

Программное и техническое обеспечение хранения и сохранения информации

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ХРАНЕНИЯ ДАННЫХ, ИНФОРМАЦИИ	4
2 СИСТЕМА ХРАНЕНИЯ ДАННЫХ (СХД)	6
3 ФАКТОРЫ, СПОСОБСТВУЮЩИЕ РАЗВИТИЮ СХД	10
ЗАКЛЮЧЕНИЕ	12
СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ	13

ВВЕДЕНИЕ

Информация играет решающую роль в любой сфере деятельности. Современный человек имеет доступ к бесчисленному множеству сведений, понятий и фактов. В этой ситуации возникает вопрос: Где сохранить нужные материалы, имея при этом удобный доступ к ним? Для этой цели существуют различные виды носителей.

Хранение информации осуществляется записью на носителях и накопителях информации. Носителем данных может являться любой материальный предмет, используемый человеком для записи, хранения, чтения и передачи информации (например, книги, диски, фотографии, flash-карты, облачные хранилища и так далее). Накопителями считаются приспособления позволяющие хранить и дополнять информацию.

Сегодня существует большое количество способов хранения информации, имеющих свои плюсы и минусы. Технологии, связанные с информацией, в наше время стремительно развиваются и вместе с ними прогрессируют средства её хранения. Каждый человек может выбрать удобный для себя способ.

Ни для кого не секрет, что в настоящее время из-за обилия различного рода цифровой информации, начинают возникать проблемы с ее упорядочиванием и хранением. Самой ценной для всех, без исключения, является личная или семейная цифровая информация: фотографии, видеозаписи, какие либо документы, цифровые коллекции музыки, редких фильмов и т.п.

Никакие современные методы хранения различного рода цифровых данных не дают нам 100% гарантии ее сохранности.

Из этого следует, что необходимо создавать резервные копии той информации, потеряв которую, нельзя восполнить никогда.

1 ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ХРАНЕНИЯ ДАННЫХ, ИНФОРМАЦИИ

Хранение информации необходимо для того, чтобы: иметь в памяти ЭВМ системные и другие, необходимые пользователям программы и данные; осуществлять различные виды работ на компьютере; её можно было в любой момент предоставить пользователю. Различные виды информации, данных и знаний хранятся на разнообразных носителях электронных данных (жёстких, гибких магнитных и лазерных дисках, микросхемах и др.). Она может редактироваться, удаляться, копироваться на другие носители, пересылаться на другие компьютеры, архивироваться с разной степенью регулярности.

Хранение – это базовая основа обеспечения сохранности [1].

Если документ повреждён, разрушен и может быть утрачен, то говорить об обеспечении сохранности бессмысленно.

Сохранность — это состояние документа, программы или технических средств, характеризуемое степенью удержания их эксплуатационных свойств.

Обеспечение сохранности информации производится путём применения специальных мер организации хранения, восстановления (регенерации) информации, специальных устройств резервирования. Качество обеспечения сохранности информации зависит от её целостности (точности, полноты) и готовности к постоянному использованию. Для долговременного хранения информации важным является выбор соответствующего носителя [1].

С первой половины прошлого века надёжными носителями информации считались фотоматериалы, способные в специальных условиях долговременно её сохранять. При этом используется технология микрофильмирования. Микрофильмирование - это совокупность процессов изготовления, хранения и использования носителей микроизображений информации. Микроизображением считается изображение, которое можно прочесть только с помощью оптических средств с увеличением до 40 крат (40х). Микроформа – это или полноразмерная, или уменьшенная в 9–30 раз (масштаб 1:9–1:30) копия оригинала.

По виду изображения выделяют негативные или позитивные микроформы. В микрофильмировании используют микрофильмы рулонные, микрофильмы в отрезках, микрофиши, микрокарты и др. Хотя микроформы относятся к машиночитаемым носителям информации, непосредственно использовать их в компьютерных технологиях затруднительно [2].

В процессе эволюции компьютерных технических средств информация хранилась на машинных носителях: перфокартах, перфолентах, магнитных лентах, магнитных дисках и дискетах. Затем появляются компактные

оптические диски (CD, DVD и др.) и твердотельная флеш-память. Для осуществления операций записи и хранения на всех этих видах электронных носителей данных используются соответствующие устройства и технологии.

Для хранения больших объёмов электронной информации создаются специальные локальные и распределённые хранилища. Доступ к распределённым хранилищам может осуществляться из любого конца планеты. Одним из родоначальников теории хранилищ был Уильям Г. Инмон (William H. Inmon). В 1988 году он, определил хранилища данных, как: “предметно ориентированные, интегрированные, неизменяемые, поддерживающие хронологию наборы данных, организованные для целей поддержки управления”. Обычно данные в хранилище находятся от одного года до пяти лет. Если в информационном хранилище не требуется присутствие данных большей давности, то их, как правило, переносят в архив (например, на магнитные ленты или CD-ROM) [2].

Существуют и индивидуальные хранилища данных. Ранее к ним обычно относили персональные коллекции файлов на дискетах. Сейчас эти ненадёжные и малой ёмкости носители практически не применяются. Часто вместо них используют компакт-диски типа CD и DVD.

Многие данные требуется сохранять для последующего их использования. Для этого создают локальные, распределённые и удалённые базы данных, информационные хранилища (репозитории) или хранилища данных, содержащие большие объёмы, как правило, взаимосвязанных данных. Всё это делается для того, чтобы пользователи могли быстро находить необходимую им информацию, рассматривать её с различных точек зрения, анализировать и создавать новые знания. Характерной особенностью сетевого хранилища данных является то, что одновременно к нему с одним и тем же запросом могут обратиться несколько пользователей. В результате проведенного поиска им будут доставлены одинаковые сведения [3].

Для хранения и надёжного сохранения огромных массивов данных на одном сервере и организации доступа к ним используют RAID-массивы, “роботизированные библиотеки” (CD и DVD) и другие системы, а в информационных сетях – информационные хранилища. Такие хранилища, как правило, являются распределёнными базами данных или сетями хранения данных. Они формируются из множества различных внешних и внутренних источников. Информационные хранилища электронной информации – это специальные программно-технические комплексы, в т.ч. специальные сети хранения данных, получившие название Storage Area Network (SAN), а в корпоративных сетях – специализированные Network Attached Storage (NAS-серверы). Они осуществляют совместимость, интеграцию и администрирование серверов общего назначения, а также

хранение огромных массивов данных.

2 СИСТЕМА ХРАНЕНИЯ ДАННЫХ (СХД)

Система хранения данных (СХД) — это совокупность аппаратных средств и программного обеспечения, предназначенная для сохранения и обработки цифровой информации, а также ее резервного копирования. Данные в такой системе хранения могут быть структурированными (как БД) и неструктурированными (как big data), в их роли часто выступают файлы, архивы преимущественно большого объема, резервные копии и т. п.

Носителями информации в СХД являются жесткие диски — обычные HDD и/или твердотельные SSD-накопители.

Как правило, система хранения данных содержит следующие подсистемы и компоненты [4]:

- устройства хранения (дисковые массивы, ленточные библиотеки);
- инфраструктуру доступа к устройствам хранения;
- подсистему резервного копирования и архивирования данных;
- программное обеспечение управления хранением;
- систему управления и мониторинга.

Чаще всего СХД монтируют в стандартный 19-дюймовый шкаф.

Пример структурной схемы СХД приведен на рисунке 1:

1. Контроллер, включающий центральный процессор (или несколько), интерфейсы для коммутации с жёсткими дисками и внешними портами, кэш-память.
2. Внешний интерфейс, в данном случае Fibre Channel.
3. Жёсткие диски – ёмкость расширяется дополнительными полками.
4. Кэш-память обычно зеркалируется, чтобы не потерять данные при выходе из строя одного из модулей.

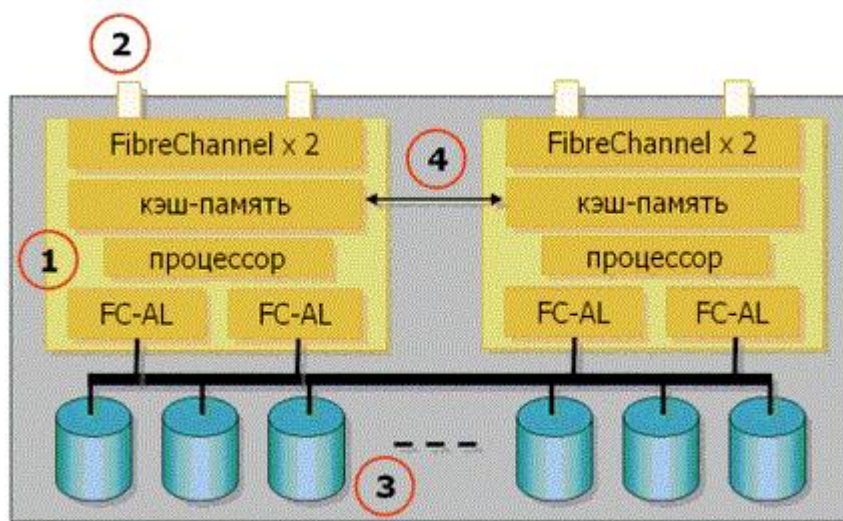


Рисунок 1 — Пример структурной схемы СХД

Имеющиеся в системе диски можно разбивать на группы и объединять в RAID различных уровней. Получившееся дисковое делится на логические блоки (LUN) – к ним получают доступ хосты и «видят» их как локальные жёсткие диски. Количество RAID-групп, LUN-ов, логика работы кэша, доступность LUN-ов конкретным серверам настраивается администратором.

Основным отличием такой системы от просто жестких дисков считается ее высокая производительность. Она имеет более сложную архитектуру, но благодаря специальному ПО работает намного быстрее.

СХД могут быть трех типов: файловыми, блочными и объектными. Такая классификация основана на уровнях хранения информации. Каждая из разновидностей также определяет, в каком виде хранятся данные и как к ним осуществляется доступ. В результате это помогает оценить, насколько просто управлять системой и насколько быстро она откликается на запрос [5]:

1. Файловые. В хранилищах этого типа данные имеют вид файлов, которые собраны в каталоги/папки. Все они организуются с помощью подробных метаданных, указывающих, где расположен определенный файл. Извлекается информация благодаря им же. Как правило, файловая СХД используется для хранения «холодных» данных, т. е. информации, не требующей операционных вычислений. Файлы зависят друг от друга, поскольку составляют иерархическую структуру. По этой причине при росте объема хранимой информации скорость работы системы заметно снижается: иерархия усложняется, путь к определенному файлу занимает все больше времени и ресурсов, отклик становится медленнее.

2. Блочные. В такой системе информация делится на отдельные самостоятельные блоки, каждый из которых обладает собственным идентификатором. Он позволяет СХД размещать данные в любой части дискового пространства и быстро находить их при необходимости. Блочное хранилище применяется как жесткий диск: его можно форматировать, устанавливать на него операционную систему, создавать виртуальные диски.

Использовать блочную систему сложнее в плане настройки и обслуживания, но она отлично подходит для выполнения высокопроизводительных вычислений, хранения больших баз данных, развертывания сред разработки/тестирования.

3. Объектные. Хранилища объектного типа по структуре схожи с БД: файлы разделяются на «объекты», которые хранятся в едином пространстве в сгруппированном виде и классифицируются по уникальным идентификаторам и метаданным. Подобные СХД предназначены для работы с огромными массивами неструктурированных данных. Они применяются в машинном обучении, big data, аналитике, облачных технологиях, позволяют хранить объемные мультимедийные файлы, создавать бэкапы и

организовывать хостинг. Функционирует такая система немного медленнее блочной.

Существующие на рынке решения для систем хранения данных отражают цепочку связей между приложениями, которые используют информацию, и устройствами-носителями этой информации. В современных СХД эта цепь состоит из трех звеньев: создания RAID-массивов, обработки метаданных (позволяет интерпретировать биты в файлы и записи) и предоставления данных приложению. Места и способы реализации этих звеньев в СХД меняются в зависимости от используемых аппаратных средств, ПО и протоколов. Всего выделяют три варианта топологий систем хранения [5]:

1. NAS. Network Attached Storage (NAS) — это, условно, хранилище, подключенное к сети. Оно представляет собой отдельно стоящую дисковую систему, которая интегрирована в локальную сеть. Как правило, это высокопроизводительный сервер с собственной ОС, поэтому все вышеперечисленные процессы-звенья происходят внутри него. Он обрабатывает данные на уровне файлов и работает с протоколами NFS/CIFS. NAS дает одновременный доступ к хранимой информации всем пользователям, а кроме того, его можно безболезненно расширять и изменять, подключая виртуальные машины, дисковые станции и другие серверы. Подходит для работы с файлами, к которым требуется коллективный доступ.

2. DAS. Direct Attached Storage (DAS) переводится как хранилище с прямым подключением. Такая СХД подключается к серверу по тому же принципу, что условная флэш-карта к персональному компьютеру. В этом случае она выполняет только одно из звеньев, а именно предоставляет данные. Остальные процессы происходят на стороне сервера. Пользователи подключаются к хранилищу через него. В итоге клиенты имеют файловый доступ, а сам сервер — блочный.

Особенности архитектуры DAS-системы обеспечивают простоту ее развертывания и управления, но требуют выделенного хоста и ограничивают возможности расширения, поскольку количество возможных подключений зависит от количества имеющихся портов.

3. SAN. Storage Area Networks (SAN) — это сети хранения данных. Такое хранилище можно назвать промежуточным вариантом между двумя предыдущими. Здесь СХД отвечает за обеспечение RAID и предоставление данных, но метаданными по-прежнему занимается сервер. SAN предназначена для использования хранилищ разных типов (от обычных дисков до ленточных библиотек), которые ОС воспринимает как один сетевой логический диск. Это целая инфраструктура, объединенная оптическими каналами связи. В системе применяются FC и iSCSI протоколы,

для нее характерен блочный тип хранения.

Прежде чем выбирать определенный тип хранилища, стоит определить, для чего предназначена система хранения данных конкретно в вашем случае. Какие задачи она будет решать, какая информация будет на ней храниться и насколько она важна? Эти и другие критерии необходимо учесть заранее [6]:

1. Тип данных. Во-первых, нужно ориентироваться на тип данных. Большие медиафайлы, рабочая документация, неструктурированный массив для обработки нейросетью или любые другие данные будут требовать от СХД соответствующих технологий обработки, компрессии, скорости доступа и т. п. Система для хранения корпоративной информации в компании будет отличаться от системы, работающей в области машинного обучения.

2. Объем данных. Во-вторых, важную роль играет объем информации, которую необходимо хранить и обрабатывать. Он влияет на выбор дисковых накопителей, а также — на выбор архитектурного решения, поскольку возможности масштабирования в них отличаются.

3. Отказоустойчивость. Также необходимо оценить важность самих данных, установить цену их потери за определенный промежуток времени. Чем выше эта стоимость, тем более надежную СХД нужно выбирать. Кроме того, этот шаг позволит рассчитать показатели точки и времени восстановления (RPO и RTO) данных, т. е. узнать допустимые для бизнеса пределы по объему и продолжительности сбоя.

4. Производительность. Если есть готовый проект, под который нужно подобрать СХД, определить необходимый уровень производительности несложно. Если же система закупается под новый проект, оценить его нагрузку будет труднее. В этом случае рекомендуется провести предварительный тест СХД или хотя бы поговорить с коллегами/поставщиками, которые уже решали схожие задачи.

5. Вендор. Подобрать качественного поставщика — не менее важная задача. Главное — найти вариант, который будет удовлетворять потребности по функционалу, надежности, удобству и цене. Иногда даже для сложных и высоконагруженных проектов подходят бюджетные решения.

3 ФАКТОРЫ, СПОСОБСТВУЮЩИЕ РАЗВИТИЮ СХД

Рост объемов данных, возросшие требования к надежности хранения и быстродействию доступа к данным делают необходимым выделение средств хранения в отдельную подсистему вычислительного комплекса. 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 восстановлению [4].

Происходит заметное развитие потребности не только в приобретении СХД корпоративными клиентами, но и в строгом учете, аудите и мониторинге использования дорогостоящих ресурсов. Нет ничего хуже остановки бизнес-процессов из-за невозможности своевременно получить необходимые данные (или полной их утраты), а ведь это может повлечь за собой необратимые последствия.

Еще один фактор — появление на рынке множества компаний, которые предлагают свои решения для поддержки бизнеса предприятий: ERP, биллинговые системы, системы поддержки принятия решений и т. д. Все они позволяют собирать детальные данные самого разного характера в огромных объемах. При наличии в организации развитой ИТ-инфраструктуры эти данные можно собрать вместе и проанализировать их [4].

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 СУБД. Пользователей уже не волнует, кто сделал для их компьютера процессор или оперативную память, — они рассматривают хранилище данных как некую услугу. И это важнейший сдвиг в сознании.

Технологии, которые позволяют использовать хранилища данных для оптимизации операционных бизнес-процессов практически в реальном времени не только для высококвалифицированных аналитиков и топ-менеджеров, но и для сотрудников фронт-офиса, в частности для сотрудников офисов продаж и контактных центров. Принятие решений делегируется сотрудникам, стоящим на более низких ступенях корпоративной лестницы. Необходимые им отчеты, как правило, просты и кратки, но их требуется очень много, а время формирования должно быть

НЕВЕЛИКО.

ЗАКЛЮЧЕНИЕ

СХД успешно применяются в разных сферах человеческой деятельности: от банковского обслуживания и телекоммуникаций до образовательных услуг и электронных СМИ. Но наиболее активно они внедряются в сегменте разработки программного обеспечения, в информационных и мультимедийных сервисах, а также в здравоохранении.

Все существующие СХД делятся на использующие аппаратный RAID и специализированное ПО для расчета RAID – программный RAID. Последние системы являются более экономичными. Теперь многие задачи обработки и хранения данных значительно эффективнее решаются в рамках СХД с программным RAID. Например - резервирование системных дисков и виртуальных машин, хранение и обработка видео, работа с крупными файлами в системах документооборота.

Крупные IT-компании одновременно работают с большим количеством проектов, каждый из которых необходимо постоянно модернизировать и тестировать, поэтому они вынуждены хранить десятки и сотни версий одного и того же продукта.

При разработке современных компьютерных приложений часто бывает необходимо создавать специальные мультиоблачные структуры, запускать ресурсоемкие виртуальные машины, обрабатывать большие данные. Производительность СХД и возможность высокоскоростного доступа к глобальным массивам информации упрощают решение этих задач.

В отраслях, которые ориентированы на максимально качественное удовлетворение потребностей клиентов, огромную роль играет скорость обработки их запросов. Это очень важно в телекоммуникационных сервисах, при выполнении банковских транзакций, в розничной торговле.

Компании, предоставляющие эти услуги, используют мощные CRM-системы, рассчитанные на одновременную работу сотен пользователей, в том числе с поддержкой технологий виртуализации рабочих мест. То есть система хранения данных должна обеспечивать выполнение тысяч параллельных запросов практически с нулевой задержкой. Решения на базе технологии разделения трафика отлично справляются с такой нагрузкой.

Средства массовой информации предоставляют интерактивный контент тысячам пользователей. Зачастую это огромные видеофайлы, которые перед загрузкой на сервер редактируют сотрудники, находящиеся в разных частях земного шара. Очевидно, что для этого необходимы значительные емкости хранения информации и высокая скорость коммуникаций между ними.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Классификация информационных технологий [Электронный ресурс] / Сайт преподавателя математики, информатики и физики. — Режим доступа: https://bosov.ucoz.ru/2022/F325/lekcija_2.pdf. — Дата доступа: 10.12.2024.
2. Изучение методов хранения информации [Электронный ресурс] / Портал информационных продуктов «Инфо-мастер». — Режим доступа: <https://info-master.su/>. — Дата доступа: 10.12.2024.
3. Устройства для хранения информации и работа с ними [Электронный ресурс] / ООО «Союз пенсионеров России». — Режим доступа: <https://sprno.ru>. — Дата доступа: 10.12.2024.
4. Система хранения данных [Электронный ресурс] / Информационный портал «Государство. Бизнес. Технологии». — Режим доступа: <https://www.tadviser.ru/>. — Дата доступа: 10.12.2024.
5. Что такое СХД [Электронный ресурс] / Хостинг «Евробайт». — Режим доступа: <https://eurobyte.ru/articles/chto-takoe-sistema-khroneniya-dannykh-i-kakie-problemy-ona-reshaet/>. — Дата доступа: 10.12.2024.
6. Что такое СХД [Электронный ресурс] / Провайдер IT-структуры Selectel. — Режим доступа: <https://selectel.ru/blog/what-is-shd/>. — Дата доступа: 10.12.2024.