FUJITSU

shaping tomorrow with you

How to run applications on Aziz supercomputer

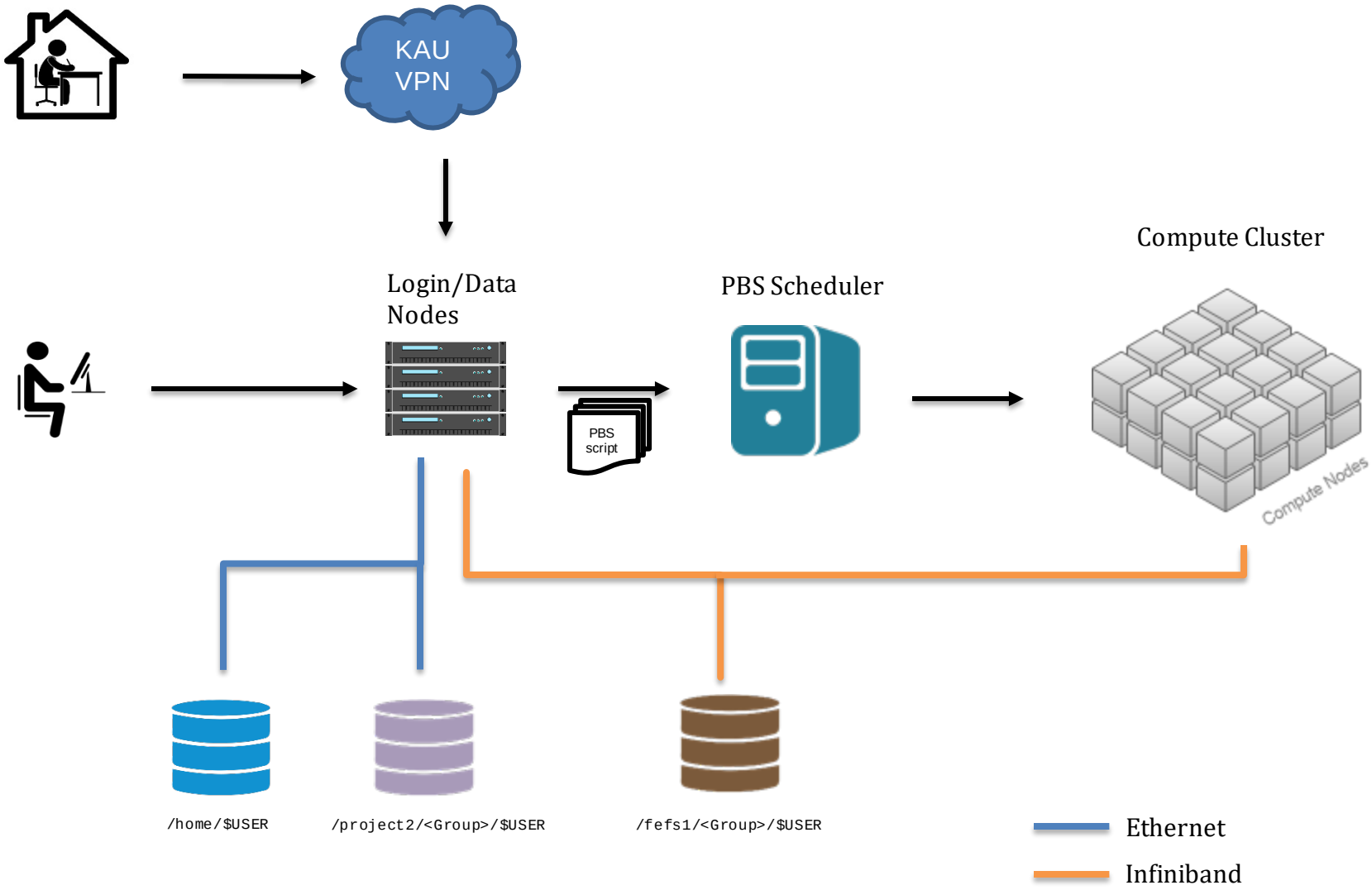
Mohammad Rafi
System Administrator
Fujitsu Technology Solutions

Agenda

FUJITSU

- Overview
- Compute Nodes
- Storage
- Infrastructure Servers
- Cluster Stack
- Environment Modules
- Interactive vs Batch Jobs
- PBS Scripts

Overview



For account request and queries, please write to hpc@kau.edu.sa

Compute Nodes

FUJITSU

Thin Nodes (thin queue)

- 380 Compute Nodes
- **96 GB** memory on each node
- Total of 9120 cores

Fat Nodes (fat queue)

- 112 Compute Nodes
- **256 GB** memory on each node
- Total of 2688 compute cores

GPU Nodes (k20 queue)

- 2 GPGPU Nodes
- 96 GB memory on each node
- Total of 4992 cuda cores

Intel Xeon Phi Nodes (phi queue)

- 2 Intel Xeon Phi Nodes
- 96 GB memory on each node
- Total of 120 cores

Storage

FUJITSU

/home/$USER

- Holds user specific files, scripts and applications
- Default quota limit is 5GB
- Accessible from login nodes and all compute nodes

/fefs1/<group>/$USER

- High-speed parallel file system with an aggregate of 2 PB of storage space
- Users infiniband interconnect to provide high-speed low-latency data transfer
- Accessible from login nodes and all compute nodes

/project2/<group>/$USER

- A total of 6 PB storage with archival system
- Accessible only from kdata1 and kdata2
- Completed projects and related data can be moved to this storage area for long term storage

Infrastructure Servers

FUJITSU

Login Nodes

- Access Aziz supercomputing facility through these nodes using ssh
- Should not be used to run user applications

Data Servers

- For data transfer purposes
- Can be used to transfer data from user system to Aziz or between different file systems in Aziz
- Project storage is accessible only from data servers

Head Nodes

- Provides management and job scheduling (PBS Scheduler) to the cluster

Management Nodes

- Provides licenses to applications running on Aziz supercomputer
- Monitoring tools and management software run on these servers

Cluster Stack

FUJITSU

Applications	Climate	Aeronautics	Genome	Chemistry	Physics
Debugging / performance tuning	Allinea DDT		GDB	Intel Vtune Amplifier	
Compilers	Intel Cluster Studio		PGI	GNU	CUDA
Message Passing	OpenMPI			Intel MPI (Cluster Studio)	
Workload Manager	Altair PBS Professional				
File System	NFS		FEFS (Fujitsu Exabyte File System)		
Cluster Manager / Monitoring	HCS	Ganglia	Server View	IPMI	
Connectivity	Infiniband		Ethernet		
Operating System	Centos 6.4		Redhat Linux 6.4		

Environment Modules

FUJITSU

- Environment modules provides dynamic modification to user's environment
- Loading and unloading an environment module is analogues to installing and uninstalling an application

Examples

To load Intel MPI module on Aziz

```
$ module load impi/5.0.3.048
```

unload Intel MPI

```
$ module unload impi/5.0.3.048
```

Load different version of Intel MPI

```
$ module load impi/5.1.1.109
```

Interactive vs Batch Jobs

FUJITSU

Batch jobs do not require user intervention during course of execution. Once job is submitted to the scheduler, resources are allocated to the job as requested in PBS script and the job runs on compute node. A batch job may contain one or more instructions which will be executed in sequence.

Interactivate jobs are usually used so that a person can become familiar with the computing environment and test or debug their code before attempting a long production run as a batch job. Applications like R is usually launched in interactive mode.

To run an interactive job, use "-I" switch of qsub command and shown in example below.

```
$ qsub –I –l select=1:ncpus=1 –q thin
```

PBS Script

FUJITSU

Example 01: PBS script to run an application

```
#!/bin/bash
#PBS –N example01
#PBS –l select=1:ncpus=1:mpiprocs=1
#PBS –q thin

cd /home/user/examples
./example01
```

Command to submit pbs script to scheduler

```
$qsub <Name of the script>
```

Command to view jobs

```
$qstat
$qstat –u <user name>
```

Command to delete job

```
$qdel <job id>
```

PBS Script

FUJITSU

Example 02: PBS script with error and output directories

```
#!/bin/bash
#PBS -N example02
#PBS -l select=1:ncpus=1:mpiprocs=1
#PBS -q thin
#PBS -e /home/user/examples/logs
#PBS -o /home/user/examples/logs

cd /home/user/examples
./example02
```

PBS Script, cont…

FUJITSU

Example 03: PBS script to run an MPI application

```
#!/bin/bash
#PBS –N example03
#PBS –l select=1:ncpus=2:mpiprocs=2
#PBS –q thin
#PBS –e /home/user/examples/logs
#PBS –o /home/user/examples/logs
#PBS –l walltime=00:20:00

cd $PBS_O_WORKDIR
module load impi/5.1.1.109
mpirun ./example03
```

PBS Script, cont…

FUJITSU

Example 04: PBS script to run an MPI application with more number of CPUS

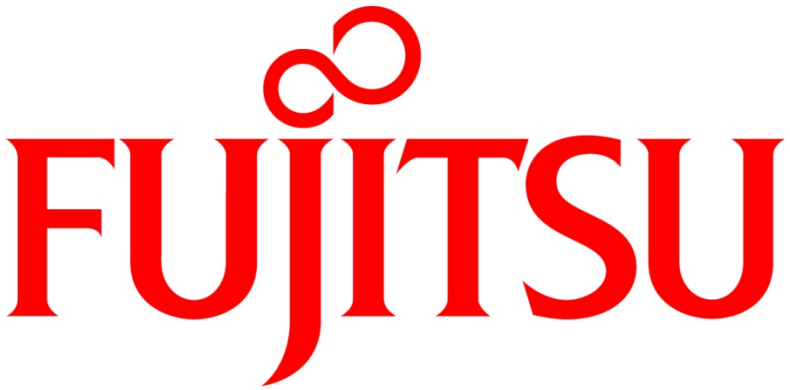
```
#!/bin/bash
#PBS –N example04
#PBS –l select=1:ncpus=24:mpiprocs=24
#PBS –q thin
#PBS –e /home/user/examples/logs
#PBS –o /home/user/examples/logs
#PBS –l walltime=00:20:00

cd $PBS_O_WORKDIR
module load impi/5.1.1.109
mpirun ./example04
```

PBS Script, cont…

FUJITSU

Example 05: PBS script to run an MPI application with multiple nodes

```
#!/bin/bash
#PBS –N mpi_example
#PBS –l select=2:ncpus=24:mpiprocs=24
#PBS –q thin
#PBS –e /home/user/examples/logs
#PBS –o /home/user/examples/logs
#PBS –l walltime=00:20:00

cd $PBS_O_WORKDIR
module load impi/5.1.1.109

echo “Job start time: $(date)”

mpirun ./example05

echo “Job end time: $(date)”

./cleanup.sh
```



شكراً, Thankyou, Merci, Danke schön, Gracias, ありがとう, 谢谢您