

---

# Информационно-аналитические системы

Хранилища данных. OLAP



---

## ▶ Хранилища данных

# Основные определения

---

- ▶ **Хранилище данных (ХД)** — предметно-ориентированный, интегрированный, неизменчивый, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений.
- ▶ **Оперативные источники данных (ОИД)** – источники импорта данных для ХД.

# Предметная ориентированность

---

- ▶ Это **наиболее фундаментальное отличие ХД от ОИД** (оперативного источника данных).
- ▶ Разные ОИД могут содержать данные, описывающие одну и ту же предметную область с различных точек зрения (например бух. учет, склад, плановый отдел и т.д.).
- ▶ ХД позволяет интегрировать информацию отображающую разные точки зрения на одну предметную область.
- ▶ Также это **позволяет хранить только необходимые для анализа данные**, не дублируя лишнюю информацию.

# Интеграция

---

- ▶ ОИД, как правило, разрабатываются в разное время несколькими коллективами с собственным инструментарием.
- ▶ Это приводит к тому, что данные, отражающие один и тот же объект реального мира в разных системах, описывают его по-разному.
- ▶ Обязательная интеграция данных в ХД позволяет решить эту проблему, приведя данные к единому формату.

# Поддержка хронологии

---

- ▶ Данные в ОИД могут **не иметь жесткую привязку** ко времени, поскольку операции над ними происходят в текущий момент.
- ▶ Для полноценного анализа как правило четкая хронология событий является необходимой.
- ▶ ХД решают эту проблему за счет приведения всех дат к единому формату, тем самым обеспечивая возможность хронологического сопоставления событий.

# Неизменяемость

---

- ▶ Во многих ОИД для минимизации объема хранимых данных зачастую устанавливается определенный срок, после которого исторические данные могут быть совсем удалены из системы.
- ▶ В ХД необходимо хранить весь накопленный массив информации.
- ▶ После загрузки данные в ХД только считываются, но не изменяются.
- ▶ Это позволяет в том числе существенно увеличить скорость доступа к данным за счет исключения операций по их модификации.

# Организация ХД

---

- ▶ Хранилища данных разделяются на 3 основные категории:
  - ▶ детальные данные;
  - ▶ агрегированные данные;
  - ▶ метаданные.



# Детальные данные

---

- ▶ **Детальные данные** переносятся непосредственно из ОИД и соответствуют элементарным событиям фиксируемым OLTP-системами.
- ▶ Принято делить все данные на:
  - ▶ **измерения** – наборы данных, необходимые для описания события. Например – площадь квартиры, удаленность от метро и центра города, этаж, кол-во комнат и т.п.;
  - ▶ **факты** – данные отражающие суть события, например, цена продажи квартиры.
- ▶ В процессе эксплуатации если потребность в детальных данных снижается они могут храниться в архивах в сжатом виде на отдельных носителях.

# Агрегированные данные

---

- ▶ На основании детальных данных в ХД могут быть получены агрегированные (обобщенные) данные.
- ▶ В зависимости от возможности агрегации данные делятся на:
  - ▶ **аддитивные** – фактические числовые данные, которые могут просуммированы по всем измерениям;
  - ▶ **полу-аддитивные** – числовые фактические данные, которые могут быть просуммированы только по определенным измерениям;
  - ▶ **неаддитивные** – фактические данные которые не могут быть просуммированы.

# Использование агрегированных данных

---

- ▶ Для обеспечения максимально оперативного доступа к агрегированным данным **часть из них хранится непосредственно в ХД.**
- ▶ Если проводить все расчеты налету, то это очень длительная по времени и накладная с точки зрения загрузки ресурсов операция.
- ▶ Очевидно, что это **ведет к избыточности информации** и увеличению размеров ХД – **важно добиться оптимального соотношения** между вычисляемыми и хранящимися агрегированными данными.
- ▶ **Данные, которые требуются более часто, должны храниться в ХД.**

# Метаданные – данные о данных (1)

---

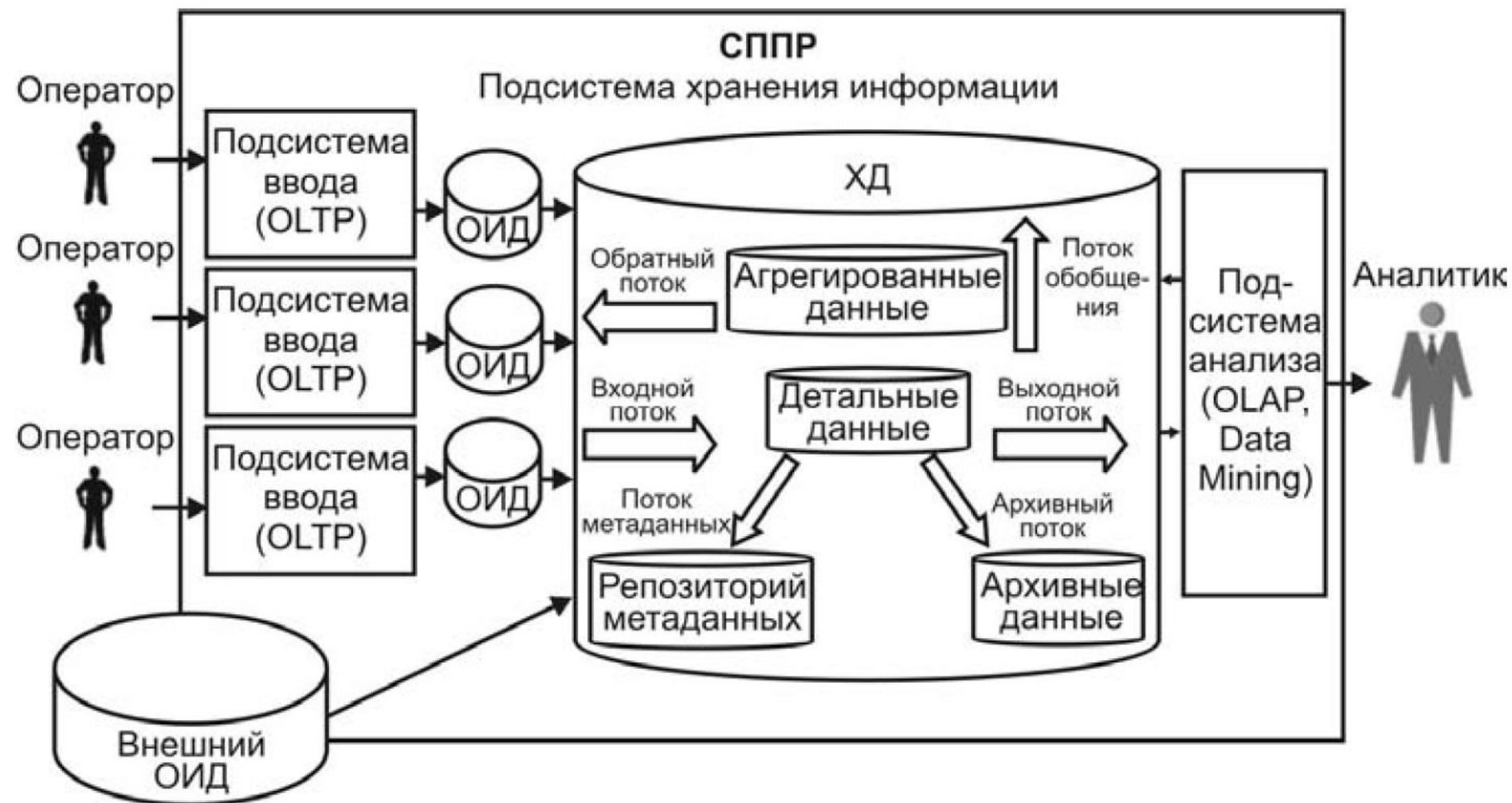
- ▶ Для удобства работы с ХД необходимо также иметь информацию о данных в ХД.
- ▶ Такая информация называется **метаданными** (данные о данных).
- ▶ Метаданные должны отвечать на следующие вопросы [Дж. Захман]:

# Метаданные – данные о данных (2)

---

- ▶ **что?** – описывают объекты предметной области, информация о которых хранится в ХД;
- ▶ **кто?** – описывают категории пользователей использующие данные;
- ▶ **где?** – описывают места хранения (расположение серверов, рабочих станций, ОХ, размещенных на них ПО и т.д.);
- ▶ **как?** – описывают возможные действия над данными;
- ▶ **когда?** – описывают время выполнения операции над данными;
- ▶ **почему?** – причины повлекшие выполнение определенных операций над данными (условия выполнения).

# Схема организации ХД



# Перенос данных в ХД

---

- ▶ При разработке ХД не менее 60% всех затрат связано именно с вопросами переноса данных. Процесс переноса включающий в себя этапы извлечения, преобразования и загрузки называется **ETL-процессом**:
  - ▶ E-extraction – извлечение данных;
  - ▶ T-transformation – преобразование;
  - ▶ L-loading – загрузка.

# Преобразование данных (1)

---

- ▶ **Обобщение данных (aggregation)** — замена многочисленных детальных данных относительно небольшим числом агрегированных данных.
  - ▶ Например, данные о продажах за год занимают в нормализованной базе данных несколько тысяч записей.
  - ▶ После обобщения преобразуются в несколько сотен записей;



# Преобразование данных (2)

---

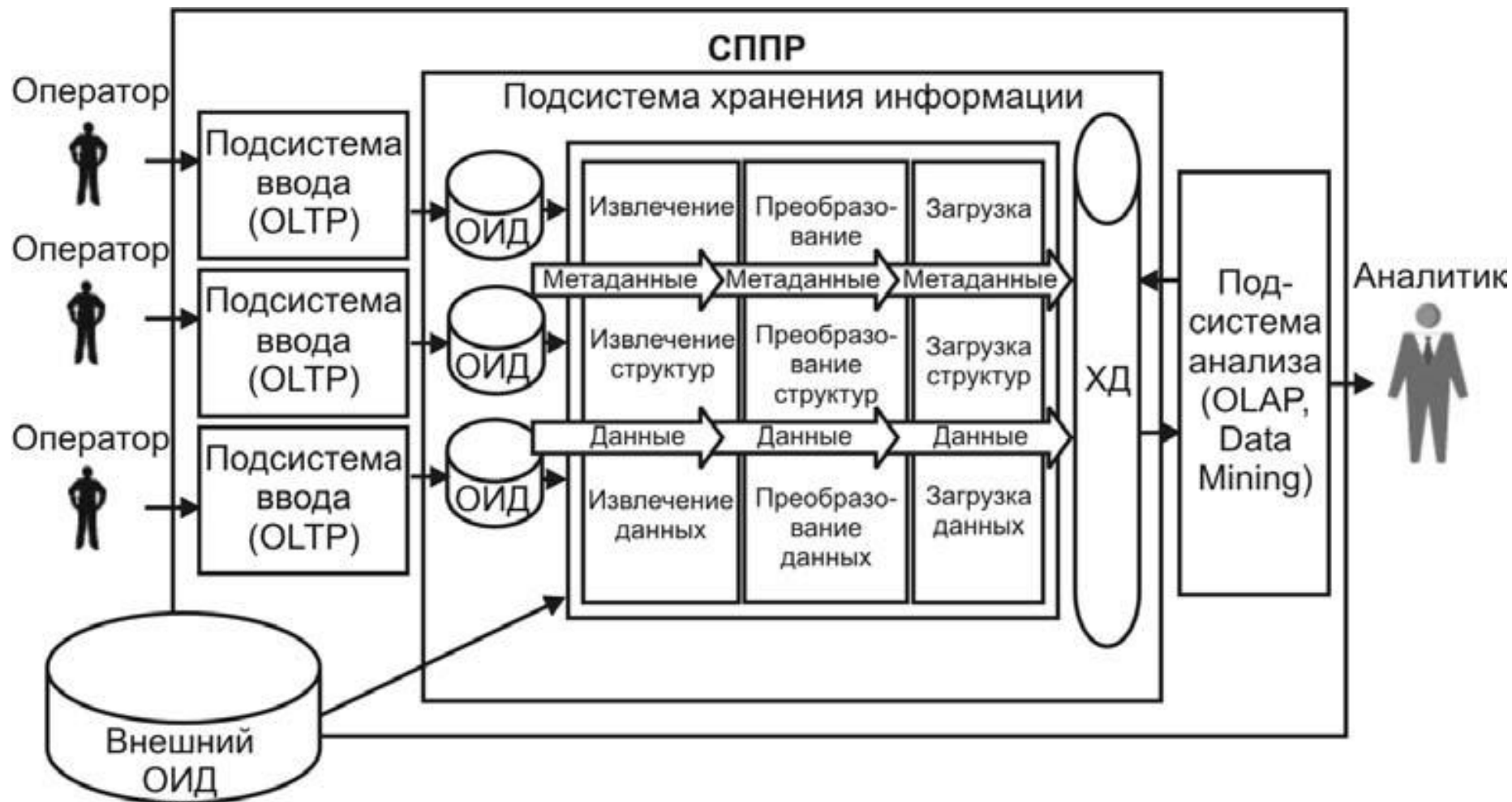
- ▶ **Перевод значений (value translation)** — в ОИД данные часто хранятся в закодированном виде для того, чтобы сократить избыточность данных и память для их хранения.
  - ▶ Названия товаров, городов, специальностей и т. п. могут храниться в ХД в сокращенном виде.
  - ▶ Их заменяют на более понятные описания.

# Преобразование данных (3)

---

- ▶ **Создание полей (field derivation)** — при создании полей для конечных пользователей создается и новая информация.
  - ▶ Например, ОИД содержит два поля: для указания количества проданных товаров и для цены одного экземпляра.
  - ▶ Можно создать специальное поле для хранения суммы.
- ▶ **Очистка данных (cleaning)** — направлена на выявление и удаление ошибок и несоответствий в данных с целью улучшения их качества.

# Схема процесса извлечения данных

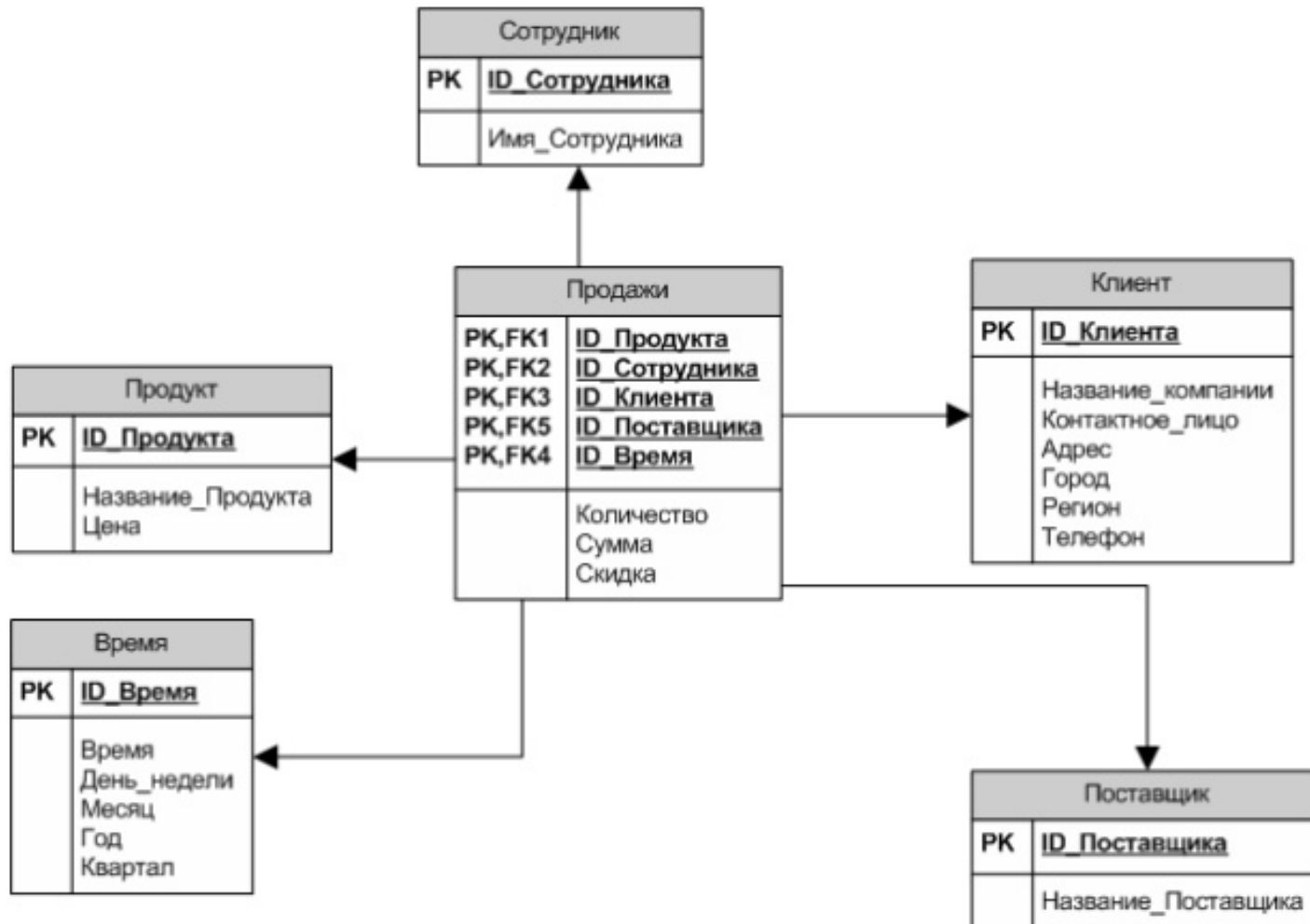


# За что НЕ отвечает концепция ХД

---

- ▶ **Это не концепция анализа данных** – это концепция подготовки данных для анализа;
- ▶ **Не предопределяет архитектуру** целевой аналитической системы.
- ▶ Концепция ХД указывает на то, какие процессы должны выполняться в системе, но не **где конкретно и как они будут выполняться**.
- ▶ В концепции ХД **нет постановки вопросов**, связанных с организацией **эффективного анализа данных** и предоставления доступа к ним. Эти задачи решаются подсистемами анализа.

# Пример схемы хранилища данных



---

## ▶ Многомерный анализ



# Анализ и возникновение гипотез

---

- ▶ Что представляет собой сам процесс принятия решения?
- ▶ У человека, который работает в системе, **возникают некие предположения (гипотезы)** – толи на основании изучения самих данных, толи на основании информации полученной извне.
- ▶ **Что при этом представляет из себя гипотеза?**
- ▶ Как правило, это не что иное как **некая закономерность выявленная в данных.**

# Примеры зависимостей для построения гипотез

---

- ▶ Зависимость объемов продаж товара от: региона, времени, категории товара и т. п.
- ▶ Зависимость количества выздоравливающих пациентов от: применяемых средств, лечения, возраста и т.п.
- ▶ Зависимость цены квартиры от: расстояния до метро , центра города, этажности здания, метража, количества комнат, сезона, курсов валют и т.п.



# Как проверить гипотезу

---

- ▶ Для принятия решения о том, что с ними делать дальше и стоят ли они вообще внимания к себе, **необходима их тщательная проверка, и подтверждение конкретными данными.**
- ▶ Для подтверждения гипотезы **аналитику необходимо максимально аккуратно перелопатить огромные объемы информации**, чтобы подтвердить или опровергнуть свое первоначальное предположение.
- ▶ Очевидно, что **необходим инструмент** который позволит сделать это максимально быстро и наглядно.

# Проверка гипотезы

---

- ▶ Гипотеза может возникнуть **на детальном уровне** – взаимосвязь нескольких фактов, или **на макро уровне** – некая тенденция.
- ▶ Для того чтобы ее подтвердить или опровергнуть необходимо убедиться, что в данных реально проявляются зависимости от различных параметров (обычно не 1, а значительно больше), которые соответствуют нашей гипотезе, или нет.

# Пример гипотезы (1)

---

- ▶ Предположим, выдвигается гипотеза о том, что в предвыборный период количество операций по снятию наличных средств в южных регионах нашей страны увеличилось.
- ▶ Традиционно средства анализа оперируют данными, представленными в виде набора плоских реляционных таблиц.
  - ▶ Причем объемы у нас значительные!

## Пример гипотезы (2)

---

- ▶ Не сложно проверить как себя ведут **потоки денежных средств во времени**.
- ▶ Отдельно посмотреть **в каких регионах** они больше тоже особых проблем нет.
- ▶ Проанализировать **данные в разрезе типов операций** также будет не сложно.
- ▶ Но вот **провести анализ поведения данных сразу по всем трем параметрам** в пределах плоской таблицы будет несколько затруднительно.

# Проблема анализа (1)

---

- ▶ Таким образом: в процессе анализа данных, поиска решений часто **возникает необходимость в построении зависимостей между различными параметрами.**
- ▶ Число таких параметров может варьироваться в широких пределах.
- ▶ Провести анализ поведения данных сразу по всем трем параметрам **в пределах плоской таблицы будет несколько затруднительно.**

# Измерения (1)

---

- ▶ **Измерение** — это последовательность значений одного из анализируемых параметров.
  - ▶ Например, для времени это последовательность календарных дней.
- ▶ Каждое измерение может быть в свою очередь представлено в виде иерархической структуры:
  - ▶ время: год-месяц-день.

## Измерения (2)

---

- ▶ Множественность измерений предполагает представление данных в виде многомерной модели.
- ▶ По измерениям в многомерной модели откладывают параметры, относящиеся к анализируемой предметной области.
- ▶ **Многомерный анализ** – одновременный анализ данных по нескольким измерениям.

# Факты – данные отражающие суть события (1)

---

- ▶ Факты, связанные с транзакциями (Transaction facts).
  - ▶ основаны на отдельных событиях: телефонный звонок или снятие денег со счета с помощью банкомата;
- ▶ факты, связанные с "моментальными снимками" (Snapshot facts).
  - ▶ основаны на состоянии объекта: банковского счета в определенные моменты времени, например, на конец дня или месяца.



## Факты – данные отражающие суть события (2)

---

- ▶ факты, связанные с элементами документа (Line-item facts).
  - ▶ основаны на том или ином документе: счет за товар или услуги и содержат подробную информацию об элементах этого документа, например, о количестве, цене, проценте скидки;

# Модель данных

---

- ▶ Многомерная модель данных обычно организуется вокруг выбираемой заранее общей «темы», например, «продажи». Это отражается в таблице фактов.
- ▶ Рассмотрим пример куба «Sales» («Продажи»). Данные в нем могут быть рассмотренные в разных измерениях:
  - ▶ таблицами измерений выступают, например, item (item\_name, brand, type), или time(day, week, month, quarter, year);
  - ▶ таблица фактов может содержать поле dollars\_sold или units sold и внешние ключи.

# Кубоиды. 2D-куб

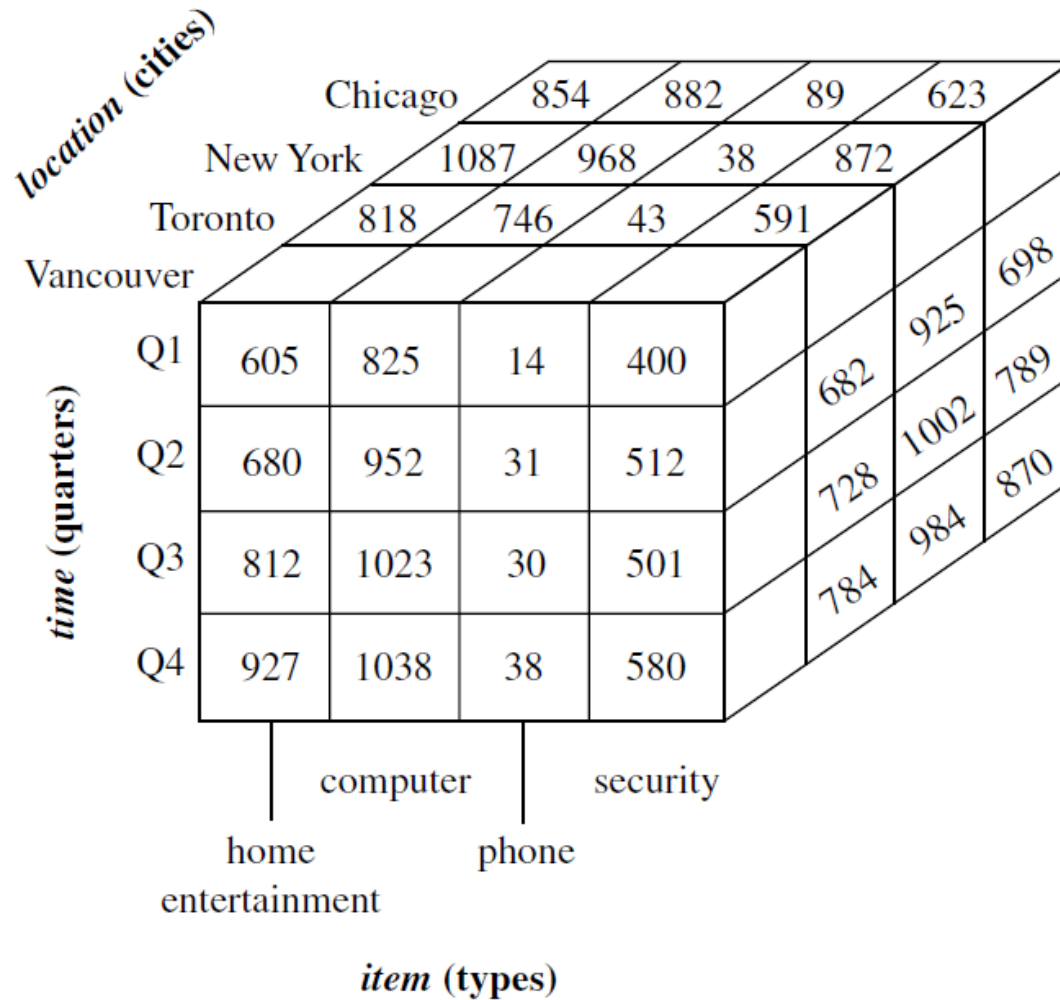
<b>location</b> = “Vancouver”				
<b>time</b> (quarter)	<b>item</b> (type)			
	home	computer	phone	security
Q1	605	825	14	400
Q2	680	952	31	512
Q3	812	1023	30	501
Q4	927	1038	38	580

# Кубоиды. 3D-куб (1)

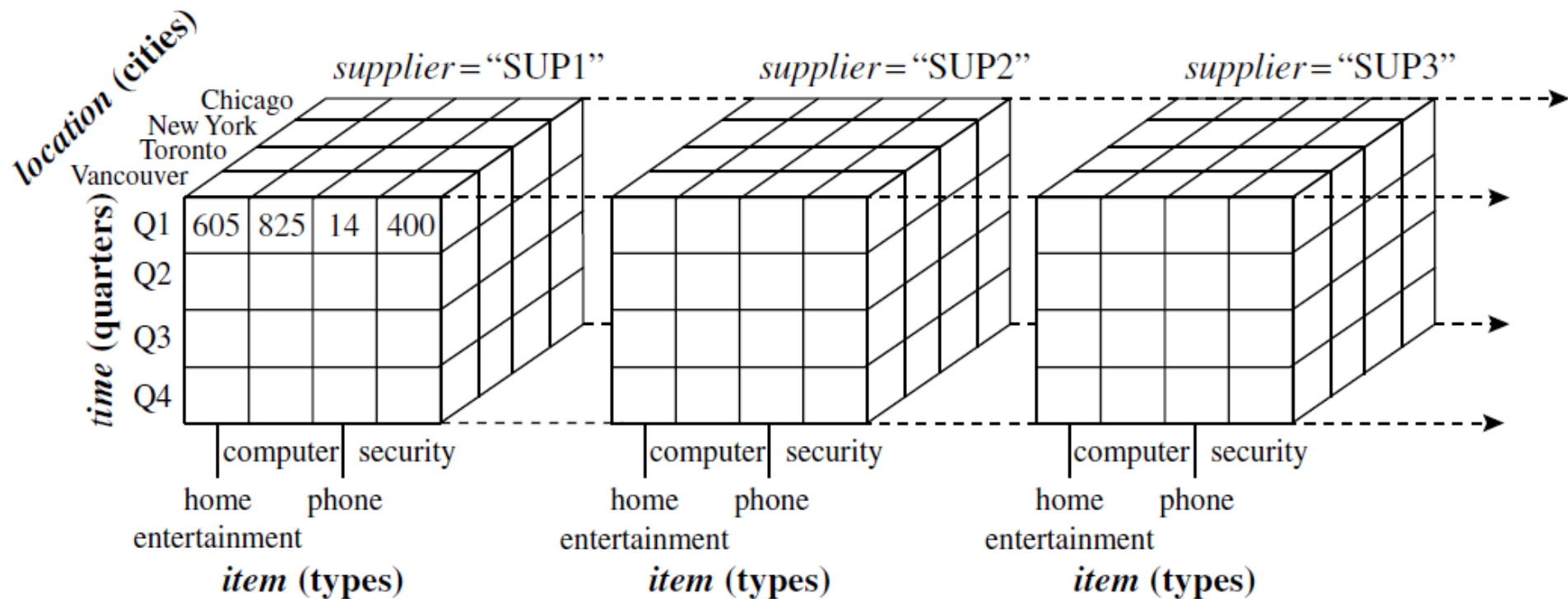
<i>location</i> = “Chicago”					<i>location</i> = “New York”					<i>location</i> = “Toronto”					<i>location</i> = “Vancouver”				
<i>item</i>					<i>item</i>					<i>item</i>					<i>item</i>				
<i>home</i>					<i>home</i>					<i>home</i>					<i>home</i>				
<i>time</i>	<i>ent.</i>	<i>comp.</i>	<i>phone</i>	<i>sec.</i>	<i>ent.</i>	<i>comp.</i>	<i>phone</i>	<i>sec.</i>		<i>ent.</i>	<i>comp.</i>	<i>phone</i>	<i>sec.</i>		<i>ent.</i>	<i>comp.</i>	<i>phone</i>	<i>sec.</i>	
Q1	854	882	89	623	1087	968	38	872		818	746	43	591		605	825	14	400	
Q2	943	890	64	698	1130	1024	41	925		894	769	52	682		680	952	31	512	
Q3	1032	924	59	789	1034	1048	45	1002		940	795	58	728		812	1023	30	501	
Q4	1129	992	63	870	1142	1091	54	984		978	864	59	784		927	1038	38	580	



## Кубоиды. 3D-куб (2)



# Кубоиды. 4D-куб



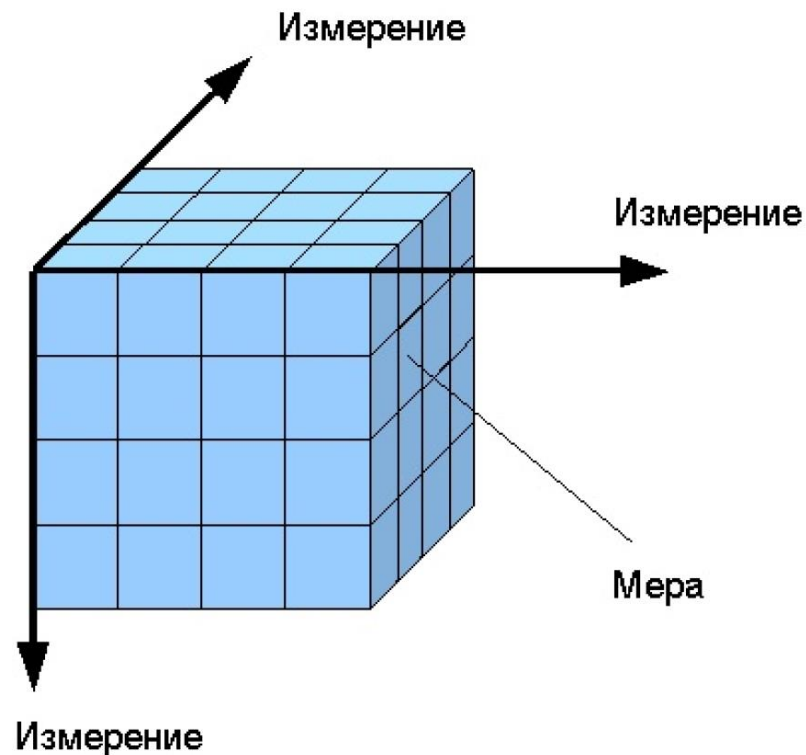
# Кубоиды

---

- ▶ Выше были показаны кубоиды разной степени обобщения.
- ▶ **Базовый кубоид** –  $n$ -D куб ( $n$ -мерный куб) имеющий наименьшую степень обобщения.
- ▶ **«Apex cuboid»** – самый верхний 0-D кубоид, содержащий самый верхний уровень обобщения.
  - ▶ В приведенных примерах, это будет сумма всех продаж по всем измерениям.
- ▶ При заданном наборе измерений можно на основе базового кубоида сгенерировать кубоиды для всех возможных подмножеств набора измерений.

# Многомерный анализ. Гиперкубы

- ▶ На пересечении осей измерений располагаются данные, количественно характеризующие **анализируемые факты (меры)**.
- ▶ Множественность измерений предполагает представление данных в виде **многомерной модели**.
- ▶ Таким образом многомерную модель данных можно представить как **ГИПЕРКУБ**.



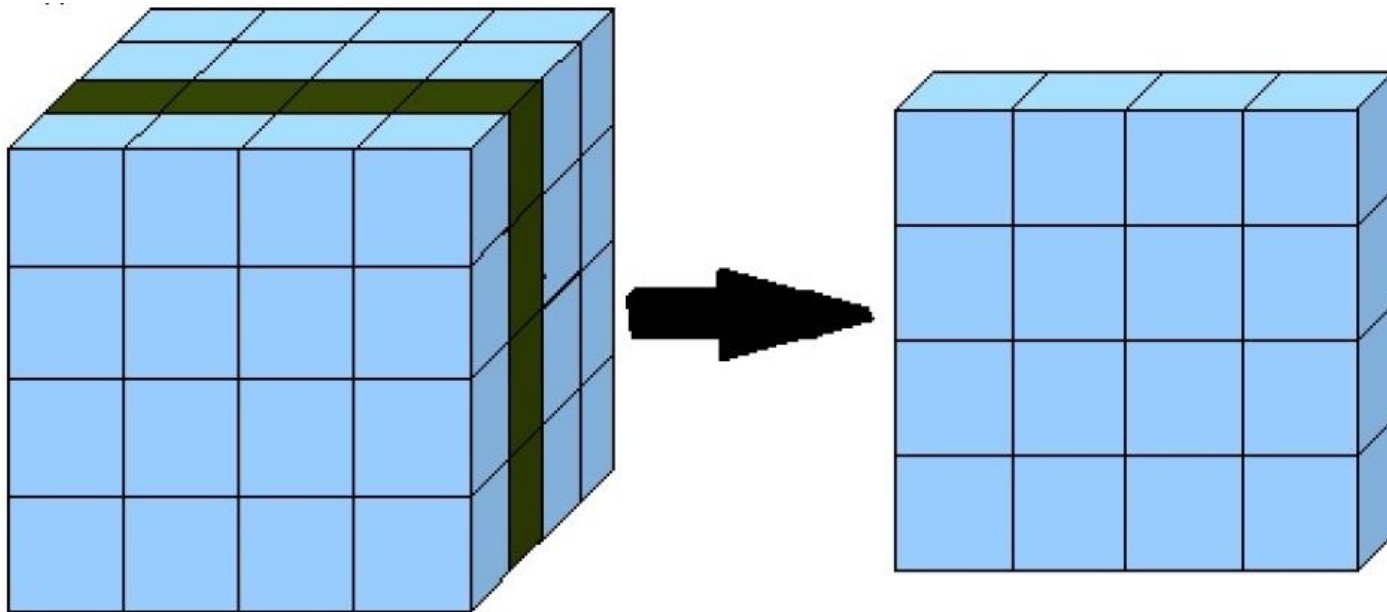


# Операции с гиперкубом.

## Срез

---

- ▶ **Срез** – формирование подмножества многомерного массива данных, соответствующего единственному значению одного или нескольких элементов измерений.

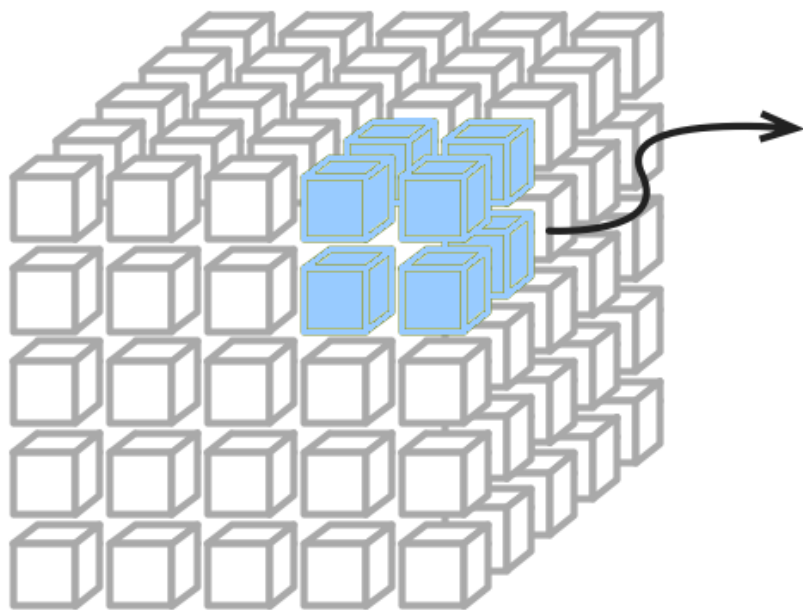


# Общий случай среза – ограничение.

---

Часто также используется термин **dicing**.

- ▶ В некоторых случаях бывает полезно наложить некоторые ограничения на часть или все измерения гиперкуба.
- ▶ Это позволяет сократить его размер, а также изолировать часть данных от аналитика, которые он либо не может просматривать, либо они для него не интересны.



---

## ▶ OLAP-системы

# Определения

---

- ▶ **OLAP (On-Line Analytical Processing)** – оперативная аналитическая обработка – технология оперативной аналитической обработки данных, использующая методы и средства для сбора, хранения и анализа многомерных данных в целях поддержки процессов принятия решений.
- ▶ **OLAP** – это совокупность концепций, принципов и требований, лежащих в основе программных продуктов, облегчающих аналитикам доступ к данным.

# OLAP-системы

---

- ▶ **Основное назначение OLAP-систем** — поддержка аналитической деятельности, произвольных запросов пользователей-аналитиков.
- ▶ **Цель OLAP-анализа** — проверка возникающих гипотез.
- ▶ У истоков технологии OLAP стоит основоположник реляционного подхода Э. Кодд.
- ▶ В 1993 г. он опубликовал статью под названием "OLAP для пользователей-аналитиков: каким он должен быть".
- ▶ В своей работе он предложил 12 требований, которым должны удовлетворять продукты, позволяющие выполнять оперативную аналитическую обработку.

# Основные вехи развития OLAP

---

- ▶ 1962 - Публикация "**A Programming Language**" Кена Айверсона (IBM) - Первый многомерный язык программирования;
- ▶ 1970 - Появился **Express**. Первый многомерный продукт, ныне - собственность компании Oracle;
- ▶ 1982 - **Comshare System W**. Первый OLAP инструмент, ориентированный на финансовые приложения. На рынке больше не предлагается, позднее Essbase использовал многие его концепции;
- ▶ 1984 - Запущен **Metaphor**. Первый ROLAP;
- ▶ 1990 - **Cognos PowerPlay**. Первый OLAP для Windows и первым настольным OLAP. Сегодня лидирует в секторе настольных OLAP;
- ▶ 1992 - Выпущен **Essbase**. Первый OLAP продукт, имеющий хороший рынок;
- ▶ 1993 - Напечатана **статья Кодда** с определением OLAP;
- ▶ 1994 - **MicroStrategy DSS Agent**. Первый ROLAP без многомерной СУБД, почти вся обработка выполняется с помощью множества SQL-запросов;
- ▶ 1995 - Создан Holos 4.0. **Первый HOLAP**;
- ▶ 1999 - Выпущен **Microsoft OLAP Services** .

# Примеры задач анализа

---

- ▶ Какие наши самые прибыльные клиенты?
- ▶ Кто наши клиенты и какие продукты они покупают?
- ▶ Как новый продукт повлияет на выручку и прибыль?
- ▶ Какой вид рекламы больше всего влияет на прибыль?
- ▶ Какой канал продаж наиболее эффективен?



# Применение OLAP на практике (1)

---

- ▶ **Анализ финансовых показателей** деятельности предприятия
  - ▶ динамика остатков и оборот организации, структура доходов и их распределение по клиентам, товарам, дням недели, месяцам, кварталам, за год и т.д.
- ▶ **Корпоративная отчетность**
  - ▶ В распределенной организации филиалы регулярно передают данные в центральный офис. Здесь данные попадают в единое Хранилище. Над ними выполняются дополнительные расчеты, для которых в филиалах нет данных.



# Применение OLAP на практике (2)

---

## ► Анализ бюджетных данных

- Сотрудники бюджетно-аналитических и планово-экономических подразделений выпускают «план-факт» отчеты об исполнении бюджетов за месяц, квартал, год, анализируют бюджетные планы в разрезе центров финансовой ответственности и бизнес-направлений, детализируют значения бюджетных статей и т.д.

# Применение OLAP на практике (3)

---

- ▶ **Анализ клиентской базы.** На основе данных о клиентах строится отраслевой и географический портрет покупателей товаров и потребителей услуг, оценивается динамика клиентской базы, влияние маркетинговых мероприятий на увеличение числа клиентов, выявляются клиенты, которые приносят 80% прибыли, исследуется история продаж и т.д.
- ▶ **Анализ складских данных**
  - ▶ «Сколько продукции было закуплено заказчиком Ивановым в третьей декаде сентября?»;
  - ▶ «Каков оптимальный объем активных и резервных запасов по данной товарной позиции?»;
  - ▶ «Существуют ли сезонные колебания по данному типу товаров и какова их амплитуда?» и т.п.

# Применение OLAP на практике (4)

---

- ▶ Анализ продаж
  - ▶ «Сколько товара продано?», «На какую сумму продано?»;
  - ▶ Эти запросы расширяются по мере усложнения бизнеса и накопления исторических данных:
  - ▶ «... в Москве, в Сибири?»;
  - ▶ «... в прошлом квартале, по сравнению с нынешним?»;
  - ▶ «... через магазин А, по сравнению с магазином Б?».
- ▶ Анализ закупок и цен
- ▶ Анализ посещаемости Web-сайта
- ▶ Публикация маркетинговых исследований
- ▶ Создание информационного сервиса

# Требования к продуктам OLAP (1)

---

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

# Требования к продуктам OLAP (2)

---

- ▶ **Доступность.** OLAP-система должна предоставлять пользователю единую, согласованную и целостную модель данных, обеспечивая доступ к данным независимо от того как и где они хранятся.
- ▶ **Постоянная производительность при разработке отчетов.**
  - ▶ Производительность OLAP-систем не должна значительно уменьшаться при увеличении количества измерений, по которым выполняется анализ.

# Требования к продуктам OLAP (3)

---

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

# Требования к продуктам OLAP (4).

## Динамическое управление разреженными матрицами

---

- ▶ OLAP-система должна обеспечить 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.

# Требования к продуктам OLAP (5)

---

- ▶ **Гибкие возможности получения отчетов.** OLAP-система должна поддерживать различные способы визуализации данных, т.е. средства формирования отчетов должны представлять синтезируемые данные или информацию, следующую из модели данных, в ее любой возможной ориентации.
- ▶ **Неограниченная размерность и число уровней агрегации.** Исследование о возможном числе необходимых измерений, требующихся в аналитической модели, показало, что одновременно могут использоваться до 19 измерений.



# Сравнение подходов OLTP и OLAP (1)

КАТЕГОРИЯ	ONLINE TRANSACTIONAL PROCESSING (OLTP)	ONLINE ANALYTICAL PROCESSING (OLAP)
<b>Источник данных</b>	Операционная информация приложения	Исторические и архивные данные
<b>Назначение</b>	Обновление данных	Отчеты и поиск информации
<b>Приложения</b>	Управление, операционные задачи, веб-сервисы, клиент-серверные	Системы управления. Отчеты. Принятие решений
<b>Пользователи</b>	Менеджеры, и другой рядовой персонал	Аналитики, управляющие, исследователи данных, маркетологи
<b>Число пользователей</b>	Тысячи	Десятки
<b>Назначение данных</b>	«Снимок» текущих бизнес-процессов	Многомерные представления различных видов бизнес-деятельностей.
<b>Обновление данных</b>	Вставка, обновление, удаление. Исполняются быстро. Выдают результаты сразу же.	Обновление данных большими наборами записей. Занимает много времени. Носит случайных характер



# Сравнение подходов OLTP и OLAP (2)

КАТЕГОРИЯ	ONLINE TRANSACTIONAL PROCESSING (OLTP)	ONLINE ANALYTICAL PROCESSING (OLAP)
<b>Модель данных</b>	Entity-relationship on databases	Одно- или многомерные данные
<b>Схема данных</b>	Нормализованные таблицы, много таблиц и отношений.	Звезда, снежинка, созвездие. Немного не нормализованных таблиц
<b>Доступ к данным</b>	Индексы, первичные ключи	Сканирование
<b>Резервное копирование</b>	Постоянное резервное копирование, архивирование. Данные являются критически важными и не должны быть потеряны	Простая перезагрузка данных
<b>Горизонт хранения</b>	Текущий день, недели, месяцы	Длительные периоды, годы.
<b>Запросы</b>	Простые	Комплексные запросы для сбора информации и составления отчетов
<b>Скорость</b>	Большая скорость. Требует построения индексов на больших данных	Низкая, зависит от объемов данных, требует больше индексов
<b>Объем данных</b>	Операционные данные, обычно не большие объемы	Большие объемы исторических данных



# Тест FASMI

---

- ▶ Тогда же в 1995 году Пендсом и Критом на основе правил Кодда был создан тест **FASMI** (Fast of Analysis Shared Multidimensional Information). В этом тесте **OLAP определяется через 5 ключевых понятий.**
- ▶ Таким образом, они определили OLAP следующими пятью ключевыми словами: **FAST** (Быстрый), **ANALYSIS** (Анализ), **SHARED** (Разделяемой), **MULTIDIMENSIONAL** (Многомерной), **INFORMATION** (Информации).

# Fast (Быстрый)

---

- ▶ Система должна обеспечивать выдачу большинства ответов пользователю в пределах 5 секунд. Причем простые запросы должны обрабатываться не более 1 секунды, и лишь наиболее сложные и редкие – более 20 секунд.
- ▶ если система не реагирует на запрос более 30 секунд, пользователи воспринимают это как отсутствие результата или зависание программы.
- ▶ Если система заранее не предупреждает пользователя о необходимости длительного ожидания результата, то скорее всего работа программы будет остановлена на пол пути.

# Analysis (анализ)

---

- ▶ Система должна без проблем справляться с любым логическим и статистическим анализом, характерным для конкретного приложения.
- ▶ Результаты должны сохраняться в виде понятном и доступном для конечных пользователей.
- ▶ Система должна позволять пользователю самому настраивать правила производимых при анализе вычислений и формы представления итоговой информации.
- ▶ Без необходимости программирования, только путем работы с пользовательским интерфейсом.

# Shared (разделяемой или распределенной)

---

- ▶ Система должна удовлетворять всем требованиям защиты информации в многопользовательской системе, вплоть до защиты каждой конкретной ячейке информации.
- ▶ Если к записи необходим множественный доступ, то необходимо обеспечить защиту информации от модификации.
- ▶ Также необходимо обеспечить корректный механизм обработки множественных модификаций.

# Multidimensional (многомерной)

---

- ▶ Технология должна обеспечить концептуальное многомерное представление данных, включая полную поддержку для иерархий и множественных иерархий, обеспечивающих наиболее логичный способ анализа.
- ▶ Это требование не устанавливает минимальное число измерений, которые должны быть обработаны, поскольку этот показатель зависит от приложения.
- ▶ Оно также не определяет используемую технологию БД, если пользователь действительно получает многомерное концептуальное представление информации.

# Information (информации)

---

- ▶ OLAP должна обеспечить получение ВСЕЙ необходимой информации в условиях реальной системы.
- ▶ В этом плане системы OLAP различаются по своей мощности.
- ▶ Но Мощность систем измеряется не объемом хранимой информации, а количеством входных данных которые они могут обработать.
- ▶ Причем разница в случае с OLAP система изменяется не в разы, а на порядки.



# Клиент-серверная архитектура

---

- ▶ **сервер**, который обеспечивает хранение данных, выполнение над ними необходимых операций и формирование многомерной модели на концептуальном уровне;
- ▶ **клиент**, который представляет пользователю интерфейс к многомерной модели данных и обеспечивает возможность удобно этими данными манипулировать.

# Сервер OLAP

---

- ▶ Необходимо помнить, что сервер при реализации скрывает от пользователя способ реализации многомерной модели.
- ▶ Но тем не менее способ формирования гиперкуба из исходных данных может значительно влиять на производительность системы, необходимые для работы вычислительные ресурсы и т.д.
- ▶ Рассмотрим 3 основных подхода к реализации серверов OLAP:

# Виды архитектур серверов OLAP

---

- ▶ **MOLAP** — многомерный (multivariate) OLAP. Для реализации многомерной модели используют многомерные БД;
- ▶ **ROLAP** — реляционный (relational) OLAP. Для реализации многомерной модели используют реляционные БД;
- ▶ **HOLAP** — гибридный (hybrid) OLAP. Для реализации многомерной модели используют и многомерные, и реляционные БД.

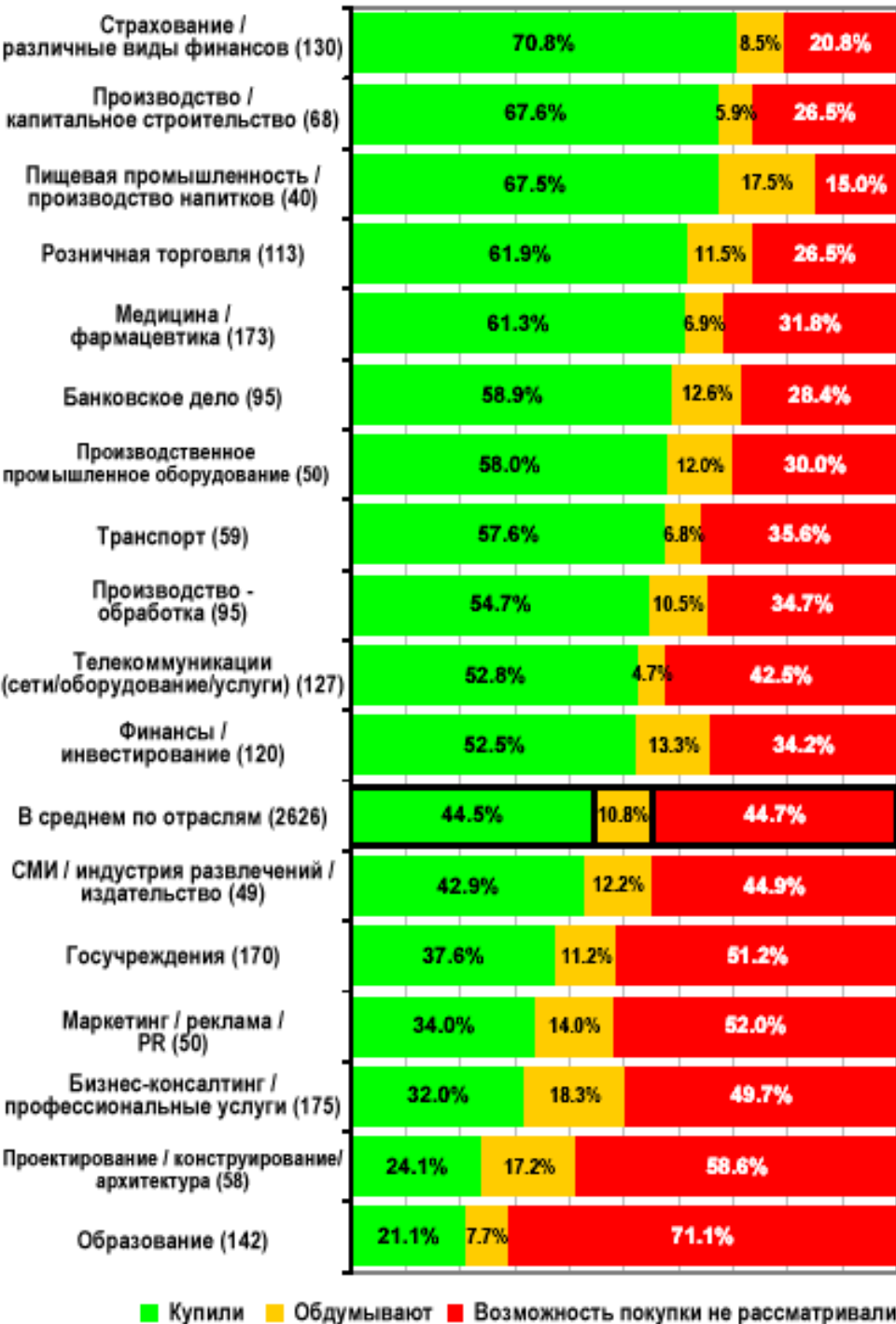
# Примеры клиентских OLAP-решений

---

В состав **Microsoft Office** входят три отдельных OLAP-компонента:

- ▶ Клиент извлечения и обработки данных для Excel позволяет создавать проекты извлечения и обработки данных на базе служб **SQL Server Analysis Services (SSAS)** и управлять ими из **Excel**;
- ▶ Средства анализа таблиц для приложения Excel позволяют использовать встроенные в службы SSAS функции извлечения и обработки информации для анализа данных, хранящихся в таблицах Excel;
- ▶ Шаблоны извлечения и обработки данных для приложения **Visio** позволяют визуализировать деревья решений, деревья регрессии, кластерные диаграммы и сети зависимостей на диаграммах Visio.

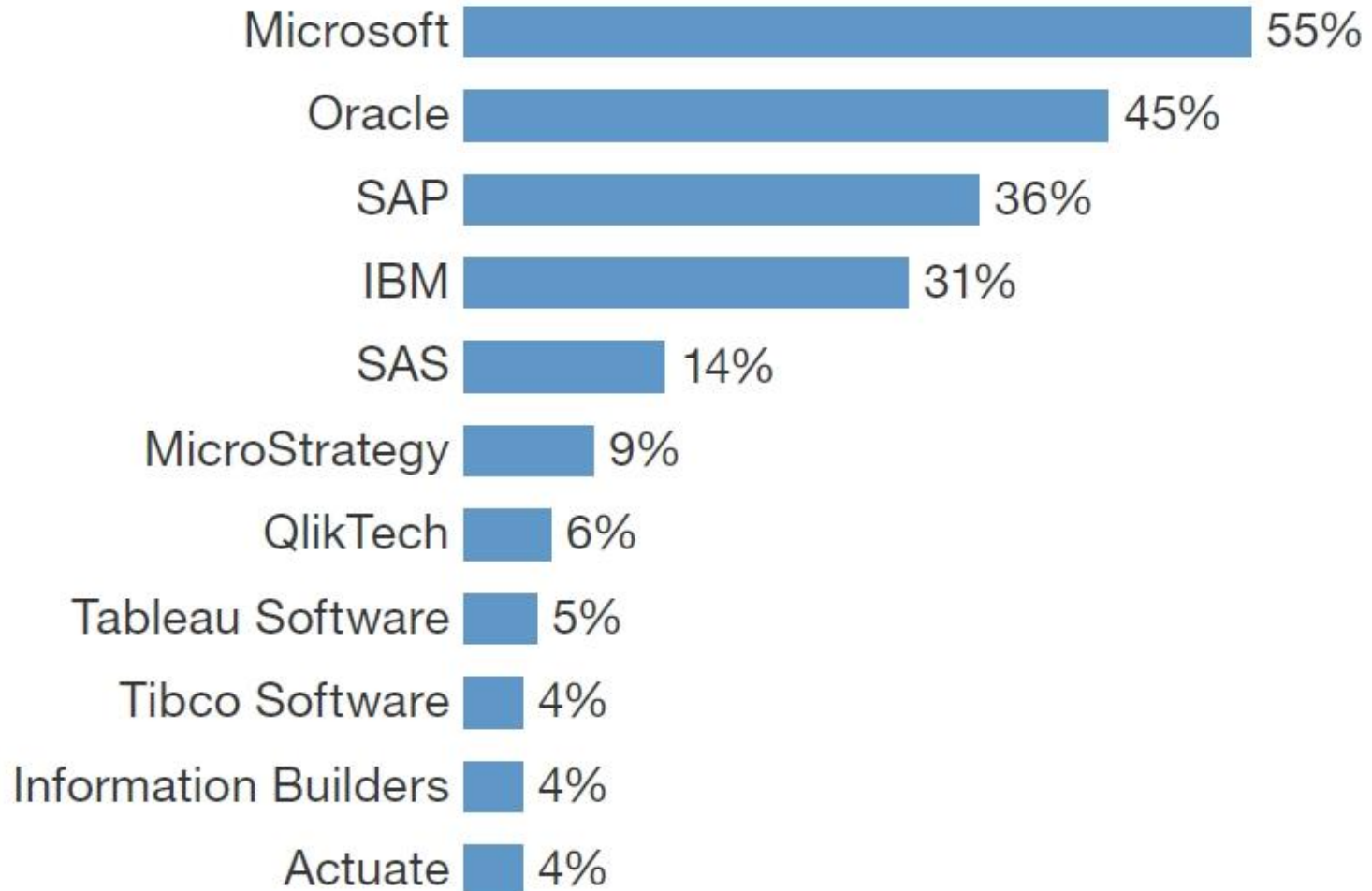
# Интенсивность использования OLAP по отраслям



1. Страхование
2. Производство
3. Пищевая промышленность
- .....
15. Консалтинг/проф.услуги
16. Архитектура/проектирование
17. Образование

# Доли производителей на рынке OLAP (2010-е)

---



# Российские производители инструментов класса OLAP

---

- ▶ Intersoft Lab «Контур»



- ▶ BaseGroup Labs «Deductor»

