

1. Архитектура информационных систем

1.1. Модуль 1: Архитектурные основы

1.1.1. Лекция 1: Введение. Принципы построения приложения

- Оптимизация системы требует выбора критерия оптимизации
- Принципы SOLID
- Dependency Injection
- Clean Architecture

1.1.2. Лекция 2: Альтернативные архитектурные паттерны

- Архитектура на основе фреймворка (Rails, Django)
- Модель акторов (Erlang/Elixir/Akka)
- Проектирование, ориентированное на данные (Data-oriented design)

1.1.3. Лекция 3: Протоколы взаимодействия

- REST: стандартный веб-протокол
- GraphQL: клиент запрашивает только нужные данные
- gRPC: бинарный протокол, высокая производительность
- WebSocket: двусторонний канал

1.2. Модуль 2: Фундаментальные структуры хранилищ данных

1.2.1. Лекция 4: Базовые механизмы хранения данных

- CSV-файлы как минимальная форма персистентности
- CSV с добавлением записей и журналированием изменений
- Хеш-индексы как первый уровень ускорения выборки

1.2.2. Лекция 5: Структурированные индексы

- Отсортированная таблица строк (Sorted String Table, SSTable)
- Дерево слияния структурированного журнала (Log-Structured Merge-Tree, LSM-Tree)
- Когда использовать LSM-деревья и B-деревья (LSM хороша для записи, B-Tree для чтения)

1.2.3. Лекция 6: ACID, уровни изоляции, аналитические и транзакционные хранилища

- ACID (Atomicity, Consistency, Isolation, Durability)
- Уровни изоляции
 - Read Uncommitted
 - Read Committed
 - Repeatable Read
 - Serializable
- Аналитические базы данных (OLAP) и колоночные хранилища
- Сжатие колонок
- Сравнение подходов

1.3. Модуль 3: Другие хранилища данных

1.3.1. Лекция 7: Аналитика в оперативной памяти

- Vectorized execution (блочная обработка)
- DuckDB
 - Чтение из различных источников
 - Трансформация данных
 - Запись

1.3.2. Лекция 8: Кэш в оперативной памяти

- Skip List
- Incremental Rehashing
- HyperLogLog

- Redis
 - Архитектура (однопоточность, event-driven)

1.4. Модуль 4: Методы поиска похожих/близких элементов

1.4.1. Лекция 9: Специализированные индексы для поиска

- Inverted Index (перевёрнутые индексы)
 - Постинг-листы (posting lists)
 - Применение: Elasticsearch, Lucene
- Trie (префиксные деревья)
 - Автодополнение, префиксный поиск
- Bitmap indexes
 - Для категориальных данных

1.4.2. Лекция 10: Векторные индексы для поиска подобия

- Основы векторных представлений
- HNSW (Hierarchical Navigable Small World)
 - Граф как индекс
 - Поиск соседей
- LSH (Locality-Sensitive Hashing)
 - Вероятностный поиск
- Примеры: Weaviate, Pinecone, Qdrant

1.4.3. Лекция 11: Геопространственные структуры

- R-Tree
 - Bounding boxes, MBR
 - Применение: PostGIS
- QuadTree и KD-Tree
 - Разбиение пространства
 - Когда какое дерево эффективнее

1.5. Модуль 5: Операционные компоненты

1.5.1. Лекция 12: Прокси и Rate Limiting

- Forward Proxy / Reverse Proxy
- Распределение нагрузки
- Ограничение частоты запросов (Token Bucket, Sliding Window)

1.5.2. Лекция 13: Безопасность и управление доступом

- Аутентификация и авторизация
- RBAC и ABAC
- Single Sign-On

1.5.3. Лекция 14: Логирование, мониторинг и планирование задач

- Сбор логов
- Сбор метрик (Prometheus)
- Мониторинг и диагностика (Grafana)
- Планировщики (Celery, Airflow, Dagster)

1.6. Модуль 6: System Design

1.6.1. Лекция 15: Разбор дизайна конкретной системы

- Применение всех концепций на практике

2. Распределённые системы обработки данных

2.1. Модуль 1: Компоненты распределённых систем

2.1.1. Лекция 1: Основы распределённых систем

- CAP-теорема
 - Следствия компромиссов между согласованностью, доступностью и устойчивостью к разделению сети
 - Примеры CA, AP, CP систем
- Механизмы репликации данных
 - Архитектура “ведущий-подчинённый” (Leader-Follower)
 - Репликация с использованием кворумов (quorum-based replication)
- ZooKeeper: распределённое решение проблем согласованности и отказоустойчивости
 - Механизмы решения классических проблем распределённых систем
 - Применение в системах Kafka и ClickHouse

2.1.2. Лекция 2: Гарантии консистентности и координация распределённых транзакций

- Алгоритмы достижения консенсуса
 - Raft
 - Paxos
- Модель конечной консистентности (Eventual consistency) и требование идемпотентности операций
- Протокол двухфазного коммита (Two-Phase Commit, 2PC)
 - Этапы выполнения
 - Гарантии надёжности

2.2. Модуль 2: Распределённые сервисы

2.2.1. Лекция 3: Асинхронное взаимодействие сервисов

- Event Bus и Event-Driven Architecture
- Publish-Subscribe vs Message Queue
- Saga Pattern

2.2.2. Лекция 4: Kafka

- Внутреннее устройство

2.2.3. Лекция 5: Kubernetes

- Роль и место в современных архитектурах
- Области использования
- Внутреннее устройство

2.2.4. Лекция 6: ClickHouse

- Архитектура системы хранения и обработки данных
 - Принципы организации хранения
 - Методы компрессии данных
- Партиционирование
 - Выбор ключа
 - Управление партициями (добавление, удаление)
- Таблицы серии MergeTree как основной механизм персистентности
 - Базовая таблица MergeTree и её специализированные варианты
 - ReplacingMergeTree для управления версионированием
 - ReplicatedMergeTree для обеспечения отказоустойчивости

2.2.5. Лекция 7: Apache Spark

- Роль и место в современных архитектурах
- Области использования
- Внутреннее устройство
- API

- Dataframe
- SparkSQL

2.2.6. Лекция 8: Erlang

- Роль и место в современных архитектурах
- Erlang VM (BEAM) и процессы
- Модель акторов
- Fault tolerance и “Let it crash” философия
- OTP фреймворк (Supervisor, GenServer)
- Синхронизация и распределённые вычисления
- Elixir

2.3. Модуль 3: Распределённые БД

2.3.1. Лекция 9: Большие данные

- Свойства
 - Объем
 - Скорость
 - Разнообразие
 - Достоверность
- Архитектуры
 - Modern Data Architecture
 - Lambda
 - Lakehouse

2.3.2. Лекция 10: Full Text Search в распределённых системах

- Elasticsearch и Solr
 - Sharding: hash-based, range-based
 - Replication и replica factor
 - Eventual consistency при индексировании
- Распределённый поиск
 - Broadcast query, merge результатов
 - Search After вместо offset/limit
- Масштабирование
 - Hot shards проблема
 - Index rollover, segment merging

2.3.3. Лекция 11: Документные хранилища

- MongoDB: sharding strategies, write concerns
- Object storage: MinIO

2.3.4. Лекция 12: Распределённые пространственные и графовые БД

- Графовые БД
- Пространственные БД
- Шардирование графов и пространственных данных