

КАЗАХСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ ИМЕНИ АЛЬ-ФАРАБИ

А.Т. Агишев

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ ОБРАБОТКИ ДАННЫХ

**Сборник лекций для студентов магистратуры,
обучающихся по образовательной программе
«7М07125 - Электроника и системы управления»**

Алматы, 2025

Лекция 10. Технологии хранения данных: распределённые системы и безопасность

Цель лекции

Изучить принципы построения современных систем хранения данных, их архитектуру, виды и особенности распределённых хранилищ. Рассмотреть вопросы отказоустойчивости, защиты информации и роли технологий хранения в интеллектуальных вычислительных системах.

Основные вопросы:

1. Архитектура систем хранения данных.
2. Типы хранилищ: локальные, сетевые и распределённые.
3. Концепции масштабируемого и отказоустойчивого хранения.
4. Технологии организации данных: RAID, NAS, SAN, SDS.
5. Распределённые файловые системы.
6. Механизмы резервирования и защиты данных.
7. Безопасность и целостность информации в интеллектуальных системах.

Краткие тезисы:

1. Архитектура систем хранения данных. Современные системы хранения представляют собой многоуровневую инфраструктуру, обеспечивающую доступ, обработку и защиту данных. Основные уровни:

- **Физический уровень** - устройства хранения (жёсткие диски, SSD, NVMe).
- **Логический уровень** - файловые и блочные структуры.
- **Уровень управления** - механизмы распределения, резервирования и контроля доступа. Хранилища интегрируются с вычислительными узлами через высокоскоростные интерфейсы (SAS, NVMe, InfiniBand).

2. Типы систем хранения.

- **DAS (Direct Attached Storage):** локально подключённое хранилище к серверу. Простое и быстрое решение, но без масштабируемости.
- **NAS (Network Attached Storage):** доступ по сети через файловые протоколы (NFS, SMB). Подходит для совместной работы пользователей и хранения документов.
- **SAN (Storage Area Network):** выделенная сеть для блочного доступа к данным. Используется в высокопроизводительных системах.
- **SDS (Software-Defined Storage):** программно-управляемая система хранения, не зависящая от конкретного оборудования.

Такая классификация позволяет выбирать архитектуру под конкретные задачи - от персональных серверов до суперкомпьютеров.

3. Масштабируемость и отказоустойчивость. Для обеспечения надёжности применяются механизмы:

- **RAID (Redundant Array of Independent Disks):** дублирование и распределение данных по дискам;
- **Мульти-репликация:** хранение копий данных на нескольких узлах;
- **Erasure Coding:** распределённое кодирование для восстановления при сбоях. Эти подходы минимизируют потери данных и обеспечивают непрерывность работы систем.

4. Распределённые системы хранения. Распределённое хранилище объединяет множество серверов, действующих как единая система. Данные автоматически делятся на блоки и распределяются по узлам с возможностью параллельного доступа. Основные примеры:

- **Ceph** - отказоустойчивая система для блочного и объектного хранения;
- **GlusterFS** - масштабируемое файловое хранилище;
- **HDFS (Hadoop Distributed File System)** - основа для анализа больших данных;
- **Lustre** - высокопроизводительная система для научных и инженерных вычислений.

Такие системы активно используются в центрах обработки данных, НРС-кластерах и облачных инфраструктурах.

5. Безопасность хранения данных. Безопасность данных обеспечивается на трёх уровнях:

- **Физическая защита:** контроль доступа к серверам и устройствам хранения.
- **Логическая защита:** разграничение прав пользователей, аутентификация и шифрование.
- **Сетевая защита:** использование VPN, межсетевых экранов и защищённых протоколов передачи (HTTPS, SFTP).

Для защиты конфиденциальной информации применяются методы шифрования на уровне томов и файловых систем (AES, BitLocker, LUKS).

6. Резервное копирование и восстановление. Резервирование - ключевой элемент безопасности данных. Основные схемы:

- **Полное копирование (Full Backup):** полная копия всех данных.
- **Инкрементное копирование:** сохраняются только изменения после последнего резервирования.
- **Дифференциальное копирование:** копируются данные, изменённые с момента последнего полного бэкапа. Для крупных систем используются автоматизированные решения и политики хранения с временными интервалами и географическим распределением копий.

7. Хранение данных в интеллектуальных системах. В системах искусственного интеллекта и анализа данных хранение играет ключевую роль. Необходимо обеспечивать:

- высокую пропускную способность при обучении моделей;
- надёжность и консистентность данных;
- безопасность при совместной работе пользователей и сервисов.

Интеллектуальные системы часто используют гибридные хранилища: сочетание локальных SSD для быстрых вычислений и облачных платформ (S3, Azure Blob, Google Storage) для долговременного хранения.

Вопросы для контроля, изучаемого материал:

- 1) Какова структура современной системы хранения данных?
- 2) В чём различие между DAS, NAS и SAN?
- 3) Как обеспечивается отказоустойчивость и масштабируемость хранилищ?
- 4) Назовите основные примеры распределённых систем хранения.
- 5) Какие механизмы применяются для защиты и резервирования данных?
- 6) Почему безопасность и целостность данных критичны для интеллектуальных систем?

Рекомендуемый список литературных источников:

1. Hopgood A. A. Intelligent Systems for Engineers and Scientists. — 3rd ed. — Boca Raton : CRC Press / Taylor & Francis, 2012. — 682 p.
2. Sterling T., Anderson M., Brodowicz M. High Performance Computing: Modern Systems and Practices. — Amsterdam : Elsevier / Morgan Kaufmann, 2017. — 728 p.
3. Russell, S., Norvig, P. Artificial Intelligence: A Modern Approach. 4th Edition. Pearson, 2021.