

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ХРАНИЛИЩА ДАННЫХ И БАЗЫ ДАННЫХ

Иванов А.А.

*Иванов Андрей Александрович – студент,
направление: прикладная информатика,
Институт магистратуры*

Санкт-Петербургский государственный экономический университет, г. Санкт-Петербург

Аннотация: технология хранилищ данных включает в себя набор концепций и методов, которые предоставляют пользователям полезную информацию для принятия решений. Необходимость создания хранилища данных возникает из-за потребности повышения качества информации в организации. Дата, исходящая из разных источников, имеющих различные формы - как структурированные, так и неструктурированные, отфильтровывается в соответствии с бизнес-правилами и интегрируется в единый большой массив данных. Используя решения информационных технологий, менеджеры поняли, что данные, хранящиеся в операционных системах, включая базы данных, являются информационным золотым прииском, который должен быть использован. Хранилища данных были разработаны для удовлетворения растущих потребностей в комплексном анализе, которые не могут быть надлежащим образом достигнуты с помощью оперативных баз данных. Данная статья подчеркивает некоторые критерии, которые разработчики информационных приложений могут использовать для выбора между решением базы данных или хранилищем данных.

Ключевые слова: хранилище данных, базы данных, системы управления базами данных, информационные системы, бизнес-аналитика.

1. Потребность в базах данных и хранилищах данных.

Возможность для пользователей получать эффективный доступ к данным с помощью аналитических запросов имеет большое значение для конкурентных преимуществ компаний. Не менее большое значение имеет также передача и совместное использование данных внутри организации, между департаментами и различными подразделениями, а также между деловыми партнерами. Решения становятся все более многочисленными благодаря множеству систем, которые могут быть интегрированы с системами поддержки принятия решений: базами данных, хранилищами данных, витринами данных, решениями бизнес-аналитики, приложениями на уровне предприятия.

Менеджерами, которые преуспеют, станут те, кто будет внедрять системы поддержки принятия решений, приложения для мониторинга эффективности бизнеса, исполнительные информационные системы или решения бизнес-аналитики [3, с. 80]. Благодаря этим технологиям компании узнают, что произошло с их бизнесом, почему это произошло и что может произойти; все это, а также опыт и интуиция пользователей создают конкурентные преимущества.

Организации должны хранить и обрабатывать все больше данных, которые становятся все более разнообразными. Необходимость использовать эти данные в качестве ресурса для организации, в качестве поддержки принятия решений, привела к постоянному совершенствованию информационных систем. Чем лучше данные компании организованы, тем лучше результаты компании. Организация больших объемов данных превратилась из файлов в базу данных, а затем в хранилища данных. Предпосылка хранения и обработки больших объемов данных привела к созданию аналитических систем на основе хранилищ данных. Цель такой системы - предоставить аналитикам комплексный и последовательный взгляд на все данные, релевантные для компании. На основе системы данных, собранной и консолидированной в таких хранилищах данных, может быть выведен всесторонний анализ эффективности компании, могут быть определены различные корреляции данных, а также тенденции, которые прогнозируют будущие разработки, а также решения для улучшения бизнеса. Возрастающая потребность предвидеть изменения в рыночных условиях и выборе клиентов требует разработки интеллектуальных бизнес-планов, которые предполагают доступ к необходимой информации. Большая часть этой информации может быть найдена в транзакционных системах, включая реляционные базы данных. Способность преобразовывать данные в информацию, информацию в знания и знания в действия является обязательным условием, чтобы компании могли быть конкурентоспособными в постоянно меняющейся экономической среде. Решение всех подобных проблем - это создание хранилища данных.

Транзакционные базы данных обеспечивают ответ на операционные требования, в то время как хранилища данных обеспечивают ответ на требования к анализу, тем самым предоставляя возможность для высококачественного анализа и сложных специальных запросов через удобные для пользователя интерфейсы. Основным критерием организации данных в хранилищах данных является предмет (область деятельности), а основным критерием для баз данных является приложение. Концепция хранилища

данных представляет собой логический архитектурный подход к извлечению оперативных данных и преобразованию их в точную историческую информацию для поддержки процесса принятия решений.

2. Сравнительный анализ особенностей хранилищ данных и баз данных.

Особенности двух режимов организации данных во внешней памяти, хранилищах данных и базах данных можно увидеть из приведенных ниже определений. База данных представляет собой ориентированный на приложения набор данных, который организован, структурирован, согласован, с минимальной и контролируемой избыточностью, к которым в свое время могут обращаться несколько пользователей.

Хранилище данных - это предметно-ориентированный набор данных, который является интегрированным, зависящим от времени, энергонезависимым, который может использоваться для поддержки процесса принятия решений [4, с. 10].

Таким образом, мы можем заключить основные характеристики хранилищ данных и баз данных, которые будут описаны ниже.

Тематическая ориентация

Организация данных в хранилищах данных основана на областях, представляющих интерес, по основным предметам организации: клиентам, продуктам, видам деятельности и т. д. Базы данных организуют данные на основе корпоративных приложений, полученных в результате выполнения своих функций. Главная цель хранилища данных - поддерживать систему принятия решений, ориентируясь на субъекты организации. Все элементы данных, относящиеся к одному и тому же предмету или событию в реальном мире, связаны и данные ориентации для обработки - очевидны в содержимом базы данных. Хранилище данных включает только информацию, которая должна использоваться при обработке информации и анализа, тогда как операционная база данных включает в себя подробные данные, необходимые для целей обработки, но не имеющие отношения к управлению или анализу. Субъектная ориентация хранилища данных позволяет разрабатывать процесс принятия решений посредством поэтапного процесса, который объединяет разные субъекты в единую структуру. Например, когда клиент включен в несколько операционных баз данных, где он по-разному определяется, клиент определяется только один раз в хранилище данных и просматривается всеми пользователями в одном и том же пути.

Интеграция

Операционные базы данных разрабатываются в разное время различными командами по-разному. Таким образом, с функциональной точки зрения база данных не может использоваться для целей анализа и отчетности. Хранилище данных - это корпоративный проект. Он включает в себя данные из всех или большинства оперативных баз данных организации, которые хранятся в соответствии с тем, чтобы аналитики могли сосредоточиться на использовании данных, а не на его надежности и согласованности. Согласованность данных очень важна для баз данных и обеспечивается задачей системы управления базами данных - СУБД в отношении целостности данных. Согласованность также применяется к хранилищам данных в отношении: имен полей, систем кода, представлений дат, переменных измерений, физических атрибутов и т. д., Так что отчеты, созданные для различных отделов или разного времени, должны включать одинаковые результаты.

Временной фактор

Значение оперативных данных в базах данных периодически обновляется и показывает текущий статус. С другой стороны, для информационных потребностей экономического анализа, основанного на хранилищах данных, исторические данные имеют существенное значение, поскольку в нем показаны тенденции точного прогнозирования. Регулярная загрузка данных из оперативных баз данных делает данные в хранилищах данных временным вариантом. Данные в хранилищах данных точно отображают статус в разные моменты, тем самым обеспечивая исторический взгляд на дату. Это делает хранилища данных отличными от операционных баз данных, где данные должны показывать статус во время доступа. В базах данных данные обновляются с каждой новой транзакцией, а прежние значения обычно теряются. В оперативных базах данных редко хранятся исторические данные, и это происходит лишь в течение коротких периодов времени, поскольку их целью является сохранение текущих данных. В отличие от этих систем хранилища данных не обновляются, но данные периодически загружаются для отображения истории данных. Это позволяет выявлять тенденции, а также сравнивать между различными периодами времени. Временной горизонт хранилищ данных значительно больше по сравнению с базами операционных данных, предоставляя информацию с исторической точки зрения (5-10 лет) [4, с. 117]. Поэтому любая структура хранилища данных включает в себя явно или неявно элемент времени для идентификации определенной функции в определенное время, что не является обязательным для баз данных.

Отсутствие изменчивости

Данные в хранилищах данных являются статическими, а не динамическими, как в случае с операционными системами. Поскольку хранилища данных показывают оперативные данные в

определенное время, данные не будут обновляться после загрузки в хранилищах данных. В результате идентичный запрос, сделанный после одного года на основе одних и тех же ссылочных данных, даст тот же результат. В операционных базах данных информация нестабильна, так как запросы сосредотачиваются на текущих данных. Данные обновляются на постоянной основе, как правило, на основе транзакций. Любая обрабатываемая транзакция включает в себя обновление: добавление новых записей, изменение или удаление существующих других.

3. Различия между хранилищами данных и операционными базами данных

Операционные базы данных и хранилища данных в основном базируются на одной и той же технологической поддержке: это сбор данных, обе функции основаны на ключах, индексах и представлениях, причем оба основаны на модели данных. Тем не менее, две системы отличаются друг от друга, как показывают критерии, описанные ниже.

1) С функциональной точки зрения: операционные базы данных обрабатывают транзакции, обеспечивая ответы на операционные требования, в то время как хранилища данных используются на основе ad hoc-запросов, в основном для целей управления.

2) Функциональные требования различны: операционные базы данных в основном сосредоточены на защите и согласованности данных, что делает запросы медленными, специальными. Эти запросы, специфичные для экономического анализа, могут значительно снизить производительность операционной системы из-за отсутствия прогнозируемых индексов, как это имеет место в хранилищах данных.

3) Хотя большинство операционных систем и хранилищ данных построены на реляционных технологиях, их дизайн существенно отличается, поскольку их назначение также различно. Операционные базы данных предназначены для онлайн обработки транзакций, и их главная цель связана с эффективным хранением большого количества транзакционных данных. Они включают текущую информацию о повседневных действиях и информацию о процессах, подлежащую обновлению. В результате данные являются динамическими и, следовательно, очень изменчивыми. Задачи таких систем являются структурированными и повторяющимися и состоят из текущих, коротких и изолированных транзакций, которые включают подробные данные. Эти транзакции считывают или обновляют несколько записей - в основном десятки, главным образом, на основе их первичных ключей. Операционные базы данных достигают размеров от сотен мегабайт до гигабайт. Их согласованность важна и относится к быстрой обработке транзакций.

4) Стратегии резервного копирования и восстановления отличаются для двух типов систем. Большинство данных в хранилищах данных - это исторические данные, которые не являются вариантами и не требуют многократного сохранения. Новые данные могут быть сохранены во время загрузки. В некоторых случаях рекомендуется сохранять данные из промежуточных баз данных, чтобы минимизировать влияние на производительность хранилищ данных. Политики восстановления также могут отличаться в случае хранилищ данных, в отличие от оперативных баз данных, в зависимости от того, насколько для организации необходим постоянный непрерывный доступ к хранилищам данных. В реальной задаче резервного копирования и восстановления баз данных для СУБД. В фактическом хранилище данных эта задача предназначена для администратора базы данных.

5) Другое различие между этими двумя типами систем связано с механизмами, требуемыми для одновременного доступа пользователей. Так как хранилища данных не обновляются, управление транзакциями, управление параллельным доступом и другие подобные механизмы, интегрированные в систему управления базами данных, используются только на начальной стадии загрузки и для последующего добавления, из-за того, что они дороги с точки зрения времени отклика. Эти механизмы могут быть отключены во время текущего использования хранилищ данных. Созданная таким образом свобода может быть использована для оптимизации доступа к данным путем: денормализации, суммирования, статистики доступа к данным, динамической реорганизации индекса и т. д.

4. Выводы

Различия, показанные выше, являются одной из причин, по которым хранилища данных создаются отдельно от оперативных баз данных. Разделение двух систем обеспечивает масштабируемость решений бизнес-аналитики, а также их способность быстро и эффективно отвечать на запросы в компании. Хранилища данных позволяют проводить всесторонний анализ, поскольку структуры коллекций данных более просты - только необходимая информация сохраняется, стандартизируется - структуры хорошо документированы и денормализованы, существует меньше объединений между наборами данных.

Список литературы

1. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных OLAP и Data Mining. СПб: БХВ-Петербург, 2004. 336 с.

2. *Гарсиа-Молина Гектор, Ульман Д. Джеффри, Уидом Дженифер.* Системы баз данных. Полный курс. М.–СПб.–Киев: Издательский дом «Вильямс», 2004. 1088 с.
3. *Паклин Н.Б., Орешков В.И.* Бизнес-аналитик: от данных к знаниям // Учебное пособие, СПб.: Питер, 2009. 642 с.
4. *Спирли Э.* Корпоративные хранилища данных: планирование, разработка, реализация. Т. 1. М.–СПб.–Киев: Издательский дом «Вильямс», 2001. 396 с.