

## НА ПУТИ К РЕШЕНИЮ ПРИКЛАДНЫХ ЗАДАЧ ИНТЕГРАЦИИ НАУЧНОЙ РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНОЙ СРЕДЫ С МИРОВОЙ GRID-ИНФРАСТРУКТУРОЙ

*Т.Л. Гайдук* – инженер-программист  
*Д.С. Естигнеев* – стажер-исследователь  
*А.В. Лесников* – инженер-программист  
*В.Е. Матюков* – мл. научн. сотр.  
*И.В. Сафронов* – мл. научн. сотр.

---

One of the case development single node of distributed data-processing network base on grid computing have been described.

**Введение.** Объектом исследования Научной станции Российской академии наук является геологическая, геоэлектрическая и геодинамическая структуры коры Тянь-Шаня и прилегающих территорий, входящих в состав одного из самых сейсмоактивных регионов мира. С 1982 г. Научная станция активно ведет комплексное изучение геодинамических процессов на данной территории, собран обширный банк данных по магнитометрии, магнитотеллурике, GPS-наблюдениям, экспериментальным данным акустической эмиссии и др. Проводится геофизический мониторинг, данные которого отражают геодеформационные процессы. В то же время развитие и совершенствование информационных технологий позволяет справиться со стремительным ростом объемов получаемой, анализируемой и используемой информации с одновременным повышением требований к ее точности и надежности. Это непосредственно связывает эффективность научных исследований с использованием достижений прогресса в области информационных технологий. Организация различного рода электронных библиотек и баз данных позволяет не только упростить анализ результатов выполненных экспериментов, но

помочь в планировании и подготовке новых. Широкое использование систем непрерывного мониторинга в реальном времени предоставляет не только ценную информацию для научных исследований, но и вносит большой вклад в системы предупреждения стихийных бедствий, что особенно важно для регионов с высоким уровнем риска. Развитие и совершенствование методов обработки, использование современных вычислительных систем и технологий позволяет добиться большей эффективности в процессах моделирования и проверки гипотез. Стремительное развитие в последние годы сетевых технологий, в том числе и сети Интернет, приближает специализированные базы данных и вычислительные мощности к конкретному пользователю, что создает благоприятные условия для реализации многих ранее недоступных возможностей. Таким образом, использование современных достижений информационных технологий создает основу для проведения как фундаментальных, так и прикладных исследований на качественно новом уровне. В связи с этим Научная станция принимает активное участие в 21-й программе фундаментальных исследований Президиума

РАН “Разработка фундаментальных основ создания научной распределенной информационно-вычислительной среды на основе технологии GRID”.

**Система хранения геофизических данных.** Согласно 21-й программе фундаментальных исследований Президиума РАН Научная станция осуществляет разработку и организацию единой системы хранения геофизических данных, результатов исследований и организацию доступа к ним. С этой целью были определены задачи, которые должны быть решены в результате создания данной системы. Это:

- разработка и реализация базы данных геофизических измерений и научных работ на ее основе;
- разработка программного обеспечения для конвертирования существующих архивов данных в новые форматы с последующим экспортом в создаваемую базу данных;

- создание программного обеспечения для организации работы лабораторий Научной станции с разрабатываемой системой;
- разработка модулей интеграции сетей мониторинга разрабатываемой системой;
- создание сетевых приложений для организации доступа к разрабатываемой системе через сети TCP/IP;
- формирование локального GRID-узла для GRID-инфраструктуры научной распределенной информационно-вычислительной среды;
- внедрение программно-аппаратных средств для обеспечения функционирования разрабатываемой системы.

Общая структурная схема разрабатываемой системы хранения данных (рис. 1) представляет собой трехуровневую клиент-серверную систему с расширенным числом интер-

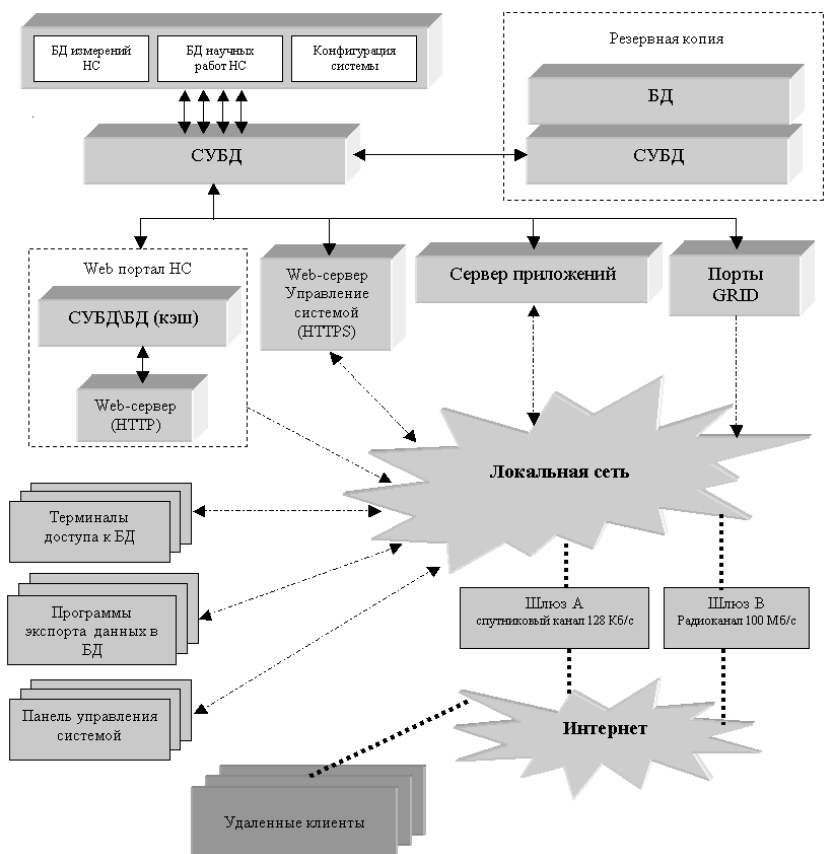


Рис. 1. Структурная схема системы хранения данных.

фейсов. Первый уровень – это система управления базами данных, которая поддерживает репликацию с резервным сервером. Второй уровень состоит из нескольких компонент; они реализуют ряд вариантов доступа к данным, в том числе сервер-приложений, web-сервер и интерфейс для GRID (GRID-узел). Третий уровень – клиентские части, они исполняются в двух вариантах: это специализированные платформу-независимые программы для доступа к системе через сети TCP/IP и cgi-приложения для стандартного интернет-проводника с доступом к порталу Научной станции через сеть Интернет. Такая структура позволяет не только организовать доступ ко всему объему информации в одной сетевой точке, но и разрабатывать и использовать программные системы с прямым доступом для осуществления работы с данными в автоматическом режиме в точках, географически удаленных от Научной станции; в том числе позволяет сформировать полноценный GRID-узел для GRID-инфраструктуры научной распределенной информационно-вычислительной среды.

Для организации максимальной совместимости с формируемой GRID-

инфраструктурой было принято решение о выделении создаваемого GRID-узла в отдельный блок разрабатываемой системы. Таким образом, формирование локальной системы сбора и обработки разнородных геофизических данных не влияет на представление информации и форму организации хранения данных в формируемом узле GRID- системы (рис. 2).

Создается специальная процедура экспорта данных из единой базы данных в GRID-узел, благодаря чему существует возможность проводить усовершенствование и расширение системы хранения и обработки данных геодинамического полигона Научной станции без оказания влияния на GRID-узел. В итоге получается следующая схема передачи данных: лаборатории Научной станции, пользуясь специальными приложениями, производят обработку и анализ данных, опираясь непосредственно на единую базу данных; web-портал импортирует данные измерений и научных результатов в режиме чтения из этой базы данных в собственный кэш (для увеличения быстродействия работы системы) и публикует их в сети Интернет. Доступ к базе данных для сотрудников Научной станции осуществляется в пределах

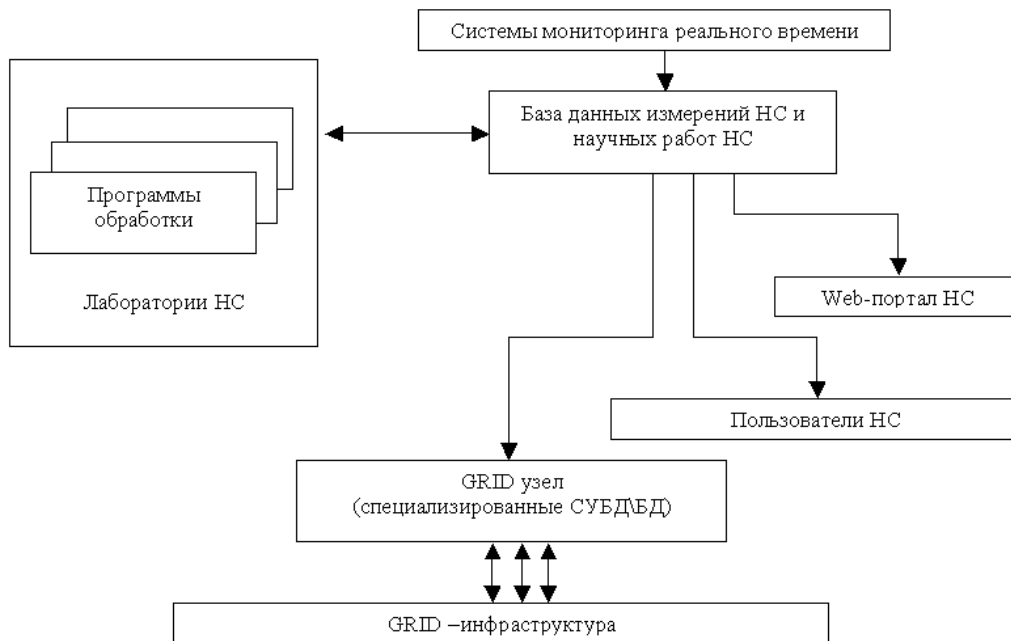


Рис. 2. Информационные потоки системы хранения данных.

локальной сети через клиентские приложения, а формируемый GRID-узел, используя процедуру экспорта, отображает в собственной системе хранения весь набор данных из единой базы данных с параллельным, при необходимости, конвертированием в форматы, определенные GRID-инфраструктурой.

Одним примером поставщика данных в разрабатываемую систему является цифровая сейсмологическая сеть KNET, работающая в режиме реального времени. В 1991 г. на территории Бишкекского геодинамического полигона была установлена сеть из десяти автоматизированных широкополосных станций STS-2 (сейсмометры типа Streckeisen). Семь станций установлены в предгорьях Киргизского хребта и северной части Чуйской долины, а три станции – в труднодоступных районах Тянь-Шаня – Джумгалский хребет, юго-западная часть Суусамырской впадины (рис. 3).

С 1998 г. сейсмологическая сеть KNET работает в режиме реального времени и в настоящее время поставляет данные в Институт сейсмологии НАН КР, на Научную

станцию РАН и в сейсмологический центр IRIS (рис. 4).

По радиоканалам данные с сейсмографов поступают на концентратор, который отправляет полученные пакеты на центральный сервер сбора данных. Далее эти пакеты записываются в буфер, организованный в виде кольцевой очереди. В данной сети рабочие станции, работающие под управлением ОС Solaris, формируют основу по автоматической передаче, обработке данных в формате CSS 3.0. Программное обеспечение, реализующее этот процесс, входит в состав пакета Antelope. Суточный объем получаемой информации из сейсмологической сети KNET составляет порядка 300 Мб. Предоставление этой информации в GRID-инфраструктуру наряду с данными других систем сбора и обработки геофизической информации, получаемой Научной станцией, является важной задачей, что в конечном итоге и определило необходимость разработки настоящей системы.

Для создания подобной системы требуется определить используемые программные средства для реализации всех

компонент системы. Для этого были выделены пять пунктов:

1. Выбор системы управления базами данных (СУБД).
2. Выбор web-сервера.

3. Выбор языка программирования (компиляторы\средства разработки).
4. Выбор операционных систем (ОС).
5. Проблемы разработки GRID-узла.

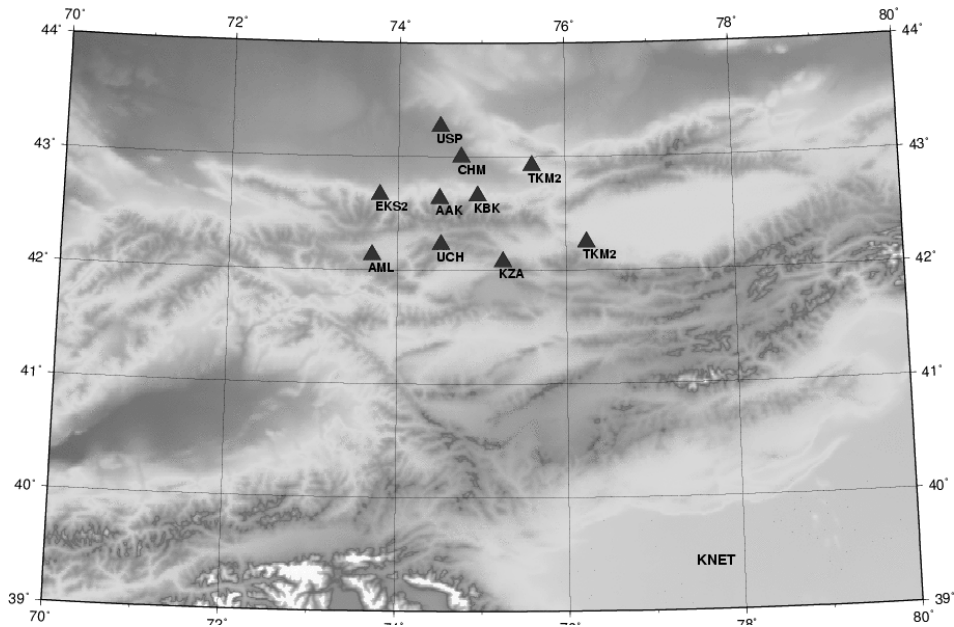


Рис. 3. Расположение станций сейсмической сети KNET.

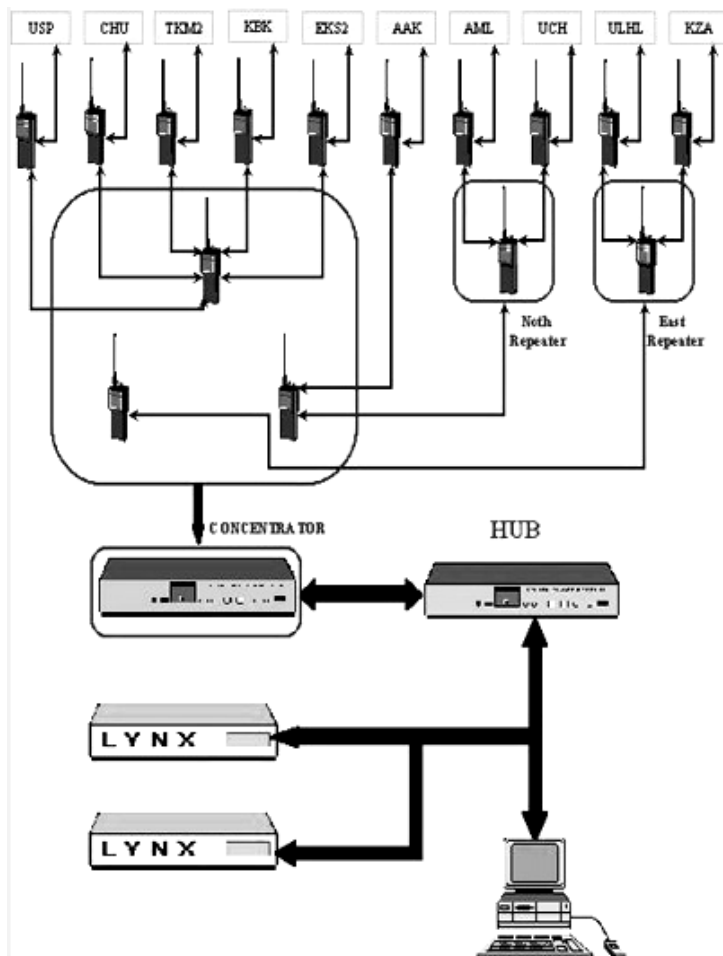


Рис. 4. Схема передачи данных в сейсмической сети KNET.

**1. Выбор СУБД.** Для обоснования выбора СУБД были рассмотрены различные реализации СУБД, а именно, наиболее широко используемые из них: DB2, Oracle, Microsoft SQL Server, PostgreSQL, MySQL [1–7]. При этом учитывалась стоимость продукта, предоставляемые им возможности и сложность использования/разработки в нем. В результате была выбрана СУБД Microsoft SQL Server 2000 [1, 2].

Microsoft SQL Server 2000 – это законченное предложение в области баз данных и анализа данных для быстрого создания масштабируемых решений электронной коммерции, бизнес-приложений и хранилищ данных. Оно позволяет значительно сократить время выхода этих решений,

одновременно обеспечивая масштабируемость, отвечающую самым высоким требованиям. В сервер SQL Server 2000 включены: поддержка языка XML [8] и протокола HTTP [9], средства повышения быстродействия и доступности, позволяющие распределить нагрузку и обеспечить бесперебойную работу; функции для улучшения управления и настройки, снижающие совокупную стоимость владения.

Microsoft SQL Server 2000 обладает следующими характеристиками:

1. Масштабируемость для хранилищ данных:
  - распределение нагрузки по серверам для обеспечения требуемой масштабируемости при

- последовательной модели программирования;
  - автоматическая синхронизация нескольких систем;
  - эффективное использование систем с симметричной многопроцессорной обработкой.
2. Бесперебойная работа и надежность:
- выполнение критически важных операций без перехода в автономный режим;
  - ускорение архивирования за счет копирования только измененных данных.
3. Упрощенное управление и настройка:
- централизованное управление базами данных SQL Server;
  - экономия времени и усилий благодаря динамическим средствам автоматического управления и настройки;
  - возможность простого перемещения или копирования базы данных с компьютера на компьютер или между экземплярами сервера без перехода в автономный режим.

**2. Выбор web-сервера.** Использование web-интерфейса для доступа к данным и контроля работы разрабатываемой системы предполагает использование web-сервера и предъявляет к нему серьезные требования. Для выбора используемого в реализации web-сервера были рассмотрены следующие варианты: Apache, iPlanet, Internet Information Server, Zeus, Netscape Server и Novell Web Server. В результате был выбран web-сервер Apache [9].

Apache – это стабильный и высокоскоростной кросс платформенный web-сервер. Он выполнен в модульной структуре и поддерживает протокол HTTP/1.1., является одним из самых распространенных web-серверов, хорошо документирован, обладает гибкими возможностями конфигурации.

Краткие характеристики web-сервера Apache:

- Поддержка операционных систем: NetBSD, Digital UNIX, BSDI, AIX, OS/2, SCO, HPUX, Novell NetWare, Macintosh, Be OS,

Windows NT, Linux [11], Windows 95, FreeBSD, Windows 98, IRIX, Solaris [13].

- Поддерживает Windows CGI, HTTP/1.1, в том числе HTTP/1.1 PUT; есть функция автоматического ответа при изменении документа; поддерживает Microsoft ISAPI, поддерживает настройку на несколько портов.
- Возможен запрет доступа с конкретных адресов, к конкретным документам, запрет запуска CGI скриптов, доступ конкретных пользователей; допускаются изменения без перезапуска сервера.
- Поддерживает безопасные подключения по протоколу HTTPS.
- Бесплатный, открытый код.

### 3. Выбор языка программирования.

Разработка серверного и клиентского программного обеспечения и требования к нему, как правило, определяют выбор языка программирования. В настоящем же случае главным требованием является разработка платформо-независимых приложений. В связи с этим, учитывая, что протоколы соединения приложений всех уровней являются стандартными, предполагается использование двух языков программирования: Java [14, 15] и C# [16, 17, 18]. Выбор в пользу того или иного варианта отдан разработчику программного продукта. Это позволяет, опираясь на единые стандарты, максимально эффективно использовать уже имеющиеся наработки и опыт разработчиков без ущерба для разрабатываемой системы.

Java – это объектно-ориентированный язык программирования, разработанный компанией Sun Microsystems [14, 15], функционирующий в среде Java. Java – платформо-независимая, мультипроцессорная программная среда, созданная для разработки программ и приложений для Интернета и локальных сетей.

Краткие характеристики Java:

- Объектная ориентированность. Концепция объектно-ориентированного программирования.
- Средства работы в сети, использующие протокол TCP/IP. Имеется специальная библиотека классов для создания сетевых приложений.

- Безопасность. Полученный по сети байт-код проверяется на правильность инструкций (принадлежность к байт-коду Java), на безопасность команд для компьютера и локальной сети, на соответствие разрешенным операциям.
- Переносимость и нейтральность относительно архитектуры. Байт-код языка Java един для всех платформ и архитектур. Благодаря наличию виртуальных машин Java-приложения могут разрабатываться и исполняться на самых разных вычислительных платформах.
- Многопоточность. В языке Java предусмотрены механизмы создания легковесных параллельных процессов (threads), которые могут работать в режиме квантования времени, используя синхронизацию на основе мониторов Хоара.
- Автоматическая сборка мусора.
- Высокая производительность. Оптимизация производится как на этапе генерации байт-кода, так и перед непосредственным исполнением на конкретной машине.

C# – это объектно-ориентированный язык программирования, предназначенный для разработки разнообразных корпоративных приложений, выполняемых в среде .NET Framework [16–18]. NET Framework представляет модель программирования для платформ .NET [19]. Основными ее компонентами являются общезыковая среда выполнения (CLR) и библиотека классов .NET Framework, включающая ADO.NET, ASP.NET и формы Windows Forms.

Краткие характеристики языка C#:

- Разработанная на нем программа является независимой от платформы.
  - Автоматическая сборка.
  - Все переменные автоматически инициализируются средой и обладают типовой защищенностью.
  - Поддержка интерфейсов и наследования интерфейсов.
  - Позволяет использовать типизированные, расширяемые метаданные, которые могут быть прикреплены к объекту.
- Является мощным инструментом по созданию компонентов пользовательского интерфейса.

#### 4. Выбор операционной системы.

Выбранная СУБД определяет выбор ОС для сервера. Учитывая требования СУБД Microsoft SQL Server [10], была выбрана для сервера базы данных ОС Microsoft Windows 2000 Server. В то же время в качестве платформы для размещения web-сервера и брандмауэра (шлюз) выбор был сделан в пользу ОС FreeBSD [12]. Данная операционная система свободно распространяема и не требовательна к ресурсам компьютера. Это UNIX-подобная операционная система для платформ i386, IA-64, PC-98, Alpha/AXP и UltraSPARC.

FreeBSD обладает следующими характеристиками [12]:

- Приоритетная многозадачность с динамическим регулированием приоритетов грамотно распределяет ресурсы компьютера между приложениями и пользователями, даже при больших нагрузках.
- Многопользовательская доступность, которая позволяет использовать FreeBSD совместно для ряда задач.
- Мощный TCP/IP стек с поддержкой промышленных стандартов, таких, как SLIP, PPP, NFS, DHCP и NIS. Это говорит о том, что FreeBSD может легко функционировать с другими системами предприятия, предоставляя жизненно важные функции, как NFS (удаленный доступ к файлам) и услуги электронной почты, или представить организацию в Internet с сервисами WWW, FTP, маршрутизации и брандмауэра (сетевая безопасность).
- Защита памяти гарантирует, что приложения не смогут чинить препятствия друг другу. Фатальная ошибка в выполнении одного приложения не скажется на работоспособности всей системы.
- Промышленный стандарт X Window System (X11R6) предоставляет графический интерфейс пользователя (GUI) для большинства VGA карт и мониторов и



поставляется с полными исходными текстами.

- Поддержка симметричной многопроцессорности (SMP) для машин с несколькими процессорами.

**5. Проблемы разработки GRID-узла.** На данный момент только формируются правила организации GRID-инфраструктуры и форматы некоторых геофизических данных. В связи с этим и была выполнена данная структуризация системы хранения данных, позволяющая вынести GRID-узел в отдельный блок. Конечно, это вносит некоторые неудобства, связанные с высокими требованиями к аппаратному обеспечению при предоставлении ресурсов сетей мониторинга реального времени в GRID-инфраструктуре. В то же время очевидны и преимущества подобной организации, в том числе и возможности автономной разработки GRID-узла, что позволяет реализовывать остальные элементы разрабатываемой системы хранения данных до определения программного обеспечения и протоколов GRID-инфраструктуры.

В результате определения необходимого программного обеспечения и выбора языков разработки для системы хранения геофизических данных была получена следующая схема взаимодействия программного обеспечения (рис. 5). Основной и резервный сервера баз данных разворачиваются на ОС Microsoft Windows 2000 Server с СУБД Microsoft SQL Server. Web-портал и брандмауэры реализуются на ОС FreeBSD 4.x\5.x; при этом web-сервер строится на базе Apache, а правила файэрвола разрешаются утилитой IPFW. Для обеспечения взаимодействия клиентских приложений и системы хранения данных, они реализуются на платформено-независимых решениях Java и .Net(C#). Подобная организация системы позволяет обеспечить взаимодействие с клиентами независимо от используемых ими платформ и сохранить высокий уровень производительности за счет разделения общей нагрузки на отдельные сервера и использования распределенных приложений.

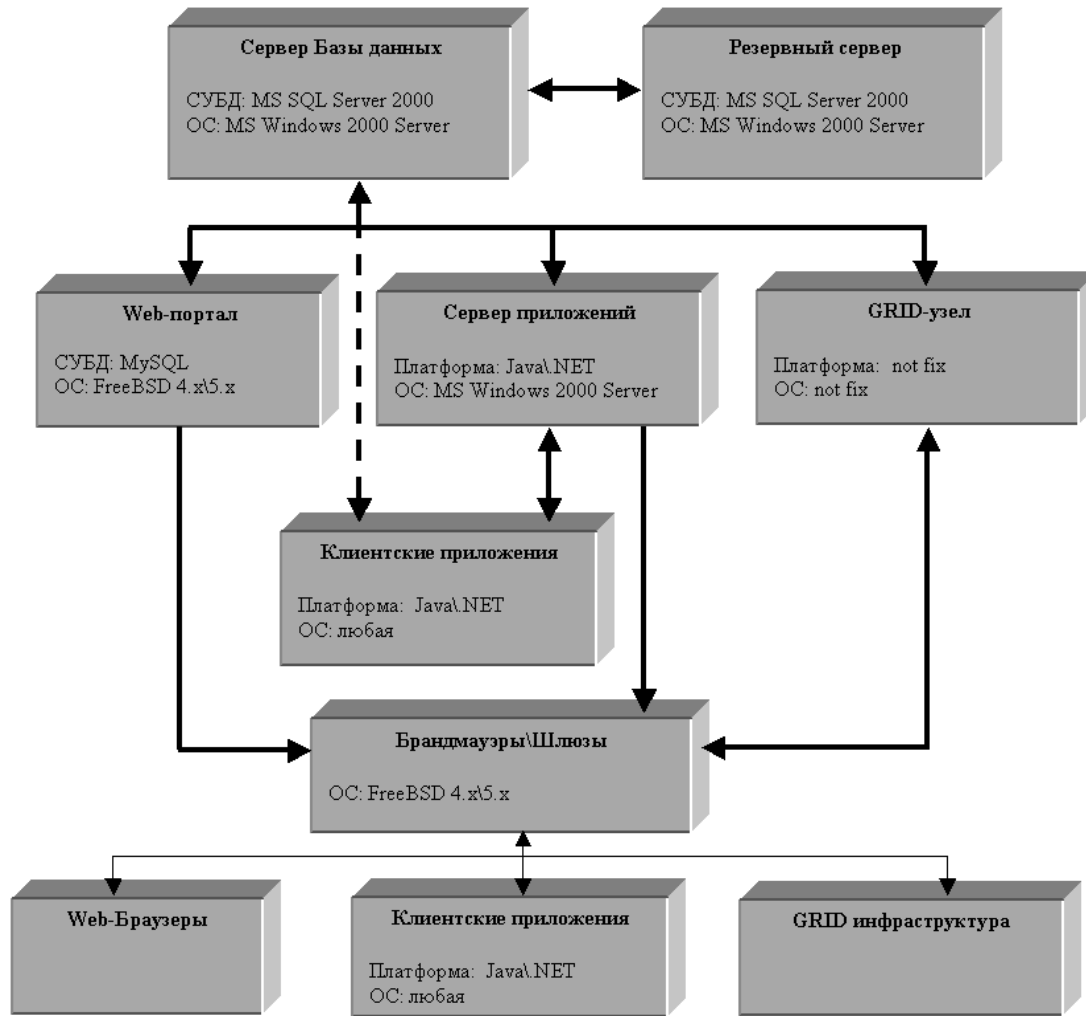


Рис. 5. Схема взаимодействия программного обеспечения.

**Заключение.** Использование нескольких высокоскоростных каналов передачи данных, систем защиты сетей и резервного копирования базы данных совместно с размещением программных систем на мощных серверах с RAID-массивами делает данную систему хранения геофизических данных надежным поставщиком данных для GRID-инфраструктуры научной распределенной информационно-вычислительной среды и предоставляет новые возможности доступа к данным как сотрудников и партнеров Научной станции, так и научной общественности.

Настоящая работа выполнялась в совместном учебно-научном центре ИС РАН и

КРСУ как часть проекта "Оптимизация научной интрасети и развитие сетевых приложений для анализа пространственно-временных геоданных Тянь-Шаня и прилегающих территорий (как возможного GRID-полигона)" по Программе № 21 фундаментальных исследований Президиума РАН "Разработка фундаментальных основ создания научной распределенной информационно-вычислительной среды на основе технологий GRID".

#### Литература

1. *Хоторн Роб.* Разработка Microsoft SQL Server 2000 на примерах / Пер. с англ. – М.: Издат. дом “Вильямс”, 2001. – 464 с.
2. *Полякова Л.Н.* Основы SQL: Курс лекций. – М.: ИНТУИТ.РУ, Интернет-Университет информ. технол., 2004. – 368 с.
3. *Ланг, Курт, Чоу Д.* Публикация баз данных в Интернете. – СПб.: Символ-Плюс, 1998. – 475 с.
4. *Гаршин И.К.* Практика работы с Oracle: генерация, администрирование, репликация. – М.: ПОЛТЕКС, 2000. – 250 с.
5. “DB-2” – <http://emanual.ru/download/701.html>
6. “Документация по MySQL” – <http://www.mysql.ru/docs/>
7. *Стоунз Р., Мэттью Н.* PostgreSQL Основы. – СПб.: Символ-Плюс, 2002. – 640 с.
8. *Холзнер С.* XML. Энциклопедия. – СПб.: Питер, 2004. – 1100 с.
9. *Скотт Хокинс.* Администрирование Web-сервера Apache и руководство по электронной коммерции – М.: Издат. дом “Вильямс”, 2001. – 336 с.
10. *Таллоч М.* Windows Server 2003. Справочник. – СПб.: Питер, 2005. – 748 с.
11. *Колисниченко Д.Н.* Linux-сервер своими руками. – 3-е изд. – СПб.: Наука и техника, 2005. – 744 с.
12. *Федорчук А.В., Торн А.В.* FreeBSD: установка, настройка, использование. – СПб.: БХВ-Петербург, 2004. – 630 с.
13. PC-NFS Solaris User’s Guide. – U.S.A.: SunSelect, 1993. – 194 p.
14. *Дейтел Х.М., Дейтел П.Дж.* Как программировать на Java: Кн. 1. Основы программирования / Пер. с англ. – М.: Бином-Пресс, 2003. – 848 с.
15. *Дейтел Х.М., Дейтел П.Дж.* Как программировать на Java: Кн. 2. Файлы, сети, базы данных / Пер. с англ. – М.: Бином-Пресс, 2005. – 672 с.
16. *Джейсон Прайс, Майк Гандэрлой.* Visual C# .NET. Полное руководство / Пер. с англ. – Киев: ВЕК+, 2004. – 960 с.
17. *Чарльз Петцольд.* Программирование для Microsoft Windows на C# / Пер. с англ.: В 2-х т. – М.: Издат.-торг. дом “Русская Редакция”, 2002. – Т. 1. – 576 с.
18. *Чарльз Петцольд.* Программирование для Microsoft Windows на C# / Пер. с англ.: В 2-х т. – М.: Издат.-торг. дом “Русская Редакция”, 2002. – Т. 2. – 624 с.
19. *Дон Бокс, Крис Селлз.* Основы платформы NET. Общеязыковая исполняющая среда. – М.: Издат. дом “Вильямс”, 2003. – Т. 1. – 288 с.
20. Сети TCP/IP. Ресурсы Microsoft Windows 2000 Server / Пер. с англ. – М.: Издат.-торг. дом “Русская Редакция”, 2001. – 784 с.