# A Grid of Grids using Condor-G

R. Walker[1], M. Vetterli[1,2], A. Agarwal, D. Vanderster[3],
R.J. Sobie[3,4], M. Grønager[5]

[1]Simon Fraser University, [2]TRIUMF, [3]University of Victoria,
[4]Institute of Particle Physics of Canada, [5]Uni-C

Mumbai, Febuary 2006

# Outline

1. Introduction
2. Condor and CondorG
   - Canadian GridX1: A CondorG Grid
3. LCG: A CondorG Grid
   - Federating Grids
4. ATLAS Production System
5. Conclusions and Future Work

## Outline

1. Introduction
2. Condor and CondorG
   - Canadian GridX1: A CondorG Grid
3. LCG: A CondorG Grid
   - Federating Grids
4. ATLAS Production System
5. Conclusions and Future Work

# Introduction

- Workload Management System(WMS)
  - Everything between the user and a WN
  - Submit, match to ’best’ resource, run, retrieve output
  - Example is EDG/GLite WMS characterized by the Resource Broker
  - ARC system in NorduGrid
- There is another way ...

# ATLAS preparations for 2007

- Test the Computing Model and stress the systems
    - a series of Data Challenges(DC) increasing in scale
    - Monte Carlo production and data consolidation
    - on 3 Grids: LCG, NorduGrid, Grid3
- DC2 production exposed scaling issues
    - LCG resources could not be fully filled - low WMS submission rate
    - much manpower required to operate

# Canadian Concerns: What are the problems?

- Several large shared facilities
    - cannot install LCG software: manpower, intrusion
    - must be used in ATLAS data challenges
- Local physicists not using available Grid resources
    - ease of use and transparent access to all resources
    - LCG WMS very awkward and not performant
- LCG submission rate: LCG WMS is the problem - new approach.

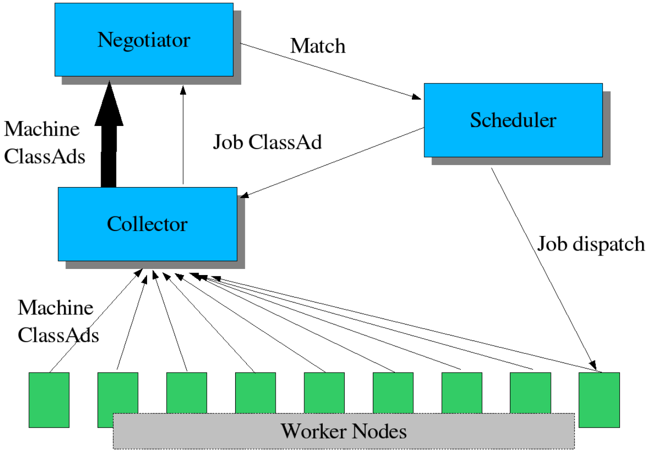
# Outline

1. Introduction
2. Condor and CondorG
   - Canadian GridX1: A CondorG Grid
3. LCG: A CondorG Grid
   - Federating Grids
4. ATLAS Production System
5. Conclusions and Future Work

# Condor Batch System Interlude

- Grid is often described as a “big batch system” so let’s look at a small one. In order to schedule jobs ...
    - need to know things about the batch nodes, e.g.OS, RAM, status
    - need to know what the job requires and prefers
- Condor represents both these as ClassAds(Classified Ads)
- The *Collector* gathers machine and job ClassAds
- The *Negotiator* matches jobs to machines
- The *Scheduler* then sends the job to the matched machine

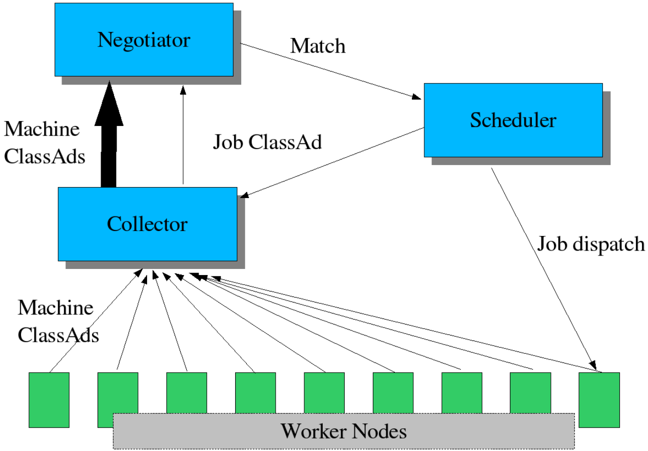
# Condor Batch System



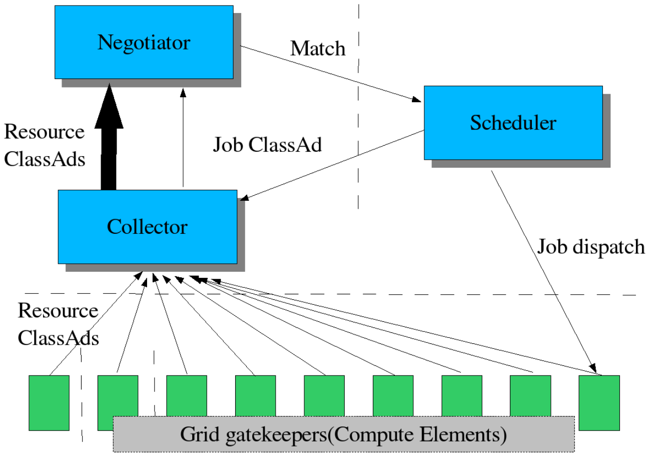
## CondorG Overview

- How does this apply to Grids?
- CondorG is an extension of the Condor batch system to the grid world
- Gatekeepers to remote clusters are the ’batch machines’
- The actual batch machines controlled by a normal batch system - LSF,PBS,..

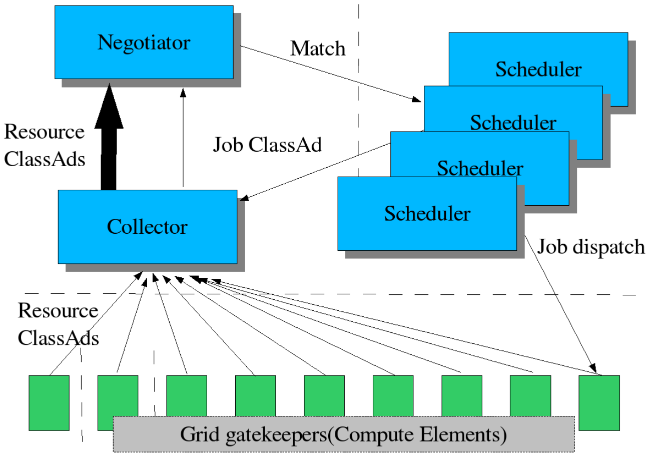
# Condor Batch System: Reminder



# CondorG Architecture



# CondorG: multiple schedulers



# GridX1 Overview

- Currently have 4 clusters: UVic, UAlberta, NRC, and WestGrid with 2000 cpus
- Shared facilities - no manpower to install LCG middleware
- They have gatekeepers so CondorG can form Grid
  - ClassAd is produced by probing Batch System
  - pushed to TRIUMF Collector
- Used heavily during dc2/rome via an interface from LCG WMS.

Edmonton
Victoria
Ottawa
● Up
● Down

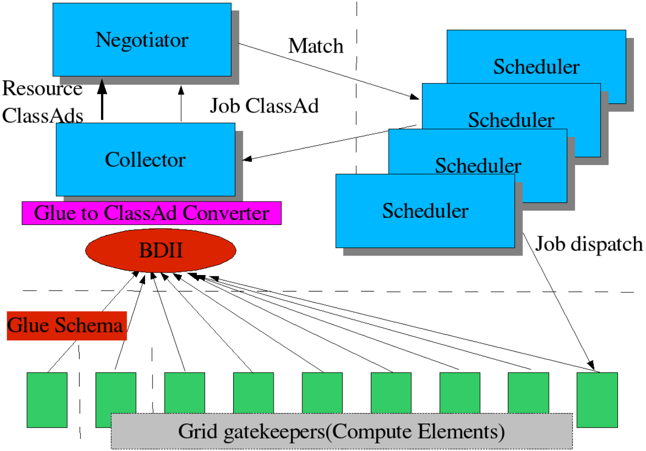
# Outline

1. Introduction
2. Condor and CondorG
   - Canadian GridX1: A CondorG Grid
3. LCG: A CondorG Grid
   - Federating Grids
4. ATLAS Production System
5. Conclusions and Future Work

# LCG: A CondorG Grid

- Success of GridX1 - try this on a larger scale
- LCG has 100+ sites and 10,000 cpus
- Need a ClassAd for each LCG CE
  - can’t create and push it from the sites without help
  - LCG has a central information service(BDII)
  - convert this info into 1000+ ClassAds, one per queue

# LCG: A CondorG Grid



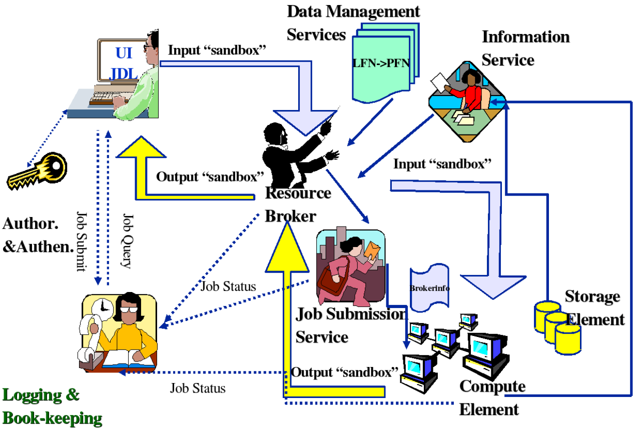
# Matchmaking

- Requirements - job and resource must be true
  - job requires:- OS, 512MB RAM, 24hr walltime
  - resource requires:- no job starts 08:00-17:00
- Rank - for job-resource pairs passing Requirements
  - job prefers:- few queued jobs, Canada
- Expressions mostly formed of simple attributes from job or resource ClassAd
  - can have arbitrary functions depending on external information
  - example is data co-location where function queries replica catalogue - fold in bandwidth from closest replica to CE
  - dynamic info - CurMatches per CE increments for each match between info updates
    - (CurMatches+gluecewaitingjobs) used in Rank/Requirements

# Condor Development

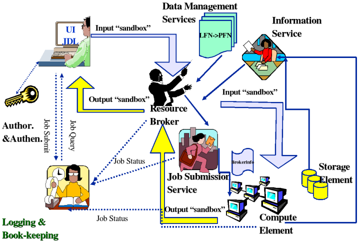
- This hadn't been tried before
  - single Scheduler easily scaled up to 2000 running jobs
  - scheduler blocked by status queries
  - requirements evaluation for 1000 queues slow (1s)
- Very good contact with Condor team in Wisconsin
  - 30+ computer scientists
  - Quill is DB frontend to Scheduler - no blocks.
  - implement short cut in Requirements logic test
- Use off-the-shelf technology and fraction of an FTE
  - reproduce LCG WMS functionality
  - scalable architecture, pseudo-dynamic info(CurMatches), flexible external matchmaking functions
  - increased performance and usability
- If LCG doesn't want it, ATLAS and Canada do

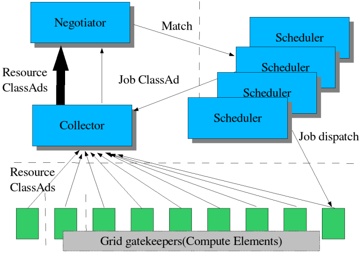
# EDG Workload Management System



# EDG vs CondorG

- EDG RB is Negotiator, Collector and Scheduler
- CondorG scales with the number of Schedulers





# EDG vs CondorG

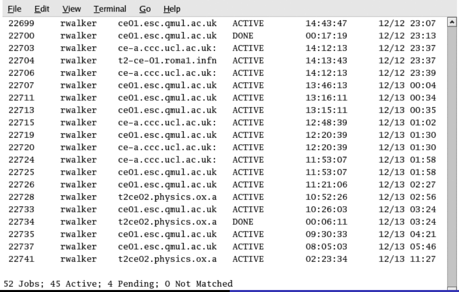
- Criticism: CondorG has no central logging and bookkeeping service
  - logging at the Scheduler level in postgres Db
  - multiple RB's have no central service either

# EDG vs CondorG

- Criticism: CondorG has no central logging and bookkeeping service
    - logging at the Scheduler level in postgres Db
    - multiple RB's have no central service either
- CondorG provides batch system like commands to *submit*, *monitor*, and *cancel* jobs with instant response

# EDG vs CondorG

- Criticism: CondorG has no central logging and bookkeeping service
  - logging at the Scheduler level in postgres Db
  - multiple RB's have no central service either
- CondorG provides batch system like commands to *submit*, *monitor*, and *cancel* jobs with instant response



```
22699   rwalker   ce01.esc.qmul.ac.uk  ACTIVE      14:43:47    12/12 23:07
22700   rwalker   ce01.esc.qmul.ac.uk  DONE        00:17:19    12/12 23:13
22703   rwalker   ce-a.ccc.ucl.ac.uk:  ACTIVE      14:12:13    12/12 23:37
22704   rwalker   t2-ce-01.roma1.infn  ACTIVE      14:13:43    12/12 23:37
22706   rwalker   ce-a.ccc.ucl.ac.uk:  ACTIVE      14:12:13    12/12 23:39
22707   rwalker   ce01.esc.qmul.ac.uk  ACTIVE      13:46:13    12/13 00:04
22711   rwalker   ce01.esc.qmul.ac.uk  ACTIVE      13:16:11    12/13 00:34
22713   rwalker   ce01.esc.qmul.ac.uk  ACTIVE      13:15:11    12/13 00:35
22715   rwalker   ce-a.ccc.ucl.ac.uk:  ACTIVE      12:48:39    12/13 01:02
22719   rwalker   ce01.esc.qmul.ac.uk  ACTIVE      12:20:39    12/13 01:30
22720   rwalker   ce-a.ccc.ucl.ac.uk:  ACTIVE      12:20:39    12/13 01:30
22724   rwalker   ce-a.ccc.ucl.ac.uk:  ACTIVE      11:53:07    12/13 01:58
22725   rwalker   ce01.esc.qmul.ac.uk  ACTIVE      11:53:07    12/13 01:58
22726   rwalker   ce01.esc.qmul.ac.uk  ACTIVE      11:21:06    12/13 02:27
22728   rwalker   t2ce02.physics.ox.a  ACTIVE      10:52:26    12/13 02:56
22733   rwalker   ce01.esc.qmul.ac.uk  ACTIVE      10:26:03    12/13 03:24
22734   rwalker   t2ce02.physics.ox.a  DONE        00:06:11    12/13 03:24
22735   rwalker   ce01.esc.qmul.ac.uk  ACTIVE      09:30:33    12/13 04:21
22737   rwalker   ce01.esc.qmul.ac.uk  ACTIVE      08:05:03    12/13 05:46
22741   rwalker   t2ce02.physics.ox.a  ACTIVE      02:23:34    12/13 11:27

52 Jobs; 45 Active; 4 Pending; 0 Not Matched
```

# Extend LCG CondorG Grid

- CondorG just needs 'standard' ClassAd plus gatekeeper
- CondorG supports several gatekeeper types inc. GT2 & NorduGrid (and of course Condor-C - glite CE)
- NorduGrid info converted to Glue in BDII, and LCG conversion script produces ClassAds
- Scalable interoperability. Transparent to the user.
- Making progress ... already ran first jobs to match across LCG, NG and GridX1

# Outline

1. Introduction
2. Condor and CondorG
   - Canadian GridX1: A CondorG Grid
3. LCG: A CondorG Grid
   - Federating Grids
4. ATLAS Production System
5. Conclusions and Future Work

# ATLAS Production System

- Distributed simulation and data reprocessing
  - central Db containing job definitions
  - executors for each Grid take jobs, run, update status
  - main problem was sluggish submission rate to LCG resources
- Enter CondorG: immediately double production rate
  - central CondorG services at TRIUMF
  - single local Scheduler at UVic or TRIUMF
    - tiny latency on job submission 0.1s cf 15s for LCG
    - status request also fast
  - single instance and operator slashed manpower
    - LCG had 4 operators and 4 RB's

# Outline

1. Introduction
2. Condor and CondorG
   - Canadian GridX1: A CondorG Grid
3. LCG: A CondorG Grid
   - Federating Grids
4. ATLAS Production System
5. Conclusions and Future Work

# Conclusions

- Connected Canadian resources with WMS
- Non-LCG shared resources were used in DC2
- Same technology applied to LCG resources
    - scales, flexible, nicer for users
- CondorG use arose from practical need
    - same functionality as LCG, also for users, batch system-like
    - outperforms LCG WMS - since Nov'05 85000 cf 20000

# Future Work

- Develop recipes to enable user analysis via CondorG
    - a few power users already exist
- Users limited by middleware usability/perfomance
    - taking that away would lead to grid carnage
    - scheduling and fair-share are important and under-developed

# Future Work

- Develop recipes to enable user analysis via CondorG
  - a few power users already exist
- Users limited by middleware usability/perfomance
  - taking that away would lead to grid carnage
  - scheduling and fair-share are important and under-developed
- Acknowledgements
  - SAMGrid team at FNAL pushed original concept, 2002.
  - LCG information service, CE's and deployment expertise are crucial

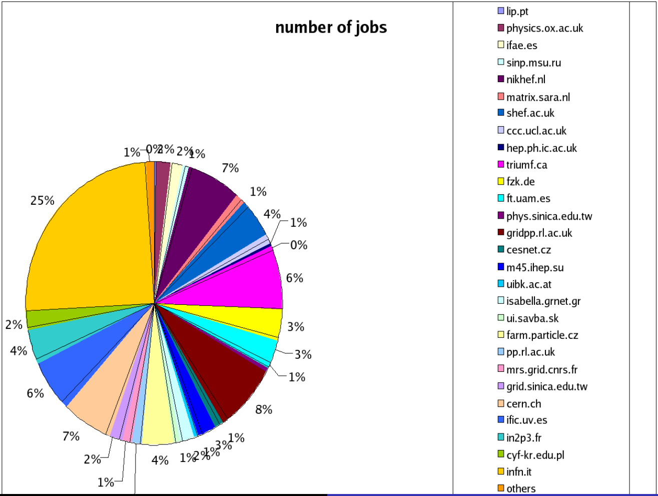
# Future Work

- Develop recipes to enable user analysis via CondorG
    - a few power users already exist
- Users limited by middleware usability/perfomance
    - taking that away would lead to grid carnage
    - scheduling and fair-share are important and under-developed
- Acknowledgements
    - SAMGrid team at FNAL pushed original concept, 2002.
    - LCG information service, CE's and deployment expertise are crucial
- Deployment team of «1! LCG deployment of CondorG alongside gLite WMS would benefit all

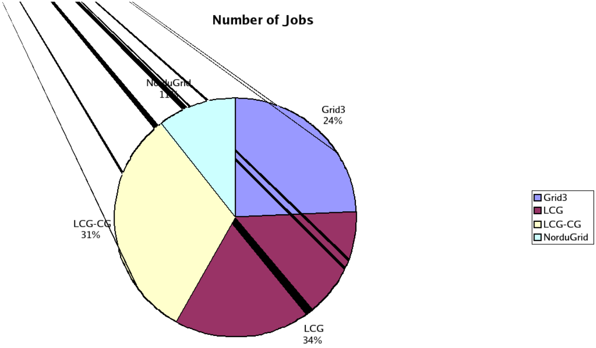
# Back-up Slides

- Jobs per site
- Jobs per grid
- Jobs per grid per day
- Authorization for 2nd GRAM Submission

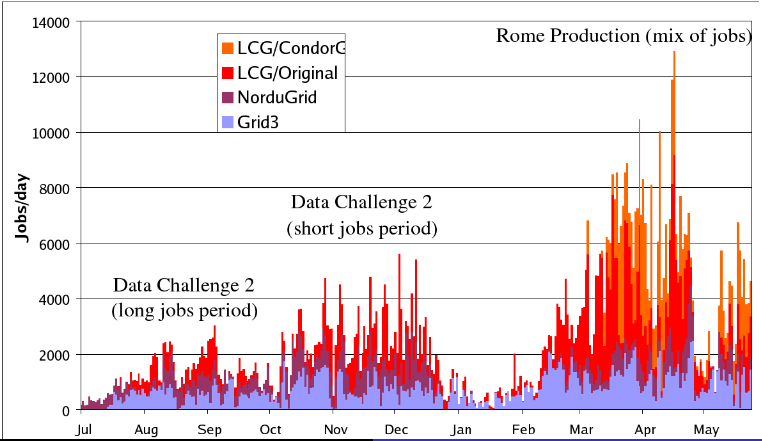
# Jobs per site in DC2



# Jobs per grid in DC2



# Jobs per grid per day



# Authorization for 2nd GRAM Submission

- Having a 2nd GRAM submission creates a proxy issue
- GRAM submission from the LCG RB delegates a *limited* proxy
  - This proxy can be used for GridFTP, but not a further GRAM submission
- We need to acquire a *full* proxy for the 2nd submission

- We could delegate a full proxy via GRAM, but we have chosen a different solution
- For the ATLAS application: user must store her credentials in a known MyProxy server
- The limited proxy is used to delegate a full proxy via MyProxy