

Хранилища данных

- Причины возникновения
 - Концепция хранилищ данных
 - Применение
-

Причины возникновения технологии DW

- Потребовался совокупный анализ информации, который не могли осуществить БД
 - Для работы с информацией возникла необходимость хранить её в интегрированном виде
 - Принципы работы БД не подходили для этих целей
-

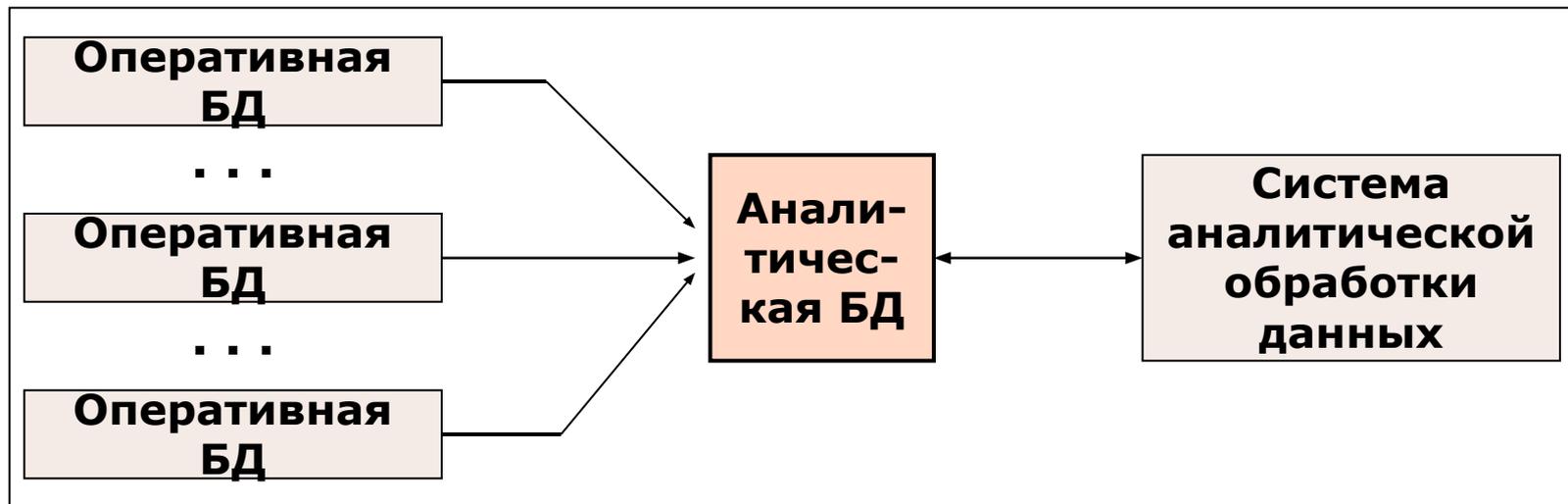
Сравнение баз данных и хранилищ данных

Характеристика	DB	DW
Модифицируемость данных	Есть	Нет
Ориентация системы	Узкая	Широкая
Упорядоченность данных	Не строгая	Строгая
Задачи системы	Оперативную работу, частые запросы	Аналитическую работу, несложные отчёты
Фильтрация данных	Не жёсткая	Жёсткая
Нормализация	Есть	Нет

Аналитическая информационная система

Для достижения этих целей была создана аналитическая информационная система. Её компоненты:

- Оперативные БД как источники информации
- Система аналитической обработки данных для их анализа
- И, самое главное, аналитическая БД для сбора, хранения и предоставления необходимой информации



Понятие хранилищ данных

Хранилище данных (Data Warehouse):
это предметно-ориентированный,
интегрированный, неизменяемый,
поддерживающий хронологию набор
данных, организованный для целей
поддержки управления.

Концепция хранилищ данных по С.Кузнецову

- Интеграция ранее разъединенных детализированных данных в едином DW, их согласование и агрегация
 - Разделение наборов данных для операционной обработки и наборов данных для решения задач анализа
-

Свойства хранилищ данных

- Предметная ориентация
 - Интегрированность данных
 - Инвариантность во времени
 - Стабильность информации
 - Минимизация избыточности информации
-

Модели хранилищ данных

- Схема «звезда»
 - Схема «снежинка»
 - ER-диаграмма
-

Схема «звезда»

Главная таблица – таблица фактов, с ней связаны таблицы размерностей.

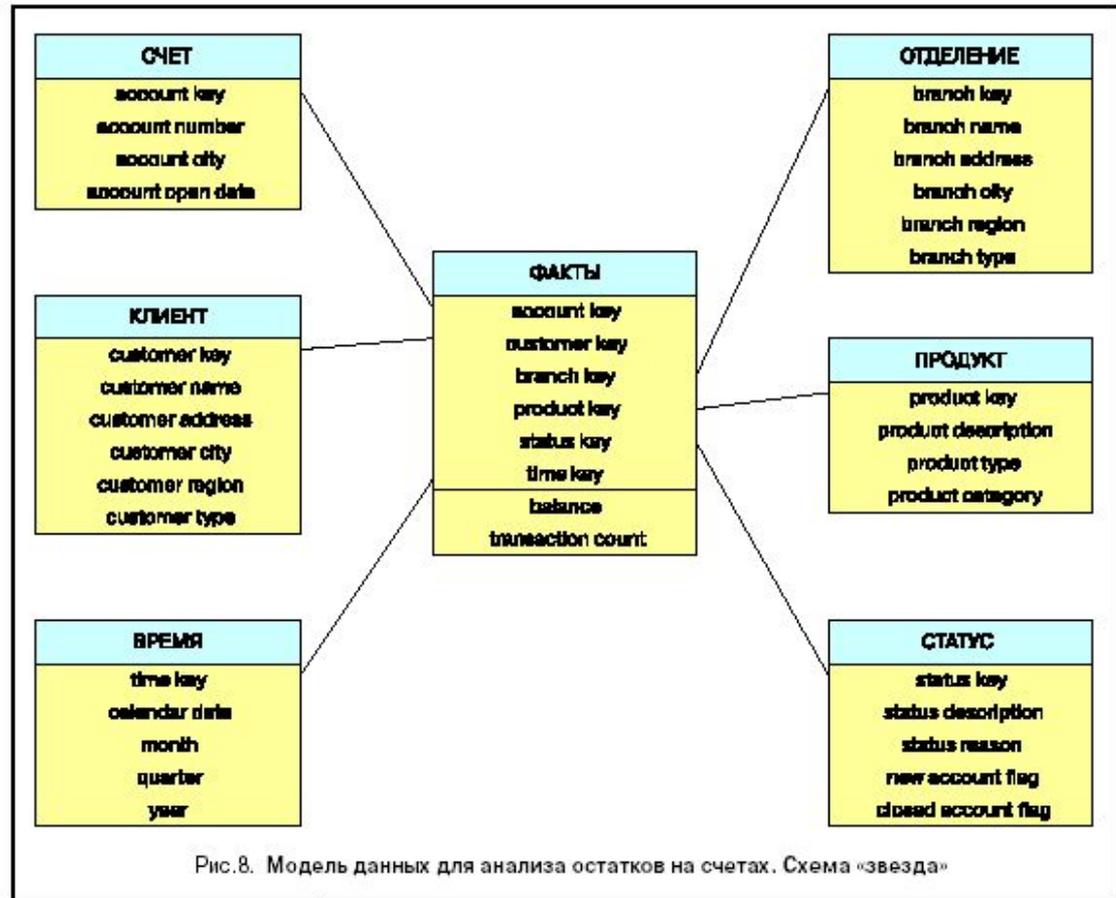
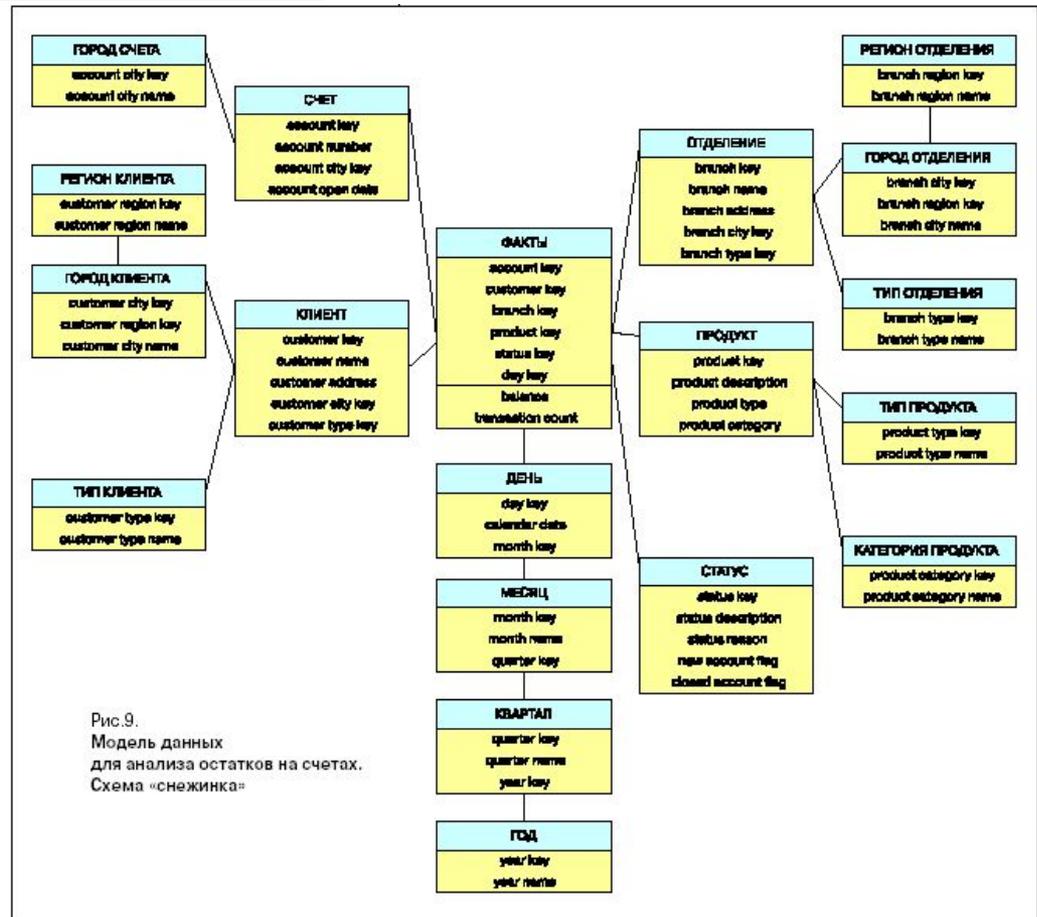


Схема «снежинка»

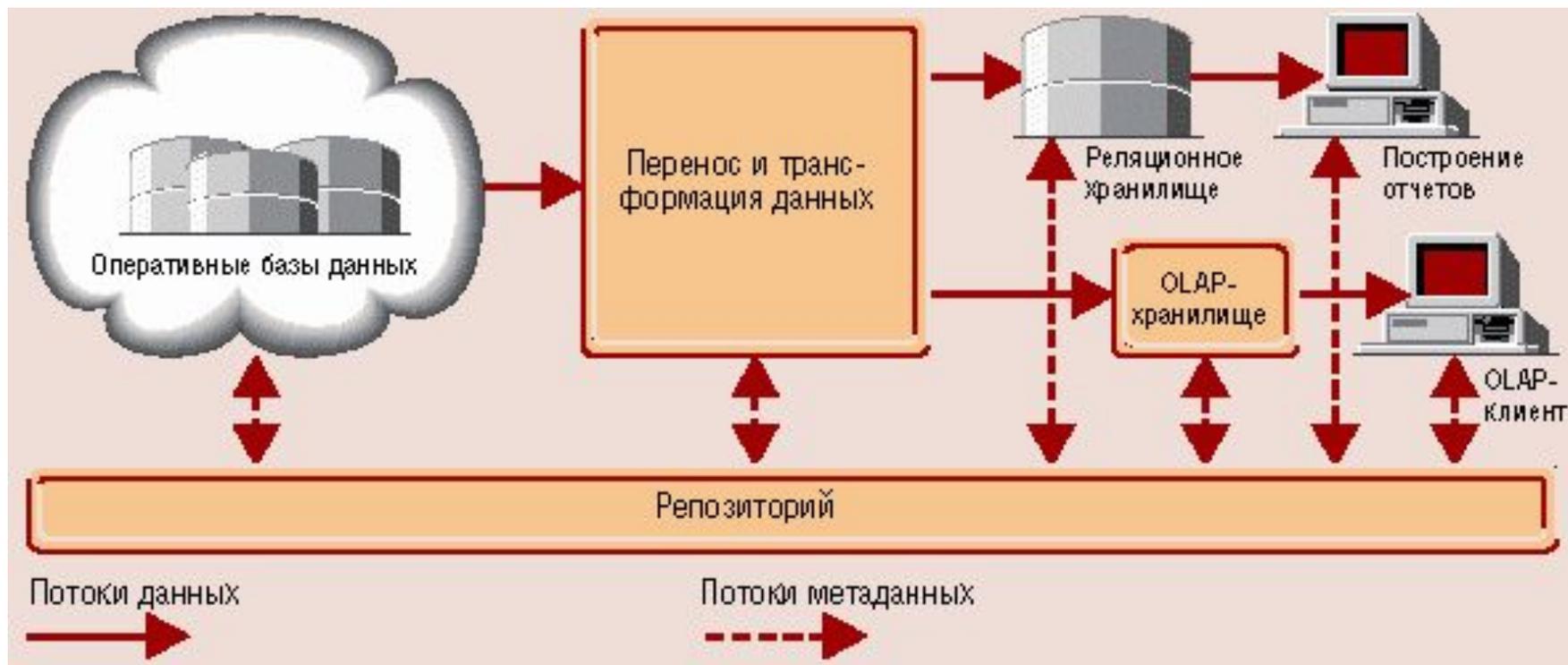
Таблицы размерностей нормализованы, что снижает избыточность информации и ускоряет выполнения некоторых запросов.



Основные компоненты хранилищ данных

- ❑ Оперативные источники данных
 - ❑ Средства переноса и трансформации данных
 - ❑ Метаданные
 - ❑ Реляционное хранилище
 - ❑ OLAP-хранилище и OLAP-клиенты
 - ❑ Средства доступа и анализа данных
-

Функционирование хранилищ данных



Пример использования OLAP-технологии

В компании, занимающейся продажей различных товаров, может быть получен и проанализирован с помощью OLAP следующий многомерный куб:

	США	Канада	Мексика
Напитки	10 000	2000	1 000
Продукты питания	5000	500	250
Прочие товары	5000	500	250

Применение хранилищ данных

- Финансовые DW
 - DW в области страхования
 - DW для управления людскими ресурсами
 - DW в области телекоммуникаций
 - и прочие
-

Инструментальные средства: решения в области хранилищ данных

- ❑ **IBM:** A Data Warehouse Plus
 - ❑ **Oracle:** Warehouse Technology Initiative
 - ❑ **Hewlett Packard:** OpenWarehouse
 - ❑ **NCR:** Enterprise Information Factory
 - ❑ **Informix Software:** On-Line Dinamic
Parallel Server
 - ❑ **SAS Institute:** ряд решений
 - ❑ **Sybase:** Warehouse WORKS
-

Примеры успешного внедрения хранилищ данных

Сеть супермаркетов Hyundai

Интеграция сведений о клиентах, продажах и спросе на товары для выявления тенденций, предпочтений и формирования стратегии

Корпорация NCR

Для повышения эффективности деятельности, управления компанией и для обеспечения общего представления об эффективности собственной работы

Компания Sekisui

DW упростило учёт финансовой деятельности, объединило базы данных управления и бухгалтерии, позволило анализировать текущее положение компании гораздо оперативнее, чем раньше.

Особенности, тенденции и проблемы: технологии хранилищ данных

□ Технологии хранилищ данных

- «Взрывание» объемов данных
- Активные DW
- Распространение источников данных

□ Новые требования к DW

- Интегрирование представления о клиенте
 - Потребность в более сложных запросах
 - Слияние с CRM
-

Выводы

Итак, хранилища данных предназначены для:

- хранения, загрузки и преобразования данных из разных источников
 - накопления истории изменения данных во времени
 - функционирования в качестве источника данных для аналитических систем
-