

ИНФОРМАТИКА

УДК 004.056.2

АРХИТЕКТУРНЫЕ ОСОБЕННОСТИ СИСТЕМ ХРАНЕНИЯ ДАННЫХЮ.Г. ВАШИНКО¹, А.В. БУДНИК², В.М. БОНДАРИК²

¹Центр обеспечения безопасности информационных технологий
Вооруженных Сил Республики Беларусь
Коммунистическая, 1, Минск, Беларусь

²Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, Беларусь

Поступила в редакцию 5 мая 2010

Выполнен обзор существующих подходов к реализации систем хранения данных, а также рассмотрены основные критерии эффективности их использования. На основании анализа различных архитектурных решений составлена обобщенная схема систем хранения данных. Дано пояснение ключевых связей между компонентами данной схемы.

Ключевые слова: система хранения данных, файл-сервер, дисковый массив, архитектура, инфраструктура доступа, жесткий диск.

Введение

С учетом быстрого роста накапливаемой информации в крупных компаниях и предприятиях особую значимость приобретают высокопроизводительные системы хранения. Следует отметить, что сохраняемая информация (данные) на сегодняшний день уже рассматриваются не как нечто, располагающееся на периферийных устройствах вычислительных машин, а как самостоятельный ресурс, который нуждается в надежном хранении, централизованном управлении и имеет жизненный цикл, по продолжительности значительно превышающий компьютерные платформы [1].

Теоретический анализ

Одним из наиболее распространенных решений хранения большого объема информации являются системы хранения данных (СХД) и, в частности, сетевые СХД. При оценке и анализе эффективности использования таких систем основными критериями являются:

- производительность, измеряемая в мегабайтах в секунду и зависящая от объема кеш-памяти, числа внутренних дисков, количества портов, контроллеров;
- надежность как способность системы выполнять свои функции с заданным качеством;
- функциональность, или степень соответствия возможностей СХД с возлагаемыми задачами;
- масштабируемость, т.е. способность к наращиванию функциональности и характеристик с сохранением всех существующих возможностей;
- управляемость и возможность "бесшовной" интеграции в существующую ИТ-инфраструктуру с поддержкой централизованного управления на основе открытых стандартов, а также обеспечение поддержки перспективных стандартов;
- совместимость с существующими технологиями;

– эксплуатационная эффективность как возможность обеспечения минимальной стоимости владения информацией;

– эргономичность как простота управления и удобство администрирования.

В общем случае СХД представляют собой жесткие диски или другие устройства хранения, подключенные к сети и позволяющие нескольким потенциальным пользователям работать с общей информацией [2]. Наиболее важной частью архитектуры СХД является инфраструктура доступа к данным, которая на сегодняшний день строится с использованием одного из следующих подходов:

– DAS (Direct Attached Storage) — прямое подключение дисковых массивов к серверам. Основное преимущество DAS-систем — относительная простота и малая стоимость. Однако оборотной стороной являются недостаточная управляемость и неоптимальная утилизация ресурсов. Системы DAS следует использовать при необходимости увеличения дискового пространства одного сервера и вынесения его за корпус;

– NAS (Network Attached Storage) — представляет собой готовый файл-сервер, использует в качестве транспорта IP-сети и оперирует на уровне файлов. Доступ к NAS-устройствам осуществляется по локальной сети на уровне протоколов передачи файлов (NFS, CIFS), и со стороны пользователя работа с таким устройством выглядит как подключаемый дисковый сетевой ресурс. NAS следует использовать, когда компании необходимо быстро и без дополнительных затрат добавить дисковое пространство в локальной сети для клиентов сети. Не рекомендуется использовать это решение в качестве дискового хранилища для серверов приложений, а также при наличии "узких мест" в сети;

– SAN (Storage Area Network) — технология, использующая высокоскоростную выделенную сеть на основе протокола Fibre Channel, связывающую один или несколько серверов с одной или несколькими системами хранения. Доступ к данным осуществляется на уровне блоков. Фактически SAN — это комбинация аппаратных средств и ПО, позволяющая большому числу пользователей хранить и совместно использовать информацию. При использовании технологии SAN сервер не задействует ресурсы на обращение к дискам, а высвобожденные ресурсы направляются на работу выполняемых на нем приложений. По мере роста объема данных применение SAN становится наиболее популярным решением. К SAN можно подключать серверы различных производителей, использующих разные операционные системы. [3]

Кроме того, для повышения уровня надежности, функциональности и эргономичности, СХД, как правило, включают в себя систему резервного копирования и архивирования, а также программное обеспечение управления хранением данных.

Разработка обобщенной архитектуры

На основании изученных подходов к созданию СХД можно выделить компоненты, которые в обязательном порядке должны входить в такого рода системы с целью соответствия критериям, приведенным в начале данной статьи. СХД должна включать в себя:

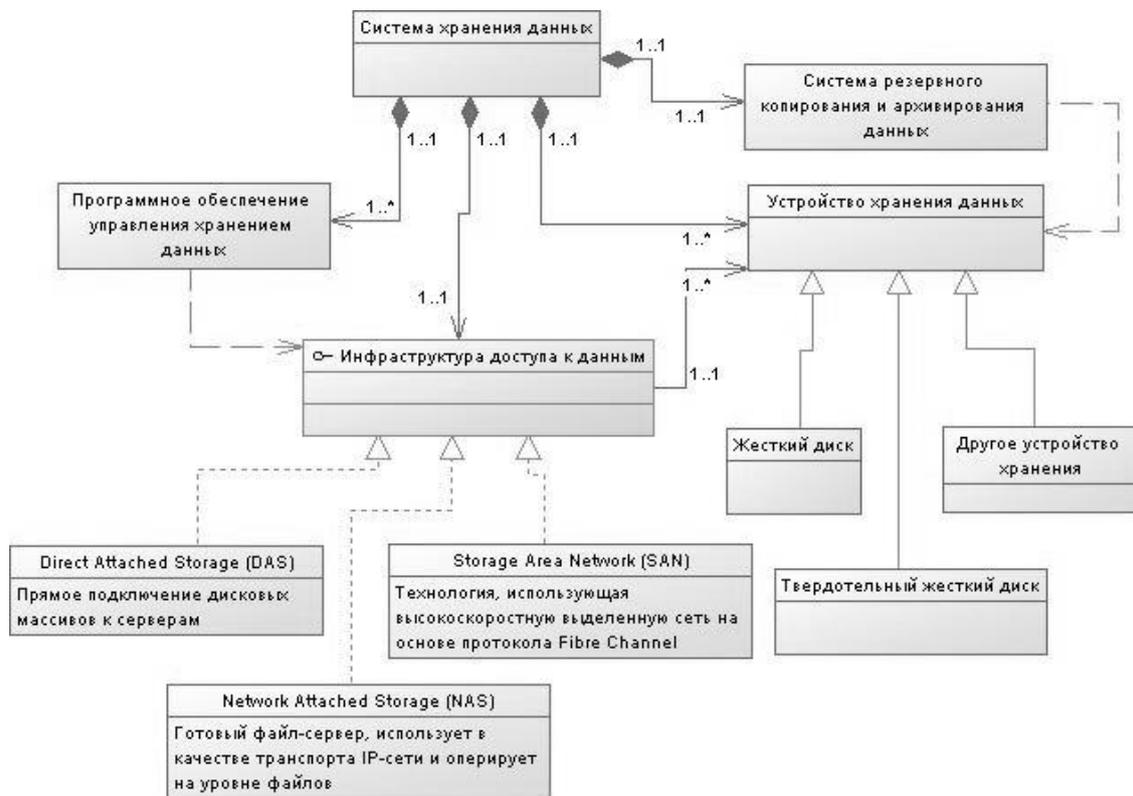
- устройства хранения данных;
- систему резервного копирования и архивирования данных;
- программное обеспечение управления хранением данных;
- инфраструктуру доступа к данным.

При анализировании взаимосвязи вышеперечисленных компонентов различных СХД выделена схема обобщенной архитектуры СХД. На рисунке данная архитектура изображена с использованием языка UML 2.0.

Основными связями для отображения взаимоотношений компонентов являются: композиция, ассоциация, наследование, зависимость. Отношение композиции показывает, что один из элементов является составной частью второго. Например, композиция между системой хранения данных и системой резервного копирования показывает, что система резервирования является составной частью СХД, причем обозначение "1..1" указывает на то, что данная система должна присутствовать обязательно и только в единственном экземпляре.

Особенно необходимо отметить компонент "Инфраструктура доступа к данным". Единственность экземпляра данного компонента обуславливается тем, что он является отличительной особенностью каждого вида СХД. Именно к инфраструктуре доступа применяют большин-

ство критериев эффективности СХД, поскольку она определяет способы и методы доступа к информации, которую хранит в себе система. Обобщенной реализации данного компонента на сегодняшний день не существует, поэтому на рисунке инфраструктура доступа к данным изображена в виде интерфейса, который определяет только правила реализации. Основные виды реализации данного компонента были описаны выше: DAS, NAS, SAN.



Обобщенная архитектура СХД

Композиция между СХД и устройством хранения данных показывает, что данные устройства могут присутствовать в системе во множественном числе (1..*). Нужно также отметить, что реализациями данного устройства могут выступать как обычные, так и твердотельные жесткие диски, а также ленточные устройства. В схеме дополнительно введен класс "другое устройство хранения". Это сделано с целью повышения гибкости обобщенной архитектуры СХД, поскольку устройства хранения данных постоянно модифицируются и видоизменяются. Отношение между жестким диском и устройством хранения данных называется наследованием. Данная связь означает, что жесткий диск наследует все свойства, которые есть у класса родителя (устройство хранения), а также дополнительно реализует свой собственный функционал.

Важную роль в данной схеме занимают отношения зависимости (пунктирная линия со стрелкой). Система резервного копирования зависит от устройства хранения данных, поскольку часть этих устройств данная система использует для записи резервных копий информации. Существует также зависимость между программным обеспечением (ПО) управления хранением данных и инфраструктурой доступа к ним, которая определяет возможности ПО.

Стоит обратить особое внимание на связь между инфраструктурой доступа к данным и устройством хранения информации. Эта связь в языке UML называется ассоциацией и в данном случае она показывает, что инфраструктура доступа использует в своей работе устройство хранения данных.

Таким образом, разработанная схема обобщенной архитектуры СХД (рисунок) отражает основные связи между компонентами СХД, а также показывает архитектуру такого рода систем на достаточно высоком абстрактном уровне.

Заключение

Как было отмечено выше, существует три основных подхода к построению СХД: DAS, NAS, SAN. Однако, несмотря на различные технологии, используемые в СХД, архитектурное построение таких систем весьма схоже. Это, в свою очередь, позволило обобщить существующие системы и представить в виде одной схемы.

Обобщенная архитектура СХД позволяет проводить анализ данных систем независимо от внутреннего устройства каждого компонента, а также может быть использована в качестве отправной точки для усовершенствования существующих подходов хранения больших объемов данных и создания нового типа СХД. Кроме того, разработанная схема позволяет выявлять недостатки и преимущества на уровне проектирования.

ARCHITECTURE FEATURES OF DATA STORAGE SYSTEMS

Y.G VASHINKO, A.V. BUDNIK, V.M. BONDARIK

Abstract

Data storage system implementation ways and the main effectiveness criteria are described. The generalized architecture based on the analysis of existing systems is designed. The key relations between architecture components are showed.

Литература

1. Фарли М. // Сети хранения данных. М., 2004.
2. Дайлин Н. Системы хранения данных в Microsoft Windows // М., 2005.
3. Ковалев В. Безопасность в системах хранения данных // LAN Magaine. 2005. № 6. С.56-62.