# Порядок использования учебного сегмента суперкомпьютера УГАТУ при выполнении лабораторных работ

Юлдашев А.В. (arthur@mail.rb.ru)
ст. преподаватель каф. ВВТиС

# Характеристики использовавшихся ранее учебных кластерных систем

## Alpha-кластер (2000/2005/2009)

На базе процессоров Alpha EV5
12 вычислительных узлов (500/533 MHz) + управляющий узел (633 MHz)
Суммарный объем оперативной памяти 3,25 Gb (256 Mb на узел)
Суммарный объем дисковой памяти 270 Gb
(включая внешний дисковый массив RAID5)
Коммуникационная среда Fast Ethernet (100 Mbit/s)
ОС Debian Sarge
Библиотека MPI – MPICH1
Инсталляция ~ несколько недель

## Athlon-кластер (2005/2008/2011)

На базе процессоров Athlon XP 3000+
12 вычислительных узлов (2,1 GHz) + управляющий узел (1,2 GHz)
Суммарный объем оперативной памяти 19,5 Gb (1,5 Gb на узел)
Суммарный объем дисковой памяти 480 Gb
Коммуникационная среда Gigabit Ethernet (1 Gbit/s)
ОС OpenSuSE 10
Библиотеки MPI – MPICH2, OpenMPI, Intel MPI
Инструменты ПП – Allinea DDT, OPT
Система очередей – Torque
Инсталляция ~ несколько дней средствами пакета OSCAR

# Характеристики суперкомпьютера УГАТУ (2007/2011/…)

На базе процессоров Intel Xeon 5300 QC, 2.33 GHz
266 двухпроцессорных узлов
IBM Blade Server HS21 XM
Пиковая производительность 19,832 Tflops
Суммарный объем оперативной памяти 2.15 ТБ
Коммуникационная среда Infiniband (10 Gbit/s)
Система хранения IBM GPFS
Суммарный объем дисковой памяти 26.7 ТБ
Ленточная библиотека 8.8 ТБ
Потребляемая мощность 100 КВт
ОС  Red Hat Enterprise Linux
Библиотеки MPI – Intel MPI и др.
Инструменты ПП – Allinea DDT, OPT, Intel SD Tools
Системы очередей – IBM LoadLeveler и Torque

# Выход на кластер

ip-адрес (головного узла) кластера – 194.190.227.29
имя пользователя – student

1. Передача данных (ftp/sftp/ssh)

   Linux: средствами файлового менеджера mc
   (F9->Left/Right->FTP Link; student@194.190.227.29)

   Windows: средствами файлового менеджера
   Far Manager (Alt+F1/F2->FTP)

2. Терминальный доступ (ssh)
   Linux: консольный ssh-клиент (команда ssh), например:
   ssh student@194.190.227.29

   Windows: средствами терминального клиента PuTTY

Домашняя папка пользователя student – /gpfs/home/student
В ней необходимо создать папку группы, а в ней индивидуальную папку, в которую складывать свои данные.

# Компиляция и запуск MPI-программ

1. Компиляция (используется реализация MPI – Intel MPI 4.x):

   #исполняемый файл будет называться a.out
   **> mpicc mpiprog.c**

   или

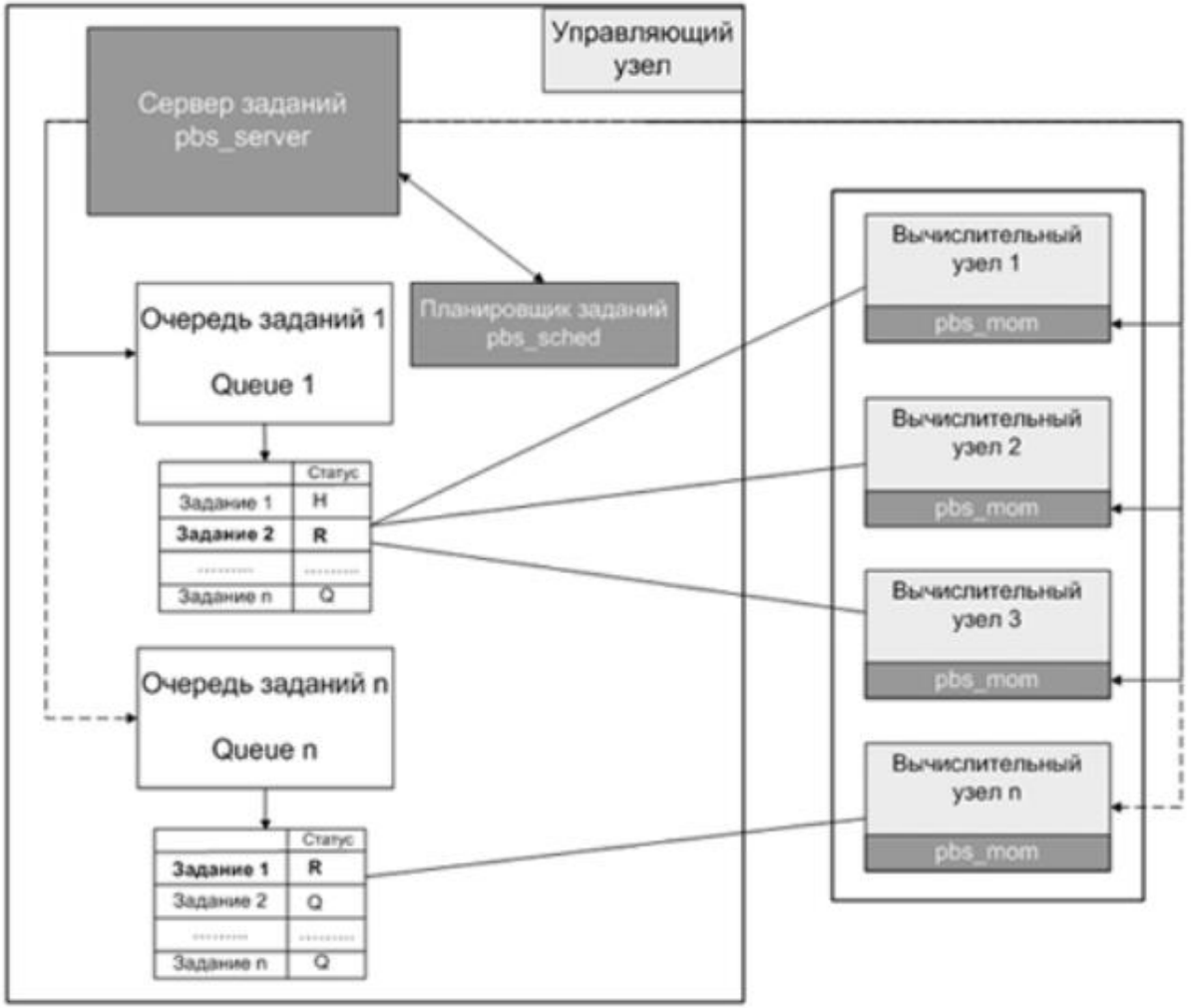
   #исполняемый файл будет называться mpiprog
   **> mpicc -o mpiprog mpiprog.c**

   или

   #для максимальной производительности
   **> mpiicc -o mpiprog mpiprog.c**

2. **Запускать MPI-программы на головном узле запрещено!** Для проведения расчетов используется система очередей/менеджер ресурсов/система пакетной обработки TORQUE.

# Система очередей TORQUE




http://hpc.ssau.ru/files/doc/torque_manual.pdf

# Постановка задач в TORQUE

Запуск осуществляется посредством постановки задачи, описанной в специальном файле - “сабмит-скрипте”, в систему очередей TORQUE командой **qsub**:

**> qsub job.pbs**

где **job.pbs** – имя “сабмит-скрипта” (пример приведен на следующем слайде).

При успешной постановке в очередь задача получает уникальный идентификатор (**JOBID**), который выводится после выполнения **qsub**, к примеру:

393827.login.nodes

Далее можно управлять задачей по JOBID – 393827.login.nodes
или короче – 393827

После запуска задачи в папке, откуда она была запущена, создаются файлы с расширением .oJOBID и .eJOBID. В них содержится информация, выводимая задачей в потоки stdout и stderr.

# Пример сабмит-скрипта

Приведенный скрипт расположен по адресу /gpfs/home/student/test/torque/job.pbs

```
#!/bin/bash
#PBS -l nodes=4:ppn=8:a:student:mpi      #ресурсный запрос, в данном примере
                                         #запрашивается 4 узла, на каждом 8 ядер
#PBS -A art                              #имя владельца задачи
#PBS -N art-test                         #имя задачи

cd ${PBS_O_WORKDIR}                      #переход в рабочую папку

date                                     #вывод текущего времени

HOSTFILE=${PBS_JOBID}.hosts              #здесь формируется файл
cat ${PBS_NODEFILE} | sort | uniq > ${HOSTFILE}   # со списком узлов для mpirun
NUM_NODES=`cat ${HOSTFILE} | wc -l`; cat ${HOSTFILE}

mpirun -r ssh -f ${HOSTFILE} -ppn 8 -n 32 ./a.out   #запуск MPI-программы
```

# Работа с системой очередей

Дополнительные команды работы с системой очередей TORQUE:

- **qstat** - просмотр состояния поставленных задач

  примеры:

  **> qstat -f JOBID** #просмотр подробной информации о задаче JOBID

  **> qstat -n1** #форматированный вывод списка задач

- **qdel JOBID** - отмена задачи
- **qalter JOBID** – изменение параметров задачи
- **qhold JOBID** - блокировка задачи
- **qrls JOBID** - вывод из состояния блокировки

- **pbsnodes** - просмотр состояния вычислительных узлов

  пример:

  **> pbsnodes -l free :student** #просмотр списка свободных узлов

По всем командам доступна справочная информация по команде **man**, например,
**> man qstat**

# Состояния задач в очереди

При просмотре состояния поставленных задач командой qstat можно увидеть следующие статусы:

C - Job is completed after having run.

E - Job is exiting after having run.

H - Job is held.

Q - Job is queued, eligible to run or routed.

R - Job is running.

T - Job is being moved to new location.

W - Job is waiting for its execution time (-a option) to be reached.