



Хранилища данных

База данных

- совокупность данных, организованных по определенным правилам, предусматривающим общие принципы описания, хранения и манипулирования данными.

ХРАНИЛИЩЕ ДАННЫХ (*DATA WAREHOUSE*)

- предметно - ориентированная информационная база данных, специально разработанная и предназначенная для подготовки отчётов и бизнес - анализа с целью поддержки принятия решений в организации.

ХРАНИЛИЩЕ ДАННЫХ

- Строится на базе систем управления базами данных Строится на базе систем управления базами данных и систем поддержки принятия решений Строится на базе систем управления базами данных и систем поддержки принятия решений. Данные, поступающие в хранилище данных, как правило, доступны только для чтения. Данные из OLTP Строится на базе систем управления базами данных и систем поддержки принятия решений. Данные, поступающие в хранилище данных, как правило, доступны только для чтения. Данные из OLTP-системы копируются в хранилище данных таким образом, чтобы построение отчётов и OLAP-анализ не использовал ресурсы транзакционной системы и не нарушал её стабильность. Как правило,

Расхождения в требованиях к хранению данных в БД и ХД

Традиционные данные, хранимые в БД	Данные для принятия решений
Детализированы	Обобщены либо очищены
Точны в момент доступа	Представляют значения на указанное время
Могут корректироваться	Не корректируются, если введены в Хранилище
Требования к способам дальнейшей обработки выясняются заранее	Требования к способам дальнейшей обработки не имеют первостепенного значения
Строятся на основе обычного цикла разработки систем	Совершенно иной цикл разработки систем
Чувствительны к производительности БД и поэтому предъявляют к ним жесткие требования	Мягкие требования к производительности БД

Продолжение таблицы

Обрабатывается один элемент данных за один запрос	Обрабатывается множество элементов данных за один запрос
Управляются транзакциями	Управляются аналитическими запросами
Ориентированы на приложения	Ориентированы на анализ
Высокая степень доступности	Относительная доступность
Контролируется целостность всех данных	Контролируется целостность подмножества данных
Данные не избыточны	Данные избыточны
Статическая структура, произвольное содержание	Гибкая структура
Массивы данных редко используются в процессе обработки	Массивы данных широко используются в процессе обработки
Поддерживают ежедневные операции	Поддерживают периодический анализ

OLTP (ONLINE TRANSACTION PROCESSING),

- **транзакционная система** — обработка транзакций в реальном времени. Способ организации БД — обработка транзакций в реальном времени. Способ организации БД, при котором система работает с небольшими по размерам транзакциями — обработка транзакций в реальном времени. Способ организации БД, при котором система работает с небольшими по размерам транзакциями, но идущими большим потоком, и при этом клиенту требуется от системы минимальное время отклика.
- Термин OLTP применяют также к системам (приложениям). OLTP-системы предназначены для ввода, структурированного хранения и обработки информации (операций, документов) в

OLAP (*ONLINE ANALYTICAL PROCESSING*)

- технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу.

Системы поддержки принятия решений



Аналитические приложения (BI, Business Intelligence)



Хранилище данных (DW, Data Warehouse)



Системы планирования ресурсов (ERP)

Корпоративные информационные системы (CIS)

Клиент-серверные приложения (тонкий клиент)

DATA MINING (ДОБЫЧА ДАННЫХ, ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ, ГЛУБИННЫЙ АНАЛИЗ ДАННЫХ)

- собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

новые

затраченные усилия на открытие знаний, которые уже известны пользователю, не окупаются, поэтому ценность представляют именно новые, ранее неизвестные знания

нетривиальные

результаты анализа должны отражать неочевидные неожиданные закономерности в данных (скрытые знания); результаты, которые могли бы быть получены более простыми способами (например, визуальным просмотром), не оправдывают привлечение мощных средств ИА

**практически
полезные**

найденные знания должны быть применимы в том числе и на новых данных, с достаточно высокой степенью достоверности, полезность заключается в том, чтобы эти знания могли принести определенную выгоду при их применении

**доступные для
понимания**

добытые знания должны быть логически объяснимы и представлены в понятном для пользователя виде

Методы теории вероятностей и
математической статистики

Регрессионный, дискриминантный
анализ

Метод дерева решений

Метод анализа иерархий

Система сбалансированных показателей

Карты восприятия методы выявления
логических закономерностей

Факторный, кластерный анализ

Методы теории нечетких множеств

Методы теории полезности

Нейронные сети

Денормализованные,

пространственные базы данных

- Одним из направлений развития РБД в интересах систем принятия решений является разработка таблиц с денормализованной формой (модификации схемы организации данных типа звезда).
- Структура такой базы данных не будет реляционной - это будет пространственная база данных с целью анализа данных, а не выполнения транзакций.

Как проектировать ненормализованную БД?

- Большинство Case – средств проектирования БД поддерживает методологию моделирования хранилищ благодаря использованию специальной нотации для физической модели – Dimensional.

Методология Dimensional

- Нормализация данных в реляционных СУБД приводит к созданию множества связанных между собой таблиц. В результате, выполнение сложных запросов неизбежно приводит к объединению многих таблиц, что существенно увеличивает время отклика.
- Создание хранилища данных подразумевает создание денормализованной структуры данных (допускается избыточность данных и возможность возникновения аномалий при манипулировании данными), ориентированной в первую очередь на высокую производительность при выполнении аналитических запросов.
- Нормализация делает модель хранилища слишком сложной, затрудняет ее понимание и ухудшает эффективность выполнения запроса.

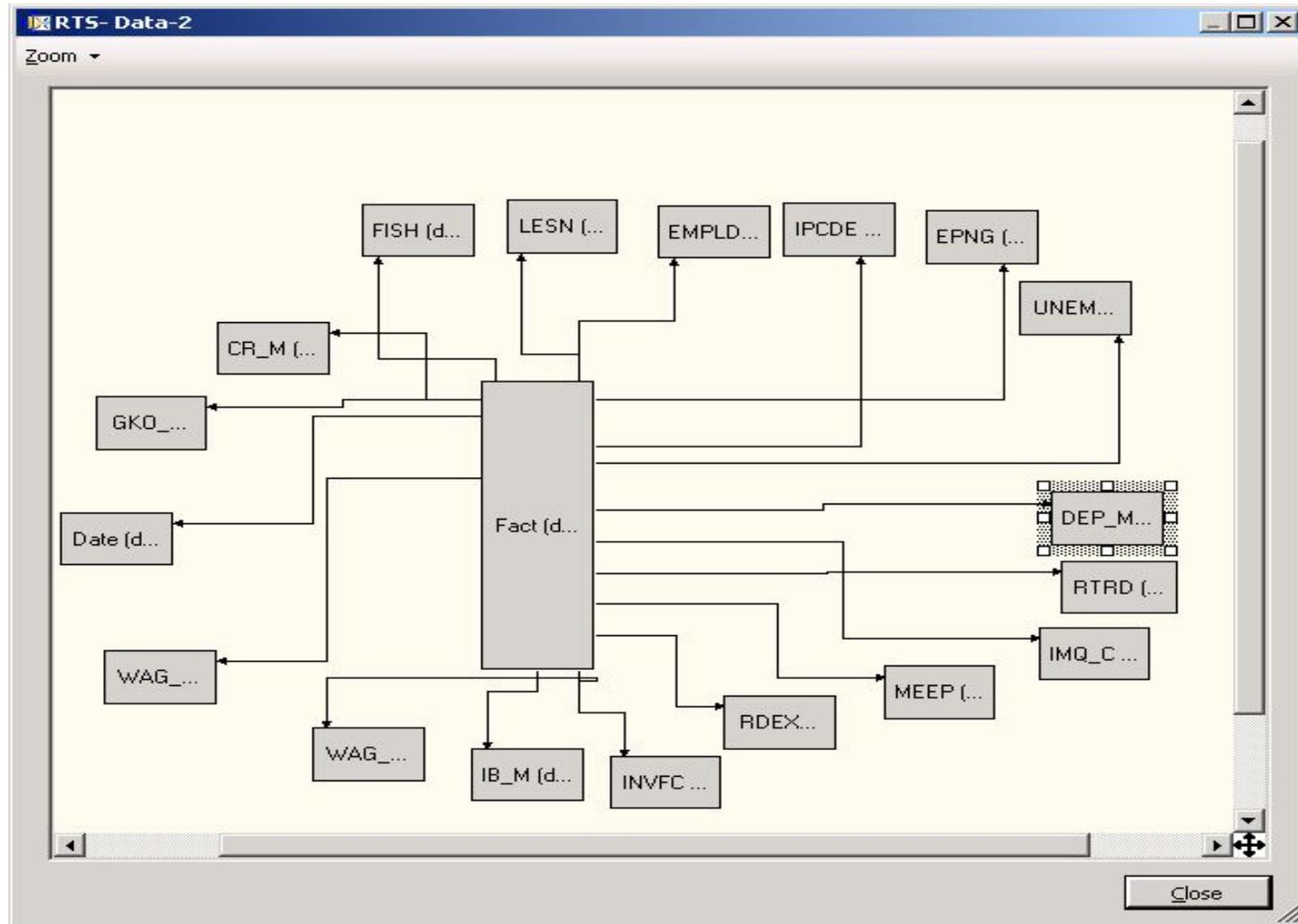
Особенности проектирования

- Моделирование Dimensional сходно с моделированием связей и сущностей для реляционной модели, но отличаются целями.
- Реляционная модель акцентируется на целостности и эффективности ввода данных.
- Размерная (Dimensional) модель ориентирована в первую очередь на выполнение сложных запросов к БД.

О схеме звезда

- В размерном моделировании принят стандарт модели, называемый **схемой звезда (star schema)**, которая обеспечивает высокую скорость выполнения запроса посредством денормализации и разделения данных.
- Невозможно создать универсальную денормализованную структуру данных, обеспечивающую высокую производительность при выполнении любого аналитического запроса. Поэтому схема звезда строится так, чтобы обеспечить наивысшую производительность при выполнении одного самого важного запроса, либо для группы похожих запросов.

Структура ХД - звезда



Основные составляющие структуры хранилищ данных

- Схема звезда обычно содержит одну большую таблицу, называемую таблицей факта (*fact table*), помещенную в центр, и окружающие ее меньшие таблицы, называемые таблицами размерности (*dimensional table*), соединенные с таблицей факта в виде звезды радиальными связями. В этих связях таблицы размерности являются родительскими, таблица факта - дочерней.
- Схема звезда может иметь также консольные таблицы (*outrigger table*), присоединенные к таблице размерности. Консольные таблицы являются родительскими, таблицы размерности - дочерними.

Таблица(ы) фактов

- Прежде чем создать DW со схемой типа звезда, необходимо проанализировать бизнес-правила предметной области с целью выяснения центрального вопроса, ответ на который наиболее важен. Все прочие вопросы должны быть объединены вокруг этого основного вопроса и моделирование должно начинаться с него. Данные, необходимые для ответа на этот вопрос, должны быть помещены в центральную таблицу модели - таблицу факта

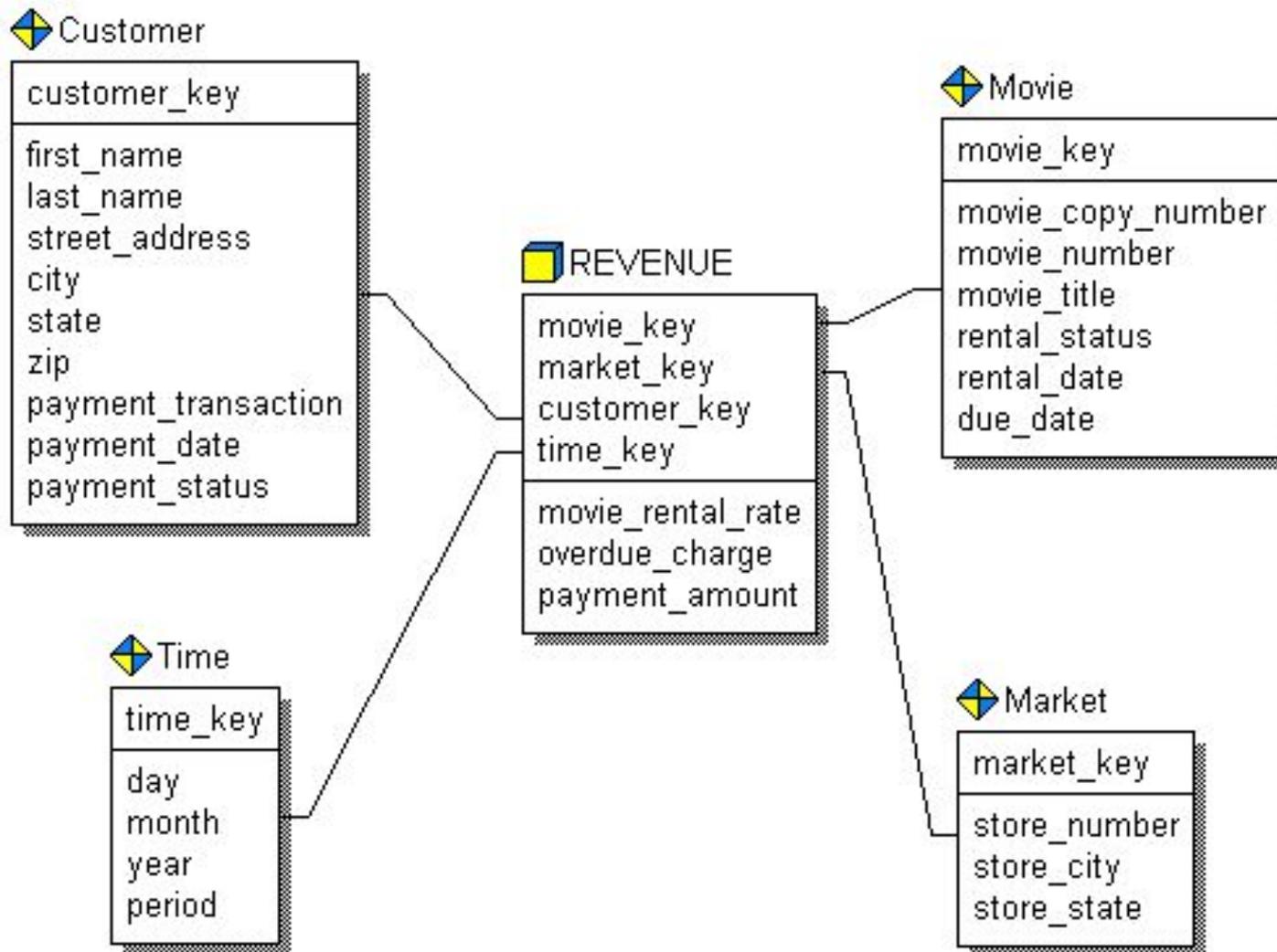
О связи таблицы фактов с таблицами измерений

- Таблица факта является центральной таблицей в схеме звезда. Она может состоять из миллионов строк и содержать суммирующие или фактические данные, которые могут помочь ответить на требуемые вопросы. Она соединяет данные, которые хранились бы во многих таблицах традиционных реляционных базах данных. Таблица факта и таблицы размерности связаны идентифицирующими связями, при этом первичные ключи таблицы размерности мигрируют в таблицу факта в качестве внешних ключей. В размерной модели направления связей явно не показываются – они определяются типом таблиц.

О связи таблицы фактов с таблицами измерений

- Таблица фактов, как правило, содержит уникальный составной ключ, объединяющий первичные ключи таблиц измерений. Чаще всего это целочисленные значения либо значения типа «дата/время» — ведь таблица фактов может содержать сотни тысяч или даже миллионы записей, и хранить в ней повторяющиеся текстовые описания, как правило, невыгодно — лучше поместить их в меньшие по объему таблицы измерений.

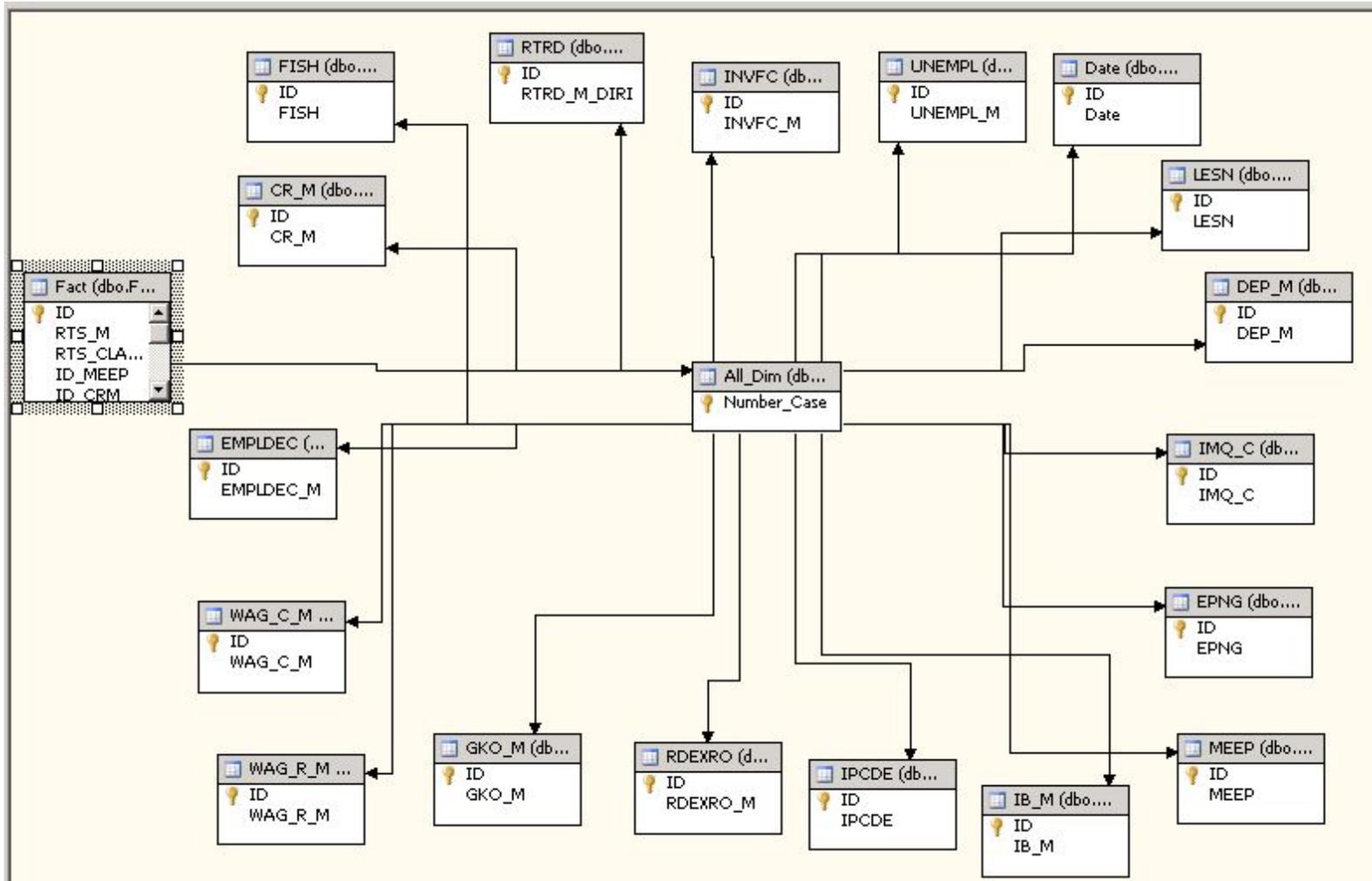
Первичный ключ (таблица факта “REVENUE”) составлен из четырех внешних ключей: movie key. market key. customer key и time_key



Отличие от схемы «звезда»

- Если хотя бы одно измерение содержится в нескольких связанных таблицах, такая схема хранилища данных носит название «снежинка» (snowflake schema).
- Дополнительные таблицы измерений в такой схеме, обычно соответствующие верхним уровням иерархии измерения и находящиеся в соотношении «один ко многим» в главной таблице измерений, соответствующей нижнему уровню иерархии, иногда называют консольными таблицами (outrigger table).

Структура ХД - снежинка



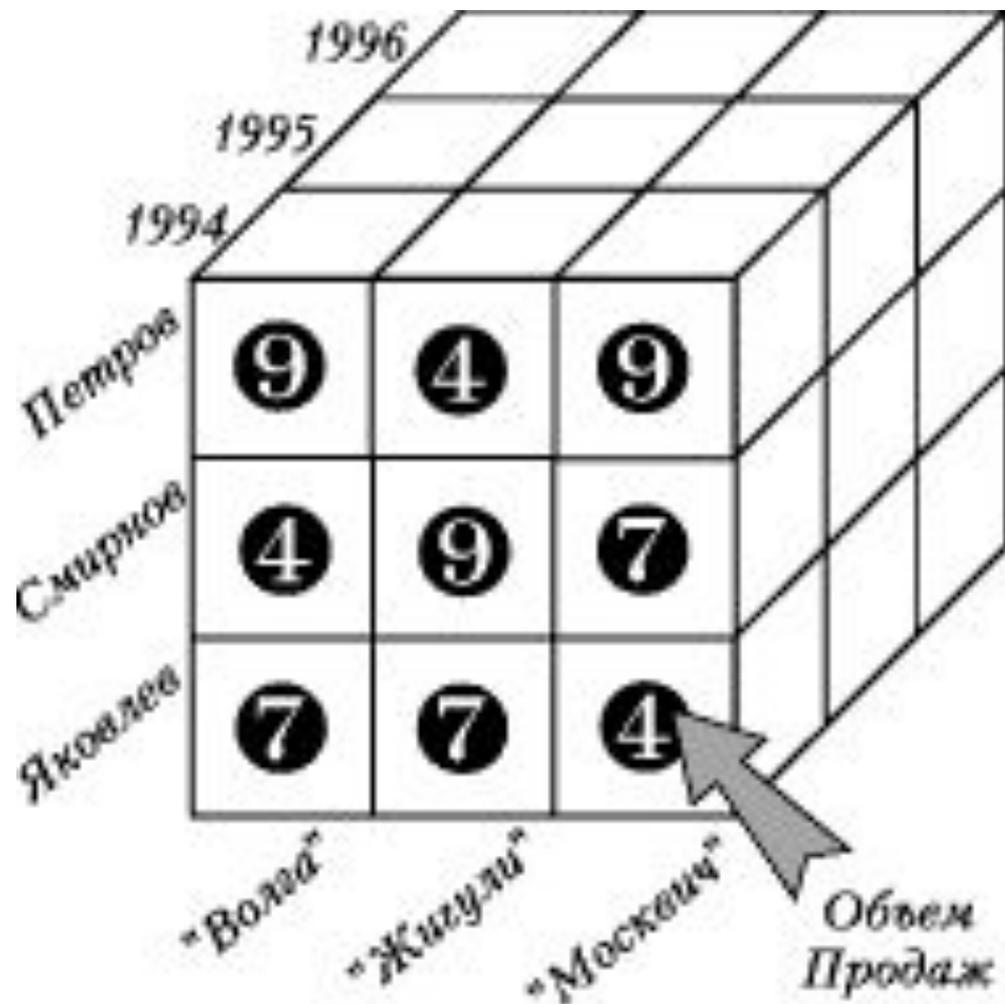
Таблицы измерений

- Таблицы измерений содержат неизменяемые либо редко изменяемые данные (типа справочник). В подавляющем большинстве случаев эти данные представляют собой по одной записи для каждого члена нижнего уровня иерархии в измерении.
- Таблицы измерений также содержат как минимум одно описательное поле (обычно с именем члена измерения) и, как правило, целочисленное ключевое поле (обычно это суррогатный ключ) для однозначной идентификации члена измерения.
- Если будущее измерение, основанное на данной таблице измерений, содержит иерархию, то таблица измерений также может содержать поля, указывающие на «родителя» данного члена в этой иерархии.

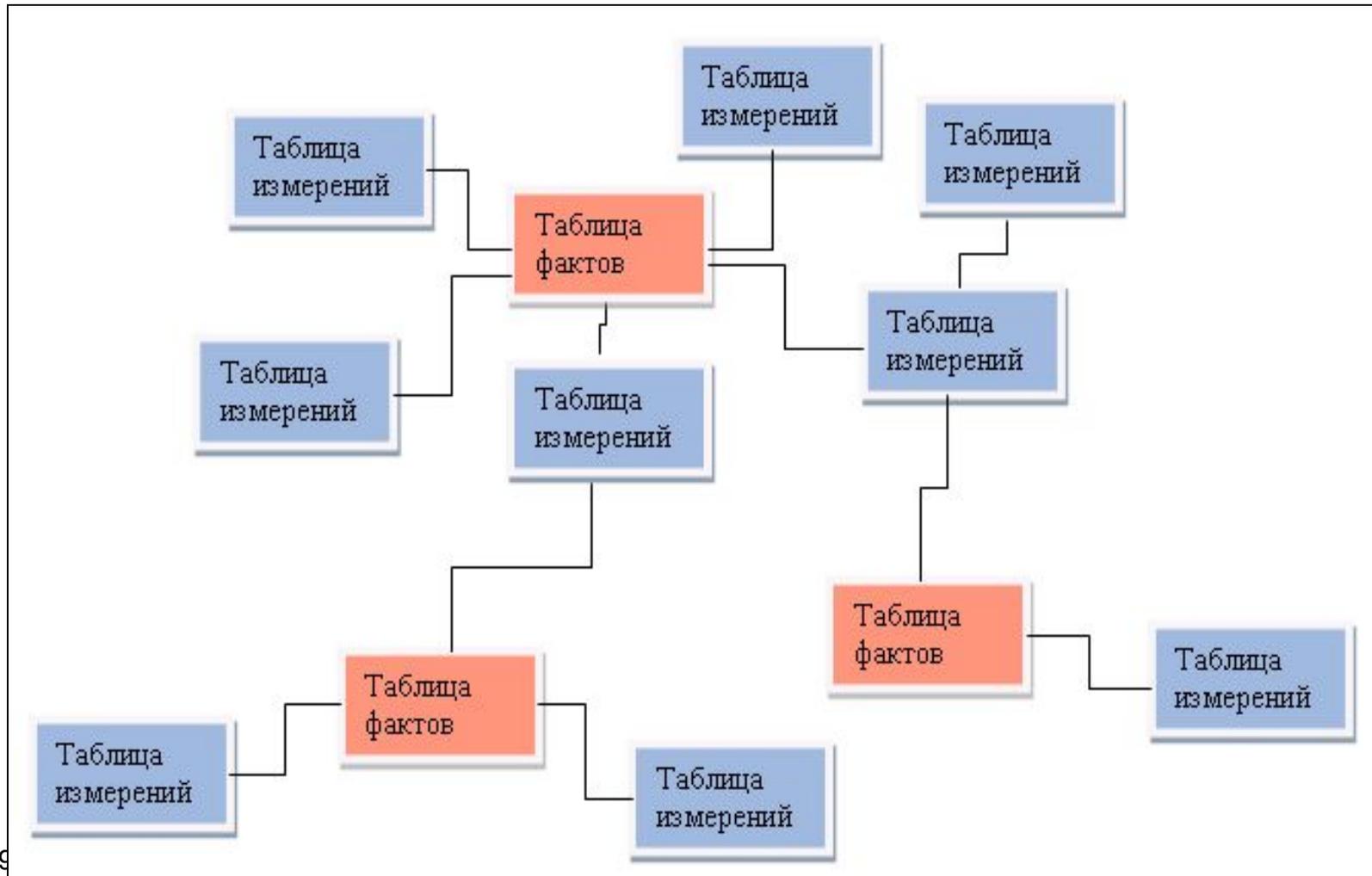
Понятие о кубах

- Куб OLAP - это структура, в которой хранятся совокупности данных, полученные из базы данных OLAP путем всех возможных сочетаний измерений с фактами продаж в таблице фактов.
- Исходя из этого, создание окончательного отчета выполняется гораздо эффективнее, поскольку не требует выполнения никакого сложного запроса.

Вид трехмерного куба



Многомерный куб с несколькими таблицами фактов



Роль измерений в кубе

- Измерения играют роль индексов, используемых для идентификации значений показателей, находящихся в ячейках гиперкуба. Комбинация членов различных измерений играют роль координат, которые определяют значение определенного показателя. Поскольку для куба может быть определено несколько показателей, то комбинация членов всех измерения будет определять несколько ячеек со значениями каждого из показателей. Поэтому для однозначной идентификации ячейки необходимо указать комбинацию членов всех измерений и показатель.

Иерархии в измерениях необходимы для возможности агрегации и детализации значений показателей

Существуют следующие типы иерархий:

- **сбалансированные (balanced);**
- **несбалансированные (unbalanced);**
- **Неровные (balanced).**

Сбалансированные иерархии

- Это - иерархии, в которых число уровней определено её структурой и неизменно, и каждая ветвь иерархического дерева содержит объекты каждого из уровней. Каждому производителю автомобилей может соответствовать несколько марок автомобилей, а каждой марке - несколько моделей автомобилей, поэтому можно говорить о трёхуровневой иерархии этих объектов. В этом случае на первом уровне иерархии располагаются производители, на втором - марки, а на третьем - модели.
- Как видно, для формирования сбалансированной иерархии необходимо наличие связи "один-ко-многим" между объектами менее детального уровня по отношению к объектам более детального уровня. В принципе каждый уровень сбалансированной иерархии можно представить как отдельное простое измерение, но тогда эти измерения окажутся зависимыми, в значит неизбежно повышение разреженности куба.

Несбалансированные иерархии

- Это - иерархии, в которых число уровней может быть изменено, и каждая ветвь иерархического дерева может содержать объекты, принадлежащие не всем уровням, только нескольким первым.
- Необходимо заметить, что все объекты несбалансированной иерархии принадлежат одному типу.
- Типичный пример несбалансированной иерархии - иерархия типа "начальник-подчиненный", где все объекты имеют один и тот же тип - "Сотрудник".

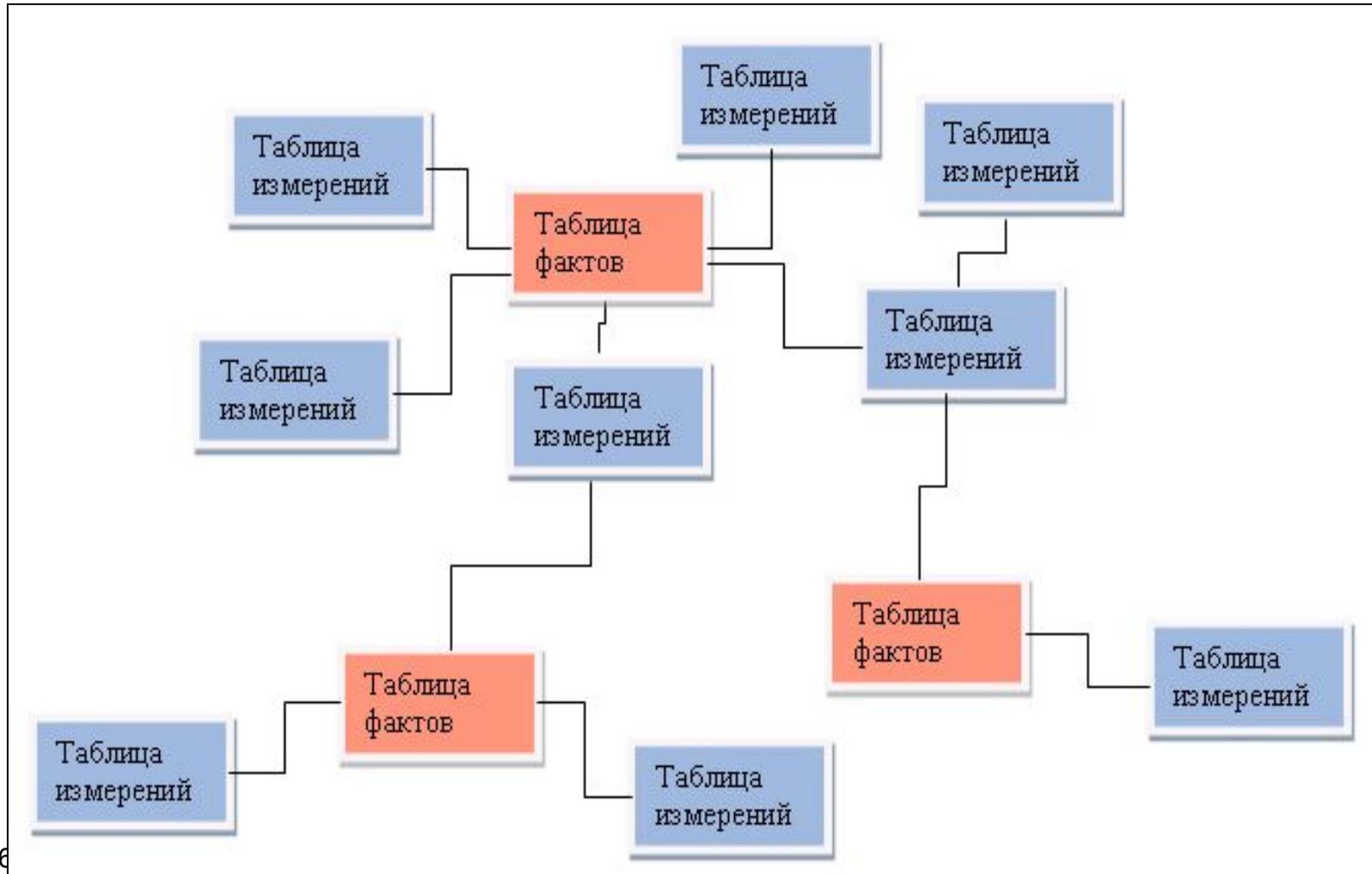
Неровные иерархии

- Это- иерархии, в которых число уровней определено её структурой и постоянно, однако в отличие от сбалансированной иерархии некоторые ветви иерархического дерева могут не содержать объекты какого-либо уровня.
- Иерархии такого вида содержат такие члены, логические "родители" которых не находятся на непосредственно вышестоящем уровне.
- Типичным примером является географическая иерархия, в которой есть уровни "Страны", "Штаты " и "Города", но при этом в наборе данных имеются страны, не имеющие штатов или регионов между уровнями "Страны" и "Города".

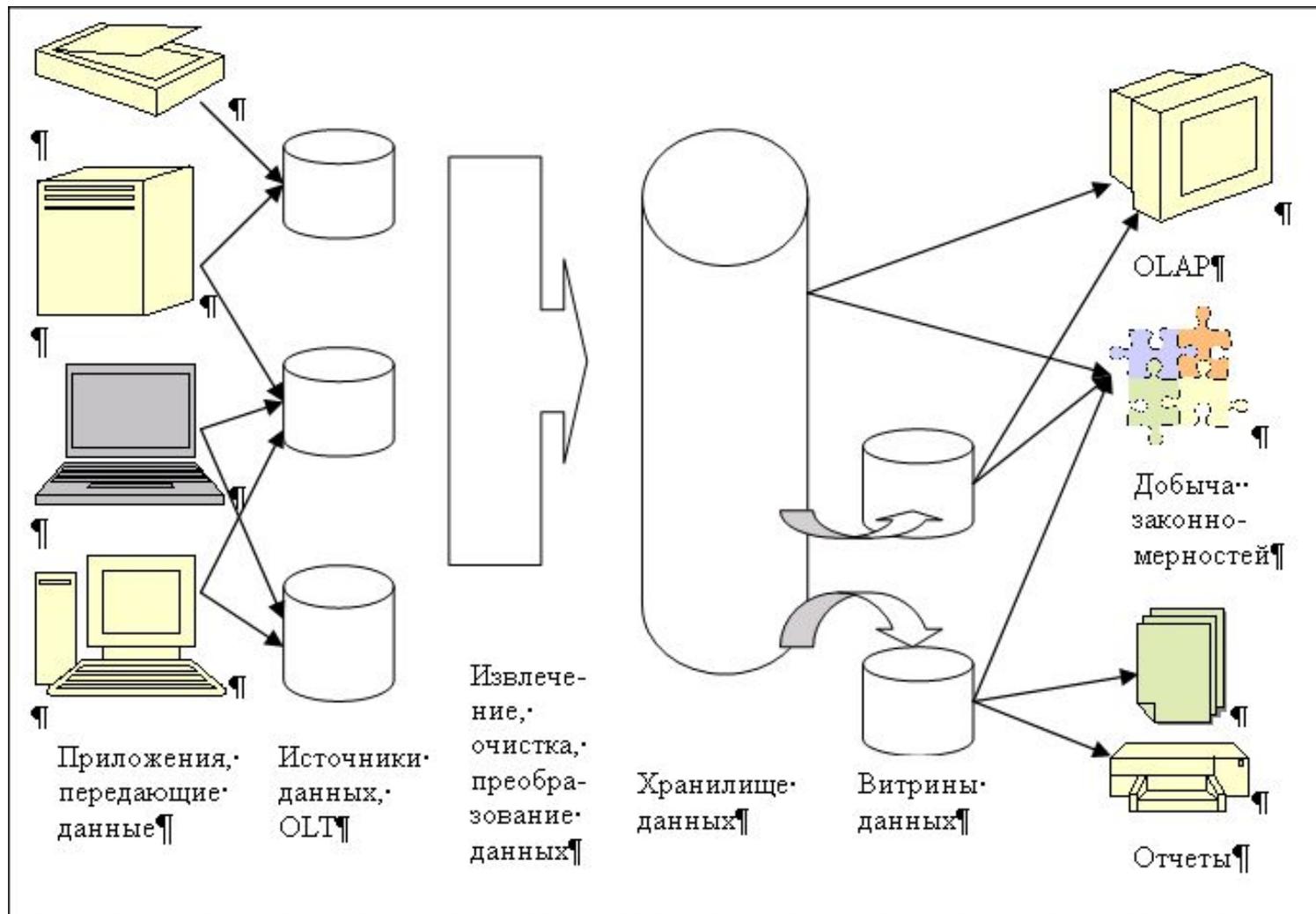
Агрегаты

- Агрегатами называют агрегированные по определенным условиям исходные значения показателей. Обычно под агрегацией понимается любая процедура формирования меньшего количества значений (агрегатов) на основании большего количества исходных значений. В дальнейшем под терминами агрегирование и агрегация будем понимать исключительно процесс суммирования данных.
- Заблаговременное формирование и сохранение агрегатов с целью уменьшения времени отклика на пользовательский запрос является основным свойством систем поддержки оперативного анализа.

Многомерный куб с несколькими таблицами фактов



DW с витринами данных





Основные сведения о MS SQL Server 2005

Microsoft SQL Server

система управления реляционными базами данных (СУБД) система управления реляционными базами данных (СУБД), разработанная корпорацией Microsoft система управления реляционными базами данных (СУБД), разработанная корпорацией Microsoft. Основной используемый язык запросов — Transact-SQL система управления реляционными базами данных (СУБД), разработанная корпорацией Microsoft. Основной используемый язык запросов — Transact-SQL, создан совместно Microsoft и Sybase. Используется для работы с

Платформа данных MS SQL Server

- MS SQL Server 2005 является интегрированным решением, которое наделяет пользователей безопасной, надежной, и продуктивной платформой для обработки информации и приложений, касающихся интеллектуальных ресурсов предприятия.
- MS SQL Server 2005 предоставляет инструменты для профессионалов информационных технологий так же, как и для работников информационной сферы, уменьшая сложность создания, развёртывания, управления и использования данных предприятия и аналитических приложений на платформах от мобильных устройств до информационных систем предприятия.

Платформа данных MS SQL Server включает следующие инструменты:

- **Реляционная база данных:** безопасное, надёжное, масштабируемое, высокодоступное ядро с улучшенной производительностью и поддержкой структурированных и неструктурированных (XML) данных.
- **Replication Services:** репликация данных для распределённых и мобильных приложений обработки данных, высокая доступность систем, масштабируемый параллелизм со вторичными хранилищами данных для отчётных решений предприятия и интеграция с разнородными системами, включая существующие базы данных Oracle.
- **Notification Services:** развитые возможности уведомлений для разработки и внедрения масштабируемых приложений, способных доставлять персонализированные, своевременные обновления информации множеству соединённых и мобильных устройств.

Платформа данных MS SQL Server включает следующие инструменты (продолжение):

- **Integration Services:** возможности извлечения, преобразования и загрузки для хранилищ данных и интеграции данных в масштабе предприятия.
- **Analysis Services:** аналитическая обработка в реальном времени (OLAP) для быстрого, сложного анализа больших и смешанных наборов данных, использующая многомерное хранение.
- **Reporting Services:** исчерпывающее решение для создания, управления и доставки как традиционных бумажных отчётов, так и интерактивных, основанных на технологии WWW отчётов.

Платформа данных MS SQL Server включает следующие инструменты (продолжение):

- **Инструменты управления:** MS SQL Server включает средства управления для развитого управления и настройки баз данных, также как и тесную интеграцию с такими инструментами, как Microsoft Operations Manager (MOM) и Microsoft Systems Management Server (SMS). Стандартные протоколы доступа к данным существенно уменьшают время, необходимое для интеграции данных SQL Server с существующими системами. Поддержка Web служб позволяет обеспечить взаимодействие с другими приложениями и платформами.
- **Инструменты разработки:** MS SQL Server предлагает интегрированные инструменты разработки для ядра базы данных, извлечения, трансформации и загрузки данных, извлечения информации, OLAP и отчётности, которые тесно интегрированы с Microsoft Visual Studio® для предоставления сквозных возможностей разработки приложений.

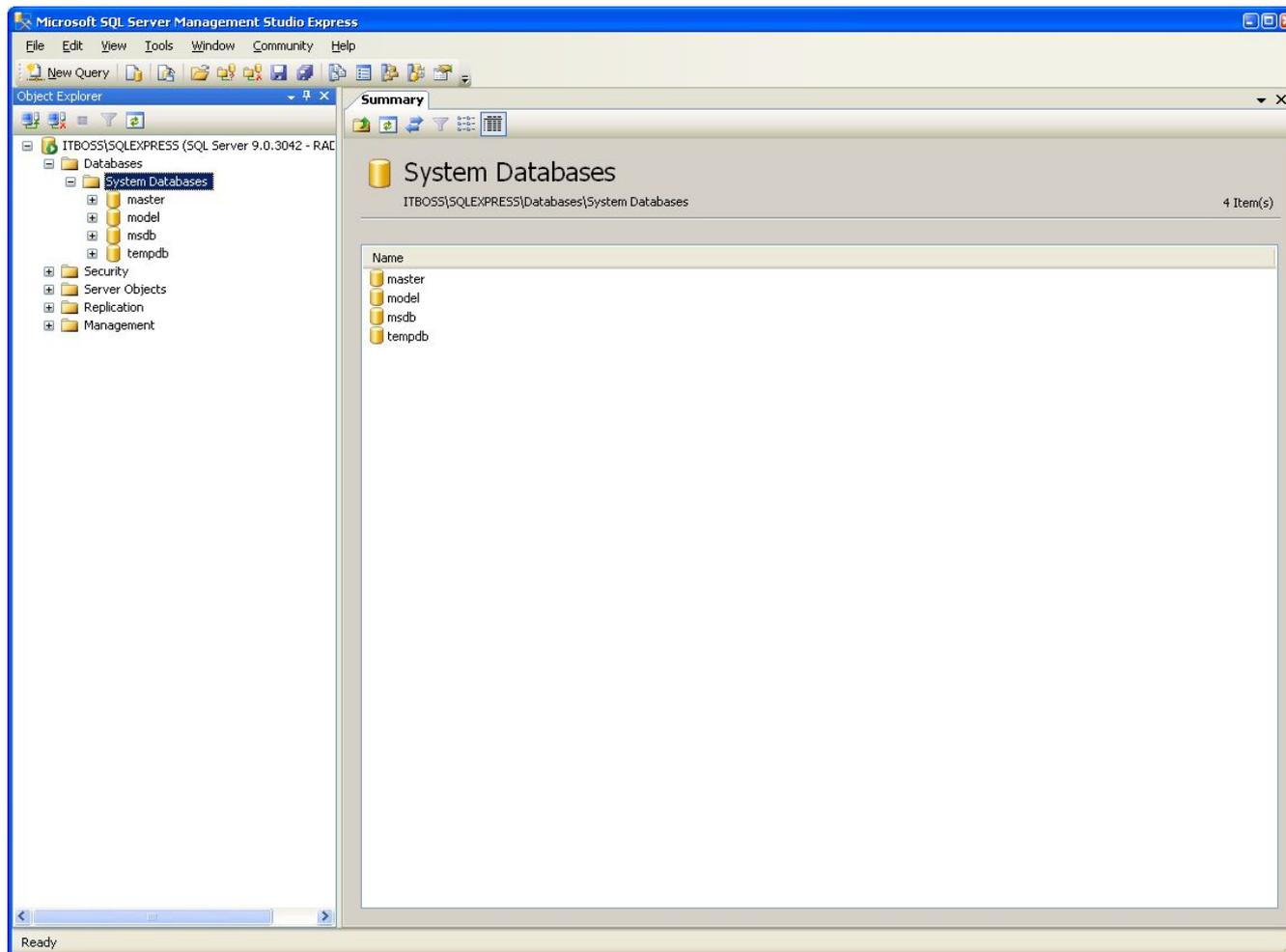
Управляемость

- MS SQL Server 2005, как платформа управления данными предприятия, предоставляет единую консоль управления, которая позволяет администраторам данных, отслеживать, управлять и настраивать все базы данных и связанные службы по всему предприятию.
- Он предоставляет расширяемую инфраструктуру управления, которая может быть легко запрограммирована при помощи SQL Management Objects (SMO), позволяя пользователям переделывать и расширять их среду управления и независимым поставщикам программных продуктов (ISV) создавать дополнительные инструменты и функциональность для дальнейшего расширения возможностей, поставляемых по умолчанию.

MS SQL Server Management Studio

- SQL Server 2005 упрощает управление средствами единой интегрированной консоли управления для мониторинга и управления реляционной базой данных SQL Server, Integration Services, Analysis Services, Reporting Services, Notification Services и SQL Mobile на большом числе распределённых серверов и баз данных.
- Администратор баз данных может выполнять несколько задач одновременно, включая следующие:
 - создание и выполнение запроса,
 - просмотр серверных объектов,
 - управление объектом,
 - отслеживание активности системы,
 - просмотр оперативной справки.
- SQL Server Management Studio содержит среду разработки для создания, редактирования и управления сценариев и хранимых процедур, используя Transact-SQL, многомерные выражения (MDX), XMLA и SQL Server Mobile Edition.

MS SQL Server Management Studio

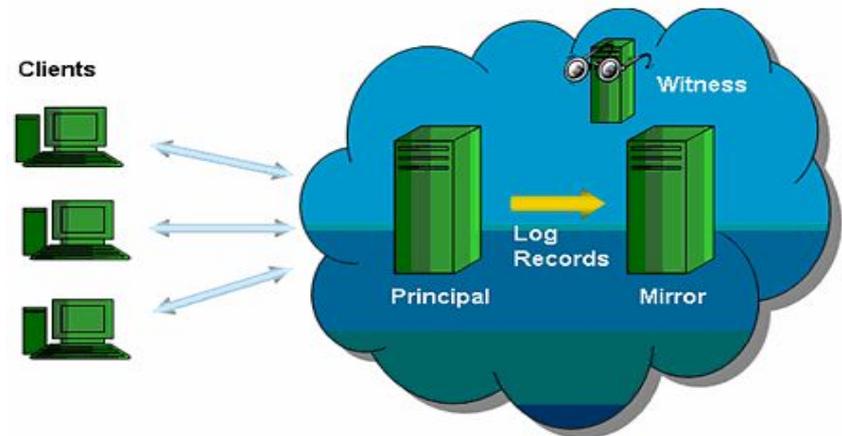


Доступность

- MS SQL Server 2005 позволяет использовать передовые функции высокой доступности, такие как:
 - зеркалирование базы данных,
 - кластеры с восстановлением после отказа,
 - моментальные снимки базы данных,
 - улучшенные оперативные операции.
- Данные средства позволяют снизить время простоя и помогут гарантировать, что важнейшие системы предприятия останутся доступными.

Зеркалирование баз данных

- Зеркалирование базы данных позволяет непрерывно передавать содержимое журнала транзакций исходного сервера на один целевой сервер.
- В случае сбоя основной системы приложения могут немедленно соединиться с базой на вспомогательном сервере.
- Зеркалирование базы данных работает на стандартном серверном оборудовании и не требует специального хранилища и контроллеров.



Кластеры с восстановлением после отказа

- Кластеры с восстановлением после отказа являются решением высокой готовности, которое использует службы кластеризации Microsoft Windows® для создания устойчивых к сбоям виртуальных серверов, обеспечивающих быстрое восстановление в случае сбоя сервера баз данных.
- В SQL Server 2005 поддержка кластеров с восстановлением после отказа была расширена для Analysis Services, Notification Services и служб репликации SQL Server.
- Максимальное количество узлов кластера было увеличено до восьми.

Моментальные снимки базы данных

- MS SQL Server 2005 даёт администраторам баз данных возможность создавать мгновенные, доступные только для чтения снимки базы данных.
- **Моментальный снимок базы данных** предоставляет устойчивый снимок без накладных расходов на время и хранение на создание полной копии базы данных. Когда основная база данных расходится со снимком, последний добавляет свои собственные копии страниц, по мере того, как они изменяются.
- Снимок может быть использован для быстрого восстановления от случайного изменения базы данных путём простого применения исходных страниц из снимка к основной базе данных.

Оперативные операции (индексные операции и восстановление)

- Опция оперативного индекса позволяет одновременно производить модификацию (вставки, изменения, удаления) таблицы или данных кластерного индекса и любых связанных индексов во время выполнения команды DDL.
- С поддержкой оперативных индексных операций можно добавлять индексы без помех доступу к таблицам или другим существующим индексам.
- В MS SQL Server 2005 существует возможность выполнять операцию восстановления во время работы экземпляра SQL Server.
- Возможность оперативного восстановления улучшает доступность SQL Server, так как недоступны только восстанавливаемые данные. Остальная часть базы данных остаётся доступной.

Секционирование таблиц и индексов

- Секционирование таблиц и индексов облегчает управление большими базами данных посредством управления малыми, более управляемыми частями базы данных.
- MS SQL Server 2005 предоставляет новую возможность для секционирования таблиц по файловым группам в базе данных.
- Горизонтальное секционирование разделяет таблицу на меньшие группы, основываясь на схеме секционирования.
- Секционирование таблиц создано для очень больших баз данных, размером от сотен гигабайт до терабайт и более.

Уровень изоляции "снимок"

- Когда данные скопированы, преобразованы и помещены в базу данных для анализа, она должна поддерживаться и/или периодически перестраиваться.
- На создание индекса могут уйти часы и это может оказаться не тем, что требуется пользователям на самом деле. Здесь и используется уровень изоляции "снимок".
- Он позволяет пользователям получить доступ к последней подтвержденной записи, используя транзакционно непротиворечивый снимок базы данных.
- Новый уровень изоляции имеет следующие выгоды:
 - Увеличенная доступность данных для только читающих приложений
 - Неблокирующие операции чтения, доступные в среде OLTP
 - Автоматическое обязательное определение конфликтов при записи транзакций

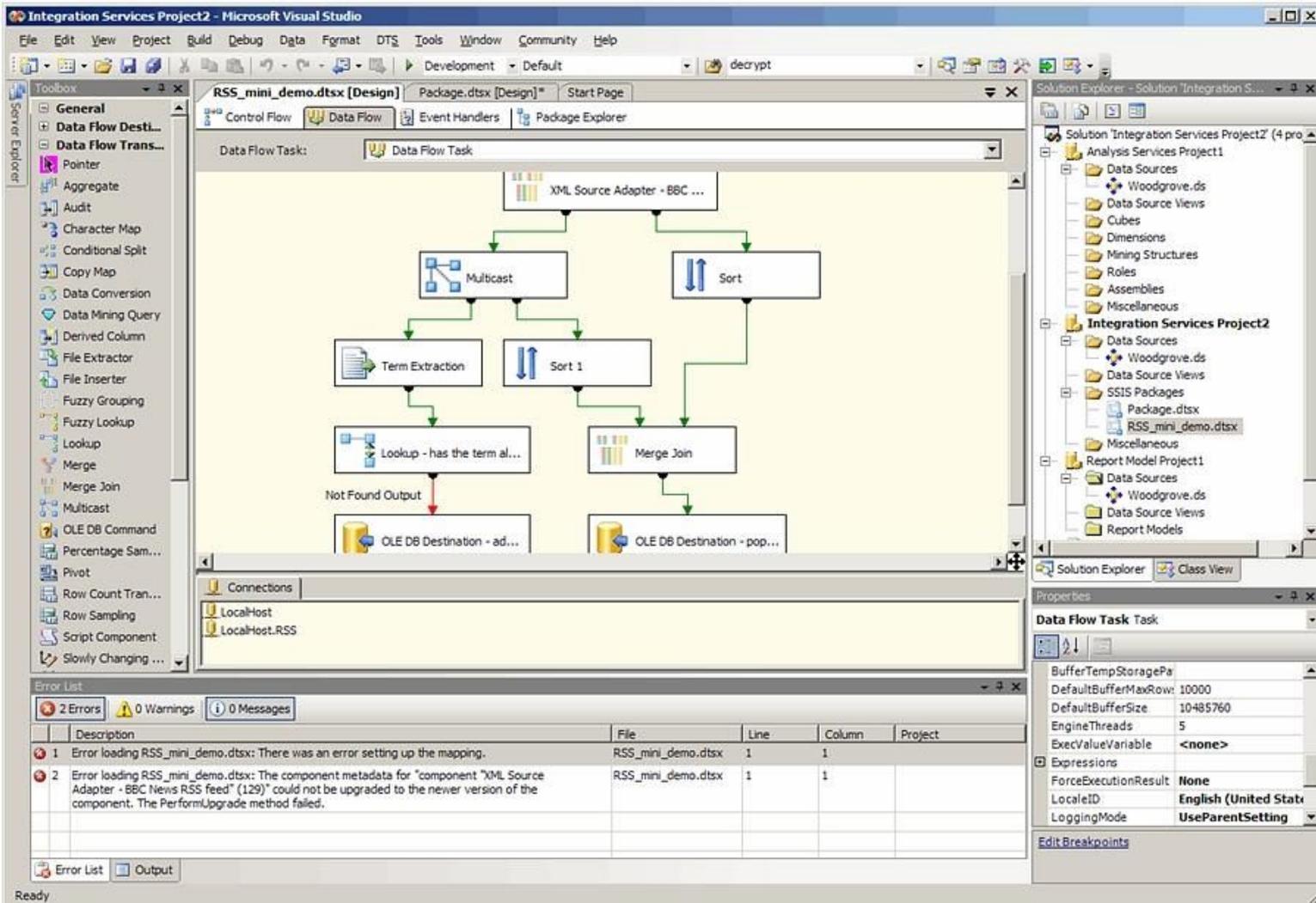
Безопасность

- MS SQL Server 2005 имеет существенные улучшения в модели безопасности платформы базы данных, с возможностью предоставить более точный и гибкий контроль для обеспечения безопасности данных:
 - Применение политик для паролей учётных записей SQL Server в области аутентификации;
 - Обеспечение большей модульности для указания разрешений на различных уровнях в области авторизации;
 - Разделение владельца и схемы в области управления безопасностью.

Встроенное шифрование

- SQL Server 2005 поддерживает шифрование внутри базы данных, полностью интегрированное с инфраструктурой управления ключами.
- По умолчанию, коммуникации между клиентом и сервером шифруются.
- Для того, чтобы концентрировать гарантии безопасности, может быть определена серверная политика, которая будет отвергать незашифрованные коммуникации.

Интерфейс Business Intelligence Development Studio в Visual Studio



Integration Services Project2 - Microsoft Visual Studio

File Edit View Project Build Debug Data Format DTS Tools Window Community Help

Development - Default decrypt

Toolbox

- General
- Data Flow Dest...
- Data Flow Trans...
- Pointer
- Aggregate
- Audit
- Character Map
- Conditional Split
- Copy Map
- Data Conversion
- Data Mining Query
- Derived Column
- File Extractor
- File Inserter
- Fuzzy Grouping
- Fuzzy Lookup
- Lookup
- Merge
- Merge Join
- Multicast
- OLE DB Command
- Percentage Sam...
- Pivot
- Row Count Tran...
- Row Sampling
- Script Component
- Slowly Changing ...

Server Explorer

RSS_mini_demo.dtsx [Design] Package.dtsx [Design]* Start Page

Control Flow Data Flow Event Handlers Package Explorer

Data Flow Task: Data Flow Task

XML Source Adapter - BBC ...

Multicast

Sort

Term Extraction

Sort 1

Lookup - has the term al...

Merge Join

Not Found Output

OLE DB Destination - ad...

OLE DB Destination - pop...

Connections

- LocalHost
- LocalHost.RSS

Solution Explorer - Solution Integration S...

- Solution Integration Services Project2 (4 pro...
- Analysis Services Project1
 - Data Sources
 - Woodgrove.ds
 - Data Source Views
 - Cubes
 - Dimensions
 - Mining Structures
 - Roles
 - Assemblies
 - Miscellaneous
- Integration Services Project2
 - Data Sources
 - Woodgrove.ds
 - Data Source Views
 - SSIS Packages
 - Package.dtsx
 - RSS_mini_demo.dtsx
 - Miscellaneous
- Report Model Project1
 - Data Sources
 - Woodgrove.ds
 - Data Source Views
 - Report Models

Solution Explorer Class View

Properties

Data Flow Task Task

BufferTempStoragePa

DefaultBufferMaxRow: 10000

DefaultBufferSize: 10485760

EngineThreads: 5

ExecValueVariable: <none>

Expressions

ForceExecutionResult: None

LocaleID: English (United Stat

LoggingMode: UseParentSetting

Edit Breakpoints

Error List

2 Errors 0 Warnings 0 Messages

Description	File	Line	Column	Project
1 Error loading RSS_mini_demo.dtsx: There was an error setting up the mapping.	RSS_mini_demo.dtsx	1	1	
2 Error loading RSS_mini_demo.dtsx: The component metadata for "component "XML Source Adapter - BBC News RSS feed" (129)" could not be upgraded to the newer version of the component. The PerformUpgrade method failed.	RSS_mini_demo.dtsx	1	1	

Error List Output

Ready

Сквозная интегрированная платформа для работы с интеллектуальными ресурсами предприятия

- Набор BI инструментов SQL Server 2005 предоставляет сквозную интеграцию BI приложений:
 - **Проектирование** : Business Intelligence Development Studio является первой интегрированной средой разработки, созданной для разработчиков BI. Основанная на Visual Studio 2005, Business Intelligence Development Studio является богатой, интегрированной, профессиональной платформой разработки.
 - **Интеграция** : SQL Server Integration Services (SSIS) были переписаны для выполнения сложной интеграции данных, преобразований, и синтеза на высокой скорости для очень больших объёмов данных. Создание и отладка пакетов становится интересной в Business Intelligence Development Studio. Integration Services, Analysis Services и Reporting Services взаимодействуют для предоставления цельного вида данных, полученных из разнородных источников.
 - **Анализ** : В Data Mining добавлены новые важные алгоритмы, включая правило ассоциации, временные ряды, регрессионные деревья, кластеризацию последовательностей, нейронные сети, простой Байес. Можно хранить данные в реляционной базе данных, в многомерной базе данных или использовать функцию упреждающего кэширования для получения преимуществ обоих вариантов. Новые важные аналитические возможности были также добавлены в кубы Analysis Services, включая инфраструктуру Ключевых Индикаторы Производительности (KPI), сценарии MDX и прочие встроенные расширенные бизнес аналитики. Доставка отчётов Reporting Services и инфраструктура управления легко распределять сложную аналитику среди широчайшей аудитории.
 - **Отчёты** : Reporting Services расширяют платформу BI Microsoft до уровня потребителей, которые используют результаты анализа. Reporting Services являются управляемой средой отчётов предприятия, встроенной и управляемой через Web службы. Отчёты могут быть персонализированы и доставлены во множестве форматов, с диапазоном интерактивных опций и опций печати. Нововведением SQL Server 2005 является средство создания отчётов – Построитель Отчётов.
 - **Управление** : SQL Server Management Studio объединяет в себе функции управления всеми компонентами SQL Server 2005.

Analysis Services

- Analysis Services в SQL Server 2005 впервые предоставляют единообразное и интегрированное представление всех бизнес данных, как основы для всей традиционной отчётности, OLAP анализа, и Data Mining.
- Соединяя лучшие 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).
- UDM является центральным хранилищем метаданных, определяющим бизнес сущности, бизнес логику, вычисления, и метрики, служащее источником для всех отчётов, электронных таблиц, программ просмотра OLAP, KPI и аналитических приложений.

Data Mining

- Data Mining в Microsoft SQL Server 2005 является интеллектуальной технологией предприятия, которая помогает создавать сложные аналитические модели и объединять их с бизнес операциями.
- Microsoft SQL Server 2005 Analysis Services задают новую основу для извлечения данных.
- Глубокая интеграция с семейством BI инструментов SQL Server, богатый набор инструментов, интерфейсов прикладного программирования и алгоритмов в SQL Server позволяет создавать новый тип BI приложений, повышающих производительность и прибыли и снижающих издержки через создание специальных решений, работающих с данными, для широкого круга проблем бизнеса.

Построитель Отчетов

The screenshot displays the Microsoft Report Builder interface. The main window shows a report titled "My New Report" with a table of sales data. The table has four columns: Product Category, Product Subcategory, Product Name, and Sum Sales Amount. The data is summarized in a total row.

Product Category	Product Subcategory	Product Name	Sum Sales Amount
xxxxxxx	xxxxxxx	xxxxxxx	\$0.00
		Total	\$0.00
	Total		\$0.00
Total			\$0.00

The interface includes a menu bar (File, Edit, View, Format, Report, Help), a toolbar with various report-building tools, and a Data Content pane on the left. The Data Content pane shows a Navigator with a tree view of data sources: Product, Reseller Sales as Product, Internet Sales as Product, Product Subcategory, and Product Category. Below the Navigator is a Fields list containing various data fields, including Product, Product Key, Product Name, and Sum Sales Amount.

