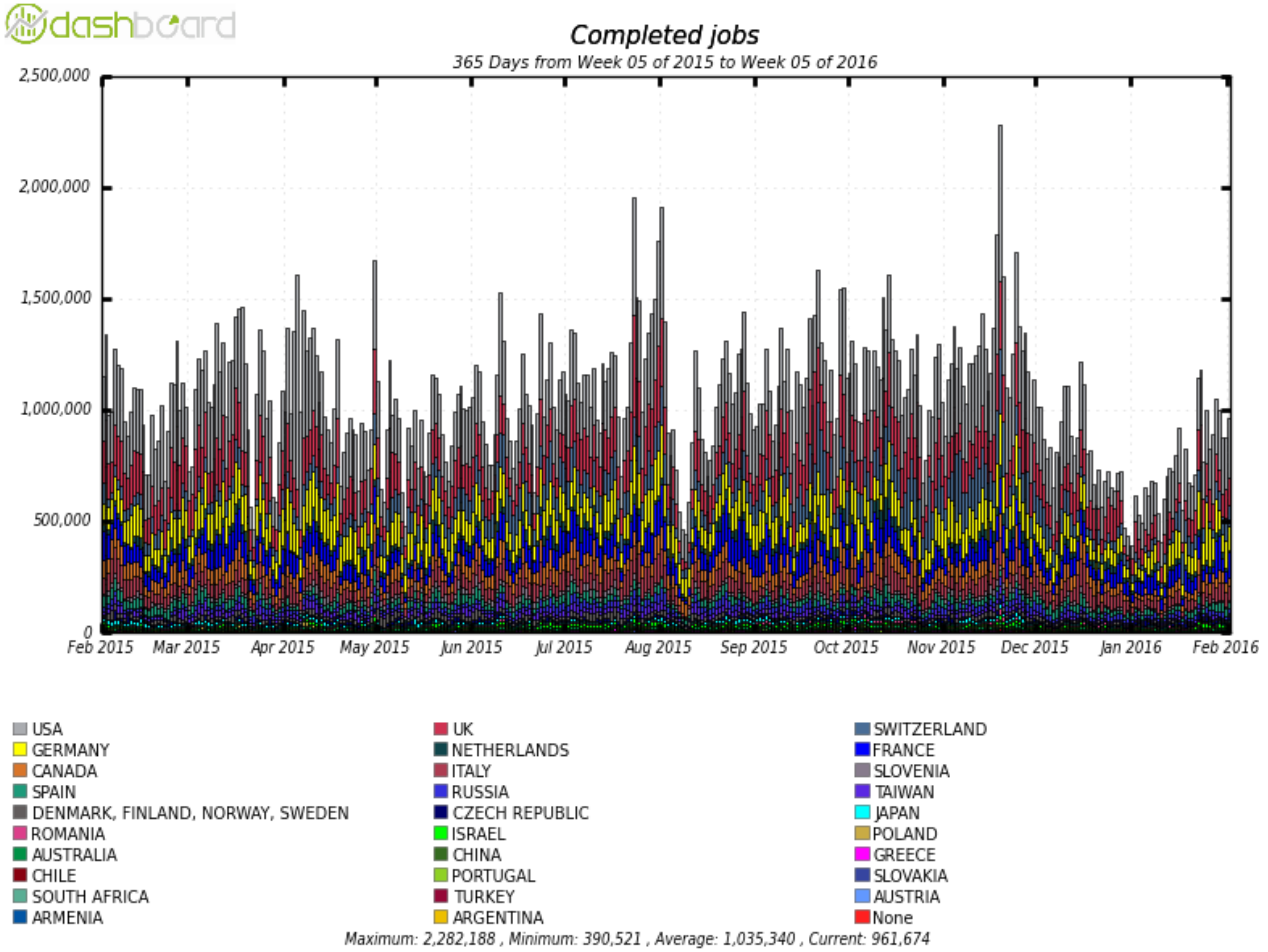
200 PB

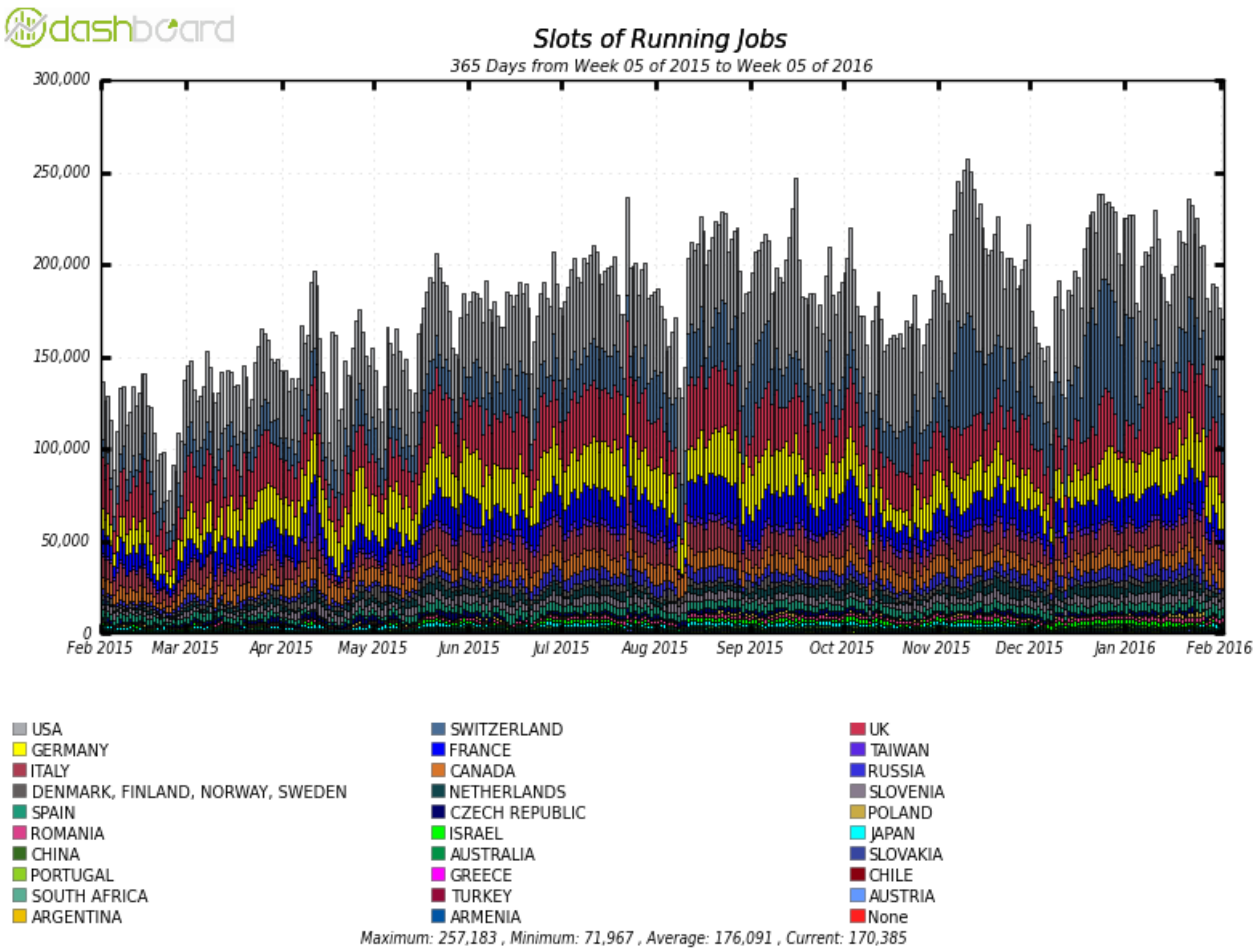
PanDA

*The **PanDA P**roduction **AN**d **D**istributed **A**nalysis system has been developed by ATLAS since summer 2005 to meet ATLAS requirements for a data-driven workload management system for production and distributed analysis processing capable of operating at LHC data processing scale. ATLAS processing and analysis places challenging requirements on throughput, scalability, robustness, efficient resource utilization, minimal operations manpower, and tight integration of data management with processing workflow.*

PanDA throughput has increased continuously over the years. In 2009, a typical PanDA processing rate was 50k jobs/day and 14k CPU wall-time hours/day for production at 100 sites around the world, and 3-5k jobs/day for analysis. In 2015, PanDA processed about a million jobs per day, with about 100,000 jobs running at any given time. The PanDA analysis user community numbers over 1400.

PanDA

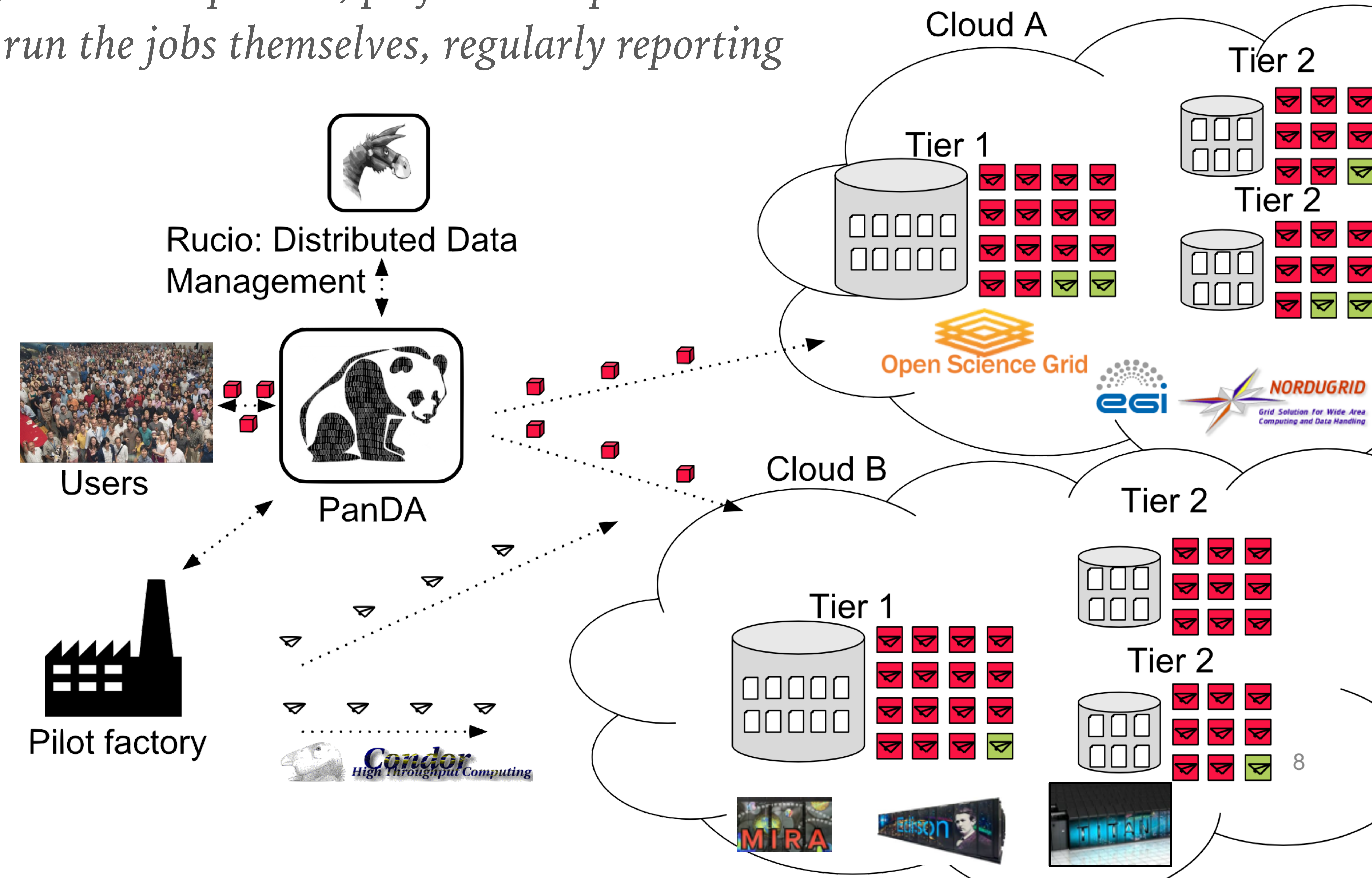

Completed jobs
365 Days from Week 05 of 2015 to Week 05 of 2016
2,500,000
2,000,000
1,500,000
1,000,000
500,000
0
Feb 2015 Mar 2015 Apr 2015 May 2015 Jun 2015 Jul 2015 Aug 2015 Sep 2015 Oct 2015 Nov 2015 Dec 2015 Jan 2016 Feb 2016
USA
GERMANY
CANADA
SPAIN
DENMARK, FINLAND, NORWAY, SWEDEN
ROMANIA
AUSTRALIA
CHILE
SOUTH AFRICA
ARMENIA
UK
NETHERLANDS
ITALY
RUSSIA
CZECH REPUBLIC
ISRAEL
CHINA
PORTUGAL
TURKEY
ARGENTINA
SWITZERLAND
FRANCE
SLOVENIA
TAIWAN
JAPAN
POLAND
GREECE
SLOVAKIA
AUSTRIA
None
Maximum: 2,282,188 , Minimum: 390,521 , Average: 1,035,340 , Current: 961,674


Slots of Running Jobs
365 Days from Week 05 of 2015 to Week 05 of 2016
300,000
250,000
200,000
150,000
100,000
50,000
0
Feb 2015 Mar 2015 Apr 2015 May 2015 Jun 2015 Jul 2015 Aug 2015 Sep 2015 Oct 2015 Nov 2015 Dec 2015 Jan 2016 Feb 2016
USA
GERMANY
ITALY
DENMARK, FINLAND, NORWAY, SWEDEN
SPAIN
ROMANIA
CHINA
PORTUGAL
SOUTH AFRICA
ARGENTINA
SWITZERLAND
FRANCE
CANADA
NETHERLANDS
CZECH REPUBLIC
ISRAEL
AUSTRALIA
GREECE
TURKEY
ARMENIA
UK
TAIWAN
RUSSIA
SLOVENIA
POLAND
JAPAN
SLOVAKIA
CHILE
AUSTRIA
None
Maximum: 257,183 , Minimum: 71,967 , Average: 176,091 , Current: 170,385

PanDA

PanDA Pilot - the execution environment (effectively a wrapper) for PanDA jobs. Pilots request and receive job payloads from the dispatcher, perform setup and cleanup work surrounding the job, and run the jobs themselves, regularly reporting status to PanDA during execution

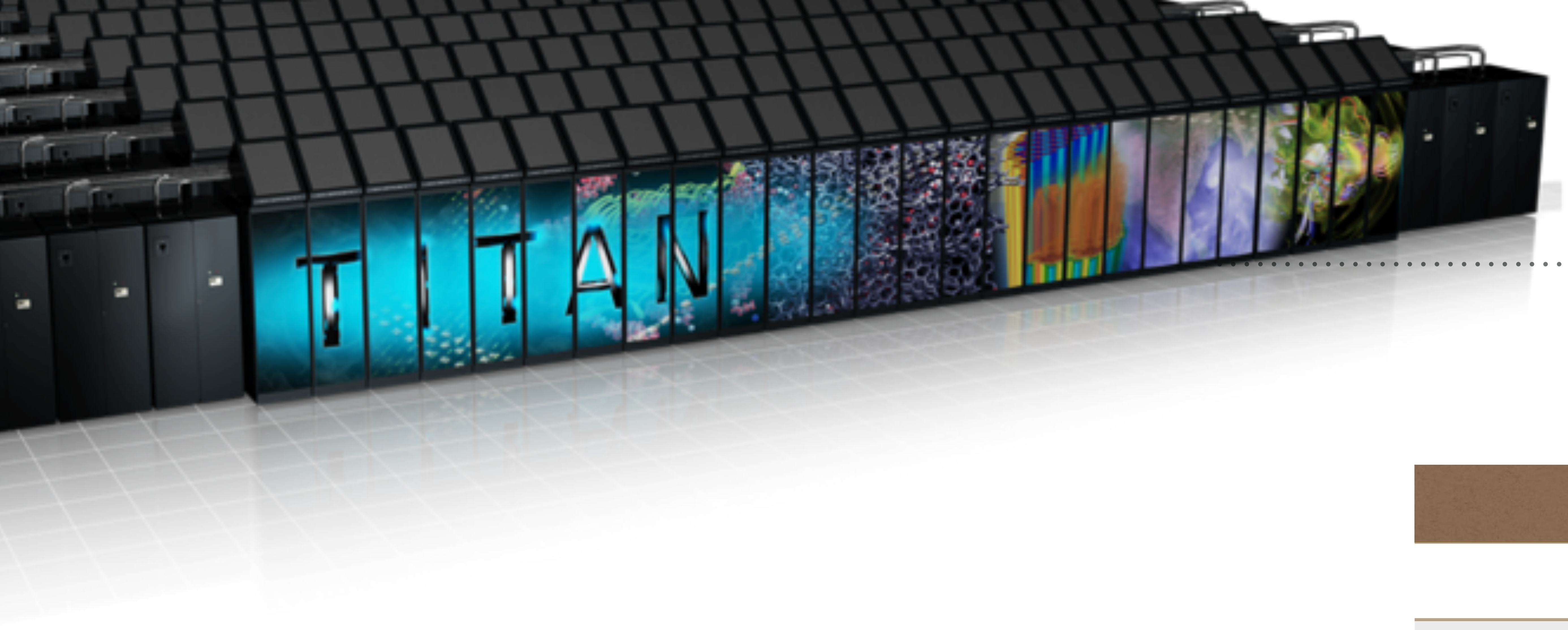
The interest in PanDA by other big data sciences brought the primary motivation to generalize PanDA, aka the BigPanDA project, providing location transparency of processing and data management, for High Energy Physics community and other data-intensive sciences, and a wider exascale community.



supercomputer, it is hard to define,
but you know it when you see it
Seymour Cray

SUPERCOMPUTERS IN ATLAS

- *The ATLAS collaboration have members with access to supercomputers, like opportunistic resources: project or personal allocations*
- *Two types of architectures (according to TOP500):*
  - *Cluster - may be served like GRID site, from PanDA point of view;*
  - *MPP (Massive parallel processing) - require special treatments due to architecture and policy restrictions*
- *Each supercomputer is unique*
  - *Unique architecture and hardware*
  - *Specialized Operating System, “weak” worker nodes, limited memory per worker node*
  - *Own usage policy*



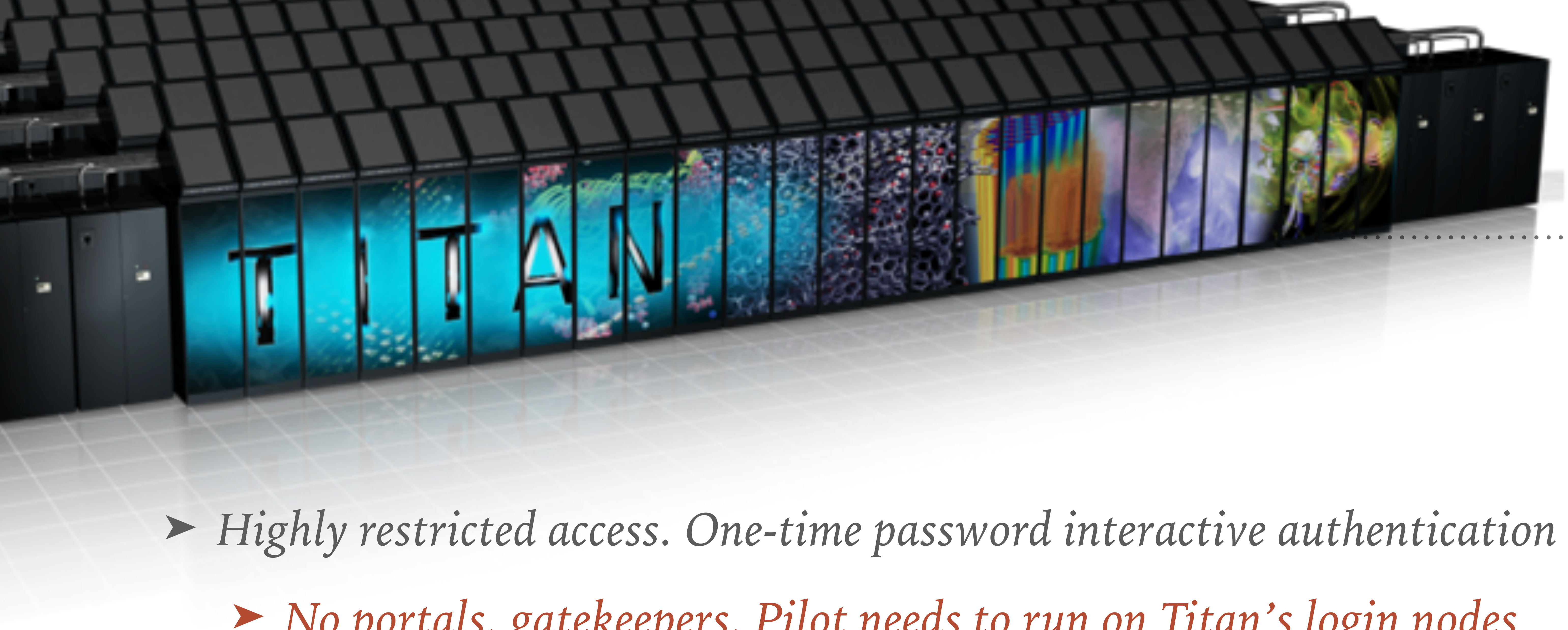
If you were plowing a field, which would you rather use? Two strong oxen or 1024 chickens?

Seymour Cray

#3 TOP 500 The List.

RMAX (TFLOP/S)	RPEAK (TFLOP/S)
17173.2	20132.7

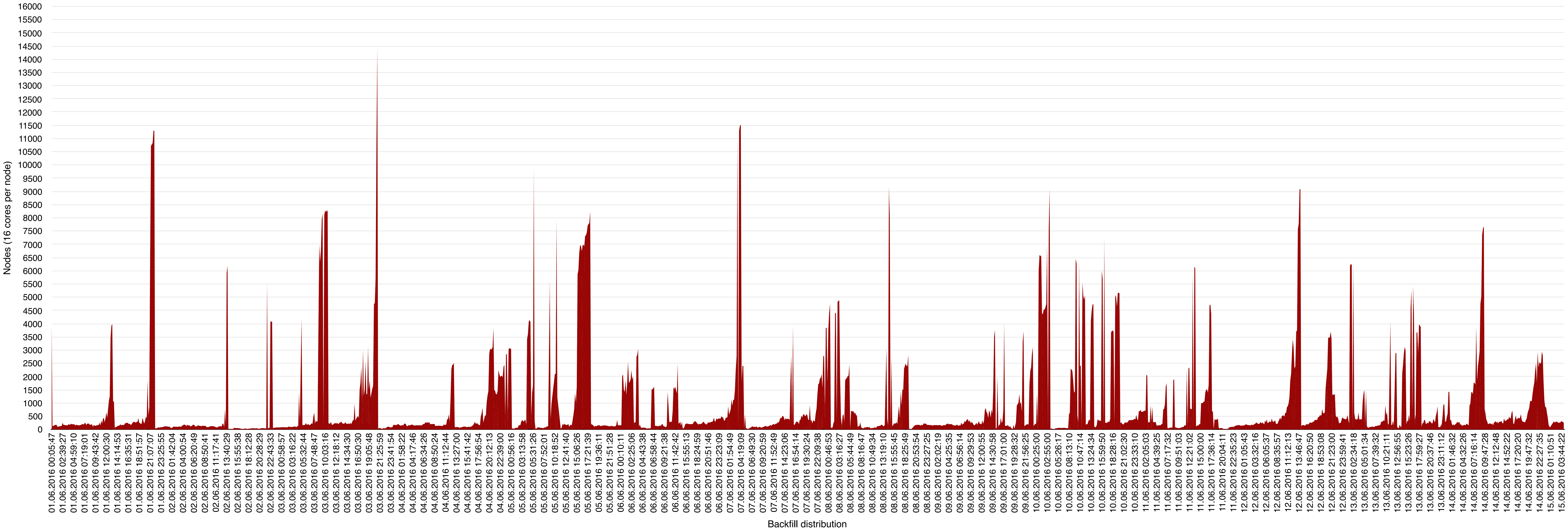
Titan System Configuration	
Architecture:	Cray XK7
Processor:	16-Core AMD
Cabinets:	200
Nodes:	18,688 AMD Opterons
Cores/node:	16
Total cores:	299,008 Opteron Cores
Memory/node:	32GB + 6GB
Memory/core:	2GB
Interconnect:	Gemini
GPUs:	18,688 K20X Keplers
Speed:	27 PF
Square Footage	4,352 sq feet



- *Highly restricted access. One-time password interactive authentication*
  - *No portals, gatekeepers. Pilot needs to run on Titan's login nodes*
- *No network connectivity from worker nodes to the outside world*
  - *Pilot can not run on worker nodes, needs a new mechanism for batch workload management*
- *Limit on number of submitted jobs in batch queue per user and limit on number of running jobs per user*
  - *Sequential submissions of single node jobs is not an option*
  - *Have to use MPI in some form!*
- *Specialized OS (SUSE based CNL) and software stack*
- *Highly competitive time allocation. Geared toward leadership class projects and very big jobs*
  - *Creates opportunity for backfill*

«BACKFILL»

- ***Typical LCF facility is running on average at ~90% occupancy***
  - *Necessary outcome of prioritizing large jobs execution*
  - *On a machine of the scale of Titan that translates into ~300M unused core hours per year*
- ***Anything that helps to improve this number consistent with LCF mission is very useful***

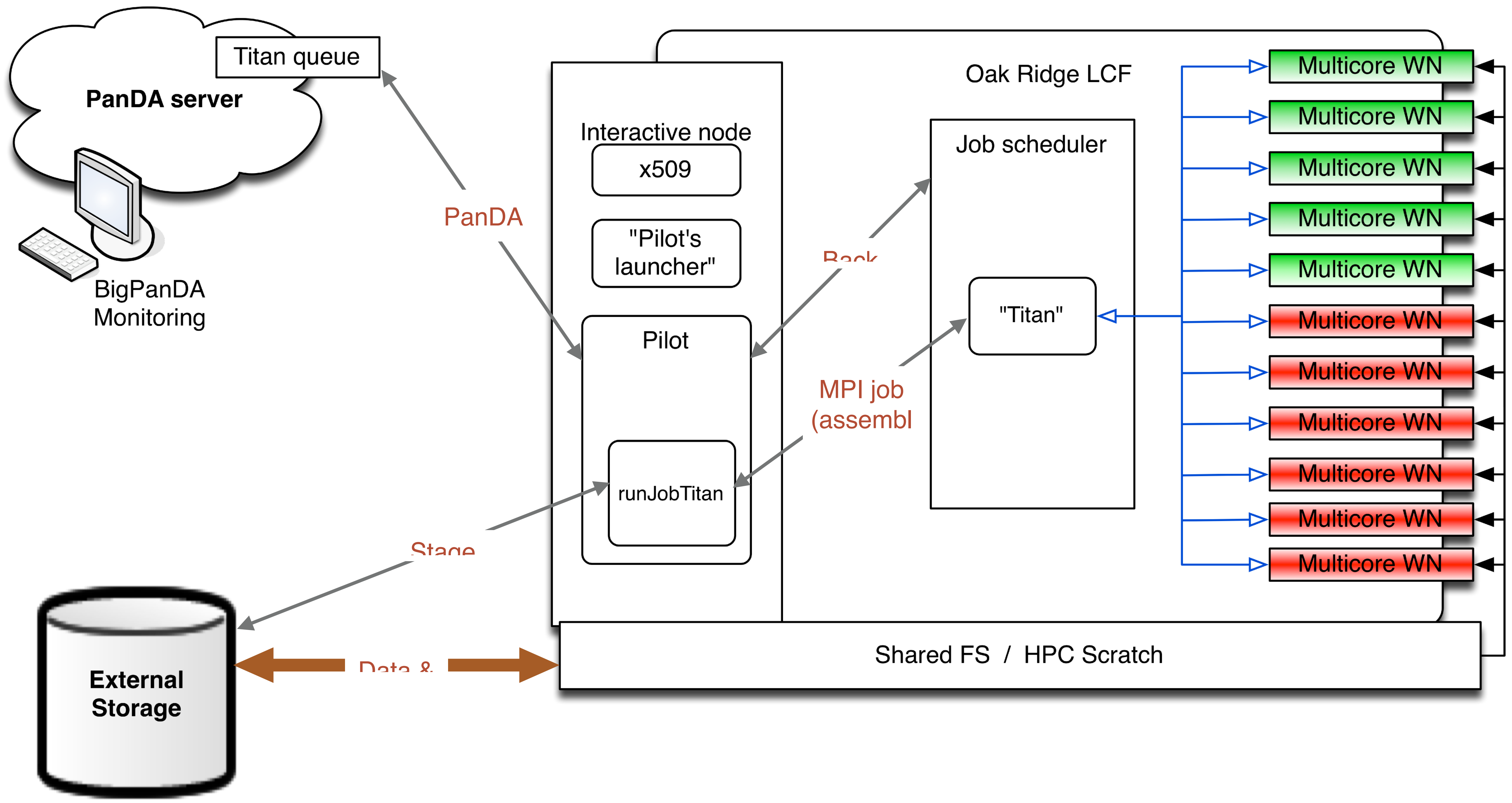


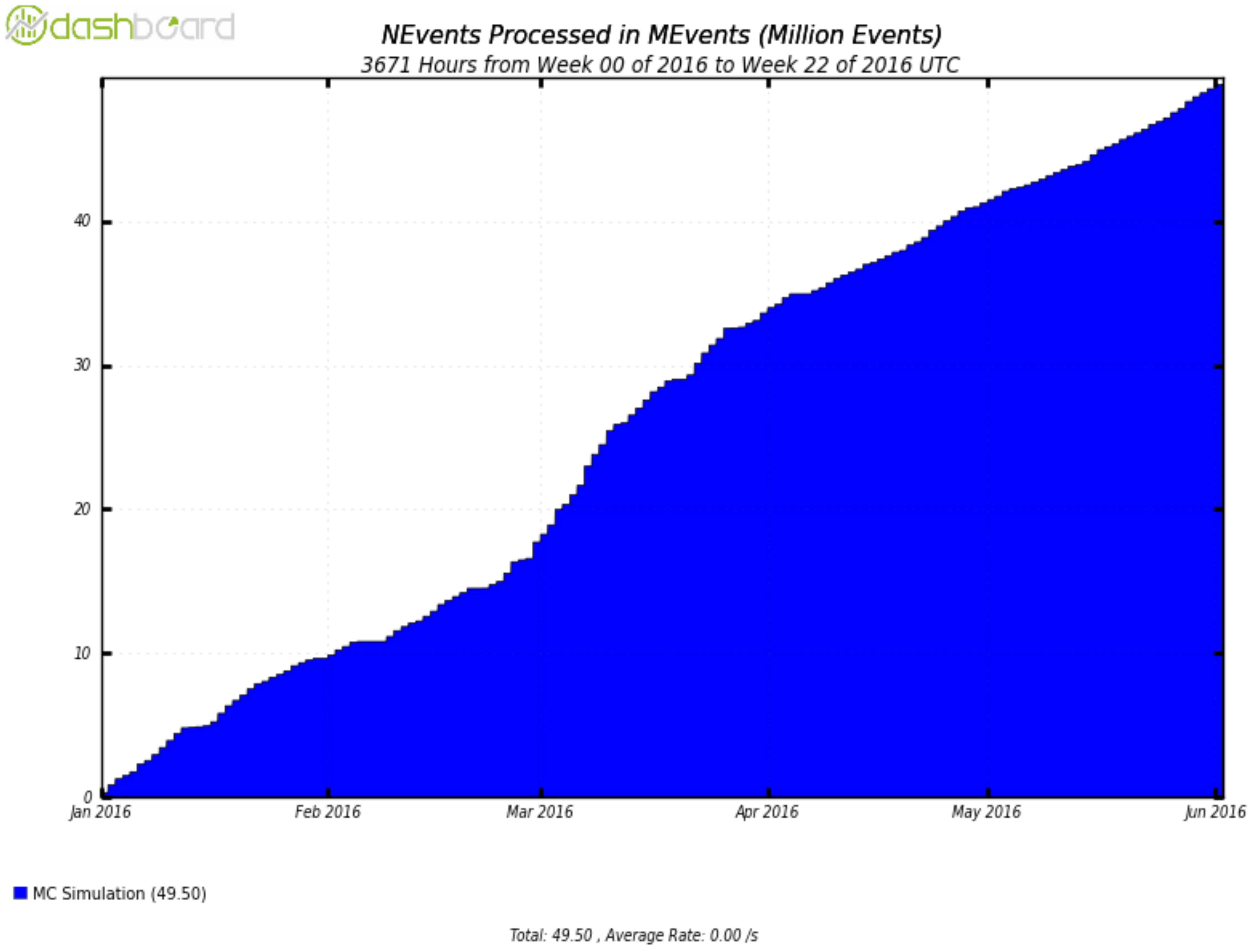
«BACKFILL» TREATMENT THROUGH PANDA

- *PanDA Pilot was instrumented with capability to collect, in near real time, information about current free resources on Titan*
  - *Both number of free worker nodes and time of their availability*
- *Based on that information Pilot can define job submission parameters when forming PBS script for Titan, thus tailoring the submission to the available resource.*
  - *Takes into account Titan's scheduling policies*
  - *Can also take into account other limitations, such as workload output size, etc*
  - *Modular architecture, adaptable to other HPC facilities*
- ***OLCF allow ATLAS to use Titan in «over allocation» mode, while ATLAS jobs executes only on «backfill» resources.***

PanDA pilot workflow for Titan (MultiJob Pilot)

In production since June 2015

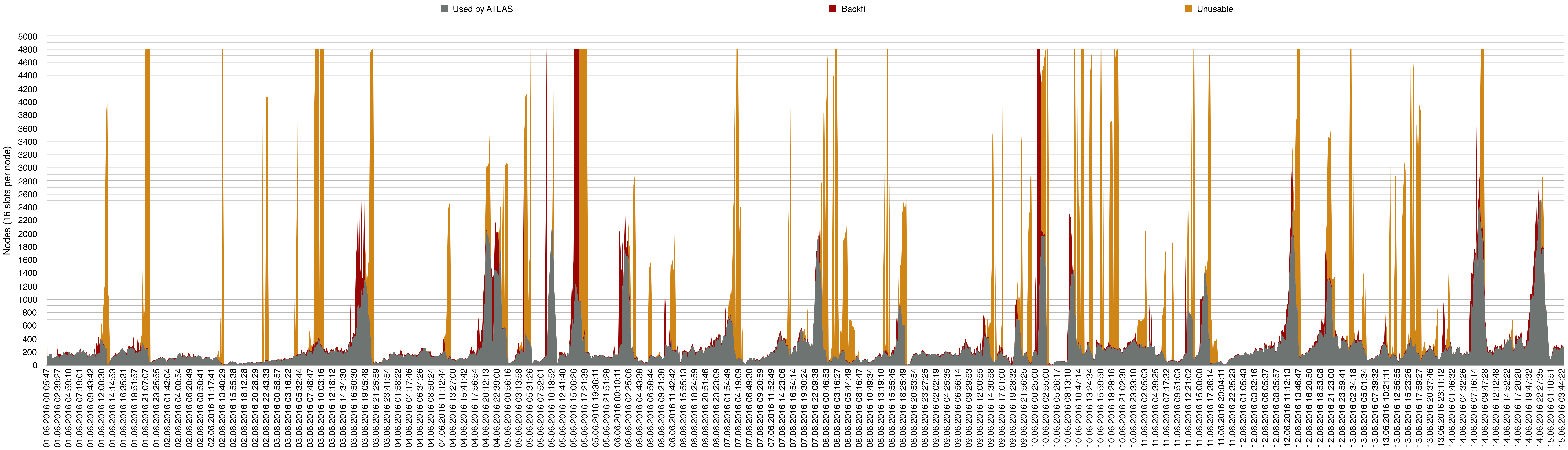




~50 000 000 events processed on Titan for ATLAS during first 6 month 2016

> 3M CPU/hours consumed each month

HOW CAN WE IMPROVE?



- *Current implementation allow us consume ~15% of available backfill on Titan (still miss millions of CPU/hours)*
- *Why we miss resources:*
  - *Stage In/Out latency*
  - *Current ATLAS payloads do not fit optimally in available transient free resources on Titan.*
    - *we can't use gaps less than 2 hours (for the moment)*

TRADITIONAL WAY OF DATA PROCESSING

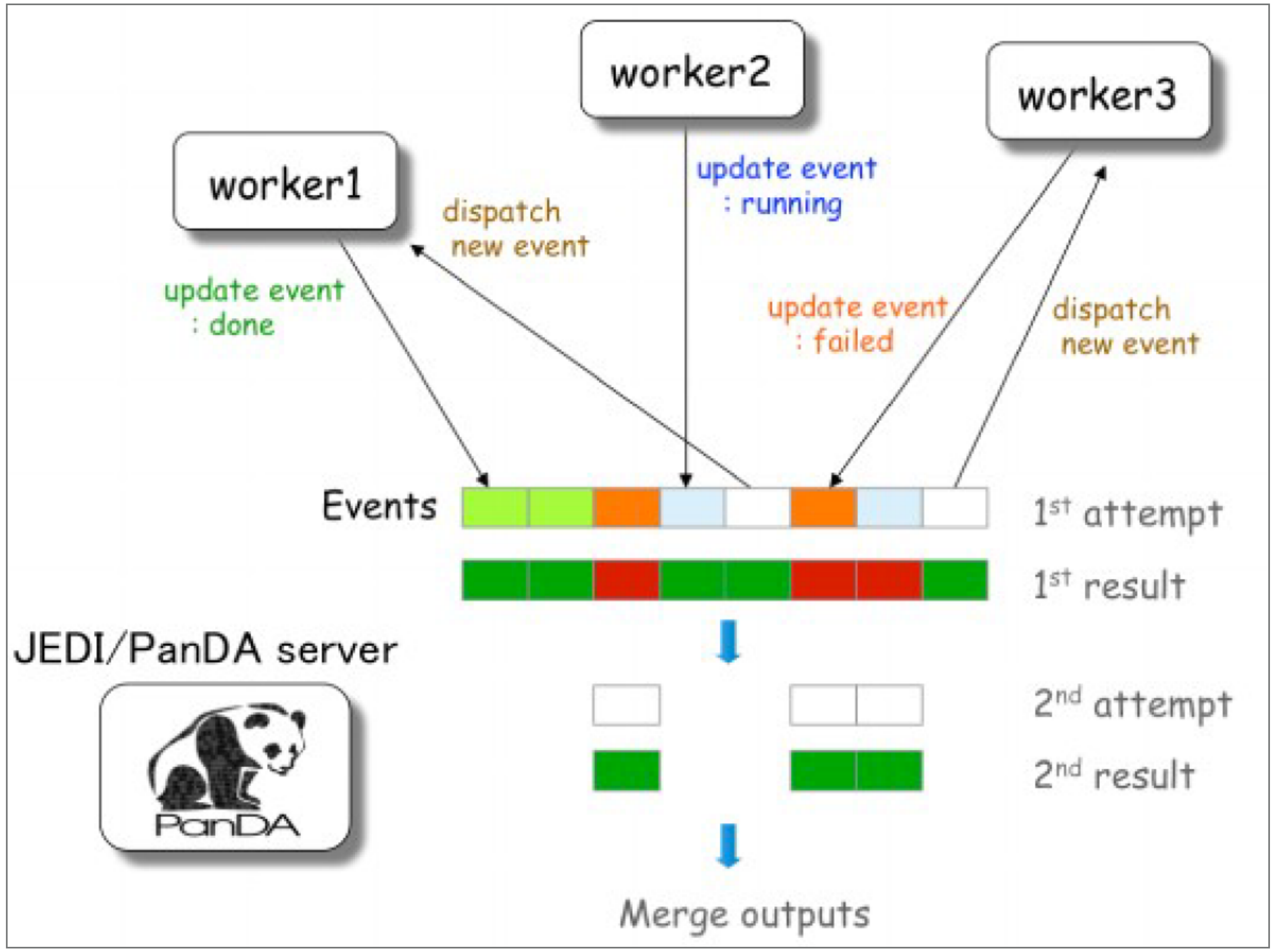
- *A typical data processing job (e.g. simulation, reconstruction, data analysis) is required to process a number of events placed continuously in one or more input files*
  - *Event – a minimal data processing unit*
- *Pilot assigns all input events to the given job at configuration time, launches the job and waits until the job is completed*
- *If processing of one event fails*
  - *The entire job is terminated*
  - *All events processed so far are discarded*
  - *The job is retried*

THE ATLAS EVENT SERVICE (AES)

- *Event Service – a new approach to HEP data processing*
- *Job granularity changes from files to individual events*
  - *Quasi-continuous event streaming through worker nodes*
  - *Agile, dynamic tailoring of workloads to fit the scheduling opportunities of the moment*
  - *Termination with minimal losses*
- *Real time delivery of fine-grained workloads to running application allows us to exploit event processors efficiently through their lifetime*
- *By promptly streaming away the outputs we can minimize local storage demands*

BUILDING BLOCKS OF THE ES

- *The ES Engine: PanDA and its extension JEDI*
  - *JEDI supports new functionality of flexible task management and fine-grained dynamic job management*
- *Parallel payload (event processing component)*
  - *Efficient usage of CPU and memory resources on the compute node*
  - *Whole-node scheduling*
- *Remote I/O*
  - *Efficient delivery of event data to compute nodes*
- *Object stores*
  - *Efficient management of outputs produced by the ES*



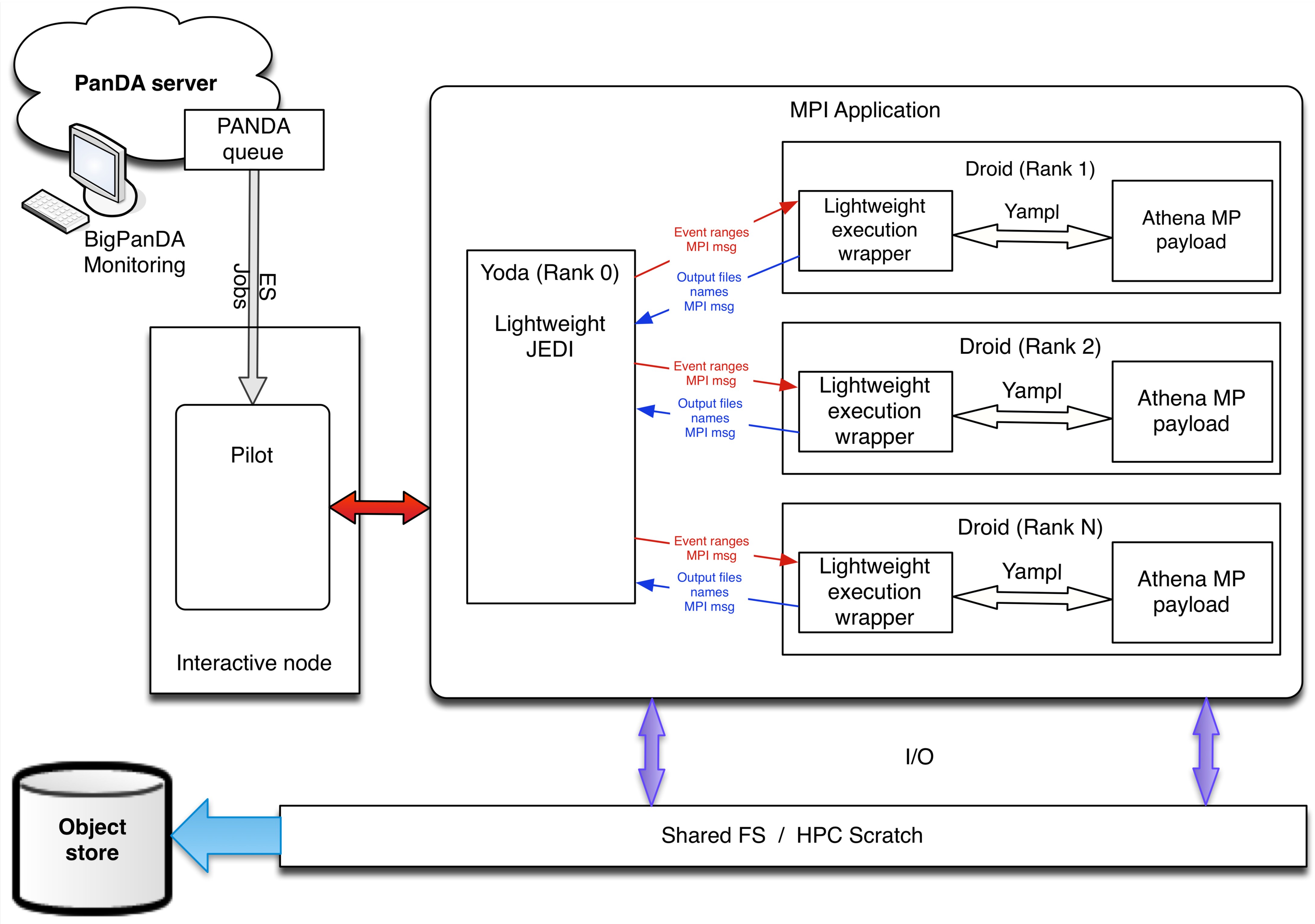
YODA: EVENT SERVICE ON HPC

- ***Yoda*** *– HPC-internal version of the Event Service.*
  - *Reuses the code of the conventional Event Service wherever possible*
  - *Otherwise implements lightweight analogs of the ES components*
    - ***Yoda*** *– a lightweight JEDI*
    - ***Droid*** *– a lightweight Pilot*
- *Yoda has been commissioned on Edison and is now running production workloads on Edison.*

Edison is a **Cray XC30**, with a peak performance of 2.57 petaflops/sec (#49 in TOP500)
133,824 compute cores (Intel Xeon E5-2695v2 12C 2.4GHz),
357 terabytes of memory,
7.56 petabytes of disk.

YODA. SCHEMATIC VIEW

- *MPI application implementing master – slave architecture*
- *Rank 0 (Yoda, master)*
  - *Distributes workload between slave ranks*
- ***Fine grained workload**: individual events or event ranges*
- *Rank N (Droid, slave)*
  - *Occupies entire compute node;*
  - *Processes assigned workload;*
  - *Saves outputs to the shared file system;*
  - *Asks for the next workload …*
- ***Payload component**: AthenaMP – multi-process version of the ATLAS simulation, reconstruction and data analysis framework Athena*



ISSUES

- *Scale (it's normal when you moves from «chicken» to «oxen»)*
    - *Each ATLAS job generates a lot of IO, especially during startup.*
    - *Stage-in and Stage-out consecutive operations and for big amounts of jobs may take significant time. During this operation some backfill resources will be missed*
- *Initial implementation of PanDA components was not oriented for serving bunches of jobs*
    - *Bulk operations with jobs was not implemented on server side (since it was not needed)*

PLANS

- ***Commissioning of Yoda on Titan.***
- *Valuable volume of backfill resources appears mostly on LCF class machines due to scheduling policy. Usage of backfill was not effective at NERSC, so this functionality should be adopted for Yoda@Titan*
- *New PanDA service **Harvester** will be implemented to cover HPC related workflows (and something more):*
  - *Implementation of complicated workflows, like MultiJob and Yoda in PanDA pilot, significantly increase complexity of application. This dramatically affects future support of Pilot.*
  - *«Late binding» on supercomputers is a tricky*
  - *Proper Data management for thousands of simultaneously executed jobs day by day requires special attention*

CONCLUSION

- *PanDA WMS was successfully integrated with most powerful computing resources on the world*
  - *but scale and complexity will increase, so we should move forward with proper solution*
- *ATLAS experiment started production usage of this resources with valuable results*
  - *up to 7% of simulated events were produced on HPC this year*
- *Still have room for improvements!*
  - *increasing of computational scale will affect storage management and data transfers*