

РЕШЕНИЕ ЗАДАЧИ ПОВЫШЕНИЯ ПРОИЗВОДИТЕЛЬНОСТИ БАЗЫ ДАННЫХ С ПРИМЕНЕНИЕМ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

В. А. Липницкий, Л. В. Синкевич

Кафедра информатики, Белорусский государственный университет информатики и радиоэлектроники

Кафедра высшей математики, Военная академия Республики Беларусь

Минск, Республика Беларусь

E-mail: oktiabrina.sinkevich@gmail.com, valipnitski@yandex.ru

В данной статье представлен анализ сочетания подходов к оптимизации баз данных корпоративных систем. Процесс повышения производительности получается более эффективным, если принимать во внимание не только метаматематические подходы, но и особенности сферы применения той или иной информационной системы.

Решение задачи повышения производительности информационной системы можно свести к решению задачи по повышению производительности пользовательских операций, требующих оптимизации.

Процесс оптимизации является циклическим и включает в себя следующие шаги:

- выявление перечня пользовательских операций, требующих оптимизации;
- получение диагностических данных, относящихся к некоторому исследуемому периоду или исследуемой пользовательской операции и позволяющих определить причины задержек во времени;
- выполнение оптимизации тех операций, которые дают высокий совокупный эффект и переход к первому шагу.

Наиболее важно на первоначальном этапе определить некоторый перечень пользовательских операций, улучшение производительности по которым влияет на систему в целом [1].

Важность операций можно определить по следующим признакам:

- критичность выполнения операции для пользователей;
- продолжительность операции;
- частота вызова операции;
- потребление ресурса, который нужно экономить.

Закон Амдала гласит, что увеличение производительности, достигаемое некоторым системным усовершенствованием, ограничено потребляемой усовершенствованным компонентом долей в общем времени. Другими словами, повышение производительности пропорционально доле времени, занимаемой улучшенным компонентом, поэтому, в первую очередь следует стремиться уменьшать наиболее весомые составляющие во времени отклика в важных и критичных пользовательских операциях. Этот закон объясняет, почему составляющие времени отклика следует рассматривать в порядке убывания и как

достичь наибольшей отдачи от вложенных в процесс оптимизации усилий [2].

После определения базисных операций, наступает время для сбора диагностических данных. Наиболее рациональным источником диагностических данных является расширенная трассировка. Данные о трассировке представляют собой упорядоченную историю действий ядра базы данных, предпринятых в ответ на запросы, посылаемые приложениями.

После обнаружения в файле расширенной трассировки события, чье время ожидания занимает по меркам системы много времени, следует определить, какие выполняемые команды можно изменить с тем, чтобы уменьшить расход времени. Для решения этой задачи необходимо сопоставить длительность каждого события первому вызову базы данных, который следует за ним. Этот прием поможет определить, какая из команд отвечает за возникновение каждого периода ожидания и внутри вызовов в базе данных и за ее пределами [3].

Собрав диагностические данные, можно приступать к выбору путей решения проблемы. Для комплексного решения вопроса оптимизации производительности не достаточно только технической оценки, кроме этого нужно учитывать финансовые последствия от предполагаемых действий.

Аналитическая работа, состоит из двух частей:

- определение компонентов, которые могут дать наибольший экономический эффект при оптимизации и выделение диагностических данных по этому компоненту из общих данных о времени отклика;
- составление предположений о путях сокращения доли отклика по компоненту и проверка предположения на тестовой системе, при этом следует рассматривать достаточное количество идей, для того, чтобы убедиться в том, что учтены все возможности.

Для получения ответов на вопросы касательно моделирования производительности применяют теорию массового обслуживания, модели которой позволяют предсказать время отклика системы в гипотетических ситуациях и проследить взаимосвязь между разнообразными параметрами оптимизации. Таким образом, весь анализ производительности баз данных можно свести к отношению между поставщиками и потребителями, при этом учитывая острую конкуренцию за ресурсы.

Фундаментальное отношение теории массового обслуживания – математическое определение времени отклика [1]:

$$R = S + W,$$

где S – время обслуживания;

W – время задержки в очереди.

В целях обобщения, систему массового обслуживания можно рассматривать, как «черный ящик», который принимает что-то на входе, обрабатывает это и формирует вывод каким-то образом преобразованных входных данных. Компьютерные системы можно рассматривать в качестве модели $M/M/n$. Интенсивность поступления запросов во многих из них подчиняется распределению Пуассона [4], имеющего обратную связь с экспоненциальным распределением и функция плотности вероятности которого есть

$$f(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!},$$

$$\lambda > 0,$$

$$x = 0, 1, 2, \dots,$$

где λ – средняя интенсивность поступления запросов.

Модель $M/M/n$ – это модель, удовлетворяющая следующим условиям:

- экспоненциальное время между поступлениями запросов;
- экспоненциальное время обслуживания;
- n однородных параллельных независимых каналов обслуживания;
- обслуживание в порядке поступления (first-come, first-served).

Основная сложность работы с реальными системами массового обслуживания состоит в том, что невозможно точно предсказать время поступления запросов. Время обслуживания системы также является случайной величиной. Именно поэтому так важна возможность прогнозирования времени задержки и максимальной пропускной способности, которую предоставляет теория массового обслуживания. Для систем $M/M/n$ прогнозирование времени задержки [2] выглядит как

$$W = \frac{C(n, \rho)}{n \cdot \mu \cdot (1 - \rho)},$$

$$C(n, \rho) = \frac{\frac{(n \cdot \rho)^n}{n!}}{(1 - \rho) \cdot \sum_{k=0}^{n-1} \frac{(n \cdot \rho)^k}{k!} + \frac{(n \cdot \rho)^n}{n!}},$$

где $C(n, \rho)$ – вероятность того, что запрос будет поставлен в очередь на предоставление обслуживания (рассчитывается по формуле Эрланга); n – количество параллельных каналов обслуживания внутри системы; μ – скорость обслуживания; ρ – средний коэффициент использования одного канала.

Максимальная пропускная способность системы массового обслуживания λ_{max} это наибольшая частота поступлений запросов, которые система обрабатывает без превышения допустимого времени отклика. Иными словами, это тот максимум производительности, которого можно требовать от системы, не опасаясь чрезмерного ухудшения времени отклика. Рассчитать величину λ_{max} можно с помощью метода биекций, основанном на теореме о промежуточных значениях [4]. Опираясь на вышеприведенную информацию, можно вычислить вероятность того, что время отклика системы окажется не больше некоторого допустимого значения пользовательского времени. Для систем $M/M/n$ вероятность успешности некоторой гипотезы по оптимизации [2] выглядит как

$$P(R \leq r) = \frac{n \cdot (1 - \rho) - W_q(0)}{n \cdot (1 - \rho) - 1} \cdot (1 - e^{-\mu \cdot r}) -$$

$$\frac{1 - W_q(0)}{n \cdot (1 - \rho) - 1} \cdot (1 - e^{-(n \cdot \mu - \lambda) \cdot r}),$$

$$W_q(0) = 1 - \frac{(n \cdot \rho)^n \cdot \rho_0}{n! \cdot (1 - \rho)},$$

$$\rho_0 = \left(\sum_{m=0}^{n-1} \frac{(n \cdot \rho)^m}{m!} + \frac{(n \cdot \rho)^n}{n! \cdot (1 - \rho)} \right)^{-1},$$

где n – количество параллельных каналов обслуживания внутри системы;

ρ – средний коэффициент использования одного канала;

μ – скорость обслуживания;

λ – интенсивность входного потока системы.

Рассмотренное сочетание подходов к оптимизации баз данных позволяет эффективно применять математическое моделирование, учитывая интересы пользователей. Моделирование с помощью теории массового обслуживания позволяет не только учесть зависимость между скоростью обслуживания, нагрузкой системы и длительностью ожидания, но и, что немаловажно, измерить вероятность успешности проводимой операции.

1. Льюис, Дж. Oracle. Основы стоимостной оптимизации / Дж. Льюис // Издательство: Питер, 2007. – 528 с.
2. Миллсап, К. Oracle. Оптимизация производительности / К. Миллсап // Издательство: Символ-Плюс, 2006. – 464 с.
3. Oracle Technology Network [Electronic resource] / Oracle. – Cambridge, 1998. – Mode of access: <http://www.oracle.com/technetwork/>. – Date of access: 10.09.2015.
4. Теория массового обслуживания / Г. И. Ивченко [и др.]. – М.: Либроком, 2012. – 304 с.