

Все главы переработаны: одни кардинально, другие чуть меньше. В тексте отражены технологические изменения, произошедшие за последнее десятилетие, **полный рефакторинг под реалии 2025 года**.

Наибольшие технические изменения связаны с **развитием ИИ и облачных архитектур**. Добавлены разделы о системах обработки данных, поддерживающих ИИ и машинное обучение, включая векторные индексы (используемые в семантическом поиске), объекты DataFrame (используемые для обучающих наборов данных) и системы пакетной обработки для подготовки больших объемов обучающих данных.

Добавлено обсуждение механизмов синхронизации и ПО с приоритетом локальных операций, механизмов координации и обеспечения надежности, формальных методов и рандомизированного тестирования, GraphQL и некоторых других технологий, о которых стоит знать. В книгу включено обсуждение некоторых юридических аспектов. Исключено обсуждение парадигмы MapReduce, которая считается устаревшей, а **глава, посвященная пакетной обработке, полностью переписана**.

1. Кардинально новая архитектура и структура книги

- Глава 1 теперь не просто введение в термины (надежность, масштабируемость и т.д.), а полноценный обзор компромиссов в архитектуре информационных систем. Она сразу погружает читателя в дилеммы: аналитика vs. оперативка, облако vs. on-premise (по лицензии), распределенные vs. одноузловые системы, а также этические аспекты.
- Остальные главы полностью переписаны, перегруппированы и дополнены.

2. Сквозная тема: облачные технологии и их влияние

Если в первом издании упор делался на распределенные системы и Hadoop, то второе издание в основном об облачных архитектурах.

- Разделение хранения и вычислений: одна из центральных идей. Автор объясняет, как устроены современные облачные БД (Snowflake, BigQuery), используя объектные хранилища (S3) как единый слой хранения, а вычислительные ресурсы — как эфемерные.
- Архитектура без дисков (Zero-Disk Architecture): новая концепция, где узлы не имеют постоянного состояния, а все данные читаются из объектного хранилища. Это позволяет достичь гибкости и дешевизны.
- Добавлены обсуждения балансировщиков нагрузки, DNS, etcd, ZooKeeper и концепции сервисных сетей (service meshes).

3. Полное переосмысление пакетной обработки (самое радикальное изменение)

- Во 2м издании полностью удалено подробное описание MapReduce, автор считает его устаревшим.
- Затем автор переходит к распределенным файловым системам (HDFS, S3), оркестраторам (Kubernetes, YARN) и движкам потоков данных (Spark, Flink).

4. Интеграция AI/ML и новых типов данных

- Векторные индексы и семантический поиск: большой новый раздел о том, как работают векторные БД (HNSW, IVF) для задач AI.
- Подробно обсуждаются таблицы DataFrame (Pandas, Spark) как мост между реляционными БД и матрицами для ML.
- ИИ и большие языковые модели (LLM): добавлены разделы о том, как пакетная обработка используется для подготовки данных (ETL) для обучения LLM (токенизация, фильтрация).

5. Эволюция моделей данных

- Впервые добавлен подробный разбор GraphQL как языка запросов для клиент-серверных приложений, с примерами.
- Регистрация событий: выделена в отдельную, большую тему, а не просто упоминается.
- Графовые БД: глава дополнена обсуждением ISO-стандарта GQL и языка Datalog.

6. Этические и правовые аспекты

Первое издание заканчивалось главой «Будущее информационных систем», где этике был посвящен небольшой раздел.

- Второе издание выделяет эти проблемы в отдельную главу 14.
- Подробно обсуждаются предвзятость алгоритмов, дискриминация, слежка.
- Появляется обсуждение «алгоритмической тюрьмы» и права на объяснение решений.

7. Инструментарий и надежность (новые технические концепции)

- Тестирование распределенных систем: введено понятие детерминированного симуляционного тестирования (DST) и внедрения сбоев (Chaos Engineering) как обязательных практик.
- Формальные методы: упоминание TLA+ и рандомизированного тестирования стало значительно подробнее.