

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ  
«САМАРСКИЙ ГОСУДАРСТВЕННЫЙ АЭРОКОСМИЧЕСКИЙ  
УНИВЕРСИТЕТ имени академика С.П. КОРОЛЁВА  
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»

# **Методическое пособие по работе пользователей с системой хранения данных грид-среды СГАУ**

**Самара 2010**

УДК 004.75

Составители:

Баскаков Александр Викторович, Бриль Валентина Евгеньевна,  
Кузьмичёв Венедикт Степанович, Симановский Евгений Аркадьевич

Методическое пособие позволяет изучить системы хранения данных, технологии организации хранения данных и их подключение к суперкомпьютеру «Сергей Королёв». Пособие содержит инструкцию пользователю по работе с системами хранения данных грид-среды СГАУ.

УДК 004.75

## Оглавление

1.	Понятие ГРИД-системы .....	4
2.	Технологии подключения устройств хранения данных.....	7
3.	Интерфейс подключения устройств хранения данных .....	9
4.	Системы хранения данных .....	10
4.1.	Несетевые системы хранения данных.....	10
4.2.	Сетевые системы хранения данных .....	14
4.2.1.	Общие сведения .....	14
4.2.2.	Технология Network Attached Storage(NAS).....	15
4.2.3.	Технология Storage Area Network (SAN).....	19
4.2.3.1.	Топология SAN-сети .....	20
4.2.3.2.	Преимущества SAN сетей.....	21
4.2.3.3.	Оборудование, используемое для создания SAN сетей .....	22
4.2.3.4.	Совместное использование СХД.....	27
4.3.	Сравнение технологий обмена данными .....	28
5.	Файловые системы.....	28
6.	Параллельная файловая система GPFS .....	30
7.	Подключение СХД к суперкомпьютеру «Сергей Королёв».....	34
8.	Протоколы передачи данных .....	39
9.	Инструкция пользователю по работе с системами хранения данных .....	43

# 1. Понятие ГРИД-системы

В настоящее время вычислительные ресурсы персональных компьютеров, подключенных к Интернет, используются примерно на 2%. Степень использования процессорной мощности персональных компьютеров очень низка. В связи с этим в компьютерном мире наметились новые тенденции развития: переход от экстенсивной модели развития к интенсивной.

Концепция ГРИД (от англ. grid — «сетка, решетка») подразумевает создание глобальной компьютерной инфраструктуры, обеспечивающей интеграцию географически распределенных информационных и вычислительных ресурсов. Родоначальниками концепции ГРИД являются американские ученые Ян Фостер и Карл Кессельман.

Грид — это согласованная, открытая и стандартизованная компьютерная среда, которая обеспечивает гибкое, безопасное, скоординированное разделение вычислительных ресурсов и ресурсов хранения информации, которые являются частью этой среды.

Имя инфраструктуры сродни термину "электросеть" которая в начале XX века обеспечила повсеместный доступ к источникам электроэнергии. GRID охватывает самые разнообразные ресурсы, от вычислительных циклов, данных, датчиков до людей.

Вычислительная среда ГРИД призвана объединить компьютерные ресурсы всех для блага каждого. Эта технология открывает небывалые возможности для любого рода исследований, требующих обработки большого объема данных и продолжительных вычислений.

Грид – это инфраструктура, объединяющая множество ресурсов разных типов (процессоры, долговременная и оперативная память, хранилища и базы данных, сети), доступ к которым пользователь может получить из любой точки, независимо от места их расположения.

Идея грид-технологии возникла вместе с распространением персональных компьютеров, развитием Интернета, технологий пакетной передачи данных на основе оптического волокна (SONET, SDH и ATM) а также технологий локальных сетей(Gigabit Ethernet). Полоса пропускания коммуникационных средств стала достаточной, чтобы при необходимости привлечь ресурсы другого компьютера. Учитывая, что множество подключенных к глобальной сети компьютеров большую часть рабочего времени простаивает и располагает ресурсами, большими, чем необходимо для решения их повседневных задач,

возникает возможность применить их неиспользуемые ресурсы в другом месте.

В большинстве прикладных областей решаются относительно простые вычислительные задачи, не требующих уникальных вычислительных ресурсов. Эти задачи можно выполнять, используя вычислительные ресурсы и программное обеспечение массового производства.

Однако существуют задачи, которые невозможно решить на компьютерах массового производства с использованием общедоступного программного обеспечения. Это задачи аэро- и гидродинамики (расчет крыла самолета или корпуса быстроходного судна), моделирование сложных динамических систем (ядерного взрыва или образования нейтронной звезды), задачи предсказания погоды и создание модели климатических изменений. Здесь требуется специальная организация работы многих сотен и тысяч процессоров для решения одной задачи. С этой целью создаются многопроцессорные компьютеры специальной архитектуры и особое программное обеспечение. К таким вычислительным ресурсам обычно применяется термин «*суперкомпьютер*».

В рамках какого-либо фундаментального исследования, прикладной проблемы или даже коммерческого проекта приходится решать множество задач, каждая из которых в отдельности не является сложной. Конечно, поток или набор относительно простых и однотипных задач легко распараллеливается, и использование суперкомпьютеров несомненно даст огромный эффект по производительности. Однако такой же эффект можно получить и используя простые наборы персональных компьютеров, объединенные локальной сетью, — *кластеры*, в которых один из компьютеров занимается распределением задач по принципу «одна задача — один процессор». В физике высоких энергий такие кластеры получили название *компьютерных ферм*. Практика последнего десятилетия показала, что использование ферм в несколько раз дешевле, чем применение суперкомпьютеров, и дает такой же эффект производительности для большинства прикладных задач. И в этом плане ГРИД, как глобальное объединение процессоров, скорее, является фермой, нежели суперкомпьютером.

В мире уже сейчас работают сотни миллионов персональных компьютеров, как на рабочих местах, так и в составе кластеров (ферм). ГРИД-технология позволит объединять эти мощности в глобальные географически распределенные фермы. В результате такого объединения пользователь получает возможность запуска своих задач на глобальной ферме, которая будет на много порядков мощнее, нежели доступные ему локальные ресурсы. При этом собственные компьютеры

будут, в свою очередь, включены в состав этой глобальной фермы и на них будут выполняться задачи других пользователей. Вычисление, где задания, запущенные на распределенном вычислительном устройстве, выполняются двумя и более административными доменами, называют термином Grid computing (грид-компьютинг).

Кроме увеличения производительности появляется также возможность использования географически распределенных данных для решения поставленной задачи без их транспортировки в одно место. Программа может обращаться к различным сайтам в соответствии с тем, где необходимые данные хранятся, а окончательный результат будет доставлен пользователю после завершения всех вычислений.

ГРИД — это набор стандартизированных сервисов, выполняющих свои функции в фоновом режиме в соответствии с универсальными оптимизирующими алгоритмами через использование специальных протоколов и стандартных программных средств промежуточного уровня (middleware). В этом он напоминает Всемирную паутину, где запрос на информацию к удаленному сайту происходит по протоколу (HTTP) и с использованием программ — Web-браузеров, сама же информация должна быть записана с использованием определенного стандарта (HTML). В случае ГРИД набор таких протоколов, стандартов и служб оказывается значительно шире.

Реализация идеи «вычислений через Интернет» не может быть простым развитием Web-технологии. Здесь требуются много принципиально новых решений. Тем не менее, вполне закономерно считать, что «ГРИД — это следующий революционный этап развития WWW в XXI веке», более того уже появился новый термин «World Wide Grid — WWG».

Сейчас в ГРИД-инфраструктуры включаются ресурсы только компьютерных центров — научных или производственных. Хотя попытки задействовать такие ресурсы, как компьютеры интернет-кафе, все же предпринимаются.

Можно назвать два основных направления развития ГРИД-технологий — вычислительное (computational) и интенсивно работающее с данными (data intensive GRID).

В вычислительном направлении создаваемая инфраструктура нацелена на достижение максимальной скорости расчётов за счет глобального распределения вычислений. В таких случаях выгоднее доставлять требуемые данные к мощному компьютеру для выполнения задачи. Одним из таких проектов является европейский проект DEISA (deisa.org), в котором предпринимается попытка объединить суперкомпьютерные центры.

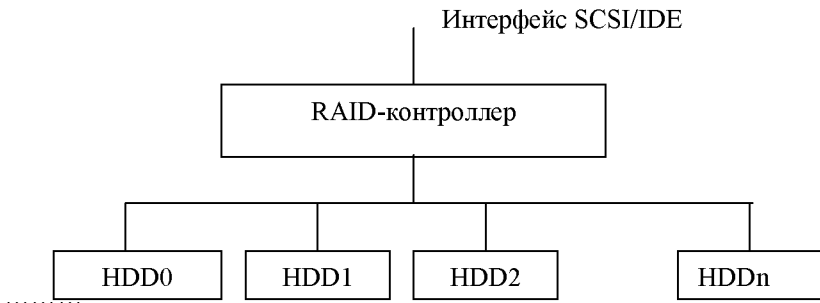
В случае же второго направления транспортировка данных представляет собой гораздо более сложную задачу, чем сами

вычисления, — такие задачи подпадают под понятие ГРИД для интенсивных операций с данными. Здесь задаче выгоднее пройти по серверам, где хранятся обрабатываемые данные.

## 2. Технологии подключения устройств хранения данных

Наиболее распространённые устройства хранения данных в настоящее время – это жёсткие магнитные диски или магнитооптические диски. Если раньше жёсткие диски находились внутри компьютера (сервера), то дальнейшее развитие привело к созданию так называемых RAID-массивов. В переводе с английского «RAID» (Redundant Arrays of Inexpensive Disks) означает «избыточный массив независимых дисков». Этот перевод не совсем дословный, но именно содержащийся в нем смысл является правильным. Дисковый массив – это набор дисковых устройств, работающих вместе. Этим набором устройств управляет специальный Raid-контроллер, который выполняет функции размещения данных по массиву, а для всей остальной системы представляет весь массив как одно логическое устройство ввода/вывода. За счёт параллельного выполнения операций чтения и записи на нескольких дисках массив обеспечивает повышенную скорость обменов по сравнению с одним большим диском (Рис 1).

Рисунок 1



Первоначальное предназначение RAID – создание на базе нескольких винчестеров диска большого объема с увеличенной скоростью доступа. Но затем к двум основным целям добавилась третья – сохранение данных в случае отказа части оборудования.

В основе теории RAID лежат пять основных принципов: массив (Array), зеркалирование (Mirroring), дуплекс (Duplexing), чередование (Striping) и чётность (Parity).

**Массивом** называют несколько накопителей, которые централизованно настраиваются, форматируются и управляются. Логический массив – это уже более высокий уровень представления, на котором не учитываются физические характеристики системы. Соответственно, логические диски могут по количеству и объёму не совпадать с физическими. Но лучше всё-таки соблюдать соответствие: физический диск – логический диск. Наконец, для операционной системы вообще весь массив является одним большим диском.

**Зеркалирование** – технология, позволяющая повысить надёжность системы. В RAID-массиве с зеркалированием все данные одновременно пишутся не на один, а на два жестких диска. То есть создается «зеркало» данных. При выходе из строя одного из дисков вся информация остаётся сохраненной на втором.

**Дуплекс** – развитие идеи зеркалирования. В этом случае так же высок уровень надёжности и требуется в два раза больше жестких дисков. Но появляются дополнительные затраты: для повышения надёжности в систему устанавливаются два независимых RAID контроллера. Выход из строя одного диска или контроллера не сказывается на работоспособности системы. Столь дорогое решение используется только во внешних RAID-массивах, предназначенных для ответственных приложений.

**Чередование** – отличная возможность повысить быстродействие системы. Очевидно, если чтение и запись вести параллельно на нескольких жестких дисках, можно получить выигрыш в скорости. Как это делается? Записываемый файл разбивается на части определенного размера и посылается одновременно на все имеющиеся накопители. В таком фрагментированном виде файл и хранится. Считывается он тоже «по кусочкам».

Размер «кусочка» может быть минимальным – 1 байт, но чаще используют более крупное дробление – по 512 байт (размер сектора).

**Чётность** является альтернативным решением, соединяющим в себе достоинства зеркалирования (высокая надёжность) и чередования (высокая скорость работы). Используется тот же принцип, что и в контроле чётности оперативной памяти.

Каждый из уровней RAID использует уникальный метод записи данных на диски, и поэтому все уровни обеспечивают различные преимущества. Уровень RAID 0 использует технологию *распределения данных* (информация разбивается на куски, которые записываются на диски и считываются параллельно). Эта технология повышает



производительность, но не обеспечивает избыточности данных. RAID 1, 2, 3, 4 и 5 обеспечивают избыточное хранение данных. Уровень RAID 1 использует *зеркалирование*, т. е. хранение резервной копии. Уровни RAID 2, 3, 4 и 5 обеспечивают *хранение битов чётности*. Обе эти технологии позволяют восстанавливать информацию в случае сбоя одного из дисков.

С точки зрения пользователя или ПО, скорость определяется не только пропускной способностью системы (Мбайт/с), но и числом транзакций – то есть числом операций ввода-вывода в единицу времени (IOPS). Увеличению IOPS способствует, что вполне логично, большее число дисков и те методики повышения производительности, которые предоставляет контроллер RAID (к примеру, кэширование).

Если для просмотра потокового видео или организации файл-сервера больше важна общая пропускная способность, то для СУБД-приложений критично именно число транзакций, которые способна обрабатывать система.

### 3. Интерфейс подключения устройств хранения данных

Интерфейс (англ. interface) — совокупность линий связи, сигналов, посылаемых по этим линиям, технических средств, поддерживающих эти линии, и правил (протокола) обмена. Серийно выпускаемые внутренние жёсткие диски могут использовать интерфейсы ATA (он же IDE и PATA), SATA, eSATA, SCSI, SAS, FireWire, SDIO и Fibre Channel.

Можно выделить два основных интерфейса подключения устройств хранения данных (серверов):

- **SCSI** (Small Computer Systems Interface)– это параллельный интерфейс, допускающий до 16 устройств на одной шине. Недостатки SCSI – толстые неудобные кабели, имеющие ограничение на максимальную длину (до 25 м), не обладающие помехозащищённостью.
- **FibreChannel** (FC) — семейство протоколов для высокоскоростной передачи данных. Изначальное применение FC – в области суперкомпьютеров, но впоследствии практически полностью перешло в сферу сетей хранения данных, где FC используется как стандартный способ подключения к системам хранения данных уровня предприятия. . FibreChannel обеспечивает значительно более высокую скорость передачи данных. Расстояния практически не ограничены – от стандартных 300 метров на самом «обычном» оборудовании до 2000 км для мощных коммутаторов, так

называемых «директоров». Главный плюс интерфейса FibreChannel – возможность объединения многих устройств хранения и хостов (серверов) в единую сеть хранения данных. Другие преимущества – большие расстояния (чем со SCSI), возможность агрегирования каналов, возможность резервирования путей доступа, «горячего подключения» оборудования, большая помехозащищенность. Физически используются двухжильные многомодовые или одномодовые оптические кабели (с коннекторами типа LC или SC) и так называемые SFP – оптические трансмиттеры на основе светодиодных или лазерных излучателей – от этих двух компонент зависит скорость передачи и максимальное расстояние между устройствами.

Большинство сетей хранения данных использует протокол SCSI для связи между серверами и устройствами хранения данных. Так как протокол SCSI не предназначен для формирования сетевых пакетов, в сетях хранения данных используются низкоуровневые, например протоколы:

- **Fibre Channel Protocol (FCP)**, транспорт SCSI через Fibre Channel. Наиболее часто используемый на данный момент протокол. Существует в вариантах 1 Gbit/s, 2 Gbit/s, 4 Gbit/s, 8 Gbit/s и 10 Gbit/s. Является основой построения сетей хранения данных;
- **iSCSI**, транспорт SCSI через TCP/IP.

## 4. Системы хранения данных

### 4.1. Несетевые системы хранения данных

В случае отдельного ПК под **системой хранения данных (СХД)** можно понимать отдельный внутренний жесткий диск или систему дисков.

Традиционно можно выделить три технологии организации хранения данных: Direct Attached Storage (DAS), Network Attach Storage (NAS) и Storage Area Network (SAN). Они различаются порядком организации доступа к системам хранения:

- DAS – система хранения, непосредственно подключаемая к серверу;
- NAS – система хранения, подсоединяемая к сети;
- SAN – сеть хранения данных. Основой SAN является выделенная специализированная сеть, которая служит исключительно для организации доступа к данным.

Основное отличие технологий связано с сетевыми интерфейсами. В случае с DAS все три компонента (приложения, файловая система и диски) размещены в одной системе.

Трафик данных в NAS-решениях идет по сети Ethernet, что с точки зрения безопасности является не безупречным вариантом. А поскольку по локальной сети идет и другой трафик, кроме NAS, то скорость передачи данных ниже, чем у DAS, и при обращении приложений к данным возможны заметные потери производительности. Дополнительные нагрузки, связанные с протоколом TCP/IP, также негативно влияют на быстродействие. Однако, будущие реализации iSCSI (SCSI over Ethernet) обещают улучшенное быстродействие.

В сети SAN физические диски сосредоточены в единой сети. Они могут консолидироваться в огромные фермы, которые могут предоставлять необходимое дисковое пространство по требованию.

Каждая из трех технологий хранения имеет свои плюсы и минусы, однако часто невозможно найти компромисс между высокой готовностью и низкими ценами, поэтому при построении систем с заделом на будущее необходимо тщательно проработать вопросы масштабируемости и производительности. В таблице 1 отражены особенности технологий хранения.

Технологии	DAS	NAS	SAN
Сетевая технология	Нет	Да	да
Максимальное количество поддерживаемых устройств /НВА	15	-	126
Защита от внешних помех	Нет (медь)	Нет (медь) / да (оптика)	Да (оптика)
Уровень цен	Низкий	Высокий	Очень высокий
Масштабируемость	Плохая	Хорошая	Очень хорошая
Максимальное расстояние до сервера	25м	-	10км
Основной используемый интерфейс	SCSI	Ethernet	FCP

Приведенные характеристики технологий хранения отражают типичные показатели существующих сетей хранения данных.

Разница между технологиями NAS и SAN сетевого хранения постепенно стирается благодаря новым блочным методам на основе IP-протоколов (iSCSI, FCIP, iFCP), которые позволяют использовать обе

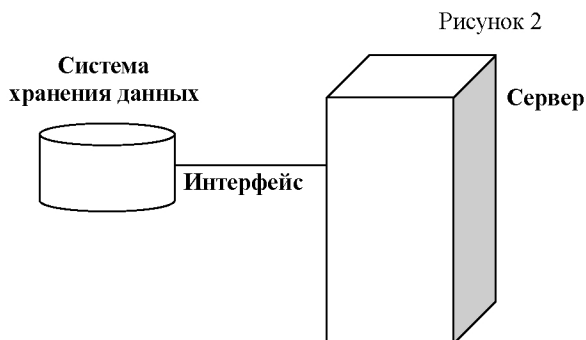
технологии вместе. Кроме того, с новыми технологиями виртуализации, такими как MPFS (Multipath File System Celerra HighRoad), или технологиями подключения, подобными FAS (Fibre Attached Storage), системы SAN и NAS практически всегда используются вместе, поэтому сегодня между ними уже трудно провести четкую границу.

Технология DAS подразумевает прямое (непосредственное) подключение накопителей к серверу или к ПК. При этом накопители (жесткие диски, ленточные накопители) могут быть как внутренними, так и внешними. Простейший случай DAS-системы — это один диск внутри сервера или ПК. Кроме того, к DAS-системе можно отнести и организацию внутреннего *RAID-массива* с использованием RAID-контроллера.

В небольших локальных сетях доминирующее положение занимает технология DAS, в соответствии с которой хранилище находится внутри сервера, обеспечивающего объем хранения и необходимую вычислительную мощность.

Запросы ввода-вывода (называемые также командами или протоколами передачи данных) непосредственно обращаются к этим устройствам. Однако такие системы плохо масштабируются, что приводит к необходимости с целью расширения объема хранилища приобретать дополнительные серверы (под масштабируемостью понимается возможность наращивания дополнительных ресурсов без структурных изменений центрального узла системы). Эта архитектура очень дорогая и может использоваться только для создания небольших по объему хранилищ данных.

Стоит отметить, что, несмотря на формальную возможность использования термина DAS-системы по отношению к одиночному диску или к внутреннему массиву дисков, под DAS-системой принято понимать внешнюю стойку или корзину с дисками, которую можно рассматривать как автономную систему хранения данных (рис 2).



Пример DAS- системы хранения данных

Кроме независимого питания, такие автономные DAS-системы имеют специализированный контроллер (процессор) для управления массивом накопителей. К примеру, в качестве такого контроллера может выступать RAID-контроллер с возможностью организации RAID-массивов различных уровней.

В качестве интерфейсов для подключения накопителей (внутренних или внешних) в технологии DAS могут выступать интерфейсы SCSI, SATA, PATA и *Fibre Channel*. Если интерфейсы SCSI, SATA и PATA используются преимущественно для подключения внутренних накопителей, то интерфейс *Fibre Channel* применяется исключительно для подключения внешних накопителей и автономных СХД. Преимущество интерфейса *Fibre Channel* заключается в данном случае в том, что он не имеет жёсткого ограничения по длине и может использоваться в том случае, когда сервер или ПК, подключаемый к DAS-системе, находится на значительном расстоянии от неё. Интерфейсы SCSI и SATA также могут использоваться для подключения внешних СХД (в этом случае интерфейс SATA называют eSATA), однако данные интерфейсы имеют строгое ограничение по максимальной длине кабеля, соединяющего DAS-систему и подключаемый сервер.

К основным преимуществам DAS-систем можно отнести их низкую стоимость (в сравнении с другими решениями систем хранения данных), простоту развертывания и администрирования, а также высокую скорость обмена данными между системой хранения и сервером. Собственно, именно благодаря этому они завоевали большую популярность в сегменте малых офисов и небольших корпоративных сетей. В то же время DAS-системы имеют и свои недостатки, к которым можно отнести слабую управляемость и неоптимальную утилизацию ресурсов, поскольку каждая DAS-система требует подключения выделенного сервера.

В настоящее время DAS-системы занимают лидирующее положение, однако на смену DAS-системам постепенно приходят либо универсальные решения с возможностью плавной миграции к NAS-системам, либо системы, предусматривающие возможность их использования как в качестве DAS, так и NAS и даже SAN-систем.

Системы DAS следует использовать при необходимости увеличения дискового пространства одного сервера и вынесения его за корпус. Также DAS-системы можно рекомендовать к применению для рабочих станций, обрабатывающих большие объемы информации (например, для станций нелинейного видеомонтажа).

## 4.2. Сетевые системы хранения данных

### 4.2.1. Общие сведения

Дальнейшим шагом на пути эволюции систем хранения стали сети хранения данных (СХД). **Сеть хранения данных (СХД)** представляет собой архитектурное решение для подключения внешних устройств хранения данных (таких как дисковые массивы, ленточные библиотеки, оптические накопители) к серверам таким образом, чтобы операционная система распознала подключённые ресурсы как локальные.

Сети хранения данных обладают широким спектром преимуществ по сравнению с системами хранения, подключенными к серверам. Они позволяют решить множество задач, связанных с системами хранения. Наиболее популярными задачами, решаемыми с использованием сетей хранения данных, являются: объединение дисковых подсистем, повышение надежности и предоставление услуг хранения данных.

Последние несколько лет большая часть информации на предприятиях хранилась на серверах или на системах хранения данных, которые подключались непосредственно к серверам.

По мере увеличения объема данных и количества накопителей подобная архитектура становилась дорогой, слабо масштабируемой, сложной в управлении и недостаточно надежной. Движущей силой для развития сетей хранения данных стал взрывной рост объема деловой информации (такой как электронная почта, базы данных и высоконагруженные файловые сервера), требующей высокоскоростного доступа к дисковым устройствам на блочном уровне.

Ранее на предприятии возникали «острова» высокопроизводительных дисковых массивов SCSI. Каждый такой массив был выделен для конкретного приложения и виден ему как некоторое количество «виртуальных жёстких дисков» (LUN'ов).

Сеть хранения данных позволяет объединить эти «острова» средствами высокоскоростной сети. Сети хранения помогают повысить эффективность использования ресурсов систем хранения, поскольку дают возможность выделить любой ресурс любому узлу сети.

Для повышения надёжности систем хранения данных применяют устройства резервного копирования, которые также подключаются к СХД. Существуют как промышленные ленточные библиотеки (на несколько тысяч лент) от ведущих брендов, так и low-end решения для малого бизнеса. Сети хранения данных позволяют подключить к одному хосту несколько приводов таких библиотек,

обеспечив таким образом хранилище данных для резервного копирования от сотен терабайт до нескольких петабайт.

Использование сетей хранения данных позволяет существенно сократить затраты на обслуживание и масштабирование систем хранения данных. Сети хранения данных представляют комбинацию аппаратных и программных продуктов, позволяющих не только обеспечить взаимодействие вычислительных подсистем с подсистемами хранения данных, но и предоставить удобные средства для управления всеми компонентами.

Сетевое хранение данных позволяет решить многие текущие задачи, связанные с хранением информации, а именно:

- универсальный и совместный доступ к ресурсам;
- поддержание непредсказуемого, взрывного роста системы ИТ;
- обеспечение непрерывной доступности при сохранении экономичности;
- обеспечение масштабируемости и высочайшей скорости работы хранилища данных;
- создание необходимых условий для работы новых приложений, например, приложений резервного копирования, без участия сервера и LAN;
- упрощение управления ресурсами, связанного с их централизацией;
- повышение уровня защиты информации и отказоустойчивости.

Как уже упоминалось, сети хранения данных (СХД) разделяются на сети, в которых:

- системы хранения данных (диски) соединяются с вычислительными ресурсами (серверами) через специализированную сеть (SAN) – обеспечивает блочный доступ к дисковым подсистемам;
- системы хранения данных, взаимодействие с которыми осуществляется по обычным сетям передачи данных (NAS) – предоставляет файловый доступ.

Технологии хранения данных NAS и SAN, ориентированны на крупные и глобальные сети.

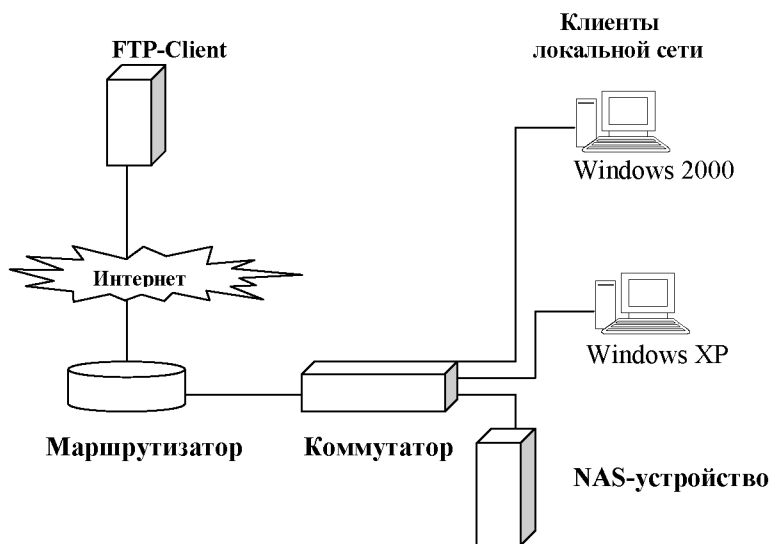
Следует обратить внимание, что категорическое разделение вида «SAN — это только сетевые диски, NAS — это только сетевая файловая система» является искусственным: с появлением iSCSI началось взаимное проникновение технологий с целью повышения гибкости и удобства их применения.

#### **4.2.2. Технология Network Attached Storage(NAS)**

NAS-системы — это сетевая система хранения данных, которая предусматривает подключение накопителей информации

непосредственно к компьютерной сети. Фактически NAS-системы представляют собой эволюцию файл-серверов. Для того, чтобы понять разницу между традиционным файл-сервером и NAS-устройством, вспомним, что традиционный файл-сервер представляет собой выделенный компьютер (сервер), на котором хранится информация, доступная пользователям сети. Для хранения информации могут использоваться жёсткие диски, устанавливаемые в сервер (как правило, они устанавливаются в специальные корзины), либо к серверу могут подключаться DAS-устройства. Администрирование файл-сервера производится с использованием серверной операционной системы. Такой подход к организации систем хранения данных в настоящее время является наиболее популярным в сегменте небольших локальных сетей, однако он имеет один существенный недостаток. Дело в том, что универсальный сервер (да еще в сочетании с серверной операционной системой) — это отнюдь не дешёвое решение. В то же время большинство функциональных возможностей, присущих универсальному серверу, в файл-сервере просто не используется. Идея заключается в том, чтобы создать оптимизированный файл-сервер с оптимизированной операционной системой и сбалансированной конфигурацией. Именно эту концепцию и воплощает в себе NAS-устройство. В этом смысле NAS-устройства можно рассматривать как «тонкие» файл-серверы, или, как их иначе называют, файлеры (filers).

Рисунок 3



Пример NAS-системы хранения данных



Зачастую в инфраструктуре предприятия сервер выполняет ряд функциональных задач, что непосредственно отражается на скорости прямого доступа к данным и, как следствие, вызывает большую загруженность системы в целом. Кроме того, при отказе сервера нет доступа к данным. Нарращивание ёмкости в сервере стоит немалых денег, и нередки случаи, когда на сервере физически не хватает места для очередного добавления накопителей на жёстких дисках.

Сетевая система хранения данных более гибка в этом плане. Архитектура NAS максимально оптимизирована для конкретной задачи: файловый сервис. Вся вычислительная мощность сосредоточена на единственной и главной задаче – обслуживании и хранении файлов. Ограничившись ключевой задачей, NAS-продукты позволяют организовать работу группы пользователей с общими файлами максимально эффективно с точки зрения быстродействия и затрат.

Кроме оптимизированной ОС, освобожденной от всех функций, не связанных с обслуживанием файловой системы и реализацией ввода-вывода данных, NAS-системы имеют оптимизированную по скорости доступа файловую систему. NAS-системы проектируются таким способом, что вся их вычислительная мощь фокусируется исключительно на операциях обслуживания и хранения файлов. Сама операционная система располагается во флэш-памяти и предустанавливается фирмой-производителем. Естественно, что с выходом новой версии ОС пользователь может самостоятельно «перепрошить» систему. Подсоединение NAS-устройств к сети и их конфигурирование представляет собой достаточно простую задачу и по силам любому опытному пользователю, не говоря уже о системном администраторе.

Таким образом, в сравнении с традиционными файловыми серверами NAS-устройства являются более производительными и менее дорогими. В настоящее время практически все NAS-устройства ориентированы на использование в сетях Ethernet (Fast Ethernet, Gigabit Ethernet) на основе протоколов TCP/IP.

Доступ к устройствам NAS производится с помощью специальных протоколов доступа к файлам. Наиболее распространенными протоколами файлового доступа являются протоколы CIFS, NFS и DAFS.

**CIFS** (Common Internet File System — общая файловая система Интернета) — это протокол, который обеспечивает доступ к файлам и сервисам на удаленных компьютерах (в том числе и в Интернет) и использует клиент-серверную модель взаимодействия. Клиент создает запрос к серверу на доступ к файлам, сервер выполняет запрос клиента и возвращает результат своей работы. Протокол CIFS традиционно используется в локальных сетях с ОС Windows для доступа к файлам.

Для транспортировки данных CIFS использует TCP/IP протокол. CIFS обеспечивает функциональность, похожую на FTP (File Transfer Protocol), но предоставляет клиентам улучшенный контроль над файлами. Он также позволяет разделять доступ к файлам между клиентами, используя блокирование и автоматическое восстановление связи с сервером в случае сбоя сети.

Протокол NFS (Network File System — сетевая файловая система) традиционно применяется на платформах UNIX и представляет собой совокупность распределенной файловой системы и сетевого протокола. В протоколе NFS также используется клиент-серверная модель взаимодействия. Протокол NFS обеспечивает доступ к файлам на удалённом хосте (сервере) так, как если бы они находились на компьютере пользователя. Для транспортировки данных NFS использует протокол TCP/IP. Для работы NFS в Интернете был разработан протокол WebNFS.

Протокол DAFS (Direct Access File System — прямой доступ к файловой системе) — это стандартный протокол файлового доступа, который основан на NFS. Данный протокол позволяет прикладным задачам передавать данные в обход операционной системы и её буферного пространства напрямую к транспортным ресурсам. Протокол DAFS обеспечивает высокие скорости файлового ввода-вывода и снижает загрузку процессора благодаря значительному уменьшению количества операций и прерываний, которые обычно необходимы при обработке сетевых протоколов.

DAFS проектировался с ориентацией на использование в кластерном и серверном окружении для баз данных и разнообразных Интернет-приложений, ориентированных на непрерывную работу. Он обеспечивает наименьшие задержки доступа к общим файловым ресурсам и данным, а также поддерживает интеллектуальные механизмы восстановления работоспособности системы и данных, что делает его привлекательным для использования в NAS-системах.

Для подсоединения NAS-устройств к сети не требуется каких-либо специфических интерфейсов или специальных устройств. Достаточно подключить Network Attached Storage к сети, как все его ресурсы становятся доступны для пользователей, у которых NAS появляется в виде дополнительных дисков.

Гибкость NAS-систем: простое и удобное администрирование, возможность наращивания объёма без отключения основной системы, использование стандартных коммуникационных протоколов и сетевых подключений позволяет использовать их в мультиплатформенных сетях в случае, когда требуется сетевой доступ к файлам. Прекрасным примером является применение NAS в качестве файл-сервера в офисе небольшой компании.

### 4.2.3. Технология Storage Area Network (SAN)

Собственно, SAN — это уже не отдельное устройство, а комплексное решение, представляющее собой специализированную сетевую инфраструктуру для хранения данных. Сети хранения данных интегрируются в виде отдельных специализированных подсетей в состав локальной (LAN) или глобальной (WAN) сети.

По сути, SAN-сети связывают один или несколько серверов (SAN-серверов) с одним или несколькими устройствами хранения данных. SAN-сети позволяют любому SAN-серверу получать доступ к любому устройству хранения данных, не загружая при этом ни другие серверы, ни локальную сеть. Кроме того, возможен обмен данными между устройствами хранения данных без участия серверов. Фактически SAN-сети позволяют очень большому числу пользователей хранить информацию в одном месте (с быстрым централизованным доступом) и совместно использовать её. В качестве устройств хранения данных могут использоваться RAID-массивы, различные библиотеки (ленточные, магнитооптические и др.), а также JBOD системы (массивы дисков, не объединенные в RAID).

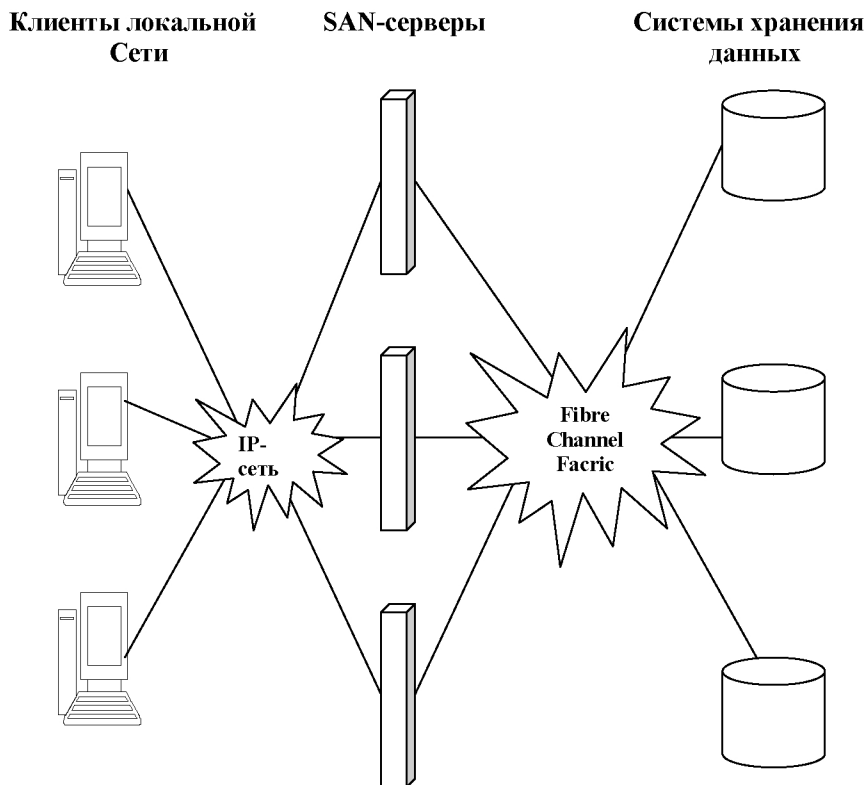
Подобно тому, как локальные сети в принципе могут строиться на основе различных технологий и стандартов, для построения сетей SAN также могут применяться различные технологии. Но точно так же, как стандарт Ethernet (Fast Ethernet, Gigabit Ethernet) стал стандартом де-факто для локальных сетей, в сетях хранения данных доминирует стандарт *Fibre Channel* (FC). Собственно, именно развитие стандарта *Fibre Channel* привело к развитию самой концепции SAN. В то же время необходимо отметить, что все большую популярность приобретает стандарт iSCSI, на основе которого тоже возможно построение SAN сетей.

Наряду со скоростными параметрами одним из важнейших преимуществ *Fibre Channel* является возможность работы на больших расстояниях и гибкость топологии. Концепция построения топологии сети хранения данных базируется на тех же принципах, что и традиционные локальные сети на основе коммутаторов и маршрутизаторов, что значительно упрощает построение многоузловых конфигураций систем.

Стоит отметить, что для передачи данных в стандарте *Fibre Channel* используются как оптоволоконные, так и медные кабели. При организации доступа к территориально удаленным узлам на расстоянии до 10 км используется стандартная аппаратура и одномодовое оптоволокно для передачи сигнала. Если же узлы разнесены на большее расстояние (десятки или даже сотни километров), применяются специальные усилители.

### 4.2.3.1. Топология SAN-сети

Рисунок 4



Типичная схема SAN-сети

Типичный вариант SAN-сети на основе стандарта Fibre Channel показан на (рис. 4). Инфраструктуру такой SAN-сети составляют устройства хранения данных с интерфейсом Fibre Channel, SAN серверы (серверы, подключаемые как к локальной сети по интерфейсу Ethernet, так и к SAN сети по интерфейсу *Fiber Channel*) и коммутационная фабрика (Fibre Channel Fabric), которая строится на основе Fibre Channel коммутаторов (концентраторов) и оптимизирована для передачи больших блоков данных. Доступ сетевых пользователей к системе хранения данных реализуется через SAN-серверы. При этом важно, что трафик внутри SAN сети отделен от IP трафика локальной сети, что, безусловно, позволяет снизить загрузку локальной сети.

### 4.2.3.2. Преимущества SAN сетей

К основным преимуществам технологии SAN можно отнести высокую производительность, высокий уровень доступности данных, отличную масштабируемость и управляемость, возможность консолидации и виртуализации данных.

Коммутационные фабрики *Fiber Channel* с неблокирующей архитектурой позволяют реализовать одновременный доступ множества SAN-серверов к устройствам хранения данных.

В архитектуре SAN данные могут легко перемещаться с одного устройства хранения данных на другое, что позволяет оптимизировать размещение данных. Это особенно важно в том случае, когда нескольким SAN-серверам требуется одновременный доступ к одним и тем же устройствам хранения данных. Отметим, что процесс консолидации данных невозможен в случае использования других технологий, как, например, при применении DAS-устройств, то есть устройств хранения данных, непосредственно подсоединяемых к серверам.

Другая возможность, предоставляемая архитектурой SAN, — это виртуализация данных. Идея виртуализации заключается в том, чтобы обеспечить SAN-серверам доступ не к отдельным устройствам хранения данных, а к ресурсам. То есть серверы должны «видеть» не устройства хранения данных, а виртуальные ресурсы. Для практической реализации виртуализации между SAN-серверами и дисковыми устройствами может размещаться специальное устройство виртуализации, к которому с одной стороны подключаются устройства хранения данных, а с другой — SAN-серверы. Кроме того, многие современные FC-коммутаторы и HBA-адаптеры предоставляют возможность реализации виртуализации.

Следующая возможность, предоставляемая SAN сетями, — это реализация удалённого зеркалирования данных. Принцип зеркалирования данных заключается в дублировании информации на несколько носителей, что повышает надежность хранения информации. Примером простейшего случая зеркалирования данных может служить объединение двух дисков в raid-массив уровня 1. В данном случае одна и та же информация записывается одновременно на два диска. Недостатком такого способа можно считать локальное расположение обоих дисков (как правило, диски находятся в одной и той же корзине или стойке). Сети хранения данных позволяют преодолеть этот недостаток и предоставляют возможность организации зеркалирования не просто отдельных устройств хранения данных, а самих SAN-сетей, которые могут быть удалены друг от друга на сотни километров.

Ещё одно преимущество SAN-сетей заключается в простоте организации резервного копирования данных. Традиционная технология резервного копирования, которая используется в большинстве локальных сетей, требует выделенного Backup-сервера и, что особенно важно, выделенной полосы пропускания сети. Фактически во время операции резервного копирования сам сервер становится недоступным для пользователей локальной сети. Собственно, именно поэтому резервное копирование производится, как правило, в ночное время.

Архитектура сетей хранения данных позволяет принципиально по иному подойти к проблеме резервного копирования. В этом случае Backup-сервер является составной частью SAN-сети и подключается непосредственно к коммутационной фабрике. В этом случае Backup-трафик оказывается изолированным от трафика локальной сети.

#### **4.2.3.3. Оборудование, используемое для создания SAN сетей**

Как уже отмечалось, для развертывания SAN-сети требуются устройства хранения данных, SAN-серверы и оборудование для построения коммутационной фабрики. Коммутационные фабрики включают как устройства физического уровня (кабели, коннекторы), так и устройства подключения (Interconnect Device) для связи узлов SAN друг с другом, устройства трансляции (Translation devices), выполняющие функции преобразования протокола Fibre Channel (FC) в другие протоколы, например SCSI, FCP, FICON, Ethernet, ATM или SONET.

- **Кабели**

Как уже отмечалось, для соединения SAN-устройств стандарт *Fibre Channel* допускает использование как волоконно-оптических, так и медных кабелей. При этом в одной SAN-сети могут применяться различные типы кабелей. Медный кабель используется для коротких расстояний (до 30 м), а волоконно-оптический — как для коротких, так и для расстояний до 10 км и больше. Применяют как многомодовый (Multimode), так и одномодовый (Singlemode) волоконно-оптические кабели, причем многомодовый используется для расстояний до 2 км, а одномодовый — для больших расстояний.

Существование различных типов кабелей в пределах одной SAN-сети обеспечивается посредством специальных конверторов интерфейсов GBIC (Gigabit Interface Converter) и MIA (Media Interface Adapter).

В стандарте *Fibre Channel* предусмотрено несколько возможных скоростей передачи (см. таблицу 2). Отметим, что в настоящее время

наиболее распространены FC-устройства стандартов 1, 2 и 4 GFC. При этом обеспечивается обратная совместимость более скоростных устройств с менее скоростными, то есть устройство стандарта 4 GFC автоматически поддерживает подключение устройств стандартов 1 и 2 GFC.

Таблица 2. Скорость передачи данных в стандарте *Fibre Channel*

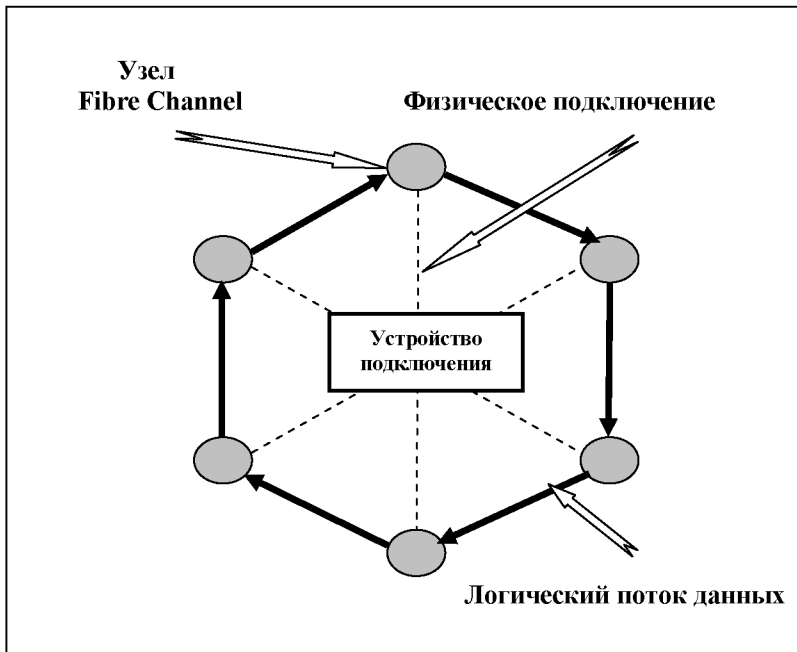
Обозначение	Полоса пропускания в дуплексном режиме (Мбайт/с)	Доступность на рынке
1 GFC	200	1997
2 GFC	400	2001
4 GFC	800	2005
8 GFC	1600	2008
16 GFC	3200	2011

- **Топологии подключения устройств**

В стандарте *Fibre Channel* допускается использование различных сетевых топологий подключения устройств, таких как «точка-точка» (Point to Point), кольцо с разделяемым доступом (Arbitrated Loop, FC\_AL) и коммутируемая связная архитектура (switched fabric).

Топология «точка-точка» может применяться для подключения сервера к выделенной системе хранения данных. В этом случае данные не используются совместно с серверами SAN-сети. Фактически данная топология является вариантом DAS-системы.

## Система хранения данных



Топология «точка-точка»

Для реализации топологии «точка-точка», как минимум, необходим сервер, оснащенный адаптером *Fibre Channel*, и **устройство хранения данных** с интерфейсом *Fibre Channel*.

Топология кольца с разделенным доступом (FC\_AL) подразумевает схему подключения устройств, при котором данные передаются по логически замкнутому контуру. При топологии кольца FC\_AL в качестве устройств подключения могут выступать концентраторы или коммутаторы *Fibre Channel*. При использовании концентраторов полоса пропускания делится между всеми узлами кольца, в то время как каждый порт коммутатора предоставляет протокольную полосу пропускания для каждого узла.

На рис. 5 показан пример кольца *Fibre Channel* с разделением доступа.

Конфигурация аналогична физической звезде и логическому кольцу, используемым в локальных сетях на базе технологии Token



Ring. Кроме того, как и в сетях Token Ring, данные перемещаются по кольцу в одном направлении, но, в отличие от сетей Token Ring, устройство может запросить право на передачу данных, а не ждать получения пустого маркера от коммутатора. Кольца *Fibre Channel* с разделением доступа могут адресовать до 127 портов, однако, как показывает практика, типичные кольца FC\_AL содержат до 12 узлов, а после подключения 50 узлов производительность катастрофически снижается.

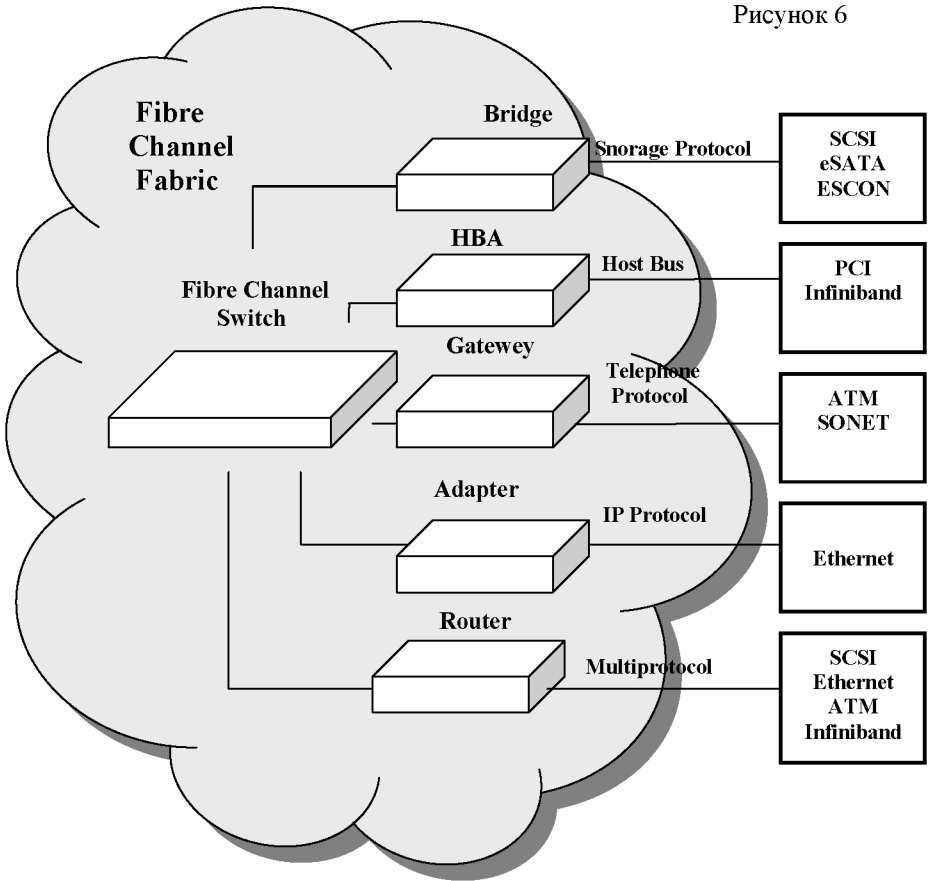
Топология коммутируемой связной архитектуры (*Fibre Channel switched\_fabric*) реализуется на базе *Fibre Channel*-коммутаторов. В данной топологии каждое устройство имеет логическое подключение к любому другому устройству. Фактически *Fibre Channel*-коммутаторы связной архитектуры выполняют те же функции, что и традиционные Ethernet-коммутаторы. Напомним, что, в отличие от концентратора, коммутатор — это высокоскоростное устройство, которое обеспечивает подключение по схеме «каждый с каждым» и обрабатывает несколько одновременных подключений. Любой узел, подключенный к *Fibre Channel*-коммутатору, получает протокольную полосу пропускания.

В большинстве случаев при создании крупных SAN-сетей используется смешанная топология. На нижнем уровне применяются FC AL кольца, подключенные к малопроизводительным коммутаторам, которые, в свою очередь, подключаются к высокоскоростным коммутаторам, обеспечивающим максимально возможную пропускную способность. Несколько коммутаторов могут быть соединены друг с другом.

- **Устройства трансляции**

Устройства трансляции являются промежуточными устройствами, выполняющими преобразование протокола *Fibre Channel* в протоколы более высоких уровней (рис. 6). Эти устройства предназначены для соединения *Fibre Channel* сети с внешней WAN сетью, локальной сетью, а также для присоединения к *Fibre Channel* сети различных устройств и серверов. К таким устройствам относятся мосты (Bridge), *Fibre Channel* адаптеры (Host Bus Adapters (HBA)), маршрутизаторы, шлюзы и сетевые адаптеры. Классификация устройств трансляции показана на (рис. 5).

Рисунок 6



Классификация устройств трансляции

Наиболее распространенными устройствами трансляции являются HBA адаптеры с интерфейсом PCI, которые применяются для подключения серверов к сети Fibre Channel. Сетевые адаптеры позволяют подключать локальные Ethernet сети к сетям Fibre Channel. Мосты используются для подключения устройств хранения данных с SCSI интерфейсом к сети на базе Fibre Channel. Следует отметить, что в последнее время практически все **устройства хранения данных**,

которые предназначены для применения в SAN, имеют встроенный Fibre Channel и не требуют использования мостов.

- **Устройства хранения данных**

В качестве **устройств хранения данных** в SAN-сетях могут использоваться как жёсткие диски, так и ленточные накопители. Если говорить о возможных конфигурациях применения жёстких дисков в качестве устройств хранения данных в SAN-сетях, то это могут быть как массивы JBOD, так и raid-массивы дисков. Традиционно устройства хранения данных для SAN-сетей выпускаются в виде внешних стоек или корзин, оснащённых специализированным *RAID-контроллером*. В отличие от NAS или DAS-устройств, устройства для SAN-систем оснащаются Fibre Channel интерфейсом. При этом сами диски могут иметь как SCSI, так и SATA интерфейс.

Кроме устройств хранения на основе жестких дисков, в SAN-сетях широкое применение находят ленточные накопители и библиотеки.

- **SAN-серверы**

Серверы для сетей SAN отличаются от обычных серверов приложений только одной деталью. Кроме сетевого Ethernet-адаптера, для взаимодействия сервера с локальной сетью они оснащаются HBA-адаптером, что позволяет подключать их к SAN-сетям на основе Fibre Channel.

#### **4.2.3.4. Совместное использование СХД**

1. Совместное использование систем хранения, как правило, упрощает администрирование и добавляет изрядную гибкость, поскольку кабели и дисковые массивы не нужно физически транспортировать и перекоммутировать от одного сервера к другому.
2. Другим преимуществом является возможность загружать сервера прямо из сети хранения. При такой конфигурации можно быстро и легко заменить сбойный сервер, переконфигурировав SAN таким образом, что сервер-замена, будет загружаться с LUN'a сбойного сервера.
3. Дополнительным преимуществом является возможность на хосте собрать RAID-зеркало с LUNов, которые презентованы хосту с двух разных дисковых массивов. В таком случае полный отказ одного из массивов не навредит хосту.

4. Также сети хранения помогают более эффективно восстанавливать работоспособность после сбоя. В SAN может входить удаленный участок с вторичным устройством хранения. В таком случае можно использовать репликацию — реализованную на уровне контроллеров массивов, либо при помощи специальных аппаратных устройств. Поскольку каналы WAN на основе протокола IP встречаются часто, были разработаны протоколы Fibre Channel over IP (FCIP) и iSCSI с целью расширить единую SAN средствами сетей на основе протокола IP.

### 4.3. Сравнение технологий обмена данными

Порой сравнивают SAN и NAS, говоря на самом деле о разнице между сетевым диском и сетевой файловой системой (ФС), которая состоит в том, кто обслуживает файловую систему, хранящую данные.

В случае сетевого диска (также «блочного устройства», англ. block device):

- обмен данными с ним по сети осуществляется блоками подобно тому, как и с локальным SCSI- или SATA-дискон;
- файловая система, если нужна, создаётся и управляется клиентом и, как правило — используется им одним.

В случае сетевой файловой системы («расширенный ресурс» — не хранит, а только передаёт данные):

- обмен данными по сети происходит с применением более высокоуровневых понятий «файл» и «каталог», соответствующих объектам подлежащей «настоящей» ФС на физических дисках (либо логических поверх них в случае применения RAID, LVM);
- эта файловая система создаётся и обслуживается в рамках удалённой системы, при этом может одновременно использоваться на чтение и запись множеством клиентов.

Доступ к системам хранения может быть как блочным, так и файловым. Услуги могут предоставляться как через инфраструктуру публичных сетей (интернет), так и по выделенным оптическим сетям. Независимо от типа используемой сети, инфраструктура должна обеспечивать высокую производительность, надежность и защищенность данных.

## 5. Файловые системы

Файловая система (англ. file system) — порядок, определяющий способ организации, хранения и именования данных на носителях информации IT-оборудования и компьютерной техники. Она определяет

формат содержимого и физического хранения информации, которую принято группировать в виде файлов. Конкретная файловая система определяет размер имени файла (папки), максимальный возможный размер файла и раздела, набор атрибутов файла. Некоторые файловые системы предоставляют сервисные возможности, например, разграничение доступа или шифрование файлов.

Файловая система связывает носитель информации с одной стороны и средства для доступа к файлам — с другой. Когда прикладная программа обращается к файлу, она не имеет никакого представления о том, каким образом расположена информация в конкретном файле, так же, как и на каком физическом типе носителе он записан. Всё, что знает программа — это имя файла, его размер и атрибуты. Эти данные она получает от драйвера файловой системы. Именно файловая система устанавливает, где и как будет записан файл на физическом носителе (например, жёстком диске).

С точки зрения операционной системы (ОС), весь диск представляет собой набор кластеров (как правило, размером 512 байт и больше). Драйверы файловой системы организуют кластеры в файлы и каталоги (реально являющиеся файлами, содержащими список файлов в этом каталоге). Эти же драйверы отслеживают, какие из кластеров в настоящее время используются, какие свободны, какие помечены как неисправные.

Однако файловая система не обязательно напрямую связана с физическим носителем информации. Существуют виртуальные файловые системы, а также сетевые файловые системы, которые являются лишь способом доступа к файлам, находящимся на удалённом компьютере.

По предназначению файловые системы можно классифицировать на нижеследующие категории:

- **Дисковые файловые системы**

Дисковые файловые системы обычно являются поток-ориентированными. Файлы в поток-ориентированных файловых системах представляются последовательностью битов, часто предоставляющие такие функции, как чтение, запись, изменение данных и произвольный доступ.

- **Файловые системы для флеш-дисков / твердотельных носителей**

Твердотельные носители, такие, как флеш-диски, своим интерфейсом данных похожи на обычные жёсткие диски, но имеют свои проблемы и недостатки. Когда проходит время поиска, они нуждаются в особой обработке такими алгоритмами как, Wear leveling и Error detection and correction.

- **Запись-ориентированные файловые системы**

В запись-ориентированных файловых системах файлы хранятся как коллекция записей. Такие файловые системы ассоциируются, прежде всего, со старыми мейнфреймами и операционными системами для миникомпьютеров. Программы считывают и записывают целыми записями, вместо байт, записанных, в определенном порядке.

- **Файловые системы для сетевых хранилищ**

Файловые системы для общих дисков (также известные как Файловые системы для сетевых (общих) хранилищ (файловая система SAN) или кластерные файловые системы) в основном используются в сетевых хранилищах, где все узлы сети имеют прямой доступ к блоковому устройству хранения, где расположена эта файловая система. Такие файловые системы функционируют, даже при поломке одного из узлов. Данные файловые системы, обычно, используются в кластерах высокой доступности вместе с аппаратным RAID. Файловые системы для сетевых хранилищ обычно не расширяются больше 64 или 128 узлов.

Могут быть симметричными, когда метаданные распределены между узлами, или асимметричными — с централизованными хранилищами метаданных.

- **Распределённые файловые системы**

Распределенные файловые системы известны и как сетевые файловые системы.

- **Распределенные параллельные файловые системы с защитой от сбоев**

Распределенные файловые систем, являющиеся параллельными и с защитой от сбоев, разделяют и реплицируют данные на многие сервера для высокой производительности и обеспечения целостности данных. Даже когда сервер дает сбой, данные не теряются. Данные файловые системы используются в высокоскоростных вычислениях и кластерах высокой доступности.

К таким системам относится GPFS — General Parallel File System, виртуальная файловая система для параллельной обработки данных, разработанная IBM. Поддерживает репликацию между подсоединенными блочными устройствами хранения. Доступна в AIX, Linux и Windows. Синхронная или асинхронная репликация (настраиваемо).

## **6. Параллельная файловая система GPFS**

Любая файловая система — это способ организации файлов и данных, обеспечивающий доступ к информации и поиск. Сетевая

файловая система, в свою очередь, позволяет осуществить удаленный доступ к файлам.

Традиционные системы сетевого хранения данных обладают рядом недостатков. В первую очередь это ограничения масштабируемости — при достижении предела производительности или ёмкости приходится переходить на другую, более мощную систему. Такое вертикальное масштабирование можно продолжать, пока не будет достигнут предел наиболее мощных существующих систем хранения.

Альтернативным подходом является построение кластерных систем для доступа к данным. В этом случае используется программно-аппаратное решение, предоставляющее пользователю доступ к одному виртуальному узлу, за которым скрывается кластерная система хранения. Увеличение размеров такой системы носит характер горизонтального масштабирования и позволяет обойти ограничения обычных систем хранения.

Решение IBM SoFS позволяет преодолеть существующие пределы производительности и масштабирования сетевых систем хранения. Аббревиатура SoFS расшифровывается IBM как Scale-out File Services — "неограниченно масштабируемые файловые сервисы" в глобально распределенной среде.

SoFS сочетает уникальную кластерную файловую систему и масштабируемые службы доступа к данным, включающие Samba/CIFS (доступ из сетей Windows), NFS и другие протоколы, такие как FTP или HTTP. Уникальность решения IBM SoFS в том, что оно сочетает в себе свойства системы хранения для высокопроизводительных вычислений (кластерная организация, масштабируемость и скорость доступа) и классических сетевых систем хранения (простота настройки и использования).

Для обеспечения высоких требований к производительности сетевое хранилище должно распределять все файлы между максимальным числом дисков, контроллеров и узлов системы хранения. Другое широко распространенное решение заключается в разделении файлов по различным системам хранения в зависимости от их назначения. Но это увеличивает сложность администрирования, добавляет новые пространства имен, тогда как ресурсы в таких системах распределяются не оптимально.

SoFS базируется на параллельной файловой системе IBM General Parallel File System (GPFS), одна из ключевых возможностей которой — распространение блоков одного файла по всем узлам кластера, что позволяет объединить производительность всех подключенных систем хранения и достигнуть скорости передачи данных, измеряемой сотнями гигабайт в секунду. Но возможности GPFS этим не ограничиваются: для управления жизненным циклом данных

используется специальная гибкая система политик, существуют средства репликации данных и кросс-мониторинга между удаленными локациями. Используя GPFS, можно объединять множество систем хранения разного класса и скорости в единую виртуальную файловую систему. Единое глобальное пространство имен в системе хранения SoFS достигается за счет технологий виртуализации и перенаправления данных, реализованных в GPFS-кластере. Каждый узел в кластере имеет доступ ко всем блокам данных одновременно и может заменить любой другой узел кластера в случае его неисправности. Таким образом, конечные приложения, обращаясь к какому-то конкретному узлу кластера, работают виртуально со всем массивом данных. Ни одна из существующих на данный момент кластерных технологий не обеспечивает прозрачный непрерывный доступ через сетевую файловую систему в случае неисправности серверного узла, с которым идет обмен. Решение SoFS базируется на чрезвычайно масштабируемой кластерной технологии, которая учитывает семантику используемых сетевых протоколов доступа к данным и обеспечивает минимальное время восстановления работы системы для клиентской стороны при очень большой скорости доступа.

Решение SoFS — это аппаратно-программный комплекс. Кластерными узлами системы могут являться блэйд-серверы IBM или серверы линейки IBM System x., а в качестве систем хранения поддерживаются все продукты из IBM System Storage.

Сердцем SoFS является кластерная файловая система IBM GPFS, которая предоставляет всем узлам кластера общую файловую систему с единым пространством имен. GPFS базируется на модели разделяемых дисков: все узлы имеют одновременный доступ на чтение и запись к группам блочных устройств (например, RAID-массивам, подключенным через Fibre Channel). GPFS реализует интерфейс файловой системы, совместимый с POSIX, так что файловые серверы в SoFS — Samba и NFS — обращаются к данным на когерентной файловой системе, совершенно аналогичной локальной файловой системе. Разделяемые диски при этом содержат как данные, так и метаданные файловой системы.

В отличие от многих других кластерных файловых систем, в GPFS нет выделенного сервера для метаданных. Узлы GPFS- или SoFS-кластера обращаются к метаданным непосредственно, синхронизируя свою работу посредством менеджера блокировок GPFS. В случае SoFS менеджер блокировок GPFS реализует две основные функции: он управляет целостностью кэша, чтобы обновление данных и метаданных могло происходить одновременно на множестве узлов, и он управляет блокировками NFS, чтобы клиенты NFS могли восстанавливать блокировки в случае неисправности одного из узлов кластера.



В GPFS есть ряд возможностей, обеспечивающих высокую надежность и отказоустойчивость. В системах хранения отказоустойчивость обычно реализуется на аппаратном уровне (технология RAID, дублирование контроллеров и т.п.), но в GPFS имеется собственная репликация блоков, которая может использоваться вместо (или в дополнение к) RAID. Для того, чтобы реагировать на отказ узла кластера, GPFS хранит журнал всех изменений метаданных на разделяемом диске. Когда узел кластера выходит из строя, он первым делом ограждается от разделяемых дисков для предотвращения возможности некорректной записи. После этого один из оставшихся узлов обновляет данные согласно журналу, чтобы восстановить все прерванные файловые операции. В GPFS всего несколько разделяемых служб: конфигурации, блокировок, кворума и квоты, все они могут работать на любом из узлов кластера. Все команды по администрированию системы (такие как добавление и удаление узлов, балансировка данных с добавлением новой системы хранения и т.п.) выполняются на лету, без остановки кластера. Длительные операции, такие как повторная балансировка, могут быть запущены повторно. То же касается и обновлений программного обеспечения — GPFS позволяет устанавливать обновления без остановки работы кластера.

GPFS разрабатывалась для очень больших файловых систем, одна файловая система может включать в себя до 2000 массивов, каждый из которых может быть, к примеру, группой четности RAID. В 2007 г. самая большая система хранения под управлением GPFS содержала 2 петабайта. GPFS использует кэширование имен файлов в директориях, что повышает производительность поиска даже для большого числа файлов.

GPFS разработана для высокопроизводительных вычислительных кластеров и обеспечивает возможность одновременного высокоскоростного доступа к файлам и приложениям, находящимся на нескольких физических узлах. Файловая система предназначена для использования на крупных предприятиях, а масштабируемость и высокая производительность GPFS позволяет оперативно обрабатывать данные в таких областях как техническое проектирование, анализ данных, медиаиндустрия, а также в научных исследованиях. С помощью технологии GPFS создаются масштабируемые параллельные файловые системы, которые имеют возможность надежно поддерживать сотни терабайт хранимой информации в рамках одной файловой системой. Компания IBM сделала возможным поддержку GPFS технологии на аппаратных средствах других производителей.

С недавнего времени технологии GPFS используют в построении решений для обработки цифрового видео в телекомпаниях,

что позволяет редактировать файлы большого объема несколькими командами одновременно. Например, готовить из снятого материала сюжет для новостей и полноценную передачу параллельно двум монтажным студиям. В IBM рассчитывают на то, что технология найдет применение и в изготовлении передач телевидения высокого разрешения.

Эта файловая система используется во многих крупнейших суперкомпьютерах.

## 7. Подключение СХД к суперкомпьютеру «Сергей Королёв»

**Суперкомпьютер** (англ. supercomputer, СуперЭВМ) — вычислительная машина, значительно превосходящая по своим техническим параметрам большинство существующих компьютеров. Как правило, современные суперкомпьютеры представляют собой большое число высокопроизводительных серверных компьютеров, соединённых друг с другом локальной высокоскоростной магистралью для достижения максимальной производительности в рамках подхода распараллеливания вычислительной задачи.

Сервером называется компьютер, выделенный из группы персональных компьютеров (или рабочих станций) для выполнения какой-либо сервисной задачи без непосредственного участия человека. Компьютеры в локальной сети подразделяются на рабочие станции и серверы. На рабочих станциях пользователи решают прикладные задачи (работают в базах данных, создают документы, делают расчёты). Сервер обслуживает сеть и предоставляет собственные ресурсы всем узлам сети, в том числе и рабочим станциям. Сервер и рабочая станция могут иметь одинаковую аппаратную конфигурацию, так как различаются лишь по участию в своей работе человека за консолью.

Некоторые сервисные задачи могут выполняться на рабочей станции параллельно с работой пользователя. Такую рабочую станцию условно называют невыделенным сервером.

Консоль (обычно — монитор/клавиатура/мышь) и участие человека необходимы серверам только на стадии первичной настройки, при аппаратно-техническом обслуживании и управлении в нестандартных ситуациях (штатно, большинство серверов управляются удаленно). Для нестандартных ситуаций серверы обычно обеспечиваются одним консольным комплектом на группу серверов (с коммутатором, например KVM-переключателем, или без такового).

В результате специализации (см. ниже), серверное решение может получить консоль в упрощенном виде (например,

коммуникационный порт), или потерять ее вовсе (в этом случае первичная настройка и штатное управление могут выполняться только через сеть, а сетевые настройки могут быть сброшены в состояние по умолчанию).

Специализация серверного оборудования идет несколькими путями, выбор того, в каком направлении идти, каждый производитель определяет для себя сам. Большинство специализаций удорожают оборудование.

Конструктивно аппаратные серверы могут исполняться в настольном, напольном, стоечном и потолочном вариантах. Последний вариант обеспечивает наибольшую плотность размещения вычислительных мощностей на единицу площади, а также максимальную масштабируемость. Система называется *масштабируемой*, если она способна увеличивать производительность пропорционально дополнительным ресурсам. Масштабируемость можно оценить через отношение прироста производительности системы к приросту используемых ей ресурсов. Чем ближе это отношение к единице, тем лучше. Также под масштабируемостью понимается возможность наращивания дополнительных ресурсов без структурных изменений центрального узла системы.

С конца 1990-х всё большую популярность в системах высокой надёжности и масштабируемости получили так называемые блэйд-серверы (от англ. *blade* — лезвие) — компактные модульные устройства, позволяющие сократить расходы на электропитание, охлаждение, обслуживание и т. п. Блэйд-серверы — компьютерные серверы с компонентами, вынесенными и обобщёнными в корзине для уменьшения занимаемого пространства. Корзина (англ. *enclosure*) — обойма для блэйд-серверов, предоставляющая им доступ к общим компонентам, например, блокам питания и сетевым контроллерам. Блэйд-серверы называют также ультракомпактными серверами.

Для вычислений компьютеру требуются как минимум следующие части (машина Тьюринга):

- память, содержащая исходные данные,
- процессор, выполняющий команды,
- память для записи результатов.

Остальные компоненты, типичные для компьютера, выполняют вспомогательные для вычислений функции, такие как ввод и вывод, обеспечение питания. Внутри сервера они представляют собой дополнительные потребители энергии, источники тепла, причины сбоев (особенно компоненты с движущимися частями). Концепция блэйд-сервера предусматривает замену их внешними агрегатами (блоки питания) или виртуализацию (порты ввода-вывода, консоли

управления), тем самым значительно упрощая и облегчая сам сервер, а также делая его производство (теоретически) дешевле.

В блэйд-сервере отсутствуют или вынесены наружу некоторые типичные компоненты, традиционно присутствующие в компьютере. Функции питания, охлаждения, сетевого подключения, подключения жёстких дисков, межсерверных соединений и управления могут быть возложены на внешние агрегаты. Вместе с ними набор серверов образует т. н. блэйд-систему.

Блэйд-системы состоят из набора блэйд-серверов и внешних компонентов, обеспечивающих невычислительные функции. Как правило, за пределы серверной материнской платы выносят компоненты, создающие много тепла, занимающие много места, а также повторяющиеся по функциям между серверами. Их ресурсы могут быть распределены между всем набором серверов. Деление на встроенные и внешние функции варьируется у разных производителей.

Преобразователь напряжения питания, как правило, создается общим для блэйд-системы. Он может быть, как вмонтирован внутрь неё, так и вынесен в отдельный блок. По сравнению с суммой отдельных блоков питания, необходимых серверам формата 1U, единый источник питания блэйд-систем — один из самых весомых источников экономии пространства, энергопотребления и числа электронных компонентов.

Традиционный дизайн серверов пытается сбалансировать плотность размещения электронных компонентов и возможность циркуляции охлаждающего воздуха между ними. В блэйд-конструкциях количество выступающих и крупных частей сведено к минимуму, что улучшает охлаждение модулей.

Блэйд-серверы особенно эффективны для решения специфических задач: веб-хостинга, организации кластеров. Серверы в стойке, как правило, поддерживают горячую замену.

При всех достоинствах, эту технологию нельзя считать решением всех серверных проблем. Крупные задачи требуют все же применения более масштабных систем для своего решения, таких как мэйнфреймы и кластеры. Также могут быть использованы кластеры, состоящие из блэйд-серверов. Такая структура особенно подвержена проблеме перегрева ввиду плотной компоновки электроники в каждом из них.

В настоящее время суперкомпьютерами принято называть компьютеры с огромной вычислительной мощностью («числодробилки» или «числогрызы»). Такие машины используются для работы с приложениями, требующими наиболее интенсивных вычислений (например, прогнозирование погодно-климатических условий, моделирование ядерных испытаний и т. п.), что в том числе отличает их от серверов и мэйнфреймов (англ. mainframe — компьютеров с высокой

общей производительностью, призванных решать типовые задачи (например, обслуживание больших баз данных или одновременная работа с множеством пользователей).

Иногда суперкомпьютеры используются для работы с одним-единственным приложением, использующим всю память и все процессоры системы; в других случаях они обеспечивают выполнение большого числа разнообразных приложений.

Наиболее распространёнными программными средствами суперкомпьютеров, также как и параллельных или распределённых компьютерных систем, являются интерфейсы программирования приложений (API) на основе MPI и PVM, и решения на базе открытого программного обеспечения, наподобие Beowulf и openMosix, позволяющего создавать виртуальные суперкомпьютеры даже на базе обыкновенных рабочих станций и персональных компьютеров. Для быстрого подключения новых вычислительных узлов в состав узкоспециализированных кластеров применяются технологии наподобие ZeroConf.

В настоящее время границы между суперкомпьютерным и общепотребимым программным обеспечением сильно размыты и продолжают размываться ещё более вместе с проникновением технологий параллелизации и многоядерности в процессорные устройства персональных компьютеров и рабочих станций. Исключительно суперкомпьютерным программным обеспечением сегодня можно назвать лишь специализированные программные средства для управления и мониторинга конкретных типов компьютеров, а также уникальные программные среды, создаваемые в вычислительных центрах под «собственные», уникальные конфигурации суперкомпьютерных систем.

**Суперкомпьютерный центр СГАУ (СКЦ)** включает в себя следующие ресурсы:

- Кластер суперкомпьютера «Сергей Королёв»
- Учебный кластер
- Системы хранения данных

#### 1) **Суперкомпьютер IBM 1350 «Сергей Королёв»**

Построен на базе линейки оборудования IBM BladeCenter с использованием блейд-серверов HS22 и обеспечивает пиковую производительность более 10 триллионов операций с плавающей точкой в секунду.

Конфигурация системы:

- 112 BladeCenter HS22 вычислительных блейд-серверов:
  - Процессоры: 2x Intel Xeon X5560, 2.80GHz, 8MB Cache;

- Общее число процессоров / вычислительных ядер: 224/896;
- Общая оперативная память: 1,3125 Тб;
- Жесткий диск: 76Гб или 146Гб SAS 10K 16Mb Cache;
- Тип системной сети: QLogic/Voltaire InfiniBand DDR, QDR;
- Тип управляющей вспомогательной сети: Gigabit Ethernet.
- 1 IBM x3650M2 управляющий сервер:
  - Процессоры: 2x Intel Xeon X5560, 2.80GHz, 8MB Cache;
  - Оперативная память: 12 Гб.
- Сетевой интерконнект:
  - QLogic 9080 Infiniband DDR 20Гбит/с;
  - QLogic 12800 Infiniband QDR 40Гбит/с.
- Операционная система: Red Hat Enterprise Linux 5.5

Производительность кластера, замеренная на стандартном тесте Linpack составляет 8,542 Тфлопс. Сегодня это самый мощный суперкомпьютер в Самарской области.

## 2) Учебный вычислительный кластер

Кластер построен на базе класса параллельных вычислений межвузовского медиацентра. Он состоит из 10 рабочих станций Dell Precision 690, объединенных локальной сетью Ethernet 1Гбит/с.

Основные характеристики компьютеров:

- Процессор: Quad-Core Intel Xeon E5355 2.66ГГц, 2x4Мб L2 кэш
- Оперативная память: 8Гб DDR2 667МГц
- Видеокарта: ATI FireGL V7300 1Гб
- Жесткий диск: SAS, Seagate, 146Гб
- Сеть Ethernet: 1Гбит/с

Рабочие станции подключены к локальной сети медиацентра интерфейсами GigabitEthernet.

Кластер состоит из управляющего и девяти вычислительных компьютеров. Работают они под управлением операционной системы GNU/Linux, дистрибутив CentOS.

Управляется кластер с помощью системы xCAT.

## 3) Системы хранения данных (СХД) СКЦ СГАУ

Состоят из серверов, дисковых хранилищ и сетевой инфраструктуры для высокопроизводительных систем. Системы хранения СКЦ СГАУ предоставляют сервис общей домашней директории на всех вычислительных системах центра. Каждый пользователь имеет квоту в 10Гб данных и 100000 файлов в своей домашней директории. Для пользователей, которым требуется более 10Гб и/или 100000 файлов, возможно выделение дополнительного объема дискового пространства. Для получения дополнительных

ресурсов, необходимо обратиться в службу технической поддержки СКЦ..

Для передачи данных на вычислительные системы СКЦ, используйте программное обеспечение поддерживающее scp или sftp протоколы, используя любой управляющий сервер кластеров. Так домашняя директория объединена со всеми системами, можно использовать любую из них для передачи файлов. Локальные жесткие диски, установленные на вычислительных узлах, используются для хранения временных данных (/tmp). Некоторые серверы имеют больший объем временного хранилища данных, необходимый для специфических задач. Подробная информация находится на страницах использования конкретных вычислительных систем.

**Текущая конфигурация СХД** состоит из:

- СХД EMC CLARiiON ёмкостью 2Тб, расположенная в корпусе научно-технической библиотеки;
- дисковый массив кластера «Сергей Королев» на 1.2Тб;
- СХД HP на SCSI дисках, емкостью 2x410Гб.

Параллельная файловая система на данный момент не доступна. Для всех файлов в домашней директории ежедневно создаются резервные копии. Файлы в локальном временном каталоге (/tmp) не сохраняются.

## 8. Протоколы передачи данных

**Протокол передачи данных** — набор соглашений логического уровня интерфейса, который определяет обмен данными между различными программами. Эти соглашения задают единообразный способ передачи сообщений и обработки ошибок при взаимодействии программного обеспечения разнесённой в пространстве аппаратуры, соединённой тем или иным интерфейсом.

Стандартизированный протокол передачи данных также позволяет разрабатывать интерфейсы (уже на физическом уровне), не привязанные к конкретной аппаратной платформе и производителю (например, USB, Bluetooth).

**Сетевой протокол** — набор правил, позволяющий осуществлять соединение и обмен данными между двумя и более включёнными в сеть устройствами.

Наиболее распространённой системой классификации сетевых протоколов является так называемая модель OSI, в соответствии с которой протоколы делятся на 7 уровней по своему назначению — от физического (формирование и распознавание электрических или других

сигналов) до прикладного (интерфейс программирования приложений для передачи информации приложениями).

Сетевые протоколы предписывают правила работы компьютерам, которые подключены к сети. Они строятся по многоуровневому принципу. Протокол некоторого уровня определяет одно из технических правил связи. В настоящее время для сетевых протоколов используется модель OSI (Open System Interconnection — взаимодействие открытых систем, ВОС).

Модель OSI — это 7-уровневая логическая модель работы сети. Модель OSI реализуется группой протоколов и правил связи, организованных в несколько уровней:

- на физическом уровне определяются физические (механические, электрические, оптические) характеристики линий связи;
- на канальном уровне определяются правила использования физического уровня узлами сети;
- сетевой уровень отвечает за адресацию и доставку сообщений;
- транспортный уровень контролирует очередность прохождения компонентов сообщения;
- задача сеансового уровня — координация связи между двумя прикладными программами, работающими на разных рабочих станциях;
- уровень представления служит для преобразования данных из внутреннего формата компьютера в формат передачи;
- прикладной уровень является пограничным между прикладной программой и другими уровнями — обеспечивает удобный интерфейс связи сетевых программ пользователя.

Наиболее известные протоколы, используемые в сети Интернет:

- **HTTP** (Hyper Text Transfer Protocol) — это протокол передачи гипертекста. Протокол HTTP используется при пересылке Web-страниц с одного компьютера на другой.
- **FTP** (File Transfer Protocol) — это протокол передачи файлов со специального файлового сервера на компьютер пользователя. FTP дает возможность абоненту обмениваться двоичными и текстовыми файлами с любым компьютером сети. Установив связь с удаленным компьютером, пользователь может скопировать файл с удаленного компьютера на свой или скопировать файл со своего компьютера на удаленный.



- **POP** (Post Office Protocol) — это стандартный протокол почтового соединения. Серверы POP обрабатывают входящую почту, а протокол POP предназначен для обработки запросов на получение почты от клиентских почтовых программ.
- **SMTP** (Simple Mail Transfer Protocol) — протокол, который задает набор правил для передачи почты. Сервер SMTP возвращает либо подтверждение о приеме, либо сообщение об ошибке, либо запрашивает дополнительную информацию.
- **uucp** (Unix to Unix Copy Protocol) — это ныне устаревший, но все еще применяемый протокол передачи данных, в том числе для электронной почты. Этот протокол предполагает использование пакетного способа передачи информации, при котором сначала устанавливается соединение клиент-сервер и передается пакет данных, а затем автономно происходит его обработка, просмотр или подготовка писем.
- **telnet** — это протокол удаленного доступа. TELNET дает возможность абоненту работать на любой ЭВМ сети Интернет, как на своей собственной, то есть запускать программы, менять режим работы и так далее. На практике возможности лимитируются тем уровнем доступа, который задан администратором удаленной машины.
- **DTN** — протокол дальней космической связи, предназначенный для обеспечения сверхдальней космической связи.

**FTP** (англ. **File Transfer Protocol** — протокол передачи файлов) — протокол, предназначенный для передачи файлов в компьютерных сетях. FTP позволяет подключаться к серверам FTP, просматривать содержимое каталогов и загружать файлы с сервера или на сервер; кроме того, возможен режим передачи файлов между серверами (см. FXP).

FTP является одним из старейших прикладных протоколов, появившимся задолго до HTTP, в 1971 году. До начала 90-х годов на долю FTP приходилось около половины трафика в сети Интернет. Он и сегодня широко используется для распространения ПО и доступа к удалённым хостам.

Протокол FTP относится к протоколам прикладного уровня и для передачи данных использует транспортный протокол TCP. Команды и данные, в отличие от большинства других протоколов передаются по разным портам. Порт 20 используется для передачи данных, порт 21 для передачи команд. В случае, если передача файла была прервана по каким-либо причинам, протокол предусматривает средства для докачки файла, что бывает очень удобно при передаче больших файлов.

Протокол не шифруется, при аутентификации передаются логин и пароль открытым текстом. В случае построения сети с использованием хаба злоумышленник при помощи пассивного sniffера может перехватывать логины и пароли находящихся в том же сегменте сети пользователей FTP, или, при наличии специального ПО, получать передаваемые по FTP файлы без авторизации. Чтобы предотвратить перехват трафика, необходимо использовать протокол шифрования данных SSL, который поддерживается многими современными FTP-серверами и некоторыми FTP-клиентами.

**FTPS (File Transfer Protocol + SSL, или FTP/SSL)** — защищённый протокол для передачи файлов. При этом стандартный протокол FTP надстраивается над протоколом, передающим данные по зашифрованному каналу протокола SSL, что обеспечивает безопасную передачу данных. Защита SSL при такой схеме подвергается вся информация — как команды FTP, так и принимаемые или передаваемые данные.

Наиболее часто используются:

- *AUTH TLS, Explicit FTPS* или *FTPES* — более новая версия протокола (rfc 4217). Соединение происходит посредством обычного для FTP порта. Клиент запрашивает TLS посылая команду AUTH.
- *AUTH* см. (rfc 2228).
- *Implicit FTPS* — более старая версия протокола. Клиент соединяется с сервером используя SSL, поэтому для него надо открывать другой, отличный от стандартного порт. Обычно используется порт 990.

**SFTP** (англ. SSH File Transfer Protocol) — протокол прикладного уровня, предназначенный для копирования и выполнения других операций с файлами поверх надёжного и безопасного соединения. Как правило, в качестве базового протокола, обеспечивающего соединение, и используется протокол SSH2, но это не обязательно.

Сам по себе протокол SFTP не обеспечивает безопасность работы; это делает нижележащий протокол. Как правило, SFTP используется в сочетании с протоколом SSH2. Можно использовать SFTP и с другими протоколами, например SSH1, но это сопряжено с дополнительными трудностями.

При загрузке (upload) файлов наряду с данными на сервер могут копироваться и метаданные файла, в частности временные метки (timestamps), что при желании даёт возможность сохранить в загруженном на сервер файле время модификации неизменным.

В настоящий момент протокол SFTP не принят в качестве официального стандарта Интернет. Наиболее новая спецификация есть в черновике (draft), который описывает версию 6 протокола. Самой распространённой версией является 3-я, которая реализована в SSH-сервере OpenSSH. В большинстве Windows-клиентов реализована версия 4. SFTP, начиная с версии 4, становится более платформонезависимым и работает одинаково на Unix и Windows.

SFTP-сервер встроен в OpenSSH. Он реализуется с помощью программы sftp-server.

SFTP-клиент sftp встроен в пакет OpenSSH.

Разные протоколы, зачастую, описывают лишь разные стороны одного типа связи; взятые вместе, они образуют стек протоколов. Названия «протокол» и «стек протоколов» также указывают на программное обеспечение, которым реализуется протокол.

*Стек протоколов TCP/IP* — это два протокола нижнего уровня, являющиеся основой связи в сети Интернет. Протокол TCP (Transmission Control Protocol) разбивает передаваемую информацию на порции и нумерует все порции. С помощью протокола IP (Internet Protocol) все части передаются получателю. Далее с помощью протокола TCP проверяется, все ли части получены. При получении всех порций TCP располагает их в нужном порядке и собирает в единое целое.

## **9. Инструкция пользователю по работе с системами хранения данных**

Удаленный доступ на кластеры, возможен только с тех ip-адресов, которые были указаны в заявке на регистрацию пользователя.

Для входа на кластер необходимо использовать программу эмулятор терминала с поддержкой протокола SSH версии 2, например PuTTY.

- Настройка соединения в PuTTY

### **Удаленный доступ на кластеры по методу открытого и закрытого ключа**

Классический метод доступа к UNIX системам с вводом имени пользователя и пароля не всегда бывает удобен. Пользователи, чтобы запомнить пароль, делают его простым, что ведет к проблемам в безопасности. Существует метод доступа по методу закрытого и открытого ключа, не требующий ввода пароля.

- Настройка доступа по открытому ключу в PuTTY

## Передача файлов на кластер

Для передачи файлов на кластер используйте программы работающие по протоколу SFTP, например WinSCP.

- Настройке соединения в WinSCP

В операционной система Линукс можно использовать программу *scp*. В приведенном ниже примере копируется файл *file.dat* из домашней директории пользователя в домашнюю директорию на кластере "Сергей Королев"

```
[user@host ~]$ scp file.dat user@sk.ssau.ru:/
```

## Домашняя директория

Домашняя директория пользователя физически находится на системе хранения данных подключенная, с помощью сетевой файловой системы NFS, ко всем кластерам и вычислительным нодам СКЦ СГАУ. Скопированные данные становятся доступны на всех вычислительных системах СКЦ. По-умолчанию квота на дисковое пространство пользователя равна 10Гб. Ее можно увеличить при необходимости.

## Dot файлы

В домашней директории пользователей находится несколько специальных шаблонных файлов. Эти файлы часто называются *dot* файлы, потому что имя файла начинается с точки. По умолчанию эти файлы скрыты, их можно увидеть, набрав команду *ls -a*.

*.forward* - файл содержащий Ваш почтовый адрес. На этот адрес будут перенаправляться сообщения посылаемые Вам системой или СПО.

*.bash\_profile* - в конце этом файле можно определять свои переменные окружения и прочие настройки рабочей среды.