

РОССИЙСКАЯ АКАДЕМИЯ НАУК
ДАЛЬНЕВОСТОЧНОЕ ОТДЕЛЕНИЕ



ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР

Гибридный вычислительный кластер ВЦ ДВО РАН: архитектура, программное обеспечение, особенности использования

Мальковский С.И., н.с., ЛИТС ВЦ ДВО РАН



Гибридный вычислительный кластер



Кластер состоит из 5 вычислительных узлов со следующими характеристиками (каждый узел):

- 2 десятиядерных процессора IBM POWER8 2.86 ГГц (всего 160 потоков);
- память ECC, 256 ГБ;
- 2 x 1 ТБ 2.5" 7K RPM SATA HDD;
- 2 x NVIDIA Tesla P100 GPU, NVLink.



Сеть передачи данных: EDR InfiniBand.

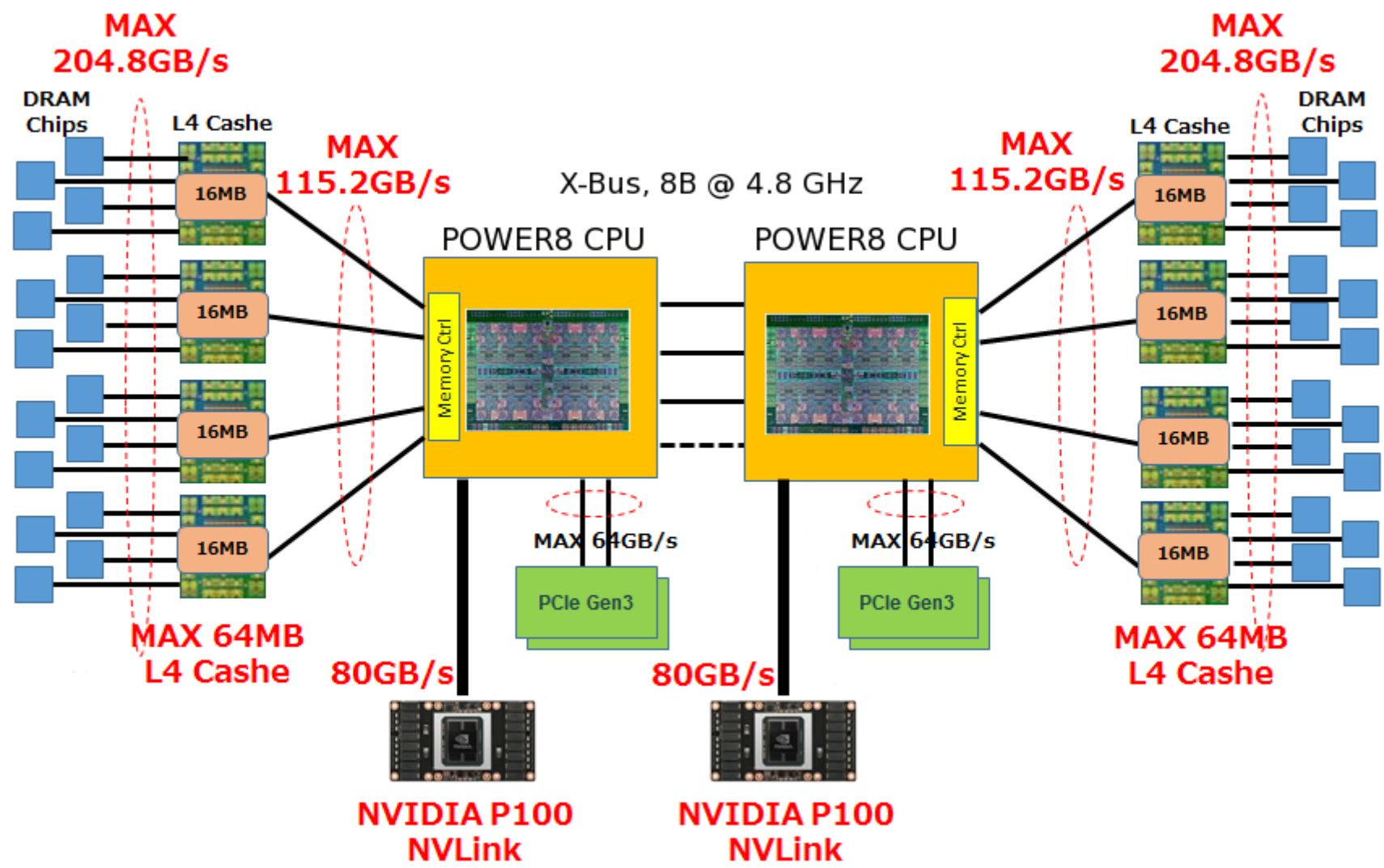
Управляющая сеть: Gigabit Ethernet.

Хранилище: 15+ ТБ

Энергопотребление: 10 кВт



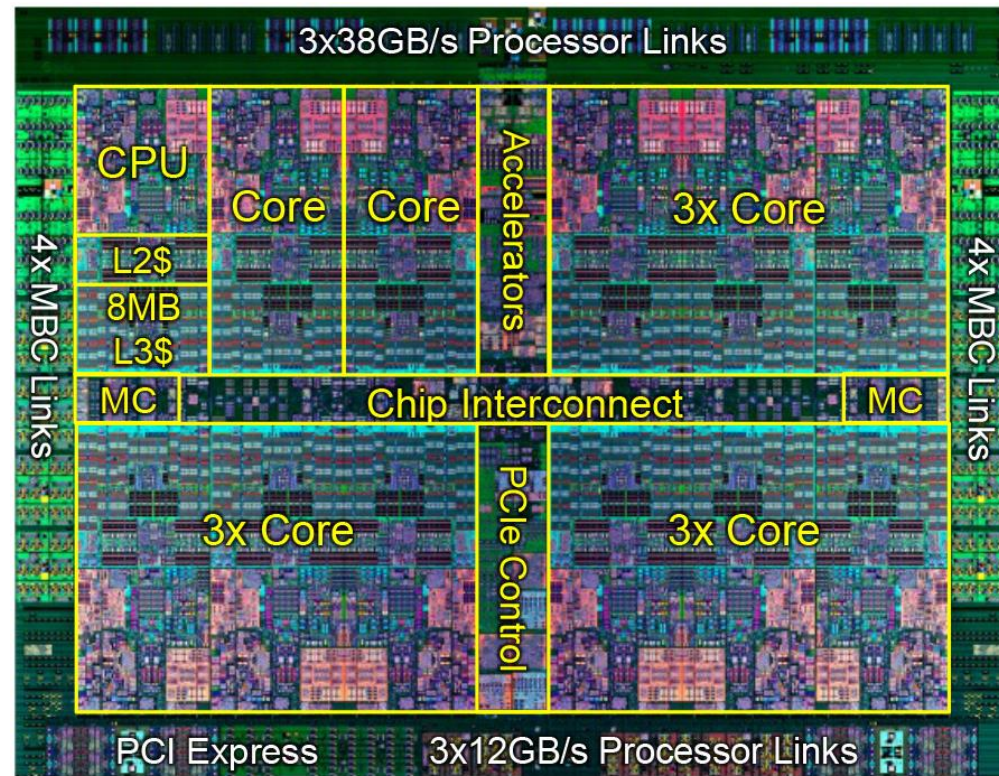
Архитектура вычислительных узлов





Процессор IBM POWER8

- 10 вычислительных ядер с частотой 2,83/4,023 ГГц
- Поддержка SMT – до 8 потоков на ядро
- Кэш: L1 – 64К (данные), 32К (инструкции) на ядро; L2 – 512 КБ на ядро; L3 (96 МБ) и L4 (128 МБ) – разделяются между ядрами
- 16 исполнительных блоков; 4 FPU, 2 VMX (SIMD, 128 бит)
- Пиковая производительность – 0,322 ТФлопс





Сопроцессор NVIDIA P100

- 1792 вычислительных ядер с частотой 1,48 ГГц (двойная точность)
- 16 ГБ памяти HBM2 с пропускной способностью 732 ГБ/с
- Шина обмена данными с центральным процессором - NVLINK
- Пиковая производительность при вычислениях с двойной точностью – 5,3 ТФлопс
- Поддерживаемые API: CUDA, OpenCL, OpenACC, DirectCompute





Программное обеспечение

1. Операционная система – Linux CentOS 7.3
2. Программные средства параллельных вычислений стандарта MPI – библиотека IBM Spectrum MPI, OpenMPI
3. Компиляторы языков программирования – IBM XL C/C++, IBM XL Fortran, GNU C/C++, GNU Fortran, PGI C/C++, PGI Fortran
4. NVIDIA CUDA Toolkit 8.0
5. Математическая библиотека – IBM ESSL и PESSL
6. Система диспетчеризации заданий – PBS Professional



Производительность вычислительной системы

$R_{peak} = 55,83$ ТФлопс

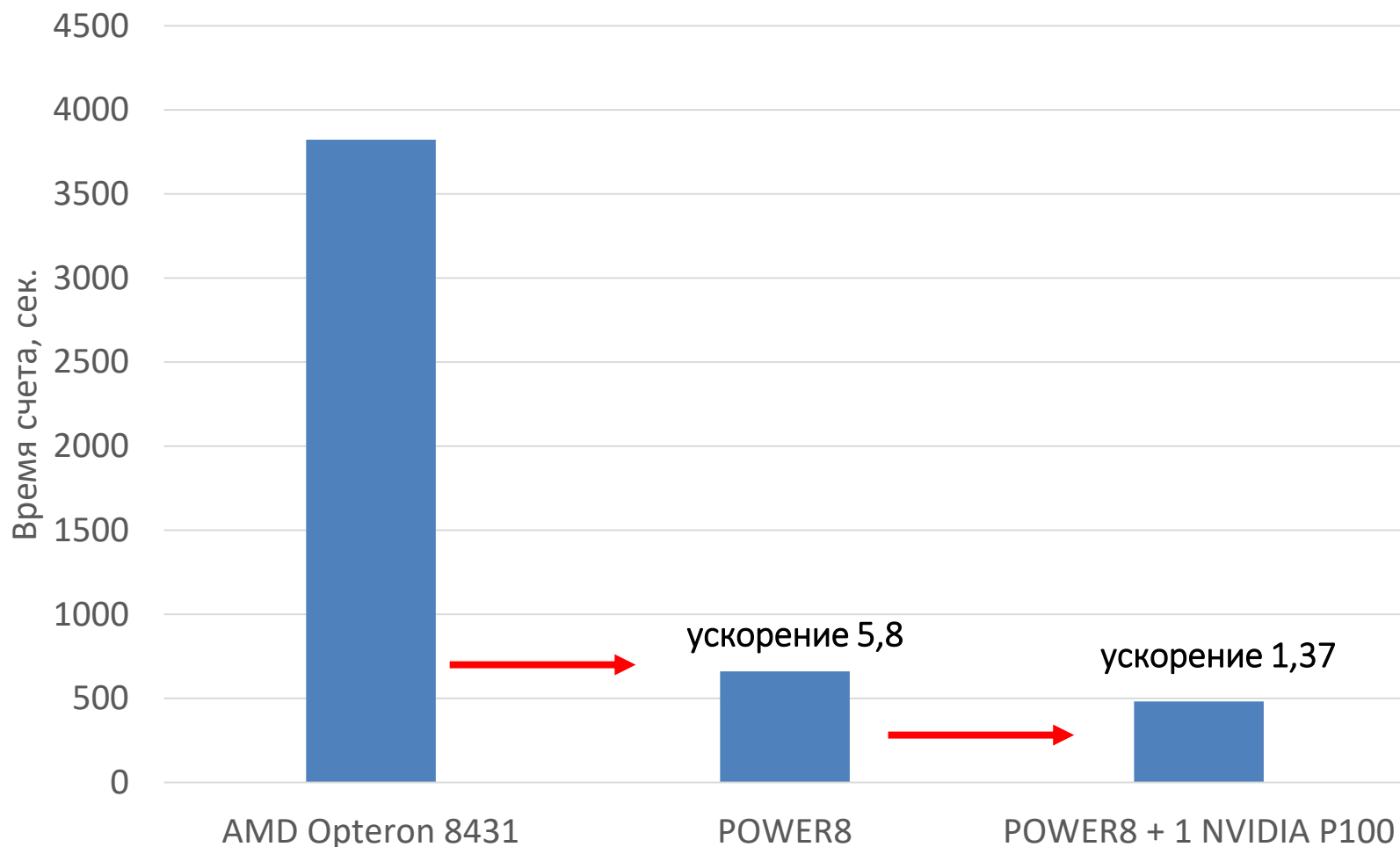
$L_{inpack} = 40,39^*$ ТФлопс (~ 72%)

* - “40 место” в 26-й редакции TOP 50 от 04.04.2017г.



Пакеты прикладных программ

Quantum ESPRESSO 5.4, бенчмарк AUSURF112, 10 MPI процессов (меньше лучше)





Программирование на GPU

Гибридная модель вычислений:

- Последовательная часть кода выполняется на CPU;
- Массивно-параллельные вычисления выгружаются на GPU (функции-ядра).

Основные техники:

- Использование оптимизированных библиотек (ESSL)
- Использование библиотек из состава CUDA Toolkit
- Директивы (OpenACC)
- CUDA/OpenCL расширения C/C++/FORTRAN

С
Л
О
Ж
Н
О
С
Т
Ь

С
К
О
Р
О
С
Т
Ь



Компиляторы C/C++ и Fortran

	IBM XL C/C++ 13.1.5, IBM Fortran 15.1.5	GNU C/C++, Fortran 4.8.5	PGI C/C++, Fortran 16.10
Стандарты	C11, C++11, C++14, Fortran 2008	C90, C99, C11, C++98, C++11, Fortran 2003, Fortran 2008	C89, C99, C11, C++11, Fortran 2003
OpenMP	3.1 (частично 4.5)	3.1	3.1
MPI	+	+	+
OpenACC	-	-	+
CUDA	+	+	+
OpenCL	+	+	+



Математическая библиотека ESSL 5.5

Состав:

- библиотека линейной алгебры (BLAS);
- решение СЛАУ (LAPACK);
- быстрое преобразование Фурье (FFTW3);
- сортировки и поиск;
- интерполяция;
- генераторы псевдослучайных чисел и т.д.

Режимы работы:

- однопоточный;
- многопоточный;
- многопоточный с выгрузкой вычислений на GPU (GPU или GPU+CPU)



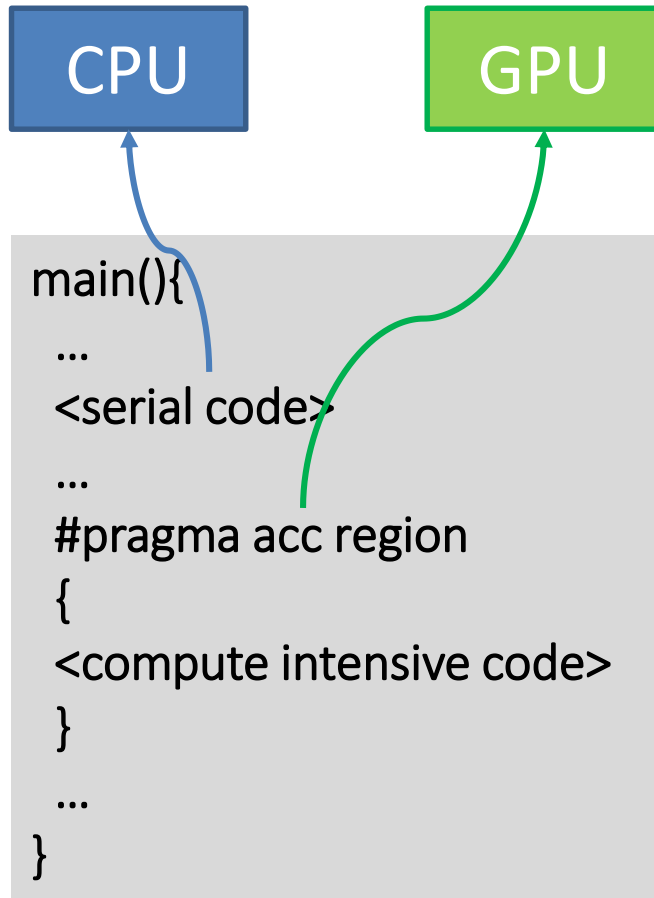
Математическая библиотека CUDA

- cuFFT – быстрое преобразование Фурье;
- cuBLAS – библиотека линейной алгебры;
- cuRAND – генераторы псевдослучайных чисел;
- cuSPARSE – работа с разреженными матрицами;
- cuSOLVER – решение СЛАУ;

Полный список: <https://developer.nvidia.com/gpu-accelerated-libraries>



OpenACC



Метки для компилятора

Компилятор параллелизует код

Работает на многоядерных CPU и массивно-параллельных GPU

Поддержка стандарта реализована в компиляторе PGI (C/C++ и Fortran)

Подробнее:

<https://developer.nvidia.com/pgi-accelerator-fortran-and-c-compilers>



Программирование на CUDA

Свой код для CPU и GPU (вычислительные функции-ядра)

Полный контроль над потоками данных и вычислений

Функции-ядра выполняются на GPU (только NVIDIA), остальной код – на CPU

Для компиляции кода используется компилятор nvcc из состава CUDA Toolkit

Подробнее: C/C++ - <https://developer.nvidia.com/how-to-cuda-c-cpp>
Fortran - <https://developer.nvidia.com/cuda-fortran>



Информационная система “Центр коллективного пользования”

ИНФОРМАЦИОННАЯ СИСТЕМА
ЦКП ЦЕНТР
КОЛЛЕКТИВНОГО ПОЛЬЗОВАНИЯ

Регистрация

Фамилия

Имя

Отчество

Место работы
Укажите место работы

Должность
Укажите должность

Email

Телефон

Введите текст с рисунка

О системе Контакты Новости Лаборатория информационно-телекоммуникационных систем ВЦ ДВО РАН

ИС ЦКП Заполнить заявку sergey.malkovsky@gmail.com Выйти

Хотите оставить заявку на услуги ЦКП?

Последняя активность

- ЦКП Центр данных ДВО РАН**
Новая заявка #69 Уравнения Максвелла с сингулярностью
Пользователь mscuster@rambler.ru
Организация ВЦ ДВО РАН
- ЦКП Центр данных ДВО РАН**
Новая заявка #68 Запрос на GPU-ресурс для расчета распространения цунами
Пользователь loskutov-imgg@yandex.ru
Организация ИМИГ ДВО РАН
- ЦКП Центр данных ДВО РАН**
Новая заявка #67 vehicle_routing
Пользователь o.dolgova@live.ru
Организация Нет организации (Частное лицо)
- ЦКП Центр данных ДВО РАН**
Новая заявка #66 12312312
Пользователь karmanno@gmail.com
Организация ИЗИ ДВО РАН
- ЦКП Центр данных ДВО РАН**
Новая заявка #65 fggjnt
Пользователь sergey.malkovsky@gmail.com
Организация ВЦ ДВО РАН

О СИСТЕМЕ КОНТАКТЫ НОВОСТИ Лаборатория информационно-телекоммуникационных систем ВЦ ДВО РАН

ИС ЦКП Заполнить заявку sergey.malkovsky@gmail.com Выйти

1. Выберите ЦКП и тип услуги

Выберите ЦКП
Центр данных ДВО РАН

Выберите услугу
Проведение численных расчетов на многопроцессорной вычислительной технике

Описание услуги
Выделение вычислительных ресурсов на вычислительных системах для проведения расчетов в различных предметных областях.

Выберите оборудование, на базе которого будет предоставляться услуга
Гибридный вычислительный кластер на базе архитектуры OpenPOWER

О СИСТЕМЕ КОНТАКТЫ НОВОСТИ Лаборатория информационно-телекоммуникационных систем ВЦ ДВО РАН

ИС ЦКП Заполнить заявку sergey.malkovsky@gmail.com Выйти

Заявка №68

Подтвердить Перевести на рассмотрение Отклонить

Назначить пользователя
loskutov-imgg@yandex.ru

Заявитель	Артём Владимирович Лоскутов
Задача	Запрос на GPU-ресурс для расчета распространения цунами
Детальное описание:	Выполнение расчета распространения волн цунами от сейсмических источников в глобальном масштабе.
Срок использования ресурсов (в месяцах):	бессрочно
Организация:	ИМИГ ДВО РАН
Статус задачи:	подтверждена
Статус электронной копии договора	подтверждена
Статус печатной копии договора	не подтверждена
Название проекта:	0
Финансирующая организация:	ФАНО России
Оборудование, на котором должна быть выполнена услуга:	Гибридный вычислительный кластер на базе архитектуры OpenPOWER

О СИСТЕМЕ КОНТАКТЫ НОВОСТИ Лаборатория информационно-телекоммуникационных систем ВЦ ДВО РАН



Запуск заданий на исполнение

Контроль за исполнением заданий на кластере осуществляется при помощи менеджера заданий **PBS Professional**. Для запуска заданий используется команда `qsub`.

Пример скрипта описания задания:

```
#!/bin/sh
#PBS -N job_name
#PBS -q workq
#PBS -j oe
#PBS -m abe
#PBS -M user@mail.com
#PBS -l select=2:ncpus=160:mpiprocs=20

module add spectrum_mpi

cd /home/user/app_dir

mpirun -np 40 -npernode 20 --hostfile $PBS_NODEFILE app_name

exit 0
```



Environment Modules

Предназначены для управления переменными среды. Используются для выбора требуемого программного обеспечения.

```
[user@jupiter ~]$ module avail
```

```
----- /etc/modulefiles -----
```

```
cuda                                openmpi/gcc/2.0.2a1/4.8.5          openmpi/xl/1.10.6
pgi/16.10(default)                  spectrum_mpi                       cudnn/5.1
openmpi/xl/2.0.2a1                   pgi/17.4                           essl
openmpi/pgi/1.10.2/2016              openmpi/xl/2.1.0                   pgi/2016
openmpi/gcc/1.10.6/4.8.5             openmpi/pgi/1.10.2/2017           pbs
openmpi/gcc/2.1.0/4.8.5              pgi/2017
```

```
[cepra@jupiter ~]$ module list
```

```
Currently Loaded Modulefiles:
```

```
1) pbs          2) cuda        3) essl        4) spectrum_mpi
```

```
[cepra@jupiter ~]$ module add pgi/16.10
```



СПАСИБО ЗА ВНИМАНИЕ!