Introduction to the PDC environment

ROYAL INSTITUTE
OF TECHNOLOGY

PDC Center for High Performance Computing
KTH, Sweden

Basic introduction

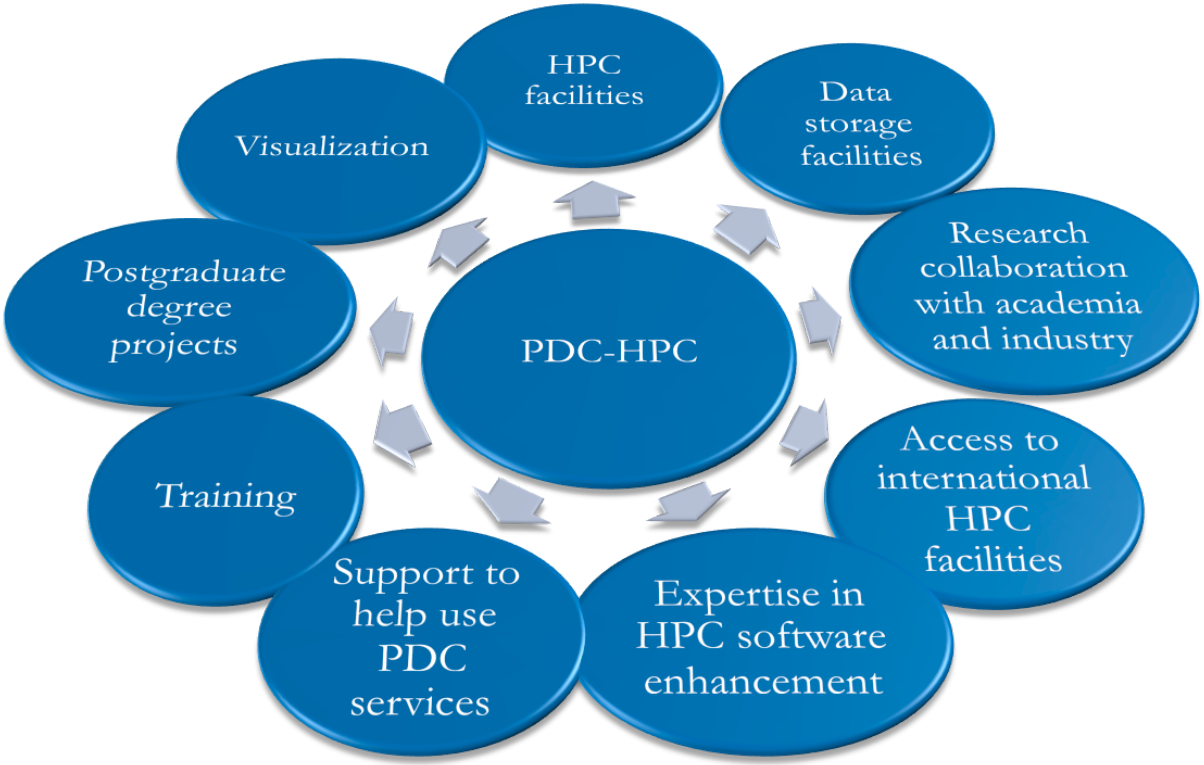
1. General information about PDC
2. Infrastructure at PDC
3. How to apply to PDC resources
4. File systems, permissions and transfer
5. How to login
6. Modules
7. Available software
8. How to run jobs
9. Compilers
10. Conclusion

General Information about PDC

SNIC Centra

The Swedish National Infrastructure for Computing (SNIC) is a 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). The resources are made available through open application procedures such that the best Swedish research is supported.

PDC Offers


PDC-HPC
HPC facilities
Data storage facilities
Research collaboration with academia and industry
Access to international HPC facilities
Expertise in HPC software enhancement
Support to help use PDC services
Training
Postgraduate degree projects
Visualization

PDC Key Assets: First-Line Support and System Staff

First-line support

Helps you have a smooth start to using PDC's resources and provides assistance if you need help while using our facilities

System staff: System managers/administrators

Ensure that PDC's HPC and storage facilities run smoothly and securely

PDC's Key Assets: HPC Application Experts

PDC-HPC 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.

Thor Wikfeldt

Computational Chemistry

Jing Gong

Scientific Computing

Cristian Cira

Code Optimization

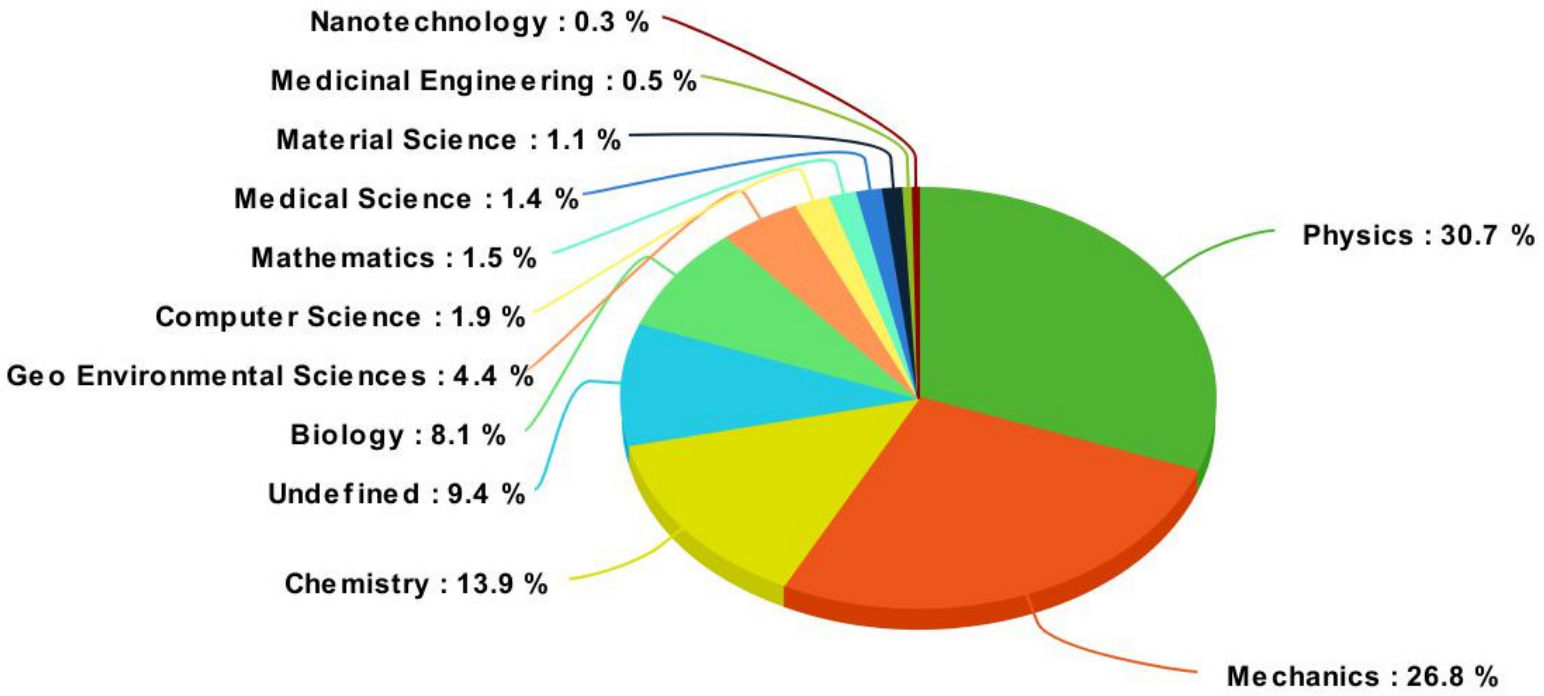
Henric Zazzi

Bioinformatics/Genetics

Jaime Rosal Sandberg

Computational Chemistry

Research areas at PDC



Usage of Beskow by research areas, March 2017

Infrastructure at PDC

Beskow

- 32nd place on the top500 (Q4 2015)
- Fastest in Scandinavia
- Intended for very large jobs (>512 core/job)
- Allocated through SNIC
- Queue limit is 24 hours
- Runs the SLURM queue system
- Partially reserved for PRACE, SCANIA, INCF
- Lifetime: Q4 2018

Tegner

- Intended for Beskow pre/post processing
- Not allocated via SNIC
- Only for academia within the Stockholm area
- Has large RAM nodes
- Has nodes with GPUs
- Runs the SLURM queue system
- Lifetime: Q4 2018

Summary of PDC resources

Computer	Beskow	Tegner
Core/node	32	48/24
Nodes	1676	**50**: 24 Haswell/GPU **10**: 48 Ivy bridge
RAM (Gb)	64	**50**: 512 **5**: 1000 **5**: 2000
Small allocations	5000	5000
Medium allocations	200000	80000
Large allocations	>200000	
Allocations via SNIC	yes	no
Lifetime	Q4 2018	Q4 2018
AFS	login node only	yes
Lustre	yes	yes

How to apply for PDC resources

Access to PDC resources

- User account (SUPR/PDC)
- Time allocation
  - A measure for how many jobs you can run per month (corehours/month)
  - Which clusters you can access
- Time allocation requirements
  - Can be personal or shared within a project
  - Every user must belong to at least one time allocation

How to get a time allocation

- PDC resources are free for swedish academia
- You can apply for a SUPR account at http://supr.snic.se
- In SUPR send in a proposal for your project
- More information at http://www.snic.vr.se/apply-for-resources

File systems, permissions and transfer

File systems at PDC

- AFS (Andrew File System)
  - distributed
  - global
  - backup
- Lustre (Linux cluster file system)
  - distributed
  - high-performance
  - no backup

AFS

- Andrew File System
- Named after the Andrew Project (Carnegie Mellon University)
- Distributed file system
- Security and scalability
- Accessible "everywhere" (remember that when you make your files readable/writeable!)
- Not available on Beskow compute nodes
- Access via Kerberos tickets and AFS tokens

AFS

- Your PDC home directory is located in AFS, example:

```
/afs/pdc.kth.se/home/u/user
```

- OldFiles mountpoint (created by default) contains a snapshot of the files as they were precisely before the last nightly backup was taken.

```
/afs/pdc.kth.se/home/u/user/OldFiles
```

- By default you get a limited quota (5 GB)

Lustre

- Parallel distributed file system
- Large-scale cluster computing
- High-performance
  - /cfs/klemming
- UNIX permissions
- No personal quota. Move your data when finished
- Not global

Lustre

- Always start and run your programs in lustre if possible
- Default home directory:

```
# Not backed up
/cfs/klemming/nobackup/[username 1st letter]/[username]
# Files older than 30 days will be deleted
/cfs/klemming/scratch/[username 1st letter]/[username]
```

File transfer between PDC machines

- /afs is mounted and visible on all machines (at least on login node)
- No need to "transfer" files which are on /afs
- You can share files between machines via /afs

How to install AFS

- Install AFS client and copy directly then AFS is mounted just like another disk on your computer
- https://www.pdc.kth.se/resources/software/file-transfer/file-transfer-with-afs/linux
- https://www.pdc.kth.se/resources/software/file-transfer/file-transfer-with-afs/mac
- https://www.pdc.kth.se/resources/software/file-transfer/file-transfer-with-afs/windows

scp, an alternative for AFS

```
# from my laptop to Beskow
$ scp myfile user@beskow.pdc.kth.se:~/Private
# from Beskow Lustre to my laptop
$ scp user@beskow.pdc.kth.se:/cfs/klemming/scratch/u/user/file.txt .
```

If the username is the same on source and destination machine, you can leave it out

For large files use the transfer nodes on Tegner:

```
t04n27.pdc.kth.se, t04n28.pdc.kth.se
```

```
# from my laptop to klemming
$ scp file.txt user@t04n27.pdc.kth.se:/cfs/klemming/scratch/u/user
```

How to login

Kerberos

Is an authentication protocol originally developed at MIT
PDC uses kerberos together with **SSH** for login

- **Ticket**
  - Proof of users identity
  - Users use password to obtain tickets
  - Tickets are cached on users computer for a specified duration
  - Tickets should be created on your local computer
  - As long as tickets are valid there is no need to enter password

Kerberos

- **Realm**
  - all resources available to access
  - example: NADA.KTH.SE
- **Principal**
  - Unique identity to which kerberos can assign tickets.
  - example: [username]@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 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

1. Get a 7 days forwardable ticket on your local system

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

2. Forward your ticket via ssh and login

```
$ ssh username@clustername.pdc.kth.se
```

3. Replace clustername...
    1. beskow login node: beskow.pdc.kth.se
4. You will have reached the cluster

Always create a kerberos ticket on your local system

Login from any computer

- You can reach PDC from any computer or network
- The kerberos implementation heimdal can be installed on most operating systems
  - Linux *heimdal, openssh-client*
  - Windows *Network Identity Manager, PuTTY*
  - Mac
- Follow the instructions for your operating system http://www.pdc.kth.se/resources/software/login-1

Modules

What are Modules

Used to load a specific software, and versions, into your environment

What modules do

```
$ module show fftw/3.3.4.0
-------------------------------------------------------------------
/opt/cray/modulefiles/fftw/3.3.4.0:

setenv           FFTW_VERSION 3.3.4.0
setenv           CRAY_FFTW_VERSION 3.3.4.0
setenv           FFTW_DIR /opt/fftw/3.3.4.0/haswell/lib
setenv           FFTW_INC /opt/fftw/3.3.4.0/haswell/include
prepend-path     PATH /opt/fftw/3.3.4.0/haswell/bin
prepend-path     MANPATH /opt/fftw/3.3.4.0/share/man
prepend-path     CRAY_LD_LIBRARY_PATH /opt/fftw/3.3.4.0/haswell/lib
setenv           PE_FFTW_REQUIRED_PRODUCTS PE_MPICH
prepend-path     PE_PKGCONFIG_PRODUCTS PE_FFTW
setenv           PE_FFTW_TARGET_interlagos interlagos
setenv           PE_FFTW_TARGET_sandybridge sandybridge
setenv           PE_FFTW_TARGET_x86_64 x86_64
setenv           PE_FFTW_TARGET_haswell haswell
setenv           PE_FFTW_VOLATILE_PKGCONFIG_PATH /opt/fftw/3.3.4.0/@PE_
prepend-path     PE_PKGCONFIG_LIBS fftw3f_mpi:fftw3f_threads:fftw3f:fft
module-whatis    FFTW 3.3.4.0 - Fastest Fourier Transform in the West
-------------------------------------------------------------------
```

Module commands

module add *software[/version]*:
loads *software[/version]*

module avail: Lists available softwares

module show *software*:
shows information about *software*

module list: Lists currently loaded softwares

module swap *frommodule tomodule*:
swaps *frommodule* to *tomodule*

How to use modules

```
$ module list # on Milner
Currently Loaded Modulefiles:
1) modules/3.2.6.7
2) nodestat/2.2-1.0501.47138.1.78.ari
3) sdb/1.0-1.0501.48084.4.48.ari
4) alps/5.1.1-2.0501.8471.1.1.ari
5) MySQL/5.0.64-1.0000.7096.23.2
6) lustre-cray_ari_s/2.4_3.0.80_0.5.1_1.0501.7664.12.1-1.0501.14255.11
7) udreg/2.3.2-1.0501.7914.1.13.ari
8) ugni/5.0-1.0501.8253.10.22.ari
9) gni-headers/3.0-1.0501.8317.12.1.ari
10) dmapp/7.0.1-1.0501.8315.8.4.ari
11) xpmem/0.1-2.0501.48424.3.3.ari
...
```

Available software

On our cluster, we have already installed a number of software with their different versions.

More information about the software, how they were installed and how to run them at PDC is available at

https://www.pdc.kth.se/software

How to run jobs

SLURM queue system

1. Allocates exclusive and/or non-exclusive access to resources (computer nodes) to users for some duration of time so they can perform work.
2. Provides a framework for starting, executing, and monitoring work (typically a parallel job) on a set of allocated nodes.
3. Arbitrates contention for resources by managing a queue of pending work
4. Installed on Beskow, Tegner
5. Installed by default, no need to load module

Difference between Beskow and Tegner

To launch a parallel program...

- On Tegner use **mpirun**
- On Beskow Cray Linux Environment (CLE) use **aprun**

Using salloc

- To book and execute on a dedicated node

```
$ salloc -t <min> -N <nodes> -A <myCAC> mpirun -n cores ./MyPrgm
```

- To run interactively

```
$ salloc -A <myCAC> -t <min>
$ mpirun -A <myCAC> -n <cores> [-N <nodes] ./MyPrgm
$ mpirun -A <myCAC> -n <cores> [-N <nodes] ./MyPrgm
$ exit
```

Requesting a specific type of node

It is also possible in SLURM to request a specific type of node, e.g. if there is a mix of large or small memory nodes e.g.

```
# Request a node with at least 1 TB RAM
salloc -t 1:00:00  -A <myCAC> -N 1 --mem=1000000
# Request a node with at least 24 logical CPUs
salloc  -A <myCAC> -N 1 -t 300 --mincpus=24
# Request a node with a K80 GPU
salloc  -A <myCAC> --gres=gpu:K80:2
```

If the cluster does not have enough nodes of that type then the request will fail with an error message.

Using sbatch

```
$ sbatch <script>
```

```
#!/bin/bash -l
#SBATCH -J myjob
# Defined the time allocation you use
#SBATCH -A <myCAC>
# 10 minute  wall-clock time will be given to this job
#SBATCH -t 10:00
# Number of nodes
#SBATCH --nodes=2
# set tasks per node to 24 to disable hyperthreading
#SBATCH --ntasks-per-node=24
# load intel compiler and mpi
module load i-compilers intelmpi
# Run program
mpirun -n 48 ./hello_mpi
```

Other SLURM commands

- To remove a submitted job

```
$ scancel jobid
```

- Show my running jobs

```
$ squeue [-u <username>]
```

Other commands for looking at job and time allocations

Projinfo:

```
$ projinfo -h
Usage: projinfo [-u <username>] [-c <clustername>] [-a] [-o]
-u [user] : print information about specific user
-o : print information about all (old) projects, not just current
-c [cluster] : only print allocations on specific cluster
-a : Only print membership in projects
-d : Usage by all project members
-p [DNR] : only print information about this project
-h : prints this help
```

Statistics are also available at...
https://vetinari.pdc.kth.se/dashboard/

Course Allocation

- Allocation:

```
edu17.XXXX
```

- XXXX is the name of the course

Compilers

Compiling serial code on Tegner

```
# GNU
$ gfortran -o hello hello.f
$ gcc -o hello hello.c
$ g++ -o hello hello.cpp
# Intel
$ module add i-compilers
$ ifort -FR -o hello hello.f
$ icc -o hello hello.c
$ icpc -o hello hello.cpp
```

Compiling MPI/OpenMP code on Tegner

```
# GNU
$ module add gcc/5.1 openmpi/1.8-gcc-5.1
$ mpif90 -FR -fopenmp -o hello_mpi hello_mpi.f
$ mpicc -fopenmp -o hello_mpi hello_mpi.c
$ mpic++ -fopenmp -o hello_mpi hello_mpi.cpp
# Intel
$ module add i-compilers intelmpi
$ mpiifort -openmp -o hello.f90 -o hello_mpi
$ mpiicc -openmp -o hello_mpi hello_mpi.c
$ mpiicpc  -openmp -o hello_mpi hello_mpi.cpp
```

Conclusion

PDC support

- A lot of question can be answered via our web
  http://www.pdc.kth.se/support
- The best way to contact us is via e-mail
  http://www.pdc.kth.se/about/contact/support-requests
- The support request will be tracked
- Use a descriptive subject in your email
- Give your PDC user name.
- Provide all necessary information to reproduce the problem.
- For follow ups always reply to our emails