

Atlas Computing at IU

Frederick Luehring

Indiana University

Saturday March 25, 2006

Grad Day 2006

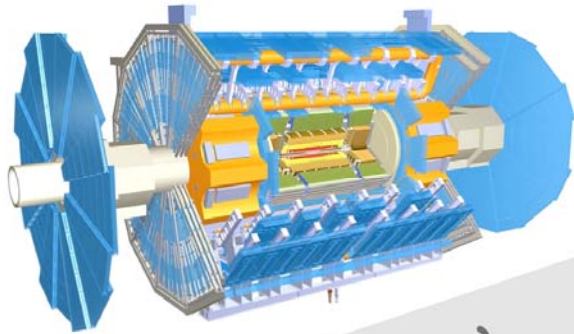
Atlas Computing at IU

- The IU HEP group is heavily into Grid Computing.
 - A simple definition of a computing grid is a system of distributed Computing Elements (CEs) and Storage Elements (SEs) connected by high-speed networks.
 - This implies establishing trust between geographically distributed sites and having software (middleware) that allows the efficient usage of distributed computer clusters.
- IU and the University of Chicago are 1 of 3 Tier 2 sites in the US. A Tier 2 site is a grid computing cluster with substantial resources.
 - By 2009, this Tier 2 center will have ~1000 kSi2k of CPU power, and 800 TB of disk for use in Atlas data analysis.
 - We will have 5 people working on grid operations, 2 people on our Tier 2, 2 people on Atlas Analysis Support so there will be lots of help for data analysers.

The Scale of Data Production

- The LHC beams cross every 25 ns. Each bunch crossing produces ~ 23 interactions at the design luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$.
 - Initial LHC operation will be at much lower luminosity ($\sim 10^{31} - 10^{32} \text{ cm}^{-2}\text{s}^{-1}$).
- The Atlas trigger selects 200 bunch crossings per second for full read-out.
 - The readout data size ranges from 1.7 MB/event initially to about 3.4 MB/event at full luminosity.
- When LHC is fully operational, the LHC beams will be available to the experiments for 10^7 s/year.
 - Simple arithmetic gives 3.4 to 6.8 PB of readout data a year (depending on the exact luminosity).

Data Producer



40 MHz (80 TB/sec)

level 1 (hardware trigger)

O(75) KHZ (150 GB/sec)

level 2 (RoI analysis)

O(2) KHZ (4 GB/sec)

level 3 (Event Filter)

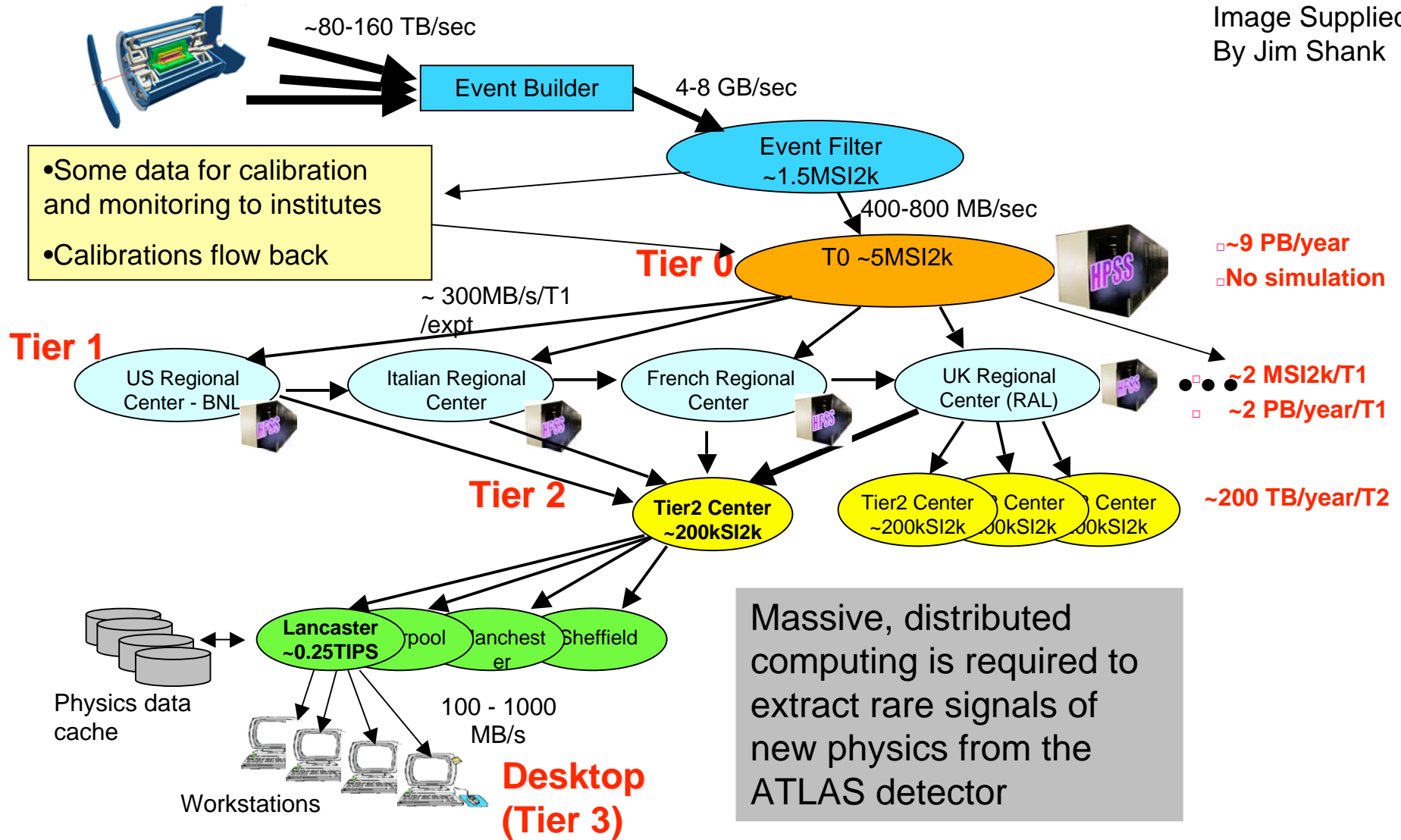
200 HZ

O(400) MB/sec

- RAW
 - 1.6MB/event
- ESD
 - Event Summary Data
 - .5MB/event
- AOD
 - Analysis object data
 - .1 MB/event
- TAG
 - Event=level metadata tags
 - 1 KB/event
- DPD
 - Derived physics data
 - N-tuple type, user based

The Atlas Data Flow

Image Supplied
By Jim Shank



•Some data for calibration and monitoring to institutes
•Calibrations flow back

Massive, distributed computing is required to extract rare signals of new physics from the ATLAS detector

Projected T2 Hardware Growth

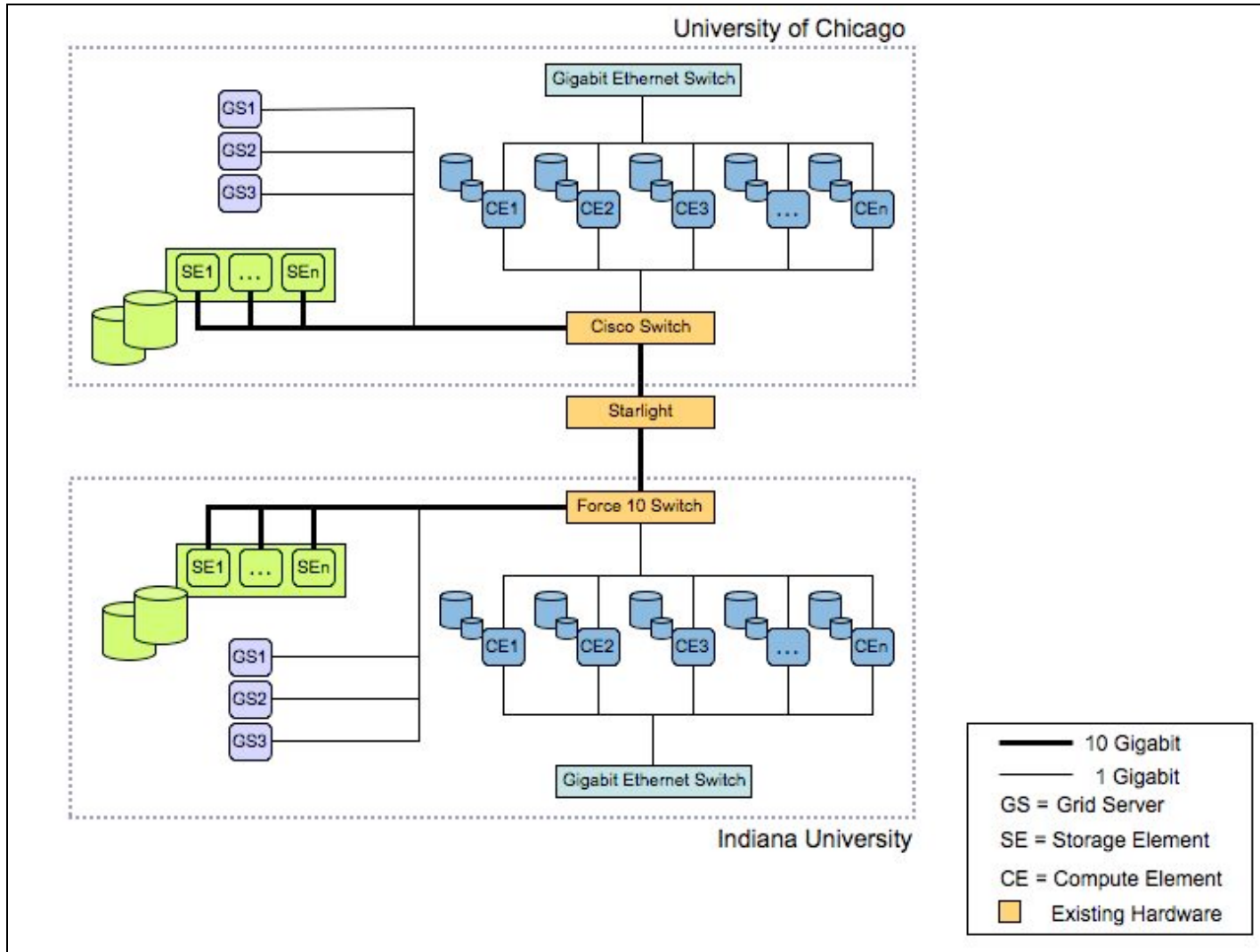
J. Shank, NSF Review 7/05

| Tier 2 Center | 2005 | 2006 | 2007 | 2008 | 2009 |
|---------------|------|-------|-------|-------|-------|
| Northeast | | | | | |
| CPU (kSi2k) | 210 | 350 | 730 | 1,090 | 1,600 |
| Disk (TB) | 40 | 170 | 370 | 480 | 630 |
| Southwest | | | | | |
| CPU (kSi2k) | 600 | 1,000 | 1,600 | 1,700 | 2,100 |
| Disk (TB) | 60 | 200 | 380 | 540 | 700 |
| Midwest | | | | | |
| CPU (kSi2k) | 100 | 240 | 465 | 700 | 1,050 |
| Disk (TB) | 50 | 130 | 260 | 465 | 790 |

- Assumes Moore's law doubling of CPU and disk capacity every 3 years at constant cost
- Assumes replacement of hardware every 3 years

MWT2 Architecture

Drawing by Kristy Kallback-Rose



Conclusions

- IU will be a great place to pursue analysis of data Atlas/LHC using the grid.
 - The LHC project is the largest scientific instrument ever built.
 - We will have both a large computing center here at IU and a very strong presence at CERN.
 - We will have a large pool of expertise both on using grids and on physics analysis.
 - We enjoy working with graduate students.
- You are in the perfect place in your careers to take advantage of the this wonderful opportunity.
 - You will have completed your first year of course work just as the first Atlas data is being taken.
 - Just as Atlas goes into year round operation in 2008, you will be finishing your course work be able to take full advantage of the Atlas data.