

Учебный кластер МФТИ

Н. И. Хохлов

МФТИ, Долгопрудный

14 сентября 2016 г.

- Адрес

- head.vdi.mipt.ru
- Доступ по протоколу ssh

- Характеристики

- 1 головной узел (head, head.vdi.mipt.ru) и 7 вычислительных узлов
- Все узлы идентичны, имеют 4 ядра и 15 ГБ оперативной памяти
- Адреса узлов n01, n02 ... n07
- Кластер построен на виртуалках
- Операционная система – CentOS
- Система очередей – Torque/PBS
- Общая файловая система (NFS)

Example (Доступ)

```
ssh mylogin@head.vdi.mipt.ru
```

- **Portable Batch System (PBS)** – система управления распределенными вычислениями
- **TORQUE** – менеджер распределенных ресурсов для вычислительных кластеров из машин под управлением Linux, одна из современных версий PBS
- Запуск задания осуществляется с головного узла (head), вычисления производятся на вычислительных узлах
- PBS автоматически раскидывает задания (использует ssh/rsh) по узлам и распределяет ресурсы
- В качестве задания выступает shell-скрипт со специальными вставками

Example (Пример PBS задания (job.sh))

```
#!/bin/bash  
  
#PBS -l walltime=00:01:00,nodes=7:ppn=1  
#PBS -N example_job  
#PBS -q batch  
  
hostname
```

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

- **-l walltime=00:01:00,nodes=7:ppn=1**
 - Задает размер запрашиваемых ресурсов – процессоро-часов
 - walltime=00:01:00 – время работы приложения в формате чч:мм:сс
 - nodes=7:ppn=1 – число аллоцированных ядер CPU.
 - Параметр nodes в нашем случае всегда может варьироваться от 1 до 7, параметр ppn может варьироваться от 1 до 4
 - Общее число ядер (потоков) есть $nodes \times ppn$
- **-N example_job**
 - Название задачи. Под таким названием она будет видна в планировщике и такое название будут иметь выходные файлы
- **-q batch**
 - Название очереди, в нашем случае не меняется

После специальных конструкций задается скрипт, который будет выполняться на узлах.

Постановка задания в очередь

qsub – команда для постановки задачи в очередь

Example (Постановка задания в очередь)

```
qsub job.sh
```

Каждое задание имеет уникальный целочисленный идентификатор. По завершению работы задания будут созданы два выходных файла в текущей директории, под названиями **example_job.oID** и **example_job.eID**, где **example_job** – название задания, указанное в скрипт файле, **ID** – уникальный целочисленный идентификатор задания, назначенный ему на этапе запуска. Файл **example_job.oID** содержит в себе stdout работы скрипта, **example_job.eID** – stderr.

Мониторинг заданий в очереди

qstat – просмотр текущих заданий в очереди

```
[kolya@head mpi]$ 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 выполняется
- **C** – задание завершено

Удаление задания из очереди

qdel – удаление задания из очереди, принимает на вход ID задания

Example (Удаление задания)

```
qdel 25
```


- Один пользователь может ставить максимум 5 заданий
- Время выполнения одного задания максимум 10 минут
- Ограничение на память одного задания составляет 1 ГБ

Запуск MPI приложений

- Компиляция осуществляется на головной узле
- `cd $PBS_O_WORKDIR` – перейти в папку с заданием
- Число потоков задается вручную

Example (Пример задания)

```
#!/bin/bash

#PBS -l walltime=00:01:00,nodes=1:ppn=3
#PBS -N my_job
#PBS -q batch

cd $PBS_O_WORKDIR
mpirun --hostfile $PBS_NODEFILE -np 3 ./hello
```

Запуск OpenMP приложений

- Компиляция осуществляется на головной узле
- `cd $PBS_O_WORKDIR` – перейти в папку с заданием
- Число потоков задается вручную
- Параметр `nodes` всегда равен 1

Example (Пример задания)

```
#!/bin/bash

#PBS -l walltime=00:01:00,nodes=1:ppn=3
#PBS -N my_job
#PBS -q batch

cd $PBS_O_WORKDIR
export OMP_NUM_THREADS=$PBS_NUM_PPN
./hello
```

Запуск гибридных OpenMP/MPI приложений

- Компиляция осуществляется на головной узле
- `cd $PBS_O_WORKDIR` – перейти в папку с заданием
- Число потоков задается вручную
- Параметр `nodes` равен числу MPI потоков
- Параметр `ppn` равен числу OpenMP

Example (Пример задания)

```
#!/bin/bash
```

```
#PBS -l walltime=00:01:00,nodes=3:ppn=3
```

```
#PBS -N my_job
```

```
#PBS -q batch
```

```
cd $PBS_O_WORKDIR
```

```
export OMP_NUM_THREADS=$PBS_NUM_PPN
```

```
mpirun --hostfile $PBS_NODEFILE -pernode ./hello
```

Спасибо за внимание! Вопросы?