EXPERIENCE WITH GLOBUS ONLINE AT FERMILAB

Thu Apr 12, 2012

Gabriele Garzoglio
Computing Sector
Fermi National Accelerator Laboratory

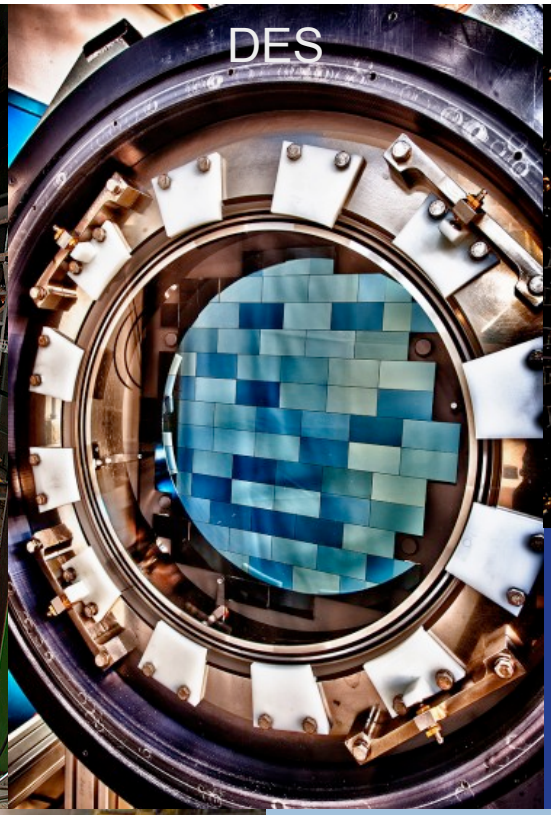
U.S. DEPARTMENT OF ENERGY

Fermilab

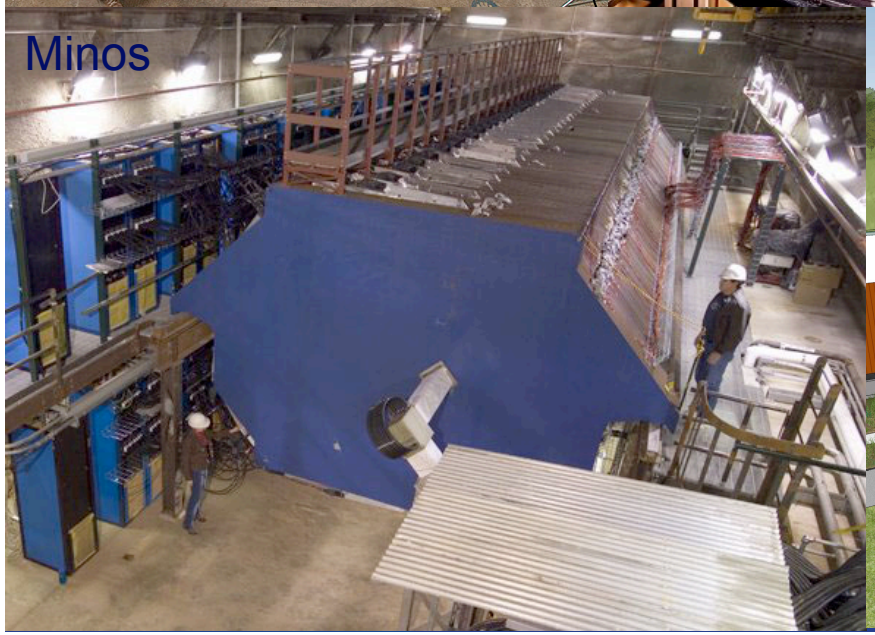
Why am I here?

- The Fermi National Accelerator Laboratory serves a large number of communities in High Energy Physics
  - Through the partnership with the Open Science Grid Consortium, Fermilab collaborates with communities from a variety of scientific fields
- It is in the mission of the Scientific Computing at Fermilab (and my department - Grid and Cloud Computing) to support their computations on highly distributed resources
- We cannot (…and are not interested in…) reinventing all services necessary to do this
- To achieve our mission, we create a network of partnerships with international groups of computer scientists
- These partnerships make us and our partners stronger because of our users
- Disclaimer: this talk only covers our work on GO


Fermi National Accelerator Laboratory

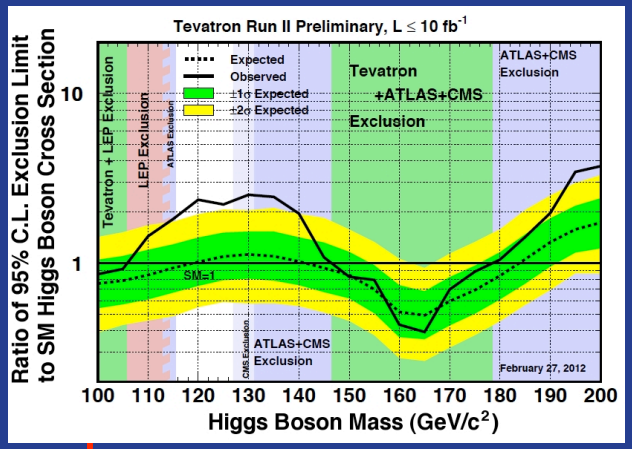


Fermilab is evolving to match the needs of future science



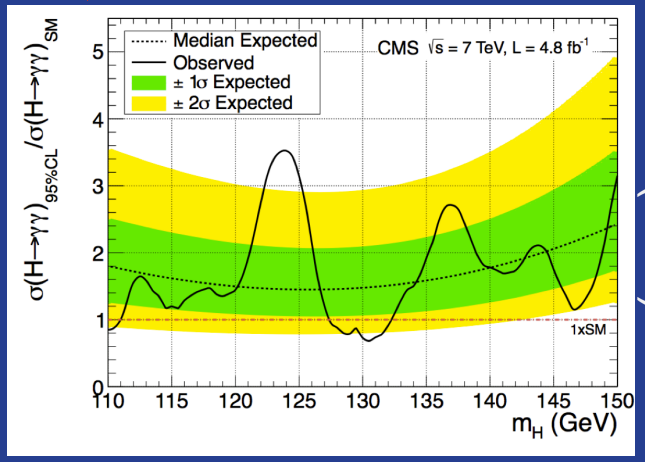
It's about the science!

- The D0 and CDF examples:
  - CDF & D0 - Top Quark Asymmetry Results.
  - CDF - Discovery of Ξ_b^0.
  - CDF - $\Lambda_c(2595)$ baryon.
  - Combined CDF & D0 Limits on standard model higgs mass.
- Hundreds of publications per year!

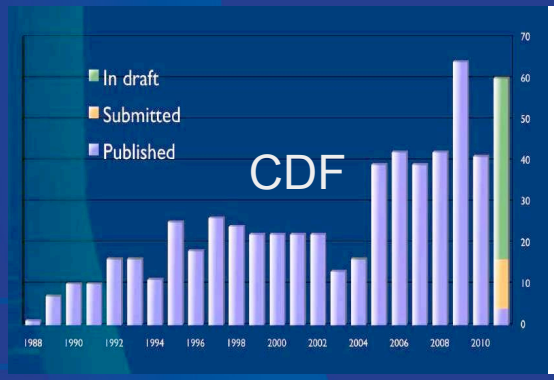


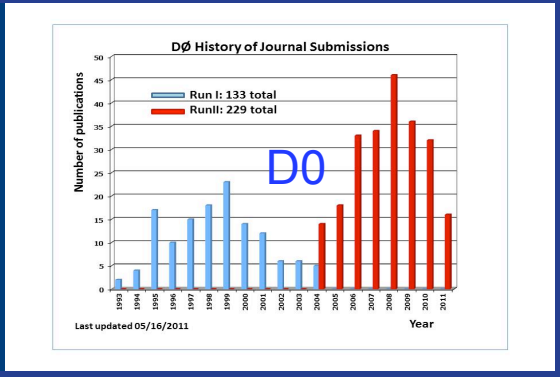
CDF & D0

Fermilab Collaborates with CERN on the LHC

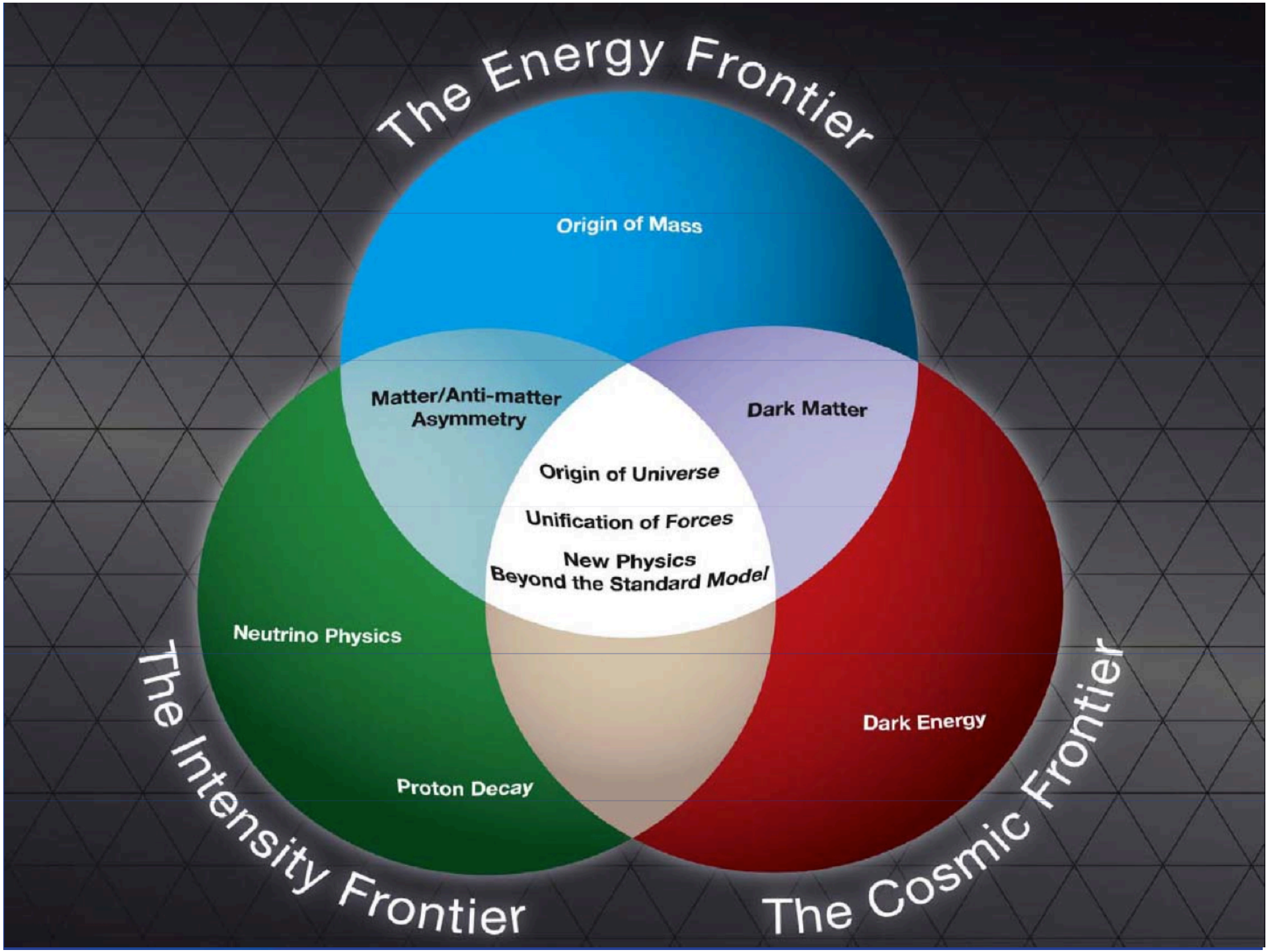


CMS (LHC)



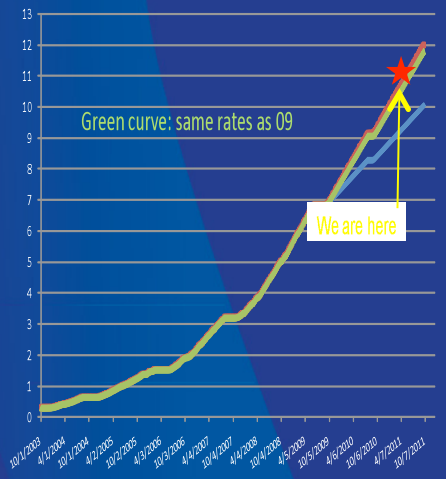


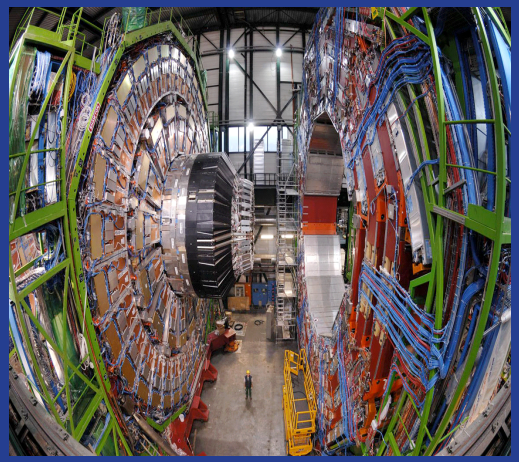
- Data recorded over O(10) yrs: CDF = 8.63 PB; Dzero 7.54 PB. Now we expect 5+ PB every year for LHC
- Fermilab will support analysis for 5+ years and data access for 10+ years.


The Energy Frontier
Origin of Mass
Matter/Anti-matter Asymmetry
Dark Matter
Origin of Universe
Unification of Forces
New Physics Beyond the Standard Model
Neutrino Physics
Dark Energy
Proton Decay
The Intensity Frontier
The Cosmic Frontier

Roles: energy frontier

Now	2013	2016	2019	2022
Tevatron LHC	LHC	LHC Upgrades ILC??	LHC ILC, CLIC or Muon Collider	





From P. Oddone, Secretary of Energy's Visit, June 2nd, 2011

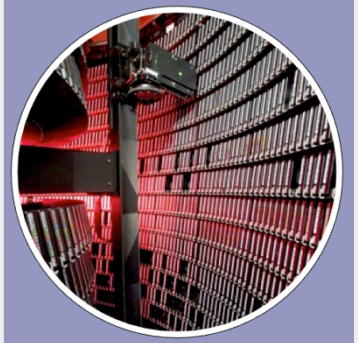
Energy frontier: the legendary Tevatron

Accelerator Innovations

- First major SC synchrotron
- Industrial production of SC cable (MRI)
- Electron cooling
- New RF manipulation techniques

Detector innovations

- Silicon vertex detectors in hadron environment
- LAr-U238 hadron calorimetry
- Advanced triggering



Analysis Innovations

- Data mining from Petabytes of data
- Use of neural networks, boosted decision trees
- Major impact on LHC planning and developing
- GRID pioneers

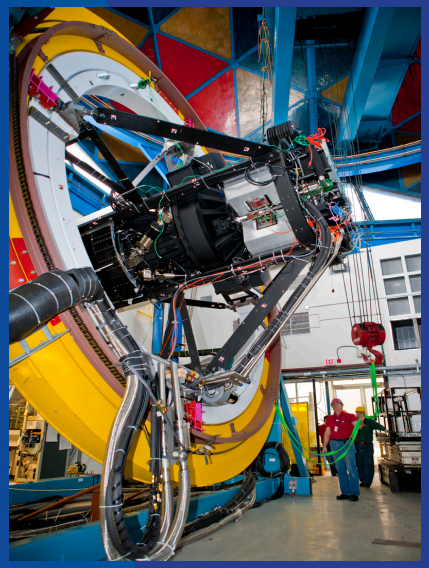
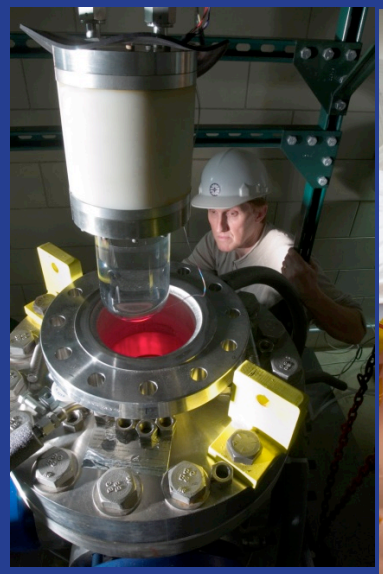
Major discoveries

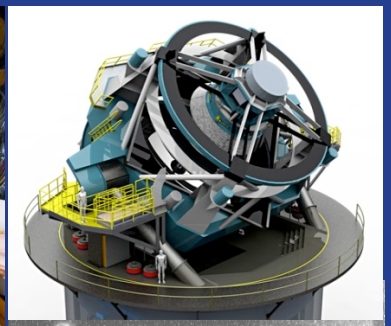
- Top quark
- B_s mixing
- Precision W and Top mass → Higgs mass prediction
- Direct Higgs searches
- Ruled out many exotica

The next generation

- Fantastic training ground for next generation
- More than 500 Ph.D.s
- Produced critical personnel for the next steps, especially LHC

Roles: cosmic frontier

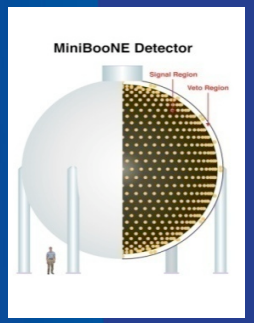



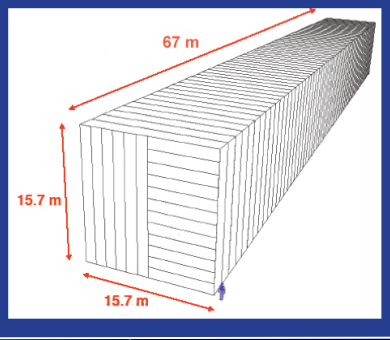


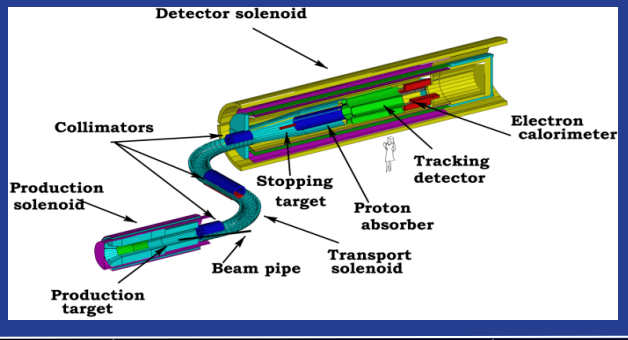

DM: ~10 kg DE: SDSS P. Auger	DM: ~100 kg DE: DES P. Auger Holometer?	DM: ~1 ton DE: LSST WFIRST?? BigBOSS??	DE: LSST WFIRST??
Now	2013	2016	2019	2022

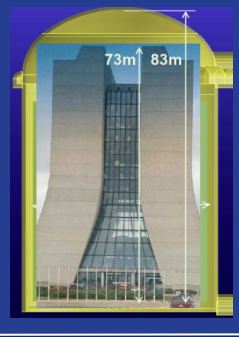
From P. Oddone, Secretary of Energy's Visit, June 2nd, 2011

Roles: intensity frontier

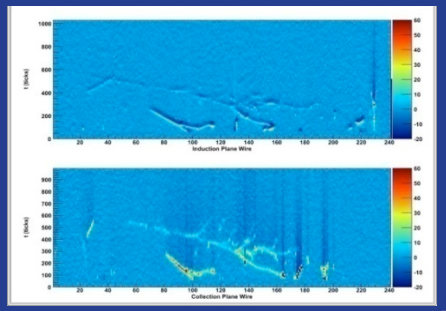


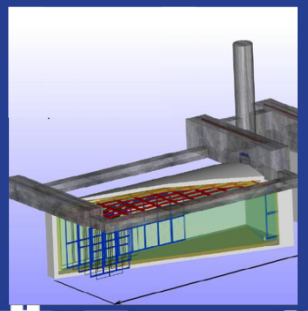






Now	2013	2016	2019	2022
MINOS MiniBooNE MINERvA SeaQuest	NOvA MicroBooNE g-2 MINERvA MINOS SeaQuest	NOvA g-2 LBNE Mu2e	Project X+LBNE μ, K, nuclear, … ν Factory ??	





From P. Oddone, Secretary of Energy's Visit, June 2nd, 2011

Work done to date on GO

- Integration of Workload Management and Data Movement Systems with GO
  1. Center for Enabling Distributed Petascale Science (CEDPS): GO integration with glideinWMS
  2. Data Handling prototype for Dark Energy Survey (DES)
- Performance tests of GO over 100 Gpbs networks
  3. GO on the Advanced Network Initiative (ANI) testbed
- Data Movement on OSG for end users
  4. Network for Earthquake Engineering Simulation (NEES)

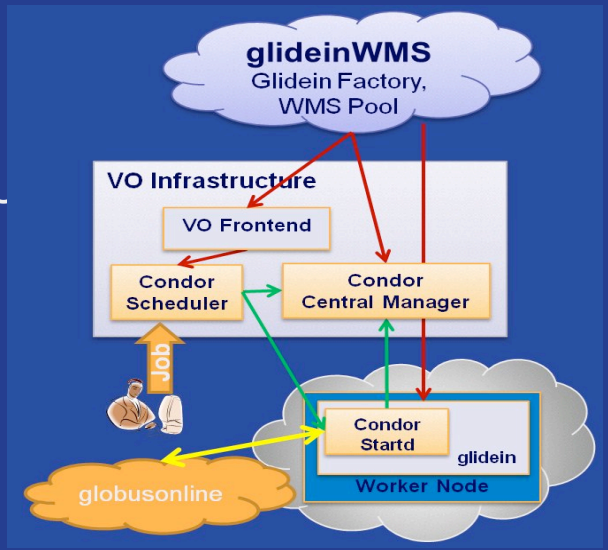
Communities potentially interested in follow ups with GO

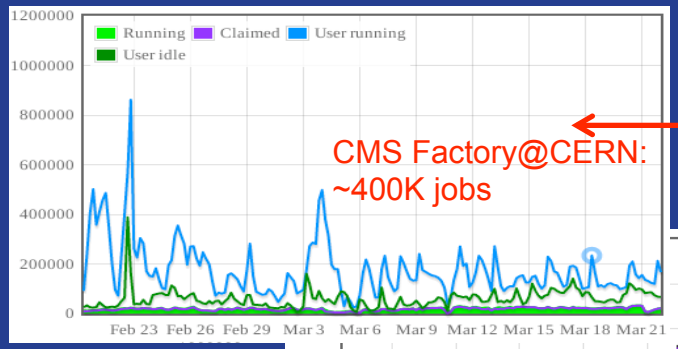
- Intensity Frontier, QCD, Accelerator modeling, Computational Cosmology, et al.

1. CEDPS: Integrating glideinWMS with GO

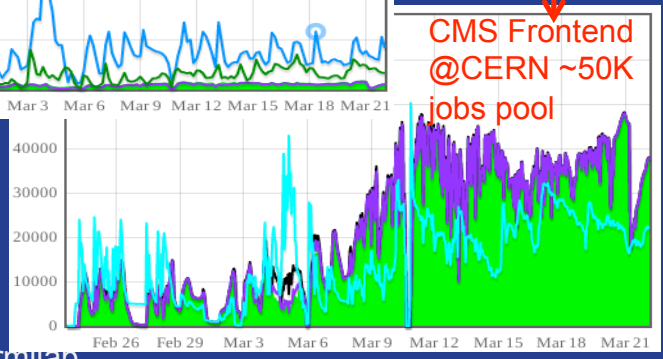
Parag Mhashilkar, Fermilab; Condor Team, U. Wisconsin Madison; GO team, ANL

- Goals:
  - Middleware handle data movement, rather than the application
  - Middleware optimize use of computing resources (CPU does not block on data movement)
- Users provide data movement directives in the Job Description File (e.g. storage services for IO)
- glideinWMS procures resources on the Grid and run jobs using Condor
- Data movement is delegated to the underlying Condor system
- globusconnect is instantiated and GO plug-in is invoked using the directives in the Job Description File
- Condor optimizes resources





Required Scale

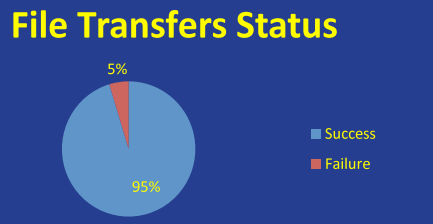


U.S. DEPARTMENT OF ENERGY

Validation Test Results

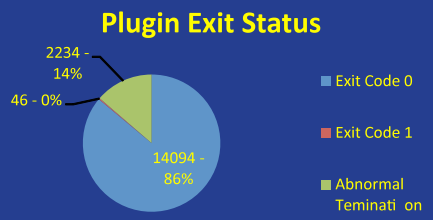
- Intensity Frontier (Minerva) jobs transfer output sandbox to GO endpoints
  - Jobs: 2636 (500 running at a time)
  - Total files transferred: 16359
- Up to 500 dynamically created GO endpoints at a given time.
- 95% transfer success rate.
- Stressing scalability of GO in new way
- Main scalability issue identified:
  Endpoint management not sufficiently responsive at this scale
- GO team working to increase scalability by
  - reusing GO endpoints and
  - transferring multiple files at once.

Transfer Result	Count
Successful - Plugin exit code 0 + scp (success)	15594
Failed	765
Total	16359



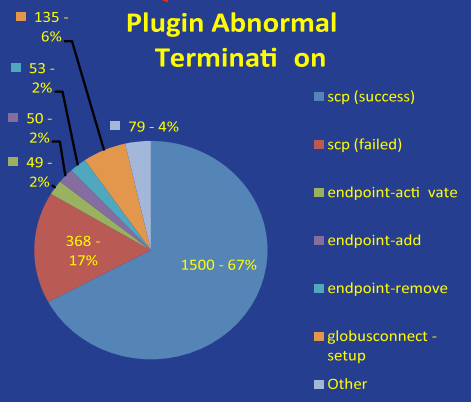
Plugin Exit Status	Count
Exit Code 0 (Success)	14094
Exit Code 1 (Failure)	46
Abnormal Terminati on	2234
Total	16374

Analysis based on exit code of the plugin in logs -
- Plugin terminati ons analyzed: 16374
- Duplicate terminati ons in same log fi le(ignored): 15
- Total plugin log fi les analyzed: 16359



Plugin Abnormal Terminati on	Count
globusconnect -setup	135
endpoint-remove	53
endpoint-add	50
endpoint-acti vate	49
scp (Transfers success)	1500
scp (Transfers failed)	368
endpoint-remove (cleanup)	0
Other	79
Total	2234

Analysis based on the last acti on tried by the plugin
- Status of the acti on never reported back to the plugin



2. Prototype integration of GO with DES Data Access Framework

See Don Petravick's talk on Wed

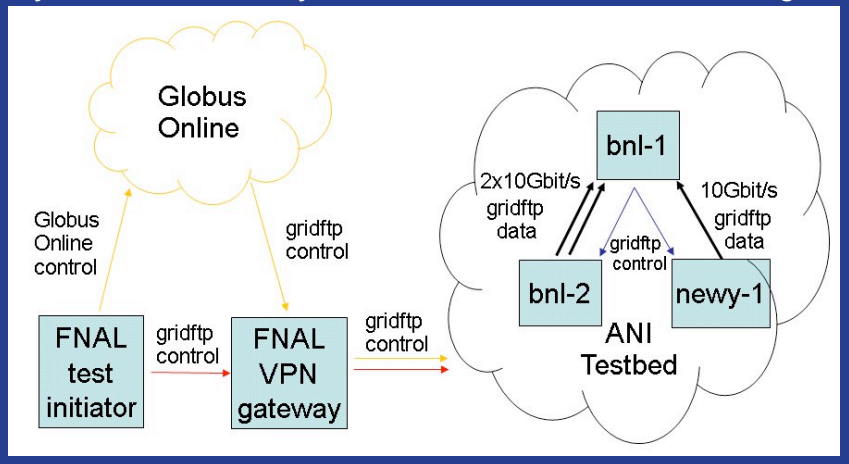
- Motivation
  - The Dark Energy Survey is an experiment on the cosmological frontier at Fermilab. Getting ready for data taking…
- DES Data Access Framework (DAF) uses a network of GridFTP servers to reliably move data across sites.
  1. DAF data transfer parameters were not optimal for both small and large files.
  2. Reliability was implemented inefficiently by sequentially verifying real file size with DB catalogue.

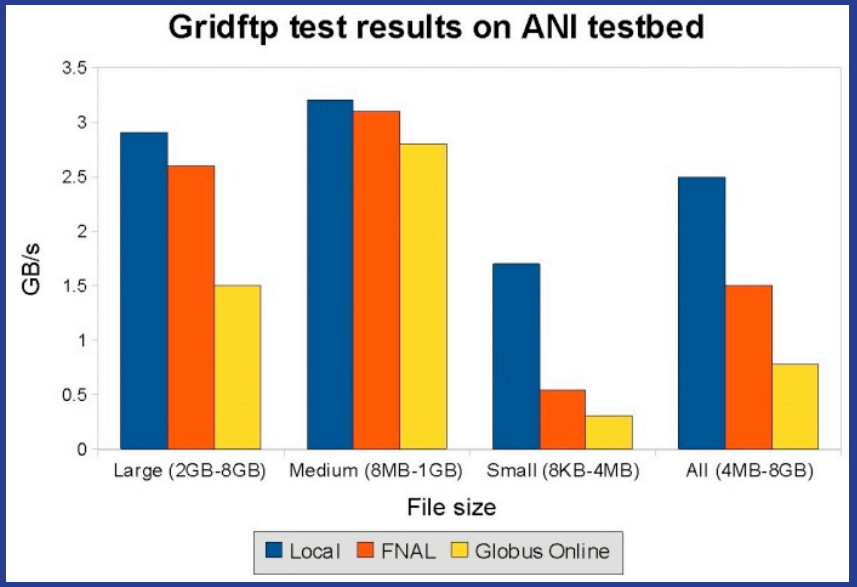
- Tested DAF moving 31,000 files (184 GB) with GO vs. UberFTP
  - Time for Transfer + Verification is the same (~100 min)
  - Transfer time is 27% faster with GO than with UberFTP
  - Verification time is 50% slower with GO than sequentially with UberFTP
- Proposed Improvements:
  - Allow specification of src / dest transfer reliability semantics (e.g. same size, same CRC, etc.) – *Implemented for size*
  - Allow finer-grain failure model (e.g. specify number of transfer retries)
  - Provide interface for efficient (pipelined) ls of src / dest files.

3. GO on the ANI Testbed

Work by Dave Dykstra w/ contrib. by Raman Verma & Gabriele Garzoglio

- Motivation:
  Testing Grid middleware readiness to interface 100 Gbits links on the Advanced Network Initiative (ANI) Testbed.
- Characteristics:
  - 300GB of data split into 42432 files (8KB – 8GB) (small, medium, large, all sizes)
  - Network: aggregate 3 x 10Gbit/s to bnl-1 test machine
- Results
  - GO (yellow) does almost as well as direct GridFTP practical max (red) for medium-size files.
  - Increasing concurrency and pipelining on small files improves throughput by 30%.
  - GO auto-tuning works better for medium files than for the large files



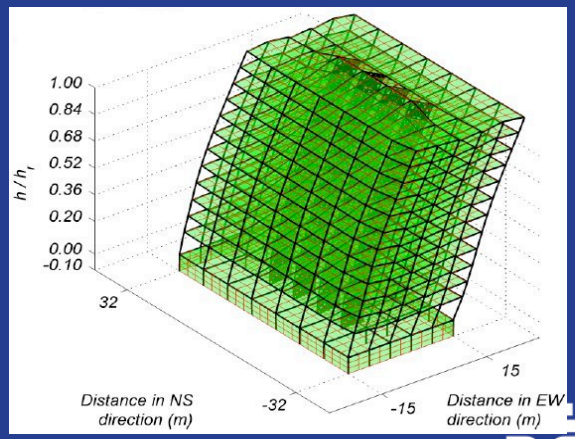


4. Data Movement on OSG for NEES

A. R. Barbosa, J. P. Conte, J. I. Restrepo, UCSD

- Motivation
  - supporting NEES group at UCSD to run parametric studies of nonlinear models of structure systems on the Open Science Grid
- NEES scientist ran at 20 OSG sites, then moved 12 TB from the RENCI submission server to the user's desktop at UCSD using GO
- Note: there is still no substitute for a good network administrator
  - Initially, we had 5 Mbps → eventually 200 Mbps (over 600 Mbps link).
  - Improvements: Upgrade eth card on user desktop; Migrate from Windows to Linux; Work with the user to use GO; Find a good net admin to find and fix broken fiber at RENCI, when nothing else worked.
- **Better use of GO on OSG: Integrate GO with the Storage Resource Broker (SRM)**

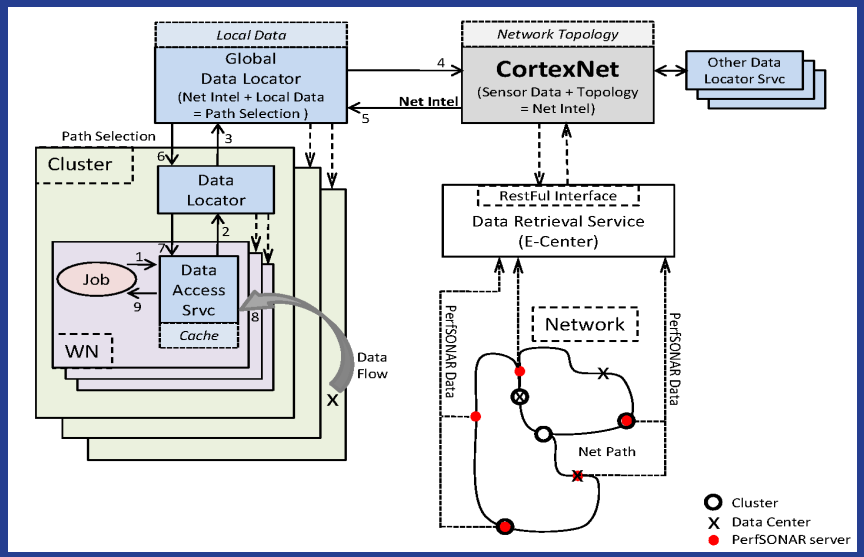
Number of NLTH analyses per parameter set realization	180
Average duration of NLTH analysis	12 hours
Average size of output data	1.5 GB
Parameters considered	6
Perturbations considered	4
Estimated clock time (180x12x[(6x4x2)+1])	**106,800 hours (12.2 years)**
Estimated output data (180x1.5x[(6x4x2)+1])	**12 TB**



30 days on OSG vs. 12 yrs on Desktop

Sample future work and ideas (1)

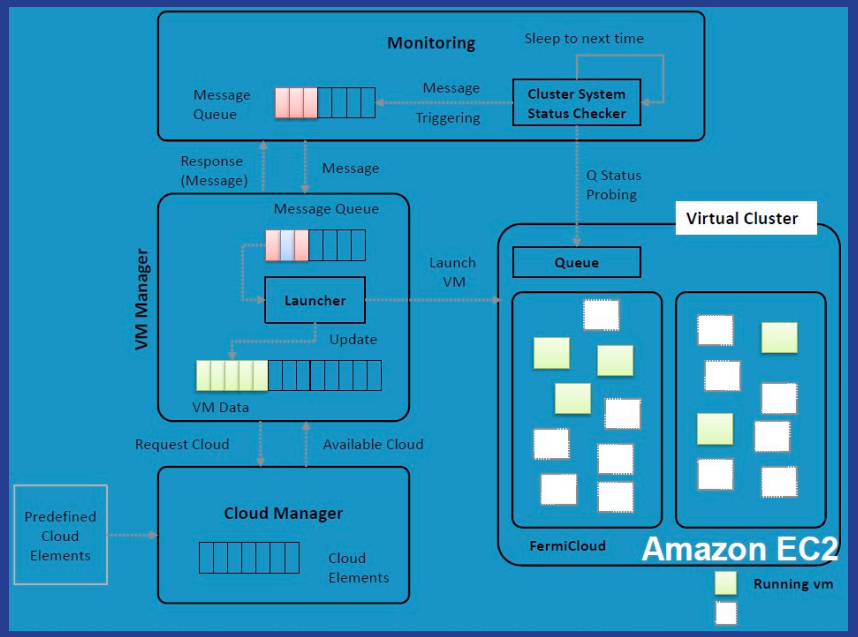
- **CortexNet**: Integration of network intelligence with data management middleware
- How do I select the highest-throughput source of a data replica considering the network conditions ?
- How do I integrate GO with CortexNet?
- There are interesting algorithmic challenges: if you are a network expert with background on AI or statistics, let's talk!



Sample future work and ideas (2)

Work by Seo-Young Noh and the FermiCloud team

- **VCluster**: Transparent access to dynamically instantiated resources through standard Grid interfaces
- Collaborating with KISTI, South Korea, to dynamically instantiate OSG-compatible Grid Clusters from Cloud resources



- Focusing now on idle VM detection and monitoring to automate Grid cluster expansion
- Can I use GO to distribute the VM images to the collaborating Cloud resources?

Summary

- Fermilab developers and computer scientists support distributed large-data intensive computing for HEP and other scientific communities via OSG.
- We collaborate closely with a wide variety of external research and development groups.
- We have worked with the GO team and improved the system:
  - Stressed the “many-globusconnect” dimension.
  - New requirements on reliability semantics.
  - Auto-tuning at extreme scale.
  - Usability.
- Fermilab science has many needs and ideas that are future opportunities for collaboration.

References

1. CEDPS Report: GO Stress Test Analysis
   - https://cd-docdb.fnal.gov:440/cgi-bin/RetrieveFile?docid=4474;filename=GlobusOnline%20PluginAnalysisReport.pdf;version=1
2. DES DAF Integration with GO
   - https://www.opensciencegrid.org/bin/view/Engagement/DESIntegrationWithGlobusonline
3. GridFTP & GO on the ANI Testbed
   - https://docs.google.com/document/d/1tFBg7QVVFu8AkUt5ico01vXcFsgyIGZH5pqbbGeI7t8/edit?hl=en_US&pli=1
4. OSG User Support of NEES
   - https://www.opensciencegrid.org/bin/view/Engagement/EngageOpenSeesProductionDemo