

Дисциплина: «Базы данных»
Специальность: №08080165
«Прикладная информатика (в экономике)»

Информационные хранилища

Институт информатики, инноваций и бизнес систем
Кафедра Информационных систем и прикладной
информатики
Старший преподаватель Богданова О.Б.

Хранилища данных

Системы операционной обработки способны выполнять **тривиальный анализ данных** – вычислять максимальные, минимальные и средние значения атрибутов.

Система поддержки принятия решений

Под **системой поддержки принятия решений** понимают вычислительный комплекс, ориентированный на анализ данных и обеспечивающий получение информации, необходимой для разработки решений в сфере управления.

Система поддержки принятия решений

К числу задач, которые традиционно решают системы поддержки принятия решений, относятся:

оценка альтернатив решений,

прогнозирование,

классификация,

кластеризация,

выявления ассоциаций

и др.

Система поддержки принятия решений

Для того, чтобы извлекать полезную информацию из данных, они должны быть организованы особым, отличным от принятого в OLTP-системах образом.



Система поддержки принятия решений

Во-первых, для выполнения аналитических запросов необходима обработка больших информационных массивов.

Во-вторых, выполнение некоторых аналитических запросов, например, анализ тенденций и прогнозирование, требует технологической упорядоченности данных.

В-третьих, данные, используемые для целей анализа, как правило, отличаются от данных операционных систем.

Хранилища данных

Концепция хранилищ данных –

**это концепция подготовки данных для последующего
анализа.**

Она предполагает выполнение следующих положений

Хранилища данных

- 1) Интеграция и согласование данных из различных источников: традиционных систем операционной обработки, информационной обработки, информации из внутренних и внешних по отношению к организации электронных архивов.
 - 2) Разделение наборов данных, используемых системами выполнения транзакций и системами поддержки принятия решений.
-

Хранилища данных

Хранилище данных

«предметно-ориентированный, интегрированный, неизменяемый и поддерживающий хронологию набор данных, предназначенный для обеспечений принятия управленческих решений.»

Хранилища данных

Исходные данные для анализа производятся системами операционной обработки, поступают из электронных архивов и от поставщиков информации, например, онлайн-информационных агентств.

Эти источники слабо связаны между собой, поэтому и данные, которые они предоставляют, имеют различную структуру и форматы представления.

Хранилища данных

Необходимо произвести согласование данных разных источников, чтобы ими было удобно оперировать при анализе.

Это подразумевает приведение их к **единому формату**, а также устранение дублирующихся и некорректных значений.

Хранилища данных

Подготовленные данные загружаются в хранилище.

Пользователи-аналитики осуществляют доступ к нему через клиентские приложения.

В отличие от систем операционной обработки в СППР, использующих концепцию ХД, критерии поиска и состав выдаваемой в виде отчета информации не фиксируются при ее разработке, пользователи оперируют в основном заранее не регламентированными запросами (ad-hoc query)

Концепция хранилища данных

OLTP: Есть ли свободные места в купе поезда Москва — Сочи, отправляющегося 20 августа в 23:15?

OLAP: Каким будет объем продаж железнодорожных билетов в денежном выражении в следующих трех месяцах с учетом сезонных колебаний

Концепция хранилища данных

Использование концепции данных в системе поддержки принятия решений преследует следующие цели:

- 1) Своевременное обеспечение аналитиков всей информацией, необходимой для выработки решений;
 - 2) Создание единой модели данных организации;
 - 3) Создание интегрированного источника данных.
-

Концепция хранилища данных

Ориентация на предметную область.

Хранилище должно разрабатываться с учетом специфики предметной области, а не приложений, оперирующих данными.

Структура хранилища должна отражать представления аналитика об информации, с которой ему приходится работать.

Концепция хранилища данных

Интегрированность.

Информация в хранилище загружается из приложений, созданных разными разработчиками.

Необходимо объединить данные этих приложений, приведя их к единому синтаксическому и семантическому виду.

Концепция хранилища данных

Неизменяемость данных.

Данные после загрузки в аналитических системах остаются неизменными.

Внесение каких-либо изменений, кроме добавления записей не предполагаются.

Концепция хранилища данных

Поддержка хронологии.

Учет хронологии достигается введением ключевых атрибутов «ДАТА» и/или «ВРЕМЯ» в структуры хранилища данных

Свойства данных в СППР и системах операционной обработки (OLTP)

Характеристика	OLTP	OLAP
Преобладающие операции	Ввод данных, поиск	Анализ данных
Характер запросов	Много простых транзакций	Сложные транзакции
Хранимые данные	Оперативные, детализированные	Охватывающие большой период времени, агрегированные
Вид деятельности	Оперативная, тактическая	Аналитическая, стратегическая
Тип данных	Структурированные	Разнотипные

OLAP

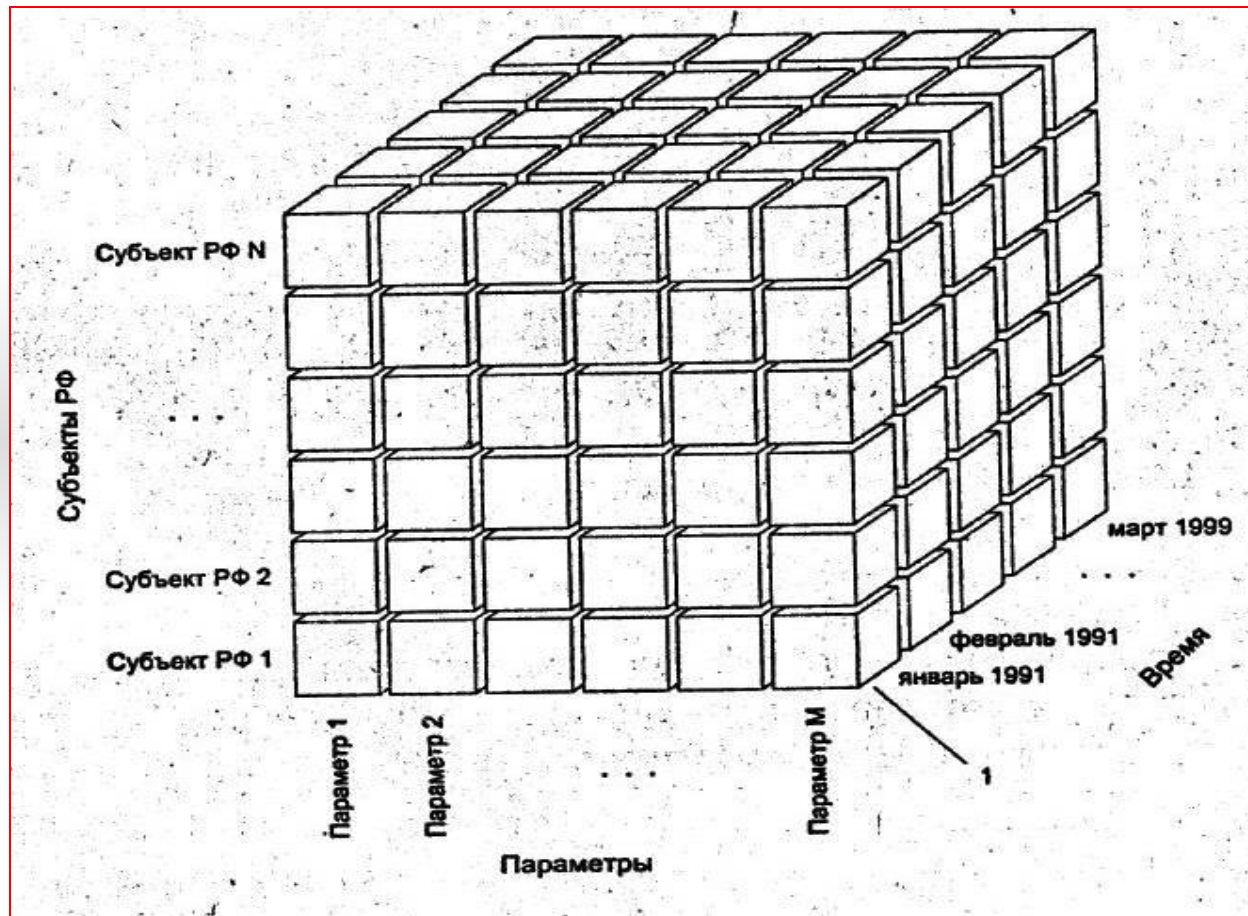
Под **OLAP-системой** принято понимать СППР, основанную на концепции хранилища данных и обеспечивающую малое время выполнения аналитических запросов.

Модели данных, используемые для построения хранилищ

В настоящее время два чем-то конкурирующих, а в чем-то взаимодополняющих друг друга подхода к построению хранилищ данных:

- подход, основанный на использовании многомерной модели БД (Multidimensional OLAP - **MOLAP**)
 - подход, использующий реляционную модель БД (Relational OLAP - **ROLAP**)
-

Многомерная модель хранилища



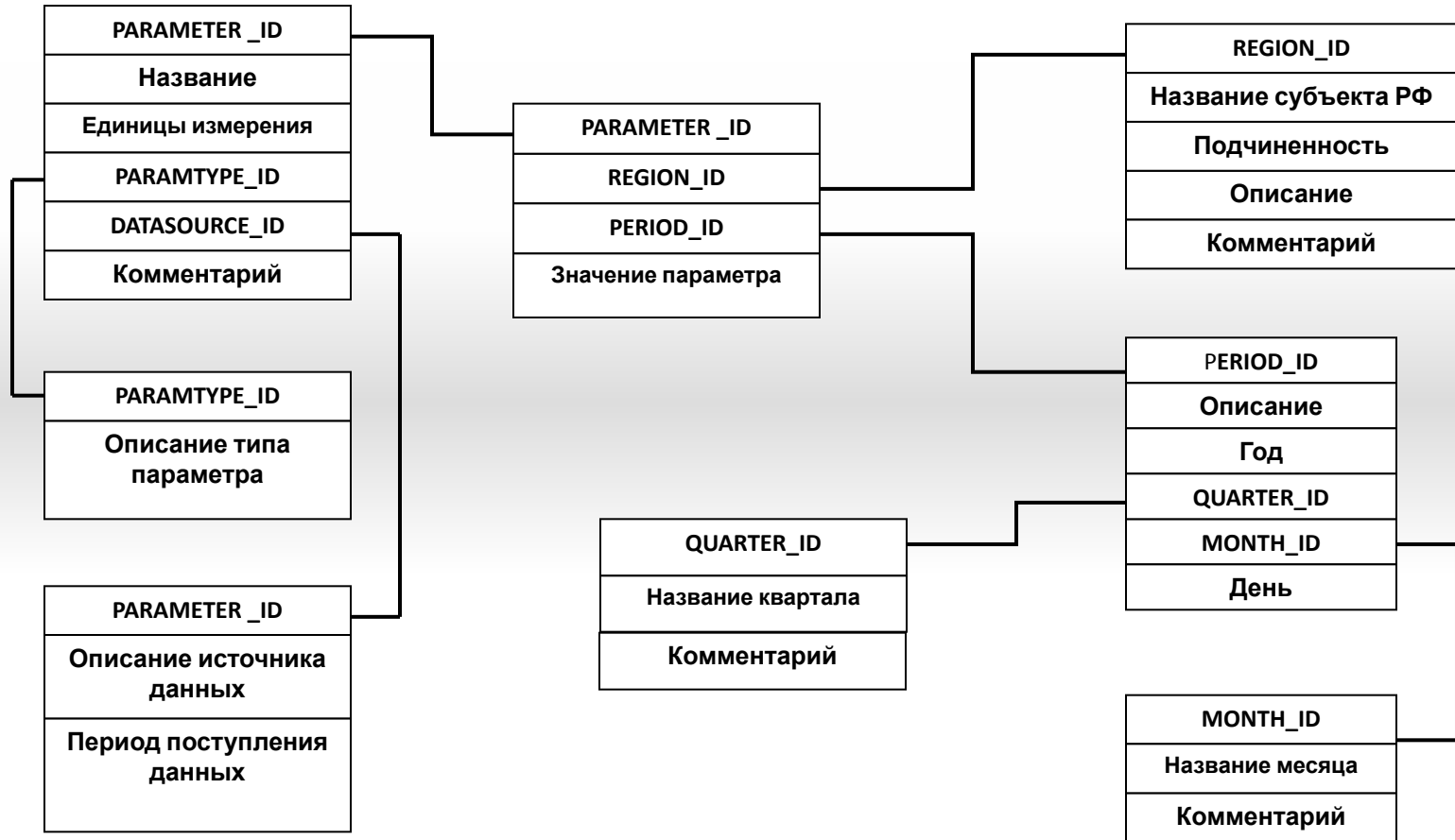
Реляционная модель хранилища данных



(Схема звезда)

Реляционная модель хранилища данных

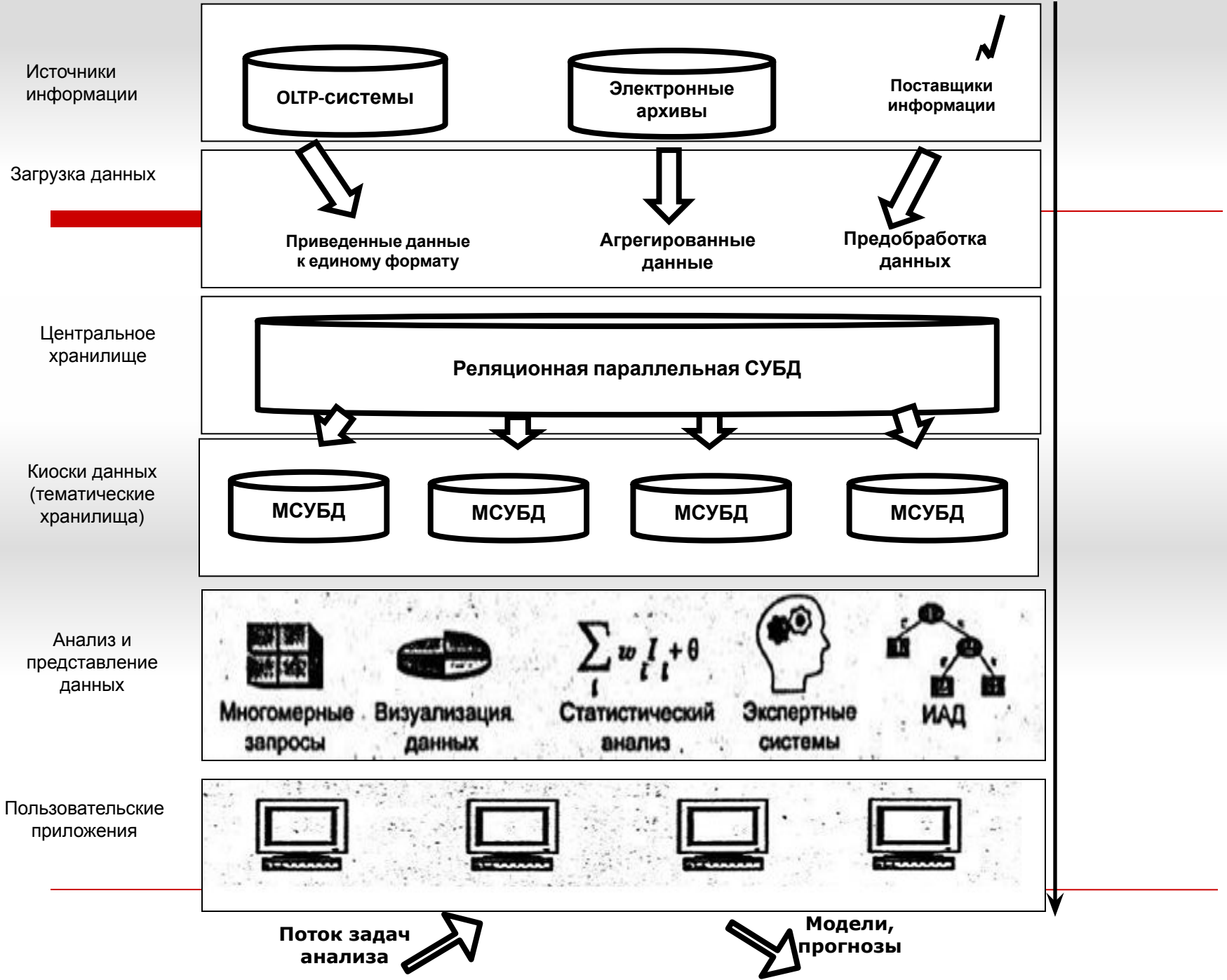
(Схема «Снежинка»)



Киоски данных (Data Marts)

Комбинация многомерного и реляционного
подходов:

Киоски данных — специализированное тематическое хранилище, обслуживающее одно из направлений деятельности организации



Источники информации



Поставщики информации

Загрузка данных

Приведенные данные к единому формату

Агрегированные данные

Предобработка данных

Центральное хранилище



Киоски данных (тематические хранилища)



Анализ и представление данных



Многомерные запросы



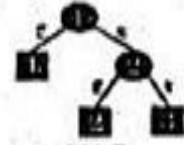
Визуализация данных

$$\sum_l w_l I_l + \theta$$

Статистический анализ



Экспертные системы



ИАД

Пользовательские приложения



Поток задач анализа

Модели, прогнозы

Классификация хранилищ данных

- Маленькие
 - Средние
 - Большие
 - Сверхбольшие
-