

## **Лекция 1**

# **ОСНОВЫ ПОСТРОЕНИЯ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИХ СИСТЕМ И СИСТЕМ ПОДДЕРЖКИ И ПРИНЯТИЯ РЕШЕНИЙ**

### **Учебные вопросы:**

- 1 Структура и архитектура информационно-аналитических систем и систем поддержки принятия решений**
- 2 Информационные системы, ориентированные на операционную (транзакционную) обработку данных (OLTP)**
- 3 Информационные системы оперативного анализа данных (OLAP)**
- 4 Структура и задачи интеллектуального анализа данных**

# Литература

1. Тейлор Д., Рейден Н. Почти интеллектуальные системы. Как получить конкурентные преимущества путём автоматизации принятия решений. – Пер. с англ. – СПб: Символ Плюс, 2009. – 448 с.

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

3. Ларсон Б. Разработка бизнес-аналитики в SQL Server 2005. – СПб.: Питер, 2008. – 684 с.

4. Малыхина М.П. Базы данных: основы, проектирование, использование. – Спб.: БХВ. – Петербург, 2004. – 512 с.

# 1 Структура и архитектура информационно-аналитических систем и систем поддержки принятия решений

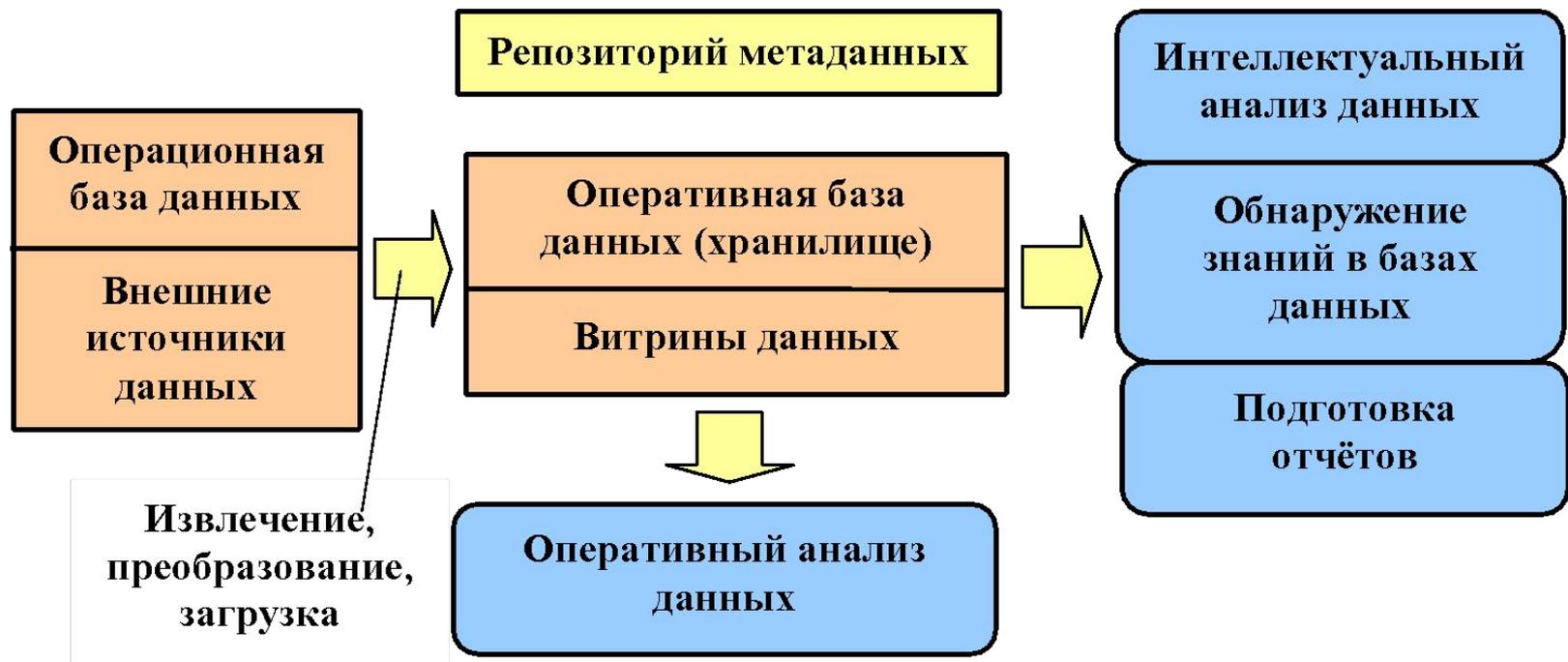


Рисунок 1 – Архитектура систем поддержки принятия решений

ХД содержат информацию, собранную из нескольких операционных баз данных (БД). Объем данных в ХД, как правило, на порядок больше объёма операционных БД и может достигать сотен гигабайт или нескольких терабайт. Как правило, хранилище данных поддерживается независимо от оперативных баз данных организации, поскольку требования к функциональности и производительности аналитических приложений отличаются от требований к транзакционным системам. ХД создаются специально для приложений поддержки принятия решений и предоставляют накопленные за определенное время, сводные и консолидированные данные, которые более приемлемы для анализа, чем детальные индивидуальные записи. Рабочая нагрузка ХД состоит из нестандартных, сложных запросов к миллионам записей. При выполнении запросов осуществляется огромное количество операций сканирования, соединения и агрегирования.

## 2 Информационные системы, ориентированные на операционную (транзакционную) обработку данных (OLTP)

Информационные системы, ориентированные на операционную (транзакционную) обработку данных, называют термином «ON-Line Transaction Processing, OLTP», что в переводе означает «оперативная транзакционная обработка данных».

В соответствии с современными требованиями к OLTP должны удовлетворять следующим показателям:

**Производительность и масштабируемость**, позволяющие предприятиям создавать БД, обеспечивающие эффективность применения информационных систем;

**высокая доступность данных**, обеспечивающая непрерывность функционирования приложений, работающих с БД и минимум административных издержек;

**безопасность хранения данных**, обеспечиваемая шифрованием важных данных, аудитом модификации данных и метаданных, применением внешних криптографических ключей;

**управляемость данных**, позволяющая снизить затраты времени и средств на управление инфраструктурой обработки и хранения данных за счет применения инновационных технологий администрирования

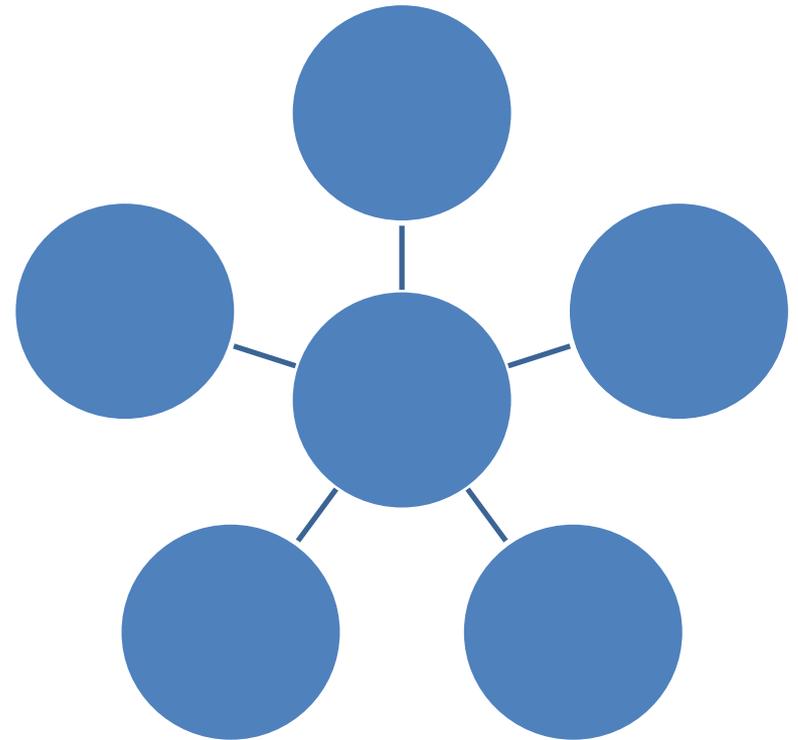
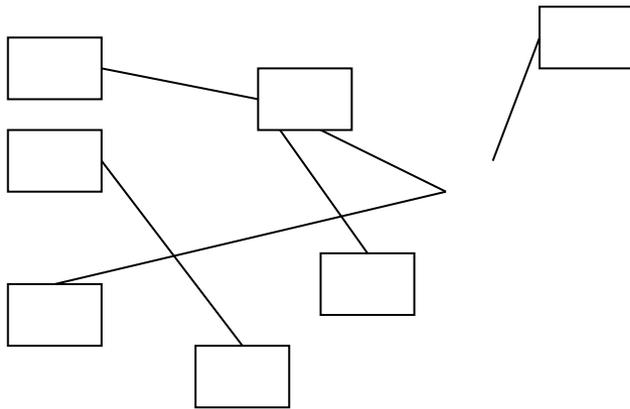
### 3 Информационные системы оперативного анализа данных (OLAP)

Накопление больших объемов данных в последнее время сделали актуальными прикладные задачи, предназначенные для извлечения, сбора и представления конечному пользователю информации, необходимой для анализа текущего состояния дел и прогноза будущего решения. Такие ИС получили название **систем поддержки принятия решений**. Исторически первыми такими системами стали **ИС руководителя** (EIS — Executive Information Systems).

Существует два подхода к интеграции корпоративной информации:

- децентрализованное объединение источников (схема спагетти) (рис.2а)
- централизованное объединение источников (рис.2б)

## Рисунок 2 - Подходы к интеграции корпоративной информации



децентрализованное объединение источников (схема спагетти)  
(рис.2а)

централизованное объединение источников (рис.2б)

**Хранилища данных (Datawarehouse) и оперативный анализ данных (On-Line Analytical Processing, OLAP) – новые информационные технологии,** которые обеспечивают аналитикам, управленцам и руководителям высшего звена возможность изучать большие объемы взаимосвязанных данных при помощи быстрого интерактивного отображения информации на разных уровнях детализации с различных точек зрения в соответствии с представлениями пользователя о предметном пространстве.

Основная цель хранилищ — создание единого логического представления данных, содержащихся в разнотипных БД или в единой модели корпоративных данных.

**Хранилище данных** — ориентированная на поддержку управленческих решений автоматизированная система, состоящая из организационной структуры, технических средств, базы или совокупности базы данных (БД) и ПО, которое выполняет, как правило, следующие функции:

- извлечение данных из разрозненных источников, их трансформация и загрузка в хранилище;
- администрирование данных и хранилища;
- извлечение данных из хранилища, аналитическая обработка и представление данных конечным пользователям

## Основные требования к хранилищам данных:

- поддержка высокой скорости получения данных из хранилища;
- поддержка внутренней непротиворечивости данных;
- возможность получения и сравнения так называемых срезов данных (slice and dice);
- наличие удобных утилит просмотра данных в хранилище;
- полнота и достоверность хранимых данных;
- поддержка качественного процесса пополнения данных.

В соответствии с данной концепцией **хранилище данных** содержит данные, поступающие от разных источников, и интегрированные данные, получаемые в результате обработки первичных данных. Кроме того, для поддержки концепции ХД требуются специальные средства управления процессами хранения и обработки данных.

Концепция OLAP (On-line analytical processing) разработана автором реляционных БД Е.Ф.Содд в 1993 году. В 1995 году на основе требований, изложенных Коддом, сформулирован тест FASMI (Fast Analysis of Shared Multidimensional Information), который переводится как «быстрый анализ разделяемой многомерной информации».

Тест FASMI включает следующие требования к приложениям для многомерного анализа [2]:

- предоставление пользователю результатов анализа за приемлемое время (не более 5 с) при допустимом уровне детализации анализа;
- возможность осуществления любого логического и статистического анализа, поддерживаемого используемым приложением, и сохранением результатов в доступном для пользователя виде;
- многопользовательский доступ к данным с поддержкой соответствующих механизмов блокировки и средств автоматизированного доступа;
- многомерное концептуальное представление данных, включая полную поддержку для иерархий и множественных иерархий измерений (ключевое требование OLAP);
- возможность обращаться к любой нужной информации независимо от её объёма и места хранения.

## 4 Структура и задачи интеллектуального анализа данных

Интеллектуальный анализ данных (ИАД) определяется как «извлечение зёрен знаний из гор данных» или «разработка данных – по аналогии с разработкой полезных ископаемых» [2]. В английском языке существует два термина, переводимые как ИАД: Knowledge Discovery in Databases (KDD) и Data Mining (DM). Таким образом, ИАД рассматривается как процесс поддержки принятия решений, основанный на поиске в данных скрытых закономерностей [1, 2].

Модели интеллектуального анализа данных могут применяться в условиях торгово-закупочной, производственной и другой деятельности:

- прогнозирования продаж и определения продуктов, которые с высокой долей вероятности могут быть проданы вместе;
- определения продуктов, которые с высокой долей вероятности могут быть проданы вместе;
- выявление последовательностей в том порядке, в котором клиенты добавляют продукты в корзину для покупок;
- контроля и управления учебной, научной и воспитательной работой преподавателей учебного заведения.

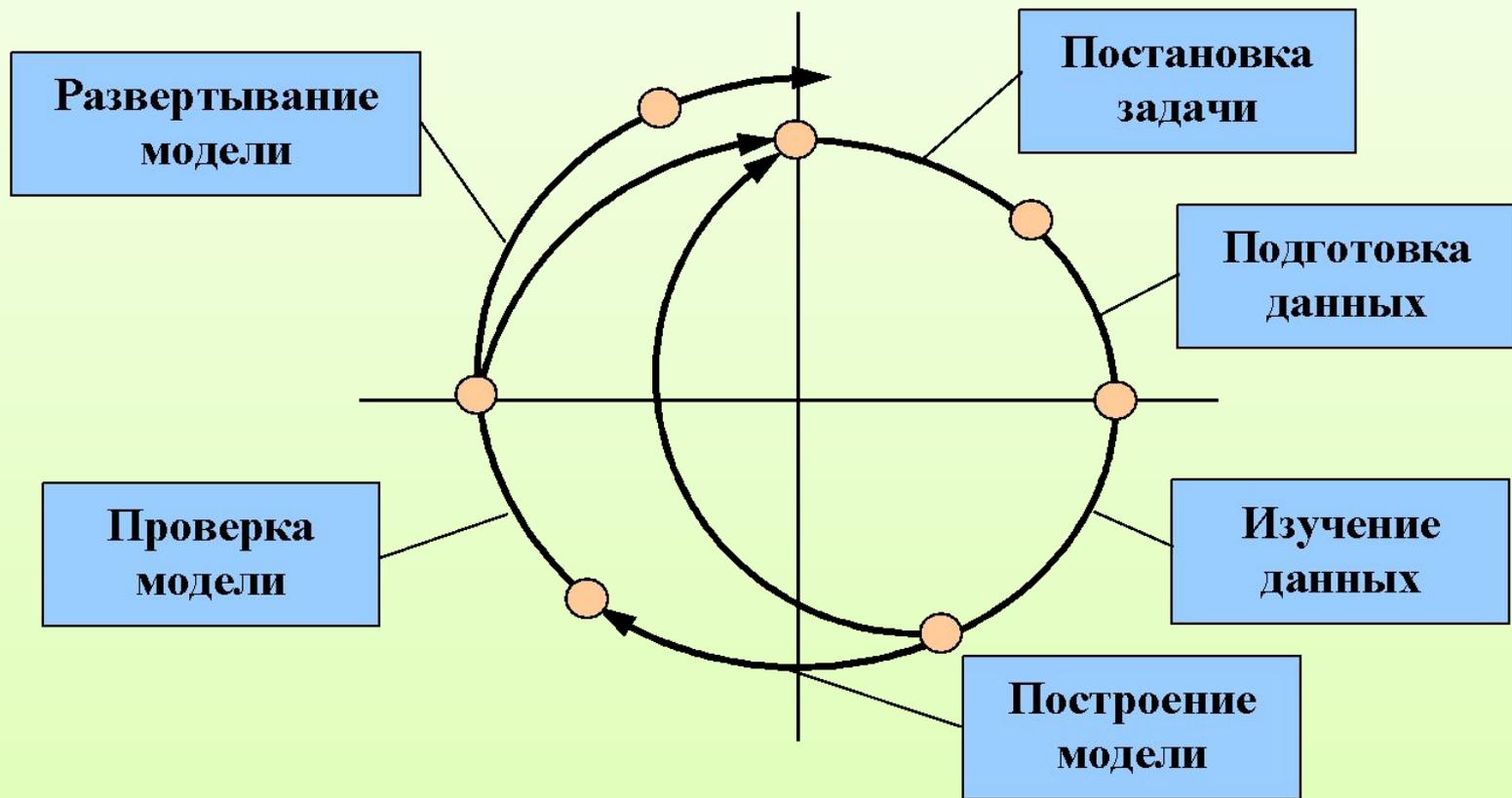


Рисунок.3 – Диаграмма анализа данных

Основными задачами ИАД являются:

- **Классификация** (Classification). Наиболее распространенная задача ИАД. В результате решения задачи классификации обнаруживаются признаки, которые характеризуют группы объектов (классы). По этим признакам новый объект можно отнести к тому или иному классу. Для решения задачи классификации могут использоваться методы: ближайшего соседа (Nearest Neighbor); k-ближайшего соседа (k-Nearest Neighbor);
- **Кластеризация** (Clustering). Кластеризация является логическим продолжением идеи классификации. Это задача более сложная, особенность кластеризации заключается в том, что классы объектов изначально не predeterminedены. Результатом кластеризации является разбиение объектов на группы.
- **Ассоциация** (Associations). В ходе решения задачи поиска ассоциативных правил отыскиваются закономерности между связанными событиями в наборе данных. Отличие ассоциации от двух предыдущих задач Data Mining: поиск закономерностей осуществляется не на основе свойств анализируемого объекта, а между несколькими событиями, которые происходят одновременно. Наиболее известный алгоритм решения задачи поиска ассоциативных правил – алгоритм Apriori.

- **Последовательность** (Sequence). Последовательность позволяет найти временные закономерности между транзакциями. Задача последовательности подобна ассоциации, но ее целью является установление закономерностей не между одновременно наступающими событиями, а между событиями, связанными во времени. Другими словами, последовательность определяется высокой вероятностью цепочки связанных во времени событий. Фактически, ассоциация является частным случаем последовательности с временным лагом, равным нулю.
- **Прогнозирование** (Forecasting). В результате решения задачи прогнозирования на основе особенностей исторических данных оцениваются пропущенные или же будущие значения целевых численных показателей. Для решения таких задач широко применяются методы математической статистики..
- **Анализ отклонений** (Deviation Detection). Данная задача решается с целью обнаружение и анализ данных, наиболее отличающихся от общего множества данных, т.е выявления нехарактерных шаблонов.

# Контрольные вопросы

1. Приведите структуру и поясните назначение элементов системы поддержки принятия решений.
2. Перечислите назначение хранилищ данных и витрин данных.
3. Дайте определение системы поддержки принятия решений и охарактеризуйте существующие подходы к интеграции корпоративной информации.
4. Охарактеризуйте требования, предъявляемые к OLTP-системам.
5. Дать определение хранилища данных и перечислить его основные функции.
6. Перечислите и поясните основные требования, предъявляемые к хранилищам данных.
7. Перечислите и поясните существующие подходы к интеграции корпоративной информации
8. Дайте определение интеллектуального анализа данных и приведите его диаграмму.
9. Приведите и поясните основные задачи интеллектуального анализа данных.