

УДК 004

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ХРАНИЛИЩА И БАЗЫ ДАННЫХ

Нафикова А.Р.

Доцент, кандидат физико-математических наук

Стерлитамакский филиал БашГУ,

Стерлитамак, Россия

Ахметгалеева Д.Ф.

Студент,

Стерлитамакский филиал БашГУ,

Стерлитамак, Россия

Аннотация. Технология хранения данных включает в себя набор концепций и методов, которые предоставляют пользователям информацию, помогающую им принимать решения. Необходимость создания хранилища данных вытекает из необходимости повышения качества информации в организации. Даты из разных источников с разными форматами, как структурированными, так и неструктурированными, фильтруются в соответствии с бизнес-правилами и объединяются в один большой массив данных. Используя решения в области информационных технологий, администратор осознал, что данные, хранящиеся в операционной системе, включая базу данных, являются золотым прииском информации для использования. Хранилище данных проектируется для удовлетворения растущей потребности во всестороннем анализе, который не может быть полностью достигнут с производственной базой данных. В этой статье освещаются некоторые критерии, которые разработчики информационных приложений могут использовать для выбора между решениями для баз данных и хранилищами данных.

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

COMPARATIVE ANALYSIS OF DATA STORAGE AND DATABASE

Nafikova A.R.

Associate Professor, Candidate of Physical and Mathematical Sciences

Sterlitamak branch of Bashkir State University,

Sterlitamak, Russia

Akhmetgaleeva D.F.

Student,

Sterlitamak branch of Bashkir State University,

Sterlitamak, Russia

Annotation: Data warehousing technology includes a set of concepts and methods that provide users with information to help them make decisions. The need to create a data warehouse stems from the need to improve the quality of information in the organization. Data from different sources with different formats, both structured and unstructured, are filtered according to business rules and combined into a single large data array. Using information technology solutions, the administrator has realized that the data stored in the operating system, including the database, is a gold mine of information to be using. Data warehousing is designed to meet the growing need for comprehensive analysis that cannot be fully achieved with a production database. This article highlights some criteria that information application developers can use to choose between database solutions and data warehousing.

Keywords: data warehouse, databases, database management systems, information systems, business analytics.

Потребность в базах данных и хранилищах данных.

Организациям необходимо хранить и обрабатывать все более разнообразные данные. Чем более упорядочены данные компании, тем выше ее производительность. Из-за предпосылки хранения и обработки больших объемов данных была построена система анализа на основе хранилища данных. Цель такой системы – предоставить аналитикам всесторонний и последовательный взгляд на все данные, относящиеся к компании. Большая часть этой информации находится в транзакционных системах, включая реляционные базы данных. Способность превращать данные в информацию, информацию в знания и знания в поведение является необходимым условием конкурентоспособности компаний в постоянно меняющейся экономической среде [3, с. 80]. Решением всех этих задач является создание хранилища данных.

Сравнительный анализ функциональности хранилища данных и базы данных.

Базы данных – это ориентированные на приложения наборы данных, которые организованы, структурированы, согласованы и одновременно доступны нескольким пользователям с минимальной контролируемой избыточностью.

Хранилище данных – это интегрированный, зависящий от времени, энергонезависимый, специфичный для предметной области набор данных, который можно использовать для поддержки процесса принятия решений [4, с. 10].

Основными характеристиками хранилищ данных и баз данных являются:

Тематическая ориентация

Организация данных в хранилище данных основана на областях, представляющих интерес для клиентов, продуктов, видов деятельности и т.д., которые являются основным субъектом организации. База данных организует данные на основе корпоративных приложений, полученных в результате выполнения ее функций. Основная цель хранилищ данных – поддерживать системы принятия решений, сосредоточившись на организационных темах. Хранилище данных содержит только ту информацию, которая должна использоваться для обработки и анализа информации, а производственная база данных содержит подробные данные, необходимые для целей обработки, но это не относится к управлению или анализу.

Интеграция

Производственная база данных в разное время разрабатывается различными командами по-разному. Следовательно, с функциональной точки зрения, база данных не может быть использована для целей анализа и отчетности. Хранение данных включает все или большую часть данных в производственной базе данных организации, так что аналитики могут сосредоточиться на использовании данных. Целостность данных очень важна для базы данных и обеспечивается задачами системы управления базами данных. Согласованность применяется к хранилищам данных, которые относятся к именам полей, системам кодов, представлениям дат, переменным измерениям, физическим атрибутам и так далее. Поэтому отчеты, генерируемые отделом или часовым поясом, также должны содержать те же результаты.

Фактор времени

Значения оперативных данных в базе данных регулярно обновляются, чтобы указать текущее состояние. С другой стороны, историческая информация показывает тенденции для точного прогнозирования. Данные в

хранилище данных точно отображают ситуацию в разное время, обеспечивая тем самым историческое представление даты. В базе данных данные обновляются с каждой новой транзакцией, а старые значения обычно теряются.

Отсутствие изменчивости

Данные в хранилище являются статическими, а не динамическими, как в операционной системе. Хранилище данных отображает текущие данные в определенное время, поэтому данные не обновляются после загрузки в хранилище данных. В результате идентичные запросы, сделанные через год на основе одних и тех же справочных данных, дадут те же результаты. В производственной базе данных информация нестабильна, потому что запрос фокусируется на текущих данных. Данные постоянно обновляются, как правило, на основе транзакций.

Разница между хранилищем данных и операционной базой данных

Производственная база данных и хранилище данных в основном основаны на одной и той же технической поддержке. Это сбор данных, и обе функции основаны на ключах, индексах и представлениях, обе основаны на моделях данных. Критерии отличия систем друг от друга:

- 1) с функциональной точки зрения производственная база данных обрабатывает транзакции и предоставляет ответы на производственные требования. Хранилище данных, с другой стороны, в основном используется для целей управления, исходя из дополнительных требований;
- 2) разные функциональные требования;
- 3) большинство операционных систем и хранилищ данных построены на реляционных технологиях, но их назначение различно.

- 4) стратегии резервного копирования и восстановления различны для двух систем;
- 5) Другое различие между этими двумя типами систем связано с механизмом, необходимым для одновременного доступа пользователя.

Заключение.

Вышеуказанное различие является одной из причин, по которой хранилище данных создается отдельно от производственной базы данных. Разделение двух систем обеспечивает масштабируемость решения бизнес-аналитики и возможность быстрого и эффективного реагирования на внутренние запросы. Хранилище данных позволяет проводить всесторонний анализ, поскольку структура сбора данных более проста – только необходимая информация хранится и стандартизируется.

Библиографический список

1. Барсегян А.А. Методы и модели анализа данных OLAP и Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод – СПб: БХВ-Петербург, 2004. – 336 с.
2. Гарсиа-Молина Гектор. Системы баз данных / Гектор Гарсиа-Молина, Джеффри Д. Ульман, Дженнифер Уидом. – М. СПб. Киев: Издательский дом «Вильямс», 2004. – 1088 с.
3. Паклин Н.Б. Бизнес-аналитик: от данных к знаниям: учебное пособие / Н.Б. Паклин, В.И. Орешков. – СПб.: Питер, 2009. – 642 с.
4. Спирли Э. Корпоративные хранилища данных: планирование, разработка, реализация / Э. Спирли – Т.1. М. СПб. Киев: Издательский дом «Вильямс», 2001. – 396 с.

Оригинальность 80%