Partnership for Advanced Computing in Europe

PRACE overview and Lindgren Introduction

Dr. Lilit Axner

SNIC coordinator in PRACE Infrastructure
Labs and Lunches

Restaurant: Syster o Bror

Names of Lab rooms: Red and (Orange)
Dinner:

When: 4th of December at 18:30

Where: Fem Små Hus (at Gamla Stan T-bana station)

NOTE: Registration is required
Realizing the ESFRI Vision for an HPC RI

- **European HPC-facilities at the top of an HPC provisioning pyramid**
  - Tier-0: European Centres for Petaflop/s
  - Tier-1: National Centres
  - Tier-2: Regional/University Centres

- **Creation of a European HPC ecosystem**
  - HPC service providers on all tiers
  - Scientific and industrial user communities
  - The European HPC hard- and software industry
  - Other e-Infrastructures

Slides from PRACE office
PRACE-1IP: a Cornerstone in PRACE History

HPC part of the ESFRI Roadmap; creation of a vision involving 15 European countries

- Creation of the Scientific Case
- Signature of the MoU
- Creation of the PRACE Research Infrastructure

Timeline:
- 2004: HPCEUR
- 2005: HET
- 2006-2007: PRACE Initiative
- 2008-2009: PRACE Preparatory Phase Project
- 2010: PRACE-1IP
- 2011-2012: PRACE-2IP
- 2013: PRACE-3IP

Slides from PRACE office
April, 23rd 2010 creation of the legal entity (AISBL) PRACE with seat location in Brussels, Belgium

25 PRACE Members

67+ Million € from EC FP7 for preparatory and implementation phases
Grants INFSO-RI-211528, 261557, 283493, and 312763
Complemented by ~ 50 Million € from PRACE members

Slides from PRACE office
PRACE is building the top of the pyramid...

First production system available:
1 Petaflop/s
IBM BlueGene/P (JUGENE) at GCS (Gauss Centre for Supercomputing) partner FZJ (Forschungszentrum Jülich)

Second production system available:
Bull Bullx CURIE at GENCI partner CEA. Full capacity of 1.8 Petaflop/s reached by late 2011.

Third production system available by the end of 2011:
1 Petaflop/s
Cray (HERMIT) at GCS partner HLRS (High Performance Computing Center Stuttgart).

Fourth production system available by mid 2012:
3 Petaflop/s
IBM (SuperMUC) at GCS partner LRZ (Leibniz-Rechenzentrum).

Fifth production system available by August 2011:
1 Petaflop/s
IBM BG/Q (FERMI) at CINEC.

Sixth production system available by January 2013:
1 Petaflop/s
IBM (MareNostrum) at BSC.
The HPC European e-infrastructure (ESFRI)

**PRACE 1.0**

- **25 members, AISBL since 2010**
- **530 M€ for 2010-2015**
- **6 supercomputers in 4 hosting countries, different architectures**
- **research and industrial access (open R&D) for all disciplines based on excellence in science, free of charge**
- **Nearly 15 Pfoles/s**
- **5 billion hours granted since 2010**

Slides from PRACE office
PRACE open calls

There are three types of call:

• **Tire-0 call for applications** – open every 6 months
  To get only **CPU time** on one of the PRACE six **Tier-0** systems

• **Preparatory access calls (Type A, B, C)** – constantly open and a chance to get an **expert help** for up to 6 months to scale the code on **Tier-0** systems.

• **DECI (Tier-1) call for applications** – open every 6 months
  To get both **CPU time and expert help** up to 6 months on one of the many **Tier-1** systems across PRACE partners.
Outline

- Lindgren
- How to login
- Where to store files
- How to compile and run a program
Lindgren

- Cray XE 6 system
- Theoretical peak performance 305.6 TeraFLOPS
- 47.38 TB Memory (32 Gbyte per node)
- [http://www.pdc.kth.se/resources/computers/lindgren](http://www.pdc.kth.se/resources/computers/lindgren)
Lindgren Cabinet

- 3 chassis per cabinet
- 8 blades per chassis
- 4 nodes per blade
- 24 cores per node
- 16 Cabinets, 1516 compute nodes, 36,384 cores
Lindgren Nodes

- Login nodes — don’t run jobs here!
- Interactive nodes — for test runs; alone or shared among users
- Dedicated nodes (compute nodes) — for running final programs
Now this Cerberus had three heads of dogs, the tail of a dragon, and on his back the heads of all sorts of snakes.

- PDC uses **ssh together with kerberos** for login.
- Kerberos — **system for authenticating users and services on a network.**
- **Kerberos server**, trusted by users and services.
- A **Kerberos principal** (*username@NADA.KTH.SE*) is a user’s or service’s username for a certain **Kerberos realm** (*NADA.KTH.SE*).
Kerberos Commands

- `kinit` — proves your identity
- `klist` — list your Kerberos tickets
- `kdestroy` — destroy your Kerberos ticket file
- `kpasswd` — change your Kerberos password

```
kinit -f -l 7d username@NADA.KTH.SE
klist -Tf
```

Credentials cache: FILE:/tmp/krb5cc_500

```
Principal: username@NADA.KTH.SE

<table>
<thead>
<tr>
<th>Issued</th>
<th>Expires</th>
<th>Flags</th>
<th>Principal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar 25 09:45</td>
<td>Mar 25 19:45</td>
<td>FI</td>
<td>krbtgt/NADA.KTH.SE@NADA.KTH.SE</td>
</tr>
<tr>
<td>Mar 25 09:45</td>
<td>Mar 25 19:45</td>
<td></td>
<td>afs/pdc.kth.se@NADA.KTH.SE</td>
</tr>
</tbody>
</table>
```
Login using Kerberos tickets

- Get a **forwardable** Kerberos ticket on your **local system**:
  
  ```
  kinit -f -l 7d username@NADA.KTH.SE
  ```

- **Forward** your ticket via ssh:
  
  ```
  ssh username@lindgren.pdc.kth.se
  ```

- The prompt changes when logged in. Check your tickets on the **remote system**:
  
  ```
  klist -Tf
  ```

- You have reached the login node of the cluster
Login from any computer

- You can reach PDC from any computer and network!
- The Kerberos implementation Heimdal can be installed on most operating systems (Mac, Windows, Linux)
- An SSH command that knows GSSAPI key-exchange can forward Kerberos tickets
- Names of the commands depend on your operating system!
- Follow the instructions for your operating system: [http://www.pdc.kth.se/resources/software/login-1](http://www.pdc.kth.se/resources/software/login-1)
File Systems

- There is NO local disk on Lindgren
- Your **home directory** is located in AFS.
- **Oldfiles** contain ~, as it was yesterday, i.e. a **backup**.
- Save **your code** in AFS to get backup
- You can access AFS from any computer [http://www.pdc.kth.se/resources/software/file-transfer/](http://www.pdc.kth.se/resources/software/file-transfer/)
- You can not run jobs from AFS on Lindgren
Parallel file system

- Lustre — massively parallel distributed file system
- No backup
- No personal quota — move data off!
  
  /cfs/klemming/nobackup/u/username
  
  where username is your username and u is the first letter in your username.
  
  /cfs/klemming/scratch/u/username

- **Always start and run your programs in Lustre!**
Modules

- Modification of the user’s environment; handles PATH, MANPATH ...

  module list
  module show modulename
  module avail
  module load modulename
  module swap fromModulename toModulename

Common PDC modules are: PrgEnv-pgi, afsws and heimdal. Essential modules are preloaded at login.
Cray compilers and libraries

- Always use the wrappers `cc` (C code), `cc` (C++) or `ftn` (FORTRAN).
- The PrgEnv module loaded (PrgEnv-pgi, PrgEnv-intel or PrgEnv-gnu), decides if the PGI (default), Intel or GNU compiler is used by cc, CC and ftn.

```
module swap PrgEnv-pgi PrgEnv-gnu
```

- The wrappers **automatically** links with math libraries if their modules are loaded (no flags for linking required).

```
module load xt-libsci fftw
```

- to link your code against lapack, blas, scalapack, blacs, fftw...
How to run

- Not Possible to log into the nodes directly
- Always use aprun to launch jobs on the compute nodes
- Use aprun even if you want to run a serial job
Running jobs on Lindgren

- Get Kerberos tickets on your own computer
- Login on lindgren.pdc.kth.se
- Check the lifetime of your ticket
- Go to your directory /cfs/klemming/nobackup/u/username
- Compile:
  
  ```
  ftn hello_mpi.f90 -o hello_mpi_f90
  cc hello_mpi.c -o hello_mpi_c
  ```
- Submit your job:
  
  ```
  qsub ./mpi_job.pbs
  ```
- `mpi_job.pbs` is a **job script** that the queue can understand.
#PBS -N hello_mpi_job
# 5 minute wall-clock time will be given to this job
#PBS -l walltime=5:00
# Number of MPI tasks. Should always be a multiple of 24
# i.e. an entire node on Lindgren
#PBS -l nppwidth=24
# Change to the work directory
echo $PBS_O_WORKDIR
cd $PBS_O_WORKDIR
# Run the program, always use aprun!
aprun -n 12 ./hello_mpi

- If you forget aprun in the script, the program will start on the login node!
Finding Information

- `man`, `-help`, `-h`, `apropos` ...
- [http://www.pdc.kth.se/education/tutorials/summer-school](http://www.pdc.kth.se/education/tutorials/summer-school)
- Flash News
- Mailinglists
- Course assistants
- `support@pdc.kth.se`

www.pdc.kth.se
Thank you!