

Аспекты аналитической подготовки принятия решений:

Сбор и хранение информации

- **Data Warehouse**

Анализ -
оперативный и
интеллектуальный

- **OLAP**
- **Data mining**

Подготовка
результатов
оперативного и
интеллектуального
анализа

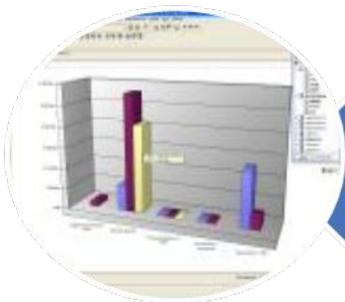
- развитые средства
визуального двумерного (2 D)
и трёхмерного (3 D)
представления информации

РЫНОК ИНСТРУМЕНТАЛЬНЫХ СРЕДСТВ

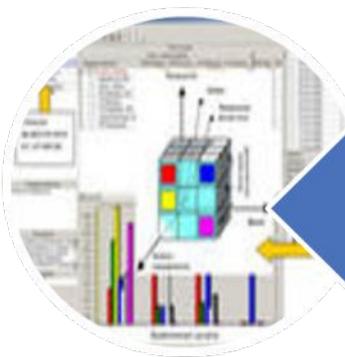
ИАС :



Средства массового применения к которым относятся Excel , Statistica .



Встроенные в интегрированные ЭИС инструментальные средства создания и поддержки ИХ и анализа.



Специализированные инструментальные средства создания ИХ и обеспечения различного вида анализа.

Инструменты для разработки отчетов

Microsoft

- Excel: таблицы и диаграммы (сводные)
 - VBA for Application
 - PowerQuery, PowerPivot, PowerView
 - Excel Services for SharePoint
- Microsoft Access
- Reporting Services
- PerformancePoint (SharePoint)
- Power BI: PowerQuery, PowerPivot, PowerView

Табло

QlikView

STATISTICA: Непараметрическая статистика

Файл Правка Вид Анализ Графика Сервис Окно ?

Описательные статистики (factor.sta)

NONPAR STATS	mean	valid N	median
WORK_1	97.02958	100	94.90816
WORK_2	98.16886	100	97.89211
WORK_3	98.94298	100	98.26256

Данные: FACTO...

This file contains random variables based on two factors

ЧИСЛ ЗНАЧ	1	2
	WORK_1	WORK_2
1	105.126	101.659
2	77.049	72.933
3	86.017	82.206
4	91.425	106.107
5	113.714	92.029
6	86.606	87.817
7	95.067	94.505
8	113.500	104.607

График2: Матричная диаграмма

Матричная диаграмма (FACTOR.STA 10n * 100n)

STATISTICA: Основные статистики и таблицы

Файл Правка Вид Анализ Графика Сервис Окно ?

Параметры Столбцы Строки

Данные: FACTOR.STA 10n * 100n

This file contains random variables based on two factors

ЧИСЛ ЗНАЧ	1	2
	WORK_1	WORK_2
1	105.126	101.659
2	77.049	72.933
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		

График5: Матричная диаграмма

Матричная диаграмма (FACTOR.STA 10n * 100n)

Корреляции (factor.sta)

ОСНОВНЫЕ СТАТИСТ. Отмеченные корреляции значимы N=100 (Построчное удаление ПД)

Перемен.	КОРРЕЛЯЦИИ			
	WORK_1	WORK_2	MISCEL_1	MISCEL_2
WORK_1	1.00	.65	.61	.55
WORK_2	.65	1.00	.71	.68
WORK_3	.65	.73	.70	.67
HOME_2	.15	.18	.64	.59
HOME_3	.14	.24	.59	.52
MISCEL_1	.61	.71	1.00	.84

Кнопки автозаг: dotplot, caseplot, Настр

Готово Вывод:ОКНО

SPSS - статистика и анализ данных

tv-survey.sav - SPSS Data Editor

File Edit View Data Transform Analyze

9:

	any	bored	critic
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15	YES	YES	
16	YES	YES	
17	YES	YES	
18	YES	YES	
19	YES	YES	
20	NO	NO	
21	NO	YES	
22	NO	NO	
23	NO	NO	
24	NO	YES	

Frequencies

- Any reason [any]
- There are no other po...
- Other people still watc...
- The original directors s...
- The original cast stays...

Display frequency tables

Data View Variable View

Graphboard Template Chooser

Basic Detailed Titles Options

Natural Name Type

life
sibs
childs
age
educ
paeduc
maeduc
speduc
prestg80
occcat80
tax
usintl
obey
popular
thnkself

Visualization of:
life
sibs

Summary: Sum

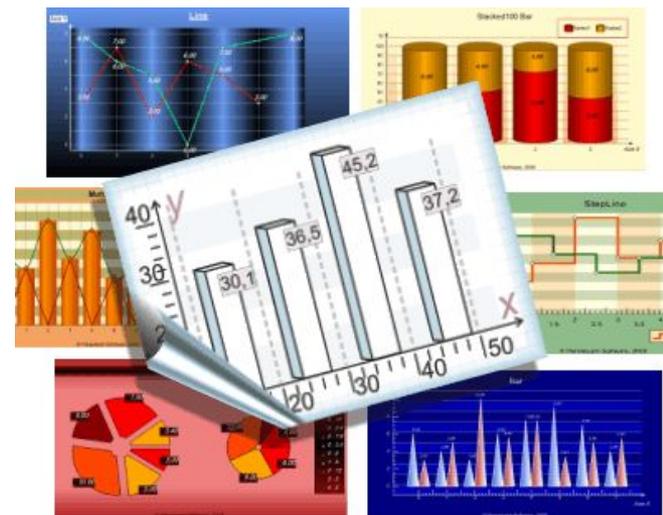
Manage... Location ... Local Machine

OK Paste Reset Cancel Help

The Graphboard Template Chooser window displays a grid of chart templates. The 'Basic' tab is selected. The 'Natural' radio button is chosen. A list of variables is shown on the left, with 'life' and 'sibs' selected for visualization. The 'Summary' dropdown is set to 'Sum'. The chart grid includes: Bar (a bar chart with three bars), Pie (a 2D pie chart), 3-D Pie (a 3D pie chart), Line (a line graph with multiple lines), Area (an area chart with a shaded region), Path (a path plot with a blue line and points), Ribbon (a ribbon plot with a blue ribbon), Scatterplot (a scatter plot with black dots), 2-D Dot Plot (a dot plot with three categories), and a box plot (a box plot with a blue box and whiskers).

OLAP – технологии

- ❖ **OLAP** (On - Line Analytical Processing оперативный анализ данных (информации))
- ❖ **Основной задачей** оперативного или OLAP -анализа является быстрое (в пределах секунд) извлечение необходимой аналитики или ЛПР для обоснования или принятия решения информации.
- ❖ **Основная функция OLAP** - управление измерениями, которые применяются для моделирования основных характеристик бизнеса.



Требования, предъявляемые к OLAP-системам

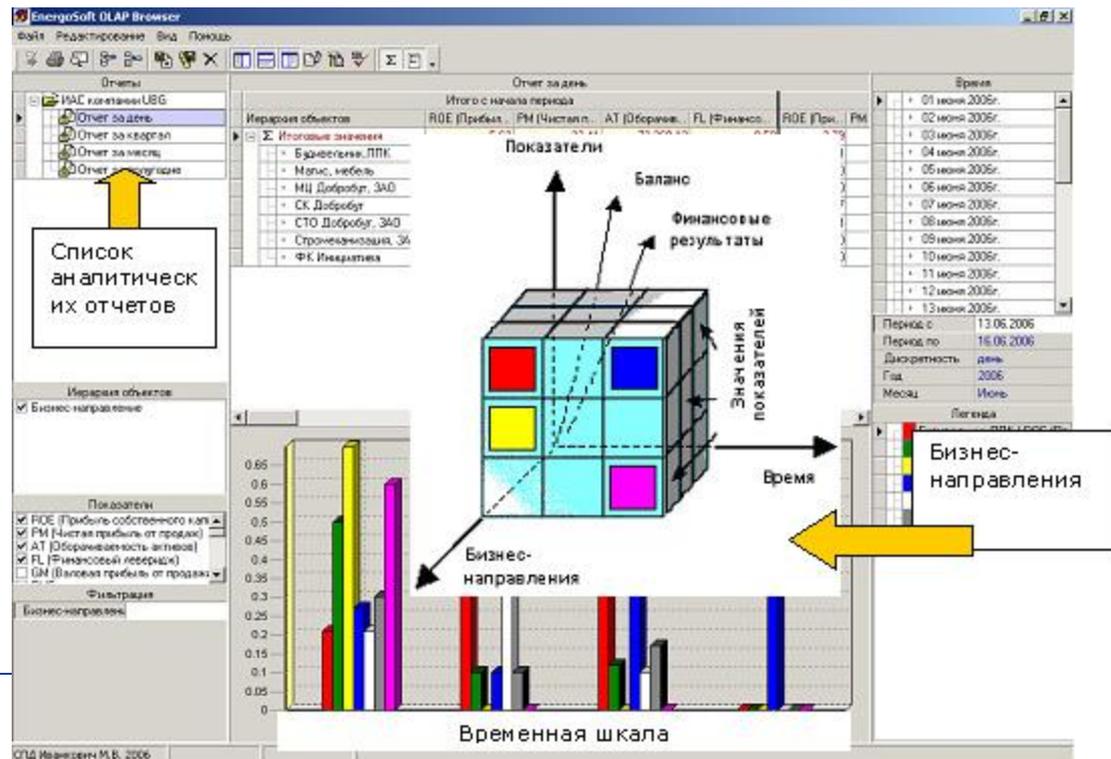
- ❖ **1. Многомерное представление данных.**
- ❖ **2. Прозрачность.**
- ❖ **3. Доступность.**
- ❖ **4. Согласованная производительность.**
- ❖ **5. Поддержка архитектуры «клиент-сервер».**
- ❖ **6. Равноправность всех измерений.**
- ❖ **7. Динамическая обработка разреженных матриц.**
- ❖ **8. Поддержка многопользовательского режима работы с данными.**
- ❖ **9. Поддержка операций на основе различных измерений.**
- ❖ **10. Простота манипулирования данными.**
- ❖ **11. Развитые средства представления данных.**
- ❖ **12. Неограниченное число измерений и уровней агрегации данных.**

Признаки OLAP – систем

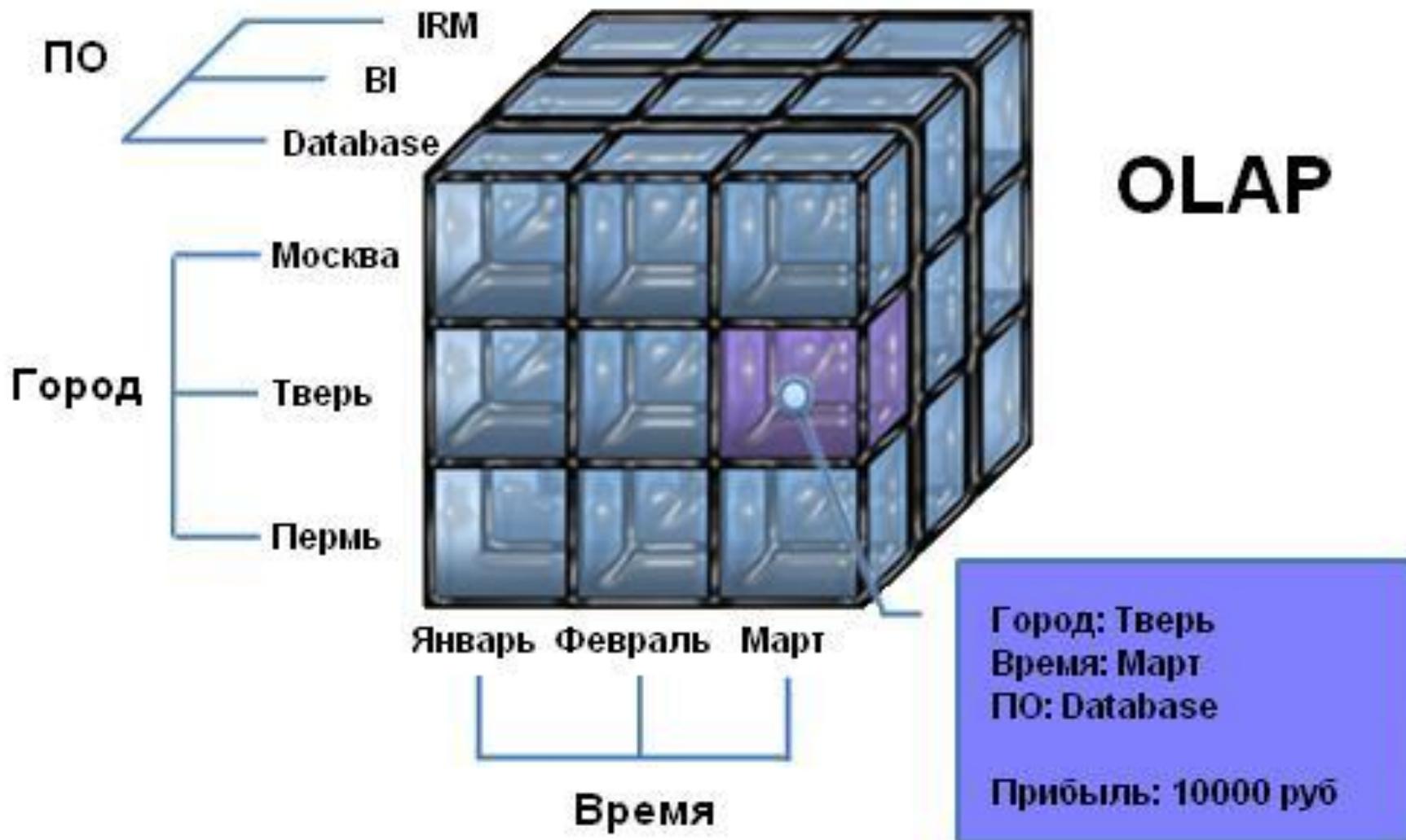
В конце 90-х годов получил распространение свод требований к информационно-аналитическим системам в виде «теста FASMI» — аббревиатуры английских слов, определяющих требования к OLAP-системам:

Fast **A**nalysis **S**hared **M**ultidimensional **I**nformation — русский перевод:

Быстрый
Анализ
Разделяемой
Многомерной
Информации



Представление исследуемого объекта в виде многомерного куба



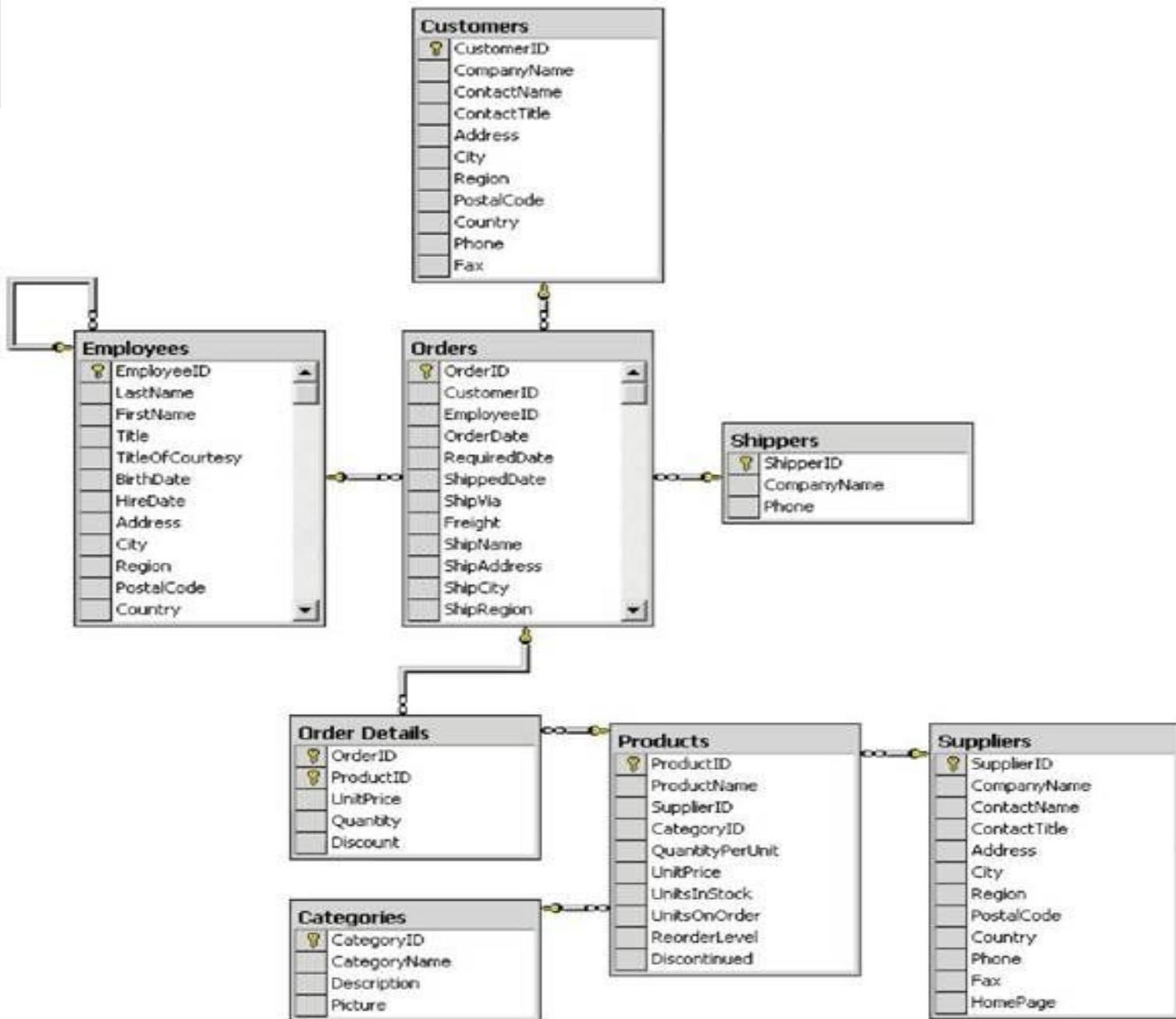


Рис.5.2. Схема базы данных Northwind

Формат: Анализ: период: структура: от: по: группа: контроль: единицы измерения: валюта:

3 квартал 19 2001

Статус: Товар - Контроль: ед. - Материал: ед. -

3 квартал

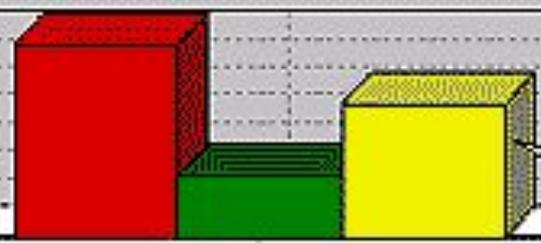
Итого	Июль		Август		Сентябрь		Итого		Итого	
	Сумма	Количество	Сумма	Количество	Сумма	Количество	Сумма	Количество	Сумма	Количество
Битовая техника			31 350.95	14.00	17 355.01	12.00	48 705.96	26.00	48 705.96	
Кухонные электроприборы			4 470.00		39 150.00	15.00	43 620.00	15.00	43 620.00	
Телевизоры					80 122.97	5.00	80 122.97			
Железные изделия, мебель			175 082.74	5.00			175 082.74			
Итого			211 913.69	22.00	136 627.98	32.00	348 541.67	57.00	348 541.67	
Инвентарь			7 800.00	5.00			7 800.00	5.00	7 800.00	
Грабли			483.45	10.00			483.45	10.00	483.45	
Матос			240.00	20.00			240.00	20.00	240.00	
Матоблок			102 000.00	2.00			102 000.00	2.00	102 000.00	
Итого			110 523.45	37.00			110 523.45	37.00	110 523.45	
Обуви			75 054.15	37.00	95 610.00	44.00	170 664.15	81.00	170 664.15	
Женская										
Итого			75 054.15	37.00	95 610.00	44.00	170 664.15	81.00	170 664.15	
Итого			75 054.15	37.00	418 047.14	104.00	205 723.35	96.00	699 824.64	197.00

Динамическая таблица

Итого

Аналитические разрезы - измерения

150000
100000
50000
0



■ Битовая техника
■ Инвентарь
■ Обуви

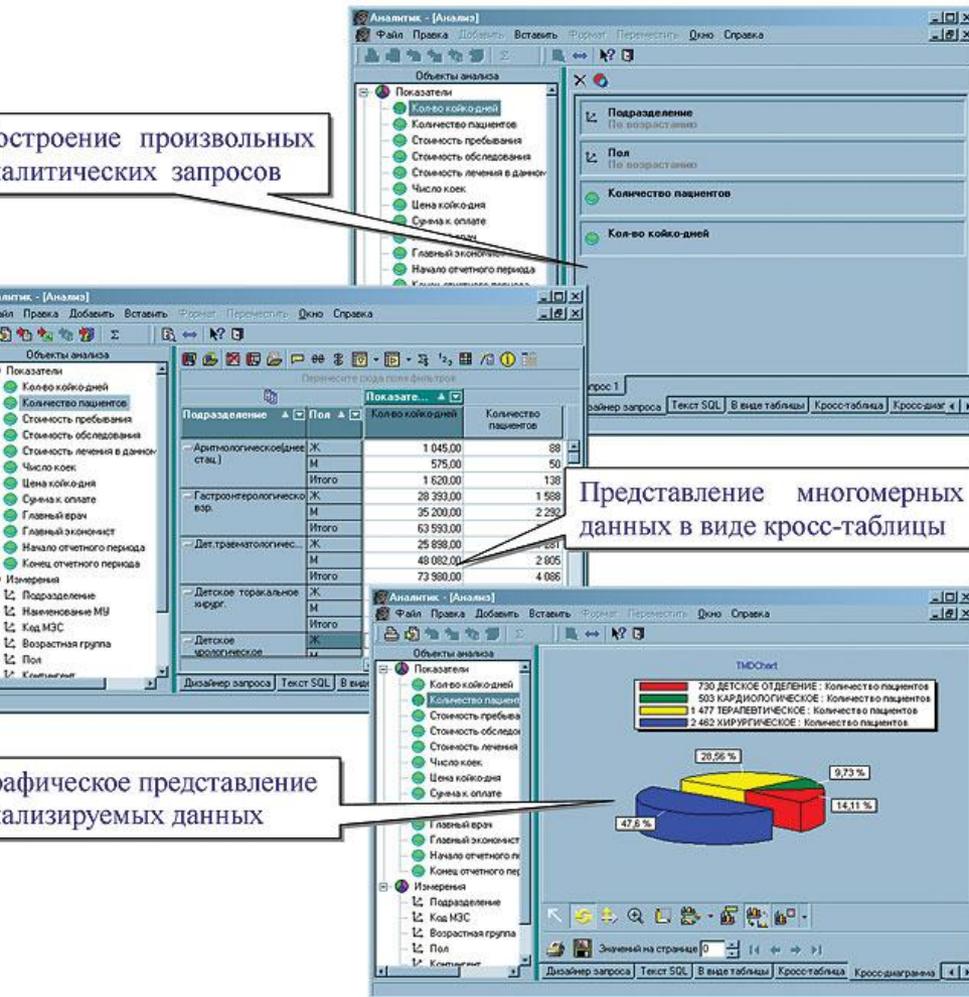
Синхронная графическая иллюстрация

Пример «Бюджет продаж»

					Дата 2005							
					январь 2005	февраль 2005	март 2005	апрель 2005	май 2005	июнь 2005		
Строки	Справочник1	Справочник2	Справочник3		План	План	План	План	План	План		
Объем продаж	Продуктовая группа 1	Продукт 1.1	Филиал 1	Розница	1 010,00	1 020,00	1 030,00	1 040,00	1 050,00	1 060,00		
				Опт	Городской опт	1 250,00	1 260,00	1 270,00	1 280,00	1 290,00	1 300,00	
					Областной опт	1 500,00	1 510,00	1 520,00	1 530,00	1 540,00	1 550,00	
		+ Филиал 2	2 850,00	2 880,00	2 910,00	2 940,00	2 970,00	3 000,00				
		+ Филиал 3	2 750,00	2 780,00	2 810,00	2 840,00	2 870,00	2 900,00				
		+ Филиал 4	2 450,00	2 480,00	2 510,00	2 540,00	2 570,00	2 600,00				
		+ Филиал 5	2 350,00	2 380,00	2 410,00	2 440,00	2 470,00	2 500,00				
		+ Продукт 1.2	14 870,00	15 030,00	15 190,00	15 350,00	15 510,00	15 670,00				
		+ Продукт 1.3	18 360,00	18 510,00	18 660,00	18 810,00	18 960,00	19 110,00				
		Продуктовая группа 2	+ Продукт 2.1	27 650,00	27 800,00	27 950,00	28 100,00	28 250,00	28 400,00			
	+ Продукт 2.2		29 210,00	29 375,00	29 540,00	29 705,00	29 870,00	30 035,00				
	+ Продукт 2.3		29 410,00	29 570,00	29 730,00	29 890,00	30 050,00	30 210,00				
	Продуктовая группа 3	+ Продукт 3.1	39 400,00	39 550,00	39 700,00	39 850,00	40 000,00	40 150,00				
		+ Продукт 3.2	39 150,00	39 300,00	39 450,00	39 600,00	39 750,00	39 900,00				
		+ Продукт 3.3	41 150,00	41 310,00	41 470,00	41 630,00	41 790,00	41 950,00				
	Выручка от реализации	Продуктовая группа 1	Продукт 1.1	Филиал 1	Розница	101 000,00	102 000,00	103 000,00	104 000,00	105 000,00	106 000,00	
					Опт	Городской опт	87 500,00	88 200,00	88 900,00	89 600,00	90 300,00	91 000,00
						Областной опт	120 000,00	120 800,00	121 600,00	122 400,00	123 200,00	124 000,00
+ Филиал 2			252 000,00	254 500,00	257 000,00	259 500,00	262 000,00	264 500,00				
+ Филиал 3			227 000,00	229 500,00	232 000,00	234 500,00	237 000,00	239 500,00				
+ Филиал 4			194 000,00	196 500,00	199 000,00	201 500,00	204 000,00	206 500,00				
+ Филиал 5			185 000,00	187 500,00	190 000,00	192 500,00	195 000,00	197 500,00				
+ Продукт 1.2			1 393 400,00	1 409 100,00	1 424 800,00	1 440 500,00	1 456 200,00	1 471 900,00				
+ Продукт 1.3			1 922 800,00	1 938 800,00	1 954 800,00	1 970 800,00	1 986 800,00	2 002 800,00				
Продуктовая группа 2			+ Продукт 2.1	1 092 500,00	1 099 500,00	1 106 500,00	1 113 500,00	1 120 500,00	1 127 500,00			
		+ Продукт 2.2	840 600,00	847 000,00	853 400,00	859 800,00	866 200,00	872 600,00				
		+ Продукт 2.3	1 190 250,00	1 198 500,00	1 206 750,00	1 215 000,00	1 223 250,00	1 231 500,00				
Продуктовая группа 3		+ Продукт 3.1	611 500,00	615 250,00	619 000,00	622 750,00	626 500,00	630 250,00				
		+ Продукт 3.2	589 750,00	593 500,00	597 250,00	601 000,00	604 750,00	608 500,00				
		+ Продукт 3.3	383 500,00	386 650,00	389 800,00	392 950,00	396 100,00	399 250,00				

Типы многомерных OLAP-систем:

строение произвольных аналитических запросов



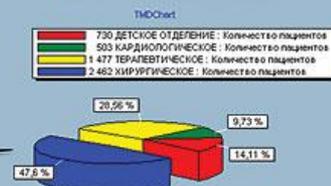
Представление многомерных данных в виде кросс-таблицы

графическое представление анализируемых данных

многомерный (Multidimensional) OLAP — **MOLAP**

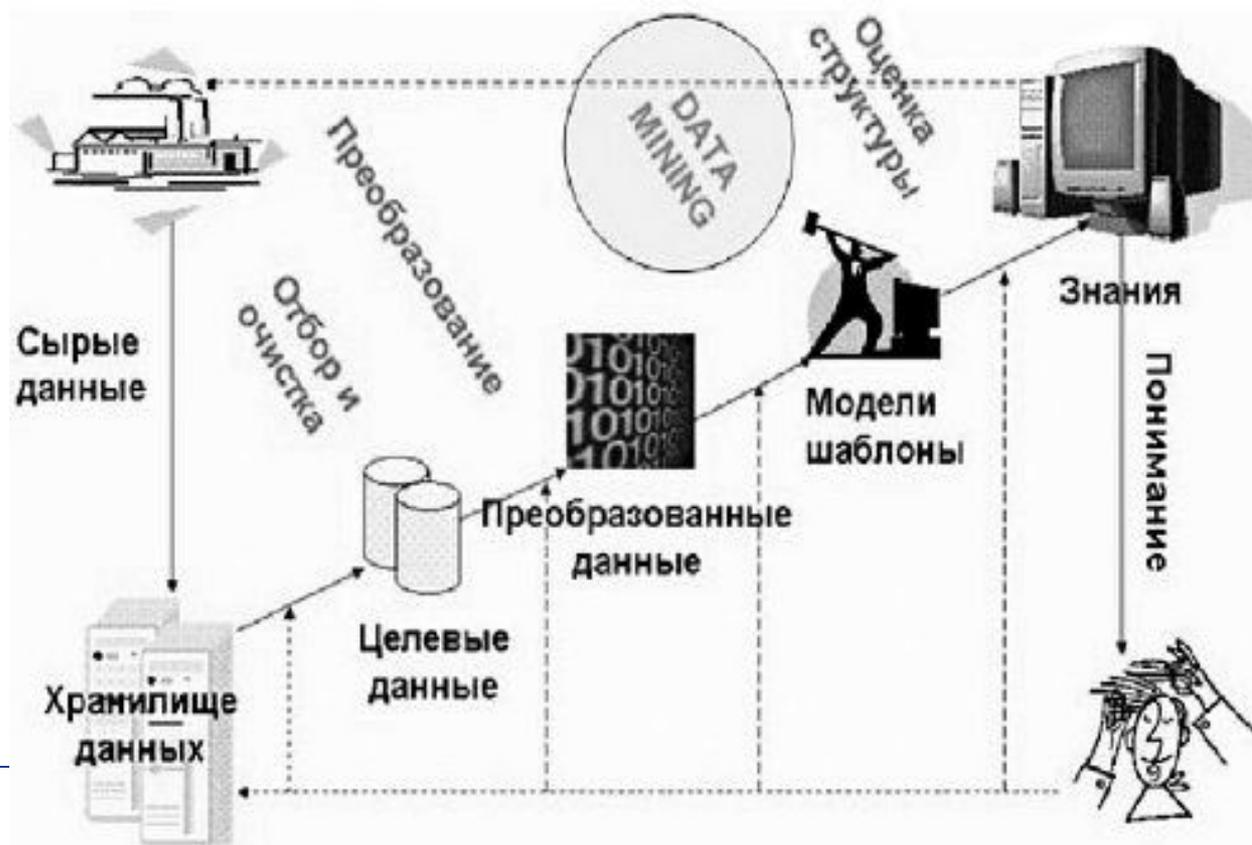
реляционный (Relation) OLAP — **ROLAP**

смешанный или гибридный (Hibrid) OLAP — **HOLAP**



Технологии Data Mining

- ◆ **Data Mining** - это процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.



OLAP

Каковы средние показатели рентабельности предприятий в регионе?

Каковы средние размеры счетов клиентов банка — физических лиц?

Какова средняя величина ежедневных покупок по украденной или фальшивой кредитной карточке?

Data Mining

Какова характерная совокупность значений показателей финансово-хозяйственной деятельности предприятий в регионе?

Каков типичный портрет клиента — физического лица, отказывающегося от услуг банка?

Существуют ли стереотипные схемы покупок для случаев мошенничества с кредитными карточками?

Основные задачи интеллектуального анализа данных:

- выявление взаимозависимостей, причинно-следственных связей, ассоциаций;
- классификация событий и ситуаций, определение профилей различных факторов
- прогнозирование хода процессов, событий

Схема преобразования данных с использованием технологии Data Mining



Области применения технологии Data Mining



Технологии Data Mining

- ❖ Можно назвать пять стандартных типов закономерностей, выявляемых с помощью методов Data Mining:

Ассоциация

Последовательность

Классификация

Кластеризация

Прогнозирование

Некоторые бизнес-приложения Data Mining

Розничная торговля

- анализ покупательской корзины
- исследование временных шаблонов
- создание прогнозирующих моделей

Банковское дело

- выявление мошенничества с кредитными карточками
- сегментация клиентов
- прогнозирование изменений клиентуры.

Телекоммуникации

- анализ записей о подробных характеристиках вызовов.
- выявление лояльности клиентов

Страхование

- выявление мошенничества
- анализ риска

Главные тенденции российского рынка ВІ

- ◆ Увеличение доли отечественных решений
- ◆ Развитие Open Source инструментов ВІ
- ◆ Интерес среднего бизнеса к «легким» ВІ-системам
- ◆ Спрос среди госструктур
- ◆ ВІ в облаке
- ◆ Новые игроки
- ◆ Другие тренды



Примеры визуализации данных от Visiology



ИЗБРАННОЕ



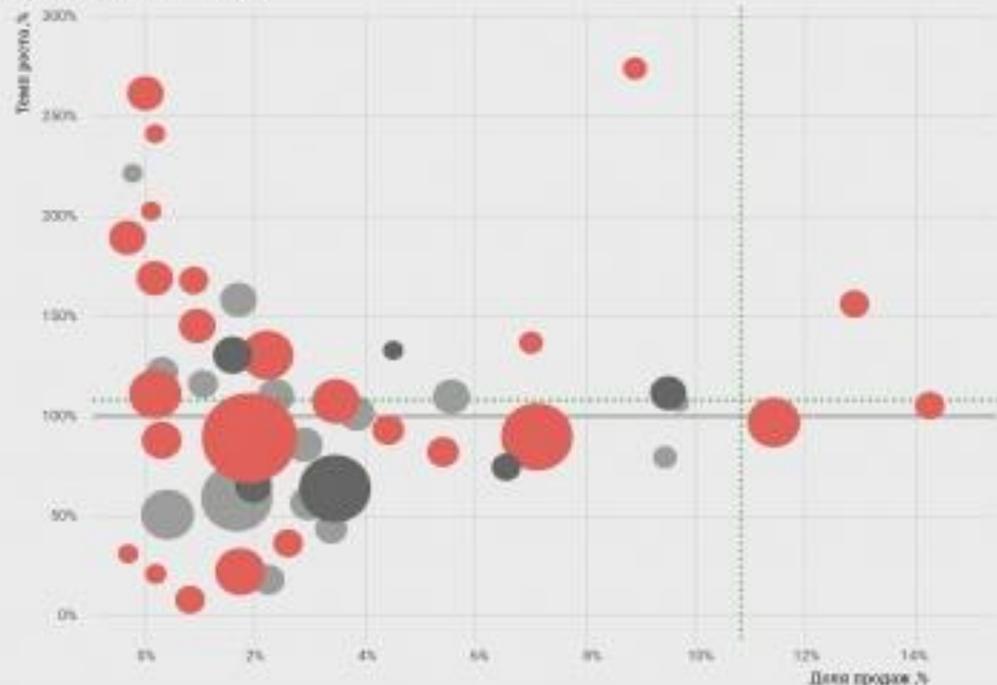
Сравнительный анализ товаров по категориям

Территория: Москва | Исполнитель: Все | Категория: Овощи, фрукты | Контакт: Петрова А.

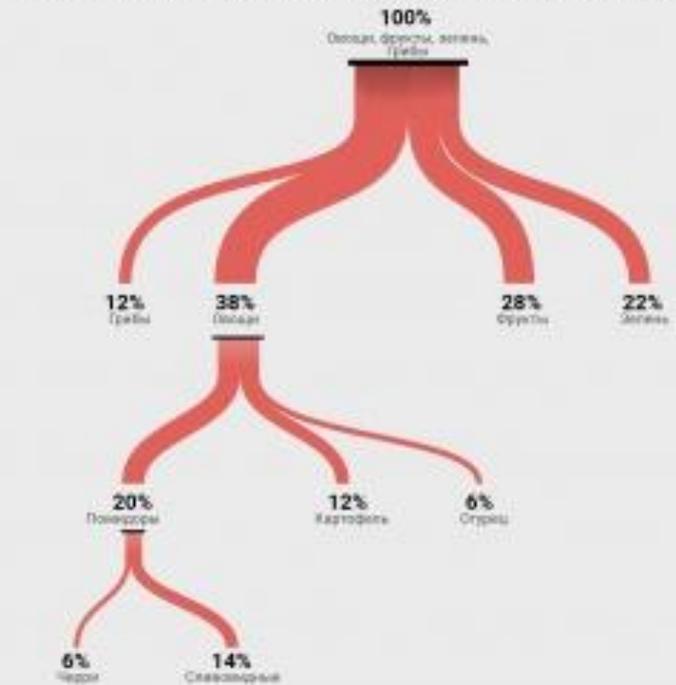
Год | Квартал | Месяц | Неделя | День

Показатель: Объем продаж

Бостонская матрица



Распределение выбранного показателя по категориям



Примеры визуализации данных от Visiology



ИЗБРАННОЕ



Анализ сезонности и распределение выбранного показателя по времени

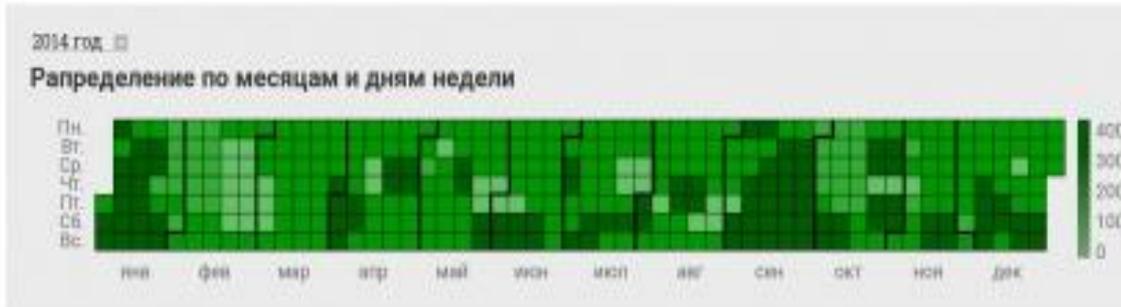
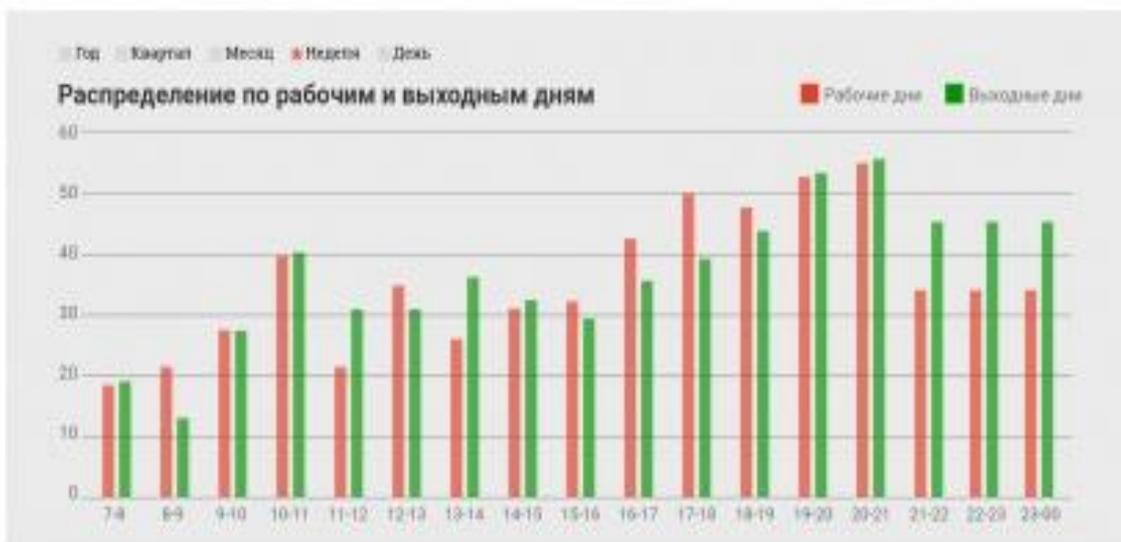
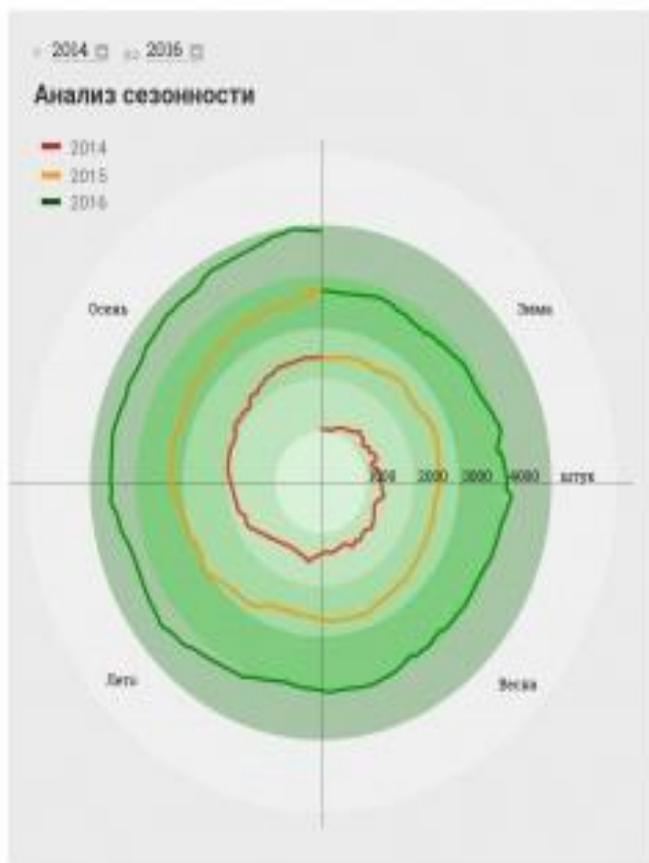
Регион: Москва

Магазин: Верный #63

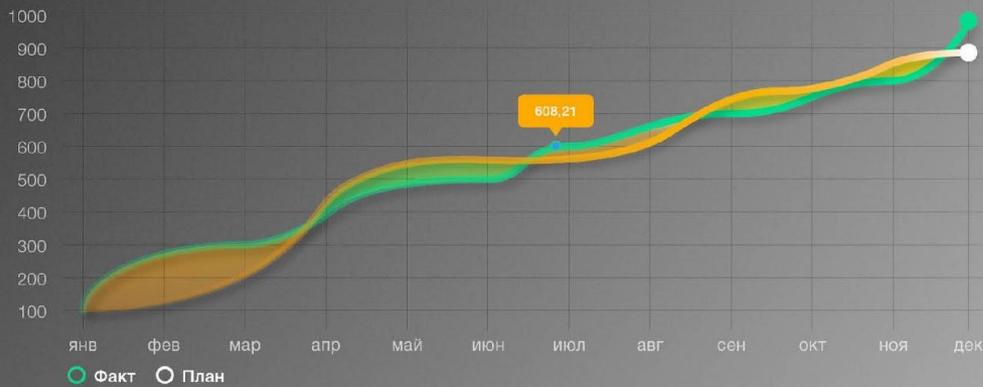
Категория: Кодд. изд.

Клиент: Петрова А.

Показатель: Объем продаж



Выручка, млрд. руб.



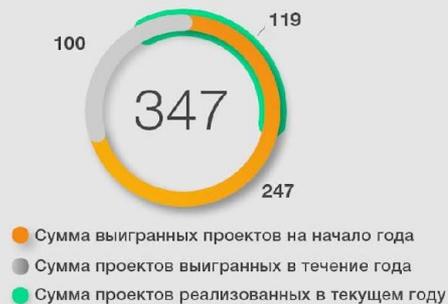
Валовая прибыль, млрд. руб.



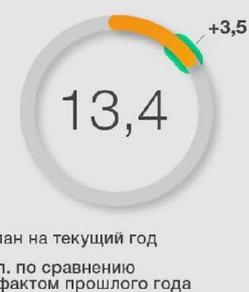
План-факт по реализации проектов, млрд. руб.



Портфель проектов, млрд. руб.



Доля рынка, %



Алгоритмы обработки данных

