

Университет ИТМО, факультет программной инженерии и компьютерной техники
Двухнедельная отчётная работа по «Информатике»: аннотация к статье

Дата прошедшей лекции	Номер прошедшей лекции	Название статьи/главы книги/видеолекции	Дата публикации	Размер статьи	Дата сдачи
10.09.25	1	Использование изменения системы отсчета для улучшения результатов анализа с помощью закона бенфорда	31.03.2025	1208	24.09.25
24.09.25	2	Метод регенерационного блочного кодирования	13.10.2023	3356	08.10.25
08.10.25	3	Многоцелевая фильтрация текста с использованием регулярных выражений	10.06.2024	2124	05.11.25
22.10.25	4	Формат PARQUET: Обзор структуры и методы оптимизации для высокой производительности в распределённых системах	11.04.2025	3651	05.11.25
05.11.25	5				
19.11.25	6				
03.12.25	7				

Выполнил(а) Бых Даниил Максимович, № группы P3109, оценка

<p>Прямая полная ссылка на источник или сокращённая ссылка https://cyberleninka.ru/article/n/format-parquet-obzor-struktury-i-metody-optimizatsii-dlya-vysokoy-proizvoditelnosti-v-raspredelyonnyh-sistemah/viewer</p>
<p>Теги, ключевые слова или словосочетания Распределённые системы хранения, большие данные, формат файлов.</p>
<p>Перечень фактов, упомянутых в статье 1) Эффективное представление информации очень важно для масштабируемости систем. 2) В отличие от строковых форматов, Parquet хранит данные по колонкам. 3) Параллельная обработка данных эффективно распределяет нагрузку в кластерах, существенно ускоряя выполнение запросов. 4) Существует проблема в требовании точного совпадения паттернов и логарифмическую зависимость скорости фильтрации от сложности шаблонов.</p>
<p>Позитивные следствия и/или достоинства описанной в статье технологии 1) Значительное снижение объёма хранимых данных за счёт колонного хранения. 2) Ускорение выполнения аналитических и выборочных запросов за счет быстрого доступа к необходимым колонкам. 3) Высокая гибкость за счет поддержки различных алгоритмов сжатия, что позволяет оптимизировать работу.</p>
<p>Негативные следствия и/или достоинства описанной в статье технологии 1) Вычислительные затраты могут быть значительными при использовании алгоритмов с высокой степенью сжатия. 2) Слишком маленькие файлы вызывают излишние накладные расходы на файловую систему, а большие - к неэффективному использованию памяти. 3) Неудачная конфигурация параллельной обработки может привести к избыточному распределению ресурсов.</p>
<p>Ваши замечания, пожелания преподавателю или анекдот о программистах :)</p>