

Проектирование баз данных, ч.2

ФИО преподавателя: Смирнов Михаил Вячеславович
e-mail: smirnov.mirea@gmail.com

Лекция 9

ОСНОВЫ ИСПОЛЬЗОВАНИЯ ХРАНИЛИЩ ДАННЫХ

Вопросы лекции

- хранилище данных и финансовая информация бизнеса, типовая трехтировая архитектура хранилища
- архитектура хранилища данных по Kimball и Inmon
- варианты имплементации хранилищ данных и группа реализации проекта

Типовые проблемы крупного бизнеса



1. Важная для бизнеса информация распределена среди разнообразных, зачастую разнородных систем.
2. Поиск информации, требуемой для принятия бизнес-решений это очень затратное по времени мероприятие, к тому же с высокой вероятностью недостоверности.
3. Очень сложно ответить на фундаментальные вопросы бизнеса

Фундаментальные вопросы бизнеса (Forbes)

В чем ценность моего продукта предложения)?

Будет ли спрос на мой товар? Какой его объем?

Насколько масштабируется мой бизнес?

Какая оптимальная (максимальная) цена у моего товара?

Как продавать товар?

Какой капитал нужен для...?

Прогноз по операционной окупаемости...

Прогноз по инвестиционной окупаемости

Конечная цель (достижимость?)

Определение хранилища данных

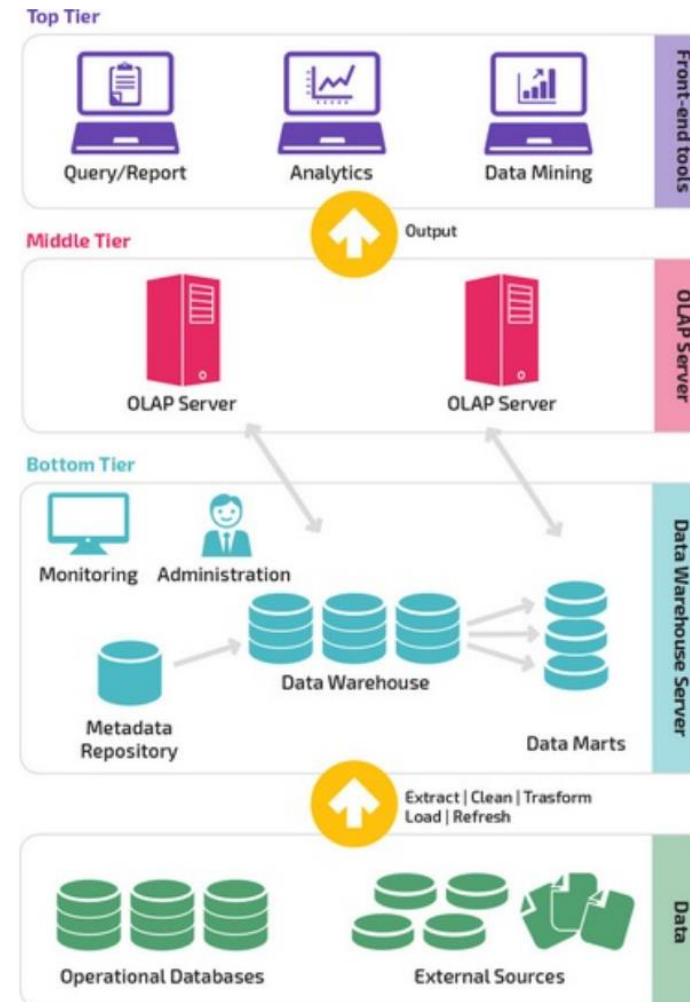
1. Централизованное хранилище бизнес информации, необходимой для отчетности и анализа.
2. Обычно, хранилище данных:
 - содержит большие массивы архивных (исторических данных);
 - оптимизировано под запросы к данным (а не на добавление-изменение, как РМД);
 - пополняется новой бизнес информацией через определенные интервалы времени;
 - представляет собой основу для решения корпоративных задач.

Классическая трехуровневая архитектура хранилища данных

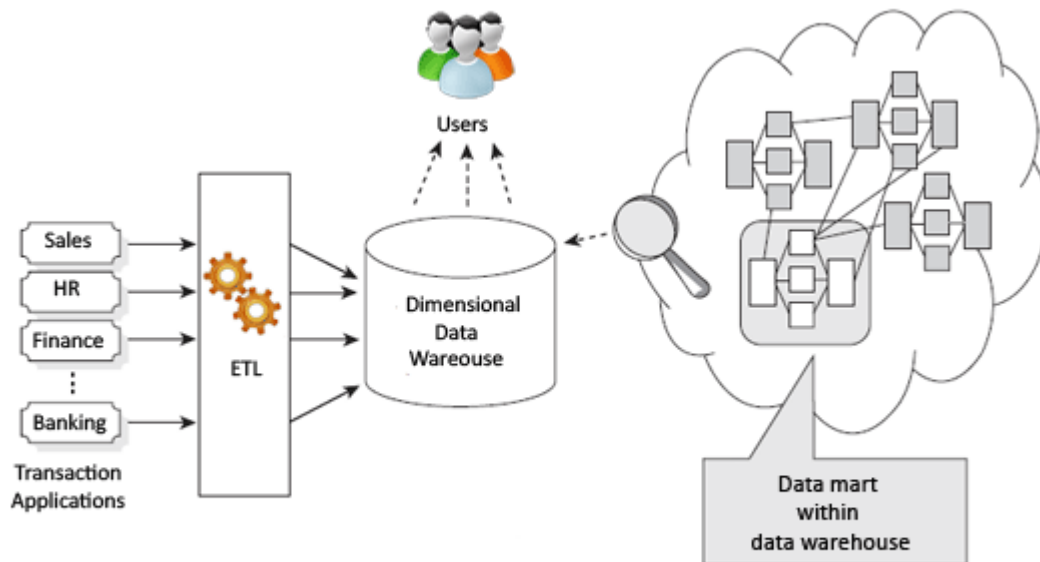
Нижний уровень: сервер базы данных, используемый для извлечения данных из множества различных источников.

Средний уровень: средний уровень содержит сервер OLAP, который преобразует данные в структуру, лучше подходящую для анализа и сложных запросов.

Верхний уровень: уровень клиента. Инструменты, используемые для высокоуровневого анализа данных, создания отчетов и анализа данных.

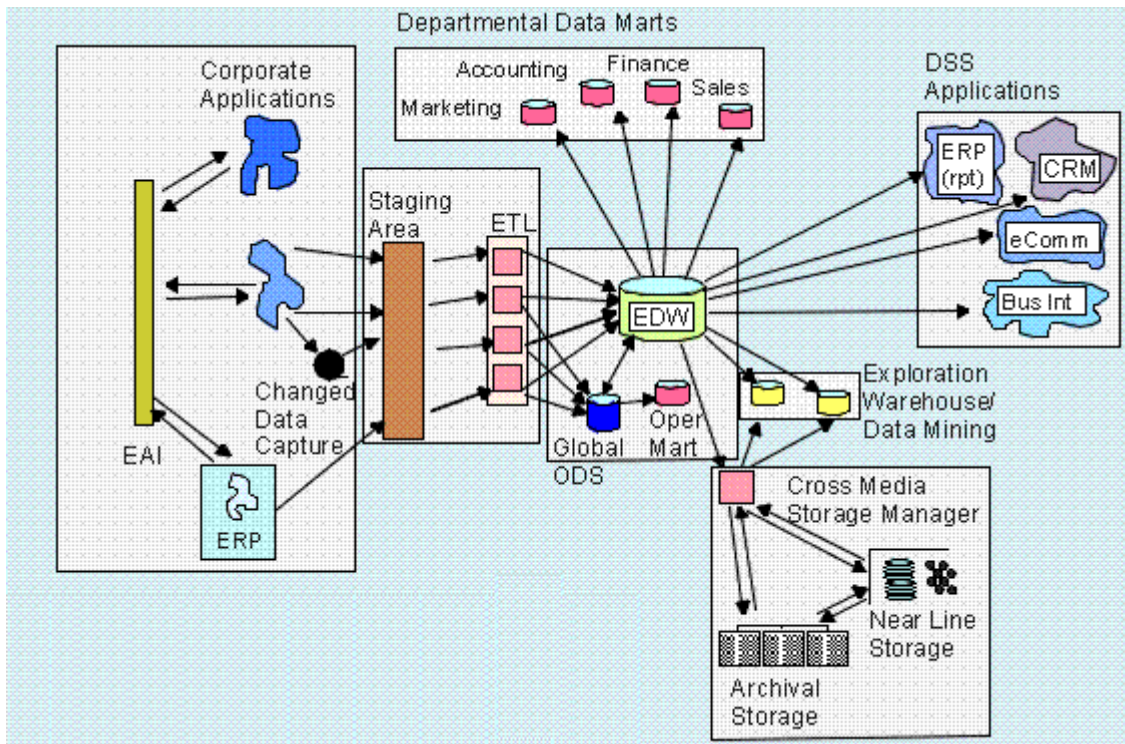


Модель хранилища мистера Кимбалла (Kimball)



- Transaction applications – системы или сервисы, созданные для обработки бизнес транзакций. Результатом обработки являются данные, сохраненные в РМД или в файлах.
- ETL – сервисы приведения данных к стандартному виду. Расшифровывается как достать, изменить, загрузить.
- A dimensional data warehouse. Содержит корпоративные данные в виде, наиболее удобном для гранулирования. Используется модель измерений. Как правило имеет форму звездочки. Аналитические системы способны извлекать данные напрямую из модели.
- Datamart. Витрина данных. Существует за пределами хранилища и представляет собой логический концепт.

Модель хранилища мистера Инмона (Inmon)



- Corporate Applications. В данной модели называются также системами источника, и предоставляют данные в хранилище.
- ETL Processes. Также называются data services. После обработки данных в рамках установленного формата, они загружаются в корпоративное хранилище. Процессы ETL могут выполняться в режиме batch или в режиме real-time.
- Enterprise data warehouse. Центральный элемент архитектуры, объединенный репозиторий атомарных данных, находящихся в третьей нормальной форме (см. реляционную модель).
- Data marts. Выбирают данные из хранилища с помощью инструментов агрегации. Аналитические приложения обращаются не к хранилищу, а к заранее подготовленным витринам данных.

Критерии выбора Kmb-Inm.

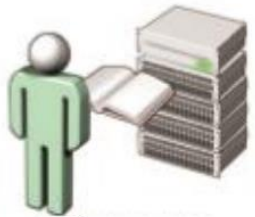
Характеристика	Kmb	Inm
Уровень бизнес-решений	Тактический	Стратегический
Требования по интеграции данных	Отдельные бизнес-требования	Корпоративные бизнес-требования
Структура данных	Простые количественные данные, KPI	Многослойные данные, качественные данные
Стабильность данных	Источники редко меняются	Источники часто меняются
Сложность исполнения	Маленькая команда проекта	Большая команда проекта
Время на развертывание	Срочно требуется хранилище	Есть существенное время на проектирование и релиз
Затраты	Сравнительно низкие затраты	Высокие затраты со старта

Варианты имплементации хранилищ данных



Сборка на заказ

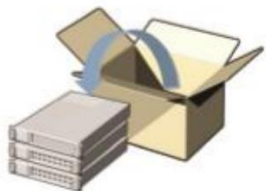
Хранилища можно построить с использованием типового инструментария многопользовательских реляционных баз данных (Oracle, PostgreSQL, MS SQL Server).



Эталонные архитектуры

Также, может быть использовано специальное программное обеспечение для построения хранилищ данных (Autonomous Data Warehouse, IBM Netezza Performance Server).

Такие продукты, помимо софта могут также комплектоваться и железом.



Готовые решения по ХД

Типовой состав участников проекта реализации хранилища данных



менеджер проекта

архитектор решения

инженер данных

АД

специалист по инфраструктуре

разработчик ETL процедур

бизнес-аналитик

тестировщик

специалисты по обслуживанию данных

Чтение на дом

- <https://www.oracle.com/ru/database/what-is-a-data-warehouse/#link1>
- Английский Кренке, стр. 573-587.

Спасибо за внимание!