Учебный кластер МФТИ
Часть 2

Характеристики

- 1 головной узел и 7 вычислительных узлов
- Все узлы идентичны, имеют 4 ядра и 15 ГБ оперативной памяти
- Имена узлов: head, n01, n02, … n07
- Кластер построен на виртуальных машинах
- Операционная система – CentOS
- Система очередей – TORQUE/PBS
- Общая файловая система (NFS)

Система очередей

- Portable Batch System (PBS) – система управления распределенными вычислениями
- TORQUE – менеджер для распределенных ресурсов для вычислительных кластеров из машин под управлением Linux
- Запуск задания производится с головного узла (head), вычисления осуществляются на вычислительных узлах (n01…n07)
- PBS автоматически раскидывает задания по узлам и распределяет ресурсы
- В качестве задания выступает shell-скрипт со специальными вставками.

PBS задание

Пример PBS задания

```
#!/bin/bash

#PBS -l walltime=00:01:00,nodes=3:ppn=2
#PBS -N example_job
#PBS -q batch

cd $PBS_O_WORKDIR
mpirun  --hostfile $PBS_NODEFILE -np 6 ./a.out
```

#!/bin/bash – указание какой shell использовать.

PBS задание

Пример PBS задания

```
#!/bin/bash

#PBS -l walltime=00:01:00,nodes=3:ppn=2
#PBS -N example_job
#PBS -q batch

cd $PBS_O_WORKDIR
mpirun  --hostfile $PBS_NODEFILE -np 6 ./a.out
```

Строки, начинающиеся с #PBS, являются служебными и задают опции PBS очереди.

PBS задание

Пример PBS задания

```
#!/bin/bash

#PBS -l walltime=00:01:00,nodes=3:ppn=2
#PBS -N example_job
#PBS -q batch

cd $PBS_O_WORKDIR
mpirun  --hostfile $PBS_NODEFILE -np 6 ./a.out
```

-l walltime=00:01:00,nodes=3:ppn=2

- Задает запрашиваемый ресурс – ядро-часы
- walltime=00:01:00 – время работы приложения в формате чч:мм:сс (у нас – не более 10 мин)

PBS задание

Пример PBS задания

```
#!/bin/bash

#PBS -l walltime=00:01:00,nodes=3:ppn=2
#PBS -N example_job
#PBS -q batch

cd $PBS_O_WORKDIR
mpirun  --hostfile $PBS_NODEFILE -np 6 ./a.out
```

-l walltime=00:01:00,nodes=3:ppn=2

- Задает запрашиваемый ресурс – ядро-часы
- nodes=3 – число запрошенных вычислительных узлов (у нас – от 1 до 7)

PBS задание

Пример PBS задания

```
#!/bin/bash

#PBS -l walltime=00:01:00,nodes=3:ppn=2
#PBS -N example_job
#PBS -q batch

cd $PBS_O_WORKDIR
mpirun  --hostfile $PBS_NODEFILE -np 6 ./a.out
```

-l walltime=00:01:00,nodes=3:ppn=2

- Задает запрашиваемый ресурс – ядро-часы
- ppn=2 – число запрошенных ядер на каждом узле (у нас – от 1 до 4)

PBS задание

Пример PBS задания

```
#!/bin/bash

#PBS -l walltime=00:01:00,nodes=3:ppn=2
#PBS -N example_job
#PBS -q batch

cd $PBS_O_WORKDIR
mpirun  --hostfile $PBS_NODEFILE -np 6 ./a.out
```

-N example_job

Имя задачи. Под таким именем она будет видна в планировщике и это имя будет использоваться в названии выходных файлов. Задается пользователем.

PBS задание

Пример PBS задания

```
#!/bin/bash

#PBS -l walltime=00:01:00,nodes=3:ppn=2
#PBS -N example_job
#PBS -q batch

cd $PBS_O_WORKDIR
mpirun  --hostfile $PBS_NODEFILE -np 6 ./a.out
```

-q batch

Имя очереди. Изменению не подлежит.

PBS задание

Пример PBS задания

```
#!/bin/bash

#PBS -l walltime=00:01:00,nodes=3:ppn=2
#PBS -N example_job
#PBS -q batch

cd $PBS_O_WORKDIR
mpirun  --hostfile $PBS_NODEFILE -np 6 ./a.out
```

После служебных строк следует скрипт, который, когда подойдет очередь, будет выполняться на узле

Постановка задания в очередь

Команда qsub

Пример:

qsub job.sh

Каждое задание имеет уникальный целочисленный идентификатор ID.

По завершению работы задания будут созданы два выходных файла в текущей директории с именами example_job.oID и example_job.eID, где example_job – название задания, указанное в скрипт-файле.

Файл example_job.oID содержит в себе stdout работы скрипта, а файл example_job.eID - stderr работы скрипта.

Мониторинг заданий в очереди

Команда qstat

```
Job id          Name             User      Time Use S Queue
--------------- ---------------- --------- -------- - -----
25.localhost    my_job           kolya            0 R batch
26.localhost    my_job           kolya            0 R batch
27.localhost    my_job           kolya            0 R batch
28.localhost    my_job           kolya            0 R batch
29.localhost    my_job           kolya            0 R batch
```

Колонка **S** – статус задания.

- **Q** – задание поставлено в очередь
- **R** – задание исполняется
- **C** – задание завершено

Удаление задания из очереди

Команда qdel

Пример:

qdel 25

Один пользователь может держать в очереди не более 5 заданий.

Ограничение на память одного задания – 1 ГБ.

Компиляция mpi программ:

для C: mpicc имя_исходника –o имя_исполняемого

для C++: mpic++ имя_исходника –o имя_исполняемого