

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ РАЗРАБОТКИ СИСТЕМ УПРАВЛЕНИЯ БОЛЬШИМИ ДАННЫМИ

Соколинский Леонид Борисович

№	Тема
1.	Общая схема обработки запроса в реляционной СУБД. Реляционная алгебра и реляционные операции. Представление реляционного выражения в виде логического плана. Общая схема компиляции запроса.
2.	Разбор запроса. Синтаксический анализ запроса. Грамматика SQL. Построение дерева разбора. Разрешение представлений. Верификация запроса.
3.	Конверсия запроса в логический план.
4.	Логическая оптимизация. Алгебраические законы. Оптимизация операций выборки, проекции и удаления дубликатов. Оптимизация композиции операций выборки и прямого произведения.
5.	Система хранения данных. Поля, записи, блоки. Использование буферного пула.
6.	Индексы. В-деревья.
7.	Оптимизация физического плана. Оценка стоимости реляционных операций. Статистические характеристики данных. Гистограммы. Выбор порядка соединений.
8.	Исполнитель запросов. Итераторы. Алгоритмы соединений.
9.	Модель параллельной обработки больших данных MapReduce.
10.	Скетчинг данных. Фильтры Блума. Count-Min скетч. Определение числа уникальных элементов с помощью HyperLogLog.

Литература:

1. Гарсиа-Молина Г., Ульман Дж., Уидом Дж. Системы баз данных. Полный курс. М.: Издательский дом "Вильямс", 2004. 1088 с.
2. Лесковец Ю., Раджараман А., Ульман Д.Д. Анализ больших наборов данных. Москва: ДМК Пресс, 2016. 498 с.
3. Garcia-Molina H., Ullman J.D., Widom J. Database Systems: The Complete Book. 2nd ed. Upper Saddle River, New Jersey: Pearson Prentice Hall, 2002. 1203 p. DOI:10.1192/bjp.112.483.211-a.
4. Leskovec J., Rajaraman A., Ullman J.D. Mining of massive datasets. 2nd ed. New York, NY, USA: Cambridge University Press, 2014. 565 p. DOI:10.1017/CBO9781139058452.