

7 причин для реализации хранилища данных

В этой короткой заметке приводится список основных причин, из-за которых компании внедряют хранилища данных. Поводом, побудившим собрать их в одном месте, стало то, что слишком много материалов о хранилищах данных посвящены преимуществам, которые связаны с хранилищами косвенно и являются прямым следствием решения базовых задач. Потратьте немного времени на знакомство с маркетинговыми материалами и вы узнаете об использовании хранилищ данных для «того, чтобы быть ближе к заказчику», «превращения данных в интеллект бизнеса», «принятия решений, основанных на фактах, а не на интуиции», и, вероятно, самой часто упоминающейся цели, «получении преимущества над конкурентами». На самом деле, в 99% проектов собственно хранилище данных является только одним из шагов на пути к достижению декларируемых целей.

На самом деле, основными причинами, побуждающими организации внедрять хранилища данных, являются:

Необходимость выполнения аналитических запросов и генерации отчетов на не задействованных основными информационными системами вычислительных ресурсах.

Большинство компаний хотят, чтобы обработка заданий регистрирующих систем была настроена таким образом, чтобы с высокой вероятностью транзакции завершались в приемлемое время. Отчеты и запросы могут потребовать гораздо больше ресурсов процессоров/дисковой памяти, поэтому их выполнение на системах, занятых обработкой транзакций существенно снижает вероятность своевременного завершения последних. Это в свою очередь грозит проблемами с проведением бизнес-операций.

Иными словами, выполнение аналитических запросов и генерация отчетов на серверах, занятых регистрирующими системами, делает задачу обработки транзакций в приемлемое время достаточно сложной. Компании приходят к выводу, что наименее затратный и/или организационно наиболее простой и быстрый способ обеспечения высокой скорости работы основных систем заключается во внедрении хранилища данных на выделенном сервере с собственной дисковой памятью.

Необходимость использования моделей данных и технологий, ускоряющих процесс выполнения запросов и подготовки отчетности, но не предназначенных для обработки транзакций

Необходимость использования моделей данных и технологий, ускоряющих процесс выполнения запросов и подготовки отчетности, но не предназначенных для обработки транзакций

Существует несколько способов проектирования структуры данных, которые обычно ускоряют выполнение запросов (схема «звезда» и производные от нее), но с другой стороны, не пригодны для регистрирующих систем, т.к. снижают скорость обработки транзакций. Существует также ряд технологий, ускоряющих выполнение запросов и не подходящих для процессинга транзакций (например, битовые индексы) и наоборот, применимых только в регистрирующих системах (восстановление транзакций). Обратите внимание что способы и технологии, помогающие или препятствующие аналитической/транзакционной обработке, зависят от производителей ПО и конкретной ситуации.

Создание среды, в которой даже относительно небольших знаний основ СУБД достаточно для создания запросов и подготовки отчетов. Это означает сокращение времени, требуемого от персонала ИТ-департамента для сопровождения системы.

Как правило, хранилище данных (а точнее схема «звезда» и ее производные) упрощает сами запросы, следовательно, требуется меньше знаний от сотрудников, работающих с системой. Несмотря на то, что конечные пользователи все равно сталкиваются с проблемами при подготовке отчетов и требуют помощи от специалистов ИТ-департамента, последним гораздо проще и быстрее подготовить необходимый отчет на основе данных хранилища, нежели БД регистрирующей системы. Следует заметить, что большая роль в повышении продуктивности работы ИТ-персонала происходит от сокращения бюрократических проволочек, возникающих при взаимодействии конечных пользователей с ИТ-департаментом.

Создание источника с предварительно очищенной информацией

Хранилище данных предоставляет возможность повышения качества информации без изменения данных в регистрирующих системах. Очистка данных осуществляется на стадии загрузки хранилища. Более того, стоит заметить, что некоторые реализации хранилищ подразумевают возможность корректировки данных в первоисточниках на основе исправлений, осуществленных на стадии загрузки информации в хранилище.

Упрощение процесса подготовки отчетов на основе информации из нескольких транзакционных систем и/или внешних источников данных и/или данных, используемых исключительно для генерации отчетов.

Компаниям, которым требуется готовить отчеты по нескольким источникам данных (а так оно как правило и бывает), необходимо сохранять выгрузки, запускать операции объединения выгрузок, пересортировывать данные и только после этого строить отчет по полученной выборке. В ряде случаев это совершенно адекватная стратегия. Однако если компания имеет большие объемы информации, которые требуется часто объединять, если данные, полученные из нескольких транзакционных систем необходимы для генерации отчетов, и, наконец, если данные необходимо «чистить», хранилище данных будет наиболее верным решением задачи.

Создание выделенного источника в тех случаях, когда возможности транзакционной системы не соответствует требуемому бизнесом сроку хранения данных и/или необходимо иметь возможность подготовки отчетов на определенные моменты времени в прошлом («as was» reporting).

В целях ускорения отклика на запросы данные после определенного периода времени часто удаляются из регистрирующих систем. Для целей обеспечения выполнения запросов и подготовки отчетности исторические и текущие данные могут храниться в выделенном хранилище, что обеспечит необходимую производительность, как аналитической, так и транзакционной систем.

Подготовка отчетов на определенный момент времени (“as was” reporting) в ряде случаев крайне затруднительна или даже невозможна. Например, если вам требуется получить отчет о зарплатах

сотрудников с уровнем образования «4» по некоторой корпоративной шкале за каждый месяц 1997 года, вы не сможете этого сделать, т.к. в справочнике системы хранится только текущий уровень образования. Для решения подобных задач компании необходимо создавать хранилище данных, с помощью которого проблема наличия медленно изменяющихся размерностей (slowly changing dimension) может быть закрыта.

Защита конечных пользователей от необходимости в какой бы то ни было степени вникать в структуру и логику работы БД регистрирующей системы

Использование хранилища данных и системы бизнес-анализа позволяет скрыть от конечных пользователей весь механизм обработки и хранения данных. Это является толчком к более плотной аналитической работе с информацией со стороны менеджмента и аналитиков, что означает повышение степени эффективности использования корпоративных информационных активов.

А как же бизнес-цели?

Некоторые компании создают хранилища для решения только одной из вышеназванных проблем, другие сталкиваются с полным перечнем. Но ни в коем случае нельзя говорить, что проект по созданию хранилища данных решает только сугубо технические проблемы и не преследует бизнес-цели.

Если вы посмотрите на список, то обратите внимание, что в основном потребность в хранилище происходит от ограничений, накладываемых транзакционными системами. Ограничения в определенных условиях использования могут и не проявиться. Но очевидно одно: хранилище данных, в той или иной степени, необходимо каждой компании и его внедрение – вопрос времени.

Возвращаясь к тому, с чего начинали. Компания, стремящаяся к лучшей поддержке принятия решений, получению преимуществ перед конкурентами, желающая стать ближе к заказчику, и для этого решившая «склепать» хранилище данных может быть очень удивлена результатом (отрицательным). Для достижения этих целей требуется, чтобы компания поняла, обычно методом проб и ошибок, как изменить бизнес-практики для наиболее эффективного использования информационных активов и хранилища данных. И это может оказаться более сложной задачей, чем задача по созданию хранилища.