

Измерение производительности в реальных условиях

Мэри Финн
Менеджер по маркетингу продукции
Корпорация InterSystems

Резюме

Лучший способ сравнить производительность систем управления базами данных (СУБД) — протестировать их на конкретных приложениях, и лучше всего, если это будут реальные приложения, на которых предполагается эксплуатировать СУБД, а еще лучше – одно из используемых вами. Это особенно важно при оценке производительности постреляционной СУБД Caché, поскольку стандартные методики тестирования рассчитаны только на простейшие двумерные базы данных со строками и столбцами. Подобные тесты не позволяют адекватно оценить производительность СУБД на реальных приложениях, особенно когда речь заходит о сложных моделях данных.

Из-за ряда ограничений, установленных разработчиками распространенных СУБД, на сегодня опубликовано крайне мало данных о производительности таких реляционных систем, как Oracle и Microsoft SQL Server, в реальных условиях. Компании, проводящие подобные тесты по собственной инициативе, в большинстве случаев приходят к выводу, что Caché работает на 100-500% быстрее реляционных баз данных даже без оптимизации приложений, исключительно на стандартном коде SQL. Выигрыш в производительности становится еще больше при оптимизации приложений под постреляционную технологию Caché.

Введение

Многие компании проводят сравнение Caché с другими системами управления базами данных. В данном отчете приведены результаты тестирования СУБД, проведенные независимыми аналитиками, а также клиентами и потенциальными клиентами InterSystems в процессе сравнительного анализа различных баз данных. Из-за ряда ограничений со стороны разработчиков широко известных СУБД мы опустили названия компаний, проводивших тестирование, а также названия СУБД, принимавших участие в тестировании вместе с Caché.

Все тесты проводились на реальных приложениях и в реальных эксплуатационных условиях. Результаты тестов не оставляют места для сомнений. При работе с базами данных средствами SQL по производительности Caché превосходит Oracle, SQL Server и другие реляционные базы данных в 5 раз или больше.

В большинстве тестов, рассматриваемых в данном отчете, в качестве языка запросов применялся SQL. Такой выбор языка был обусловлен максимальной простотой сравнения Caché и реляционных СУБД. Во всех тестах превосходство Caché по производительности объясняется ее многомерной архитектурой, избавляющей от необходимости проведения сложных операций объединения таблиц. Однако SQL — не единственный способ работы с базами данных Caché. Если код приложений оптимизирован для работы с многомерными структурами данных Caché, скорость Caché становится еще выше.

Пример №1

Эксперты компании, занимающейся энергоснабжением, пришли к выводу, что Caché работает в 5 раз быстрее известной реляционной СУБД на их внутреннем приложении класса ETL (Extract, Transform, Load). Базы данных тестировались путем обработки 13 миллионов записей на реальном модуле приложения, предназначенном для обработки заказов. Базы данных тестировались на одинаковом наборе запросов и определений таблиц. Однако в модуле приложения, на котором проводилось тестирование, интенсивно использовались сохраненные процедуры, язык которых зависит от конкретной базы данных, поэтому эти процедуры были переписаны на языке сохраненных процедур Caché. На это преобразование потребовалось 40 человеко-часов работы.

В следующей таблице приведены итоги тестирования:

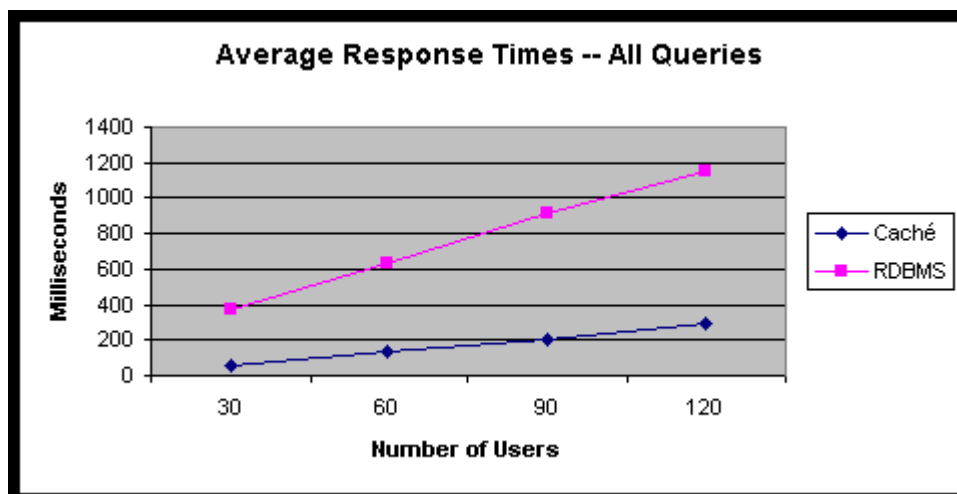
Задача	Затраченное время		Выигрыш в скорости Cache : РСУБД
	РСУБД	Cache	
Загрузка данных в тестовую таблицу	138 мин.	6 мин.	23 : 1
Загрузка второй тестовой таблицы и фильтрация данных	245 мин.	23 мин.	10 : 1
Создание таблицы измерений для заказов	168 мин.	24 мин.	7 : 1
Создание таблицы фактов для заказов	890 мин.	233 мин.	4 : 1
Общая продолжительность загрузки и фильтрации данных	1441min.	286 min.	5 : 1

Пример №2

Компания-поставщик решений в области здравоохранения, сравнила скорость обработки запросов SQL у Cache и одной из распространенных реляционных СУБД. На наборе данных о состоянии пациентов за длительное время (7 таблиц, свыше 6,5 миллиона записей) были протестированы 8 запросов. Во время теста имитировалась одновременная работа 30, 60, 90 и 120 пользователей.

В следующей таблице указано среднее время выполнения запроса в миллисекундах для обеих баз данных. Обе базы данных продемонстрировали хорошие результаты, однако Cache была в среднем в 4-6 раз быстрее. Помимо этого, кривая зависимости времени выполнения запроса от количества пользователей наглядно демонстрирует превосходство Cache по масштабируемости. Например, увеличение количества пользователей с 90 до 120 замедляет отклик Cache всего на 84 мс, тогда как отклик реляционной СУБД замедляется на 231 мс, что в 2,7 раза больше.

Количество пользователей	Среднее время отклика (по всем восьми запросам)		Выигрыш в скорости Cache : РСУБД
	РСУБД	Cache	
30	375,125 мс	59,125 мс	6.3 : 1
60	637,25 мс	137,75 мс	4.6 : 1
90	915,625 мс	206,875 мс	4.2 : 1
120	1146,375 мс	290,125 мс	3.9 : 1



Пример №3

Независимый консультант сравнил производительность Caché и распространенной реляционной базы данных в рамках сценария по организации хранилища данных. Скорость проверки и загрузки данных вместе со временем отклика замерялись на таблице с миллионом реальных записей, хранящих информацию об обработке заказов.

Хотя структуры данных Caché не оптимизировались для достижения максимальной производительности, Caché выполнила проверку и загрузку 1 000 000 записей в три раза быстрее (за 2681 с), чем реляционная СУБД (за 8597 с). Более того, итоговая база данных Caché заняла 146 Мб дисковой памяти, тогда как база данных реляционной СУБД — 216 Мб.

Скорость выполнения запросов SQL приведена в следующей таблице:

Запрос	Время отклика		Выигрыш в скорости Caché : PCУБД
	PCУБД	Caché	
Вычисление агрегата целой таблицы со всеми измерениями	638 с	180 с	3.5 : 1
Вычисление агрегата по всем измерениям с ограничением по 2 измерениям	3 с	< 1 с	>3 : 1
Поиск неиспользуемых «кодов причины» (поле заказа) по всей таблице	8 с	3 с	4 : 1
Поиск записей, в которых одно из измерений превышает заданное значение	50 с	16 с	3.1 : 1

Пример №4

По результатам теста на реальном электронном телефонном справочнике крупная телефонная компания пришла к выводу, что Caché в среднем работает в 10 раз быстрее, чем распространенная реляционная СУБД.

На загрузку данных из текстовых файлов в реляционную базу данных потребовалось около 10 часов, а получившаяся база данных заняла 60 Гб дисковой памяти. На загрузку данных в Caché потребовалось 45 минут, а сама база данных заняла 2 Гб. Поскольку база данных Caché заняла существенно меньше места, при работе с ней наблюдался максимальный эффект от кэширования данных. Как следствие, запросы в Caché исполнялись гораздо быстрее, чем в реляционной СУБД:

Запрос	Время отклика		Выигрыш в скорости Caché : PCУБД
	PCУБД	Caché	
Поиск по имени и улице	0,33 с	0,07 с	4 : 1
Поиск по имени	0,46 с	0,02 с	23 : 1
Поиск по имени, улице и почтовому индексу	0,76 с	0,04 с	19 : 1
Поиск по номеру дома	0,14 с	0,02 с	7 : 1
Поиск по имени, фамилии, улице и почтовому индексу	0,01 с	0,01 с	1 : 1
Поиск по административно-территориальной единице	0,02 с	0,01 с	2 : 1
Среднее значение по всем запросам	0,287 с	0,028 с	10 : 1

Для повышения скорости выполнения запросов в обеих базах данных использовались индексы, и, несмотря на то, что реляционная СУБД работала медленнее, чем Caché, ей удалось выполнить все запросы менее чем за секунду. В то же время для обновления реляционной СУБД (включая необходимую операцию перестроения индексов) требуется прекращать ее работу во избежание возникновения серьезных проблем с производительностью. Фактически, приложение рассчитано на работу с двумя базами данных, одна из которых находится в рабочем режиме, а вторая в это время обновляется. При этом базы сменяют друг друга раз в день. В противоположность этому, базу данных Caché можно обновлять одновременно с обслуживанием клиентов, т.к. обновление практически не влияет на производительность.

Пример №5

Компания-разработчик решений в сфере телекоммуникаций получила заявку на разработку системы контроля передачи коротких сообщений (SMS). Данная система должна была получать

большие объемы данных из различных источников, индексировать полученные данные и записывать их в централизованную базу данных на жестком диске. Компания-разработчик проверила производительность Caché и распространенной реляционной СУБД на прототипе системы.

В ходе теста базы данных выполняли один и тот же программный код. Другой нагрузки на систему не было. Результаты приведены ниже:

	Среднее число обработанных записей в секунду		Выигрыш в скорости Caché : PCСУБД
	PCСУБД	Caché	
Загрузка, индексирование и сохранение данных	315	1700	5 : 1

По результатам этого тестирования первоначальный проект реализации системы на основе реляционной СУБД был отклонен. После оптимизации кода для многомерных структур данных Caché превосходство в скорости стало шестикратным. При обработке данных из нескольких источников и наличии фоновой нагрузки на систему оптимизированное приложение для Caché в среднем обрабатывало 10 600 записей в секунду.

Заключение

Лучший индикатор эффективности СУБД — это тест на реальных приложениях. В данном отчете приведены результаты сравнения Caché с широко распространенными реляционными базами данных. Тесты проводились независимыми экспертами, а также клиентами и потенциальными клиентами InterSystems на реальных приложениях. Результаты тестов демонстрируют, что в среднем Caché выполняет запросы SQL в 5 раз быстрее, чем реляционные СУБД. Оптимизация кода приложений для максимально эффективной работы с многомерными структурами данных Caché позволяет добиться еще большего выигрыша в производительности.

Другие отчеты

Другие отчеты о сравнении Caché с реляционными базами данных приведены в следующих публикациях:

- § [“A Summary of Application/Database Performance”, KLAS Enterprises](#)
- § [Отчет IDC: “Breaking the Relational Barrier: User Data Management Triumphs With InterSystems’ Caché”, автор — Карл Олофсон](#)

Эти отчеты опубликованы на сайте www.InterSystems.ru; также их можно запросить по почте info@InterSystems.ru.

Проведите собственный тест

InterSystems часто помогает клиентам и заинтересованным компаниям в проведении тестов, подобных упомянутым в данном обзоре. Если вы заинтересованы в тестировании Caché на предмет производительности или масштабируемости, InterSystems поможет Вам в проведении теста и бесплатно предоставит временные лицензии на Caché. За дополнительной информацией обращайтесь по телефону +7 (095) 967 00 88.