This is the LHC: Where is the Data?
Outline

- LHC Background
- Recent News ... Colliding at Last!
- So ... Where is All the Data?
- Connectivity
- What to expect
LHC Background

- Large Hadron Collider (LHC) - World's largest and highest-energy particle accelerator
- Synchrotron intended to collide opposing particle beams
- Beneath the Franco-Swiss border near Geneva, Switzerland
  - Operated by CERN (home of the Internet!)
- Testing various predictions of high-energy physics:
  - hypothesized Higgs boson (‘God’ particle)
  - “Supersymmetry”
- Operational History
  - 9/10/2008 - First beam circulation … then the magnets failed
  - 11/20/2009 - Another beam circulation and collision at low energy (1.18 TeV) … then it was shut down for the winter
  - 3/30/2010 – Collisions at 3.5 TeV, operating currently
LHC Background - Tiers

- **Tier 0**
  - CERN Computing center, permanent home of all data

- **Tier 1s (11 Total)**
  - TRIUMF - Canada
  - KIT - Germany
  - PIC - Spain
  - IN2P3 - France
  - INFN-CNAF - Italy
  - NDGF - Scandinavia
  - NL-T1 - Netherlands
  - ASGC - Taiwan
  - RAL - United Kingdom
  - Fermilab - USA
  - Brookhaven Lab - USA

- **Tier 2s**
  - Numerous, depends on the project.

- **Tier 3s**
  - Even more numerous. Note that ‘Tier3’ status is normally not an official designation so anyone can be a Tier3
LHC Background – Data Requirements

• Data requirements (raw):
  – Roughly 1 billion proton-proton collision events per second
  – Heavily filtered - 100 events per second will be recorded
  – Each event = a few MB
  – Expected data rate from the experiment(s) = 1 Gbps.
  – Data stored to CERN computing center (Tier-0)

• Data requirements (including raw, processed, and simulated data):
  – 15 Petabytes of data per year
  – Copies will be stored at the Tier-1s
  – Analysis will be carried out by Tier-2s
  – Computing power equivalent to ~100,000 CPUs in the distributed LHC Grid
LHC Background – Network Overview

- LHCOPN – Private optical network(s):
  - Tier-0 to Tier-1s
  - Some Tier-1s to other Tier-1s
- Tier-1s to Tier-2s, connectivity varies:
  - Private connectivity
    - Ultralight, USLHCnet
    - Other private (direct) links
  - R&E Networks
    - ESnet, Internet2, NLR
- Tier-2s to other Tier-2s
  - Sometimes private (see above)
  - Mostly R&E
- Tier3s
  - R&E in most cases
  - Commercial Internet (smaller facilities)
LHC Background – LHCOPN
Recent News ... Colliding at Last!

- March 30, 2010 - 13:06 CEST
  - Crossing the streams at 3.5 TeV (7 TeV total - record setting)
  - 3rd attempt of the day (previous two runs failed)
- LHC will run for 18–24 months (expected shutdown during the winters)
- Some collisions since first
- Lots of processing
  - Tier-2 centers are processing
  - Some early results
- Could take a while to find Higgs, other results expected as processing continues
After a collision, the data does not flow directly from Tier-0 to Tier-1, etc. There are some steps first:

- “Calibration”. Up to 48 hours where the data is reconstructed. This is done at Tier-1 and Tier-2 centers.
- Tier-0 will re-construct the original data using the calibration data.
- Data flows from Tier-0 to Tier-1s (couple of hours latency)
- Data flows from Tier-1s to Tier-2s (couple more hours latency)

Each site (ideally) gets a small piece of the entire set

- Tier-1s get the complete copy
- For processing, follow the ‘Grid’ model (everyone processes a small set)
- Sets are downloaded at Tier-2s/Tier-3s as needed (and resources allow)
So ... Where is All the Data?

• What about the data? Example taken from CMS experiment
  – Other experiments have results too
    • ATLAS
    • ALICE
    • LHCb

• First data set was “small”
  – 90 Billion “events” (after processing)
  – About 9 TB total
  – 9 TB @ 10 Gbps = ~2hrs to transfer (at best).

• Data Set Specifics
  – 2728 ‘Files’
  – 292 ‘Blocks’ - A block is the unit of operation for much of the analysis code. A Tier-2/Tier-3 would download a couple of blocks at a time and start processing.
  – Currently (as of 4/21) replicated completely at 19 sites
• Data dissemination (ideally)
  – Propagation from Tier-0 to Tier-1s via the LHCOPN
  – Tier-2s *should* be grabbing from a designated Tier-1
  – Tier-3s *should* be grabbing from certain Tier-2s
  – Organization is done per experiment (e.g. USATLAS as a VO manages Tier-1s, Tier-2s, and Tier-3s).

• Data dissemination (currently)
  – Early days of the experiment traded simulated (Monte Carlo) data
  – Tools exist (PhEDEx) to download sets in an organized manner (CMS Model)
  – Not much attention was given to the details in the Tier model

• When the data first hit – there was excitement
  – Some sites got data from CERN directly
  – Tier-3s may have downloaded form Tier-1s when available
  – Sharing between Tier-2s that received it 1st
So ... Where is All the Data?

- Aggregate traffic from Fermilab/BNL on Internet2 Network
- Dates: 3/30 to 4/2 (First collision through data dissemination)
- Note the ‘peaks’ of around 3-5G. Didn’t last too long, supports the maximum size of the data set.
- Graph courtesy of Chris Robb.
So ... Where is All the Data?

- Aggregate traffic from Fermilab/BNL on Internet2 Network
- Dates: 3/30 to 4/2 (First collision through data dissemination)
- Note the ‘peaks’ of around 3-5G. Didn’t last too long, supports the maximum size of the data set.
- Graph courtesy of Chris Robb.

Possible Transfers?
So ... Where is All the Data?

- Despite these facts on data size and where it came from, did we see the data on Internet2?
  - “Some”, but not all
  - A little later than first availability (more with Tier2 and Tier3 transfers)
- Who saw the data?
  - Private nets (Ultralight, USLHCNet)
  - ESnet (into and out of Fermilab/BNL)
  - Internet2/NLR
    - From Tier1s
    - Between Tier2s
    - To Tier3s
Connectivity

- Minor experiment to see how Tier-2s route to each other, and the Tier-1 for USATLAS.
- pS Performance Toolkit (http://psps.perfsonar.net/toolkit/) available at Tier-1 and most Tier-2s.
  - Co-allocated near the rest of the processing/storage
  - Using performance tools, analyze the routes
  - Determine how the data is flowing
  - Check the times/data stores to find evidence of the transfers
  - ‘Reverse Traceroute’ Tool – Developed by SLAC
Connectivity – BNL (Tier 1)

- Tier 1 for USATLAS
- Connectivity to other sites (Tier-2s)
  - MSU/UMich – Ultralight
  - Indiana - ESnet
  - U of Chicago – Private Network/Peering
  - Boston Univ. – Private Network/Peering
  - Oklahoma - ESnet
  - U of Texas at Arlington – Esnet/NLR
  - SMU – ESnet/Internet2
  - U of Wisconsin - ESnet
  - LBNL/NERSC – ESnet
- As expected for a Tier1, there is not much touching Internet2
- Surprising that there were several private networks
Connectivity – Indiana (Tier 2)

• Tier 2 (MWT2 w/ U of C), equipment in Indianapolis
• Connectivity to other sites
  – BNL (Tier-1) - ESnet
  – MSU/UMich (Tier-2) – Internet2
  – U of Chicago (Tier-2) – Private Network/Peering (internal routing)
  – Boston Univ. (Tier-2) – Internet2
  – Oklahoma (Tier-2) - NLR
  – U of Texas at Arlington (Tier-2) - NLR
  – SMU (Tier-3) - NLR
  – U of Wisconsin (Tier-3) - NLR
  – LBNL/NERSC (Tier-3) – Esnet
• Very well connected (diversity in routing)
Connectivity – U of Chicago (Tier 2)

- Tier 2 (MWT2 w/ Indiana)
- Connectivity to other sites
  - BNL (Tier-1) – Private Network/Peering
  - MSU/UMich (Tier-2) – Private Network/Peering
  - Indiana (Tier-2) – Private Network/Peering (internal routing)
  - Boston Univ. (Tier-2) – Internet2
  - Oklahoma (Tier-2) – Internet2
  - U of Texas at Arlington (Tier-2) – Internet2
  - SMU (Tier-3) – Internet2
  - U of Wisconsin (Tier-3) – Private Networking/Peering
  - LBNL/NERSC (Tier-3) – ESnet
- Several private peerings, well connected.
- Major partner in OSG development
Connectivity – Boston Univ. (Tier 2)

- Tier 2 (NET2 w/ Harvard)
- Connectivity to other sites
  - BNL (Tier-1) – Private Network/Peering
  - MSU/UMich (Tier-2) – Internet2
  - Indiana (Tier-2) – Internet2
  - U of Chicago (Tier-2) – Internet2
  - Oklahoma (Tier-2) – Internet2
  - U of Texas at Arlington (Tier-2) – Internet2
  - SMU (Tier-3) – Internet2
  - U of Wisconsin (Tier-3) – Internet2
  - LBNL/NERSC (Tier-3) – ESnet
- Private connectivity to the Tier1 (shared with NET2 partner Harvard)
Connectivity – Oklahoma (Tier 2)

- Tier 2 (SWT2 w/ U of T at A)
- Connectivity to other sites
  - BNL (Tier-1) - ESnet
  - MSU/UMich (Tier-2) - NLR
  - Indiana (Tier-2) - NLR
  - U of Chicago (Tier-2) – Internet2
  - Boston Univ. (Tier-2) – Internet2
  - U of Texas at Arlington (Tier-2) - NLR
  - SMU (Tier-3) - NLR
  - U of Wisconsin (Tier-3) - NLR
  - LBNL/NERSC (Tier-3) – ESnet
- Well connected site.
What to Expect

• Frequent collisions will generate more (and larger) data sets
  – Activity through November, followed by winter shutdown
  – Startup again around March 2011
  – Switch over to ‘Heavy ION’ activities once a year will change the Tier models a little (different sites in charge of data)

• Excitement over startup will wane – return to “business as usual” with the Tier model
  – Expect more large transfers between Tier-2s and the Tier-1 (and each other – particularly in the CMS model…)
  – Tier-3s from all over will start getting regular data

• Transfer of ‘entire’ datasets at the Tier-2 is unexpected, but still large quantities

• Tier-3 participation will start slow, become more significant over time
What to Expect – 2 Examples

- Tier-2 to Tier-2 Transfers
- International Tier-2 Transfers, sometimes International (CMS)
Example of T2 – T2: 4/26 7 EDT

- Inbound to Internet2 from GPN (UNL – A CMS Tier-2)
Example of T2 – T2: 4/26 7 EDT

- Outbound to CalREN (Caltech – A CMS Tier-2)
Example of T2 – T2: 4/26 7 EDT

- Backed up by CMS PhEDEx data
Example of T2 – T2: 4/26 7 EDT

- Backed up by CMS PhEDEx data
Other Example of a T2: 4/15 to date

- Backbone traffic heating up (NEWY-CHIC)
Other Example of a T2: 4/15 to date

- Tracked to University of Wisconsin (USCMS Tier-2)
Other Example of a T2: 4/15 to date

- PhEDEx confirms (into UofWisc)
Other Example of a T2: 4/15 to date

- Some (not all) coming out of a Tier-1 in Germany (KIT/GridKA)
This is the LHC: Where is the Data?
April 27th 2010, Internet2 Spring Member Meeting
Network Member and Connector BoF
Eric Boyd, Deputy Technology Officer

For more information, visit www.internet2.edu/performance
Connectivity – UMich/MSU (Tier 2)

- Tier 2 facility (ALGT2), share common routing infrastructure
- Connectivity to other sites
  - BNL – Ultralight
  - Indiana -
  - U of Chicago -
  - Boston Univ. -
  - Oklahoma -
  - U of Texas at Arlington -
  - SMU -
  - U of Wisconsin -
  - LBNL/NERSC –
- Ultralight does the heavy lifting. Still very well connected.
Connectivity – U. Texas at Arlington (Tier 2)

- Tier 2 (SWT2 w/ U of O)
- Connectivity to other sites
  - BNL – ESnet/NLR
  - MSU/UMich - NLR
  - Indiana - NLR
  - U of Chicago - NLR
  - Boston Univ. – Internet2
  - Oklahoma – NLR
  - SMU - LEARN
  - U of Wisconsin - NLR
  - LBNL/NERSC – ESnet
- Well connected site.
Connectivity – Southern Methodist (Tier 3)

- Tier 3
- Connectivity to other sites
  - BNL – ESnet/Internet2
  - MSU/UMich – Internet2
  - Indiana – Internet2
  - U of Chicago – Internet2
  - Boston Univ. – Internet2
  - Oklahoma – Internet2
  - U of Texas at Arlington – LEARN
  - U of Wisconsin – Internet2
  - LBNL/NERSC – ESnet/Internet2
- Connected via LEARN, expected to get most data from close Tier-2s (SWT2 partners UTA and OU)
Connectivity – U of Wisconsin (Tier 3)

• Tier 3
• Connectivity to other sites
  – BNL - ESnet
  – MSU/UMich – Private Network/Peering
  – Indiana - Private Network/Peering
  – U of Chicago - Private Network/Peering
  – Boston Univ. – Internet2
  – Oklahoma – Internet2
  – U of Texas at Arlington – Internet2
  – SMU – Internet2
  – LBNL/NERSC – ESnet
• Several private peerings, very well connected
• Major partner in OSG development
Connectivity – LBNL/NERSC (Tier 3)

- Tier 3
- Connectivity to other sites
  - BNL - ESnet
  - MSU/UMich – ESnet/Ultralight
  - Indiana - ESnet
  - U of Chicago - ESnet
  - Boston Univ. - ESnet
  - Oklahoma - ESnet
  - U of Texas at Arlington – ESnet/NLR
  - SMU – ESnet/Internet2
  - U of Wisconsin – ESnet
- DOE site, connectivity is ESnet