OSCAR Grid Engine Package
a.k.a
Build a SGE-HPC Cluster in Minutes

Babu Sundaram

High Performance Computing and Tools group

Computer Science Department

University of Houston

Mentor:

Bernard Li, Chair, OSCAR

Motivation

- UH is a CoE in Geosciences
- Interactions with core developers of GE
- Interactions with Oil Sector Companies
- Exploit idling cluster resources
- Google Summer of Code 2005

Sun Grid Engine

- Distributed resource management and batch job queuing software
- Increase cluster utilization to maximum
- Widely deployed at major institutions
  - UH HPCC has a SGE-managed cluster
- gridengine.sunsource.net

Overview of the Presentation

- Introduction to Summer of Code and OSCAR
- Motivations for SGE / OSCAR
- OSCAR framework and Internals
- Details of the work in SGE package creation
- Software, implementation information
- How to install and use the SGE Package

Google Code

Google Code Home

APIs

Open Source Projects

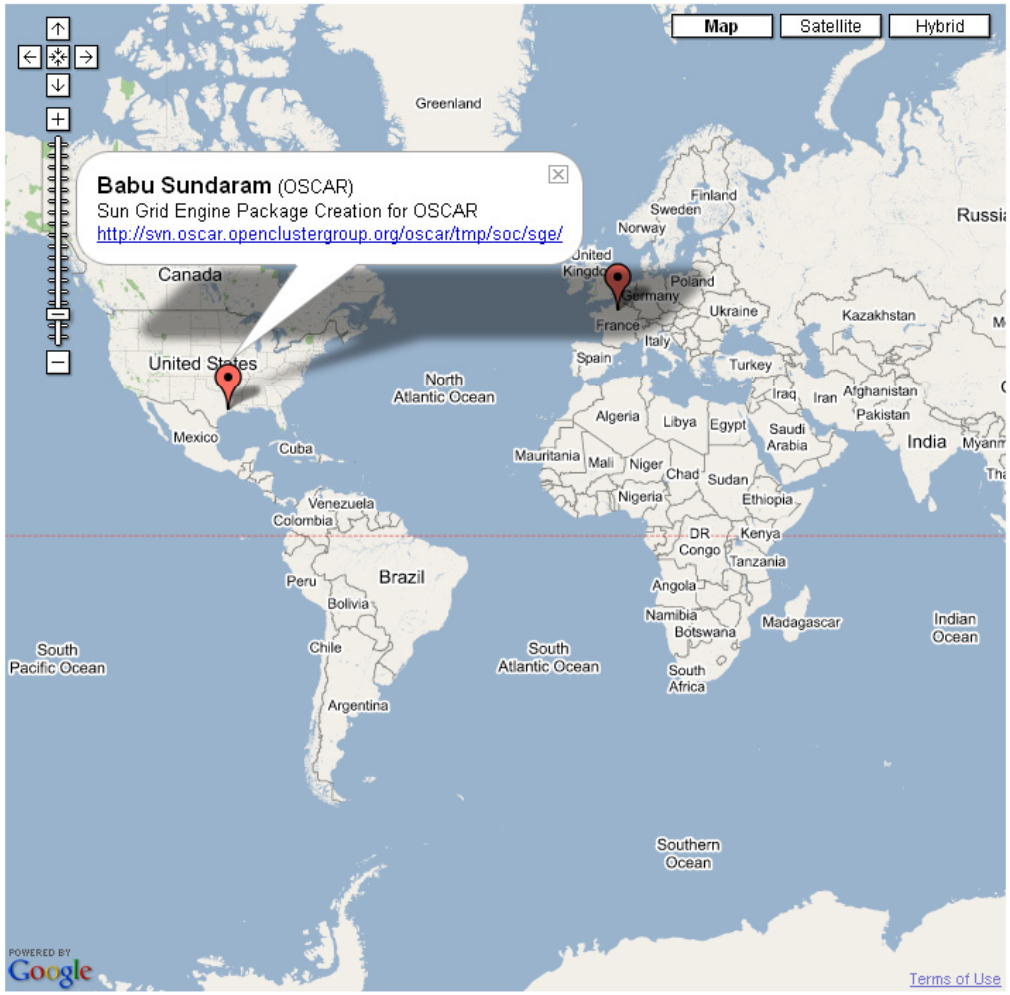
Organizations

FAQ

Summer of Code

Summer of Code: Where are they now?

To show you the global reach of the Summer of Code, we've put together a map showing the locations of both our students and mentors around the world (or at least the students and mentors that told us where there are!). Click the checkboxes to see which project participants are where, and click any marker to find out who it belongs to.



Show: All Mentors Students

- [ ] **Toggle all**
- [ ] Apache
- [ ] Blender
- [ ] Codehaus
- [x] OSCAR
- [ ] Drupal
- [ ] Fedora
- [ ] FreeBSD
- [ ] Gaim
- [ ] Gallery
- [ ] Google
- [ ] Handhelds
- [ ] Horde
- [ ] Inkscape
- [ ] Internet2
- [ ] JXTA
- [ ] Jabber
- [ ] Joomla
- [ ] KDE
- [ ] LispNYC
- [ ] LiveJournal
- [ ] Monotone
- [ ] Mozdev
- [ ] NMap
- [ ] NetBSD
- [ ] OpenOffice
- [ ] Other
- [ ] PSU
- [ ] Project Looking Glass
- [ ] Python
- [ ] Samba
- [ ] Semedia
- [ ] GNOME
- [ ] Mono
- [ ] Perl
- [ ] Subversion
- [ ] Wine
- [ ] Ubuntu
- [ ] WinLibre
- [ ] XWiki

☐ **Map Students to Mentors**
(Not all lines will be shown since not all students and mentors are mapped)

The OSCAR Project

- *"…a snapshot of the best known methods for building, programming, and using HPC clusters"*
- Easy to install software bundle
- Everything needed to install, build, maintain and use a Linux cluster
- Supports various distros such as Red Hat Enterprise Linux (and clones), Fedora Core, Mandriva Linux on x86, ia64, x86_64 architectures
- Debian port was another SoC project

OSCAR Latest Release – 4.2

- http://oscar.openclustergroup.org/news.oscar-4.2
- Many enhancements over previous releases
  - x86_64 support for RHEL 3 and 4
  - Support for FC3, RHEL 4 and Mandriva 10.1
  - atftp-server for better scalability
  - Support for more recent versions of Ganglia (v3.0.1), PVM (v3.4.5+4), Torque(v1.2.0p5), SystemImager(v3.5.3), MPICH(v1.2.7)
  - …

Motivations for SGE-OSCAR

- Medium-to-large sized clusters provide an inexpensive HPC platform
- Identical installation and configuration operations get repeated on all client nodes
- Increased productivity if the usual installation and maintenance tasks can be automated
- Resource management systems are important to operate clusters; SGE support in OSCAR

OSCAR framework

- Simple-to-manage package structure to easily and flexibly install and manage software
- OSCAR cluster = Server/Head node + Client nodes
- Model fits most of high-performance computing software for clusters – MPICH, SGE, C3, …
- System Installation suite (SIS) used as the cluster installation mechanism (http://sisuite.org)
  - High-quality, open source…
  - Heterogeneous client support, No need for pre-installed linux on clients
  - Fine-grained control

OSCAR Head node

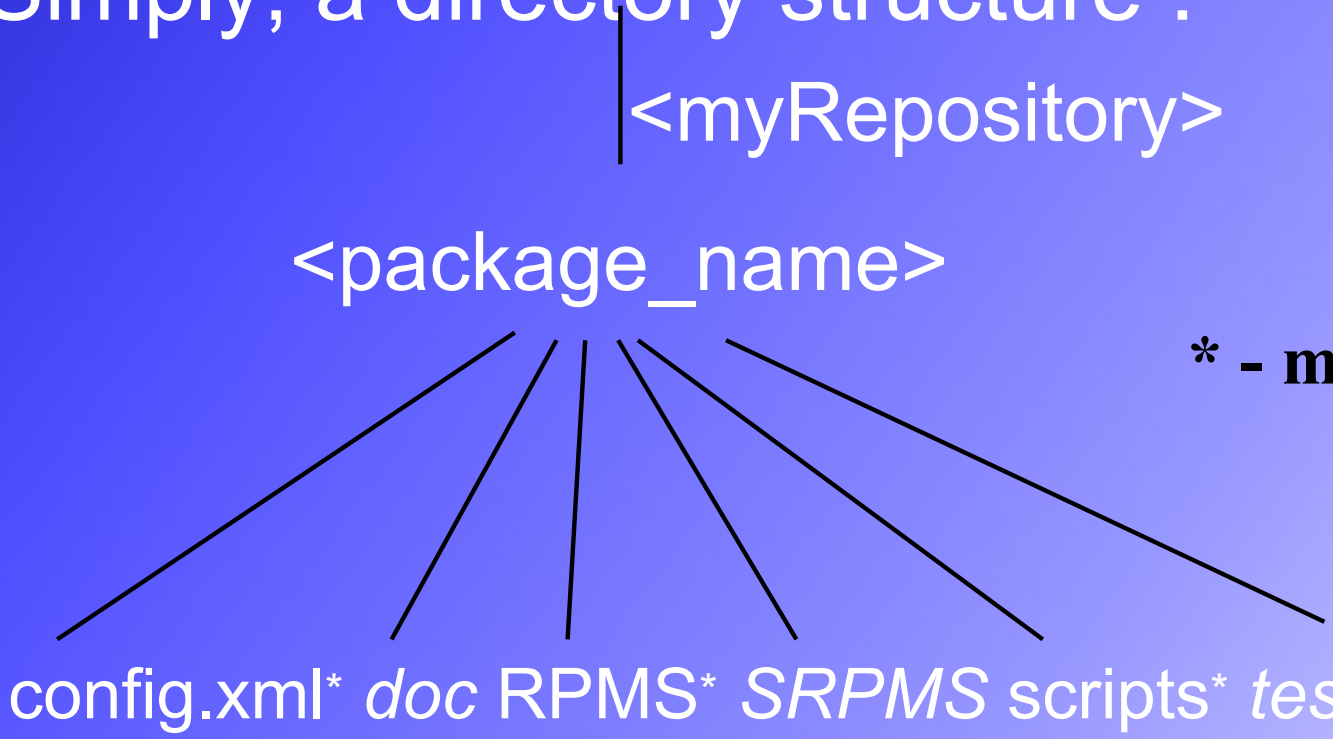
- Any supported Linux distro installed
  - A basic workstation-type install will suffice
- All RPMs are found under /tftpboot/rpm
- Later, images are built for clients
- Acts as Boot (image) server for the clients and services their requests
- Can also act as dhcp server
- Packages can be added / modified on an existing OSCAR head node and later propagated to clients

OSCAR clients

- Should have PXE-enabled BIOS or support AutoInstall CD/etherboot on floppy
- They boot over the network and image is obtained from the image server (head node)
- The images contain the requisite software portions for the clients
- Config steps are propagated to the clients

Whats in an OSCAR package?

- Simply, a directory structure :



*** - mandatory**

OSCAR directory structure

- config.xml – XML file indicating package details, its version, dependencies (e.g., sge, ksh) and OS-, client-specific rpmlists
- doc – Mostly help and README files
- RPMS – pre-compiled binaries as RPMs
- SRPMS – to allow building on other platforms
- Testing – tests after package installation

OSCAR *scripts*

- OSCAR framework recognizes a standard set of scripts and they have definitive purpose

Seq#	Script Name	Description
1	setup	Perform any package setup
2	pre_configure	Prepare package config (dynamic user input)
3	post_configure	Process results from package config
4	post_server_rpm_install	Perform "out of RPM" operations on server
5	post_client_rpm_install	Perform "out of RPM" operations on client
6	post_clients	Perform configurations with knowledge about cluster nodes
7	post_install	Perform final config with fully install/booted cluster nodes

SGE Package for OSCAR

- Lot of interest for SGE OSCAR package
- SGE fits the standard “master node – worker nodes” model
  - qmaster orchestrates the cluster and assigns jobs to execd nodes
- Wont it be nice to get a functioning SGE installation in a matter of few minutes?
  - Of course, you can fix queues and alter configurations anytime

Tasks in SGE package creation

- Source RPM generation
- Binary RPM generation
  - Server-, client- and GUI-specific RPMs
- Develop OSCAR configuration and scripts
- Implementation, Licensing, Documentation

RPM generation for SGE

- Source RPM generation was our first step
- SGE source rpm for version 6.0 update 4
  - At that time, ScalableSystems had a release ready
  - Now, we have SRPM and RPM based on update 6
- Some patches were identified earlier on and some were added later for correct compilation
  - qtcsh, inst_sge, aimk, distinst, qmon icons
- Spec file modification and SGE binary RPM generation

Scripts for SGE-OSCAR

- Automates SGE install on the OSCAR cluster
- All perl scripts; cd sge/scripts/; ls
- post_server_install
  - Configures the overall SGE setup; Sets up SGE master with various values for the options
    - **SGE_ROOT, CELLNAME, FULLSERVER, GIDRANGE, SPOOLTYPE, PORTS…**
    - **myInstall.conf is a file that gets generated at this stage to drive "inst_sge –m –auto"**
  - User input/customization happens at this stage (configurator.html)
  - At the end of this step, the qmaster is up and running on the OSCAR head node
- post_clients
  - Gets executed after clients are defined (not installed)
  - Adds exec clients as admin hosts; Needed from SGE POV
  - **get_machine_listing();** then, **qconf –ah $hostname;**

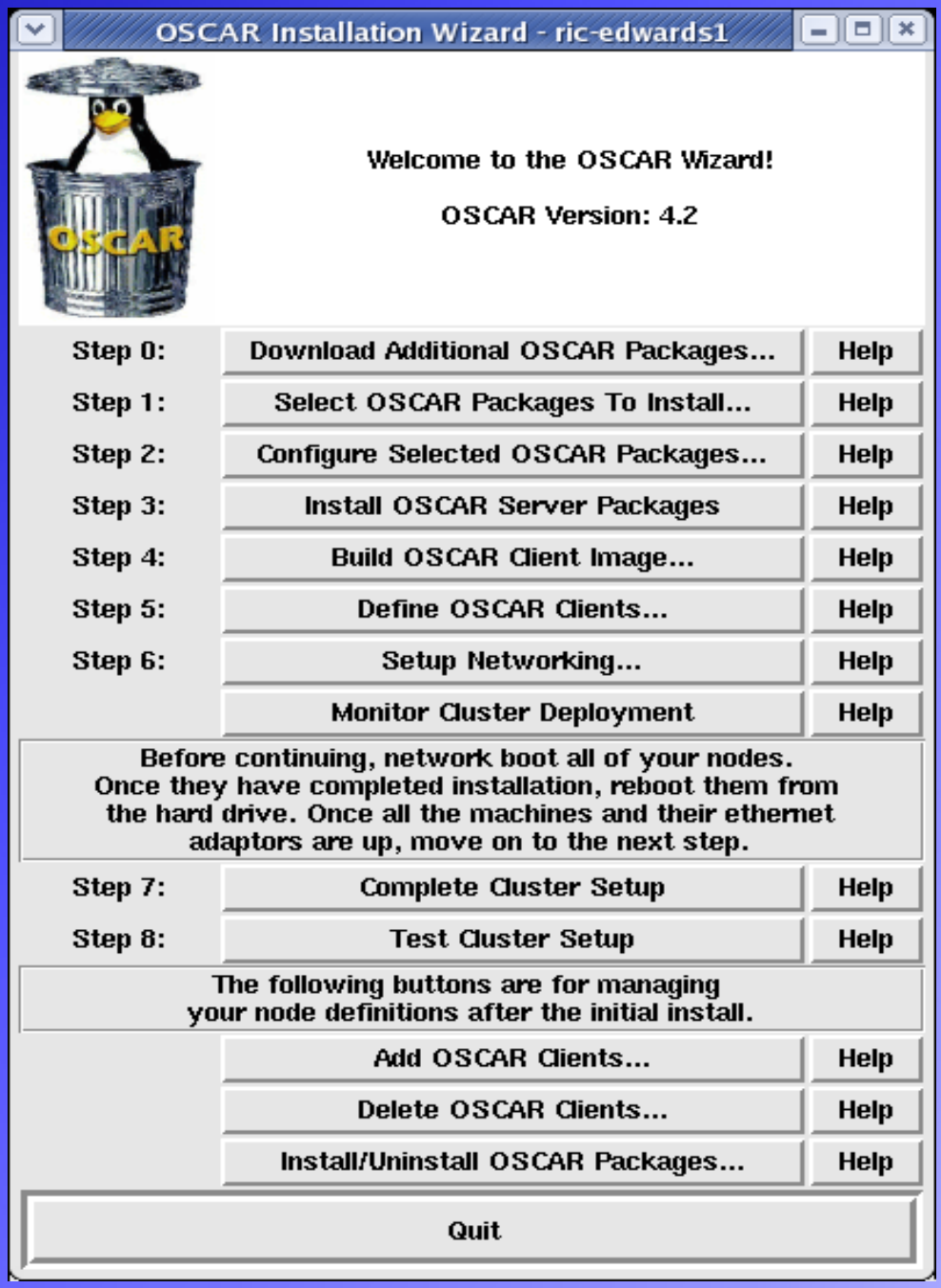
SGE OSCAR scripts – cont…

- post_install
  - All actions that can be done only after a full cluster install happen in this step
  - qmaster already knows about the clients (from the definition step) and they are already admin hosts
  - All settings (dir: cell_name) gets tarred and ready to get pushed to the clients during post_install
  - Cannot assume NFS; So, the *cell_name_dir*.tar gets pushed to the clients and untarred
  - Clients now know about the qmaster details
  - Automated install of inst_sge –x (patched in spec); Executed via cexec over ssh
  - Checks for parallel environment availability and adds PE support to SGE
    - Right now, checks for LAM/MPI availability only; Will add support for all PEs soon

- post_server_rpm_uninstall, post_client_rpm_uninstall
  - Not much SGE-specific functionality, but there to allow clean SGE uninstall

Implementation details

- OSCAR's Subversion repository for code revision control
  - Current package is available in the trunk
- Initial implementation was on FC2, x86
  - Tested on RHEL3 and FC3; 64-bit RPMs
- Basic tools involved: rpm, make, perl, tk, diff/patch
- OSCAR-specific code is under GPL; SGE under SISSL


OSCAR Installation Wizard - ric-edwards1
OSCAR
Welcome to the OSCAR Wizard!
OSCAR Version: 4.2
Step 0:
Download Additional OSCAR Packages...
Help
Step 1:
Select OSCAR Packages To Install...
Help
Step 2:
Configure Selected OSCAR Packages...
Help
Step 3:
Install OSCAR Server Packages
Help
Step 4:
Build OSCAR Client Image...
Help
Step 5:
Define OSCAR Clients...
Help
Step 6:
Setup Networking...
Help
Monitor Cluster Deployment
Help
Before continuing, network boot all of your nodes.
Once they have completed installation, reboot them from
the hard drive. Once all the machines and their ethernet
adaptors are up, move on to the next step.
Step 7:
Complete Cluster Setup
Help
Step 8:
Test Cluster Setup
Help
The following buttons are for managing
your node definitions after the initial install.
Add OSCAR Clients...
Help
Delete OSCAR Clients...
Help
Install/Uninstall OSCAR Packages...
Help
Quit

RPM installation,OSCAR installation; core, non-core

Choose additional packages; customize them and install;

Install packages and build image for clients;




HPCTools

Google Summer of Code 2005

OSCAR open source cluster application resources

OSCAR head node; runs sge_master and sge_sched d

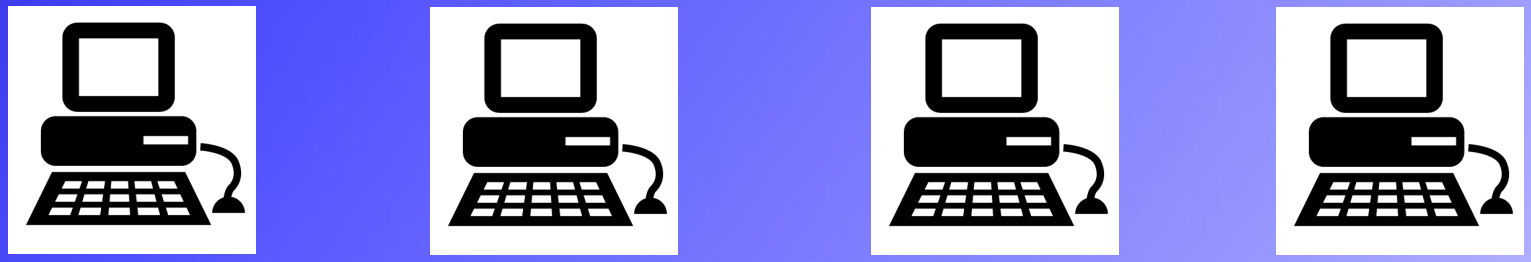
Clients are defined; qmaster adds them as admin hosts

Tars up the cell_name dir and stores for pushing to clients




Qmaster does inst_sge –x on clients after full boot

SGE setup is ready to run jobs; Do qsub ☺



All are exec hosts in the cluster; run sge_execd

Current Work

- Support for 64 bit architectures
- Support for parallel environment integrations
- Add better user- and admin- level testing for SGE;
  - Mimic SGE tests via OSCAR
- Porting OSCAR to other OS (Solaris10?)

Acknowledgements

- Google Inc.,
- OSCAR developers
- SGE developers (Ron, Fritz, Andreas…)
- Chandler Wilkerson, LAN admin, CS, UH
- Dr. Barbara M. Chapman
- ScalableSystems