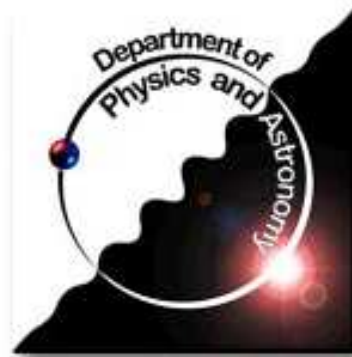


# A Canadian Computational Grid for Particle Physics Applications

Daniel Vanderster

`dvanders@engr.uvic.ca`

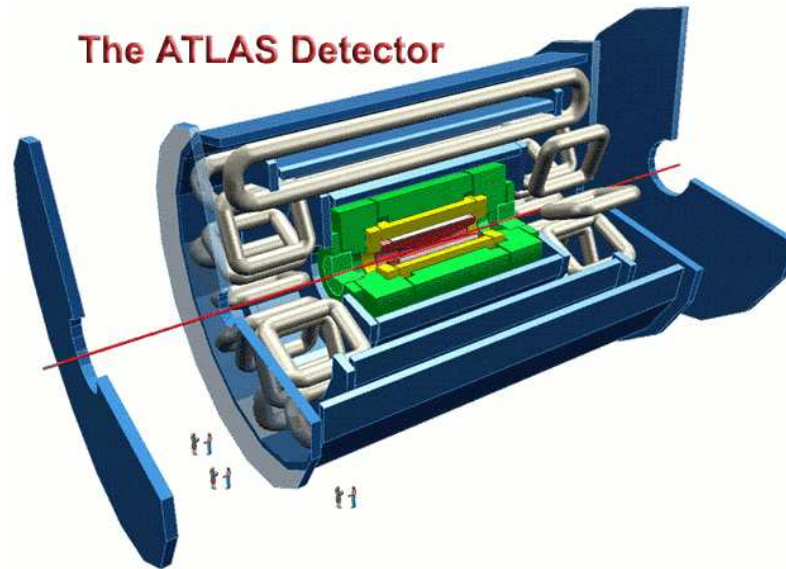
Supervisors: R. J. Sobie and N. J. Dimopoulos



# Outline

- LHC: A New Scale of Computing
- Introduction to Grid Computing
- Globus: Grid Middleware
- Engineering the Canadian Computational Grid
- Using the Canadian Computational Grid
- The Future of Grid Computing

# A New Scale of Computing



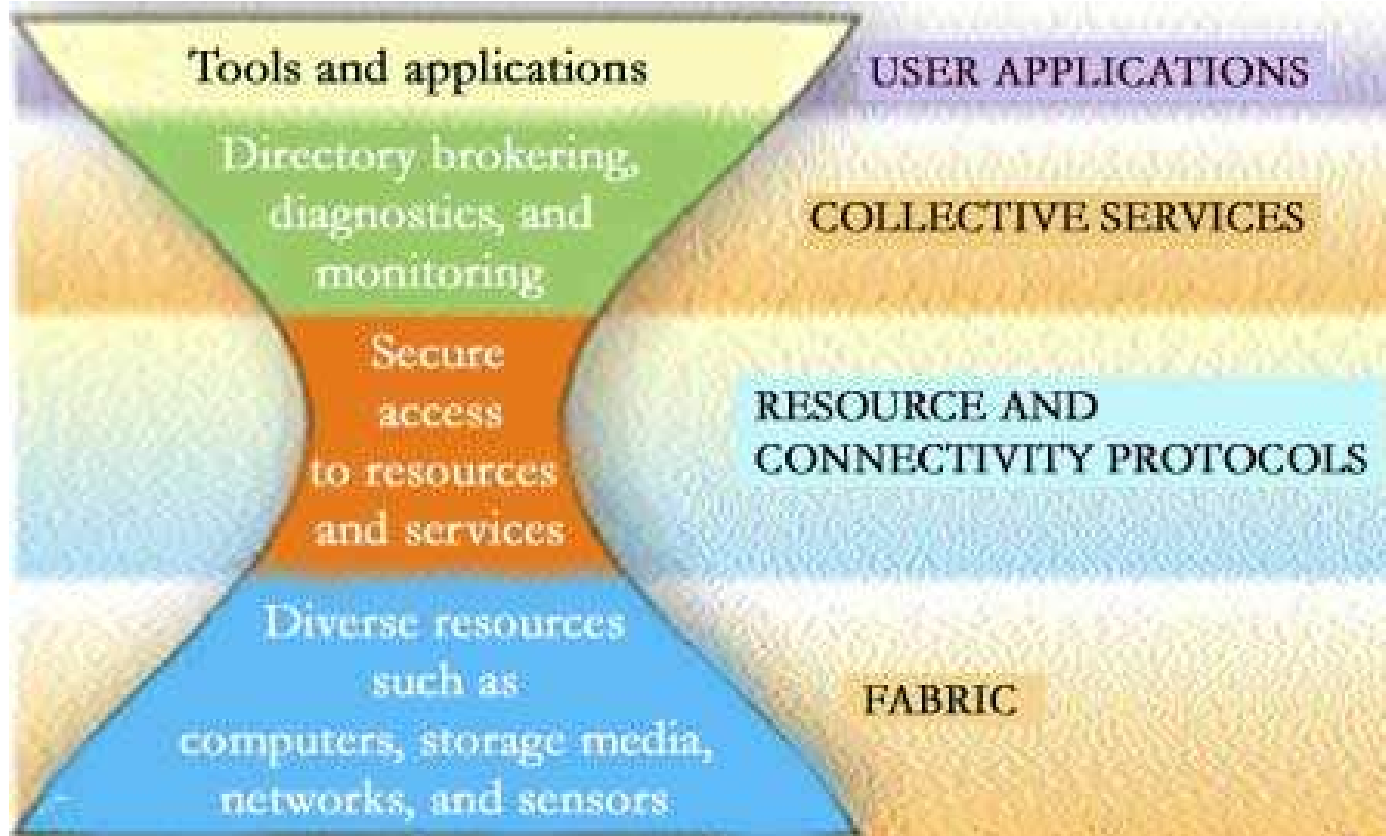
- ATLAS will begin recording in 2007
- $10^9$  events per year \* 1 MB each = 1 PB per year
- $10^5$  commodity processors are needed to process LHC data
- Must be accessible to  $\sim 2000$  users at  $\sim 150$  institutes

# Computational Grids are the Solution



- We must combine the resources of the Institutes world-wide.
- A Grid allows for transparent and secure access to remote storage, compute, and other resources.

# Introduction to Globus



- The Globus Toolkit is the de facto standard Grid middleware.

# Introduction to Globus

- Provides the infrastructure for combining compute resources across administrative domains.
  - Security:
    - Provides single-sign-on access to all available resources.
  - Job Management:
    - Globus provides job submission and monitoring
  - Data Management:
    - GridFTP utilizes security mechanism to allowing data transfers across the Grid.
  - Info Management:
    - Info Service stores a directory of information about a Grid: resource capabilities and states.

# Introduction to Grid Canada



- UVic: >140, UAlberta: >110, NRC: >50 CPUs
- Combined storage capacity = >20 TB

# Grid Canada Job Submission

```
[adimopou@imp atlas]$ ./gcsb2.pl -v -l /bin/echo -a "job %1" -p 1-10/1
Running parameter sweep...
Processes to be run:
    /bin/echo job 1
    /bin/echo job 2
    /bin/echo job 3
    /bin/echo job 4
    /bin/echo job 5
    /bin/echo job 6
    /bin/echo job 7
    /bin/echo job 8
    /bin/echo job 9
    /bin/echo job 10
```

- gcsb script for job submission
  1. Constructs parametric-sweep task list.
  2. Query information service for available resources.
  3. Filter resources for requirements and authorization.
  4. Submits tasks to available resources.



# Grid Canada Job Monitoring

Status of Globus Job Queue

User	Filename	Arguments	Staged	Time Submitted	Status	Time Started	Time Finished	Resource
Alexandros Dimopoulos	/bin/echo	job 10	no	2004-02-10 09:31:54	queued			
Alexandros Dimopoulos	/bin/echo	job 9	no	2004-02-10 09:31:54	pending	2004-02-10 09:32:52		mercury.uvic.ca
Alexandros Dimopoulos	/bin/echo	job 8	no	2004-02-10 09:31:54	pending	2004-02-10 09:32:49		mercury.uvic.ca

- GC website shows the state of the GC job queue.
- MySQL DB keeps track of jobs running on the GC Grid.
- Jobs have states  
*Queued, Pending, Active, Failed, Done*

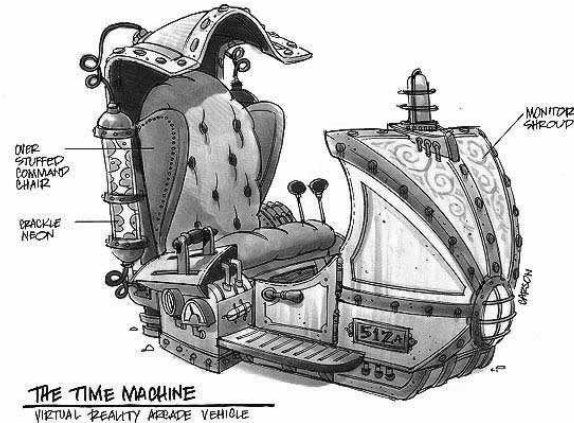
# Grid Canada Data Management

- Goal is to store and retrieve vast amounts of data.
- Data catalog stores the locations of data.
- GridFTP is used to transfer data from point to point.
- Engineering the network to optimize bulk data transfers; dynamic bandwidth.
  - Network lightpaths are a Grid resource.



# Grid Computing in the Future

- Grid solutions are in flux now: by 2007 we will likely be using a new system.
- Where can we improve now?
  - Job management:
    - error recovery, optimized job scheduling
  - Monitoring:
    - convergence of web-based services with Grid services
  - Data management:
    - Locate job to the data, or move the data to the job?



# The end.

- Websites for more information:
  - Grid Canada: <http://www.gridcanada.ca/>
  - UVic Grid: <http://grid.phys.uvic.ca/>
  - GC Production Grid: <http://grid.phys.uvic.ca/gcprod/>
- Any questions or comments?

