

На правах рукописи

Гантимуров Андрей Павлович

**Анализ и синтез распределенных систем хранения данных**

Специальность 05.13.15

Вычислительные машины, комплексы и компьютерные сети  
(технические науки)

Автореферат  
диссертации на соискание ученой степени  
кандидата технических наук

Москва - 2022

Работа выполнена в Федеральном государственном бюджетном учреждении высшего образования «Московский государственный технический университет имени Н. Э. Баумана (национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана).

Научный  
руководитель:

**Иванов Игорь Потапович**  
доктор технических наук, доцент

Официальные  
оппоненты:

**Соловьев Сергей Юрьевич**  
доктор физико-математических наук, профессор  
Федеральное государственное образовательное  
учреждение высшего образования «Московский  
государственный университет имени М.В.Ломоносова»,  
заведующий кафедрой

**Городничев Михаил Геннадьевич**

кандидат технических наук, доцент

Ордена Трудового Красного Знамени федеральное  
государственное бюджетное образовательное  
учреждение высшего образования «Московский  
технический университет связи и информатики», декан

Ведущая  
организация

Межведомственный суперкомпьютерный центр  
Российской академии наук – филиал Федерального  
государственного учреждения «Федеральный научный  
центр Научно-исследовательский институт системных  
исследований Российской академии наук»

Защита состоится «22» сентября 2022 года в 13 часов 00 минут на  
заседании диссертационного совета Д 999.216.02 при МАИ и МГТУ им. Н.Э.  
Баумана по адресу: 105005 Москва, 2-я Бауманская ул., д. 5, стр. 1, малый зал  
Ученого совета МГТУ (ауд. 316).

С диссертацией можно ознакомиться в библиотеке МГТУ им. Н.Э. Баумана  
и на сайте <http://bmstu.ru>.

Отзывы на автореферат в двух экземплярах, заверенные печатью  
организации, просьба направлять по вышеуказанному адресу на имя ученого  
секретаря диссертационного совета.

Автореферат разослан «\_\_» 2022 года.

Ученый секретарь  
диссертационного совета  
д.т.н., доцент

А.Н. Алфимцев

## Общая характеристика работы

**Актуальность проблемы.** В последние годы экономика стал присущ инновационный характер, т.е. управление интеллектуальным капиталом предприятия теперь играет ключевую роль. Благодаря информационным технологиям происходит глубокая интеграция различных сфер деятельности: исследования, производство, коммерциализация, маркетинг, управление, логистика и т.д. Сейчас информационные технологии без которых невозможно добиться успеха, приобретают глобальный характер практически во всех областях. Грамотное внедрение информационных технологий может привести к улучшению позиции на рынке и увеличить конкурентоспособность предприятия.

Информационные технологии сегодня становятся единой сферой деятельности и жизненно необходимой отраслью. В крупных предприятиях России стала проявляется проблема организации информационного взаимодействия, так как характерные для производства организационные структуры – холдинги, компании и т.д. предполагают сложные взаимосвязи между отдельными, часто финансово независимыми предприятиями.

Одной из основных и важных проблем при разработке и внедрении информационных систем управления производственно-экономической деятельностью всегда являлось значительное увеличение объемов информации с высокой степенью важности, которую конечно же следует тщательно отбирать и аккумулировать, достаточно надежно хранить и оперативно обрабатывать. Разработка информационных систем от всемирно известных вендоров: HP, IBM, BMC, Oracle, SAP, 1С, Microsoft, стала решением выше отмеченной проблемы.

Однако, в ходе развития экономики, основанного на выстраивании взаимодействия многих производителей и поставщиков в рамках систем управления жизненным циклом продукции и соответствующих цепей поставок, создания на этой основе групп предприятий (ГПР), актуализировалась проблема обработки множества различных информационных потоков и часто неформализованной информации как внутри такой группы, так и при взаимодействии с внешними предприятиями. Указанное привело к тому, что анализировать, хранить эти данные и принимать правильные эффективные управленческие решения становится трудно ввиду переизбытка данных и неструктурированности информации.

В настоящее время переход к «цифровой» экономике — это не только инвестиции в ИТ-инфраструктуру. Для интенсивного инновационного развития экономики необходимо создания эффективных комплексных информационных систем на национальном уровне, обеспечивающих высокую реакцию при сборе, анализе и обмене данными о введении новых продуктов и процессов, а также производственных технологиях.

Прогнозируемый мировой объем хранимых данных на 2022 год ориентировочно составляет порядка 40 зеттабайт. При этом доля данных, генерируемых и хранимых в Российской Федерации, составляет порядка 980 экзабайт. Такой рост ставит перед разработчиками задачу нахождения решений, позволяющих осуществлять долгосрочное надежное хранение больших массивов

данных с быстрой скоростью отклика и минимальными задержками передачи информации.

Очевидная сложность информационных систем в группах промышленных предприятий (ГПР) - междисциплинарный характер решаемых задач. Преобладающее наличие западных информационных технологий и инструментов указывают на отсутствие системного подхода к проектированию систем хранения данных (СХД) и недостаточность отечественных методов для решения задачи эффективной организации информационного взаимодействия, надежного хранения и обработки информации в крупной ГПР. Таким образом, вышесказанное обуславливает актуальность темы диссертационного исследования.

**Цели и задачи исследований.** Цель: Создание научных основ для проектирования и разработки распределенных систем хранения данных (РСХД).

Задачи:

- модернизация принципов построения СХД с учетом современных тенденций в архитектуре информационно-коммуникационных систем;
- анализ основных характеристик РСХД, влияющих на ее производительность;
- разработка методики синтеза РСХД для заданного профиля нагрузки.

**Методы исследований.** Для решения поставленных задач используются современные методы теории информации и теоретической информатики, математические методы теории вероятностей, теории очередей и массового обслуживания, планирования эксперимента и статистической обработки результатов экспериментальных данных.

**Научная новизна результатов.** Заключается в:

- предложении новых принципов организации СХД, аналогичных принципам построения коммутируемых компьютерных сетей технологии Ethernet на иерархическом уровне доступа. При этом в качестве клиента сетей рассматривается связка «модуль управления – модуль хранения»;
- разработке методик анализа производительности структурных элементов РСХД и системы в целом с учетом объема дополнительного служебного трафика;
- создании методики синтеза РСХД как многоканальной системы массового обслуживания с отказами для заданного профиля нагрузки на нижнем иерархическом уровне ее построения.

**Положения, выносимые на защиту.** В соответствии с пунктами области исследований паспорта специальности 05.13.15 «Вычислительные машины, комплексы и компьютерные сети», «Разработка научных методов и алгоритмов организации параллельной и распределенной обработки информации, многопроцессорных, многомашинных и специальных вычислительных систем» на защиту выносятся следующие положения:

- технология минимизации времени обработки данных в подсистеме кодирования;

- алгоритм раннего диагностирования отключения жесткого диска в РСХД позволяющий, заблаговременно минимизировать потери общей производительности РСХД;
- методика проектирования РСХД для заданного профиля нагрузки, в частности расчета количества модулей управления и числа модулей хранения, позволяющую рассчитать минимальную конфигурацию модулей управления и модулей хранения;
- методика расчета характеристик служебного информационного потока и коэффициента полезности информационного обмена в РСХД удовлетворяющих потребности пользователей;
- технология подбора оптимальных значений коэффициентов групп приоритетности для предлагаемой структуры РСХД.

**Достоверность полученных результатов подтверждается:**

- Использованием известных методов теории систем массового обслуживания, теории очередей и процессов восстановления;
- Сопоставлением полученных решений с общеизвестными достоверными результатами, опубликованными в литературе;
- Результатами собственных экспериментальных исследований, проведённых в ООО «СК-СХД».

**Практическая значимость** заключается в разработке и совершенствовании моделей, методов и создания инструментария разработки и проектирования корпоративного хранилища данных для информационных систем, интегрирующих предприятия в группы. По сравнению с другими известными решениями данное решение предлагает новый подход к построению корпоративной информационной системы, которая основывается на распределенной системе обработки и хранения данных. Изложены новые научно обоснованные методы и инструментарий создания распределенного корпоративного хранилища данных для группы предприятий, имеющие существенное значение для экономики страны.

**Результаты работы реализованы**, апробированы и внедрены в Федеральной Налоговой Службе (ФНС) России и Центральном Банке Российской Федерации (ЦБ).

**Основанием для выполнения работы** является ряд НИР и ОКР, проведенных ООО «СК-СХД» по исследованию перспективной распределенной системы хранения данных для применения в коммерческих и государственных учреждениях.

**Апробация работы и публикации.** Основные результаты диссертации были представлены на всероссийских и международных научно-практических конференциях: III Международная заочная научно-практическая конференция «Глобальные и локальные проблемы экономики: новые вызовы и решения» (Краснодар, 2012), VI ежегодная всероссийская с международным участием научно-практическая конференция студентов «Человек. Общество. Экономика: проблемы и перспективы взаимодействия» (Пермь, 2012), Materiály IX mezinárodní vědecko - praktická conference «Moderní výmoženosti vědy – 2013» / Díl 13. Ekonomické vědy: Praha. Publishing House «Education and Science» (Прага,

2013), XXXI Международная научно-практическая конференция для студентов, аспирантов и молодых ученых «Актуальные проблемы экономики в XXI веке: причины и решения» (Санкт-Петербург, 2015), XXVI Международная (заочная) научно-практическая конференция «Тенденции развития мировой экономики в XXI веке (Э-26)» (Казань, 2015).

**Публикации.** Основные результаты, полученные при пополнении диссертационной работы, опубликованы в 13 научных работах; из них 12 статей опубликованы в изданиях, рекомендованных ВАК и 1 статья опубликована в издании Web of Science.

**Структура и объем работы.** Диссертация состоит из введения, четырех глав, заключения и списка литературы. Работа изложена на 94 страницах, проиллюстрирована 36 рисунками, 9 таблицами. Список использованной литературы составляет 46 наименований.

### Содержание работы

**Во введении** обоснована актуальность работы, рассмотрена ее необходимость и своевременность, определены цели и задачи исследования, сформулирована научная новизна и выносимые на защиту положения, изложена структура диссертации.

**В первой главе** диссертации показано, что существующая концепция построения СХД имеет следующие преимущества: низкое время вывода продуктов на рынок, возможность проведения большого числа изменений в сжатые сроки, использование открытого аппаратного и программного обеспечения.

Наибольшее распространение имеет архитектура кластера СХД EMC Isilon и IBM XIV, представленная на Рисунке 1.

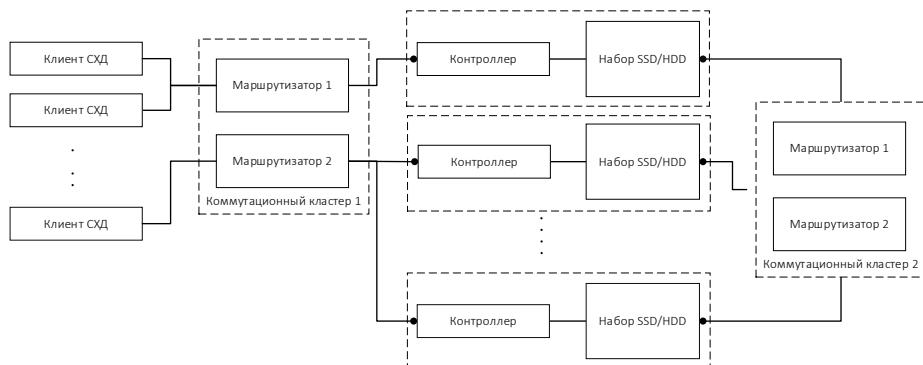


Рисунок 1. Архитектура кластера СХД EMC Isilon и IBM XIV

Вместе с тем, повышение производительности и масштабируемости СХД может быть достигнуто переходом на РСХД. Архитектурное решение РСХД с учетом предлагаемого подхода показано на Рисунке 2.

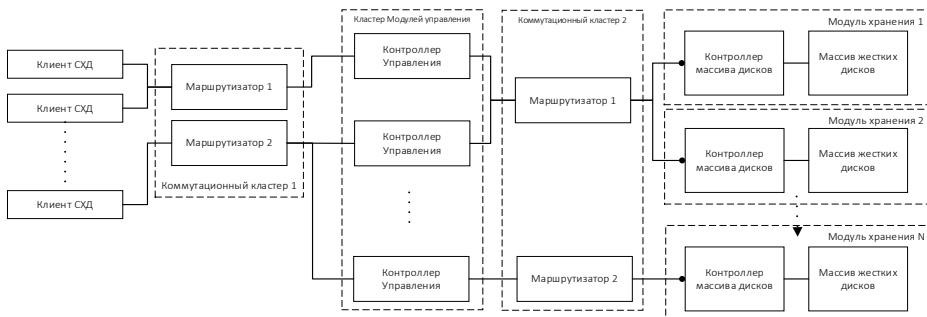


Рисунок 2. Архитектура кластера проектируемой системы хранения данных

Введение дополнительного уровня распределения увеличивает сетевой трафик, снижая производительность системы. Компенсация снижения осуществляется переходом к технологии Ethernet, характеризуемой высокой пропускной способностью интерфейсов, а также возможностью параллелизма и приоритетностью обслуживания на всех уровнях иерархии проектируемой системы.

**Во второй главе** предлагаются методы повышения производительности РСХД:

1. В подсистеме кодирования информации сокращение времени обработки может быть осуществлено использованием недвоичного циклического расширенного кода Рида-Соломона на поле Галуа  $GF(2^8)$ , т.е. осуществлением кодирования байтов информации.

Согласно способу формирования матрицы, увеличивая размер логической области значимой информации, можно добиться размещения блока поступившей от клиента информации в матрице меньшего размера, но при этом будет снижен уровень защиты информации. На Рисунке 3 показана зависимость количества столбцов матрицы от размера области значимой информации.

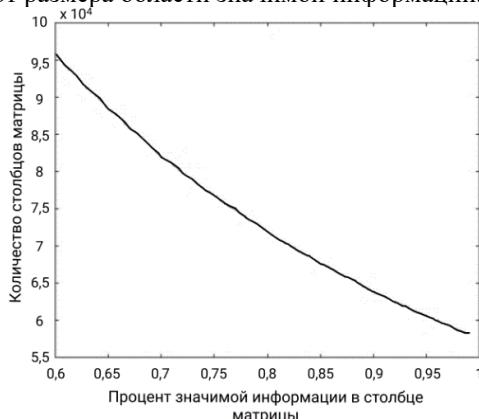


Рисунок 3. Зависимость размерности матрицы от размера области значимой информации в столбце

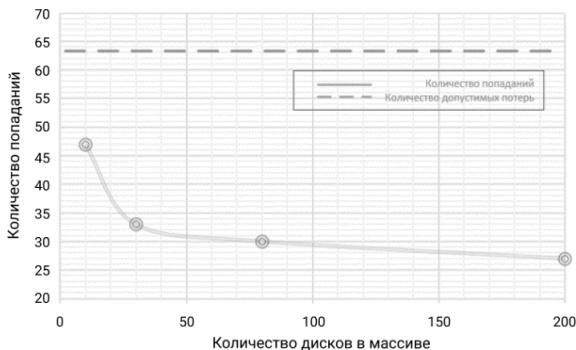


Рисунок 4. Частота попадания данных из 1-го столбца матрицы на диски 1-ого модуля хранения данных, при распределении по алгоритму CRUSH, в системе с десятью модулями хранения данных

Как можно видеть по графику приведенному на Рисунке 4, размер логической области значимой информации 75% является достаточным для обеспечения защиты информации, поскольку потери данных при выходе из строя, например, 1-го модуля хранения не превышают возможности корректирующего алгоритма восстановления информации.

Путем расчета определенной архитектуры РСХД можно добиться снижения затрат времени на кодирование блока информации, полученного от клиента и, тем самым повысить общую производительность РСХД.

2. Увеличение количества модулей хранения в РСХД существенно повысило актуальность обеспечения надежности функционирования жестких дисков, поэтому дополнительно к имеющимся методам предлагается отслеживание динамики заполнения буфера обмена данными между модулем управления и модулем хранения.

Если отслеживать не только заполнение буфера, но и скорость этого процесса, то ситуация появления дефектного диска в массиве жестких дисков выявляется более явно.

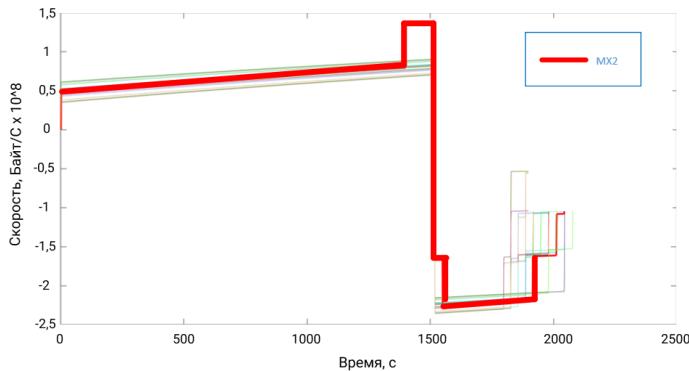


Рисунок 5. Графики скорости заполнения буфера обмена данных между модулем управления и модулем хранения

Ситуация появления дефектного диска в массиве выявляется весьма явно. Прогнозирование появления дефектного диска в массиве предложено осуществлять с использованием математического аппарата нечетких множеств.

3. Производительность РСХД может быть установлена, если рассматривать ее как многоканальную систему массового обслуживания с отказами. При этом заявками на обслуживание являются команды чтения и записи блоков информации.

На Рисунке 6 изображены изменения вероятности отказа  $otk p_i$ , при некотором фиксированном значении интенсивности входящего потока заявок  $\lambda$  для фиксированном числе каналов  $n$ .

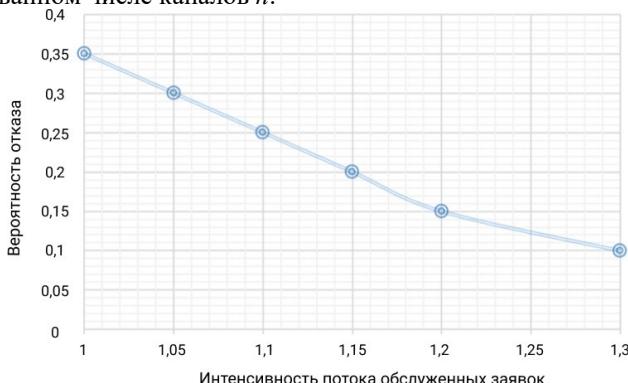


Рисунок 6. График изменения вероятности отказа при некотором фиксированном значении интенсивности входящего потока заявок  $\lambda$  для фиксированном числе каналов  $n$

На Рисунке 7 изображен график изменения абсолютной пропускной способности системы массового обслуживания  $A$ , при некотором фиксированном значении интенсивности входящего потока заявок  $\lambda$  и фиксированном числе каналов  $n$ .

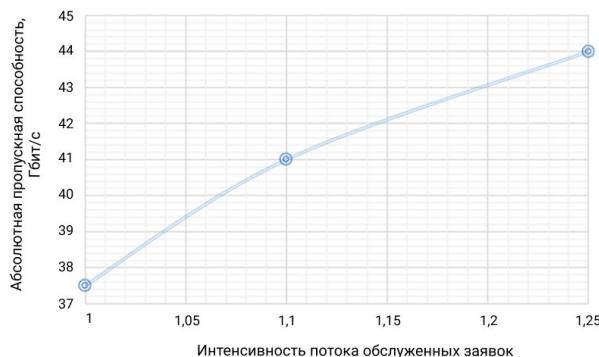


Рисунок 7. График изменения абсолютной пропускной способности входящего потока заявок  $\lambda$  при фиксированном числе каналов  $n$

Далее определим величину интенсивности потока обслуженных заявок на 1 канал,  $\mu$ .

$$\mu = \frac{1}{R * {}^R Lat_{mean} + W * {}^W Lat_{mean} + M(T_{rtt})}, \quad (1)$$

где:  $R$  - процент запросов на чтение данных;

$W$  - процент запросов на запись данных;  ${}^W Lat_{mean}$  - средняя величина задержки записи;  ${}^R Lat_{mean}$  - средняя величина задержки чтения;  $M(T_{rtt})$  - математическое ожидание случайной величины времени прохождения пакета в сети  $T_{rtt}$ .

Определив согласно пользовательского профиля нагрузки среднюю величину интенсивности входящего потока заявок  $\lambda$ , на базе теории многоканальных систем массового обслуживания с отказами можно получить характеристики эффективности системы для определенных групп приоритетности путем оптимизации настроек пропускной способности для выделенной группы приоритетности и архитектурных решений, позволяющих влиять на величину интенсивности потока обслуженных заявок.

4. Соответствие структуры РСХД известному профилю нагрузки

При формировании профиля нагрузки учитываются следующие параметры:

- процентное соотношение операций чтения;
- процентное соотношение операций записи;
- режимы процессов чтения/записи (последовательный или случайный);
- размеры информационного блока данных,
- объем проводимых операций чтения и записи.

Кривая зависимости минимально необходимых жестких дисков, требуемых для обеспечения среднего времени ожидания по профилю, от величины средней скорости чтения блоков данных с жесткого диска показана на Рисунке 8.

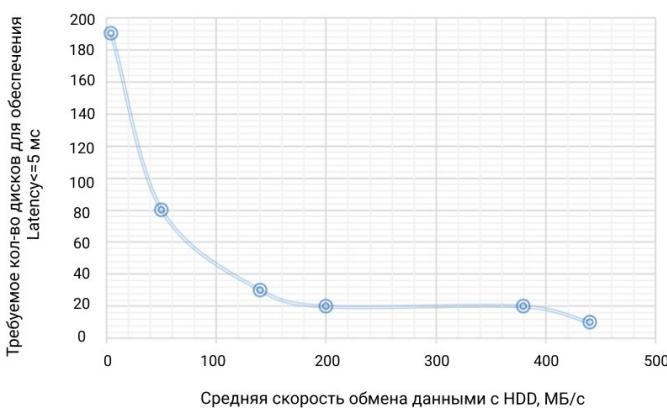


Рисунок 8. Зависимость минимально необходимого количества жестких дисков от величины средней скорости чтения блоков данных с жесткого диска

График зависимости производительности обмена данными между клиентом и РСХД приведен на Рисунке 9.

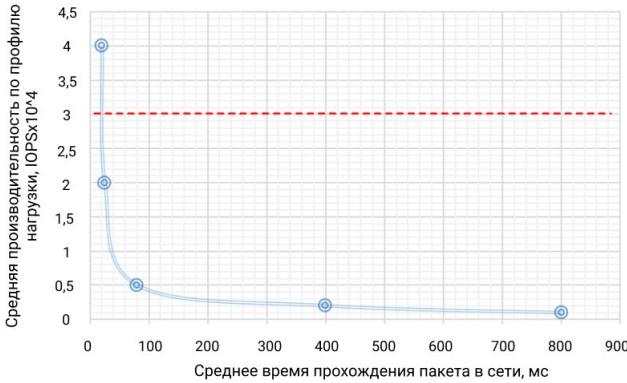


Рисунок 9. Зависимость производительности обмена данными между клиентом и распределенной системой хранения данных с 100 SSD от величины среднего времени прохождения пакета

РСХД наиболее эффективна на высокопроизводительных дисковых носителях, каковыми могут являться, например, SSD-диски. Производительность процессов может измеряться как в Мегабайт/секунду, так и в единицах iops (in-out per second). Примем за единицу измерения производительности информационного обмена данных – iops.

При синтезе РСХД одним из важных параметров является средняя величина задержки в процессе чтения пользовательских блоков с дисков.

С учетом архитектуры РСХД, величина  ${}^RLat_{mean}$  может быть определена следующим образом:

$${}^RLat_{mean} = M(T_{connect}) + \max_i \left( \frac{255 * k_{RS} * blk_{SWARM}}{\min(N_{disk} * N_{module}, 255 * k_{RS}) * disk_{v_i}} \right) + M(T_{repair}) + M(T_{conv}) + M(T_{CRwait}) + Lat_{serial} , \quad (2)$$

где:  $M(T_{connect})$  - математическое ожидание случайной величины времени установки соединения между клиентом и модулем управления  $T_{connect}$ ;  $M(T_{repair})$  - математическое ожидание случайной величины времени работы процедуры восстановления данных в матрице преобразования  $T_{repair}$ ;  $M(T_{conv})$  - математическое ожидание случайной величины времени работы процедуры преобразования хранимых блоков данных в пользовательский блок данных  $T_{conv}$ ;  $M(T_{CRwait})$  - математическое ожидание случайной величины времени пребывания пользовательского блока данных в «Кэше Чтения» модуля управления РСХД  $T_{CRwait}$ ;  $T_{readblk}$  - время чтения блоков данных с жестких дисков и сбора матрицы преобразования  $A_{RS}$ ;  $blk_{SWARM}$  - размер блока РСХД;  $k_{RS}$  - безразмерный коэффициент, с помощью которого устанавливается размерность матрицы преобразований  $A_{RS}$ ;  $N_{disk}$  - количество жестких дисков, подключенных к модулю хранения;  $Lat_{serial}$  - задержка сериализации коммутатора;  $disk_{v_i}$  - скорость записи и чтения данных с диска.

Принимая значение  $disk_v_i$  постоянным и варьируя значением величины  $N_{module}$  можно получить серию калибровочных кривых зависимости величины  ${}^RLat_{mean}$  от величины  $N_{module}$  при различных фиксированных величинах  $k_{RS}$ .

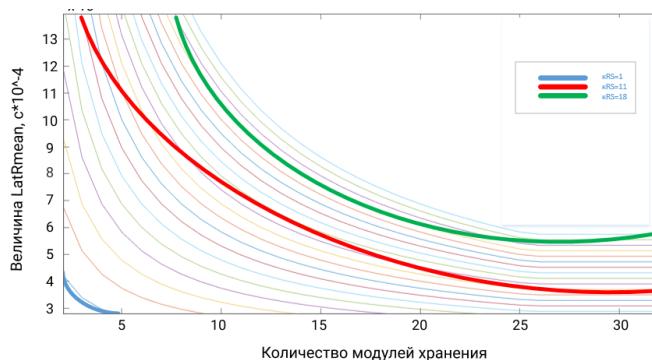


Рисунок 10. Зависимость  ${}^RLat_{mean}$  от количества модулей хранения

Количество модулей управления определяется по формуле:

$${}_{cont}^{\lambda}N = \frac{thread_{user}}{thread_{cont}} , \quad (3)$$

где:  $thread_{user}$  - минимальное необходимое количество независимых потоков данных;  $thread_{cont}$  - Количество независимых потоков обработки данных модуля управления.

Поскольку предполагается, что количество модулей управления должно быть парным, в качестве величины  ${}_{cont}^{\lambda}N$  берется ближайшее четное число, ограничивающее эту величину сверху.

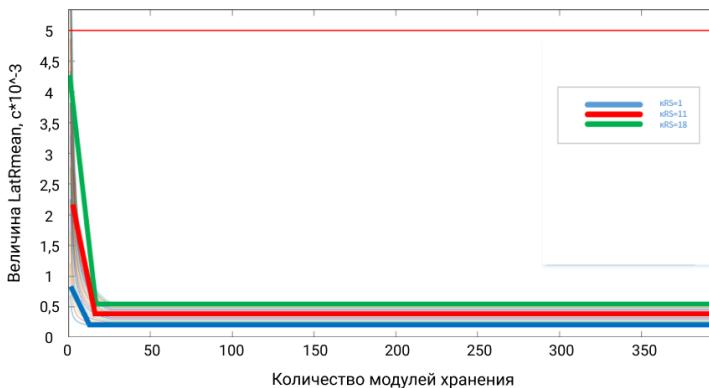


Рисунок 11. Графики зависимости  ${}^RLat_{mean}$  от величины  $N_{module}$  и визуализация требований по задержке в 5 мс

Соответственно, при фиксации величины  $k_{RS} = i$  получим формулу для вычисления количества модулей хранения:

$$N_{module} = {}^{Lat_i}N_{module}$$

Желательную величину  $\frac{sec}{A_N} N$  можно оценить через величину интенсивности входного потока заявок  $\lambda$ , по следующей формуле

$$A_{RS}^{\sec}N \geq \frac{\lambda^{*} \frac{blk_{user}}{blk_{SWARM}}}{(255 - \frac{addN}{elementN}) * k_{RS}}, \quad (4)$$

где:  $element^{add}N$  - количество требуемых элементов «избыточных» данных в столбце.

В нашем случае, фиксируя величину  $loose module N = 1$  (ожидаемое количество одновременно вышедших из строя серверов хранения, входящее в  $add element N$ ), получим следующие кривые величины  $add element N$

**В третьей главе** изложена методика синтеза РСХД для заданного профиля нагрузки, блок схема которой приводится на Рисунке 12.

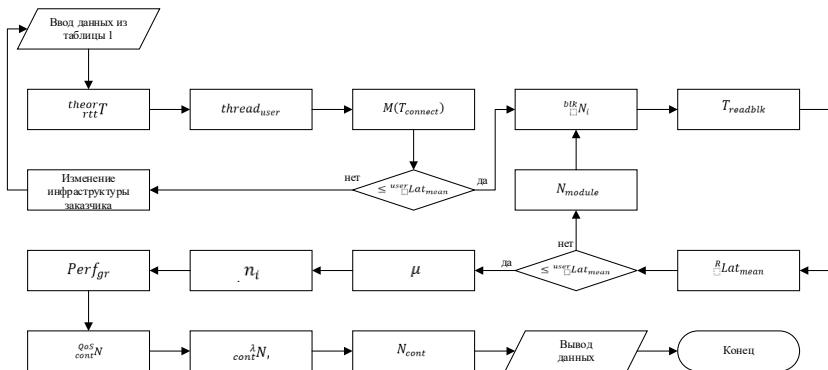


Рисунок 12. Блок-схема методики синтеза РСХД для заданного профиля нагрузки

Установлен перечень дополнительных параметров инфраструктуры заказчика, требуемых разработанными методиками анализа и синтеза РСХД.

РСХД взаимодействует с различными клиентами и объём информации между системой и этими клиентами можно характеризовать как полезный трафик. Также существует и служебный трафик, под которым понимается количество дополнительной информации на каждый пользовательский блок данных для обеспечения требований надежности и отказоустойчивости.

На Рисунке 13 представлена блок-схема предлагаемой методики расчета.

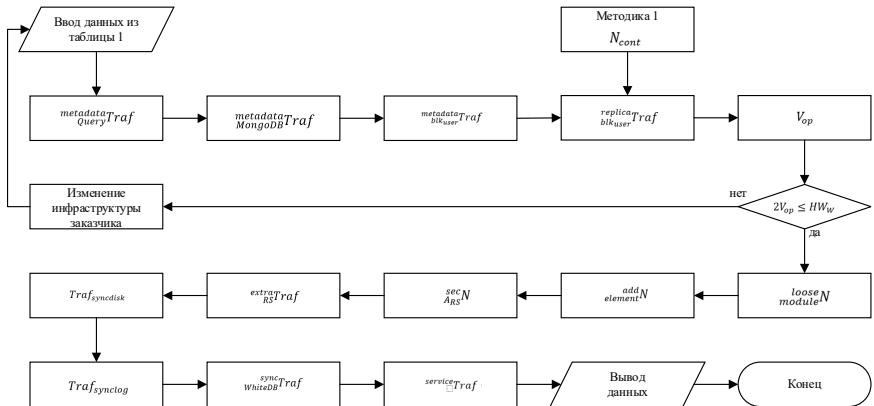


Рисунок 13. Блок-схема расчета характеристик служебного информационного потока в РСХД

В Таблице 1 приведено описание параметров блок-схем.

Таблица 1.

Описание параметров блок схемы

1.	$theor_{rtt}T$	Минимальное необходимое время обмена tcp-пакетом данных по сети, требуемое для обеспечения производительности информационного обмена
2.	$thread_{user}$	Минимальное необходимое количество независимых потоков данных
3.	$Lat_i N_{module}$	Минимальное необходимое количество модулей хранения
4.	$\mu$	Величина интенсивности потока обслуженных заявок на один канал
5.	$n_i$	Минимальное необходимое количество независимых потоков (каналов) $n_i$ обслуживания, для нагружочных групп потоков данных
6.	$thread_{Perfgr}$	Количество независимых потоков, зарезервированных в РСХД, необходимых для обработки заявленного входного потока заявок
7.	$Perfgr N_{cont}$	Минимальное необходимое количество модулей управления, для обеспечения требований производительности
8.	$N_{cont}$	Минимальное количество модулей управления РСХД
9.	$N_{module}$	Минимальное количество модулей хранения РСХД

Предложена оценка производительности РСХД коэффициентом полезности информационного обмена, т.е. долей трафика полезной информации в общем его объёме, которую расчитать по следующей формуле:

$$E = \frac{SWARM_R}{SWARM_R + serviceTraf}, \quad (4)$$

где:  $SWARM_R$  - ресурс, затрачиваемый на поддержание общего информационного обмена;  $serviceTraf$  - дополнительный служебный трафик.

**В четвертой главе** описана схема проведения эксперимента, а также перечень программного и аппаратного его обеспечения.

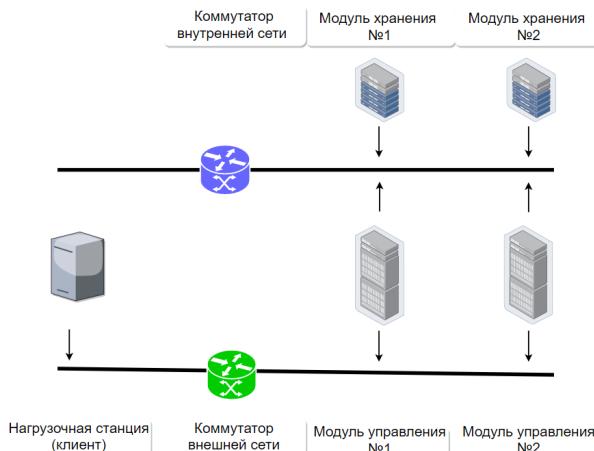


Рисунок 14. Схема подключения аппаратного обеспечения эксперимента

Основным программным обеспечением был выбран пакет vdbench.

Запуск vdbench осуществляется командой :

`vdbench -f vdbench.cfg -o log_file.log`

где vdbench.cfg - это имя файла конфигурации нагрузки, содержание которого представлено ниже:

`report_run_totals=yes;sd=sd1,lun=/dev/INSERT_DEV_NAME,openflags=o_d  
irect,threads=8,size=1T;wd=wd1,sd=sd*,seekpct=random,xfersize=4k,rdpct=65;rd=rd  
1,wd=wd1,iorate=max,elapsed=12h,interval=5,warmup=0,threads=8.`

В результате тестирования величина IOPS оказалось равной 290 000. Погрешность определения заданной при синтезе производительности РСХД (равной 300 000 iops) от экспериментально установленной составляет 3.3 %, что дает основание заменить дорогостоящие натурные испытания экспериментами на предлагаемой цифровой модели РСХД, которая включила в себя все исследования и разработки данной диссертации. Получено свидетельство о государственной регистрации программы для ЭВМ № 2017612060 «Компьютерная программа для организации распределенной системы хранения и анализа информации «BAUMAN ALYTICS».

На подтвержденной цифровой модели РСХД возможен анализ процесса обеспечения качества обслуживания (QoS) информационного обмена.

На Рисунке 15 представлены упрощённые результаты математического моделирования работы РСХД с QoS информационного обмена в сети с максимальной пропускной способностью канала связи «Клиент – Модуль» управления равной 50 Гбит/с. QoS в данной работе понимается как приоритет выделения ресурсов РСХД для определенной задачи. При моделировании были заданы 3 группы приоритетности с коэффициентами 0.6, 0.25 и 0.15. Критичной считалась только первая группа приоритетности с коэффициентом 0.6. Из Рисунка 15 можно видеть, каким образом происходит выделение ресурсов РСХД, согласно технологии QoS, на начальном этапе и перераспределение ресурсов между группами приоритетности по мере высвобождение мощностей.

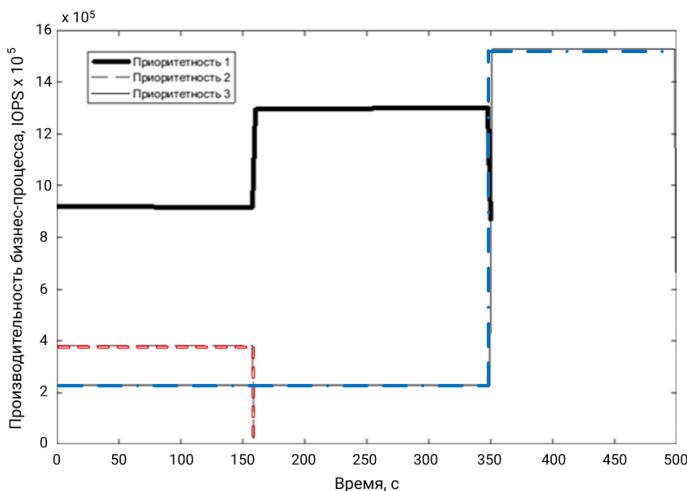


Рисунок 15. Результаты математического моделирования работы РСХД с QoS

Возможность повышения производительности системы путем подбора оптимальных значений коэффициентов групп приоритетности разработанной технологии QoS, предсказываемая цифровой моделью РСХД, подлежит подтверждению экспериментами с реально функционирующими системами, планируемыми в ООО «СК-СХД».

### Основные результаты и выводы

Предлагаемая архитектура РСХД базируется на концепции Web-Scale IT с акцентом на горизонтальное масштабирование при наличии большого количества вычислительных и коммуникационных устройств, объединенных в кластер. Недостаточная гибкость и масштабируемость существующих СХД может быть устранена разделением в одном решении модулей управления и модулей хранения. Увеличение сетевого трафика, неизбежное при подобном разделении, компенсируется внедрением технологии Ethernet с высокой пропускной способностью каналов связи и интерфейсов узлов РСХД. Переход к технологии Ethernet не только даст возможность применения имеющегося на рынке оборудования для построения компьютерных сетей, но и внедрения механизмов обеспечения QoS для повышения надежности и производительности

разрабатываемой РСХД. Этот вывод подтвержден результатами экспериментов с цифровой моделью, разработанной для новой архитектуры РСХД в программной среде MatLab. Эти же результаты подтвердили возможность повышения производительности системы путем расчета коэффициентов групп приоритетности в технологии QoS.

Разделение модулей управления и модулей хранения потребовало проведение анализа динамики заполнения буферов обмена данными с использованием математического аппарата теории нечетких множеств, дающего возможность раннего диагностирования ситуации появления дефектного диска в массиве жестких дисков в процессе записи хранимой информации в режиме реального времени с заблаговременной инициализацией механизма минимизации потерь общей производительности РСХД.

Рассмотрение РСХД как многоканальной системы массового обслуживания с применением теории очередей позволило установить минимально необходимое количество независимых потоков обслуживания для выделенных групп приоритетности.

При заданном профиле нагрузки предложена методика проектирования РСХД, базирующаяся на определении конфигурации множества данных инфраструктуры пользователя, в соответствии с которой, устанавливается количество модулей управления и число модулей хранения с их минимальной конфигурацией. Разработана также методика расчета характеристик служебного информационного потока, обеспечивающего функционирование РСХД. При этом предложено оценивать эффективность системы коэффициентом полезности информационного обмена.

#### **Список основных публикаций**

1. Nelub V., Gantimurov A., Borodulin A. Economic analysis of data protection in systems with complex architecture using neural network methods. *Economic Annals-XXI*, 185(9-10), P. 178-188. URL: <https://doi.org/10.21003/ea.V185-17> (дата обращения: 24.02.2022) (0,5 п.л. / 0,4 п.л.).
2. Гантимуров А.П., Разработка методов и инструментов информационного взаимодействия промышленных предприятий на основе коммуникативной системы // Вестник ВУИТ. 2012. URL: <https://cyberleninka.ru/article/n/razrabotka-metodov-i-instrumentov-informatsionnogo-vzaimodeystviya-promyshlennyh-predpriyatiy-na-osnove-kommunikativnoy-sistemy-1/viewer> (дата обращения: 24.02.2022) (0,5 п.л. / 0,4 п.л.).
3. Гантимуров А.П., Бром А.Е. Исследование и анализ информационного взаимодействия промышленных предприятий // Журнал «Глобальный научный потенциал». 2014. С. 124-126. (0,5 п.л. / 0,4 п.л.).
4. Гантимуров А.П., Бром А.Е. Принципы организации современного информационного взаимодействия промышленного предприятия // Журнал «Глобальный научный потенциал». 2015. С. 73-75. (0,5 п.л. / 0,4 п.л.).
5. Горлачева Е.Н., Гантимуров А.П. Создание информационного пространства промышленного предприятия // Техника машиностроения РИНЦ. 2015. №3. С. 37-41. (0,5 п.л. / 0,4 п.л.).

6. Гантимуров А.П. Разработка методов и инструментов для построения и анализа информационного взаимодействия промышленных предприятий // Вестник ВУиТ. 2016. №1. URL: <https://cyberleninka.ru/article/n/razrabotka-metodov-i-instrumentov-dlya-postroeniya-i-analiza-informatsionnogo-vzaimodeystviya-promyshlennyh-predpriyatiy> (дата обращения: 24.02.2022). (0,5 п.л. / 0,4 п.л.).
7. Гантимуров А.П., Бром А.Е. Надежность при построении информационного взаимодействия промышленных предприятий на основе распределенной системы хранения данных // Научно-аналитический журнал «Инновации и Инвестиции». 2018. №11. С. 261. (0,5 п.л. / 0,4 п.л.).
8. Бром А. Е., Гантимуров А. П. Методы оптимизации информационных потоков в бизнес-процессах с распределенной системой хранения данных // Проблемы теории и практики управления. 2019. №3-4. С. 49-55. (0,5 п.л. / 0,4 п.л.).
9. Выбор архитектурных решений построения системы хранения данных в технологии Wes-Scale IT / А.П. Гантимуров [и др.] // Перспективы науки. 2019. №8(119). С. 6-7. (0,5 п.л. / 0,4 п.л.).
10. Методы оптимизации информационных потоков в бизнес-процессах с распределенной системой хранения данных / А.П. Гантимуров [и др.] // Перспективы науки. 2019. №9(120). С. 35-39. (0,5 п.л. / 0,4 п.л.).
11. Метод повышения общей производительности информационной системы с распределенной системой хранения данных путем решения оптимизационной задачи подбора параметров технологии QoS на базе математической модели информационной системы / А.П. Гантимуров [и др.] // Перспективы науки. 2019. №12(123). С. 78-82. (0,5 п.л. / 0,4 п.л.).
12. Способ снижения времени обработки данных в подсистеме кодирования информации распределенной системы хранения данных / А.П. Гантимуров [и др.] // Промышленные АСУ и контроллеры. 2019. №12. С. 36-39. (0,5 п.л. / 0,4 п.л.).
13. Метод раннего диагностирования отключения жесткого диска в распределенной системе хранения данных / А.П. Гантимуров [и др.] // Радиотехника. 2019. Т. 83. №12(19). С. 5-11. (0,5 п.л. / 0,4 п.л.).