Harnessing Computing Power

Grid, Xgrid: A complementary approach

Dr. Massimo Marino
ARTS Project Leader
Apple Scientific & Research Programs
Apple Europe, Ltd
marino.m@euro.apple.com

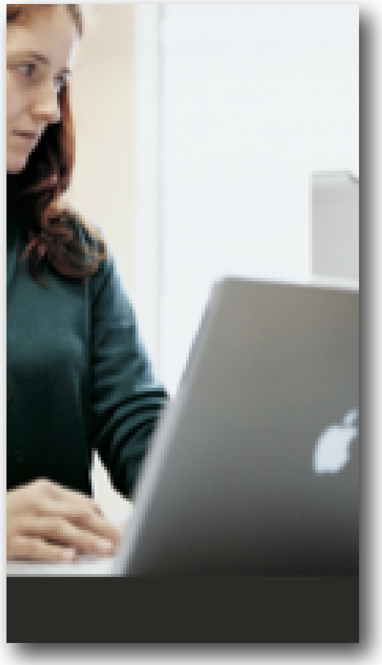
Overview

The man on stage

An old dream

Xgrid: Ready to share

In the real world



On the web

Where do I come from?

Physicist/Computer Scientist with 17 years presence in the field

1988 - 1997
CERN Laboratory - Switzerland

- Detector R&D
- RD41
- LHC/CMS experiment - Computing Group

1997 - 2005
Lawrence Berkeley National Laboratory - USA

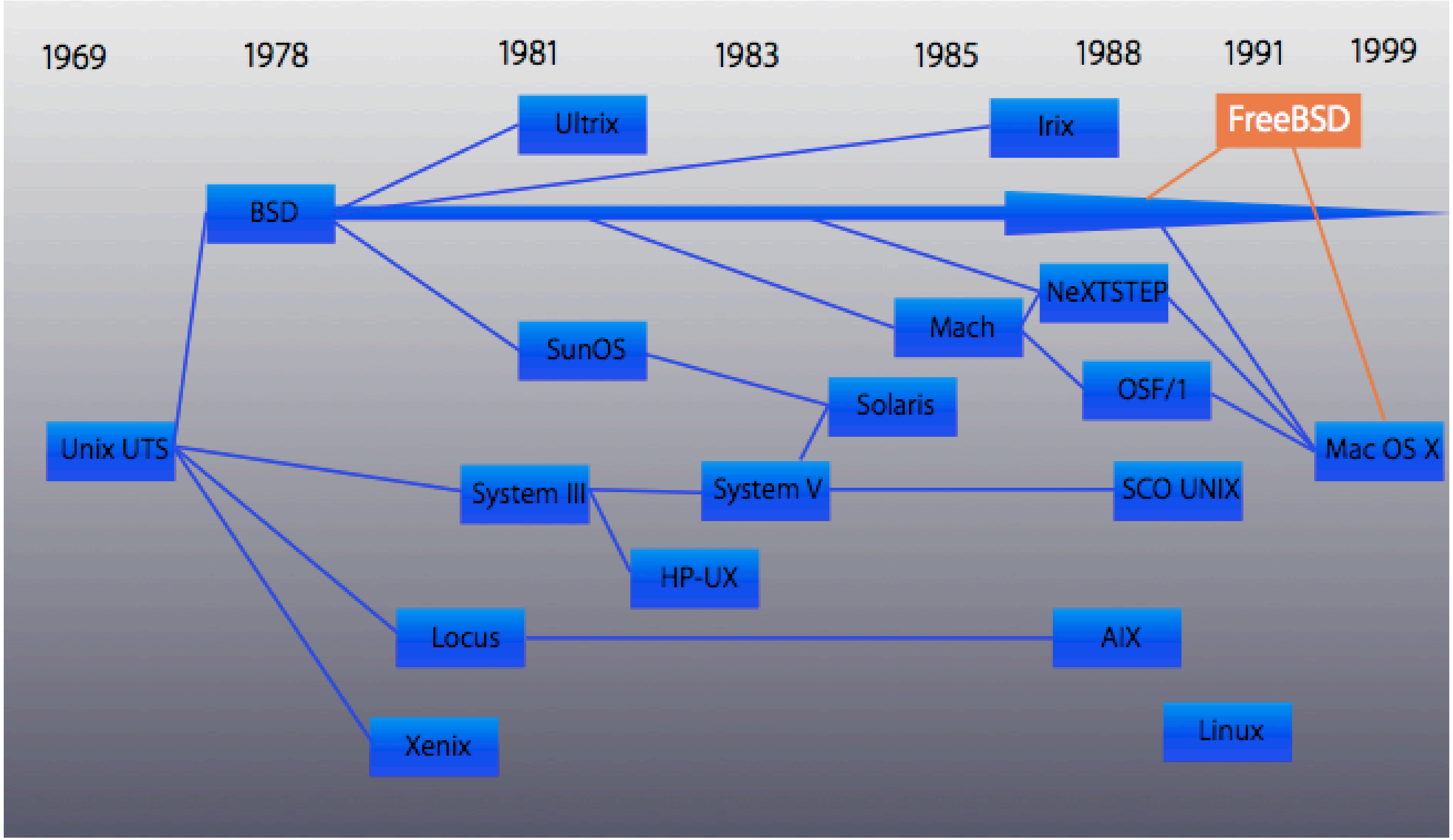
- NERSC (National Energy Research Scientific Computing - DOE)
- BaBar experiment @ SLAC
- LHC/ATLAS experiment @ CERN

Computing exposure

- Various Unix flavors
  - Solaris
  - Scientific Linux (SL)
  - Red Hat
  - HP-UX
  - AIX
  - Mac OS X
- Various languages
  - Fortran, Smalltalk, Eiffel, C++, Python,...
- Mac OS
  - HEP fully into Unix workstations
    - Mac mainly platform of choice for graphics and papers
  - On radar screens once Apple had a real OS for scientists: Mac OS X

Unix Family Tree

Ancestors of Mac OS X



Dr Massimo Marino, marino.m@euro.apple.com

Why Unix was the right move

- Highly "compose-able" as operating systems go
  - It's an onion, not a potato
- Gives Apple a **huge amount of open source** to leverage
  - critical to the implementation process and evolution progress
- Instant **portability** for a huge number of important applications (and important users) in SciTech and other fields
- **Interoperability** with *BSD, Linux, Solaris and other UNIX-derivatives
  - came almost for free
- Development community is active, innovative and a well-established track record on **OS design and security**

The next Unix move

Pushing forward with Mac OS X 10.5 Leopard

Second Mac OS X version to run natively on intel processors

- **64-bit OS**
  - can seamlessly run 32-bit applications and extensions
  - unlike other OSes, only one version of the software
    - anything, be it 64 or 32-bit, **runs natively and without penalty**
      - Apache2, MySQL, Postfix and Cyrus, iChat Server, QuickTime Streaming Server
- **Certified Unix 03** (The Open Group)
  - not *just* Unix-based
    - Conforming to the Single UNIX Specification: SUS version 3
  - runs any Unix-certified application after recompilation for the Mac platform
    - no changes to the program APIs, no changes to the code
- **DTrace** (Open Source) & **Xray**

Grid: an old* dream comes true

*1996: proposal to NSF; 1998: The Grid: Blueprint for a New Computing Infrastructure

GRID

Older than that

1969 UCLA press release

"As now, computer networks are still in their infancy. But as they grow up and become more sophisticated, we will probably see the spread of computing utilities, which, like present electric and telephone utilities, will service individual homes and offices across the country."

Dr. Leonard Kleinrock

Moore’s Law

it has all the blame/merit

- computing power and storage grew enormously
- $/GFlops dropped dramatically

Famous last words

- “The world will only need five computers”
  Thomas J. Watson, IBM
- “640 KB is all the memory you will ever need”
  Bill Gates, Microsoft
- “There is absolutely no need for a computer at home”
  Ken Olsen, DEC

A Paradigm shift

From

- costly centralized data centers
  - scientists share financial resources

To

- generalized institutions computing power
  - capable local IT infrastructures
  - scientists share (local) access to several and powerful computers

Share CPUs rather than $$

- how-to and in an efficient way

Grid computing

it’s all about sharing

- heterogeneous resources
  - different platforms (hw/sw architectures, languages), tools, ...
- different locations
  - belonging to different administrations

Functional taxonomy

- Computational GRIDs (and CPU harnessing ones)
- Data GRIDs
- Equipment GRIDs

The new problem

which GRID has to answer to

- Develop a true “sharing” technology and on a global scale
  - CPU power
  - Storage
  - Databases
  - Services
- A secure technology
- Load balancing
- Network
- Open Standards

A global and huge effort

Global scale GRID projects

- Standards and middleware
- Services
- Applications and scheduling tools
- Networking

Very often overlapping

A de-facto standard

- Globus Alliance Toolkit

A huge effort

  - LCG: 389 FTE-years over 3.5 years (at 2004)

The reason for “Local” approaches

vast and powerful IT assets

- CPUs are not fully utilized across same department
- Large computing resources are idling in one dept while high demand (and unsatisfied) is experienced on the next
- Compute and applications still exceed capabilities of a single group

“Local” solutions not mutually exclusive

- harness idle computing power over LAN/WAN
- distributed computing systems

Xgrid:
Cluster-ready architecture

Built-in “gridification”

Apple Xgrid - Distributed computing the easy way

- Cluster-ready architecture
  - fast easy configuration
  - accessible to non IT specialists
- Harness computing power across the network
- Bonjour (ZeroConf) and DNS lookups support
  - automatic agents/clients/controllers/ discovery
- Both local and remote users
- XML-based open protocol for network comms
- Fault-tolerance features
- Kerberized access

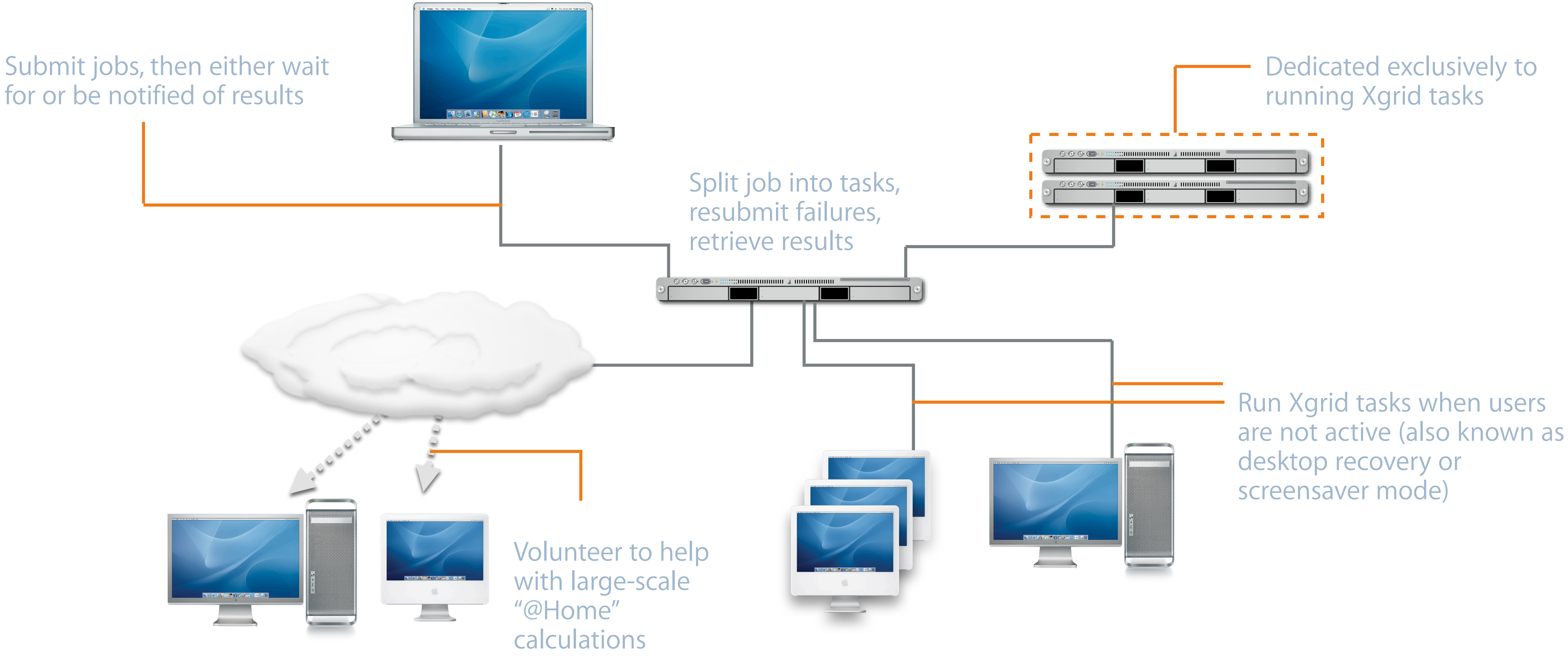
Xgrid architecture

Three-tier architecture

- clients
  - MPI apps, CLI/GUI tools
    - describe/submit/retrieve jobs
- controllers
  - distribute jobs, manage comms
- agents
  - system daemons

Three Tier Architecture - Xgrid 1.0

- Client, Controller, and Agent



Submit jobs, then either wait for or be notified of results

Dedicated exclusively to running Xgrid tasks

Split job into tasks, resubmit failures, retrieve results

Run Xgrid tasks when users are not active (also known as desktop recovery or screensaver mode)

Volunteer to help with large-scale “@Home” calculations

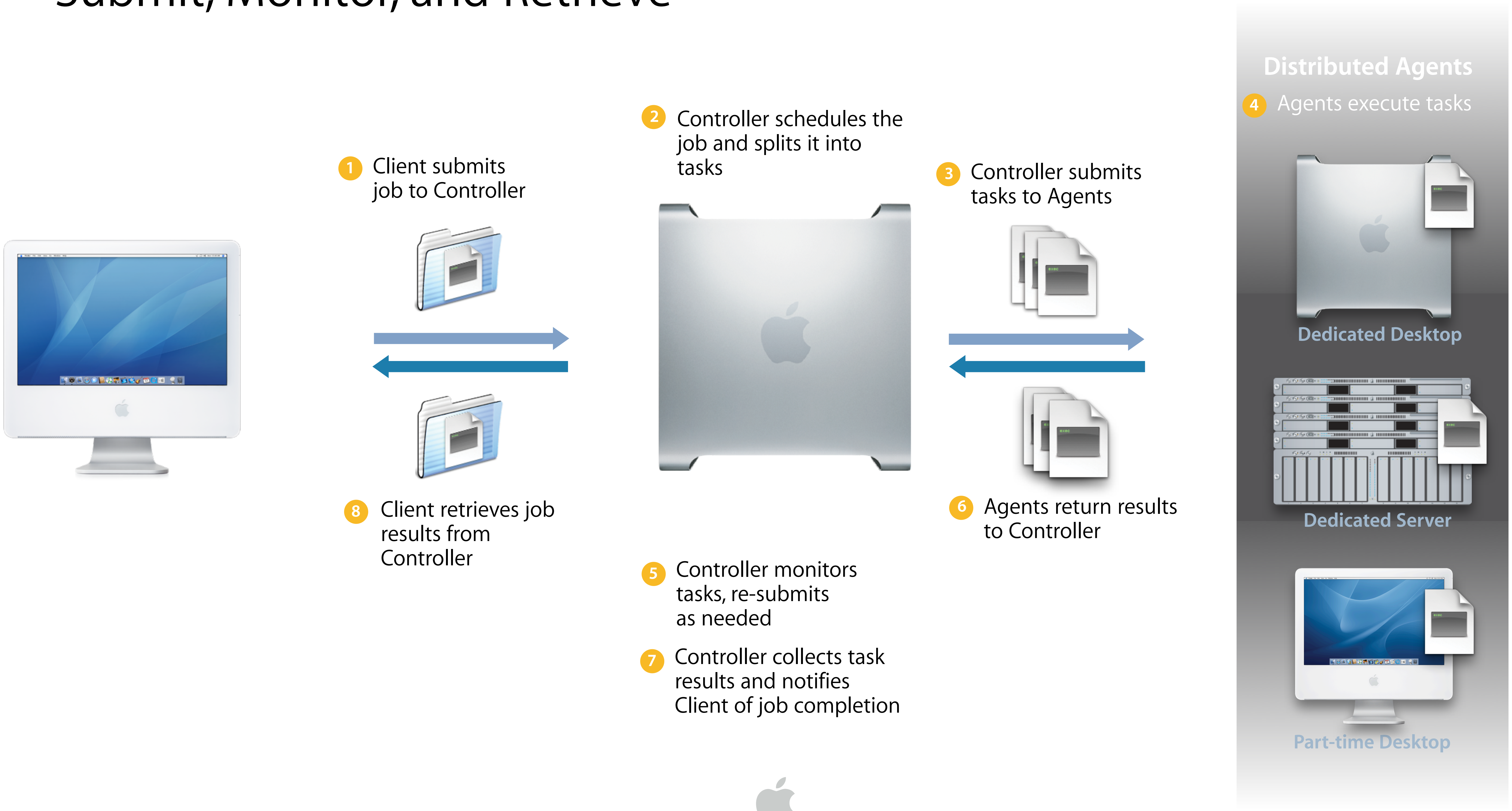
Xgrid Security

Authentication

- MD5 hashes pass protocol
  - agents run jobs as user ‘nobody’
- Kerberos
  - agents run jobs with submitter privileges
- SSH tunneling
  - agents/clients connect to “localhost”
- No ports to be opened on clients or agents

Xgrid Workflow

- Submit, Monitor, and Retrieve



Distributed Agents

4 Agents execute tasks

1 Client submits job to Controller

2 Controller schedules the job and splits it into tasks

3 Controller submits tasks to Agents

5 Controller monitors tasks, re-submits as needed

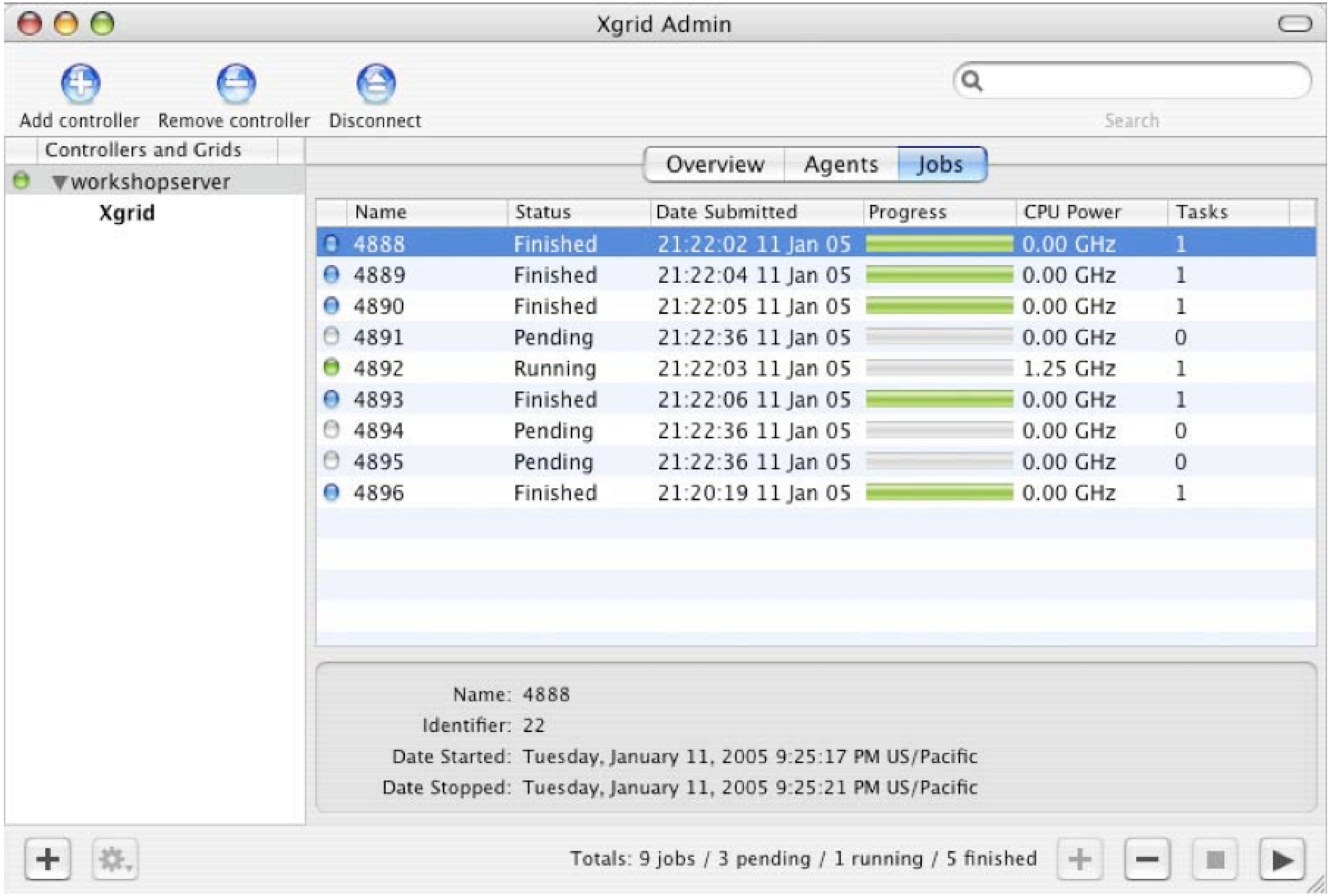
6 Agents return results to Controller

7 Controller collects task results and notifies Client of job completion

8 Client retrieves job results from Controller

Xgrid Admin tool

- manages multiple Xgrid controllers
- surveys/manages agents activities and jobs status
- manages logical agents sub-pools
- monitors dedicated CPU power



Name	Status	Date Submitted	Progress	CPU Power	Tasks
4888	Finished	21:22:02 11 Jan 05		0.00 GHz	1
4889	Finished	21:22:04 11 Jan 05		0.00 GHz	1
4890	Finished	21:22:05 11 Jan 05		0.00 GHz	1
4891	Pending	21:22:36 11 Jan 05		0.00 GHz	0
4892	Running	21:22:03 11 Jan 05		1.25 GHz	1
4893	Finished	21:22:06 11 Jan 05		0.00 GHz	1
4894	Pending	21:22:36 11 Jan 05		0.00 GHz	0
4895	Pending	21:22:36 11 Jan 05		0.00 GHz	0
4896	Finished	21:20:19 11 Jan 05		0.00 GHz	1

Name: 4888
Identifier: 22
Date Started: Tuesday, January 11, 2005 9:25:17 PM US/Pacific
Date Stopped: Tuesday, January 11, 2005 9:25:21 PM US/Pacific

Totals: 9 jobs / 3 pending / 1 running / 5 finished

Xgrid - Distributed computing the easy way

Since Tiger pre-installed on all Macs

- Xgrid handles the hard work of:
  - connecting nodes into a cluster
  - managing a queue of jobs and subtasks
  - monitoring node availability
  - scheduling tasks on the nodes
  - copying executables and input data to nodes
  - staging output data and collecting results
- Security can be handled via ad-hoc mutual authentication (MD5 hash pass, Kerberos) or managed via Open Directory. No ports to be opened at clients side
- See www.apple.com/acg/xgrid for more info

How easy?

Lowering the technology barrier

- Kentucky Dataseam Initiative (KDSI)
  - First such collaboration between K-12 schools and a university lab in the U.S.
    - Goal: over 5000 Mac platforms at schools. 2600+ so far participating already.
    - Supported by dedicated Apple back-end systems
    - Mac OS X (client & Server), Xserve, Xgrid
  - Dedicated to cancer research
    - James Graham Brown Cancer Center @ University of Louisville
  - Machines are used 24/7 with no special IT infrastructure but school's own

"We've reduced data processing jobs that used to take 50 years of CPU time down to 20 days — and we're speeding up our drug discovery by orders of magnitude."

Dr. John Trent, Director of Molecular Modeling and Associate Professor, Departments of Medicine, Biochemistry, and Molecular Biology, and Chemistry; James Graham Brown Cancer Center.

www.apple.com/education/profiles/louisville/index.html

Apple Advanced Computation Group

originators of Xgrid

- Researches algorithms and high-performance issues relevant to Apple technology
- ACG is interested in feedback about Xgrid
  - including, for example, how far the tachometer can be pushed in an actual clustered computation
- ACG research focuses on
  - Mac OS X with scientific applications
  - Vectorization
  - Tutorial materials for science customers and developers
  - Algorithm implementation/optimization for specific Apple products
  - Joint R&D with outside parties

Inquire about ACG research with Dr Ernest Prabhakar: prabhaka@apple.com

Xgrid mailing list: http://lists.apple.com/faq/pub/xgrid_users

MPI on Mac OS X

Available MPI Software for Mac OS X

Best implementations

- Argonne National Labs
  - MPICH-1.2.7
  - Myrinet enabled: MPICH-GM, MPICH-MX
  - Infiniband enabled MVAPICH
  - New: MPICH-2.1—the latest from Argonne
- LAM/MPI
  - Includes native Myrinet and InfiniBand support
- Open MPI
  - Joint venture by LANL, Oak Ridge,HLR Stuttgart, ICL/UT, Livermore, ZIH Dresden, Sandia, ...
  - Is Xgrid enabled
  - Includes native Myrinet and InfiniBand support

Cluster Interconnect Technology

The fabric that links nodes together

- Has a major impact on overall cluster performance
  - Does not use TCP/IP stack like Gigabit Ethernet
  - Data flows directly from the network to memory
  - Processors do not have to wait for data
  - Also have high bandwidth capability
- Current options for Apple-based clusters include: Myrinet, InfiniBand
  - Uses external interface cards
  - Link is either fiber optic or copper based
  - Connected to purpose-built high performance switches

When to Use High Performance Interconnects

Interconnect selection influences performance

- Often Gigabit ethernet provides good performance
- Parallel code with lots of messages require low latency
- Parallel code with large messages require high bandwidth
- A combination of the two

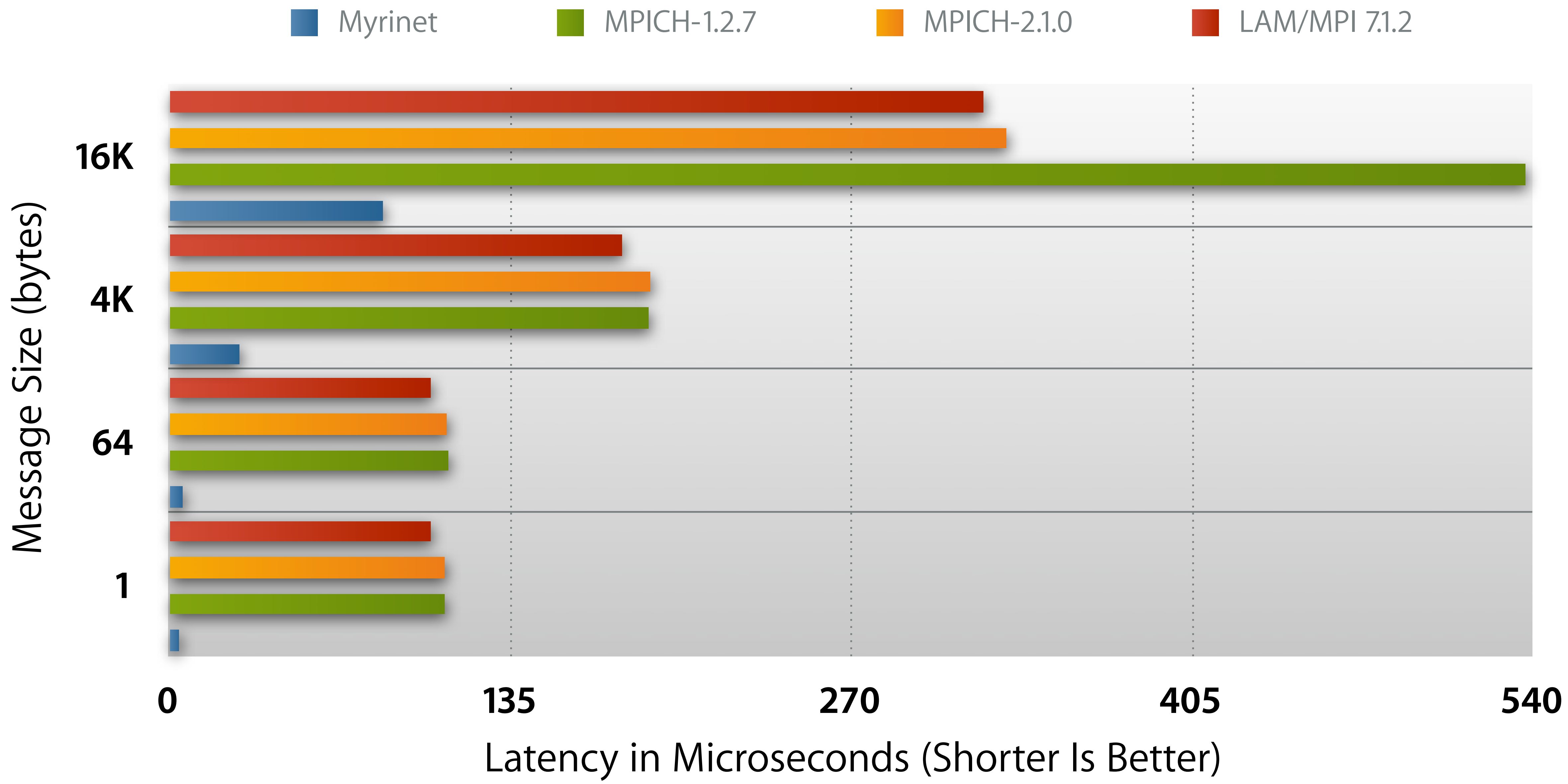
Shared compute environment

- A high performance interconnect attracts more users
- Variety of users with broad range of requirements

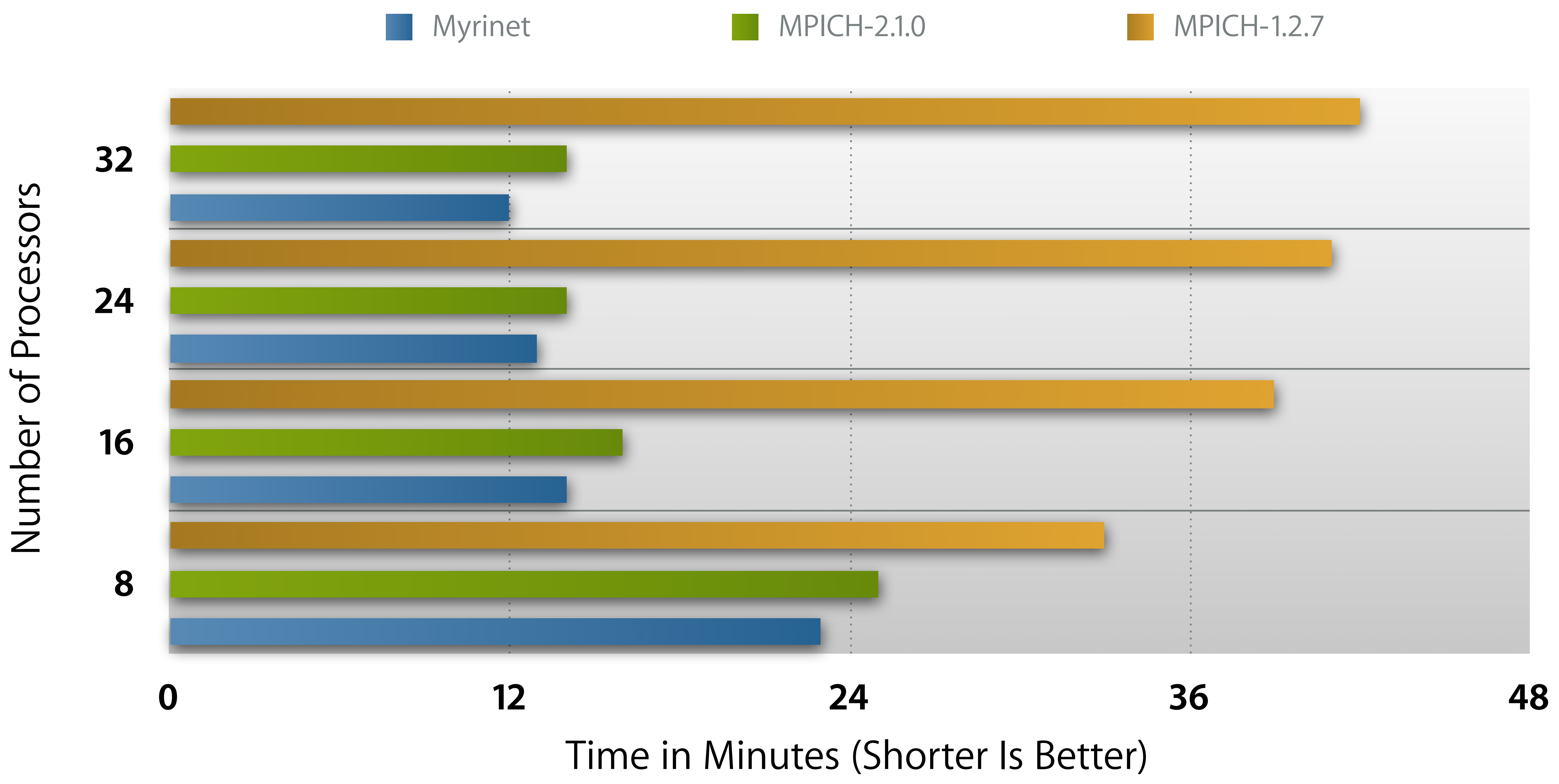
Testing MPI Performance

- MPI Ping-Pong Performance Benchmark on Mac OS X
  - Measure MPI software and fabric performance
  - Is set not to run on two cores of the same node
  - Benchmark executed on two processes
    - Message (ping) from the client sent to the server process
    - The message is bounced back to the client (pong)
    - Message size is variable
    - Communication time of the message is measured for performance
- MPI software impact on real world applications
  - Compare an application with different MPI software

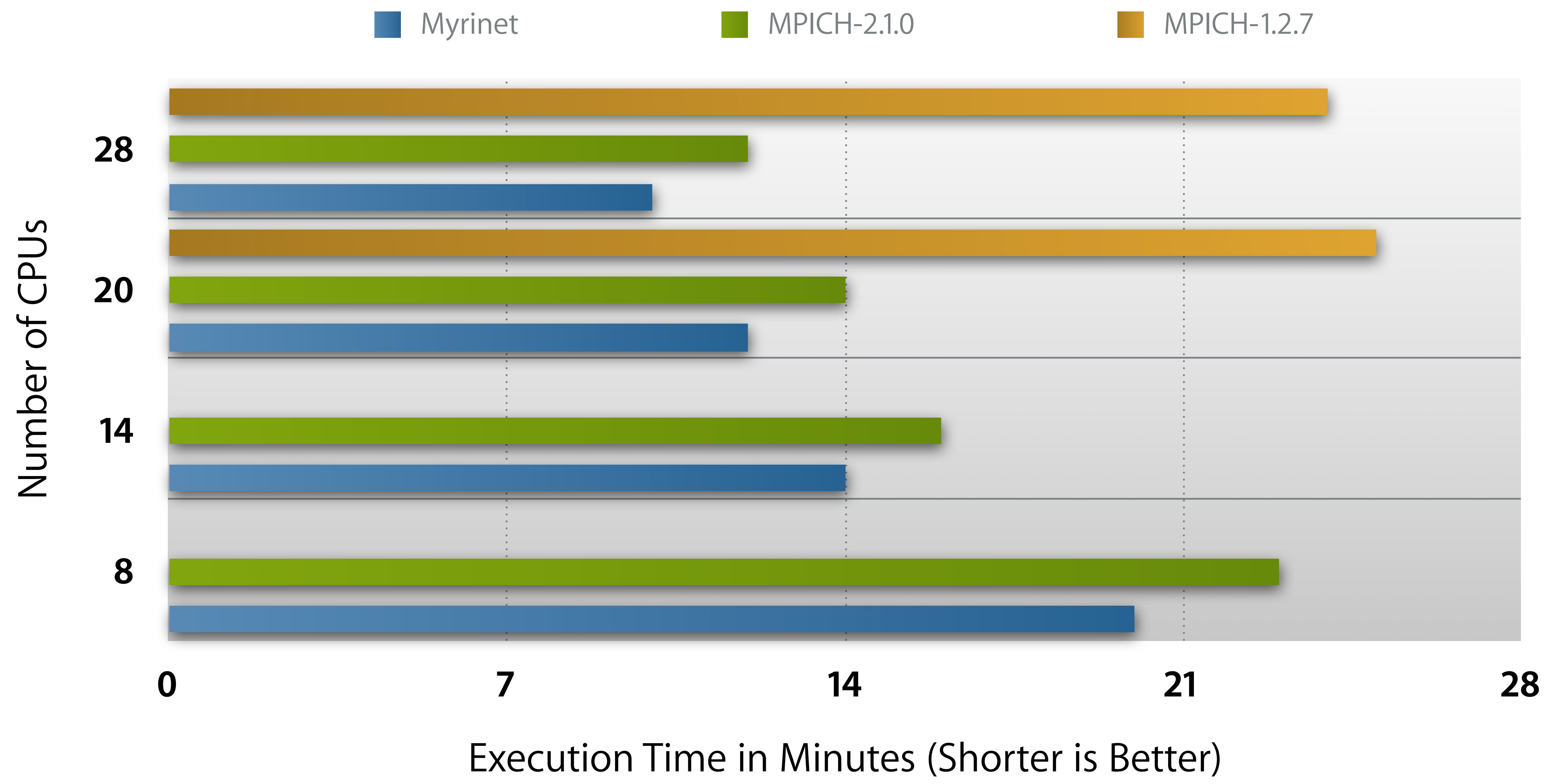
MPI Ping-Pong Benchmark



Gromacs 3.3.1 Benchmark



WRF 2.0.31 Benchmark



Dr Massimo Marino, marino.m@euro.apple.com

MPI Summary

- Choose your MPI software wisely
  - MPI software can have a major effect on performance
    - MPICH-1.2.x should not be used
    - Myrinet MPICH and MVAPICH is the exception
    - MPICH-2.0.x is a much better alternative to MPICH-1.2.x
  - LAM/MPI provides excellent performance
    - Compatible with different communication fabrics
  - OpenMPI
    - Excellent alternative to all other MPI software
    - Automatically selects fastest fabric at runtime
    - Can integrate with Xgrid as a basic job scheduler

Using Xgrid with MPI

- OpenMPI 1.0 Supports Xgrid (Support is Beta)
  - Compiling OpenMPI on Mac OS X automatically builds Xgrid Support
  - MPI jobs will automatically submit to Xgrid if environment is set

```
$> export XGRID_CONTROLLER_HOSTNAME=mycontroller.example.com
$> export XGRID_CONTROLLER_PASSWORD=pass
```

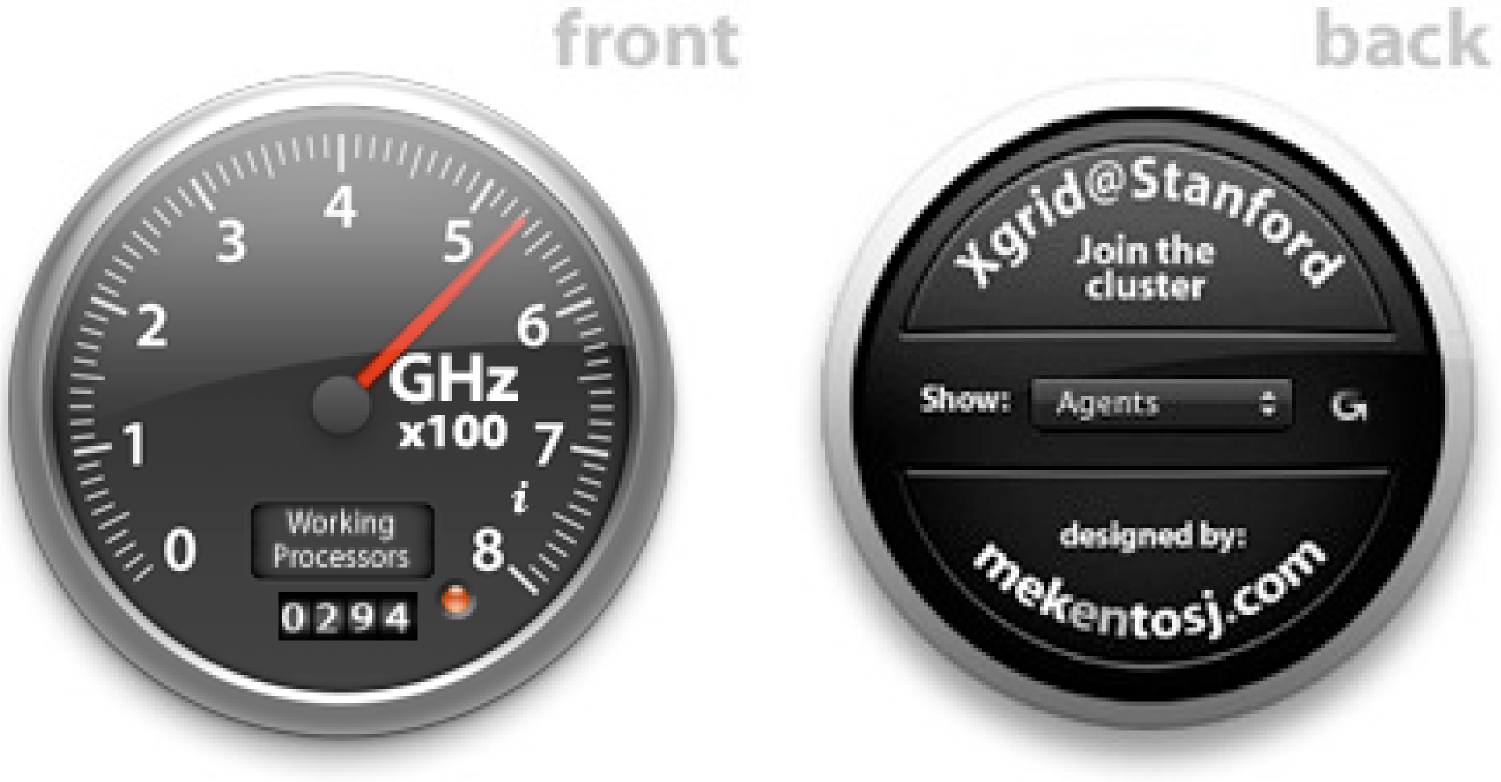
  - Requirements for using Xgrid with MPI applications
    - Open-MPI must be installed on all nodes
    - NFS shared work space where user ‘nobody’ has read/write permissions
    - Set MPI path, e.g. ‘export PATH =/usr/local/ompi/bin:$PATH’
    - Submit Xgrid MPI job using ‘mpirun’

Using Xgrid with serial applications

```
$> export XGRID_CONTROLLER_HOSTNAME=mycontroller.example.com
$> export XGRID_CONTROLLER_PASSWORD=pass

$> xgrid -job submit /usr/bin/cal 2005
{jobIdentifier = 24; }

$> xgrid -job list
{jobList = (24); }

$> xgrid -job attributes -id 24
{
  jobAttributes = {
      activeCPUPower = 2000;
      applicationIdentifier = “com.apple.xgrid.cli”;
      dateNow = 2005-06-24 16:58:32 +0200;
      dateStarted = 2005-06-24 16:58:28 +0200;
      dateSubmitted = 2005-06-24 16:58:27 +0200;
      jobStatus = Running;
      name = “/usr/bin/cal”;
      percentDone = 0;
      taskCount = 0;
      undoneTaskCount = 1;
   };
}

$> xgrid -job results -id 24 -so job.out -se job.err -out job-outdir

$> xgrid -job delete -id 24
```

Xgrid
notable examples

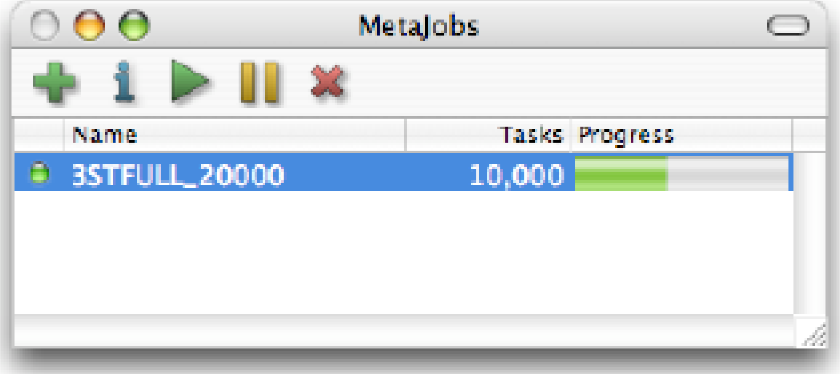
Xgrid@Stanford Widget



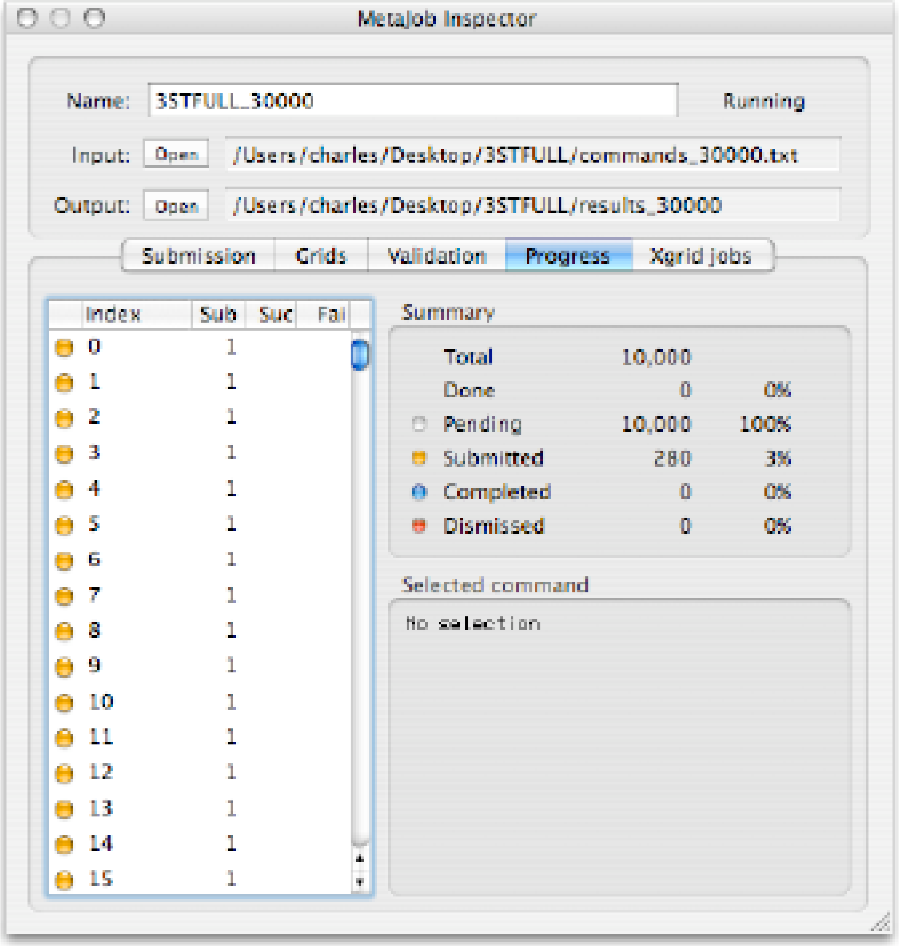
Running on a cluster of more than 500 computers connected to the internet somewhere in the world. 200 on average connected continuously. “This allows us to run a calculation in 1 week instead of a year!! The cluster is happily running past 200GHz” - Universal binary available

http://cmgm.stanford.edu/~cparnot/xgrid-stanford/

- Gridstuffer
  - Cocoa application to submit multi-task jobs
    - add MetaJob concept
      - several Xgrid tasks combined
      - tasks can run several times
      - be validated
      - rescheduled on failure
      - ...
  - GUI based
  - Uses Core Data to store jobs info
    - Can restart between reboots


MetaJobs

Name	Tasks	Progress
● 3STFULL_20000	10,000	


Metajob Inspector

Name: 3STFULL_30000 Running

Input: Open /Users/charles/Desktop/3STFULL/commands_30000.txt

Output: Open /Users/charles/Desktop/3STFULL/results_30000

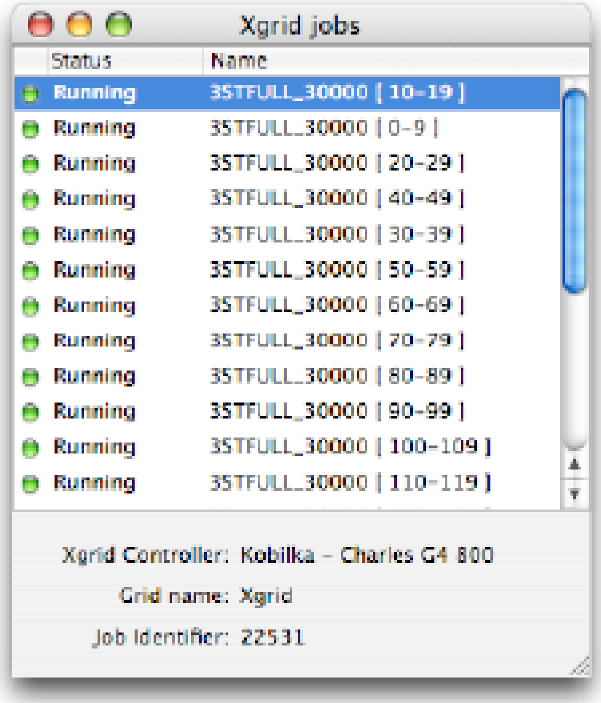
Index	Sub	Suc	Fai
● 0	1		
● 1	1		
● 2	1		
● 3	1		
● 4	1		
● 5	1		
● 6	1		
● 7	1		
● 8	1		
● 9	1		
● 10	1		
● 11	1		
● 12	1		
● 13	1		
● 14	1		
● 15	1		

Summary

Total	10,000	
Done	0	0%
○ Pending	10,000	100%
● Submitted	280	3%
● Completed	0	0%
● Dismissed	0	0%

Selected command

No selection


Xgrid jobs

Status	Name
● Running	3STFULL_30000 [10-19]
● Running	3STFULL_30000 [0-9]
● Running	3STFULL_30000 [20-29]
● Running	3STFULL_30000 [40-49]
● Running	3STFULL_30000 [30-39]
● Running	3STFULL_30000 [50-59]
● Running	3STFULL_30000 [60-69]
● Running	3STFULL_30000 [70-79]
● Running	3STFULL_30000 [80-89]
● Running	3STFULL_30000 [90-99]
● Running	3STFULL_30000 [100-109]
● Running	3STFULL_30000 [110-119]

Xgrid Controller: Kobilka – Charles G4 800

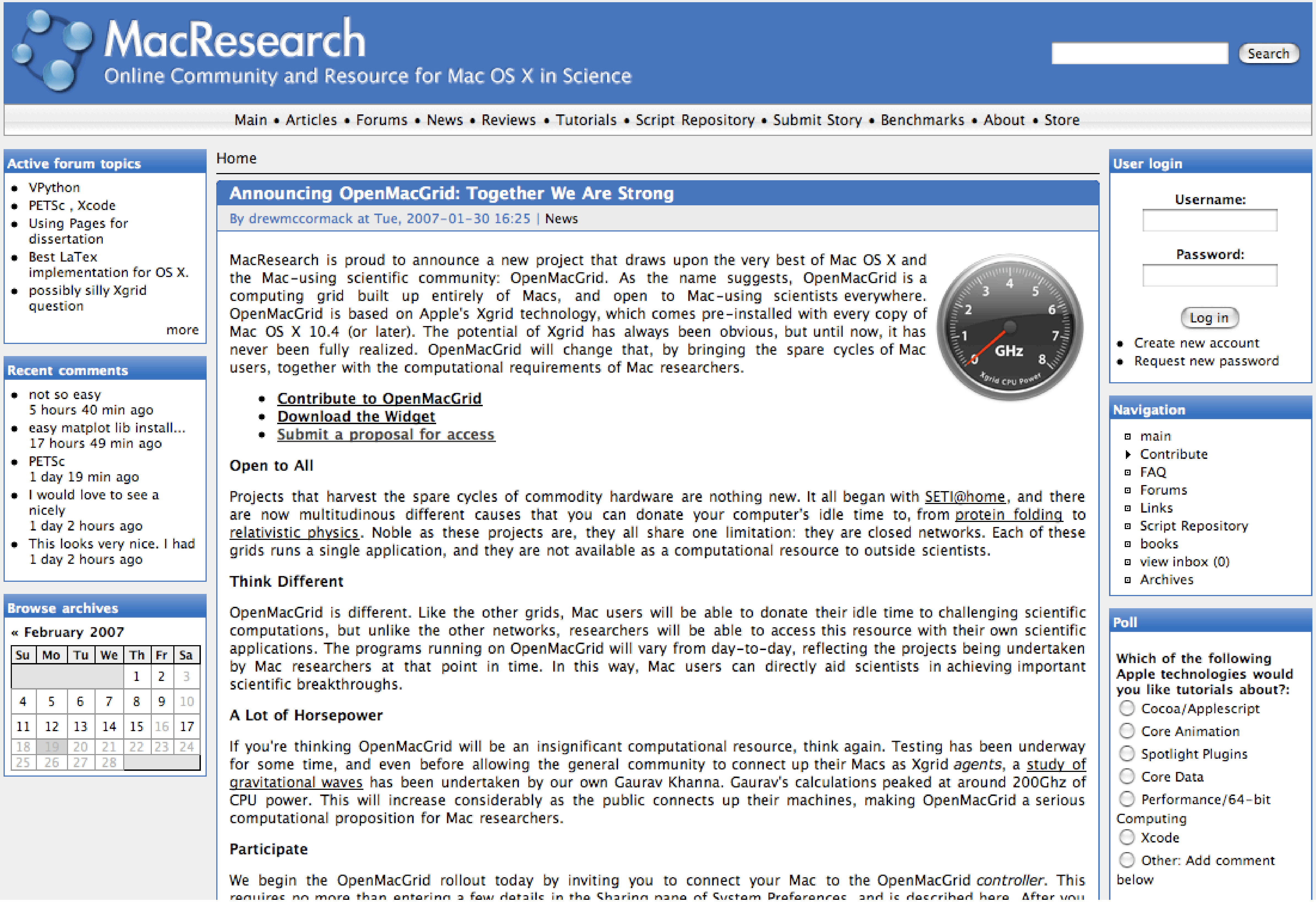
Grid name: Xgrid

Job Identifier: 22531

http://cmgm.stanford.edu/~cparnot/xgrid-stanford/html/goodies/GridStuffer-info.html

OpenMacGrid

over 1,000 GHz available already



Main • Articles • Forums • News • Reviews • Tutorials • Script Repository • Submit Story • Benchmarks • About • Store

Active forum topics

- VPython
- PETSc , Xcode
- Using Pages for dissertation
- Best LaTex implementation for OS X.
- possibly silly Xgrid question

more

Recent comments

- not so easy
  5 hours 40 min ago
- easy matplot lib install...
  17 hours 49 min ago
- PETSc
  1 day 19 min ago
- I would love to see a nicely
  1 day 2 hours ago
- This looks very nice. I had
  1 day 2 hours ago

Browse archives

« February 2007

Su	Mo	Tu	We	Th	Fr	Sa
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28			

Home

Announcing OpenMacGrid: Together We Are Strong

By drewmccormack at Tue, 2007-01-30 16:25 | News

MacResearch is proud to announce a new project that draws upon the very best of Mac OS X and the Mac-using scientific community: OpenMacGrid. As the name suggests, OpenMacGrid is a computing grid built up entirely of Macs, and open to Mac-using scientists everywhere. OpenMacGrid is based on Apple's Xgrid technology, which comes pre-installed with every copy of Mac OS X 10.4 (or later). The potential of Xgrid has always been obvious, but until now, it has never been fully realized. OpenMacGrid will change that, by bringing the spare cycles of Mac users, together with the computational requirements of Mac researchers.

- Contribute to OpenMacGrid
- Download the Widget
- Submit a proposal for access

Open to All

Projects that harvest the spare cycles of commodity hardware are nothing new. It all began with SETI@home, and there are now multitudinous different causes that you can donate your computer's idle time to, from protein folding to relativistic physics. Noble as these projects are, they all share one limitation: they are closed networks. Each of these grids runs a single application, and they are not available as a computational resource to outside scientists.

Think Different

OpenMacGrid is different. Like the other grids, Mac users will be able to donate their idle time to challenging scientific computations, but unlike the other networks, researchers will be able to access this resource with their own scientific applications. The programs running on OpenMacGrid will vary from day-to-day, reflecting the projects being undertaken by Mac researchers at that point in time. In this way, Mac users can directly aid scientists in achieving important scientific breakthroughs.

A Lot of Horsepower

If you're thinking OpenMacGrid will be an insignificant computational resource, think again. Testing has been underway for some time, and even before allowing the general community to connect up their Macs as Xgrid *agents*, a study of gravitational waves has been undertaken by our own Gaurav Khanna. Gaurav's calculations peaked at around 200Ghz of CPU power. This will increase considerably as the public connects up their machines, making OpenMacGrid a serious computational proposition for Mac researchers.

Participate

We begin the OpenMacGrid rollout today by inviting you to connect your Mac to the OpenMacGrid *controller*. This requires no more than entering a few details in the Sharing pane of System Preferences, and is described here. After you

User login

Username:

Password:

Log in

- Create new account
- Request new password

Navigation

- main
- Contribute
- FAQ
- Forums
- Links
- Script Repository
- books
- view inbox (0)
- Archives

Poll

Which of the following Apple technologies would you like tutorials about?:

○ Cocoa/Applescript
○ Core Animation
○ Spotlight Plugins
○ Core Data
○ Performance/64-bit Computing
○ Xcode
○ Other: Add comment below

KDSI Kentucky Grid

Lowering the technology barrier

- First such collaboration between K-12 schools and a university lab in the U.S.
  - Macs located at over 40 K-12 school districts
  - No special IT infrastructure in place but schools' own
  - Macs in schools screen millions of chemical compounds for lung, prostate, and breast cancer therapeutics - daily

"But — and this is where the schools' computers come in — it's a linear relationship: If you use 100 machines, you get results 100 times faster," Trent continues. "We have more than 1000s machines, so we can work more than 1000 times faster."

The Kentucky Cabinet for Economic Development, through the Department of Commercialization and Innovation, has supported the Kentucky Dataseam Initiative with over $2 million in grants.

www.apple.com/education/profiles/louisville/index.html

Xgrid RL examples

- Spatial biogeochemical modeling and sensitivity analysis: University of Wisconsin
- Natural Language Processing
- Cryptography and Monte Carlo molecular transport
- Black Hole Astrophysics & Quantum Cosmology - UMass, Dartmouth
- Low autocorrelation binary sequences - Fraser University, Burnaby, British Columbia
- XGrid BLAST - Genentech
- "Jet3D": Jet noise prediction code - NASA Langley Research Center, Hampton, Va.
- Military command and control research - Australian Department of Defence
- AstroVision's Xgrid enabled cluster - live satellite image processing
- Numerical relativity, fluid dynamics and scientific visualization - Nemeaux Xgrid cluster, LSU
- OpenMacGrid - over 1THz (1,000GHz) reached. Open to everyone. Macresearch.org

Google: about 150,000 for "Xgrid research" - about 411,000 for "Xgrid"

Documentation

- The primary Xgrid documentation is the Xgrid Administration manual for Mac OS X Server:
  - http://images.apple.com/server/pdfs/Xgrid_Admin_v10.4.pdf
- The ADC Developer library contains a reference description of the Xgrid Foundation API for Cocoa developers:
  - http://developer.apple.com/documentation/Performance/Conceptual/XgridDeveloper/index.html
- In addition to this FAQ, there are numerous Apple web sites that deal with Xgrid:
  - http://www.apple.com/macosx/features/xgrid/
  - http://www.apple.com/server/macosx/features/xgrid.html
  - http://developer.apple.com/hardware/hpc/
  - http://www.apple.com/science/solutions/clustercomputingresources.html
- There are also man pages for the command-line tools:
  - $ man xgrid # submit and monitor jobs and results
  - $ man xgridctl # adminster xgrid daemons
- The 'xgrid' man page in particular contains a detailed description of keys used by the job specification.

Documentation elsewhere

Xgrid, a 'just do it' grid solution

- MacResearch has an good tutorial by Charles Parnot:
  - http://www.macresearch.org/the_xgrid_tutorials

There are several helpful third-party sites that discuss Xgrid

- though they may not be completely accurate or updated

  http://www.macdevcenter.com/pub/a/mac/2005/08/23/xgrid.html

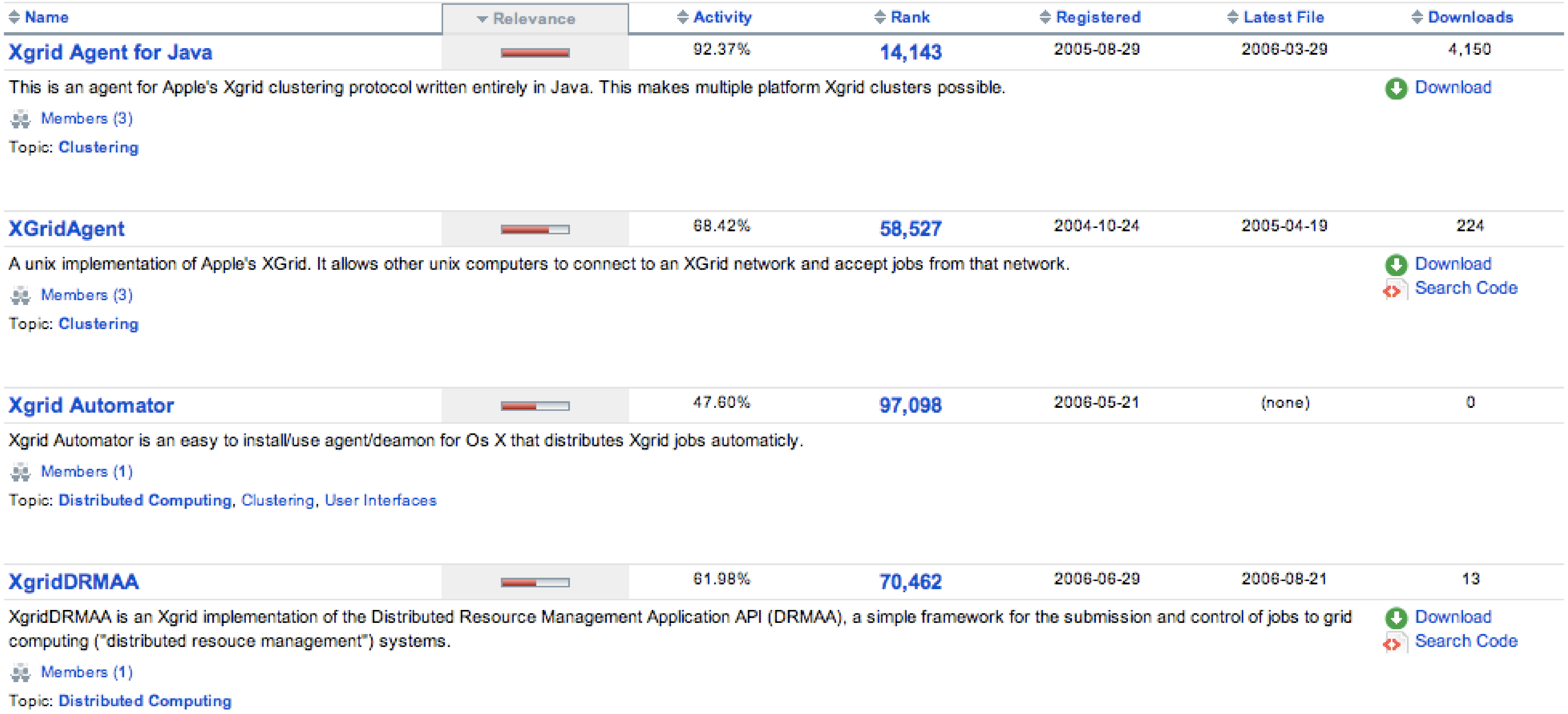
  http://www.macos.utah.edu/Documentation/xgrid/

  http://pyxg.scipy.org

  http://cmgm.stanford.edu/~cparnot/xgrid-stanford/

  http://unu.novajo.ca/simple/

Xgrid on sourceforge



◆ Name	▼ Relevance	◆ Activity	◆ Rank	◆ Registered	◆ Latest File	◆ Downloads
Xgrid Agent for Java This is an agent for Apple's Xgrid clustering protocol written entirely in Java. This makes multiple platform Xgrid clusters possible. Members (3) Topic: Clustering		92.37%	14,143	2005-08-29	2006-03-29	4,150 Download
XGridAgent A unix implementation of Apple's XGrid. It allows other unix computers to connect to an XGrid network and accept jobs from that network. Members (3) Topic: Clustering		68.42%	58,527	2004-10-24	2005-04-19	224 Download Search Code
Xgrid Automator Xgrid Automator is an easy to install/use agent/deamon for Os X that distributes Xgrid jobs automaticly. Members (1) Topic: Distributed Computing, Clustering, User Interfaces		47.60%	97,098	2006-05-21	(none)	0
XgridDRMAA XgridDRMAA is an Xgrid implementation of the Distributed Resource Management Application API (DRMAA), a simple framework for the submission and control of jobs to grid computing ("distributed resouce management") systems. Members (1) Topic: Distributed Computing		61.98%	70,462	2006-06-29	2006-08-21	13 Download Search Code

Q&A

...and thank you!