Introduction to PDC environment

Cristian Cira

PDC Center for High Performance Computing
KTH Royal Institute of Technology

PDC April 2016
Outline

1. PDC Overview
2. Infrastructure
   - Beskow
   - Tegner
3. Accounts
   - Time allocations
   - Authentication
4. Job submission and monitor
   - SLURM commands
5. Conclusions
SNIC
Swedish National Infrastructure for Computing

National research infrastructure that provides a balanced and cost-efficient set of resources and user support for large scale computation and data storage to meet the needs of researchers from all scientific disciplines and from all over Sweden (universities, university colleges, research institutes, etc).
Access to EU Facilities and Experts

- EUDAT
- PRACE
- CRESTA
- EPiGRAM
- EGI
- E-infrastructure

Cristian Cira (PDC)
PDC and Industry

Working with industrial researchers and developers on major international projects that push high-performance computing to the next level.

Recently established a **business development unit** that provides consultancy and HPC services to industries.
Broad Range of Training

**Summer School**  Introduction to HPC held every year

**Specific Courses**  Programming with GPGPU, Recent Advances in Distributed and Parallel Computing and/or Cloud Computing, Software Development Tools, etc

**PDC User Days**  PDC Open House and Pub Afternoon
First-Line Support and System Staff

First-line support
Provides specific assistance to PDC users
Answers general questions from the public

System staff
System managers/administrators ensure that computing and storage resources run smoothly and securely
Application Experts

Hold PhD degrees in different scientific fields and are experts in HPC. Together with researchers, they optimize, scale and enhance scientific codes for the next generation supercomputers.

Jaime Rosal Sandberg
Computational Chemistry

Henric Zazzi
Bioinformatics/Genetics

Jing Gong
Scientific Computing

Cristian Cira
Performance Analysis

Michael Djurfeldt
Computational Physics

Thor Wikfeldt
Computational Chemistry
Services

- Access to supercomputers
- HPC training
- Postgraduate degree projects
- Visualization
- Support
- Expertise in HPC software
- Access to international HPC facilities
- Data storage
Outline

1. PDC Overview
2. Infrastructure
   - Beskow
   - Tegner
3. Accounts
   - Time allocations
   - Authentication
4. Job submission and monitor
   - SLURM commands
5. Conclusions
What is a cluster?

- Cluster
- Racks
- Nodes
- Processors
- Cores
- Login nodes
- Compute nodes
- Dedicated nodes
- Transfer nodes
- Service nodes
**Beskow - Cray XC40 system**

**Fastest machine in Scandinavia**

- in production until Q4 2018
- 9 racks 1676 nodes
- Intel Xeon Processor E5-2698 v3
  40M Cache, 2.30 GHz
- 53.632 cores - 32 cores/node
- Aries Dragonfly network topology
- 104.7 TB memory - 64 GB/node
**Tegner**

*pre/post processing for Beskow*

- **5 x 2TB Fat nodes**
  - 4 x 12 core Ivy Bridge
  - 2TB RAM
  - 2 x Nvidia Quadro K420
  - Used for pre/post processing data
  - Not allocated through SNIC

- **5 x 1TB Fat nodes**
  - 4x 12 core Ivy Bridge
  - 1TB RAM
  - 2 x Nvidia Quadro K420
  - Only for academia within the Stockholm area
  - Has large RAM nodes
  - Has nodes with GPUs

- **55 Thin Nodes**
  - 2 x 12 core Haswell
  - 512GB RAM
  - Nvidia Quadro K420 GPU
  - Lifetime: Q4 2018

---

Cristian Cira  (PDC)
## Summary of PDC resources

<table>
<thead>
<tr>
<th>Computer</th>
<th>Beskow</th>
<th>Tegner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core/node</td>
<td>32</td>
<td>48/24</td>
</tr>
<tr>
<td>Nodes</td>
<td>1.676</td>
<td>50 x 24 Haswell/GPU</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 x 48 Ivy bridge</td>
</tr>
<tr>
<td>RAM (GB)</td>
<td>1.676 x 64</td>
<td>50 x 512</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 x 1000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 x 2000</td>
</tr>
<tr>
<td>Allocations (core hours)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>&lt; 5.000</td>
<td>&lt; 5.000</td>
</tr>
<tr>
<td>Medium</td>
<td>&lt; 200.000</td>
<td>&lt; 80.000</td>
</tr>
<tr>
<td>Large</td>
<td>≥ 200.000</td>
<td></td>
</tr>
<tr>
<td>Allocation via SNIC</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>AFS</td>
<td>login node only</td>
<td>yes</td>
</tr>
<tr>
<td>Lustre</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
File Systems

Andrew File System (AFS)
- Global distributed file system accessible from computer around the world running an AFS client
- Home directory
  /afs/pdc.kth.se/home/[initial]/[username]
- Access via Kerberos tickets and AFS tokens

Lustre
- Massively parallel distributed file system
- Very high performance
- NO backup
- NO personal quota
Outline

1. PDC Overview

2. Infrastructure
   - Beskow
   - Tegner

3. Accounts
   - Time allocations
   - Authentication

4. Job submission and monitor
   - SLURM commands

5. Conclusions
Access requirements

User account  either SUPR or PDC
Time allocation  set the access limits

Apply for SUPR account

- http://supr.snic.se
- SNIC database of persons, projects, project proposals and more
- Apply and link SUPR account to PDC

Apply for PDC account

- http://www.pdc.kth.se/support/accounts/user
- Electronic copy of your passport
- Valid post address for password
- Membership of specific time allocation
## Time Allocations

### Small allocation
- Applicant can be a PhD student or higher
- Evaluated on a technical level only
- Limits is usually $5K$ corehours/month

### Medium allocation
- Applicant must be a senior scientist in Swedish academia
- Evaluated on a technical level only
- On large clusters: $200K$ corehours/month

### Large allocation
- Applicant must be a senior scientist in Swedish academia
- Need evidence of successful work at a medium level
- Evaluated on a technical and scientific level
- Proposal evaluated by SNAC twice a year
Using resources

- All resources are free of charge for Swedish academia
- Acknowledgement are taken into consideration when applying
- Please acknowledge SNIC/PDC when using these resources:

### Acknowledge SNIC/PDC

The computations/simulations/[SIMILAR] were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at [CENTERNAME (CENTER-ACRONYME)]

### Acknowledge people

NN at [CENTER-ACRONYME] is acknowledged for assistance concerning technical and implementation aspects [OR SIMILAR] in making the code run on the [OR SIMILAR] [CENTER-ACRONYME] resources.
Authentication

Kerberos Authentication Protocol

Ticket

- Proof of users identity
- Users use passwords to obtain tickets
- Tickets are cached on the user’s computer for a specified duration
- Tickets should be created on your local computer
- No passwords are required for during the ticket’s lifetime

Relam

Sets boundaries within which an authentication server has authority NADA.KTH.SE

Principal

Refers to the entries in the authentication server database username@NADA.KTH.SE
Kerberos commands

- `kinit` proves identity
- `klist` lists tickets
- `kdestroy` destroys ticket file
- `kpasswd` changes password

```
$ kinit -f username@NADA.KTH.SE
$ klist -Tf

Credentials cache : FILE:/tmp/krb5cc_500
          Principal: username@NADA.KTH.SE
          Issued Expires Flags Principal
       Mar 25 09:45 Mar 25 19:45 FI krbtgt/NADA.KTH.SE@NADA.KTH.SE
       Mar 25 09:45 Mar 25 19:45 FA afs/pdc.kth.se@NADA.KTH.SE
```
Login using Kerberos tickets

Get a 7 days forwardable ticket on your local system

$ kinit -f -l 7d username@NADA.KTH.SE

Forward your ticket via ssh and login

$ ssh username@clustername.pdc.kth.se

Always create a kerberos ticket on your local system
Outline

1. PDC Overview
2. Infrastructure
   - Beskow
   - Tegner
3. Accounts
   - Time allocations
   - Authentication
4. Job submission and motitor
   - SLURM commands
5. Conclusions
SLURM queueing system
Simple Linux Utility for Resource Management

- Open source, fault-tolerant, and highly scalable cluster management and job scheduling system
  - Allocates exclusive and/or non-exclusive access to resources for some duration of time
  - Provides a framework for starting, executing, and monitoring work on the set of allocated nodes
  - Arbitrates contention for resources by managing a queue
- Job Priority computed based on
  - **Age** the length of time a job has been waiting
  - **Fair-share** the difference between the portion of the computing resource that has been promised and the amount of resources that has been consumed
  - **Job size** the number of nodes or CPUs a job is allocated
  - **Partition** a factor associated with each node partition
  - **QOS** a factor associated with each Quality Of Service
salloc

Request for an interactive allocation or resources

$ salloc -t <min> -N <nodes> -A <myCAC> [script/command]

Once the resource is allocated use

for Beskow

$ aprun -n <cores> [-N <nodes>] ./MyPrgm

for Tegner

$ mpirun -n <cores> ./MyProg
`$ sbatch <script>`

The script should contain all necessary data to identify the account and requested resources.

**Example of request to run myexe for 1 hour on 4 nodes**

```bash
# set !/bin/bash -l

#SBATCH -A 201X-X-XX
#SBATCH -J myjob
#SBATCH -t 1:00:00
#SBATCH --nodes=4
#SBATCH --ntasks-per-node=32

#SBATCH -e error_file.e
#SBATCH -o output_file.o

aprun -n 128 ./myexe > my_output_file 2>&1
```
squeue and scancel
allow the monitoring and/or canceling of (all) running jobs

**squeue**

```bash
$ squeue -u <username>
```

Displays all queue and/or running jobs that belong to the user

**scancel**

```bash
$ scancel <jobid>
```

Stops a running job or removes a pending one from the queue
1. PDC Overview
2. Infrastructure
   - Beskow
   - Tegner
3. Accounts
   - Time allocations
   - Authentication
4. Job submission and monitor
   - SLURM commands
5. Conclusions
How to start your project

- Proposal for a small allocation
- Develop and test your code
- Run and evaluate scaling
- Proposal for a medium (large) allocation