

1. Предварительная формулировка темы.

Разработка оптимизатора запросов для колонно-ориентированной in-темогу базы данных.

2. Цель ВКР.

Целью выпускной квалификационной работы является разработка оптимизатора запросов для колонно-ориентированной in-темогу базы данных

3. Краткое описание прикладного процесса, для автоматизации которого будет предназначен разрабатываемый компонент.

Разрабатываемый компонент является частью базы данных общего назначения, специализированной для поиска по большому количеству критериев. Подобную базу можно использовать для большого количества разнообразных задач, в первую очередь для организации поиска по товарам в больших торговых площадках (wikimart.ru - 1.8 млн товаров, Яндекс.Маркет - 57 млн. предложений, amazon.com - 36 млн книг).

Для подобных задач характерен поиск по большому количеству критериев(около ста для одного предложения Яндекс.Маркета, двадцати - для amazon.com) и разные шаблоны поиска по критериям (поиск по диапазону, по совпадению, полнотекстовый поиск). Задача осложняется неравномерным распределением критериев в данных и потенциально большими размером результирующих выборок. При этом, к базе данных предъявляются высокие требования к производительности(как и к нагрузке, так и к скорости получения результата). Стандартные реляционные базы данных не удовлетворяют таким требованиям, NoSQL - так же не реализуют необходимую функциональность.

Система управления базами данных Vindur предназначена для решения задач быстрого многокритериального поиска и представляет собой хранящееся в памяти колонно-ориентированное хранилище. Колонно-ориентированное представление данных предназначено для задач сложного поиска (см. статью “C-store: a column-oriented DBMS.” Mike Stonebraker, Daniel J. Abadi, Adam Batkin, Xuedong Chen, Mitch Cherniack, Miguel Ferreira, Edmond Lau, Amerson Lin, Sam Madden, Elizabeth O’Neil, Pat

O'Neil, Alex Rasin, Nga Tran, Stan Zdonik. August 2005. VLDB '05: Proceedings of the 31st international conference on Very large data bases)

4. Требования к разрабатываемому компоненту.

Функциональные требования:

1. Генерация плана выполнения запроса
2. Возможность модификации плана в ходе исполнения
3. Возможность подстройки оптимизатора под конкретные запросы
4. Возможность настройки сложности оптимизации
5. Возможность подстройки оптимизатора под конкретное железо

Формальные показатели:

1. Увеличение скорости выполнения запроса в два раза относительно отсутствия оптимизатора

Нефункциональные требования:

1. Нетребовательность к ресурсам (не составлять план запроса дольше, чем исполнять его)

5. Руководитель. В каком объеме предоставлена информация для выполнения ВКР.

Научный руководитель и заказчик системы - Дельгадо Филипп Игоревич, куратор академических программ, JetBrains. Имеется возможность непосредственного взаимодействия, техническая документация Oracle Java и стандарты кодирования Java Coding Conventions

6. Описание технологий, которые будут использованы при разработке.

Язык реализации - Java. Некоторые фреймворки и библиотеки - Spring, JUnit, JProfiler, Apache Lucene, Java EWAH. Система контроля версий - Git. Репозиторий - github.com

7. Описание методов тестирования и оценки качества разработанного компонента.

Тестирование - Performance, Unit testing. Так же, для оценки качества оптимизатора и базы в целом предполагается использование генератора тестовых данных и реальных данных, взятых из открытого источника.

8. Описание результатов работы.

1. Артефакты проектирования:
 - (a) Различные алгоритмы оптимизации
 - (b) Диаграммы UML программной архитектуры
2. Артефакты разработки:
 - (a) Исходный код в репозитории
3. Документация по системе
4. Описание различных подходов к оптимизации.
5. Сравнение различных алгоритмов оптимизации на тестовых данных