МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Тольяттинский государственный университет»

Кафедра Прикладная математика и информатика
(наименование)

09.04.03 Прикладная информатика
(код и наименование направления подготовки)

Управление корпоративными информационными процессами
(направленность (профиль))

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)

на тему Математическое и программное обеспечение системы управления хранением данных на основе RAID-массивов

Обучающийся М.В. Олексеенко
(Инициалы Фамилия) (личная подпись)

Научный руководитель д.т.н., доцент, С.В. Мкртычев
(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Тольятти 2025

Оглавление

Введение

Актуальность данной магистерской диссертации заключается в том, что ИТ-технологии стали неотъемлемой частью нашей жизни. Каждый день человечество генерирует экзабайты цифровых данных в виде разного рода контента. Это может быть видео снятое на телефон, финансовая транзакция и т.д. Любые обработанные данные представляют собой информацию, которая на сегодняшний день является самым ценным ресурсом в мире. Именно поэтому большинство современных компаний имеют целью получить способность хранить и обрабатывать огромные объемы данных, из которых в будущем они смогут получить финансовую выгоду.

Аппаратное решение, которое способно хранить и управлять большими объемами данных, называется системой хранения данных (СХД). Как и любой прибор, системы хранения данных имеют ограничения, обусловленные физическими параметрами самого оборудования. Как правило, для СХД это количество дисков, которые могут работать в рамках одной системы. В этой связи большинство систем хранения данных способны использовать алгоритмы компрессии и дедубликации данных. Данный функционал позволяет увеличить количество хранимой информации за счет того, что данные занимают меньшее пространство на накопителе. В то же время, пользователь должен понимать, что использование методов экономии емкости данных оказывает прямое влияние на производительность СХД и их использование может негативно влиять на работу бизнес процессов.

Объектом исследования в диссертации является корпоративный информационный процесс хранения данных.

Предметом исследования является система управления хранением данных.

Целью магистерской диссертации является разработка математического и программного обеспечения системы управления хранением данных на базе RAID-массивов.

Поставленная цель достигается за счет выполнения следующих задач:

– проанализировать современное состояние проблемы исследования математического и программного обеспечения систем управления хранением данных на основе RAID-массивов;
– разработать математическое и программное обеспечение системы управления хранением данных на базе RAID-массивов;
– выполнить апробацию и оценку эффективности проектных решений.

Гипотеза исследования заключается в том, что разработанное математическое и программное обеспечение системы управления хранением данных на базе RAID-массивов позволит повысить производительность и надежность хранения данных [2].

Научная новизна, полученная в ходе работы над магистерской диссертацией, заключается в разработке методики конфигурирования системы хранения данных, что позволит пользователю выбрать оптимальную СХД для решения поставленных перед ним задач.

На защиту выносятся:

– модель эффективной системы управления хранением данных на основе RAID-массивов;
– результаты проверки адекватности предлагаемой модели системы управления хранением данных.

По теме исследования опубликованы 2 статьи [15], [16].

Диссертация состоит из введения, трех глав, заключения, списка используемой литературы и используемых источников.

Работа изложена на 93 страницах и включает 56 рисунков, 32 таблицы и 30 источников.

Глава 1 Анализ современного состояния проблемы управления хранением данных

1.1 Типы RAID-систем и их характеристики

Изначально аббревиатура RAID расшифровывалась как -Redundant Arrays of Inexpensive Disks- (избыточные массивы недорогих дисков). Однако со временем значение изменилось: теперь «I» означает -Independent- (независимый), а не -Inexpensive-, что отражает эволюцию технологии в сторону повышения надежности и автономности дисковых систем.

Основная идея RAID заключается в объединении нескольких физических дисков в единый массив, что позволяет увеличить объем хранилища, ускорить чтение данных, обеспечить избыточность и, как следствие, повысить производительность и отказоустойчивость системы. Существует несколько уровней RAID, каждый из которых предлагает разную степень резервирования, контроля ошибок, емкости и стоимости.

Принцип работы RAID основан на распределенном хранении данных: доступное пространство разделяется на компьютерно-генерируемые тома, которые равномерно распределяются между дисками. Операции ввода-вывода выполняются на уровне блоков, а не отдельных дисков, что значительно ускоряет восстановление данных в случае сбоя [1].

Как показано на рисунке 1, RAID реализуется за счет параллельного чтения данных и дублирования информации на двух или более дисках. Для оптимизации передачи данных в основной памяти резервируются несколько буферов.

Кроме того, применение двойной буферизации позволяет сократить среднее время доступа к последовательным блокам диска.

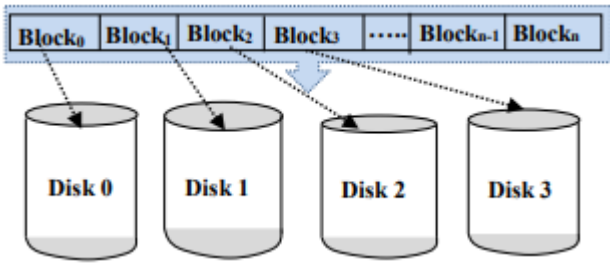

Рисунок 1 – Картографирование данные файл в другой диск

Методы RAID: Disk Striping и Mirroring.

Дисковое разделение (stripe) – это метод распределения данных, при котором информация разбивается на блоки (обычно от 32 КБ до 128 КБ) и записывается циклически на несколько физических носителей. В качестве устройств хранения могут использоваться не только жесткие диски, но и, например, ленточные накопители. Каждый сегмент данных размещается на отдельном диске, что позволяет ускорить операции чтения и записи за счет параллельной работы устройств.

Зеркалирование (или теневое копирование) – это метод повышения отказоустойчивости, при котором одни и те же данные дублируются на двух или более идентичных дисках. Логически эти диски воспринимаются как единое устройство. Если один из них выходит из строя, система продолжает работу с резервного носителя, а данные восстанавливаются после замены поврежденного диска.

Преимущества зеркалирования:

- высокая надежность и возможность восстановления данных;
- параллельное чтение с нескольких дисков увеличивает производительность.

Недостатки:

- запись данных на несколько дисков снижает скорость записи;
- высокая стоимость из-за необходимости дублирования дискового пространства.

Для снижения затрат иногда применяются альтернативные методы, использующие избыточные данные для восстановления информации при сбоях.

RAID 0 (страйпинг данных) – это технология, при которой данные разбиваются на блоки (стрипы) и распределяются по нескольким дискам в циклическом порядке. Поскольку дублирование не применяется, этот уровень обеспечивает максимальную скорость записи.

Особенности RAID 0:

– высокая производительность за счет параллельной работы дисков;
– простота реализации (поддерживается Windows, Linux, macOS).

Недостатки:

– отсутствие отказоустойчивости – при выходе из строя одного диска все данные теряются;
– не является «настоящим» RAID, так как не обеспечивает избыточности.

Пример работы показан на рисунке 2.

Данные разбиваются на блоки (Блок 0, Блок 1, Блок 2…).

Четные блоки записываются на один диск, нечетные – на другой.

RAID 0 подходит для задач, требующих высокой скорости, но не обеспечивает надежности хранения данных.

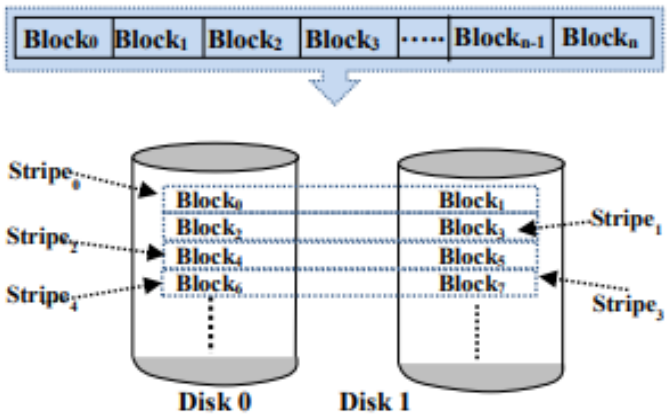

Рисунок 2 – RAID уровень 0 (Техника страйпинга дисков)

Таблица 1 демонстрирует распределение первых 28 блоков в 4-дисковой системе RAID 0 с использованием 7 стрипов (обозначения: B — блок, S — стрип).

Таблица 1 – Полная Stripe из 5-дисковый массив в RAID. Уровень 0

Stripe #	Диск 0	Диск 1	Диск 2	Диск 3
С 0	Б 0	Б 1	Б 2	Б 3
С 1	Б 4	Б 5	Б 6	Б 7
С 2	Б 8	Б 9	Б 10	Б 11
С 3	Б 12	Б 13	Б 14	Б 15
С 4	Б 16	Б 17	Б 18	Б 19
С 5	Б 20	Б 21	Б 22	Б 23
С 6	Б 24	Б 25	Б 26	Б 27

RAID 1 – это уровень избыточности, основанный на полном дублировании данных. В этой схеме каждый файл записывается одновременно на два (или более) идентичных диска, создавая точные копии информации. Если исходный диск заполняется, система позволяет добавлять новые диски парами, сохраняя принцип зеркалирования.

Принцип работы:

- данные синхронно записываются на основной и зеркальный диски;
- при отказе одного носителя система автоматически переключается на резервный, предотвращая потерю данных.

Преимущества RAID 1:

- 100% избыточность – данные сохраняются даже при выходе из строя одного диска;
- повышенная надежность – идеален для критически важных систем;
- ускоренное чтение – данные можно считывать параллельно с обоих дисков.

Недостатки:

- высокая стоимость – требует вдвое больше дискового пространства;
- снижение скорости записи – данные должны записываться на оба

диска одновременно.

RAID 1 обеспечивает максимальную отказоустойчивость, но за счет увеличения затрат на хранилище. Он оптимален для задач, где надежность важнее экономии ресурсов.

На рисунке 3 показана схема, где информация полностью дублируется между двумя дисками.

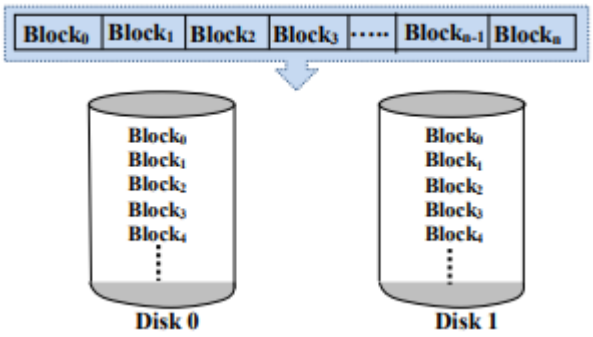

Рисунок 3 – Блокировка шаблона RAID уровня 1

Хотя RAID 1 позволяет параллельное чтение данных, его производительность записи снижается из-за необходимости одновременной записи одинаковых данных на две пары дисков (основную и резервную). По сравнению с RAID 0, этот уровень требует вдвое больше дискового пространства для хранения того же объема данных. Однако он обеспечивает значительно более высокую надежность, так как система продолжает работать при сохранении хотя бы одного исправного диска]. Как и в случае с простым зеркалированием, это решение удваивает затраты на хранение данных.

RAID 2, несмотря на внешнее сходство с RAID 0, имеет принципиальное отличие в организации данных. Если в RAID 0 используется чередование блоков, то в RAID 2 применяется битовое чередование с дополнительным использованием кода коррекции ошибок (ECC) для контроля целостности данных. Как показано на рисунке 4, данные разбиваются на мелкие сегменты, для которых вычисляются биты четности, после чего и данные, и контрольная

информация распределяются по дискам. При возникновении ошибки на любом диске, информация восстанавливается с использованием дисков четности.

Преимущество RAID 2 перед RAID 1 заключается в более эффективном механизме исправления ошибок. Однако этот уровень имеет проблемы с производительностью: операции чтения требуют одновременного доступа как к данным, так и к ECC-кодам с проверочных дисков. Ситуация усложняется при записи, когда необходимо синхронно обновлять данные, проверочные диски и диски четности.

Дополнительная проблема RAID 2 связана с ограничением на параллельное чтение нескольких файлов из-за битового чередования данных. Это существенный недостаток, требующий поиска оптимальных решений для конкретных задач хранения информации (рисунок 4).

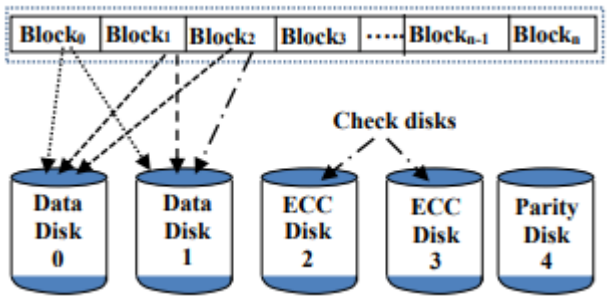

Рисунок 4 – Блокировка шаблона RAID уровня 2

Для преодоления ограничений предыдущих уровней и поддержки параллельного чтения с дополнительными функциями была разработана технология RAID Level-3. В этой конфигурации данные распределяются по нескольким дискам, а отдельный диск выделяется исключительно для хранения битов четности. Современные диски с встроенными кодами коррекции ошибок (ECC) автоматически обнаруживают поврежденные секторы и восстанавливают данные с помощью диска четности, используя информацию с остальных дисков массива.

Для решения проблем параллельной работы, характерных для уровней 2 и 3, были созданы RAID Level-4 и Level-5. Благодаря экспоненциальному росту производительности процессоров и памяти, RAID 4 обеспечивает возможность параллельного чтения с существенным увеличением скорости по сравнению с предыдущими уровнями.

На рисунке 5 представлена архитектура 5-дискового массива RAID 4, где:

– «S» обозначает полосу данных (стрип);
– блоки в одной строке образуют полосу (например, S0-S7 - первые восемь полос).

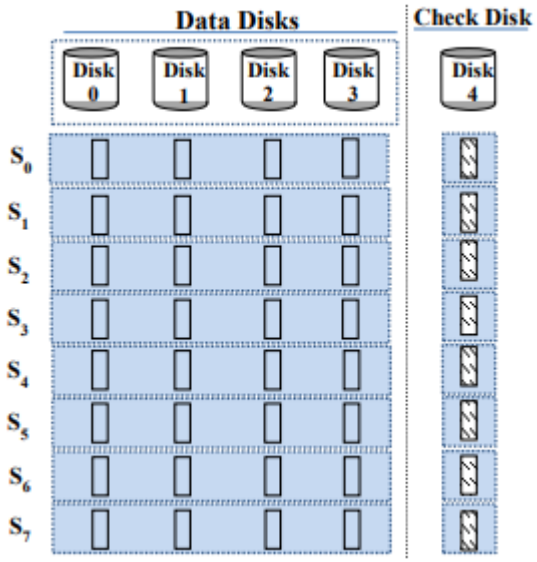

Рисунок 5 – Блокировка шаблона уровня RAID- 4

Однако RAID 4 имеет существенный недостаток - необходимость сложных вычислений четности RAID-контроллером делает операции записи относительно медленными. Это ограничение привело к разработке более совершенного RAID Level-5, где блоки четности распределяются по всем дискам массива, устраняя узкое место в виде выделенного диска для четности.

Таблица 2 демонстрирует распределение первых 32 блоков RAID-4 с обозначениями:

– «S» – полоса;

– «B» – блок данных;

– «P» – блок четности.

Таблица 2 – Полная Stripe из 5-дискового множества в RAID-4

Stripe #	Disk0	Disk1	Disk2	Disk3	Disk4
S0	B0	B1	B2	B3	P0
S1	B4	B5	B6	B7	P1
S2	B8	B9	B10	B11	P2
S3	B12	B13	B14	B15	P3
S4	B16	B17	B18	B19	P4
S5	B20	B21	B22	B23	P5
S6	B24	B25	B26	B27	P6
S7	B28	B29	B30	B31	P7

Для преодоления ограничений RAID-4 была разработана усовершенствованная технология RAID Level-5, основанная на принципе распределенной информации о четности.

В отличие от RAID-4, где данные проверки концентрировались на одном диске, в RAID-5 блоки четности равномерно распределяются по всем дискам массива (рисунок 6).

Это решение представляет собой блочное чередование (stripe) с децентрализованным хранением контрольных сумм.

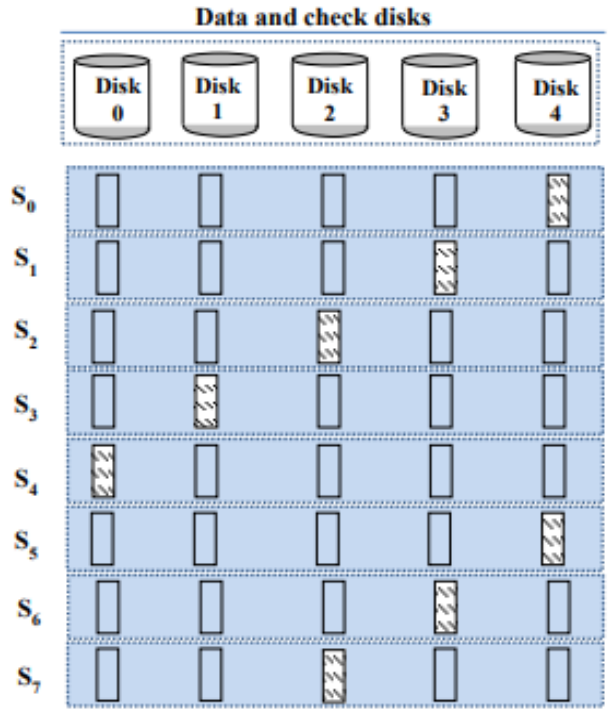

Рисунок 6 – RAID-5 эволюционировал из RAID-4

Таблица 3 наглядно демонстрирует организацию первых 32 блоков в 5-дисковом массиве RAID-5 с восемью стрипами, где:

– «В» обозначает блоки данных;

– «Р» обозначает блоки четности.

Таблица 3 – Полный Stripe из 5-дискового множества в RAID-5

Stripe #	Диск 0	Диск 1	Диск 2	Диск 3	Диск 4
S 0	Б 0	Б 1	Б 2	Б 3	П 0
S 1	Б 5	Б 6	Б 7	П 1	Б 4
S 2	Б 10	Б 11	П 2	Б 8	Б 9
S 3	Б 15	П 3	Б 12	Б 13	Б 14
S 4	П 4	Б 16	Б 17	Б 18	Б 19
S 5	Б 20	Б 21	Б 22	Б 23	П 5
S 6	Б 25	Б 26	Б 27	П 6	Б 24
S 7	Б 30	Б 31	П 7	Б 28	Б 29

Ключевые преимущества RAID-5 включают:

– восстановление данных: при отказе одного диска информация

восстанавливается с использованием распределенных данных четности;

– требования к оборудованию: минимальная конфигурация требует 3 дисков, причем емкость одного диска резервируется для обеспечения отказоустойчивости.

Технология Rotated Parity в RAID-5 эффективно устраняет «узкое место» RAID-4, связанное с использованием выделенного диска для четности. Однако важно отметить, что RAID-5 обеспечивает защиту только от единичного отказа диска. В случае повреждения накопителя:

– операции чтения выполняются с использованием распределенной проверочной информации;

– после замены диска все данные полностью восстанавливаются.

RAID Level-5 представляет собой усовершенствованную версию RAID Level-4, обеспечивающую более эффективную обработку операций записи за счет оптимального распределения блоков четности по всем дискам массива. Хотя прирост производительности для операций чтения остается умеренным, технология значительно улучшает обработку множественных запросов на запись благодаря параллельным методам работы и сбалансированной нагрузке между дисками.

RAID 5 заслуженно считается наиболее распространённым и надёжным решением среди уровней RAID. Эта технология, как и RAID 4, относится к категории систем с избыточностью на основе чётности, требуя минимум три диска и поддерживая массивы до 16 накопителей.

Развитием этой концепции стал RAID-6, который сохраняет принципы чередования дисков и распределённой чётности, но выводит отказоустойчивость на новый уровень благодаря использованию двух независимых блоков чётности для каждой полосы данных. Такая архитектура обеспечивает беспрецедентную защиту информации - система продолжает работать даже при одновременном отказе двух дисков, что делает её идеальным выбором для работы с конфиденциальными данными.

Для развёртывания RAID-6 требуется как минимум четыре диска. В этой системе применяются две группы блоков чётности, обозначаемые как P (P0, P1, P2...) и Q (Q0, Q1, Q2...). Хотя необходимость обработки дополнительных контрольных сумм несколько снижает скорость записи по сравнению с RAID 5, эта технология обеспечивает исключительную надёжность - данные остаются доступными даже при потере трети дискового массива. При этом сохраняются высокие показатели скорости чтения и эффективность обработки множества небольших операций ввода-вывода, что делает RAID 6 оптимальным решением для критически важных систем, где на первом месте стоит гарантированная сохранность данных (рисунок 7).

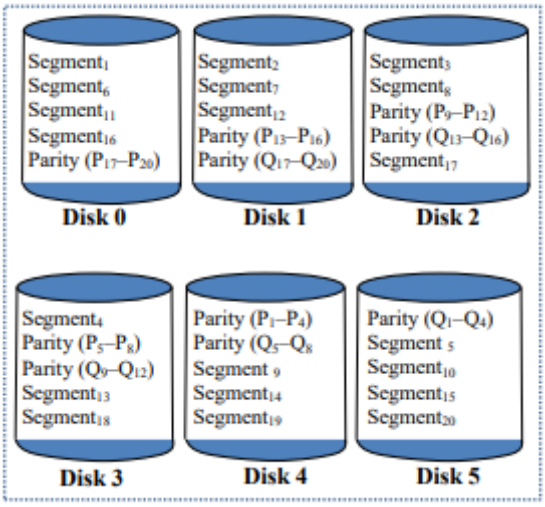

Рисунок 7 – Блокировка шаблона RAID- 6

RAID Level-10 (также известный как RAID 1+0) представляет собой гибридную технологию, объединяющую преимущества RAID 0 (страйпинг) и RAID 1 (зеркалирование). Эта архитектура, ранее называвшаяся «вложенный RAID», обеспечивает одновременно высокую производительность и надежную отказоустойчивость.

В основе RAID-10 лежит принцип использования двух идентичных массивов RAID 0, которые содержат полные дублирующие копии данных (рисунок 8).

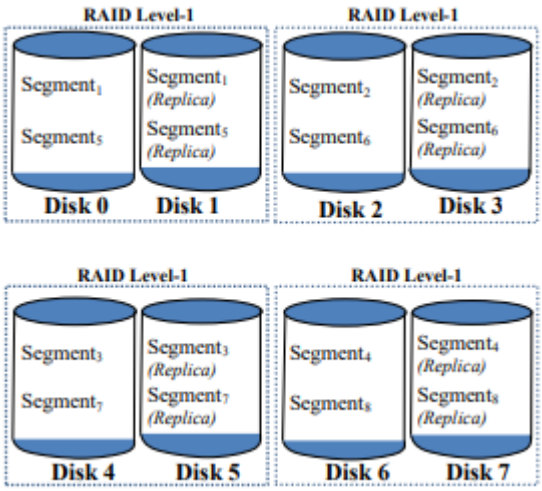

Рисунок 8 – Блокировка шаблона RAID-10

RAID Level-60 представляет собой виртуальный дисковый массив, объединяющий преимущества RAID-0 (страйпинг) и RAID-6 (двойная четность) в единую систему. Эта технология сочетает распределенное хранение данных с повышенной отказоустойчивостью, обеспечивая как высокую производительность, так и надежную защиту информации.

«Как показано на рисунке 9, RAID-60 реализуется на основе двух идентичных массивов RAID-6, между которыми равномерно распределяются данные. Каждая группа дисков хранит полную копию информации, включая два независимых набора блоков четности (обозначенных как P и Q). Это позволяет системе выдерживать одновременный отказ любых двух дисков в каждой группе RAID-6 без потери данных, сохраняя работоспособность даже при потере до половины накопителей в массиве» [30].

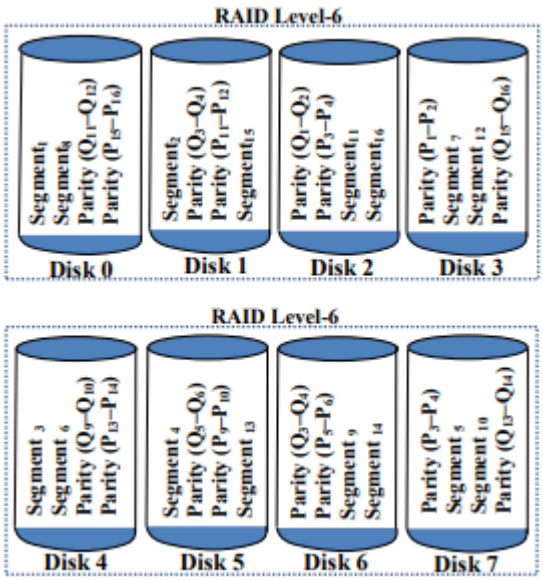

Рисунок 9 – Блокировка шаблона RAID-60

Ключевые особенности RAID-60:

– двойная защита четности – данные остаются доступными при множественных сбоях благодаря распределенным контрольным суммам;
– параллельная обработка – страйпинг ускоряет операции чтения/записи за счет одновременного использования всех дисковых групп;
– масштабируемость – поддерживает крупные хранилища с высокой нагрузкой, подходящие для корпоративных сред.

На схеме видно, что данные разбиваются на блоки и равномерно распределяются между группами RAID-6, а четность рассчитывается отдельно для каждого подмассива. Такой подход минимизирует риски потери информации и обеспечивает гибкость при восстановлении после сбоев.

Помимо основных конфигураций, существуют и другие, менее распространенные уровни RAID, представленные в Таблице 4.

Эти варианты встречаются реже либо применяются в узкоспециализированных сценариях.

Таблица 4 – Другие уровни RAID

RAID - уровень	Описание
Уровень- 00	Это обеспечивает преимущества из RAID-0 такой как независимый диски и полосатый.
Уровень -01	Уровень-01 или (RAID 0 + 1) поддерживает комбинацию из два большой массивы из (RAID- 0) что являются отражено (RAID-1).
Уровень- 15	По своей концепции он аналогичен RAID-10, но чередование процесс является выполненный с паритет чеки.
Уровень -30	Это поддерживает сочетание из оба RAID -0 и RAID- 3.
Уровень- 50	Это поддерживает сочетание из оба RAID -0 и RAID-5

Некоторые компании разрабатывают собственные модификации RAID, комбинируя различные технологии для достижения специфических преимуществ – например, повышенной отказоустойчивости, скорости или экономии ресурсов [17].

Для обеспечения максимальной надежности системы рекомендуется использовать горячие резервные диски, которые автоматически подключаются при выходе из строя основного накопителя. Такие диски немедленно начинают реконструкцию данных без необходимости остановки системы или вмешательства пользователя, существенно сокращая время простоя. Эти автономные накопители, известные как JBOD (Just a Bunch of Disks), функционируют как независимые элементы вне структуры RAID, обеспечивая дополнительный уровень защиты критически важных данных.

Существует два основных метода организации RAID-массивов:

– программная реализация использует многоуровневую архитектуру, где каждый уровень предоставляет сервисы вышестоящему уровню, скрывая технические детали реализации. Такой подход, поддерживаемый ОС Windows Server и Mac OS, идеален для простых конфигураций (RAID 0, 1) и основан на комбинации методов чередования данных и вычисления избыточной информации;
– аппаратное решение предполагает использование специализированных контроллеров с дополнительной кэш-памятью. Несмотря на более высокую стоимость, этот вариант обеспечивает

лучшую производительность для сложных конфигураций (RAID 5, 6) и может быть интегрирован непосредственно в материнскую плату.

RAID не заменяет систему резервного копирования. Эти технологии дополняют друг друга в комплексной стратегии защиты данных:

- резервные копии остаются единственным решением при вирусных атаках, физических повреждениях (наводнения, пожары) или краже оборудования;
- RAID не защищает от человеческих ошибок (случайное удаление файлов) или целенаправленного взлома системы;
- обнаружение потери данных может произойти с задержкой, и только архивные копии позволят восстановить информацию;
- идеальная стратегия включает географически распределенные резервные копии в сочетании с локальным RAID-массивом;
- хотя система может функционировать без RAID, отсутствие резервных копий создает критическую уязвимость.

Таким образом, RAID обеспечивает непрерывность работы и защиту от аппаратных сбоев, тогда как резервное копирование гарантирует долгосрочную сохранность данных в любых сценариях. Оптимальное решение сочетает оба подхода с учетом специфики конкретной информационной системы.

1.2 Аппаратное и программное обеспечение систем управления хранением данных на основе RAID-массивов

«Избыточный массив независимых дисков (RAID) представляет собой технологию, предназначенную для достижения двух ключевых целей: повышения надежности хранения данных и увеличения производительности операций ввода-вывода. При объединении нескольких физических накопителей в RAID-массив они функционируют как единое логическое устройство с точки зрения операционной системы и конечного пользователя.

Распределение данных между дисками осуществляется RAID-адаптером в соответствии с выбранным уровнем организации массива» [8].

RAID может быть реализован как на аппаратном, так и на программном уровне.

Аппаратный RAID базируется на специализированном контроллере, который может быть интегрирован в материнскую плату или выполнен в виде отдельного PCI-устройства. Данный подход обеспечивает высокую отказоустойчивость за счет использования интеллектуального контроллера и избыточного хранения данных. Аппаратные решения поддерживают различные уровни RAID, выбор которых осуществляется пользователем через графический интерфейс системы хранения.

Программный RAID реализуется средствами операционной системы и, как правило, поддерживает ограниченный набор уровней (чаще всего RAID 0 и RAID 1). Несмотря на отсутствие необходимости в специализированном оборудовании, программные решения уступают аппаратным в производительности и надежности.

Наиболее распространенными уровнями RAID являются:

- RAID 0 (страйпинг) – обеспечивает высокую производительность за счет распределения данных между дисками без избыточности. Минимальное количество накопителей – 2. Не подходит для критически важных систем из-за отсутствия отказоустойчивости;
- RAID 1 (зеркалирование) – дублирует данные на нескольких дисках, обеспечивая высокую доступность и защиту от сбоев. Оптимален для workloads с преобладанием операций чтения. Требует минимум 2 диска;
- RAID 5 использует распределенную четность, что позволяет восстанавливать данные при отказе одного диска. Минимальная конфигурация – 3 накопителя;
- RAID 6 – аналогичен RAID 5, но с двойной четностью, что повышает отказоустойчивость (минимум 4 диска).

– RAID 10 (1+0) – комбинация зеркалирования и страйпинга, обеспечивающая высокую производительность и надежность (минимум 4 диска);
– RAID 50 (5+0) и RAID 60 (6+0) – гибридные уровни, сочетающие страйпинг с распределенной четностью, требующие 6 и 8 дисков соответственно.

Выбор уровня RAID зависит от требований к производительности, отказоустойчивости и экономической эффективности. Аппаратные решения предпочтительны для высоконагруженных систем, тогда как программные могут использоваться в менее критичных сценариях [20].

Архитектура RAID-системы состоит из следующих компонентов:

Физические диски (hard drives): это основные носители данных, которые объединяются в массивы RAID. Физические диски могут быть обычными жесткими дисками (HDD) или твердотельными накопителями (SSD).

«Аппаратное обеспечение включает в себя специализированные RAID-контроллеры, которые обеспечивают аппаратную реализацию RAID-уровней и функций управления массивом. Они могут быть представлены в виде отдельных устройств или встроенных в серверы и хранилища данных. Аппаратные RAID-контроллеры имеют свои собственные процессоры, кэш-память и порты подключения дисков. Они обеспечивают аппаратную обработку операций RAID, таких как расчет контрольной суммы и восстановление данных» [13].

Преимущества использования аппаратных RAID-контроллеров включают высокую производительность, низкую нагрузку на центральный процессор сервера, возможность независимой работы от операционной системы и поддержку специфических функций, таких как кэширование чтения/записи, аппаратное ускорение шифрования и другие.

Программное обеспечение для управления RAID-массивами может быть реализовано на уровне операционной системы или представлять собой специализированное программное решение. В некоторых случаях,

программное обеспечение RAID может быть встроено непосредственно в аппаратный RAID-контроллер [27].

Программное обеспечение RAID, работающее на уровне операционной системы, предоставляет драйверы и утилиты для управления и мониторинга RAID-массивами. Оно позволяет конфигурировать массивы, выполнять операции обслуживания, мониторить состояние массива и предоставлять интерфейс для взаимодействия с другими приложениями и сервисами.

Специализированное программное обеспечение RAID может предоставлять расширенные функции управления и мониторинга массивов. Оно может включать графические интерфейсы для конфигурирования и управления, возможности автоматического обнаружения и восстановления отказавших дисков, а также функции мониторинга состояния массива и оповещения о сбоях [28].

На рынке существует множество программных решений для управления RAID-массивами, которые предоставляют широкий спектр функций и возможностей. Некоторые из наиболее популярных программных решений включают [29], [30]:

- Dell OpenManage Storage Manager: предоставляет централизованный интерфейс для управления и мониторинга RAID-массивами, включая функции конфигурирования, мониторинга состояния и оповещения;
- HP Smart Storage Administrator: позволяет управлять массивами HP Smart Array и предоставляет функции мониторинга, конфигурирования и обслуживания, такие как пересборка массива и замена дисков;
- IBM ServeRAID Manager: предоставляет возможность управления и мониторинга массивами IBM ServeRAID и функции, такие как конфигурирование, пересборка и расширение массива.

Общая схема взаимосвязи между компонентами архитектуры RAID-системы может быть представлена следующим образом (рисунок 10):

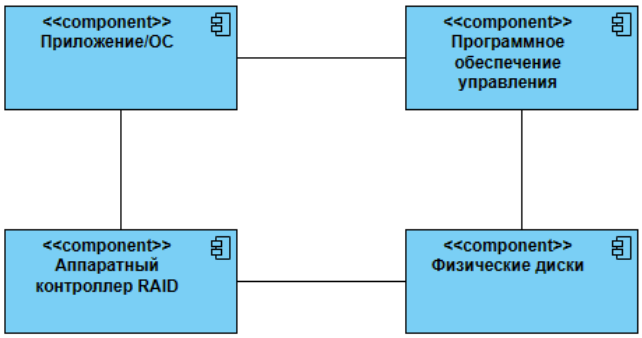

Рисунок 10 – Архитектура системы хранения данных на основе RAID-массивов

Программное обеспечение управления позволяет настраивать параметры и свойства массивов, мониторить состояние дисков и массивов, а также выполнять операции обслуживания, такие как пересборка массива, замена дисков, расширение массива и т. д. Приложение или операционная система взаимодействуют с программным обеспечением управления для доступа к данным на RAID-массиве.

1.3 Методологические основы построения систем управления хранением данных на основе RAID-массивов

«Система хранения данных предназначена для организации надежного хранения данных, а также отказоустойчивого, высокопроизводительного доступа серверов к устройствам хранения. Существующие в настоящее время методы по обеспечению надежного хранения данных и отказоустойчивого доступа к ним можно охарактеризовать одним словом – дублирование.

Так, для защиты от отказов отдельных дисков используются технологии RAID, которые (кроме RAID-0) применяют дублирование данных, хранимых на дисках. Уровень RAID-5 хотя и не создает копий блоков данных, но все же сохраняет избыточную информацию, что тоже можно считать дублированием. Для защиты от логического разрушения данных (разрушение целостности

базы данных или файловой системы), вызванных сбоями в оборудовании, ошибками в программном обеспечении или неверными действиями обслуживающего персонала, применяется резервное копирование, которое тоже является дублированием данных. Для защиты от потери данных вследствие выхода из строя устройств хранения по причине техногенной или природной катастрофы, данные дублируются в резервный центр» [30].

Отказоустойчивость доступа серверов к данным достигается дублированием путей доступа (рисунок 11).

Рисунок 11 – Конфликт целей

«Применительно к SAN дублирование заключается в следующем: сеть SAN строится как две физически независимые сети, идентичные по функциональности и конфигурации. В каждый из серверов, включенных в SAN, устанавливается как минимум по два FC-HBA. Первый из FC-HBA подключается к одной «половинке» SAN, а второй – к другой. Отказ оборудования, изменение конфигурации или регламентные работы на одной из частей SAN не влияют на работу другой. В дисковом массиве отказоустойчивость доступа к данным обеспечивается дублированием RAID-контроллеров, блоков питания, интерфейсов к дискам и к серверам. Для защиты от потери данных зеркалируют участки кэш-памяти, участвующие в операции записи, а электропитание кэш-памяти резервируют батареями. Пути

доступа серверов к дисковому массиву тоже дублируются. Внешние интерфейсы дискового массива, включенного в SAN, подключаются к обоим ее «половинкам». Для переключения с вышедшего из строя пути доступа на резервный , а также для равномерного распределения нагрузки между всеми путями, на серверах устанавливается специальное программное обеспечение, поставляемое либо производителем массива (EMC CLARiiON – PowerPath, HP EVA – AutoPath, HDS – HDLM), либо третьим производителем (VERITAS Volume Manager)» [8].

«Необходимую производительность доступа серверов к данным можно обеспечить созданием выделенной высокоскоростной транспортной инфраструктуры между серверами и устройствами хранения данных (дисковым массивом и ленточными библиотеками). Для создания такой инфраструктуры в настоящее время наилучшим решением является SAN. Использование современных дисковых массивов с достаточным объемом кэш-памяти и производительной, не имеющей «узких мест» внутренней архитектурой обмена информацией между контроллерами и дисками, позволяет осуществлять быстрый доступ к данным. Оптимальное размещение данных (disk layout) по дискам различной емкости и производительности, с нужным уровнем RAID в зависимости от классов приложений (СУБД, файловые сервисы и т.д.), является еще одним способом увеличения скорости доступа к данным» [3].

Следует подчеркнуть, что грамотная настройка программного обеспечения - как приложений, так и операционной системы - часто дает более значительный прирост производительности, чем простое обновление аппаратного обеспечения. Это объясняется тем, что оптимизация программных параметров позволяет устранить узкие места в обработке данных, тогда как замена оборудования лишь незначительно расширяет пропускную способность системы (хотя в некоторых случаях этого бывает достаточно). Для эффективной оптимизации может применяться специализированное программное обеспечение, учитывающее особенности

взаимодействия между аппаратной частью, ОС и прикладными программами. Ярким примером является технология Quick I/O файловой системы VxFS, доступная в составе пакета VERITAS DataBase Edition for ORACLE. Эта технология позволяет СУБД Oracle использовать асинхронный ввод-вывод на уровне ядра (KAIO), что значительно повышает производительность операций с данными.

Помимо обеспечения требуемых показателей производительности, отказоустойчивости и надежности хранения данных, важной задачей является минимизация совокупной стоимости владения (TCO) системой хранения данных. Внедрение современных систем управления позволяет сократить операционные расходы на администрирование и более эффективно планировать модернизацию оборудования. Консолидация технических ресурсов также способствует снижению эксплуатационных затрат.

При создании системы хранения данных необходимо найти оптимальный баланс между:

- производительностью;
- доступностью (надежностью хранения и отказоустойчивостью);
- совокупной стоимостью владения.

Одним из распространенных методов обеспечения высокой доступности является дублирование компонентов, что неизбежно увеличивает стоимость системы. Без учета реальных бизнес-требований к доступности это может привести к неоправданному удорожанию. Аналогично, погоня за избыточной производительностью ведет к использованию дорогостоящего оборудования. Помимо этих факторов, необходимо обеспечить требуемые функциональные характеристики: объем хранилища и количество поддерживаемых серверов.

На практике заказчики часто затрудняются сформулировать требования к производительности в общепринятых количественных показателях (пропускная способность в МБ/с или количество операций ввода-вывода в секунду - IOPS). Поэтому проектировщику важно определить, какие именно приложения будут работать в системе, чтобы понять характер нагрузки [3], [4].

Разные типы приложений создают различную нагрузку на систему хранения. Например, СУБД могут работать в двух основных режимах:

- OLTP (онлайн-обработка транзакций) – интенсивный поток запросов на ввод-вывод небольших порций данных;
- DSS (системы поддержки решений) – небольшое количество запросов на чтение крупных блоков информации.

Характер нагрузки напрямую влияет на выбор методов распределения данных и объема кэш-памяти. Для OLTP-систем кэширование может значительно повысить производительность, тогда как для DSS-задач, работающих с большими объемами данных, эффективность кэша существенно ниже, так как вероятность повторного обращения к тем же данным минимальна.

Аналогичные закономерности наблюдаются и для других типов приложений:

- почтовые системы на базе MS Exchange или Lotus Domino создают нагрузку, аналогичную OLTP;
- почтовые сервисы на базе sendmail нагружают систему частыми изменениями метаданных;
- системы резервного копирования работают с последовательным чтением данных, подобно DSS-задачам.

Особый тип нагрузки характерен для процессов журналирования (журналы транзакций СУБД, системные журналы ОС) и загрузки данных в хранилища. В этих случаях кэш-память для записи не дает значительного прироста производительности, так как при обработке больших объемов данных или записи в новые области кэш быстро заполняется, и системе приходится освобождать место, сбрасывая данные на диск перед обработкой новых запросов (рисунок 12).

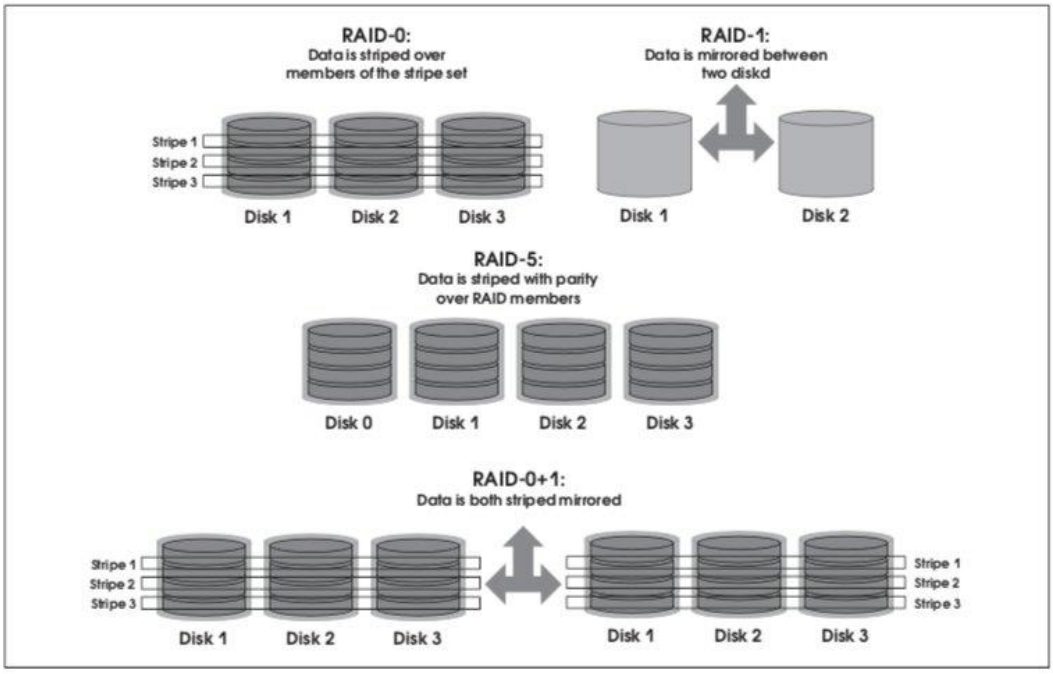

Рисунок 12 – Уровни RAID

Определить классы задач, которые будет обслуживать СХД, необходимо не только для выработки политики работы с кэш, но также для распределения данных по дискам (disk layout). Для обеспечения надежного хранения информации диски организуются в уровни RAID 1, 0+1/1+0 или 5. Самым экономичным с точки зрения использования дополнительного (дублирующего) дискового пространства является уровень RAID 5. Однако производительность RAID 5 ниже, чем у RAID 1 или 0+1 при частых случайных изменениях данных, характерных для OLTP-задач, и высока при считывании данных, что характерно для DSS-задач.

Также разные уровни RAID обеспечивают различные уровни отказоустойчивости дисковой памяти к сбоям отдельных дисков. Так RAID 5 не спасает от двух последовательных отказов дисков, впрочем, как и RAID 0+1, если это диски разных половин «зеркала». Наиболее отказоустойчивым является уровень RAID 1+0 (попарное «зеркалирование» дисков и затем их «striping4»), поэтому его рекомендуется применять для хранения критичных данных, например, журналов транзакций СУБД. Практика показывает, что для хранения файловых систем и данных DSS-задач достаточно использовать

RAID 5. Однако, если файловая система часто изменяется как, например, в почтовых системах sendmail, то имеет смысл использовать журналированную файловую систему или файловую систему с отдельно хранимыми метаданными, например, файловую систему Sun QFS. Тогда для хранения журналов или отдельных метаданных лучше применять RAID 1 или 1+0.

Для «больших» систем актуальной становится оптимизация настроек СХД, направленная на повышение быстродействия для достижения заданного уровня сервиса. Под «большой» понимается такая система, в которой обрабатывается значительный объем данных – единицы и десятки терабайт, и/или к СХД подключены десятки серверов. Для небольших систем проблемы с быстродействием можно решить применением более производительной аппаратуры. В «больших» системах такой подход может оказаться неприемлемым либо в связи с тем, что потребуется очень дорогая аппаратура, либо в связи с тем, что уже достигнут предел аппаратной производительности. Единственным решением в данном случае является оптимизация. Для оптимизации производительности СХД желательно иметь возможность управлять настройками на всем пути следования данных от процессора к дискам и обратно. Для СУБД ORACLE такая оптимизация заключается в возможности использовать KAIO, а также возможности отключить кэш файловой системы для файлов данных ORACLE (поскольку СУБД ORACLE кэширует данные в собственной области памяти SGA) [14]. Для этой цели можно использовать упомянутый ранее пакет VERITAS DBE.

Если в системе эксплуатируются OLTPи DSS-задачи (что является типичным для большинства систем), то для оптимизации производительности дискового массива желательно иметь возможность управлять настройками кэш-памяти для каждого логического диска (LUN) в отдельности. Это необходимо для того, чтобы для тех дисков, где расположены данные OLTP-задачи, использовать кэш (и желательно большого объема), а для дисков с данными DSS-задачи использование кэшпамяти отключить. Однако, если для OLTPи DSS-задач используются одни и те же таблицы данных, то такие

настройки теряют смысл до тех пор, пока не будет решен вопрос о физическом разнесении данных задач по разным дискам, а выполнение самих задач перенесено на разные серверы. Это можно реализовать средствами СУБД, например, с помощью экспорта данных в файл и импорта их в другую базу [21]. Если объем данных велик и синхронизацию баз данных OLTPи DSS-задач надо проводить достаточно часто, этот вариант может оказаться неэффективным. Здесь может помочь технология создания PITкопий данных, реализованная в большинстве современных дисковых массивов.

Выводы по главе 2

Выше говорилось, что СХД является важным звеном в обеспечении заданного уровня сервисов, предоставляемых информационной системой.

Уровень сервиса зависит не только от производительности СХД, вопросы обеспечения которой только что обсуждались, но также от готовности и надежного хранения данных, другими словами, от доступности СХД.

Исходя из бизнес-требований к информационной системе, необходимо определить режимы ее работы (24x7, 12x5 и т.п.), степень критичности данных в зависимости от степени критичности сервисов, использующих эти данные, требования к готовности и надежности данных в зависимости от степени их критичности и режимов работы системы.

Глава 2 Анализ и выбор методологии построения эффективной системы управления хранением данных на основе RAID-массивов

2.1 Обзор и анализ технологий RAID, их преимущества и недостатки

Избыточный массив независимых дисков (RAID) включает в себя две ключевые цели проектирования: повышение надежности данных и повышение производительности ввода/вывода. Когда несколько физических дисков настроены для использования технологии RAID, они, как говорится, находятся в массиве RAID. Хотя сам массив распределен между несколькими дисками в системе хранения данных, он рассматривается конечным пользователем и операционной системой как единый диск. Теперь операционная система получает доступ к единому логическому диску, и RAID-адаптер обрабатывает распределение данных на нескольких дисках в массиве на основе выбранного уровня RAID.

RAID можно настроить с помощью аппаратных RAID-адаптеров или с помощью программного обеспечения операционной системы, что обеспечивает функциональность RAID без необходимости использования аппаратного адаптера RAID. Аппаратный RAID реализован с помощью адаптера ввода-вывода, который может быть встроенным устройством на системной плате или внешним адаптером RAID на основе PCI. Аппаратный RAID использует интеллектуальный и надежный контроллер и избыточный массив дисков для защиты данных в случае отказа накопителей. Аппаратные адаптеры RAID поддерживают разные уровни RAID, которые можно настроить на основе индивидуальных потребностей системы и приложений. Уровень для хранения данных выбирается выбираемым самостоятельно пользователем в графическом интерфейсе системы хранения.

Программный RAID настраивается с помощью функций RAID, предоставленных операционными системами. Поддерживаемые уровни RAID могут отличаться для разных операционных систем. Наиболее часто

конфигурированными программными уровнями RAID являются RAID 0 (полосатый) и RAID 1 (зеркальное отображение). Программный RAID предоставляет функции без использования аппаратного RAID-адаптера. Однако в целом аппаратный RAID предоставляет лучшие возможности, чем программный RAID, при этом достигая одинаковых или более высоких уровней защиты от неисправностей дисков и адаптеров.

Наиболее часто используемые уровни RAID – это 0, 1, 5, 6.

RAID0, также называемый полосным, разделяет данные на всех дисках без какой-либо информации о паритете, избыточности или устойчивости к неисправностям. Он может быть сконфигурирован на несколько дисков с разными размерами. Для настройки RAID 0 требуется минимум два диска. RAID 0 обычно является хорошим выбором, когда производительность наиболее важна, а целостность данных менее важна. Он обеспечивает высокую производительность ввода-вывода для чтения и записи. Поскольку этот уровень RAID не обеспечивает избыточности данных, он обычно не используется для критически важных сред с жесткими требованиями по доступности данных.

RAID 1. Этот уровень записывает одинаковую копию данных на все диски, являющиеся частью массива RAID 1. RAID 1 обеспечивает наивысшую доступность данных, поскольку блоки отображаются на нескольких дисках и могут защищать от одной или нескольких одновременных неисправностей дисков в зависимости от общего количества дисков/копий в массиве. Это обеспечивает хорошую производительность для операций ввода-вывода с записью и лучшую производительность для операций ввода-вывода чтения. RAID 1 обычно настраивается, когда целостность данных наиболее важна, и рабочая нагрузка больше считывает данные, чем записывает. Для зеркального отображения нужно отразить минимум два диска в массиве.

RAID 2 выделяет данные на уровне битов, а не на уровне блоков. Диски синхронизируются контроллером для вращения в одинаковой угловой ориентации. Поскольку все диски настроены по собственной коррекции кода

ошибки и другим сложным конфигурациям, данный тип RAID редко реализуется. RAID 2 требует по меньшей мере четыре необработанных диска в массиве.

RAID 3 состоит из выделения на уровне байтов с выделенным четным диском. Одной из характеристик RAID 3 является то, что он обычно не может обслуживать несколько запросов одновременно, поскольку любой отдельный блок данных по определению будет распределен по всем членам набора и будет находиться в одном и том же месте. Поэтому любая операция ввода-вывода добавляет использование процессора на все диски в массивах и обычно требует синхронизированных шпинделей. И RAID 3, и RAID 4 заменены на RAID 5. Для настройки RAID 3 требуется по крайней мере четыре выходных диска.

RAID 4 сконфигурирован с помощью специального парного диска. Это обеспечивает хорошую производительность по сравнению с RAID 2 и RAID 3. 4 требует не менее трех необработанных дисков для настройки массива RAID. Он используется редко.

RAID 5 состоит из полосового уровня с распределенной четностью. Информация о четности распределяется между всеми дисками в массиве RAID. Если один диск в массиве выходит из строя, потеря данных не происходит, все данные можно восстановить на новый диск. RAID 5 обеспечивает хорошую производительность ввода-вывода для чтения и записи. Для настройки RAID 5 требуется минимум три диска.

RAID 6 состоит из полосового уровня с тем же распределенным диском четности, что и RAID 5, и добавляет дополнительный или расширенный диск четности в конфигурации массива. Этот уровень может обеспечить лучший уровень защиты от отказа диска, чем RAID 5. По сравнению с производительностью RAID 5, RAID 6 предлагает лучшую производительность для операций чтения, операции записи хуже из-за дополнительной обработки, необходимой дополнительному диску четности. RAID 6 требует не менее четырех выходных дисков для настройки RAID 6.

RAID 10 – это один из гибридных уровней RAID, также известный как RAID 1+0. Это сочетание зеркального отображения диска с полоской диска. Для настройки RAID 10 требуется минимум четыре диска. Он обеспечивает лучшую производительность ввода-вывода для чтения и записи в массиве. Однако это не экономично, поскольку обеспечивает всего 50% полезного дискового пространства в общей емкости массива. Для этого уровня нужно настроить по крайней мере четыре необработанных диска в массиве [23].

RAID 50 – это нестандартная конфигурация, которая также известна как RAID 5+0. Это сочетание дискового полосования с одним парным диском. Для настройки RAID 50 требуется минимум шесть дисков. Он обеспечивает высокую защиту данных даже при многих неисправностях диска.

RAID 60 – это нестандартная конфигурация, которая также известна как RAID 6+0. Это сочетание дисковой полоски с несколькими дисками четности. Для настройки RAID 60 требуется минимум восемь дисков.

Во многих программах наблюдается значительный перекос в распределении рабочей нагрузки ввода-вывода. Небольшая часть хранилища отвечает за непропорционально большую долю от общей нагрузки ввода-вывода среды. Easy Tier действует для обнаружения этого перекоса и автоматического размещения данных, перемещая самые горячие данные на самый быстрый уровень хранилища, в связи с этим рабочая нагрузка на остальное хранилище значительно уменьшается, обслуживая большую часть рабочей нагрузки программы из быстрого хранилища, тем самым ускоряя работу программы. Easy Tier – это функция оптимизации производительности, которая автоматически мигрирует (или перемещает) экстенты данных, относящихся к тому между различными уровнями хранения, в зависимости от их нагрузки ввода-вывода. В результате перемещения система не сохраняет все данные на одном уровне, а скорее в двух-трех уровнях одновременно. Каждый уровень обеспечивает оптимальную эффективность для объема, как показано на рисунке 13.

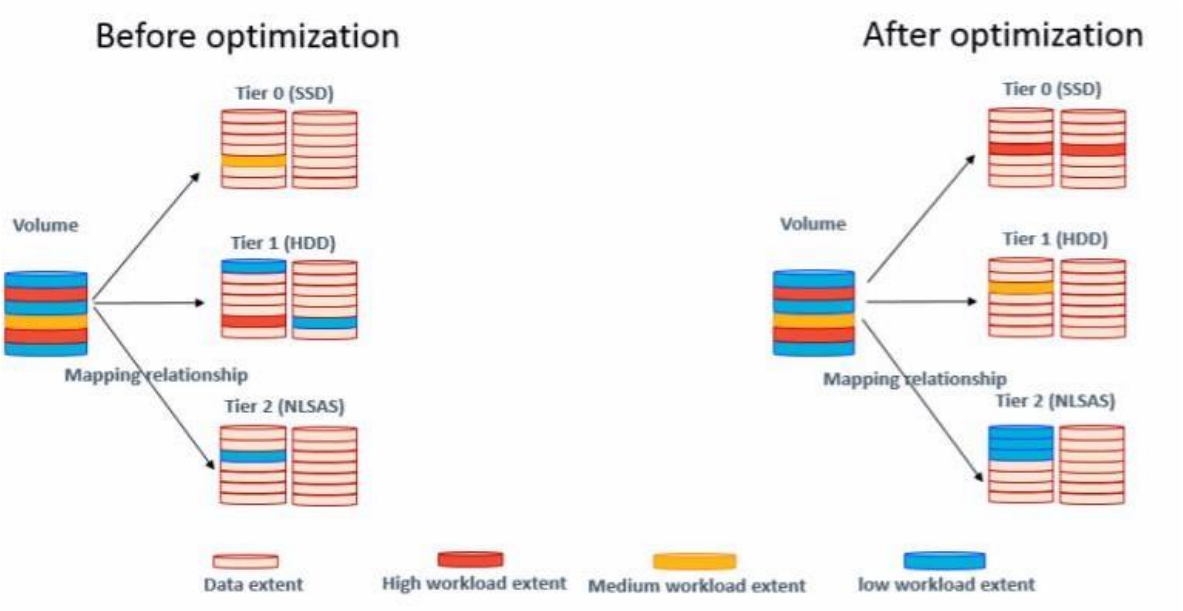

Рисунок 13 – Концепция работы tiering

Easy Tier контролирует активность ввода-вывода и задержку экстентов на всех пулах хранения для создания тепловых карт. На их основе он создает план миграции экстентов и продвигает (перемещает) высокую активность или горячие экстенты на более высокий уровень диска в том же пуле хранения. Он также переносит экстенты, активность которых упала, перемещая их с более быстрых дисков на более медленные.

Пулы хранения данных, содержащие только один уровень хранилища, также могут извлечь выгоду из использования Easy Tier, если они имеют несколько дисковых массивов. Easy Tier имеет режим балансировки: он перемещает экстенты из массивов занятых дисков в менее занятые массивы того же уровня, балансируя нагрузку ввода-вывода.

Одним из способов ускорения резервного копирования является минимизация скопированных данных. Данная технология носит название Snapshot, ее суть состоит в том, что система приостанавливается на мгновение, пока производится копия (снимок) метаданных. Этот процесс занимает значительно меньшую долю времени, чем полноценное копирование данных. Затем резервное копирование выполняется с помощью метаданных для поиска файлов. Если при других операциях ввода/вывода копируемые данные

изменяются, оригинальные метаданные также обновляются, а созданная копия снимка – нет, поэтому система не будет создавать резервные копии данных, добавленных после создания снимка.

Снимок содержит указатели на большинство данных и содержит фактические данные, только если копия повторно обновилась. Альтернативой может быть приостановка работы системы и начало полного резервного копирования. Эта альтернатива более безопасна, но занимает гораздо больше времени и остановит работу запущенных программ, пока резервное копирование не будет завершено.

Снимки значительно сокращают окно резервного копирования и особенно полезны, когда пользователь постоянно обновляет версии своих приложений, поскольку вернуть систему в состояние последнего моментального снимка довольно легко.

Функция unmap – набор примитивов SCSI, позволяющий хостам указывать системе хранения, что пространство, выделенное диапазону блоков на целевом томе хранения, больше не требуется. Эта команда позволяет контролеру хранилища использовать мероприятий да оптимизировать систему, чтобы пространство можно было использовать повторно для других целей.

К примеру, наиболее распространенным вариантом использования является программа VMware, освобождающая память в файловой системе. Затем контроллер хранения может выполнять функции для оптимизации пространства, например, реорганизовать данные на томе. Когда хост выделяет память, данные помещаются в том. Для освобождения выделенного пространства обратно в пулы хранения используется функция SCSI Unmap. Это позволяет хост-операционным системам отключить хранение на контроллере хранилища, чтобы ресурсы могли автоматически освобождаться в пулах хранения и использоваться для других целей.

Производительность систем хранения данных измеряется количеством операций ввода/вывода, которые они способны выполнять в единицу времени

(IOPS). Для разработки модели, которая будет оценивать влияние методов экономии емкости на производительность системы хранения данных, необходимо сначала четко определиться с системами, которые будут поддерживать технологии компрессии и дедубликации данных.

В данной диссертации было принято решение для моделирования использовать системы хранения данных IBM FlashSystem 5035 и IBM FlashSystem 5200, что обусловлено тем, что система 5035 поддерживает программную компрессию и дедупликацию данных, а система 5200 в свою очередь использует аппаратные ускорители для достижения этого же функционала. Технические характеристики данных систем представлены в настоящей главе.

Производительность системы хранения зависит не только от ее конфигурации, а в первую очередь от программ, выполняющих свою деятельность ввода-вывода в системе хранения. Интенсивность и характеристика деятельности ввода-вывода называется профилем нагрузки. Подробные атрибуты рабочих нагрузок, как правило, определяются заказчиком, запускающим программы [25].

Профиль погрузки состоит из следующих пунктов:

- размер блока передачи/ Данный атрибут указывает на размер передаваемых данных от сервера в систему хранения данных и наоборот;
- отношение количества считываемых операций к операциям записи. Это обусловлено тем, что время выполнения одной операции считывания значительно меньше времени выполнения операции записи, следовательно, и количество IOPS напрямую зависит от данного показателя.

Рассмотрим характеристики пулов хранения данных.

Большинство современных СХД способны включать алгоритмы компрессии и дедубликации на уровне пулов, стоит отметить, что их использование ускоряет утилизацию центрального процессора системы,

который в свою очередь на сегодняшний день является «узким местом» СХД.

Приведем определение количества и объема использованных накопителей и выбранный RAID-уровень.

Storage Modeller – это программное обеспечение, позволяющее пользователю создавать объекты рабочей нагрузки, представляющие активность ввода-вывода для различных программ, использующих смоделированную систему хранения данных. В зависимости от выбранной системы, предоставляется возможность выбирать: объем и скорость вращения жесткого диска, объем флеш-накопителей, уровней RAID и других поддерживаемых характеристик, касающихся конкретной системы хранения данных [8].

С учетом всех возможных конфигураций становится проблемой рассчитать физическую и эффективную емкость разных систем хранения. Пользователю требуется глубокое техническое понимание того, как назначаются запасные и паритетные диски, учитывая одновременное использование дисков с разной емкостью и конфигурацией [6].

Функция емкости, которая является частью программного обеспечения Storage Modeller, предлагает графический интерфейс, позволяющий вводить конфигурацию накопителя, необходимую емкость и другие характеристики. С помощью этих параметров Storage Modeller автоматически вычисляет и отображает физическую и эффективную емкость хранения.

Также данный функционал предполагает динамическое изменение конфигурации системы хранения с учетом всех применимых к ней правил.

В отличие от аналогов, данное ПО позволяет выставлять желаемый коэффициент компрессии и дедубликацию данных, это позволяет быть уверенным, что входные данные для всех моделей будут одинаковыми.

2.2 Обзор технологий RAID 5 и RAID 6. Обоснование выбранной технологии

Система IBM FlashSystem 5035 входит в семейство СХД IBM FlashSystem 5000. Для того чтобы начать построение модели, необходимо найти ее профайл в графическом интерфейсе Storage Modeller. На рисунке ниже изображено меню выбора профайлов систем хранения (рисунок 14).

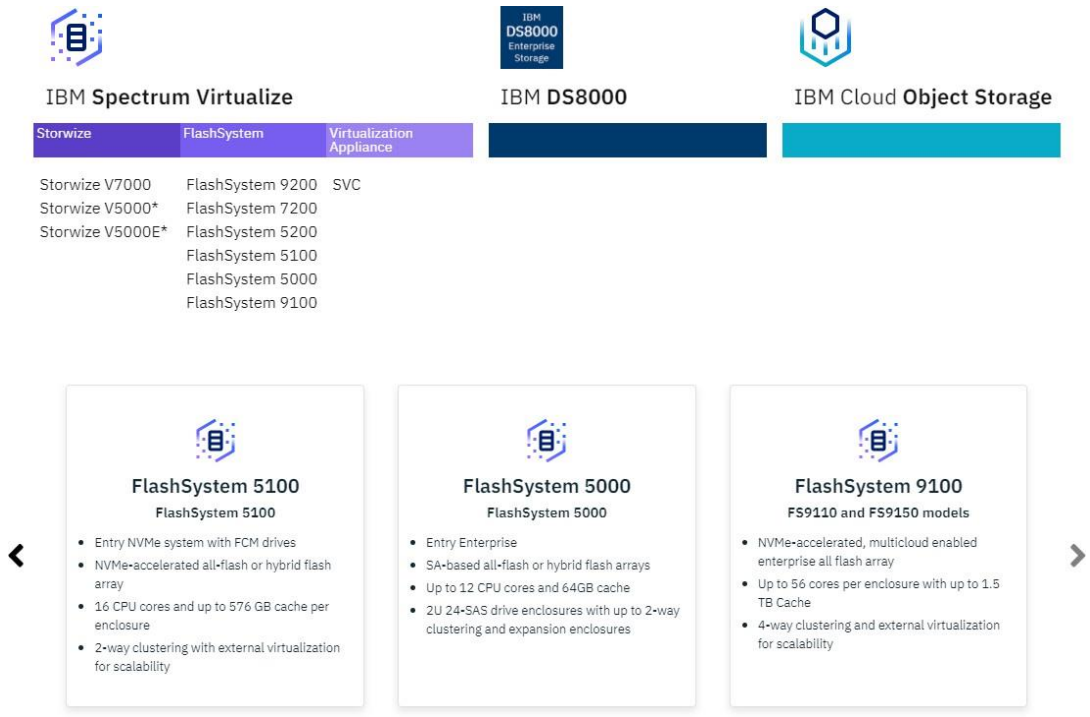

Рисунок 14 – Меню выбора профайла системы хранения

После этого будет предложено выбрать систему сохранения и настроить количество установленной кэш памяти на систему. На рисунке изображены доступные варианты выбора количества кэша для системы хранения данных

FlashSystem 5035. Для нашей модели будет использоваться конфигурация с 64 гигабайтами кэша на систему (рисунок 15).

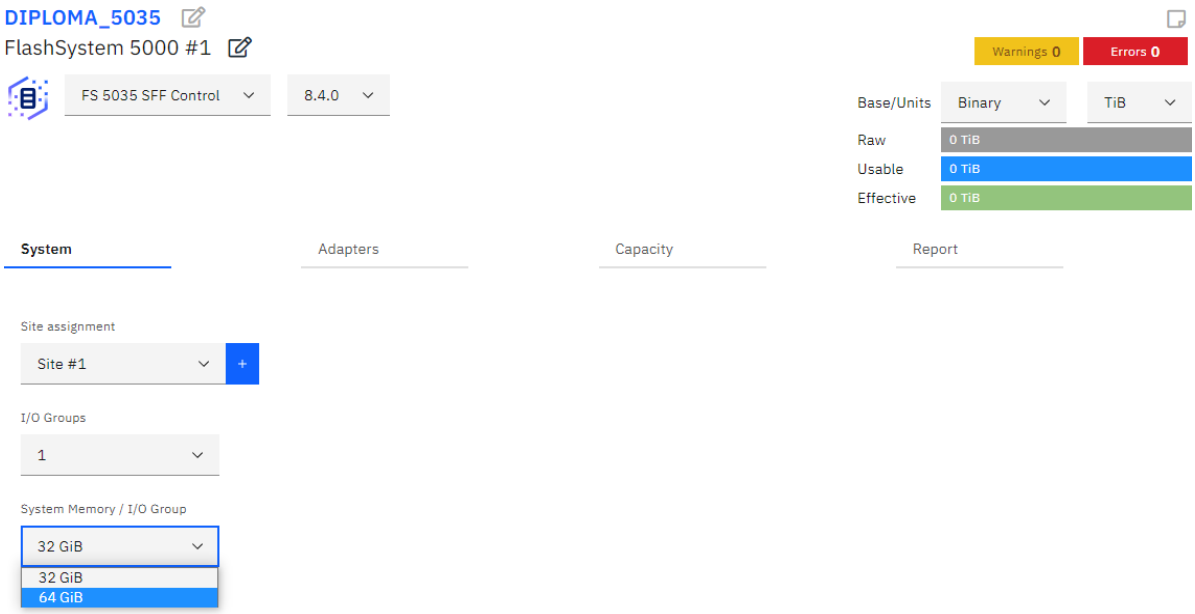

Рисунок 15 – Меню выбора количества кэш памяти

После выбора параметров можно переходить к выбору адаптеров ввода/вывода и протокола передачи данных, который будет использоваться для соединения с хостом. Для СХД 5035 доступны следующие адаптеры ввода/вывода:

– 16 Gb/s Fibre Channel;

– 12 Gb/c SAS;

– 10 Gb/sSCSI или 25 Gb/s SCSI.

Для нашей модели выбрано два 16 Gb/s Fibre Channel адаптеров, каждый из которых насчитывает по четыре порта.

Данный тип оборудования выбран в связи с тем, что они обеспечивают высокую производительность для данной системы хранения.

На рисунке 16 изображено меню выбора адаптеров.

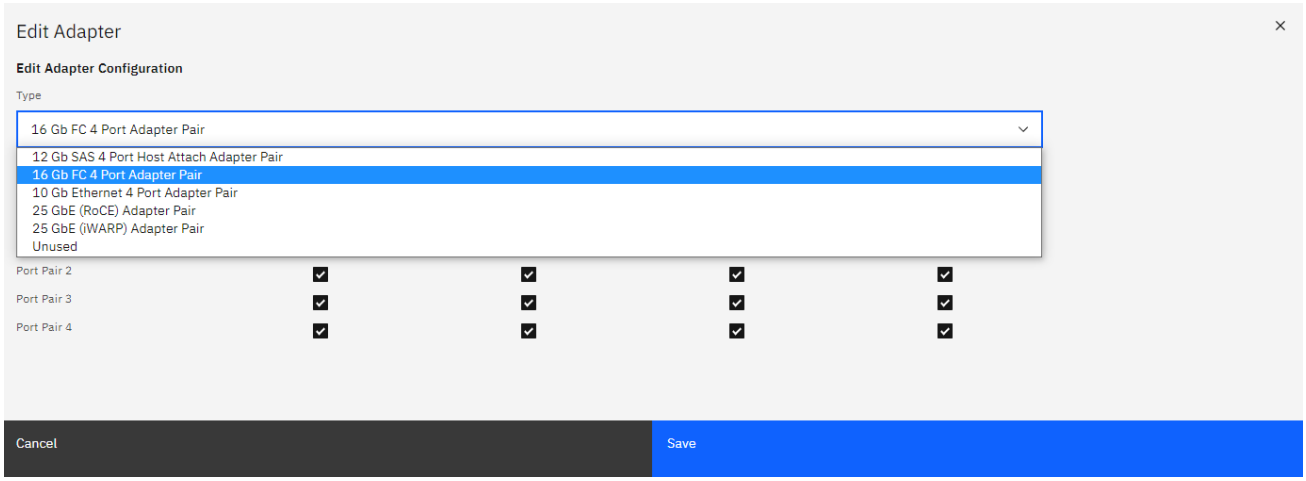

Рисунок 16 – Меню выбора адаптеров

После выбора необходимых адаптеров необходимо перейти на страницу «Capacity». Данный этап очень важен в процессе моделирования системы, поскольку для того, чтобы узнать влияние алгоритмов компрессии и дедубликации на СХД сохранения, необходимо сначала промоделировать ее работу без их использования [9]. Процесс включения алгоритмов компрессии и дедубликации достигается за счет создания специального пула хранения данных. Он носит название Data Reduction Pool (DRP). Если система создает DRP вместо обычного регулярного пула, пользователь может использовать методы сокращения емкости. При использовании пулов уменьшения данных на FlashSystem 5100 или FlashSystem 5030, данные хоста могут быть сжаты или сжаты и дедуплированы, прежде чем они будут записаны на накопители.

Сжатие DRP семейства FlashSystem базируется на методе сжатия данных без утрат Лемпеля-Зива и способе работы в режиме настоящего времени. Когда хост отправляет запрос на запись, запрос подтверждается кэшем записи системы, затем передается в DRP. В рамках индексирования запрос на запись проходит через механизм сжатия, затем сохраняется в сжатом формате. Этот процесс происходит прозрачно для сервера. Сжатие DRP поддерживается на FlashSystem 5100.

В канистрах узлов этой системы установлен аппаратный ускоритель сжатия, увеличивающее пропускную способность передач ввода-вывода

между узлами и сжатыми объемами. FlashSystem 5030 с конфигурацией кэш-памяти 64 ГБ (32 ГБ оперативной памяти на каждый узел) также поддерживает сжатие DRP. Однако эта система не имеет аппаратного ускорителя сжатия, а использует процессор канистры для сжатия данных. В этой связи нужно придерживаться четкого планирования необходимой производительности СХД. Системы FlashSystem 5010 и FlashSystem 5030 с 16 ГБ оперативной памяти для каждого узла не поддерживают компрессию и дедупликацию данных.

Дедупликацию можно настроить с помощью тонко предоставленных и сжатых томов в пулах для уменьшения данных для дополнительной экономии емкости. Процесс дедубликации определяет уникальные фрагменты данных или шаблоны байтов и сохраняет подпись фрагмента для справки при записи новых фрагментов данных. Если подпись нового фрагмента соответствует существующей подписи, новый фрагмент заменяется небольшой ссылкой, указывающей сохраненный фрагмент. Совпадения обнаруживаются при записи данных. Один и тот же шаблон байта может происходить многократно, в результате чего объем данных, который нужно хранить, значительно уменьшается.

Дедупликация поддерживается на томах DRP в системах FlashSystem 5100. Она также может быть реализована на FlashSystem 5030 с функцией кэширования 64 ГБ. Однако это может быть менее эффективно по сравнению с FlashSystem 5100, поскольку размер базы данных подписей ограничен на этой модели. Чтобы помочь профилированию и анализу существующих нагрузок, которые необходимо перенести на систему семейства FlashSystem 5000, можно использовать инструмент оценки уменьшения данных (DRET) [18]. DRET – это очень точная утилита на базе командной строки для оценки экономии уменьшения данных на блочных устройствах хранения данных. Инструмент сканирует целевые нагрузки в различных массивах данных, объединяет все результаты сканирования и обеспечивает оценку уменьшения данных. Сжатие и дедупликация не взаимоисключающие; можно включить

одну, обе функции или одну. При включении обеих функций данные сначала дедуплицируются, а затем сжимаются. Поэтому ссылки на дедупликацию создаются на сжатых данных.

Для создания модели, не использующей функционал компрессии и дедубликации, необходимо создать обычный пул хранения (RG), как изображено на рисунке 17.

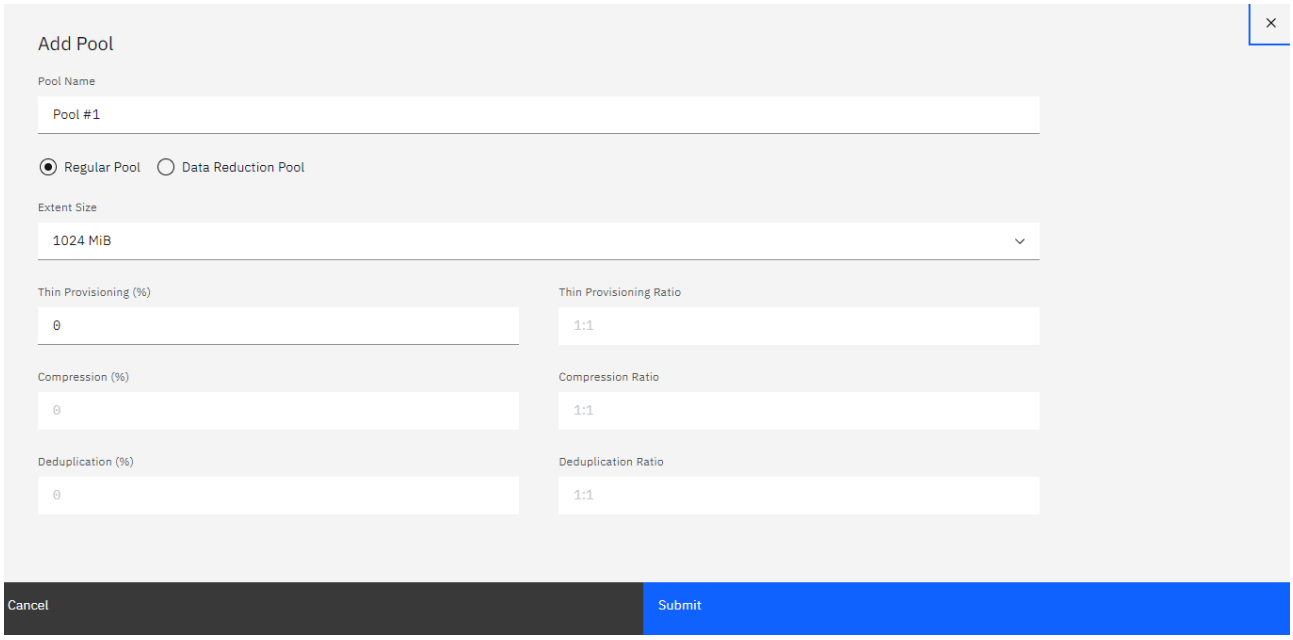

Рисунок 17 – Меню создания дискового пула

После добавления пуля RG необходимо создать том данных, который будет его частным.

На рисунке 18 изображено меню, в котором отображаются создания пулы и тома данных.

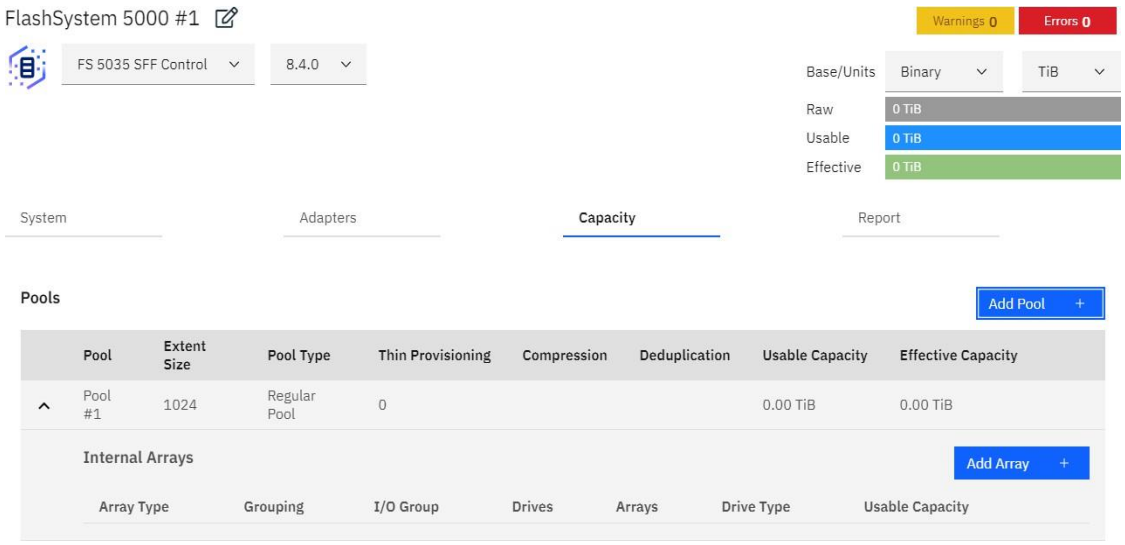

Рисунок 18 – Меню страницы Capacity

Для данной системы существуют следующие рекомендации, для получения наилучшей производительности СХД. Дисковый пул должен содержать не менее тринадцати накопителей. Тип накопителей – SSD. Луны должны быть объединены в RAID 6 [19].

Для предложенной модели создан пул из 13 дисков типа SSD, каждый из которых имеет объем 3.84 TB. На рисунке 19 изображено меню с необходимыми для создания параметров.

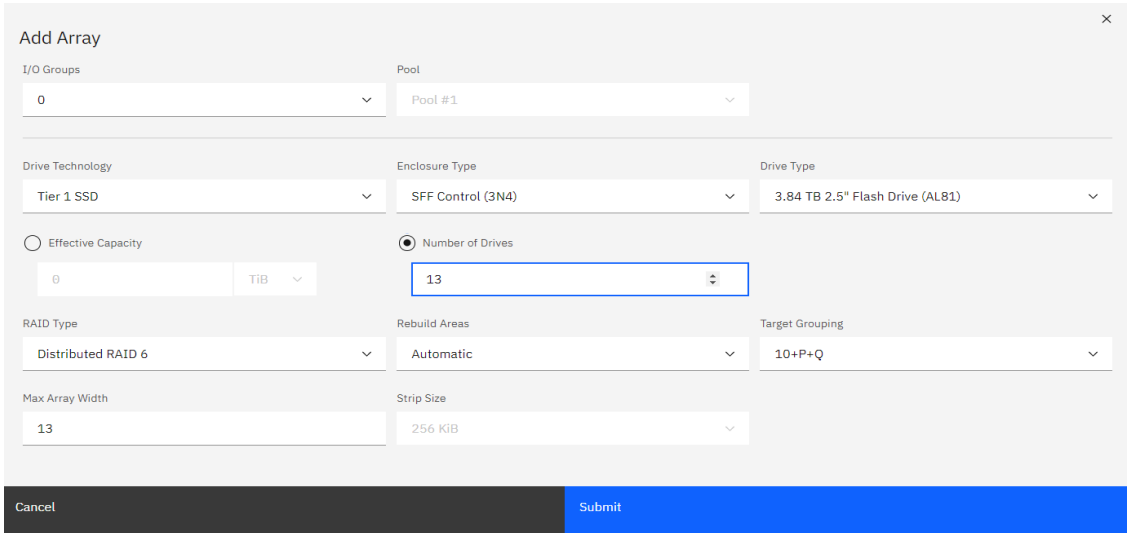

Рисунок 19 – Меню создания том хранения

В меню создания тома можно наблюдать следующие поля для заполнения [26]:

- I/O Groups. В случае, когда моделируется кластерная конфигурация СХД, она имеет несколько групп ввода/вывода. То есть если система 5035 работала в двухкластерной конфигурации, необходимо было бы выбрать на какой I/O группе будут храниться данные;
- Pool. В случае, когда моделируется работа СХД с двумя или большим количеством пулов, пользователь может выбрать, к какому из пулов ему отнести данный том;
- Drive Technology. СХД 5035 поддерживает три типа накопителей: жесткие диски со скоростью вращения шпинделя 7.2К rpm носят название Nearline HDD, диски 10К rpm называются Enterprise HDD. Поскольку твердотельные накопители являются самым быстрым носителем информации в данной системе хранения данных, их название происходит от быстрого уровня тиринга, а именно Tier 1 SSD. Для нашей модели нужно выбрать технологию Tier 1 SSD;
- Enclosure Type. Контроллерное шасси в системе 5035 максимально может вместить 24 накопителя. В случае, если пользователь захочет увеличить количество дисков, он может приобрести дополнительную полку расширения, которая будет подключена к контроллеру благодаря протоколу SAS 12 Gb/s;
- Drive Тип. Данное поле требует выбрать необходимый объем носителя, который будет использоваться для моделирования;
- Effective capacity. Данный пункт позволяет пользователю требуемый для него эффективный объем;
- Drives - номер. Позволяет указать необходимое количество накопителей, которые должны входить в группу RAID;
- RAID Type. Выбор желаемого уровня RAID, на базе которого будет работать массив;
- Rebuild Areas. Позволяет установить желаемое количество spare-

накопителей. При выходе из строя диска, который был частью RAID-группы информация автоматически восстанавливается на накопителе типа spare;

– Target Grouping. После того, как пользователь выбрал необходимый уровень RAID. Необходимо избрать его длину. Выражение 10+P+Q означает, что дисковый набор данных состоит из десяти накопителей, а условные обозначения P и Q отвечают за parity-накопители.

На рисунке 20 изображено меню страницы Capacity, где отображаются созданные пулы и LUN массива.

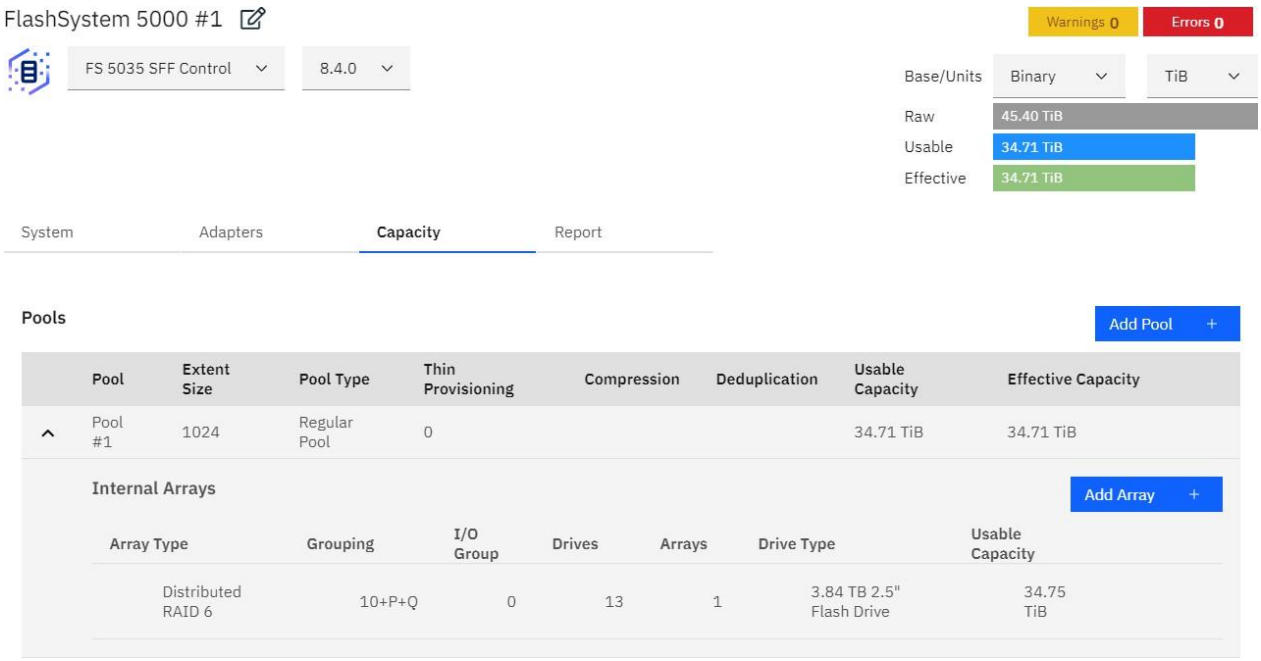

Рисунок 20 – Заполненное меню страницы Capacity

Конечный отчет емкости созданного дискового пула представлен в таблице 5.

Таблица 5 – Отчет полученного объема

Summary Report		
Номер I/O групп:	1	-
Number of pools:	1	-
Номер arrays:	1	-
Номер велосипедов:	13	-
Число external virtualized MDisks:	0	-
Номер of enclosures:	1	(0 expansion enclosures)
FS 5035 SFF Control:	1	
Raw capacity:	49,922.84 GB	49.92 TB
	46,494.27 GiB	45.40 TiB
Effective capacity:	38,159.71 GB	38.16 TB
	35,539.00 GiB	34.71 TiB

После создания дискового пула. Можно перейти к настройкам профиля загрузки. Первая часть меню Performance изображена на рисунке 21.

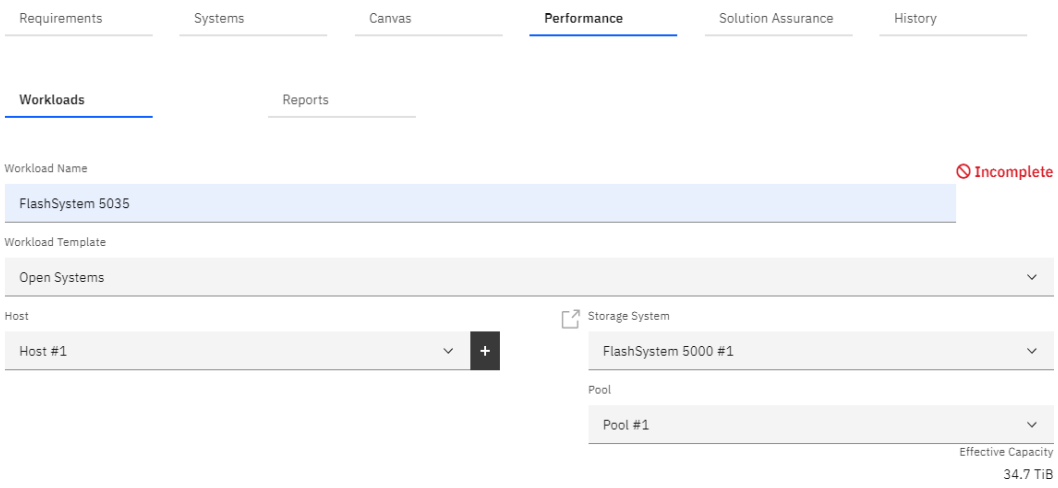

Рисунок 21– Меню Performance

Для создания профиля нагрузки сначала необходимо создать хост, который будет имитировать запросы ввода/вывода по протоколам передачи данных Fibre Chanel.

На рисунке 22 изображено меню создания сервера.

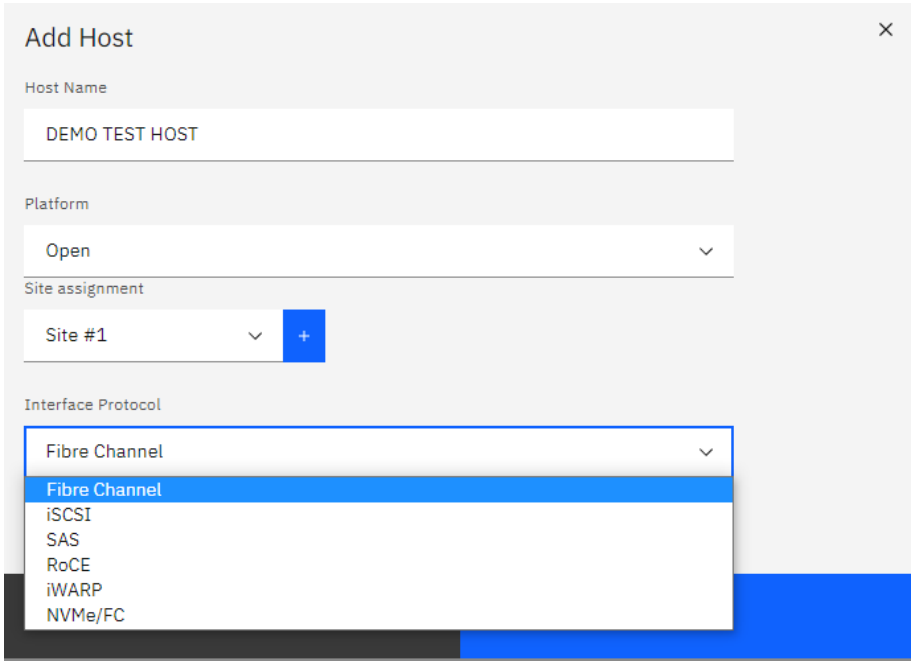

Рисунок 22 – Меню создания сервера

Особое внимание следует обратить на доступные протоколы передачи данных хосту. Кроме обычного FC-соединения пользователь может выбрать протокол NVMe/FC. NVMe over Fiber Channel – это полнофункциональная, высокоэффективная технология передачи данных на основе протокола NVMe. Поскольку система 5035 не поддерживает NVMe накопители, то и использовать вышеупомянутый протокол нельзя.

После создания хоста необходимо выбрать ранее созданный пул данных и подключить его характеристики к нашему профилю нагрузки. На выходе наша модель имеет следующие входные параметры:

– блок передачи данных – 8 Kib;
– процентное отношение количества считываемых операций к операциям записи: 70/30 %;
– процент данных, которые будут считываться с кэш памяти системы.

На рисунке 23 изображено меню параметров нагрузки.

Рисунок 23 – Меню параметров нагрузки

В ходе проведенного анализа и моделирования системы хранения данных IBM FlashSystem 5035 было обосновано применение технологии RAID 6 в составе дискового пула из 13 SSD-накопителей объемом по 3.84 ТБ каждый. Данный выбор произведен путем обоснования действия факторов:

RAID 6 обеспечивает защиту от одновременного выхода из строя двух дисков в массиве [13].

При использовании SSD-накопителей технология RAID 6 демонстрирует оптимальный баланс между отказоустойчивостью и производительностью, для заданного профиля нагрузки (70% операций чтения/30% записи).

Выбранная конфигурация 10+P+Q (10 дисков данных + 2 диска четности) позволяет эффективно использовать дисковое пространство при сохранении требуемого уровня надежности.

Система IBM FlashSystem 5035 с 64 ГБ кэш-памяти оптимально поддерживает работу с RAID 6.

RAID совместима с возможностями системы по аппаратному ускорению операций ввода-вывода через Fibre Channel адаптеры 16 Gb/s.

В конфигурации предусмотрены spare-накопители для автоматического восстановления при выходе дисков из строя. Полученная эффективная емкость массива составила 38.16 ТБ при raw-емкости 49.92 ТБ, соответствует ожидаемым показателям для RAID 6.

2.3 Концептуальное проектирование системы управления хранением данных на основе RAID-массивов

По окончании моделирования получаем следующие данные:

– таблица утилизации ресурсов в зависимости от IOPS;
– таблица с данными о времени выполнения одной операции при разных количествах IOPS;
– график зависимости времени выполнения одной команды от IOPS
– производительность СХД 5035 без использования алгоритмов компрессии и дедубликации.

В таблице 6 приведена степень утилизации системы при различных IOPS без включенного функционала компрессии и дедубликации данных.

Таблица 6 показывает время выполнения одной операции при разных IOPS.

Таблица 6 – Степень утилизации системы при различных IOPS без включенного функционала компрессии и дедубликации данных

FlashSystem 5000 #1 FS5000/8.4.0								
Peak IO/s	Total I/O Rate (I/O per second)							
Utilizations	Amber Threshold	Red Threshold	131050	135278	139505	202917	207144	211372
System Core	60%	80%	60,5%	62,4%	64,4%	93,7%	95,6%	97,6%
NVMe Drive Interface	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest SCM Drive	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest NVMe Flash	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
SAS Drive Interface	60%	80%	23,0%	23,7%	24,4%	35,5%	36,3%	37,0%
Highest Tier 0 SSD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Tier 1 SSD	60%	80%	61,9%	63,9%	65,9%	95,9%	97,9%	99,9%
Highest Enterprise HDD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Nearline HDD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest FC Adapter	60%	80%	12,9%	13,3%	13,7%	19,9%	20,3%	20,7%
Highest FC Port	60%	80%	2,1%	2,2%	2,2%	3,3%	3,3%	3,4%
Highest Ethernet Adapter	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%

Highest Ethernet Port	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%

Продолжение таблицы 6

FlashSystem 5000 #1 FS5000/8.4.0								
Total I/O Rate (I/O per second)								
Highest SAS Adapter	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest SAS Port	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Metro Mirror Write (MiB/s)			0,00	0,00	0,00	0,00	0,00	0,00
Global Mirror Write (MiB/s)			0,00	0,00	0,00	0,00	0,00	0,00

Таблица 7 – Время выполнения одной операции при различных IOPS

FlashSystem 5000 #1 FS5000/8.4.0				
FlashSystem 5035		Service Time		
I/O Rate (ops/s)	Data Rate (MiB/s)	Total (ms/op)	Read (ms/op)	Write (ms/op)
143733	1122,9	0,65	0,66	0,61
147960	1155,9	0,67	0,69	0,63
152187	1189,0	0,70	0,72	0,66
156415	1222,0	0,73	0,76	0,68
160642	1255,0	0,77	0,80	0,71
164870	1288,0	0,81	0,84	0,74
169097	1321,1	0,86	0,89	0,77
173325	1354,1	0,91	0,95	0,80
177552	1387,1	0,97	1,03	0,84
181780	1420,2	1,05	1,12	0,89
186007	1453,2	1,15	1,24	0,95
190234	1486,2	1,28	1,40	1,02
194462	1519,2	1,46	1,61	1,11
198689	1552,3	1,71	1,92	1,23
200000	1562,5	1,82	2,05	1,29
202917	1585,3	2,13	2,42	1,43
207144	1618,3	2,90	3,37	1,79
211372	1651,3	4,89	5,82	2,70

На рисунке 24 изображен график зависимости IOPS от времени обслуживания одной команды для SSD накопителей.

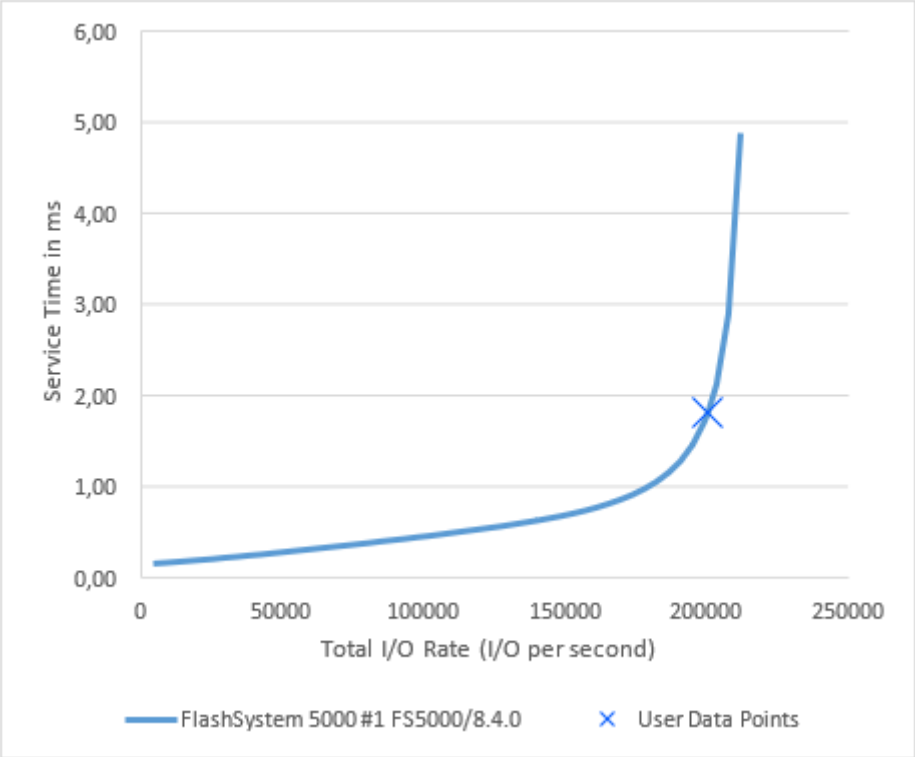

Рисунок 24 – График зависимости IOPS от времени обслуживания одной команды для SSD накопителей

Построим аналогичные модели для этой же системы хранения данных, но с включенными функциями компрессии и дедубликации поочередно. Для того чтобы включить вышеуказанный функционал необходимо создать Data Reduction Pool.

В таблице ниже 8 показаны исходные параметры емкости с учетом использования алгоритма компрессии, коэффициент сжатия 2:1.

Таблица 8 – Параметры емкости с учетом использования алгоритма компрессии, коэффициент сжатия 2:1

Summary Report		
Номер I/O групп:	1	-
Number of pools:	1	-
Номер arrays:	1	-
Номер:	13	-
Число external virtualized MDisks:	0	-
Номер of enclosures:	1	-

Продолжение таблицы 8

Summary Report		
FS 5035 SFF Control:	1	-
Raw capacity:	49,922.84 GB	49.92 TB
	46,494.27 GiB	45.40 TiB
Effective capacity:	67,422.40 GB	67.42 TB
	62,792.00 GiB	61.32 TiB

В таблице 9 приведена степень утилизации системы при различных IOPS с включенным функционалом программной компрессии данных с коэффициентом сжатия 2:1.

Таблица 10 содержит данные о времени выполнения одной операции при различных IOPS при включенной компрессии.

Таблица 9 – Степень утилизации системы при разных IOPS с включенным функционалом компрессии данных

FlashSystem 5000 #2 FS5000/8.4.0								
Peak IO/s	Total I/O Rate (I/O per second)							
Utilizations	Amber Threshold	Red Threshold	43870	45982	51034	55293	56104	62349
System Core	60%	80%	61,9%	63,9%	75,9%	81,9%	83,9%	99,9%
NVMe Drive Interface	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest SCM Drive	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest NVMe Flash	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
SAS Drive Interface	60%	80%	5,0%	5,2%	6,1%	6,6%	6,8%	8,1%
Highest Tier 0 SSD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Tier 1 SSD	60%	80%	13,5%	13,9%	16,5%	17,8%	18,3%	21,8%
Highest Enterprise HDD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Nearline HDD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest FC Adapter	60%	80%	2,8%	2,9%	3,4%	3,7%	3,8%	4,5%
Highest FC Port	60%	80%	0,5%	0,5%	0,6%	0,6%	0,6%	0,7%
Highest Ethernet Adapter	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Ethernet Port	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest SAS Adapter	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest SAS Port	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Metro Mirror Write (MiB/s)			0,00	0,00	0,00	0,00	0,00	0,00
Global Mirror Write (MiB/s)			0,00	0,00	0,00	0,00	0,00	0,00

Таблица 10 – Время выполнения одной операции при разных количествах IOPS

FlashSystem 5000 #2 FS5000/8.4.0				
flashsystem 5035 comp		Service Time		
I/O Rate (ops/s)	Data Rate (MiB/s)	Total (ms/op)	Read (ms/op)	Write (ms/op)
921	7,2	1,18	1,24	1,06
1842	14,4	1,23	1,28	1,10
2763	21,6	1,26	1,31	1,14
3684	28,8	1,29	1,34	1,16
4605	36,0	1,31	1,36	1,19
5526	43,2	1,33	1,38	1,21
6447	50,4	1,35	1,40	1,22
7368	57,6	1,36	1,42	1,24
8289	64,8	1,38	1,43	1,25
9210	72,0	1,39	1,45	1,27
10131	79,2	1,41	1,46	1,28
11052	86,3	1,42	1,48	1,30
11973	93,5	1,44	1,49	1,31
12894	100,7	1,45	1,50	1,33
13815	107,9	1,47	1,52	1,34
14736	115,1	1,48	1,53	1,35
15657	122,3	1,50	1,55	1,37

На рисунке 25 изображено график зависимости IOPS от времени обслуживания для SSD накопителей с включенным функционалом компрессии.

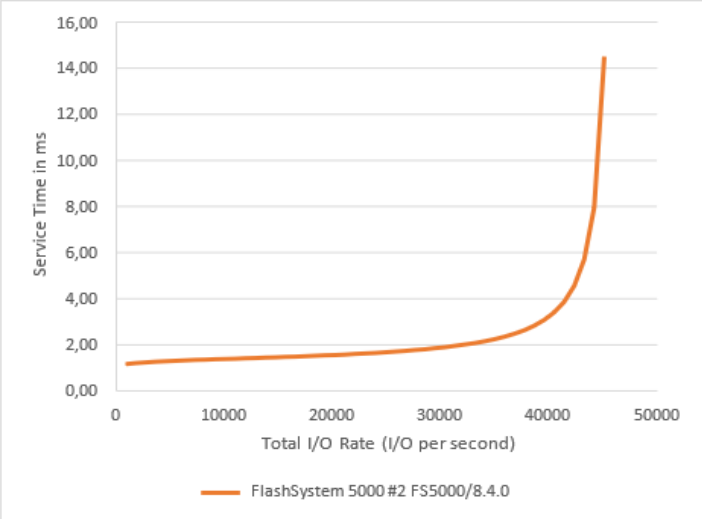

Рисунок 25 – График зависимости IOPS от времени обслуживания для SSD накопителей с использованием программного алгоритма компрессии

В таблице 11 показаны исходные параметры емкости СХД с учетом использования алгоритма дедубликации, коэффициент сжатия 1,2:1.

Таблица 11 – Параметры емкости с учетом использования алгоритма дедубликации, коэффициент сжатия 2:1

Summary Report		
Номер I/O групп:	1	
Number of pools:	1	
Номер arrays:	1	
Номер велосипедов:	13	
Число external virtualized MDisks:	0	
Номер of enclosures:	1	
FS 5035 SFF Control:	1	
Raw capacity:	49,922.84 GB	49.92 TB
	46,494.27 GiB	45.40 TiB
Effective capacity:	42,139.00 GB	42.14 TB
	39,245.00 GiB	38.33 TiB

В таблице 12 приведена степень утилизации системы при различных IOPS с включенным функционалом программной дедубликации данных с коэффициентом сжатия 1,2:1. Таблица 13 приводит данные о времени выполнения одной операции при различных IOPS.

Таблица 12 – Степень утилизации системы при различных IOPS с включенным программным алгоритмом дедубликации данных

FlashSystem 5000 #3 FS5000/8.4.0									
Peak IO/s	Total I/O Rate (I/O per second)								
Utilizations	Amber Threshold	Red Threshold	26822	27688	35475	36340	41531	42397	43262
System Core	60%	80%	61,9%	63,9%	81,9%	83,9%	95,9%	97,9%	99,9%
NVMe Drive Interface	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest SCM Drive	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest NVMe Flash	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
SAS Drive Interface	60%	80%	4,7%	4,9%	6,2%	6,4%	7,3%	7,4%	7,6%
Highest Tier 0 SSD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Tier 1 SSD	60%	80%	12,7%	13,1%	16,8%	17,2%	19,6%	20,0%	20,4%
Highest Enterprise HDD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%

Продолжение таблицы 12

FlashSystem 5000 #3 FS5000/8.4.0									
Peak IO/s	Total I/O Rate (I/O per second)								
Highest Nearline HDD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest FC Adapter	60%	80%	8,7%	8,9%	11,5%	11,7%	13,4%	13,7%	14,0%
Highest FC Port	60%	80%	0,9%	1,0%	1,3%	1,3%	1,5%	1,5%	1,5%
Highest Ethernet Adapter	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Ethernet Port	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest SAS Adapter	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest SAS Port	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%

Таблица 13 – Время выполнения одной операции при разных IOPS

Flashsystem 5035 dedup		Service Time		
I/O Rate (ops/s)	Data Rate (MiB/s)	Total (ms/op)	Read (ms/op)	Write (ms/op)
865	6,8	1,28	1,24	1,37
1730	13,5	1,34	1,29	1,46
2596	20,3	1,38	1,32	1,53
3461	27,0	1,42	1,35	1,59
4326	33,8	1,45	1,37	1,63
5191	40,6	1,48	1,40	1,67
6057	47,3	1,50	1,41	1,71
6922	54,1	1,53	1,43	1,75
7787	60,8	1,55	1,45	1,78
8652	67,6	1,57	1,46	1,81
9518	74,4	1,59	1,48	1,84
10383	81,1	1,61	1,49	1,87
11248	87,9	1,63	1,51	1,90
12113	94,6	1,65	1,53	1,93
12979	101,4	1,67	1,54	1,96
13844	108,2	1,69	1,56	1,99
14709	114,9	1,71	1,57	2,02
15574	121,7	1,73	1,59	2,06

На рисунке 26 изображен график зависимости IOPS от времени обслуживания для SSD накопителей при включенной дедубликации данных.

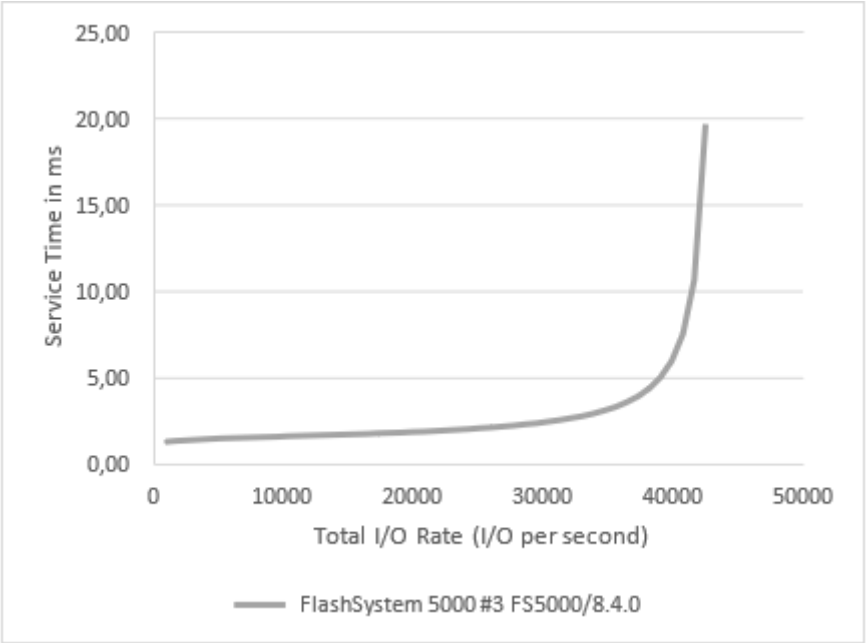

Рисунок 26 – График зависимости IOPS от времени обслуживания для SSD накопителей при включенной дедубликации данных

Система хранения данных FlashSystem 5200 в отличие от 5035 использует аппаратные ускорители, позволяющие оптимизировать работу СХД при включении алгоритмов компрессии и дедубликации. Процесс создания модели отличается тем, что выбираем другой профайл системы.

В таблице 14 приведены данные, приводящие степень утилизации системы хранения без включенного функционала компрессии и дедубликации данных при пиковых значениях IOPS. Таблица 15 приводит данные о времени выполнения одной операции при разных IOPS.

Таблица 14 – Степень утилизации системы хранения без включенного функционала компрессии и дедубликация данных при пиковых значениях IOPS

FlashSystem 5000 #1 FS5200/8.4.0								
Peak IO/s	Total I/O Rate (I/O per second)							
Utilizations	Amber Threshold	Red Threshold	208501	215226	316114	322840	329566	336291
System Core	60%	80%	61,9%	63,9%	93,9%	95,9%	97,9%	99,9%
NVMe Drive Interface	60%	80%	8,2%	8,5%	12,5%	12,7%	13,0%	13,3%

Продолжение таблицы 14

FlashSystem 5000 #1 FS5200/8.4.0								
Total I/O Rate (I/O per second)								
Highest SCM Drive	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest NVMe Flash	60%	80%	58,3%	60,2%	88,4%	90,3%	92,2%	94,0%
SAS Drive Interface	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Tier 0 SSD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Tier 1 SSD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Enterprise HDD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Nearline HDD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest FC Adapter	60%	80%	7,4%	7,6%	11,2%	11,4%	11,7%	11,9%
Highest FC Port	60%	80%	20,0%	20,6%	30,3%	31,0%	31,6%	32,2%
Highest Ethernet Adapter	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Ethernet Port			0,00	0,0%	0,0%	0,0%	0,0%	0,0%
Highest SAS Adapter			0,00	0,0%	0,0%	0,0%	0,0%	0,0%

Таблица 15 – Время выполнения одной операции ввода/вывода при различных IOPS

FlashSystem 5200 #1 FS5200/8.4.0				
5200_no comp_dedup		Service Time		
I/O Rate (ops/s)	Data Rate (MiB/s)	Total (ms/op)	Read (ms/op)	Write (ms/op)
6726	52,5	0,23	0,27	0,13
13452	105,1	0,23	0,28	0,13
20177	157,6	0,24	0,28	0,13
26903	210,2	0,24	0,29	0,13
33629	262,7	0,25	0,30	0,14
40355	315,3	0,25	0,30	0,14
47081	367,8	0,26	0,31	0,14
53807	420,4	0,26	0,32	0,14
60532	472,9	0,27	0,32	0,14
67258	525,5	0,28	0,33	0,15
73984	578,0	0,28	0,34	0,15
80710	630,5	0,29	0,34	0,15
87436	683,1	0,29	0,35	0,15
94162	735,6	0,30	0,36	0,15
100887	788,2	0,30	0,37	0,16
107613	840,7	0,31	0,37	0,16
114339	893,3	0,32	0,38	0,16
121065	945,8	0,32	0,39	0,16
127791	998,4	0,33	0,40	0,17
134517	1050,9	0,34	0,41	0,17
141242	1103,5	0,35	0,42	0,17

На рисунке 27 изображен график зависимости IOPS от времени обслуживания для SSD накопителей.

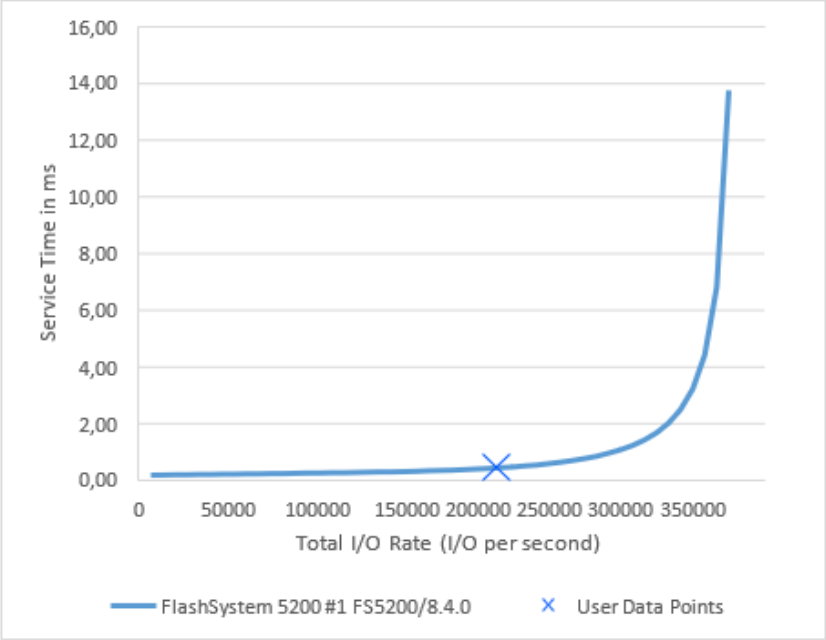

Рисунок 27 – График зависимости IOPS от времени обслуживания для SSD накопителей

В таблице 16 изображены исходные параметры емкости с учетом использования аппаратной компрессии, коэффициент сжатия 2:1.

Таблица 16 – Параметры емкости с учетом использования алгоритма аппаратной компрессии, коэффициент сжатия 2:1

Summary Report		
Номер I/O групп:	1	
Number of pools:	1	
Номер arrays:	1	
Номер велосипедов:	12	
Число external virtualized MDisks:	0	
Номер of enclosures:	1	
FS 5200 SFF Control:	1	
Raw capacity:	49,922.84 GB	49.92 TB
	46,494.27 GiB	45.40 TiB
Effective capacity:	67,422.40 GB	67.42 TB
	62,792.00 GiB	61.32 TiB

Степень утилизации системы хранения с включенным функционалом аппаратной компрессии данных приведена в таблице 17. Таблица 18 приводит данные требуемого времени для выполнения одной операции ввода/вывода при разных IOPS.

Таблица 17 – Производительность системы хранения данных с включенным программным алгоритмом компрессии данных

FlashSystem 5000 #1 FS5200/8.4.0								
Peak IO/s	Total I/O Rate (I/O per second)							
Utilizations	Amber Threshold	Red Threshold	147937	149388	221279	223774	225398	229375
System Core	60%	80%	61,9%	63,9%	93,9%	95,9%	97,9%	99,9%
NVMe Drive Interface	60%	80%	2,7%	2,8%	4,1%	4,2%	4,3%	4,4%
Highest SCM Drive	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest NVMe Flash	60%	80%	18,2%	18,8%	27,6%	28,2%	28,8%	29,4%
SAS Drive Interface	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Tier 0 SSD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Tier 1 SSD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Enterprise HDD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Nearline HDD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest FC Adapter	60%	80%	2,4%	2,5%	3,7%	3,7%	3,8%	3,9%
Highest FC Port	60%	80%	6,6%	6,8%	9,9%	10,1%	10,4%	10,6%
Highest Ethernet Adapter	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Ethernet Port			0,00	0,0%	0,0%	0,0%	0,0%	0,0%
Highest SAS Adapter			0,00	0,0%	0,0%	0,0%	0,0%	0,0%

Таблица 18 – Время выполнения одной операции ввода/вывода при различных IOPS

FlashSystem 5200 #2 FS5200/8.4.0				
5200_comp		Service Time		
I/O Rate (ops/s)	Data Rate (MiB/s)	Total (ms/op)	Read (ms/op)	Write (ms/op)
13227	103,3	1,26	1,69	0,26
15432	120,6	1,33	1,78	0,27
17636	137,8	1,39	1,86	0,28
19841	155,0	1,45	1,94	0,29
22045	172,2	1,50	2,02	0,30
24250	189,4	1,56	2,09	0,31

Продолжение таблицы 18

FlashSystem 5200 #2 FS5200/8.4.0				
5200_comp		Service Time		
26454	206,7	1,61	2,16	0,32
28659	223,9	1,67	2,24	0,33
30863	241,1	1,72	2,31	0,34
33068	258,3	1,78	2,39	0,35
35272	275,6	1,83	2,46	0,36
37477	292,8	1,89	2,54	0,37
39681	310,0	1,95	2,62	0,38
41886	327,2	2,01	2,70	0,39
44090	344,5	2,07	2,79	0,41
46295	361,7	2,14	2,88	0,42
48499	378,9	2,21	2,97	0,43
50704	396,1	2,28	3,07	0,44

На рисунке 28 изображен график зависимости количества IOPS от времени обслуживания для SSD накопителей.

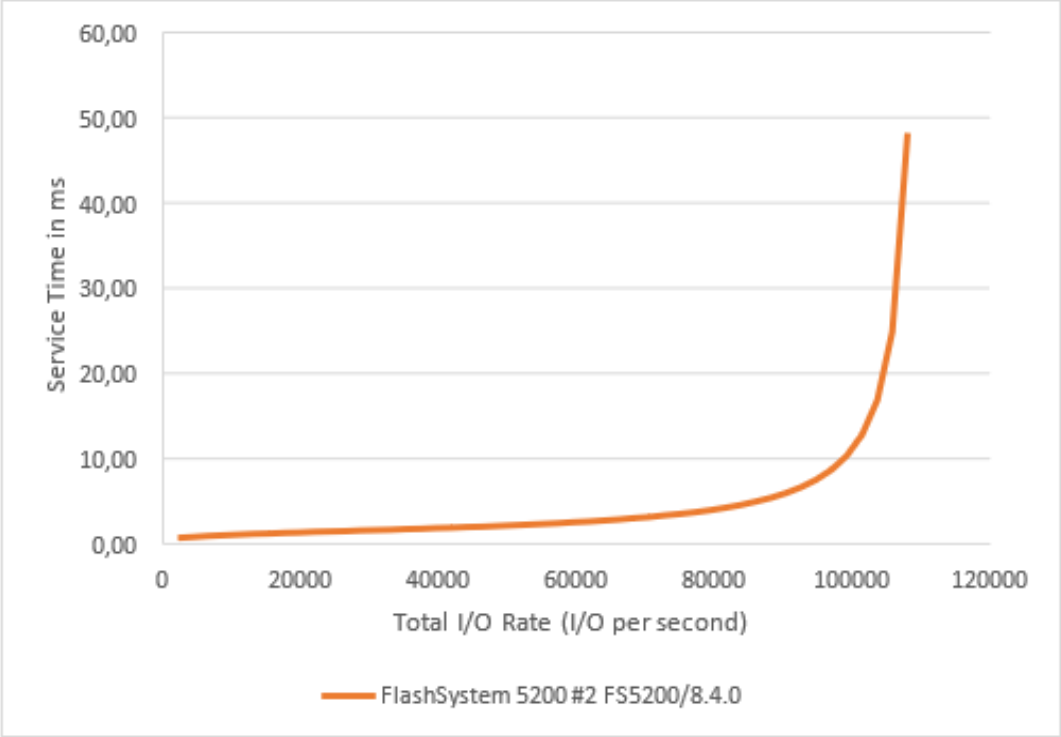

Рисунок 28 – График зависимости IOPS от времени обслуживания для SSD накопителей с использованием программного алгоритма компрессии

В таблице 19 изображены исходные параметры емкости с учетом использования аппаратной дедубликации, коэффициент сжатия 1,2:1.

Таблица 19 – Параметры емкости с учетом использования алгоритма дедубликации, коэффициент сжатия 1,2:1

Summary Report		
Номер I/O групп:	1	-
Number of pools:	1	-
Номер arrays:	1	-
Номер велосипедов:	12	-
Число external virtualized MDisks:	0	-
Номер of enclosures:	1	(0 expansion enclosures)
FS 5200	1	-
Raw capacity:	49,922.84 GB	49.92 TB
	46,494.27 GiB	45.40 TiB
Effective capacity:	42,139.00 GB	42.14 TB
	39,245.00 GiB	38.33 TiB

В таблице 20 показана степень утилизации СХД 5100 с включенным функционалом аппаратной дедубликации данных. В таблице 21 приведены данные по количеству требуемого времени для выполнения одной операции ввода/вывода.

Таблица 20 – Производительность системы хранения данных с включенной аппаратной дедупликацией данных

FlashSystem 5000 #1 FS5200/8.4.0								
Peak IO/s	Total I/O Rate (I/O per second)							
Utilizations	Amber Threshold	Red Threshold	52221	53906	79174	80858	82543	84227
System Core	60%	80%	61,9%	63,9%	93,9%	95,9%	97,9%	99,9%
NVMe Drive Interface	60%	80%	2,1%	2,1%	3,1%	3,2%	3,3%	3,3%
Highest SCM Drive	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest NVMe Flash	60%	80%	13,9%	14,4%	21,1%	21,6%	22,0%	22,4%
SAS Drive Interface	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Tier 0 SSD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Tier 1 SSD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Enterprise HDD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Nearline HDD	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%

Продолжение таблицы 20

FlashSystem 5000 #1 FS5200/8.4.0								
Peak IO/s	Total I/O Rate (I/O per second)							
Highest FC Adapter	60%	80%	7,7%	8,0%	11,7%	12,0%	12,2%	12,5%
Highest FC Port	60%	80%	8,0%	8,2%	12,1%	12,3%	12,6%	12,9%
Highest Ethernet Adapter	60%	80%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Highest Ethernet Port			0,00	0,0%	0,0%	0,0%	0,0%	0,0%
Highest SAS Adapter			0,00	0,0%	0,0%	0,0%	0,0%	0,0%

Таблица 21 – Время выполнения одной операции ввода/вывода при разных количествах IOPS

FlashSystem 5200 #3 FS5200/8.4.0				
5200_dedup		Service Time		
I/O Rate (ops/s)	Data Rate (MiB/s)	Total (ms/op)	Read (ms/op)	Write (ms/op)
1685	13,2	0,89	1,14	0,29
3369	26,3	1,04	1,35	0,33
5054	39,5	1,18	1,53	0,36
6738	52,6	1,29	1,68	0,39
8423	65,8	1,39	1,81	0,41
10107	79,0	1,49	1,94	0,44
11792	92,1	1,57	2,05	0,46
13476	105,3	1,66	2,16	0,48
15161	118,4	1,73	2,26	0,50
16845	131,6	1,81	2,36	0,52
18530	144,8	1,88	2,46	0,54
20215	157,9	1,95	2,56	0,55
21899	171,1	2,03	2,65	0,57
23584	184,2	2,10	2,75	0,59
25268	197,4	2,17	2,84	0,61
26953	210,6	2,25	2,94	0,63
28637	223,7	2,32	3,04	0,64
30322	236,9	2,40	3,15	0,66
32006	250,0	2,49	3,26	0,68
33691	263,2	2,57	3,37	0,70
35375	276,4	2,66	3,49	0,73
37060	289,5	2,75	3,61	0,75
38745	302,7	2,85	3,74	0,77
40429	315,9	2,96	3,88	0,80

На рисунке 29 изображен график зависимости количества IOPS от времени обслуживания для SSD накопителей.

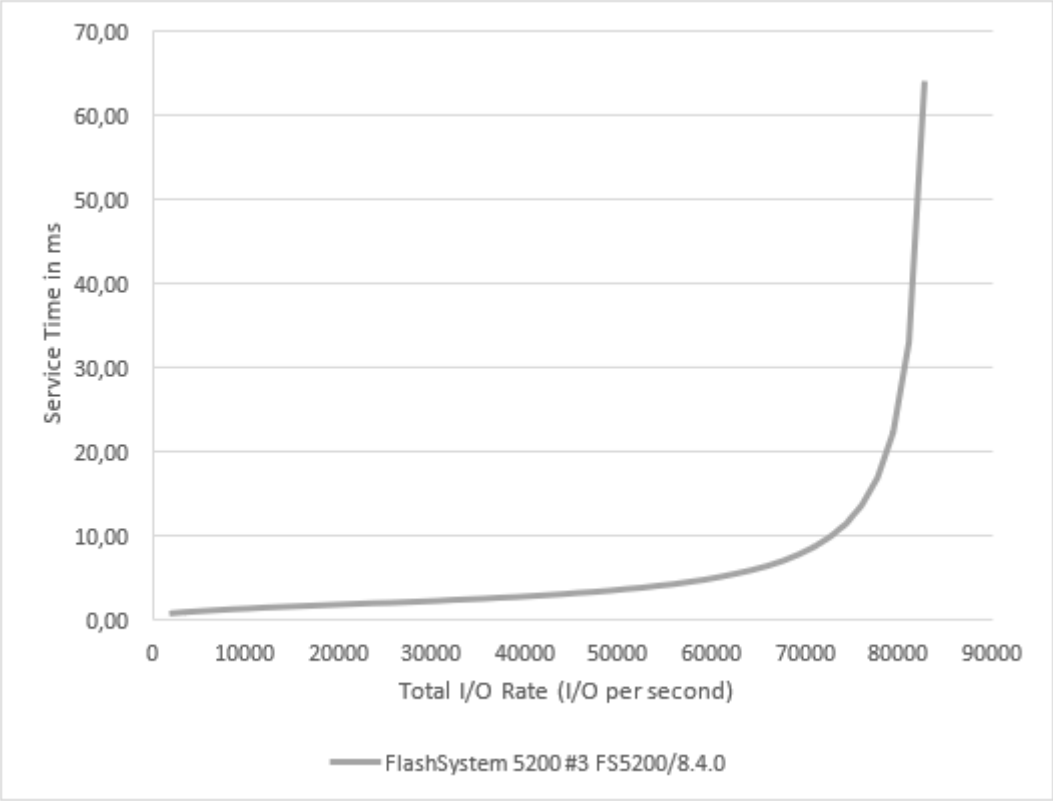

Рисунок 29 – График зависимости IOPS от времени обслуживания для SSD накопителей с использованием аппаратного алгоритма дедубликации

Выводы по главе 2

Система IBM FlashSystem 5035 с 64 ГБ кэш-памяти оптимально поддерживает работу с RAID 6.

RAID совместима с возможностями системы по аппаратному ускорению операций ввода-вывода через Fibre Channel адаптеры 16 Gb/s.

В конфигурации предусмотрены spare-накопители для автоматического восстановления при выходе дисков из строя. Полученная эффективная емкость массива составила 38.16 ТБ при raw-емкости 49.92 ТБ, соответствует ожидаемым показателям для RAID 6.

Глава 3 Разработка эффективной модели системы управления хранением данных на основе RAID-массивов

3.1 Выбор методологии моделирования системы управления хранением данных на основе RAID-массивов

Устройство, создающее запросы ввода/вывода, служит сервером Lenovo x3650 m5. Его технические свойства приведены в таблице 22.

Таблица 22 – Технические характеристики сервера x3650 m5

Компоненты	Спецификация
Форм-фактор	2U Rack.
Процессор	Два процессора Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40GHz
Кэш память	128 ГБ
Объем	4 ТВ
Сетевые интерфейсы	Четыре интегрированных порта Gigabit Ethernet 1000BASE-T (RJ-45); два встроенных порта Ethernet 10 Гбит (SFP+)
Компоненты	Спецификация
Внешние порты ввода/вывода	Четыре порта fibre chanel 16 gbps
Блоки питания	Два блока питания 550 W
Операционная система	Microsoft Windows Server 2012 R2; VMware vSphere (ESXi) 6.0.

Оборудованием для соединения служит коммутатор Brocade 6505. В таблице 23 приведены технические характеристики данного коммутатора.

Таблица 23 – Технические характеристики Brocade 6505

Компоненты	Спецификация
Форм-фактор	1U Rack.
Количество портов	24 шт.
Скорость передачи данных	16 gbps fibre channel
Количество блоков питания	2 шт.

Предметом исследования в данном эксперименте выступают две

системы хранения: Storwize 5030 и Storwize 5100, которые будут поочередно подключаться к серверу. Конфигурации которых приведены в таблицах 24 и таблице 25 соответственно.

Таблица 24 – Конфигурация системы хранения данных Storwize 5030

Компоненты	Спецификация
Форм-фактор	2U Rack.
Процессор	Два процессора Intel Broadwell 1.6GHz
Кэш память	64 ГБ
Количество установленных дисков	12 SSD накопителей объемом 3.84 ТБ
Сетевые интерфейсы	Два встроенных менеджмента интерфейса 10gbps iSCSi (RJ45)
Внешние порты ввода/вывода	Восемь внешних портов Fibre Chanel 16gbps (SFP+)
Блоки питания	Два блока питания 650 W
Операционная система	Spectrum Virtualize 8.4.0

Таблица 25 – Конфигурация системы хранения данных Storwize 5100

Компоненты	Спецификация
Форм-фактор	2U Rack.
Процессор	Два процессора Intel Skylake 1.8GHz
Кэш память	192 ГБ
Количество установленных дисков	24 NVMe SSD накопителей объемом 3.84 ТБ
Сетевые интерфейсы	Два встроенных менеджмента интерфейса 10gbps iSCSi (RJ45)
Внешние порты ввода/вывода	Восемь внешних портов Fibre Chanel 16gbps (SFP+)
Компоненты	Спецификация
Блоки питания	Два блока питания 650 W
Операционная система	Spectrum Virtualize 8.4.0

Для организации сети SAN необходимо четко определить порты ввода/вывода, которые будет использовать коммутатор, поскольку в дальнейшем необходимо будет использовать уникальный идентификатор

порта для построения зон маскировки.

На рисунке 30 изображена схема подключения в SAN среде.

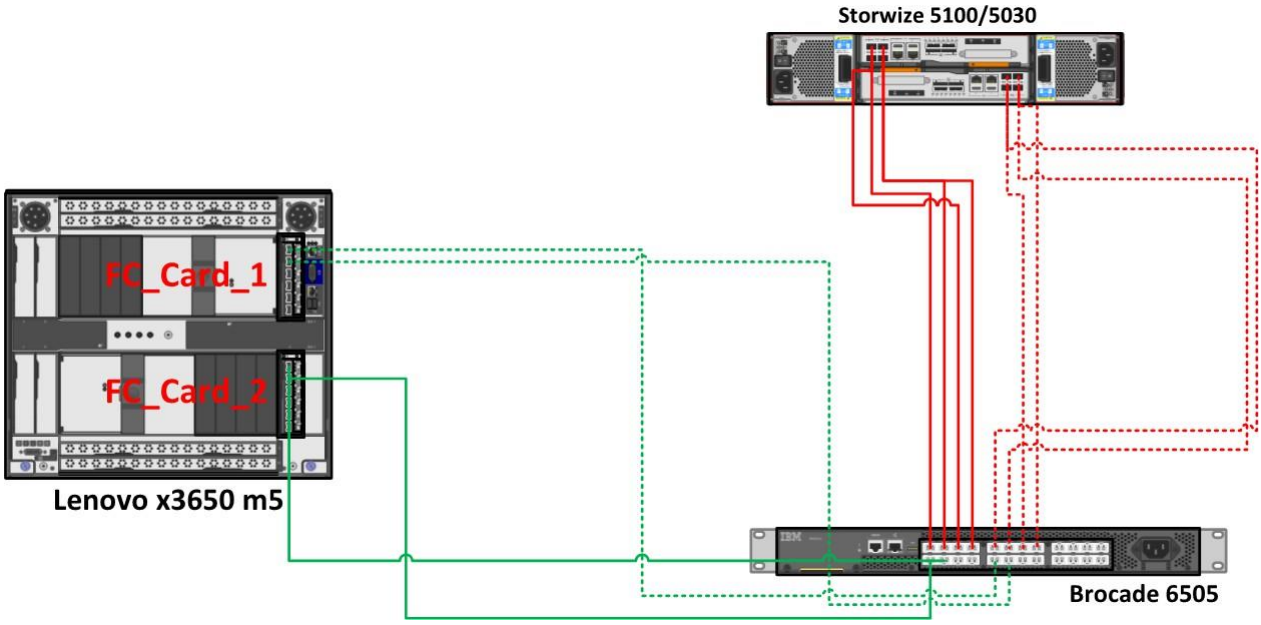

Рисунок 30 – Схема подключения

Согласно схеме, порты №1, 3, 5, 7 подключаются к первому контроллеру системы хранения данных. Порты № 9, 11, 13, 15 соответственно отсоединяются ко второму контроллеру. Первый адаптер сервера подключен к портам № 2, 4. Второй к десятому и двенадцатому порту, соответственно.

3.2 Разработка логической и физической моделей системы управления хранением данных на основе RAID-массивов

Для настройки управления интерфейсом сервера необходимо сначала настроить сетевые параметры порта IMM. Модуль встроенного управления (IMM) сочетает функции сервисного процессора, видеоконтроллера и функции удаленного присутствия в одном чипе. На рисунке 31 изображено меню авторизации IMM.

Рисунок 31 – Меню авторизации IMM

После авторизации пользователь получает доступ к WEB-интерфейсу управления сервером. Благодаря меню «remote control» пользователь может наблюдать за происходящими на сервере процессами в режиме реального времени. Меню «remote control» изображено на рисунке 32.

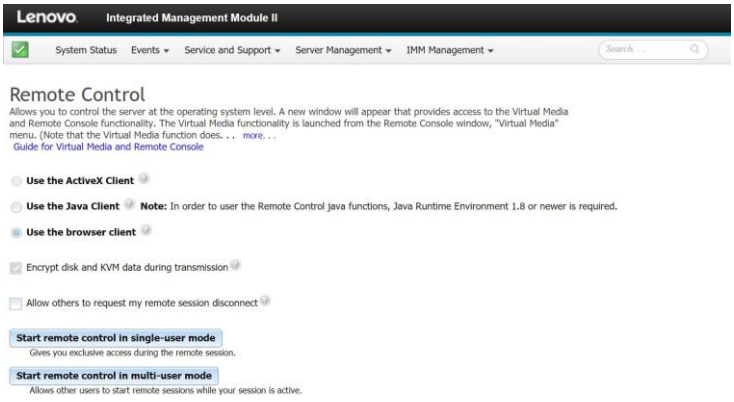

Рисунок 32 – Меню «Remote Control»

Вышеуказанный сервер использует гипервизор VMware ESXi версии 7.0, позволяющий распределять ресурсы сервера на логическом уровне, не привязываясь к аппаратным характеристикам конкретно выделенного элемента сервера. Функционал гипервизора можно использовать только по его Web-интерфейсу, доступ к которому настраивается с графического интерфейса сервера в меню «Configure Management Network». На рисунке 33 показано меню настройки доступа к Web-интерфейсу ESXi.

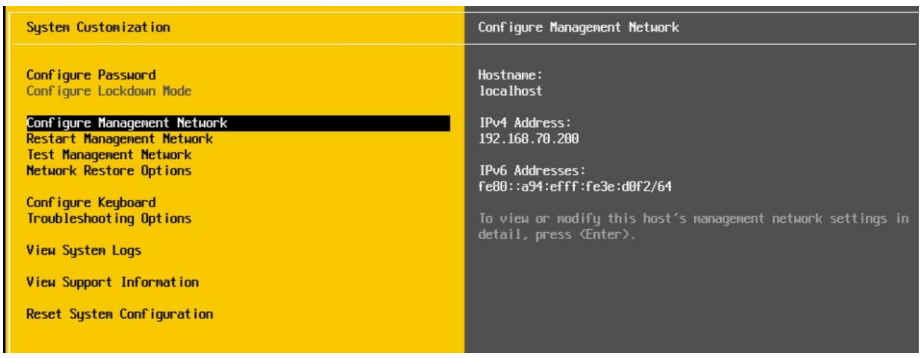

Рисунок 33 – Меню Configure Management Network

После настройки менеджмента интерфейса гипервизора переходим к его Web-интерфейсу и проверяем работу установленной поверх него операционной системы. На рисунке 34 изображен Web-интерфейс ESXi версии 7.0.

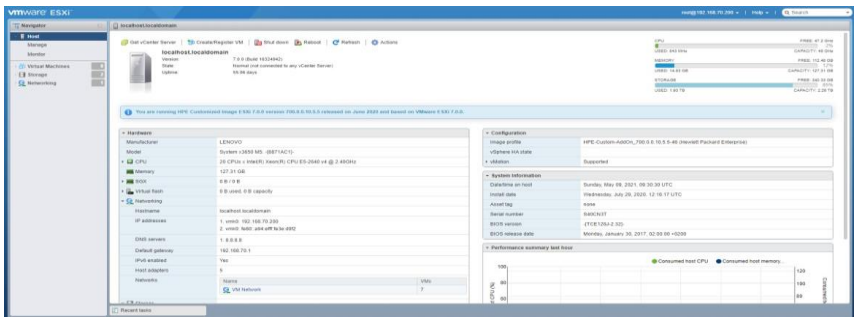

Рисунок 34 – Web-интерфейс ESXi версии 7.0

Список установленных операционных систем можно наблюдать в меню «Virtual Machines», изображенное на рисунке 35.

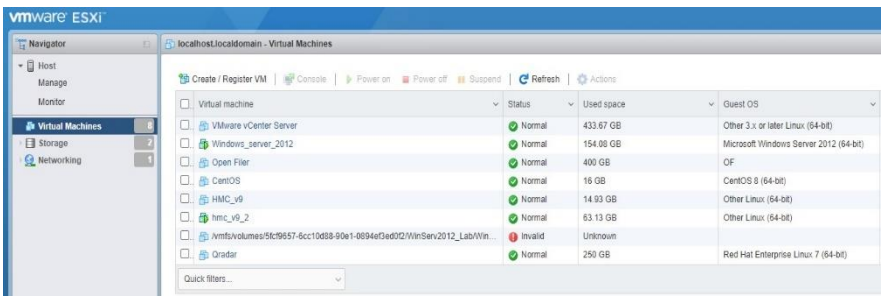

Рисунок 35 – Меню Virtual Machines

В качестве операционной системы, которая будет создавать нагрузку ввода/вывода, был использован Windows server 2012. Для того чтобы работать с данной ОС необходимо выбрать ее из списка всех установленных операционных систем и включить, используя кнопку Power on. На рисунке 36 изображено окно операционной системы.

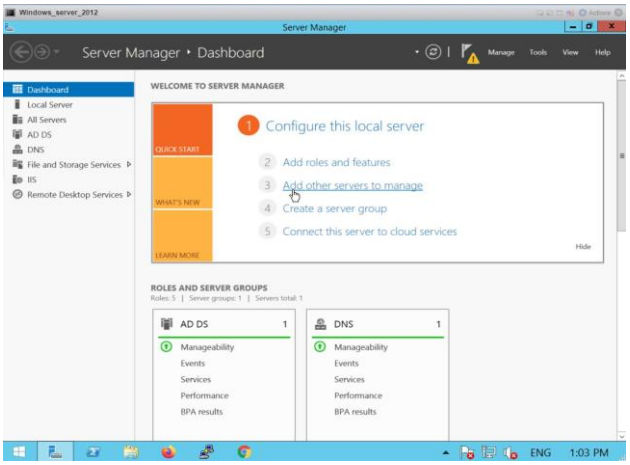

Рисунок 36 – Окно Windows Server 2012

Настройка системы хранения данных происходит при первом запуске СХД. Этот процесс называется инициализацией системы, для его начала необходимо выполнить следующие ниже шаги.

Подключите свой ПК или ноутбук к техническому порту системы хранения, он обозначен буквой T. Убедитесь, что адрес IPv4 получен с помощью DHCP.

Открыть веб-браузер и перейти по ссылкеhttp://install.Браузер автоматически перенаправит пользователя в меню инициализации системы. также можно использовать IP-адрес http://192.168.0.1. Стартовое окно инициализации изображено на рисунке 37.

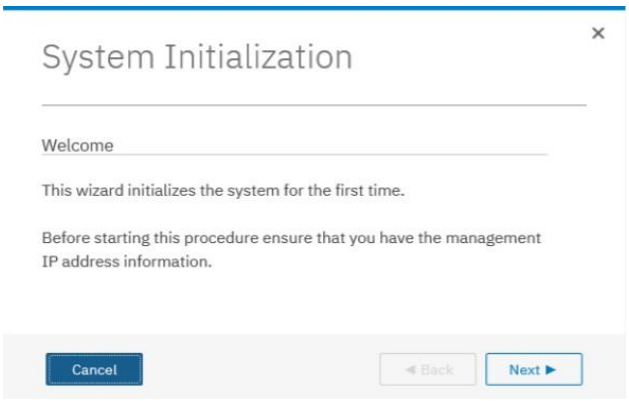

Рисунок 37 – Стартовое окно инициализации

Ввести информацию об IP-адресе управления для новой системы (рисунок 38). Необходимо ввести IP-адрес, маску сети и ее шлюз.

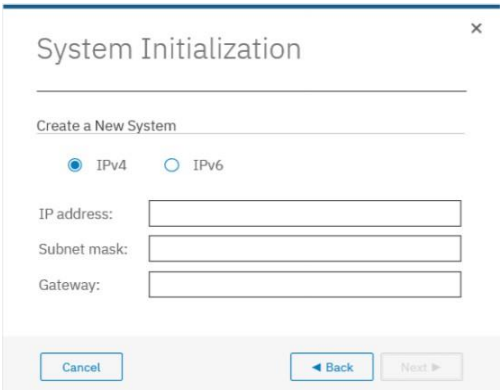

Рисунок 38 – Настройка менеджмент интерфейса

Далее отобразится окно с таймером перезапуска, по истечении которого будет показано окно окончательной инициализации с последующими инструкциями (рисунок 39).

Рисунок 39 – Завершение инициализации системы

После инициализации системы необходимо подключиться к менеджменту интерфейса, который предоставляет полный контроль над системой хранения данных. Менеджмент-интерфейс приведен на рисунке 40.

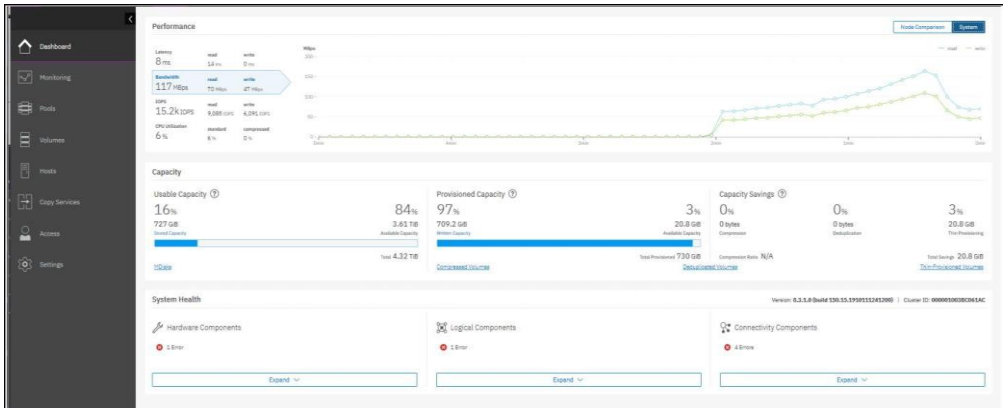

Рисунок 40 – Менеджмент-интерфейс СХД FlashSystem

Зонирование – это функционал коммутатора, позволяющий разделить сеть SAN на логические группы устройств, которые могут получить доступ друг к другу. Зоны обеспечивают контролируемый доступ к сегментам сети и устанавливают барьеры между рабочими средами. Устройство в зоне может общаться только с другими устройствами, расположенными в той же зоне. Когда зонирование включено, устройства, не включенные в одну из конфигураций зон, недоступны для всех других устройств в сети.

Для того чтобы начать создавать зоны, необходимо сначала создать соединение между ПК и коммутатором. Для этого необходимо подключиться к консольному порту коммутатора и создать терминальную сессию благодаря программному обеспечению PuTTy. В таблице 26 приведены настройки для корректной работы терминала.

Таблица 26 – Параметры настройки PuTTy

Параметры	Значение
Скорость передачи	9600
Биты данных	8
Четность	-
Количество стоп-битов	1
Управление потоком данных	-

На рисунке 41 показано окно настроек для консольного подключения.

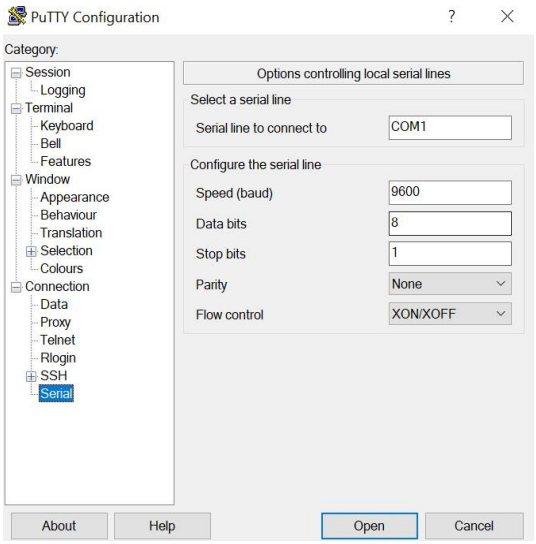

Рисунок 41 – Окно настроек терминала

В рамках данного эксперимента была создана одна зона, которая обеспечивает связь между хостом и СХД. Порты, добавленные в зону. Листинг создания зоны приведен в приложении Б. Для того чтобы создать том хранения данных необходимо сначала создать пул, в состав которого он будет

принадлежать. Процесс создания пула изображен на рисунке 42.

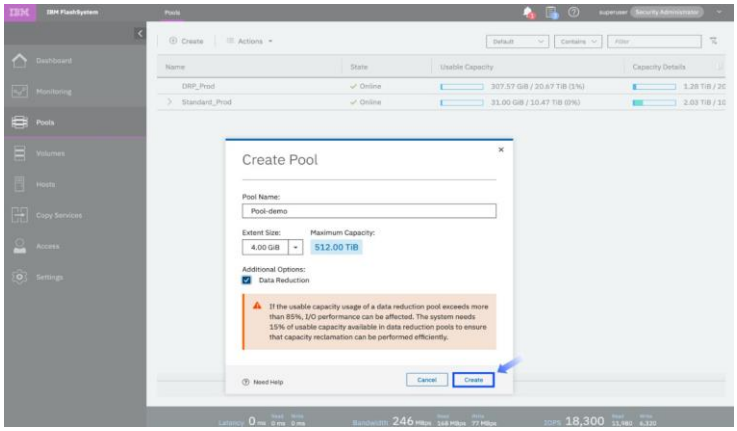

Рисунок 42 – Процесс создания пула

После этого необходимо подсоединить накопители к уже созданному пулу. Следует отметить, что к одному пулу данных можно добавлять разные типы накопителей. IBM FlashSystem 5100 в контроллерной полке поддерживает следующие накопители:

– NVMe Solid State Drive;

– IBM FlashCore module.

Данный процесс изображен на 43.

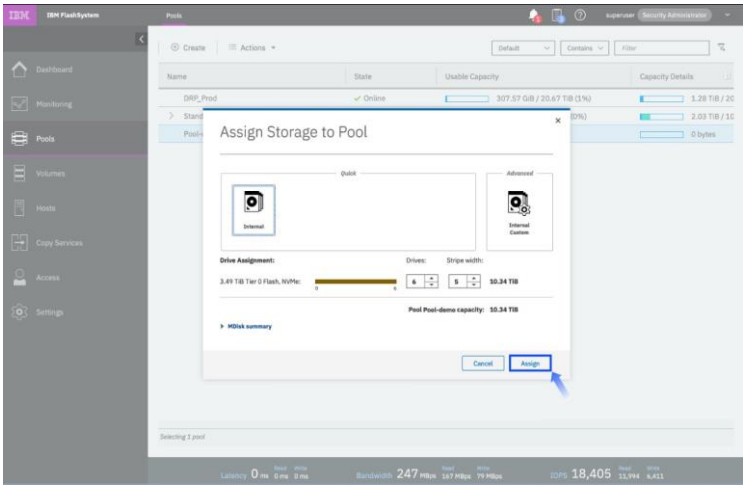

Рисунок 43 – Присоединение дисков к пулу сохранения

После создания пула его доступную емкость можно разделить среди томов хранения данных [7]. Пользователь может создать три типа томов: базовый, зеркальный и пользовательский тип тома. При создании базового тома система автоматически начнет процесс его форматирования, он может занимать довольно длительный период времени. При форматировании пользователь не сможет динамически увеличивать его объем. В случае, когда необходимо отключить вышеуказанный функционал, необходимо создавать пользовательский тип тома, при создании которого необходимо снять галочку с соответствующего чек-бокса. Зеркальный LUN позволяет создать сразу два тома, которые могут быть в разных пулах хранения. Данные, которые будут записываться на главный том, будут автоматически копироваться на второстепенный LUN. Также стоит отметить, что процесс копирования происходит в фоновом режиме, поэтому его влияние на производительность системы хранения минимальных.

На рисунке 44 показано меню создания дисковых томов.

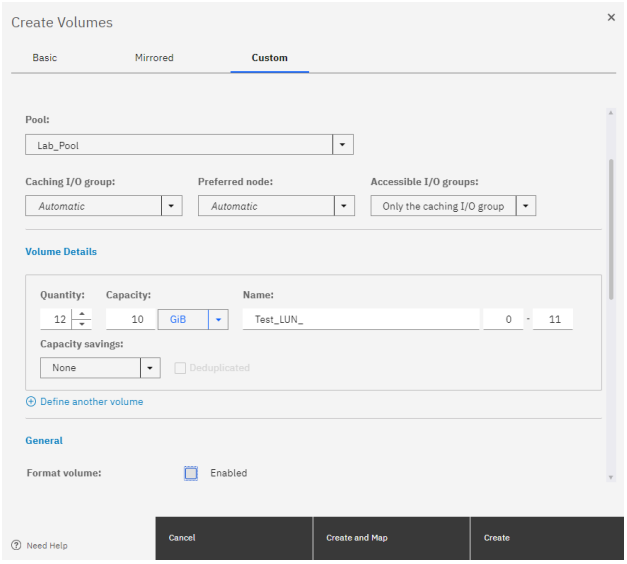

Рисунок 44 – Меню создания дисковых томов

При нажатии кнопки «Create» система выводит на передний план, в котором отображаются консольные команды создания тома. Следует также

отметить, что пользователь может самостоятельно вводить данные команды в командной строке СХД.

Для этого ему необходимо подключиться к ней по SSH протоколу (рисунок 45).

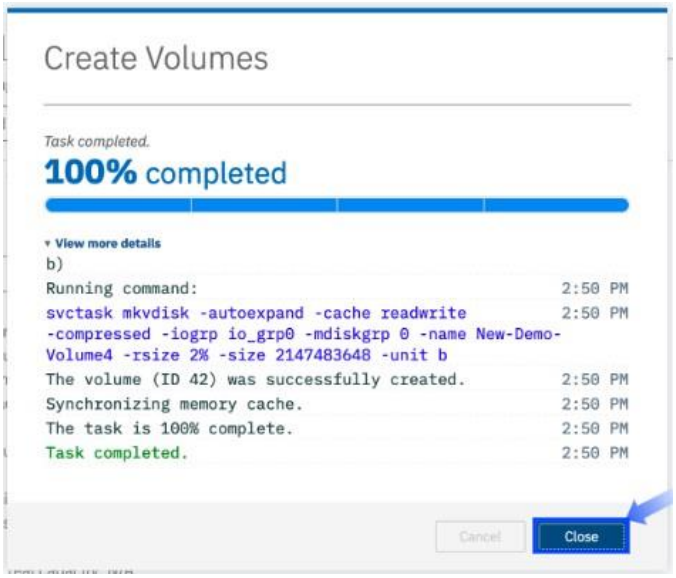

Рисунок 45 – Результат создания томов

Следующим шагом является идентификация хостов и подключение томов хранения к ним. Меню «host», благодаря которому можно презентовать серверы для системы хранения данных, изображено на рисунке 46.

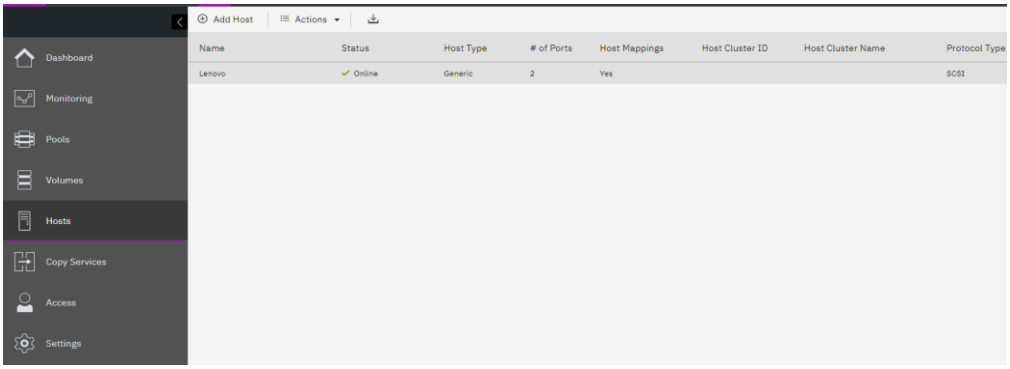

Рисунок 46 – Меню «Host»

После нажатия на кнопку «Add Host» система представляет выпадающее окно, в котором пользователь должен заполнить следующие пункты: name, host connection, host WWN. Параметр name предоставляет серверу уникальное имя, по которому пользователь может находить нужный ему сервер. Host connection отвечает за то, какой протокол передачи данных будет использоваться для обмена данными с хостом. Для параметра Host WWN необходимо указать доступные адреса портов ввода/вывода. Следует заметить, что СХД в большинстве случаев сама определяет все доступные WWN адреса хостов. Для сервера Lenovo 3650 m5 были идентифицированы следующие WWN:

– 21:00:00:24:FF:11:9F:3E;
– 21:00:00:24:FF:11:9E:BB;
– 21:00:00:24:FF:11:9C:2A;
– 21:00:00:24:FF:11:9G:4F.

На рисунке 47 показано меню настройки параметров сервера.

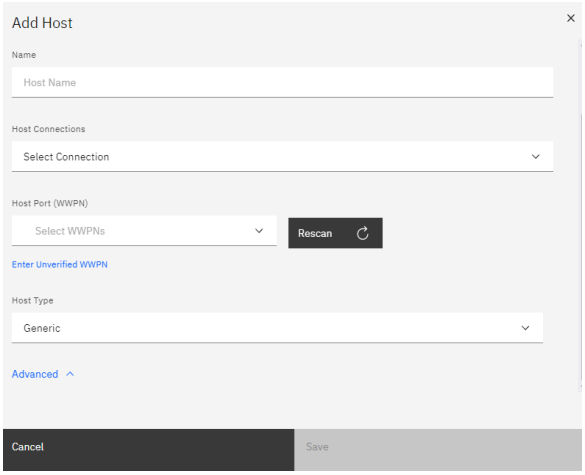

Рисунок 47 – Меню настройки параметров сервера

После того, как связь с сервером была настроена, необходимо начать процесс подключения томов хранения данных (LUN). Для этого необходимо выделить все необходимые LUN, нажав правой клавишей, тем самым будет

вызвано подменю (рисунок 48), в котором необходимо выбрать пункт «Map to host or host cluster».

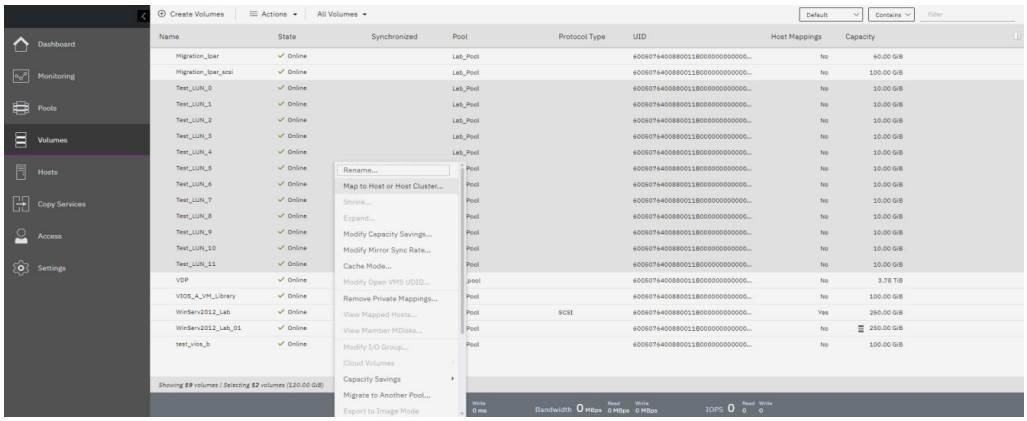

Рисунок 48 – Начало подключения томов хранения

Далее необходимо выбрать в списке ранее созданный хост с именем Lenovo и выбрать каким образом будут создаваться ID номера для томов хранения (рисунок 49). Первый вариант – это когда система в автоматическом порядке раздает ID номера томам хранения. Второй вариант предполагает, что пользователь сам будет указывать ID для каждого LUN в отдельности.

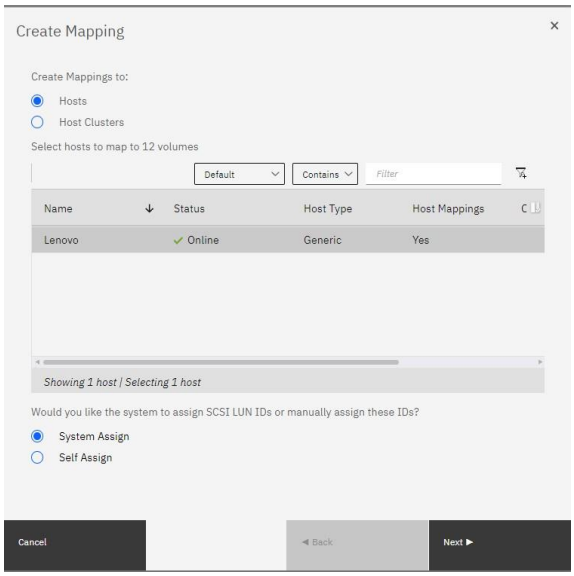

Рисунок 49 – Меню Create Mapping

Результат выполнения программы можно наблюдать в меню Volumes by host, изображенном на рисунке 50.

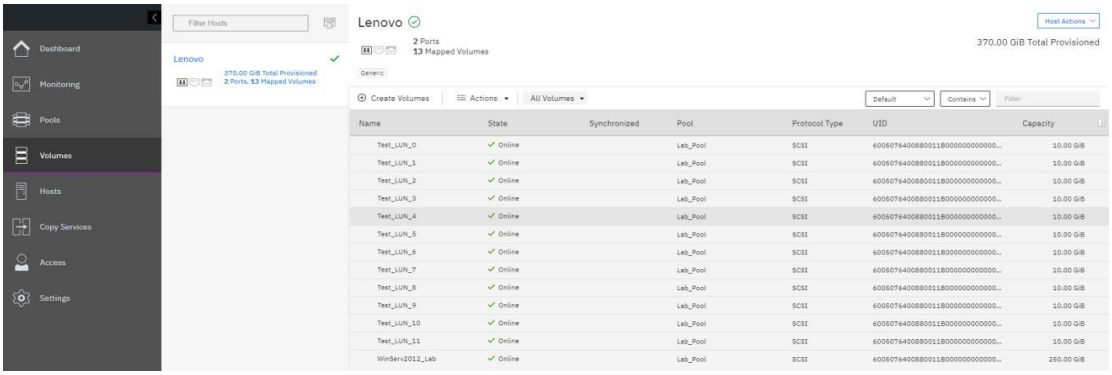

Рисунок 50 – Меню Volumes by host

После того как пользователь убедился, что со стороны системы хранения данных все созданные тома представлены серверу. Необходимо вернуться к Web-интерфейсу гипервизора, чтобы подключить уже подключенное к гипервизору LUN к уже установленной Windows Server 2012. Для этого необходимо перейти в меню настройки операционной системы и нажать кнопку «Add new raw disk» (рисунок 51) и добавить двенадцать созданных томов.

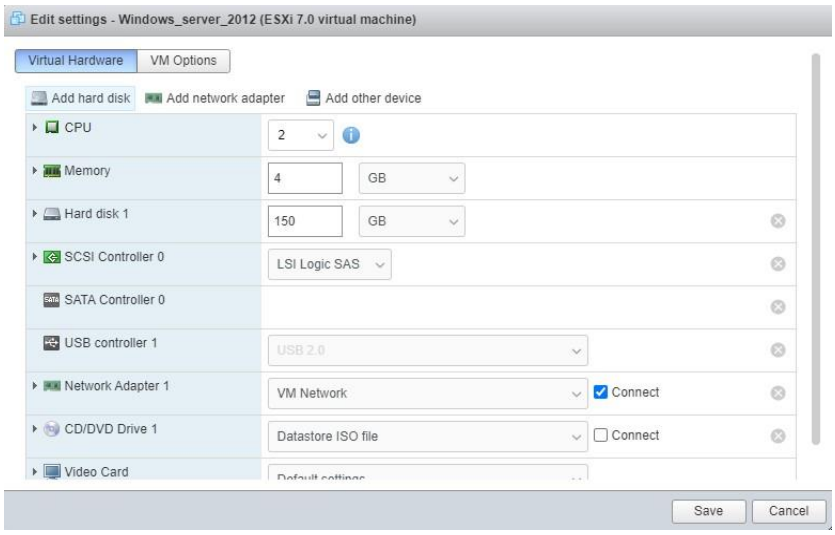

Рисунок 51 – Настройка физических параметров операционной системы

После добавления томов в операционную систему, их необходимо перевести в статус онлайн для того, чтобы программное обеспечение VDbench получило доступ к ним. Для этого необходимо в окне «run» написать следующую команду.

«Diskmgmt.msc», после чего пользователь получит доступ ко всем устройствам хранения данных, которыми способна управлять операционная система. На рисунке 52 изображены меню «Disk Management» и процесс изменения устава накопителей.

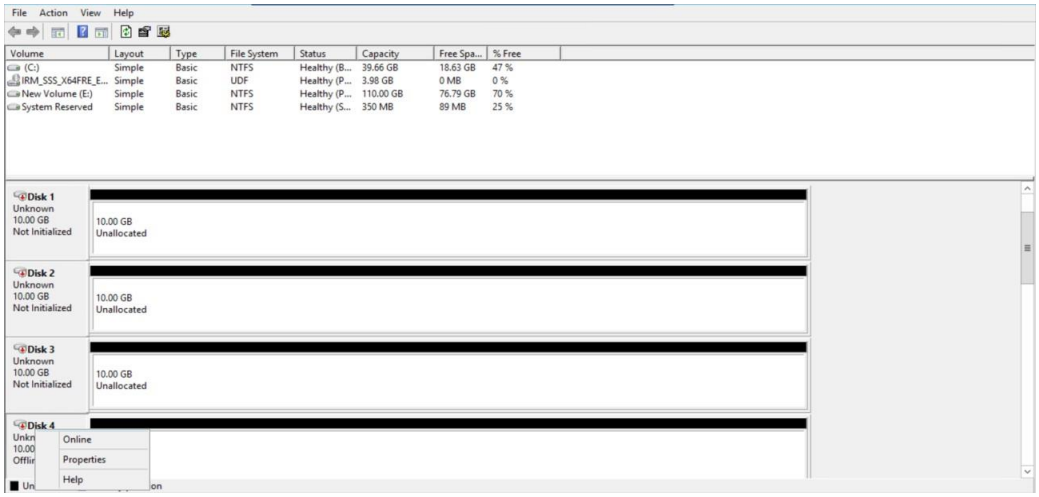

Рисунок 52 – Процесс изменения устава накопителей

После того как все накопители были переведены в требуемый статус, пользователь может начать процесс тестирования системы хранения данных.

3.3 Алгоритмы и их практическая реализация

В качестве программного обеспечения, которое будет создавать нагрузку ввода/вывода на систему хранения данных, был выбран Oracle Vdbench.

Vdbench – это дисковый генератор рабочей нагрузки ввода-вывода, используемый для тестирования и сравнительного анализа существующих и

будущих продуктов хранения. Vdbench написан на Java и поддерживает работу со следующими операционными системами: Solaris Sparc и x86, все версии Windows, HP/UX, AIX, Linux, Mac OS X, zLinux и RaspBerry Pi.

Структура скрипта складывается с трех основных параметров, от настройки которых зависит работа теста:

- storage definition (SD);
- workload definition (WD);
- run definition (RD);
- storage definition (SD).

Этот параметр идентифицирует каждый физический или логический том диспетчера томов или отдельно выделенный файл, используемый в запрашиваемой нагрузке. Конечно, при работе с файлами именно файловая система берет на себя ответственность за все операции ввода-вывода, а Vdbench не будет иметь контроль над физическим вводом-выводом. Параметр SD содержит следующие параметры:

- sd=name. Этот параметр устанавливает уникальное имя любого типа хранилища. Название хранилища в дальнейшем используется параметрами WD и RD для того, чтобы определить, какие носители использовать для своей рабочей нагрузки;
- параметр lun=name описывает имя неразмеченного диска или имя файла. Например, для операционной системы Windows название неразмеченного тома выглядит следующим образом: lun=\\.\PhysicalDrive1, а для выбора файла необходимо прописать путь к нему;
- команда size описывает размер неразмеченного диска или файла. Необходимо ввести это значение в байтах, килобайтах, мегабайтах, гигабайтах или терабайтах. Если размер не указан, данные будут взяты с необработанного диска или файла. Vdbench поддерживает емкость не менее 2 Гб;
- range=(min,max). Данный параметр позволяет ограничить активность

команд ввода вывода в указанном пользователем диапазоне;

- параметр thread определяет максимальное количество одновременных вводов-выводов, которое может быть нерешенным для конкретного тома;
- openflags. Данный параметр позволяет контролировать процесс записи данных, например: операция записи завершается, как только данные хранятся в системном кэше;
- workload definition. Параметры WD описывают, какая рабочая нагрузка должна производиться на введенных логических томах или файлах.

Основными параметрами для настройки workload definition являются:

- rdpct определяет процент выполняемых операций считывания. rdpct = 0 означает, что выполняется 100% операций записи. По умолчанию этот параметр имеет значение 100% для операций считывания;
- параметр xfersize указывает размер блока, который передается от системы хранения данных к серверу. Если этот параметр не был прописан в скрипте, размер блока будет равен четырем килобитам;
- skew определяет процент от общей скорости ввода/вывода, который будет выделен для этой рабочей нагрузки. По умолчанию общий коэффициент ввода/вывода будет равномерно распределен между всеми рабочими нагрузками;
- seekpct определяет частоту генерирования поиска случайного логического блока данных;
- iorate. С помощью специфического для рабочей нагрузки параметра iorate можно указать фиксированные IOPS для конкретной рабочей нагрузки, тогда как другие WD продолжают контролироваться другими параметрами iorate, которые указываются в части RD.

Рассмотрим Run definition (RD).

Параметры RD определяют, какую из ранее определенных рабочих нагрузок (WD) необходимо выполнить, какие IOPS нужно создать на

устройстве хранения (SD) и как долго будут выполняться рабочие нагрузки в рамках тестирования. Для RD выделяют следующие параметры:

– параметр elapsed определяет время, прошедшее в секундах для каждого цикла WD. Каждая запрашиваемая рабочая нагрузка выполняется в течение указанного количества секунд, тогда как подробная статистика интервалов производительности подается через определенное пользователем время, которое необходимо указать в параметре «interval». В конце каждого цикла выводится среднее значение IOPS, время выполнения одной операции считывания/записи, а также скорость передачи данных;
– Warmup. По умолчанию Vdbench исключает первый интервал из общего цикла работы программы. Параметр «warmup=« отвечает за разогрев СХД.

Это связано с тем, что система результатов первых циклов выполнения программы будет не соответствовать действительности, поскольку система сохранения должна сначала равномерно распределить нагрузку относительно всех WD.

3.4 Обсуждение полученных результатов

Для того чтобы начать эксперимент необходимо выполнить вход в командную строку операционной системы и перейти к необходимому каталогу путь, к которому C:\Users\Administrator\Desktop\vdbench50407. После этого пользователю необходимо прописать команду «vdbench –f file_name» для включения скрипта [22].

Результат работы скрипта представлен на рисунке 53.

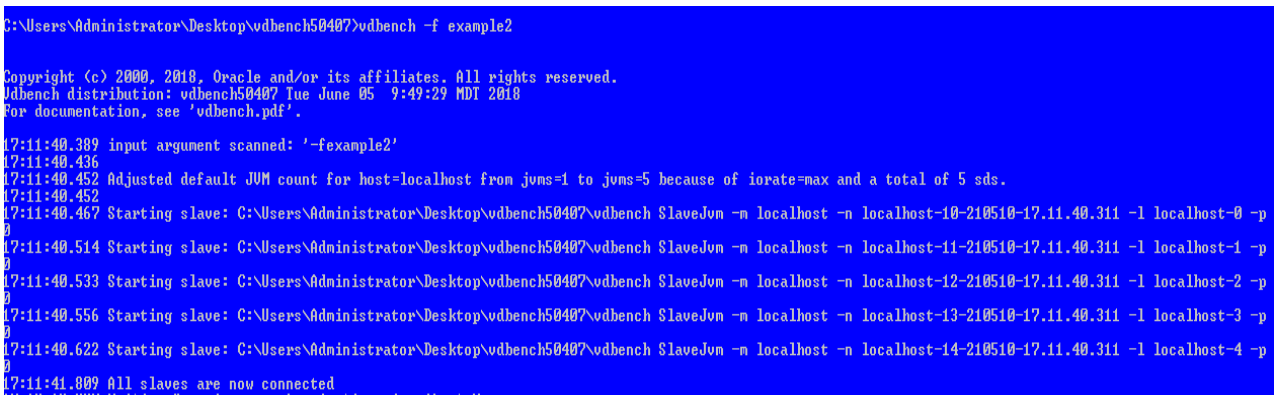

Рисунок 53 – Результат работы скрипта

Результаты теста можно просмотреть в файле totals.html, он находится в папке output (рисунок 54).

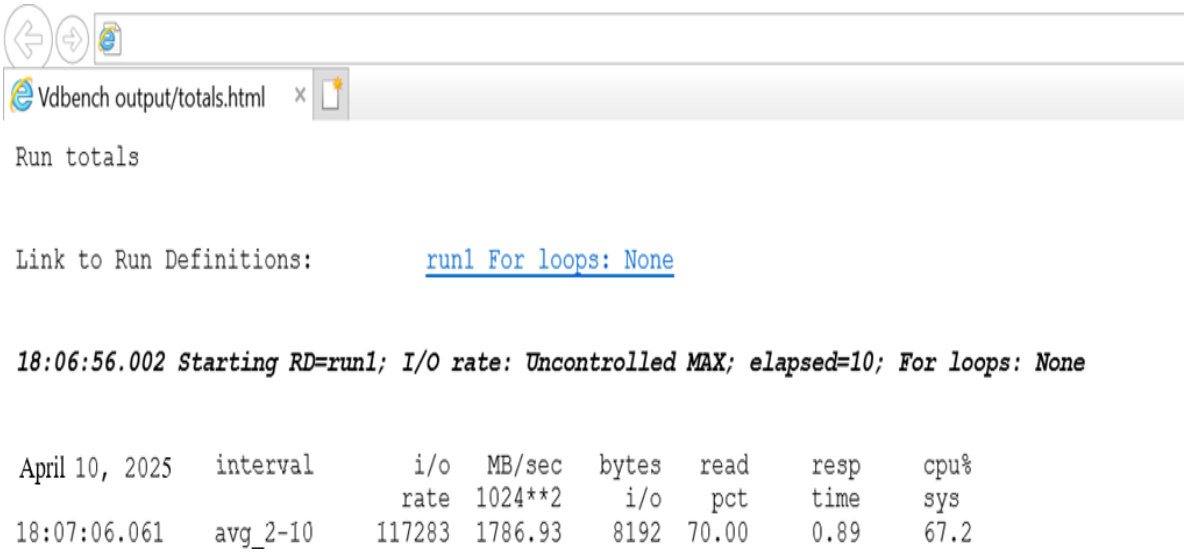

Рисунок 54 – Результаты тестирования

Эти и последующие результаты тестирования будут занесены в соответствующие таблицы, указывающие пиковые IOPS, и время, необходимое для выполнения одной операции. В таблице 27 представлены результаты тестирования системы Storwize 5030 при выключенном функционале компрессии и дедубликации данных.

Таблица 27 – Результаты тестирования системы Storwize 5030 при выключенном функционале компрессии и дедубликации данных

Storwize 5030 без компрессии и дедубликации	
Количество операций ввода/вывода за одну секунду (IOPS), кст/с	117283
Время выполнения одной операции, мс	0,89
Уровень утилизации процессора, %	63

Для того, чтобы система хранения данных имела возможность использовать методы экономии емкости, необходимо создать новый пул хранения данных, аналогично тому, как это было при работе с моделью СХД FlashSystem 5035. На рисунке 55 изображен чек бокс, который необходимо активировать пользователю для создания Data Reduction Pool.

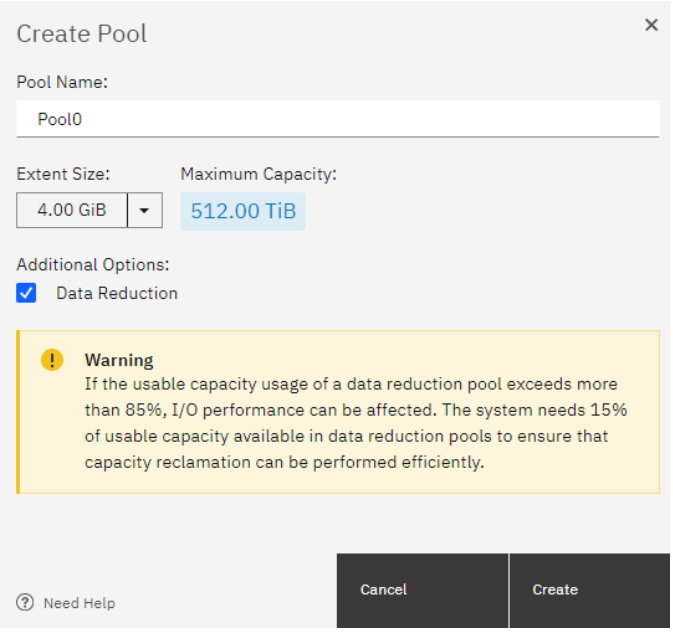

Рисунок 55 – Создание Data Reduction Pool

После того, как был создан DRP, необходимо создать двенадцать новых томов, но уже с включенным алгоритмом компрессии. На рисунке 56 показан пункт меню, который необходимо выбрать, чтобы система работала с использованием компрессии.

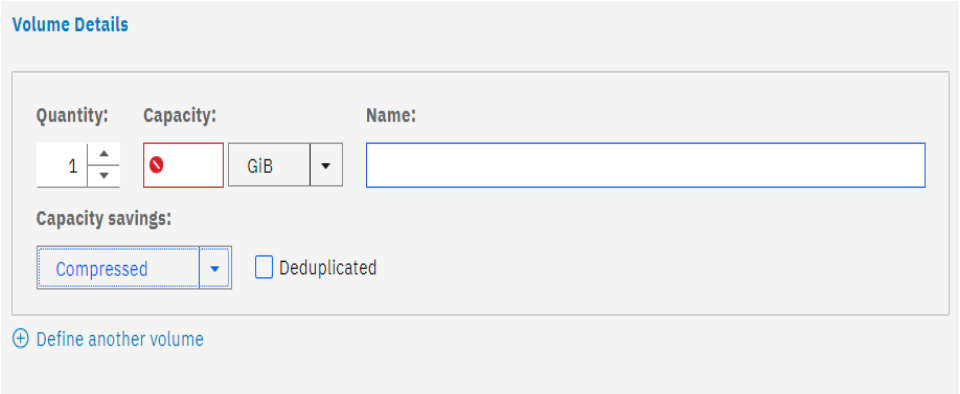

Рисунок 56 – Включение функционала компрессии

После того, как пользователь снова подключил созданные тома к серверу, необходимо повторно выполнить запуск скрипта. В таблице 28 представлены результаты теста СХД 5030 с включенным алгоритмом компрессии.

Таблица 28 – Результаты тестирования системы Storwize 5030 при включенном функционале компрессии данных

Storwize 5030 с включенной компрессией	
Количество операций ввода/вывода за одну секунду (IOPS), кст/с	39486
Время выполнения одной операции, мс	3.8
Уровень утилизации процессора, %	62%

Последним этапом тестирования для Storwize 5030 есть проверка производительности СХД с включенным функционалом компрессии данных.

Для того чтобы активировать вышеупомянутый функционал пользователю необходимо нажать на соответствующий чек-бокс, расположенный рядом с параметром, отвечающим за включение компрессии данных.

Таблица 29 представляет данные с результатом тестирования системы с включенным функционалом компрессии.

Таблица 29 – Результаты тестирования системы Storwize 5030 при включенном дедубликационном функционале данных

Storwize 5030 с включенным функционалом дедубликации данных	
Количество операций ввода/вывода за одну секунду (IOPS), кст/с	19523
Время выполнения одной операции, мс	2.7
Уровень утилизации процессора, %	64%

Для тестирования системы Storwize 5100, пользователь должен повторить эксперимент снова, но уже с другой системой хранения данных. Поскольку все этапы построения стенда уже приведены в пункте 4, можно переходить к результатам тестирования СХД 5100. В таблице 30 представлены результаты тестирования системы Storwize 5100 при выключенном функционале компрессии и дедубликации данных.

Таблица 30 – Результаты тестирования системы Storwize 5100 при выключенном функционале компрессии и дедубликации данных

Storwize 5100 без компрессии и дедубликации	
Количество операций ввода/вывода за одну секунду (IOPS), кст/с	180374
Время выполнения одной операции, мс	0,53
Уровень утилизации процессора, %	59

В таблицах 31,32 соответственно показаны результаты тестирования при работе алгоритмов дедубликации данных.

Таблица 31 – Результаты тестирования системы Storwize 5100 при выключенном функционале компрессии и дедубликации данных

Storwize 5100 без компрессии и дедубликации	
Количество операций ввода/вывода за одну секунду (IOPS), кст/с	124063
Время выполнения одной операции, мс	3.1
Уровень утилизации процессора, %	69

Таблица 32 – Результаты тестирования системы Storwize 5100 при выключенном функционале компрессии и дедубликации данных

Storwize 5100 без компрессии и дедубликации	
Количество операций ввода/вывода за одну секунду (IOPS), кст/с	47275
Время выполнения одной операции, мс	4.2
Уровень утилизации процессора, %	65

Анализируя полученные результаты, можно увидеть, что использование программных алгоритмов компрессии снижает максимальный порог IOPS примерно на 65%, в свою очередь программная дедупликация данных уменьшает производительность системы хранения данных на 75-80%. Это связано с тем, что растет скорость утилизации центрального процессора системы, в связи с дополнительной нагрузкой на него. Следует отметить, что разница между результатами работы моделей и эксперимента составляет примерно 5 - 10%. Допустимое отклонение, поскольку используется разное количество кэш памяти и процессоры разных поколений. Результаты работы системы хранения данных с использованием аппаратных ускорителей показали, что максимальный порог IOPS был снижен на 30% при активированном алгоритме компрессии, для дедубликации соответственно количество операций ввода/вывода за одну единицу времени было уменьшено на 60%.

По результатам проведенных исследований создана методика конфигурирования системы хранения данных, которая состоит из следующих пунктов [5].

Первым шагом необходимо определиться с типом сети хранения данных, поскольку от выбранной сети будет зависеть перечень необходимого оборудования.

Необходимо обратить на выбор протоколов передачи данных и интерфейсов подключения, так как существуют случаи приобретения несовместимых конфигураций. К примеру, передача данных по протоколу

iSCSI может происходить с помощью обычного медного кабеля, использующего интерфейс подключения RJ45, в то же время на сервере могут быть приобретены адаптеры ввода/вывода с интерфейсами SFP+.

Перед тем, как приобрести любую систему хранения данных, необходимо четко понимать рекомендуемые параметры для использованных пользователем приложений. Для примера возьмем базу данных Oracle, для оптимальной работы которой необходимо, чтобы скорость выполнения одной операции не превышала одной миллисекунды [10].

Выводы по главе 3

IOPS является основным показателем производительности системы хранения данных, поэтому перед приобретением СХД необходимо запрашивать построение модели ее работы, это позволит понять, подойдите ли выбранная пользователем система хранения данных для достижения поставленных перед ним целей.

Использование алгоритмов компрессии, дедубликаций для систем хранения данных начального уровня считается целесообразным в случае, если первоочередным критерием для пользователя является именно хранение данных [11]. Пользователи, которые планируют использовать методы экономии емкости высоконагруженных приложений, должны исследовать свои данные на коэффициент сжатия. Вполне вероятна ситуация, что дешевле купить СХД, которая будет иметь лучшие технические характеристики и создаст необходимый запас IOPS даже при использовании алгоритмов компрессии и дедубликации данных.

Роль при работе с системами хранения данных играет ее администрирование, поскольку большинство функционала, предоставляющего СХД, требует тонкой настройки. Поэтому, перед началом работы с системой хранения, пользователю следует пройти подготовительные курсы, которые обычно предоставляются компаниями, которые и являются разработчиками данных аппаратных решений.

Заключение

Магистерская диссертация посвящена актуальной проблеме исследования и разработки математического и программного обеспечения системы управления хранением данных на основе RAID-массивов.

В ходе выполнения магистерской диссертации был проведен анализ существующих архитектур хранения данных: Direct Attached Storage, Network Attached Storage, Storage Area Networks [24].

Отдельный раздел посвящен исследованию и анализу работы моделей, позволяющих оценить влияние использования алгоритмов компрессии и дедубликации данных на СХД. В рамках исследования моделей был приведен и описан алгоритм конфигурирования моделей [12].

В рамках диссертации были проведены эксперименты, для которых был подготовлен программно-аппаратный стенд с использованием следующего оборудования: сервера Lenovo x3650 m5, коммутатора Brocade 6505 и двух систем хранения данных Storwize 5030 и Storwize 5100. Также в ходе экспериментов было настроено зонинг хранения данных.

Особое внимание в эксперименте было уделено конфигурированию программного обеспечения Oracle VDbench, создававшему запросы ввода/вывода на системы хранения данных. Роль при работе с системами хранения данных играет ее администрирование, поскольку большинство функционала, предоставляющего СХД, требует тонкой настройки.

Поэтому, перед началом работы с системой хранения, пользователю следует пройти подготовительные курсы, которые обычно предоставляются компаниями, которые и являются разработчиками данных аппаратных решений.

По полученным результатам была разработана методика конфигурирования системы хранения данных, позволяющая пользователю выбрать оптимальную СХД для решения поставленных перед ним задач.

Гипотеза исследования подтверждена.

Список используемой литературы и используемых источников

1. Арлоу, Д. UML 2 и Унифицированный процесс: практический объектно-ориентированный анализ и проектирование / Д. Арлоу, А. Нейштадт. - М.: Символ, 2015. - 624 с. URL: https://www.litres.ru/book/ayla-neyshtadt/uml-2-i-unificirovannyy-process-prakticheskiy-obektno-orien-24500462/ (дата обращения: 25.04.2025).

2. Баранов А.В., Рыбалко В.А., Жуков В.А. Математическое обеспечение распределенных вычислительных систем. М.: Наука, 2015. - 456 с. (дата обращения: 25.04.2025).

3. Белов, В.В. Проектирование информационных систем: Учебник / В.В. Белов. - М.: Академия, 2018. - 144 с. URL: https://academia-moscow.ru/ftp_share/_books/fragments/fragment_22893.pdf (дата обращения: 25.04.2025).

4. Гвоздева, Т.В. Проектирование информационных систем. Стандартизация: Учебное пособие / Т.В. Гвоздева, Б.А. Баллод. - СПб.: Лань, 2019. - 252 с. URL: https://e.lanbook.com/book/103082 (дата обращения: 25.04.2025).

5. ГОСТ 20886-85 Организация данных в системах обработки данных. Термины и определения.

6. ГОСТ 34.003-90 Информационная технология (ИТ). Комплекс стандартов на автоматизированные системы. Термины и определения.

7. Дадян, Э.Г. Методы, модели, средства хранения и обработки данных: Учебник / Э.Г. Дадян, Ю.А. Зеленков. - М.: Вузовский учебник, 2019. - 176 с URL: https://publications.hse.ru/books/214789491 (дата обращения: 25.04.2025).

8. Жданов А.А., Панкратов А.И. Программное обеспечение RAID-массивов. М.: ДМК Пресс, 2014. - 208 с. (дата обращения: 25.04.2025).

9. Кнут Д. Искусство программирования. Том 3. Сортировка и поиск. М.: Вильямс, 2016. - 800 с. URL: https://www.labirint.ru/books/695679/ (дата

обращения: 25.04.2025).

10. Коннолли, Т. Базы данных. Проектирование, реализация и сопровождение. Теория и практика / Т. Коннолли. - М.: Вильямс И.Д., 2017. – 1440. URL: https://www.labirint.ru/books/579845/ (дата обращения: 25.04.2025).

11. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы: построение и анализ. М.: Издательство «Вильямс», 2019. - 1392 с. URL: https://e-maxx.ru/bookz/files/cormen.pdf (дата обращения: 25.04.2025).

12. Кузнецов И.В., Малашенкова Ю.А. Математические модели и методы в информационных технологиях. М.: Логос, 2019. - 344 с.

13. Макеев А.А., Куликов А.В., Столяров А.В. Распределенные хранилища данных и RAID-технологии. М.: Информационные технологии, 2012. - 344 с.

14. Мартишин, С.А. Проектирование и реализация баз данных в СУБД MySQL с использованием MySQL Workbench: Методы и средства проектирования информационных систем и техноло / С.А. Мартишин, В.Л. Симонов, М.В. Храпченко. - М.: Форум, 2018. - 61 с.

15. Олексеенко М.В., Математическое и программное обеспечение систем управления хранением данных на основе RAID-массивов // 2023 - 7 С.

16. Олексеенко М.В., Системы управления хранением данных на основе RAID-массивов // 2023 - 6 С.

17. Перлова, О.Н. Проектирование и разработка информационных систем: Учебник / О.Н. Перлова. - М.: Академия, 2018. - 272 с. URL: https://academia-library.ru/catalogue/4831/480245/ (дата обращения: 25.04.2025).

18. Петров И.Г., Крюков А.В., Ковалев А.С. Анализ и синтез информационных систем. М.: Бином, 2018. - 448 с.

19. Савин И.В. Особенности применения технологии RAID при создании файловых хранилищ // Журнал «Известия Тульского государственного университета. Технические науки», 2018 - 5 с. URL: https://cyberleninka.ru/article/n/osobennosti-primeneniya-tehnologii-raid-pri-sozdanii-faylovyh-hranilisch (дата обращения: 25.04.2025).

20. Сидоров В.В., Ильин Д.В. Хранение данных в распределенных хранилищах. М.: Издательский дом «Лань», 2016. - 248 с.

21. Стружкин, Н.П. Базы данных: проектирование. практикум: Учебное пособие для академического бакалавриата / Н.П. Стружкин, В.В. Годин. - Люберцы: Юрайт, 2016. – 291 с. URL: https://urait.ru/book/bazy-dannyh-proektirovanie-praktikum-512160 (дата обращения: 25.04.2025).

22. Таненбаум Э., Ван Стивенс В. Распределенные системы. Принципы и парадигмы. М.: Вильямс, 2012. - 1008 с.

23. Терентьева Д.И. Исследование дисковых массивов RAID по параметрам надежности и быстродействия // Международный журнал экспериментального образования. – 2015. – № 3 (часть 3) – С. 423-427. URL: https://expeducation.ru/ru/article/view?id=7191 (дата обращения: 25.04.2025).

24. Шилов Г.Е. Введение в математический анализ. М.: Издательство Московского университета, 2016. - 720 с.

25. Almeida F, Simões J (2019) Moving from waterfall to agile: perspectives from IT Portuguese companies. IJSSMET, 10(1):30–43.

26. Bishop D, Deokar A (2014) Toward an understanding of preference for agile software development methods from a personality theory perspective. In: 47th Hawaii international conference on system sciences, Waikoloa, USA. IEEE, pp 4749–4758.

27. Martin R (2013) Agile software development, principles, patterns, and practices. Pearson, London, 10–17.

28. Mishra A, Abdalhamid S, Mishra D, Ostrovska S (2021) Organizational issues in embracing agile methods: an empirical assessment. Int J Syst Assur Eng Manag 12(6):1420–1433.

29. UML 2 Tutorial - Package Diagram [Электронный ресурс]. URL: https://sparxsystems.com/resources/tutorials/uml2/package-diagram.html (дата обращения 05.04.2025).

30. What is RAID Storage? [Электронный ресурс]. URL: https://www.westerndigital.com/solutions/raid (дата обращения 05.04.2025).