

# Хранилища данных

Исполняется впервые

- Процесс управления сводится к решению 3 задач:
- Где мы находимся?
- Куда мы хотим прийти?
- Как мы туда попадем?

# Проблемы (в бизнесе):

- динамичное изменение экономической ситуации, что мешает применять накопленный опыт, не успевают выработаться интуиция.
- в условиях свободного рынка нет возможности проводить целенаправленные эксперименты.

# Определение 1

- **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.**

# Хранилище данных создается с **целью:**

- Интеграции в одном месте, согласования и, возможно, агрегации ранее разъединенных детализированных данных:
- Исторических архивов
- Данных из оперативных систем
- Данных из внешних источников

# Определение 2

- **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.

# 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.

- Технология комплексного многомерного анализа данных получила название OLAP (On-Line Analytical Processing).
- Концепция OLAP была описана в 1993 году Эдгаром Коддом, известным исследователем баз данных и автором реляционной модели данных.

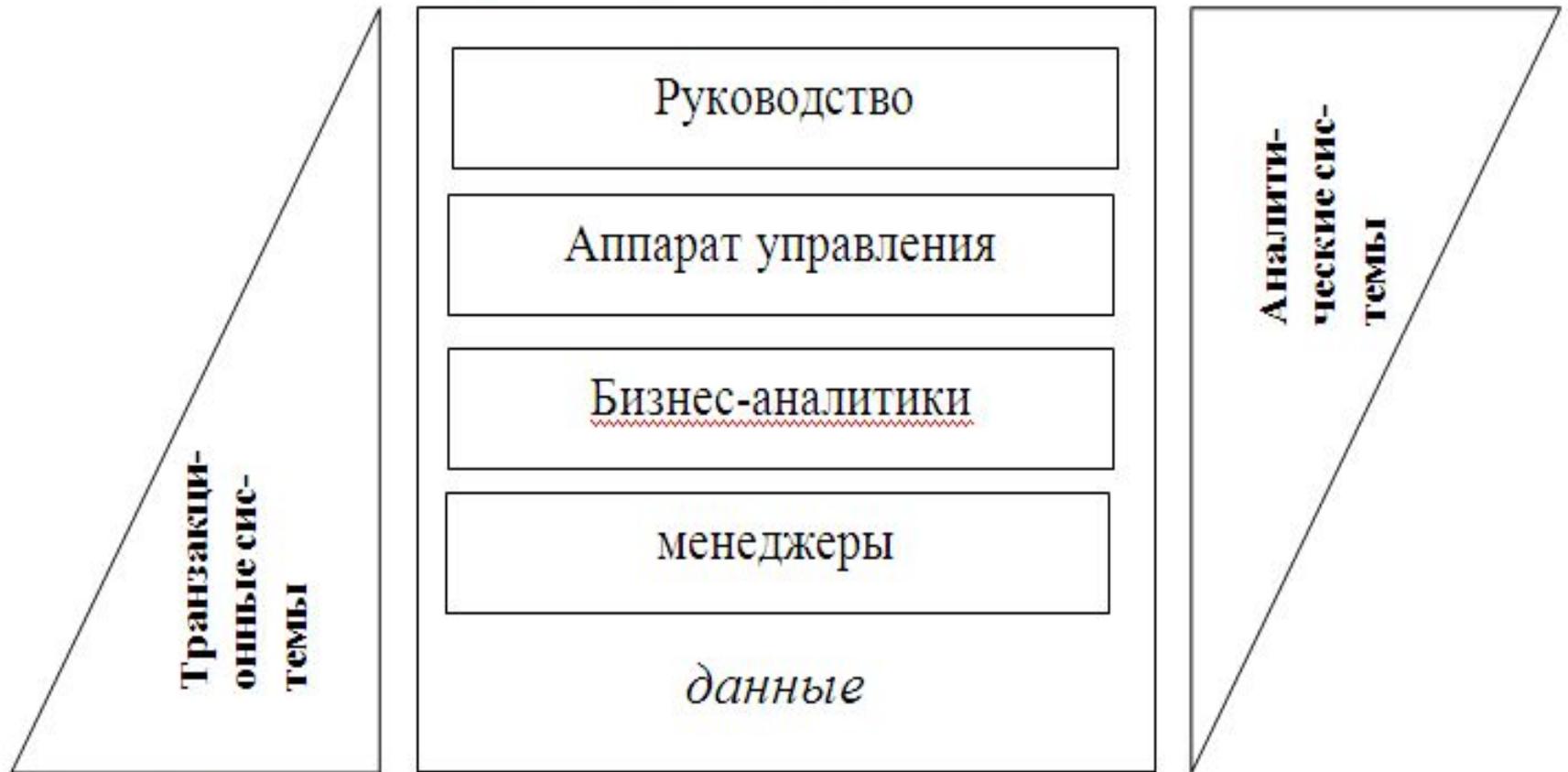
# требования к приложениям для многомерного анализа:

- предоставление пользователю результатов анализа за приемлемое время (обычно не более 5 с), пусть даже ценой менее детального анализа;
- возможность осуществления любого логического и статистического анализа, характерного для данного приложения, и его сохранения в доступном для конечного пользователя виде;
- многопользовательский доступ к данным с поддержкой соответствующих механизмов блокировок и средств авторизованного доступа;
- многомерное концептуальное представление данных, включая полную поддержку для иерархий и множественных иерархий (это — ключевое требование OLAP);
- возможность обращаться к любой нужной информации независимо от ее объема и места хранения.

# В контуре управления взаимосвязаны 5 функций

- Учет
- Кон-троль
- Анализ
- Приня-тие решения
- Плани-рование

# . Распределение функционала между аналитическими и транзакционными информационными системами



# Сравнительные характеристики хранилищ данных и оперативных систем

Системы хранилищ данных	Оперативные системы
Используются руководством	Используются работниками «переднего края»
Стратегическое значение	Тактическое значение
Поддерживают стратегические направления развития бизнеса	Поддерживают повседневную деятельность
Используются для интерактивного анализа	Используются для обработки транзакций
Предметно-ориентированные	<u>Ориентированны на приложения</u>
Хранят исторические данные	Хранят только текущие данные
Непредсказуемые запросы	Предсказуемые запросы

# Общие свойства хранилищ

- ориентированность на предметную область или ряд предметных областей,
- интегрированность,
- зависимость от времени (поддержка хронологии),
- ПОСТОЯНСТВО.

# Ориентированность на предметную область

- Приложения всегда оперируют функциями, такими, например, как открытие сделки, кредитование, выписка накладной, зачисление на счет и т.д. Хранилище данных организовано вокруг фактов и предметов, таких, как сделка, сумма кредита, покупатель, поставщик, продукт и т.д.

# Интегрированность

- Наиболее важный аспект хранилища данных состоит в том, что данные, находящиеся в хранилище, интегрированы.
- Интегрированность проявляется во многих аспектах:
- в согласованности имен,
- в согласованности единиц измерения переменных,
- в согласованности структур данных,
- в согласованности физических атрибутов данных и др.
- Контраст между интеграцией данных в хранилище данных и в прикладном окружении иллюстрируется следующим образом.

# Зависимость от времени

- проявление зависимости хранилища данных от времени состоит в неукоснительном выполнении правила, что данные, однажды корректно в хранилище записанные, не могут быть обновлены.
- Хранилище данных с точки зрения практического использования представляет собой большую серию моментальных снимков.
- Естественно, если моментальный снимок данных был сделан некорректно, он может быть изменен. Но если был получен корректный моментальный снимок, то, однажды сделанный, он в последующем изменению не подлежит.

# ПОСТОЯНСТВО

- При передаче данных из оперативной среды в хранилище данных данные фильтруются.
- Многие данные вообще никогда не выгружаются из оперативной среды.
- В хранилище данных передается только информация, используемая для обработки в системе поддержки принятия решений.
- Временной горизонт в средах существенно различается. Данные в оперативной среде всегда являются текущими. Данные в хранилище имеют хронологию.
- С точки зрения временного горизонта пересечение между оперативной средой и средой хранилища данных минимально.
- Хранилище данных содержит агрегированные (итоговые) данные, которые никогда не включаются в оперативную среду.
- Передача данных из оперативной среды в хранилище данных сопровождается фундаментальными преобразованиями. Большинство данных при поступлении в хранилище видоизменяется.

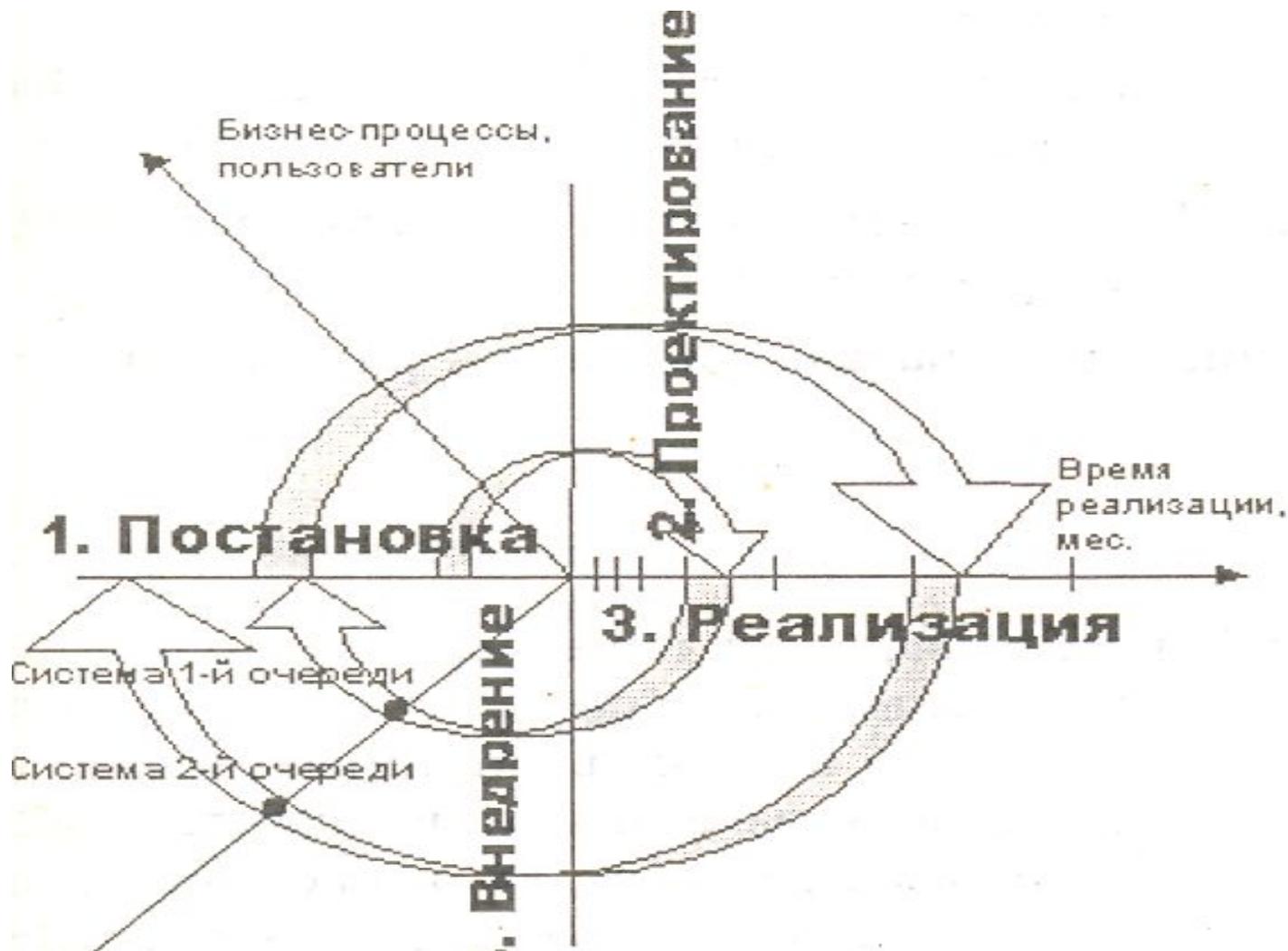
# Данные хранилища

- В общем случае модель данных современных Систем Поддержки Принятия Решений (СППР) строится на основе **пяти классов данных**:
  - источники данных,
  - хранилища данных (в узком смысле),
  - оперативный склад данных,
  - витрины данных,
  - метаданные.

# Компоненты хранилища

- Хранилище на самом верхнем уровне **состоит**, как правило, из трех подсистем:
  - подсистемы загрузки данных,
  - подсистемы обработки запросов и представления данных,
  - подсистемы администрирования хранилища.

# Методика (методология) построения хранилищ данных



# Постановка задачи

- проводятся интервью с основными участниками проекта со стороны компании-заказчика и лицами, ответственными за принятие управленческих решений;
- уточняется организационная структура, фиксируются организационные и функциональные рамки проекта;
- выявляются и документируются особенности и недостатки существующих информационных решений;
- формализуется схема бизнеса компании с учетом функциональных рамок;
- производится сбор существующих отчетных материалов и прочих официальных документов, имеющих непосредственное отношение к реализации проекта.

# Проектирование

- **Архитектура информационной системы** рассматривается в четырех аспектах:
- Логическая архитектура. Представляет архитектуру системы с точки зрения пакетов базовых классов и их взаимосвязей.
- Архитектура процессов. Применительно к СППР, определяет информационное обеспечение системы – состав и содержание процессов преобразования и передачи данных.
- Компонентная архитектура. Представляет архитектуру ПО системы, ее декомпозицию на подсистемы и компоненты.
- Техническая архитектура. Описывает физические узлы системы и связи между ними.

# Проектирование данных

- Сбор данных.
- Преобразование данных:
  - - Очистка данных.
  - - Согласование данных.
  - - Унификация данных.
  - - Агрегирование данных.
- Хранение данных:
  - - Промежуточное хранение данных.
  - - Накопление исторических данных.
- Предоставление данных потребителям.
- Сопровождение метаданных.

# ОСНОВНЫЕ ТИПЫ ДАННЫХ

- **Персональная информация** – эта информация, используемая пользователями со строго определенными обязанностями и информационными потребностями.
- **Информацию по бизнес-темам** – информация, относящаяся к определенной тематике, например, как финансовая деятельность организации. Для организаций имеющих близкие функциональные и организационные структуры, ее можно определить как информацию для подразделения (например, для финансовой службы).
- **Детальные данные** – самая подробная информация, доступная в хранилище данных. Обычными пользователями применяется весьма редко, только в случае необходимости подробного уточнения информации. Обычно является полем деятельности аналитиков по добыче знаний (или поиску скрытых зависимостей в больших объемах информации).

