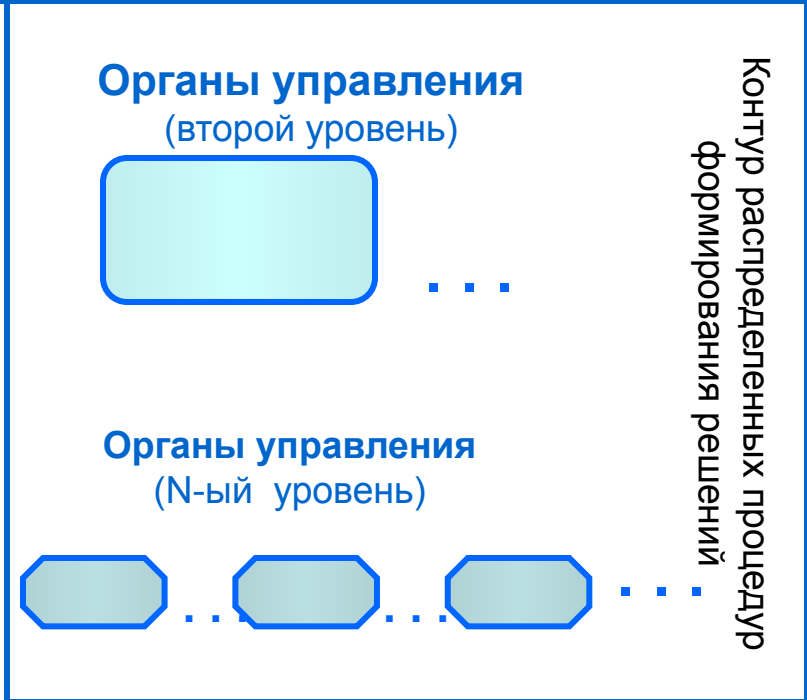
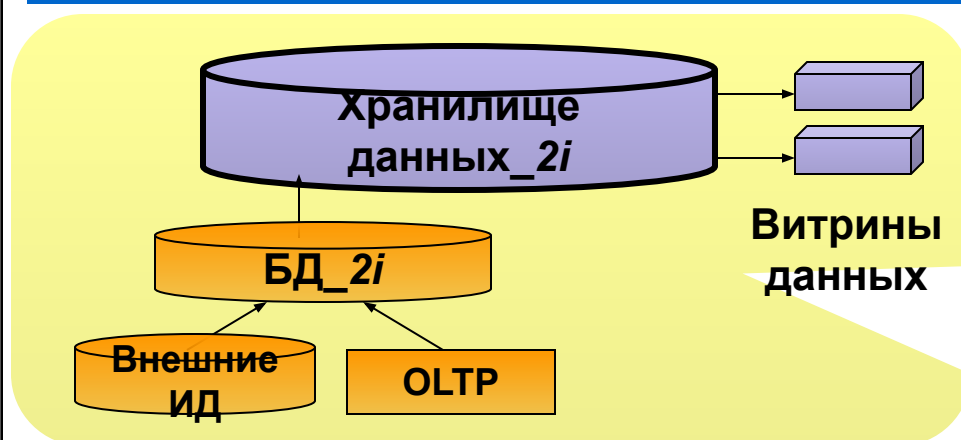
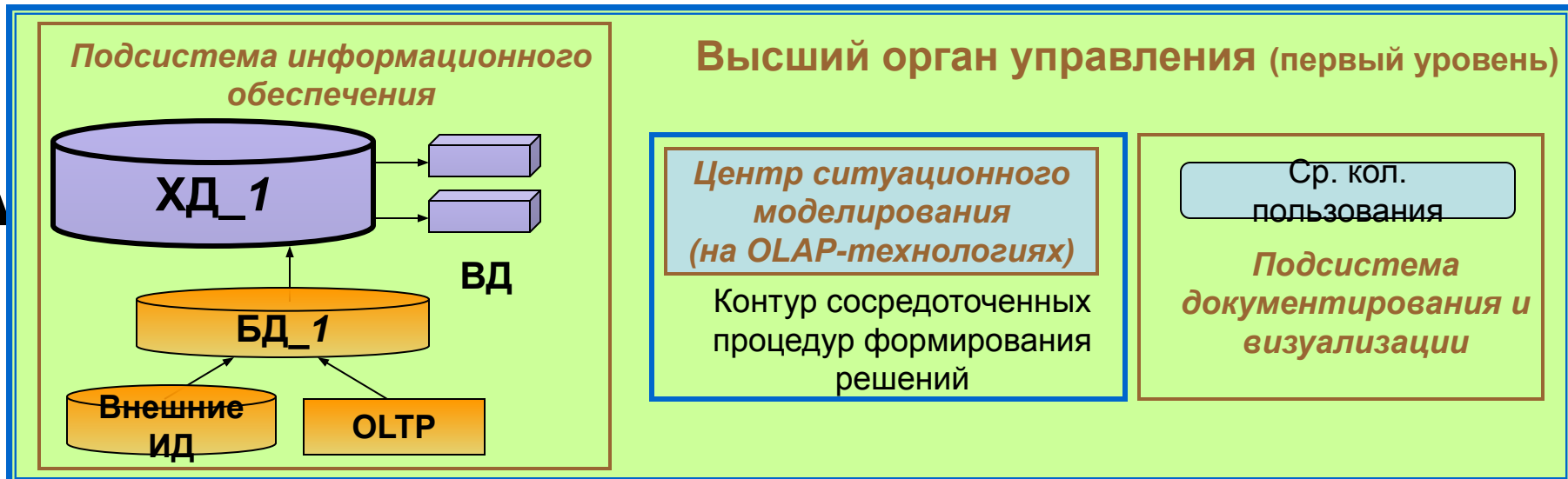


# Базы и хранилища данных

# Построение единого информационного пространства в административно-технических системах

Подъем и агрегирование информации



# Информационная поддержка в системах автоматизированных исследований



**Подсистема ввода данных.** В таких подсистемах, называемых OLTP (Online transaction processing), реализуется операционная (транзакционная) обработка данных. Для их реализации используют обычные системы управления базами данных (СУБД).

**Подсистема анализа.** Данная подсистема может быть построена на основе:

- подсистемы информационно-поискового анализа на базе реляционных СУБД и статических запросов с использованием языка SQL (Structure Query Language);
- подсистемы оперативного анализа. Для реализации таких подсистем применяется технология оперативной аналитической обработки данных OLAP (On-line analytical processing), использующая концепцию многомерного представления данных;
- подсистемы интеллектуального анализа. Данная подсистема реализует методы и алгоритмы Data Mining ("добыча данных").

# Реляционные базы данных \_1

Реляционный подход известен благодаря работам Е. Кодда, опубликованных в 1970 году, в которых он сформулировал следующие **12** правил для реляционной БД:

**Данные представляются в виде таблиц** — БД представляет собой набор таблиц. Таблицы хранят данные, сгруппированные в виде рядов и колонок. Ряд представляет собой набор значений, относящихся только к одному объекту, хранящемуся в таблице, и называется записью. Колонка представляет собой одну характеристику для всех объектов, хранящихся в таблице, и называется полем.

**Данные доступны логически** — реляционная модель не позволяет обращаться к данным физически, адресуя ячейки по номерам колонки и ряда нет возможности получить значение в ячейке колонка 2, ряд 3). Доступ к данным возможен только через идентификаторы таблицы, колонки и ряда.

**NULL трактуется как неизвестное значение** — если в ячейку таблицы значение не введено, то записывается NULL.

**БД должна включать в себя метаданные** — БД хранит два вида таблиц: пользовательские таблицы и системные таблицы. В пользовательских таблицах хранятся данные, введенные пользователем. В системных таблицах хранятся метаданные: описание таблиц (название, типы и размеры колонок), индексы, хранимые процедуры и др.

**Должен использоваться единый язык для взаимодействия с СУБД** — для управления реляционной БД должен использоваться единый язык. В настоящее время таким инструментом стал язык структурных запросов SQL.

**СУБД должна обеспечивать альтернативный вид отображения данных** — СУБД не должна ограничивать пользователя только отображением таблиц, которые существуют. Пользователь должен иметь возможность строить виртуальные таблицы — представления (View).

**Должны поддерживаться операции реляционной алгебры** — записи реляционной БД трактуются как элементы множества, на котором определены операции реляционной алгебры. СУБД должна обеспечивать выполнение этих операций. В настоящее время выполнение этого правила обеспечивает язык SQL.

# Реляционные базы данных\_2

Реляционный подход известен благодаря работам Е. Кодда, опубликованных в 1970 году, в которых он сформулировал следующие **12** правил для реляционной БД:

**Должна обеспечиваться независимость от физической организации данных** — приложения, оперирующие с данными реляционных БД, не должны зависеть от физического хранения данных (от способа хранения, формата хранения и др.).

**Должна обеспечиваться независимость от логической организации данных** — приложения, оперирующие с данными реляционных БД, не должны зависеть от организации связей между таблицами (логической организации). При изменении связей между таблицами не должны меняться ни сами таблицы, ни запросы к ним.

**За целостность данных отвечает СУБД** — под целостностью данных в общем случае понимается готовность БД к работе. Различают следующие типы целостности:

*физическая целостность* — сохранность информации на носителях и корректность форматов хранения данных;

*логическая целостность* — непротиворечивость и актуальность данных, хранящихся в БД.

Различают два способа обеспечения целостности: декларативный и процедурный.

**Целостность данных не может быть нарушена** — СУБД должна обеспечивать целостность данных при любых манипуляциях, производимых с ними

**Должны поддерживать распределенные операции** — реляционная БД может размещаться как на одном компьютере, так и на нескольких — распределенно. Пользователь должен иметь возможность связывать данные, находящиеся в разных таблицах и на разных узлах компьютерной сети.

## Теория нормализации данных (исключение избыточности):

**БД имеет 1-ю НФ**, если каждое значение, хранящееся в ней, неразделимо на более примитивные (неразложимость значений);

**БД имеет 2-ю НФ**, если она имеет 1-ю НФ, и при этом каждое значение целиком и полностью зависит от ключа (функционально независимые значения);

**БД имеет 3-ю НФ**, если она имеет 2-ю НФ, и при этом ни одно из значений не предоставляет никаких сведений о другом значении (взаимно независимые значения) и т. д.

# Основные понятия ER-диаграмм

## Определение 1.

**Сущность** - это класс однотипных объектов, информация о которых должна быть учтена в модели.

Каждая сущность должна иметь наименование, выраженное существительным в единственном числе.

## Определение 2.

**Экземпляр сущности** - это конкретный представитель данной сущности.

## Определение 3.

**Атрибут сущности** - это именованная характеристика, являющаяся некоторым свойством сущности.

Наименование атрибута должно быть выражено существительным в единственном числе (возможно, с характеризующими прилагательными)

## Определение 4.

**Ключ сущности** - это **неизбыточный** набор атрибутов, значения которых в совокупности являются **уникальными** для каждого экземпляра сущности.

Неизбыточность заключается в том, что удаление любого атрибута из ключа нарушается его уникальность.

Сущность может иметь несколько различных ключей.

## Определение 5.

**Связь** - это некоторая **ассоциация** между **двумя** сущностями.

Одна сущность может быть связана с другой сущностью или сама с собою

# Нормализация данных

**Первая нормальная форма.** Сущность находится в 1НФ, если все ее атрибуты являются простыми (имеют единственное значение).

**Вторая нормальная форма.** Сущность находится в 2НФ, если она находится в 1НФ и каждый неключевой атрибут функционально полно зависит от первичного ключа (составного)

**Третья нормальная форма.**

Определение 1. Сущность находится в 3НФ, если она находится в 2НФ и каждый неключевой атрибут нетранзитивно зависит от первичного ключа.

Существует и альтернативное определение.

Определение 2. Сущность находится в 3НФ в том и только в том случае, если все неключевые атрибуты сущности взаимно независимы и полностью зависят от первичного ключа

# Нормализация данных





# Определение OLAP-системы\_1

**OLAP (On-Line Analytical Processing)** — технология оперативной аналитической обработки данных, использующая методы и средства для сбора, хранения и анализа многомерных данных в целях поддержки процессов принятия решений.

**Основное назначение OLAP-систем** — поддержка аналитической деятельности, произвольных (часто используется термин ad-hoc) запросов пользователей-аналитиков.

**Цель OLAP-анализа** — проверка возникающих гипотез.

В 1993 г. Э. Кодд опубликовал статью под названием "**OLAP для пользователей-аналитиков: каким он должен быть**", в которой изложил основы концепции оперативной аналитической обработки и определил **12 требований**, которым должны удовлетворять **продукты**, позволяющие выполнять **оперативную аналитическую обработку**:

**1. Многомерность** — OLAP-система на концептуальном уровне должна представлять данные в виде многомерной модели, что упрощает процессы анализа и восприятия информации.

**2. Прозрачность** — OLAP-система должна скрывать от пользователя реальную реализацию многомерной модели, способ организации, источники, средства обработки и хранения.

**3. Доступность** — OLAP-система должна предоставлять пользователю единую, согласованную и целостную модель данных, обеспечивая доступ к данным независимо от того, как и где они хранятся.

**4. Постоянная производительность при разработке отчетов** — производительность OLAP-систем не должна значительно уменьшаться при увеличении количества измерений, по которым выполняется анализ.

# Определение OLAP-системы\_2

В 1993 г. Э. Кодд опубликовал статью под названием "**OLAP для пользователей-аналитиков: каким он должен быть**", в которой изложил основы концепции оперативной аналитической обработки и определил **12 требований**, которым должны удовлетворять **продукты**, позволяющие выполнять **оперативную аналитическую обработку**:

**5. Клиент-серверная архитектура** — OLAP-система должна быть способна работать в среде "клиент-сервер", т. к. большинство данных, которые сегодня требуется подвергать оперативной аналитической обработке, хранятся распределенно.

**6. Равноправие измерений** — OLAP-система должна поддерживать многомерную модель, в которой все измерения равноправны.

**7. Динамическое управление разреженными матрицами** — OLAP-система должна обеспечивать оптимальную обработку разреженных матриц. Скорость доступа должна сохраняться вне зависимости от расположения ячеек данных и быть постоянной величиной для моделей, имеющих разное число измерений и различную степень разреженности данных.

**8. Поддержка многопользовательского режима** — OLAP-система должна предоставлять возможность работать нескольким пользователям совместно с одной аналитической моделью или создавать для них различные модели из единых данных. При этом возможны как чтение, так и запись данных, поэтому система должна обеспечивать их целостность и безопасность.

**9. Неограниченные перекрестные операции** — OLAP-система должна обеспечивать сохранение функциональных отношений, описанных с помощью определенного формального языка между ячейками гиперкуба при выполнении любых операций среза, вращения, консолидации или детализации. Система должна самостоятельно (автоматически) выполнять преобразование установленных отношений, не требуя от пользователя их переопределения.

# Определение OLAP-системы\_3

В 1993 г. Э. Кодд опубликовал статью под названием "**OLAP для пользователей-аналитиков: каким он должен быть**", в которой изложил основы концепции оперативной аналитической обработки и определил **12 требований**, которым должны удовлетворять **продукты**, позволяющие выполнять **оперативную аналитическую обработку**:

**10. Интуитивная манипуляция данными**— OLAP-система должна предоставлять способ выполнения операций среза, вращения, консолидации, детализации над гиперкубом без необходимости пользователю совершать множество действий с интерфейсом. Измерения, определенные в аналитической модели, должны содержать всю необходимую информацию для выполнения вышеуказанных операций.

**11. Гибкие возможности получения отчетов** — OLAP-система должна поддерживать различные способы визуализации данных, т. е. отчеты должны представляться в любой возможной ориентации. Средства формирования отчетов должны представлять синтезируемые данные или информацию, следующую из модели данных в ее любой возможной ориентации.

**12. Неограниченная размерность и число уровней агрегации** — исследование о возможном числе необходимых измерений, требующихся в аналитической модели, показало, что одновременно может использоваться до 19 измерений. Отсюда вытекает настоятельная рекомендация, чтобы аналитический инструмент мог одновременно предоставить хотя бы 15, а предпочтительно — 20 измерений. Более того, каждое из общих измерений не должно быть ограничено по числу определяемых пользователем-аналитиком уровней агрегации и путей консолидации.

# Определение OLAP-системы\_4

В 1995 г. Э. Кодд **12** требованиям, которым должны удовлетворять продукты, позволяющие выполнять оперативную аналитическую обработку, добавил следующие **6**:

**13. *Пакетное извлечение против интерпретации*** — OLAP-система должна в равной степени эффективно обеспечивать доступ как к собственным, так и к внешним данным.

**14. *Поддержка всех моделей OLAP-анализа*** — OLAP-система должна поддерживать все четыре модели анализа данных, определенные Коддом: категориальную, толковательную, умозрительную и стереотипную.

**15. *Обработка ненормализованных данных*** — OLAP-система должна быть интегрирована с ненормализованными источниками данных. Модификации данных, выполненные в среде OLAP, не должны приводить к изменениям данных, хранимых в исходных внешних системах.

**16. *Сохранение результатов OLAP***: хранение их отдельно от исходных данных — OLAP-система, работающая в режиме чтения-записи, после модификации исходных данных должна результаты сохранять отдельно. Иными словами, обеспечивается безопасность исходных данных.

**17. *Исключение отсутствующих значений***— OLAP-система, представляя данные пользователю, должна отбрасывать все отсутствующие значения. Другими словами, отсутствующие значения должны отличаться от нулевых значений.

**18. *Обработка отсутствующих значений*** — OLAP-система должна игнорировать все отсутствующие значения без учета их источника. Эта особенность связана с 17-м правилом.

# Определение OLAP-системы\_5

Кодд все 18 требований сгруппировал в четыре группы **B**, **S**, **R** и **D**, назвав их особенностями OLAP систем:

**Основные особенности (B)** включают следующие правила:

- многомерное концептуальное представление данных (правило 1);
- интуитивное манипулирование данными (правило 10);
- доступность (правило 3);
- пакетное извлечение против интерпретации (правило 13);
- поддержка всех моделей OLAP-анализа (правило 14);
- архитектура "клиент-сервер" (правило 5);
- прозрачность (правило 2);
- многопользовательская поддержка (правило 8).

**Специальные особенности (S):**

- обработка ненормализованных данных (правило 15);
- сохранение результатов OLAP: хранение их отдельно от исходных данных (правило 16);
- исключение отсутствующих значений (правило 17);
- обработка отсутствующих значений (правило 18).

**Особенности представления отчетов (R):**

- гибкость формирования отчетов (правило 11);
- стандартная производительность отчетов (правило 4);
- автоматическая настройка физического уровня (измененное правило 7).

**Управление измерениями (D):**

- универсальность измерений (правило 6);
- неограниченное число измерений и уровней агрегации (правило 12);
- неограниченные операции между размерностями (правило 9).

# Определение OLAP-системы\_6

На основе анализа правил Кодда, Найджелом Пендсом (Nigel Pendse) и Ричардом Критом (Richard Creeth) в 1995 г. были сформулированы **мечм FASMI (Fast of Shared Multidimensional Information)**, который определил OLAP-системы следующими пятью ключевыми признаками: Fast (Быстрый), Analysis (Анализ), Shared (Разделяемой), Multidimensional (Многомерной), Information (Информации):

**FAST** (Быстрый) — OLAP-система должна обеспечивать выдачу большинства ответов пользователям в пределах приблизительно 5 с. При этом самые простые запросы обрабатываются в течение 1 с, и очень немногие более 20 с.

**ANALYSIS** (Анализ) — OLAP-система должна справляться с любым логическим и статистическим анализом, характерным для данного приложения, и обеспечивать его сохранение в виде, доступном для конечного пользователя.

**SHARED** (Разделяемой) — OLAP-система должна выполнять все требования защиты конфиденциальности (возможно, до уровня ячейки хранения данных). Если множественный доступ для записи необходим, обеспечивается блокировка модификаций на соответствующем уровне.

**MULTIDIMENSIONAL** (Многомерной) — OLAP-система должна обеспечить многомерное концептуальное представление данных, включая полную поддержку для иерархий и множественных иерархий, обеспечивающих наиболее логичный способ анализа.

**INFORMATION** (Информации) — OLAP-система должна обеспечивать получение необходимой информации в условиях реального приложения. Мощность различных систем измеряется не объемом хранимой информации, а количеством входных данных, которые они могут обработать. В этом смысле мощность продуктов весьма различна. Большие OLAP-системы могут оперировать по крайней мере в 1000 раз большим количеством данных по сравнению с простыми версиями OLAP-систем.

# Сопоставление систем: OLTP и OLAP\_1

<b>Характеристика</b>	<b>Требования к OLTP-системе</b>	<b>Требования к системе анализа</b>
<b><i>Степень детализации хранимых данных</i></b>	Хранение только детализированных данных	Хранение как детализированных, так и обобщенных данных
<b><i>Качество данных</i></b>	Допускаются неверные данные из-за ошибок ввода	Не допускаются ошибки в данных
<b><i>Формат хранения данных</i></b>	Может содержать данные в разных форматах в зависимости от приложений	Единый согласованный формат хранения данных
<b><i>Допущение избыточных данных</i></b>	Должна обеспечиваться максимальная нормализация	Допускается контролируемая денормализация (избыточность) для эффективного извлечения данных
<b><i>Управление данными</i></b>	Должна быть возможность в любое время добавлять, удалять и изменять данные	Должна быть возможность периодически добавлять данные

# Сопоставление систем: OLTP и OLAP\_2

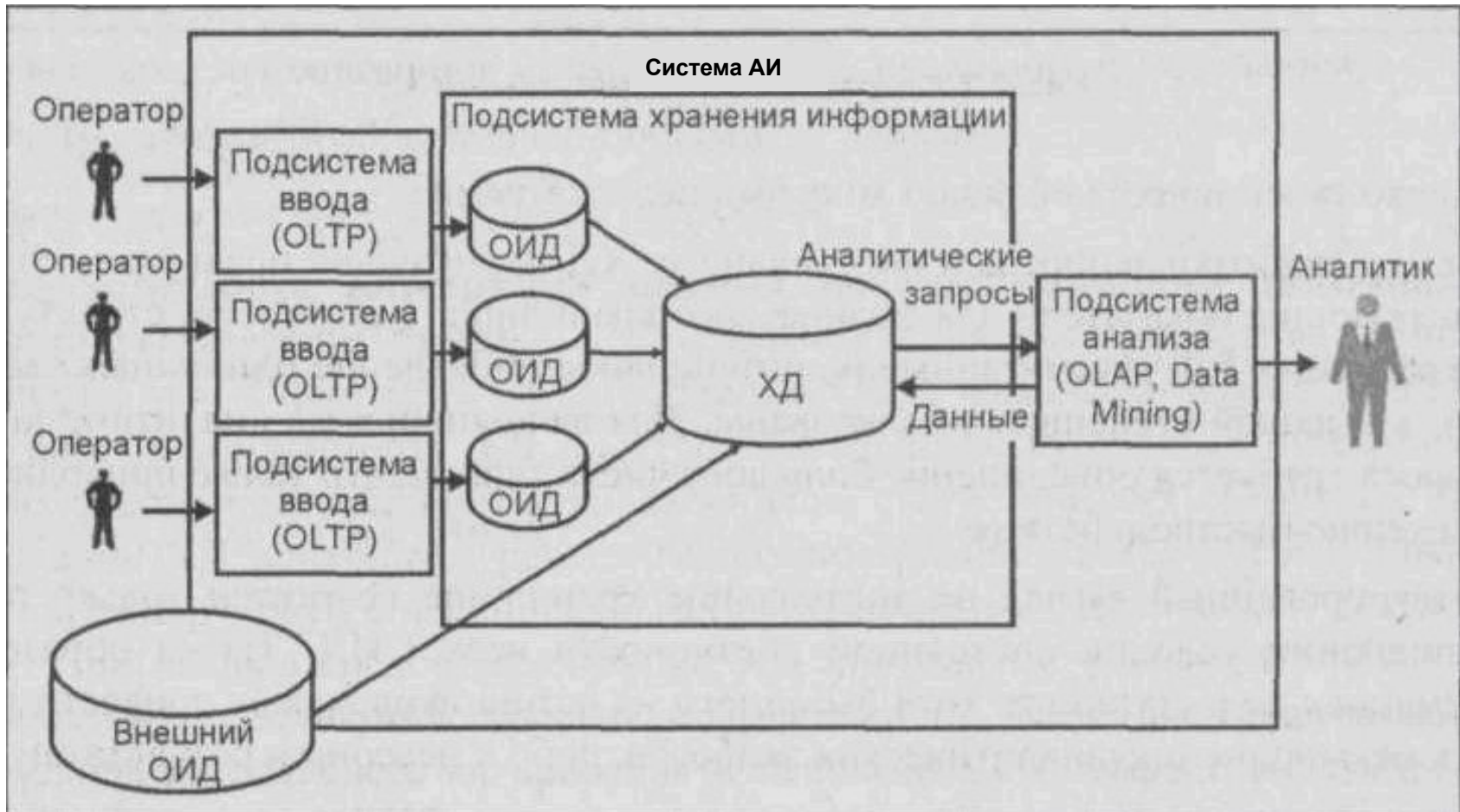
<b>Характеристика</b>	<b>Требования к OLTP-системе</b>	<b>Требования к системе анализа</b>
<i>Количество хранимых данных</i>	Должны быть доступны все оперативные данные, требующиеся в данный момент	Должны быть доступны все данные, накопленные в течение продолжительного интервала времени
<i>Характер запросов к данным</i>	Доступ к данным пользователей осуществляется по заранее составленным запросам	Запросы к данным могут быть произвольные и заранее не оформлены
<i>Время обработки обращений к данным</i>	Время отклика системы измеряется в секундах	Время отклика системы может составлять несколько минут
<i>Характер вычислительной нагрузки на систему</i>	Постоянно средняя загрузка процессора	Загрузка процессора формируется только при выполнении запроса, но на 100 %
<i>Приоритетность характеристик системы</i>	Основными приоритетами являются высокая производительность и доступность	Приоритетными являются обеспечение гибкости системы и независимости работы пользователей



# Концепция хранилища данных\_1

Систематизированное изложение концепции хранилища данных было сделано в 1992 г. Уильман Г. Инмоном в монографии *"Построение хранилищ данных"*.

*Хранилище данных* — предметно-ориентированный, интегрированный, неизменяемый, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений. В основе концепции ХД лежит идея разделения данных, используемых для оперативной обработки и для решения задач анализа.

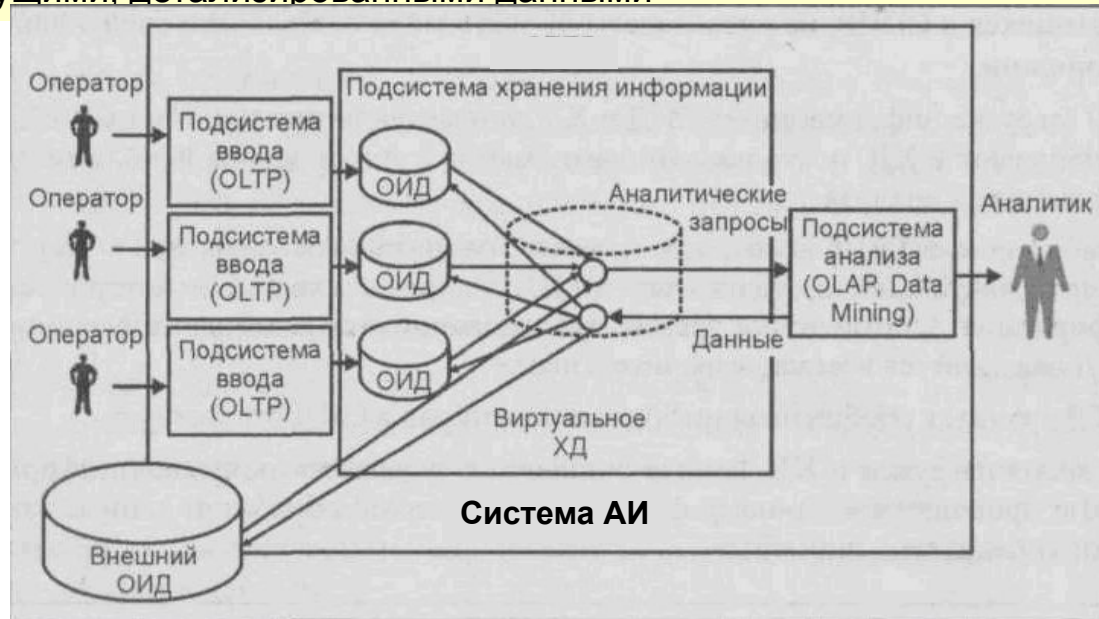


**Структура системы автоматизированных исследований с физическим ХД**

# Концепция хранилища данных\_2

**Виртуальные ХД.** Основными **достоинствами** виртуального ХД являются:

- минимизация объема памяти, занимаемой на носителе информацией;
- работа с текущими, детализированными данными



**Структура системы автоматизированных исследований с виртуальным ХД**

**Недостатки:**

- время обработки запросов к виртуальному ХД значительно превышает соответствующие показатели для физического хранилища;
- интегрированный взгляд на виртуальное хранилище возможен только при выполнении условия постоянной доступности всех ОИД;
- выполнение сложных аналитических запросов над ОИД занимает большой объем ресурсов компьютеров, на которых они работают;
- различные ОИД могут поддерживать разные форматы и кодировки данных. Часто на один и тот же вопрос может быть получено несколько вариантов ответа. Это может быть связано с не синхронностью моментов обновления данных в разных ОИД;
- главный недостаток виртуального хранилища – практическая невозможность получения данных за долгий период времени

# Концепция хранилища данных\_3

## **Основные проблемы создания ХД:**

- необходимость интеграции данных из неоднородных источников в распределенной среде;
- потребность в эффективном хранении и обработке очень больших объемов информации;
- необходимость наличия многоуровневых справочников метаданных;
- повышенные требования к безопасности данных

**Витрина данных (ВД)** — это упрощенный вариант ХД, содержащий только тематически объединенные данные

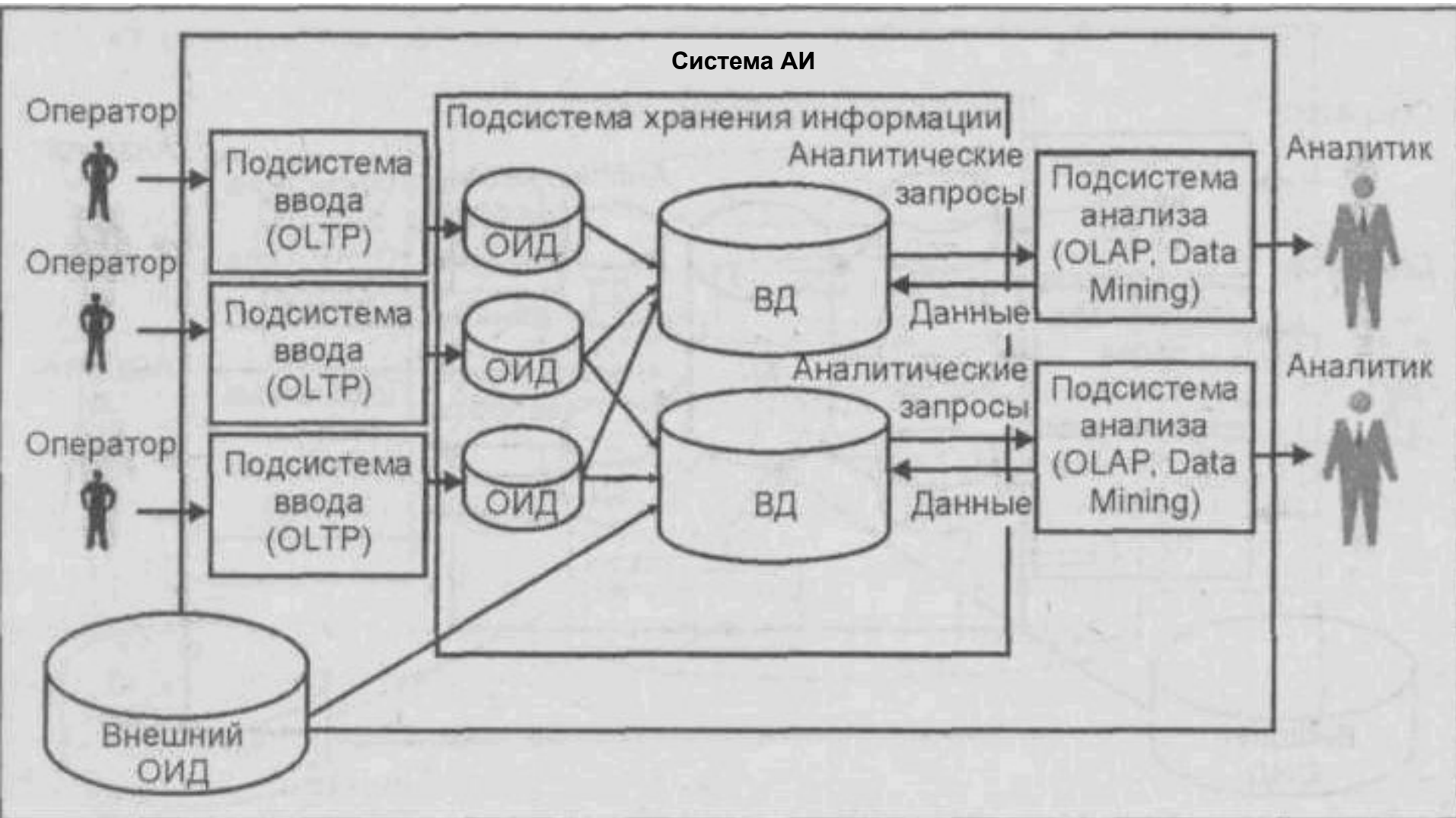
## **Достоинствами ВД:**

- проектирование ВД для ответов на определенный круг вопросов;
- быстрое внедрение автономных ВД и получение отдачи;
- упрощение процедур заполнения ВД и повышение их производительности за счет учета потребностей определенного круга пользователей

## **Недостатками автономных ВД:**

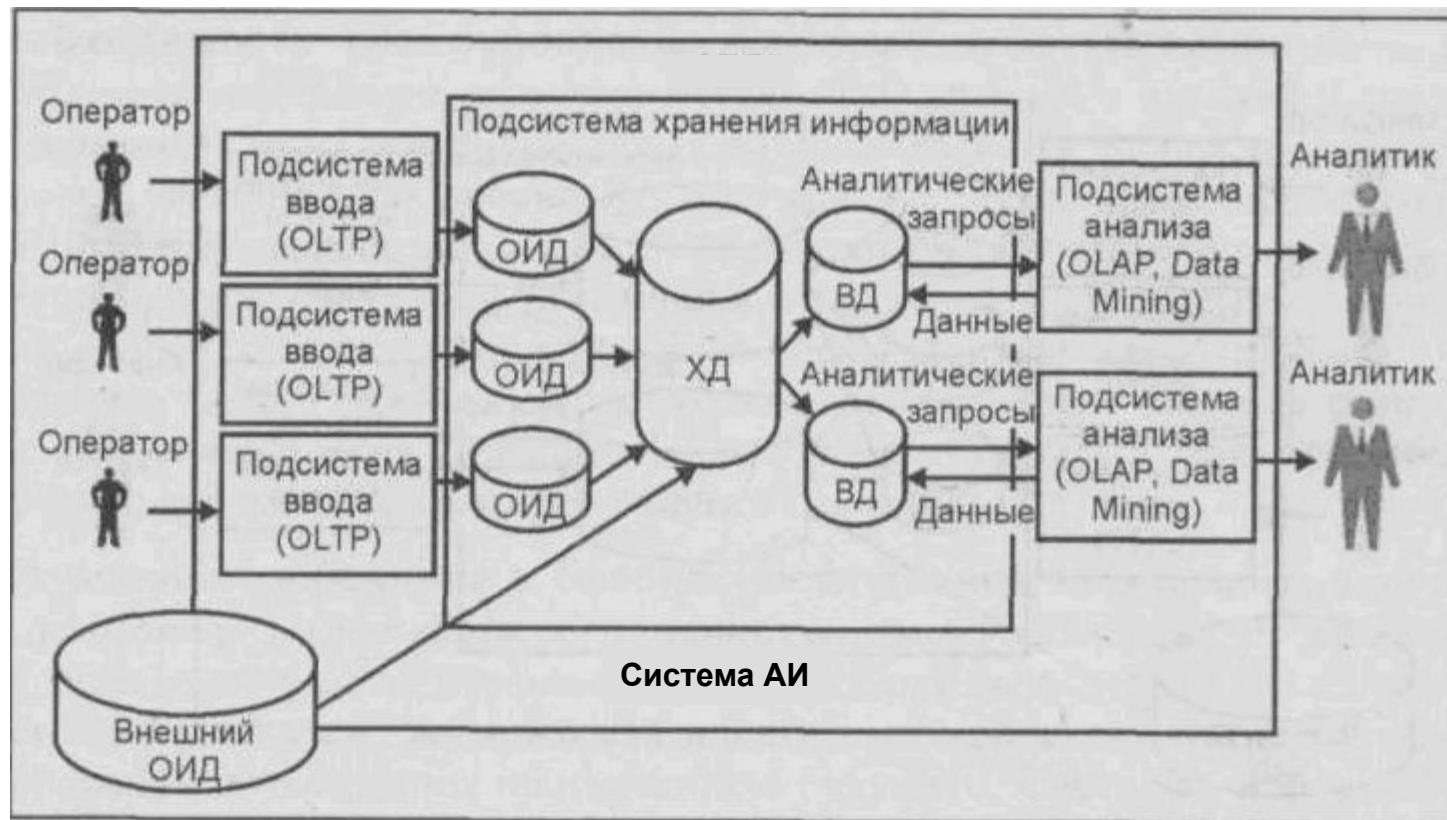
- многократное хранение данных в разных ВД, что приводит к увеличению расходов на их хранение и потенциальным проблемам, связанным с необходимостью поддержания непротиворечивости данных;
- отсутствие консолидированности данных на уровне предметной области, а следовательно — отсутствие единой картины

# Концепция хранилища данных\_4



*Структура системы автоматизированных исследований с самостоятельными ВД*

# Концепция хранилища данных\_5



**Структура системы AI с самостоятельными ХД и ВД**

## **Достоинства:**

- простота создания и наполнения ВД, поскольку наполнение происходит из единого стандартизованного надежного источника очищенных данных — из ХД;
- простота расширения системы AI за счет добавления новых ВД;
- снижение нагрузки на основное ХД.

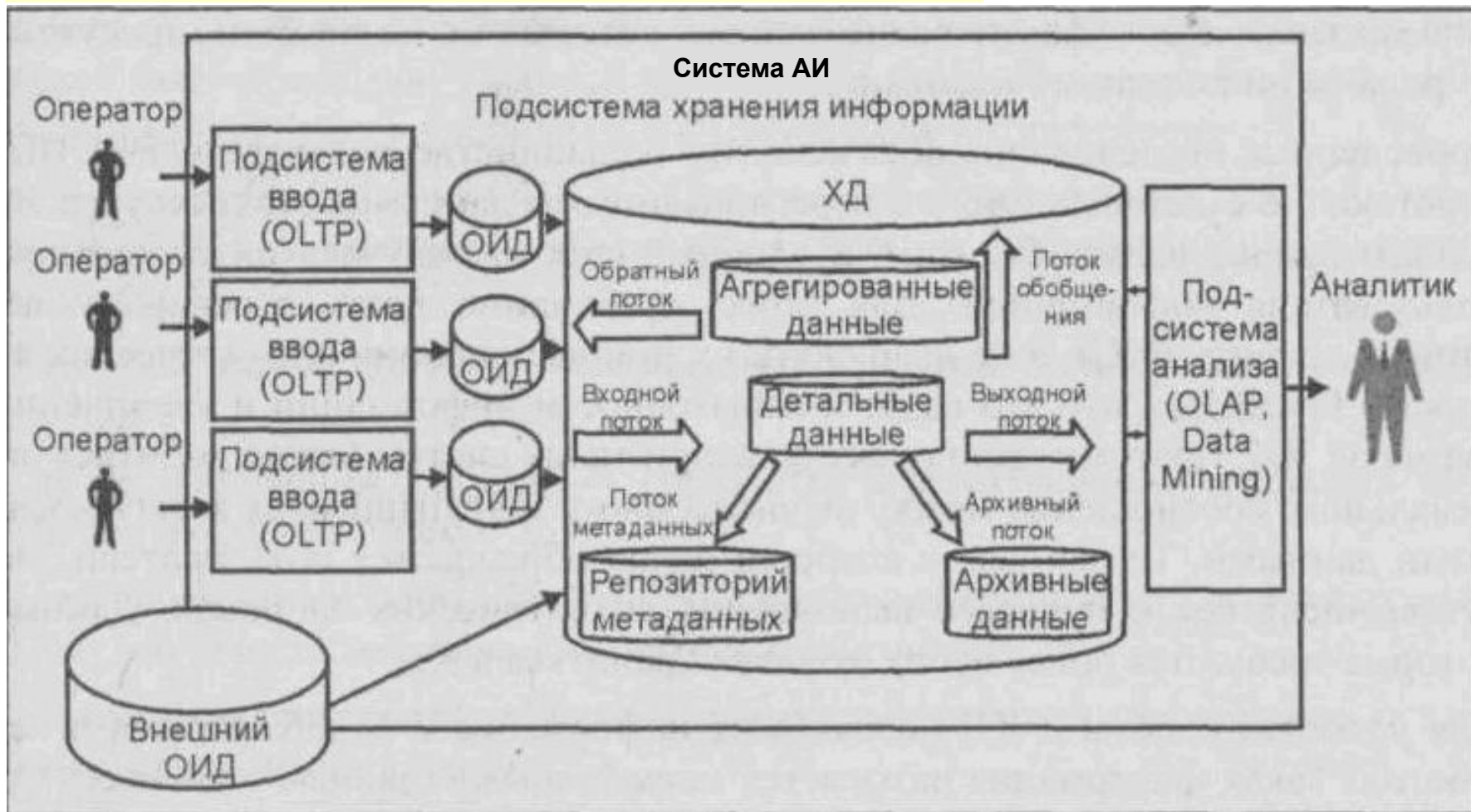
## **Недостатки:**

- избыточность (данные хранятся как в ХД, так и в ВД);
- дополнительные затраты на разработку системы AI с ХД и ВД.

# Организация хранилищ данных

Все данные в ХД делятся на **три** основные **категории**:

- детальные данные;
- агрегированные данные;
- метаданные



Архитектура ХД

# Репозиторий метаданных

Для удобства работы с ХД необходима информация о содержащихся в нем данных. Такая информация называется **метаданными (данные о данных)**.

Метаданные должны отвечать на следующие вопросы — **что, кто, где, как, когда и почему**:

- **что** (описание объектов) — метаданные описывают объекты предметной области, информация о которых хранится в ХД. Такое описание включает: атрибуты объектов, их возможные значения, соответствующие поля в информационных структурах ХД, источники информации об объектах и т. п.;

- **кто** (описание пользователей) — метаданные описывают категории пользователей, использующих данные. Они описывают права доступа к данным, а также включают в себя сведения о пользователях, выполнявших над данными различные операции (ввод, редактирование, загрузку, извлечение и т. п.);

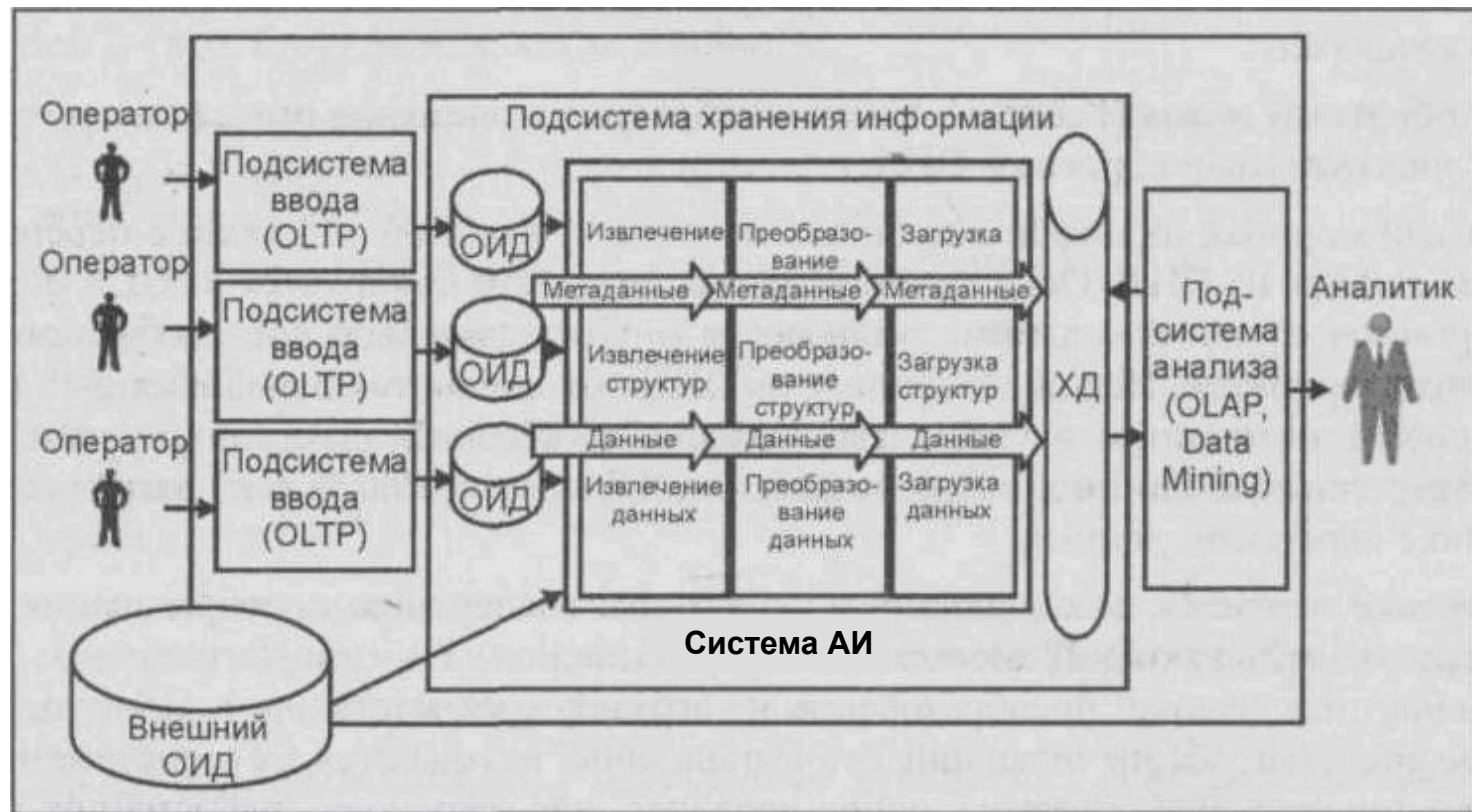
- **где** (описание места хранения) — метаданные описывают местоположение серверов, рабочих станций, ОИД, размещенные на них программные средства и распределение между ними данных;

- **как** (описание действий) — метаданные описывают действия, выполняемые над данными. Описываемые действия могли выполняться как в процессе переноса из ОИД (например, исправление ошибок, расщепление полей и т. п.), так и в процессе их эксплуатации в ХД;

- **когда** (описание времени) — метаданные описывают время выполнения разных операций над данными (загрузка, агрегирование, архивирование, извлечение и т.п.);

- **почему** (описание причин) — метаданные описывают причины, повлекшие выполнение над данными тех или иных операций. Такими причинами могут быть требования пользователей, статистика обращений к данным и т. п.

# ETL-процесс (E – extraction, T – transformation, L – loading)



## ETL-процесс

### Преобразование данных:

- обобщение данных (aggregation);
- перевод значений (value translation) — в ОИД данные часто хранятся в закодированном виде для того, чтобы сократить избыточность данных и память для их хранения;
- создание полей (field derivation);
- очистка данных (cleaning).

**Загрузка данных** — после того как данные преобразованы для размещения в ХД, осуществляется этап их загрузки. При загрузке выполняется запись преобразованных детальных и агрегированных данных. Кроме того, при записи новых детальных данных часть старых может переноситься в архив



# Очистка данных

Основные **проблемы очистки данных классифицируют** по следующим **уровням**:

- уровень ячейки таблицы;
- уровень записи;
- уровень таблицы БД;
- уровень одиночной БД;
- уровень множества БД.

В целом, **очистка данных включает несколько этапов**:

- выявление проблем в данных;
- определение правил очистки данных;
- тестирование правил очистки данных;
- непосредственная очистка данных.

**Непосредственная очистка данных.** На этом этапе выполняются преобразования в соответствии с определенными ранее правилами. Очистка **выполняется в два приема**:

- сначала устраняются проблемы, связанные с отдельными источниками данных,
- за тем устраняются проблемы множества БД.

**Над отдельными ОИД выполняются следующие процедуры**:

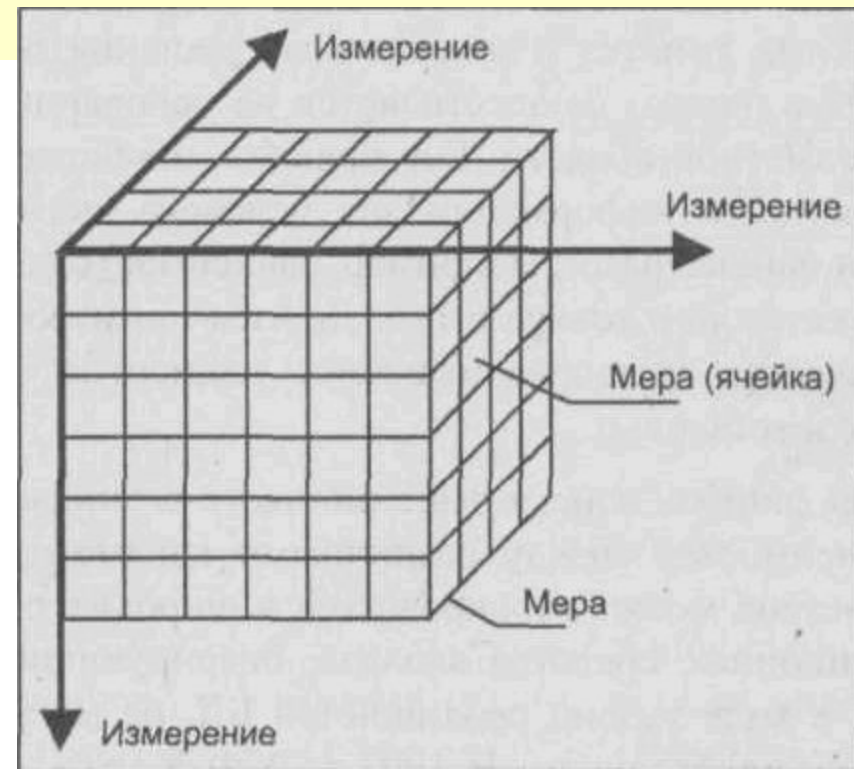
- проверка допустимости и исправления;
- стандартизация;
- исключение дубликатов

# Многомерная модель данных\_1

Э. Ф. Кодд— основоположник реляционной модели БД— рассмотрел ее недостатки, указав в первую очередь на невозможность **"объединять, просматривать и анализировать данные с точки зрения множественности измерений, т. е. самым понятным для аналитиков способом"**.

**Измерение** — это последовательность значений одного из анализируемых параметров. Каждое измерение может быть представлено в виде иерархической структуры. Например, измерение "Исполнитель" может иметь следующие иерархические уровни: "предприятие — подразделение — отдел — служащий". Более того, некоторые измерения могут иметь несколько видов иерархического представления. Например, измерение "время" может включать две иерархии со следующими уровнями: "год — квартал — месяц — день" и "неделя — день"

**На пересечениях осей измерений (Dimensions) располагаются данные**, количественно характеризующие анализируемые факты, — **меры** (Measures). Это могут быть объемы продаж, выраженные в единицах продукции или в денежном выражении, остатки на складе, издержки и т. п.

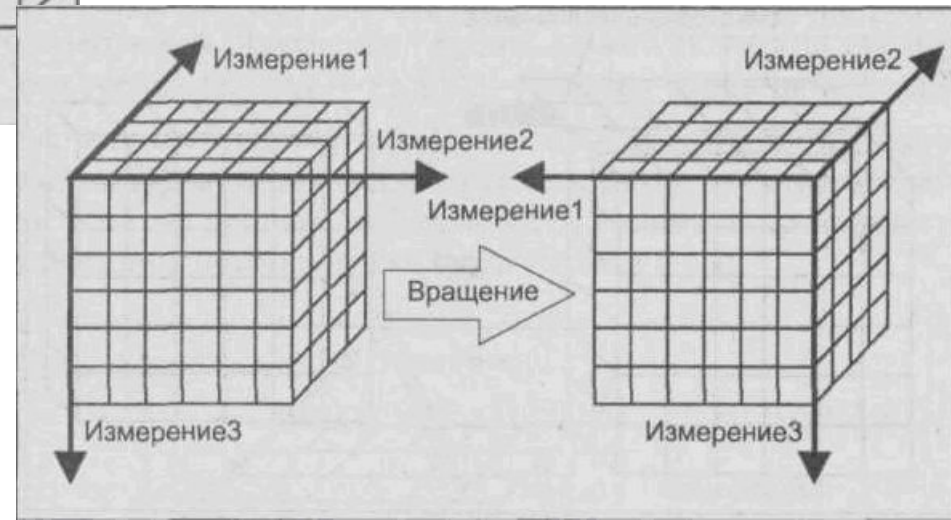
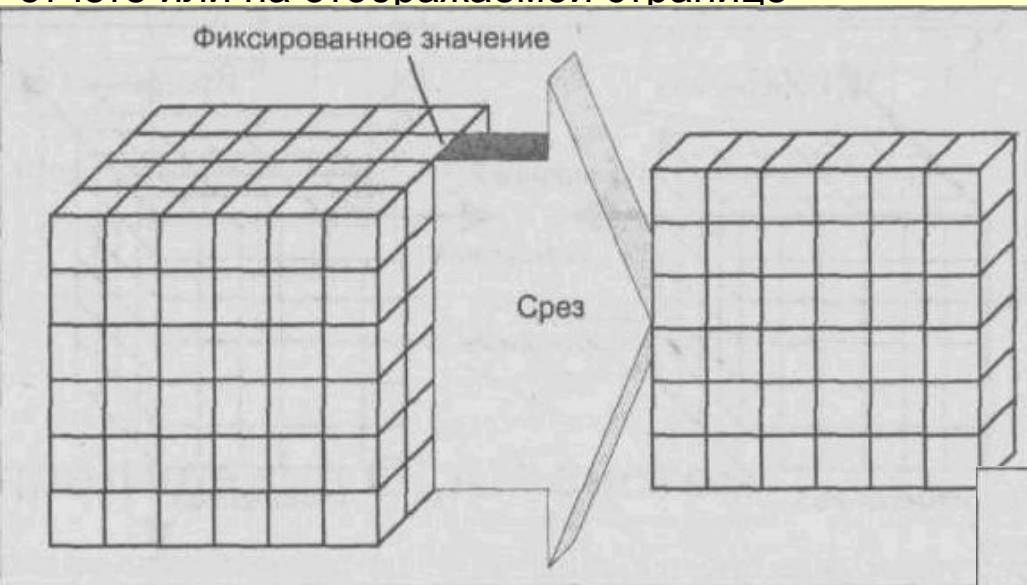


# Многомерная модель данных\_2

*Над гиперкубом могут выполняться следующие операции:*

- **Срез (Slice)** — формируется подмножество многомерного массива данных, соответствующее единственному значению одного или нескольких элементов измерений, не входящих в это подмножество;

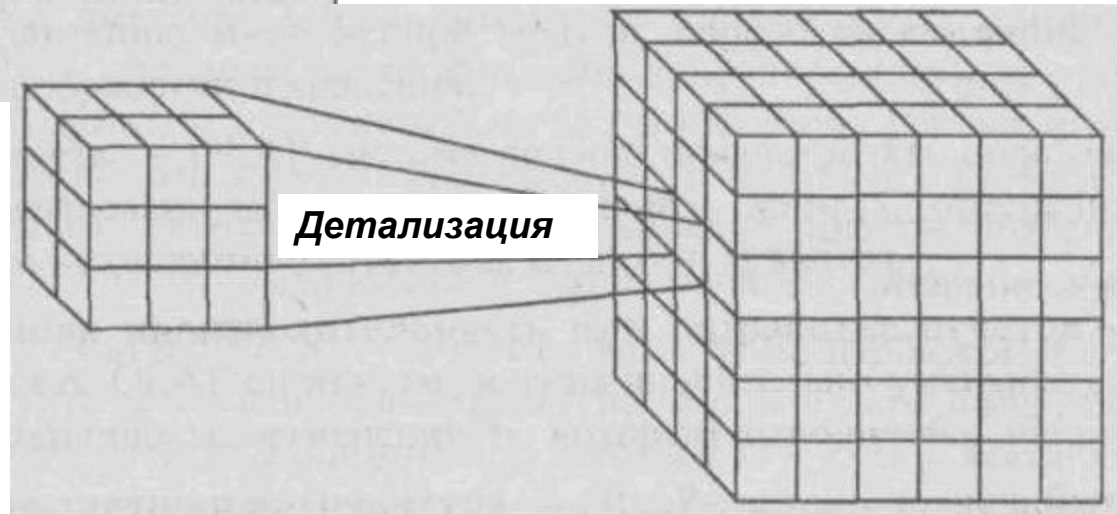
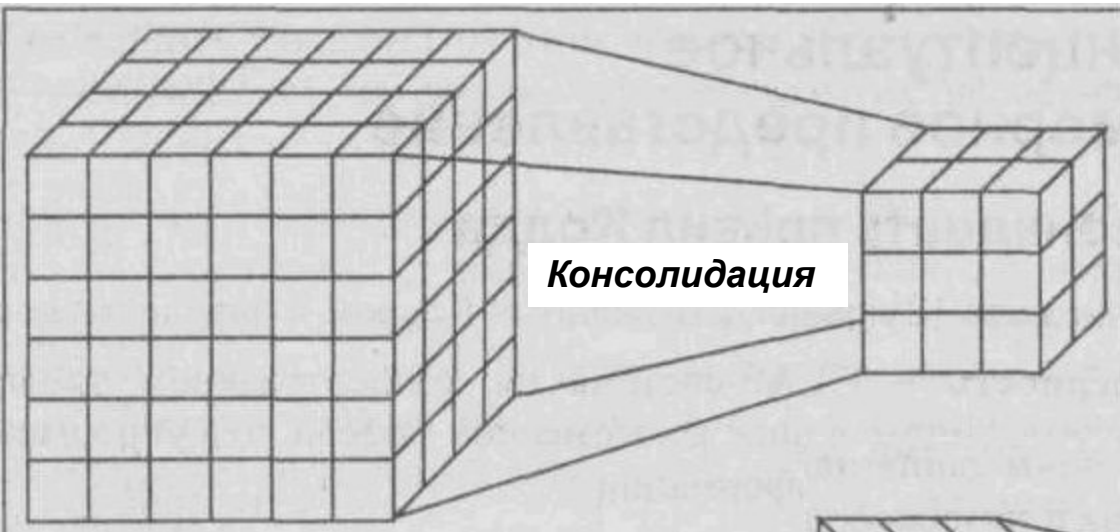
- **Вращение (Rotate)** — изменение расположения измерений, представленных в отчете или на отображаемой странице



# Многомерная модель данных\_3

*Над гиперкубом могут выполняться следующие операции:*

- **Консолидация** (Drill Up) и **детализация** (Drill Down) — операции, которые определяют переход вверх по направлению от детального (down) представления данных к агрегированному (up) и наоборот, соответственно. Направление детализации (обобщения) может быть задано как по иерархии отдельных измерений, так и согласно прочим отношениям, установленным в рамках измерений или между измерениями



# Архитектура OLAP-систем

**OLAP-система включает в себя два основных компонента:**

- **OLAP-сервер** — обеспечивает хранение данных, выполнение над ними необходимых операций и формирование многомерной модели на концептуальном уровне. В настоящее время OLAP-серверы объединяют с ХД или ВД;
- **OLAP-клиент** — представляет пользователю интерфейс к многомерной модели данных, обеспечивая его возможностью удобно манипулировать данными для выполнения задач анализа.

OLAP-серверы скрывают от конечного пользователя способ реализации многомерной модели. Они формируют гиперкуб, с которым пользователи посредством OLAP-клиента выполняют все необходимые манипуляции, анализируя данные. Между тем способ реализации очень важен, т. к. от него зависят такие характеристики, как производительность и занимаемые ресурсы.

**Выделяют три основных способа реализации OLAP-серверов :**

- **MOLAP** — для реализации многомерной модели используют многомерные БД;
- **ROLAP** — для реализации многомерной модели используют реляционные БД;
- **HOLAP** — для реализации многомерной модели используют и многомерные и реляционные БД.

# Архитектура OLAP-систем: MOLAP

**MOLAP-серверы** используют для хранения и управления данными многомерные БД. При этом данные хранятся в виде упорядоченных многомерных массивов. Такие массивы подразделяются на **гиперкубы и поликубы**.

В **гиперкубе** все хранимые в БД ячейки имеют одинаковую размерность, т.е. находятся в максимально полном базисе измерений.

В **поликубе** каждая ячейка хранится с собственным набором измерений, и все связанные с этим сложности обработки перекладываются на внутренние механизмы системы.

Физически данные, представленные в многомерном виде, хранятся в "плоских" файлах. При этом куб представляется в виде одной плоской таблицы, в которую построчно вписываются все комбинации членов всех измерений с соответствующими им значениями мер

## **Достоинства:**

- высокая скорость обработки данных.

## **Недостатки:**

- за счет денормализации и предварительно выполненной агрегации объем данных в многомерной БД в 2,5...100 раз меньше объема исходных детализированных данных;
- информационный гиперкуб является сильно разреженным;
- многомерные БД чувствительны к изменениям в многомерной модели. При добавлении нового измерения приходится изменять структуру всей БД, что влечет за собой большие затраты времени.

# Архитектура OLAP-систем: ROLAP

**ROLAP-серверы** используют реляционные БД.

В настоящее время распространены две основные схемы реализации многомерного представления данных с помощью реляционных таблиц: схема **"звезда"** и схема **"снежинка"**



# Архитектура OLAP-систем: HOLAP

**ROLAP-серверы** используют гибридную архитектуру, которая объединяет технологии ROLAP и MOLAP. В отличие от MOLAP, которая работает лучше, когда данные более-менее плотные, серверы ROLAP показывают лучшие параметры в тех случаях, когда данные довольно разрежены.

**Серверы HOLAP** применяют подход **ROLAP** для разреженных областей многомерного пространства и подход **MOLAP** — для плотных областей.

**Серверы HOLAP** разделяют запрос на несколько подзапросов, направляют их к соответствующим фрагментам данных, комбинируют результаты, а затем предоставляют результат пользователю.

## Основные производители OLAP-систем:

- **SAS Институт,**
- **Компания ORACLE,**
- **Компания Microsoft,**
- **Компания «Прогноз»**



# OLAP-система SAS-института\_0

## SAS System

- [AppDev Studio](#)
- [Base SAS](#)
- [Enterprise Guide](#)
- [Enterprise Miner](#)
- [Enterprise Reporter](#) (English)
- [JMP Software](#) (English)
- [SAS/ACCESS](#) (English)
- [SAS/AF](#) (English)
- [SAS/ASSIST](#) (English)
- [SAS/C Compiler](#) (English)
- [SAS/CONNECT](#) (English)
- [SAS/EIS](#)
- [SAS/ETS](#) (English)
- [SAS Financial Management Solutions](#)
- [SAS/FSP](#) (English)
- [SAS/GIS](#)
- [SAS/GRAPH](#) (English)
- [SAS Human Capital Management](#)
- [SAS/IML](#) (English)
- [SAS/INSIGHT](#) (English)
- [SAS Integration Technologies](#)
- [SAS/IntrNet](#)
- [SAS IT Charge Management](#) (English)
- [SAS IT Resource Management](#) (English)
- [SAS/LAB](#) (English)
- [SAS/MDDDB Server](#) (English)
- [SAS OnLineTutor](#) (English)
- [SAS/OR](#) (English)
- [SAS/QC](#) (English)
- [SAS Risk Management](#)
- [SAS/SECURE](#) (English)
- [SAS/SHARE](#)
- [SAS/SPECTRAVIEW](#) (English)
- [SAS/STAT](#) (English)
- [SAS Strategic Performance Management](#) (English)
- [SAS/TOOLKIT](#) (English)
- [SAS/TUTOR](#) (English)
- [SAS/Warehouse Administrator](#)
- [SAS Universal ODBC Driver](#) (English)
- [Scalable Performance Data Server](#)
- [StatView Software](#) (English)
- [SYSTEM 2000 software](#) (English)
- [webAF](#)
- [WebHound](#)
- [webEIS](#)

# OLAP-система SAS-института\_1

SAS включает свыше 20 различных программных продуктов, объединенных друг с другом «средствами доставки информации» (**Information Delivery System**, или IDS), так что весь пакет иногда обозначается как **SAS/IDS**.

**SAS/IDS** — это система комплексной интеграции разнообразных возможностей доступа к данным и управления ими, а также средств анализа данных, способов представления информации, генерации отчетов и визуализации результатов обработки информационных материалов

В составе системы **SAS/IDS** функционируют следующие компоненты (модули системы):

**BASE SAS** — ядро системы со встроенным языком программирования 4GL и языком работы с базами данных SQL, средства управления данными, поддержки индексов для баз данных, возможностями доступа к широкому набору форматов данных, процедуры описательной статистики и генерации отчетов.

**FSP** - обеспечивает доступ к данным, ввод, редактирование, преобразование данных, генерацию отчетов и деловую переписку.

**GRAPH** - содержит деловую, научную, рекламную графику, различные шрифты и карты.

**STAT** - включает в себя многофункциональный набор статистических процедур анализа данных.

# OLAP-система SAS-института\_2

SAS включает свыше 20 различных программных продуктов, объединенных друг с другом «средствами доставки информации» (*Information Delivery System*, или IDS), так что весь пакет иногда обозначается как **SAS/IDS**.

## **Дополнительные модули:**

**1ML** - представляет собой интерактивный матричный язык программирования для выполнения углубленных математических, инженерных и статистических расчетов. Этот язык дает возможность математику легко программировать свои собственные процедуры, используя язык, близкий к языку линейной алгебры.

**LAB** - предоставляет пользователю экспертную поддержку. В частности, здесь система подсказывает пользователю, выполняются или нет предположения, лежащие в основе того или иного метода анализа данных.

**ASSIST** - служит средством для обеспечения интерактивного доступа пользователей к различным возможностям системы SAS/IDS.

**EIS** - является меню-управляемым инструментом разработки и поддержки интерактивных исполняемых информационных систем методом объектно-ориентированной технологии. С помощью этого модуля легко настроить систему на собственные данные и формы представления результатов.

**ACCESS** - дает возможность конструировать отдельные интерфейсы для связи SAS/IDS с самыми разнообразными СУБД (ADABAS, DB2, ORACLE, SQL/DS и др.).

**INSIGHT** - представляет собой интерактивный инструмент для графического анализа данных

# OLAP-система SAS-института\_3

Стратегия SAS в области интеграции данных предполагает их преобразование в обобщенный ресурс – независимо от того, на какой платформе эти данные хранятся.

**SAS ETLQ** (Extraction, Transformation, Load integrated with Data Quality – извлечение, преобразование, загрузка интегрированные с поддержкой качества данных) представляет собой платформу интеграции, позволяющую клиентам извлекать их из любого источника – в том числе таких приложений, как Oracle, PeopleSoft, Siebel и SAP, а также их различных реляционных, нереляционных и других систем хранения данных. По существу, где бы данные ни хранились, SAS располагает инструментами для доступа к ним и их анализа.

Инструменты пакета реализуют провозглашенную SAS методологию работы с данными - методологию **SEMMA** (Sample, Explore, Modify, Model, and Assess - **выборка, исследование, изменение, моделирование и оценка**).

Компонент **Enterprise Miner** является заглавной составляющей общего семейства систем анализа данных, которые предлагает компания SAS. Он в полном объеме реализует методологию SEMMA в работе с данными.

В **Enterprise Miner** :

**функции выборки (Sample)** собраны в узлах **Input Data Source, Sampling u Data Partition** (Источник входных данных, Выборка и Разбиение данных) этого продукта.

В узле **Sampling** размещены функции простой случайной выборки, выборки всех n-ных элементов, расслоенной (стратифицированной) выборки, выборки первых n элементов и групповой выборки.

Узел **Data Partition** позволяет разбить исследуемый набор данных на подмножества для обучения, тестирования и проверки. В сочетании с **Input Data Source** этот узел предоставляет удобный в работе набор функций доступа к данным.

# OLAP-система SAS-института\_4

## В *Enterprise Miner* :

**Функции исследования (Explore)** размещаются в узлах *Distribution Explorer*, *SAS Insight*, а также в узле *Link Analysis*.

К категории *Noninteractive Explore* относятся *Multiplot*, *Association* и *Variable Selection*.

**Multiplot** — это узел простых неинтерактивных графических операций, позволяющий создавать диаграммы и гистограммы.

Узел **Association** позволяет определять традиционные правила ассоциаций и последовательных цепочек.

Узел **Variable Selection** служит для автоматического или ручного выбора входных переменных. В нем можно выполнить оценки критерия хи-квадрат и R-квадрат для автоматического определения наиболее важных входных переменных при попытке создания моделей прогноза интервальных или двоичных значений.

**Distribution Explorer** позволяет создавать многомерные гистограммы. Он оптимизирован для эффективной обработки больших наборов данных и позволяет интерактивно вращать и перемещать объемные диаграммы.

Узел **Insight** — это ссылка на систему SAS Insight, интерактивное средство исследования и анализа данных.

Узлы **Distribution Explorer** и **Insight** взаимно дополняют друг друга: **Distribution Explorer** применяется для исследования очень больших наборов данных с ограниченной графической поддержкой, а **Insight** — для тщательных изысканий в меньших, более тщательно отобранных подмножествах данных.

# OLAP-система SAS-института\_5

## В *Enterprise Miner* :

**Функции модификации (Modify)** позволяют:

**1.** Управлять атрибутами наборов данных (**Data Set Attributes**), выполнять преобразования и замену переменных (**Transform and Replace Variables**) и фильтрацию выбросов (**Filter Outliers**), а также проводить кластеризацию данных (разбиение на кластеры). Для этих операций используются программные средства **узлов Cluster и SOM/Kohonen** (самоорганизующиеся схемы Кохонена).

**2.** Преобразовывать транзакционные данные во временные ряды с помощью нового (экспериментального) узла **Time Series**.

Узел **Data Set Attribute** — это инструмент создания и использования набора данных для просеивания.

В узле **Transform** можно создавать из существующих переменных новые.

Узел **Replace** применяется для преобразования существующих или восстановления недостающих значений.

Узел **Filter Outliers** служит для работы с интервальными переменными и переменными типа категория и предоставляет автоматические и ручные методы удаления выбросов.

# OLAP-система SAS-института\_6

## В *Enterprise Miner* :

- кластеризация рассматривается как функция «модификации» данных, а не «моделирования». Возможно, это объясняется тем, что чаще всего кластеризация применяется для деления набора данных на группы для дальнейшего анализа.

- есть два узла для выполнения кластеризации — **Cluster** (рекомендуемый) и **SOM/Kohonen**.

Узел **Cluster** — это просто ссылка на процедуру SAS Fastclus, которая позволяет эффективно распределять очень большие наборы данных на взаимоисключающие кластеры.

Узел **SOM/Kohonen** используется преимущественно для выделения признаков и сокращения размерности, особенно в случаях, когда переменные характеризуются высокой степенью нелинейных отношений.

# OLAP-система SAS-института\_7

## В *Enterprise Miner* :

К функциям моделирования (*Model*) относятся узлы *Regression, Tree, Neural Network, Principal Components/DMNeural, Memory-Based Reasoning, Two-Stage, Ensemble* и *User Defined*.

Узел *Regression* позволяет выполнять стандартные линейную и логистическую регрессии, причем в нем доступны почти все функции основного модуля SAS/Stat.

Задача узла *Tree* — создание деревьев решений с использованием составного алгоритма, в который SAS включила некоторые возможности алгоритмов CHAID и CART. Это позволяет выполнять как автоматическое, так и интерактивное обучение.

Имя узла *Principal Components/DMNeural* указывает на то, что речь здесь идет о комбинации анализа главных компонент (principal component analysis, PCA) и нейронных сетей. PCA — популярная методика для выделения признаков. Поскольку алгоритмы нейронных сетей чувствительны к объемам наборов входных переменных и переменных со многими возможными значениями (большие области определения), в них обычно используется PCA или другие методы выделения признаков или сокращения размерности. Размещение указанных операций в этом узле оказывается весьма удобным для пользователя.

*Two-Stage* — это действительно составная модель, в ней объединены классовая и интервальная модели прогнозирования. Узел хорошо приспособлен для разработки моделей, предназначенных для решения задач типа: «Станет ли покупатель А приобретать товар X (да/нет), и если да, то сколько (количество/объем)?».

В узле *Ensemble* результаты отдельных моделей (апостериорная вероятность или прогнозируемые величины) объединяются в единый составной результат. Объединения могут быть как гомогенными, так и гетерогенными.



# OLAP-система SAS-института\_8

## В *Enterprise Miner* :

*Узел оценки (Assess)* делится на два подузла — *Assess* и *Reporter*.

Подузел *Assess* — это комплексный модуль для сравнения ожидаемых результатов моделей с реальной производительностью или моделей между собой.

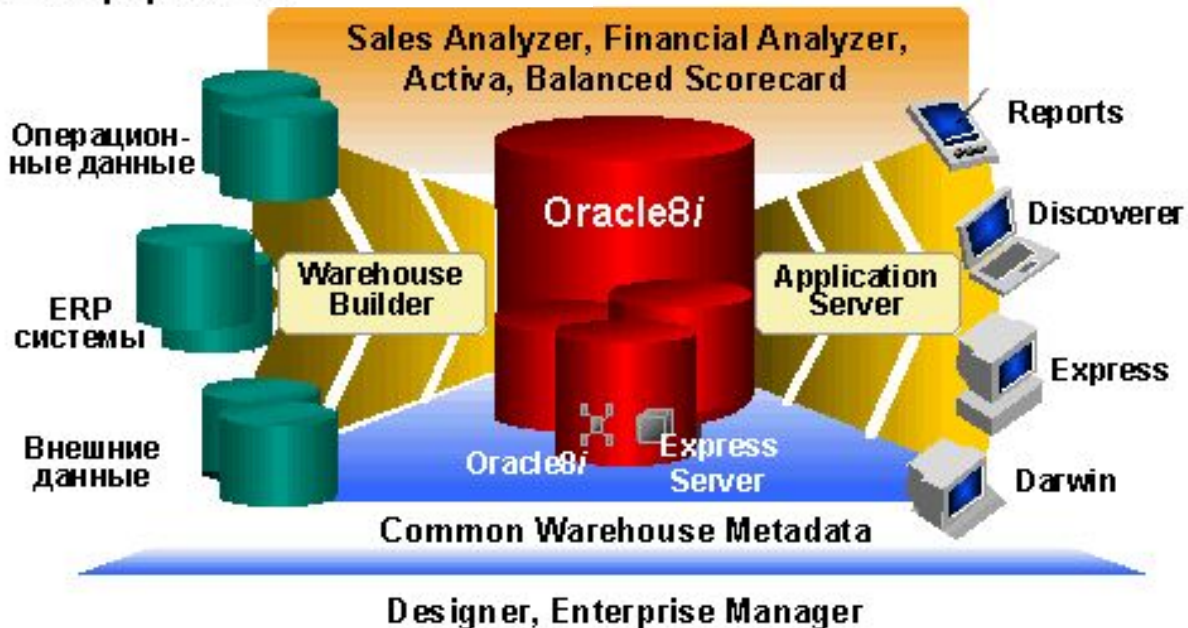
Подузел *Reporter* автоматически генерирует HTML-отчеты о потоках завершенных проектов добычи данных. Интеллектуальность этого узла заключается в этом, что он «понимает» свое место в рамках конкретного проектного потока и создает отчеты с разной детализацией в зависимости от своего местоположения и входных данных.

Помимо прочего, в *Enterprise Miner* включена база данных *Data Mining Database* (DMDB) для хранения уже добытых результатов. DMDB — это особый набор данных SAS, оптимизированный для операций по добыче данных. Хранение предварительно рассчитанных статистических данных в DMDB позволяет алгоритмам Enterprise Miner не выполнять эти операции и сэкономить на многих «проходах» по данным.

# OLAP-система компании ORACLE\_1



## Проектирование и администрирование



# OLAP-система компании ORACLE\_2

В качестве механизма хранения в реляционных Хранилищах и Витринах Данных используется сервер **Oracle9i (Oracle10i)**, в многомерных Витринах - **Express Server**.

Инструментальное средство **Oracle Warehouse Builder**, построенное на базе современной архитектуры Common Warehouse Metadata, позволяет спроектировать структуру целевого Хранилища, создать процедуры извлечения, согласования и загрузки данных из различных источников и сгенерировать метаданные для средств доступа, таких как **Discoverer** и **Express**.

Для проектирования Хранилища можно также использовать **Oracle Designer**, а затем автоматически перенести описание проекта в репозиторий метаданных **Warehouse Builder**.

Средства доступа к данным покрывают весь спектр аналитических задач: для стандартной отчетности используется **Reports**, для генерации нерегламентированных отчетов и запросов - **Discoverer**, для сложного многомерного анализа - продукты семейства **Express**, а для задач "извлечения знаний" - **Darwin Data Mining Suite**.

Существуют готовые аналитические приложения для решения специализированных задач - **Sales Analyzer, Financial Analyzer, Activa u Balanced Scorecard**. В настоящее время готовятся к выпуску новые аналитические приложения, такие как **Demand Planning, Value Based Management** и другие.

Конечные пользователи могут получать информацию как с помощью традиционных клиент-серверных технологий, так и в Интранет-архитектуре с использованием масштабируемого сервера приложений **Application Server**.

# OLAP-система компании ORACLE\_3

**Oracle Warehouse Builder** - многофункциональная расширяемая среда для разработки и развертывания корпоративных Хранилищ и Витрин Данных:

1. Продукт позволяет визуально моделировать схему Хранилища либо импортировать описывающие ее метаданные из репозитория Oracle Designer.
2. В Warehouse Builder включены возможности визуального определения отображений между источниками и Хранилищем и последующей генерации на их основе модулей загрузки данных в виде процедур PL/SQL или с использованием SQL\*Loader.
3. Используемый в сочетании с Warehouse Builder продукт Oracle Pure\*Integrate предоставляет мощные средства автоматической очистки данных, включающие алгоритмы нечеткой логики, синтаксический разбор имен и адресов, вероятностные модели и т.д.
4. Источниками данных для Warehouse Builder, помимо СУБД Oracle различных версий, могут быть плоские файлы, СУБД других производителей (доступ через шлюзы и ODBC), файловые системы мэйнфреймов (через шлюзы либо продукт Oracle Pure\*Extract), а также приложения ERP (через компоненты Integrator for Oracle Applications, SAP R/3 и др.).
5. Открытая архитектура Common Warehouse Metadata позволяет осуществлять обмен метаданными между Warehouse Builder и средствами бизнес-анализа - Discoverer и Express.

**Oracle Reports** - это инструмент создания и публикации **стандартных** отчетов с помощью визуального интерфейса. Oracle Reports позволяет распространять информацию, используя Web-архитектуру

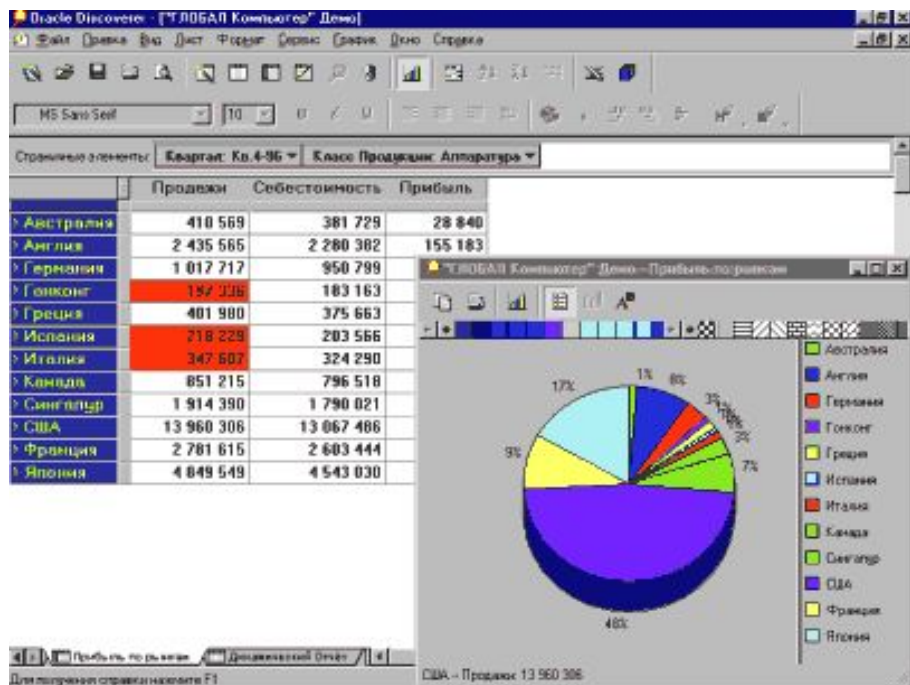
**Oracle Discoverer** - это инструмент для получения произвольных отчетов, формирования нерегламентированных запросов и анализа данных, который дает бизнес-пользователям всех уровней быстрый и удобный доступ к информации, содержащейся в реляционных Хранилищах и Витринах Данных, а также в OLTP-системах (в том числе не обязательно реализованных на СУБД Oracle).

# OLAP-система компании ORACLE\_4

Существует несколько редакций **Oracle Discoverer** для разных категорий пользователей.

**1. Discoverer Administration Edition** предназначен для создания и администрирования слоя метаданных, а также для разграничения прав доступа к данным и функциональным возможностям Discoverer на уровне каждого конечного пользователя.

**2. Discoverer User Edition** - это генератор интерактивных отчетов для конечного пользователя, позволяющий ему формулировать свои потребности в информации в терминах бизнес-области и не требующий от него знания языка SQL.



Встроенные Мастера позволяют определять условия-фильтры, итоги и подитоги, вычисляемые поля и т.д., а также представлять информацию в графическом виде.

Существует "тонкий" клиент с ограниченной функциональностью - Discoverer Viewer, предназначенный для просмотра отчетов, созданных в User Edition.

# OLAP-система компании ORACLE\_5

**Продукты Oracle Express** предназначены для решения аналитических задач высокого уровня, связанных со сложными расчетами, прогнозированием, моделированием сценариев "что-если" и т.д. Эти продукты используют особую технологию. Она реализована в виде семейства OLAP-продуктов Oracle Express, включающего в себя:

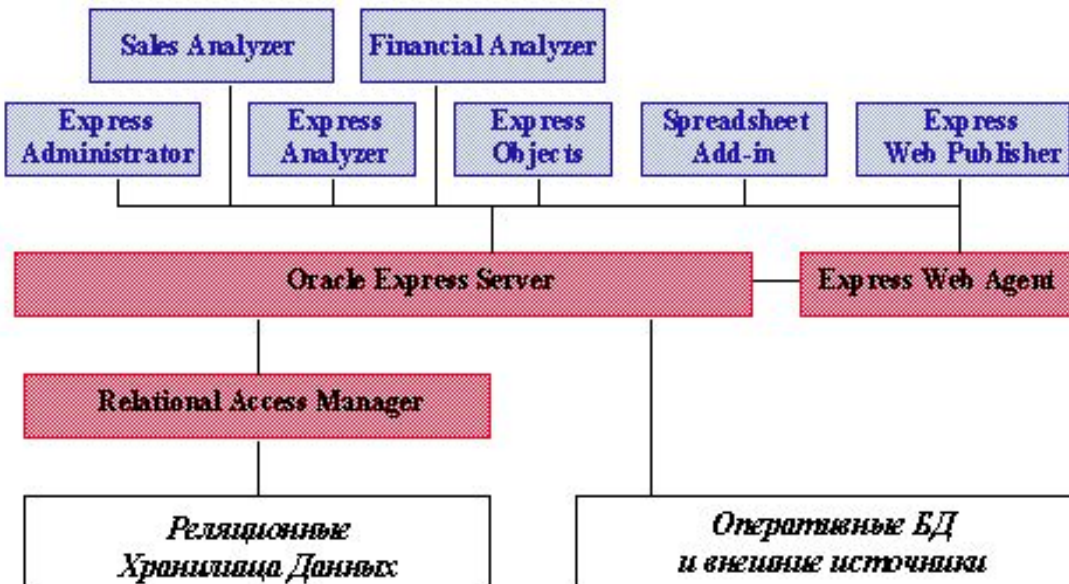
- специальный сервер многомерных БД,
- инструментарий администрирования,
- средства разработки в среде клиент-сервер и Web,
- готовые аналитические приложения, ориентированные на конкретные

функциональные задачи, такие как анализ продаж и маркетинговые исследования, финансовый анализ и бюджетное планирование.

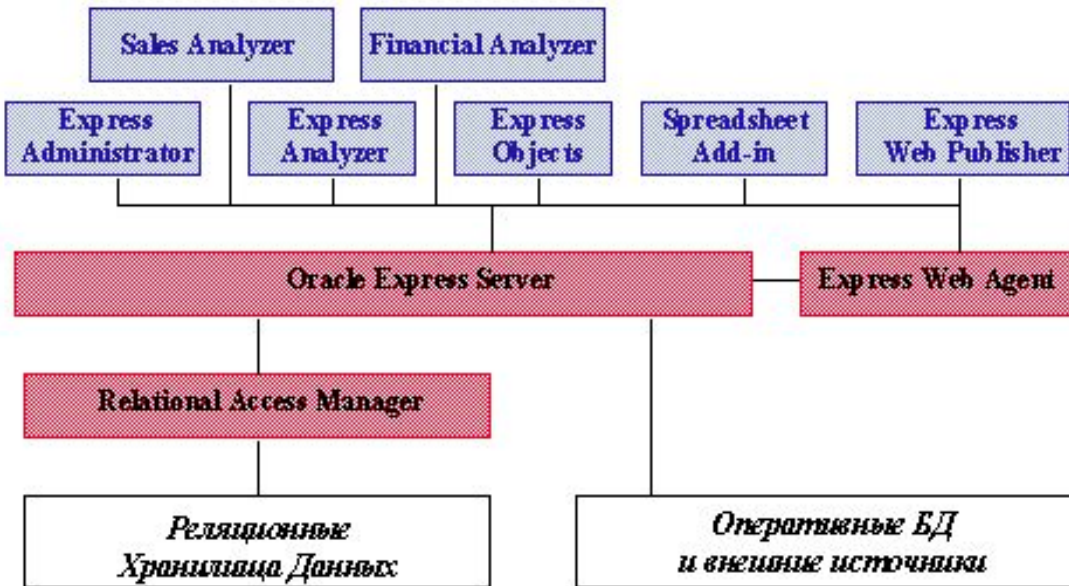
**Oracle Express Server** использует многомерную модель данных.

**Express Server** имеет механизм аналитических вычислений, включающий в себя более сотни встроенных функций - математических, статистических, функций анализа временных рядов, финансовых и других.

Разработчики могут расширять аналитические возможности путем создания собственных функций на процедурном языке **Express SPL**



# OLAP-система компании ORACLE\_6



Существует версия **Express Server**, работающая на персональном компьютере под управлением Windows 95/98/NT, - **Personal Express**.

С точки зрения архитектуры, модели данных и аналитических возможностей эти продукты абсолютно идентичны.

**Personal Express** используется прежде всего мобильными пользователями, отсоединенными от корпоративной сети.

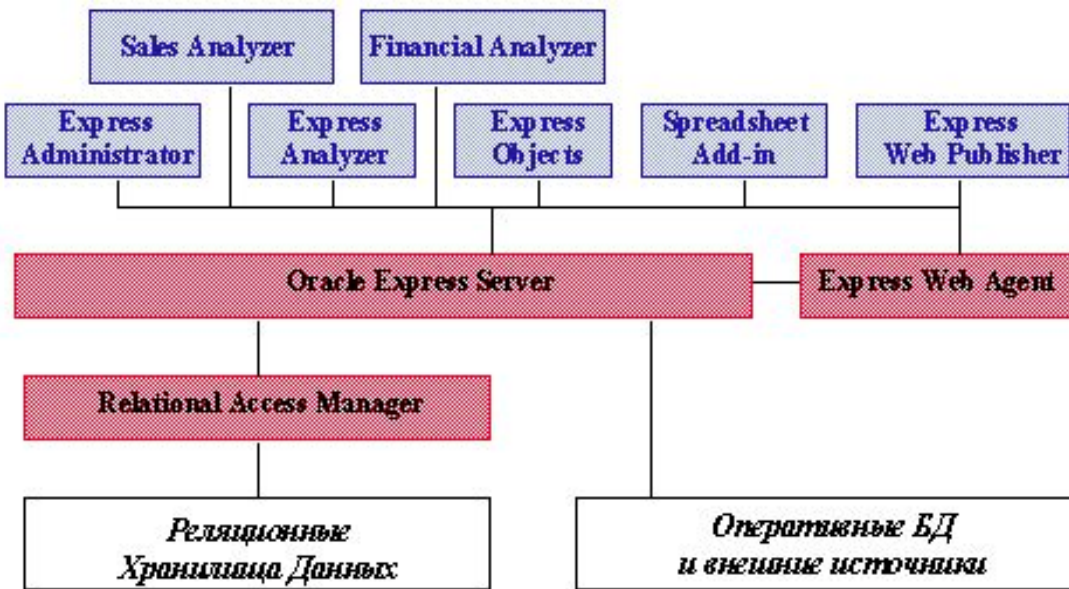
**Oracle Express Administrator** является инструментом построения и администрирования многомерных Витрин Данных. Это визуальная среда разработчика для создания, модификации и управления объектами БД Express, а также загрузки данных из различных источников.

**Express Relational Access Manager** используется для динамического доступа к реляционным Хранилищам и Витринам. Он позволяет реализовывать произвольные схемы хранения данных - ROLAP (реляционный OLAP), MOLAP (многомерный OLAP) или HOLAP (гибридный OLAP).

С помощью графической утилиты **Express Relational Access Administrator** создается структура многомерной базы данных, а затем, используя drag-and-drop, устанавливает соответствие между объектами **Express** и структурами реляционного Хранилища

**Express Relational Access Manager** – инструмент генерация метаданных непосредственно на основе структуры Хранилища, описанной в репозитории **Warehouse Builder**

# OLAP-система компании ORACLE\_7



**Oracle Express Analyzer** - средство конечного пользователя, предназначенное для самостоятельного построения отчетов, анализа многомерных данных Express и публикации результатов на Web. Основным понятием в Express Analyzer является брифинг - набор страниц, содержащих динамические интерактивные отчеты в табличном и графическом виде, кнопки с привязанными к ним действиями (например, печать или экспорт в Excel) и OLE-объекты.

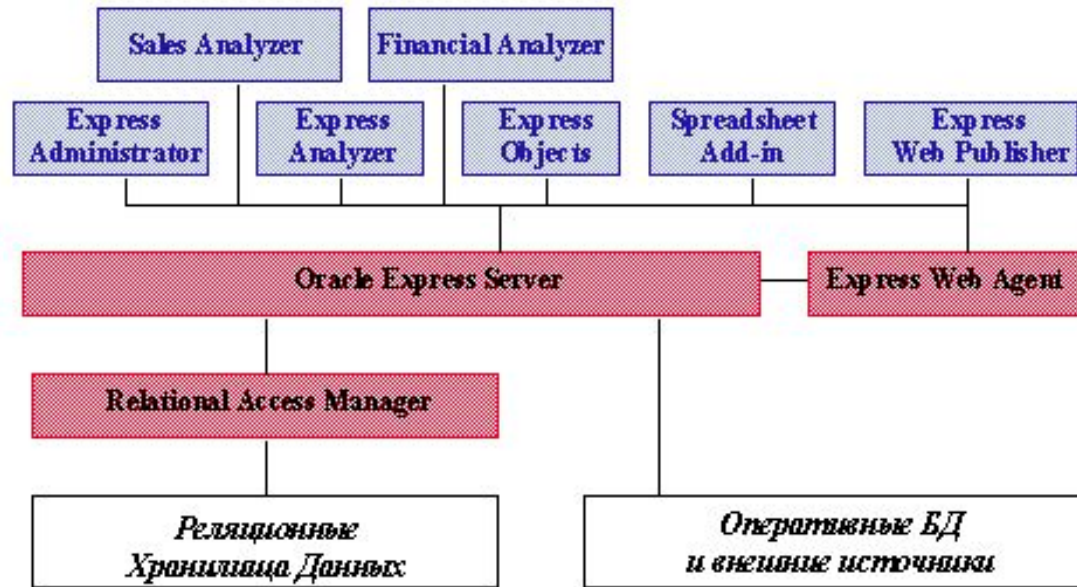
**Oracle Express Objects** - объектно-ориентированная графическая среда, предназначенная для быстрой разработки полнофункциональных OLAP-приложений в среде клиент/сервер. В качестве языка управления объектами в Express Objects используется Express Basic, синтаксически совместимый с Microsoft Visual Basic.

**Oracle Express Spreadsheet Add-In.** Кроме Express Analyzer или Express Objects, пользователи могут получать динамический доступ к многомерным данным, используя в качестве интерфейса Microsoft Excel. Специальная компонента **Express Spreadsheet Add-In** дополняет стандартные возможности этой электронной таблицы функциональностью OLAP, позволяя с помощью простого Мастера строить в среде Excel интерактивные многомерные отчеты.



# OLAP-система компании ORACLE\_8

**Технология Express** позволяет пользователям проводить интерактивный анализ не только в среде клиент-сервер, но и в архитектуре Web



Серверная компонента **Oracle Express Web Agent**, состоящая из интерфейсного модуля к **Express Server** и набора Java-апплетов для визуализации и работы с данными, предоставляет все возможности OLAP, не требуя программного обеспечения **Express** на персональном компьютере

**Oracle Financial Analyzer** - интегрированное решение для формирования финансовой отчетности, анализа, прогнозирования, составления и контроля исполнения бюджетов.

**Oracle Sales Analyzer** - приложение масштаба предприятия, предназначенное для анализа продаж и маркетинговых исследований. В продукт встроен графический инструментальный для построения индикаторов эффективности и других показателей, таких как динамика продаж, доля рынка, различные индексы, результаты нарастающим итогом, скользящее среднее, сдвиги по времени и т. д.

Инструментальная среда **Darwin** предназначена для анализа данных методами, относящимися к технологии **"data mining"** (извлечение знаний).

# OLAP-система компании «Прогноз»\_1

Аналитический комплекс "Прогноз" - [База - КАДАСТР]

База данных Объект Правка Сервис Администрирование Вид Окно ?

База данных

- Таблицы
- Измерения
- Кубы
- Запросы
- Отчеты
- Формы

Измерения

- Измерения
  - ! TMP
  - 1. Федеральное Агенство Кадастра Недвижимости
    - 1. Справочники
    - 2. ГКС
    - 3. S\_Interface
    - прочее (старое)

Наименование	Идентификация
Справочники форм отчетности	
ATO по федеральным округам	DIM_ATO_F
Календарь по кварталам	CALEND_Q
Справочник показателей (боковик таблиц)	D_POK_SPF
Справочник показателей (шапка таблиц)	D_POK_SPF
Текущая дата	DIM_DSYS

Моделирование  
Администрирование  
Личные папки

# OLAP-система компании «Прогноз»\_2

Аналитический комплекс "Прогноз" - [База - КАДАСТР]

База данных | Объект | Правка | Сервис | Администрирование | Вид | Окно | ?

База данных

- Таблицы
- Измерения
- Кубы
- Запросы
- Отчеты
- Формы

Кубы

- Кубы
  - ! TMP
  - 1. Федеральное Агенство Кадастра Недвижимости
    - Основные кубы
    - Старое
  - 2. ГКС
  - прочее (старое)

Наименование	Идентификатор	Примеч.
Данные в структуре отчетных дат - свод по субъектам	CUBE_DREP_SUMREG	
Данные форм отчетности в структуре текущих дат	CUBE_DCUR	

Моделирование  
Администрирование  
Личные папки

# OLAP-система компании «Прогноз»\_3

Аналитический комплекс "Прогноз" - [Вычисляемый куб: Для отчетов ФАКН (конечные данные)]

База данных: Окно ?

Текущий элемент: **Общая площадь**

Измерение-назначение: Справочник показателей (шапка таблиц)

- Сельскохозяйственные угодья
  - В стадии мелиоративного строи...
- Лесные земли
  - Под древесно-кустарниковой ра...
  - Из них защитного значения
  - Под водой
- Земли застройки
  - Под дорогами
  - Болота
- Нарушенные земли
- Прочие земли
  - Из всех земель: олени пастбища
- СВЕДЕНИЯ О ПРАВАХ, НА КОТОРЫХ ИСП...
- СВЕДЕНИЯ О ПРАВАХ, НА КОТОРЫХ ИСП...
- ОТЧЕТ О СОСТОЯНИИ МЕЛИОРИРУЕМЫ...
- ОТЧЕТ О СОСТОЯНИИ МЕЛИОРИРУЕМЫ...
- СВЕДЕНИЯ О НАЛИЧИИ ЗЕМЕЛЬ У ГРА...

Измерение-источник: Для отчетов ФАКН (данные в исходной структуре)

Формы отчетности ФАКН

- Отчет о составе работников, замещающих государственные должности и должности государственной гражданской с...
- Сведения о наличии и распределении земель по категориям и формам собственности
- Сведения о наличии и распределении земель по категориям и угодьям**
- Сведения о правах, на которых использовали землю граждане (объединения граждан), занимающиеся производством...
- Сведения о правах, на которых использовали землю предприятия, организации, хозяйства, общества, занимающиеся...
- Отчет о состоянии мелиорируемых земель (орошение), (приложение № 1 к форме № 22-4 /организации, граждане/)
- Отчет о состоянии мелиорируемых земель (осушение), (приложение № 2 к форме № 22-4 /организации, граждане/)
- Сведения о наличии земель у граждан (объединений граждан), занимающихся производством сельскохозяйственной...
- Сведения о наличии земель у предприятий, организаций, хозяйств, обществ, занимающихся производством сельско...
- Сведения о распределении общих площадей городских поселений по видам использования земель и формам собст...

Строки форм отчетности ФАКН

9. Из всех земель: земли природоохранного назначения

Графы форм отчетности ФАКН

Общая площадь

Разделы форм отчетности ФАКН

Раздел I

Показатель-назначение: Показатель

Показатель-источник: Показатель

Период действия формулы: 02.02.2004-02.02.2006

{  
Сведения о наличии и распределении земель по категориям и угодьям | Раздел I | 8. Итого земель в административных границах | Общая площадь(Показател  
+  
Сведения о наличии и распределении земель по категориям и угодьям | Раздел I | 9. Из всех земель: земли природоохранного назначения | Общая площадь(По

+ - \* / ( ) > < = <=> Or And Not If Nz  
Min Max Power Check MMin MMax Round Trunc

Отменить Проверить

Закреть

9:35

# OLAP-система компании «Прогноз»\_4

Аналитический комплекс "Прогноз" - [База - КАДАСТР]

База данных | Объект | Правка | Сервис | Администрирование | Вид | Окно ?

База данных

- Таблицы
- Измерения
- Кубы
- Запросы
- Отчеты
- Формы

Моделирование  
Администрирование  
Личные папки

## Отчеты

- Отчеты
  - ! TMP
  - 1. Федеральное Агентство Кадастра Недвижимости
    - 2. ГКС
      - S\_interface
      - Мои отчеты

Наименование	Идентификатор	Примеч.
разработка		
Старое		
<input checked="" type="checkbox"/> Анализ структуры земельного фонда по категориям...	R_FAKN_22U	
<input checked="" type="checkbox"/> Анализ структуры земельного фонда по категориям...	R_FAKN_22C	
<input checked="" type="checkbox"/> Данные АКН в исходной структуре	P_FAKN_PRIMARY_...	
<input checked="" type="checkbox"/> Данные показателей форм отчетности АКН	P_FAKN_DATAA	
<input checked="" type="checkbox"/> динамика	PIVOT	
<input checked="" type="checkbox"/> Оценка достоверности и согласованности данных	R_CHECK_DATA_CO...	
<input checked="" type="checkbox"/> Сведения об изменении площадей сельскохозяйств...	R_FAKN_2TB	
<input checked="" type="checkbox"/> Сводный отчет о рекультивации земель, снятии и ис...	R_FAKN_RKZ	
<input checked="" type="checkbox"/> Экспресс анализ исходных данных по формам АКН	R_FAKN_EXPRES_A...	

Пуск | ПродуктыARIS | Входящие - Outlo... | ПрезентацияАна... | Аналитический к... | ЭкраныПрогноза ... | EN | 15:13

# OLAP-система компании «Прогноз»\_5

Конструктор отчётов - [Анализ структуры земельного фонда по категориям и формам собственности в разрезе субъектов РФ \*]

Форма Правка Вид Вставить Данные Макро Окна ?

data(27,2,19,16,2)

Структура в динамике >>

Период: 2000

Территория: РОССИЙСКАЯ ФЕДЕРАЦИЯ

## Сведения о наличии и распределении земель по категориям форм собственности

за 2000 г.  
единица измерения: гектар

РОССИЙСКАЯ ФЕДЕРАЦИЯ

Наименование строки	Код строки	Общая площадь	В собственности граждан	В собственности юридических лиц	В государственной и муниципальной собственности	В собственности Российской Федерации				
						всего	из них:		в аренду	
							гражданам	юридическим лицам		
А	Б	1	2	3	4	5	6	7	8	9
1. Земли сельскохозяйственного назначения, в том числе:	101	440 165 200								
1.1. фонд перераспределения земель	102									
2. Земли поселений, в том числе:	103	18 599 400								
2.1. городских поселений	104									
2.2. сельских поселений	105									
3. Земли промышленности, энергетики, транспорта, связи, радиовещания, телевидения, информатики, земли для обеспечения космической деятельности, земли обороны, безопасности и земли иного специального назначения	106	17 415 600								
3.1 Земли промышленности	107									
3.2 Земли энергетики	108									
3.3. Земли транспорта, в том числе:	109									