SHIV NADAR UNIVERSITY

HPCC- MAGUS- Introduction.

Deepak Agrawal

Contents

- Introduction to HPC
- How to connect to Magus
- Basic Linux for day to day operation on Magus
- Compiling your own code serial and parallel
- Introduction to MPI
- JOB Schedulers
- Software and Library Locations
- Best Practices and guidelines to use Magus

Introduction

- **Introduction of Parallel Computing**
- **Example of Serial Computing.**
  - **(5+2) + (3*4) * 2*(100-(2+4))**
    - *7+(3*4)*2*(100-(2+4))*
    - *7+12*2*(100-(2+4))*
    - *7+12*2*(100-6)*
    - *7+12*2*94*
    - *7+24*94*
    - *7+2256*
    - Result 2263
    - **Total Cycles Used 7**
- **Using Parallel Computing**
  - **Using 3 nodes**

	Node 1	Node 2	Node 3
Cycle 1	(5+2)	(3*4)	(2+4)
Cycle 2	100-6	12*2	Idle
Cycle 3	24*94	idle	Idle
Cycle 4	7+2256	idle	idle

Result : 2263

Total Cycle used 4

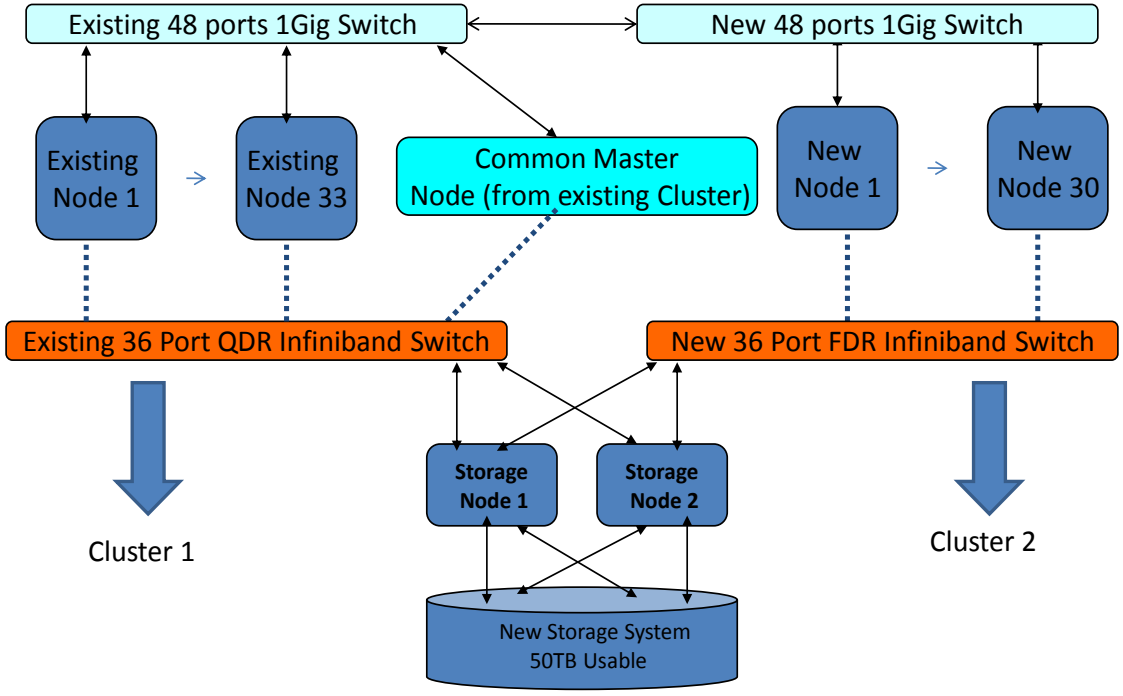
SNU HPCC- Magus – Key facts

- **Total 62 Compute Nodes.**
- **New Intel Hash well Architecture Processors on 30 Nodes**
- **Total ~ 1000 Cores**
- **Total ~ 6 TB RAM.**
- **IBM GPFS Parallel File System**
- **~30 TF Theoretical Peak Performance Total.**
- **8- High CPU & Memory Nodes**
- **Latest software upgrade for IBM Platform HPC**

SNU HPCC- Magus Architecture

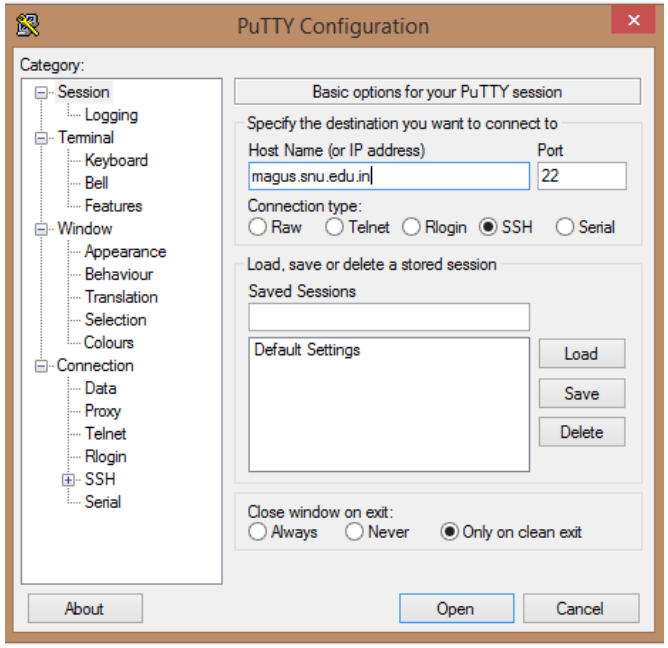
Final and revised HPC Cluster Layout for SNU

Below cluster 1 and cluster 2 will only share 50TB Storage system

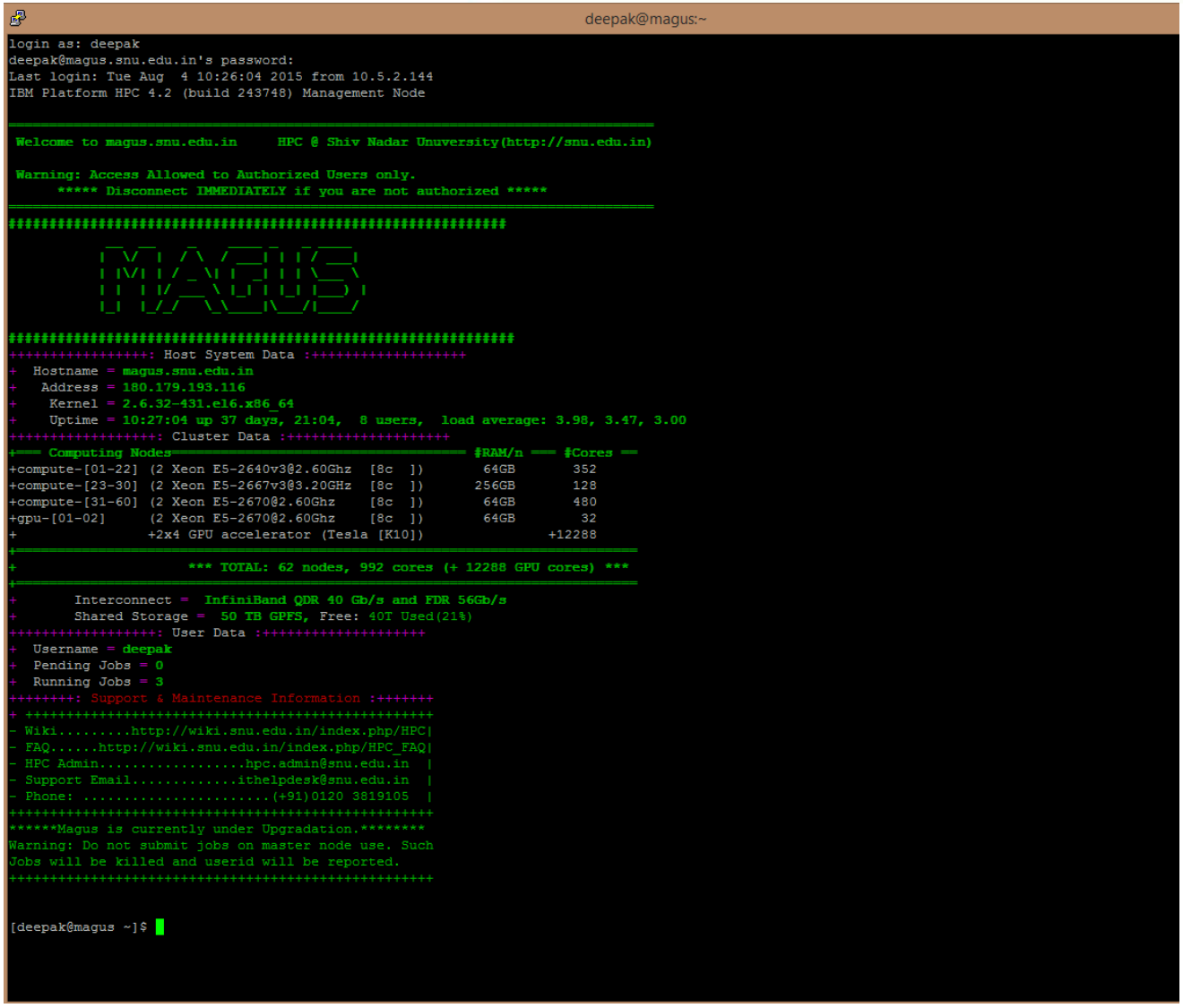


How to Connect to Magus

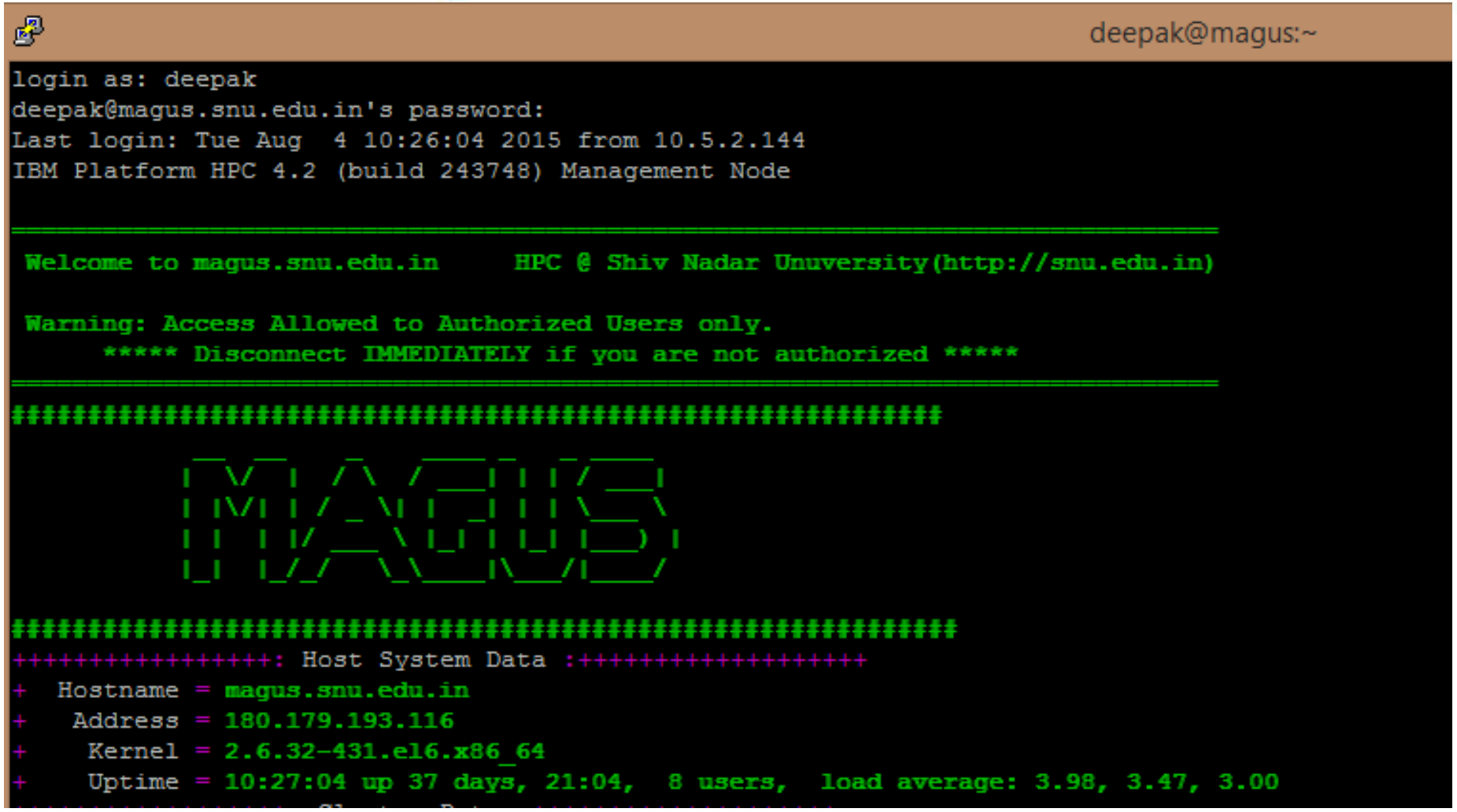
- **User Account .**
- **Request for User Account on HPC. http://hpc.snu.edu.in/hpcAccount/**
- **SSH**
  - **Host magus.snu.edu.in (default port 22)**
- **From Linux Machine**
  - **$ ssh user@magus.snu.edu.in**
- **From Windows Machine using putty**



Magus Login Screen



Login Screen Explained

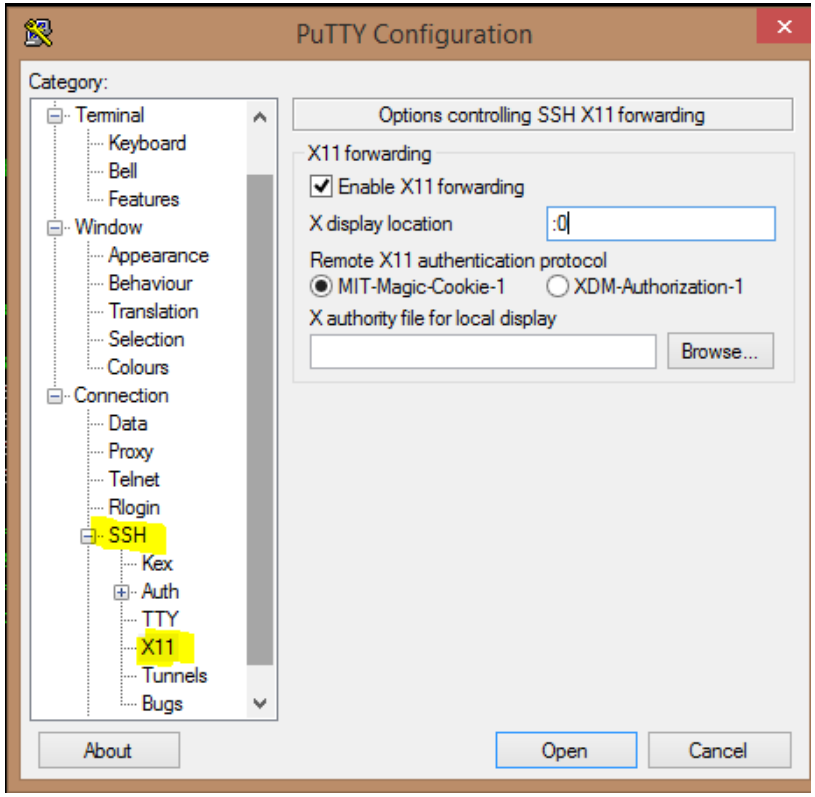


Login Screen

```
++++++++++++++++++: Cluster Data :++++++++++++++++++++
+=== Computing Nodes ================================== #RAM/n === #Cores ==
+compute-[01-22] (2 Xeon E5-2640v3@2.60Ghz   [8c  ])      64GB        352
+compute-[23-30] (2 Xeon E5-2667v3@3.20GHz   [8c  ])     256GB        128
+compute-[31-60] (2 Xeon E5-2670@2.60Ghz     [8c  ])      64GB        480
+gpu-[01-02]     (2 Xeon E5-2670@2.60Ghz     [8c  ])      64GB         32
+                +2x4 GPU accelerator (Tesla [K10])                +12288
+=========================================================================
+                    *** TOTAL: 62 nodes, 992 cores (+ 12288 GPU cores) ***
+=========================================================================
+       Interconnect =  InfiniBand QDR 40 Gb/s and FDR 56Gb/s
+       Shared Storage =  50 TB GPFS, Free: 40T Used(21%)
++++++++++++++++++: User Data :++++++++++++++++++++
+  Username = deepak
+  Pending Jobs = 0
+  Running Jobs = 3
++++++++: Support & Maintenance Information :+++++++
+ ++++++++++++++++++++++++++++++++++++++++++++++++++
- Wiki.........http://wiki.snu.edu.in/index.php/HPC|
- FAQ......http://wiki.snu.edu.in/index.php/HPC_FAQ|
- HPC Admin..................hpc.admin@snu.edu.in  |
- Support Email.............ithelpdesk@snu.edu.in  |
- Phone: .....................(+91)0120 3819105  |
++++++++++++++++++++++++++++++++++++++++++++++++++++
******Magus is currently under Upgradation.********
Warning: Do not submit jobs on master node use. Such
Jobs will be killed and userid will be reported.
++++++++++++++++++++++++++++++++++++++++++++++++++++
```

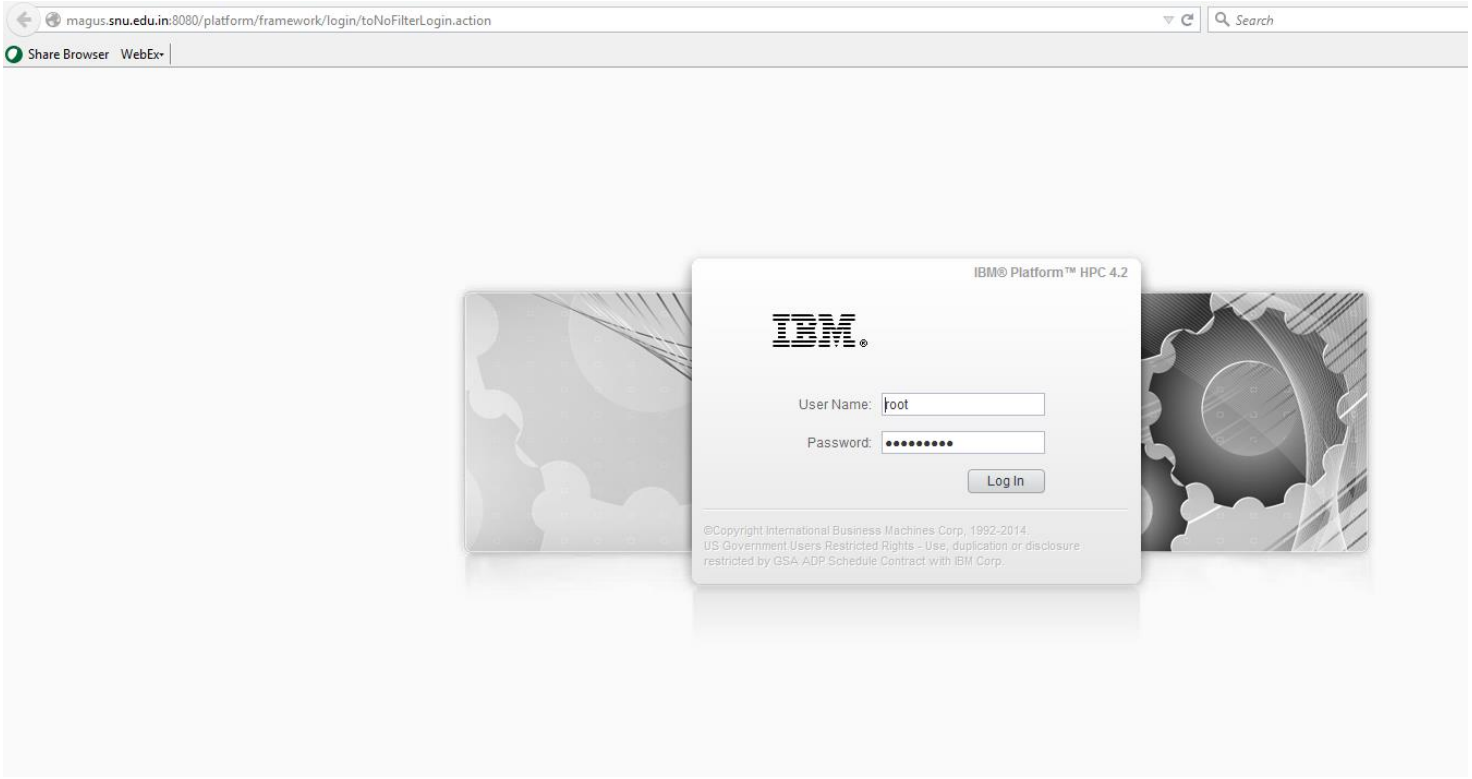
How to Connect to Magus

- **GUI Forwarding (X – Forwarding)**
  - **From Linux Machine**
    - **$ ssh –X user@magus.snu.edu.in**
  - **From Windows Machine**
    - **Xming**
    - **Putty X forwarding**

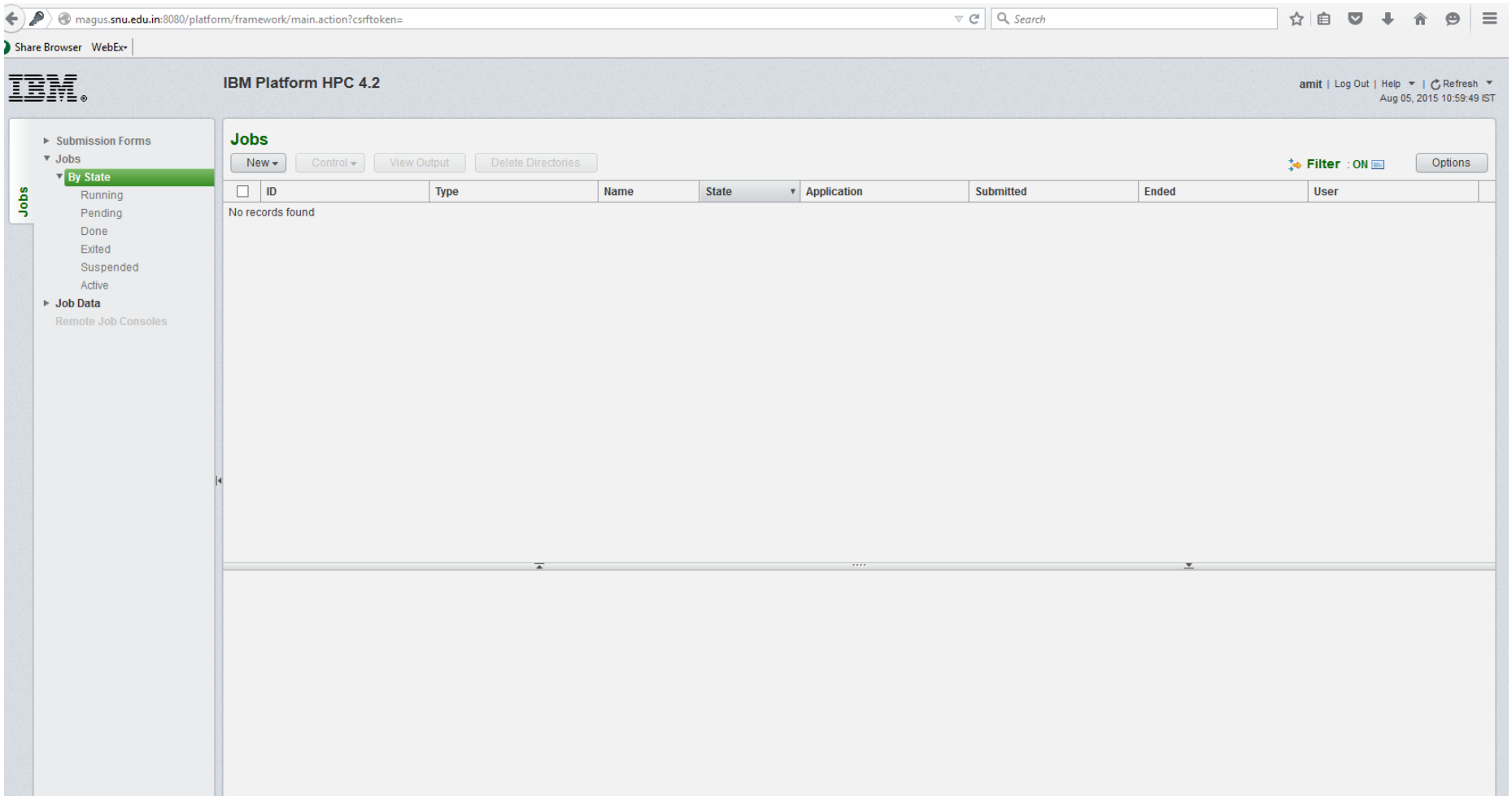


Magus Web UI

- **Web UI (Job status and monitoring)**
  - **http://magus.snu.edu.in**



Magus Web UI


magus.snu.edu.in:8080/platform/framework/main.action?csrftoken=
Search
Share Browser
WebEx
IBM Platform HPC 4.2
amit | Log Out | Help | Refresh
Aug 05, 2015 10:59:49 IST
Jobs
Submission Forms
Jobs
By State
Running
Pending
Done
Exited
Suspended
Active
Job Data
Remote Job Consoles
Jobs
New
Control
View Output
Delete Directories
Filter : ON
Options
ID
Type
Name
State
Application
Submitted
Ended
User
No records found

Basic Linux CLI Commands

- man (Manual)
- ls (list directory)
- mkdir (Make Directory)
- rmdir (Remove Directory)
- pwd (present working Directory)
- cp (Copy)
- mv (Move / Rename)
- scp (Secure copy to other machine)
- cat (Display contents of the file)
- tail (Display last 10 lines)
- head (display first 10 lines)
- chmod (Change File permissions)
- grep (Search within files)
- du (Disk Usage)
- wc (word count)

Shortcuts

- Ctrl+C (kill a running command)
- Ctrl+Z (suspend a running command)
- fg/bg (Foreground and Background)
- Ctrl+R (search recent commands)

- Standard Output Redirection using > and >>
- Standard Error Redirection 2 >
- Sending command to background using &
- Sending output to other command using Pipe |

Introduction to MPI

- **Message Passing Interface**
  - **Reference**
    https://en.wikipedia.org/wiki/Message_Passing_Interface
  - **Communication Protocol for Parallel Computing**
- **Open MPI**
- **Intel MPI**

Compiling your code

- **Compiling a simple Hello World in C on linux**
- **Compiling a MPI Hello World in C on linux**
- **How to set environment variables for location of compilers and mkl libraries.**
  - ***$ source /snufs/intel/parallel_studio_xe_2015/bin/psxevars.sh intel64***

Sample Mpi Hello World

```
#include<stdio.h>
#include<mpi.h>

int main(int argc, char *argv[])
{

        int ranks, rank, n;
        int Debugmode=1;

        MPI_Init(&argc, &argv);
                char node[MPI_MAX_PROCESSOR_NAME];
                MPI_Comm_size(MPI_COMM_WORLD, &ranks);
                MPI_Comm_rank(MPI_COMM_WORLD, &rank);
                MPI_Get_processor_name(node,&n);
                printf("Greetings: %d of %d from the node %s\n", rank, ranks, node);
        MPI_Finalize();

return 0;
}
```

Compiling and Executing

- **Setting environment variables**

  $ source /snufs/intel/parallel_studio_xe_2015/bin/psxevars.sh intel64

- **Compiling**

  $ mpiicc

- **Executing mpi jobs**

  $ mpirun –np <number of processors> –f <hostfile>

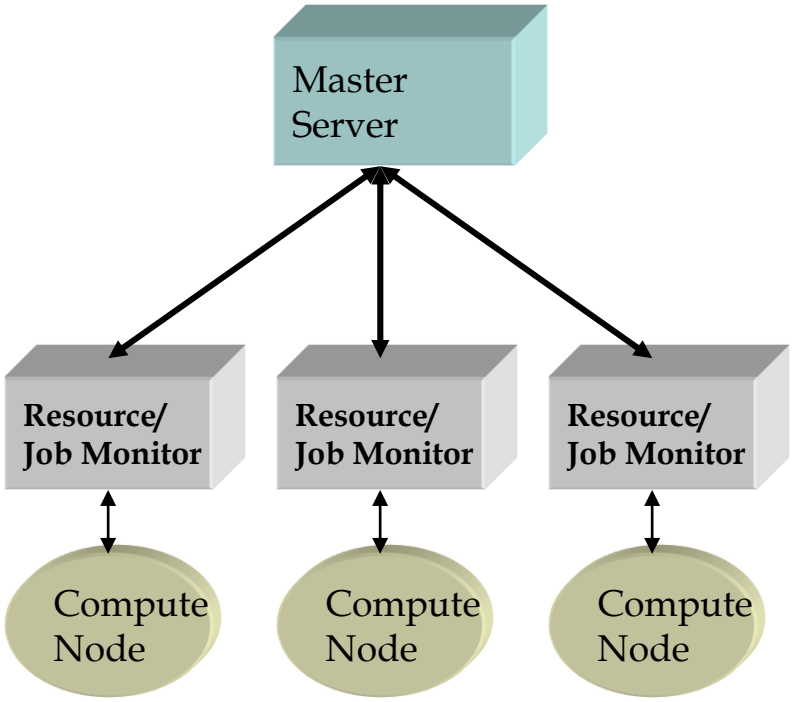
- **Mpirun options.**

  - ppn Processor per node

Job Scheduler

- Allow users to share computing resources
- Utilize resources efficiently.

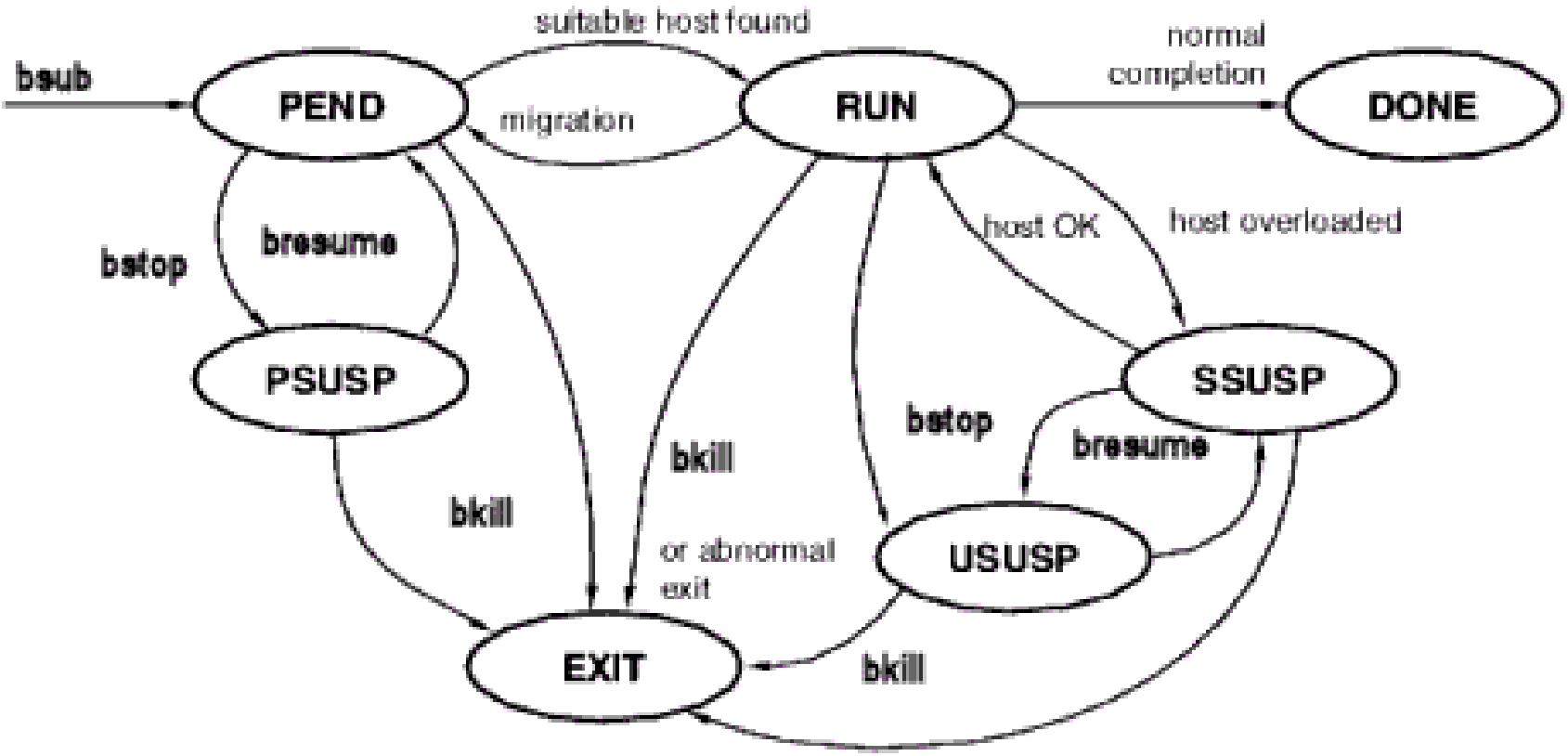
- IBM Platform LSF
  - distributed workload management solution for maximizing the performance of High Performance Computing (HPC) clusters.



JOB States

- LSF jobs have the following states:
  - PEND — Waiting in a queue for scheduling and dispatch
  - RUN — Dispatched to a host and running
  - DONE — Finished normally with zero exit value
  - EXIT — Finished with non-zero exit value
  - PSUSP — Suspended while pending
  - USUSP — Suspended by user
  - SSUSP — Suspended by the LSF system

Job Transitions



Queues

- A cluster wide container for jobs. All jobs wait in queues until they are scheduled and dispatched to hosts.
- Queues do not correspond to individual hosts; each queue can use all server hosts in the cluster, or a configured subset of the server hosts.
- When you submit a job to a queue, you do not need to specify an execution host. LSF dispatches the job to the best available execution host in the cluster to run that job.
- Queues implement different job scheduling and control policies.

Queues on Magus

S.No.	Processor Architecture	Priority	Queue Name	Min no of cores required to submit job	Max no of cores allowed per job	No of nodes in the queue	Wall Time	Max job per users in the Queue based on max cores
1	SandyBridge	50	serial_short	1	1	1	1 hour	unlimited
2	SandyBridge	50	serial_long	1	1	1	1 month	2
3	SandyBridge	50	short_sdb	4	8	12	1 day	8
4	SandyBridge	50	med_sdb	8	16	16	2 weeks	4
5	Hashwell	50	high_mem	16	128	8	1 week	1
7	Hashwell	50	med_hsw	8	16	22	2 weeks	5
8	SandyBridge	50	GPU	16	16	2	1 week	1
9	SandyBridge	20	long_sdb	8	32	20	1 month	unlimited
10	Hashwell	20	long_high_mem	8	64	8	1 month	unlimited
11	Hashwell	20	long_hsw	8	64	20	1 month	unlimited

Working with LSF

Job Submission

$ bsub < jobfile

```
#!/bin/bash
#BSUB -J Helloworld
#BSUB -n 16
#BSUB -q high_mem
#BSUB -e error.%J
#BSUB -o out.%J
#BSUB -u "deepak.agrawal@snu.edu.in"
#BSUB –N
#BSUB –W 10:00
#BSUB -m compute26
```

bsub options

-B	**Sends email when the job is dispatched**
-H	Holds the job in the PSUSP state at submission
-I	Submits a batch interactive job.
-K	Submits a job and waits for the job to finish
-N	Emails the job report when the job finishes
-x	Exclusive execution on host
-b	begin_time Dispatches the job on or after the specified date and time in the form [[month:]day:]:minute
-e	error_file Appends the standard error output to a file
-E	"pre_exec_command[arguments...]" Runs the specified pre-exec command on the execution host before running the job
-J	"job_name" Assigns the specified name to the job.
-q	queuename Submits jobs to the specified queue
-o	output_file Appends toe standard output to a file
-u	email address < sets the email where the email has to be sent >

LSF Sample Submission script

```
#!/bin/bash
#BSUB -J Hello world
#BSUB -n 16
#BSUB -q high_mem
#BSUB -e error.%J
#BSUB -o out.%J
#BSUB -R "span[ptile=16]"

MPI=16
PPN=16
MYDIR=$(pwd)

EXE=/snufs/home/deepak/helloworld/a.out
OUT_FILE=$MYDIR/out.log

# Do not change anything below this
export I_MPI_DAPL_PROVIDER=ofa-v2-mlx4_0-1
export I_MPI_FABRICS=shm:dapl
export I_MPI_FALLBACK=0
export OMP_NUM_THREADS=1
export FORT_BUFFERED=yes
export I_MPI_PIN_PROCESSOR_LIST=0-15
rm -f host.list
cat $LSB_DJOB_HOSTFILE > ./host.list
env > log.env
mpiexec.hydra -np $MPI -f ./host.list -genvall -ppn $PPN  $EXE 2>&1 | tee -a $OUT_FILE
```

Working with LSF – JOB Administration

- **Kill a running Job**
  - **$ bkill < jobid>**
- **See the status of your jobs in the Queue**
  - **$ bjobs**
- **See status of all user jobs**
  - **$ bjobs –u all**
- **See Information about Queues**
  - **$ bqueues**
- **See Resources on hosts**
  - **$ bhosts**

Software's and Libraries - VASP

- **Environment Variables in .bashrc**
  - source /snufs/intel/composer_xe_2015.2.164/bin/compilervars.sh intel64
  - source /snufs/intel/impi/5.0.3.048/intel64/bin/mpivars.sh
- **Location of VASP-5.3.5 Binaries.**
  - **Vasp for Hashwell**
    - /snufs/apps/vasp/5.3.5/intelmpi/vasp.hw
    - /snufs/apps/vasp/5.3.5/intelmpi/vasp.hw.gamma
    - /snufs/apps/vasp/5.3.5/intelmpi/vasp.hw.nc
  - **Vasp binaries for Sandybridge**
    - /snufs/apps/vasp/5.3.5/intelmpi/vasp.sdb
    - /snufs/apps/vasp/5.3.5/intelmpi/vasp.sdb.gamma
    - /snufs/apps/vasp/5.3.5/intelmpi/vasp.sdb.nc
- **Sample Job Script :** */snufs/apps/vasp/5.3.5/intelmpi/submit.lsf.example*

Vasp scaling

# cores	# nodes	Wall Time seconds (HSW, E5-2667V3)	Wall Time seconds (HSW, E5-2640V3)	Wall Time seconds (SDB, E5-2670)
64	4	10759.269	15310.481	17072.357

Software's and Libraries – Quantum Espresso

- **Environment Variables in .bashrc**
  - source /snufs/intel/composer_xe_2015.2.164/bin/compilervars.sh intel64
  - source /snufs/intel/impi/5.0.3.048/intel64/bin/mpivars.sh
- **Location of QE-5.1.1 Binaries.**
  - **QE Binaries for Hashwell**
    - /snufs/apps/qe/5.1.1/impi/espresso-hsw/bin
  - **QE Binaries for Sandybridge**
    - /snufs/apps/qe/5.1.1/impi/espresso-sdb/bin
  - **QE pseudo Directory**
    - /snufs/apps/qe/5.1.1/impi/pseudo_espresso
- **Sample Job Script :** */snufs/apps/qe/5.1.1/impi/submit.lsf.example*

Software's and Libraries

- **Other Softwares.**
  - **Gromacs**
  - **BigDFT**
  - **Lammps**
  - **Gaussian**
  - **USPEX**
  - **Matlab (single node)**

Q & A

Thank You