**УДК 004.62:004.75**

**ПОДХОДЫ К ХРАНЕНИЮ И ОБРАБОТКЕ БОЛЬШИХ ОБЪЕМОВ СТРУКТУРИРОВАННЫХ ДАННЫХ**

Быченков А.К., Матчин В.Т.

Федеральное государственное бюджетное образовательное учреждение высшего образования «МИРЭА – Российский технологический университет», Москва

Идея: рассказать про различные подходы к хранению данных, описать их плюсы и минусы.

1. Введение (появляется большое количество данных, их необходимо эффективно хранить и обрабатывать)
2. Хранение данных в СУБД (best practices): лучше хранить в колоночном, чтобы экономить место, сжимать данные, реплицировать данные для минимизации риска потери данных, шардирования для эффективности обработки данных

Недостатки:

1. Хранение данных в файловых системах

Best practices: репликация данных (кластерная архитектура хранения), сжатие данных (выбирать те кодеки сжатия, которые обеспечивают максимальное сжатие при оптимальной скорости), эффективная структура хранения (преимущественно колоночная структура), версионирование данных для возможности восстановления (использование снапшотов либо использование механизмов git), механизмы обеспечения консистентности данных (механизмы гарантирования требований ACID)

Преимущества:

отделение хранения данных от их обработки

независимость от системы хранения данных (легкая миграция из одной системы в другую);

возможность использовать любые системы обработки запросов

1. Заключение (выбор оптимальной стратегии хранения данных должен быть основан на анализе конкретных потребностей проекта, стоимости реализации и обслуживания, а также учете текущих и потенциальных требований к масштабируемости, производительности и безопасности). Можно реализовать стек самостоятельно либо используя готовые Cloud Native решения

В эпоху цифровой трансформации и массового накопления данных сбор, хранение и анализ информации становятся ключевыми аспектами успешной деятельности организаций во всех сферах. С каждым днем возрастает сложность этой задачи в связи с ростом объемов, разнообразием и скоростью поступления данных из различных источников.

3.1. Реляционные СУБД

До определенного момента, практически единственным ответом на вопрос «как хранить и обрабатывать данные?» являлась какая-нибудь реляционная СУБД. Но с увеличением объемов появились проблемы, с которыми классическая реляционная архитектура не справлялась, поэтому инженерам пришлось придумывать новые решения. Попробуем представить те шаги, которые можно предпринять, если СУБД прекращает справляться с объемом выполняемых операций.

**Список использованных источников:**

[**https://scholar.google.com/**](https://scholar.google.com/)

[**https://www.ssrn.com/index.cfm/en/**](https://www.ssrn.com/index.cfm/en/)

**© Быченков А.К., Матчин В.Т., 2024**