# Определение аналитической зависимости времени выполнения запроса от его параметров

Время выполнения запроса к БД () зависит от нескольких параметров, отраженных в формуле (1):

где – время передачи запроса от приложения к БД (умножается на 2, так как учитывается и время передачи запроса от БД к приложению) равное 0,9мс; – время диспетчерской работы координатора по распределению нагрузки между серверами, равное 1мс; – время выполнения запроса сервером, которое зависит от особенностей выполнения запроса на конкретном сервере.

Расчет времени выполнения запроса на сервере () считается следующим образом:

где – время чтения с диска в оперативную память, которое зависит от количества записей в таблице; – время выполнения запроса сервером, которое зависит от разных типов индексов; – время выполнения запроса сервером, которое зависит от категории СУБД и особенностей моделей данных.

Время выполнения запроса сервером, которое зависит от разных типов индексов () принимается равным нулю, так как заранее заготовленные индексированные таблицы не создавались. Так как в аналитической зависимости присутствует определенная погрешность, то для её уменьшения в параметре заложен поправочный коэффициент, уменьшающий расхождение , в итоге мс.

Время рассчитывается по следующей формуле:

Где n – количество строк, которые будут рассматриваться при выполнении запроса на чтение, t – количество столбцов, которые будут рассматриваться при выполнении запроса на чтение, b – смещение, которое позволяет учесть влияние количества строк и столбцов на запрос.

Смещение (b) влияние того, что строка более затратная для поиска, чем столбце из-за внутреннего устройства СУБД Cassandra. Если приравнять строку и столбцы, то для одинокого количества данных (количество строк, умноженное на количество столбцов) время выполнения будет различное. Для вычисления необходимо поделить два времени выполнения при одинаковом количестве данных, но разном количестве строк и столбцов. Смещение зависит от устройства, в данной работе оно вычислено и равно 1,8.

Для определения время выполнения запроса используется механизм трассировки в Cassandra, который сохраняет данные о выполнения запроса, включая время начала и завершения.

# Описание устройства для проведения эксперемента

Оценка времени выполнения запросов запроса зависит от производительности устройства, на котором проводятся эксперименты. Поэтому приведем характеристики устройств, на которых будут проводиться эксперименты.

Технические устройства

1. Процессор: 8 ядерный процессор с тактовой частотой 2,6 ГГц;
2. Оперативная память: DDR3 8192 МБ со скоростью 3000 Мгц;
3. Диск: Пустой SSD диск объемом 10 Гб со скоростью 1900 Мбит/c;

Программное обеспечение:

1. Операционная система Ubuntu 24.04.

Практический эксперимент проводится на операционной системе Ubuntu 24.04, работающей на базе Virtual Box на Windows 10. Во время проведения эксперимента отсутствуют фоновые процессы, помимо фоновых процессов операционной системы.

В качестве основы для эксперимента в базу данных были загружены все варианты объединения строк и столбцов следующего количества: 10000/100000/200000/300000 строк и 5/10/15 столбцов. Каждый такой вариант объединения является одной частью эксперимента.

# Методика проведения эксперимента

Для проведения исследования по оценке времени выполнения запросов на чтение с условиями требуется выполнить следующие задачи:

1. Создать базу данных и пространство ключей;
2. Создать таблицу с нужным для определённой части эксперимента количеством строк и столбцов и заполнить её тестовыми данными;
3. Выполнить запрос для получения экспериментального времени выполнения запроса. Данный пункт выполняется три раза подряд, после чего для всех результатов рассчитывается среднее значение, которое будет считаться экспериментальным временем выполнения запросов;
4. Пункты 1-3 повторяются для каждой части эксперимента.

# Экспериментальные исследования производительности

При проведении практического эксперимента соберем данные для формирования зависимости времени выполнения запроса от количества строк и столбцов.

В табл. 1 представлен результат выполнения эксперимента с последовательным увеличением количества строк и столбцов в таблицах.

Таблица 1. Результаты экспериментального выполнения запроса

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | **Столбцы** | **Строки** | | | |
| **10 000** | **100 000** | **200 000** | **300 000** |
| **t, мс** | **Теоретическое время выполнения** | **5** | 76,23372036 | 1047,965812 | 2306,638071 | 3659,371715 |
| **10** | 142,7462475 | 1962,296822 | 4319,137613 | 6852,106627 |
| **15** | 213,3932607 | 2933,463572 | 6456,736164 | 10243,30517 |
| **Экспериментальное время выполнения** | **5** | 118,698 | 819,4866667 | 3835,273 | 5686,882667 |
| **10** | 117,846 | 1125,921333 | 4461,397 | 7657,586667 |
| **15** | 154,1826667 | 3172,168333 | 5985,99 | 8049,659333 |

# Оценка результатов

Обобщив теоретические выкладки и практические исследования, получим следующие результаты, представленные в табл. 2.

Таблица 2. Анализ времени выполнения запроса

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **Время выполнения, мс** | |  | |
| **Строка** | **Столбец** | **Экспериментальное** | **Теоретическое** | **Ошибка** | **Абсолютная ошибка, мс** |
| 10 000 | 5 | 118,698 | 76,23372036 | 36% | 42,46427964 |
| 10 000 | 10 | 117,846 | 142,7462475 | 21% | 24,9002475 |
| 10 000 | 15 | 154,1826667 | 213,3932607 | 38% | 59,210594 |
| 100 000 | 5 | 819,4866667 | 1047,965812 | 28% | 228,4791453 |
| 100 000 | 10 | 1125,921333 | 1962,296822 | 74% | 836,375489 |
| 100 000 | 15 | 3172,168333 | 2933,463572 | 8% | 238,704761 |
| 200 000 | 5 | 3835,273 | 2306,638071 | 40% | 1528,634929 |
| 200 000 | 10 | 4461,397 | 4319,137613 | 3% | 142,259387 |
| 200 000 | 15 | 5985,99 | 6456,736164 | 8% | 470,746164 |
| 300 000 | 5 | 5686,882667 | 3659,371715 | 36% | 2027,510952 |
| 300 000 | 10 | 7657,586667 | 6852,106627 | 11% | 805,48004 |
| 300 000 | 15 | 8049,659333 | 10243,30517 | 27% | 2193,645837 |

Данные из таблицы 2 свидетельствуют о том, что время при увеличении количества строк в 10 и более раз увеличивается примерно в 9 раз. Увеличение количества столбцов в меньшей степени, чем строки влияет на производительность СУБД, например, при увеличении в 2 раза, время выполнения увеличивается примерно на чуть меньше 50%.

Гистограмма сравнительного анализа приведена на рис. 1. Шкала слева показывает время выполнения запросов (t, мс), шкала снизу - количество строк. Цветовая гамма столбцов на гистограмме показывает количество столбцов для экспериментального и теоретического времени выполнения.

Рисунок 1. Гистограмма сравнительного анализа