

На правах рукописи



Пойманова Екатерина Дмитриевна

**МОДЕЛИ УПРАВЛЕНИЯ РЕСУРСАМИ СИСТЕМ ХРАНЕНИЯ
ДАННЫХ**

Специальность 05.13.01 – «Системный анализ, управление и обработка
информации (технические системы)»

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Санкт-Петербург - 2020

Работа выполнена:

на кафедре Информационные системы федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)» (СПбГЭТУ «ЛЭТИ»).

Научный руководитель:

доктор технических наук, доцент **Татарникова Татьяна Михайловна**, федеральное государственное бюджетное образовательное учреждение высшего образования «Российский государственный гидрометеорологический университет» (РГГМУ), кафедра информационных технологий и систем безопасности, заведующий кафедрой, г. Санкт-Петербург.

Официальные оппоненты:

доктор технических наук, **Парамонов Александр Иванович**, федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича», кафедра сетей связи и передачи данных, профессор, г. Санкт-Петербург

кандидат технических наук **Богатырев Анатолий Владимирович**, АО «Санкт-Петербургский центр компетенций «НЕО», технический директор, г. Санкт-Петербург

Ведущая организация: АО «НИИ «Масштаб», г. Санкт-Петербург

Защита состоится «17» февраля 2020 года в 16.00 на заседании диссертационного совета Д 212.238.07, созданном при Санкт-Петербургском государственном электротехническом университете «ЛЭТИ» им. В.И. Ульянова (Ленина) по адресу: 197376, Санкт-Петербург, ул. Профессора Попова, 5.

С диссертацией можно ознакомиться в библиотеке ФГАОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)» и на сайте университета www.etu.ru в разделе «Подготовки кадров высшей квалификации» - «Объявление о защитах»

Автореферат разослан «16» декабря 2019 года.

Ученый секретарь
диссертационного совета Д 212.238.07
к.т.н., доцент

/В. В.Цехановский/

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы диссертации. Необходимость в хранении данных можно встретить на уровне отдельного предприятия, корпорации, государственных структур. Практически все организации нуждаются в формировании развитой инфраструктуры хранения данных. Существующие технологии типа дубликации, зеркалирования, виртуализации, иерархического хранения и т.д. способствуют обеспечению гарантированного хранения.

Исследование цифровой вселенной, проведенное компанией IDC (International Data Corporation) по заказу корпорации Dell EMC в 2014 г. с целью количественной оценки ежегодного объема создаваемых данных, показало, что мировой объем цифровой информации способен увеличиться с 4,4 (2014 г.) зеттабайт до 44 зеттабайт к 2020 году. По мнению экспертов, каждые два года объем данных удваивается. Поэтому, помимо задач обработки и передачи, актуальной становится задача хранения данных.

Проблема непрерывного роста объемов вычислительных ресурсов, известная как закон Гордона Мура не является новой. Объем оперативной памяти современного компьютера вырос в десять тысяч раз, емкость жесткого диска увеличилась более чем в сто тысяч раз в сравнении с первым персональным компьютером. Однако, рост объема памяти в настоящее время приближается к физическому пределу, при котором минимальная единица хранения ограничена размером атома. Поэтому рост объемов для хранения данных в последние годы решается технологиями параллелизма – проектированием систем хранения данных (СХД) сетевого типа.

Задача хранения не сводится только к выделению СХД необходимого объема для размещения данных. Актуальной становится задача рационального расходования физических ресурсов СХД, в частности емкостных. Множество сопутствующих факторов, включая федеральные нормы, неэффективность затрат на хранение данных и другие, требуют новых моделей управления ресурсами СХД. При этом необходимо учитывать следующие обстоятельства:

- разнородность СХД;
- наличие разных требований к хранению данных;
- отсутствие единых алгоритмов организации хранения данных с разными требованиями к хранению;
- осуществление параллельного выполнения основной функции СХД и функции, обеспечивающей восстановление данных без остановки выполняемых бизнес-процессов;
- увеличение стоимости хранения единицы данных в связи с модернизацией и утилизацией СХД.

Наличие такого количества глобальных проблем при проектировании СХД с заданными свойствами объясняется, в том числе, отсутствием единой модели организации физического хранения данных.

Учитывая обозначенные проблемы при организации хранения данных, основной научной задачей является разработка новых моделей и алгоритмов

управления ресурсами систем хранения данных, что определяет актуальность темы диссертационной работы.

Степень разработанности проблемы. По вопросам стандартизации и моделирования систем хранения данных опубликовано мало работ. Это объясняется спецификой области исследования. Корпорации, разрабатывающие технологии хранения данных и внедряющие их в ИТ-сферу, такие как Dell EMC, Fujitsu, IBM и другие, предоставляют свои решения в виде IaaS-услуги (Infrastructure as a Service – инфраструктура как услуга). Эти решения не являются открытыми с точки зрения реализации. Тем не менее доступны регулярные обзоры о трендах в области хранения данных.

Помимо аналитических обзоров от этих корпораций отметим труды отечественных и зарубежных авторов Р. Ландауэра, Л.Б. Киша, М. Фарли, Д. Паттерсона, Б.Я. Советова, В.В. Цехановского, В.Д. Чертовского, А.И. Водяхо, В.А. Дубенецкого, и других, составляющих теоретическую базу проектирования хранилищ данных и систем хранения данных, а также работы В.А. Богатырева, О.И. Кутузова, М.О. Колбанева, Т.М. Татарниковой, С.А. Яковлева в области моделирования сложных технических систем и их элементов, к числу которых можно отнести и системы хранения данных.

Цель диссертационной работы состоит в разработке моделей и алгоритмов управления ресурсами систем хранения данных с учетом многоуровневой организации хранения данных.

В соответствии с указанной целью в работе поставлены, обоснованы и решены следующие **задачи**:

1. Предложена многоуровневая модель хранения данных.
2. Предложена модель управления емкостью физического хранилища данных, основанная на алгоритмах размещения и миграции данных по уровням иерархии системы хранения данных.
3. Построена модель прогноза дифференцированного наращивания емкости системы хранения данных и предложена соответствующая методика прогнозирования.

Объектом исследования являются системы хранения данных центров обработки данных, построенные на разных технологиях хранения и применяемые в корпоративном и государственном секторе.

Предметом исследования является применение моделей управления ресурсами систем хранения данных.

Методы исследования. Решение сформулированной в диссертационной работе задачи разработки моделей и алгоритмов управления ресурсами систем хранения данных базируется на методах системного анализа, теории вероятности, случайных процессов и математической статистики, методов анализа данных, методов построения прогнозов временных рядов.

К основным **научным результатам диссертации**, которые определяют новизну исследования и выносятся на защиту, относятся:

1. Многоуровневая модель хранения данных отличается от других моделей открытых систем стратифицированным описанием, сочетающим уровневую организацию хранения данных, технологии хранения данных и характеристики

физического хранилища и сохраняемых файлов, что позволяет проектировщикам систем хранения данных получать решения с заданными свойствами на всем многообразии вариантов хранения данных.

2. Модель управления емкостью физического хранилища отличается применением алгоритмов размещения и миграции данных по уровням иерархии системы хранения данных, которые основаны на анализе метаданных, что позволяет реализовать проактивное управление ресурсами систем хранения данных с учетом сроков гарантированного хранения.

3. Модель прогноза дифференцированного наращивания емкости системы хранения данных и соответствующая методика прогнозирования отличаются учетом паттернов состояния, что позволяет наращивать емкости СХД исходя из реального динамического изменения состояния системы хранения данных.

Теоретическая значимость исследования состоит в дальнейшем развитии методов, моделей и технологий хранения данных в СХД.

Практическая ценность результатов диссертационной работы заключается в полученных расчетных выражениях, моделях и алгоритмах управления ресурсами систем хранения данных.

Основные научные результаты диссертации используются в учебном процессе на кафедре Информационные системы СПбГЭТУ «ЛЭТИ» по направлению обучения бакалавров «Информационные системы и технологии» по курсу «Управление данными», на кафедре информационных систем и технологий СПбГЭУ при изучении дисциплин направления 09.03.02 «Информационные системы и технологии» и направления 38.03.05 «Бизнес-информатика», а также на кафедре безопасности информационных систем СПбГУАП при изучении дисциплин направления 10.03.01 «Информационная безопасность», 11.04.02 «Инфокоммуникационные технологии и системы связи».

Достоверность полученных результатов подтверждена результатами моделирования и экспериментальных исследований. Для моделирования использовалось специализированное программное обеспечение, подкрепленное апробированными теоретическими методами исследований моделей компьютерных систем. Полученные в ходе выполнения диссертационного исследования результаты не противоречат ранее полученным данным, описанным в литературе другими авторами.

Апробация работы. Основные научные результаты диссертационной работы докладывались, обсуждались и были одобрены на X межвузовской научно-практической конференции преподавателей, аспирантов и докторантов «Проблемы и пути развития предпринимательской деятельности в современных условиях» СПбГЭУ 13-14 марта 2014 г.; Санкт-Петербургской международной конференции «Региональная информатика» в 2013-2018 гг., VIII Санкт-Петербургской межрегиональной конференции «Информационная безопасность регионов России (ИБРР-2013)», 28 октября 2013 г., II межрегиональной научно-практической конференции «Перспективные направления развития развития отечественных информационных технологий», Севастополь, 13-17 сентября 2016 г., на научной конференции аспирантов СПбГЭУ «Россия в современном

мире: экономические, правовые и социальные аспекты развития», 25 апреля 2017 г.; X Санкт-Петербургской межрегиональной конференции «Информационная безопасность регионов России (ИБРР-2017)», 1-3 ноября 2017 г.; X научно-практическая конференция молодых ученых «Программная инженерия и компьютерная техника» (Майоровские чтения), V международной научно-практической конференции ИНФОГЕО 2018 «Геоинформационное обеспечение устойчивого развития территорий», кафедральных семинарах СПбГЭУ 2014-2018 гг., XXII международная научная конференция «Волновая электроника и инфокоммуникационные системы» 3-7 июня 2019 Санкт-Петербург.

Получено свидетельство о государственной регистрации программы для ЭВМ «Программа прогноза наращивания емкости систем хранения данных» №2019661945.

Публикации. Основное содержание диссертации опубликовано в 19 печатных работах, в том числе в 4 работах, рекомендованных ВАК РФ и в 4 работах, входящих в базы цитирования Web of Science и Scopus.

Личный вклад автора. Автором выполнены основные расчеты, разработана многоуровневая модель хранения данных, построена модель управления емкостью физического хранилища, разработаны алгоритмы размещения и миграции данных по уровням иерархии системы хранения данных, предложено использование аппарата нейронной сети Кохонена в качестве инструмента решения задачи размещения объектов в соответствии с требуемыми параметрами, проведено соответствующее экспериментальное исследование, построена модель прогноза дифференцированного наращивания емкости системы хранения данных и разработана методика прогнозирования наращивания емкости системы хранения данных.

Результаты, выносимые на защиту.

1. Многоуровневая модель хранения данных.
2. Модель управления емкостью физического хранилища данных, основанная на алгоритмах размещения и миграции данных по уровням иерархии системы хранения данных.
3. Модель прогноза дифференцированного наращивания емкости системы хранения данных и соответствующая методика прогнозирования.

Объем и структура работы. Диссертация состоит из введения, четырех глав, заключения и списка литературы (всего 77 источников). Общий объем работы составляет 140 страниц машинописного текста, который включает 55 рисунков, 13 таблиц.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении сформулированы актуальность, цель и задачи исследования, определена научная новизна и практическая значимость работы.

В первой главе «Анализ особенностей технологии хранения данных» проведен анализ законодательных требований ко времени гарантированного хранения данных, технологий и ресурсов, поддерживающих жизненный цикл данных, определена задача исследования.

Проведен анализ требований законов и норм Российской Федерации, который показал, что сроки хранения информации во всех важнейших отраслях составляют от 1 года до постоянного (бессрочного) хранения. Кроме того, высокие требования предъявляются к системам хранения данных в связи с принятием Федерального закона №374-ФЗ от 06.07.2016, устанавливающего дополнительные меры противодействию терроризму и обеспечения общественной безопасности.

Определены этапы жизненного цикла процесса хранения данных:

- 1) появление у источника потребности в хранении;
- 2) формирование сообщения (преобразование сведений в сообщение);
- 3) запись сообщения на носитель (сжатие, кодирование информации, запись ее на физический носитель);
- 4) хранение носителя заданное время (создание и поддержание условий, обеспечивающих целостность данных на носителе в течение установленного срока);
- 5) считывание сообщения с носителя;
- 6) интерпретация данных (преобразование сообщения в сведения).

Каждому этапу жизненного цикла поставлена в соответствие информационная технология, реализующая процесс хранения данных. Выполнена классификация носителей данных, включая перспективные технологии записи, обеспечивающие хранение до 1 000 и более лет. Для рассмотренных видов носителей выделены интервалы сроков хранения информации и поставлены в соответствие следующие типы данных:

1. Исходные данные – время хранения до одного года.
2. Резервные копии – время хранения от 1 года до 10 лет.
3. Архивные данные – требуемое время хранения от 10 до 100 лет.
4. Данные, для передачи следующим поколениям – возможное время хранения от 100 до 1000 и более лет.

В результате анализа требований к срокам хранения информации предложено соответствие между типами данных, существующими технологиями записи и видами носителей.

Предложена иерархическая структура СХД, состоящая из трех уровней иерархии:

- 1) информационно-поисковая система;
- 2) база данных;
- 3) хранилище данных.

Рассмотрен процесс инкапсуляции данных в ходе формирования файла по цепочке: минимальная единица хранения (МЕХ) → физический блок данных (ФБ) → логический блок данных (ЛБ) → файл. Показано, что разные файловые системы поддерживают разный размер логического блока данных, максимальный размер файла, максимальный объем носителя и т.д.

Показано, что особую роль при хранении играют метаданные, которые несут в себе информацию о месте, времени хранения файлов и другие сведения, используемые при организации рационального хранения разнородных данных.

Предложена классификация метаданных файлов:

- 1) по функциональным характеристикам:

- организационные:
 - первичные – заявляются пользователем;
 - адресные – включают в себя адрес файла на носителе и адрес носителя в хранилище;
 - специализированные – включают в себя характеристики файлов с данными и характеристики носителя информации;
- 2) по необходимости формирования:
 - обязательные – необходимые для идентификации и считывания информации;
 - вариативные – заявленные пользователем;
- 3) по месту хранения:
 - хранимые на носителе вместе с данными;
 - хранимые в базе метаданных.

Выделены типы ресурсов, обеспечивающих хранение данных: интеллектуальные, информационные и физические. Определен основной тип физических ресурсов – емкость, эффективное использование которого позволяет экономить средства, затрачиваемые на хранение данных.

Выделены укрупненные параметры сохраняемых данных, от которых зависит выбор для них технологии хранения: время директивного хранения, размер логического блока данных, размер файла, частота обращения к нему. Сделан вывод, что данные параметры влияют на выбор файловой системы, работу СХД и расходование ресурсов.

Рассмотрена система управления хранилищем данных, цель которой – максимальное использование емкости хранилища, а также его своевременное наращивание.

Определена задача исследования, состоящая в разработке моделей управления ресурсами СХД.

Во второй главе «Многоуровневая модель хранения данных» предложена стратифицированная модель, учитывающая структуру файла при его записи, а также объединяющая иерархию хранения, виды носителей и характеристики данных.

Анализ моделей информационных процессов, таких как МИС и ЭМВОС показал, что в них отсутствует подробное описание записи и физического хранения битов данных, что является необходимым для описания процесса хранения.

Предложена многоуровневая модель хранения данных, имеющая стратифицированную структуру, которая объединяет уровни организации хранения, виды носителей и характеристики данных (рисунок 1).

Физический уровень предложенной модели хранения данных определяет порядок записи и адресацию битов данных файла. На данном уровне модели рассматриваются следующие характеристики входящего потока файлов на запись F : размер логического блока данных S и размер файла f .



Рисунок 1 – Многоуровневая модель хранения данных

Физический уровень состоит из четырех подуровней, соответствующих инкапсуляции данных: Минимальная единица хранения → Физический блок данных → Логический блок данных → Файл. Соответственно, на *подуровне МЕХ* реализуются функции по поддержанию устойчивого состояния минимальной единицы хранения, на *подуровне физического блока* происходит инкапсуляция МЕХ в физические блоки данных, далее на *подуровне логического блока данных* происходит инкапсуляция физических блоков данных в логические. На подуровне файла определяется адресация битов данных, физических и логических блоков и происходит логическое объединение битов данных в файл.

На структурном уровне определяются характеристики матричной структуры физического хранилища данных: уровни хранения (RAID, автоматизированные библиотеки, носители длительного хранения) и тома каждого уровня хранилища. Перед записью на определенный уровень хранилища предлагается проводить анализ данных с целью выбора файловой системы с определенным размером логического блока данных (для уровня RAID) или определенным типом архивного носителя. Таким образом каждая ячейка матрицы – это том соответствующего уровня хранилища, предназначенного для хранения данных с определенными метрическими характеристиками.

Уровень управления описывает процессы управления емкостью физического хранилища данных. Уровень содержит четыре подуровня.

На *подуровне распределения данных* проводится анализ характеристик входящего потока данных и размещение данных в матрице хранения. Предложены следующие механизмы:

– механизм вертикального размещения – выбор уровня хранилища m в зависимости от времени хранения t ;

– механизм горизонтального размещения – выбор логического тома или носителя n в зависимости от размера файла f .

На *подуровень миграции данных* предложено в зависимости от частоты обращения к данным λ осуществлять миграцию данных по уровням физического хранилища с помощью механизма динамического размещения.

Подуровень прогноза наращивания емкости предполагает составление прогноза наращивания емкости физического хранилища.

В третьей главе «Модели и алгоритмы управления ресурсами систем хранения данных» разработаны алгоритмы распределения файлов по ячейкам хранилища данных, предложена модель управления емкостью СХД, проведен эксперимент, показывающий возможность использования механизма нейронных сетей Кохонена для распределения файлов по ячейкам матрицы хранения.

С учетом архитектуры СХД предложена матричная структура хранилища данных размером $m \times n$, где m – это количество уровней хранилища, n – количество физических или логических носителей. Предложены следующие механизмы распределения файлов данных по уровням СХД (рисунок 2):

1) выбор уровня СХД в зависимости от времени хранения (вертикальное размещение);

2) выбор логического тома уровня хранения СХД в зависимости от размера файла и длины логического блока данных (горизонтальное размещение);

3) миграция данных по уровням СХД в зависимости от частоты обращения к ним (динамическое размещение).

Идея вертикального размещения файлов основана на анализе организационных метаданных, содержащих сведения о типе данных: *ind* (*initial data*) – исходные данные, которые размещаются на уровень RAID; *bck* (*backups*) – резервные копии, архивные данные, которые будут храниться на уровне автоматизированной библиотеки; *ngd* (*next generation data*) – данные бессрочного хранения.

На следующем этапе происходит горизонтальное распределение файлов. Идея горизонтального размещения основана на выборе файловой системы для уровня RAID и по типам носителей на нижних уровнях СХД.

Правило горизонтального размещения:

$$\text{Если } f \in (f_i; f_{i+1}], \text{ то } \rightarrow a_{i+1} \Leftrightarrow F \rightarrow \text{Том}_{i+1}$$

где f – размер сохраняемого файла F ;

f_i, f_{i+1} – левая и правая границы соответственно размера файла F , с которым может работать файловая система;

a_{i+1} – размер логического блока данных, которым оперирует файловая система;

Том_{i+1} – номер тома RAID, управляемый соответствующей файловой системой.

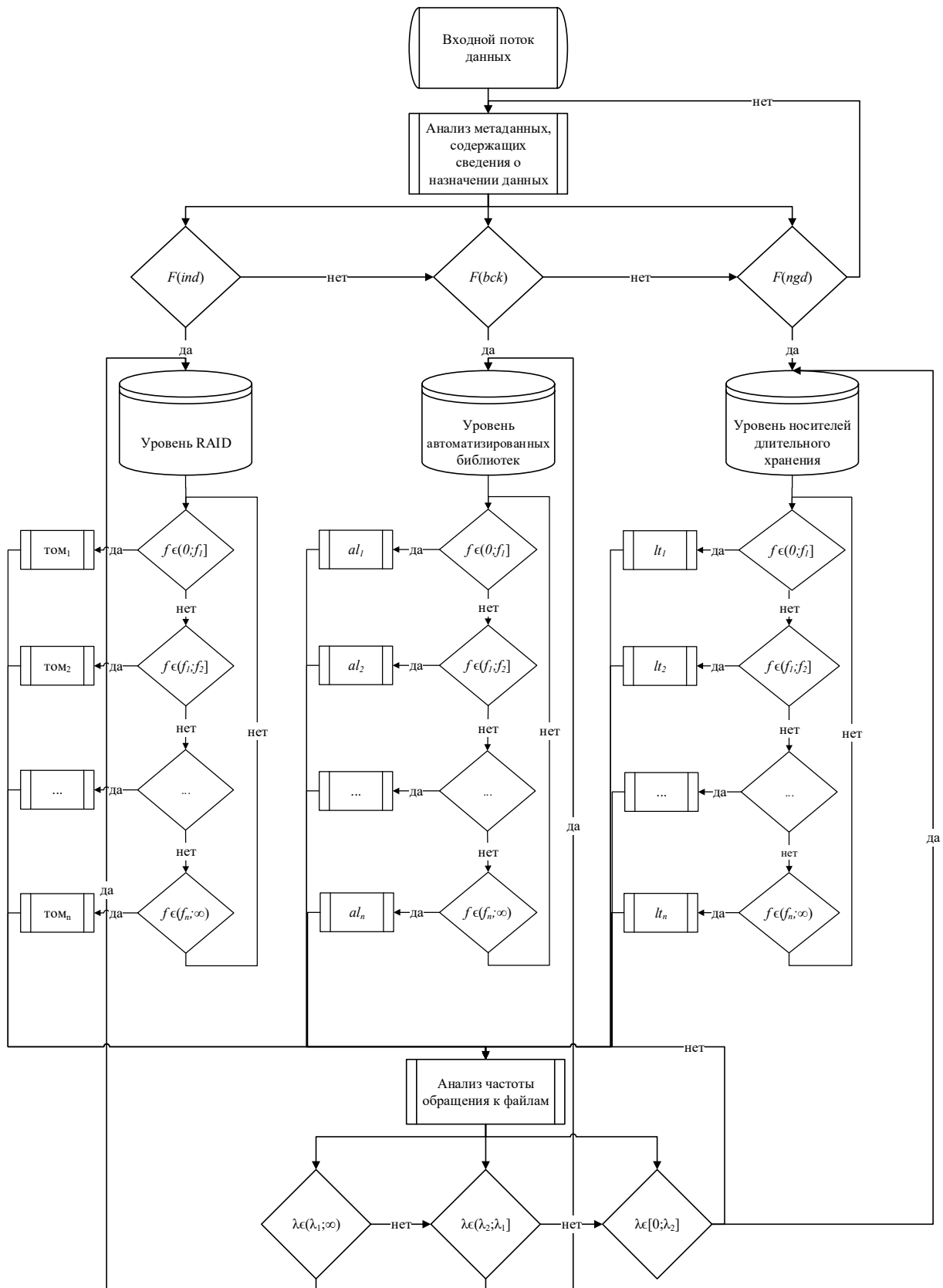


Рисунок 2 – Алгоритм управления емкостью СХД на основе анализа метаданных

На нижних уровнях СХД предлагается разделение емкости по типам носителей: стример, DVD, BD для уровня автоматизированных библиотек и М-диск, стеклянный диск и ДНК – для носителей длительного хранения:

$$\text{Если } f \in (f_i, f_{i+1}], \text{ то } F \rightarrow al_{i+1} (lt_{i+1}),$$

где al_{i+1} или lt_{i+1} – тип носителя на уровне автоматизированной библиотеки (al) или на уровне длительного хранения (lt).

Идея динамического размещения основан на миграции данных по уровням СХД в зависимости от частоты обращения к файлам данных:

$$\text{Если } \lambda_F \in (\lambda_i, \lambda_{i+1}], \text{ то } F \rightarrow l,$$

где λ_F – частота запроса файла F ;

λ_i, λ_{i+1} – левая и правая границы соответственно частоты запроса файла;

l – номер уровня СХД, на который мигрирует файл F .

Механизм миграции позволяет преодолеть недостатки субъективного выбора типа сохраняемых файлов при реализации первого этапа.

Предложено проводить распределение файлов по ячейкам матрицы хранения с использованием аппарата нейронных сетей Кохонена, что позволяет отказаться от последовательного выполнения механизмов размещения. Обученная нейронная сеть Кохонена способна решать задачу размещения файлов за один шаг.

В четвертой главе «Методика прогнозирования наращивания емкости систем хранения данных» предложена модель прогноза дифференцированного наращивания емкости физического хранилища данных, рассмотрены особенности входящего потока данных в СХД, которые положены в основу методики прогнозирования наращивания емкости СХД.

Предложенная модель прогноза основана на анализе срезов состояния физического хранилища данных и представляет собой паттерн поведения СХД, исходя из прогноза поведения каждой ячейки матрицы хранения. Ячейки имеют следующие характеристики:

– максимальное значение емкости V_{\max} , при этом емкость ячейки соответствует емкости используемого носителя (тома носителя);

– предельное значение емкости V_{\lim} ($V_{\lim} < V_{\max}$), при достижении которого необходимо производить наращивание.

Также задаются пограничные значения частоты обращения к файлам $\lambda_1, \dots, \lambda_{m-1}$, при преодолении которых осуществляется миграция файлов по уровням системы.

Совокупность указанных характеристик является нулевым *паттерном* состояния системы хранения данных.

Для актуализации физического хранилища данных СХД предложено провести анализ метаданных о размере файлов, хранимых в каждой ячейки матрицы хранения для оценивания текущей емкости каждой ячейки V_{current} . Полученная модель матрицы хранения является актуальным паттерном состояния СХД.

Поскольку данные в СХД накапливаются, необходимо производить подсчет накопившихся данных с течением времени и определять точки преодоления предельных значений емкости ячеек матрицы хранения (V_{\lim}).

Обозначим характеристики, влияющие на поведение системы:

$V_{\max mn}$ – максимальная емкость ячейки mn ;

$V_{\lim mn}$ – предельная емкость ячейки mn ;

$V_{\text{current} mn}$ – текущая емкость ячейки mn ;

введем дополнительно:

$t_{\lim mn}$ – время достижения предельной емкости ячейки mn ;

$t_{\max mn}$ – время достижения максимальной емкости ячейки mn .

Задача прогнозирования заключается в нахождении времени достижения предельной емкости $t_{\lim mn}$ и времени достижения максимальной емкости $t_{\max mn}$ каждой ячейки.

Значение V_{\lim} и V_{\max} , заданные в паттерне состояний матрицы хранения, математически можно выразить следующим образом:

$$V_{\lim} = \int_1^{t_{\lim}} f(t)dt, \quad (1)$$

$$V_{\max} = \int_1^{t_{\max}} f(t)dt, \quad (2)$$

где $f(t)$ – функция входящего потока данных.

Функция $f(t)$ математически может быть не определена, поэтому предложено выразить значения V_{\lim} и V_{\max} , исходя из метода правых прямоугольников. Тогда:

$$V_{\lim} = \int_1^{t_{\lim}} f(t)dt = T \sum_1^{t_{\lim}} f(t), \quad (3)$$

$$V_{\max} = \int_1^{t_{\max}} f(t)dt = T \sum_1^{t_{\max}} f(t), \quad (4)$$

где T – шаг разбиения, равный единице минимального выбранного масштаба времени.

Математически значения t_{\lim} и t_{\max} вычислить сложно, т.к. функция $f(t)$ не имеет первообразной. Поэтому предложено вычислять данные параметры программно методом подстановки. Т.е. необходимо найти такие t_{\lim} и t_{\max} , при которых соответственно V_{\lim} и V_{\max} будут равны установленному значению.

Необходимо отметить, что данные расчеты производятся для каждой ячейки матрицы хранения. Таким образом формируется паттерн поведения системы.

Показано, что особенностью входящего потока данных в СХД является свойство самоподобия. Определены параметры, указывающие на самоподобность входящего потока данных: долговременная зависимость, медленно затухающая дисперсия и наличие распределения с тяжелыми «хвостами» времени между соседними поступлениями.

Рассмотрены существующие модели прогнозирования временных рядов: общая линейная, авторегрессии, скользящего среднего и их комбинации.

Для анализа входящего потока на СХД выбраны:

– общая линейная модель для случая входящего потока данных без самоподобной структуры;

– модель авторегрессии и проинтегрированного скользящего среднего для самоподобного входящего потока данных.

Методика прогнозирования наращивания емкости СХД состоит из следующих этапов:

1. Анализ текущего состояния системы, в ходе которого строится нулевой и актуальный паттерны состояния СХД. Нулевой паттерн – это шаблон структуры хранилища данных, устанавливающий ограничения на значения параметров ее

ячеек. Актуальный паттерн определяет текущее состояние хранилища. В построении актуального паттерна участвуют метаданные, указывающие на размер хранимых в СХД файлов и время директивного хранения.

2. Выявление свойств входящего потока данных.

2.1. Визуализация в разных масштабах времени структуры входящего потока данных на основе графиков временного ряда.

2.2. Визуальный анализ графиков временного ряда – выявление наличия или отсутствия признаков самоподобия. При отсутствии признаков самоподобия переход на п. 3.1, иначе переход на п. 2.3.

2.3. Определение статистических показателей самоподобия входящего потока данных (параметра формы распределения, показателя Херста, дисперсии). Переход на п. 3.2.

3. Построение паттерна поведения P системы (прогнозной модели), на основе которого строится одна из моделей прогноза:

3.1. Общая линейная модель.

3.2. Модель авторегрессии и проинтегрированного скользящего среднего для каждой ячейки матрицы хранения.

4. Расчет времени преодоления пограничных значений емкости ячеек матрицы хранения.

5. Мониторинг состояния СХД с помощью систематических срезов состояний системы.

6. Проверка соответствия состояния СХД паттерну поведения.

Если разница между срезом состояния СХД и паттерном поведения меньше допустимой ошибки прогноза, то считать прогноз верным и перейти на п.1. Иначе переход на п. 3 с другими параметрами прогнозирования.

Методика прогнозирования наращивания емкости СХД автоматизирована в виде приложения, которое рекомендуется запускать при работе системы хранения данных. Пример окна приложения приведен на рисунке 3.



Рисунок 3 – Окно приложения прогнозирования наращивания емкости СХД

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

В ходе диссертационного исследования была решена важная научная задача, заключающаяся в разработке моделей и алгоритмов управления ресурсами систем хранения данных.

Применение разработанных моделей позволяет рационально использовать физические ресурсы СХД, что является актуальным в условиях современного объема создаваемых и хранимых данных.

Основные новые результаты работы заключаются в следующем:

В ходе диссертационного исследования была решена важная научная задача, заключающаяся в разработке моделей и алгоритмов управления ресурсами систем хранения данных. Особое внимание в работе уделено емкости физического хранилища данных, то есть объему памяти, который занимают файлы данных в СХД.

Проведенный в работе анализ позволил:

- установить требования, предъявляемые к срокам хранения данных, выделить интервалы сроков хранения и провести их соответствие с носителями данных в зависимости от физического типа памяти и временного ресурса;

- определить особую роль метаданных при цифровом хранении и провести их классификацию;

- выявить типы ресурсов, обеспечивающих хранение данных и определить основной тип ресурсов – емкость физического хранилища данных, эффективное использование которой позволяет экономить средства, затрачиваемые на хранение данных;

- выделить укрупненные параметры сохраняемых данных, от которых зависит выбор для них технологии хранения и сделать вывод, что данные параметры влияют на выбор файловой системы, работу СХД и расходование ресурсов;

- рассмотреть СХД как систему управления хранилищем данных, цель которой – максимальное использование емкости хранилища, а также его своевременное наращивание.

Основные результаты работы заключаются в следующем:

1. Предложена многоуровневая модель хранения данных, объединяющая уровни организации процесса хранения, технологии, реализующие этот процесс и соответствующие характеристики данных:

- описаны функции уровней и подуровней предлагаемого варианта многоуровневой модели хранения данных;

- описаны работа многоуровневой модели;

- рассмотрены проблемы реализации процесса хранения данных на каждом уровне предлагаемой модели, заключающиеся в экономии физических ресурсов и своевременном наращивании емкости СХД.

2. Предложена модель управления емкостью физического хранилища СХД, основанная на применении алгоритмов вертикального, горизонтального и динамического размещения файлов:

– предложены алгоритмы вертикального и горизонтального размещения файлов в физическом хранилище СХД.

– предложен алгоритм динамического размещения файлов в СХД, основанный на миграции данных в зависимости от частоты обращения к файлам.

– предложена матричная структура управления хранилищем данных;

– экспериментально показано, что аппарат нейронной сети Кохонена может быть инструментом решения задачи размещения объектов в соответствии с требуемыми параметрами.

3. Предложена методика прогнозирования наращивания емкости СХД на основе анализа потока входящих данных и их миграции. Предложен макет интерфейса приложения для анализа входящего потока данных и построения модели прогноза наращивания на основе предложенной методики.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

В изданиях научно-технической базы Scopus

1. Poymanova E.D., Tatarnikova T.M. Models and Methods for Studying Network Traffic//2018 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF) 26-30 Nov. 2018, DOI: 10.1109/WECONF.2018.8604470.

2. Советов Б.Я., Татарникова Т.М., Пойманова Е.Д. Организация многоуровневого хранения данных//Information & Control Systems / Informazionno-Upravlyaushie Sistemy. 2019, Vol. 99 Issue 2, p68-75.

3. Т. М. Tatarnikova; E. D. Poymanova Energy Model of Data Storage Process//2019 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF) 3-7 June 2019, DOI: 10.1109/WECONF.2019.8840111

4. E. D. Poymanova ; T. M. Tatarnikova Tiered Data Storage Model//2019 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF) 3-7 June 2019, DOI: 10.1109/WECONF.2019.8840589

В изданиях, рекомендованных ВАК:

5. Татарникова Т.М., Пойманова Е.Д. Энергетическая модель процесса хранения данных//Информация и Космос 2019, № 1. С. 89-95

6. Татарникова Т.М., Пойманова Е.Д. Управление энергозатратами процесса хранения данных при выборе размера физического блока данных//Программные продукты и системы 2019, №1(32). С. 68-72.

7. Верзун Н.А., Колбанёв М.О., Пойманова Е.Д. Энергетические характеристики процесса долговременного хранения данных//Известия вузов. Приборостроение 2017 Т.60, №2. С.158-164.

8. Колбанёв М.О., Пойманова Е.Д., Татарникова Т.М. Физические ресурсы информационного процесса сохранения данных//Известия вузов. Приборостроение 2014 Т.57, №9. С.38-42.

В других изданиях:

9. Татарникова Т.М., Пойманова Е.Д. Модель многоуровневой организации системы хранения данных //Научно-технический вестник информационных технологий, механики и оптики. 2019. Т. 19. № 2. С. 271–279.

10. Татарникова Т.М., Пойманова Е.Д. Управление энергозатратами процесса хранения данных//Волновая электроника и инфокоммуникационные системы: Двадцать вторая международная научная конференция ГУАП: сб. статей: в 2 ч. Ч. 1. СПб.: ГУАП, 2019. С.273-276

11. Пойманова Е.Д., Татарникова Т.М. Модель многоуровневого хранения данных//Волновая электроника и инфокоммуникационные системы: Двадцать вторая международная научная конференция ГУАП: сб. статей: в 2 ч. Ч. 1. СПб.: ГУАП, 2019. С.263-267

12. Пойманова Е.Д. Обзор законодательных норм в области хранения информации. «Закон Яровой»//Информационные технологии цифровой экономики: сборник статей. – СПб.: Изд-во СПбГЭУ, 2017. – 128 с. С.91-94

13. Пойманова Е.Д., Колбанёв М.О., Татарникова Т.М. Метаданные в системах долговременного хранения//Информационная безопасность регионов России (ИБРР-2017). Юбилейная X Санкт-Петербургская межрегиональная конференция. Санкт-Петербург, 1-3 ноября 2017 г.: Материалы конференции / СПОИСУ. – СПб., 2017. – 580 с.

14. Пойманова Е.Д., Колбанёв М.О., Ипатов О.С. Подуровень безопасности архитектурной модели сохранения данных//Информационная безопасность регионов России (ИБРР-2017). Юбилейная X Санкт-Петербургская межрегиональная конференция. Санкт-Петербург, 1-3 ноября 2017 г.: Материалы конференции / СПОИСУ. – СПб., 2017. – 580 с. С. 223-224.

15. Пойманова Е.Д., Колбанёв М.О., Касаткин В.В. Энергетические аспекты безопасности долговременного хранения данных//Информационная безопасность регионов России (ИБРР-2017). Юбилейная X Санкт-Петербургская межрегиональная конференция. Санкт-Петербург, 1-3 ноября 2017 г.: Материалы конференции / СПОИСУ. – СПб., 2017. – 580 с.

16. Колбанёв М.О., Касаткин В.В., Пойманова Е.Д. Энергетические аспекты обеспечения целостности данных в системах хранения информации//Технологии информационно-экономической безопасности. Сборник статей. СПбГЭУ. 2016. С. 70-76.

17. Колбанёв М.О. Пойманова Е.Д., Татарникова Т.М. Физические ресурсы информационных систем хранения данных//Материалы IX Санкт-Петербургской межрегиональной конференции "Информационная безопасность регионов России (ИБРР-2015)" С. 216-217.

18. Татарникова Т.М., Пойманова Е.Д. Роль метаданных при длительном хранении информации//Сборник научных трудов по материалам Международной научно-практической конференции «Современное общество, образование и наука». 2015. С. 141-142.

19. Татарникова Т.М., Пойманова Е.Д. Технологии долговременного хранения данных//Сборник научных трудов по материалам Международной научно-практической конференции «Наука и образование в XXI веке». 2013. С. 136-138.