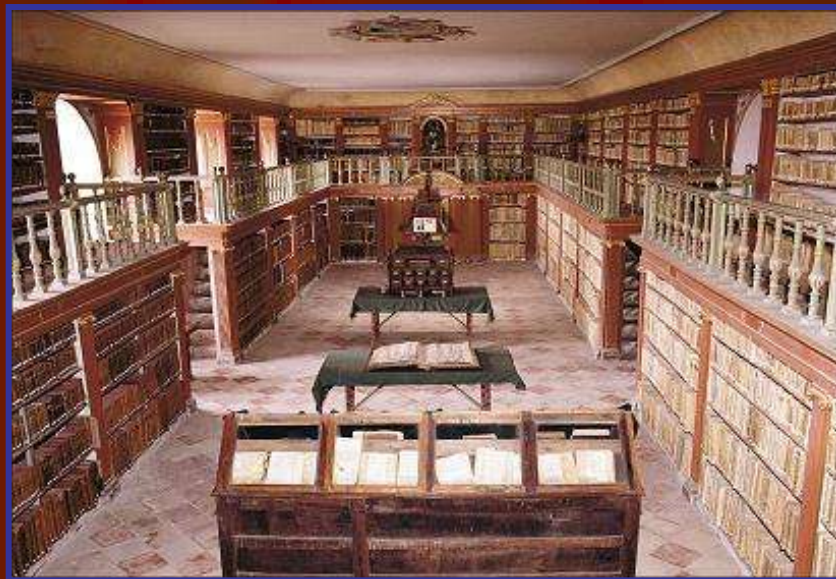
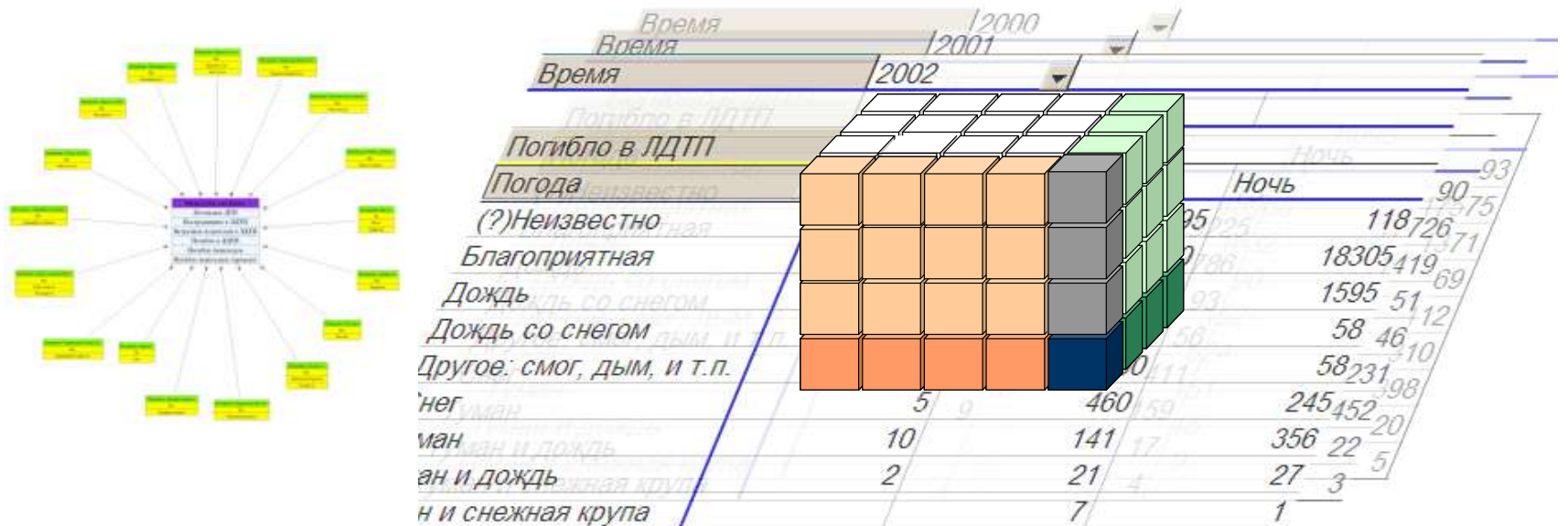


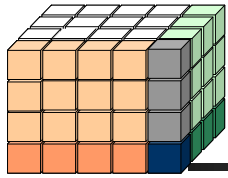
# Гарри Поттер и Хранилище Тайн



# OLAP и Информационные Хранилища

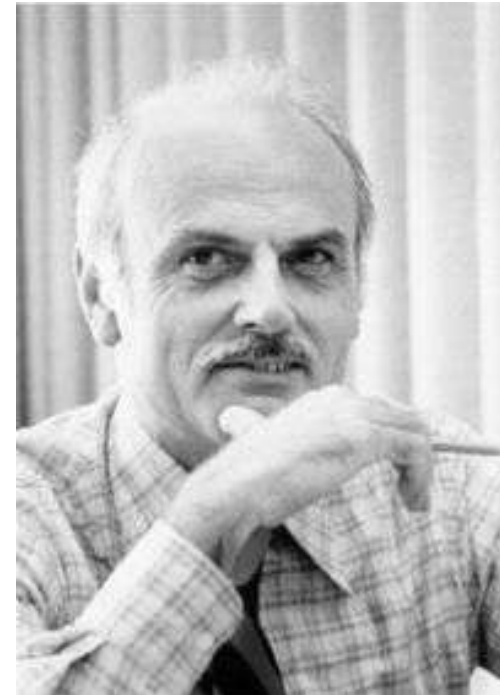


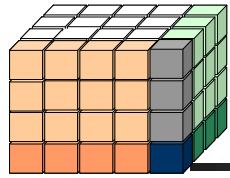
## Основные концепции



# Эдгар Франк «Тед» Кодд

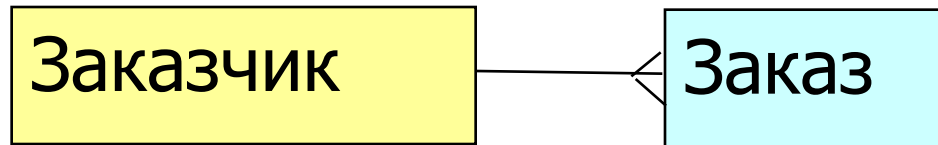
- Математик, химик, боевой пилот ВВС.
- Создатель концепций:
  - реляционной (IBM, 1970 г.) и
  - многомерной баз данных (1993 г.)
- 23 августа 1923 — 18 апреля 2003





# Реляционная модель

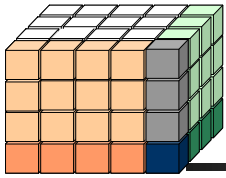
- Модель данных через двумерные таблицы



ID	Фамилия	...
001	Таранов	...
002	Фомин	...
...	...	...

Номер	Дата	Id_customer	...
01	16 ноября 2006	002	...
02	17 ноября 2006	002	...
...	...	...	...

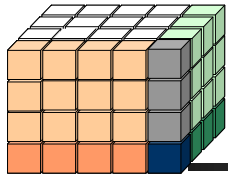




# 12 правил Кодда для РСУБД

- Система должна быть и *реляционной*, и *базой данных*, и *системой управления*.
- Явное представление данных.
- Гарантированный доступ к данным.
- Полная обработка неизвестных значений.
- Доступ к словарю данных в терминах реляционной модели.
- Полнота подмножества языка.
- Возможность модификации представлений.
- Наличие высокоуровневых операций управления данными.
- Физическая независимость данных.
- Логическая независимость данных.
- Независимость контроля целостности.
- Дистрибутивная независимость.
- Согласование языковых уровней.



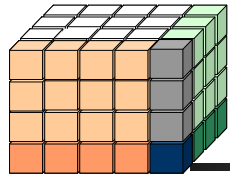


# OLAP-тезисы Кодда (1993)

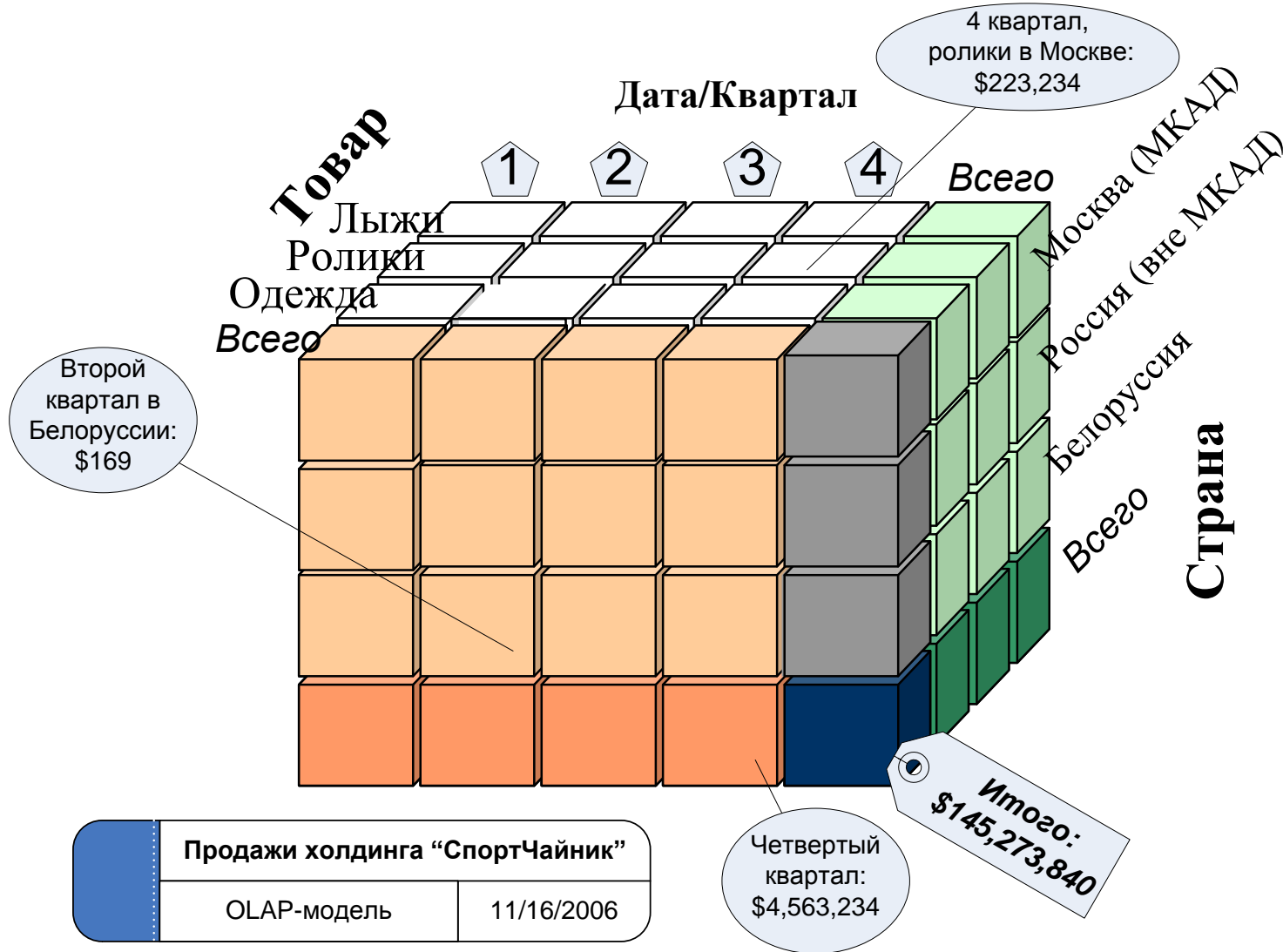
(Теперь входят в критерии FASMI)

1. Многомерность (*Multi-Dimensional Conceptual View*)
2. Прозрачность сервера (*Transparency*);
3. Доступность (*Accessibility*);
4. стабильные доступ и работа (*Consistent Reporting Performance*);
5. архитектура "клиент-сервер" (*Client-Server Architecture*);
6. видовая размерность;
7. управление разреженностью данных (*Dynamic Sparse Matrix Handling*);
8. многопользовательский режим (*Multi-User Support*);
9. операции с измерениями (*Unrestricted Cross-dimensional Operations*);
10. интуитивное манипулирование данными (*Intuitive Data Manipulation*);
11. гибкая запись и редактирование (*Flexible Reporting*);
12. Неограниченная размерность и число уровней агрегации (*Unlimited Dimensions and Aggregation Levels*)



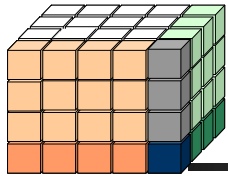


# Многомерная модель

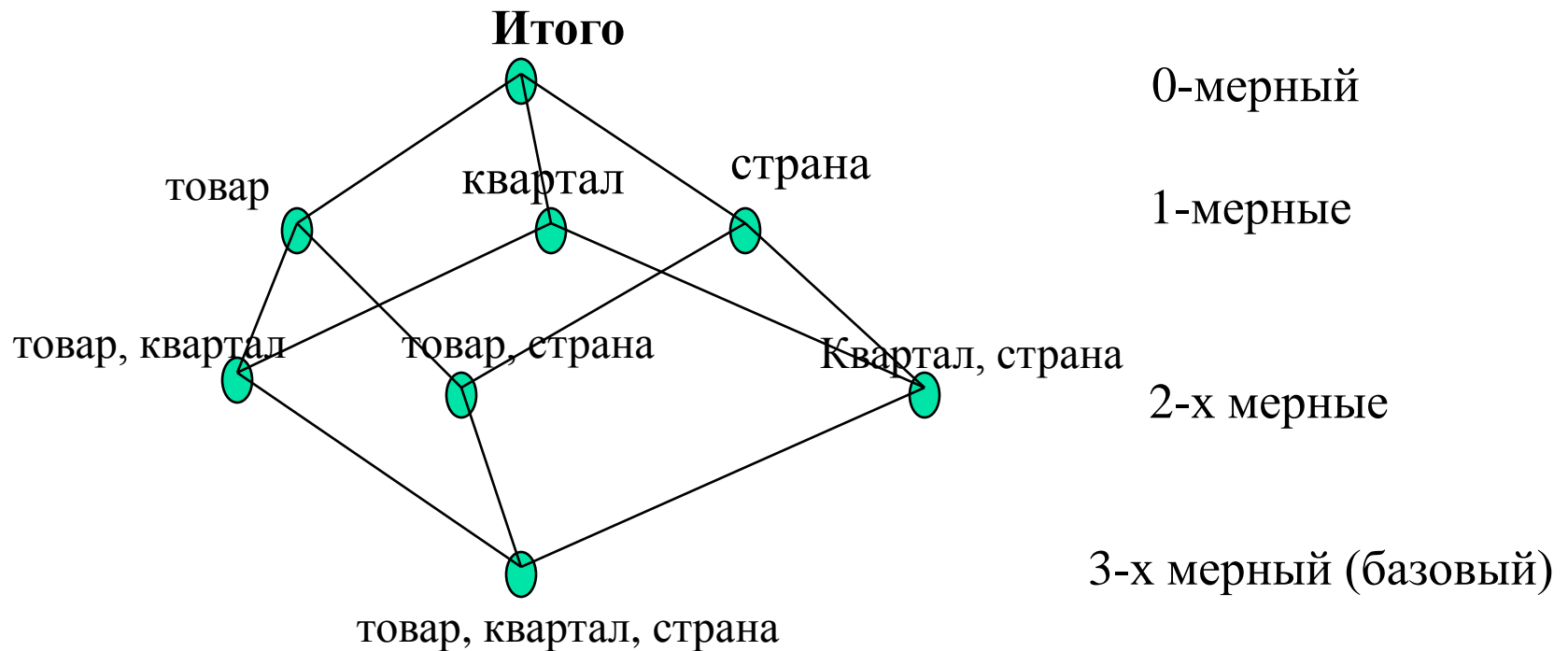


Продажи холдинга "СпортЧайник"	
OLAP-модель	11/16/2006





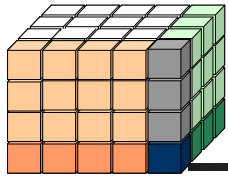
# Кубоиды в кубе



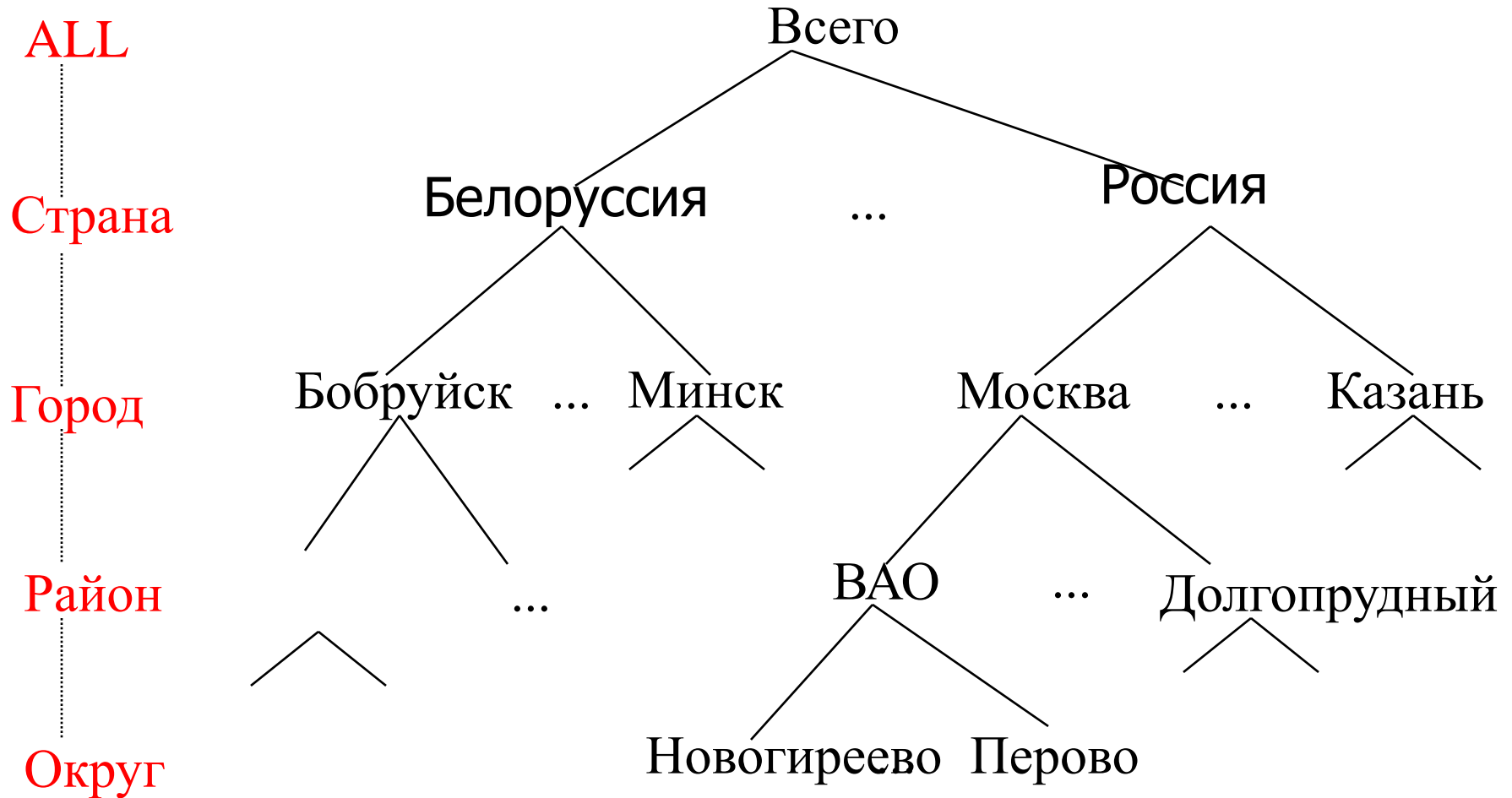
Кубоиды выделяют данные на разных уровнях агрегации.

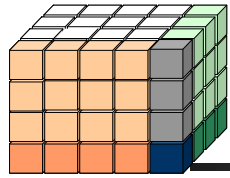




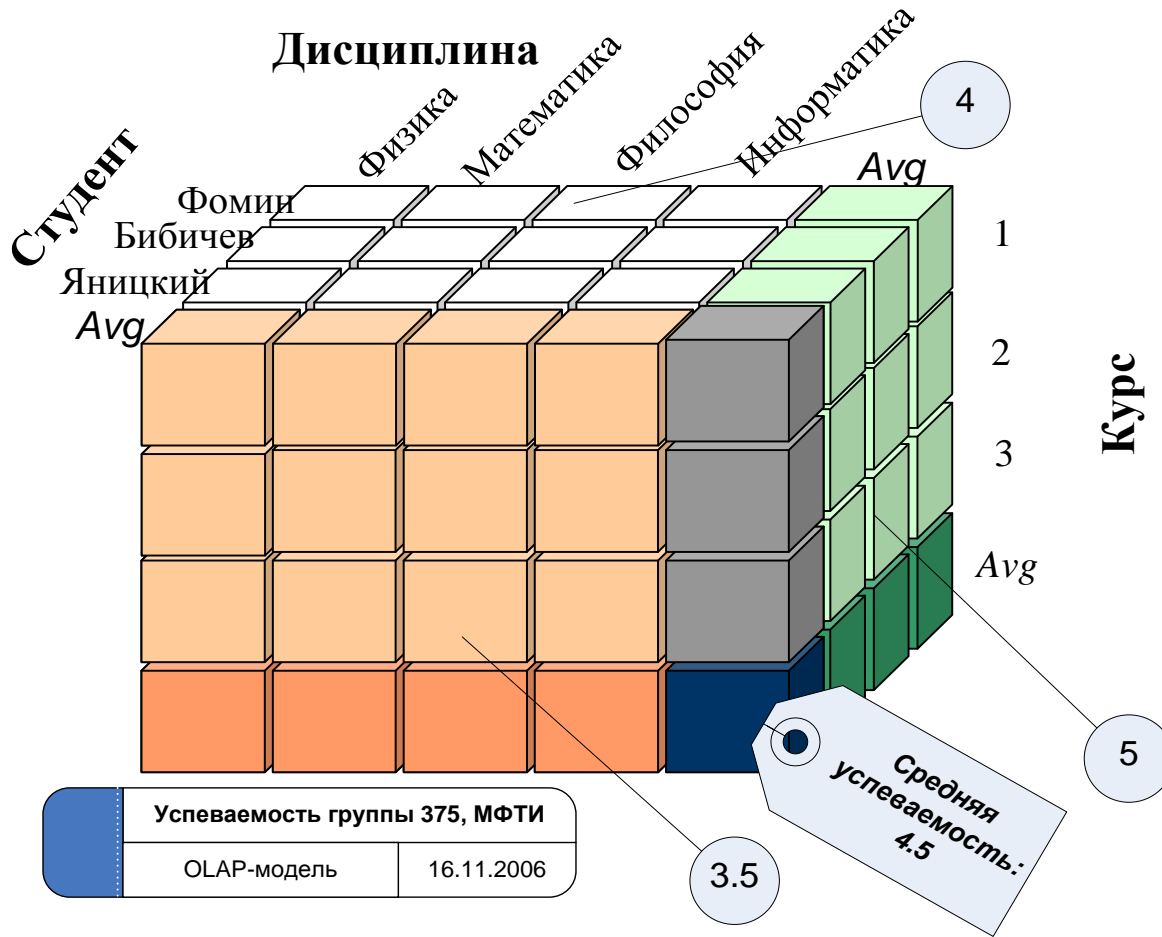


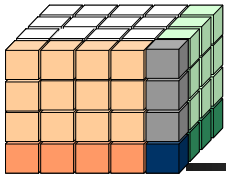
# Иерархии в Измерениях





# Варианты агрегации – AVG (среднее)

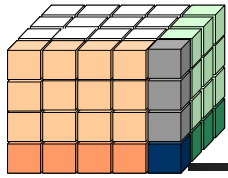




# Основные OLAP-операции

- **Roll up:** агрегация данных: по иерархии(-ям) до полного исключения измерения.
- **Drill down:** детализация: от обобщенных данных к более детальным, от верхних уровней измерений – к нижним, детализация данных по дополнительным измерениям.
- **Slice and dice:** проекции и выборки – выборка нужных “ломтей” кубика
- **Pivot (rotate):** вращение куба, визуализация, выборка и ориентация одно-, двух-, трехмерных срезов для визуального анализа
- Другие операции:
  - **drill across:** кросс-детализация (условно – смена кубов при drilldown)
  - **drill through:** переход с самого нижнего уровня детализации OLAP-куба, к фактам из выбранной ячейки (из исходной реляционной таблицы)

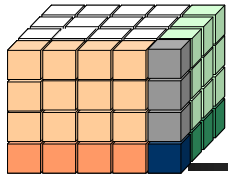




# OLAP vs. OLTP

- OLTP (on-line transaction processing)
  - Основное назначение реляционных СУБД
  - Ежедневные операции: покупки, заказы, производство, регистрация и т.п..
- OLAP (on-line analytical processing)
  - Основное назначение хранилищ данных;
  - Анализ данных и поддержка принятия рациональных решения.

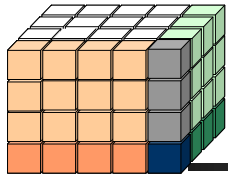




# OLTP vs. OLAP

	<b>OLTP</b>	<b>OLAP</b>
<b>Пользователи</b>	Клерки и IT-шники сопровождения	Эксперт-аналитик (предметник)
<b>Режим работы</b>	Ежедневные операции	При поиске оптимального решения
<b>Архитектура</b>	Ориентировано на приложение	Предметно-ориентированная
<b>Данные</b>	Текущие, актуальные, детализированные, реляционные, нормализованные (безизбыточные).	Исторические, агрегированные, многомерные, консолидированные, денормализованные.
<b>Использование</b>	Однородное, повторяющееся	Априори неизвестное (ad-hoc)
<b>доступ</b>	Чтение/запись, доступ по к отдельным записям по индексам.	Массовые операции над большими объемами.
<b>Элемент доступа</b>	Простые короткие транзакции	Сложные запросы
<b># строк доступа</b>	десятки	миллионы
<b># пользователей</b>	тысячи	сотни
<b>Размер базы</b>	<GB	100GB-TB
<b>Мера производительности</b>	Транзакций в секунду	Скорость выполнения аналитических запросов

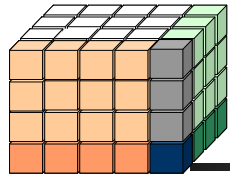




# Разделяй РСУБД и Хранилище!

- Это повысит производительность:
  - РСУБД — настроена на OLTP: методы доступа, индексирование, совместный доступ, восстановление...
  - Хранилище — для OLAP: сложные OLAP запросы, многомерные представления, консолидация данных.
- Различие содержимого и функций:
  - Отсутствующие данные: Анализ для принятия решений (АПР) требует наличие исторических данных, которых может не быть в оперативной базе.
  - Консолидация данных: АПР требует данные из различных источников, возможно включая нереляционные.
  - Качество данных: Консолидация данных из различных источников требует специальной обработки, для приведения их к целостному и совместному виду.

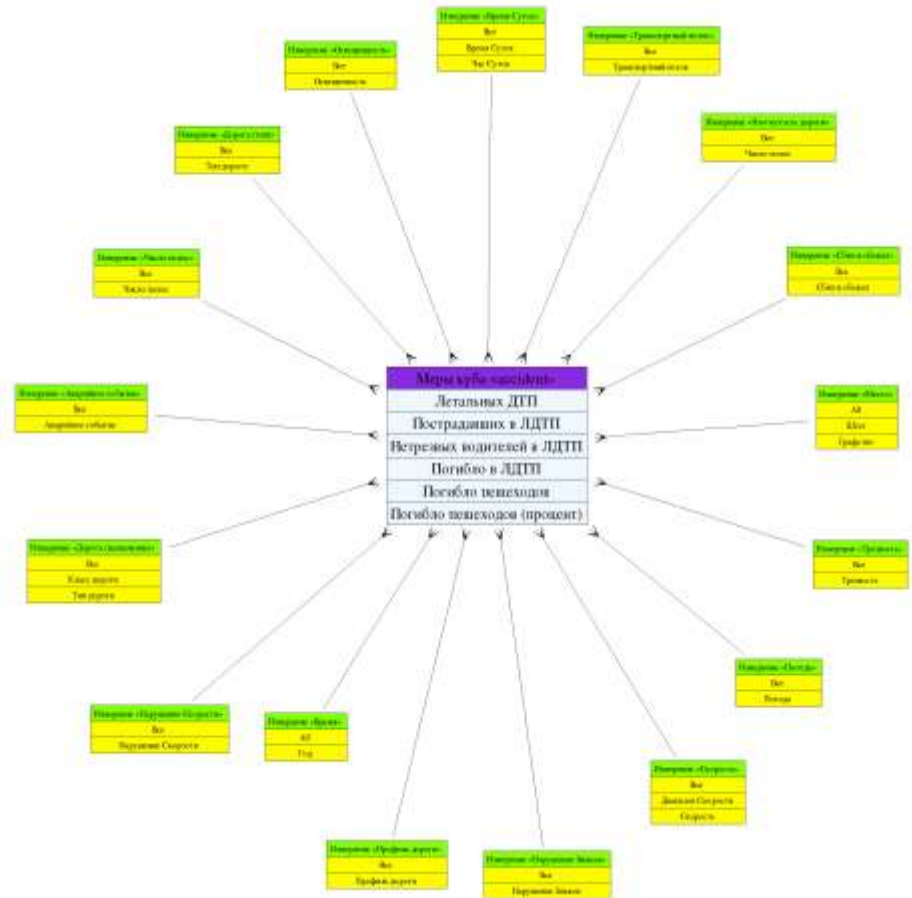


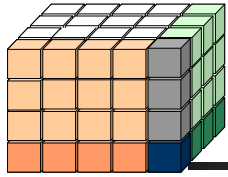


# Многомерная модель

Многомерную модели используют как Информационные хранилища, так и средства OLAP-анализа. Многомерный куб можно представить в реляционной модели, в виде:

- таблицы фактов**, каждая запись которой соответствует ячейке куба,
- и набора **таблиц измерений**, в которых каждая запись – координата в измерении.





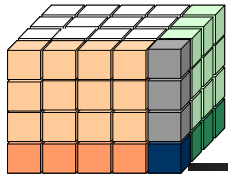
# Реляционные модели хранилища

- Схема "Звезда" ("Star"): Таблица фактов "в середине" соединяется с набором "сателлитов"-таблиц измерений. Все уровни агрегации для каждой координаты являются атрибутами соответствующей записи из таблицы измерений.
- Схема "Снежинка" ("Snowflake"): Базовый кубоид также представляется в схеме "Звезда", но уровни агрегации реляционно нормализованы, и каждый уровень хранится в своей собственной таблице.

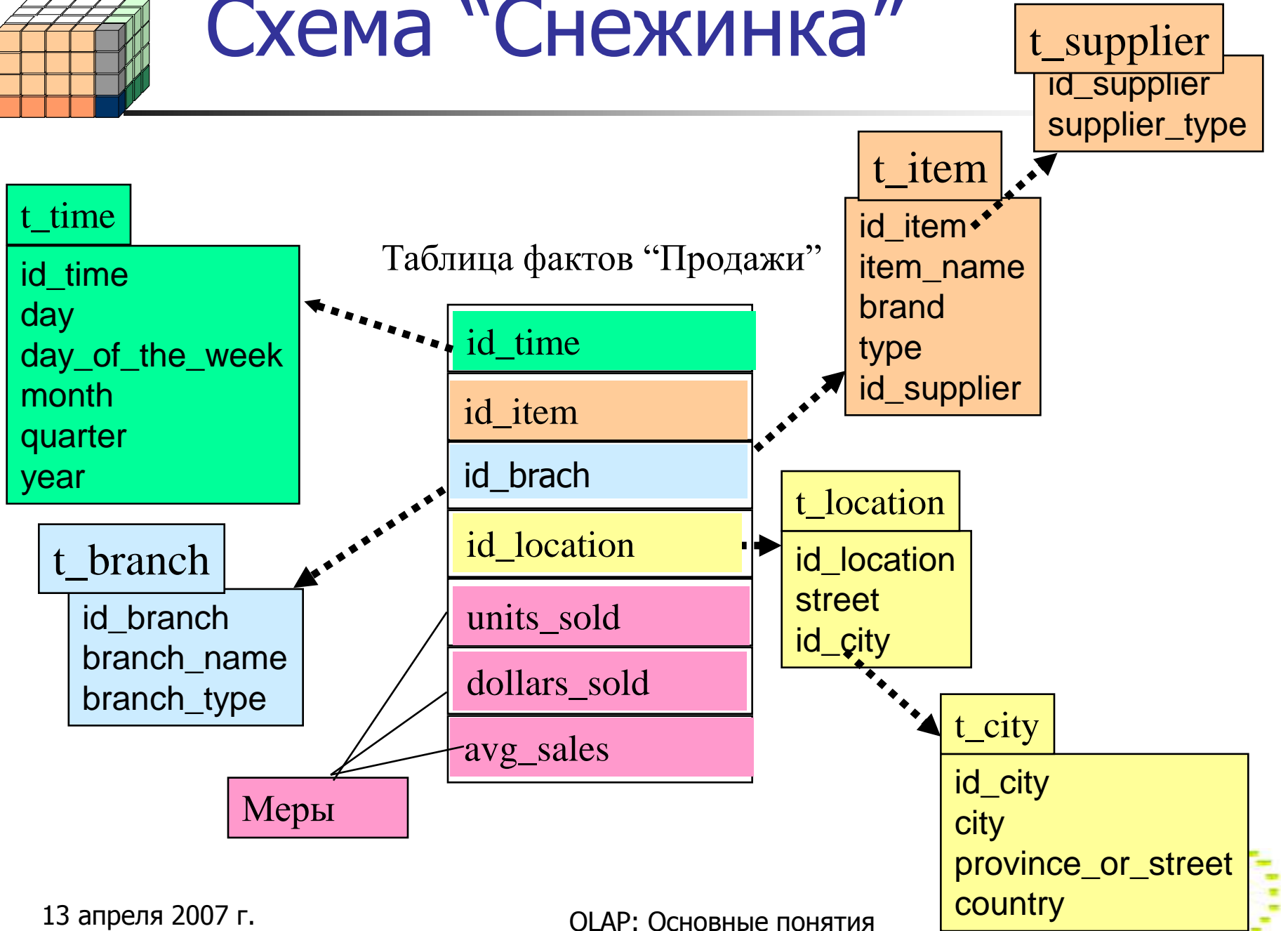


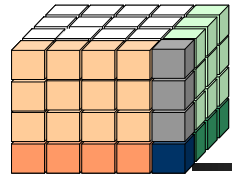




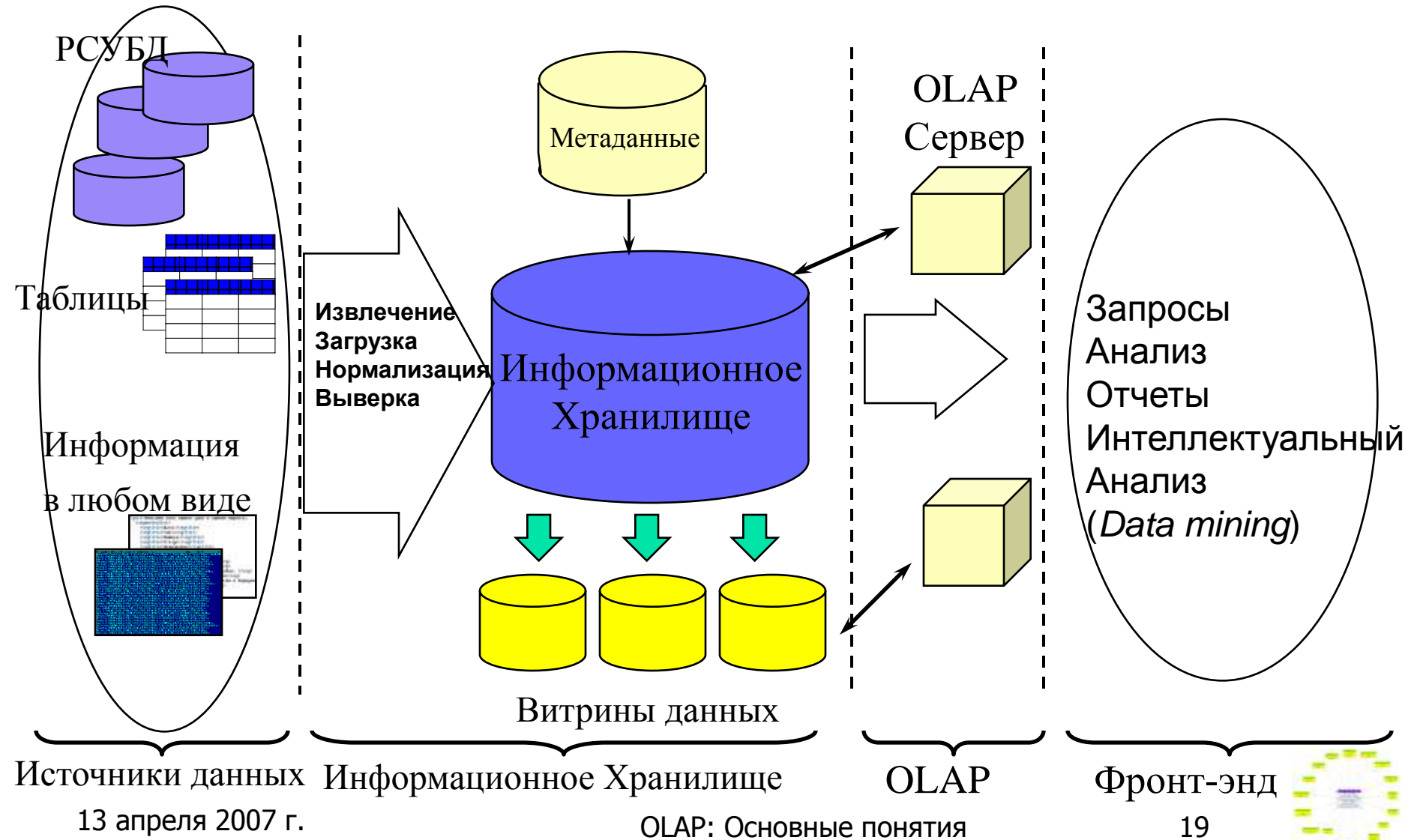


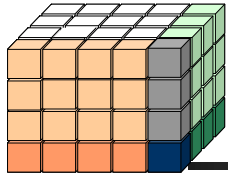
# Схема "Снежинка"





# Архитектура многоуровневого Хранилища

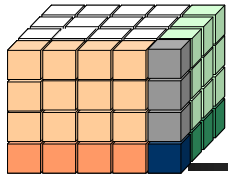




# Модели ИХ

- **Корпоративное ИХ (*Enterprise warehouse*)**
  - Информация о всех предметных областях в компании
- **Витрина данных**
  - Подмножество данных из КИХ, для определенной предметной области или группы пользователей
- **Виртуальное ИХ**
  - Набор представлений (*view*) поверх РСУБД
  - Некоторые представления могут быть материализованы (в форме *materialized views* или обновляемых таблиц).





# OLAP - Архитектуры

- Реляционный OLAP (ROLAP)
  - Используется РСУБД для хранения ИХ.
  - Оптимизируются агрегационные возможности РСУБД
  - (+) Масштабируемость
- Многомерный OLAP (MOLAP)
  - Механизм хранения многомерных массивов (как плотных так и разреженных)
  - (+) Очень быстрый доступ к любым срезам, с произвольной агрегацией
- Гибридный OLAP (HOLAP)
  - $HOLAP = ROLAP + MOLAP$  (масштабируемость+скорость)
  - Нижние уровни (факты) – в реляционной БД, верхние, агрегированные уровни – в кубах.

