



# Техническое описание продуктов: СХД ENGINE AQ Лайт

Дата: 18.07.2024

Версия: 1.0.0



## Оглавление

1. Аппаратные платформы .....	3
2. Архитектура хранения.....	4
3. Функциональность.....	5
3.1 Управление СХД.....	5
3.2 REST API .....	5
3.3 Высокая доступность.....	6
3.3.1 Отказ контроллера.....	8
3.4 Файловый и блочный доступ.....	9
3.5 Мгновенные снимки, клоны, связанные клоны.....	9
3.6 Миграция блочных устройств «на лету» .....	11
3.7 Ускорение ввода/вывода для HDD дисков .....	11
3.7.1 SSD-кэширование для RDG.....	11
3.7.2 Онлайн-tiering для RDG .....	12
3.8 VLAN и BONDING .....	14
3.9 Quality of Service.....	15
3.10 Политика перестроения.....	15
3.11 Дедупликация .....	16
3.12 Компрессия .....	17
3.13 Авто-поддержка .....	18
4. Правила лицензирования.....	19
5. Поддерживаемые операционные системы и гипервизоры:.....	20

## 1. Аппаратные платформы

Система хранения данных АЭРОДИСК ENGINE AQ Лайт – двухконтроллерная. Все основные компоненты, включая блоки питания, фронт/бэк-энд адаптеры FC и Ethernet, дублированы.

Контроллерная пара ENGINE AQ Лайт представляет собой классическое исполнение СХД. Узлы контроллерной пары физически расположены в одном корпусе/шасси, соединены интерконнектом по шине Ethernet и подключены к общему бэмплейну дисковой корзины на передней панели корпуса. Использование внешних модулей для наращивания дисковой емкости, в текущей реализации ENGINE AQ Лайт не предусмотрено.



*На базе аппаратной платформы компании "Аквариус"*

Системы хранения ENGINE AQ Лайт поставляются с программным обеспечением ВОСТОК ЛАЙТ версии 1.0.0 или выше.

Все конфигурации поставляются с портами 25Gb Ethernet. В комплектацию СХД включены оптические трансиверы 10Gb Ethernet. Опционально доступны следующие опции расширения:

- оптические трансиверы 25Gb Ethernet;
- интерфейсные модули Fibre Channel 16Gb.

Доступные конфигурации контроллерных пар приведены в документе «Техническая спецификация».

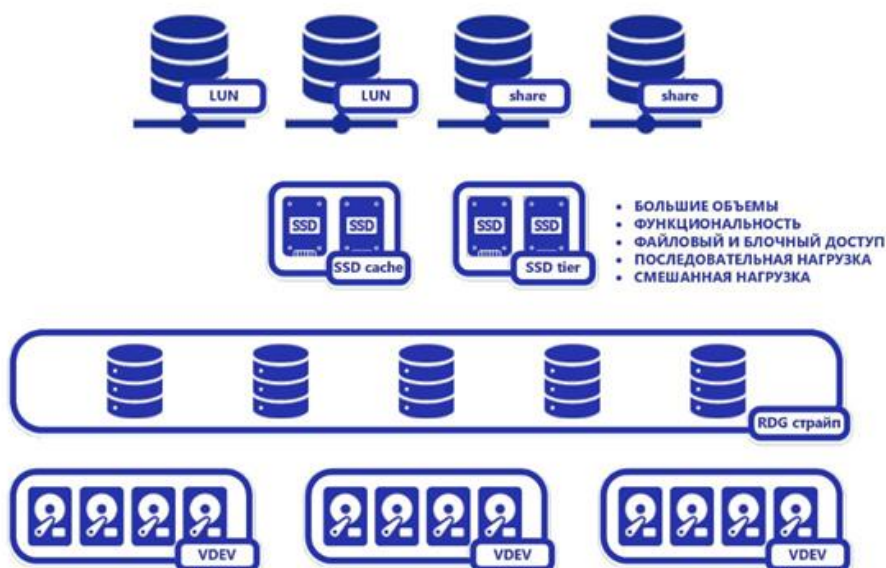
## 2. Архитектура хранения

Архитектура хранения данных в СХД АЭРОДИСК ENGINE AQ Лайт построена на использовании пулов RAID Distributed Group (RDG).

Основные особенности реализации RDG в системах АЭРОДИСК:

- RDG состоят из виртуальных устройств, каждое из которых имеет заданную структуру RAID (1/10, 5/50, 6/60, 6P/60P) (тройная четность);
- в RDG поддерживается и файловый, и блочный доступ;
- виртуальные устройства последовательно объединяются в одну виртуальную группу RDG, за счет чего количество дисков в группах (и для данных, и для четности) не ограничено;
- вне зависимости от объема тома или файловой системы все диски в группе участвуют в вводе-выводе для данного тома или файловой системы;
- диски горячей замены являются глобальными;
- используется SSD кэш для ускорения операций записи на жесткие диски;
- есть возможность использования SSD дисков для online-tiering;
- миграция LUN «на лету»;
- мгновенные снимки, снапклоны и связанные клоны;
- скорость перестроения RAID можно регулировать политикой перестроения;
- RDG наилучшим образом подходит для операций последовательного чтения/записи данных, а также для операций случайного чтения.

### АРХИТЕКТУРА RAID DISTRIBUTED GROUP

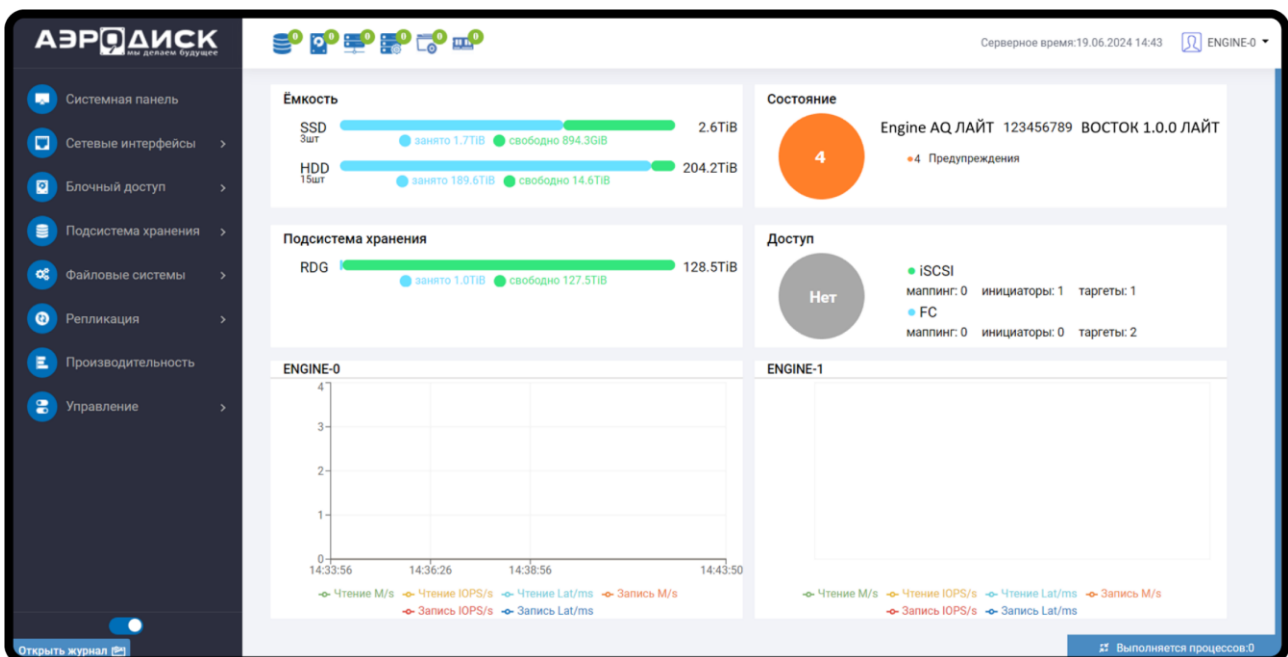


## 3. Функциональность

### 3.1 Управление СХД

Все системы хранения АЭРОДИСК используют единый интерактивный web-интерфейс на русском языке, позволяющий управлять всеми контроллерами СХД, установленными в системе, а также обеспечивающий:

- поддержку протокола HTTP и защищенного HTTPS для доступа к Web-интерфейсу;
- визуализацию контроллеров, дисков и портов ввода-вывода;
- визуализацию сенсоров и датчиков температуры;
- мониторинг состояния и нагрузки в реальном времени;
- логирование действий администратора;
- возможность выгрузки логов и статистики;
- командную строку (linux-like) для автоматизации операций;
- отправку оповещений по SMTP, SNMP;
- управление через Rest API;
- внешний мониторинг с помощью GRAFANA и Zabbix.

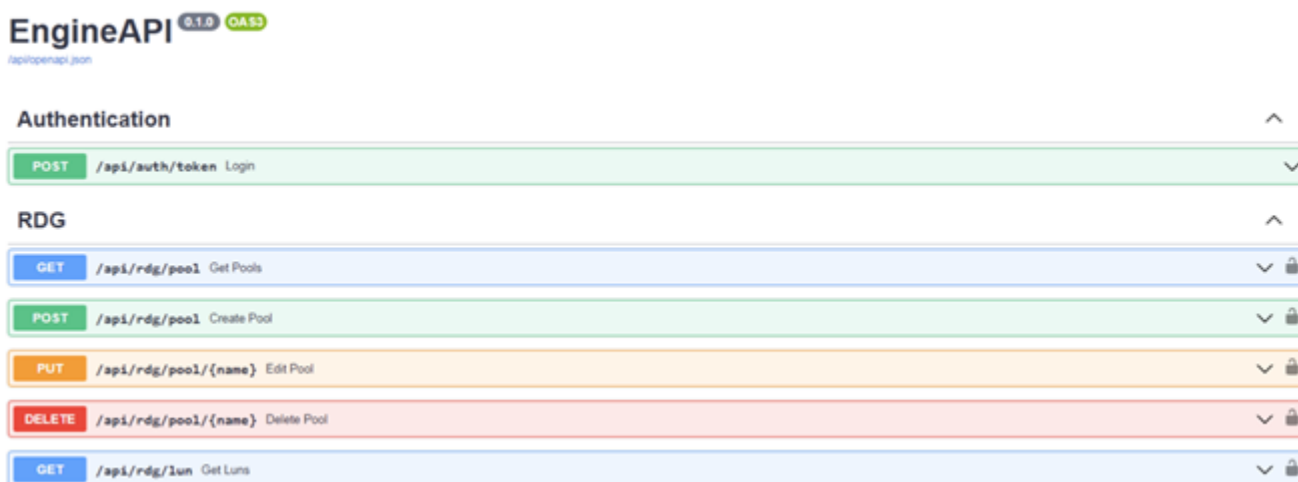


### 3.2 REST API

Интерфейс прикладного программирования (API) необходим для взаимодействия с другими программными системами, такими как системы резервного копирования и мониторинга. API позволяет автоматизировать управление СХД, интегрировать необходимые функции во

внешние системы и приложения: системы виртуализации, системы резервного копирования и восстановления, порталы самообслуживания в облачных инфраструктурах.

Реализация API в СХД АЭРОДИСК следует архитектурному подходу REST, что облегчает взаимодействие с другими информационными системами и масштабирование. Набор функций управления СХД включает создание и удаление пулов и томов, маппинг томов, управление созданием снимков данных, управление репликацией, QOS, файловыми средами и другие.



### 3.3 Высокая доступность

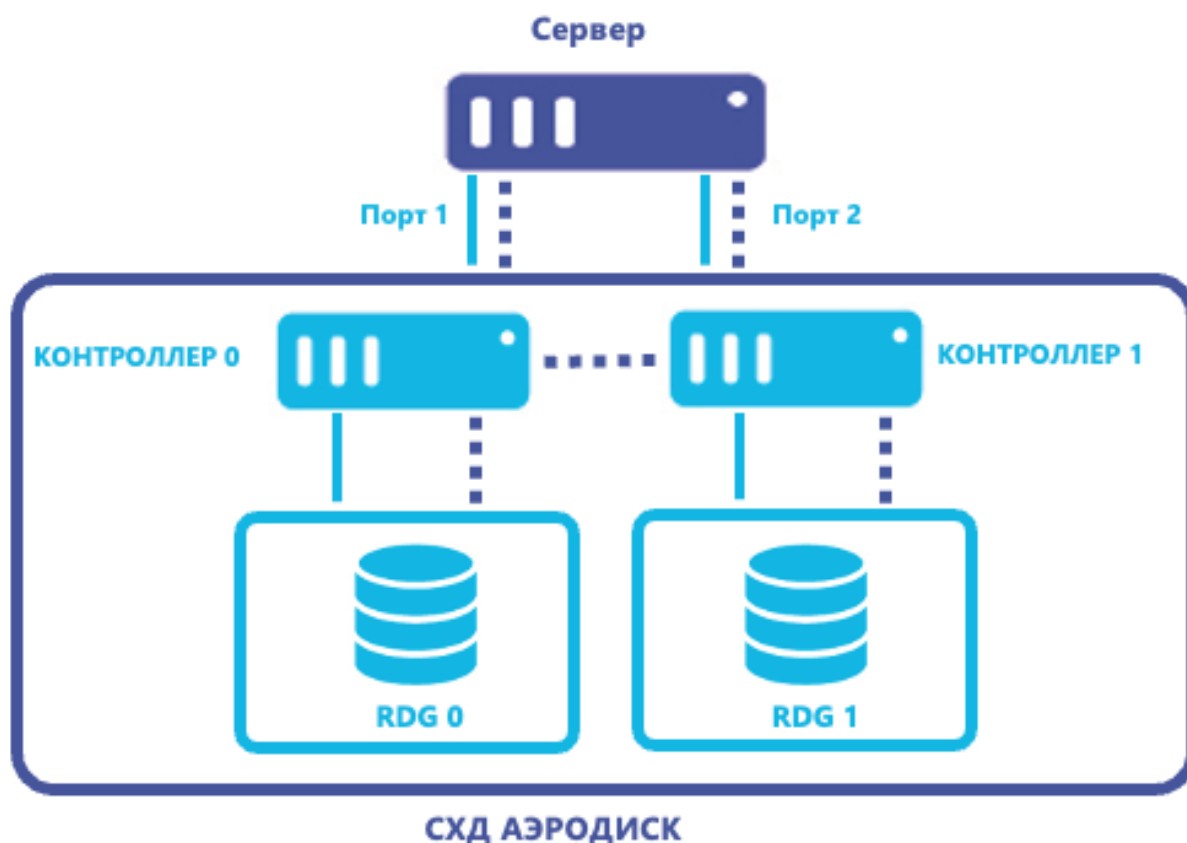
СХД АЭРОДИСК поддерживает высокую доступность Active-Active. Это означает, что все системные контроллеры всегда используются при обработке данных. Дисковые группы распределяются между всеми активными контроллерами. При этом администратор системы в случае необходимости (например, для обновления) может вручную переключать группы между контроллерами.

Кластерное ПО АЭРОДИСК работает как с блочным, так и с файловым доступом. Heartbeat между нодами выполняется с помощью интерконнекта (внутреннего или внешнего в зависимости от аппаратной реализации контроллерных пар). Кластер автоматически переключает оптимальные и неоптимальные пути, а также позволяет автоматически менять владельца групп хранения в следующих случаях:

- отказ контроллера (смена владельца);
- отказ задействованных в вводе-выводе портов СХД (смена владельца);
- отказ порта на хосте (смена путей оптимальный-неоптимальный).

На примере ниже показана двухконтроллерная конфигурация, которая подключена к 2-м портам хоста, для которых средствами ОС настроен multipath. На СХД созданы 2 группы

хранения: для одной назначен владельцем первый контроллер (контроллер-0), для другой владельцем назначен второй контроллер (контроллер-1). Оба контроллера видны обоим портам хоста. Для RDG0 владельцем назначен контроллер 0; пути через этот контроллер для данной группы являются оптимальными. Для RDG1 обратная ситуация: владельцем является контроллер-1, через него лежит оптимальный путь.

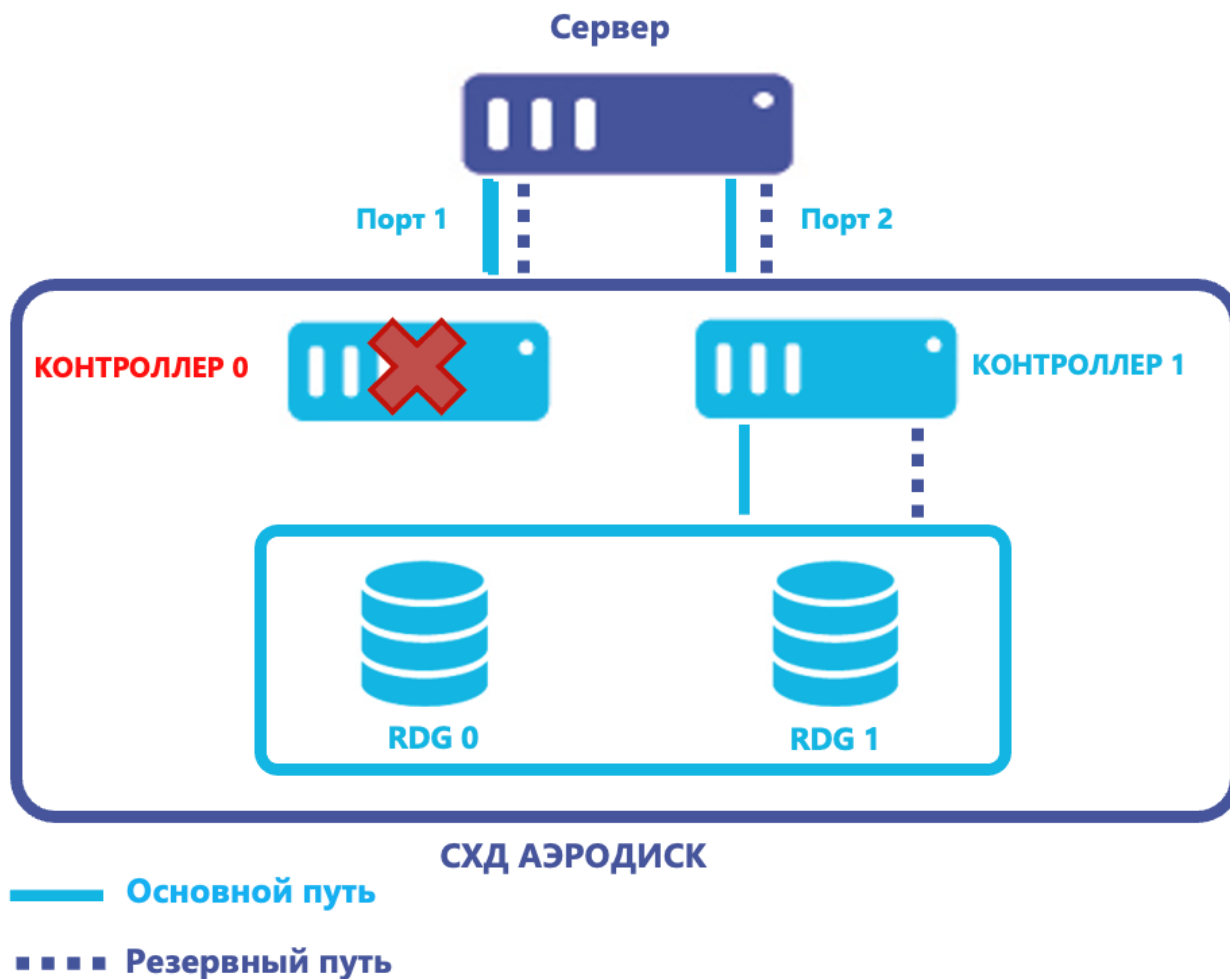


- Основной путь
- Резервный путь

В любой момент администратор СХД может сменить владельца каждой из групп. Процесс смены владельца занимает примерно 5-10 секунд и происходит без прерывания ввода-вывода. Эта же операция выполняется администратором для перевода контроллера в режим обслуживания, например, когда требуется аппаратное или программное обновление СХД.

### 3.3.1 Отказ контроллера

На схеме ниже представлена ситуация отказа контроллера. В случае физической потери контроллера (или 2-х портов ввода-вывода на контроллере) система выполнит принудительную смену владельца всех групп хранения на отказавшем контроллере. Далее произойдёт смена владельца, что происходит без прерывания ввода-вывода.



Когда контроллер-0 снова вернется в строй, администратору нужно будет вручную сменить владельца на контроллер-0 обратно.



### 3.4 Файловый и блочный доступ

**Блочный доступ** обеспечивается путем предоставления блочного устройства (LUN) конечному хосту или хостам по протоколам Fibre Channel, iSCSI с поддержкой iSER или IB.

**Файловый доступ** обеспечивается путем предоставления файловой системы по протоколам NFS и SMB (CIFS) конечному хосту или хостам. Для SMB (CIFS) может использоваться авторизация пользователей с помощью Active Directory. Файловый доступ работает только для RDG-групп.

LUN-ы и файловые системы создаются внутри RDG-групп. В рамках одной RDG-группы могут функционировать как LUN-ы, так и файловые системы. При этом размер RDG может быть динамично увеличен (т. е. в онлайн режиме) с помощью добавления дополнительных дисков в RDG.

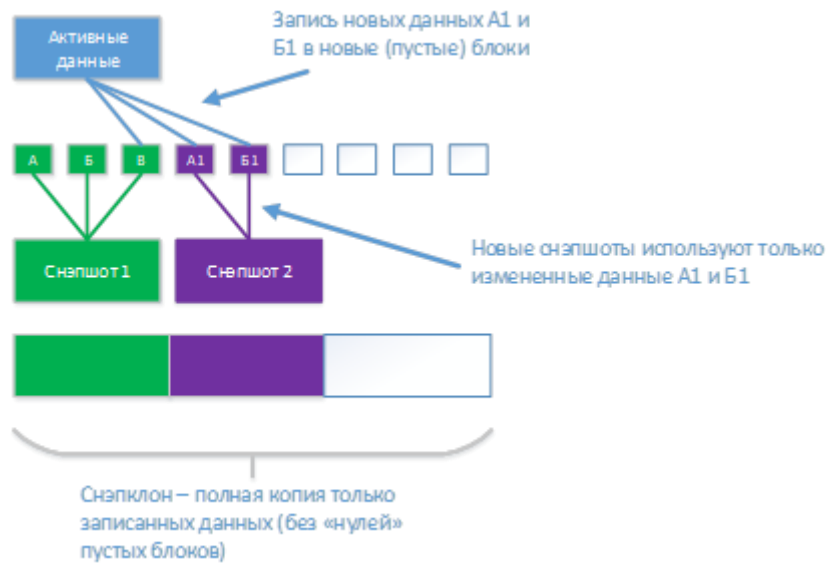
Как для файлового, так и для блочного доступа поддерживаются следующие уровни RAID для RDG:

- RAID 1/10;
- RAID 5/50;
- RAID 6/60;
- RAID 6P/60P (тройная чётность).

В системах хранения данных **АЭРОДИСК ENGINE AQ Лайт** файловый и блочный доступ можно обеспечивать с одного и того же контроллера, достаточно лишь наличия соответствующих Front-End адаптеров (FC, Ethernet). Установка дополнительных специальных модулей не требуется.

### 3.5 Мгновенные снимки, клоны, связанные клоны

Мгновенные снимки (снэпшоты) и связанные клоны в RDG используют модель переадресации при записи (Redirect-on-Write), т.е. СХД всегда пишет новые блоки данных в новое место, переставляя на них указатель, а старые блоки данных (т.е. на которые уже нет указателя) никогда не стираются, а помечаются системой как освобожденные. Этот механизм позволяет создавать любое количество снэпшотов и связанных клонов без какого-либо влияния на производительность СХД.



**Снимки** создаются мгновенно и изначально не потребляют дисковое пространство, а растут по мере изменения данных.

**Связанные клоны** создаются мгновенно и сразу могут быть доступны серверу на чтение/запись при наличии маппинга.

Полезной функцией является создание/удаление снимков по расписанию (локальная репликация). Это применимо, если требуется сохранять резервные копии данных очень часто, что невозможно сделать внешними системами резервного копирования, т.к. в силу их специфики (высокая нагрузка на каналы, долгое время записи и пр.) они резервируют данные обычно не чаще чем раз в сутки.

В этом случае есть возможность настроить расписание снимков, например, каждые 3 часа в течение суток со сроком хранения одни сутки. Через сутки снимки начнут перезаписываться заново, а данные старше суток уже будут сохранены внешней системой резервного копирования.

**Снэпклон** – это гибрид клона и снимка. Снэпклоны создаются быстрее, чем классические клоны и изначально занимают ровно ту полезную емкость, которую занимает источник. При этом снэпклон, как и классический клон, может находиться в любой группе.

Восстановление данных из снимков и связанных клонов можно выполнить двумя способами:

1. Откатить снимок/связанный клон, полностью перезаписав данные LUN/ФС. Это удобно, когда нужно быстро восстановить LUN/ФС полностью.
2. Присоединить связанный клон в виде отдельного LUN/ФС к хосту и восстановить данные с этого LUN. Такой способ подходит для ситуаций, когда не нужно

восстанавливать LUN/ФС целиком, а нужно восстановить только некоторые объекты (файлы).

Снэпклоны и связанные клоны возможно также подключать к хосту в виде отдельных LUN или файловых систем.

**СХД АЭРОДИСК не имеет ограничений по количеству созданных снэпшотов и снэпклонов, за исключением физического ограничения используемого оборудования.**

### 3.6 Миграция блочных устройств «на лету»

СХД АЭРОДИСК поддерживает функционал миграции блочных устройств «на лету» прозрачно для конечного потребителя. Функционал может быть полезен для перемещения данных на другой уровень RAID, например, с RAID5 на RAID10 для увеличения производительности. Также данные могут быть перемещены на другой тип дисков, например, с NLSAS дисков на SSD-диски. Возможны любые направления перемещения блочных устройств между типами дисков и типами рейдов.

### 3.7 Ускорение ввода/вывода для HDD дисков

Для реализации максимальной производительности и гибкости в СХД АЭРОДИСК предусмотрена функция ускорения ввода/вывода для HDD-дисков. Данная функция разделяется на три подфункции:

- SSD-кэш (SSD RW);
- SSD-кэш и хранение метаданных (SSD RW + MCACHE);
- online-tiering (SSD online-tiering).

#### 3.7.1 SSD-кэширование для RDG

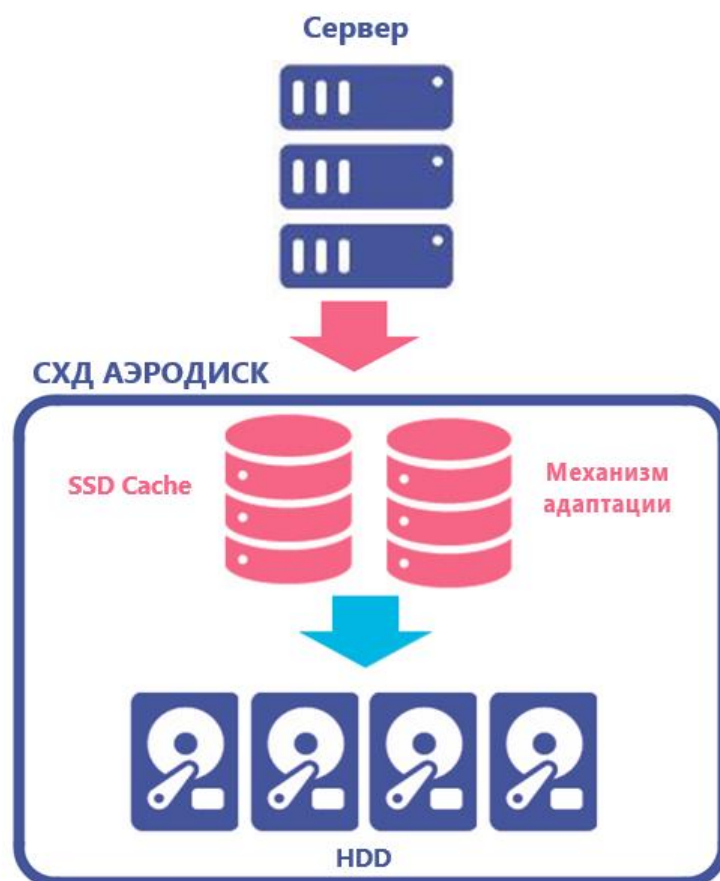
SSD-кэш добавляется к RDG-группе в целом и применяется ко всем LUN-ам и ФС, работающим в данной RDG сразу после добавления. При создании гибридного хранилища SSD-диски добавляются в кэш пул на запись/чтение (минимум 2 диска) в RAID1.

SSD-кэш применяется для всех записываемых транзакций. Подтверждение серверу о том, что данные записаны, отправляется сразу после записи данных в SSD-кэш.

При использовании технологии SSD RW+MCache все метаданные RDG-группы хранятся на SSD-дисках. Это позволяет снять часть нагрузки с шпиндельных дисков и сильно уменьшить задержки при случайных чтении и записи.

SSD-RW-кэш является достаточно экономичным вариантом повышения производительности СХД, поскольку не требует дисков большого объема. При этом, поскольку данный механизм активно использует запись, это утилизирует ресурс надежности SSD-дисков (DWPD) и для данного типа кэша рекомендуется использовать SSD-диски с высоким показателем DPWD (3+).

На рисунке ниже приведен пример логики работы SSD-кэша.



Практическая информация о конфигурировании гибридного хранилища приведена в документе «АЭРОДИСК RAID-guide».

Системы АЭРОДИСК не имеют ограничения по объему SSD и RAM кэша, за исключением физического ограничения используемого оборудования.

### 3.7.2 Онлайн-tiering для RDG

Online-tiering хранение позволяет перемещать блоки данных между различными уровнями в зависимости от нагрузки на них, позволяя тем самым размещать более «горячие» данные (т.е.

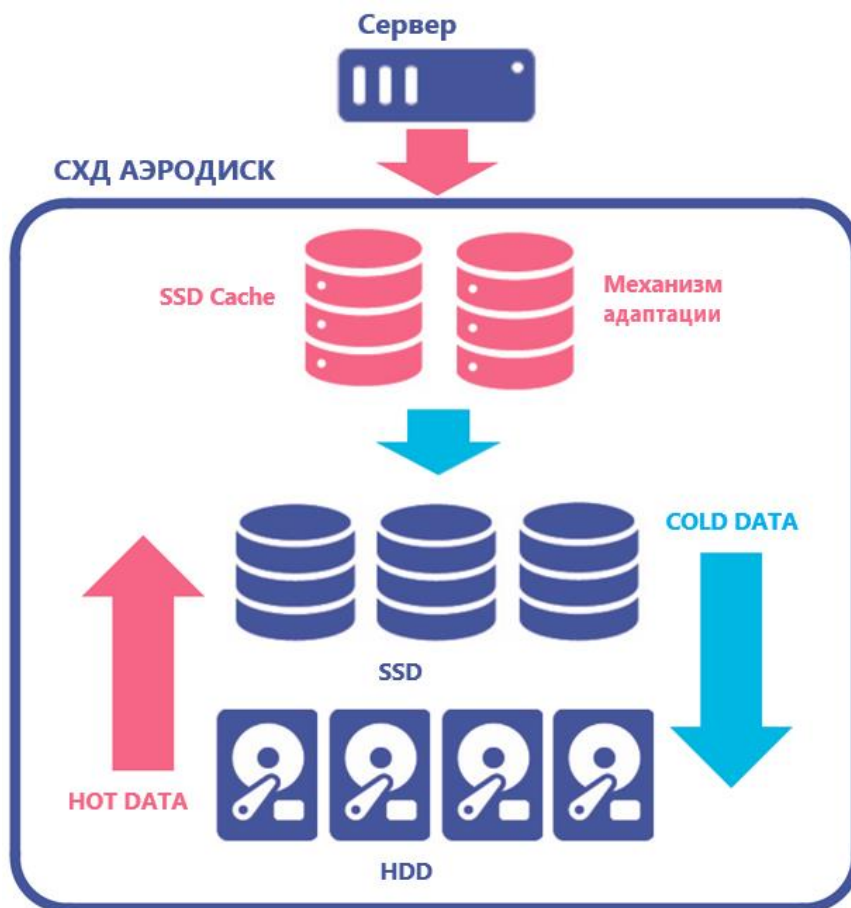
часто используемые) на быстрых дисках, а более «холодные» данные (т.е. редко используемые) на медленных.

Перемещение блоков данных между уровнями происходит в онлайн-режиме.

Диски для многоуровневого хранения также добавляются на уровне RDG-группы, после добавления дисков в online-tier группа меняет статус на «Быстрый».

Минимальное количество дисков на уровень online-tier – 2. Для эффективного использования online-tiering рекомендуется использовать SSD-диски большого объема.

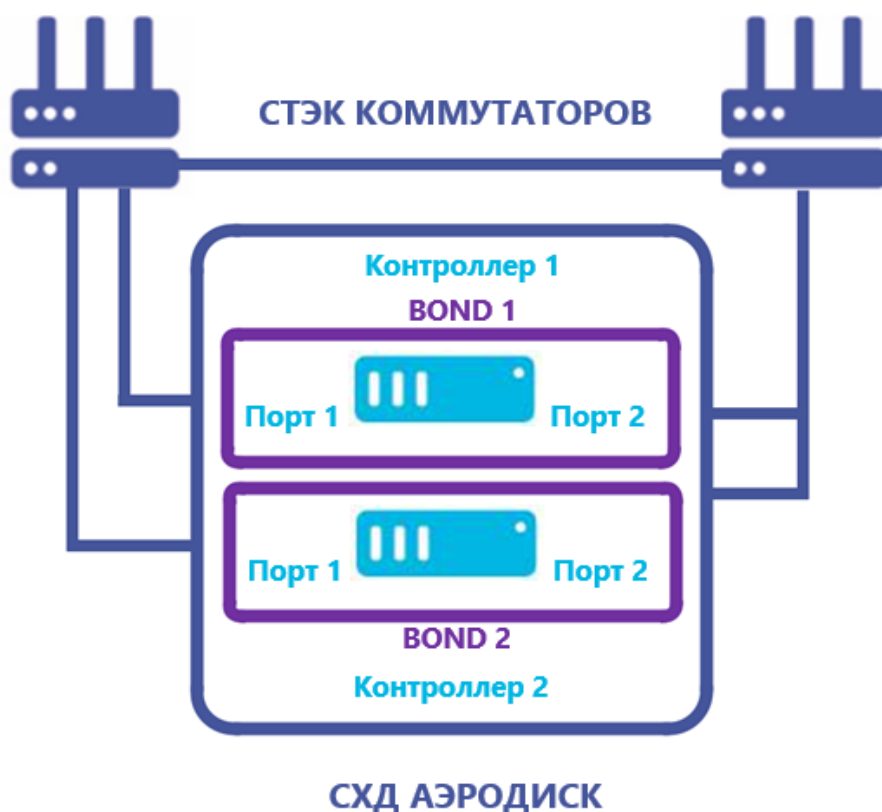
На рисунке ниже приведен пример логики работы многоуровневого хранения.



### 3.8 VLAN и BONDING

Для ускорения операций ввода/вывода можно задействовать функционал объединения нескольких физических портов в один логический порт-BOND интерфейс. Поддерживаются как независимые от настроек коммутаторов BOND-интерфейсы, так и зависимые от настроек коммутаторов BOND-интерфейсы. Объединение нескольких физических интерфейсов дает увеличение пропускной способности, а также повышает уровень отказоустойчивости, так как в рамках BOND-интерфейса физический порт может выйти из строя и обмен данными при этом не прекратится. В BOND-интерфейс можно объединить до 16 физических однотипных интерфейсов.

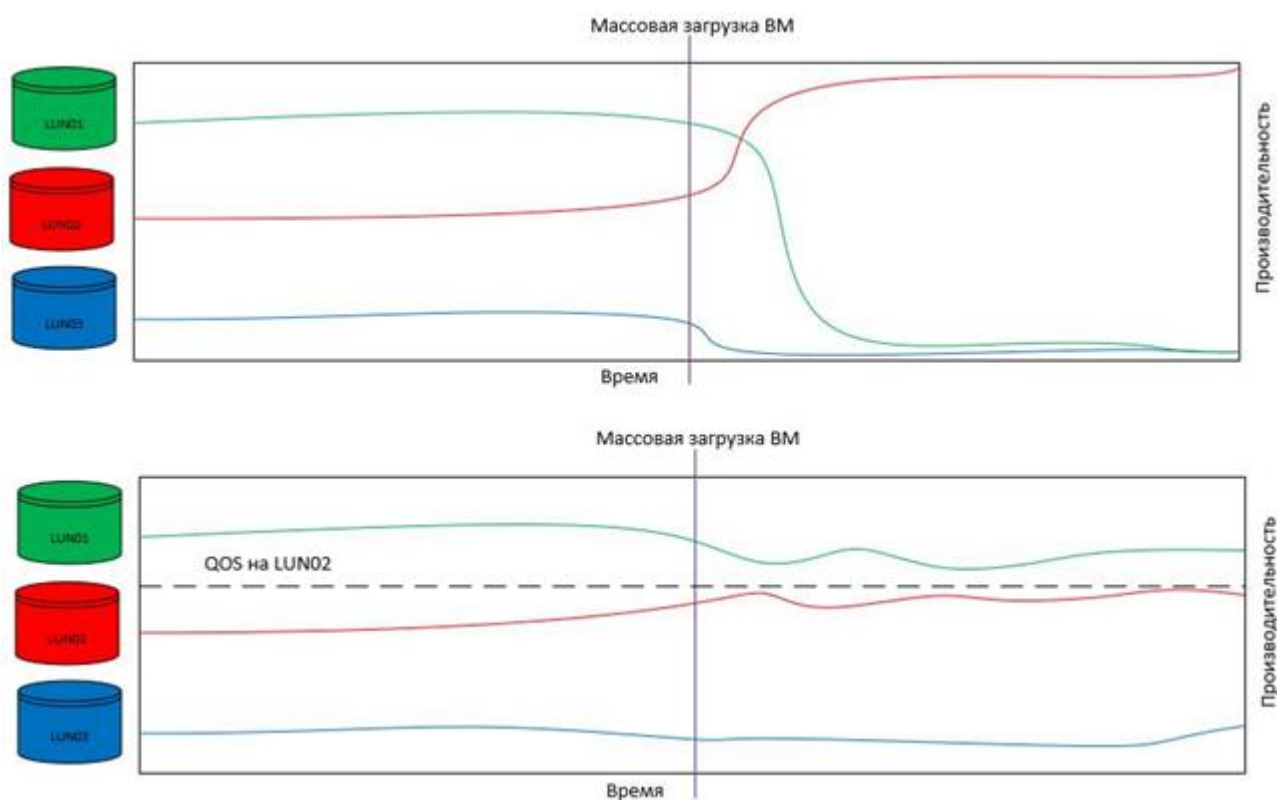
Для разграничения сетевого доступа, а также для более гибкой настройки СХД под сетевую инфраструктуру заказчика можно задействовать механизм тегирования трафика – VLAN. VLAN могут быть назначены как на физические сетевые интерфейсы, так и на BOND-интерфейсы. VLAN могут быть применены как для файловых шар для протоколов NFS/CIFS, так и для блочного доступа по iSCSI.



«ПОРТ 1» и «ПОРТ 2» объединены в BOND 2x10 Гбит/с LACP на контроллерах 1 и 2. VLAN – 10, 100, 1000 доступны на BOND-интерфейсах.

### 3.9 Quality of Service

Настройка качества обслуживания (QoS) позволяет минимизировать эффект «шумного соседа». При правильной настройке этого параметра можно гарантировать, что все потребители ресурсов СХД будут работать так, как ожидает администратор системы. QoS в СХД АЭРОДИСК назначаются на блочные устройства. Параметры качества обслуживания назначаются на уровне каждого конкретного LUN и могут ограничивать его потребление ресурсов в MB/s и IOPS. На картинке ниже приведен пример установки ограничений на LUN02, на котором по расписанию стартует множество VM, что является распространённым сценарием при использовании VDI.



### 3.10 Политика перестроения

При выходе из строя диска в RAID-группе автоматически начинается процесс ее перестроения. При перестроении, как правило, может страдать общая производительность СХД, то есть могут страдать конечные потребители ресурсов СХД. Чтобы минимизировать эффект от перестроения, администратор системы может назначить политику перестроения рейдов, в том числе заданную по расписанию. Для пулов RDG есть 3 варианта политики перестроения, чтобы можно было гранулярно управлять скоростью перестроения рейдов.

## Политика перестроения



Политика перестроения определяет, сколько ресурсов системы выделять на перестроение поврежденных дисковых групп. Выбор политики не влияет на производительность в штатном режиме работы и регулирует только поведение при перестроении дисковой группы.

Выберите желаемые политики перестроения и интервалы в течение дня в часах:

Оптимальная ⓘ +

Производительность ⓘ Начало: 9 - Конец: 18 ✖

Перестроение ⓘ Начало: 18 - Конец: 9 ✖

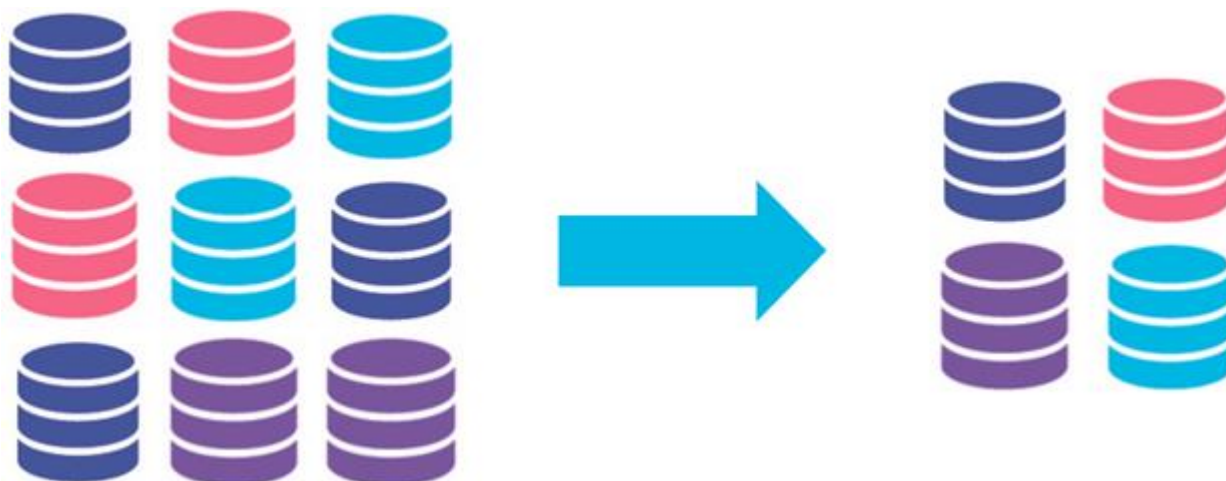
Отменить

Подтвердить

### 3.11 Дедупликация

Дедупликация – это процесс устранения дублей блоков данных при сохранении уникальных блоков для экономии дискового пространства.

На рисунке ниже приведен результат работы дедупликации.



Основное:

- в системах АЭРОДИСК применяется онлайн-дедупликация с фиксированным блоком;
- дедупликация работает на RDG, на SDD и на жестких дисках;



- для RDG можно включать на конкретный LUN;
- для файловых систем дедупликация и компрессия включаются на уровне RDG в целом;
- делит входящие данные на равные блоки;
- в СХД должен быть достаточный объем оперативной памяти для работы дедупликации;
- устраняет дубли только тогда, когда блоки на 100% совпадают.

Процесс дедупликации происходит следующим образом:

- снятие хэш-сумм с блоков данных;
- сравнение хэш-сумм;
- при условии совпадения контрольных сумм блоков данных система вместо выделения нового дискового пространства под дубль добавляет ссылку в таблицу дедупликации, которая указывает на реально существующие данные вместо того, чтобы создавать их дубли.

В зависимости от сферы применения и характера записи дедупликация может снизить потребляемый объем дискового пространства от 20% до 40%.

Дедупликация выполняется на блочном уровне, что особенно эффективно для больших объемов похожих данных. Например, при дедупликации хранилища виртуальных машин (VM) уникальными, как правило, являются только некоторые блоки данных, а идентичные данные, такие как гостевые ОС, шаблоны VM, клоны VM и пр. являются дублируемыми и, соответственно, при дедупликации не потребляют дополнительного объема.

Для 1 ТБ дедуплицируемых данных нужно резервировать не менее 1 Гб ОЗУ на хэш таблицу. Для SSD-дисков это незначительный объем, а вот для оперативной памяти – наоборот. Т.к. объем оперативной памяти СХД ограничен, то рекомендуется использовать дедупликацию при наличии SSD-дисков в СХД.

### 3.12 Компрессия

Для экономии места на СХД можно использовать механизм компрессии транзакций. Компрессия транзакций работает в онлайн-режиме, то есть данные записываются на диски уже в оптимизированном виде. Так как система оптимизирует размер хранимых данных еще до записи на диски, то в ряде случаев включение этой функции может увеличить количество операций ввода/вывода, так как физических записей/чтений на диски становится меньше.

Для выполнения компрессии транзакций «на лету» используются выделенные ядра процессора и процесс компрессии никогда не конкурирует за ресурсы. В случае если количество операций ввода/вывода велико и ресурсов выделенного ядра перестает хватать,

система «на лету» автоматически выделяет под процесс компрессии дополнительные выделенные ядра.

### 3.13 Авто-поддержка

Для обеспечения максимальной доступности систем хранения АЭРОДИСК предусмотрена функция автоматической поддержки. Данная опция обеспечивает:

- постоянный проактивный мониторинг всех компонентов СХД;
- автоматическую отправку диагностической информации в АЭРОДИСК в случае сбоя;
- автоматическое открытие обращений (тикетов) в АЭРОДИСК.

Открытие обращений производится путем отправки диагностической информации в виде почтовых уведомлений от контроллеров СХД на серверы поддержки АЭРОДИСК. После поступления данной информации сообщения автоматически преобразуются в тикет и регистрируются, далее специалист АЭРОДИСК, имея необходимую входную информацию, приступает к работе по устранению сбоя.

## 4. Правила лицензирования

В АЭРОДИСК ENGINE AQ Лайт лицензии на весь поддерживаемый функционал включены в базовую конфигурацию. Исключение составляет лицензия на подключение по протоколу Fibre Channel, которая приобретается вместе с соответствующими FC 16Gb адаптерами.

## 5. Поддерживаемые операционные системы и гипервизоры:

- Microsoft Windows Server 2012/2016/2019
- VMware 6.0/6.5/7.0
- Red Hat Enterprise Linux
- CentOS
- Debian 9/10/11
- Ubuntu 18/20
- zVirt 4.1
- AltLinux 8/9/10
- Astra Linux
- SUSE Linux
- Proxmox
- POCA
- БРЕСТ
- RedOS
- Tionix
- АЭРОДИСК vAIR

Указаны ОС, с которыми проводились тесты на момент публикации. Компания АЭРОДИСК постоянно проводит тестирование с новыми версиями ОС. Актуальное состояние уточняйте у технических специалистов производителя.

Совместимость с OEM-версиями операционных систем и сред виртуализации не гарантируется.