Campus Compute Co-operative (CCC): A service Oriented Cloud Federation

Authors
Andrew Grimshaw (UVA)
Md Anindya Prodhan (UVA)
Alexander Thomas (UVA)
Craig Stewart (IU)
Richard Knepper (IU)

Agenda

- Motivation
- What is CCC
- CCC system model
- Using the CCC
- Social, political and market aspects
- Related Work
- Final Remarks

Motivation

- The need for cyberinfrastructure (CI) is now ubiquitous and not all needs are the same
- It is not feasible to buy everything that the researchers need
- One solution is sharing
  - Sharing often leads to the tragedy of the commons
  - Hence ***trading***

Why CCC ?

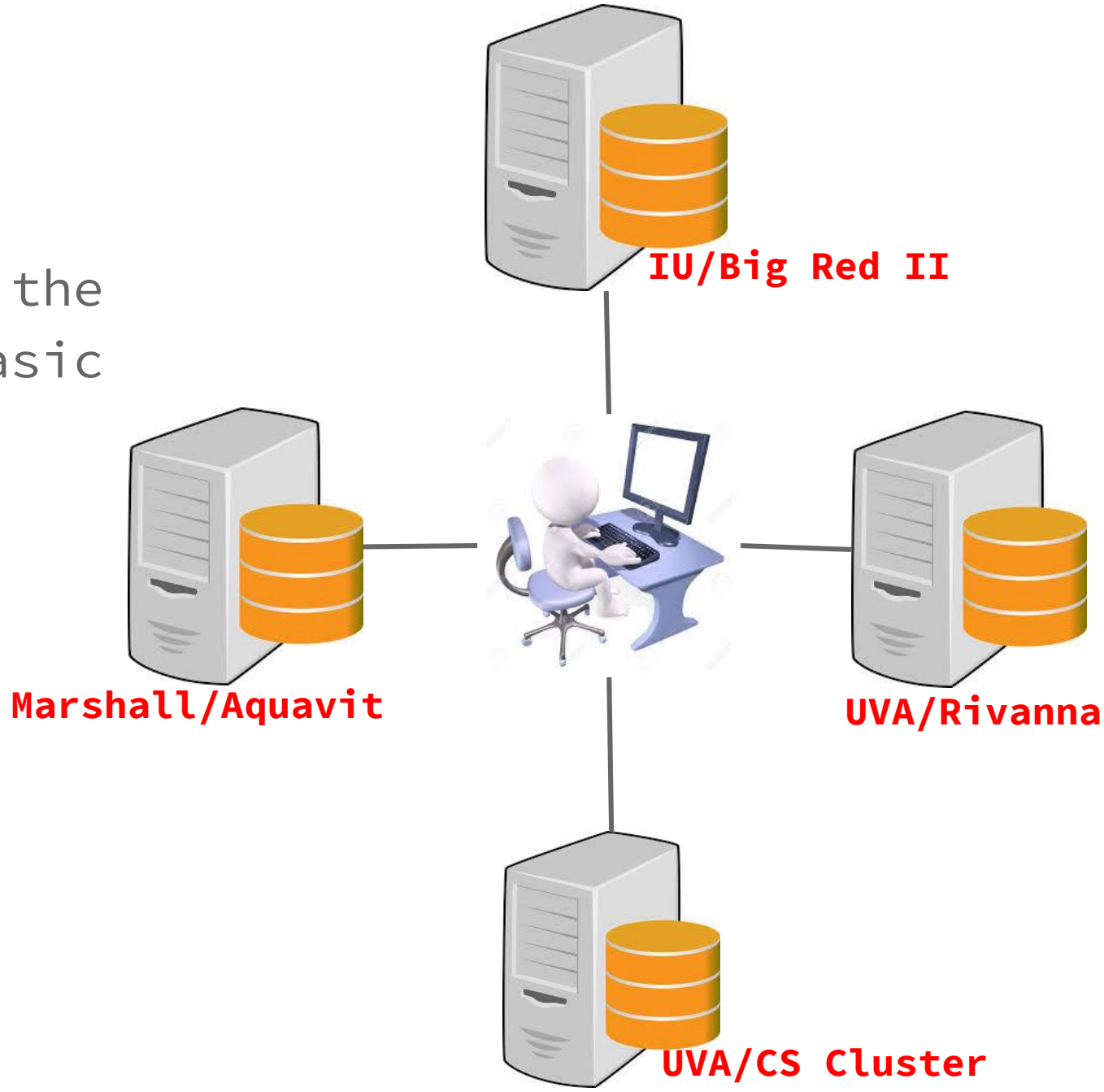
Use-cases

- urgent jobs
- Save money by being flexible
- Burst capacity
- Exchange of computational resources



What is CCC

- CCC is a pilot project in the US which combines three basic ideas into a production compute environment
  - Resource Market
  - Differentiated QoS
  - Resource Federation



What does CCC Provide

- Diversity of resources
- More resources are available to researchers ***when they need them***
- Important jobs are scheduled immediately
- Projects with less funding still have access to resources
- Fair and transparent job priority
- Familiar and easy to use paradigm
- Cloud bursting capability
- Data sharing

Current Status

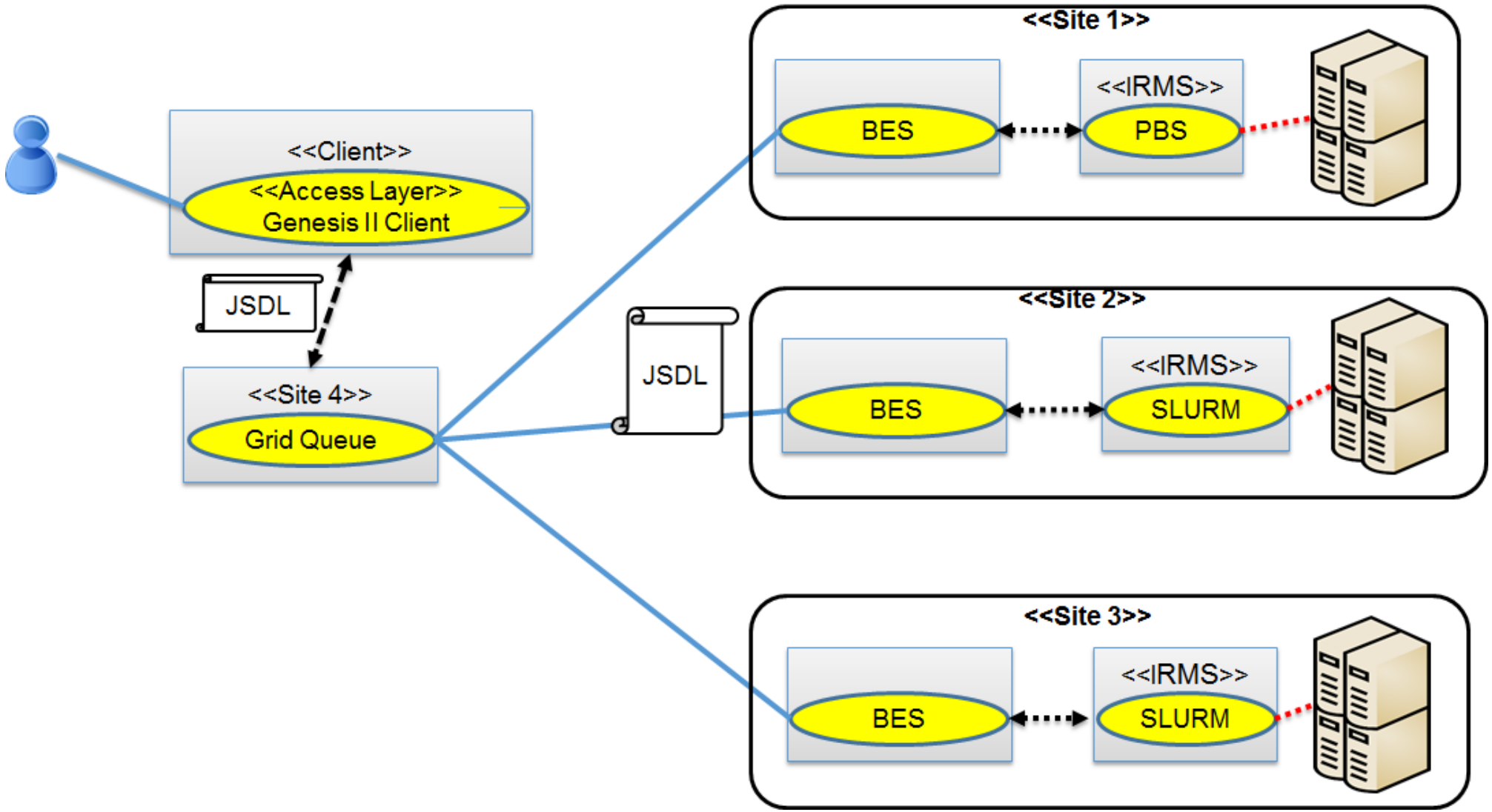
- CCC is up and running
- IU and UVA are already on-board with some of their major computing resources
  - Big-Red II (IU)
  - Rivanna (UVA)
- Marshall University is also joining the co-operative soon.

CCC System Model

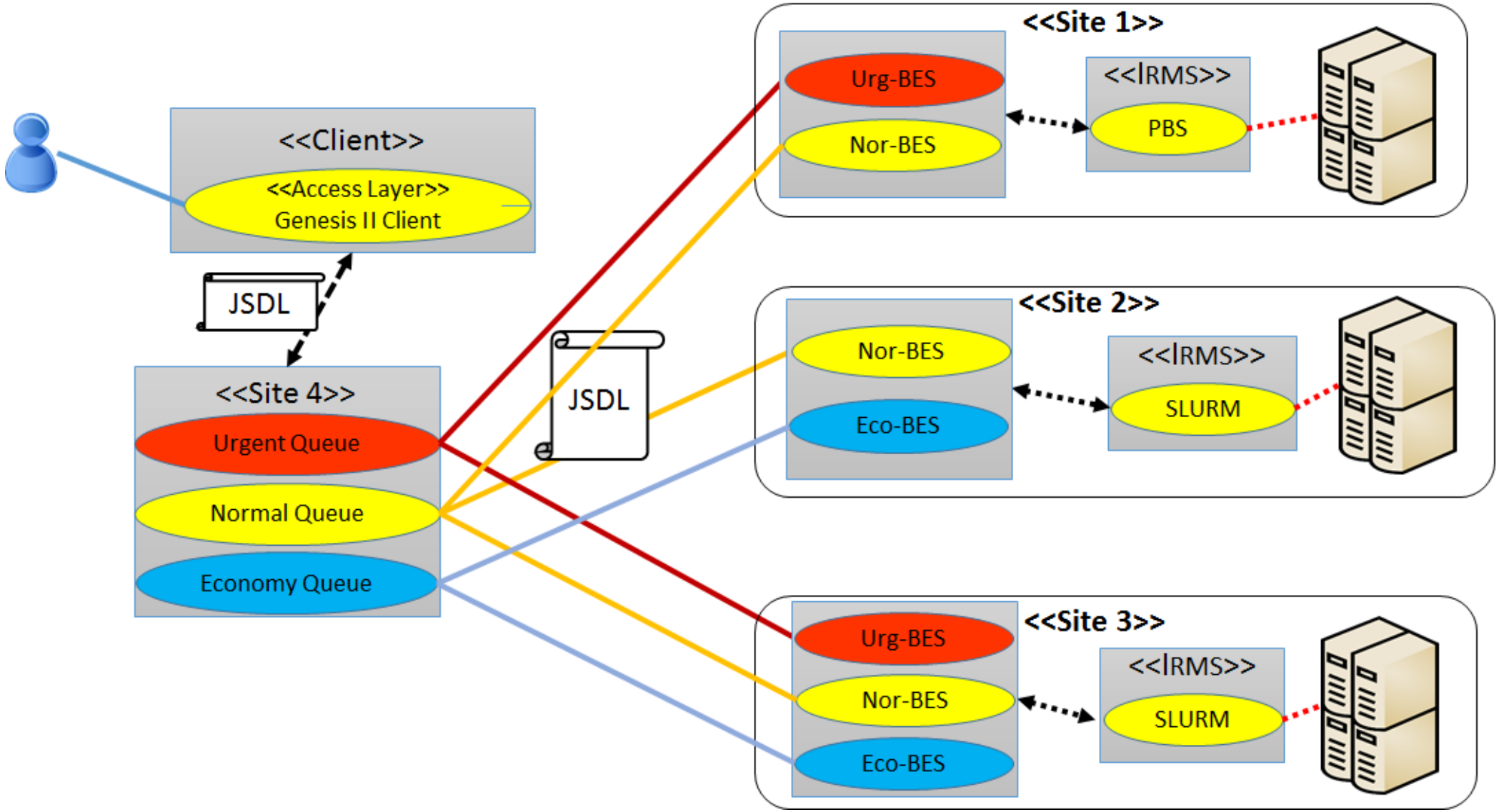
CCC System Model

- Build on Genesis II and XSEDE EMS (Execution Management Services)
- Differentiated QoS
  - Run Immediately (high priority)
  - Long Uninterrupted Run (Medium Priority)
  - Best effort (Low Priority)
- Target Jobs
  - Long Sequential Jobs
  - High-Throughput Computing Jobs (HTC) / Parameter Sweep Jobs
  - Parallel / MPI Jobs
  - GPU Jobs
- Resource Accounting

XSEDE EMS


<<Client>>
<<Access Layer>>
Genesis II Client
JSDL
<<Site 4>>
Grid Queue
JSDL
<<Site 1>>
BES
<<IRMS>>
PBS
<<Site 2>>
BES
<<IRMS>>
SLURM
<<Site 3>>
BES
<<IRMS>>
SLURM

CCC Architecture



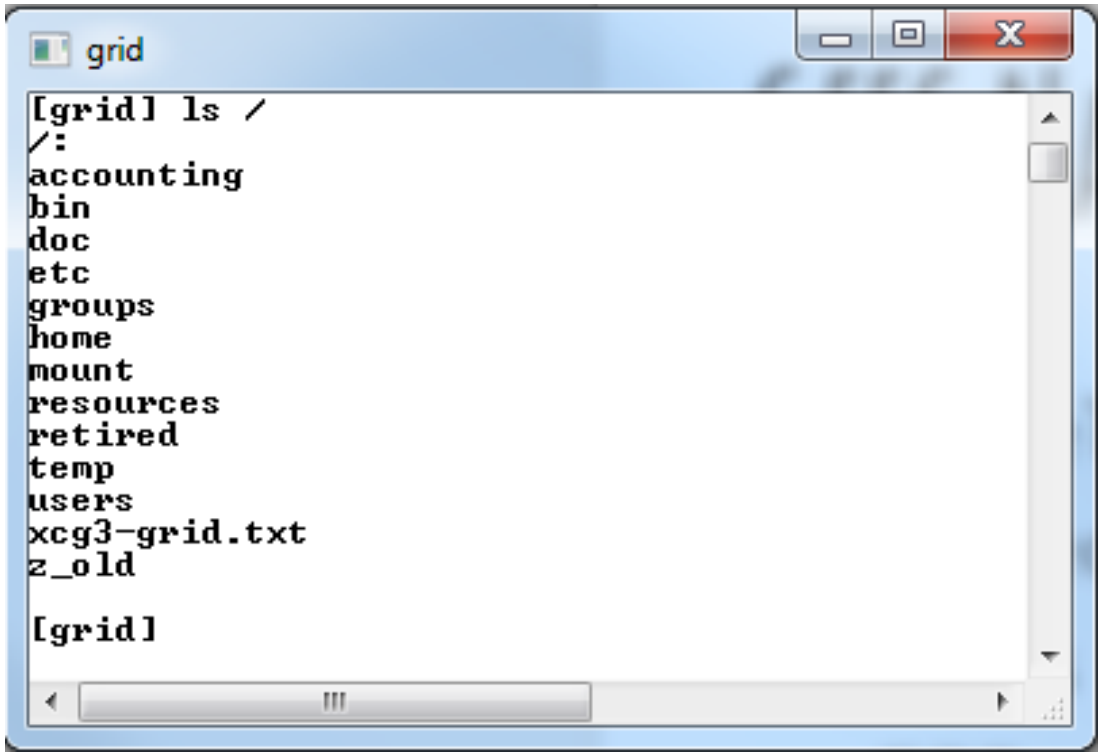
Using The CCC

Using The CCC

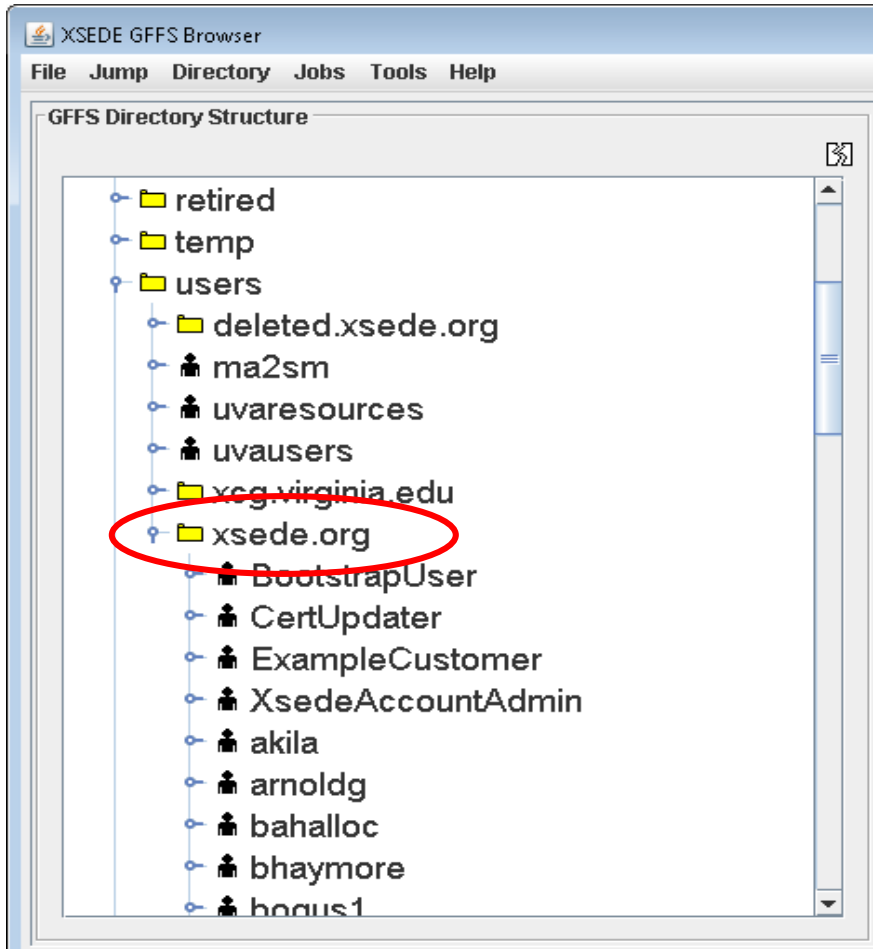
- Using CCC is very similar to what the researchers are used to with typical shared computational environment
  - There is a namespace (GFFS) similar to unix directory structure
- The steps for using CCC are as follows
  - Login to access the system
  - Use qsub to submit their job(s)
  - Use qstat to check the status of the job(s)

GFFS NameSpace

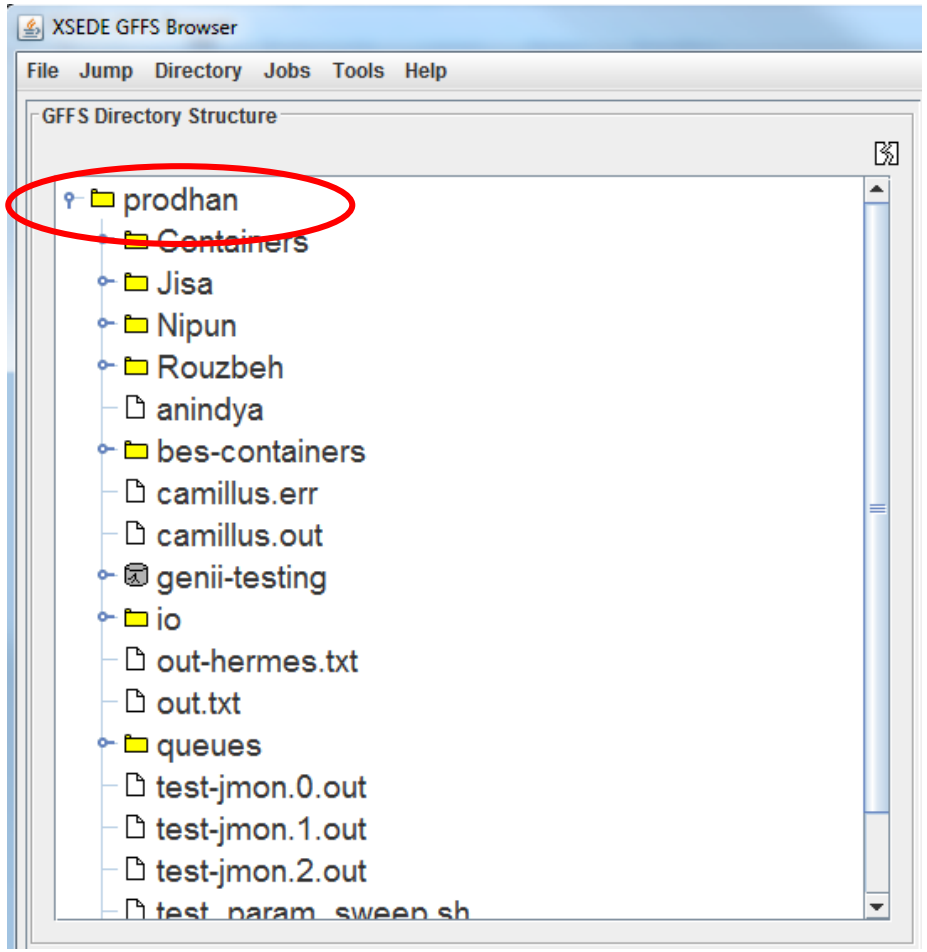
- Modeled on the Unix directory structure
- Maps file-names to resource EPRs
- Genesis II client supports access to GFFS namespace via-
  - command line interface
  - GUI
  - APIs
  - Mounting the GFFS namespace using FUSE



Users and Home Directory

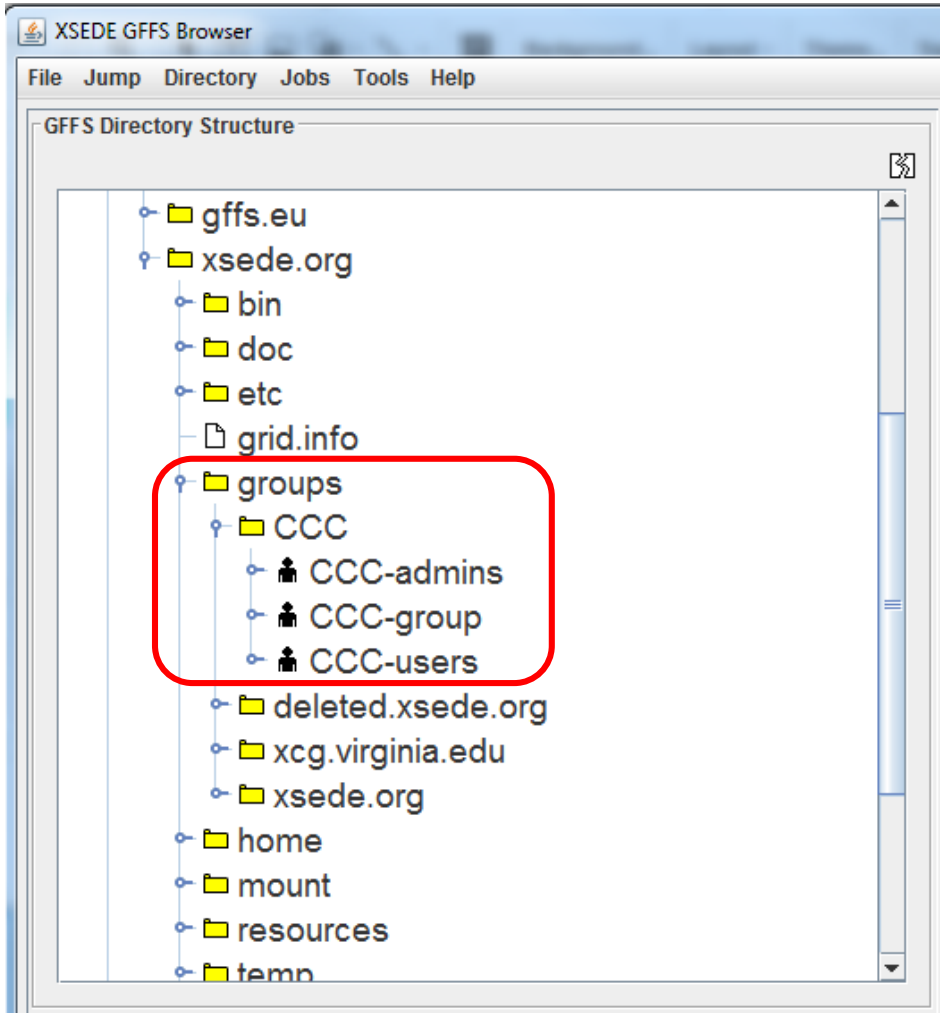


User directory for the xsede user (/users/xsede.org)



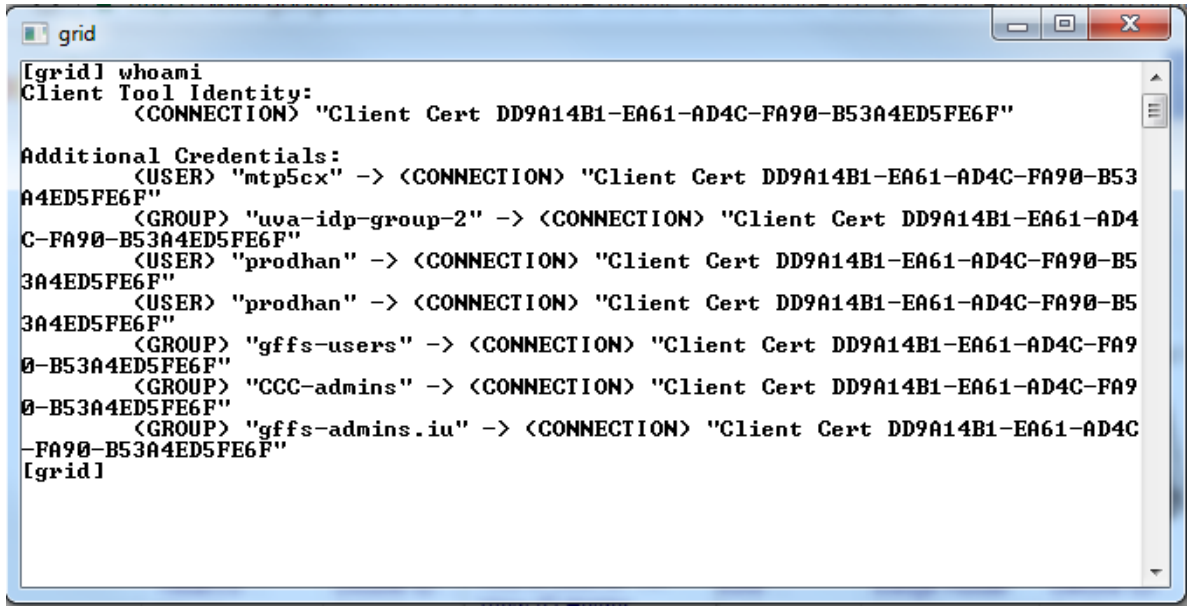
My home directory on the grid (/home/xsede.org/prodhan)

Groups



- Users are grouped into different user-groups
- Each group has their own **permissions** and **capabilities**
- Admin groups are responsible for the **administration** of different resources

Authentication-Credential Wallet

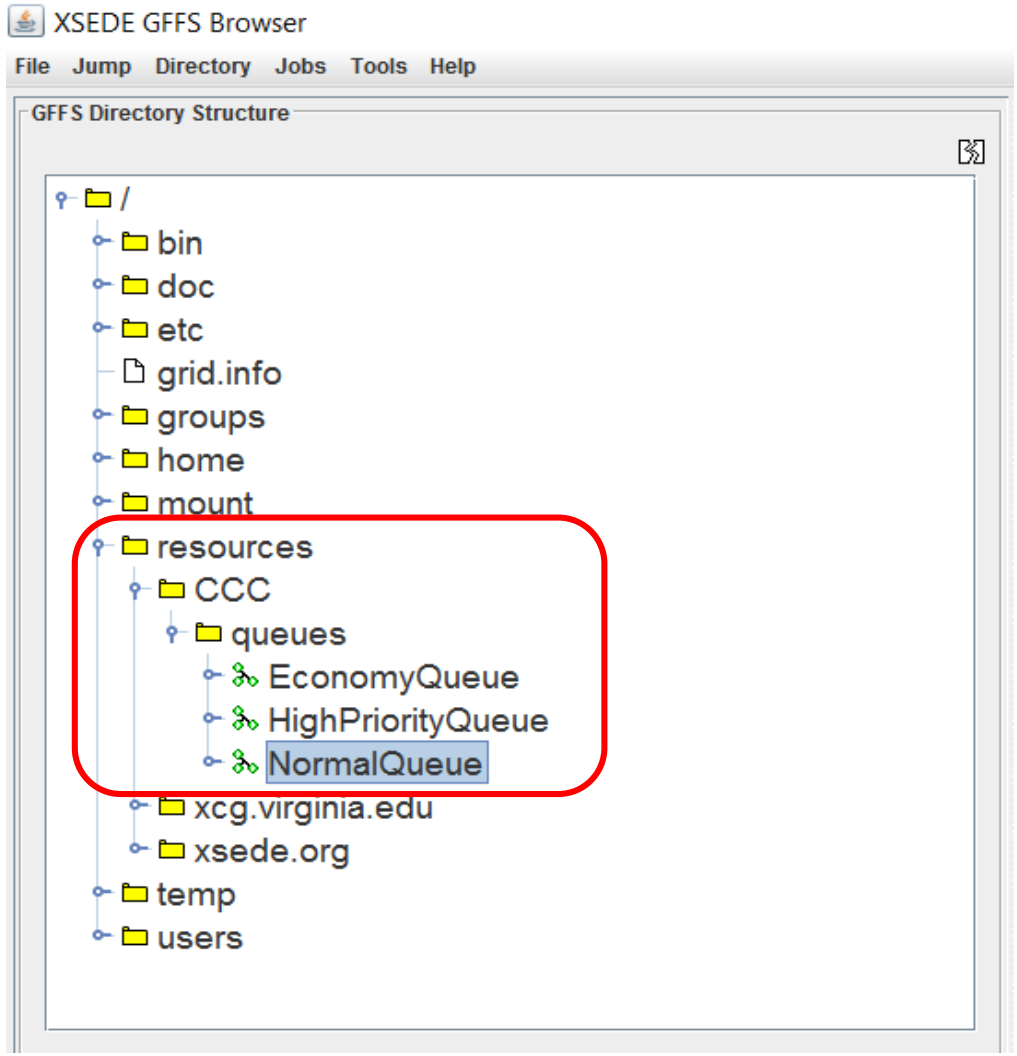


- User's credential are used to authenticate the user into the system.
- User's and User-groups create a credential wallet which can be used to run the jobs and pay for them.
- The system is build on standards

JSDL & JSDL++

- JSDL is the standard XML based language to describe jobs
- Defines-
  - Application Specification (e.g. LAMMPS)
  - Resource requirements (e.g. GPU, 32 cores, 8 nodes etc.)
  - Data staging specification (e.g. input and output files)
- JSDL++ is the non-standard extension of JSDL to allow multiple job descriptions in one jsdl file
  - Addresses the shortcomings of JSDL in a heterogeneous environment

Resources

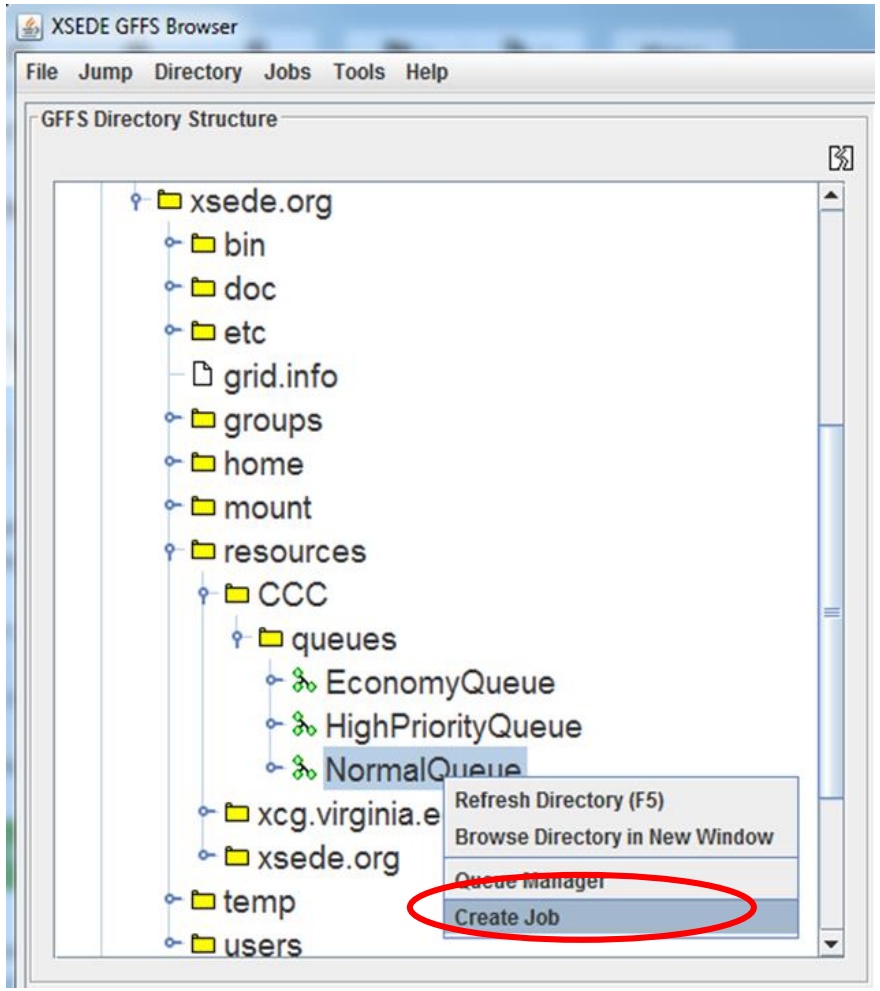


- Grid Queue(s) are mapped on the ***/resources/CCC/queues*** location.
- User(s) can submit their job(s) on one of the three priority queues based on their requirement.
- To submit a job to the queue, with a job description file we just need to run the following command and qstat command can be ised to monitor the job status

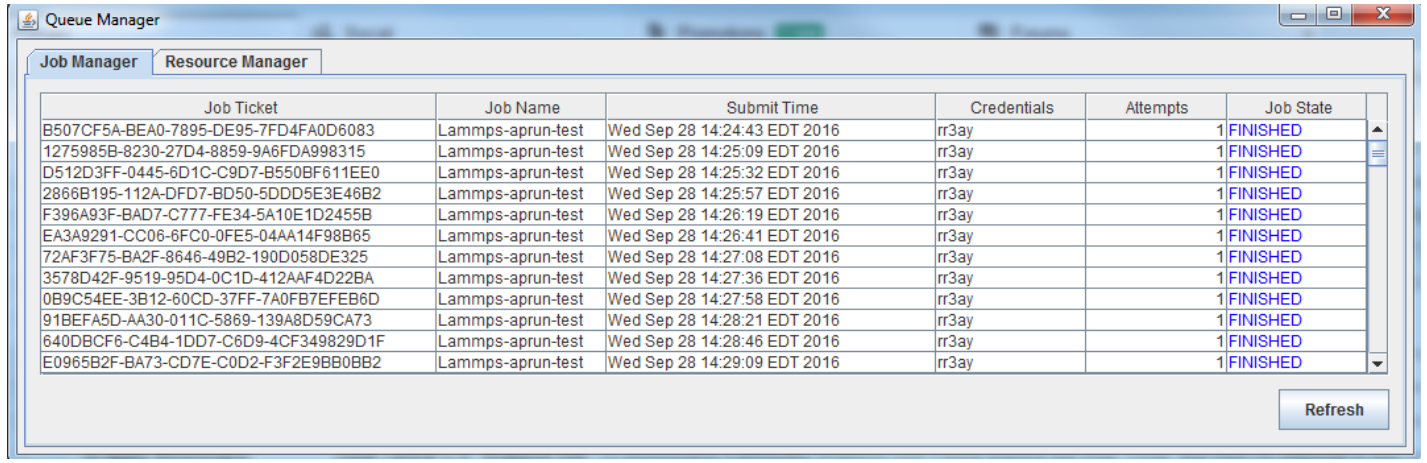
```
qsub /resources/CCC/queues/NormalQueue
local://home/drake/job.jsdl
```

```
qstat /resources/CCC/queues/NormalQueue
```

Job SubMission & Monitoring Through GUI

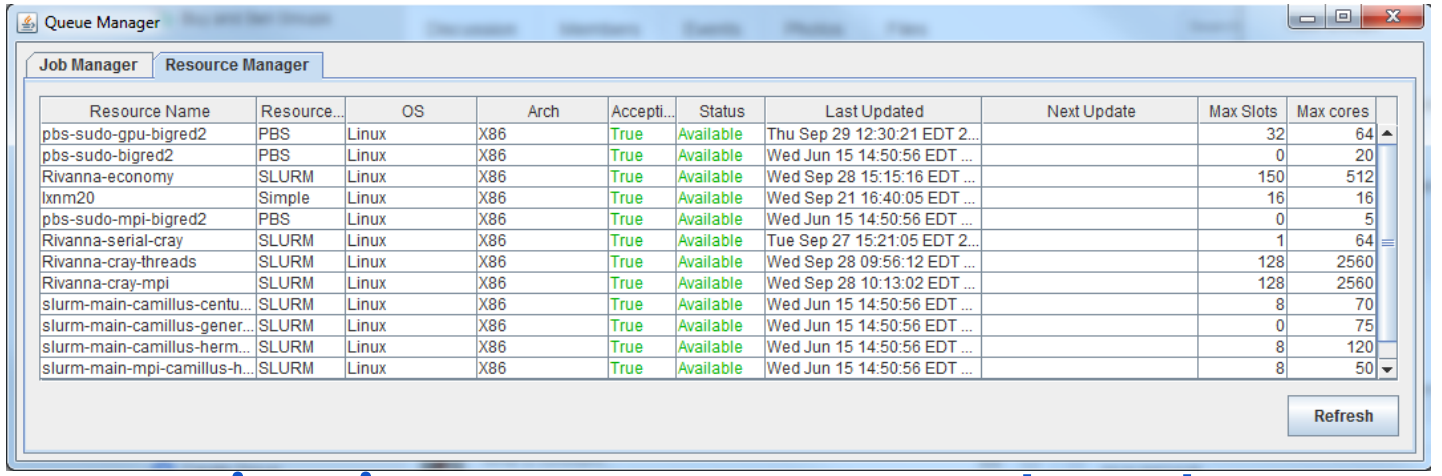


Job submission through GUI



Job Ticket	Job Name	Submit Time	Credentials	Attempts	Job State
B507CF5A-BEA0-7895-DE95-7FD4FA0D6083	Lammps-aprun-test	Wed Sep 28 14:24:43 EDT 2016	rr3ay	1	FINISHED
1275985B-8230-27D4-8859-9A6FDA998315	Lammps-aprun-test	Wed Sep 28 14:25:09 EDT 2016	rr3ay	1	FINISHED
D512D3FF-0445-6D1C-C9D7-B550BF611EE0	Lammps-aprun-test	Wed Sep 28 14:25:32 EDT 2016	rr3ay	1	FINISHED
2866B195-112A-DFD7-BD50-5DDD5E3E46B2	Lammps-aprun-test	Wed Sep 28 14:25:57 EDT 2016	rr3ay	1	FINISHED
F396A93F-BAD7-C777-FE34-5A10E1D2455B	Lammps-aprun-test	Wed Sep 28 14:26:19 EDT 2016	rr3ay	1	FINISHED
EA3A9291-CC06-6FC0-0FE5-04AA14F98B65	Lammps-aprun-test	Wed Sep 28 14:26:41 EDT 2016	rr3ay	1	FINISHED
72AF3F75-BA2F-8646-49B2-190D058DE325	Lammps-aprun-test	Wed Sep 28 14:27:08 EDT 2016	rr3ay	1	FINISHED
3578D42F-9519-95D4-0C1D-412AAF4D22BA	Lammps-aprun-test	Wed Sep 28 14:27:36 EDT 2016	rr3ay	1	FINISHED
0B9C54EE-3B12-60CD-37FF-7A0FB7EFEB6D	Lammps-aprun-test	Wed Sep 28 14:27:58 EDT 2016	rr3ay	1	FINISHED
91BEFA5D-AA30-011C-5869-139A8D59CA73	Lammps-aprun-test	Wed Sep 28 14:28:21 EDT 2016	rr3ay	1	FINISHED
640DBCF6-C4B4-1DD7-C6D9-4CF349829D1F	Lammps-aprun-test	Wed Sep 28 14:28:46 EDT 2016	rr3ay	1	FINISHED
E0965B2F-BA73-CD7E-C0D2-F3F2E9BB0BB2	Lammps-aprun-test	Wed Sep 28 14:29:09 EDT 2016	rr3ay	1	FINISHED

Monitoring a job through GUI



Resource Name	Resource...	OS	Arch	Accepti...	Status	Last Updated	Next Update	Max Slots	Max cores
pbs-sudo-gpu-bigred2	PBS	Linux	X86	True	Available	Thu Sep 29 12:30:21 EDT 2...		32	64
pbs-sudo-bigred2	PBS	Linux	X86	True	Available	Wed Jun 15 14:50:56 EDT ...		0	20
Rivanna-economy	SLURM	Linux	X86	True	Available	Wed Sep 28 15:15:16 EDT ...		150	512
lxnm20	Simple	Linux	X86	True	Available	Wed Sep 21 16:40:05 EDT ...		16	16
pbs-sudo-mpi-bigred2	PBS	Linux	X86	True	Available	Wed Jun 15 14:50:56 EDT ...		0	5
Rivanna-serial-cray	SLURM	Linux	X86	True	Available	Tue Sep 27 15:21:05 EDT 2...		1	64
Rivanna-cray-threads	SLURM	Linux	X86	True	Available	Wed Sep 28 09:56:12 EDT ...		128	2560
Rivanna-cray-mpi	SLURM	Linux	X86	True	Available	Wed Sep 28 10:13:02 EDT ...		128	2560
slurm-main-camillus-centu...	SLURM	Linux	X86	True	Available	Wed Jun 15 14:50:56 EDT ...		8	70
slurm-main-camillus-gener...	SLURM	Linux	X86	True	Available	Wed Jun 15 14:50:56 EDT ...		0	75
slurm-main-camillus-herm...	SLURM	Linux	X86	True	Available	Wed Jun 15 14:50:56 EDT ...		8	120
slurm-main-mpi-camillus-h...	SLURM	Linux	X86	True	Available	Wed Jun 15 14:50:56 EDT ...		8	50

Monitoring resource status through GUI

First Applications

- Large Sequential Jobs
  - simulate the performance of a search engine
  - used by a group in Computer Science Department
- Single/Multi-node Parallel Jobs (Lammps)
  - molecular dynamics simulation
  - used by a group in Mechanical and Aerospace Engineering Department
  - cpu and gpu acceleration
- High-Throughput Computing
  - Astro-chemical Simulation
  - used by a group in Chemistry Department
- Big **Gromacs** run upcoming

Social, political and market aspects

Social & Political Issues

- Traditionally researchers are accustomed to using the shared resources with no QoS or not fairly defined priority
- There is often no mechanism of allocating resources fairly
- And often sharing becomes very one sided
- Hence we need a resource market

Resource Pricing and Market model

- Static pricing (Initially)
- Similar to Amazon's static pricing scheme
- **Standard base pricing** for a **standard resource type**
  - 2.1 GHz CPU with 4GB mem/core
  - Ethernet or GigE network connections
- Additional features with additional cost (e.g. Large memory, InfiniBand, GPU)
- Different cost for different QoS jobs
  - Different scaling factors based on QoS
- An initial distribution of allocations to get the market flowing

Governance and Clearance

- What about the chronic debtors?
- Any obligatory exchange of real money will make it a non-starter to the potential adapters.
- MoU to be signed by each institute
  - Institute can opt-out any time
  - No way to force anyone to pay
  - Institutions will vouch for their users

Related Work

Related Work

- Open Science Grid (OSG)
- Grid Economy
- Cloud Computing
- Cloud Federation

Open Science Grid

- Developed primarily for high energy physics in the 90's
- Resources are contributed in an altruistic manner
- ***Issues***
  - No incentive for resource sharing
  - No QoS support in OSG
  - OSG is targeted for high throughput sequential job while CCC supports sequential, threaded or MPI jobs

Grid Economy

- Plethora of work in **The Grid Economy**
- Spawn (Waldspurger et al.), Nimrod (Abramson et al.), The Grid Economy (Buyya et al.), GridEcon (Altmann et al.), InterGrid (Buyya et al.)
- ***Issues***
  - Much of the existing work has been done in simulations
    - Synthesized data
    - Small grid test-beds
  - None of the existing production grids or clusters or supercomputing centers use these solutions
  - Not focused on on-Demand solutions

Cloud Computing and Federation

- “Infinite” resource on-Demand
- Amazon AWS the leader in cloud computing
- Cloud Federation: interconnecting the **cloud** computing environments of two or more service providers. i.e. Contrail (carlini et al.), Reservoir (rochwerger et al.)
- ***Issues:***
  - Designed for VMs
  - More expensive options
  - A resource consumer can’t be a resource provider

Final Remarks

Should YOu Join CCC

- If you need access to diverse resources and quick turnaround during bursts then CCC can definitely help you.
- Anyone with a small cluster can join the collaborative as a provider.

How to Join CCC

- To access resources within CCC-
  - You will just need the genesis II client to access the computational and data resources available in CCC
  - You would probably need an allocation on CCC too.
  - Identity (e.g. XSEDE id or CCC id through your institution)
- Signing an MOU
- To share your resources-
  - You will need a genesis II container installed on your server and allow CCC to submit jobs to the local queuing system
  - **No root required !!!**

Conclusion and Future Work

- Future direction
  - Dynamic pricing model
  - Desktop VMs
  - Support starting VMs for users, not just for jobs
  - Expand to more Institutions
- We believe federations like CCC can go a long way to deal with the growing need of CI resources
  - However the success of CCC really depends on the participation of users and user institutes

Questions

Thank You All

Reference (1)

1	R. Buyya, D. Abramson and S. Venugopal, "The Grid Economy," Proceedings of the IEEE, vol. 93, no. 3, 2005.
2	J. Altmann, C. Courcoubetis, G. D. Stamoulis, M. Dramitinos, T. Rayna, M. Risch and C. Bannink, "GridEcon: A market place for computing resources," Grid Economics and Business Models, pp. 185-196, 2008.
3	R. Wolski, J. S. Plank, T. Bryan and J. Brevik, "G-commerce: Market formulations controlling resource allocation on the computational grid," in *15th International Parallel and Distributed Processing Symposium*, 2001
4	P. Padala, C. Harrison, N. Pelfort, E. Jansen, M. P. Frank and C. Chokkareddy, "OCEAN: the open computation exchange and arbitration network, a market approach to meta computing," in *International Symposium on Parallel and Distributed Computing*, 2003.
5	C. Waldspurger, T. Hogg, B. Huberman, J. O. Kephart and W. S. Storn, "Spawn: A distributed computational economy," *IEEE Transactions on Software Engineering,* vol. 18, no. 2, pp. 103-117, 1992.
6	F. Berman, R. Wolski, S. Figueira, J. Schopf and G. Shao, "Application-level scheduling on distributed heterogeneous networks," in *ACM/IEEE Conference on Supercomputing*, 1996.
7	O. Regev and N. Nisan, "The popcorn market. online markets for computational resources," *Decision Support Systems,* vol. 28, no. 1, pp. 177-189, 2000.
8	D. Abramson, R. Sosic, J. Giddy and B. Hall, "Nimrod: a tool for performing parametrised simulations using distributed workstations," in *Fourth IEEE International Symposium on High Performance Distributed Computing*, 1995.

Reference (2)

9	"Amazon EC2," [Online]. Available: https://aws.amazon.com/ec2/. [Accessed 1 1 2016].
10	"Amazon AWS Instance Types," [Online]. Available: https://aws.amazon.com/ec2/instance-types/. [Accessed 1 1 2016].
11	"Open Science Grid," [Online]. Available: http://www.opensciencegrid.org/. [Accessed 1 1 2016].
12	R. Pordes, D. Petravick, B. Kramer, D. Olson, M. Livny, A. Roy, P. Avery, K. Blackburn, T. Wenaus, F. Würthwein, I. Foster, R. Gardner, M. Wilde, A. Blatecky, J. McGee and R. Quick, "The open science grid," in *Journal of Physics: Conference Series*, 2007.
13	R. Buyya, R. Ranjan and R. N. Calheiros, "Intercloud: Utility-oriented federation of cloud computing environments for scaling of application services," in *Algorithms and architectures for parallel processing*, 2001.
14	E. Carlini, M. Coppola, P. Dazzi, L. Ricci and G. Righetti, "Cloud federations in contrail," in *Euro-Par: Parallel Processing Workshops*, 2012.
15	B. Rochwerger, D. Breitgand, E. Levy, A. Galis, K. Nagin, I. M. Llorente, R. Montero, Y. Wolfsthal, E. Elmroth and J. Caceres, "The reservoir model and architecture for open federated cloud computing," *IBM Journal of Research and Development,* vol. 53, no. 4, 2010.

References (3)

16	"RightScale: Cloud Portfolio Management," [Online]. Available: http://www.rightscale.com/. [Accessed 1 1 2016].
17	"Dell Hybrid Cloud," [Online]. Available: http://www.enstratius.com/home. [Accessed 1 1 2016].
18	"Scalr Enterprise Cloud Management Platform," Scalr, [Online]. Available: http://www.scalr.com/. [Accessed 1 1 2016].
19	"Kaavo- Cloud Management Software," [Online]. Available: http://www.kaavo.com/. [Accessed 1 1 2016].
20	F. Bachmann, I. Foster, A. Grimshaw, D. Lifka, M. Riedel and S. Tuecke, "XSEDE Architecture Level 3 Decomposition," 2013.
21	M. A. T. Prodhan and A. Grimshaw, "Market-based on demand scheduling (MBoDS) in co-operative grid environment," in *XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*, St. Louis, Mo, 2015.