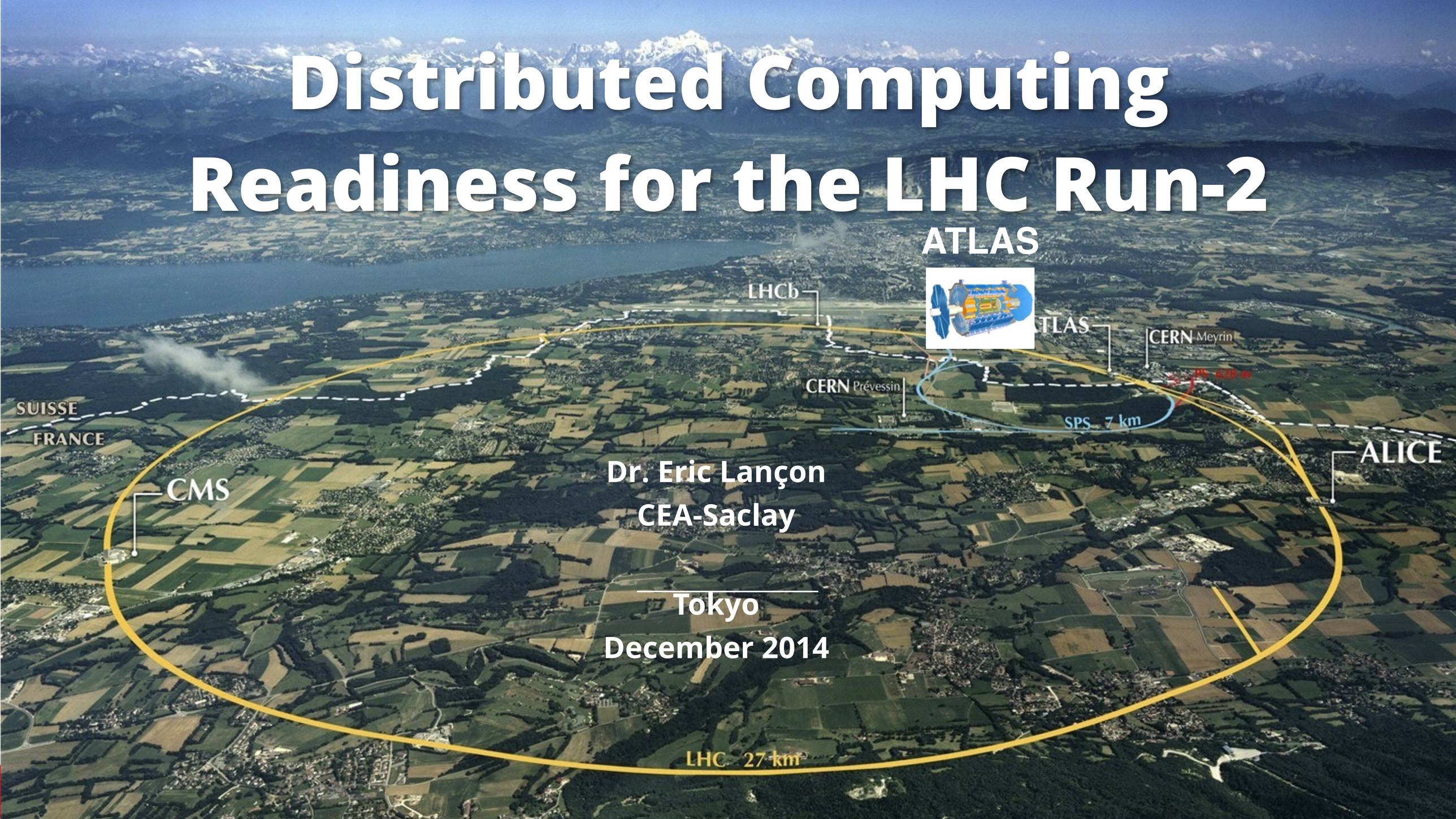
Distributed Computing Readiness for the LHC Run-2

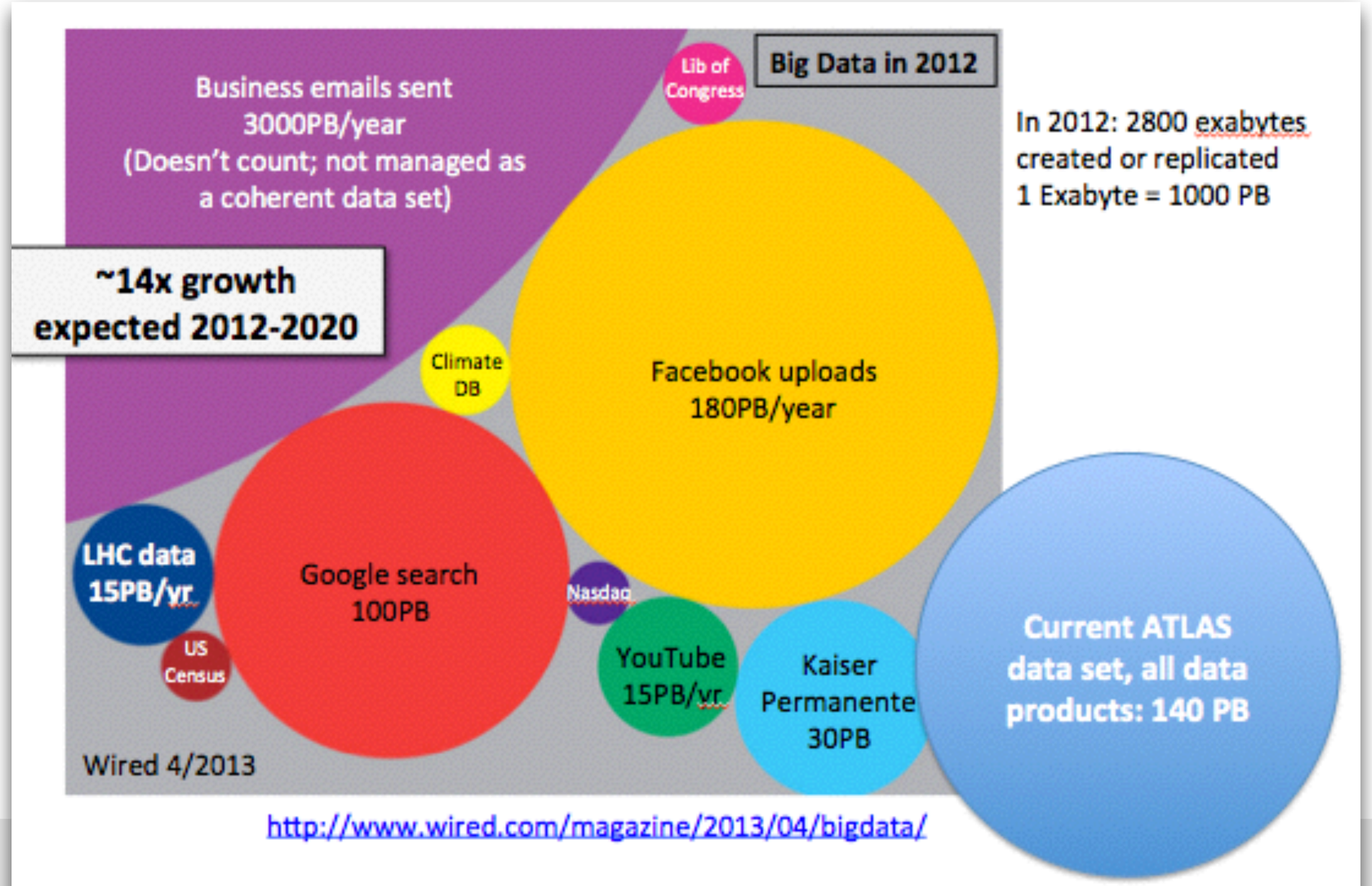


Dr. Eric Lançon
CEA-Saclay

Tokyo
December 2014

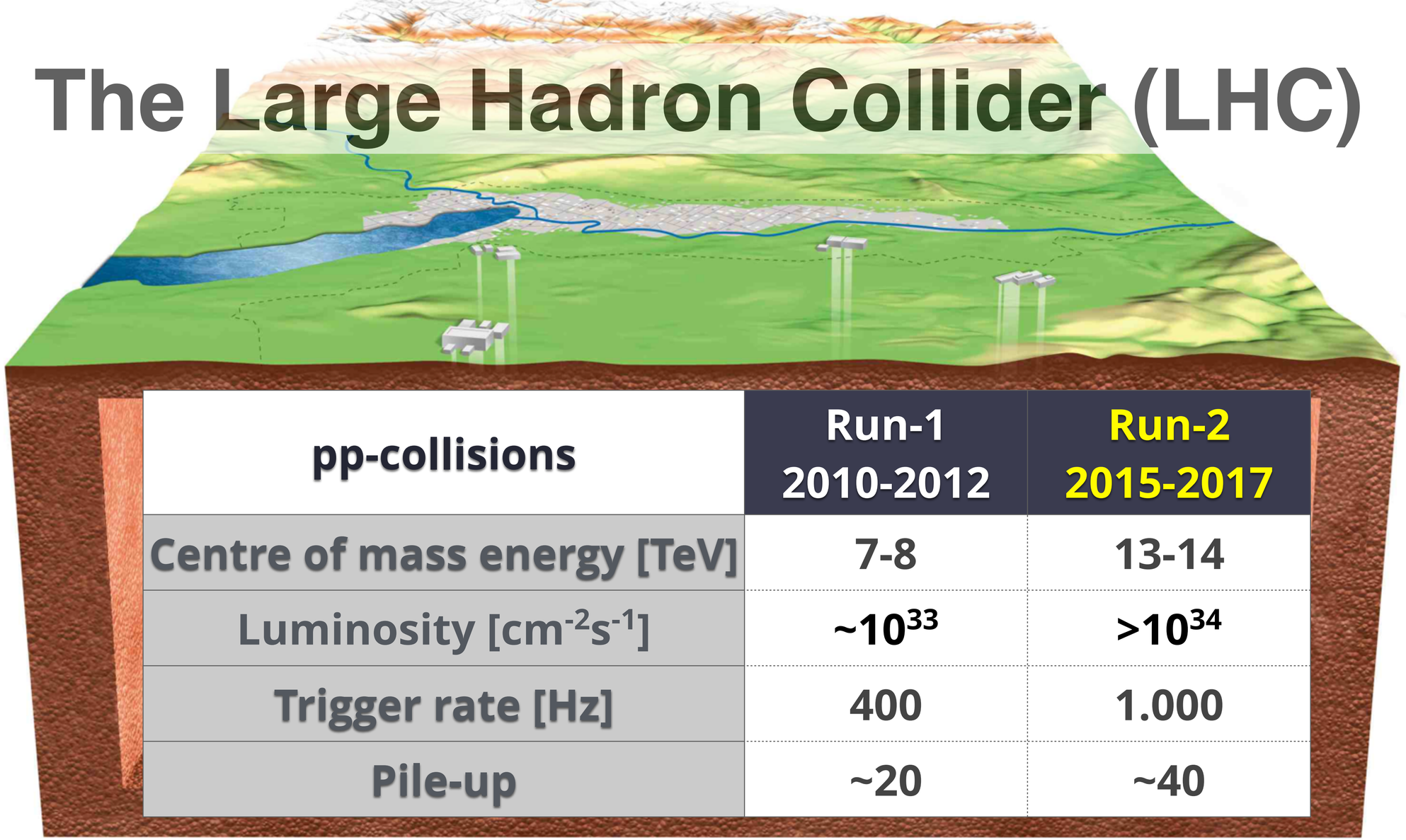

This is ATLAS

ATLAS is big data experiment



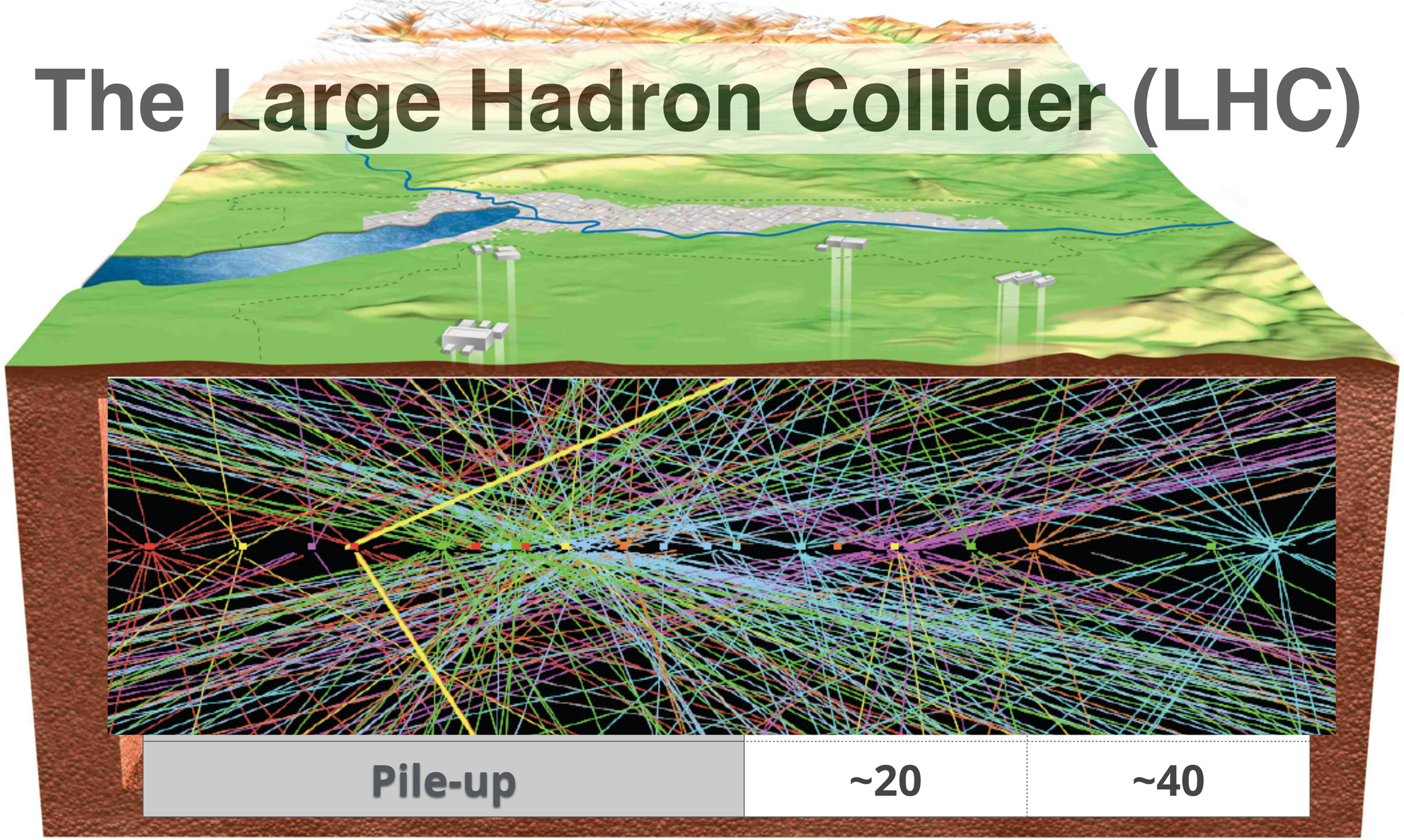
http://www.wired.com/magazine/2013/04/bigdata/

The Large Hadron Collider (LHC)



pp-collisions	Run-1 2010-2012	Run-2 2015-2017
Centre of mass energy [TeV]	7-8	13-14
Luminosity [$cm^{-2}s^{-1}$]	~10^{33}	>10^{34}
Trigger rate [Hz]	400	1.000
Pile-up	~20	~40

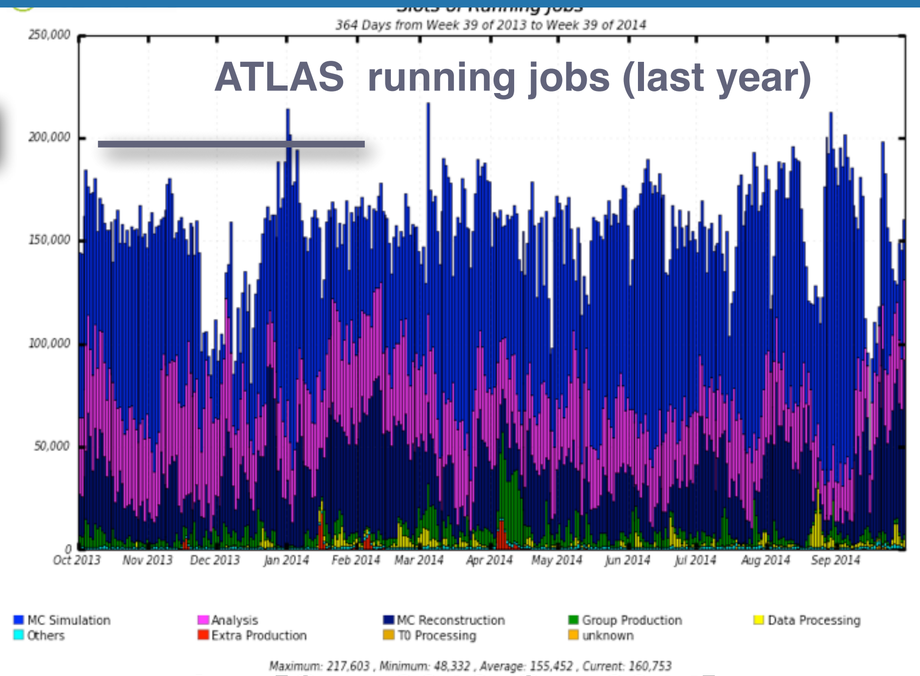
The Large Hadron Collider (LHC)

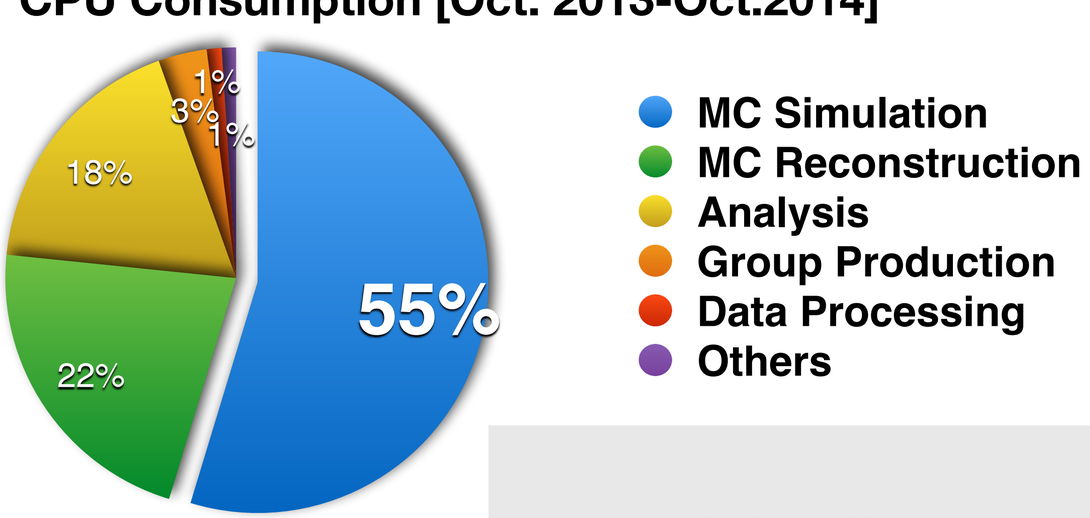


Pile-up	~20	~40

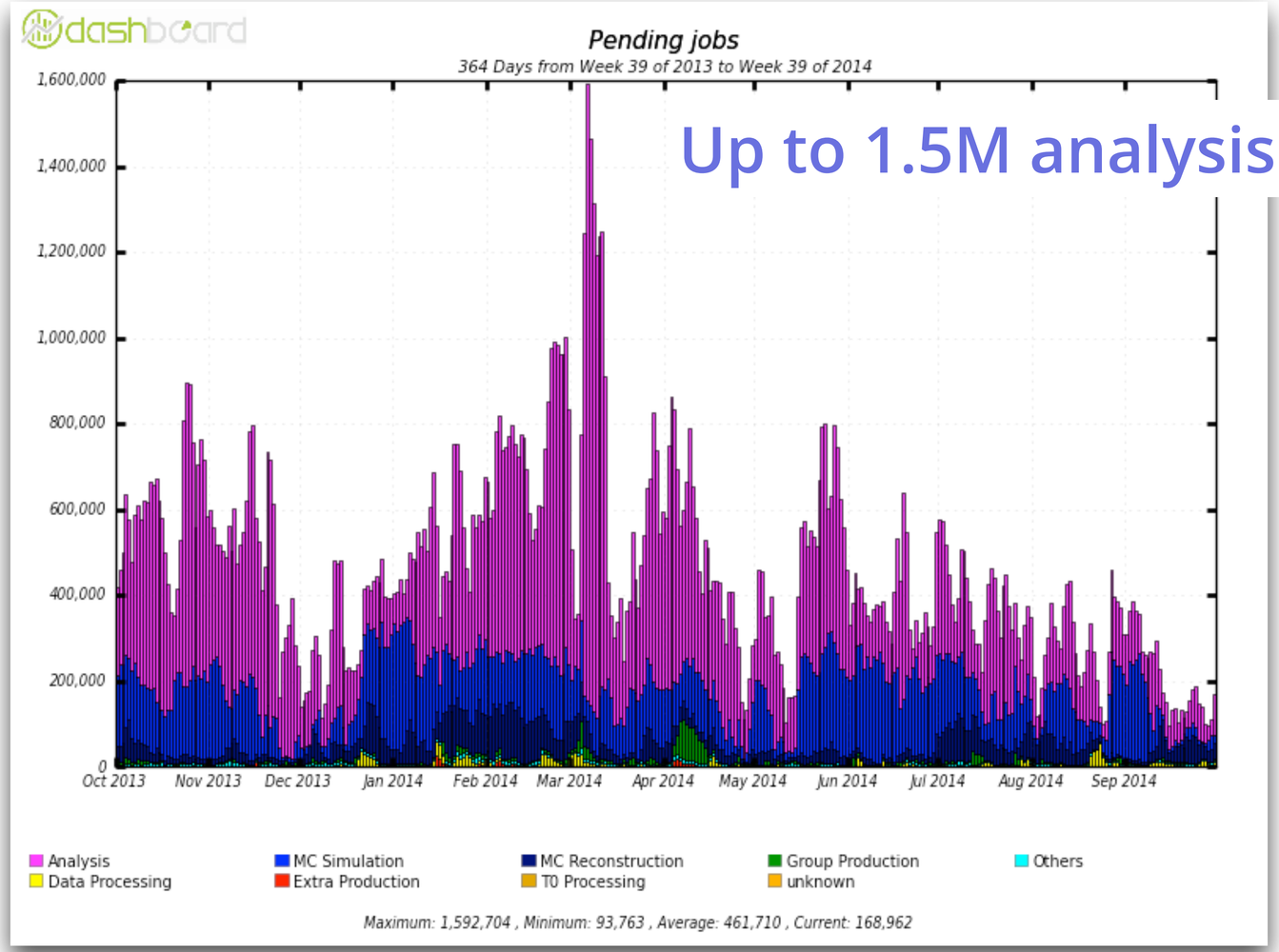
ATLAS Grid activity

- ~170K concurrent jobs running
- 350M jobs completed in 2013
- **1.2 EB** of data read-in by ATLAS grid jobs in 2013
- **Analysis** is the main driver of storage & network I/O capacity





But at the same time



Up to 1.5M analysis jobs waiting…

Improvements needed for Run-2

Some limitations of current model & tools

- Partitioning of resources
  - User analysis vs Central Production
  - T1s vs T2s
- Difficulties of current Data Distribution Management & production systems to accommodate new use cases and technologies
- Memory increase of MC pile-up simulation & reconstruction
- Full reprocessing once a year only
- Multitude of data formats for analysis

The Challenges of Run-2

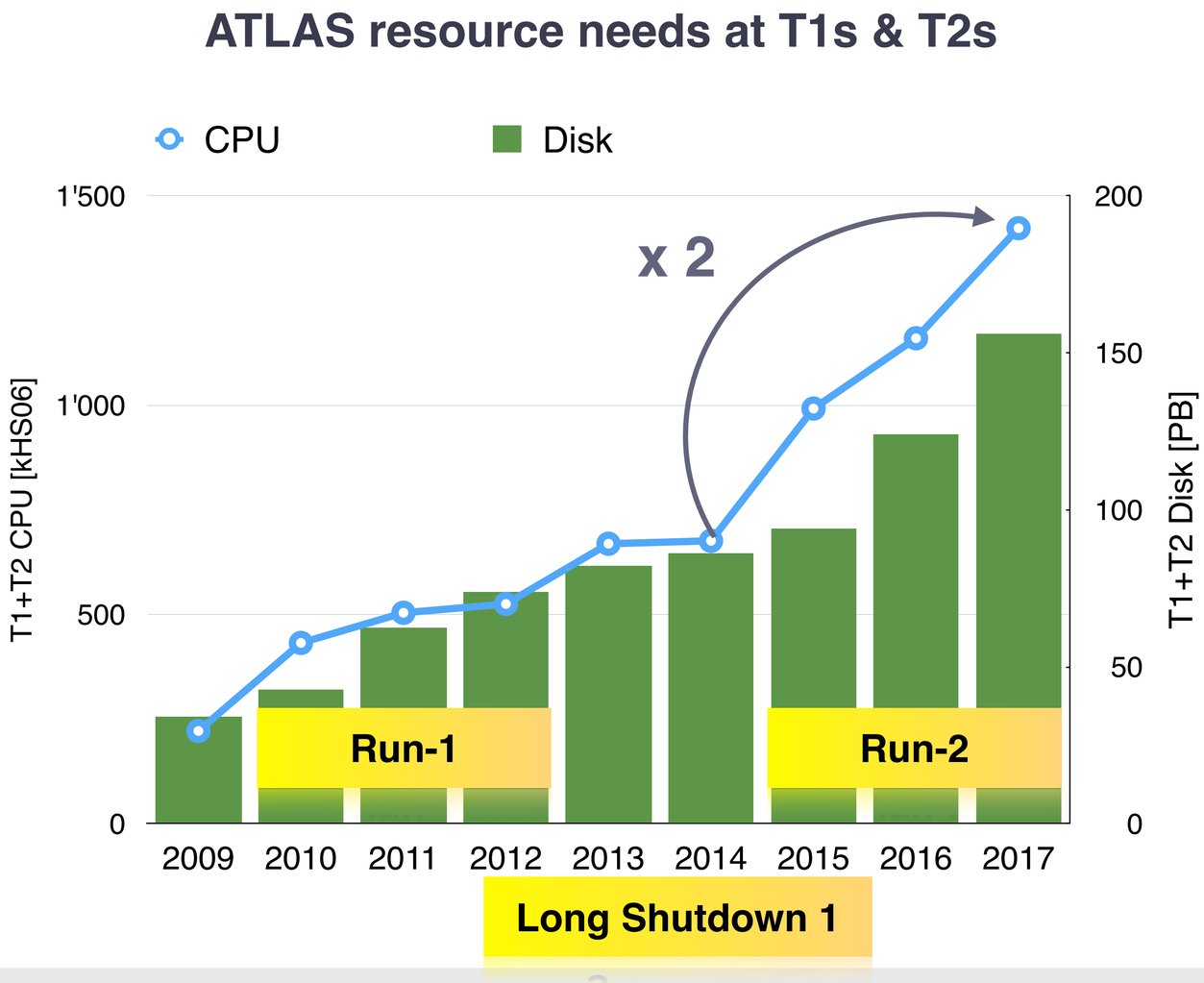
- Constraints of ‘flat budget’
  - Both for hardware and for operation and development
  - Hardware increase from Moore’s law gain only, estimated at factors of 1.2/year for CPU and 1.15/year for disk
- Data from Run1

- ~new detector
- Factor 2-3 to gain in reconstruction speed
- Reduction of memory requirement
- Increased use of fast simulation

With shrinking manpower (-10% in 2013)

Resources for Run-2

- **Resource estimates for 2016 & 2017 are still preliminary**
  - Profile of hardware replacement not taken into account in ‘flat budget’ hypothesis yet
  - Introduction of dataset lifetime both on disk and tape : more tape I/O and possibly more tape volume needed
  - Balance of disks between T1s & T2s to be optimised
  - Optimistic use of fast simulation in resource planning?

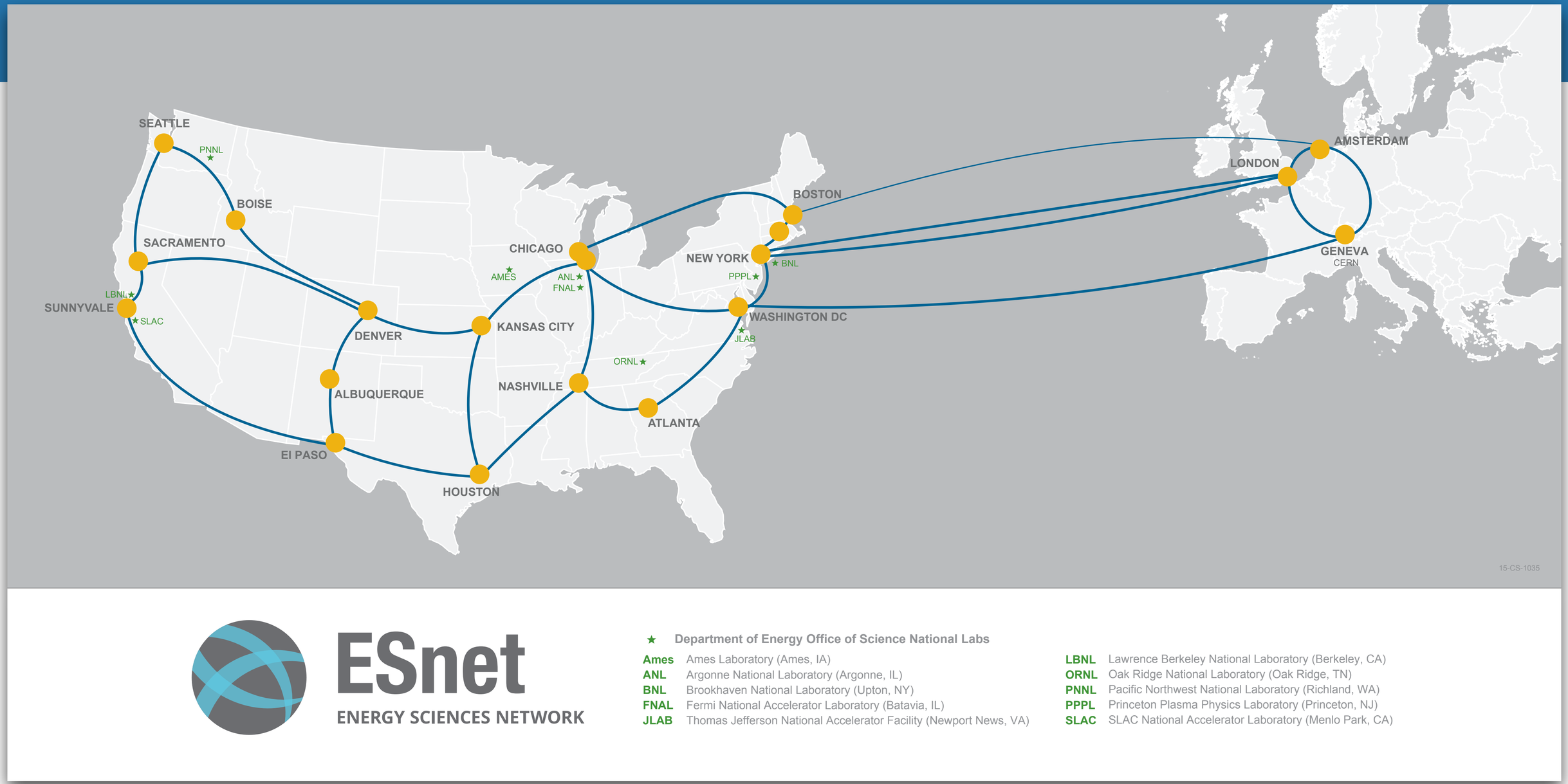


Software challenges

- Get the Integrated Simulation Framework (ISF) in production
- Speedup reconstruction by a factor 2-3
- Migration to new data format xAOD readable both from Athena and ROOT and new analysis model
- Changes/Upgrade of the infrastructure : ROOT6, CMake, Tag Collector, JIRA…

Computing Challenges

- More efficient use of resources
  - More flexibility in the computing model (Clouds/Tiers)
  - Limit avoidable resource consumption (multicore)
  - Optimize workflows (Derivation Framework/Analysis Model)
- New ATLAS distributed computing systems
  - Rucio for Data Management
  - Prodsys-2 for Workload Management
  - FAX and Event Service to optimize resource usage
- New data management strategy : each dataset has a lifetime



ESnet
ENERGY SCIENCES NETWORK

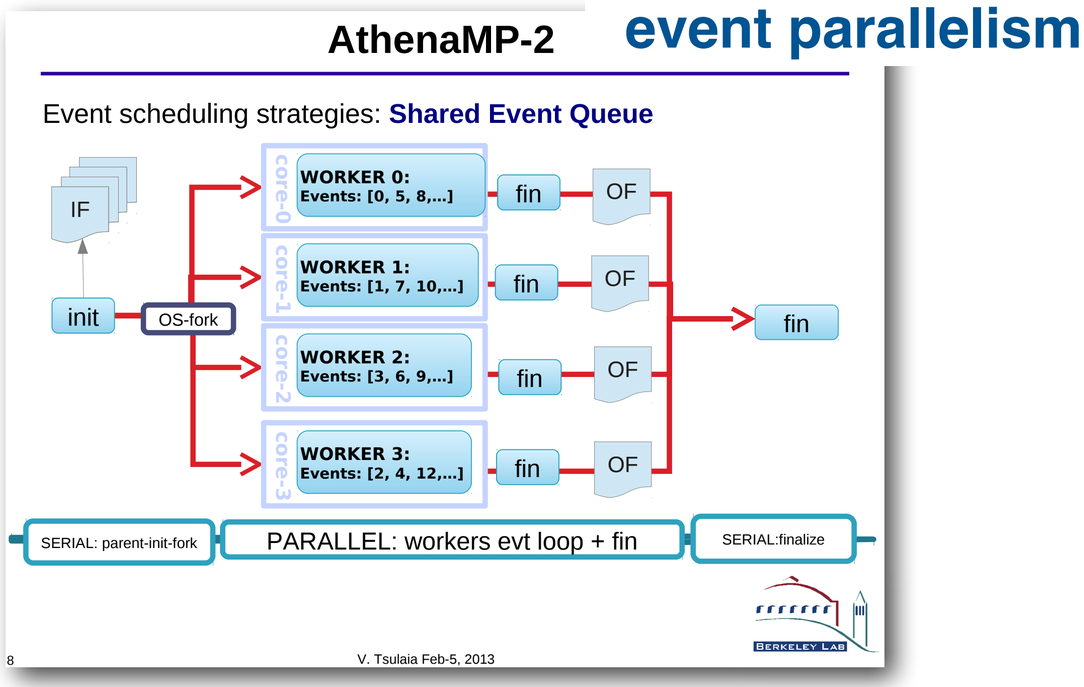
★ Department of Energy Office of Science National Labs

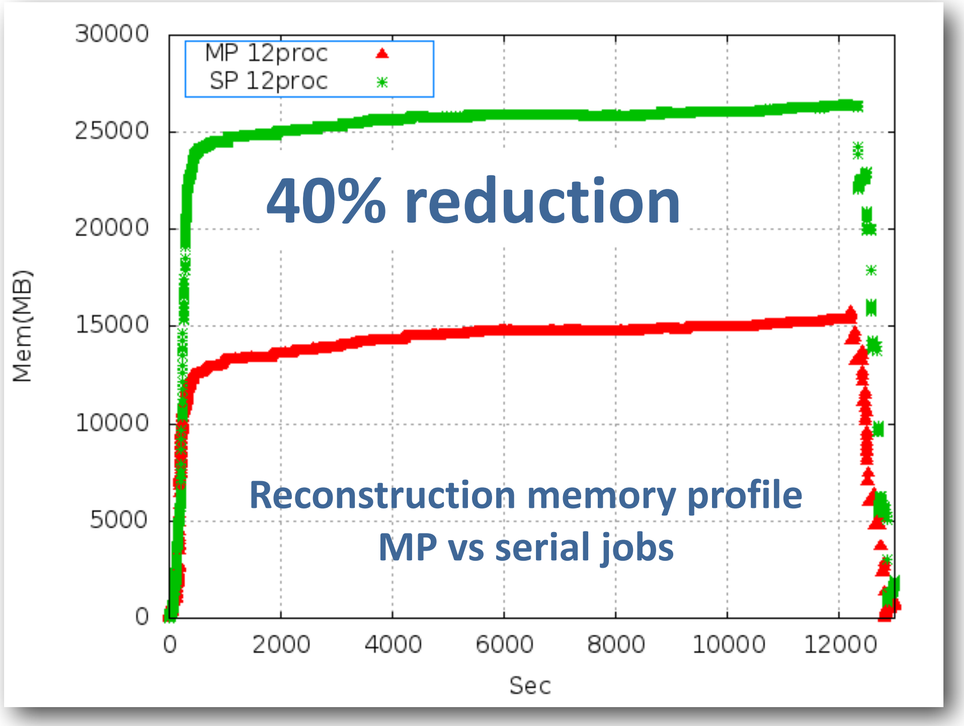
Ames	Ames Laboratory (Ames, IA)
ANL	Argonne National Laboratory (Argonne, IL)
BNL	Brookhaven National Laboratory (Upton, NY)
FNAL	Fermi National Accelerator Laboratory (Batavia, IL)
JLAB	Thomas Jefferson National Accelerator Facility (Newport News, VA)
LBNL	Lawrence Berkeley National Laboratory (Berkeley, CA)
ORNL	Oak Ridge National Laboratory (Oak Ridge, TN)
PNNL	Pacific Northwest National Laboratory (Richland, WA)
PPPL	Princeton Plasma Physics Laboratory (Princeton, NJ)
SLAC	SLAC National Accelerator Laboratory (Menlo Park, CA)

Software improvements

Memory saving

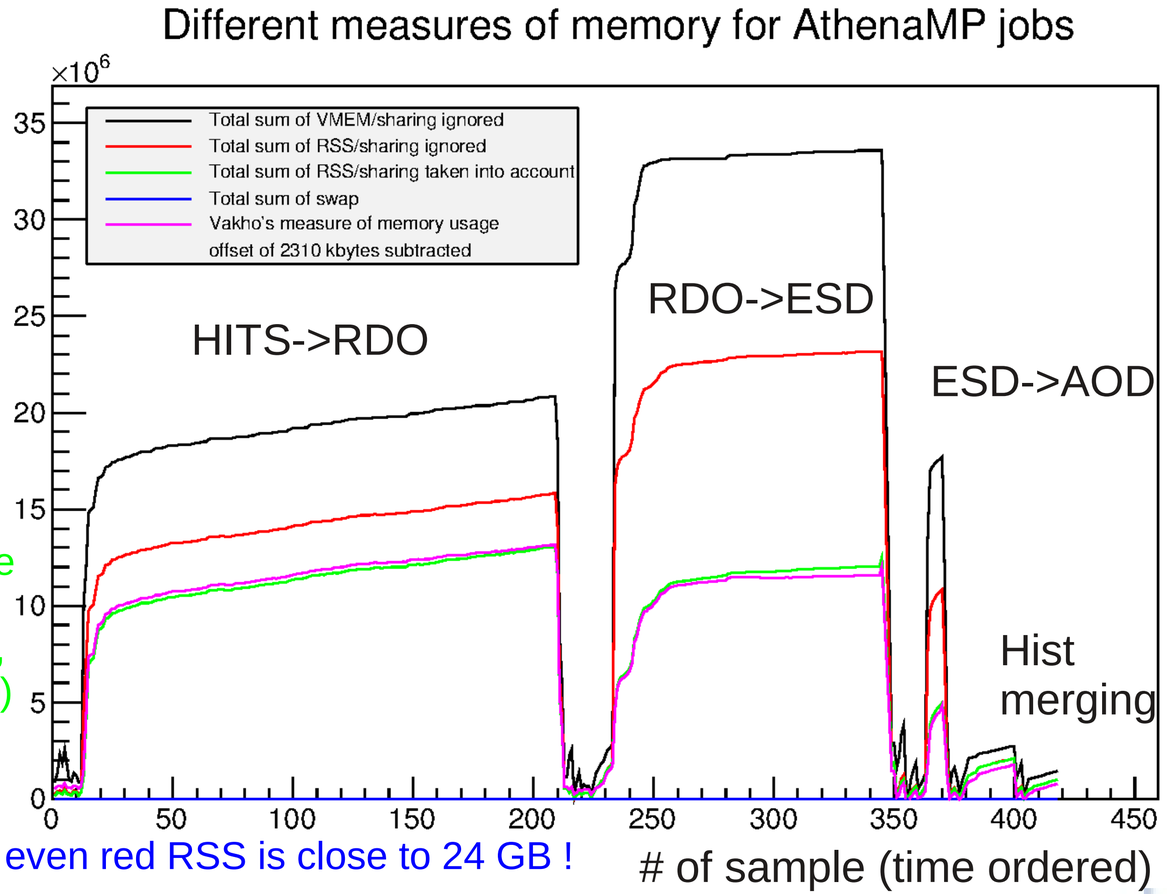
- Multi-Process Athena (**AthenaMP**) the current approach to memory saving in reconstruction and simulation jobs





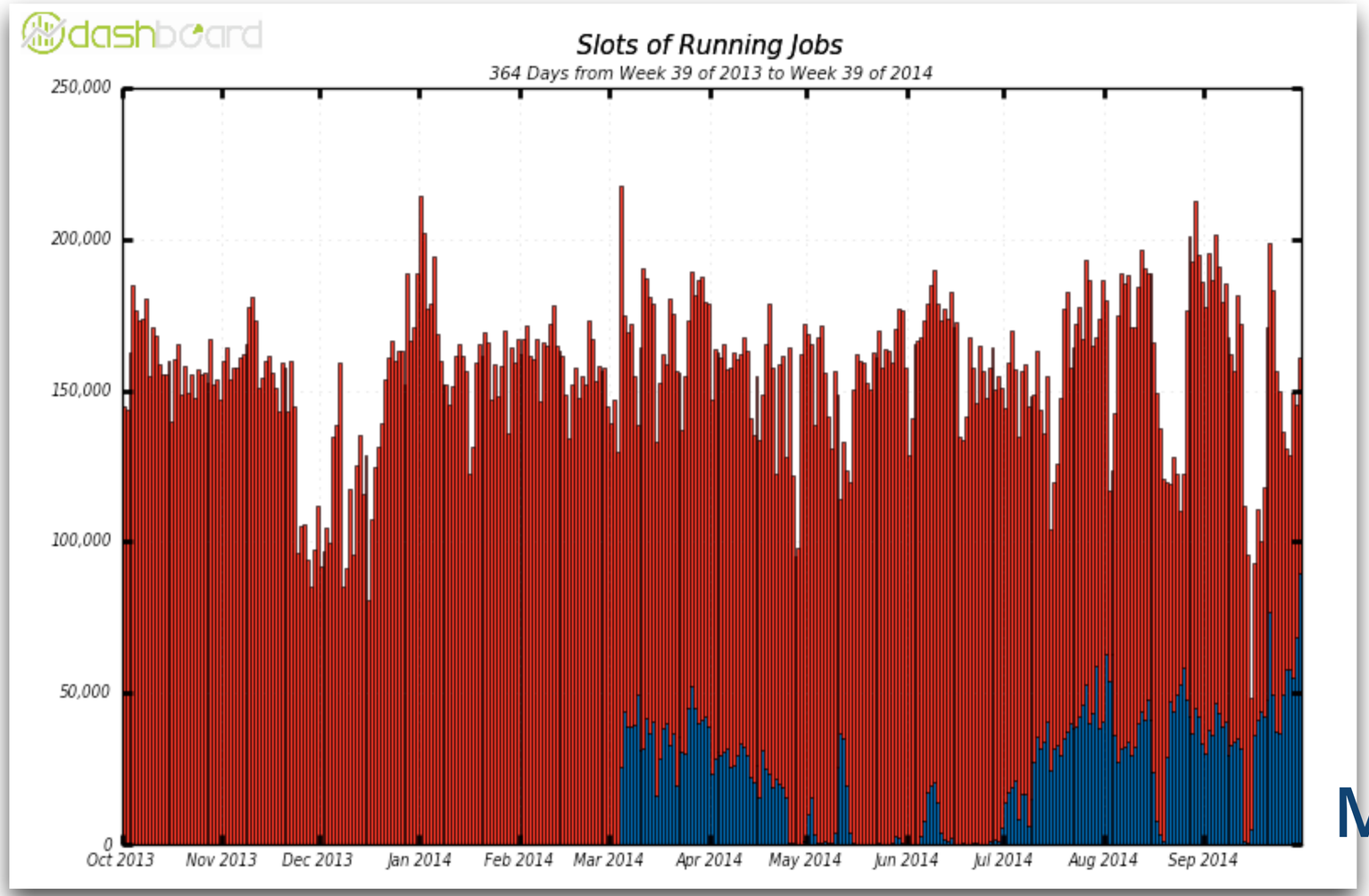
Memory usage in athenaMP in DC14 – random DC14 job

- DigiMReco with 8 workers incl. Pileup (mu = ?)
- Vmem: virtual memory, → irrelevant (some memory profilers use >16 TB of Vmem)
- RSS, sharing ignored: this is approximately how much memory N single athena jobs would use …
- RSS, accounting for sharing: the ratio of the red and the green is the gain we get from athenaMP, THIS IS USING smaps (backup)
- amount of swapping in this job; in this job no swapping, machine has 24GB of real ram, even red RSS is close to 24 GB !



- Vakho's measure of memory usage utilising 'free' to determine used, total memory of workernode – depends on what else goes on on machine so needs dedicated single user machine, might work for whole node VMs

Multi-core jobs in production



10 months

Multi-core jobs

Accounting…

Accounting

- Wallclock as itis not correctly reported in the APEL portal
  - eff > 100%
  - In WLCG accounting a mixture of cores
    - Difficult to understand what is going on
- New development portal
  - Has more selections
  - Efficiency is correct
  - In production next year

CPU Efficiency (%) by SITE and DATE

SITE	Jun 14	Jul 14	Aug 14	Sep 14	Oct 14	Nov 14	Total
EFDA-JET	0.0	15.9	84.9	89.5	90.6	66.7	18.6
RAL-LCG2	93.3	130.0	125.1	107.9	162.0	104.1	117.3
UKI-LT2-Brunel	76.9	84.7	86.5	85.6	88.5	86.8	83.7
UKI-LT2-IC-HEP	80.4	81.9	90.3	97.2	134.4	77.3	90.3
UKI-LT2-QMUL	87.1	98.7	101.1	97.5	105.9	99.4	97.4
UKI-LT2-RHUL	94.2	92.7	85.1	92.0	78.3	88.1	89.5
UKI-LT2-UCL-HEP	95.9	88.1	79.1	91.3	88.5	90.8	91.8
UKI-NORTHGRID-LANCS-HEP	91.7	102.6	115.9	162.4	261.1	123.0	136.4
UKI-NORTHGRID-LIV-HEP	97.2	97.6	101.2	106.2	124.7	128.4	105.1
UKI-NORTHGRID-MAN-HEP	95.3	96.7	108.6	90.0	81.8	92.0	95.0
UKI-NORTHGRID-SHEF-HEP	94.7	90.8	89.9	87.4	77.3	102.3	89.4
UKI-SCOTGRID-DURHAM	20.5	7.7	39.1	60.7	50.9	54.0	40.8
UKI-SCOTGRID-ECDF	85.1	82.1	84.2	87.6	80.1	75.7	83.1
UKI-SCOTGRID-GLASGOW	94.7	90.3	94.5	81.6	92.4	92.2	91.3
UKI-SOUTHGRID-BHAM-HEP	79.5	91.7	91.8	91.7	76.0	84.3	86.3
UKI-SOUTHGRID-BRIS-HEP	28.0	58.7	84.0	80.0	82.0	81.8	54.4
UKI-SOUTHGRID-CAM-HEP	92.4	91.1	91.8	92.3	106.1	85.9	93.8
UKI-SOUTHGRID-OX-HEP	93.0	92.4	95.4	96.0	105.4	102.2	96.6
UKI-SOUTHGRID-RALPP	68.4	64.7	64.7	105.5	124.2	91.1	91.7
UKI-SOUTHGRID-SUSX	88.6	96.2	54.5	93.0			84.7
Total	93.5	110.0	111.4	102.7	130.1	100.3	106.8

Click here for a CSV dump of this table

Click here for a Extended CSV dump of this table

Click here for XML encoded data

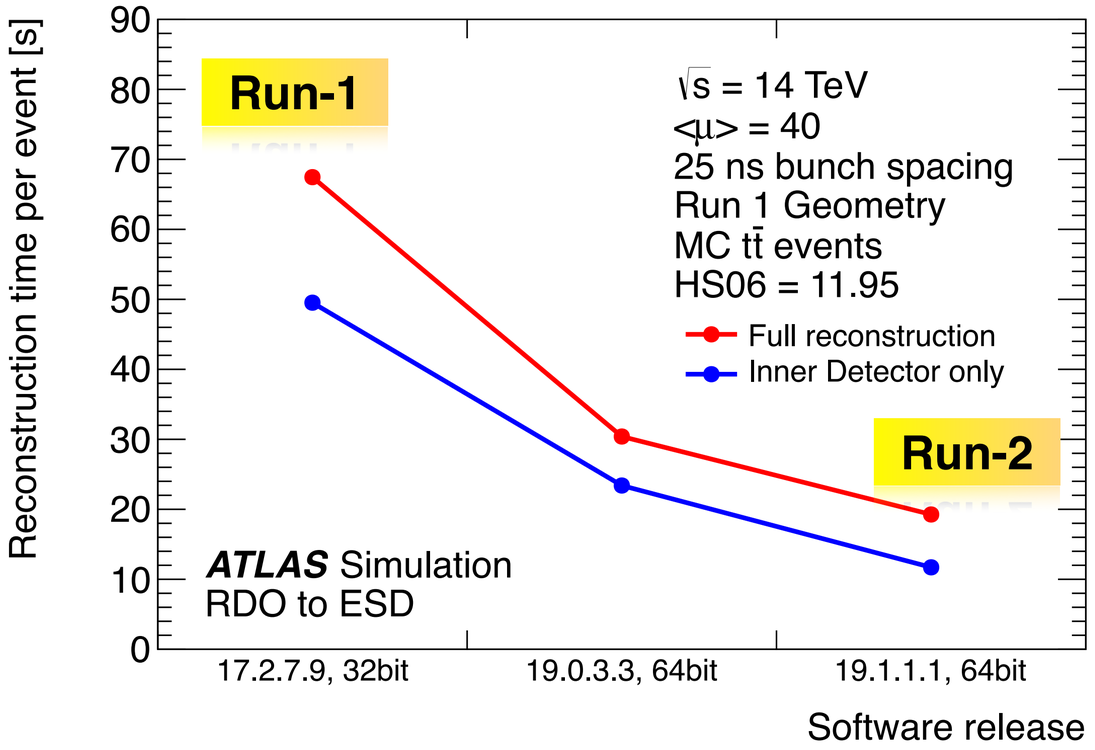
Key: 0% <= eff < 50%; 50% <= eff < 60%; 60% <= eff < 75%; 75% <= eff < 90%; 90% <= eff < 100%; eff >= 100% (parallel jobs)

- Sites should make sure they are publishing correctly.
  - ARC-CEs should work out of the box
  - For CREAM-CEs check here
  - OSG working on US sites publishing

Reconstruction speedup

ATLAS software ~6M lines of code

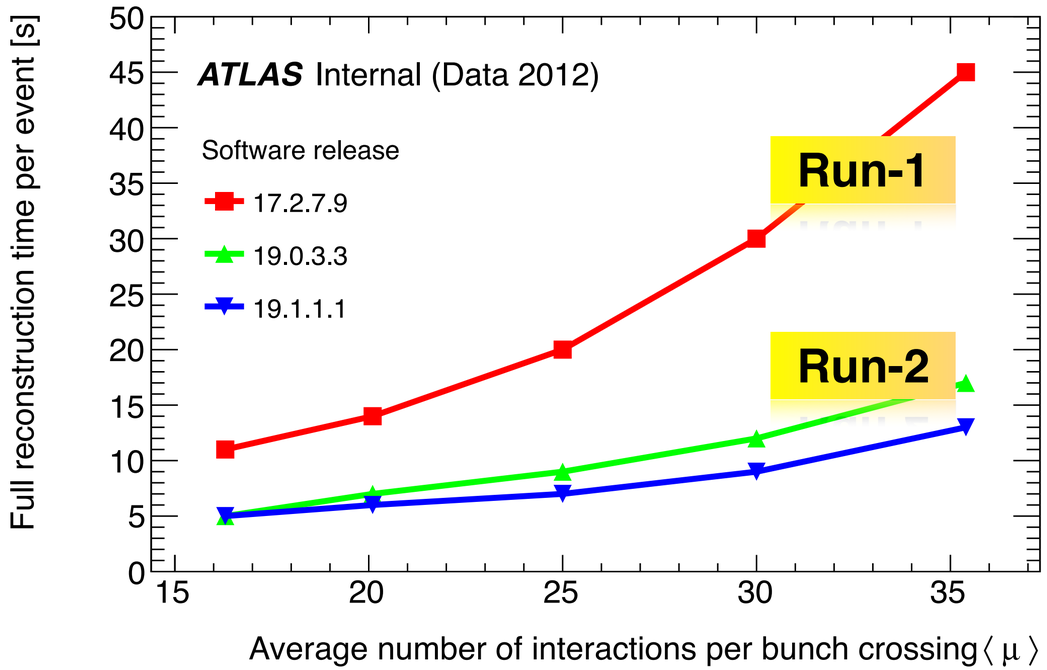
- **Factor 2 achieved**
- Large-scale software cleanup and optimisation
- Replacement of algebra library CLHEP by Eigen, …
- New Event Data Model (>1000 packages modified)
- Optimisation still ongoing…



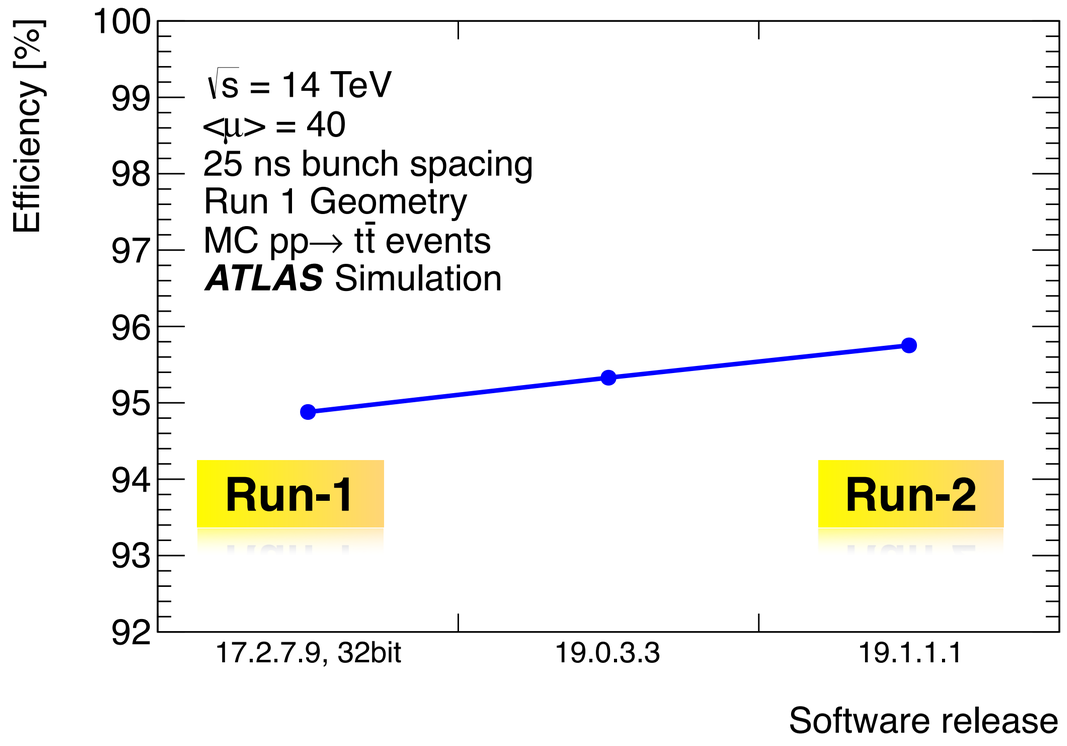
ATLAS software is SL6 64 bits

Software robustness

Reconstruction time vs pile-up

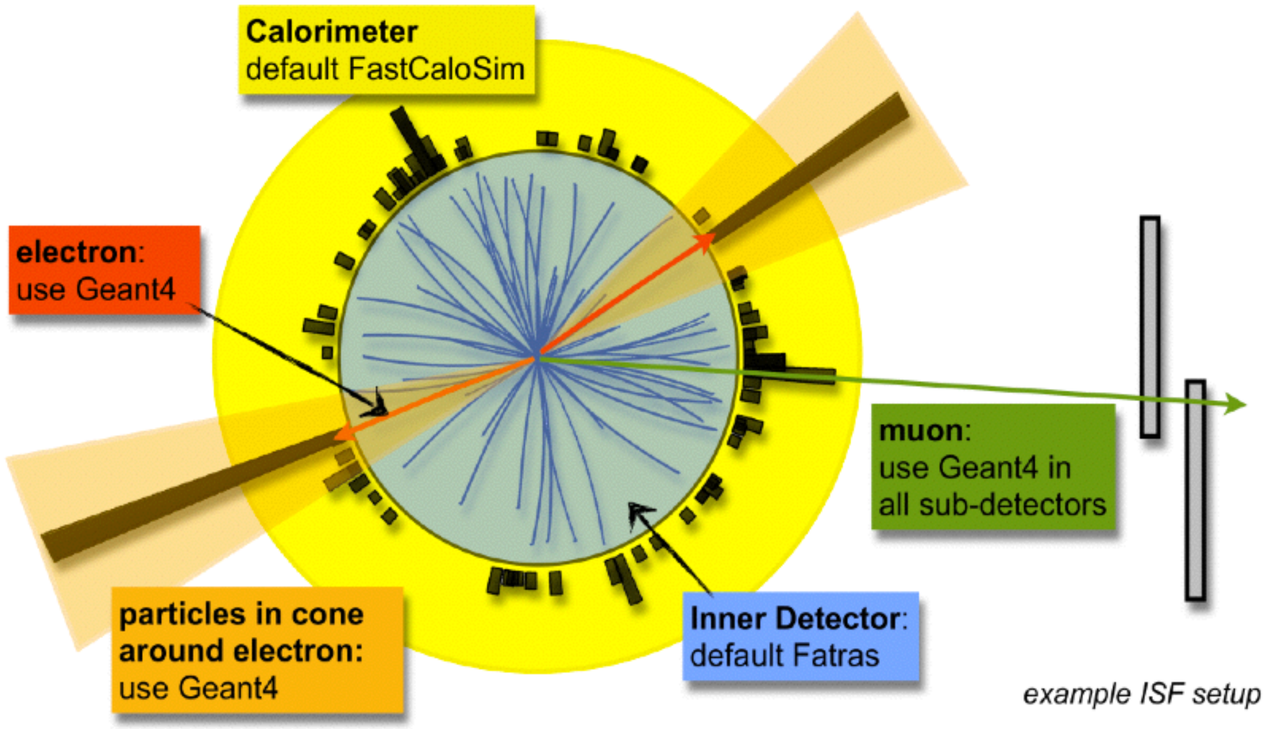


Track reconstruction efficiency



Fast Simulation

- Take advantage of fast simulation where appropriate
- Tradeoff accuracy for speed: Smearing, Frozen Showers, Parametric techniques
- ATLAS's Integrated Simulation Framework (**ISF**): clever mixing of fast and full simulation within the same event
  - Keep high precision for some particles and regions
  - Use fast simulation in areas that are not so importa
  - x100 speed ups possible, with much better results than normal fast simulation



example ISF setup

On going work... no ready before end 2015?

Computing improvements

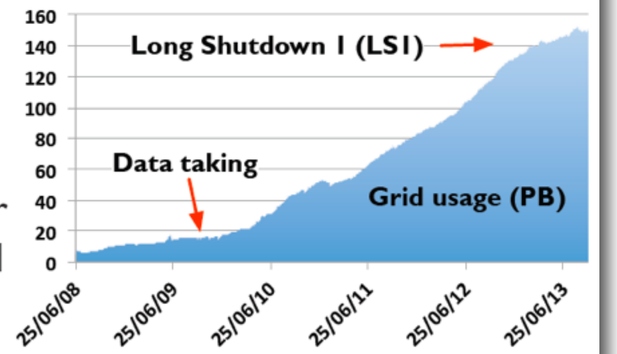
Data Distribution Management

http://rucio.cern.ch

- Rucio to replace DQ2
  - New scalable architecture
  - File level functionality instead of dataset
  - Built-in data replication policy for space and network optimisation
  - Multi-protocol (http,…)

The current DDM system Don Quijote 2 (DQ2) has demonstrated very large scale data management

- 150 PB
- 130 grid sites
- 800 users
- +40 PB per year
- +1 M files per year
- 0.6 M downloaded files per day



DQ2 will simply not continue to scale for LHC Run-2

Status :

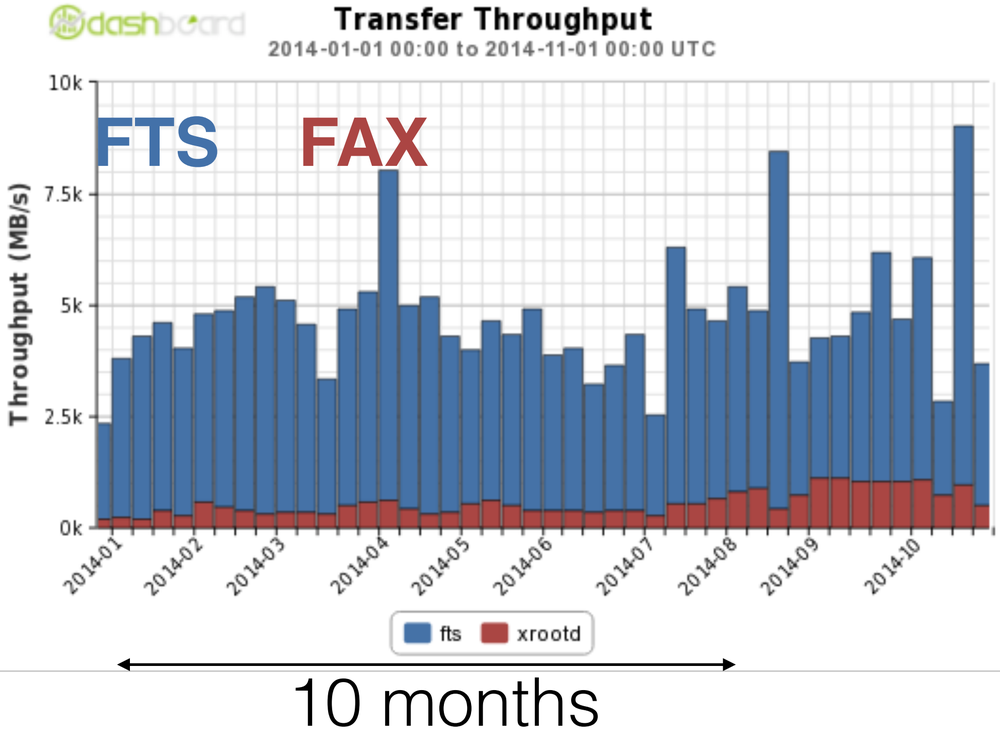
- **Rucio catalog used instead of LFC**
- **Sites to deploy WebDAV for data renaming**
- **Integration in ProdSys2 almost done**

Benefits of ProdSys-2, Rucio

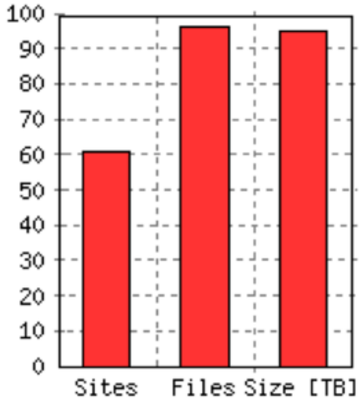
- Dynamic jobs, dynamic placement (distributed datasets)
- Each job will be sampled for resource usage:
  - Physical memory (RSS) profile
  - Swap profile
  - cpu usage profile
- Finished job reports will be registered in PanDA
- Scout job reports will be used to estimate the job resource limits for the rest of the jobs in a task
  - maxrss – maximum physical memory usage
  - maxcputime – maximum total cputime usage (multicore agnostic)
  - … + any other metrics which might be useful (e.g. network, total I/O...)
- Tasks will have job resource requirements much better defined

DISTRIBUTED STORAGE / REMOTE ACCESS

- Jobs access data on shared storages via WAN
- Better usage of storage resources (disk prices!)
- Bandwidth and network stability required
- **FAX** (Federating ATLAS data stores using Xrootd)
  - Deployment ongoing, job fail-over in case of access failure for now (also useful in case of storage downtime at site)
  - Future: generalised WAN access, throttled so that other activities/sites are not impacted
- http protocol also considered/coming



Remote data access: FAX



Goal reached ! >96% data covered

We deployed a Federate Storage Infrastructure (*): all data accessible from any location

Analysis (and production) will be able to access remote (offsite) files

Jobs can run at sites w/o data but with free CPUs. We call this “overflow”.

Throughput between major centers not what it should be

FTS Channel Performance Issues

Tier 0 to Tier 1

1. pps.lcg.triumf.ca and bunsen.ndgf.org have poor rates everywhere **<10 MB/s**
2. T0 to NIKHEF, RU is poor **<10 MB/s**
3. srm-atlas.cern.ch is poor **<10 MB/s**

Tier 1 to Tier 1

CA
1. TRIUMF is not really getting 10Gb/s to any off-site T1 **<~300 MB/s**
2. ppshead.lcg.triumf.ca **<1 MB/s**

ES
1. PIC to SARA **<3 MB/s**

FR
1. Looks OK! **Near 10Gb/s or better to all 10 Gb/s T1s**

DE
1. FZK to TRIUMF **6 MB/s**
2. FZK to NIKHEF **1 MB/s**
3. FZK to PIC **1 MB/s**

IT
1. INFN to NIKHEF **3 MB/s**

UK
1. PIC **85 MB/s**
2. FZK **112 MB/s**

US
1. Looks OK! **Near 10Gb/s or better to all 10Gb/s T1s**

ND
1. bunsen.ndgf.org **<3 MB/s**
2. srm.ndfg.org to itself!? **2 MB/s**

TW
1. No 10Gb/s to any site

RU
1. RU to TRIUMF **28 MB/s**
2. RU to NIKHEF **18 MB/s**
3. RU to TW **3 MB/s**

FTS Channel Performance Issues

Tier 2D to Tier 1

CA
1. Westgrid to IN2P3, SARA **<10 MB/s**
2. Toronto to SARA, NIKHEF **<10 MB/s**
3. McGill to SARA **<2 MB/s**
4. Wormhole to NDGF, RAL **<3 MB/s**

ES
1. uam.es to INFN **<3 MB/s**
2. PIC to TRIUMF **<5 MB/s**
3. IFIC to INFN **<3 MB/s**

FR
1. LAL to FZK, INFN **<36 MB/s**
2. LPNSE to SARA, NIKHEF, NDGF **<1 MB/s**
3. LPSC to PIC **<1 MB/s**
4. Marsellie to TRIUMF, BNL **<10 MB/s**

DE
1. DESY to SARA **<1 MB/s**
2. Wuppertal to TRIUMF, NIKHEF **<1 MB/s**
3. lcg-se0.ifh.de to INFN **<1 MB/s**
4. Goegrid to IN2P3, INFIN **<10 MB/s**

IT
1. ROMA to NIKHEF **<10 MB/s**
2. INFN to SARA, FZK, RAL **<20 MB/s**

UK
1. Manchester to TRIUMF **18 MB/s**
2. HEPLNX to INFN **4 MB/s**
3. QMUL o SARA,NIKHEF **<1MB/s**
4. Cambridge to BNL **<1MB/s**
5. Glite to NDGF **<10MB/s**
6. SCOTGRID to TRIUMF, NIKHEF **<10MB/s**

US
1. NET2 to NIKHEF **<1MB/s**
2. SWT2 to RAL **<30MB/s**
3. NET2, SWT2, AGLT2, WT2 have worse rates to SARA than to other T1s **(74,29,10,17,54 MB/s)**

Panda WAN modes

- Failover

  In the case stage-in fails due to a temporary SE related problem, the pilot will re-attempt the stage-in a second time after a few minutes. If that fails as well, the pilot has the option to attempt stage-in from a remote SE using FAX.

- Overflow

  When deciding where to broker a task, JEDI can estimate that it is better to send it to a site that does not have the input data and let it read from FAX, rather than let it sit in the queue of the site that has the data. Limited to analysis queues.

- Explicit overflow

  If a user explicitly requires CE that does not have the input data, the task will be brokered to that CE and FAX used to get the data.

Site controls

Failover (*)

Setup per Panda queue using two AGIS fields:

- **allowfax**=True will enable FAX retries.
- **faxredirector** sets the FAX access point to be used. For optimal performance it should be set to the site's closest redirector.

(*) Enabled by default for all Panda queues March 2014

Overflow

Other queue settings:

- **wansinklimit**** - limits the bandwidth that jobs overflown to the site can use.
- **wansourcelimit**** - limits the bandwidth that site's FAX endpoint can deliver to jobs overflown elsewhere.

** zero value turns off overflow in that direction.

Overflow - technical details

JEDI brokering

- Chooses a possible alternative queue based on weights defined by the “busyness” of the queue, possible input dataset replicas, and the network cost (as determined by the “cost matrix” discussed yesterday)

FAX endpoint used

JEDI sets a variable “source site” for each job, but for a technical reason the actual FAX access endpoint used comes from the faxredirector value set for the destination queue (uses FAX redirection)

Limit enforcement

JEDI first sends 10 “scout” jobs. When all of them finish it calculates average per job bandwidth used.

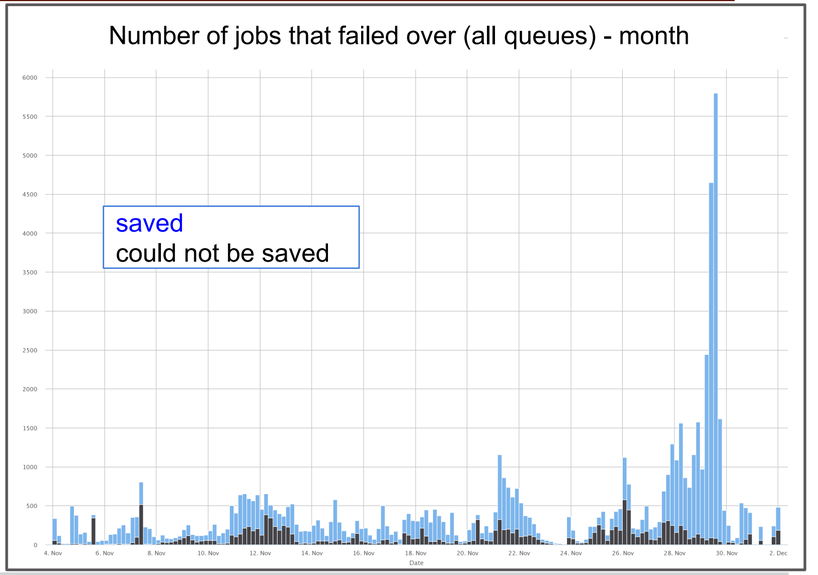
Only sends more jobs if the limit is expected not to be breached.

This is an approximate BW usage estimate; thus site WAN limits are considered to be conservative since:

- Lag between cost matrix measurement and job start
- Can’t predict how fast jobs will start running

Failover performance

- Running since March 2014.
- In average ~100 failover jobs/hour of which half finish correctly (reasons: simple job failure, or lack of a replica at another site)
- There is a standing task to get the failover monitor in Dashboard.



Enabling Overflow

Our goal: have 5-10% of all the jobs use WAN access, before Run 2. Have at least 50% of the CPU efficiency of regular jobs. Have similar error rates.

Overflows started mid August, first in US only. All the sites were both delivering data and accepting overflow jobs.

Now includes some CA, EU sites.

Most EU sites are not set to serve as the data source. Proceeding one-by-one to check for load issues and job failures.

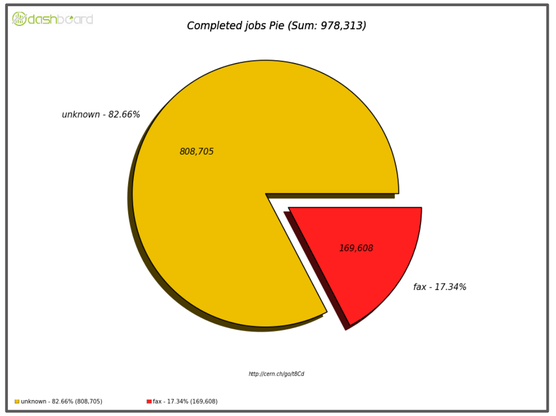
Started with UK, next week FR cloud.

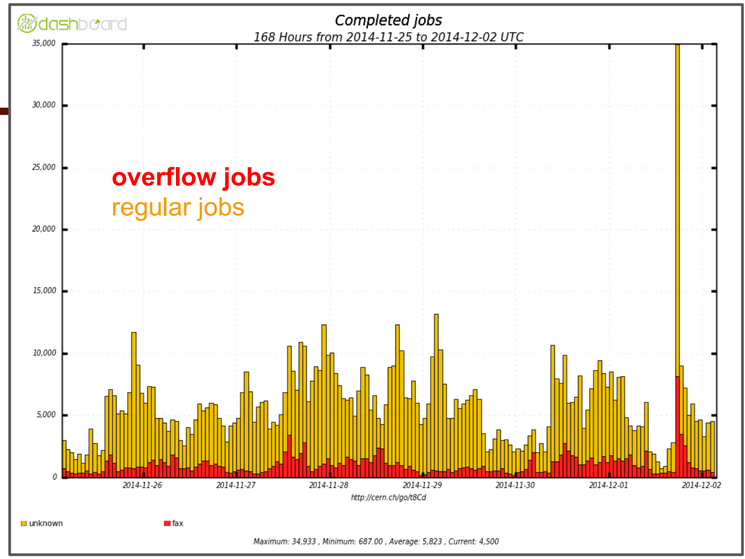
Queues currently accepting overflow jobs:

ANALY_AGLT2_SL6	ANALY_INFN-T1
ANALY_CONNECT	ANALY_IN2P3-CC
ANALY_BU_ATLAS	ANALY_MPPMU
ANALY_MWT2_SL6	ANALY_DESY-HH
ANALY_OU_OCHEP	ANALY_BNL_SHORT
ANALY_SLAC	ANALY_BNL_LONG
ANALY_SFU	ANALY_QMUL_SL6

Rate of Overflow

	regular	overflow
jobs per hour	4810	1010 **17%**



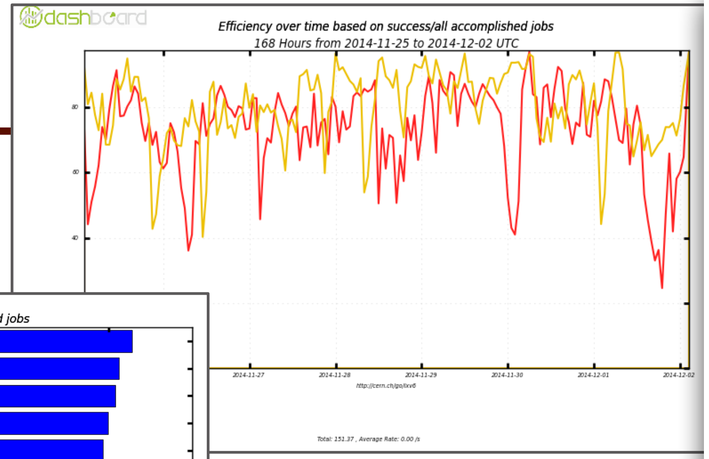


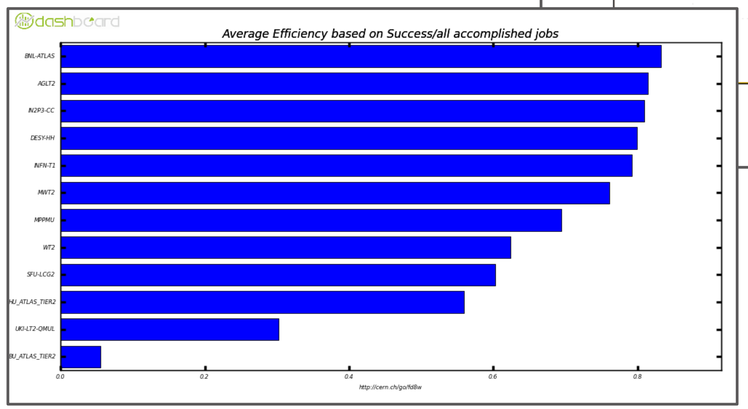
These and following plots are:

- for queues with overflow enabled
- from last the week. As soon as all the queues are overflowing I will make longer time plots.

Job efficiency

	regular	overflow
job efficiency	80%	73%

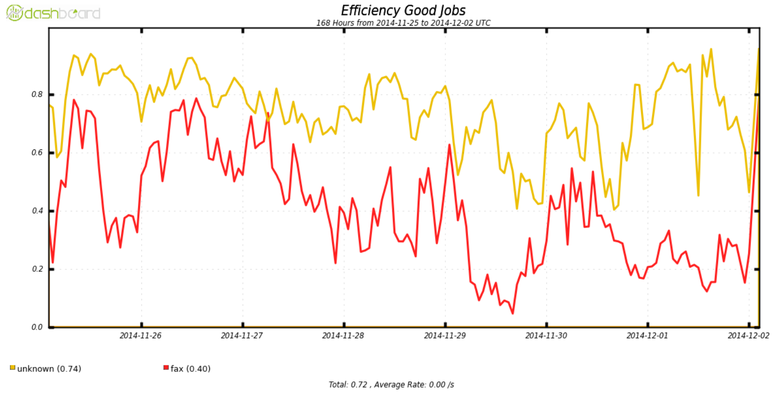




some sites handle overflow jobs better than others.

correlated with cost matrix and large-file timeouts

CPU efficiency



Strong dependence on job mix (user code).

Jobs with TTreeCache have roughly the same efficiency as local jobs.

The move to new versions of ROOT and xAODs format should improve CPU efficiency considerably.

Tests - local access and LRZ-LMU

Randomly repeated 3 times, class access mode

- local: 40/88/77 events/s
- dcap: 51/74/64 events/s
- FAX: 71/69/59 events/s
- Davix: 27/26/26 events/s (no TEvent.cxx patch)
- Davix: 46/56/56 events/s (with TEvent.cxx patch)

Randomly repeated 3 times, branch access mode

- local: 66/100/101 events/s
- dcap: 108/93/106 events/s
- FAX: 90/78/92 events/s
- Davix: 48/47/50 events/s (no TEvent.cxx patch)
- Davix: 65/81/91 events/s (with TEvent.cxx patch):

Important:

- Event rate only for particular analysis - seen factor 5-10 faster event rates (without systematics and heavy output mode)
- Davix suffers from missing buffering in start-up - event rate in later event loop otherwise fine

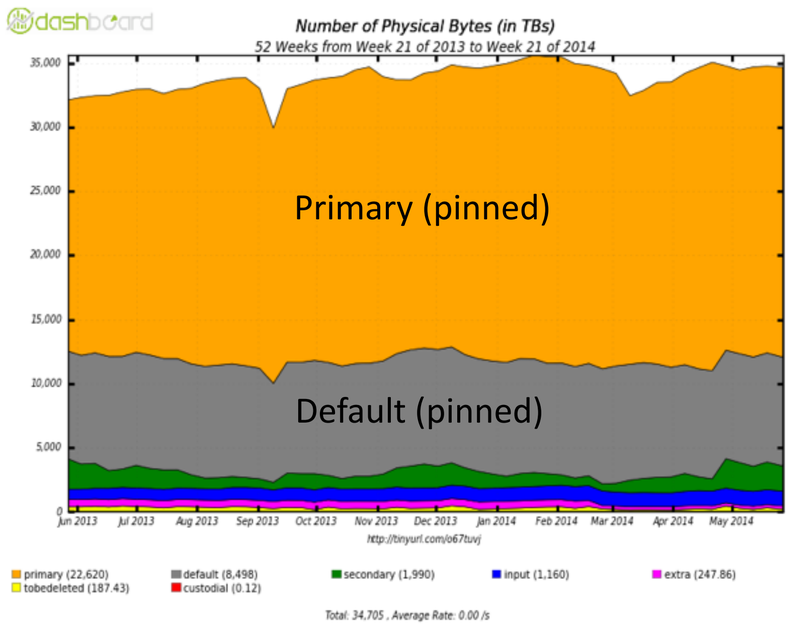
Summary of protocols

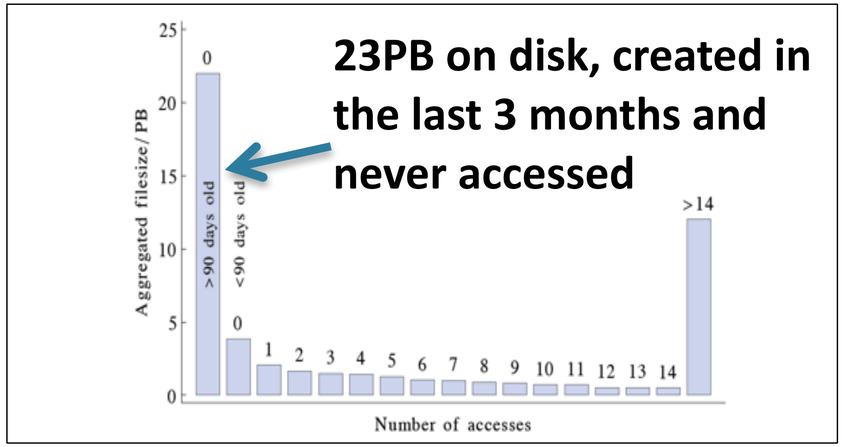
- We need SRM for tapes.
  - and this could be all if the items mentioned before (free/used space, deletion, space token accounting) are addressed properly
- We need gridFTP for 3rd party transfers.
  - Xrootd/webDav are under testing/commissioning xrootd but will take time
- We need xrootd for direct I/O. Possibly download and upload.
- WebDav can be used for all the rest (download/upload/deletions)
  - but till matures to production we need other protocols

New data lifecycle management model

Space management crisis

Disk occupancy at T1s





8 PB of data on disk never been touched

- T1 dynamically managed space (**green**) is unacceptably small
  - ➢ It compromises our strategy of dynamic replication and cleaning of popular/unpopular data
- A lot of the primary space is occupied by old and unused data

The new data lifecycle model

- Every dataset will have a lifetime set at creation
  - The lifetime can be infinite (e.g. RAW data)
- The lifetime can be extended
  - E.g. if the dataset is recently accessed. Or if there is a known exception
- Every dataset will have a retention policy
  - E.g. RAW need at least 2 copies on tape. Need at least one copy of AODs on tape.
- Lifetime being agreed with ATLAS Computing Resources management

Effect of the data lifecycle model

- Datasets with expired lifetime can disappear at any time from (data)disk and datatape
  - groupdisk and localgroupdisk exempt
- “Organized” expiration lists will be distributed to groups
- ATLAS Distributed Computing will flexibly manage data replication and reduction
  - Within the boundaries of lifetime and retention
- For example
  - Increase/reduce the number of copies based on data popularity
  - Re-distribute data at T2s rather than T1s and viceversa
  - Move data to tape and free up disk space

Further Implications

- We will use more tapes (*)
  - Both in terms of volume and number of accesses
  - Access to tape remains “centralized” (through PanDA + Rucio)
- For the first time we will “delete” tapes
  - How to do this efficiently?
- In the steady flow, we will approximately delete as much as we will write
- Access through storage backdoors is today not accounted

Impact: staging from tape

What happens if we remove all “unused” data from disk and keep it on tape?

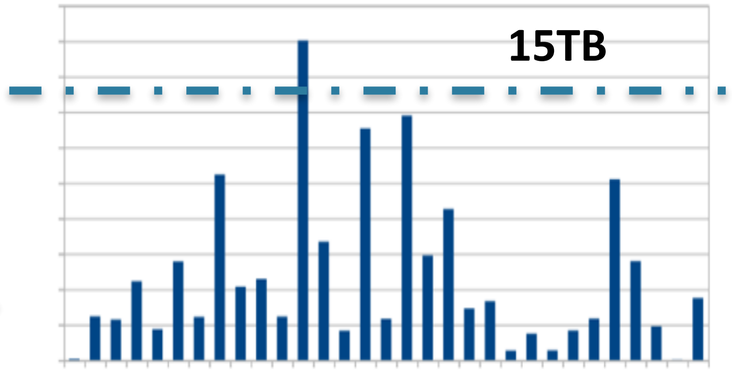
“unused” here = not accessed in 9 months

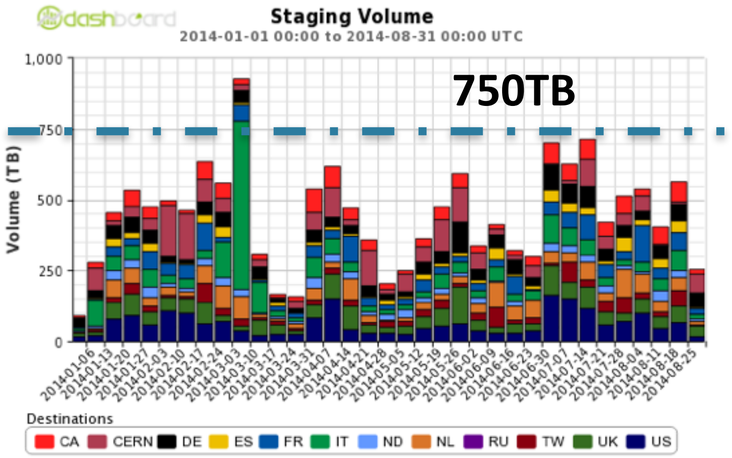
Simulation based on last year’s data access

Tape access from Reconstruction and Reprocessing in 2014

We would have to restage from tape 20TB/ week, compare with 1PB/week for reco/ repro (2% increase). In terms of number of files, it is a 10% increase

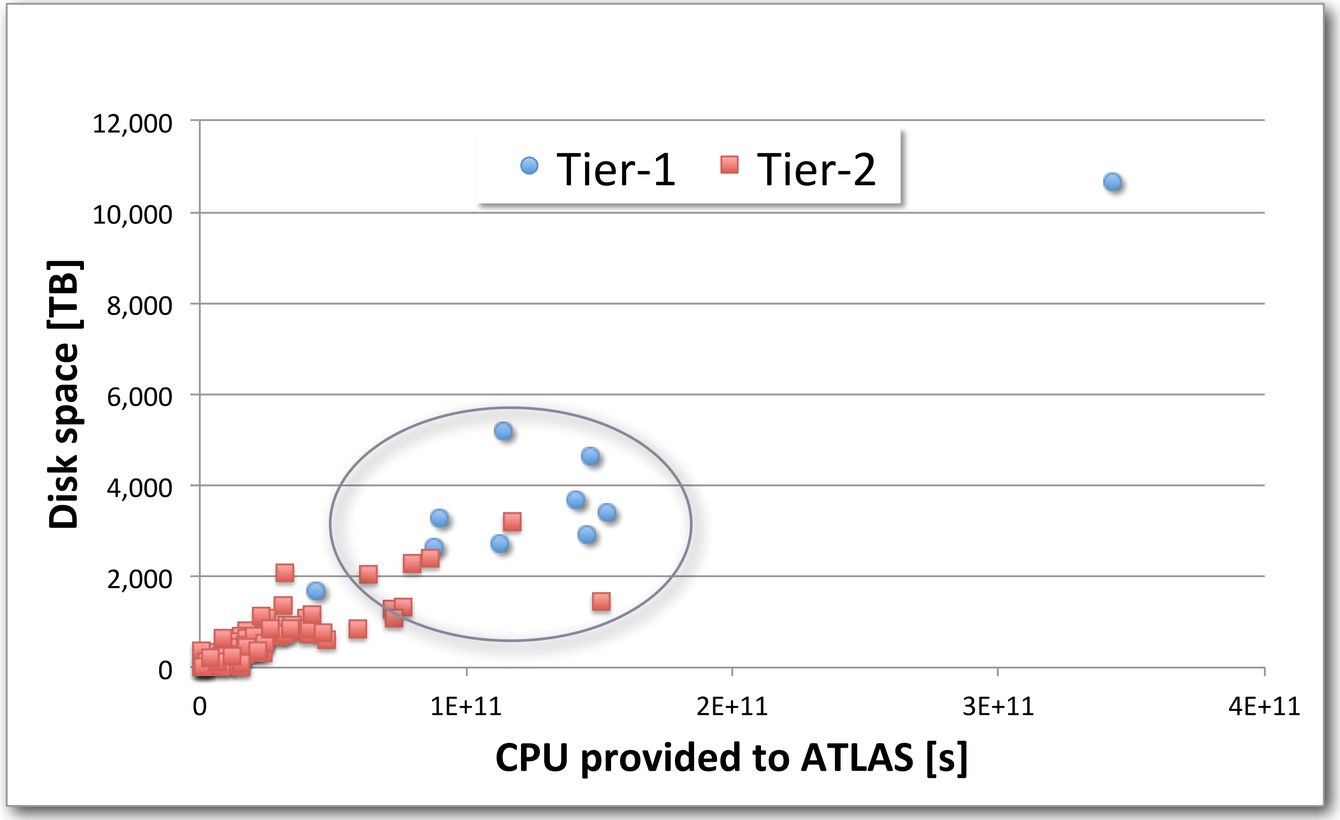
Data staged per week (TB)





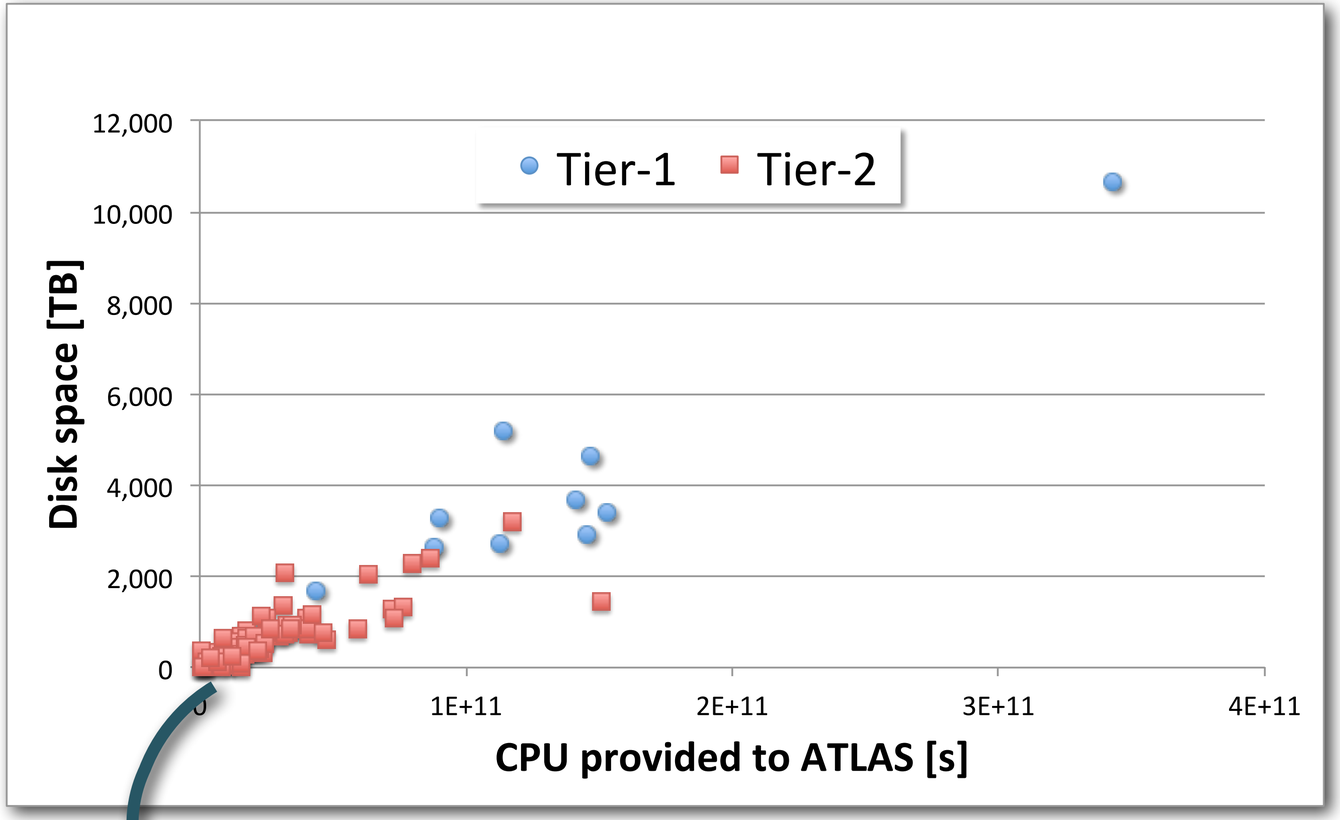
Data Processing

Data processing : Flexibility to be introduced



Some T2s are equivalent to T1s in term of disk storage & CPU power

Data processing : Operational load



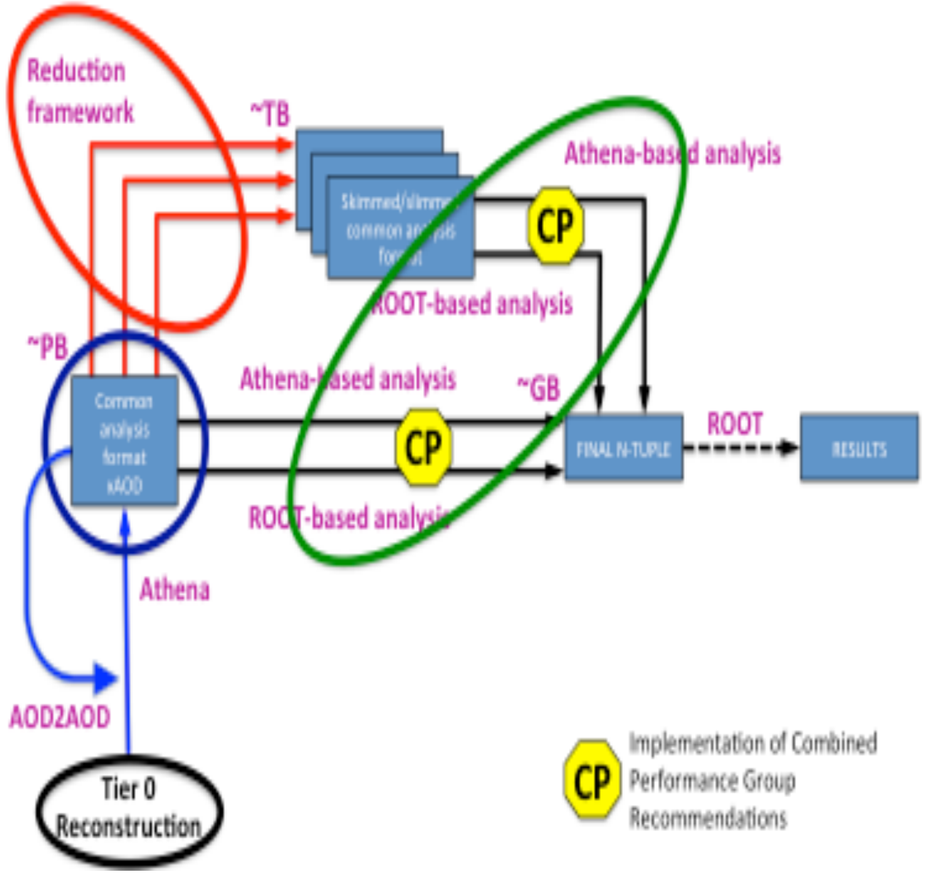
***Operational** load of many 'small' sites*
Less and larger sites would be better

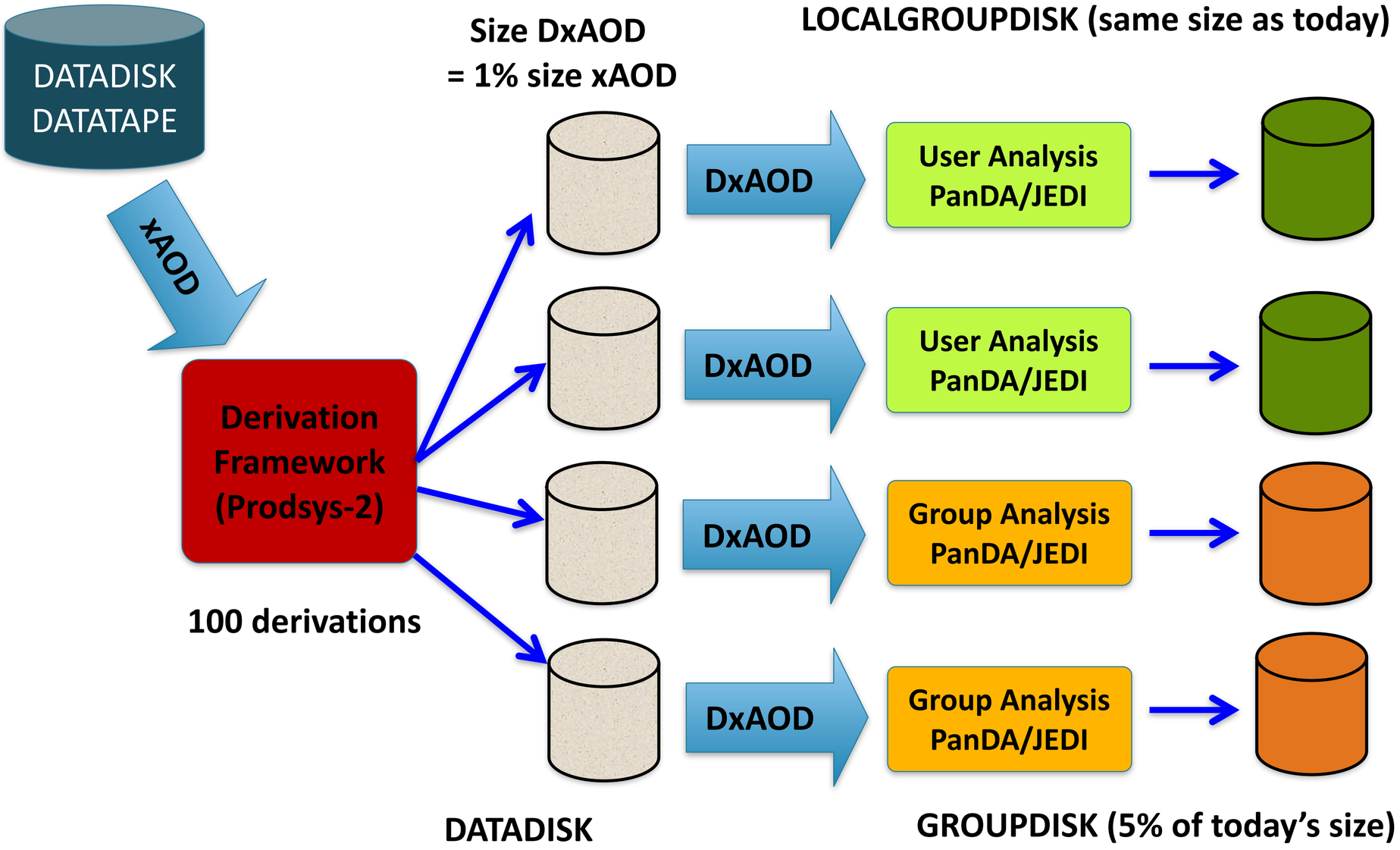
Data processing

- Optional extension of first pass processing from T0 to T1s in case of resource shortage at T0
- T1s and **some** T2s used for the most demanding workflows : high memory and I/O intensive tasks
- Data reprocessing & MC reconstruction also performed at **some** T2s
- Derivation Framework (train model) for analysis datasets

Analysis Model

- Common analysis data format: xAOD
  - replacement of AOD & group ntuple of any kind
  - Readable both by Athena & ROOT
- Data reduction framework
  - Athena to produce group derived data sample (DxAOD). Centrally via Prodsys
  - Based on train model
    - one input, N outputs / from PB to TB




DATADISK
DATATAPE
xAOD
Derivation Framework (Prodsys-2)
100 derivations
Size DxAOD
= 1% size xAOD
DxAOD
DxAOD
DxAOD
DxAOD
User Analysis PanDA/JEDI
User Analysis PanDA/JEDI
Group Analysis PanDA/JEDI
Group Analysis PanDA/JEDI
LOCALGROUPDISK (same size as today)
DATADISK
GROUPDISK (5% of today's size)

Recommendations for sites

WN Disk space requirements

- In Run-2, same as in Run-1
- But the jobs will change:
  - All the output (and thus input) files will be optimized to fit in a 2-5GB range (possible with JEDI dynamic jobs), to reduce (or get rid of) the number of merge jobs
  - Most of the cpu-intensive jobs will be multi-core
    - Effective reduction of local disk space usage on the node
    - One 8-core job is I/O friendlier than 8 single-core jobs
- Direct I/O:
  - Recommended for analysis (xrootd)
  - SE infrastructure must scale to a high number of connections and bandwidth

Memory Requirements

- Standard WN deployment: 2GB/core
- Limits are needed
  - 64-bit code not auto-protected by 3.3GB RSS, 4GB VMEM limit
  - Too many high-memory jobs can kill a node
  - Shared sites usually impose strict limits
- Recommendation:
  - Limit total RSS job usage, limit SWAP
  - RSS limit is only possible with cgroups enabled batch systems (SLURM, condor, UGE?)
    - Sites are encouraged to deploy it
  - VMEM only sites:
    - Jobs will request RSS, not VMEM → VMEM should be scaled by a factor of ~2
  - “ATLAS only” sites:
    - No limits on memory – the job-resource monitor will make sure the resources are not overused
- Switching to RSS for jobs/tasks will happen early next year

Local network requirements

- AtlasDerivation Framework:
  - Most of the common analysis processing will be driven centrally through the train production
  - Less need for CPU intensive user analysis
- Most of the user analysis is expected to:
  - Process big amount of data
  - Use less cputime
  - A single analysis job on derived datasets can use 40MB/s
- Local network:
  - 1Gb/s on fat nodes (32-core, 64-core) will be a bottleneck → consider 10Gb/s or 40Gb/s infrastructure
  - SE – WN bandwidth might need to be improved

Site utilisation

Run-1 model, briefly

- Strict hierarchical model (Monarc):
  - Clouds: T1 + T2s (+ T3s)
  - No direct transfers between foreign T2s
  - Relaxed towards the end of Run-1 (Multi-cloud production – T2s can process jobs of many clouds)
- Production organization:
  - Tasks assigned to T1s
  - T1 is the aggregation point for the output datasets of the tasks
  - T2 PRODDISK used for input/output transfers from/to T1
- T2 disk space:
  - distribute the final data to be used by analysis
  - store secondary replicas of precious datasets

Planning for Run-2 model - facts

- Network globally improved
  - Much higher bandwidth (an order of magnitude increase)
  - Most of the links between ATLAS sites provide sufficient throughput : full mesh for transfers can be used
- Many Tier-2 sites provide the Tier-1 level stability of computing, storage and WAN
  - Many in LHCONE or other high-throughput networks
  - Tape resource is the only difference between Tier-1s and large Tier-2s, as far as the usability for ATLAS is concerned
- CPU only (opportunistic) centers are fully integrated in ATLAS
  - Some run all kind of tasks, including data reprocessing
  - Have good connectivity to geographically close Storage Elements

CPU and Storage organization

- Breaking the barrier between the Storage Element and Computing Element:
  - Remote I/O, job overflow, remote fail-over of input or output file staging → storage not strictly bound to the site computing resource
  - Tier-1, Tier-2, Tier-3 storage classification does not make much sense anymore
- ATLAS Storage pool:
  - TAPE
  - STABLE disk storage – T1 + reliable T2 (former T2Ds)
  - UNSTABLE disk storage – less reliable T2s
  - VOLATILE disk storage – unreliable T2s, T3s, opportunistic storage

Job optimizations 2

- Production / Analysis
  - Run-1: 75% / 25% (slots occupancy ~ cputime usage)
  - Run-2: 90% / 10% (not even a rough estimate)
    - Bulk of analysis (Derivation) moving to (group) production
    - Remaining analysis will be shorter and I/O intensive
- Reduce the merging
  - Avoid it if possible (simulation, reconstruction)
  - Local merging – merge on the site, where the files to be merged are
- Jobs will produce bigger outputs
  - Good for tape storage
  - Bigger files transferred – good for efficient transfers (but less files to transfer)

Tier-2 site classification

- Based on ASAP metric
  - ATLAS Site availability for analysis
    - Analysis tests do all relevant checks of CE and SE availability
  - See Martina's talk later today
- 3 types of Tier-2s: AN EXAMPLE, to be refined, rediscussed
  - T2S : STABLE, ASAP > 90% in the last 3 months
  - T2U : UNSTABLE, 90% > ASAP > 80% in the last 3 months
  - T2V : VOLATILE, ASAP <80% in the last 3 months
- ICB policy:
  - T2V will be exposed to ICB which will inform the corresponding funding agency
  - IF T2V has ASAP < 80% for more than 6 months, it will be put in degraded mod
    - Storage will be removed from ATLAS
    - Can continue to contribute as Tier-3 (CPU)
- Metric might be too simple (network throughput), further experience needed

Consequences for production

- STABLE storage effectively doubles the space available for production:
  - ~50 out of ~80 Tier-2 SEs will be part of it (today's T2D, in 2015 T2S)
  - Not limited to Tier-1 disk space for brokering
- Much larger space to consolidate the production data
  - Less complex rules for data placement policies, less need for data migration
- Solving the always problematic full Tier-1 space and less used Tier-2 space which did occasionally block the production of some tasks in the past

Conclusions

- New production and data management system provides many possibilities for further improvements and dynamic optimizations
  - Unfortunately, the commissioning was delayed, to give us more time for big changes well in advance of the Run-2 startup
- Fortunately, many of the changes can be implemented before the Run-2 starts
  - Many hooks are present already, we just need to use and tune them
- And even during the Run-2 we can afford to bring drastic improvements to our distributed system
- BUT, the production STABILITY will be the FIRST PRIORITY during data taking
  - In the last 2 years, we got used to a bit relaxed modus of operandi
  - In the next few months, we need to gradually tighten the overall stability to be ready for Run-2

FLAT Hierarchy

- STABLE storage with “stable” computing resources and fast network connections – A set of reliable resources
- 2nd layer of the less reliable, sometimes unavailable, pool of computing resources
- ATLAS plans to use the STABLE layer in a completely FLAT way
  - optimizing all the workflows (cpus, transfers, storage) for fast turnaround while minimizing the resource usage (minimize the transfers, balance disk usage...)

FLAT hierarchy

- Rucio supports distributed datasets:
  - A dataset replica can be distributed over many sites
- Strict ATLAS cloud model does not make much sense any more
  - Tasks are brokered to all stable sites, the point of consolidation of the production chain output
- A task still needs to be processed by many sites – job brokering will rely on
  - Input data proximity
  - Transfer cost matrix
  - Dynamic evaluation of transfer time (number of assigned jobs, recent history of past activity)
- New Prodsys and DDM:
  - Intermediate datasets (middle of the chain) will stay unconsolidated – distributed among the sites, skipping the output transfers
  - This might have to be limited to T2S sites only
- Final datasets consolidation:
  - Primary replicas will be consolidated
  - Secondary replicas can stay distributed at the sites that produced the files

Global cloud

- Tasks do not need to be assigned to any site – global task
  - The final consolidation can be delayed
- Final (primary) datasets do not need to be consolidated at all
  - Will be evaluated
  - Might be too difficult to manage (migration to tape)
- Big global task can be managed in a better way
  - Less tasks to manage, better activity overview, clearer prioritization
  - Large production tasks have been artificially split in Run-2 to run everywhere
- Experience with the new system is needed to choose the best option

Specializing the sites for workloads

- The pre Run-1 constraints for job placement are gone
  - Frontier instead of direct DB access → data reprocessing runs anywhere
  - High priority jobs (HLT reprocessing, Tier-0 spillover) with a short deadline could run everywhere
- But not all the sites are equal
  - Tier-1 vs Tier-2 is definitely not the correct answer
- ALL the jobs are important,
  - But not all the job types run equally well on all the sites
  - Some sites are slow for analysis but they are good for data reprocessing
  - Some sites are very big but cannot run 100% of heavy I/O jobs
- Differentiation was already used during Run-1 by limiting the job types through the fairshare (AGIS settings)
  - e.g. evgensimul=60%,all=40%
- But not all the jobs are EQUALLY important:
  - Some tasks have short deadline
  - Some large activities have close deadline (physics conferences)

ATLAS T2 Site Categorization: “the” proposal

Andrej Filipcic & Alessandro Di Girolamo
as ADC Operations

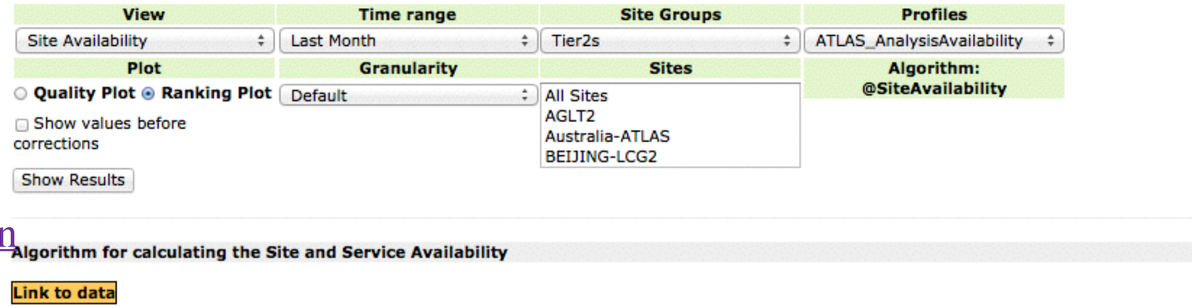
ATLAS Site Availability Performance (ASAP)

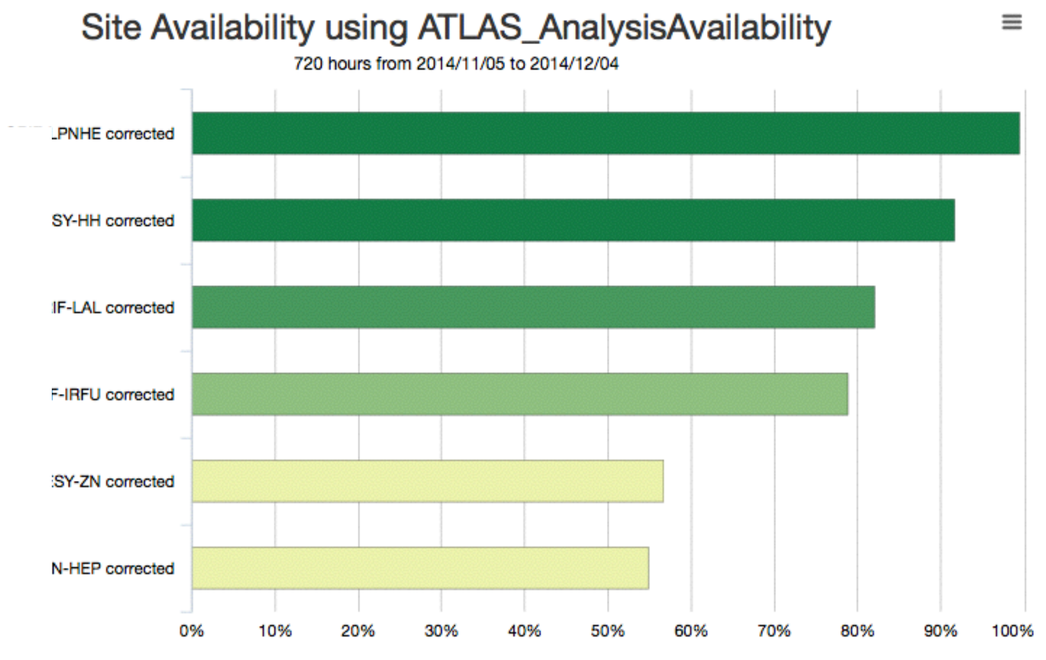
- We do have a metric to measure the analysis usability
  - Reliable metric to test the full site functionality
  - Will be described in details tomorrow morning

ASAP summary:

•We do have instructions for sites https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/ATLASSiteCategorization

- We do have possibility to recompute in case of special issue
- We are recording data since few monthshttp://cern.ch/go/N8xD





T2 categorization proposal
in a nutshell

- Presented at ICB
  - https://indico.cern.ch/event/301690/ and https://indico.cern.ch/event/330737/

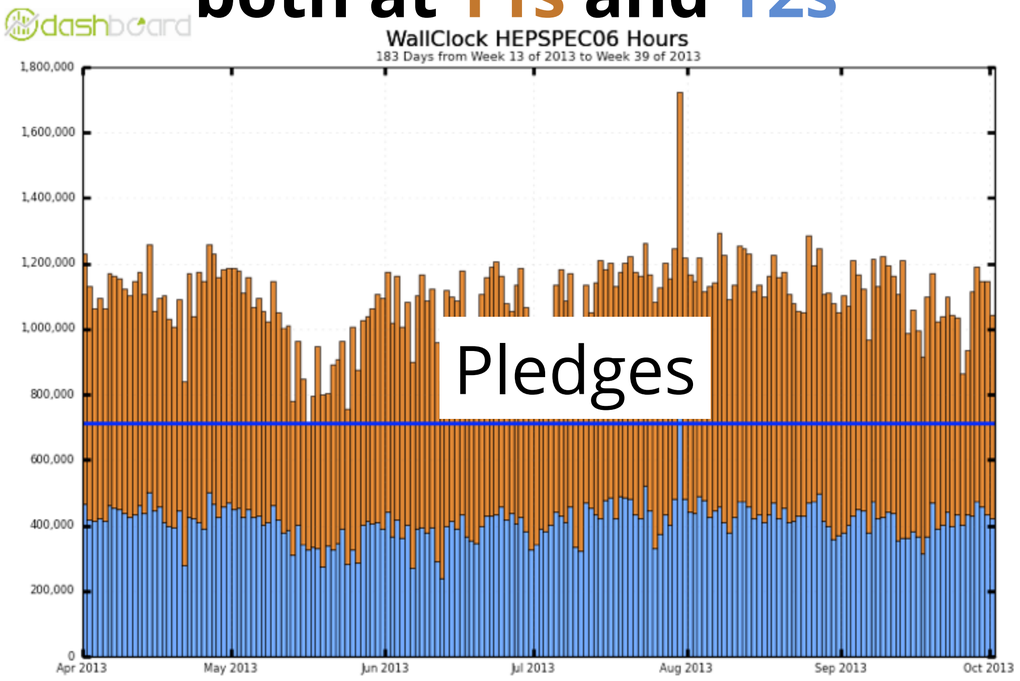
- Sites below 80% of *ASAP* over the past 3 months are not effective for ATLAS
  - ICB should contact the Funding agency to investigate the issues and understand the possible actions foreseen to improve the situation
  - If the situation persists for another trimester:
    - Remove the capacity (CPU and Storage) provided by that site from ATLAS available disk space: site will never be able to match the 95% MoU yearly average availability
    - Disable the storage from data distribution: ATLAS will use the site as a Tier3
  - If the situation persist for one full year ICB should consider to recommend WLCG to reconsider the MoU

Opportunistic resources

At Run-1 : quality of physics results and physics throughput benefited a lot from these additional resources!

- Need for additional solutions beyond pledges; 2 examples :
  - **HPC** (High Performance Computing) centres (effort on software needed to make full use of them)
  - **Volunteer computing**: ATLAS@home, also useful for university/department clusters

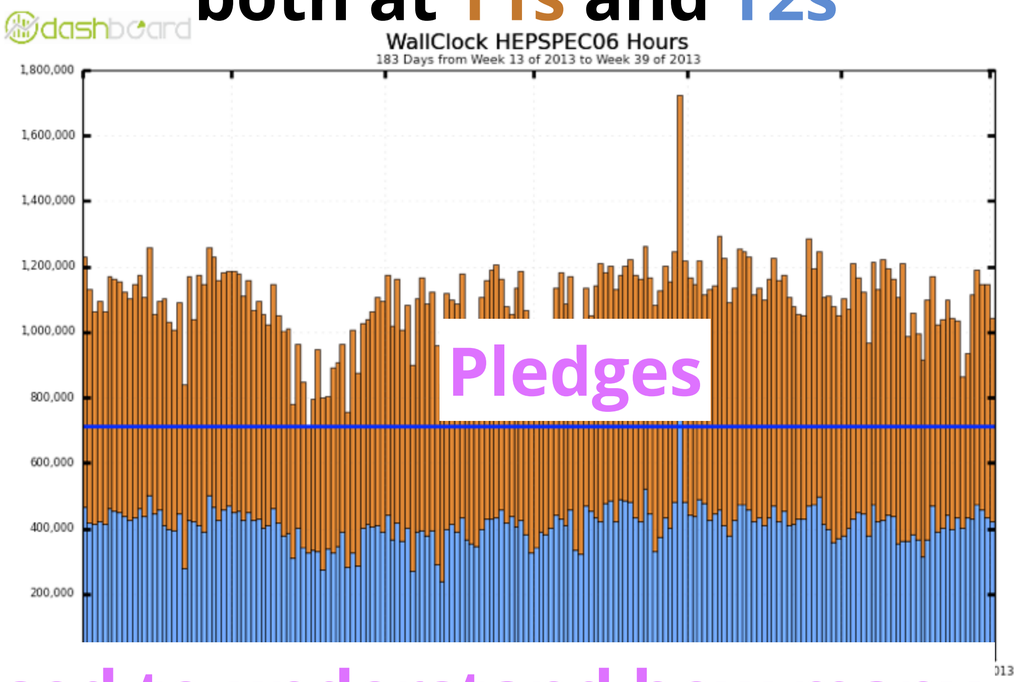
CPU consumption above pledges both at T1s and T2s



Opportunistic resources

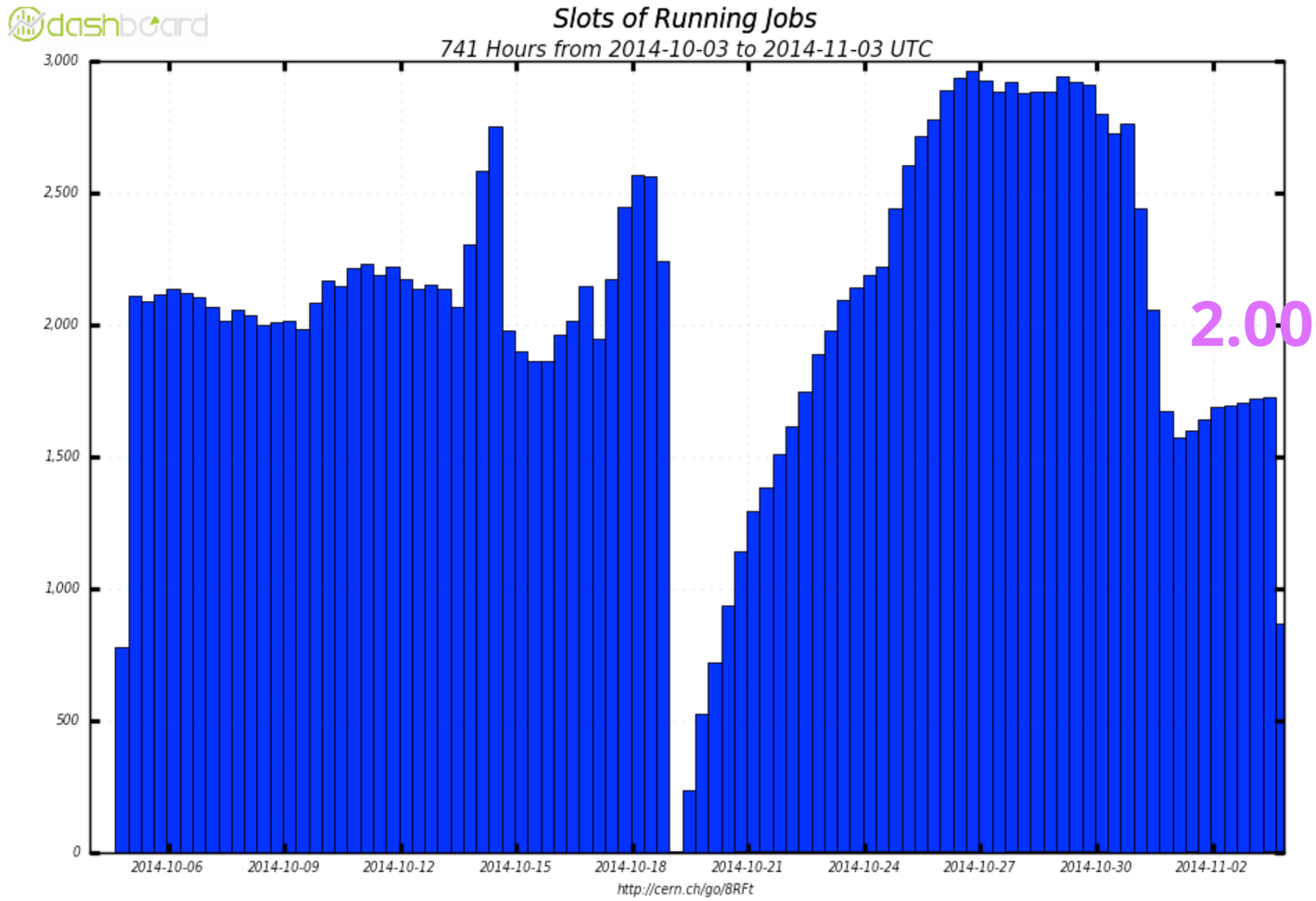
- Need for additional solutions beyond pledges; 2 examples :
  - HPC (High Performance Computing) centres (effort on software needed to make full use of them)
  - Volunteer computing: ATLAS@home, also useful for university/department clusters

CPU consumption above pledges both at T1s and T2s



ATLAS sites availability/usability will be used to understand how many pledged resources are effectively available for us

ATLAS@home



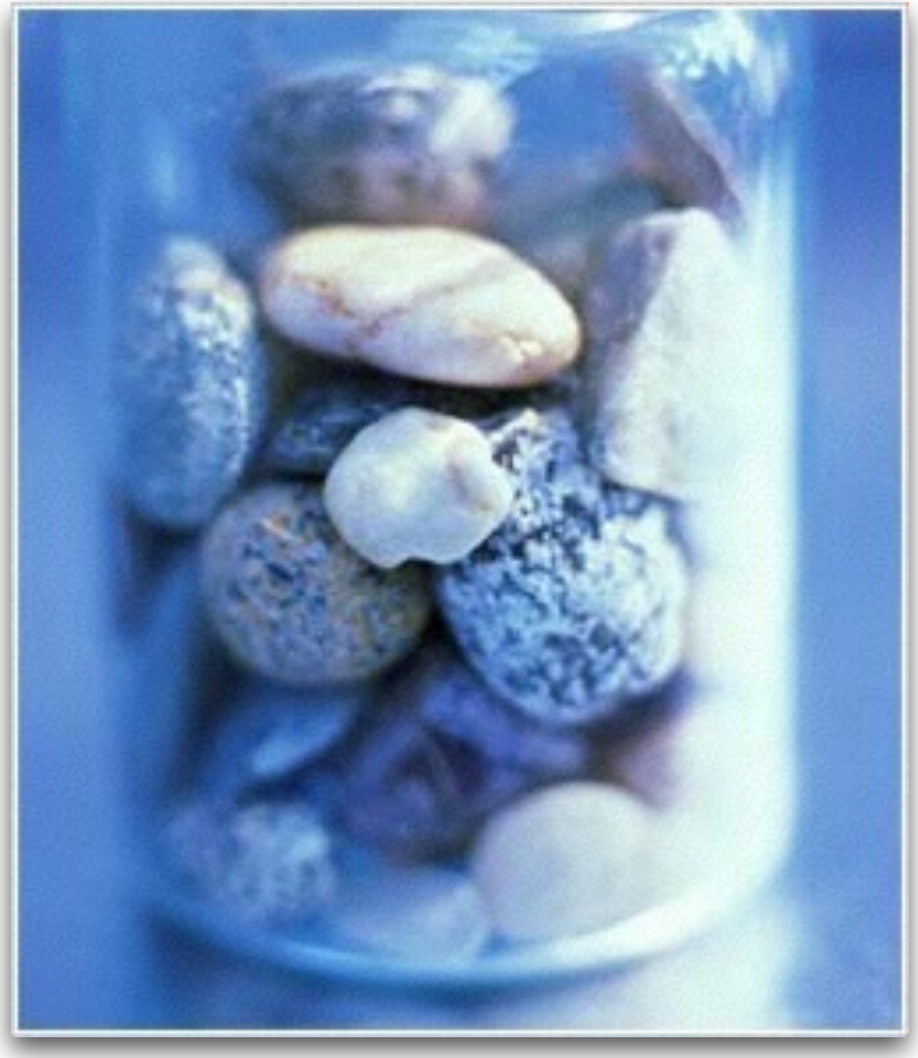
2.000 jobs

1 month

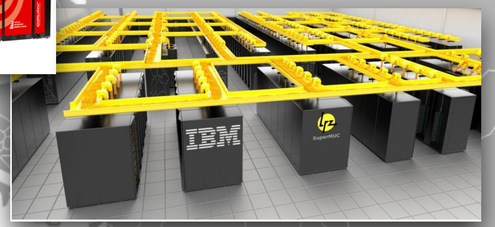
http://atlasathome.cern.ch

ATLAS at HPC centres

- The jar on the right is full of rocks
- Nevertheless it is not full
- Often when supercomputers are “full”, there are empty nodes
- ATLAS program would benefit a lot by using those empty nodes



On going ATLAS projects at HPC centres

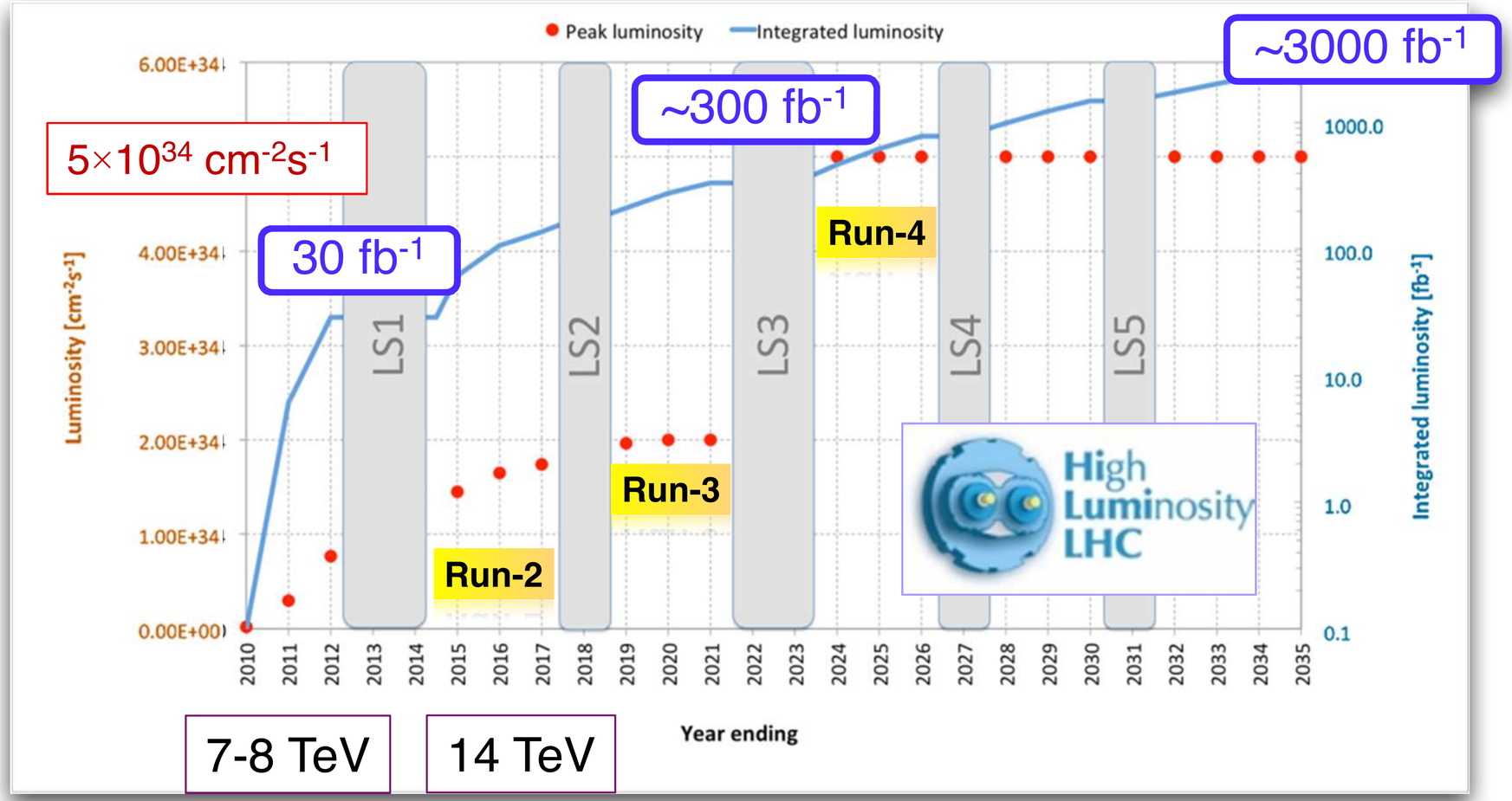


U.K.

Many more to come…

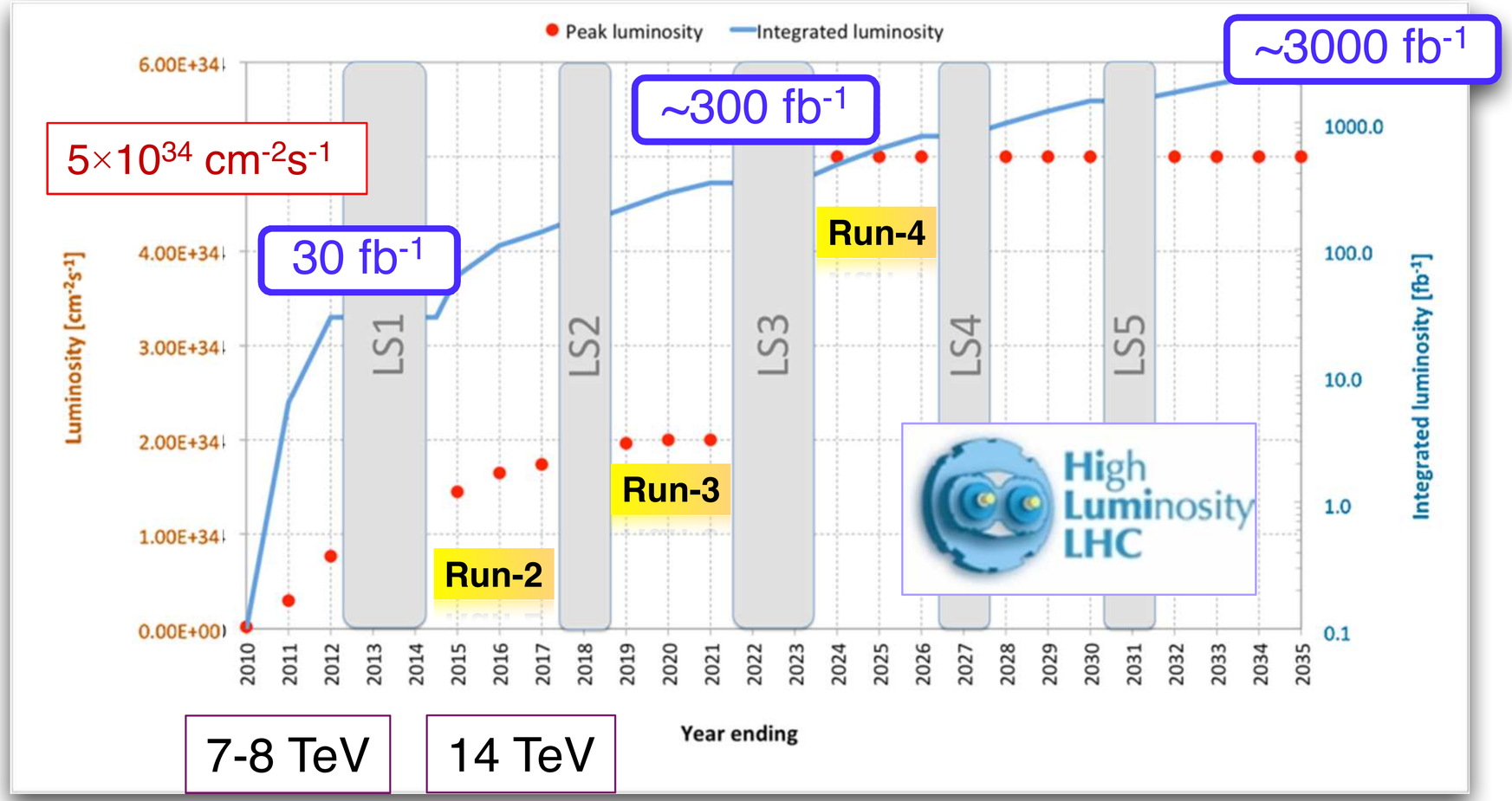
Beyond Run-2

LHC in the next years



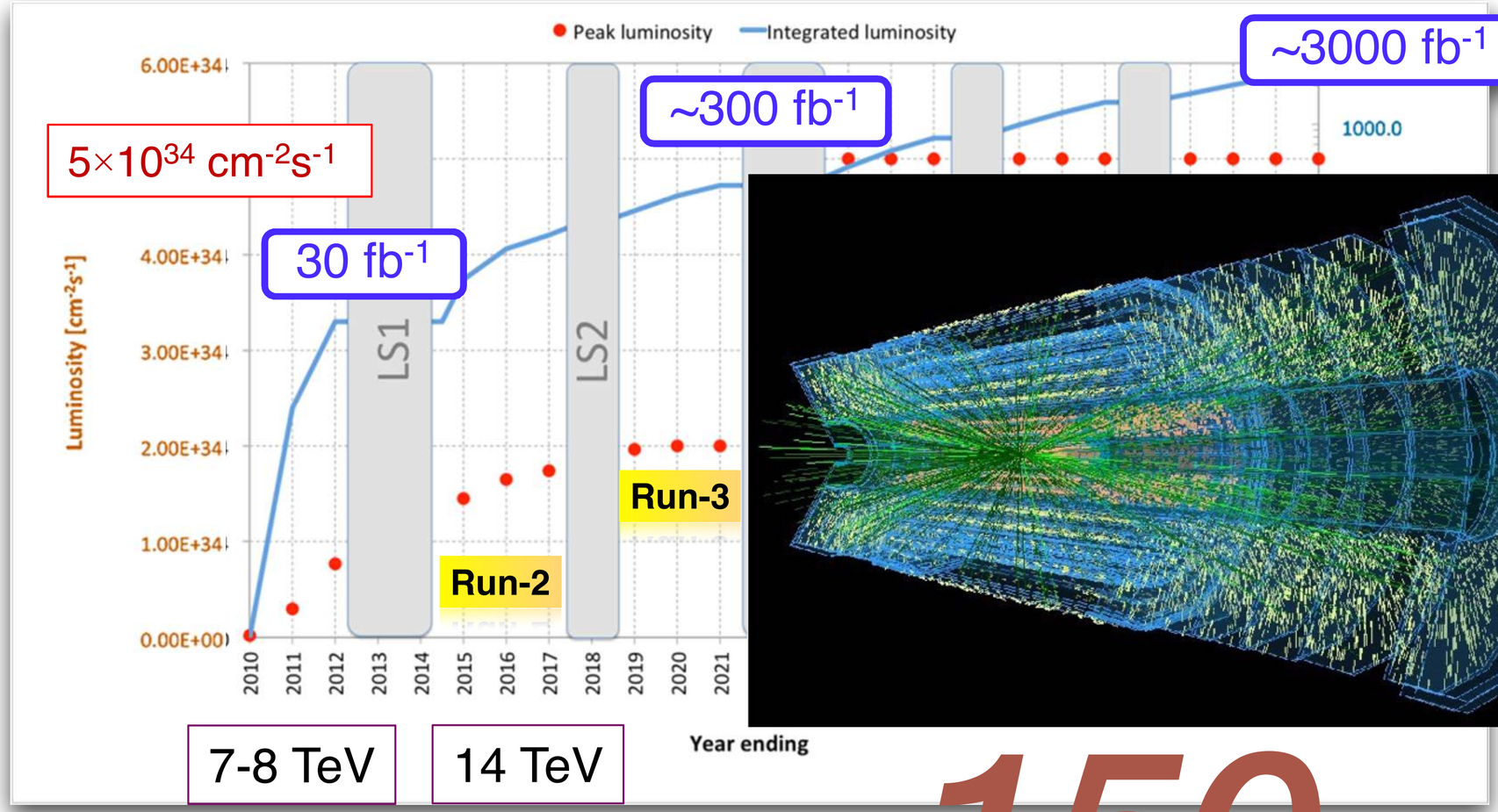
~25 ~40 *Pile-up*

LHC in the next years



~25 ~40 ~70 *Pile-up*

LHC in the next years

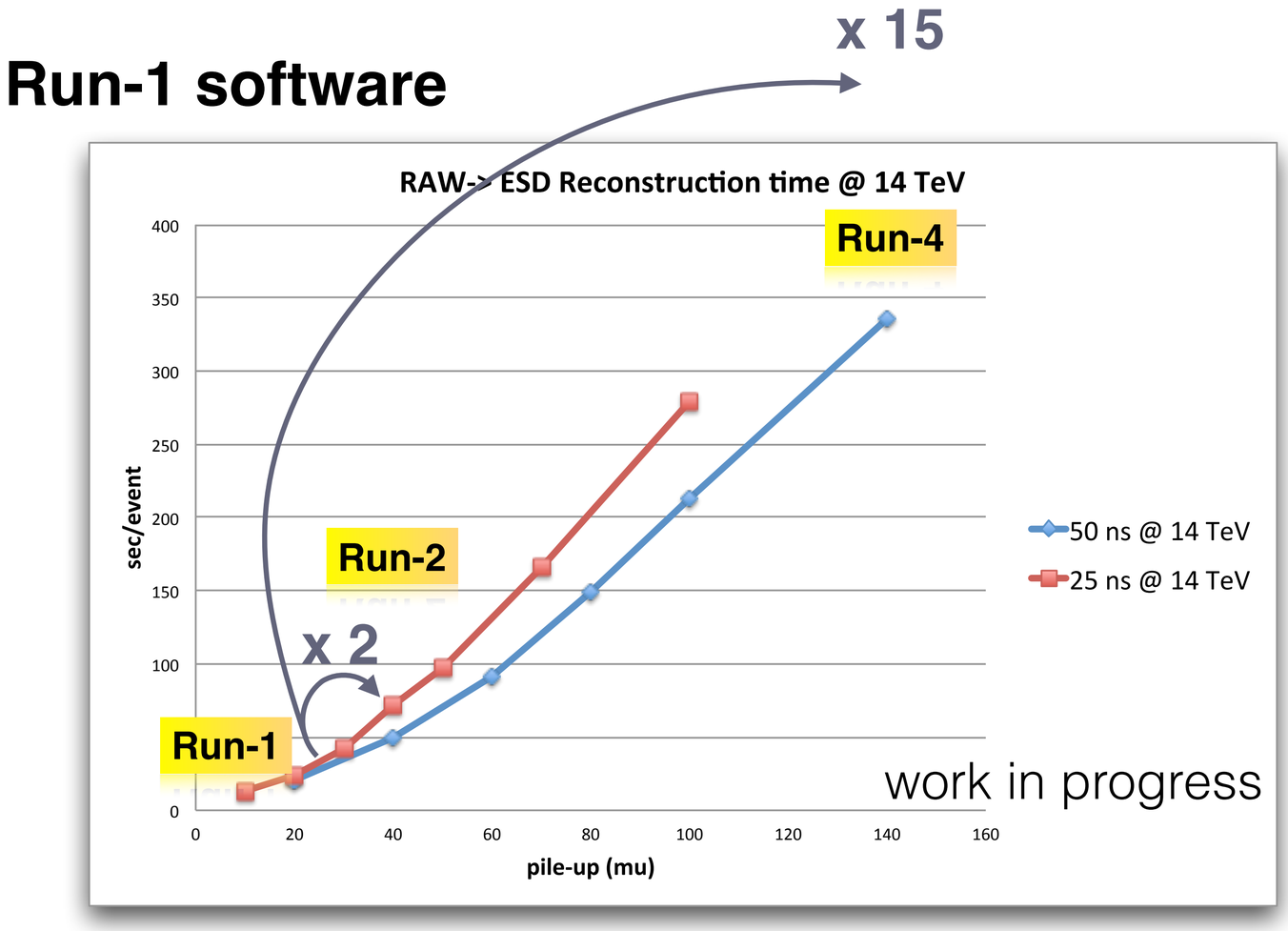


~25 ~40 ~70 ~150 Pile-up

Preparing for Run3 – a multi-year software effort

- Re-engineer for future (or even current) computer architectures: vectorisation, multi-threading,…
- Make Software Quality an integral part of software development and maintenance
- Define strategy for reconstruction and simulation of high pileup events
- Ensure that software packages on which we depend are compatibly re-engineered (Geant4, Root, Eigen, …)
- Collaboration and cooperation with external projects

Reconstruction time vs Pile-up



Summary & Outlook

- A lot of experience acquired in 3 years of LHC data taking
- Run-2 will put high pressure on hardware and human resources
- Solutions under development and manpower is critical
- New computing model and its components are being tested and commissioned

spares

Publication publish with Argonne computing help

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

CERN-PH-EP-2014-053

Submitted to: Phys. Rev. D

Search for high-mass dilepton resonances in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector

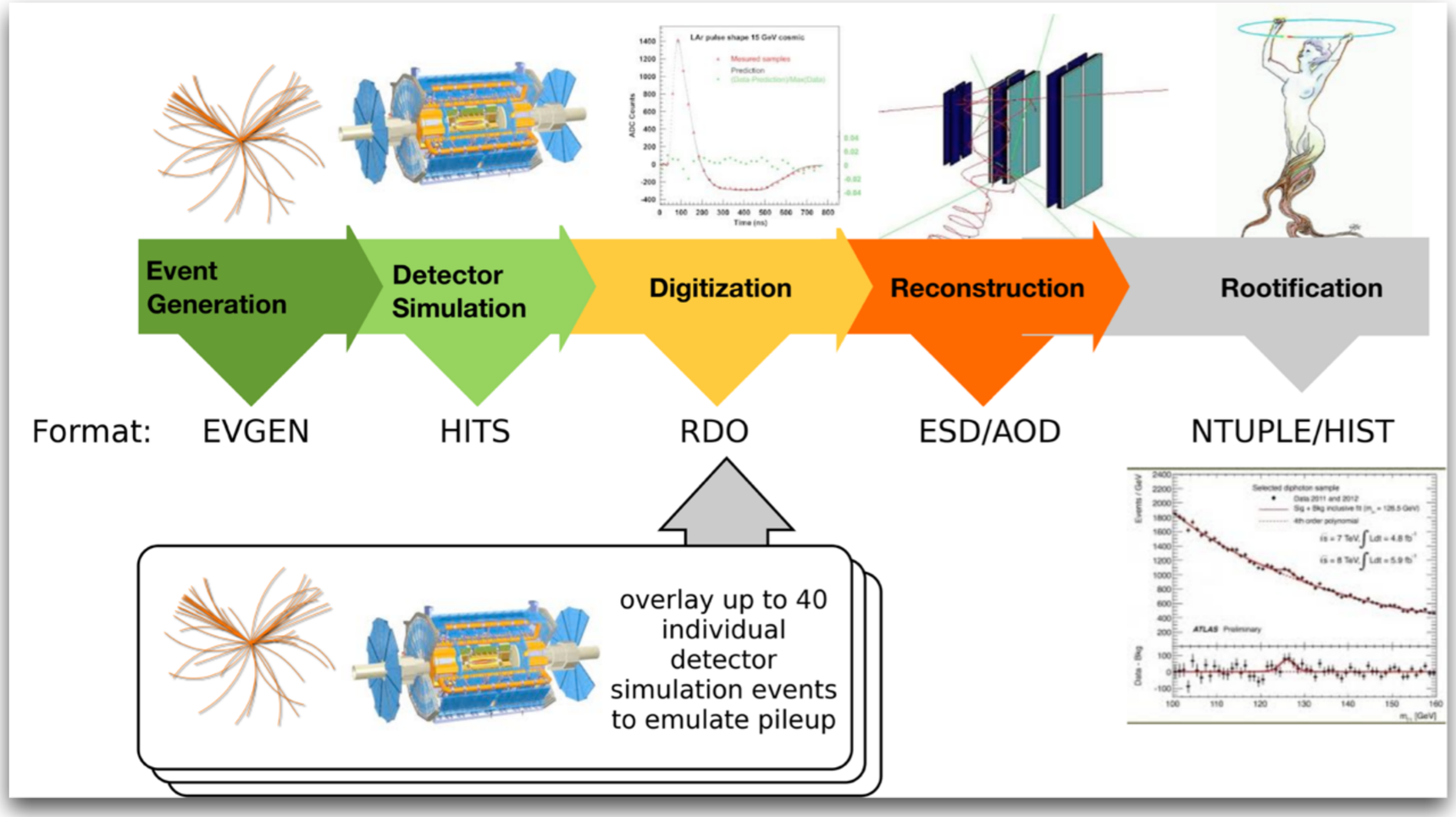
The ATLAS Collaboration

XIV. ACKNOWLEDGEMENTS

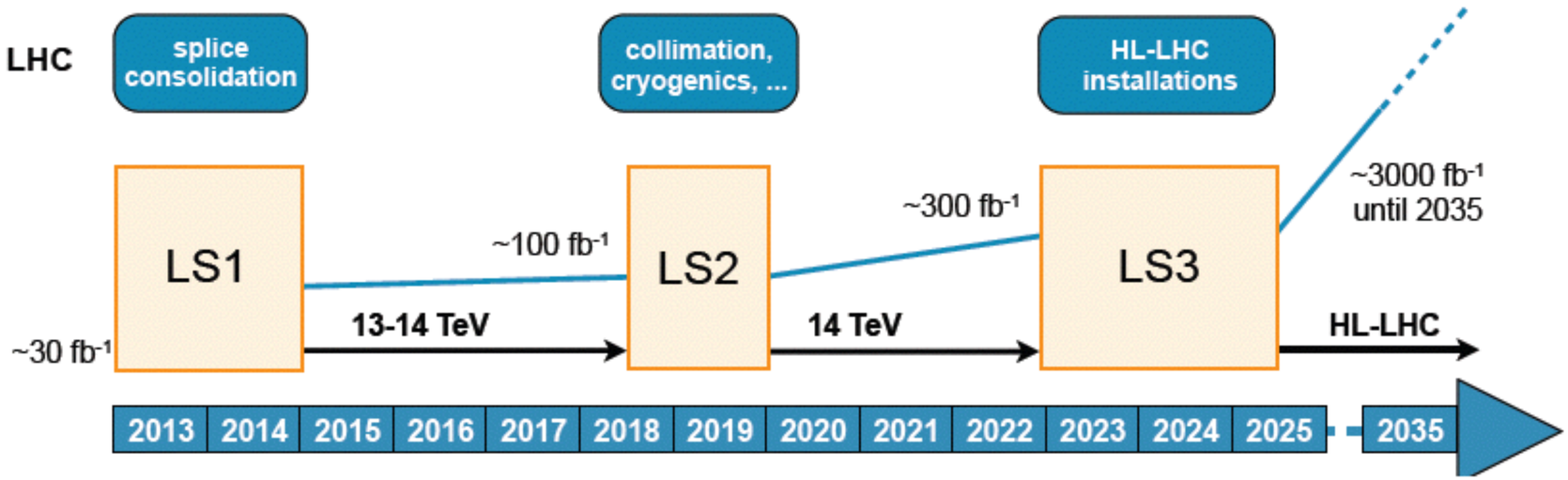
We thank T. Hapola for implementing the Minimal Walking Technicolor model using MadGraph to generate the signal and for his help with acceptance studies.

The limits shown in Section XII were calculated using computing resources provided by the Argonne Leadership Computing Facility and the National Energy Research Scientific Computing Center.

Simulation workflow



ATLAS Upgrade Roadmap



ATLAS Phase-0

New inner pixel layer
Detector consolidation
2015: FTK deployment

ATLAS Phase-1

Improve L1 Trigger, NSW and LAr electronics to cope with higher rates

ATLAS Phase-2

Prepare for 140-200 pile-up events
Replace Inner Tracker
New L0/L1 trigger scheme
Upgrade muon/calorimeter electronics
Upgrade of DAQ detector readout

A long and exciting road ahead !