Building and Refining General Purpose Computing Clusters in an Emerging HPC Oriented Research Environment

Albert Gazendam
agazendam@csir.co.za

9 June 2008
Overview

- South African HPC environment
- HPC infrastructure and OSCAR market share
- Describing the typical challenges
- Highlighting solutions to three of these
  - Comparing vendor offers
  - Partial disablement of SSH
  - Special group accounts
- Conclusion
South African HPC environment

- Many legacy SMP and vector machines collecting dust
- Major upsurge in interest and activity since early 2000's
- Currently a $10m per annum market for hardware vendors
- Set to grow to $100m per annum in the next five years
- Primarily used by scientific research community
HPC infrastructure and OSCAR market share

- One national HPC facility, CHPC
  - 2.5Tflops computing cluster: IBM software stack
  - Power4+ based 32 way SMPs
  - BlueGene/L (single cabinet) on the way
- Major facilities at CSIR and several universities
  - C4: 3 x OSCAR based computing clusters
  - UCT, UOFS, UP, etc. with substantial OSCAR based computing clusters
- OSCAR run on around 50% of the HPC clusters
Africa's largest OSCAR deployment

This diagram shows the various components of C4 and their relative positions in the high-level network architecture.
Describing the typical challenges

- **Before installed**
  - Securing funding
  - Comparing offers from competing vendors 😊

- **Once installed**
  - Management of user accounts
  - Simplifying deployment of common apps and libraries
  - Encouraging users to use the job queues 😊
  - Empowering users to 'own' and 'share' their software 😊
1. Comparing vendor offers

- Remove price as variable
- Performance: commitment on HPCC results
- Weighted comparison

\[
Score = \sum_{i=1}^{n} e \left( \frac{x_i}{\max_k x_{k,i}} \right) - 1 \times w_i
\]

Where \( k \) is collection of systems being compared and \( n \) is the number of metrics considered

- Useful weighting set:

<table>
<thead>
<tr>
<th>Element</th>
<th>Symbol</th>
<th>Weighting</th>
<th>Score</th>
<th>Unit in HPCC v1.2.0 output file (hpccoutf.txt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-HPL</td>
<td>( w_1 )</td>
<td>40%</td>
<td>( x_1 )</td>
<td>HPL_Tflops</td>
</tr>
<tr>
<td>G-FFTE</td>
<td>( w_2 )</td>
<td>30%</td>
<td>( x_2 )</td>
<td>MPIFFT_Gflops</td>
</tr>
<tr>
<td>G-RandomAccess</td>
<td>( w_3 )</td>
<td>30%</td>
<td>( x_3 )</td>
<td>MPIRandomAccess_GUPs</td>
</tr>
</tbody>
</table>
...demonstrated

System offered by Vendor A:
- G-HPL = 2.9 Tflops
- G-FFTE = 55 Gflops
- G-RandomAccess = 0.0045 GUPS

System offered by Vendor B:
- G-HPL = 2.3 Tflops
- G-FFTE = 65 Gflops
- G-RandomAccess = 0.0052 GUPS

System offered by Vendor C:
- G-HPL = 2.6 Tflops
- G-FFTE = 53 Gflops
- G-RandomAccess = 0.0065 GUPS

Weighted scores:
- A = 85.34
- B = 85.65
- C = 85.78 😞
2. Partial disablement of SSH

- Problem: User SSHing to compute nodes directly and running their software by hand
- Solution: `chmod o-x /usr/bin/ssh`
- Issue: Job manager uses SSH in the background to launch jobs from the queues
- Trick: Create special `/etc/sudoers` entries and add wrappers to job launching mechanisms of the job manager, thereby enabling the job manager to use SSH (still as the user)
1. Deactivate SSH for normal users
   chmod o-x /usr/bin/ssh

2. Use ‘visudo’ to add lines to /etc/sudoers that allows
   <USERNAME> to use ‘sudo ssh’ under specific conditions
   <USERNAME> ALL=NOPASSWD:/usr/bin/ssh -X
   <USERNAME>@<NODEMASK>
   <USERNAME> ALL=NOPASSWD:/usr/bin/ssh -X
   -p*<USERNAME>@<NODEMASK> exec
   /opt/sge/utilbin/1x26-amd64/qrsh_starter
   /opt/sge/default/spool/<NODEMASK>

3. Create /opt/sge/utilbin/qlogin_wrapper to look as follows:
   #!/bin/sh
   HOST=$1
   sudo /usr/bin/ssh -X 'whoami'@$HOST

4. Create /opt/sge/utilbin/qrsh_wrapper to look as follows:
   #!/bin/sh
   /usr/bin/sudo /usr/bin/ssh -X $1 $2
   'whoami'@$3 $4 $5 $6

5. Modify the SGE queue configuration with ‘qconf -mconf’
   qlogin_command /opt/sge/utilbin/qlogin_wrapper
   qlogin_daemon /usr/sbin/sshd -i
   rlogin_daemon /usr/sbin/sshd -i
   rsh_daemon /usr/sbin/sshd -i
   rlogin_command /usr/bin/ssh -x
   rsh_command /opt/sge/utilbin/qrsh_wrapper

6. Set the root account’s public key as an authorized key
   # of <USERNAME>
   cat /root/.ssh/id_rsa.pub >>
   /home/<USERNAME>/.ssh/authorized_keys
3. Special group accounts

- When software is of potential benefit to several users
- Create special group account and assign an administrator to it
- The administrator gets SSH keys to allow entry to the special group account
- The administrator can manage group membership with `gpasswd`
- Group members can benefit from the efforts of the group administrator and other group members
...demonstrated

**Typical scenario**

- `/home/<user_1>`
  - `software_1`
  - `software_2`
  - `dataset_A`
  - `dataset_B`

- `/home/<user_2>`
  - `software_2`
  - `software_3`
  - `dataset_A`
  - `dataset_C`

- `/home/<user_3>`
  - `software_1`
  - `software_3`
  - `dataset_B`
  - `dataset_C`

**Conventional approach**

```
gpasswd -a <user_3> <user_1>
gpasswd -a <user_1> <user_2>
gpasswd -a <user_2> <user_3>
```
...demonstrated

### Users

### Administration

### Special group accounts

- `/home/<group_1>/software_1` 750
- `/home/<group_1>/dataset_A` 740
- `/home/<group_2>/software_2` 750
- `/home/<group_2>/dataset_B` 740
- `/home/<group_3>/software_3` 750
- `/home/<group_2>/dataset_C` 740
Conclusion

- OSCAR has gained substantial market share in South Africa
- The relatively immaturity of emerging HPC communities are characterised by:
  - Limited vendor insight
  - Undisciplined users
  - Poor support structures for users
- Practical solutions were presented
Questions?