

Семинар
по суперкомпьютерному моделированию для
профессорско-преподавательского состава ЮУрГУ

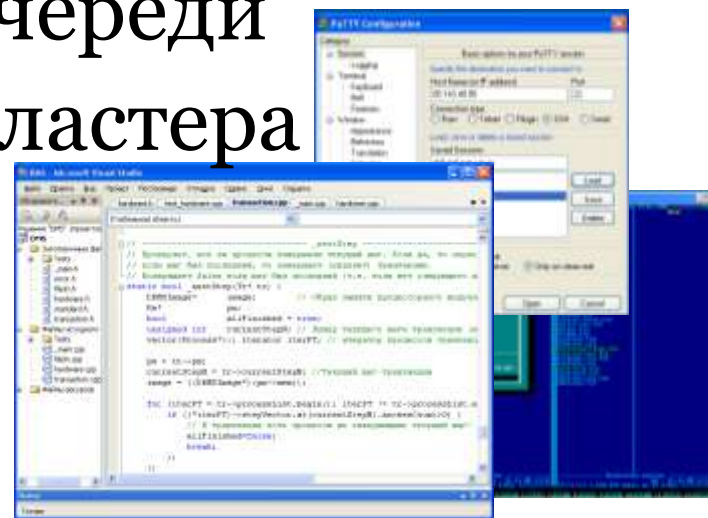
Программирование на вычислительном кластере



Зам. директора СКЦ
П.С. Костенецкий

План семинара

- Информация о кластере «СКИФ Урал»
- Подготовка программы на ПК
- Передача файлов на кластер
- Компиляция программ (C/C++/Fortran)
- Постановка задачи в очередь
- Параметры заданий в очереди
- Система мониторинга кластера



Инструкции для пользователей

[HTTP://supercomputer.susu.ru/users/instructions/](http://supercomputer.susu.ru/users/instructions/)

Суперкомпьютерный центр ЮУрГУ - Windows Internet Explorer

http://supercomputer.susu.ru/users/instructions/ Поиск "Live Search"

Суперкомпьютерный центр ЮУрГУ

СКЦ ЮУрГУ СУПЕРКОМПЬЮТЕРНЫЙ ЦЕНТР Южно-Уральского государственного университета Мир современных компьютерных технологий!

Главная Новости Информация в СМИ Ссылки Карта сайта Контакты

СУПЕРКОМПЬЮТЕРЫ

- Кластер СКИФ Урал
- Кластер Infinity

ПОЛЬЗОВАТЕЛЯМ

- Регистрация
- Техническая поддержка
- Программное обеспечение
- Конфигурация кластеров
- Справочная информация
- **Семинары**
- Вопрос - Ответ

КОММЕРЧЕСКИМ КЛИЕНТАМ

- Услуги
- Документы

О СКЦ ЮУРГУ

- История
- Задачи
- Сотрудники
- Фотографии
- Публикации

ДОКУМЕНТАЦИЯ

Правила доступа к вычислительным ресурсам СКЦ ЮУрГУ

Инструкция для пользователей кластера СКИФ Урал

- [Базовая инструкция для пользователей высокопроизводительного вычислительного кластера ЮУрГУ - СКИФ Урал](#)
- [Инструкция по работе с солвером пакета LSTC LS-Dyna](#)
- [Система управления заданиями Torque: руководство на русском языке \[в формате PDF\]](#)
- [Система управления заданиями Torque: руководство на английском языке \[в формате PDF\]](#)

Инструкция для пользователей кластера Infinity

- [Пример специализированных команд кластера](#)

Учебные пособия и инструкции

- Антонов А.С. Введение в параллельные вычисления [в формате PDF] 696 Кб
- Антонов А.С. Параллельное программирование с использованием технологии MPI: Учебное пособие [в формате PDF] 393 Кб
- Шпаковский Г.И., Серикова Н.В. Программирование для многопроцессорных систем в стандарте MPI [в формате PDF] 2167 Кб

ЮУрГУ

МехМат

Кафедра СП

WTM

Загрузчик заданий в кластер через WEB интерфейс

Вход для сотрудников СКЦ (авторизованный доступ)

© СКЦ ЮУрГУ, 2009

Регистрация пользователей СКЦ

<http://supercomputer.susu.ru/users/registration>

- Служебная записка для регистрации
- Данные для регистрации

ЗАРЕГИСТРИРОВАТЬ
 _____ 2009 г.

Директору
 Служебного центра
 ЮУрГУ
 Т.В. Сидельников

СЛУЖЕБНАЯ ЗАПИСКА

1. Целью

О регистрации пользователей
 вычислительных ресурсов
 Служебного центра ЮУрГУ

Прошу предоставить доступ к вычислительным ресурсам
 Служебного центра ЮУрГУ следующим сотрудникам:

1. _____

2. _____

3. _____

Доступ необходимо для

 (подпись)

Данные для регистрации пользователей
 вычислительных ресурсов СКЦ ЮУрГУ

Фамилия Имя Отчество	
Кафедра (подразделение)	
Должность	
Контактный телефон	
Адрес E-mail	
Предпочтительный язык (русский/английский/иной язык)	
IP-адрес	
Название учебного заведения	
Научный руководитель проекта	
Web-страница проекта (адрес)	
Полное имя	

Характеристики кластера СКИФ Урал

- Число вычислительных узлов/процессоров/ядер:
166/332/1328
- Тип процессора:
Intel Xeon E5472 (4 ядра по 3.0 GHz)
- Оперативная память: **1.33 TB**
- Дисковая память: **49.29 TB**
- Параллельная система хранения данных:
T-Platforms ReadyStorage – 20TB
- Тип системной сети:
InfiniBand (20 Gbit/s, макс. задержка 2 μ s)
- Тип управляющей сети: **Gigabit Ethernet**
- Сервисная сеть: **СКИФ ServNet**
- Пиковая производительность: **16 TFlops**
- Производительность на тесте Linpack: **12.2 TFlops**
- Операционная система:
- **SUSE Linux Enterprise Server 10**
- **Windows HPC Server 2008**
- Система бесперебойного электропитания:
APC Symmetra 160 kVA



Библиотеки MPI

Программирование для кластера производится с использованием интерфейс передачи сообщений **MPI** (Message Passing Interface).

MPI – программный интерфейс (API) для передачи информации, который позволяет обмениваться сообщениями между компьютерами, выполняющими одну задачу.

MPICH и **MPICH2** – базовые реализации стандарта MPI.

MVAPICH – реализация с поддержкой InfiniBand.

OpenMPI – новая реализация MPI со встроенной поддержкой разнообразного оборудования.

У узлов кластера нет общей памяти, поэтому для программирования используется MPI.

Запуск готового MPI-проекта в MS Visual Studio

Открываем «Мой компьютер» (через меню Пуск)
Заходим на диск M:\

Находим в папку:
«M:\Пример MPI\Перемножение матриц\»

Запускаем проект:
«MATRIX»



Обозреватель решений ...

Решение "MATRIX" (проектов: MATRIX)

- Заголовочные файлы
- Файлы исходного кода
 - matrix.cpp
- Файлы ресурсов

Окна

- ▶ Начать отладку F5
- ▶ Запуск без отладки Ctrl+F5
- ▶ Присоединиться к процессу...
- ▶ Исключения... Ctrl+Alt+E
- ▶ Шаг с заходом F11
- ▶ Шаг с обходом F10
- ▶ Точка останова F9
- ▶ Создать точку останова
- ▶ Удалить все точки останова Ctrl+Shift+F9

```

using row-wise splitting

#include <...>
#include <...>
#include <...>

#define TAG_INIT 31337
#define TAG_RESULT 42
#define DISP_MAXORDER 12

int getRowCount(int rowsTotal, int mpiRank, int mpiSize);
int matrixMultiply(double *A, double *B, double *C, int n, int n_local);

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

    /******
    int n = 1000; // Размерность перемножаемых квадратных матриц
    /******

    int n_ubound, n_local, n_sq, i;
    int mpiRank = 0, mpiSize = 1;
    double *A, *B, *C;
    double t;
    int sizeSent, sizeToBeSent;

    MPI_Init(&argc, &argv);

    MPI_Comm_rank(MPI_COMM_WORLD, &mpiRank);
    MPI_Comm_size(MPI_COMM_WORLD, &mpiSize);

    /* Get n and broadcast it to all processes */
    if (mpiRank == 0) {

```

Вывод

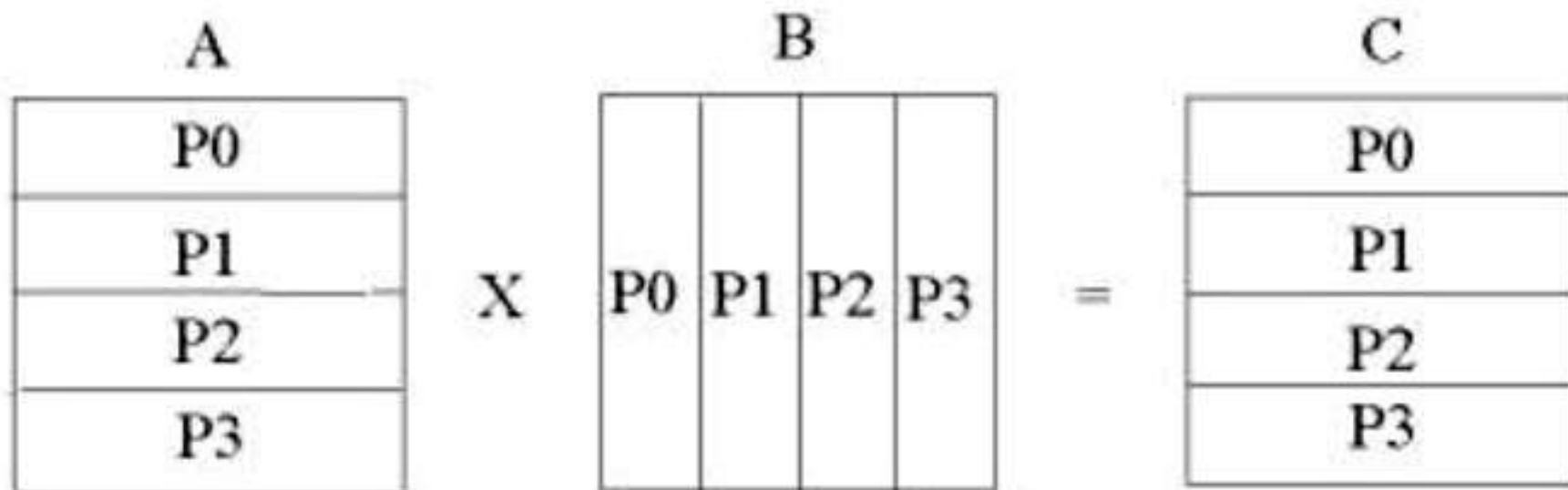
Показать выходные данные от: Отладка

```

"MATRIX.exe": Загружено "C:\Windows\System32\pnprnp.dll"
"MATRIX.exe": Загружено "C:\Windows\System32\winrnr.dll"
"MATRIX.exe": Загружено "C:\Windows\System32\rasadhlp.dll"
Первый этап обработки исключения в "0x76386e39" в "MATRIX.exe": 0x40010005: Control-C.
Программа "[532] MATRIX.exe: Машинный код" завершилась с кодом 0 (0x0).

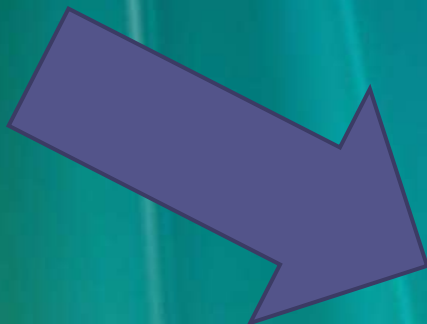
```


Параллельный алгоритм перемножения квадратных матриц

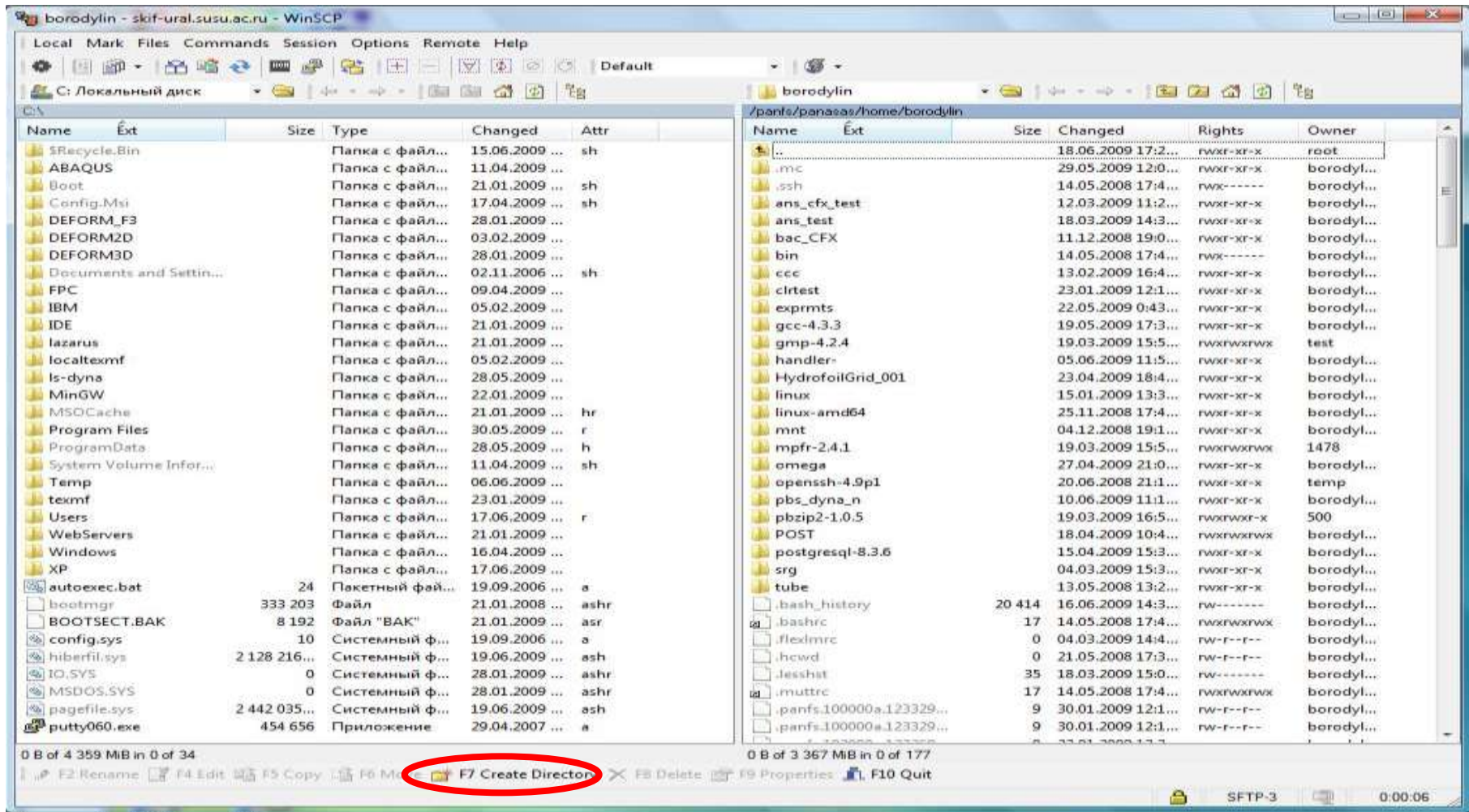


Распределение строк и столбцов матриц по процессорам.

Передача файлов на кластер производится при помощи программы WinSCP

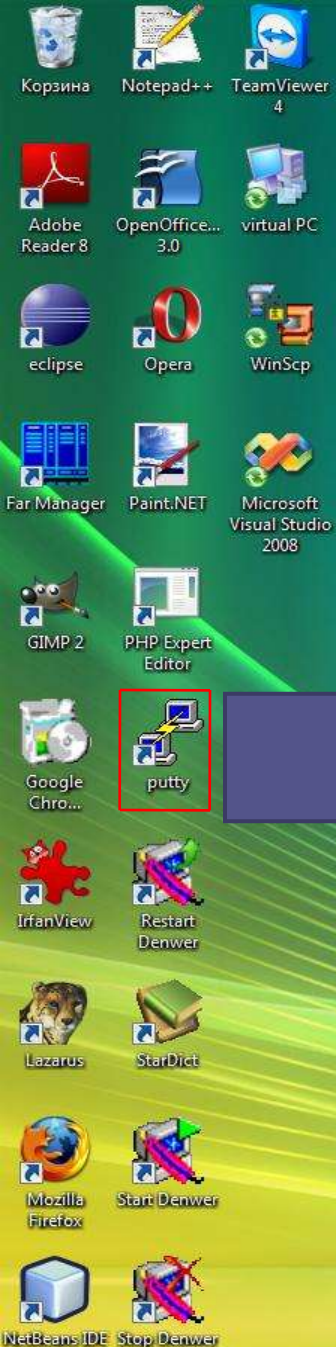


Передача файлов на кластер (и обратно) WinSCP

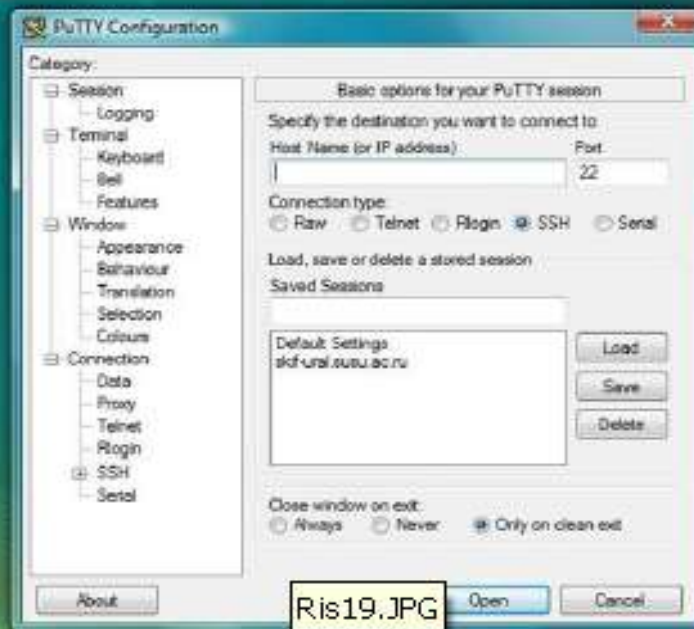


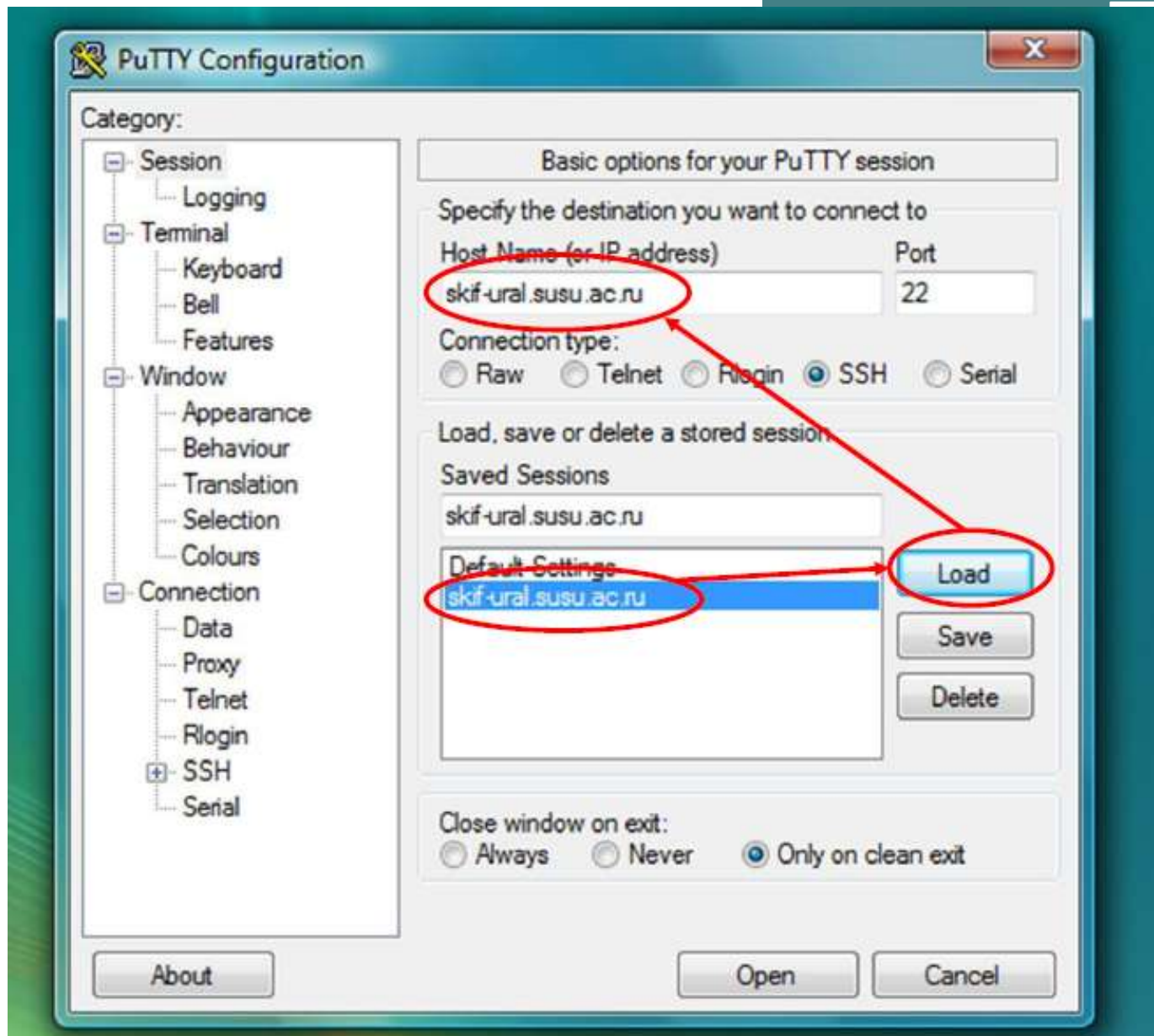
Создайте в своей директории папку projects
и скопируйте в нее файл matrix.c

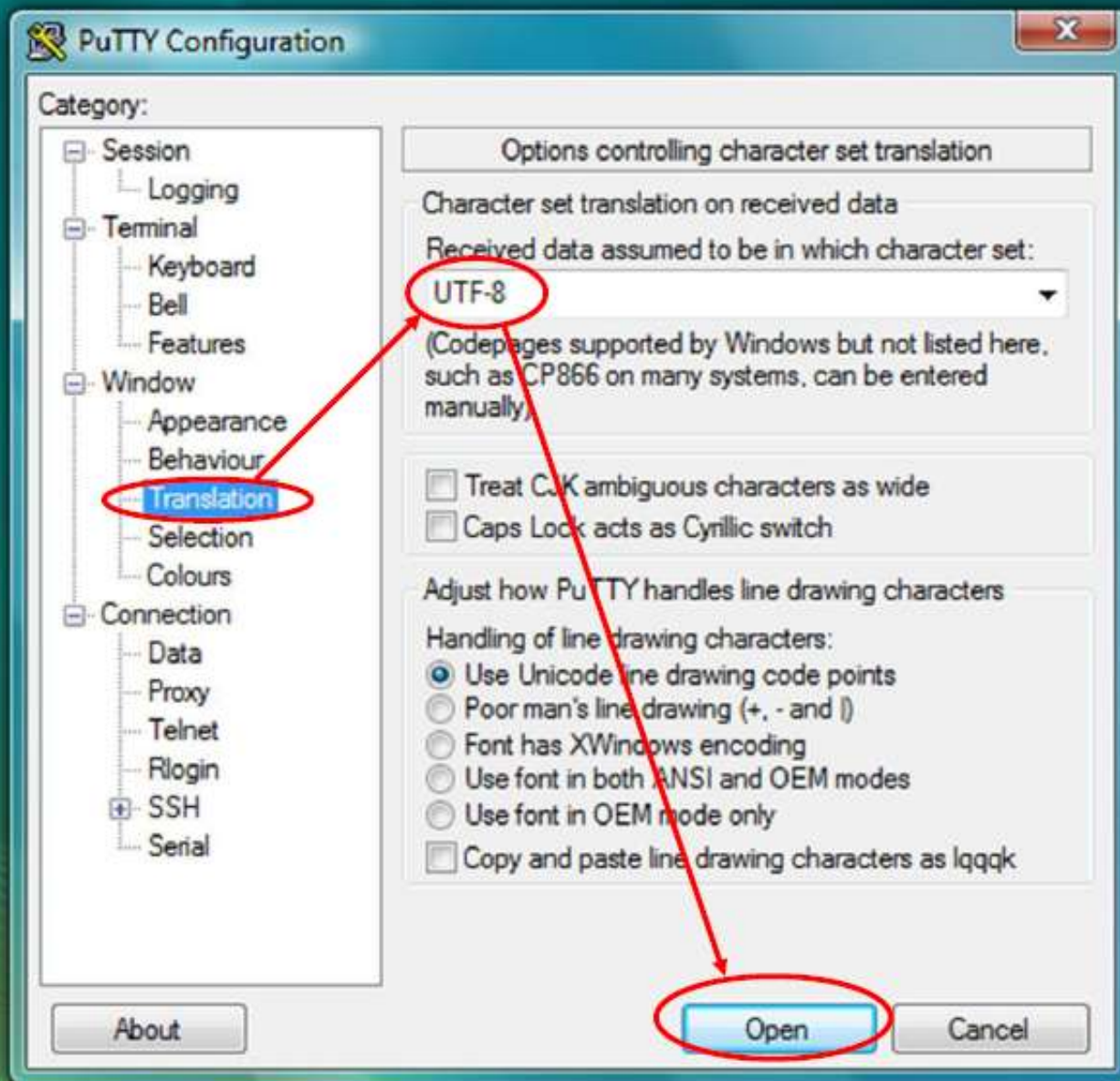
© СКМ:Пример MPI\Перемножение матриц\ex1\matrix.c



Запуск задачи на выполнение в параллельном режиме на кластере «СКИФ Урал» через систему очередей производится с помощью программы
putty







Подключение к вычислительному кластеру через удаленный терминальный доступ PuTTY

```

85.143.49.95 - PuTTY
login as:

```

Login: fpk.
Password: ...

```

85.143.49.95 - PuTTY
login as: dolganina
You are welcome to the 'SKIF Ural' cluster of the Southern Ural State University
http://supercomputer.susu.ru, e-mail: supercomputer@susu.ru, phone:(351) 2679089

Using keyboard-interactive authentication.
Password:
Last login: Thu Apr 23 14:08:59 2009 from mimi.mim.susu.ac.ru
dolganina@umaster1: ~

```

mc

```

mc - /panfs/panasas/home/dolganina
Left  File      Command  Options  Right
<- /panfs/panasas/home/dolganina -v>
Name      Size      MTime
/..      UP--DIR
/.mc      4096     Apr 23 14:13
/.ssh     4096     Mar 13 17:00
/10x10_Polotno 8192     Mar 19 14:28
/5x5_Polotno 4096     Apr 16 11:33
/Spherule 8192     Apr 16 19:15
/bin      4096     Mar 13 17:00
.bash_history 946     Apr 23 14:14
@.bashrc  17       Mar 13 17:00
.muttrc   6043     Mar 13 17:00
@.profile 18       Mar 13 17:00
/..
Note: Shell commands will not work when you are on a non-local file system.
dolganina@umaster1 /panfs/panasas/home/dolganina $
1:Help 2:Menu 3:View 4:Edit 5:Copy 6:RenMov 7:Mkdir 8>Delete 9:PullDn 10:Quit

```


mc (Midnight Commander)

The screenshot shows the Midnight Commander (mc) interface. The title bar reads "mc - /panfs/panasas/home/dolganina". The interface is split into two panes, both showing a directory listing for the same path: </panfs/panasas/home/dolganina>. The listing includes columns for Name, Size, and MTime. The files listed are: /.. (UP--DIR), /.mc (4096, Apr 23 14:13), /.ssh (4096, Mar 13 17:00), /10x10_Polotno (8192, Mar 19 14:28), /5x5_Polotno (4096, Apr 16 11:33), /Spherule (8192, Apr 16 19:15), /bin (4096, Mar 13 17:00), .bash_history (946, Apr 23 14:14), @.bashrc (17, Mar 13 17:00), .muttrc (6043, Mar 13 17:00), and @.profile (18, Mar 13 17:00). At the bottom, a status bar shows the current directory path and a list of keyboard shortcuts: 1Help, 2Menu, 3View, 4Edit, 5Copy, 6RenMov, 7Mkdir, 8Delete, 9PullDn, 10Quit.

```

mc - /panfs/panasas/home/dolganina
Left      File      Command  Options  Right
< /panfs/panasas/home/dolganina v>  < /panfs/panasas/home/dolganina v>
Name      Size      MTime    Name      Size      MTime
/..       UP--DIR   /..       4096     Apr 23 14:13
/.mc      4096     Mar 13 17:00
/.ssh     4096     Mar 13 17:00
/10x10_Polotno 8192     Mar 19 14:28
/5x5_Polotno 4096     Apr 16 11:33
/Spherule 8192     Apr 16 19:15
/bin      4096     Mar 13 17:00
.bash_history 946     Apr 23 14:14
@.bashrc  17       Mar 13 17:00
.muttrc   6043     Mar 13 17:00
@.profile 18       Mar 13 17:00
/..
/..
Note: Shell commands will not work when you are on a non-local file system.
dolganina@umaster1 /panfs/panasas/home/dolganina $
1Help 2Menu 3View 4Edit 5Copy 6RenMov 7Mkdir 8Delete 9PullDn 10Quit

```

Конвертация исходного кода из кодировки CP1251 в UTF8

```
iconv -f cp1251 -t utf8 ./matrix.cpp > m.cpp
```

Обратная конвертация

```
iconv -f utf8 -t cp1251 ./m.cpp > m2.cpp
```

Перекодировка требуется для правильного отображения русских символов на кластере.

Компиляция программы на кластере

```
mpicc ./matrix.cpp -o ./matrix
```

```
mpif90 ./matrix.f -o ./matrix
```

Процесс компиляции

```
$ mpicc ./matrix.cpp -o ./matrix  
./matrix.cpp(58): (col. 9) remark: LOOP WAS VECTORIZED.  
./matrix.cpp(86): (col. 5) remark: LOOP WAS VECTORIZED.
```

На кластере установлены оптимизирующие компиляторы Intel, позволяющие повысить производительность приложений за счет оптимизации кода для выполнения на процессорах Intel Xeon

В той же директории, где располагался исходный код программы, появился скомпилированный исполняемый файл matrix

Определение числа свободных узлов

qfree

```
$ qfree
```

```
Free nodes: 21
```

То есть свободно: $21 \text{ узел} * 8 \text{ ядер} = 168$
вычислительных ядер

Параметры команды cl-run

LS-Dyna:

```
cl-run -as dyna971 -np 80 mytestname i=./SAMPLE.k  
memory=400m memory2=20m
```

ANSYS Mechanical

```
cl-run -as ansys -np 15 testname _test -i=test.log
```

ANSYS CFX

```
cl-run -as cfx -np 15 testname -def ./test-file.def
```

ABAQUS

```
cl-run -as abaqus -np 5 testname ws_scr_tennis.inp
```

МРІ-приложения

```
cl-run -as mpi -np 120 testname ./program
```

cl-run - - help

Запуск скомпилированного приложения на кластере

```
cl-run -as mpi -np 16 test ./matrix
```

Состояние задач в очереди

qstat – список задач;

qstat -a – список задач с расширенной информацией;

qstat -f <номер задачи> – полная информация о задаче (на каких вычислительных узлах выполняется, какой командой запущена и т.д.).

qdel <номер задачи> – удаление задачи из очереди

Управление выполнением задачи

Для просмотра текущего состояния задачи используется команда **qstat** :

```
$ qstat
```

Job id	Name	User	Time	S	Queue
1510.umaster1	test	ivanov	01:15:05	R	batch
1511.umaster1	test2	ivanov	01:15:05	R	batch

Статус задачи принимает следующие значения

- **R** - задача выполняется
- **Q** - задача в очереди
- **E** - задача прервана (с ошибкой)
- **C** - задача выполнена

Для удаления задачи используется команда **qdel {номер задачи}**.

```
$ qdel 1510
```


Задания

Сравните время выполнения задачи на персональном компьютере и на разном числе ядер кластера.

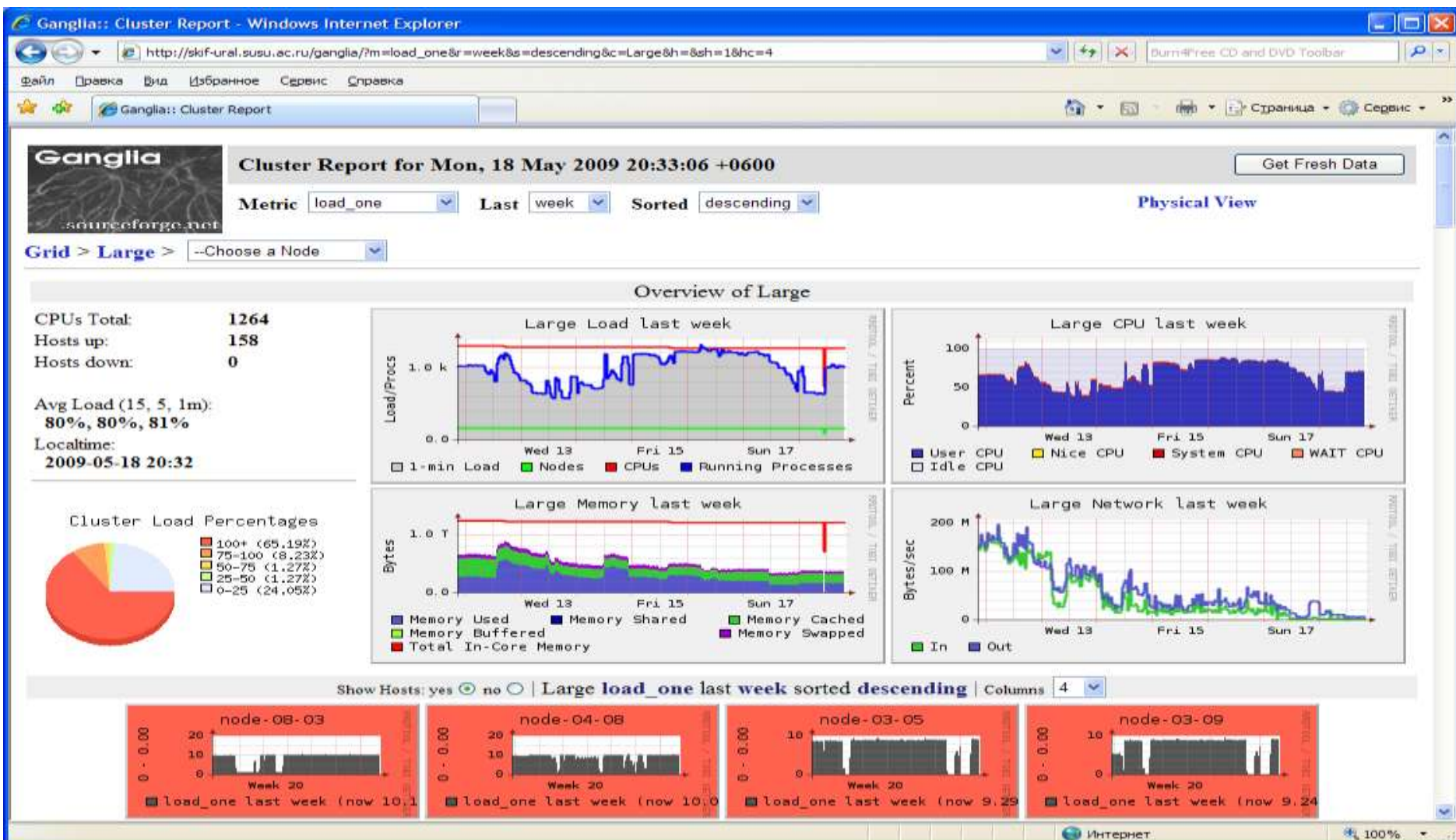
Измените размерность матриц на кластере (режим редактирования в MS включается при нажатии клавиши F4).

После изменения исходного кода не забывайте перекомпилировать программу.

В каждом узле кластера 8 ядер, так что оптимально запускать приложение на числе ядер кратном 8 (8,16,24,32,40,...)

Система мониторинга кластера

<http://skif-ural.susu.ac.ru/ganglia>



Литература

- Суперкомпьютерный центр ЮУрГУ
<http://supercomputer.susu.ru>
- Информационно-аналитический центр по параллельным вычислениям
<http://parallel.ru>
 - *А.С. Антонов. Параллельное программирование с использованием технологии MPI.*
М.: Учебное пособие, изд-во МГУ, 2004, 72 с.
 - *А.С. Антонов. Введение в параллельные вычисления.*
М.: Методическое пособие, изд-во Физического факультета МГУ, 2002, 70 с.

Спасибо за внимание!