

## ИССЛЕДОВАНИЕ МЕХАНИЗМОВ ОПТИМИЗАЦИИ ХРАНЕНИЯ И ОБРАБОТКИ ИНФОРМАЦИИ ПРИ ПОСТРОЕНИИ МНОГОМЕРНЫХ БАЗ ДАННЫХ В MS SQL SERVER

Цель работы: исследование аппарата аналитических операций и средств оптимизации скорости расчётов в рамках системы обработки данных MS Analysis Services 2005, а также выработка стратегии оптимизации быстродействия при миграции данных с предыдущей версии программного продукта - MS Analysis Services 2000.

Скорость расчета кубов данных в аналитических системах больших предприятий является важной характеристикой систем в целом и определяет область задач, разрешаемых на базе этих систем. Большие объемы анализируемых данных и сложные аналитические алгоритмы в условиях временных алгоритмических ограничений и ресурсных ограничений серверов предприятий требуют оптимального использования аналитических средств и выработки стратегий обработки и хранения данных.

Предварительный поиск обозначил три основных направления исследования:

- Перенос данных на платформу MS SQL Server 2005.
- Оптимизация структур хранения данных.
- Оптимизация процессинга данных.

Процесс переноса данных (или миграция) со старой версии на новую не вызывает затруднений и производится восстановлением базы в MS SQL Server 2005 из резервной копии, сделанной в MS SQL Server 2000. Это касается как реляционной базы данных, так и её многомерного представления для аналитической обработки в реальном времени (OLAP).

Соединяя подходы традиционного OLAP анализа и реляционной отчётности, Analysis Services 2005 предоставляют модель метаданных, покрывающую обе потребности. Набор кубов и измерений, определённый в Analysis Services 2005, называется Единообразной Пространственной Моделью (UDM). Усовершенствованная по сравнению с 2000 версией продукта архитектура позволяет достичь сокращения времени расчета многомерных кубов сразу же после миграции данных. Так на тестовом примере при прямой миграции данных наблюдаемый прирост производительности расчета кубов составил 24%.

Однако при проведении миграции следует учитывать, что служба отчётности была разработана заново и сводные таблицы и отчёты, построенные в MS Analysis Services 2000, не могут быть использованы в новой версии.

Исследования в области оптимизации и ускорения расчётов при использовании многомерных кубов при переходе на новую версию системы обработки данных показали, что наиболее эффективными механизмами для повышения производительности при работе с данными являются:

- индексирование (организация и расположение данных с помощью определённых механизмов, ускоряющих поиск и выборку нужной информации из всей базы);
- секционирование (разбиение исходных данных или их представлений на части).

В рамках индексирования стоит выделить аппарат индексированных представлений. Так как в MS Analysis Services 2000 многомерные кубы строились только на основе представлений, это нововведение позволяет ускорить производительность системы с наименьшими затратами по её доработке.

При секционировании наибольший интерес представляет возможность разместить отдельные секции таблиц исходных данных на различных физических носителях, что уменьшает время обращения к ним в случае наличия большого количества фактов.

В MS SQL Server 2005 существует ряд методов, позволяющих оптимизировать работу с многомерными данными и уменьшить нагрузку на сервер:

- Секционирование многомерных структур данных.
- Предварительный расчёт агрегатов.
- Инкрементный процессинг.

Секционирование куба по сути аналогично секционированию таблиц фактов. Предварительный расчёт агрегатов позволяет для измерений, имеющих иерархическую структуру (например, дата: день>месяц>год), на этапе процессинга куба вычислить промежуточные значения фактов для иерархий. Это позволяет ускорить обращение к данным, однако требует дополнительного времени на этапе процессинга. OLAP-запросы обрабатывались максимально быстро, когда все агрегаты заранее посчитаны. Однако добавление в хранилище агрегатов ведет к быстрому увеличению занимаемого им места на диске. Если вычислять все агрегаты, то, во-первых, это может занять продолжительное время, во-вторых, размер хранилища будет расти как степень числа измерений. Если агрегатов не вычислять, то мы не получаем никакого выигрыша в быстродействии, и все преимущество OLAP сведется только к тому, что администратору не нужно писать сложных и громоздких SQL-запросов.

Инкрементный процессинг – это эффективный способ частичной модификации многомерной структуры данных. При его выполнении не происходит пересчёта всей конструкции, изменения затрагивают только указанные администратором аспекты. Источником данных для обновления может служить как таблица или представление, так и запрос, выбирающий нужные строки из таблицы фактов. Инкрементный процессинг может быть использован для внесения новых данных в куб или для актуализации содержащейся в нём информации. При этом не важен тип данных измерения, для которого добавляются новые факты.