



GridPP

UK Computing for Particle Physics



UNIVERSITY
of
GLASGOW

Needles and Haystacks

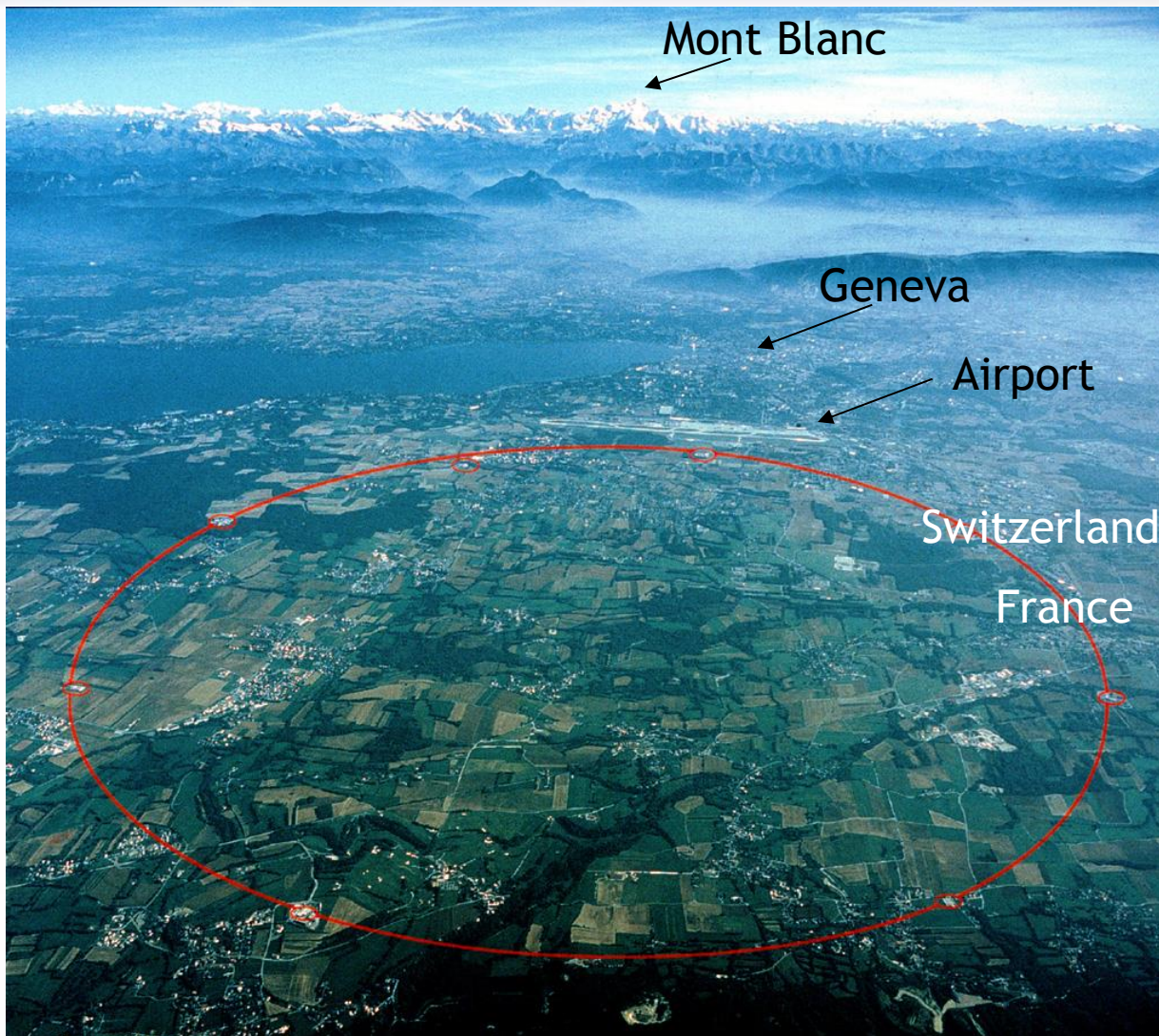
Glasgow

3rd Feb 2017

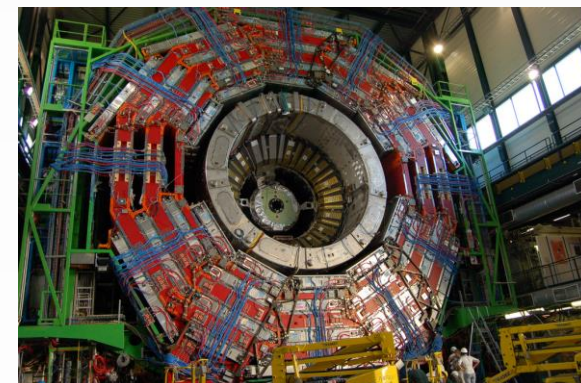
Prof. David Britton
GridPP Project leader
University of Glasgow



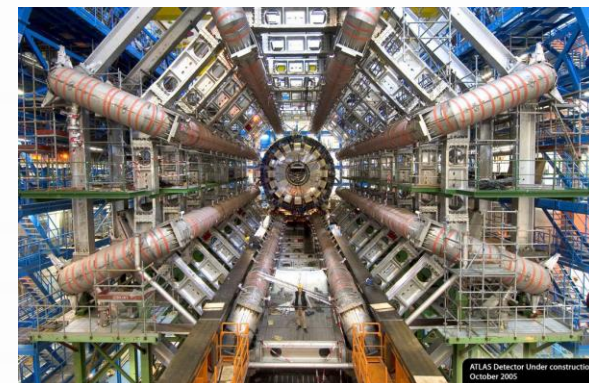
The LHC and Experiments



CMS



ATLAS





Data flow to permanent storage: 4-6 GB/sec

LHCb ~ 50 MB/sec

200-400 MB/sec

ATLAS ~ 320 MB/sec

1-2 GB/sec

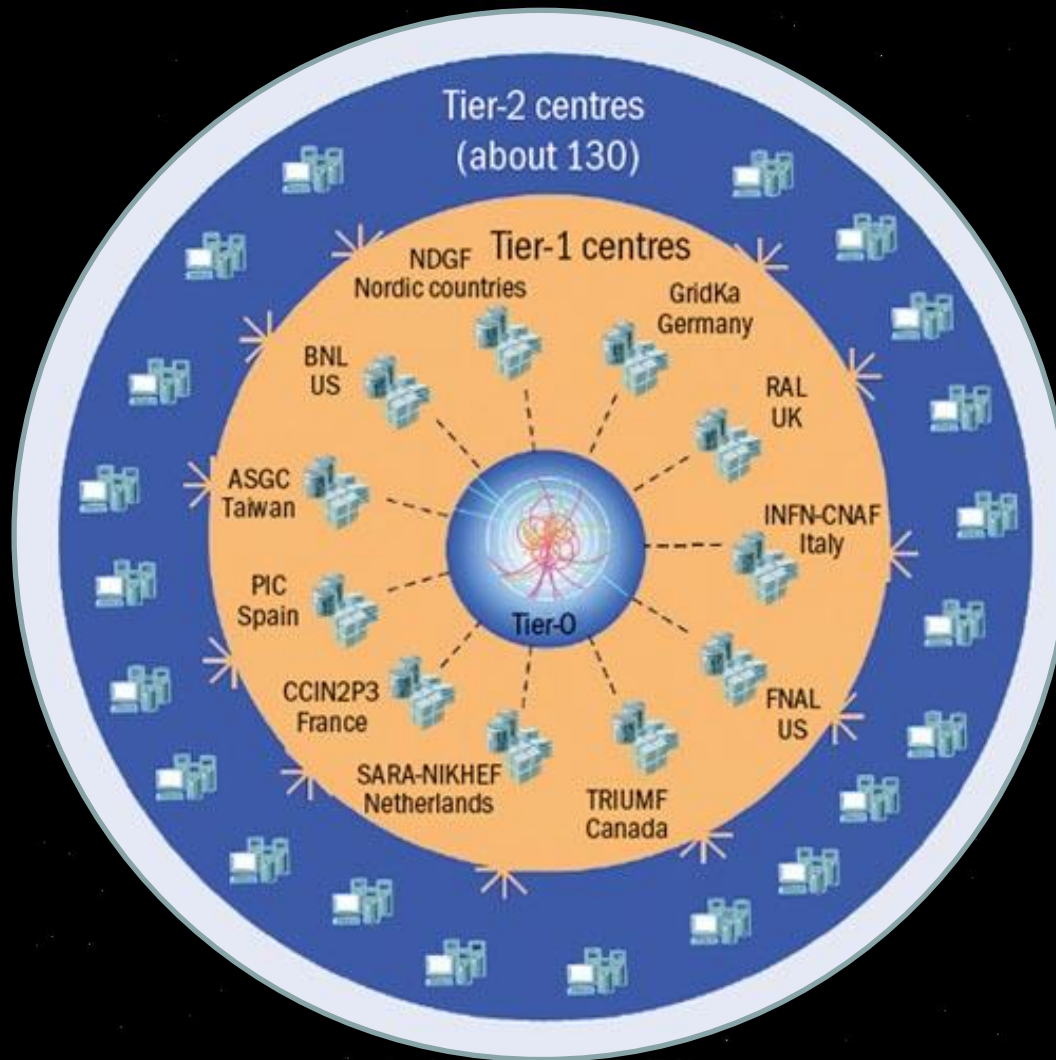
ALICE ~ 100 MB/sec

~ 4 GB/sec

CMS ~ 220 MB/sec

1-2 GB/sec

The Worldwide LHC Computing Grid



Infrastructure based predominantly on Linux machines running code written in C++.

WLCG:

An International collaboration to distribute and analyse LHC data

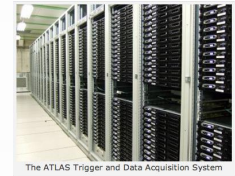
Integrates computer centres worldwide that provide computing and storage resource into a single infrastructure accessible by all LHC physicists

LHC Challenge - The Data Volume



The ATLAS Trigger and Data Acquisition System

Notionally
40TB/sec



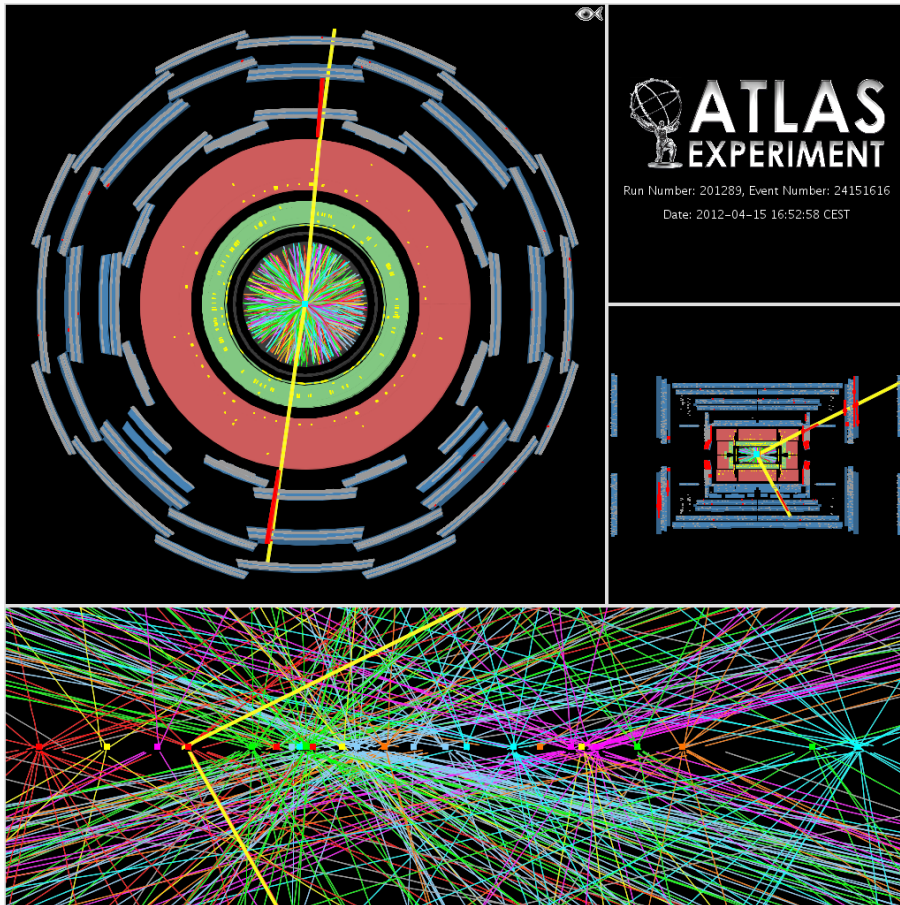
200MB/sec
recorded

Higgs?





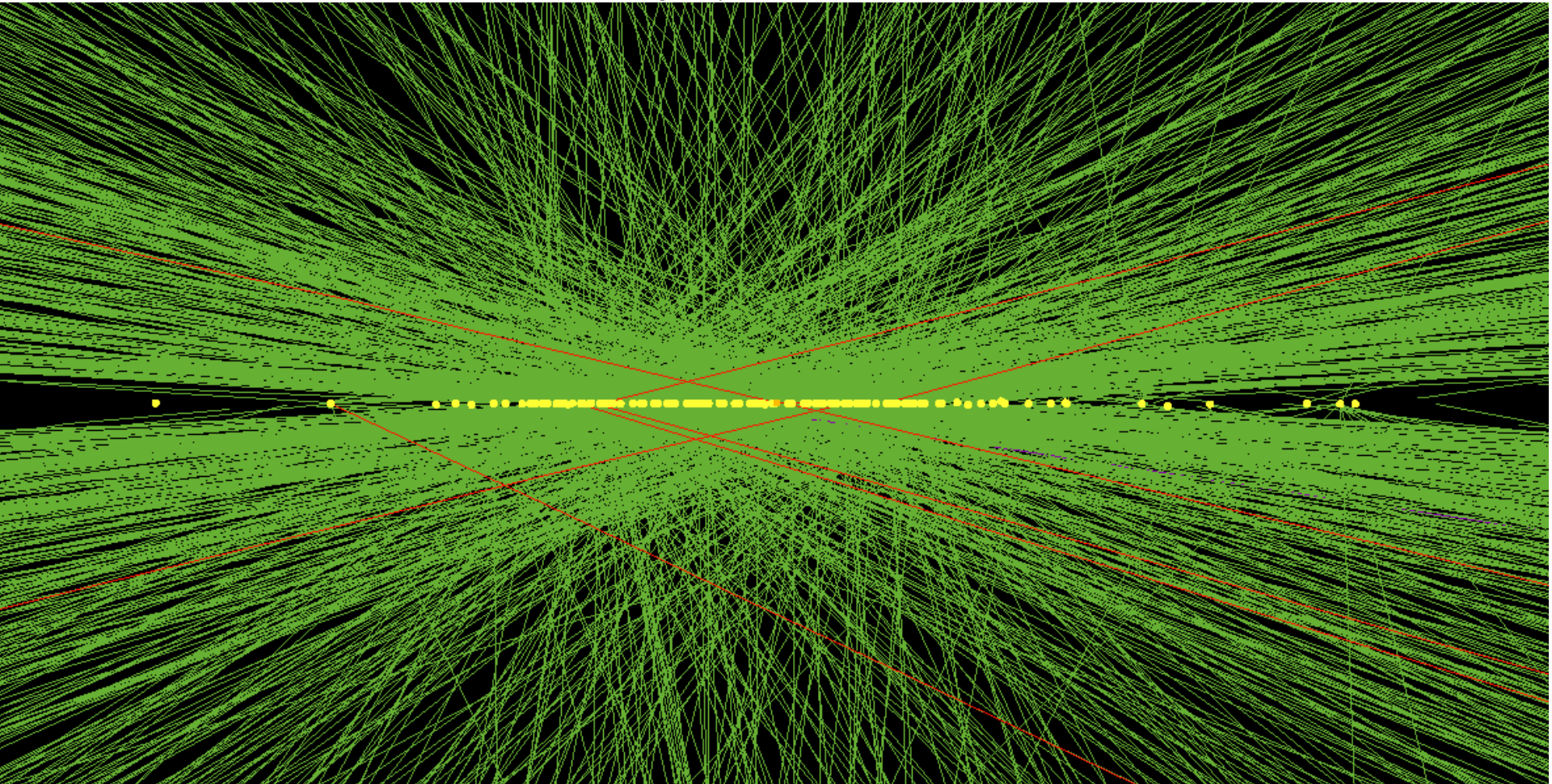
LHC Challenge: Data Complexity



- ✧ Collisions occur 40-million times a second.
- ✧ Each “event” is a composite of many separate interactions.



- Simulated Event Display at 140 PU (102 Vertices)





~~Hay!~~ Needle in the Hay-Stack

LOOKING FOR A SPECIAL PIECE OF HAY IN A HAYSTACK


by David Britton

WHY IS SO MUCH COMPUTING REQUIRED?

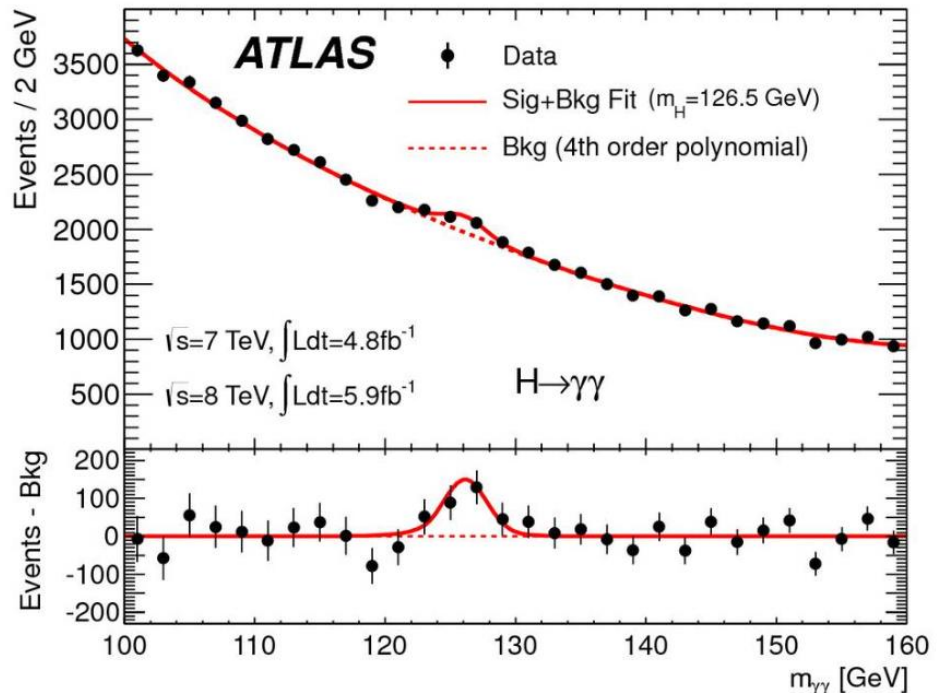
Looking for a Higgs boson is like looking for a special piece of hay in a haystack. This is a much harder problem than searching for a needle in a haystack – with a needle, you at least know when you’ve found it. At the LHC, a Higgs boson is produced in about one interaction in a billion, and an individual Higgs event is indistinguishable from the rest. To find the Higgs events, all of an experiment’s interactions need to be analysed by computer, to try to find a tiny excess in the number of particle decays at a specific, but unknown, mass. It’s like having a hundred or so pieces of hay, cut to the same length, and hidden at random places in an enormous haystack. The only way of identifying them would be to sort through all of the pieces of hay in the stack, divide them into piles by length, and identify the pile with a slight excess.

DAVID BRITTON

David Britton is a Professor of Physics at the University of Glasgow, and is leader of the GridPP project, having been a founder member in 2001. He has worked on both ATLAS, his current experiment, and CMS.



In fact, even when you find it, you can’t tell whether an individual event is a Higgs or not

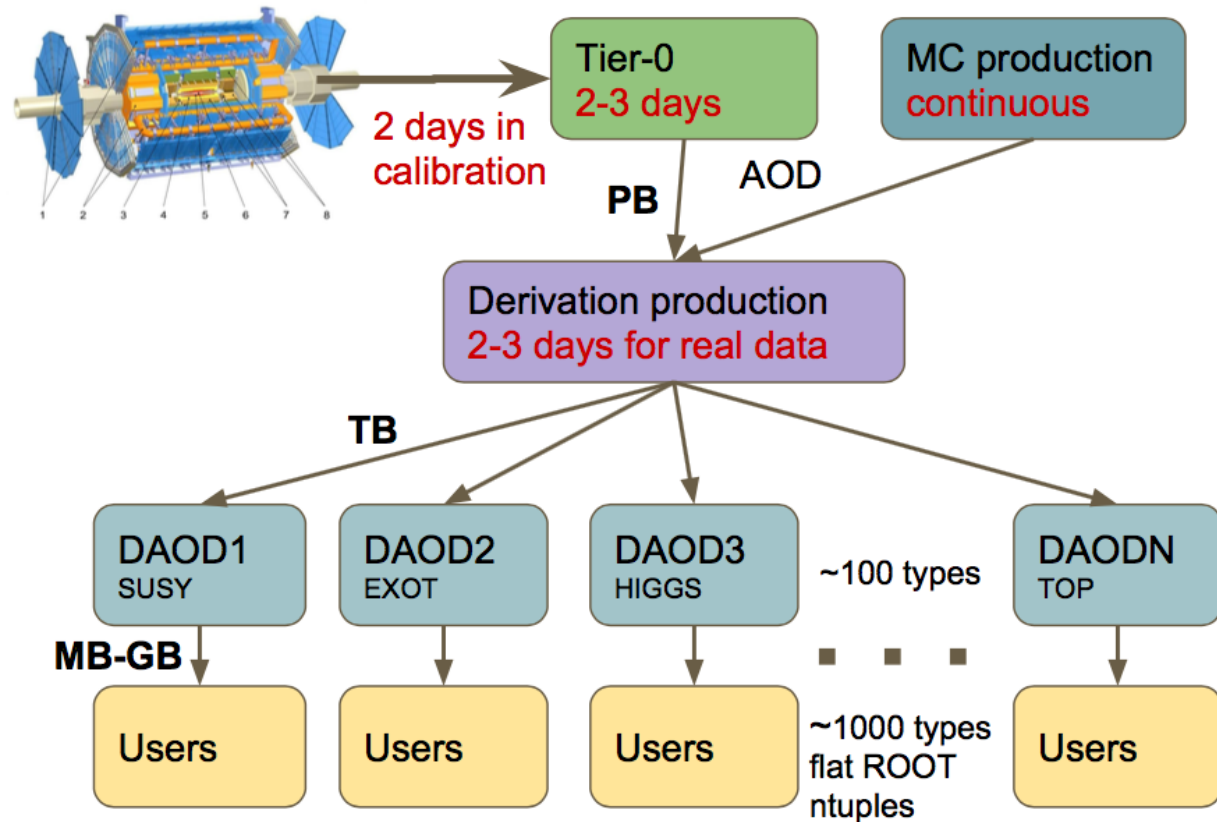




ATLAS experiment Data-Reduction

Derivations production

- Centrally managed production of analysis specific DAOD datasets (reduced data format from main AOD format)
- Real data:
 - Available ~1 week after data taking
- Several campaigns with improved analysis code on data and MC
 - 2-3 weeks to process



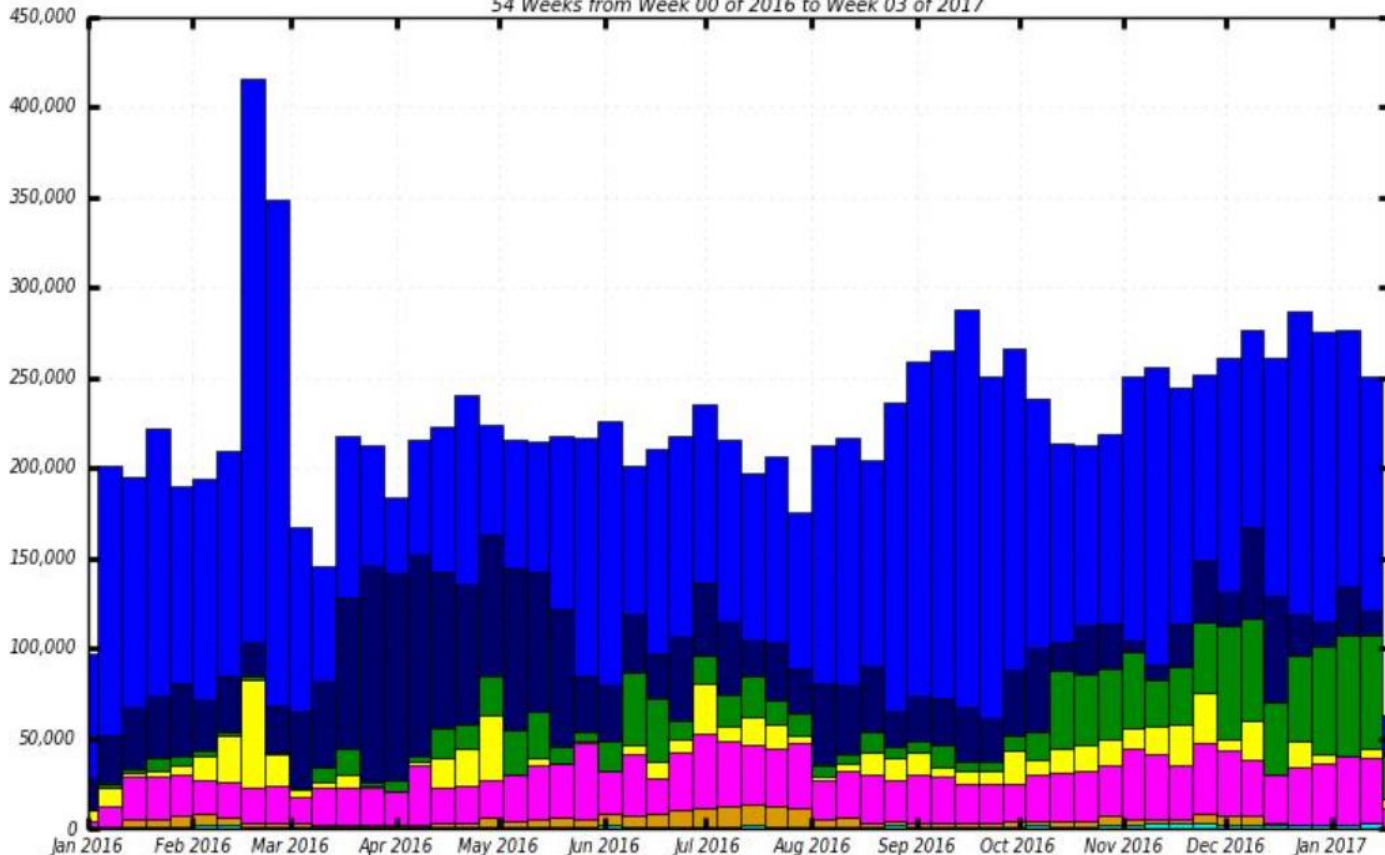


2016 ATLAS CPUs (Worldwide)



Slots of Running Jobs

54 Weeks from Week 00 of 2016 to Week 03 of 2017



Jobs

■ MC Simulation
■ TO Processing

■ MC Reconstruction
■ Others

■ Group Production
■ unknown

■ Data Processing

■ Analysis

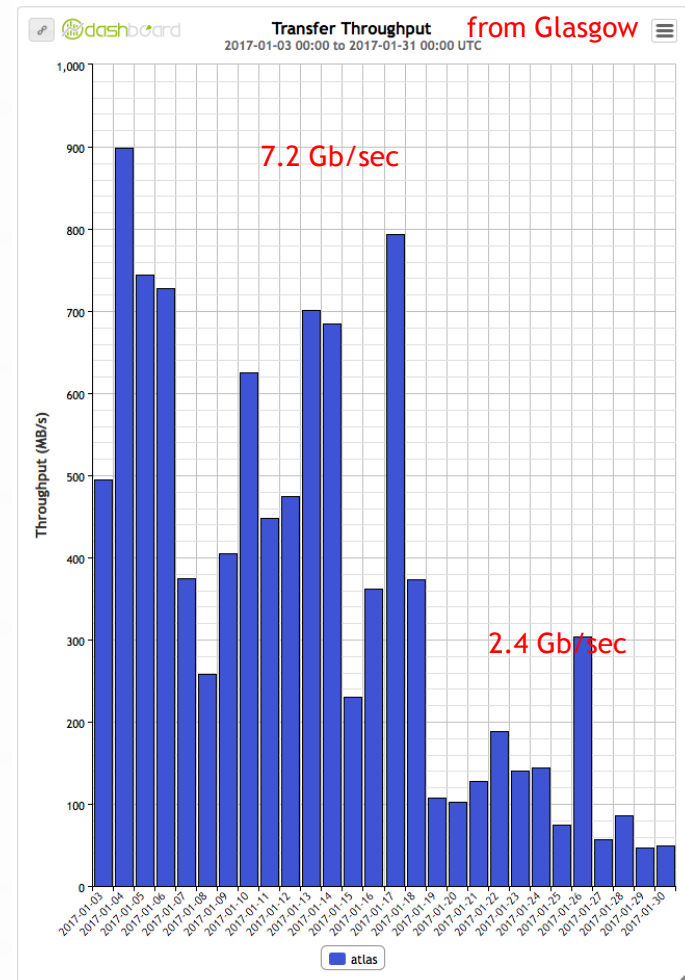
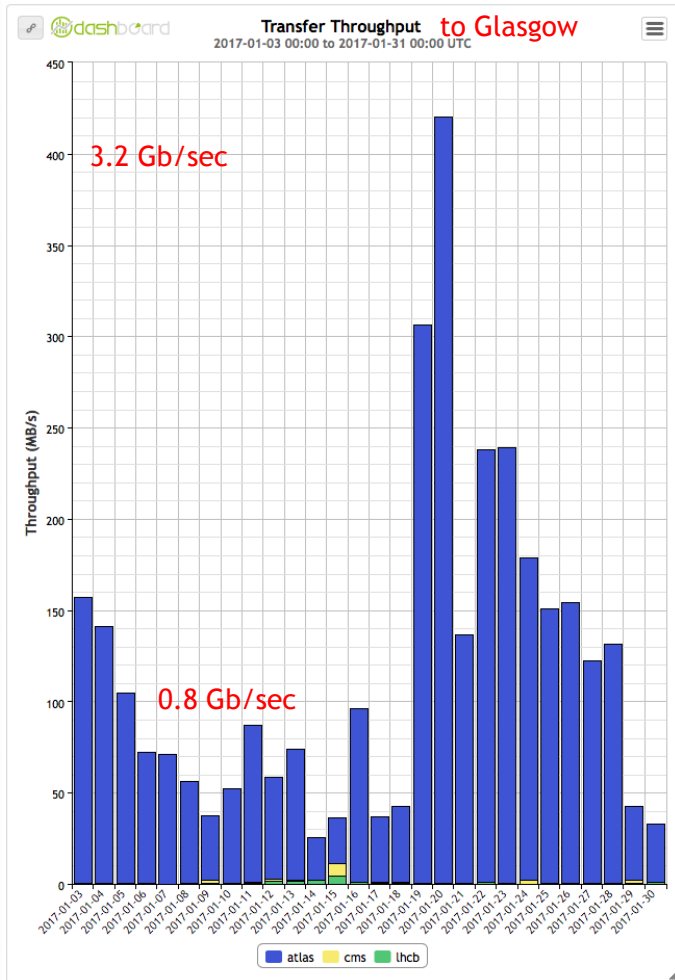
Maximum: 415,666 , Minimum: 0.00 , Average: 222,646 , Current: 62,490



- Largest scientific computing Grid in the world.
- >600,000 logical CPUs.
- >350 PB of disk storage.
- >335 PB of tape storage.
- >167 sites in 42 countries.
- Runs > 2 million jobs/day.
- Transfer ~100PB of data per month around the world.
- Pre-cursor to the development of cloud computing and big data
- UK provides about 10% of the resources via the GridPP project.
- Glasgow is biggest UK Tier-2 site and leads UK collaboration.
 - Shortly will have ~4PB storage and ~7000 cores.



HEP Data to/from Glasgow, Jan 17



Integral is ~1PB (in January!)



- We handle a lot of Particle Physics data at Glasgow.
 - Have some in-house expertise.
 - Have access to a vast amount of national and international expertise.
 - We dominate usage of the campus network link to JANET.
- Nationally, LHC data has been one of the major users of the JANET network but there are growing number of other large (and eventually larger) user groups developing.
 - What about Glasgow ?
- GridPP (nationally) supports ~30 VOs (user groups) running many different types of work. Some of these groups are small; some have nothing to do with physics.
 - We have developed tools and documentation to help people
 - We have (limited) manpower to assist.
 - We give away a small percentage (5-10%) of our CPU and storage to these groups.
- We continue to be keen to discuss new joint projects....



Examples and Case Studies

[HOME](#) / [USERS](#)

To give you an idea of what is possible with GridPP, we've selected some case studies featuring user communities that have successfully used GridPP resources in their work. Hover over the community name for a short summary, and click on it to read more.

| User Community | Sector | Compute | Storage | CernVM | CVMFS | DIRAC | Ganga |
|-------------------------------------------------|--------------|---------|---------|--------|-------|-------|-------|
| Galactic Dynamics (GalDyn) | Astrophysics | ✓ | ✗ | ✓ | ✓ | ✓ | ✗ |
| Large Synoptic Survey Telescope | Astrophysics | ✓ | ✓ | ✗ | ✗ | ✓ | ✓ |
| LUCID | Space | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ |
| PRaVDA | Healthcare | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ |
| HTC for Biology | Biology | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ |
| MoEDAL | Physics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

<https://www.gridpp.ac.uk/users/case-studies/>



Ran a small project with Brian Aitken from Arts:

- The AHRC funded SAMUELS project had produced a semantic tagger that figured out the meaning of words.
- Lancaster partners had run this over the 1.6 billion words in the Hansard parliamentary records and tagged 6.2 million records.
- Brian was categorizing these records and looking at frequencies but estimated his current set-up would take 2-3 years to process all the data.
- We helped him port his application to our computer cluster and completed the work over one weekend, producing 90GB of data to insert into a database, containing 682million rows, for future analysis.
- We are now hosting that database.