

Лабораторная работа № 3

Цели и задачи лабораторной работы

Целями выполнения лабораторной работы являются:

1. Закрепление имеющихся знаний об индексах и оптимизации запросов. Изучение методологии оптимизации базы данных.
2. Приобретение навыков настройки производительности языка запросов к базе данных.
3. Приобретение навыков оптимизации базы данных с помощью использования индексов.

Краткие теоретические сведения

- Настройка производительности базы данных относится к набору действий и процедур, предназначенных для обеспечения того, чтобы запрос конечного пользователя обрабатывался СУБД в наименьшее количество времени. Настройка производительности SQL относится к действиям на стороне клиента, которые предназначены для генерации кода SQL, который возвращает правильный ответ за наименьшее количество времени, используя минимальное количество ресурсов на стороне сервера. Настройка производительности СУБД относится к действиям на стороне сервера, которые ориентированы таким образом, чтобы СУБД была должным образом сконфигурирована для быстрого реагирования на запросы клиентов при оптимальном использовании существующих ресурсов.

- Статистика базы данных относится к количеству измерений, собранных СУБД, которые описывают моментальный снимок характеристик объектов базы данных. СУБД собирает статистику об объектах, таких как таблицы, индексы и доступные ресурсы, которые включают число используемых процессоров, скорость процессора и доступное временное пространство. СУБД использует статистику для принятия критических решений по повышению эффективности обработки запросов.

- СУБД обрабатывают запросы в три этапа. На этапе синтаксического анализа СУБД анализирует запрос SQL и выбирает наиболее эффективный план доступа/выполнения. На этапе выполнения СУБД выполняет запрос SQL, используя выбранный план выполнения. На этапе выборки СУБД выбирает данные и отправляет набор результатов обратно клиенту.

- Индексы имеют решающее значение в процессе, который ускоряет доступ к данным. Индексы облегчают поиск, сортировку и использование

агрегатных функций и операций объединения. Повышение скорости доступа к данным происходит потому, что индекс представляет собой упорядоченный набор значений, который содержит ключ индекса и указатели. Разреженность данных относится к числу различных значений, которые может иметь столбец. Индексы рекомендуются в столбцах высокой разреженности, используемых в условиях поиска.

- Во время оптимизации запросов СУБД должна выбирать, какие индексы использовать, как выполнять операции объединения, какую таблицу использовать первой и т.д. Каждая СУБД имеет свои алгоритмы определения наиболее эффективного способа доступа к данным. Двумя наиболее распространенными подходами являются оптимизация на основе правил и затрат.
- Оптимизатор на основе правил использует заданные правила и точки для определения наилучшего подхода к выполнению запроса. Оптимизатор на основе затрат использует сложные алгоритмы, основанные на статистике об объектах, к которым осуществляется доступ, для определения наилучшего подхода к выполнению запроса. В этом случае процесс оптимизатора складывает стоимость обработки, затраты на ввод/вывод и затраты ресурсов (ОЗУ и кэш), чтобы определить общую стоимость заданного плана выполнения.
- Настройка производительности SQL связана с написанием запросов, которые эффективно используют статистику. В частности, запросы должны эффективно использовать индексы. Индексы очень полезны, когда необходимо выбрать небольшое подмножество строк из большой таблицы на основе условия.
- Формулировка запроса имеет дело с тем, как преобразовать деловые вопросы в конкретный код SQL для получения требуемых результатов. Чтобы сделать это, нужно тщательно оценить, какие столбцы, таблицы и вычисления необходимы для создания желаемого результата.
- Настройка производительности СУБД включает такие задачи, как управление процессами СУБД в оперативной памяти (выделение памяти для целей кэширования) и управление структурами в физической памяти (выделение пространства для файлов данных).

Контрольные вопросы для допуска к работе

1. Что такое настройка производительности SQL?
2. Что такое настройка производительности базы данных?

3. На чем сосредоточена большая часть действий по настройке производительности и почему она существует?
4. Что такое статистика базы данных и почему она важна?
5. Какие статистические измерения базы данных типичны для таблиц, индексов и ресурсов?
6. Чем отличается обработка операторов SQL DDL (таких как CREATE TABLE) от обработки, требуемой операторами DML?
7. Если индексы так важны, почему бы не проиндексировать каждый столбец в каждой таблице?
8. В чем разница между оптимизатором на основе правил и оптимизатором на основе затрат?
9. Что такое подсказки оптимизатора и как они используются?

Порядок выполнения работы

В процессе выполнения лабораторной работы необходимо:

1. Определить столбцы для создания кластерных индексов. Обосновать выбор. Создать кластерные индексы для каждой таблицы в базе данных для заданной предметной области.
2. Определить столбцы для создания некластерных индексов. Обосновать выбор. Создать некластерные индексы.
3. Сравнить результаты выполнения запросов до и после создания индексов.