

**ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ**  
**Федеральное государственное образовательное бюджетное**  
**учреждение высшего профессионального образования**  
**«САНКТ-ПЕТЕРБУРГСКИЙ**  
**ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ**  
**им. проф. М. А. БОНЧ-БРУЕВИЧА»**

---

**А.Н.Губин**

## **Сети хранения данных**

**Краткий конспект лекций**

**Раздел 5. Системы хранения данных. Интеллектуализация систем  
хранения данных**

**САНКТ-ПЕТЕРБУРГ**  
**2017**

## **Содержание**

### **Раздел 1. Содержание курса, цели и задачи дисциплины**

- 1.1. Общее содержание курса, цели и задачи изучения дисциплины.*
- 1.2. Структура дисциплины и ее связь с другими курсами.*
- 1.3. Хранение информации как основная задача информационных центров.*

### **Раздел 2. Общие характеристики процессов хранения и управления данными**

- 2.1. Основные технологии хранения данных.*
- 2.2. Структурированные и неструктурированные данные.*
- 2.3. Общие характеристики процессов хранения и управления данными.*
- 2.4. Общая структура информационных центров.*
- 2.5. Управление хранением данных. Многоуровневое хранение данных.*

### **Раздел 3. Основные компоненты среды хранения данных**

- 3.1. Физические компоненты. Хост, дисковое устройство.*
- 3.2. Логические компоненты. Файловые системы.*
- 3.3. Производительность дисковых устройств.*

### **Раздел 4. Защита данных. RAID-массивы**

- 4.1. Распределение данных. Зеркалирование данных.*
- 4.2. Контроль четности.*
- 4.3. Основные конфигурации RAID-массивов.*
- 4.4. IOPS-операций и конфигурация дисков.*
- 4.5. Влияние наличия RAID на производительность дисковых устройств.*

### **Раздел 5. Системы хранения данных. Интеллектуализация систем хранения данных**

- 5.1. Основные компоненты интеллектуальных систем хранения информации.*
  - 5.2. Операция чтения и записи данных с использованием КЭШ-памяти.*
- Защита КЭШ- данных*
- 5.3. Основные компоненты системы хранения данных Symmetrix.*

### **Раздел 6. Сети хранения данных**

- 6.1. Архитектура сетей хранения данных. Особенности архитектуры СХД Fiber- Channel. Зонирование данных.*
- 6.2. Контекстная адресация данных.*
- 6.3. Виртуализация систем хранения данных.*

### **Раздел 7. Перспективные направления развития технологий хранения данных и управления информацией**

- 7.1. Внеполостная и внутриполостная виртуализация систем данных.*
- 7.2. Основные проблемы виртуализации систем хранения данных.*

## Раздел 5. Системы хранения данных. Интеллектуализация систем хранения данных

### 5.1. Основные компоненты интеллектуальных систем хранения информации

Интеллектуальные системы хранения данных, рассматриваемые в данном разделе, представляют собой полнофункциональные RAID массивы, обеспечивающие оптимальные параметры обработки запросов на ввод/вывод информации. Управление процессами обработки данных в таких массивах осуществляется операционной средой с использованием больших массивов Кэш-памяти.

В общем случае интеллектуальная СХД состоит из четырех основных компонентов (рис. 5.1) – внешнего блока, Кэш-памяти, внутреннего блока и массива физических дисков.

Внешний блок обеспечивает интерфейс связи между СХД и хостом, состоит из внешних портов и контроллеров, реализующих основные функции управления данными.

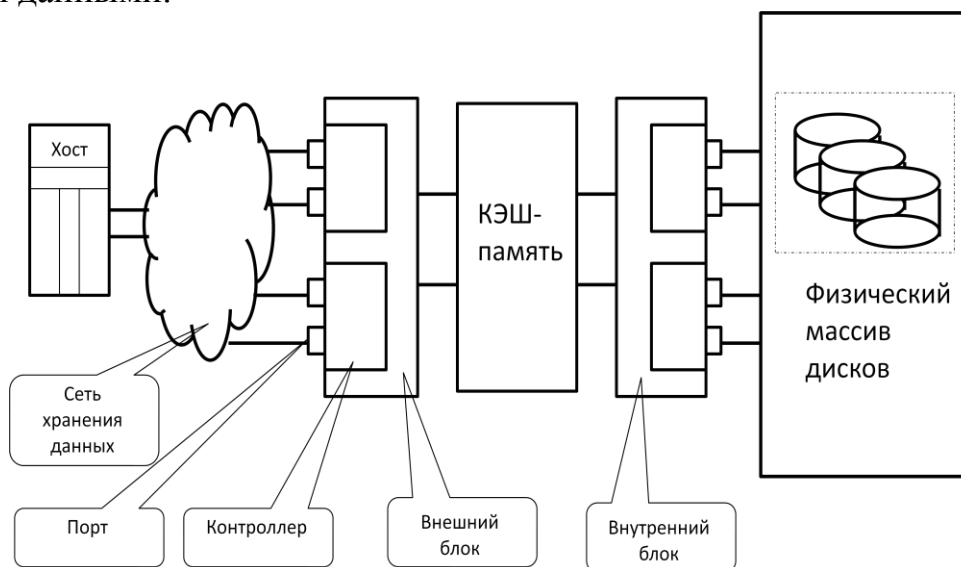


Рис. 5.1. Основные компоненты интеллектуальной системы хранения данных

Внешние порты работают в соответствии с используемыми транспортными протоколами передачи данных. Это, как правило, SCSI, Fibre Channel или iSCSI.

Для повышения доступности и надежности функционирования СХД входной блок контроллеров снабжается резервными портами.

Кэш-память представляет собой энергозависимую полупроводниковую память, в которую временно помещают данные для уменьшения времени, требуемого на обслуживание запросов хоста на выполнение операций ввода/вывода.

Внутренний блок обеспечивает выполнение интерфейсных функций между Кэш-памятью и физическими дисками, состоит из внутренних контроллеров и внутренних портов.

Алгоритмы работы внутренних контроллеров определяются необходимостью поддерживать функции RAID, а также обнаружение и исправление ошибок.

Использование нескольких внутренних контроллеров и портов повышает надежность и облегчает балансировку загрузки элементов СХД.

Физические диски предназначены для хранения данных. Перспективным направлением развития устройств хранения данных является использование энергонезависимых твердотельных накопителей (SSD - solid-state drive). В отличие от традиционных жестких дисков флэш-накопители не имеют движущихся частей и выигрывают в быстродействии, бесшумности и механической надежности. В массиве хранения данных флэш-накопители могут хранить до терабайта информации и при этом обеспечивают меньшее (более 30%) энергопотребление по сравнению с традиционными механическими дисковыми накопителями.

### Определение очередности выполнения запросов

Контроллеры внешнего блока оптимизируют обработку запросов ввода/вывода, изменяя очередность выполнения запросов.

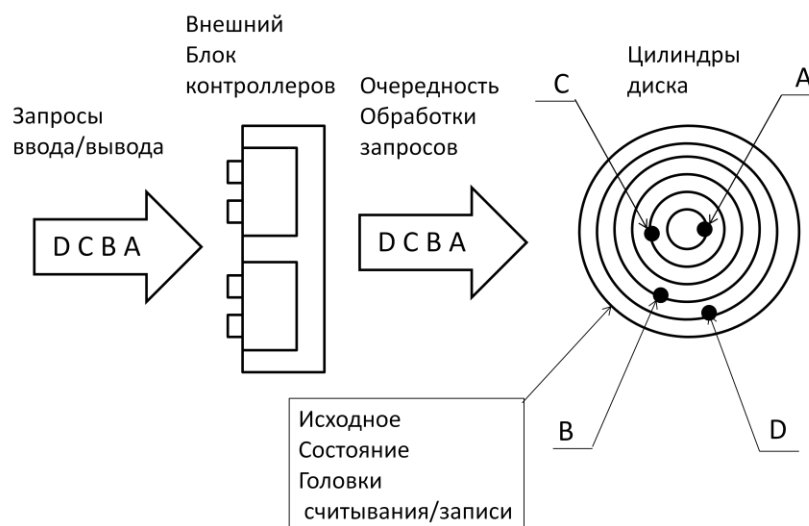


Рис. 5.2. Обработка запросов согласно алгоритму FIFO

Существует несколько наиболее распространенных алгоритмов обработки запросов в СХД.

Алгоритм "Первый пришел, первым выполнен" (FIFO - first in, first out). Это стандартный алгоритм, при котором запросы выполняются в последовательности, в которой они были получены (рис. 5.2). То есть очередность выполнения запросов не изменяется.

Оценка общего времени выполнения запросов хоста приведена на рис. 5.3. На указанном рисунке:

$t_p$  – время прохождения блока головок чтения/записи одной дорожки при выполнении операции позиционирования головок;

$t_o$  – время ожидания момента прохождения нужного блока данных под головкой чтения/записи;

$t_z$  – время чтения/записи блока данных.

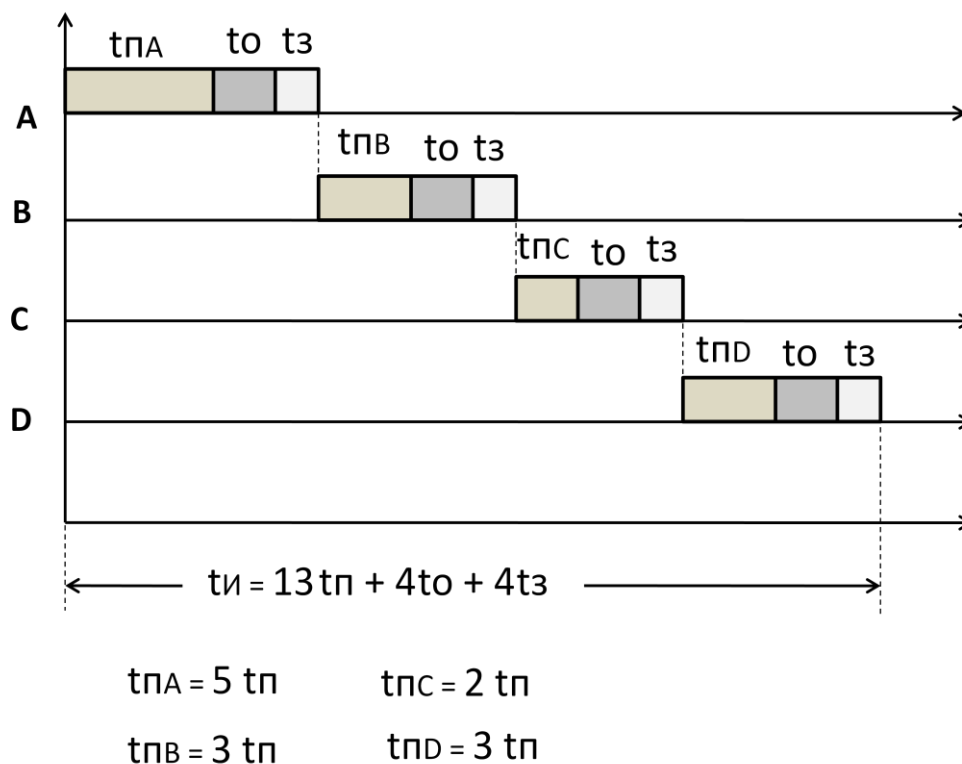


Рис. 5.3. Оценка времени выполнения запросов согласно алгоритму FIFO

Алгоритм оптимизации времени позиционирования блока головок чтения/записи. При использовании данного алгоритма порядок обработки запросов изменяется согласно рис. 5.4.

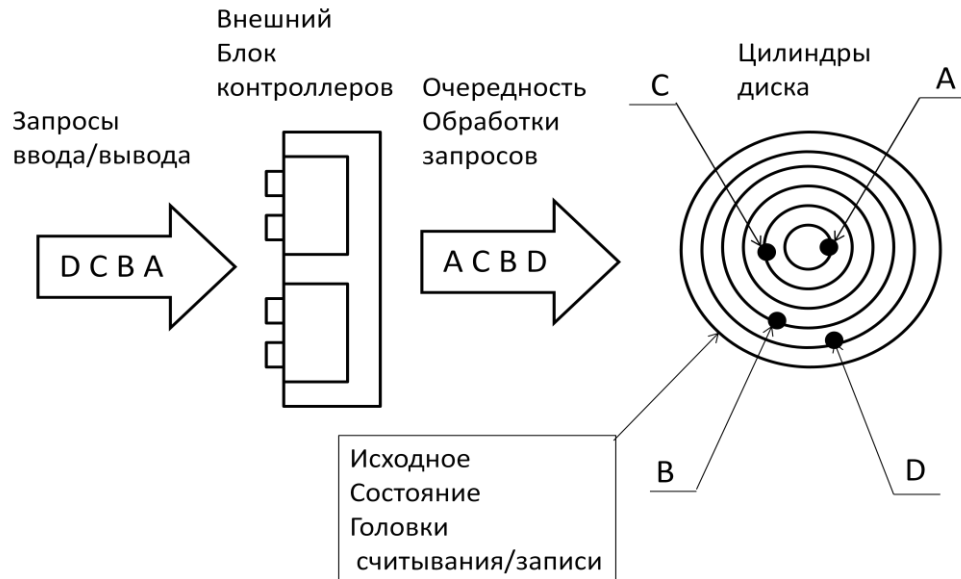


Рис. 5.4. Обработка запросов согласно алгоритму оптимизации времени позиционирования головок

Оценка общего времени выполнения запросов хоста при использовании алгоритма оптимизации времени позиционирования головок приведена на рис. 5.5.

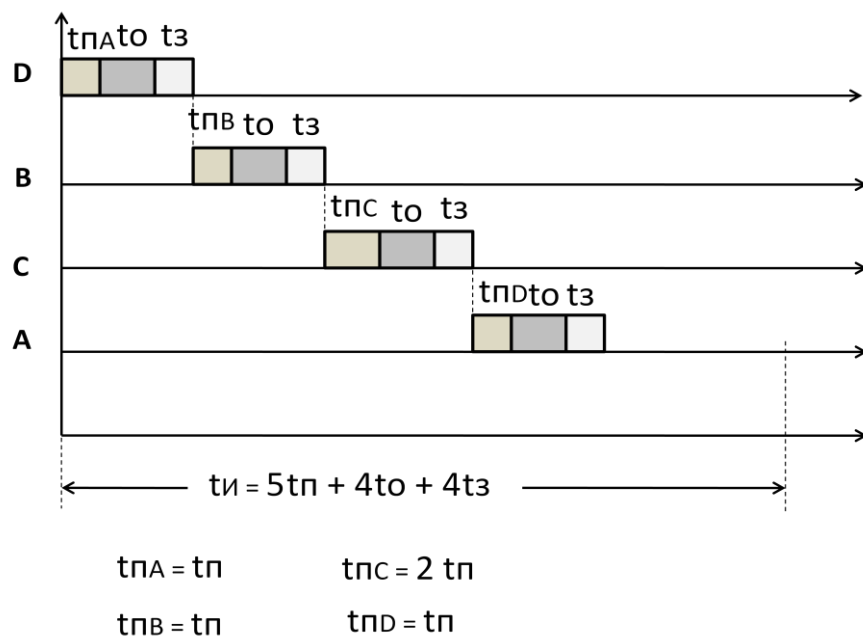


Рис. 5.5. Оценка времени выполнения запросов согласно алгоритму оптимизации времени позиционирования головок

Таким образом, общее время исполнения запросов A, B, C, D при оптимизации перемещений блока головок записи/считывания уменьшилось на величину  $8 t_{п}$ .

Существуют алгоритмы обработки запросов, которые наряду с оптимизацией времени позиционирования головок считывания/записи, учитывают время ожидания момента прохождения нужного блока под головками считывания/записи.

## 5.2. Операция чтения и записи данных с использованием КЭШ-памяти. Защита КЭШ- данных

Использование Кэш-памяти – это универсальный способ ускорения доступа к данным СХД. Улучшение быстродействия СХД обеспечивается за счет исключения из процесса доступа к данным механических задержек, связанных с физическими дисками, которые являются самыми медленными компонентами интеллектуальных систем хранения данных.

Под Кэш-памятью понимается не только способ ускорения доступа к данным, но и быстродействующее запоминающее устройство.

Обычно Кэш-память организована в виде страниц (слотов), размер которых конфигурируется в соответствии с размерами порций данных, которыми оперирует приложение (рис.5.6). Вся Кэш-память делится на память для хранения массивов данных и оперативную теговую память. В массиве данных хранятся данные приложения или дискового массива, а теговая память используется для управления процессами обработки данных и содержит, как правило, информацию о расположении данных, а также служебную информацию, характеризующую время нахождения данных в КЭШ, актуальность хранимых данных и др.

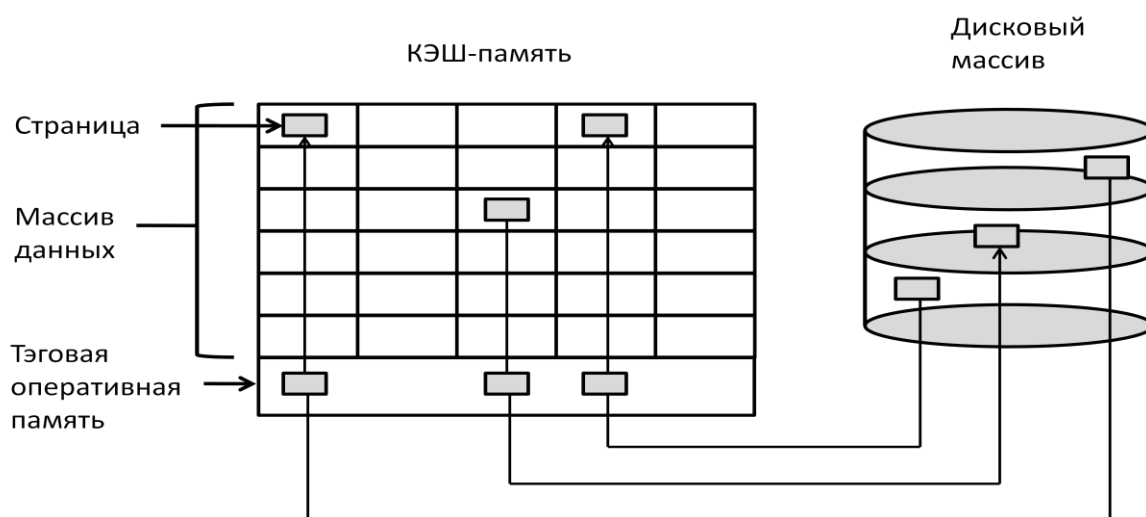


Рис. 5.6. Общая структура Кэш-памяти СХД

Когда приложение отправляет запрос на чтение информации с дискового массива, то внешний контроллер обращается к теговой оперативной памяти

КЭШ, чтобы определить, имеются ли запрошенные данные в Кэш-памяти. Если запрошенные данные найдены в Кэш-памяти, то такое событие называют КЭШ-попаданием при чтении, и данные передаются приложению без обращения к дисковому массиву. Таким образом, обеспечивается быстрый ответ (около миллисекунды) на запрос приложения.

Событие, когда запрошенные данные не найдены в Кэш-памяти называется КЭШ-промахом, при этом запрошенные приложением данные должны быть считаны с дискового массива, помещены в Кэш-память и переданы приложению. При этом время выполнения запроса существенно увеличивается.

При последовательных запросах на чтение в интеллектуальных СХД используется алгоритм предварительной выборки (предварительного считывания). В этом случае при выполнении одного запроса с дискового массива извлекается смежный набор связанных блоков. То есть, с дискового массива считывается и помещается в Кэш-память несколько блоков, которые еще не были запрошены приложением. Когда приложение в дальнейшем запросит эти блоки, то произойдет КЭШ-попадание при чтении.

В интеллектуальных системах хранения данных используются фиксированные и переменные размеры данных предварительного считывания.

Фиксированное предварительное считывание позволяет существенно уменьшить время выполнения запросов на считывание, если размеры ввода/вывода для приложения постоянны.

Переменное предварительное считывание позволяет интеллектуальной системе хранения считывать объем данных, кратный размеру запроса приложения.

Общее время доступа к данным с учетом Кэширования зависит от вероятности КЭШ-попадания.

Так, если время доступа к данным на дисковом массиве равно  $t_1$ , а время доступа к данным расположенным в Кэш-памяти равно  $t_2$ , то при вероятности КЭШ-попадания равной  $p$ , среднее время доступа к данным определится как

$$t = t_1(1-p) + t_2 \cdot p$$

Например, при  $p = 0,8$

$$t = t_1(1-0,8) + t_2 \cdot 0,8$$

Если учесть, что  $t_2 \ll t_1$ , то можно считать, что среднее время доступа к данным уменьшилось в пять раз.

Опыт использования СХД с КЭШ памятью показывает, что при решении стандартных задач процент КЭШ-попаданий достаточно высок (более 90%).



Операции записи данных на дисковый массив с использованием Кэш-памяти также позволяют получить преимущества по быстродействию по сравнению с записью данных напрямую на дисковый массив.

Операции записи данных с использованием Кэш-памяти могут выполняться согласно следующим алгоритмами.

*КЭШ с отложенной записью.* В этом случае, данные помещаются в Кэш-память и сразу же в приложение отправляется подтверждение о записи. Позднее данные из нескольких запросов на запись переносятся на дисковый массив. Реагирование приложения на запрос о записи данных происходит гораздо быстрее (без учета времени записи данных на дисковый массив).

*Запись в обход КЭШ.* При использовании этого алгоритма, данные помещаются в Кэш-память и немедленно записываются на дисковый массив с отправлением подтверждения о записи приложению. Так как данные переносятся на дисковый массив сразу после получения запроса, риск потери данных из-за сбоя Кэш-памяти снижается, но время реакции на запрос о записи со стороны приложения увеличивается.

### **Управление Кэш-памятью**

Кэш-память является дорогим и ограниченным по объему ресурсом СХД. В процессе функционирования интеллектуальной системы хранения данных происходит заполнение данными страниц Кэш-памяти. Для поддержания режимов работы СХД, позволяющих эффективно использовать возможности Кэш-памяти, необходимо обеспечивать периодическое освобождение страниц Кэш-памяти для приема новой информации.

В интеллектуальных СХД используются различные алгоритмы управления процессами освобождения страниц Кэш-памяти для поддержания некоторого количества свободных страниц Кэш-памяти, а также перечня страниц, которые могут быть освобождены в случае необходимости.

Таковыми наиболее известными и широко используемыми алгоритмами являются LRU (Least Recently Used) и MRU (Most Recently Used).

Least Recently Used – это алгоритм замещает страницы на которых находятся давно не использованные данные. Контроллер Кэш-памяти постоянно отслеживает моменты доступа к данным, определяет давно не использованные страницы и либо освобождает эти страницы, либо маркирует их для повторного использования. Работа LRU алгоритма основывается на предположении, что данные, не использованные в течение некоторого времени, не будут запрошены приложением и в ближайшее время. Если такая страница содержит метку о том, что данные этой страницы не записаны на диск, то перед повторным использованием страницы ее содержимое записывается на дисковый массив.

Most Recently Used – это алгоритм противоположный по действию LRU. Страницы Кэш-памяти, которые были только что использованы, освобождаются или помечаются для повторного использования. Работа

алгоритма основывается на предположении, что недавно использованные данные некоторое время не будут затребованы приложением.

По мере заполнения Кэш-памяти данными интеллектуальная система хранения данных должна выполнять сброс "грязных" страниц (страниц с данными, записанными в Кэш-память, но не записанными на дисковый массив). Сброс представляет собой процесс передачи на диск данных, находящихся в Кэш-памяти.

Для управления процессом сброса в Кэш-памяти задают два пороговых уровня заполнения страниц: верхний уровень и нижний уровень.

Верхний уровень – это уровень заполнения Кэш-памяти при котором начинается быстрый сброс данных, находящихся в Кэш-памяти.

Нижний уровень - это уровень заполнения Кэш-памяти при котором СХД прекращает быстрый или принудительный сброс данных и переходит на фоновый режим сброса.

Сброс в фоновом режиме происходит постоянно в небольшом объеме, когда уровень заполнения Кэш-памяти находится между высоким и нижнем уровнями заполнения.

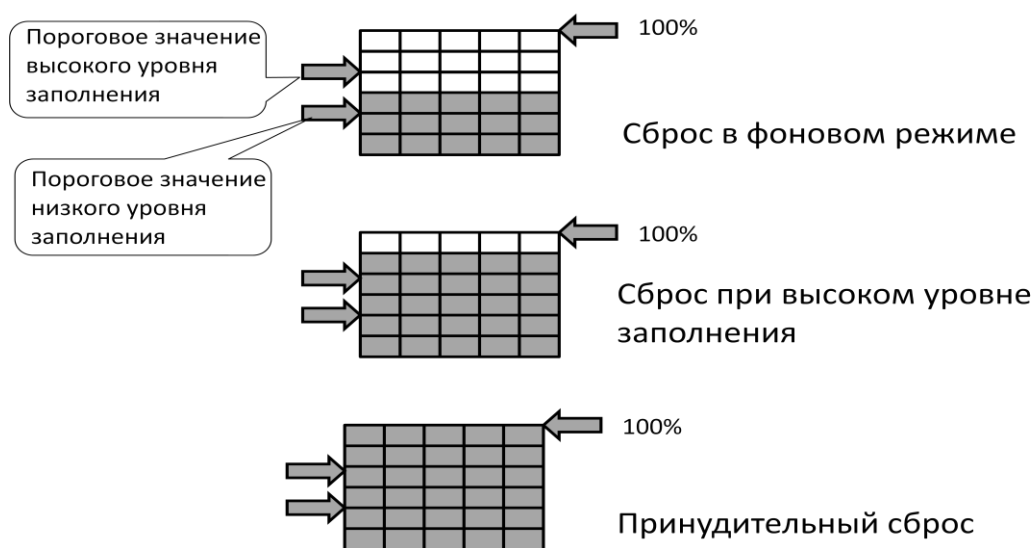


Рис. 5.7. Типы сброса Кэш-памяти

Сброс при высоком уровне заполнения Кэш-памяти активизируется, если при использовании Кэш-памяти достигается верхний уровень заполнения памяти, при этом СХД выделяет дополнительные ресурсы на выполнение сброса. Этот режим имеет минимальный уровень воздействия на процессы обработки запросов приложения.

Принудительный сброс происходит в случае получения запросов на большой объем ввода/вывода, когда Кэш-память заполняется на 100%. В этом случае данные всех грязных страниц Кэш-памяти принудительно перезаписываются на дисковый массив (рис. 5.7).

КЭШ является энергозависимой памятью и перебои в электропитании оборудования СХД может привести к потере данных, еще не переданных на дисковый массив. Такой риск потери данных можно существенно уменьшить, используя зеркальное кэширование или резервирование Кэш-памяти.

Зеркальное кэширование предусматривает хранение каждой записи данных на двух независимых картах Кэш-памяти. В случае сбоя в работе Кэш-памяти данные для записи на дисковый массив сохраняются в зеркально отображенной странице Кэш-памяти. Что касается операции считывания, то в случае сбоя в работе Кэш-памяти данные могут быть повторно считаны в Кэш-память с дискового массива.

Таким образом, зеркальному кэшированию подлежат только данные предназначенные для записи на дисковый массив.

При использовании методов зеркального кэширования возникает проблема синхронизации Кэш-памяти, то есть необходимо обеспечивать идентичность данных на двух разных страницах Кэш-памяти. Эту задачу решает операционная среда интеллектуальных систем хранения данных.

Резервирование Кэш-памяти предусматривает использование дополнительных магнитных дисков, которые используются для записи содержимого Кэш-памяти при возникновении аварийных ситуаций. При восстановлении работоспособности СХД данные с этих дисков переписываются в Кэш-память, а затем на дисковый массив.

### ***5.3. Основные компоненты системы хранения данных Symmetrix***

Системы EMC® Symmetrix® VMAXe™ с операционной средой Enginuity™ реализуют многоконтроллерную горизонтально масштабируемую архитектуру, предоставляющую предприятию возможности консолидации ресурсов и обеспечения эффективности хранения данных.

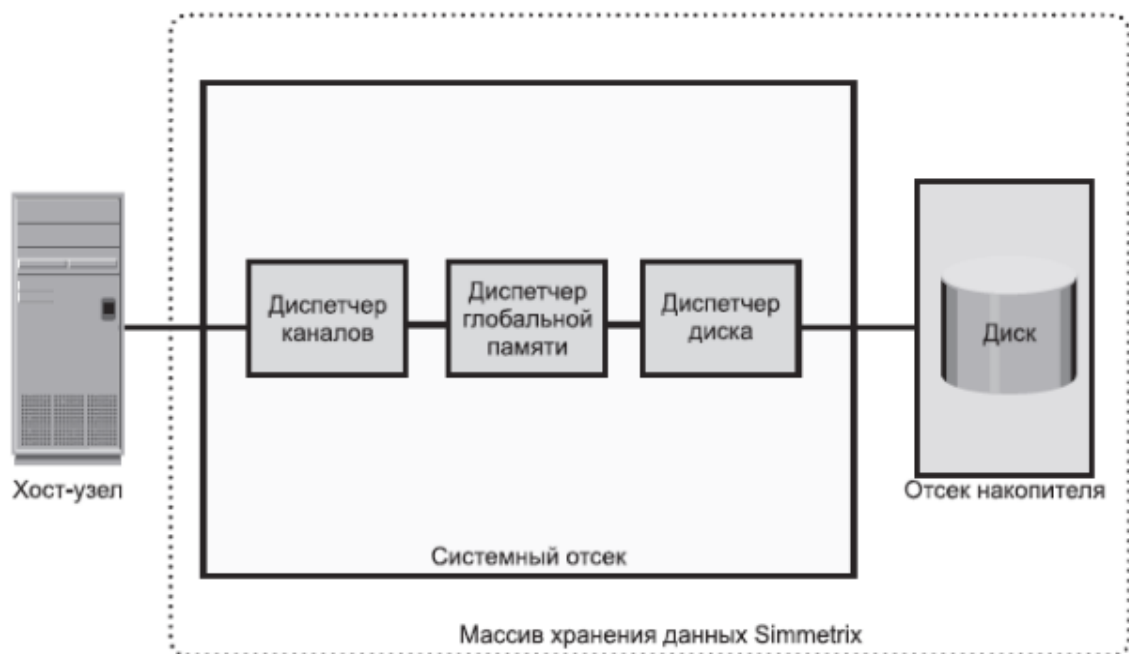
Symmetrix система масштаба корпорации позволяет хранить десятки терабайт данных, обеспечивает мгновенный доступ, быстрое резервное копирование и восстановление данных, а также зеркалирование массивов данных на другие системы Symmetrix для построения катастрофоустойчивых решений.

В отличие от традиционных систем хранения Symmetrix представляет собой специализированный массивно-параллельный сервер, содержащий несколько десятков процессоров PowerPC, до 64 гигабайт оперативной памяти и более трех с половиной сотен дисковых накопителей максимальным общим объемом около 70 терабайт.

Особое внимание при проектировании Symmetrix было уделено вопросам сохранности и целостности информации, обеспечение которых достигается за счет высокой отказоустойчивости оборудования и применения программных средств, обеспечивающих различного рода резервное копирование данных. Система выполнена по схеме полного функционального дублирования всех аппаратных узлов, исключая наличие критических точек, отказ в которых влечет за собой отказ системы в целом:

- Встроенные источники бесперебойного питания (два параллельных блока), обеспечивающие автономную работу устройства при отказе внешней системы питания.
- Четыре параллельные внутренние шины.
- Интерфейсные карты подключения внешних устройств к Symmetrix (Channel Director/Adapter), выполненные в виде двухпроцессорных систем, имеющих выход на две внутренние шины.
- Интерфейсные карты подключения дисковых накопителей (Disk Director), также выполненные по двухпроцессорной схеме.
- Дублированные SCSI-пути к каждому дисковому накопителю от различных диск-директоров.
- Возможность выбора RAID-1 либо RAID-S.
- Наряду с RAID-технологиями высокая степень сохранности данных обеспечивается за счет применения программных продуктов для резервного копирования данных TimeFinder и SRDF (Symmetrix Remote Data Facility).

Рис.5.8. Общая структура Symmetrix



Общая структура системы хранения данных Symmetrix представлена на рис. 5.8.

Системный отсек Symmetrix содержит внешние контроллеры (диспетчеры каналов) для подключения к хост-узлу, глобальную кэш-память большой емкости (диспетчер глобальной памяти) и внутренние контроллеры (диспетчеры дисков) для подключения к дисковому массиву.

Отсек накопителя представляет собой корпус, который может вмещать до 240 накопителей, каждый из которых содержит блок питания с резервированием и модули охлаждения. Предусматривается возможность использования от 4 до 15 дисковых накопителей с интерфейсом Fibre Channel на каждый корпус отсека накопителей.

