Russian Supercomputing Days 2017, Москва, Россия
25-26 сентября 2017 года

Использование зонтичного проекта распределенных вычислений в рамках учебного курса

Курочкин И.И.
Институт проблем передачи информации РАН (ИППИ РАН), Москва

Платформы для организации распределенных вычислений

- HTCondor

- Globus

- BOINC

Платформа BOINC

BOINC – Berkeley Open Infrastructure for Network Computing

Платформа для организации добровольных распределенных вычислений:

- состоит из серверной и клиентской части;
- дает возможность задействовать вычислительные мощности персональных компьютеров(ПК);
- Кросплатформенная клиентская часть;
- Гибкая настройка клиентской части для эффективного использования ресурсов ПК.

Добровольные распределенные вычисления (ДРВ) на платформе BOINC

- 4.3 миллиона участников
- 16 миллионов компьютеров
- Около 100 международных проектов добровольных распределенных вычислений

Активность пользователей по странам

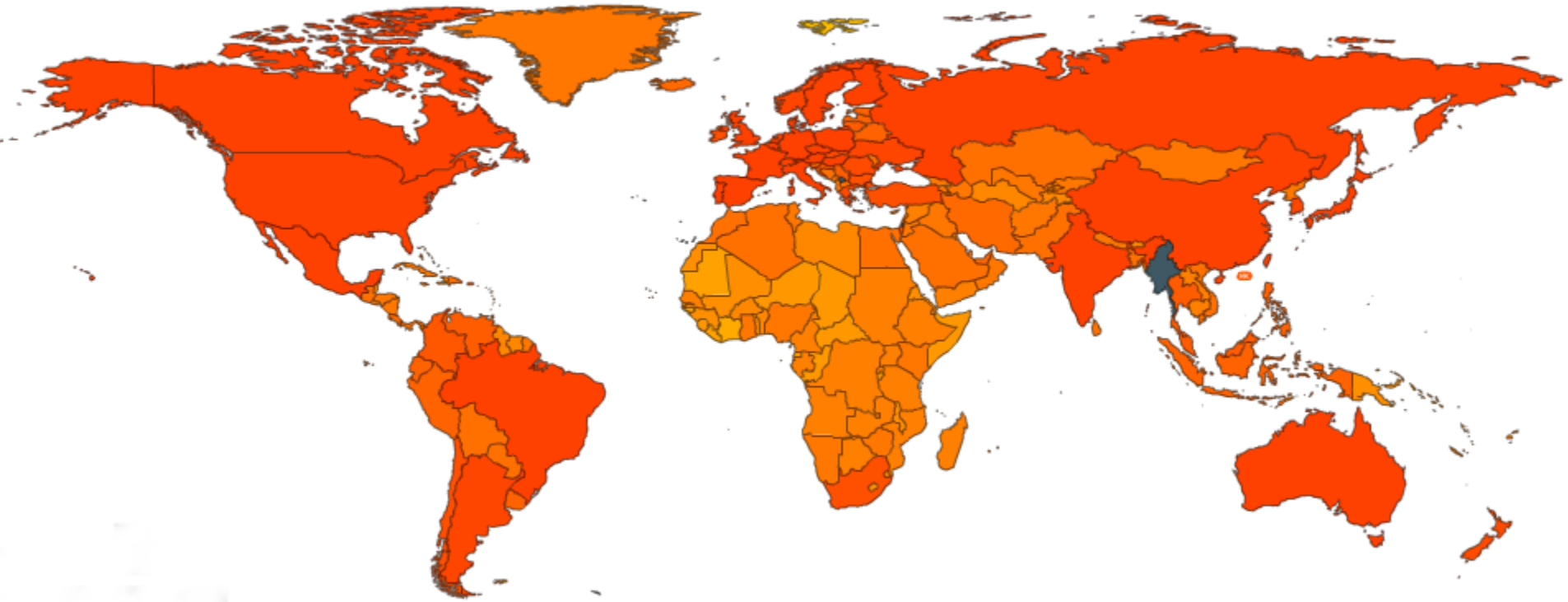
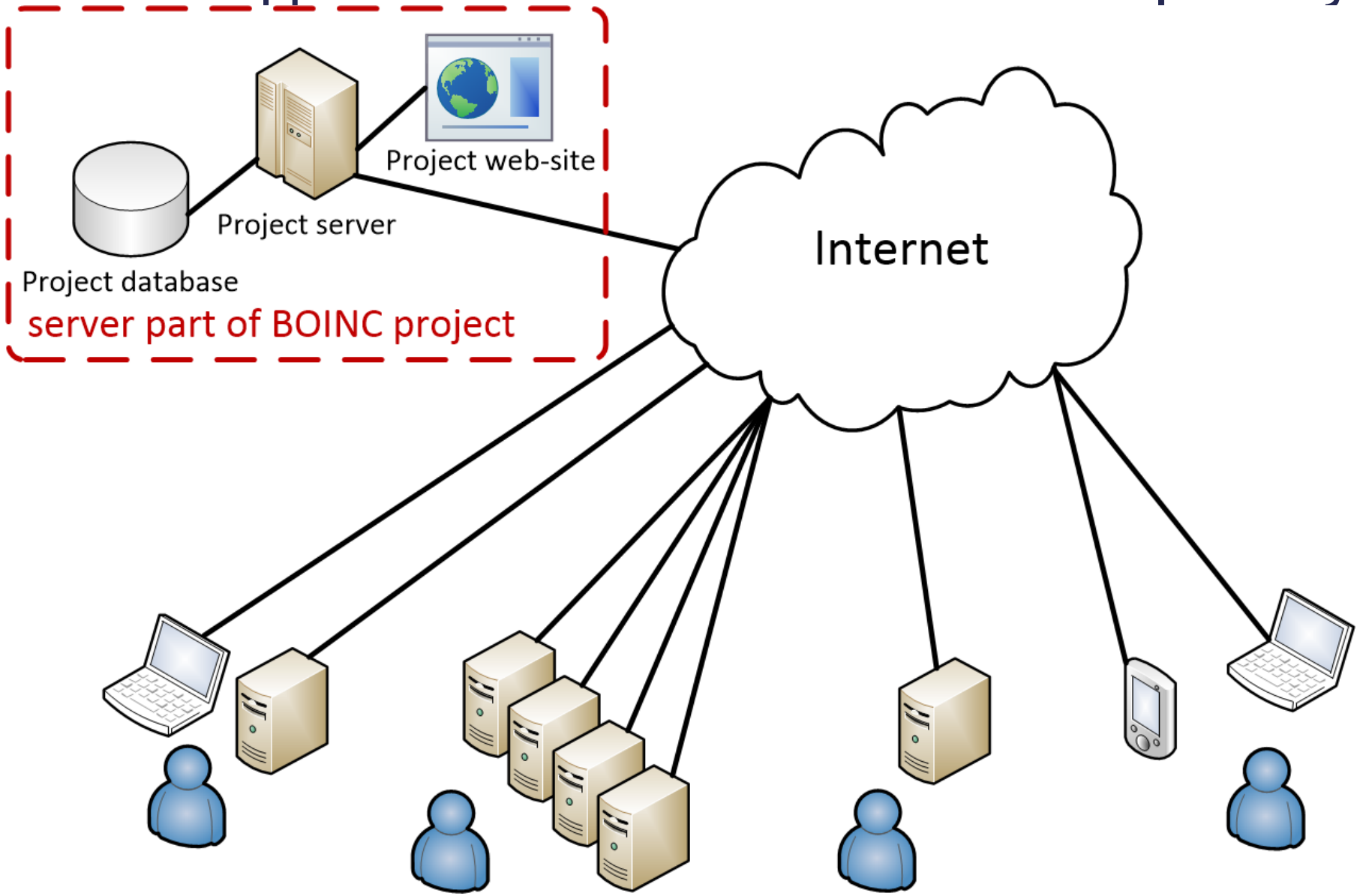
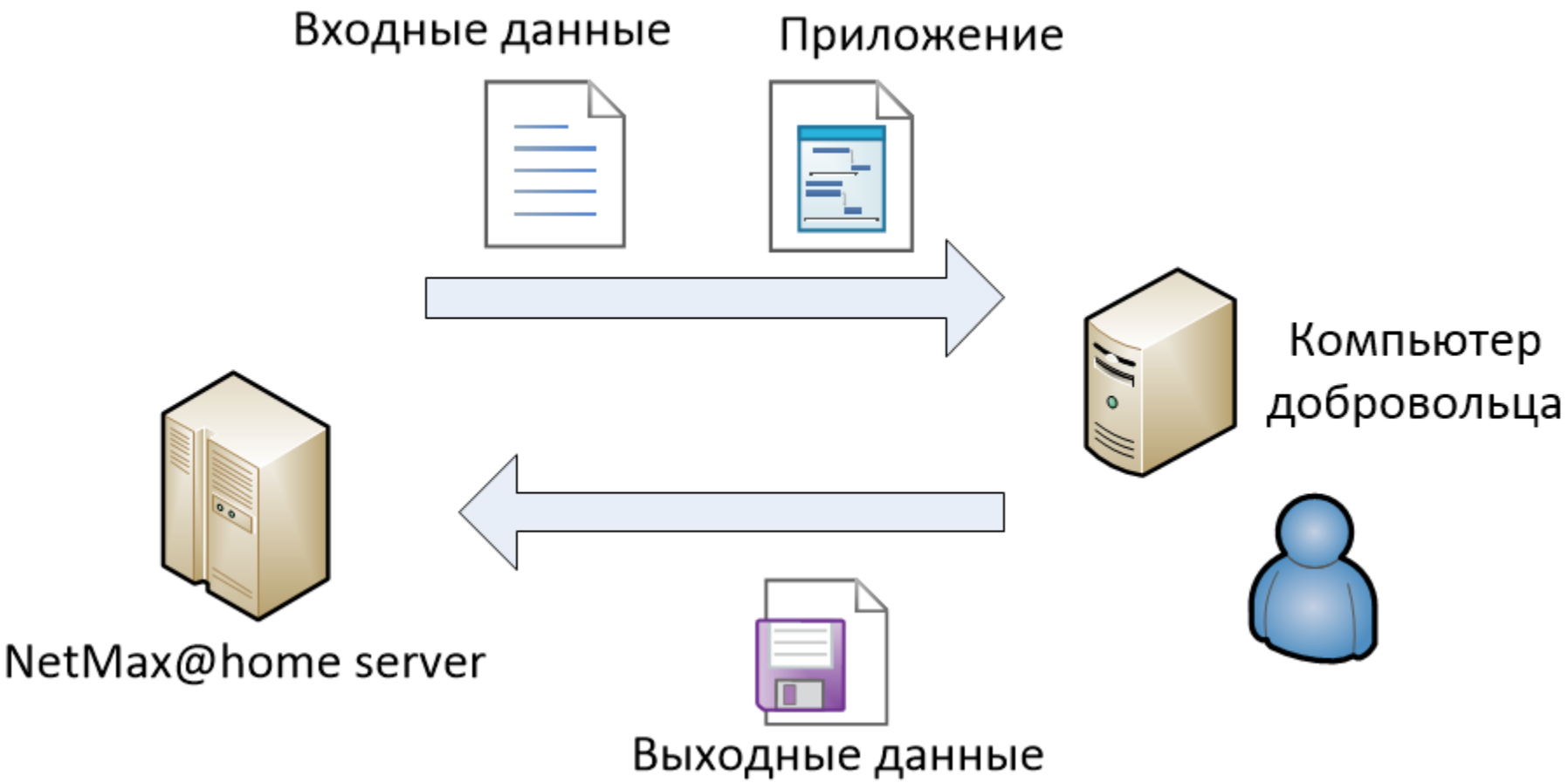


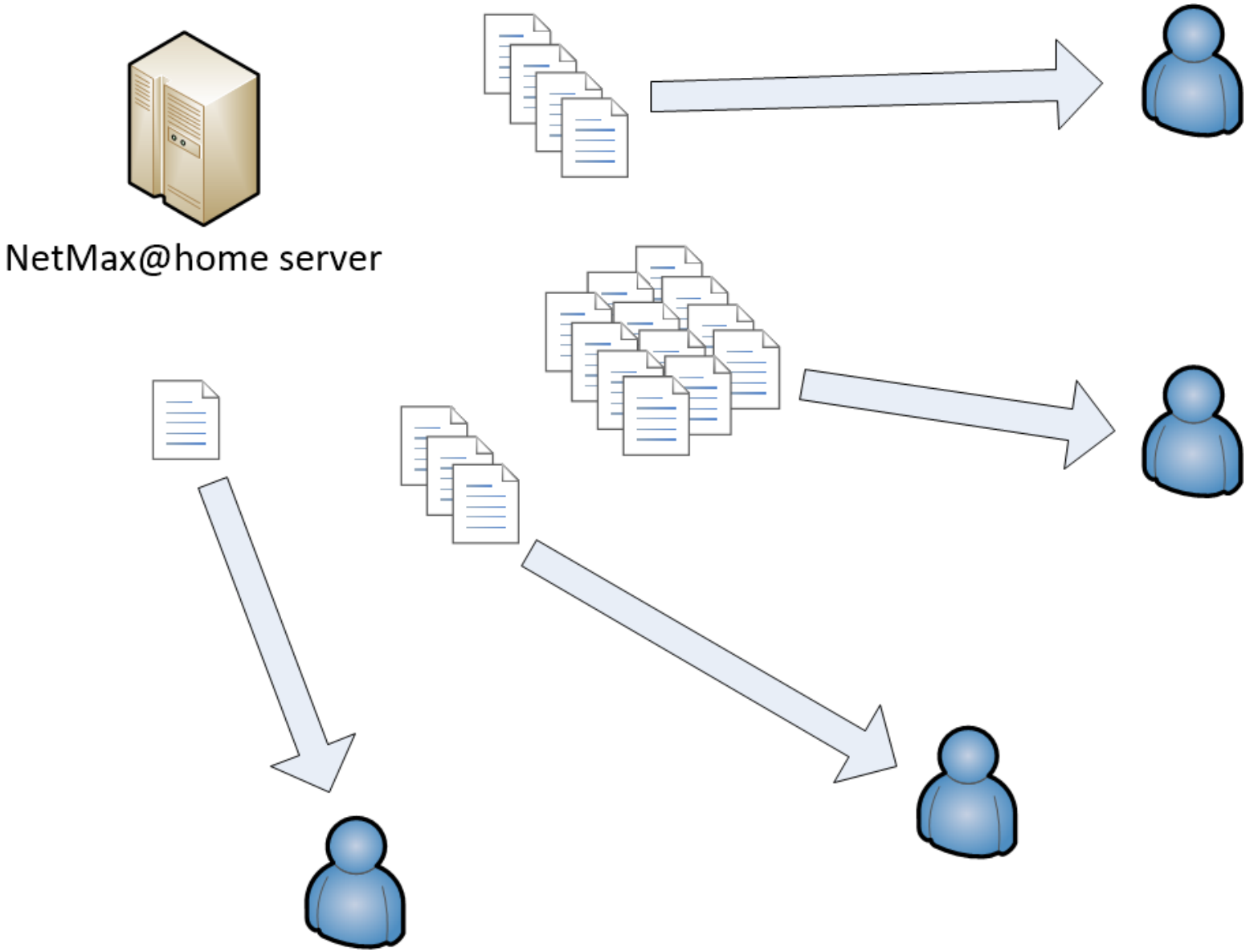
Схема подключения пользователей к проекту



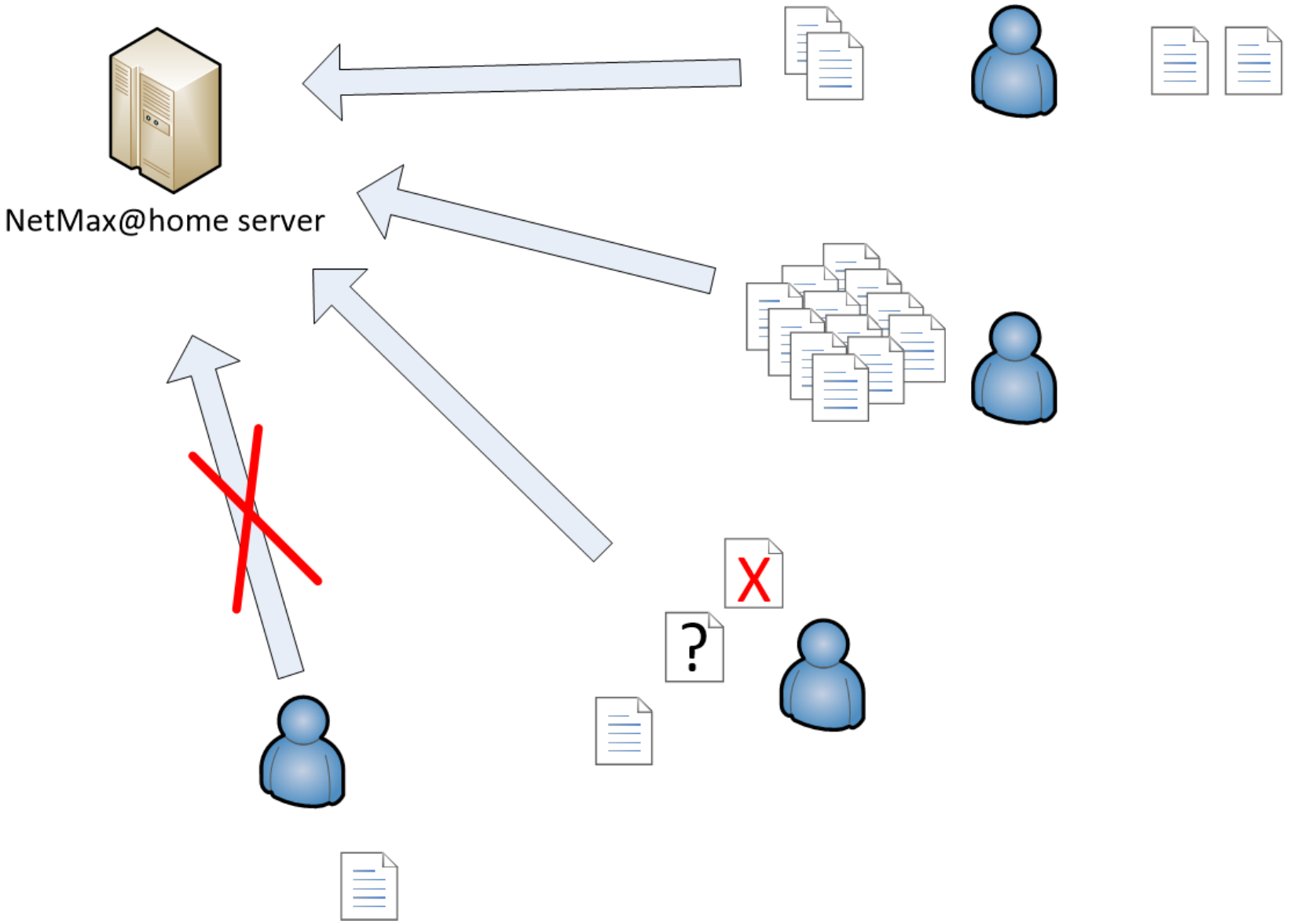
Обмен данными



Выдача заданий пользователям проекта



Получение результатов



Особенности организации экспериментов в проекте добровольных распределенных вычислений

- Необходимость взаимодействия с добровольцами
- Общение на форуме
- Научно-популярное объяснение задач проекта
- Своевременное обновление информации на сайте проекта
- Обеспечение достаточного количества заданий для постоянной работы проекта
- Большой объем результатов и входных данных
- Необходимость обработки полной серии результатов
- Минимизация вероятности ошибочного расчета подзадачи
- Необходимость начальной репликации заданий в связи ненадежностью узлов

Перечень работ по проекту ДРВ

- **1. Технические затраты**
- 1.1. Создание вычислительного приложения с сохранением промежуточных результатов
- 1.2. Создание генератора входных заданий
- 1.3. Создание валидатора и агрегатора результатов
- 1.4. Тонкая настройка параметров серверной части BOINC-проекта
- **2. Организационные затраты**
- 2.1. Регистрация домена
- 2.2. Информационный сайт проекта
- 2.3. Общее описание научной и административной группы проекта
- 2.4. Организация соревнований в проекте
- **3. Взаимодействие с добровольцами (привлечение новых ресурсов и удержание имеющихся)**
- 3.1. Научно-популярное описание научной составляющей проекта
- 3.2. Регулярное научно-популярное описание проводимых численных экспериментов
- 3.3. Регулярная публикация на сайте проекта полученных результатов
- 3.4. Взаимодействие с сообществом добровольцев
- 3.5. Ведение блога проекта и публикация ссылок на научные статьи
- **4. Дополнительные затраты**
- 4.1. Разработка и внедрение системы начисления баллов
- 4.2. Разработка настройка системы выдачи виртуальных призов
- 4.3. Дизайн сайта и информационного сайта проекта
- 4.4. Персонализация полученных промежуточных результатов, когда доброволец может узнать, что именно он считает в данный момент. Дополнительно возможна визуализация результатов.

Пример организации численного эксперимента в проекте PB

	Этап 1	Этап 2	Этап 3
Количество алгоритмов	3	5	+2 (всего 7)
Количество сетей	3696	~300	~70 (несколько групп)
Количество заданий	11 088	~1500	~240
Средний размер выходного файла	4.2 MB	534 MB	534 MB
Общий размер выходных данных	46.6 GB	801 GB	128 GB
Среднее время выполнения задания	4.7 часов	33 часа	33 часа
Общее время вычислений	5.9 лет	5.7 лет	0.9 лет

	Итог по трем этапам	Без этапов
Количество алгоритмов	7	7
Количество сетей	3696	3696
Количество заданий	12 828	25 872
Средний размер выходного файла	76 MB	534 MB
Общий размер выходных данных	0.98 TB	13.8 TB
Среднее время выполнения задания	8.5 часов	33 часа
Общее время вычислений	12.5 лет	97.4 лет

Запуск зонтичного проекта добровольных распределенных вычислений

- Возможность запускать небольшие эксперименты
- Нет необходимости разворачивания нового проекта
- Не нужно развивать проект с нуля
- Существенная вычислительная мощность на старте эксперимента

Подпроекты:

- Поиск диагональных латинских квадратов (GPU)
- Поиск диагональных латинских квадратов (Xeon Phi)
- Моделирование телекоммуникационных сетей (CPU)

Международная федерация грид-систем из персональных компьютеров

- Технологии грид-систем из персональных компьютеров (ГСПК) позволяют использовать простаивающие мощности персональных компьютеров, серверов для решения трудоемких вычислительных задач. ГСПК может использоваться в пределах одной организации или же объединять ресурсы добровольных участников из одного города, страны или даже всего мира.

- Развертывание гетерогенной инфраструктуры и разработка приложений для них являются непростыми задачами. Именно поэтому была создана Международная федерация грид-систем из персональных компьютеров.

- Международная федерация грид-систем из персональных компьютеров объединяет различные компании, университеты и исследовательские институты, а также людей заинтересованных в использовании простаивающей вычислительной мощности и желающих обменяться опытом друг с другом.

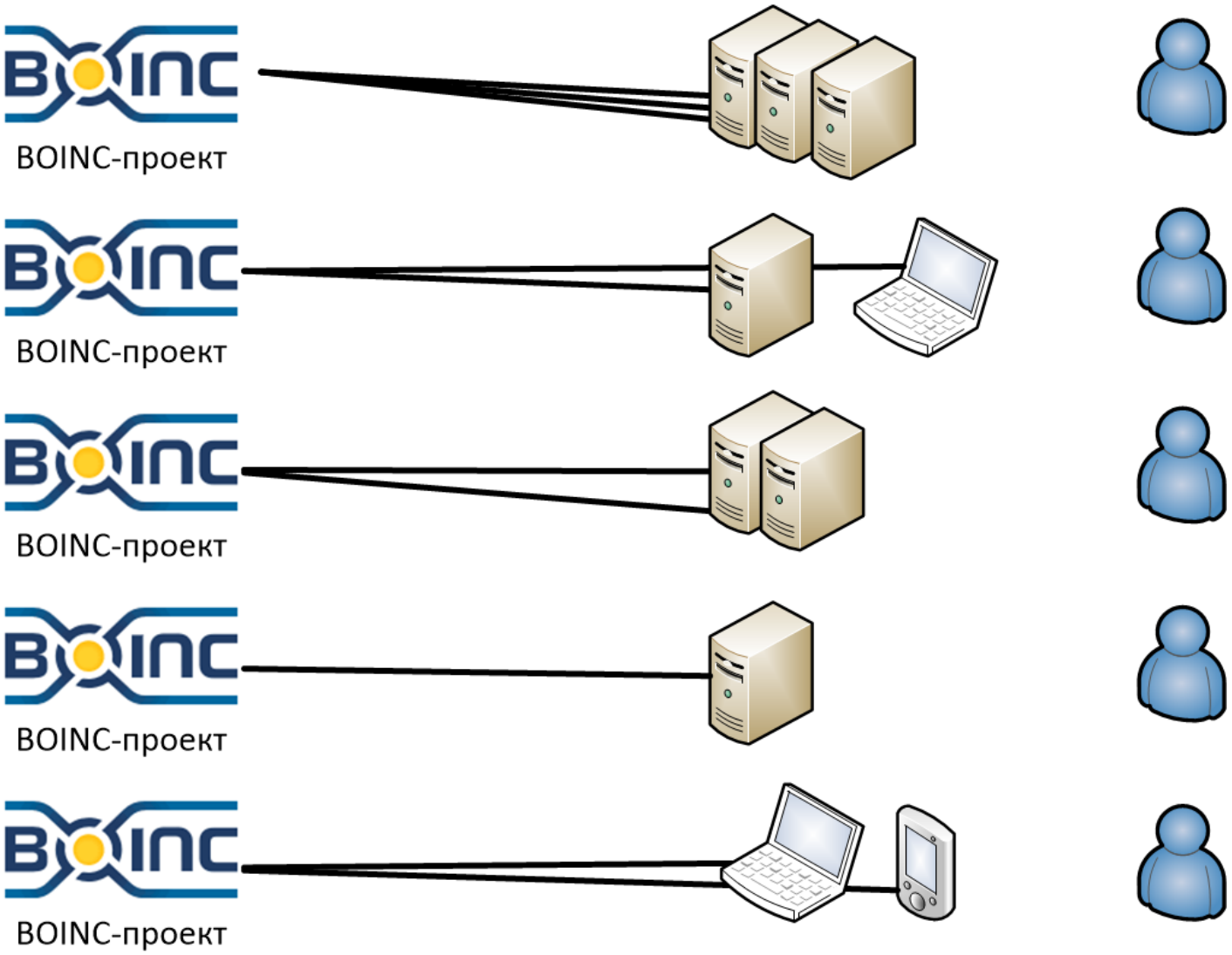
Образовательная деятельность

- Популяризация добровольных распределенных вычислений
- Обучение студентов основам распределенных вычислений на платформе BOINC
- Работа студентов и аспирантов над действующими проектами

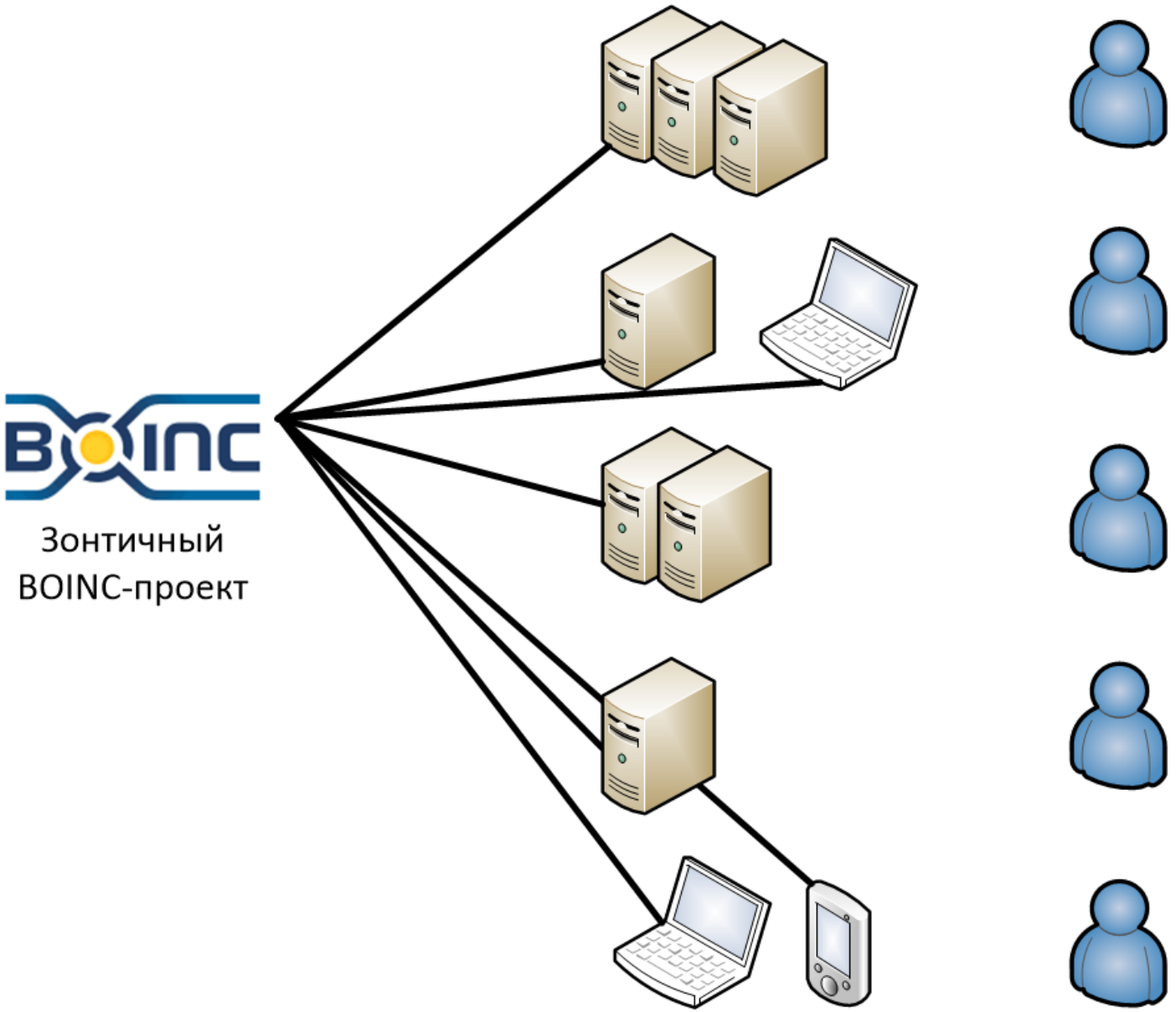
Студенты ведущих ВУЗов изучают распределенные вычислительные системы и добровольные распределенные вычисления:

- МГУ им.Ломоносова (г.Москва)
- МФТИ (г.Долгопрудный)
- МИСиС (г.Москва)
- ЮЗГУ (г.Курск)
- ИГУ (г.Иркутск)
- ПетрГУ (г.Петрозаводск)

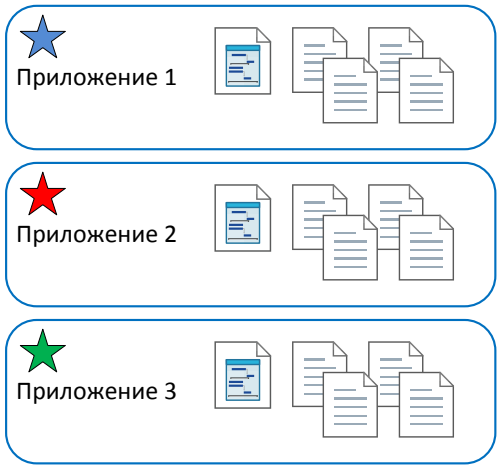
Проект РВ для каждого студента

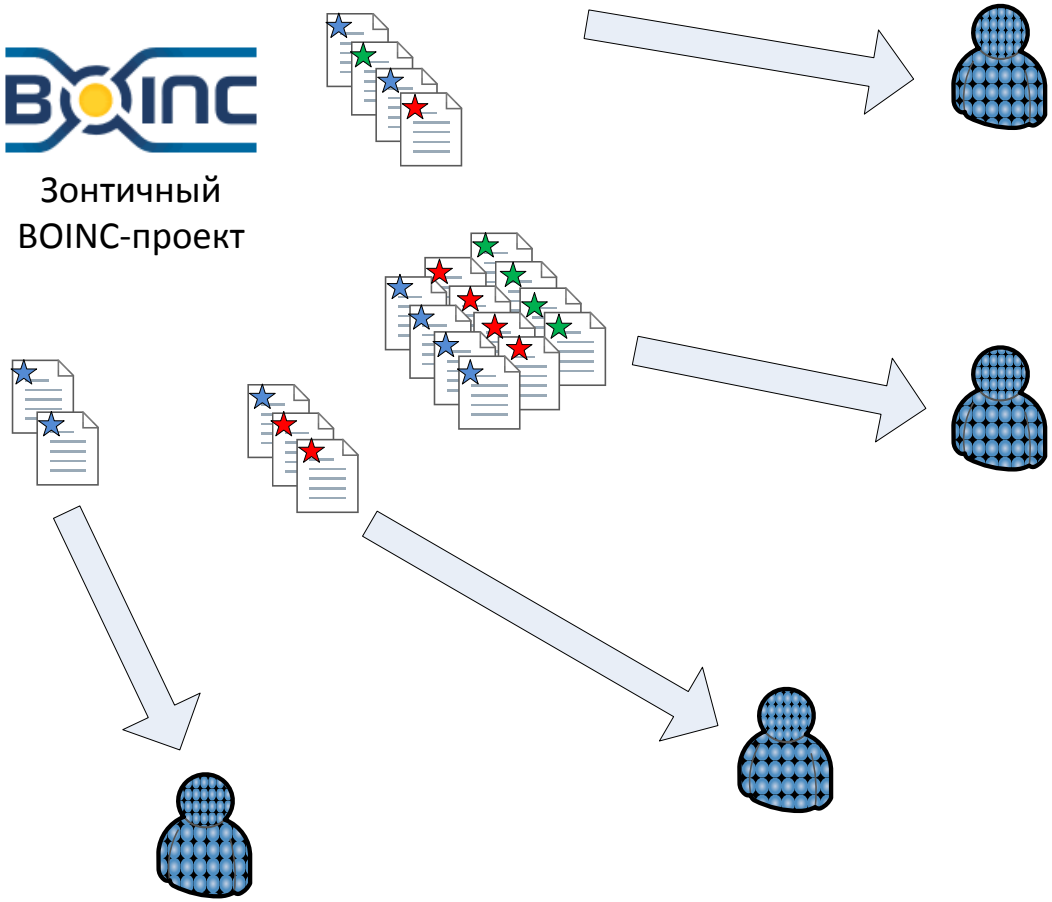


Один зонтичный проект для всех студентов



Распределение заданий в зонтичном проекте ДРВ





Этапы работы студента

- Создание автономного вычислительного приложения;
- Адаптация приложения для работы в грид-системе на платформе BOINC;
- Автоматизация генерации входных данных;
- Обработка и агрегирование результатов;
- Проверка корректности результатов и обработка ошибок;
- Проведение численного эксперимента в грид-системе.

Оценка работы студента. Критерии

- Теоретические знания по распределенным вычислениям;
- Разработка вычислительного приложения;
- Адаптация приложения для грид-систем;
- Эффективность использования ресурсов грид-системы;
- Автоматизация генерации входных данных и обработки результатов работы;
- Анализ результатов и устойчивость к ошибкам.

Заключение

- Использование зонтичного проекта добровольных распределенных вычислений на платформе BOINC позволяет студенту изучить различные аспекты функционирования грид-систем:
  - Практический опыт настройки и использования функционирующей грид-системы;
  - Особенности разработки приложений для грид-систем;
  - Опыт проведения вычислительного эксперимента на реальной распределенной системе;
  - Совместное использование одной распределенной инфраструктуры.
- Использование зонтичного проекта позволяет существенно сократить аппаратные требования для разворачивания серверной части проекта распределенных вычислений;
- Использование одного зонтичного проекта позволяет преподавателю облегчить оценку результатов работы студентов;
- Использование зонтичного проекта при дистанционном обучении.

Спасибо за внимание

Институт проблем передачи информации РАН (ИППИ РАН)

Центр распределенных вычислений

web: distributed-computing.ru

e-mail: kurochkin@iitp.ru