Санкт-Петербургский государственный университет

**Кафедра моделирования экономических систем**

**Прохоренко Филипп Александрович**

**Выпускная квалификационная работа бакалавра**

**Реализация информационной поддержки жизненного цикла товара на базе технологии блокчейн**

Направление 01.03.02

Прикладная математика и информатика

Научный руководитель,

доктор физ.-мат. наук,

профессор,

Прасолов А.В.

Санкт-Петербург

2018Содержание

Введение..................................................................................................................3

Постановка задачи..................................................................................................4

Обзор литературы................................................................................................5-7

Глава 1. Общие сведения о системах хранения данных.................................8-24

* 1. Система хранения данных............................................................8-11
	2. Типы СХД.....................................................................................12-18
	3. Типы носителей информации и протоколы взаимодействия с СХД...............................................................................................19-26

Глава 2. Блокчейн.............................................................................................26-27

 2.1 Что такое блокчейн......................................................................26-27

 2.2 Принцип работы блокчейна........................................................27-28

Глава 3. Результаты..........................................................................................29-32

 3.1 Классы баз данных......................................................................29-31

 3.2 Периодичность копирования......................................................31-32

Выводы...................................................................................................................33

Заключение.............................................................................................................34

Список литературы................................................................................................35 Введение

В нашем мире очень важна структуризация данных, позволяющая находить ту или иную информацию за достаточно короткое время. Под данными можно понимать практически любой объект повседневной жизни. Ведь если данные располагаются в хаотичном порядке, доступ к ним займет очень долгое время и не факт, что они вообще будут найдены. Таких примеров в жизни множество, начиная задачами уровня правительства страны и заканчивая нахождением нужного предмета на своем рабочем столе. Помимо сокращения времени поиска структуризация позволяет минимизировать ошибки при нахождении данных, так как правильная организация позволяет не упустить важные детали.

Одним из вариантов структуризации данных является вариант их добавления в общую базу — базу данных. Данная работа посвящена разбору систем хранения данных, прорывной технологии блокчейн и попыткам объяснить, как правильно выбрать периодичность создания реплик для базы данных, для того, чтобы защитить базы данных от различных атак. Будут рассмотрены разные методы сохранения, и для каждого будет произведена попытка анализа защиты, и нахождения оптимального сценария.

Постановка задачи

В настоящее время все люди обеспокоены проблемой сохранения информации. Одним из способов — введение хэша на каждую запись, исходные данные преобразуются в битовую строку, состоящую из случайных чисел и букв и имеющую установленную заранее длину, произвести обратное преобразование можно, используя специальный ключ, к которому имеют доступ только определенный круг лиц, назначаемый в установленном порядке. Получается, что доступ к исходным данным имеется только у ограниченного количества человек и отсутствует у тех, кто не имеет права взаимодействовать с этими данными. Алгоритмов хэширования огромное количество и каждый разработчик выбирает наиболее оптимальный для себя. Для сохранения базы данных ее нужно периодически перезаписывать, создавать копии базы данных, способы создания копий находятся в руках администраторов. Стоит задача управления сохранением информации, какими параметрами можно управлять, чтобы сохранить максимальное количество информаций. Необходимо понять, как часто нужно создавать реплики баз данных, чтобы не лишиться данных. Слишком частое сохранение приведет к необоснованным тратам, а слишком редкое может привести к утрате данных. Надлежит разобраться в том, каким образом найти оптимальную периодичность создания реплик баз данных, которая позволит сэкономить денежные средства и максимально защитить информацию.

Обзор литературы

Перейдем к обзору литературы, которая была использована для написания данной работы.

Статья Сатоши Накамото "Bitcoin: Peer-to-Peer Electronic Cash System"([1]), написанная в 2008 году рассказывает о становлении первой всемирно известной криптовалюты — биткоин и о блокчейне, как платформе, позволяющей биткоину жить. В этой статье особая роль уделена блокчейну, как технологии, которая может быть широко использована в будущем не только для поддержания криптовалюты, но и для всевозможных процессов, связанных с жизнью людей. Статья состоит из 11 глав, где каждая следующая плавно вытекает из предыдущих. Во вступлении говорится о проблеме доверия при совершении сделок, при его отсутствии могут произойти необратимые последствия. Предлагается использовать криптографическое доказательство вместо доверия, которое позволит людям избежать мошенничества. Далее, на примере биткоина разъясняется, как проводятся транзакции с использованием технологии блокчейн, идет речь о важности открытого и закрытого ключа, как с их помощью можно ее извлечь информацию, также раскрывается структура каждого блока и как происходит добавление нового блока. В 3 главе повествуется об отметках времени, которые показывают, что транзакция была проведена в конкретное время, дабы избежать проблемы двойного расходования средств. 4 глава является самой важной в этой статье. Она объясняет алгоритм, называемый "доказательство работы", который позволяет достичь консенсуса между участниками сети и признать блок, для которого этот алгоритм запускается, ликвидным. Следующая глава описывает алгоритм, необходимы для запуска сети, построенной на блокчейне. 6-7 главы позволяют понять, как сэкономить память, заменив хэши всех транзакций блока на один корневой хэш, используя Дерево Меркла, при помощи которого хэши попарно объединяются, пока не останется один хэш, которой называется корневым. Последние 4 главы посвящены проблеме верификации, а также приватности, в самом конце проводятся, показывающие с какой вероятностью сеть будет захвачена злоумышленниками, ведь если больше половины вычислительных мощностей будет в недобросовестных руках, они могут взять сеть под свой контроль и повести цепь блоков в том направлении, в котором они захотят. При помощи вычислений выяснилось, что вероятность продолжения цепи добросовестными участниками экспоненциально убывает с увеличением количества "захваченных" подряд блоков злоумышленников.

Подытожив выше описанное, можно сказать, что технология блокчейн, как технология, не основанная на доверии, а на математике имеет большие перспективы для внедрения во многие области. Конечно, она имеет существенные недостатки, но все ее преимущества их перекрывают, и при грамотном использовании она может принести огромную пользу.

В [2] статье показан краткий обзор технологии блокчейн, начиная с объяснения принципов работы данной технологии, и, заканчивая возможностями применения в разных сферах жизни. Авторы данных статей, основываясь на знаковой работе о блокчейне [1], попытались дать свою оценку возможностям и потенциалу блокчейна. Эти статьи очень полезны для ознакомления с технологией блокчейн и дальнейшем ее использованием.

 [3] и [4] статьи рассказывают о системах хранения данных, они описывают типы систем, рассказывают об их архитектуре, рассматриваются достоинства и недостатки каждого типа систем хранения данных, чтобы пользователь мог выбрать наиболее оптимальный для себя. Также приводятся примеры моделей каждого из типов. Далее, рассказывается о типах носителей информации, и о важности выбора протокола хранения данных. В конце статей рассмотрен краткий обзор имеющихся протоколов, и в табличной форме выделены их основные компоненты и особенности.

Статья [5] повествует основные сведения о репликации данных, разъясняет ее механизмы и показывает, какие пакеты можно использовать для создания реплик.

В книге [6] рассказывается о разделе теории вероятности, называемом теория массового обслуживания. Автор дает основные факты из теории вероятности и математической статистики, а также о теории марковских процессов. После ознакомления с основными фактами из вышеописанных разделов дается определения потока событий и системы массового обслуживания с ожиданием с примерами. В последней главе автор приводит некоторые модели массового обслуживания для более подробного ознакомления с предметной областью.

Глава 1. Общие сведения о системах хранении данных

 $§1.$ Система хранения данных

История возникновения баз данных начинается с 1960-ых годов, когда информация собиралась и хранилась в файлах. Это имело свои сложности, так как очень часто приводило к несогласованности данных и ,соответственно, к противоречивости самой информации. С 1970-ых годов начали появляться первые базы данных, и их развитие происходит до сих пор. В течение всех этих лет разработчики баз данных озабочены проблемой сохранения информации, каждый разработчик использует свой способ, чтобы не допустить потерю данных. Чтобы не потерять данные, необходимо правильно их хранить. Для этого появились системы хранения данных (далее СХД), являющиеся комплексным аппаратным решением, обеспечивающим надежное хранение информации и удобный доступ к ней. Остановимся на этих системах поподробнее. Почему же возникла необходимость в создании СХД? Аналитическая компания IDC Perspectives провела исследование, показывающее, что хранение данных занимает второе место среди расходов на ИТ и составляет примерно 23% от всех расходов. По общему мнению аналитиков, в организациях по всему миру ежеминутно вырастают объемы хранимой и обрабатываемой информации. Уникальная информация становится все дороже, ее объем каждый год увеличивается многократно, а ее хранение требует затрат. Ввиду этого организации стремятся не только формировать развитие инфраструктуры хранения данных, но и изыскивать возможности улучшения и повышения экономической эффективности СХД: снижения энергопотребления, расходов на сервис, общей стоимости владения и закупки систем резервного копирования и хранения. Рост объемов данных, возросшие требования к надежности хранения и быстродействию доступа к данным делают необходимым выделение средств хранения в отдельную подсистему вычислительного комплекса (ВК). Возможность доступа к данным и управления ими является необходимым условием для выполнения бизнес-процессов. Безвозвратная потеря данных подвергает бизнес серьезной опасности. Утраченные вычислительные ресурсы можно восстановить, а утраченные данные, при отсутствии грамотно спроектированной и внедренной системы резервирования уже не подлежат никакому восстановлению. Происходит заметное развитие потребности не только в приобретении СХД корпоративными клиентами, но и в строгом учете, аудите и мониторинге использования дорогостоящих ресурсов. И это касается не только бизнес-проектов, в любой сфере жизни невозможность восстановления данных приводит к необратимым последствиям. СХД появились не на пустом месте, этому способствовало огромное количество факторов. Основным фактором был рост конкуренции и усложнение ее характера во всех сегментах рынка. Для примера можно рассмотреть продажу SIM-карт для мобильных телефонов. Пять лет назад у мобильного оператора было 25-35 млн зарегистрированных SIM-карт, а сегодня — 50-70 млн. Таким образом, мобильной связью от этих компаний обеспечено в 2 раза больше людей, чем 5 лет назад, а ведь мобильных операторов достаточно много. Вот реальный уровень конкуренции: на рынке не осталось никого, кто не имел бы мобильного телефона. И теперь операторы не могут экстенсивно расти за счет продажи своих продуктов тем, у кого аналогичных продуктов еще нет. Им нужны клиенты, которые работают с конкурентами, и необходимо понять, как их получить ([4]). Чтобы провести анализ потребностей клиентов, извлечь полезную информацию из доступных данных, необходимо поместить их в хранилище. Следующий фактор — технологического характера. До некоторого времени производители приложений самостоятельно разрабатывали разные версии своих решений для разных серверных платформ или предлагали открытые решения. Важной для отрасли технологической тенденцией стало создание адаптируемых платформ для решения различных аналитических задач, которые включают аппаратную составляющую и систему управления базами данных (СУБД). Мир изменился, и теперь пользователей уже не волнует, кто сделал составляющие их компьютеров, будь то оперативную память или жесткий диск. Они стали рассматривать хранилище данных, как некую услугу.

Какие же требования должны быть соблюдены при создании СХД, ведь если какой-то аспект будет упущен, возрастет вероятность утечки данных или других неприятностей. Согласно статьи [4], группа компаний ТИМ решила провести опрос среди своих клиентов, чтобы понять, какие характеристики СХД клиенты считают наиболее значимыми. Самым важным была признана надежность, далее расположим в порядке убывания важности остальные характеристики: скорость, стоимость, объем, производитель и удобство пользования. И такое расположение выглядит вполне логичным, так как основная задача СХД состоит в сохранении данных, а без надежной защиты, ни о каком качественном сохранении не может быть и речи.

Рассмотрим подсистемы и компоненты, входящие в состав СХД:

* Устройства хранения
* Инфраструктура доступа к устройствам хранения
* Подсистема резервного копирования и архивирования данных
* Программное обеспечение управления хранением
* Система управления и мониторинга

 К устройствам хранения можно отнести дисковые массивы, ленточные библиотеки, стримеры, флэш-карты. Инфраструктура доступа к устройствам хранения необходима для того, чтобы связать сервера и устройства хранения. Системы управления и мониторинга позволят следить за системой круглосуточно, дабы избежать непредвиденных ситуаций и колоссальных убытков.

 На практике к СХД подключается не один сервер, а десятки и даже сотни, поэтому перечислим ряд ключевых требований, которые должны быть применимы к системам хранения данных и дадим их краткий обзор ([3]).

1. **Надежность и отказоустойчивость.** В СХД предусмотрено полное или частичное резервирование всех компонент – блоков питания, путей доступа, процессорных модулей, дисков, кэша и т.д. Обязательно наличие системы мониторинга и оповещения о возможных и существующих проблемах.
2. **Доступность данных.** Обеспечивается продуманными функциями сохранения целостности данных (использование технологии RAID, создание полных и мгновенных копий данных внутри дисковой стойки, реплицирование данных на удаленную СХД и т.д.) и возможностью добавления (обновления) аппаратуры и программного обеспечения в горячем режиме без остановки комплекса
3. **Средства управления и контроля.** Управление СХД осуществляется через web-интерфейс или командную строку, есть функции мониторинга и несколько вариантов оповещения администратора о неполадках. Доступны аппаратные технологии диагностики производительности.
4. **Производительность.** Определяется числом и типом накопителей, объёмом кэш-памяти, вычислительной мощностью процессорной подсистемы, числом и типом внутренних и внешних интерфейсов, а также возможностями гибкой настройки и конфигурирования.
5. **Масштабируемость.** В СХД обычно присутствует возможность наращивания числа жёстких дисков, объёма кэш-памяти, аппаратной модернизации и расширения функционала с помощью специального ПО. Все перечисленные операции производят без значительного переконфигурирования и потерь функциональности, что позволяет экономить и гибко подходить к проектированию ИТ-инфраструктуры ([3]).

§2. Типы СХД

Теперь перейдем к типам систем хранения данных. В случае отдельного ПК под системой хранения данных можно понимать внутренний жесткий диск или систему дисков (RAID массив). Если же речь заходит о системах хранения данных разного уровня предприятий, то традиционно можно выделить три технологии организации хранения данных ([3]):

* Direct Attached Storage (DAS)
* Network Attached Storage (NAS)
* Storage Area Network (SAN)
1. Устройства DAS (Direct Attached Storage) – решение, когда устройство для хранения данных подключено непосредственно к серверу, или к рабочей станции, как правило, через интерфейс по протоколу SAS ([5]). К основным преимуществам DAS систем можно отнести их низкую стоимость (в сравнении с другими решениями СХД), 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, хостинг-провайдеров и небольших корпоративных сетей. В то же время DAS-системы имеют и свои недостатки, к которым можно отнести не оптимальную утилизацию ресурсов, поскольку каждая DAS система требует подключения выделенного сервера и позволяет подключить максимум 2 сервера к дисковой полке в определенной конфигурации. В связи с этим, можно выделить достоинства и недостатки рассматриваемого типа СХД ([3]).

 Плюсы:

* Достаточно низкая стоимость. По сути эта СХД представляет собой дисковую корзину с жесткими дисками, вынесенную за пределы сервера.
* Простота развертывания и администрирования
* Высокая скорость обмена между дисковым массивом и сервером.

Минусы:

* Низкая надежность. При выходе из строя сервера, к которому подключено данное хранилище, данные перестают быть доступными
* Низкая степень консолидации ресурсов – вся ёмкость доступна одному или двум серверам, что снижает гибкость распределения данных между серверами. В результате необходимо закупать либо больше внутренних жестких дисков, либо ставить дополнительные дисковые полки для других серверных систем.
* Низкая утилизация ресурсов.

Архитектура DAS представлена на рисунке 1.



Рис. 1

  *Проблема данного типа СХД состоит в том, что ввиду особенности подключения устройств увеличивается вероятность потери данных в связи с поломкой сервера. Выбирая низкую стоимость, получаем больший риск потерять данные. В связи с этим, предлагается слегка увеличить стоимость эксплуатации, либо для увеличения количества серверов, либо для их более надежной работы, как одни из вариантов.*

2. Устройства NAS (Network Attached Storage) – отдельно стоящая интегрированная дисковая система, по-сути, NAS-сервер, со своей специализированной ОС и набором полезных функций быстрого запуска системы и обеспечения доступа к файлам ([3]). Система подключается к обычной компьютерной сети (ЛВС), и является быстрым решением проблемы нехватки свободного дискового пространства, доступного для пользователей данной сети. Технология NAS (сетевые подсистемы хранения данных, Network Attached Storage) развивается как альтернатива универсальным серверам, несущим множество функций (печати, приложений, факс сервер, электронная почта и т.п.). В отличие от них NAS-устройства исполняют только одну функцию — файловый сервер. И стараются сделать это как можно лучше, проще и быстрее. NAS подключаются к локально-вычислительным системам (ЛВС) и осуществляют доступ к данным для неограниченного количества гетерогенных клиентов (клиентов с различными ОС) или других серверов. 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. Внутри подобных серверов стоят специализированные ОС, например MS Windows Storage Server. Перейдем к достоинствам и недостаткам данного типа СХД ([3]).

Плюсы:

* Дешевизна и доступность его ресурсов не только для отдельных серверов, но и для любых компьютеров организации.
* Простота коллективного использования ресурсов.
* Простота развертывания и администрирования
* Универсальность для клиентов (один сервер может обслуживать клиентов MS, Novell, Mac, Unix)

 Минусы:

* Доступ к информации через протоколы “сетевых файловых систем” зачастую медленнее, чем как к локальному диску.
* Большинство недорогих NAS-серверов не позволяют обеспечить скоростной и гибкий метод доступа к данным на уровне блоков, присущих SAN системам, а не на уровне файлов.

 Архитектура NAS представлена на рисунке 2.



Рис. 2

 *Для данного типа СХД проблема заключается в скорости передачи данных, ведь никто не хочет ждать большое количество времени, чтобы получить доступ к данным. Несмотря на огромное количество плюсов, можно увеличить цену обслуживания для более гибкого доступа к данным, который позволит быстрее среагировать на всевозможные эксцессы, в том числе и потеря данных.*

3. Storage Area Network (SAN) –это специальная выделенная сеть, объединяющая устройства хранения данных с серверами приложений, обычно строится на основе протокола Fibre Channel или протокола iSCSI ([3]). Storage Area Network (SAN) — это специальная выделенная сеть, объединяющая устройства хранения данных с серверами приложений, обычно строится на основе протокола Fibre Channel, либо на набирающем обороты протоколу iSCSI. Постепенно отходят от протокола Fibre Channel, который несомненно имеет огромные возможности масштабирования, но является очень дорогим и требующим поддержки опытных специалистов, и из-за этого не всегда оптимально отвечает запросам потребителей. В отличие от NAS, SAN не имеет понятия о файлах: файловые операции выполняются на подключенных к SAN серверах. SAN оперирует блоками, как некий большой жесткий диск. Идеальный результат работы SAN — возможность доступа любого сервера под любой операционной системой к любой части дисковой емкости, находящейся в SAN. Оконечные элементы SAN — это серверы приложений и системы хранения данных (дисковые массивы, ленточные библиотеки и т. п.). А между ними, как и в обычной сети, находятся адаптеры, коммутаторы, мосты, концентраторы. ISCSI является более «дружелюбным» протоколом, поскольку он основан на использовании стандартной инфраструктуры Ethernet – сетевых карт, коммутаторов, кабелей. Более того, именно системы хранения данных на базе iSCSI являются наиболее популярными для виртуализированных серверов, в силу простоты настройки протокола. Обратим внимание на достоинства и недостатки данного типа([3]).

Плюсы:

* Высокая надёжность доступа к данным, находящимся на внешних системах хранения. Независимость топологии SAN от используемых СХД и серверов.
* Централизованное хранение данных (надёжность, безопасность).
* Удобное централизованное управление коммутацией и данными.
* Перенос интенсивного трафика ввода-вывода в отдельную сеть, разгружая LAN.
* Высокое быстродействие и низкая латентность.
* Масштабируемость и гибкость логической структуры SAN
* Возможность организации резервных, удаленных СХД и удаленной системы бэкапа и восстановления данных.
* Возможность строить отказоустойчивые кластерные решения без дополнительных затрат на базе имеющейся SAN.

 Минусы:

* Более высокая стоимость
* Сложность в настройке FC-систем
* Необходимость сертификации специалистов по FC-сетям (iSCSI является более простым протоколом)
* Более жесткие требования к совместимости и валидации компонентов.
* Появление в силу дороговизны DAS-«островов» в сетях на базе FC-протокола, когда на предприятиях появляются одиночные серверы с внутренним дисковым пространством, NAS-серверы или DAS-системы в силу нехватки бюджета.

В настоящий момент имеется достаточно большой выбор дисковых массивов для построения SAN, начиная от моделей для малых и средних предприятий, такие как серия DELL AX, которые позволяют создавать хранилища емкостью до 60 Тбайт, и, заканчивая дисковыми массивами для больших корпораций DELL/EMC серии CX4, они позволяют создать хранилища емкостью до 950 Тб. Есть недорогое решение на основе iSCSI, это PowerVault MD3000i – решение позволяет подключать до 16-32 серверов, в одно устройство можно установить до 15 дисков, и расширить систему двумя полками MD1000, создав массив на 45 терабайт.

Отдельного упоминания заслуживает система Dell EqualLogic на базе протокола iSCSI. Она позиционируется как СХД масштаба предприятия и сравнима по цене с системами Dell | EMC CX4, с модульной архитектурой портов, поддерживающих как FC протокол, так и iSCSI протокол. Система EqualLogic является одноранговой, т.е каждая дисковая полка имеет активные контроллеры RAID. При подключении этих массивов в единую систему, производительность дискового пула плавно растет с ростом доступного объема хранения данных. Система позволяет создать массивы более 500TB, настраивается менее, чем за час, и не требует специализированных знаний администраторов.

Архитектура SAN представлена на рисунке 3.



Рис. 3

## *SAN является самым надежным и безопасным из всех ранее описанных типов СХД, так как он предполагает централизованность, также этот вариант является намного более быстрым, чем предыдущие, но это влечет за собой более высокую стоимость и повышенную сложность в настройке. При снижении стоимости может понизиться надежность и безопасность, но если этот вариант будет финансово невыгоден, то даже надежность не сможет взять вверх, потому что это будет экономически невыгодно.*

 *Для всех трех типов СХД необходимо найти оптимальный вариант, обеспечивающий баланс между стоимостью, надежностью, сложностью построения, безопасностью и скоростью доступа к данным.*

## §3.Типы носителей информации и протокол взаимодействия с СХД

Перейдем к выбору протокола взаимодействия с СХД и выбору накопителей, которые будут использоваться в системе хранения. В настоящий момент для хранения данных в дисковых массивах используются SATA и SAS диски([4]). Какие диски выбрать в хранилище зависит от конкретных задач. Стоит отметить несколько фактов ([4]).

SATA II диски:

* Доступны объемы одного диска до 1 ТБ
* Скорость вращения 5400-7200 RPM
* Скорость ввода/вывода до 2,4 Гбит/с
* Время наработки на отказ примерно в два раза меньше чем у SAS дисков.
* Менее надежные, чем SAS диски.
* Дешевле примерно в 1,5 раза, чем SAS-диски.

 SAS диски:

* Доступны объемы одного диска до 450 ГБ
* Скорость вращения 7200 (NearLine), 10000 и 15000 RPM
* Скорость ввода/вывода до 3,0 Гбит/с
* Время наработки на отказ в два раза больше чем у SATA II дисков.
* Более надежные диски.

Важно! В прошлом году начался промышленный выпуск SAS дисков с пониженной скоростью вращения – 7200 rpm (Near-line SAS Drive) ([4]). Это позволило повысить объем хранимых данных на одном диске до 1 ТБ и снизить энергопотребление дисков со скоростным интерфейсом. При том, что стоимость таких дисков сравнима со стоимостью дисков SATA II, а надежность и скорость ввода/вывода осталась на уровне SAS дисков. Таким образом, в настоящий момент стоит действительно серьезно задуматься над протоколами хранения данных, которые вы собираетесь использовать в рамках корпоративной СХД. До недавнего времени основными протоколами взаимодействия с СХД являлись – FibreChannel и SCSI. Сейчас на смену SCSI, расширив его функционал, пришли протоколы iSCSI и SAS. Давайте ниже рассмотрим плюсы и минусы каждого из протоколов и соответствующих интерфейсов подключения к СХД. Некоторые особенности протокола Fibre Channel мы уже затронули, теперь остановимся на них более подробно.

##

## **Протокол Fibre Channel**

На практике современный Fibre Channel (FC) имеет скорости 2 Гбит/Сек (Fibre Channel 2 Gb), 4 Гбит/Сек (Fibre Channel 4 Gb) full- duplex или 8 Гбит/Сек, то есть такая скорость обеспечивается одновременно в обе стороны ([3]). При таких скоростях расстояния подключения практически не ограничены – от стандартных 300 метров на самом «обычном» оборудовании до нескольких сотен или даже тысяч километров при использовании специализированного оборудования. Главный плюс протокола FC – возможность объединения многих устройств хранения и хостов (серверов) в единую сеть хранения данных (SAN). При этом не проблема распределенности устройств на больших расстояниях, возможность агрегирования каналов, возможность резервирования путей доступа, «горячего подключения» оборудования, большая помехозащищенность ([3]). Но с другой стороны мы имеем высокую стоимость, и высокую трудоемкость инсталляции и обслуживания дисковых массивов использующих FC ([3]).

Важно! Следует разделять два термина протокол Fibre Channel и оптоволоконный интерфейс Fiber Channel. Протокол Fibre Сhannel может работать на разных интерфейсах — и на оптоволоконном соединении с разной модуляцией, и на медных соединениях ([3]).

Плюсы:

* Гибкая масштабируемость СХД
* Позволяет создавать СХД на значительных расстояниях (но меньших, чем в случае iSCSI протокола; где, в теории, вся глобальная IP сеть может выступать носителем.
* Большие возможности резервирования.

 Минусы:

* Высокая стоимость решения;
* Еще более высокая стоимость при организации FC-сети на сотни или тысячи километров
* Высокая трудоемкость при внедрении и обслуживании

*Помимо появления протокола FC 8Гб/c, ожидается появление протокола FCoE (Fibre Channel over Ethernet), который позволит использовать стандартные IP сети для организации обмена пакетами FC. Так что, вероятно в будущем этот протокол может долго еще быть актуальным. Но высокая стоимость идет не на пользу, поэтому преимущество больших расстояний нивелируется высокой стоимостью. Поэтому, возможно, следует уменьшить зону действия, но в то же время сократить расходы и уменьшить сложность.*

**Протокол iSCSI**

Протокол iSCSI (инкапсуляция SCSI пакетов в протокол IP) позволяет пользователям создать сети хранения данных на базе протокола IP с использованием Ethernet-инфраструктуры и портов RJ45. ([3]) Таким образом, протокол iSCSI дает возможность обойти те ограничения, которыми характеризуются хранилища данных с непосредственным подключением, включая невозможность совместного использования ресурсов через серверы и невозможность расширения емкости без отключения приложений. Скорость передачи на данный момент ограничена 1 Гб/c (Gigabit Ethernet), но данная скорость является достаточной для большинства бизнес-приложений масштаба средних предприятий и это подтверждают многочисленные тесты. Интересно то, что важна не столько скорость передачи данных на одном канале, сколько алгоритмы работы RAID контроллеров и возможность агрегации массивов в единый пул, как в случае с DELL EqualLogic, когда используются по три 1Гб порта на каждом массиве, и идет балансировка нагрузки среди массивов одной группы([3]).

Плюсы:

* Высокая доступность;
* Масштабируемость
* Простота администрирования, так как используется технология Ethernet;
* Более низкая цена организации SAN на протоколе iSCSI, чем на FC.
* Простота интеграции в среды виртуализации

 Минусы:

* Есть определенные ограничения по использованию СХД с протоколом iSCSI с некоторыми OLAP и OLTP приложениями, с системами Real Time и при работе с большим числом видеопотоков в HD формате
* Высокоуровневые СХД на базе iSCSI, также как и СХД с FC-протоколом, требуют использования быстрых, дорогостоящих Ethernet-коммутаторов
* Рекомендуется использование либо выделенных Ethernet коммутаторов, либо организация VLAN для разделения потоков данных. Дизайн сети является не менее важной частью проекта, чем при разработке FC-сетей ([3]).

*Более низкая стоимость делает данный протокол более привлекательным для клиентов, но вводится ряд ограничений, так как данный протокол не может работать со всеми приложениями. Выбирая более низкую стоимость, теряется способность в вариативности, которую можно исправить возможностью работать со всеми приложениями, что требует больших средств, но наличие технологии Ethernet позволяет говорить о простоте администрирования, что не имеет место для Fibre Channel. Но Fibre Channel покрывает большую дистанцию, поэтому тут не все так однозначно.*

##

## **Протокол SAS**

Протокол SAS и одноименный интерфейс разработаны для замены параллельного SCSI и позволяет достичь более высокой пропускной способности, чем SCSI. Хотя SAS использует последовательный интерфейс в отличие от параллельного интерфейса, используемого традиционным SCSI, для управления SAS-устройствами по-прежнему используются команды SCSI. SAS позволяет обеспечить физическое подключение между массивом данных и несколькими серверами на небольшие расстояния ([3]).

Плюсы:

* Приемлемая цена;
* Легкость консолидации хранилищ – хотя СХД на базе SAS не может подключаться к такому количеству хостов (серверов), как SAN конфигурации которые используют протоколы FC или iSCSI, но при использовании протокола SAS не возникает трудностей с дополнительным оборудованием для организации общего хранилища для нескольких серверов.
* Протокол SAS позволяет обеспечить большую пропускную способность с помощью 4 канальных соединений внутри одного интерфейса. Каждый канал обеспечивает 3 Гб/c , что позволяет достичь скорости передачи данных 12 Гб/с (в настоящий момент это наивысшая скорость передачи данных для СХД).

 Минусы:

* Ограниченность досягаемости – длина кабеля не может превышать 8 метров. Тем самым хранилища с подключением по протоколу SAS, будут оптимальны только тогда когда серверы и массивы будут расположены в одной стойке или в одной серверной;
* Количество подключаемых хостов (серверов) как правило, ограничено несколькими узлами.

Важно! В 2009 году ожидается появление технологии SAS со скоростью передачи данных по одному каналу – 6 Гбит/c, что позволит значительно увеличить привлекательность использования данного протокола ([3]).

*Рассматриваемый протокол имеет большое количество плюсов в виде дешевизны и высокой пропускной способности, но у него есть большой недостаток, как и у предыдущего, зона использования очень мала. Досягаемость зависит от стоимости. Оптимального варианта среди этих нет, но при развитии каждого протокола, есть все шансы стать более эффективными при грамотно найденном балансе между зоной работы, стоимостью, сложностью администрирования, безопасностью, вариативностью работы с приложениями.*

 **Сравнение протоколов подключения СХД**

|  |  |
| --- | --- |
| Параметр | Протоколы подключения СХД |
| iSCSI | SAS | FC |
| Архитектура | SCSI команды инкапсулируются в IP пакет и передаются через Ethernet, последовательная передача | Последовательная передача SCSI команд | Коммутируемая |
| Расстояние между дисковым массивом и узлом(сервер или свитч) | Ограничено лишь расстоянием IP сетей | Не более 8 метров между устройствами | 50000 метров без использование специальных репиторов |
| Масштабируемость | Миллионы устройств по протоколу IPv6 | 32 устройства | 256 устройств( 16 миллионов при использовании FC-SW архитектуры) |
| Производительность | 1Гб/с( в планах до 10Гб/с) | 3Гб/с при использовании 4-х портов, до 12Гб/с | До 8Гб/с |
| Уровень вложений(затрат на внедрение) | Незначительный( так как используется Ethernet) | Средний | Значительный |

*Таким образом, представленные решения на первый взгляд достаточно четко разделяются по соответствию требованиям заказчиков. Однако на практике все не так однозначно, включаются дополнительные факторы в виде ограничений по бюджетам, динамики развития организации (и динамики увеличения объема хранимой информации), отраслевая специфика и т.д.*

Глава 2. Блокчейн, как средство хранения данных

§1. Что такое блокчейн.

 Вначале разберемся, что же такое блокчейн. Блокчейн — это цепочка блоков, содержащих информацию, выстроенная по определенному правилу. Блокчейн является распределенной базой данных и децентрализованной, так как устройства хранения данных не подключены к общему серверу ([2]). Каждый блок, помимо транзакций, содержит метку времени (момент времени, когда он был добавлен в блок) и ссылку на предыдущий блок. Применение шифрования гарантирует, что пользователи могут изменять только те части цепочки блоков, которыми они «владеют» в том смысле, что у них есть закрытые ключи, без которых запись в файл невозможна. Кроме того, шифрование гарантирует синхронизацию копий распределенной цепочки блоков у всех пользователей ([2]).

 В технологию блокчейн сначала была заложена безопасность на уровне баз данных, о которых шла речь в первой главе. Концепцию цепочек блоков придумал Сатоши Накомото и подробно расписал в своей статье [1], описание которой разобрано в обзоре литературы. Первое широкое применение блокчейна связано с цифровой криптовалютой биткойн, где он играл роль главного общего реестра для всех операций с биткоинами. Благодаря технологии блокчейна биткоин стал первой цифровой валютой, которая решает проблему двойных расходов (в отличие от физических монет или жетонов, электронные файлы могут дублироваться и тратиться дважды) без использования какого-либо авторитетного органа или центрального сервера ([2]). Безопасность в технологии блокчейн обеспечивается через децентрализованный сервер, который проставляет метки времени, и одноранговые сетевые соединения. В результате формируется база данных, которая управляется автономно, без единого центра. Это делает цепочки блоков очень удобными для регистрации событий (например, внесения медицинских записей) и операций с данными, управления идентификацией и подтверждения подлинности источника ([2]).

§2. Принцип работы блокчейна

Блокчейн позволяет отправлять любую информацию в каждую точку мира пользователям, которые могут получить к ней доступ. Необходимо иметь закрытый ключ для того, чтобы иметь доступ к тем блокам, которыми участник сети "владеет". Закрытый ключ создается по специальному криптографическому алгоритму, где алгоритм выбирается из нужд и удобств тех, кто создает данный блокчейн ([2]). Предоставляя кому-либо ваш закрытый ключ, вы по сути передаете этому лицу денежную сумму, которая хранится в соответствующем разделе цепочки блоков. Если рассматривать биткоин, то такие закрытые ключи используются для доступа к адресам, по которым хранятся денежные суммы, в данном случае, в виде биткоинов ([2]). Таким образом реализуется функция перевода средств. Обычно, эту роль выполняют банки. Еще одна важная функция, которая устанавливается — это функция доверия между пользователями и подтверждения подлинности личности, потому как невозможно изменить цепочку блоков, не имея закрытых ключей. Так как изменения, не подтвержденные ключами отклоняются. Есть вероятность, что ключи могут быть украдены, но для того, чтобы защитить несколько строк компьютерного кода, все же необходимо намного меньше затрат, чем, например, хранение денег в банке или золотых запасов. Можно сделать вывод, что основные банковские функции, такие как идентификация личности и регистрация проведенных сделок, могут быть выполнены при помощи блокчейна гораздо быстрее и точнее.

 Блокчейн имеет встроенную устойчивость к ошибкам по причине того, что он не контролируются одним местом и не имеет единой точки отказа. Блокчейн уже на данный момент представляет собой механизм, который обеспечивает высокую степень учета и идентификации, и он продолжает развиваться ([2]). Наиболее важно то, что блокчейн помогает гарантировать законность транзакции не только в главном реестре, а в распределенной системе реестров, связанных через защищенный механизм проверки. Блокчейн может быть применим во многих сферах, таких как всевозможные финансовые операции, сделки с недвижимостью, где он будет выполнять схожую функцию, как и с биткоином, страхование, логистика, голосование (недавно были проведены первые успешные выборы с использованием данной технологии), здравоохранение, где блокчейн играет роль распределенной базы данных. Более подробную информацию о структуре блоков и о работе блокчейна можно найти в [1].

 *В Главе 2 было сказано, что блокчейн является прорывной технологией. В Главе 1 при рассмотрении типов СХД мы нашли параметры, которые можно изменять, например, стоимость обслуживания, надежность, безопасность, сложность построения. Блокчейн решает проблему надежности и безопасности благодаря своей децентрализованности. Стоимость обслуживания также будет ниже. Поэтому блокчейн может поспорить с рассматриваемыми выше типами СХД в более оптимальной защите данных, предлагая свой инновационный подход*

Глава 3. Результаты

 §1. Классы баз данных

Базы данных можно разделить на 3 больших класса. К первому классу отнесем, так называемые "тривиальные" базы данных, информация в которых не представляет большой ценности и риски убытков при утрате целой базы или ее части очень мал. К таким базам данных можно отнести базы, созданные для тестирования работы некоторого приложения, реализованные стажерами в процессе освоения нового программного компонента или базы, которые имеют аналог в сети. В связи с тем, что для таких баз сохранность информации не является приоритетным направлением, углубляться в этот класс мы не будем.

 Ко второму классу отнесем такие базы данных, для которых необходимо создавать копии, базы данных этого класса имеют большую ценность, чем представители предыдущего класса, поэтому для их сохранности нужны специальные методы, такие как копирование. Процесс копирования баз данных называется их репликацией. Данные копируются из одного источника на множество других, в копии можно добавлять новую информацию, из них можно удалять часть данных, можно оставить копии без изменений. Обновление данных в оригинале или в копии приводит к тому, что произойдет обновление данных и во всех остальных копиях, либо сразу же в текущей транзакции (существует всего одна версия данных), в случае синхронной репликации, либо в течение некоторого промежутка времени, в случае асинхронной репликации. Самое важное то, что при утрате оригинальных данных, появляется возможность восстановить целиком всю базу данных по ее копии. Репликация данных имеет свои недостатки. При синхронной репликации наблюдается дополнительная нагрузка на систему, в связи с тем, что обновление данных происходит на всех копиях, а не на одной. При асинхронной репликации существует проблема несогласованности данных, так как в определенный момент времени на разных копиях могут быть совершенно разные данные. Если учесть, что транзакции передаются на все копии за различные промежутки времени, может получится так, что одна реплика содержит определенную строку данных, а другая нет. Примерами баз данных для которых проводится репликация достаточно много, статистика футбольных матчей, включающая в себя количество ударов, фолов, желтых карточек и т.д, ее потеря не фатальна, но для присуждения некоторых индивидуальных и командных наград необходимо иметь данные сведения, и в текущей ситуации реплика поможет их получить, сведения о загрязнении воздуха, количество машин, выпущенных автомобильной компанией и т.п.

 К третьему классу баз данных относятся базы данных, которые нуждаются в наивысшей степени защиты. В таких базах находятся сведения, которые ни при каких обстоятельствах не должны попасть в чужие руки. К их числу можно отнести правительственные базы, содержащие государственную тайну, медицинские карточки о болезнях пациентов, которые в случае кражи можно использовать в корыстных целях, сведения о финансовых операциях, о заключении всевозможных сделок, потеря таких данных может привести к катастрофическим убыткам. Таким базам данных нужна децентрализация и тут им на помощь приходит блокчейн. Блокчейн является инструментом, благодаря которому можно хранить данные, он является одной из разновидностей баз данных. Основным преимуществом блокчейна является то, что он предусматривает полную прозрачность транзакций и их множественное копирование. Копирование происходит таким образом, что каждый участник транзакции имеет копию всех сделок и может отслеживать любые действия своих коллег. Доступ к данным у каждого участника совершенно разный, кто-то может видеть лишь сам факт совершения сделки, а другой может напрямую пользоваться средствами, которые эта сделка приносит. Рассмотрим на примере, пусть у нас есть 2 пользователя A и B, пользователь А хочет перевести 50 денежных единиц пользователю B. Сведения о том, что эта операции были проведены имеют все участники сети, а в случае не анонимности пользователей, могут видеть участников сделки, но доступ к средствам имеет только пользователь B, остальные играют роль наблюдателей. Блокчейн не имеет какого-либо командного центра, который можно взломать, он абсолютно децентрализован. Взлом любого компьютера в системе никак не скажется на сохранности данных, потому что на всех других компьютерах они останутся без изменений. Чтобы как-то изменить данные без согласия участников сети, необходимо взломать все компьютеры в сети, что практически не представляется возможным.

§2. Периодичность копирования

Для нахождения оптимальной периодичности копирования данных возможны численные методы решения. Также возможны математические методы решения задачи, тут можно воспользоваться теорией массового обслуживания.

**Утверждение 1.** Пусть $p$ - вероятность потери данных, $C$- стоимость убытков, $x$- длина рассматриваемого промежутка времени, $n$- число "атак" (число всевозможных атак на систему, из-за которых существует вероятность потери данных), тогда число "атак" распределено по закону Пуассона с параметром $λ$ и $P=λxe^{-λx}$, а $p=λe^{-λ}$, где λ – интенсивность атак.

Действительно, воспользуемся терминологией теории массового обслуживания. Если рассматривать под потоком событий последовательность попыток исчезновения информации, то поток будем простейшим, то есть однородным и стационарным. Поток стационарный, потому что вероятность потери зависит только от длины интервала, чем длительнее срок, тем выше вероятность того, что данные могут быть утеряны. Поток является однородным, по причине того, что каждая потеря данных равноправна, и на начальном этапе важен сам факт возможной потери без уточнения каких-либо деталей. Для потери данных достаточно одной успешной атаки на базу на рассматриваемом интервале длины $x$. Так как было объяснено, почему поток является простейшим, то распределения и вероятность будут равны $P=λxe^{-λx}$и$p=λe^{-λ}$ соответственно.

**Утверждение 2.**

Возможны разные сценарии, из-за которых будут потеряны данные, может быть совершена кража данных хакерами или недобросовестными конкурентами, возможны технические проблемы, такие как поломка сервера, которая повлечет невозможность восстановления информации.

**Утверждение 3.**

Для нахождения оптимального типа сохранения данных нужно варьировать такие параметры, как стоимостью, надежностью, сложностью построения, безопасностью и скоростью доступа к данным.

Выводы

В настоящее время проблема репликации не стоит так остро над людьми, в большинстве случаев данные удается сохранить без реальных потерь. Но технические возможности растут, и в будущем существует вероятность, что репликация так хорошо работать не будет, и необходимо будет точно определять время для копирования информации, чтобы оставить ее в сохранности. На данный момент общего правила копирования нет, некоторые компании копируют на 3-4 диверсифицированных источника, создают журналы операций, чтобы по истечении срока можно было по заявке клиента внести изменения после разбора последовательность действий, используя журнал, такое имеет место в банковской сфере. Каждая компания копируют в те промежутки времени, в которые считает нужным, но, возможно, с течением времени удастся придумать единый рабочий алгоритм для оптимальной периодичности копирования баз данных.Заключение

Проблема сохранения информации будет всегда актуальна. Управление сохранением информации является очень важной составляющей, потому что при правильном выборе факторов, можно минимизировать все риски и не допустить ни денежных потерь, ни потерь данных. Данная сфера имеет много возможностей для развития, чтобы в ближайшее время сохранность данных вышла на новый уровень, благодаря чему и поставщики всевозможной продукции, и потребители были полностью удовлетворены. В настоящий момент нет математического решения данной проблемы, но оптимизировав параметры, о которых речь шла в Выводе и Главе 1 можно из имеющихся типов или из нового созданного придумать оптимальный тип хранения данных, который бы обеспечивал максимальную защиту информации.Список литературы

1. Satoshi Nakamoto, Bitcoin: Peer-to-peer Electronic Cash System, 2008. 9 с.
2. Что такое Blockchain (блокчейн)? Технология, платформа, транзакции. <https://mining-cryptocurrency.ru/blockchain/#i>
3. Основные системы хранения данных и их особенности. http://www.itworkroom.com/main-shd/
4. Система хранения данных. [http://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F:%D0%A1%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0\_%D1%85%D1%80%D0%B0%D0%BD%D0%B5%D0%BD%D0%B8%D1%8F\_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85](http://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F%3A%D0%A1%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0_%D1%85%D1%80%D0%B0%D0%BD%D0%B5%D0%BD%D0%B8%D1%8F_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85)
5. Что такое репликация БД и что о ней нужно знать пользователю. https://www.kv.by/archive/index2008021107.htm
6. В.А. Павский, Теория массового обслуживания. Кемеровский технологический институт пищевой промышленности. - Кемерово, 2008. 116 с.