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 access PDC resources
4. File systems
5. How to login
6. Parallel computing and parallel programming
7. Modules
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 Key Assets: Access to EU Facilities and Experts

PDC is an active partner in many international and national projects.

PDC and industry

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

PDC recently established a business development unit that provides consultancy and HPC services to industries.

PDC Key Assets: Broad Range of Training

- PDC Summer School every year: Introduction to HPC
- Specific courses: Programming with GPGPU, Recent Advances in Distributed and Parallel Computing, Software Development Tools, Recent Advances in Cloud Computing, and many many more….
- PDC user days, PDC Open House and Pub Afternoon



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.

Jonathan Vincent

Computational Physics

Michael Djurfeldt

Computational Physics

Jing Gong

Scientific Computing

Cristian Cira

Code Optimization

Thor Wikfeldt

Computational Chemistry

Henric Zazzi

Bioinformatics/Genetics

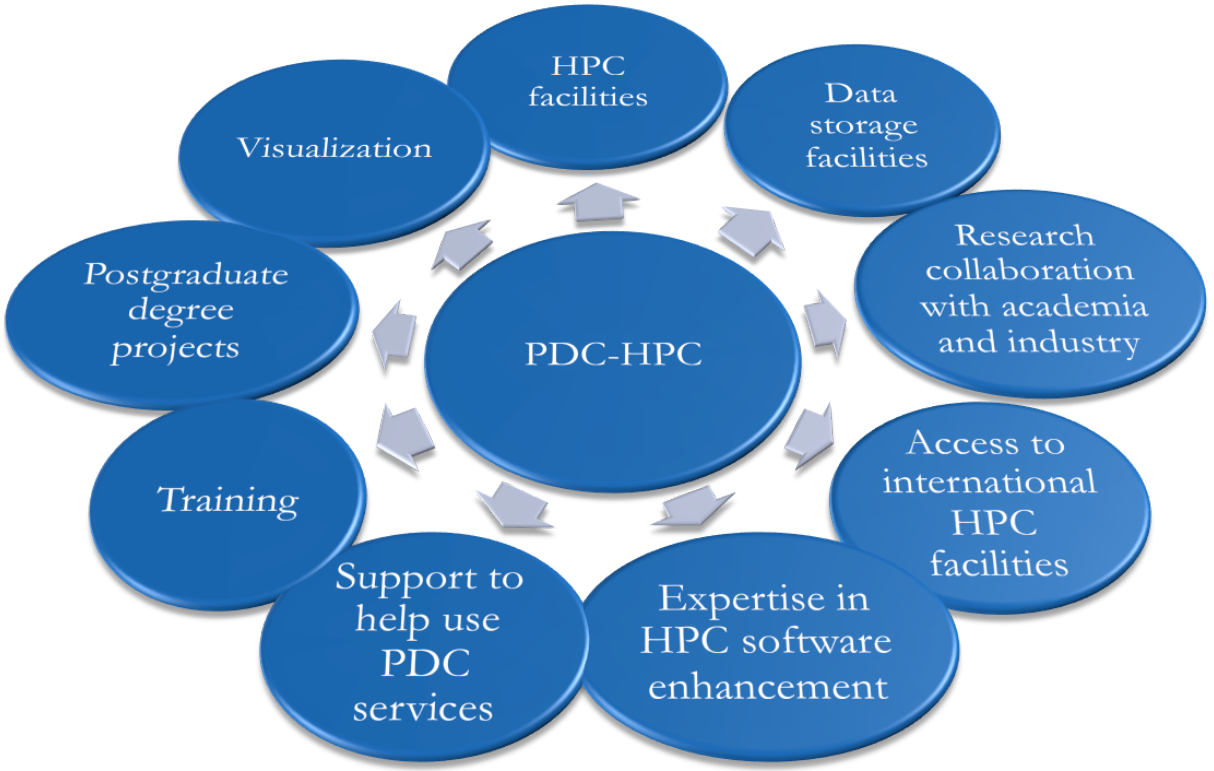
Jaime Rosal Sandberg

Computational Chemistry

What can an application expert help with

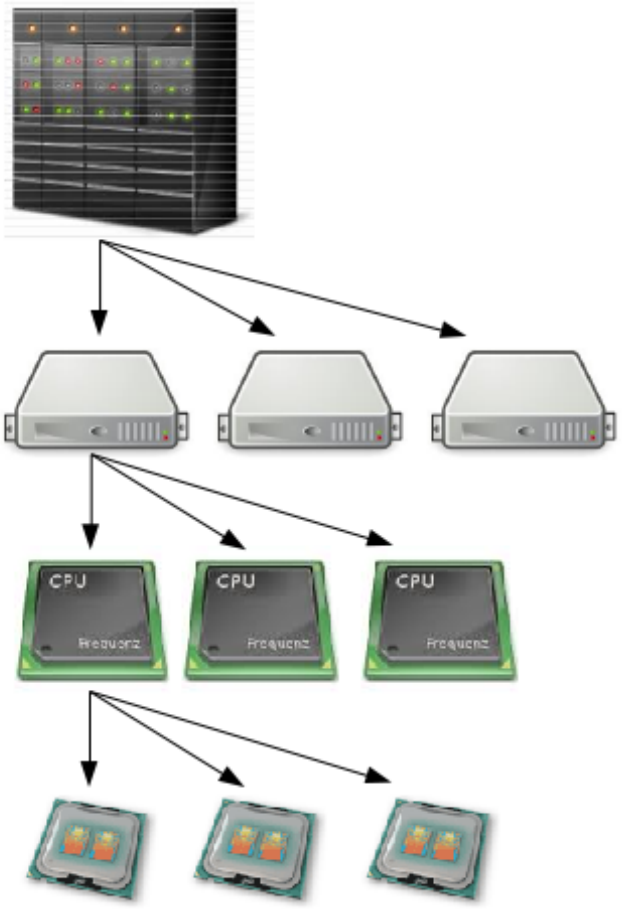
1. Support
2. Help in installing specific software
3. Help in using specific software
4. Development of software
5. Code optimization

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

Infrastructure at PDC

What is a cluster



Cluster

Node
Similar to computer without keyboard and monitor

Processor

Cores

Beskow

- 32nd place on the top500 (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
  - Large allocations: 75%
  - Medium allocations: 20%
- Lifetime: Q4 2018

Tegner

- Intended for Beskow pre/post processing
- Not allocated through 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 access PDC resources

What do you need

- User account (SUPR/PDC)
- Time allocation (CAC)
  - 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

Time allocations

Small allocation

Applicant can be a PhD student or higher

Evaluated on a technical level only

Limits is usually 5000 corehours/month

Medium allocation

Applicant must be a senior scientist in swedish academia

Evaluated on a technical level only

Limits depend on machine

On large clusters: 200 kcorehours/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

No formal limits

Use of resources

- All resources are free of charge for Swedish academia
- Please acknowledge SNIC/PDC when using these resources.

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

- Acknowledge them or include them in article if somebody has contributed

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

Acknowledgement will be taken into consideration when applying for new resources.

Apply to an account

- Electronic copy of your passport
- Examine the computer rules at PDC
- Which postal address the password should be sent
- Which time allocations the users will access
  - Not needed in case the users is applying for a time allocation
- Apply to a SUPR account http://supr.snic.se
- Apply for a PDC account http://www.pdc.kth.se/support/accounts/user

File systems

Andrew File System (AFS)

- AFS is a global file system accessible everywhere
- Your home directory is located in AFS

```
/afs/pdc.kth.se/home/[username 1st letter]/[username]
```

- Access via Kerberos tickets and AFS tokens

  Backup

- Follow the instructions for your operating system http://www.pdc.kth.se/resources/software/file-transfer/file-transfer-with-afs

Lustre

- Massively parallell distributed file system
- Very high performance
- No backup
- No personal quota. Move your data when finished
- 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]
```

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 kerberor 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

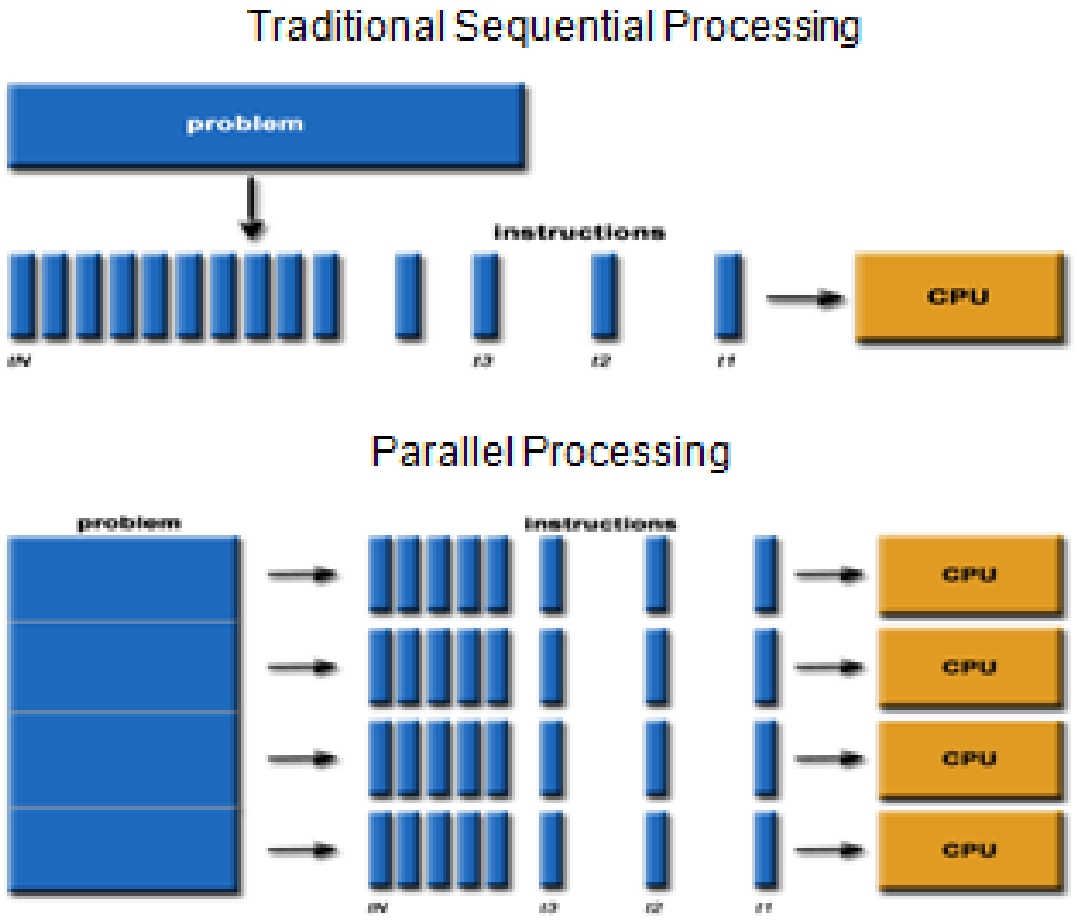
Always create a kerberos ticket on your local system

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 *PuTTY*
  - Mac
- Follow the instructions for your operating system http://www.pdc.kth.se/resources/software/login-1

Parallel computing and parallel programming

What is parallel computing



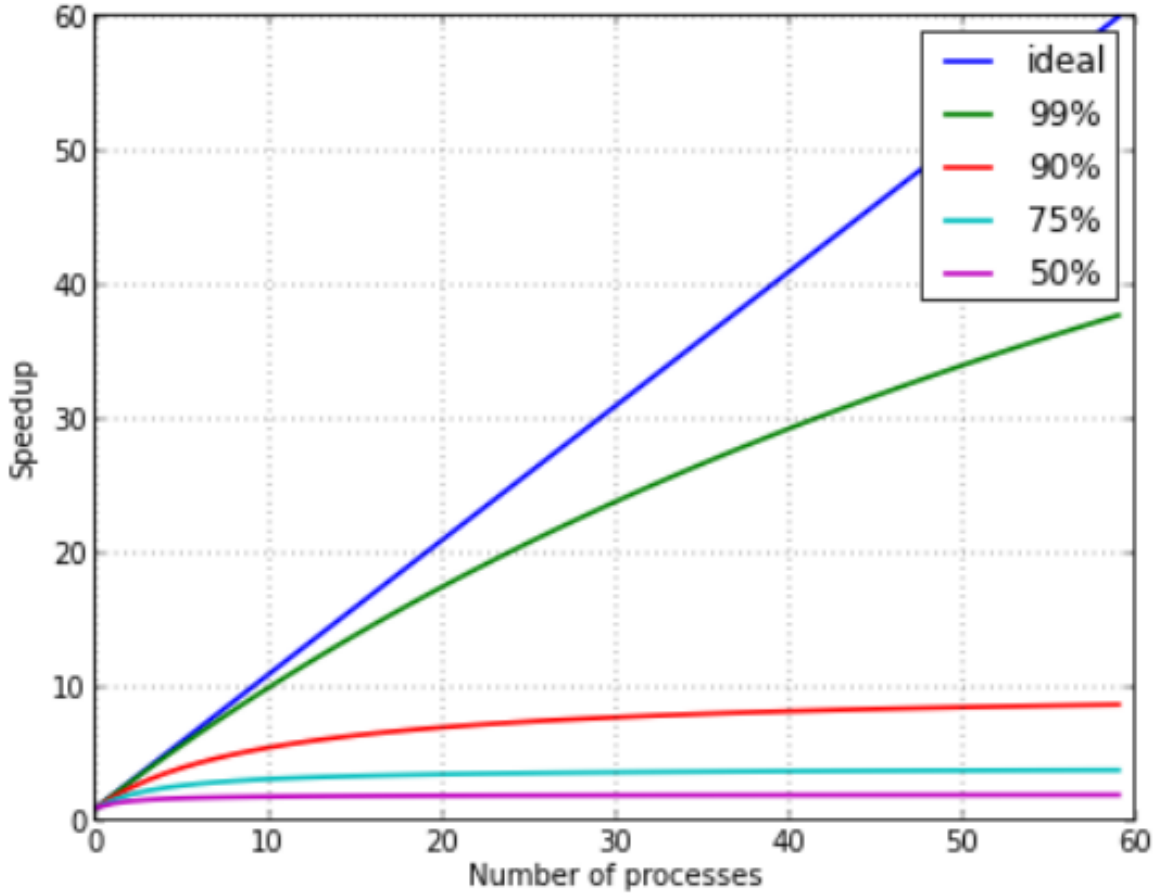
Typical parallel programming standards

- MPI (message passing interface)
  - C, C++, Fortran, R, Python
  - Distributed memory (mostly)
- OpenMP (CPU, GPU)
  - C, C++, Fortran
  - Shared memory

Typical parallel programming standards

- CUDA (GPU)
  - NVIDIA
  - C, C++, Fortran
- OpenACC (GPU and accelerators)
  - NVIDIA, Cray, PGI
- OpenCL (GPU, CPU, and accelerators)
  - AMD, Intel

Parallel performance: Ideal vs. real



Parallel performance bottlenecks

- Synchronization overhead
- False sharing
- Load imbalance
- Memory bandwidth
- File I/O

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
...
```

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, Milner, Tegner
5. Installed by default, no need to load module

Prior to starting a job

1. Get a forwardable kerberos ticket from local computer

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

2. Code and data should reside in Lustre

```
/cfs/klemming/[nobackup/scratch]/u/username
```

3. Login/Forward your ticket via ssh

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

Using salloc

- To book a dedicated node

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

- On Beskow, to run interactively

```
$ salloc -t <min>
$ aprun -n <cores> [-N <nodes] ./MyPrgm
$ aprun -n <cores> [-N <nodes] ./MyPrgm
```

- On Tegner, to allocate and run at the same time

```
$ salloc -t <min> mpirun -n <cores> ./MyPrgm
```

- myCAC edu16.SF2568

Using sbatch

```
$ sbatch <script>
```

```
#!/bin/bash -l
#SBATCH -J myjob
# 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>]
```

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

How to start your project

- Proposal for a small allocation
- Develop and test your code
- Scaling
- Proposal for a medium (large) allocation
- Analysis

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
- Write descriptive subject line
- For follow ups always include support number [SNIC support #NNNNN]
- Do not make new support cases by replying to old tickets
- Split unrelated problems into separate email requests.
- Use a descriptive subject in your email (unhelpful subject line: "problem").

PDC support

- Give your PDC user name.
- Be as specific as possible.
- For problems with scripts/jobs, give an example. Either send the example or make it accessible to PDC support.
- Make the problem example as small/short as possible.
- Provide all necessary information to reproduce the problem.
- If you want the PDC support to inspect some files, make sure that the files are readable.
- Do not assume that PDC support personnel have admin rights to see all your files or change permissions.