



УНИВЕРСИТЕТ

искусственного интеллекта

ВИТРИНЫ ДАННЫХ

План

- Витрина данных
- Витрина Данных vs Хранилище Данных
- Иммон vs Кимбалл
- Use Case
- OLAP
- BI (Business Intelligence)



Витрина Данных



Витрина Данных (Data Mart) - относительно небольшое хранилище или же его часть, предназначенную для применения конкретным подразделением организации и/или определенной группой пользователей.

Достоинства:

- Представление аналитикам только той информации, которая действительно нужна для определенного рабочего задания, профиля служебной деятельности.
- Максимальная приближенность целевой части хранилища данных к конкретному пользователю.
- Содержание тематических подмножеств заранее агрегированных специалистами данных, которые в дальнейшем проще настраивать и проектировать.
- Для реализации витрины данных (хранилища данных специализированного типа) не требуется вычислительная техника большой мощности.

Недостатки:

- Реализация информационной территориально распределенной системы, чья избыточность слабо контролируется.
- Не предполагается методик, способов, которые могли бы обеспечить целостность и непротиворечивость хранящейся в витрине данных (базе данных узкоспециальной) информации.

Витрина Данных vs Хранилище Данных 1/2

Хранилище Данных

Фокус: Промышленный репозиторий различных источников данных.

Источники данных: Множество внутренних и внешних источников данных из различных областей организации.

Размер: Начиная от 100 GB и до Петабайт

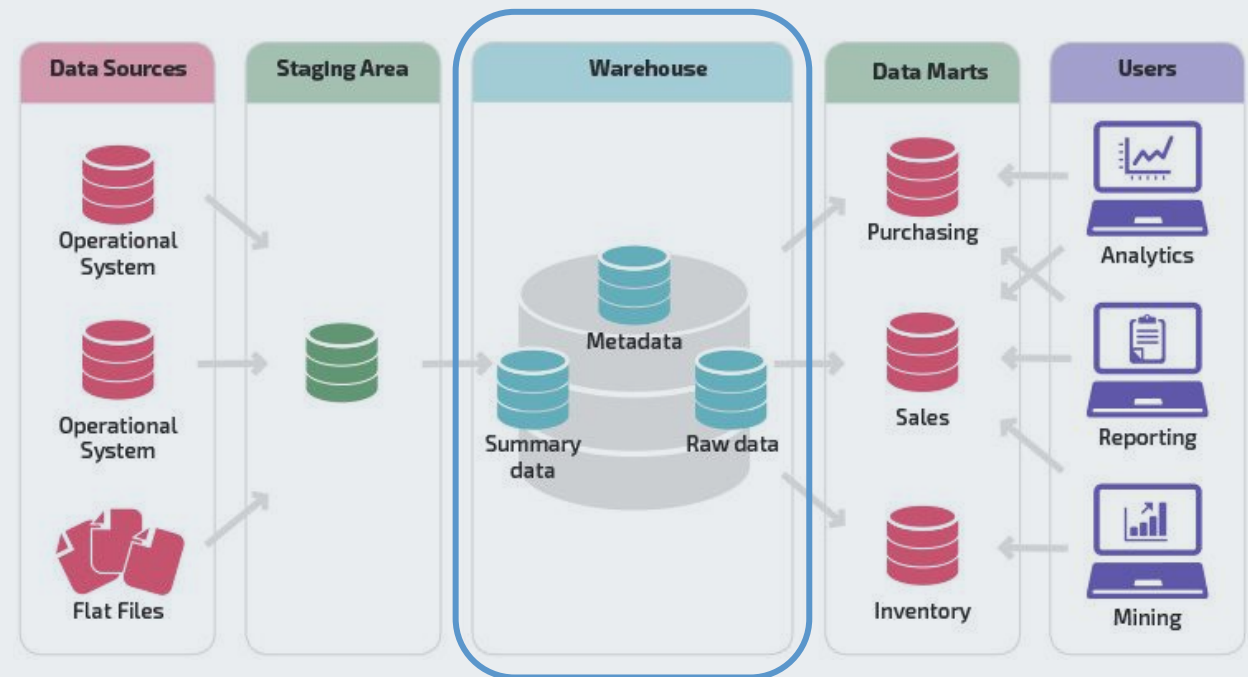
Нормализация: Современные хранилища слабо нормализованы для обеспечения наибольшей производительности.

Типы принимаемых решений: На основе данных из хранилища принимаются стратегические решения влияющие на всю организацию в целом.

Стоимость: Часто более \$100,000.

Время развертывания и установки: Не менее года, при использовании собственной инфраструктуры

Типы Данных: сырые данные, справочники, транзакционные данные



Витрина Данных vs Хранилище Данных 1/2

Витрина Данных

Фокус: Конкретный субъект в деятельности организации (например отдел продаж)

Источники Данных: Ограниченное кол-во источников данных связанных с бизнес потребителем.

Размер: Менее 100 GB

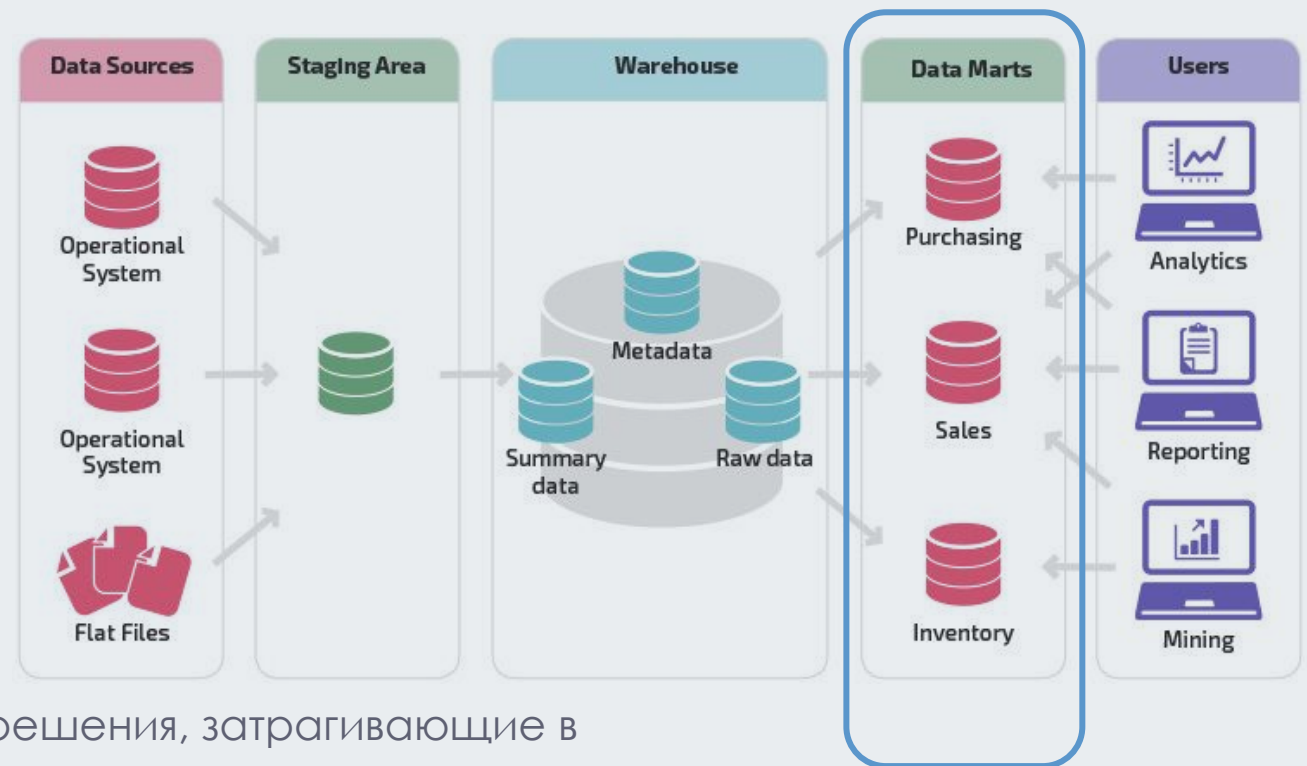
Нормализация: Часто строго нормализованы, но позволяют гибко перестраивать отношения

Типы принимаемых решений: Тактические решения, затрагивающие в основном деятельность конкретного подразделения

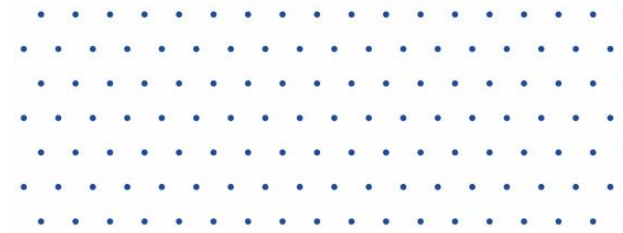
Стоимость: в районе \$10,000 и выше

Время развертывания: 3-6 месяцев

Типы Данных: Агрегаты, KPI и справочники

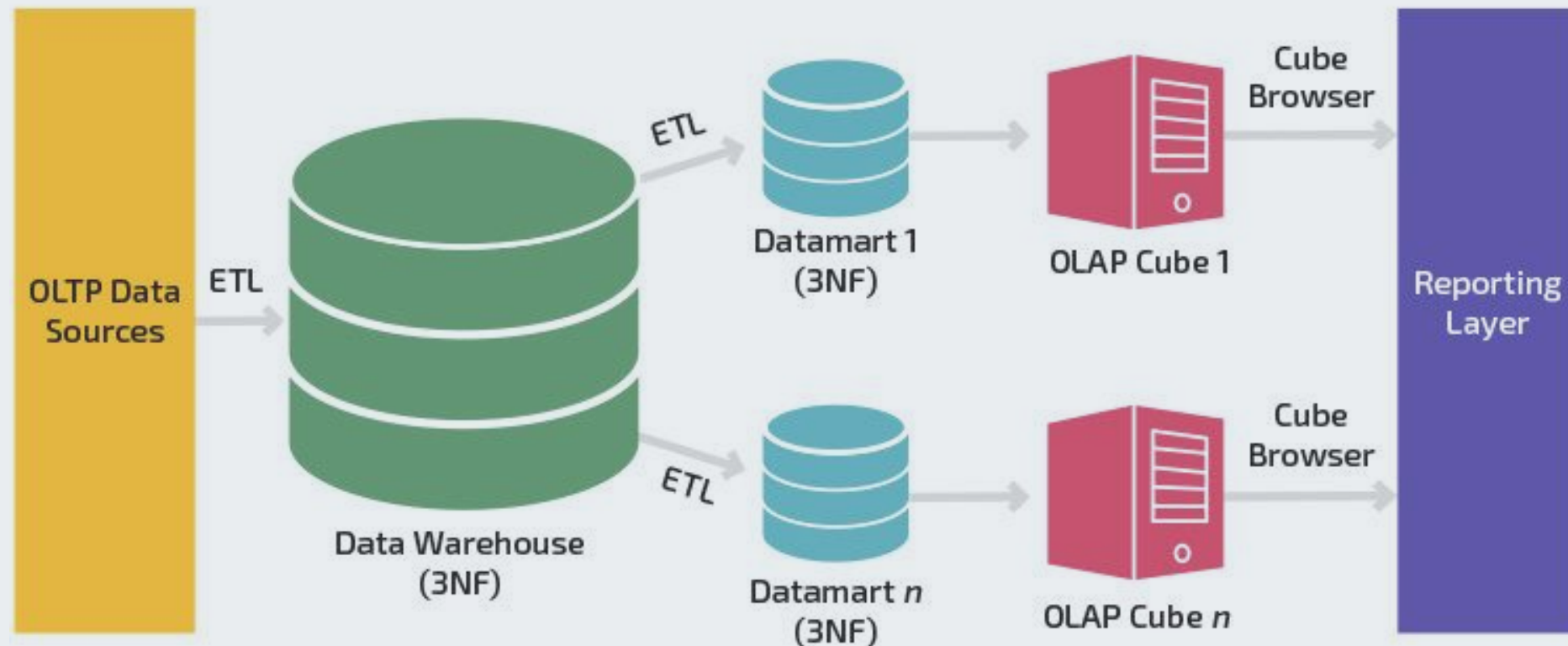


Иммон vs Кимбалл 1/2

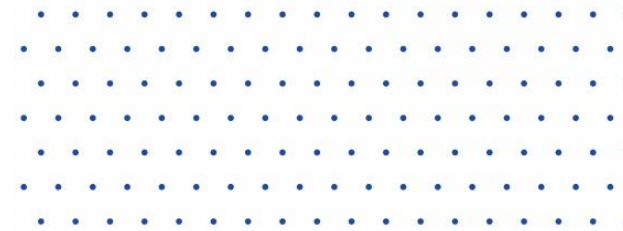


Данные жестко нормализованы даже в хранилище, витрина представляет собой физическое представление модели данных

Inmon Model

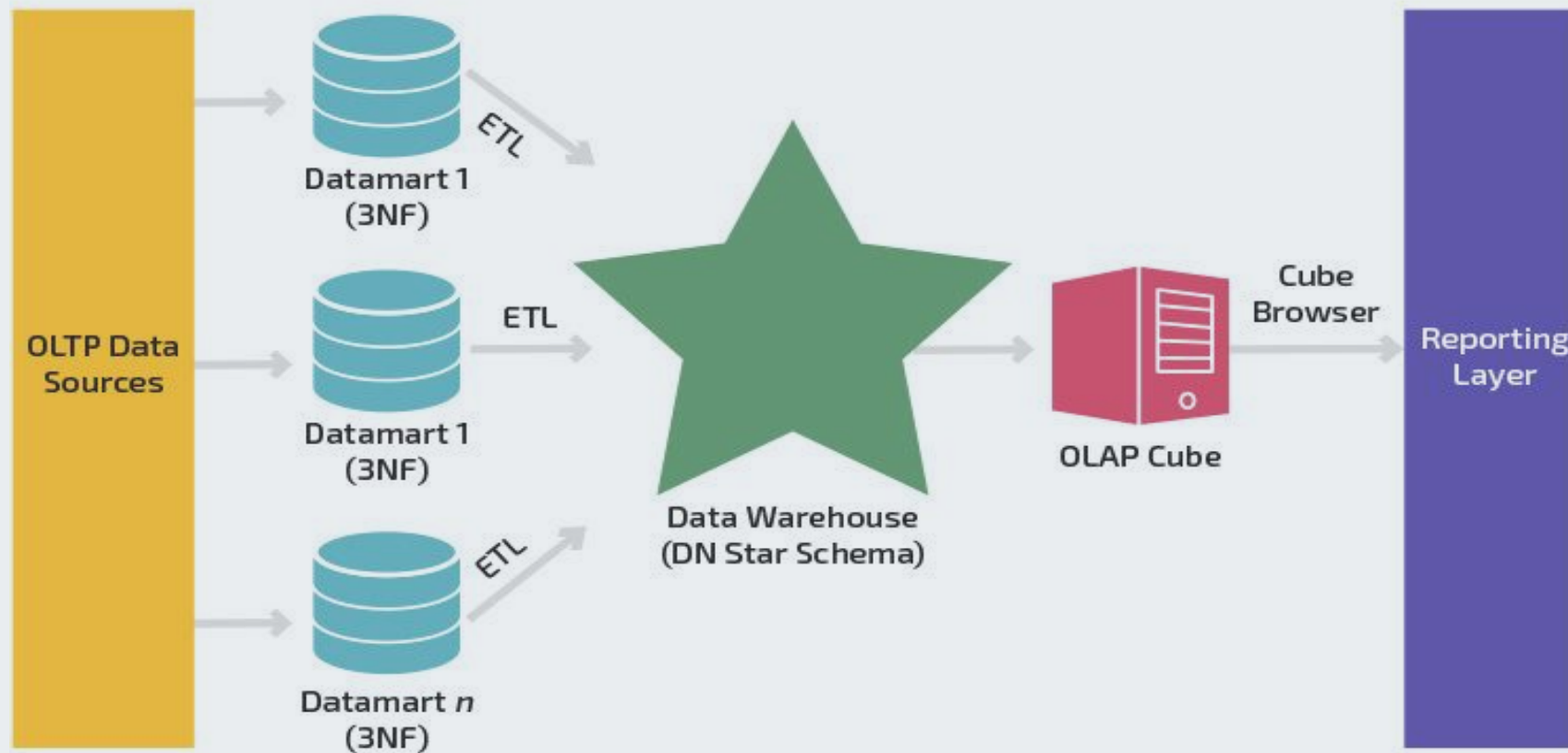


Иммон vs Кимбалл 2/2



Данные представляют собой агрегаты и Хранилище представляет собой Конгломерат витрин данных объединённых между собой

Kimball Model



Use Case

- Маркетинговый анализ эффективности маркетинговых мероприятий;
- Финансовый анализ По прибыли и расходам;
- Анализ продаж в разрезе товаров и бизнес единиц;
- Объединение различных KPI (Key performance Indicators) от различных подразделений компании в связанные отчеты для контроля эффективности
- Анализ маржинальности бизнеса за счет объединения различных бизнес данных с связанных с цепочками продаж, поставок и производства.



OLAP

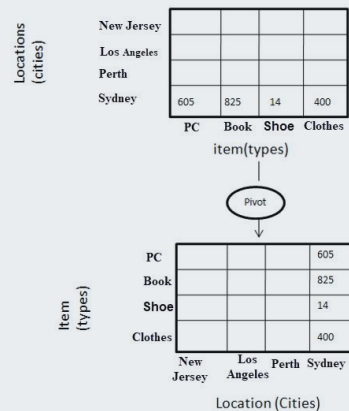
OLAP (*online analytical processing*,) — технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу.

Основной концепт OLAP – это куб данных.

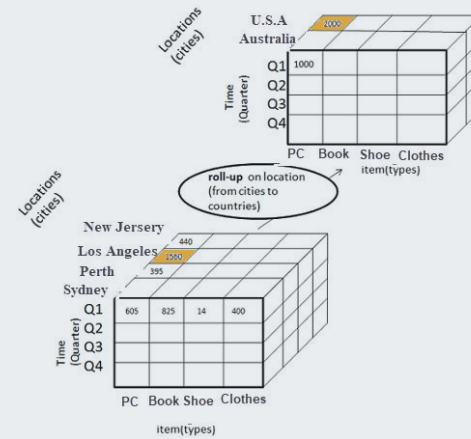
Операции над кубом:

- Roll-up
- Drill-down
- Slice and dice
- Pivot (rotate)

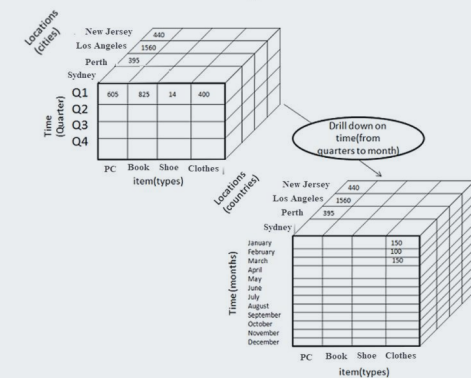
Pivot (rotate)



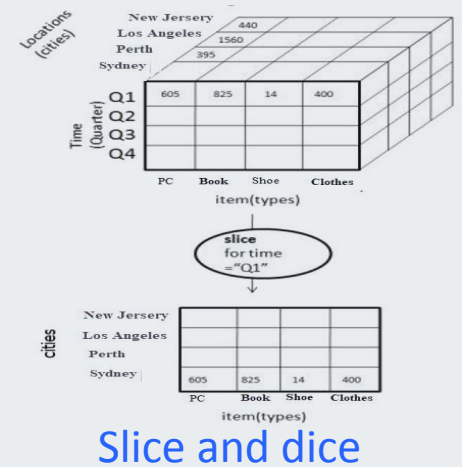
Roll-up



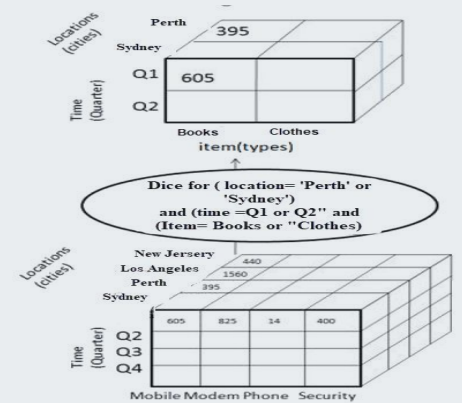
Drill-down



Slice and dice



Slice and dice



BI (Business Intelligence)

