



Решения Oracle для анализа Больших Данных

Ольга Горчинская
Oracle EE&CIS

План

- Платформа Oracle для Больших Данных
- Аналитическая in-memory машина Oracle Exalytics
- От бизнес-анализа к исследованию данных – Oracle Endeca Information Discovery
- Встроенная аналитика для статистических исследований – Oracle R Enterprise

План

- **Платформа Oracle для Больших Данных**
- Аналитическая in-memory машина Oracle Exalytics
- От бизнес-анализа к исследованию данных – Oracle Endeca Information Discovery
- Встроенная аналитика для статистических исследований – Oracle R Enterprise

О Больших Данных

- **Сверхбольшие объемы структурированных и неструктурированных данных, с которыми трудно работать с помощью традиционных средств**
- Впервые термин появился 3 сентября 2008 года, Клиффорд Линч, редактор научного журнала Nature, в связи с проблемой накопления научных данных
- Источники больших данных— интернет-документы измерительные устройства, радиочастотная идентификация, социальные сети, устройства ауди и видеорегистрации
- В 2011 году Gartner отмечает Большие Данные как тренд номер два в информационно-технологической инфраструктуре (после виртуализации)



VVV
Volume, Velocity, Veracity
(объем, скорость,
многообразие)

Технологии Больших данных

- Массово-параллельная обработка
- NO-SQL базы данных (Not Only SQL)
- Распределенное выполнение приложений (Google Map Reduce, Hadoop)...
- In-memory архитектуры
- Новые подходы к методам и инструментам анализа – «от бизнес-анализа к исследованию данных»



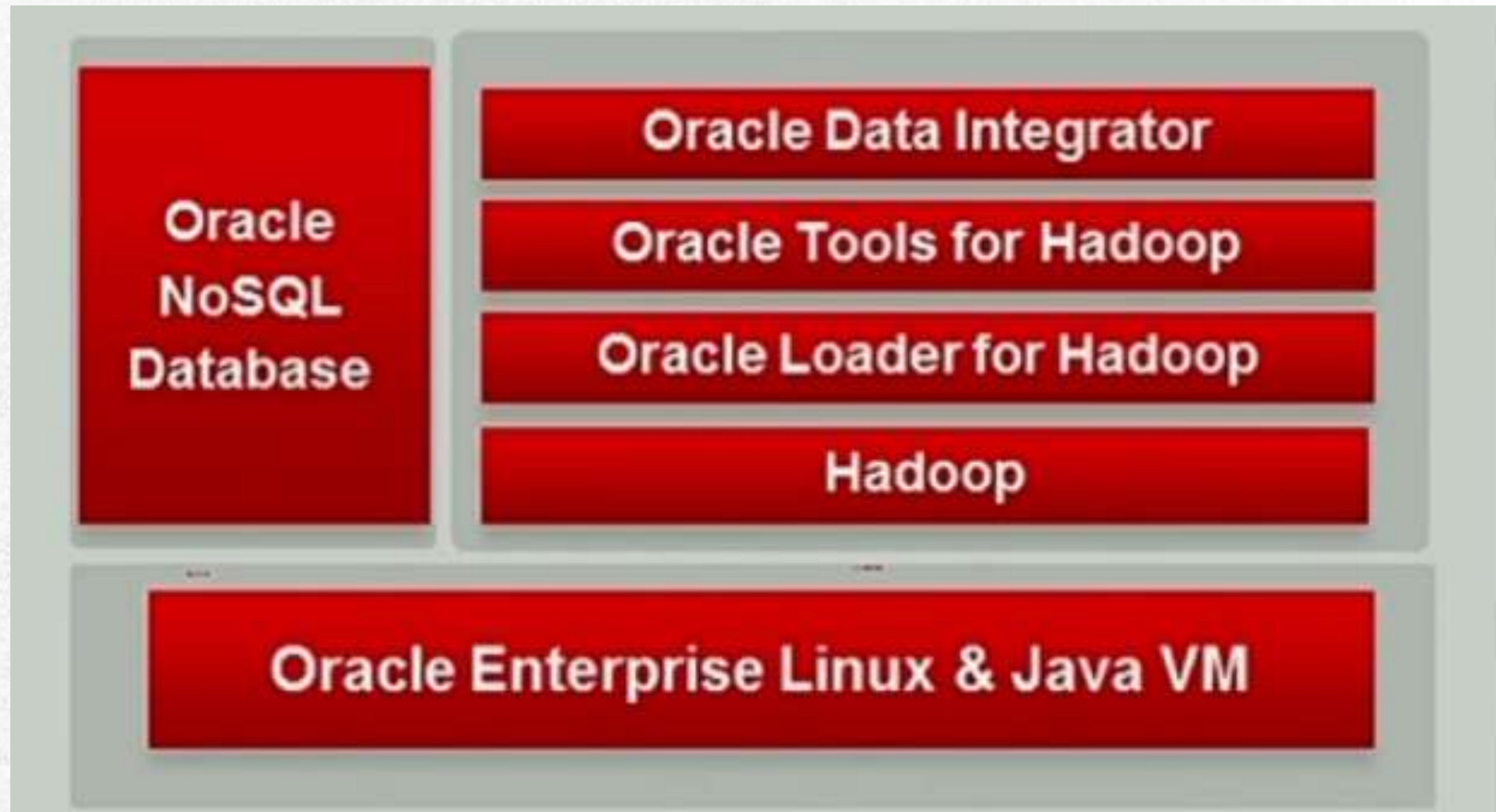
VVV
Velocity, Volume, Veracity
(скорость, объем,
многообразие)

Oracle для Больших Данных

- Программно-аппаратный комплекс Oracle Big Data Appliance
- Программно-аппаратные комплексы для поддержки экстремальной производительности и сверхбольших объемов данных
 - машина баз данных,
 - аналитическая in-memory машина
- Платформа класса Исследования Данных (Data Discovery) -- Endeca Information Discovery
- Среда статистических исследований и визуализации Oracle R Enterprise

Oracle Big Data Appliance

Программно-аппаратный комплекс для обработки и анализа «больших данных»

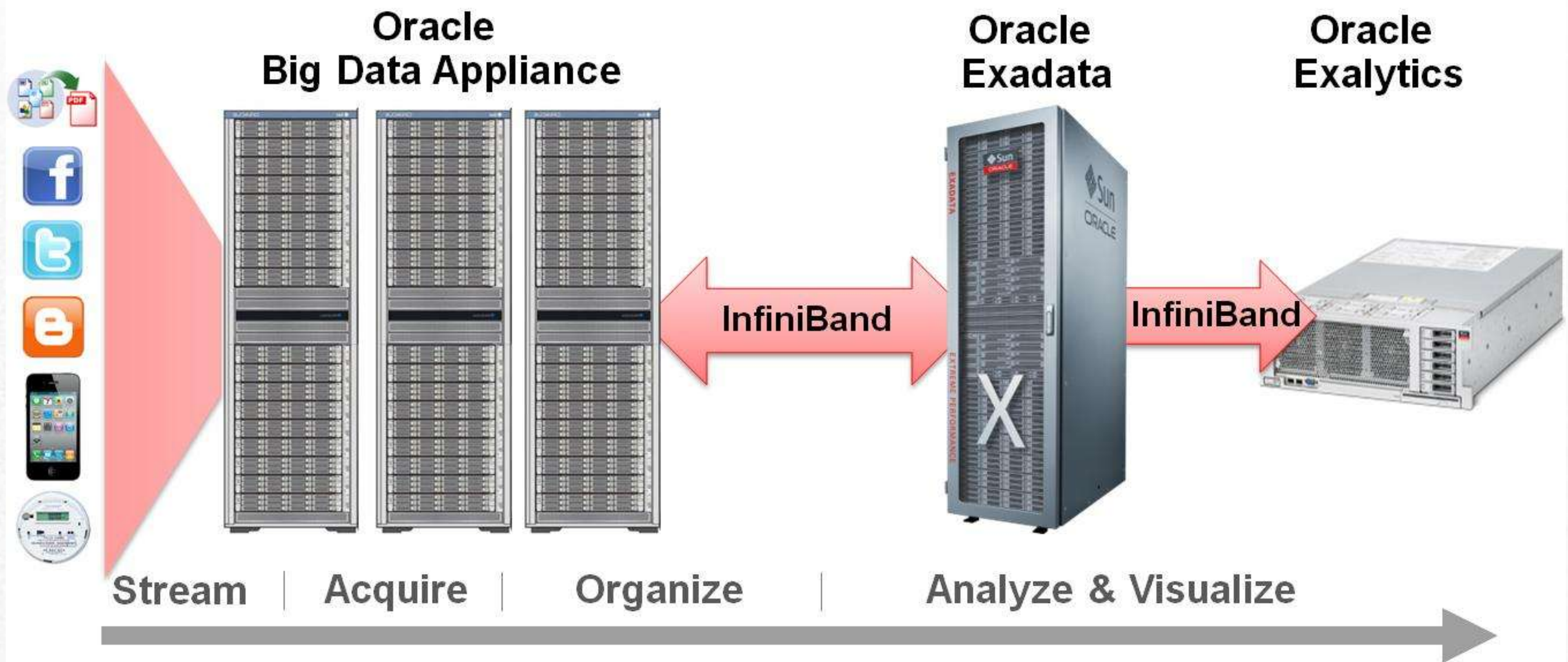


Аппаратное обеспечение

- Стандартный Sun Rack II 1242 Base
- 18 узлов – Sun X4270 M2 Servers
- Каждый сервер содержит
 - 48 Gb RAM
 - 2 Intel Xeon X5675 процессора (3.06 GHz) по 6 ядер на каждый (всего 216 ядер)
 - 12 Disks (432 TB raw disk total)
- 2 NM2-GW Sun Network QDR InfiniBand Gateway Switches with 10GigE connectivity



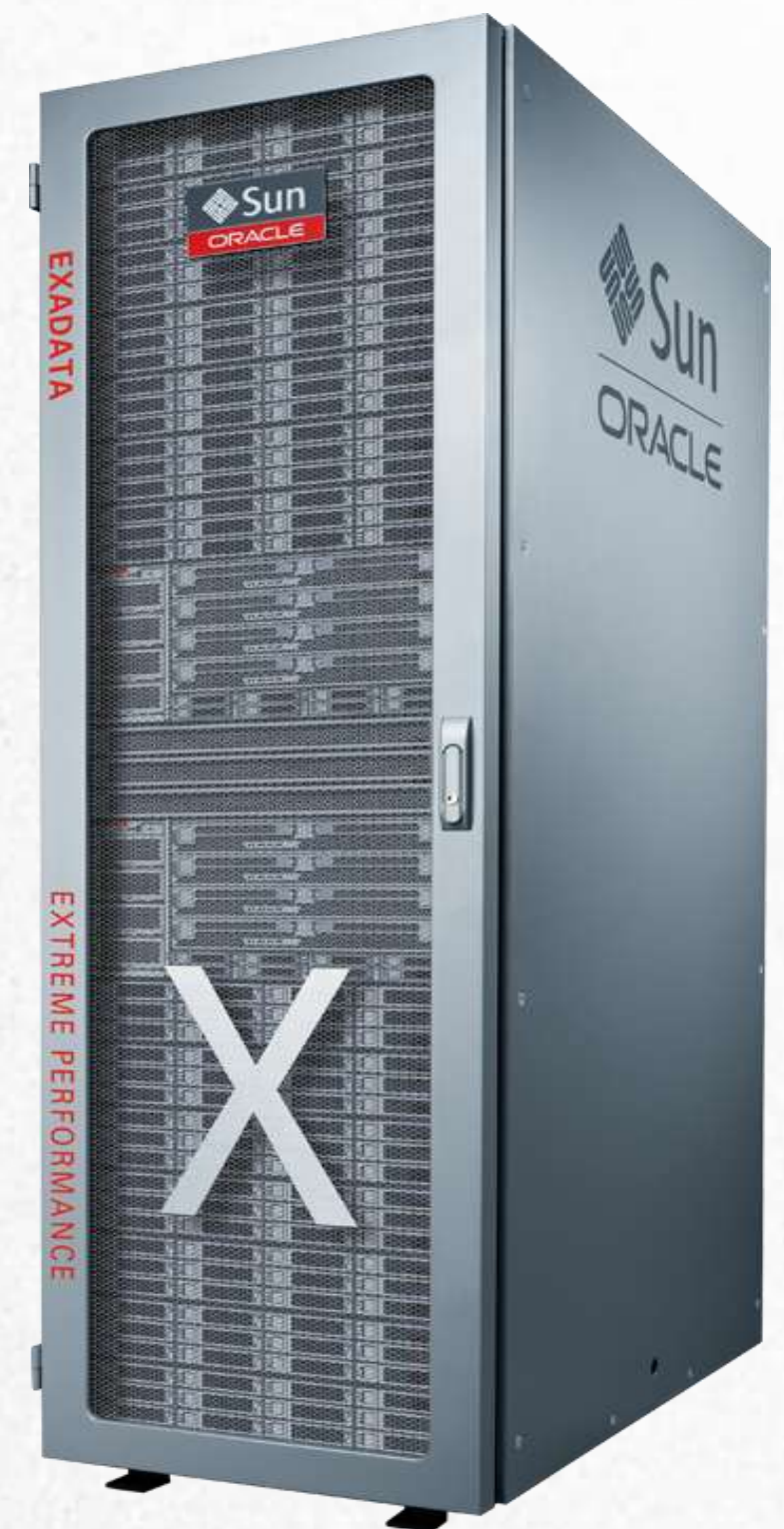
Общая архитектура Oracle для анализа Больших Данных



Exadata Database Machine

Машина баз данных

- Специализированный аппаратно-программный комплекс для хранилищ данных
 - высокопроизводительное оборудование + Oracle Database 11g
- Повышение скорости выполнения сложных аналитических запросов в десятки раз
 - перенос обработки запроса на «умные» ячейки хранения
- Сокращение объемов хранения данных
 - гибридное колоночное сжатие данных



План

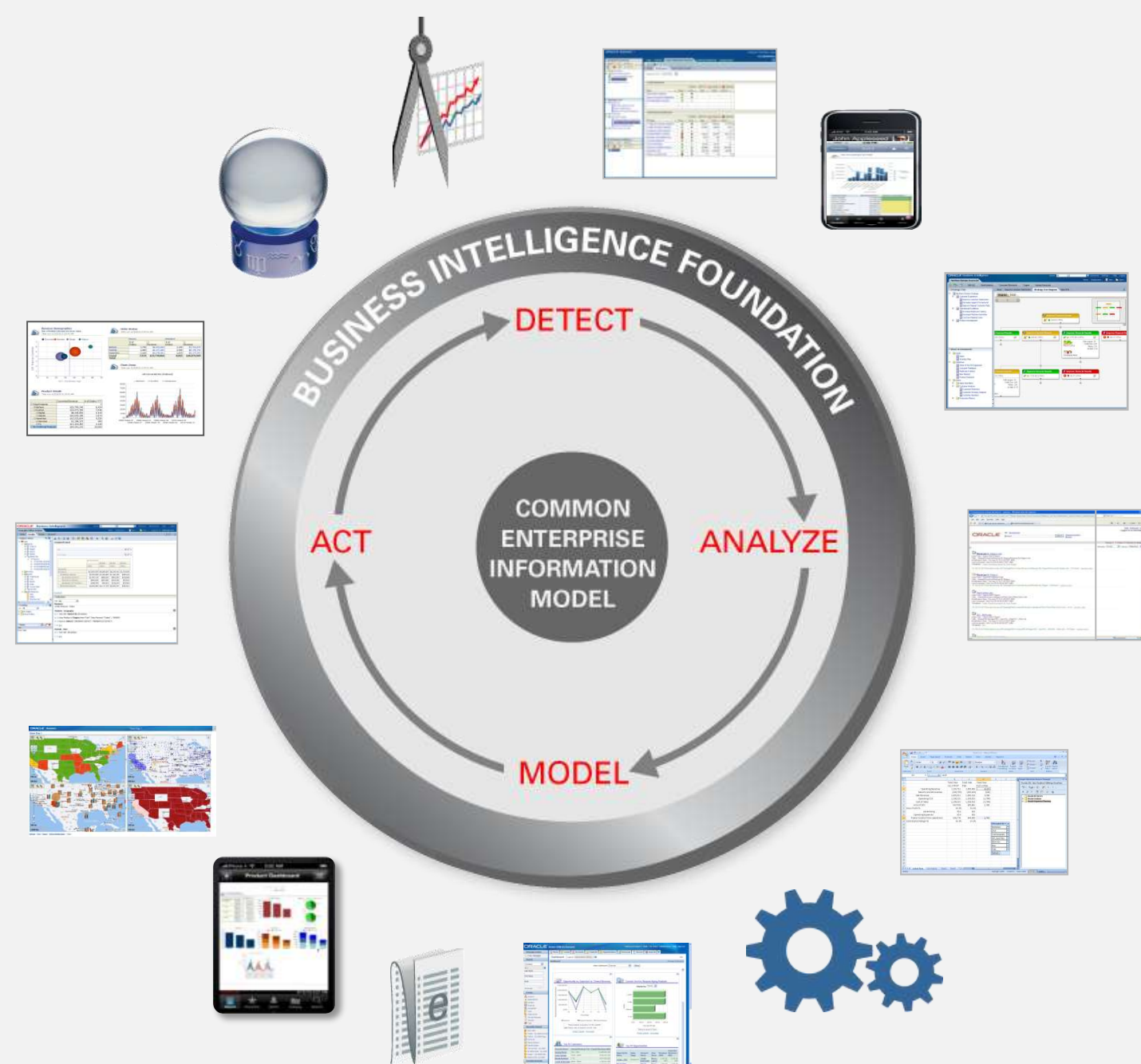
- Платформа Oracle для Больших Данных
- Аналитическая in-memory машина Oracle Exalytics
- От бизнес-анализа к исследованию данных – Oracle Endeca Information Discovery
- Встроенная аналитика для статистических исследований – Oracle R Enterprise

Аналитическая машина Oracle Exalytics

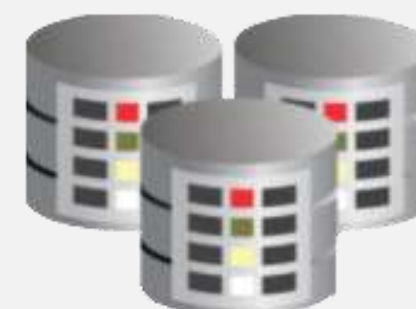
- Программно-аппаратный комплекс для бизнес-анализа
- «Экстремальная производительность и неограниченная визуализация»
- Реализация концепции «аналитика в оперативной памяти» (in-memory analytics)



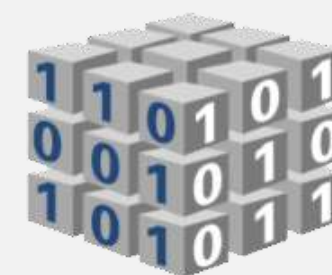
Oracle Exalytics – основные компоненты



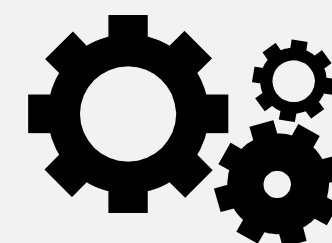
Oracle Business Intelligence Suite – специальная редакция для Exalytics



TimesTen for Exalytics



Memory Optimized Essbase



Adaptive In-Memory Tools

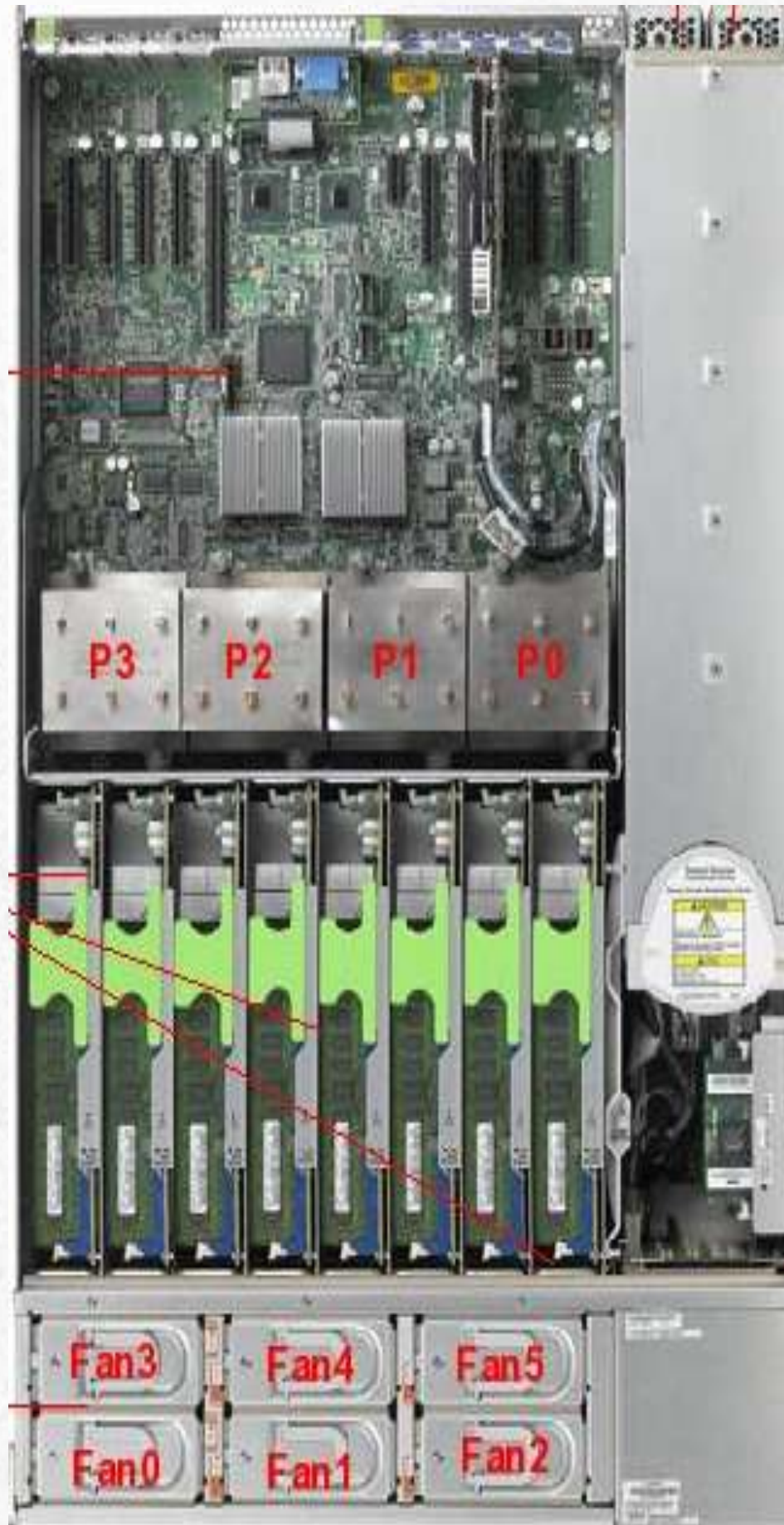
Программное обеспечение для In-Memory аналитики



1 TB RAM
40 Processing Cores
High Speed Networking

Аппаратный комплекс для In-Memory аналитики

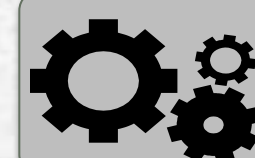
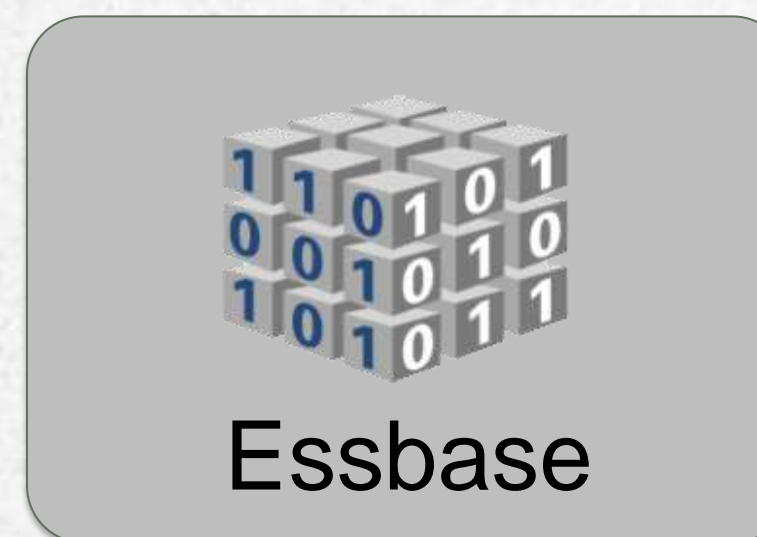
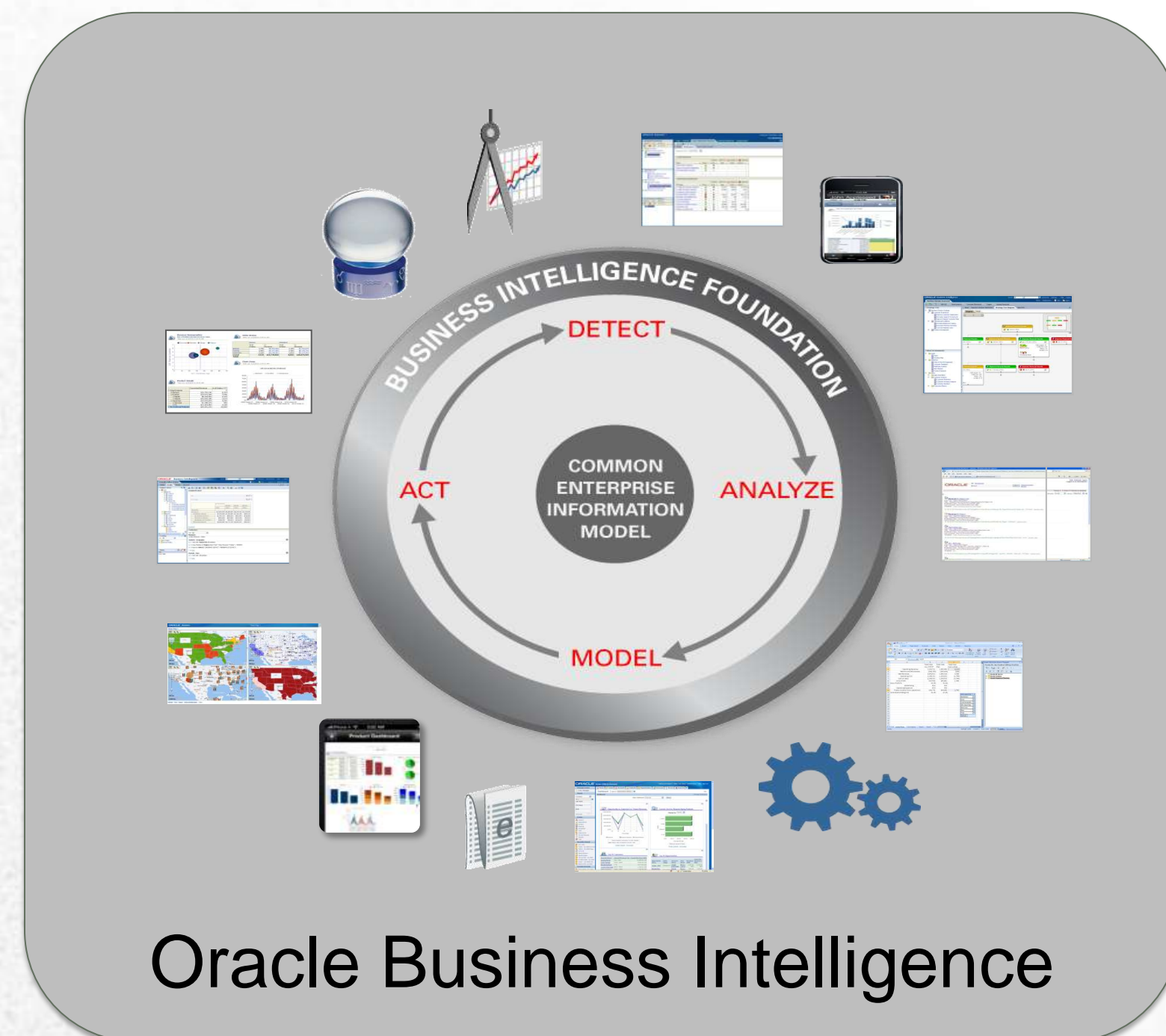
Аппаратное обеспечение



- **Оперативная память**
1 TB RAM, 1033 MHz
- **Процессоры**
4 Intel® Xeon® E7-4870, 40 cores
- **Сетевые интерфейсы**
40 Gbps InfiniBand – 2 ports
10 Gbps Ethernet – 2 ports
1 Gbps Ethernet – 4 ports
- **Дисковая память**
3.6 TB HDD Capacity

Программное обеспечение

- Oracle Business Intelligence
- Oracle Essbase
- Oracle TimesTen for Exalytics
- Адаптивные in-memory акселераторы

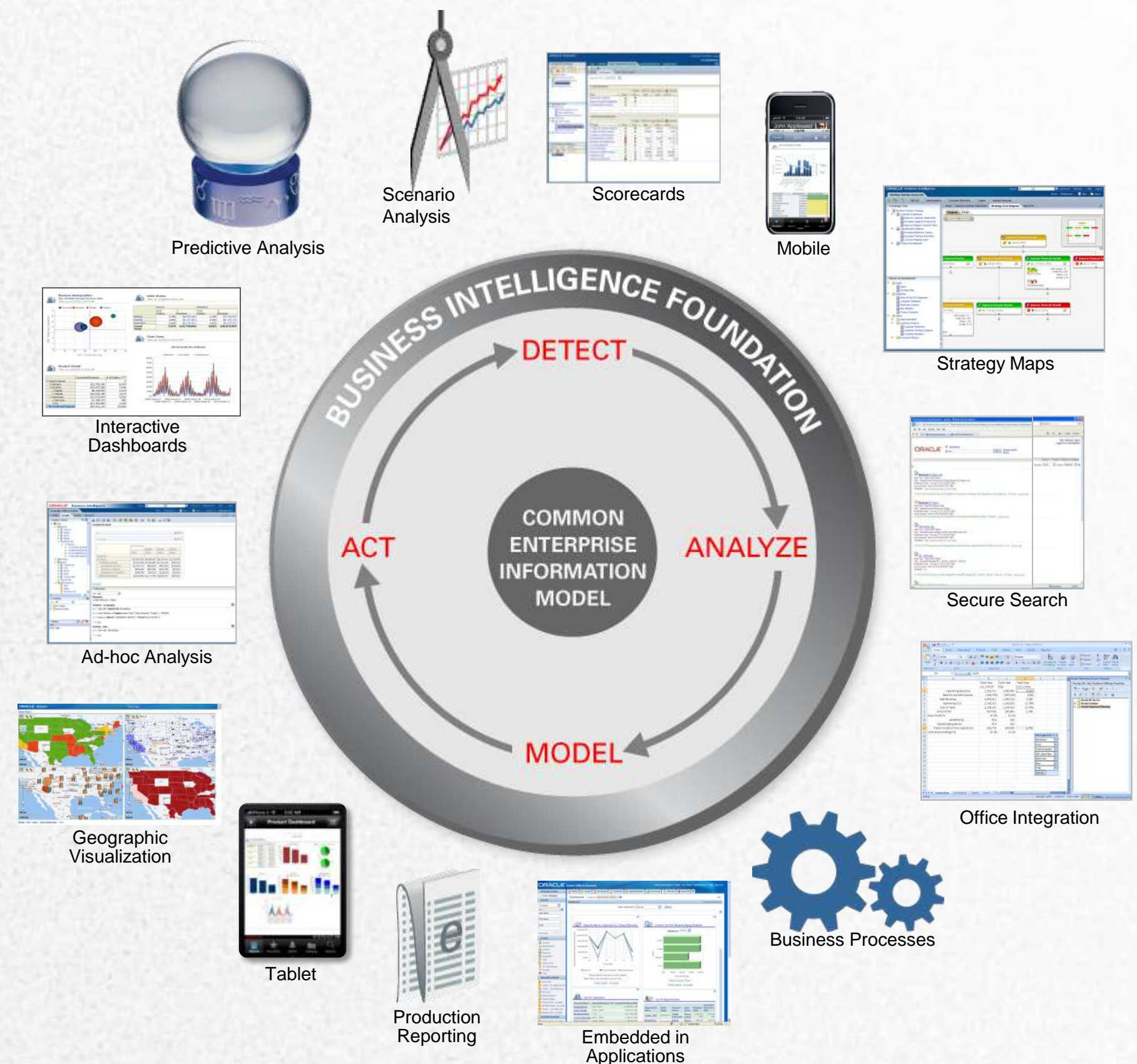


Адаптивные in-memory акселераторы

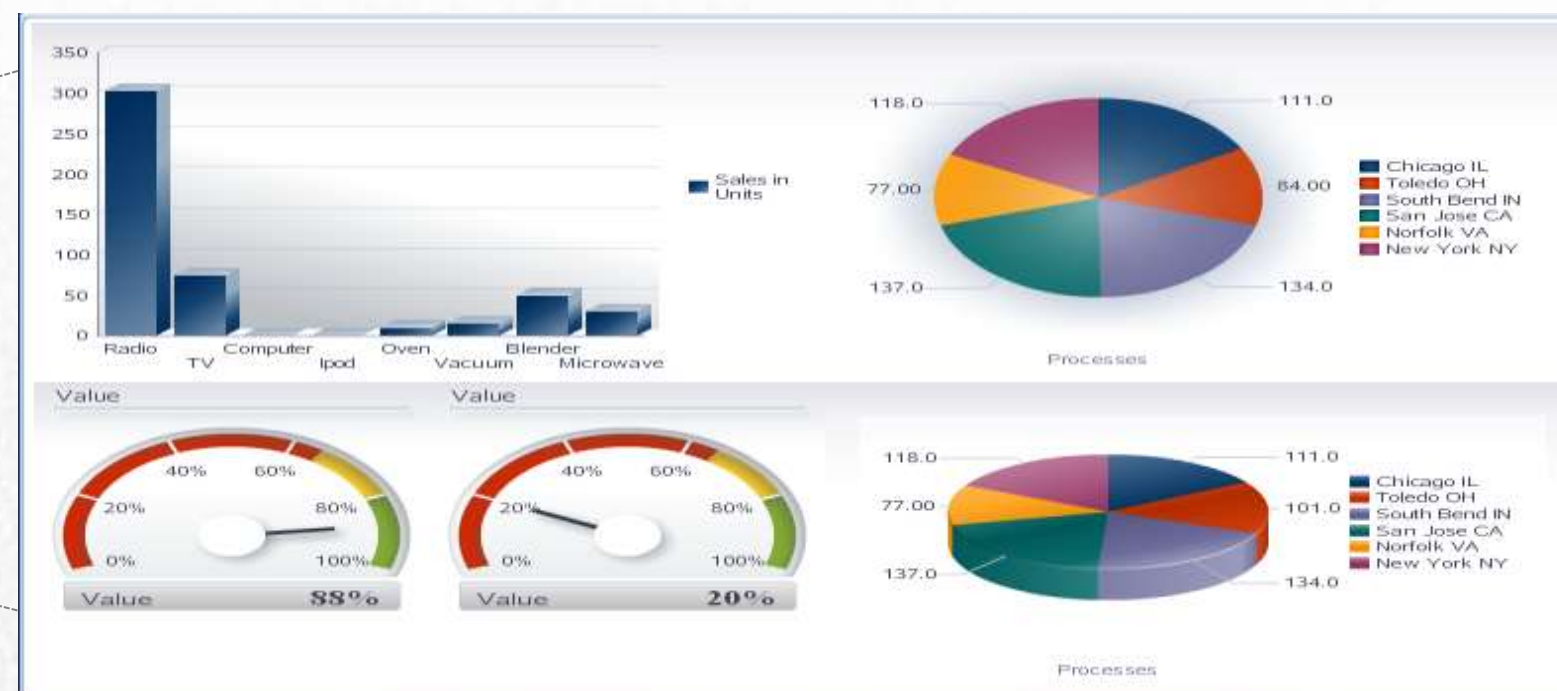
Oracle Business Intelligence

Полная интегрированная система бизнес-анализа

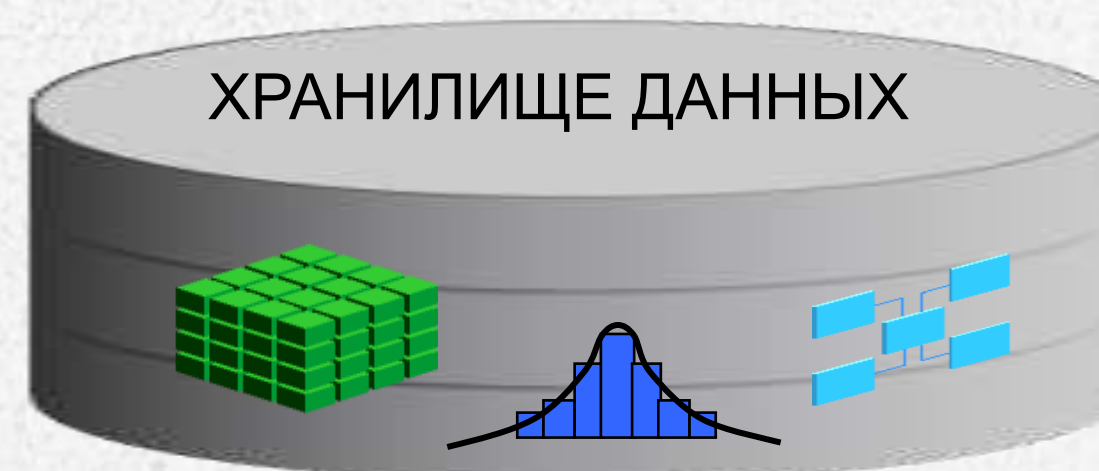
- Единая платформа для всех видов анализа
- Доступ ко всем данным предприятия и внешним источникам
- Интеграция с любыми системами, приложениями и бизнес-процессами



Инструментальная среда Oracle BI



Oracle BI



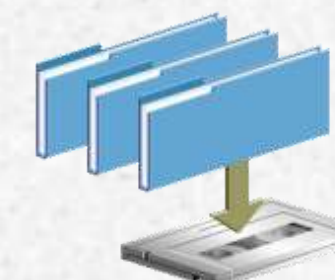
Oracle Database



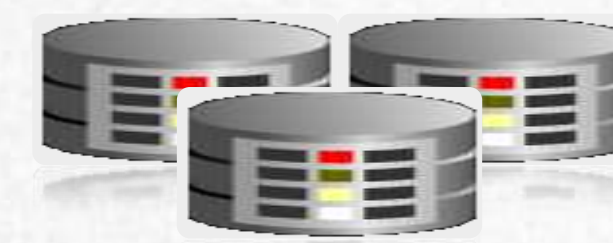
Oracle ERP, CRM



Sybase, DB2, MS SQL Server



Файлы, MS Excel



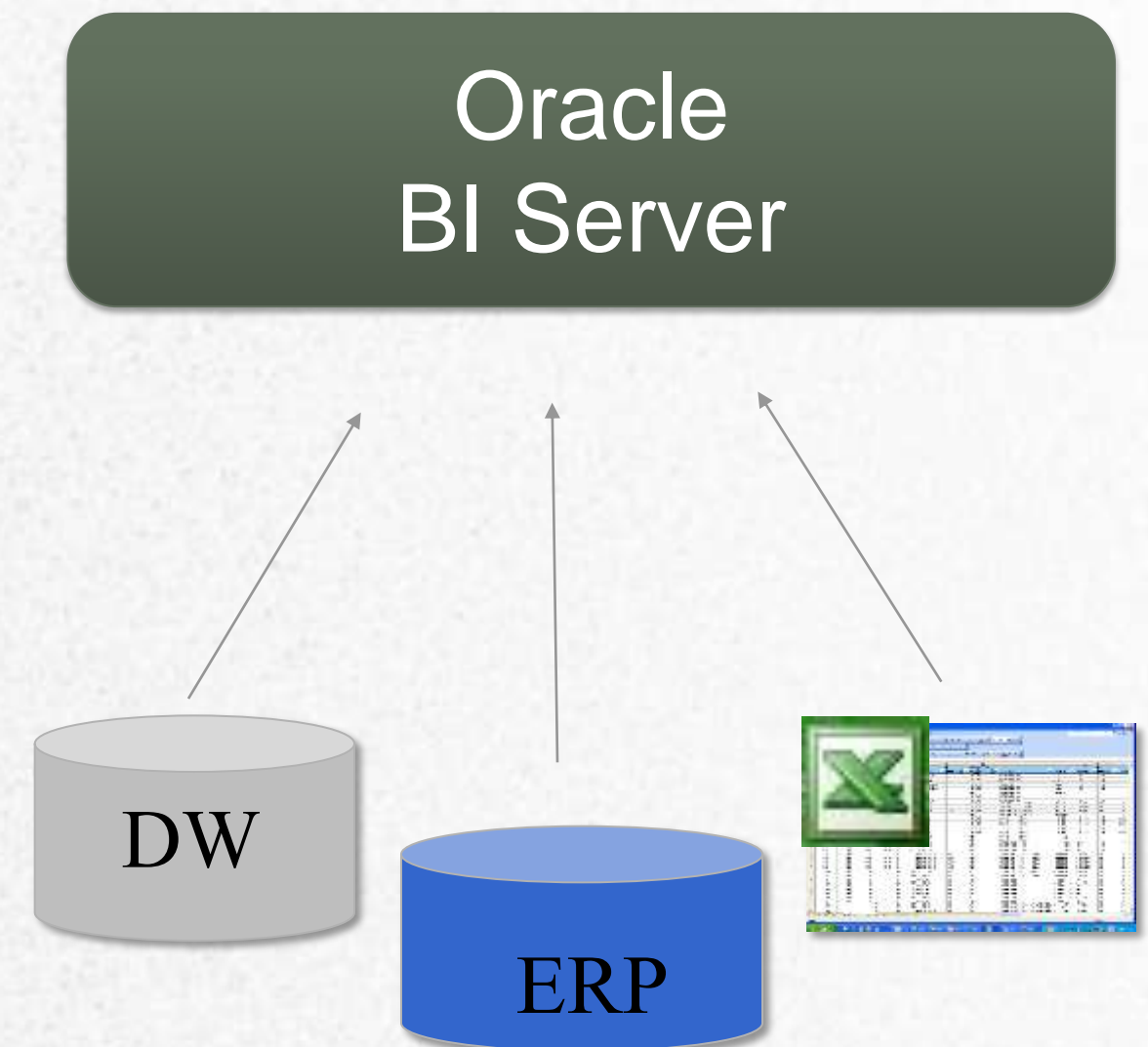
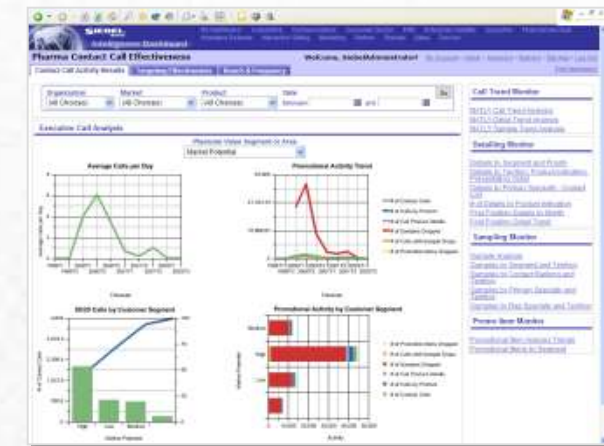
SAP

ORACLE®

Аналитический сервер

Oracle BI Server

- Генерация запросов и оптимизация доступа к информационным источникам
- Выполнение сложных вычислений и аналитическая обработка
- Доступ к любым хранилищам данных, информационным источникам и системам
 - Прямой доступ к Oracle Database, DB2, MS SQL Server, MS SQL Server Analysis Services, Teradata
 - Реляционные, многомерные, файловые системы, XML, и др.
 - Аналитические, оперативные, транзакционные, внешние
- Масштабируемость, надежность, производительность, интеллектуальное кэширование, балансировка нагрузки, работа в кластерной архитектуре



Все виды бизнес-анализа на единой платформе

Не только отчеты

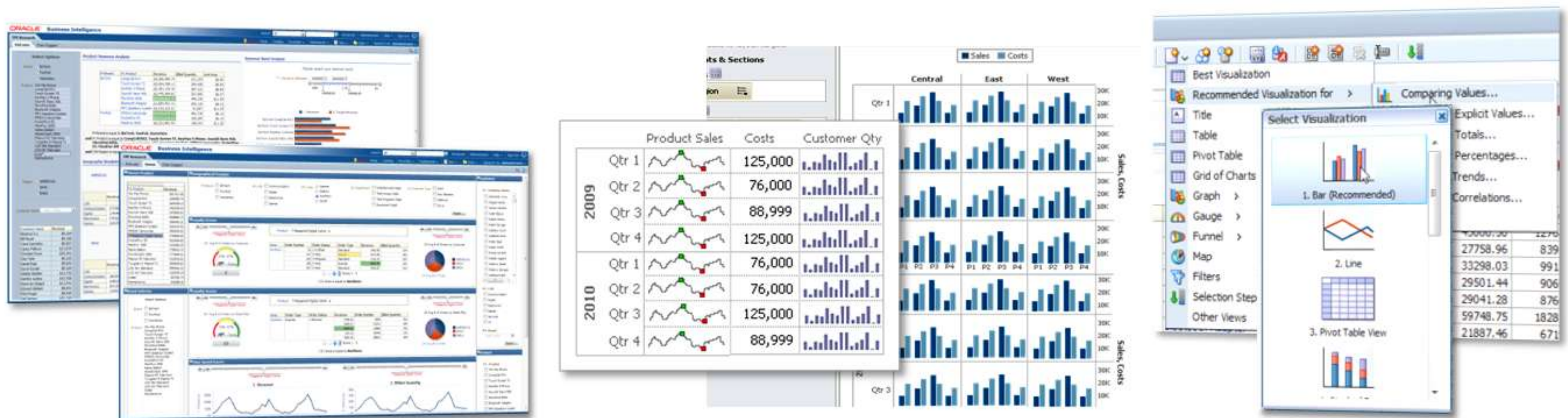
- Интерактивный анализ
- Пространственная аналитика
- Активный бизнес-анализ
- BI Collaboration
- Ключевые показатели и стратегическое управление
- Мобильная аналитика



Новые возможности бизнес-анализа

Аналитика со скоростью мысли

- Высокая степень интерактивности
- Инновационная визуализация
- Рекомендации по выбору визуального представления
- Встроенные сценарии
- Все стили и виды анализа



Vehicle Sales Analyzer

Home Catalog Favorites Dashboards New Open Signed In As prodney

Sales Overview

Dealer Demographics

Sales by Geography

Sales Daily Trend

Forecast by Geography

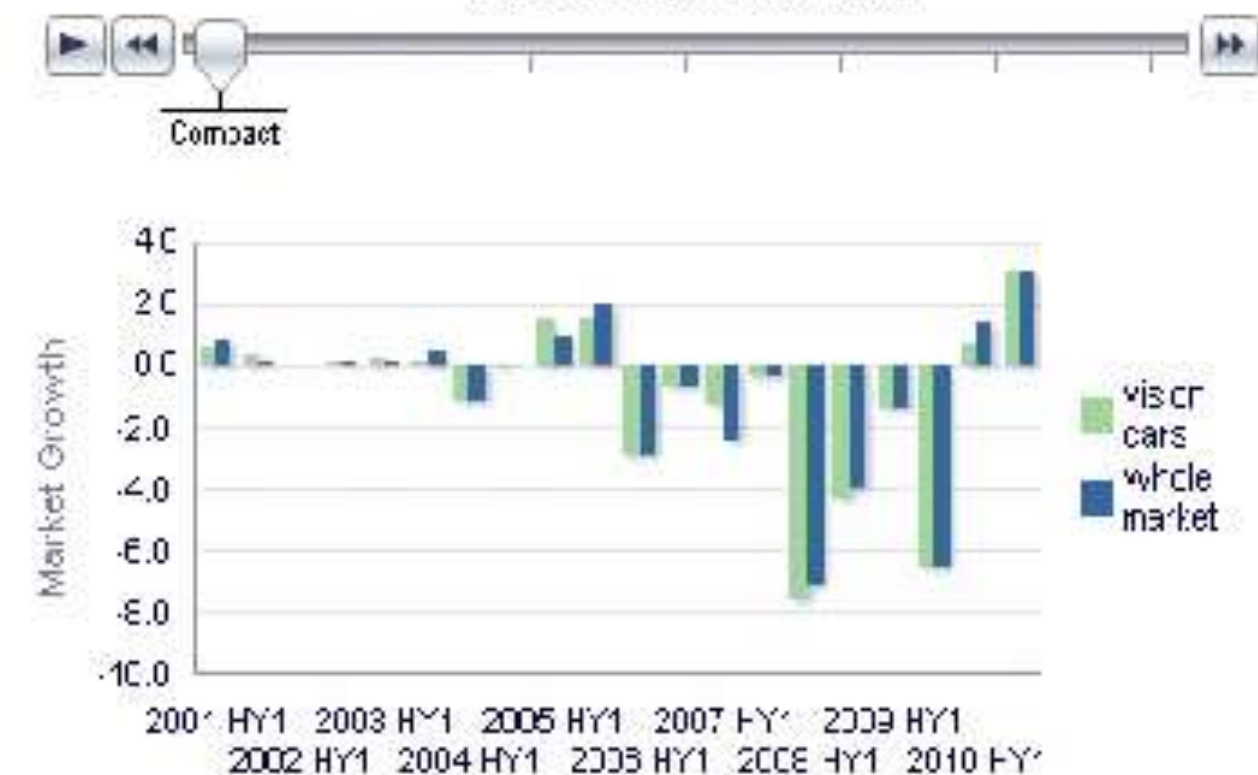
Trend Data by Vehicle Type

Records: 911,523,524

Registrations Summary

	Registrations	Trend	Growth
Compact	121,395,717		
Full Size	62,704,321		
Midsize	105,633,147		
Pickups	74,386,170		
SUVs	135,901,070		
Sports	21,718,648		
Vans	42,174,057		
Total	563,915,164		

Market Growth Over Time

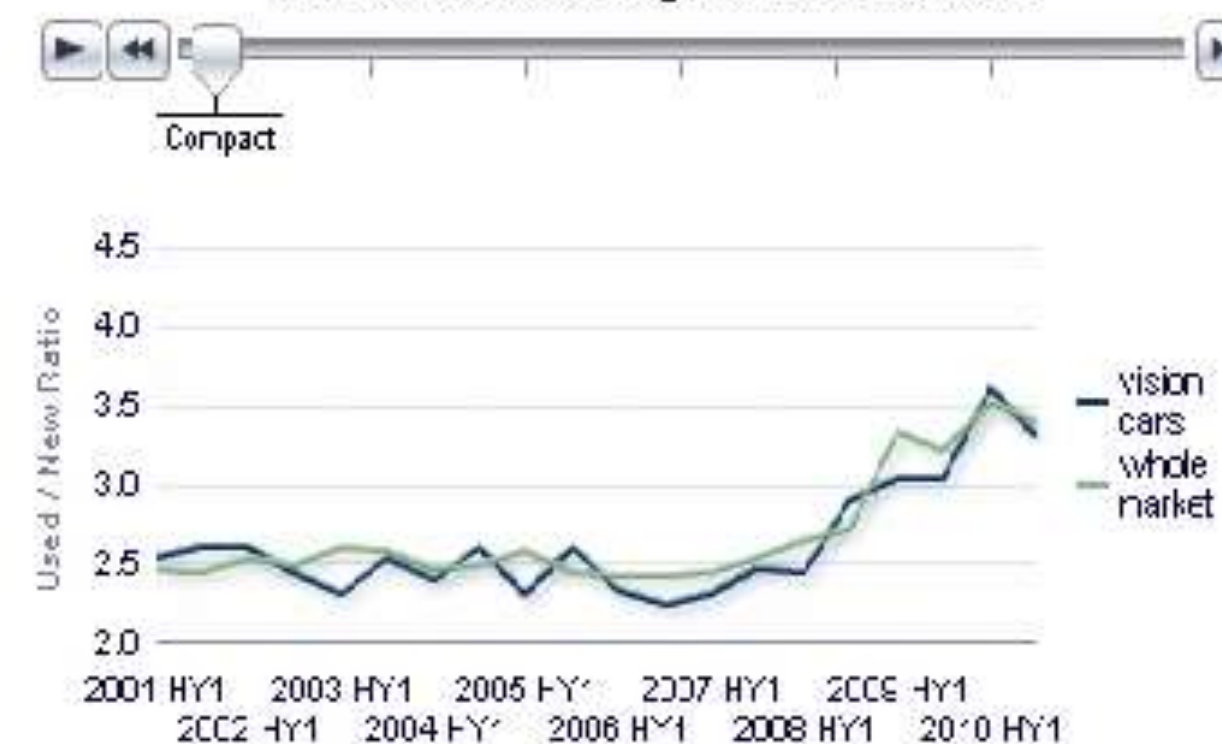


Trend Data by Registration Type

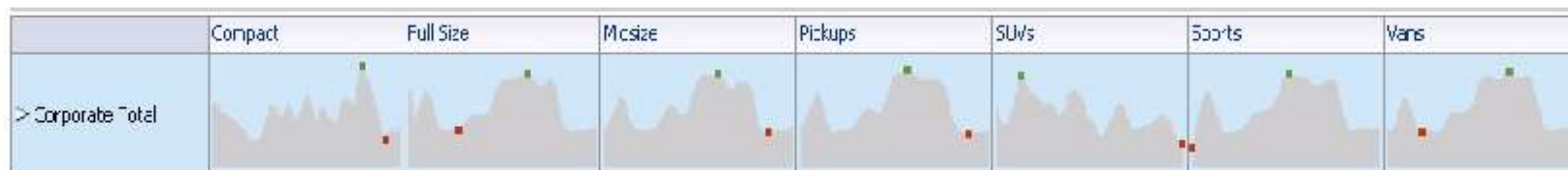
New and Pre-Owned Registration Breakdown

	Registrations	Pre-Owned	New	Pre-Owned/ New Ratio
Compact	121,396,714	88,056,037	33,340,677	2.64
Full Size	62,703,324	45,835,081	16,868,243	2.72
Midsize	105,633,147	76,690,237	28,942,910	2.65
Pickups	74,386,170	53,817,063	20,569,107	2.62
SUVs	135,999,376	96,446,765	39,552,611	2.63
Sports	21,719,545	15,797,465	5,922,080	2.67
Vans	42,174,397	30,362,636	11,811,761	2.57
Total	563,915,164	409,010,277	154,904,887	2.64

New and Pre-owned Registrations Over Time



Trend Data by Region



Company

--Select Value--

Vehicle Type

- ☒ Compact
- ☒ Full Size
- ☒ Midsize
- ☒ Pickups
- ☒ SUVs
- ☒ Sports
- ☒ Vans

Fuel Technology

- ☒ Diesel
- ☒ Gasoline
- ☒ Hybrid
- ☒ Electric

* Sales Period

Between



Financing Option

- ☒ (All Column Values)
- ☐ Lease

Vehicle Sales Analyzer

Sales Overview Dealer Demographics Sales by Geography Sales Daily Trend Forecast by Geography



Records: 41,335,900

Company

Mission Cars

Vehicle Type

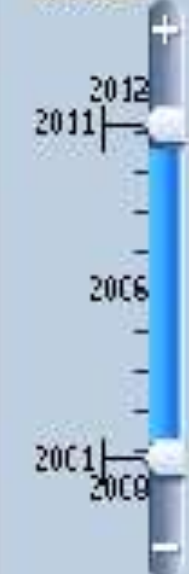
- ☒ Compact
- ☒ Full Size
- ☒ Midsize
- ☐ Pickups
- ☒ SUVs
- ☒ Sports
- ☐ Vans

Fuel Technology

- ☒ Diesel
- ☒ Gasoline
- ☒ Hybrid
- ☒ Electric

* Sales Period

Between



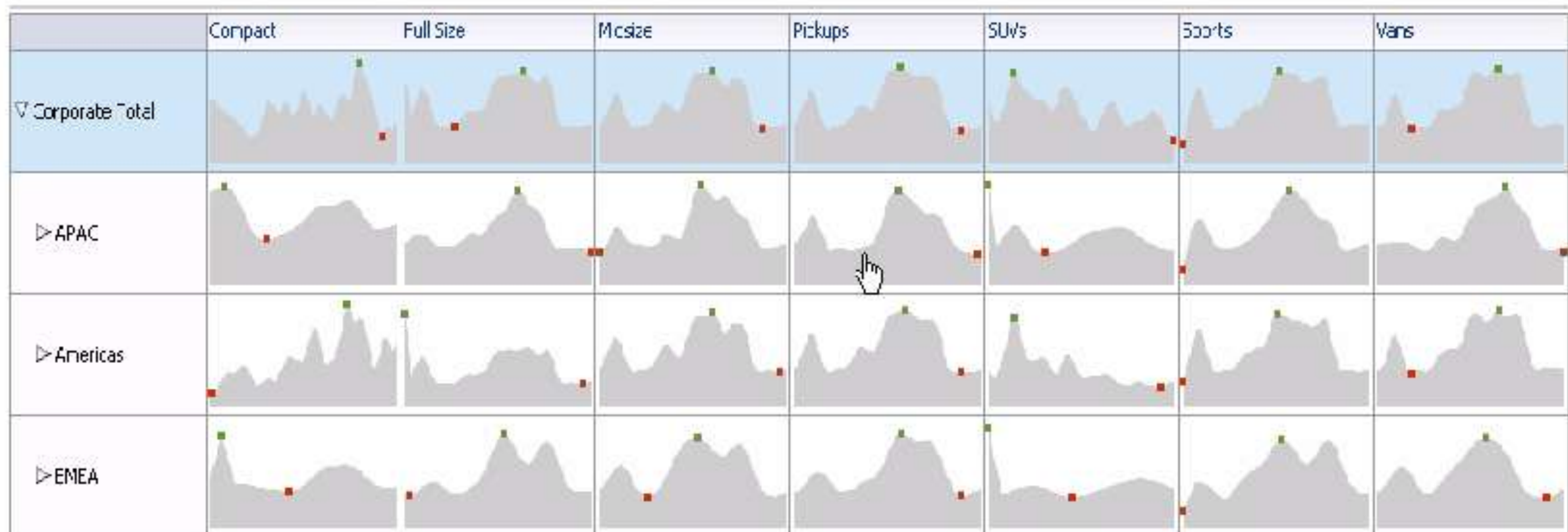
Financing Option

- ☒ (All Column Values)
- ☐ Lease

> Trend Data by Vehicle Type

> Trend Data by Registration Type

> Trend Data by Region



Records: 41,305,900

Company

Mission Cars

Vehicle Type

- ☒ Compact
- ☒ Full Size
- ☒ Midsize
- ☐ Pickups
- ☒ SUVs
- ☒ Sports
- ☐ Vans

Fuel Technology

- ☒ Diesel
- ☒ Gasoline
- ☒ Hybrid
- ☒ Electric

* Sales Period



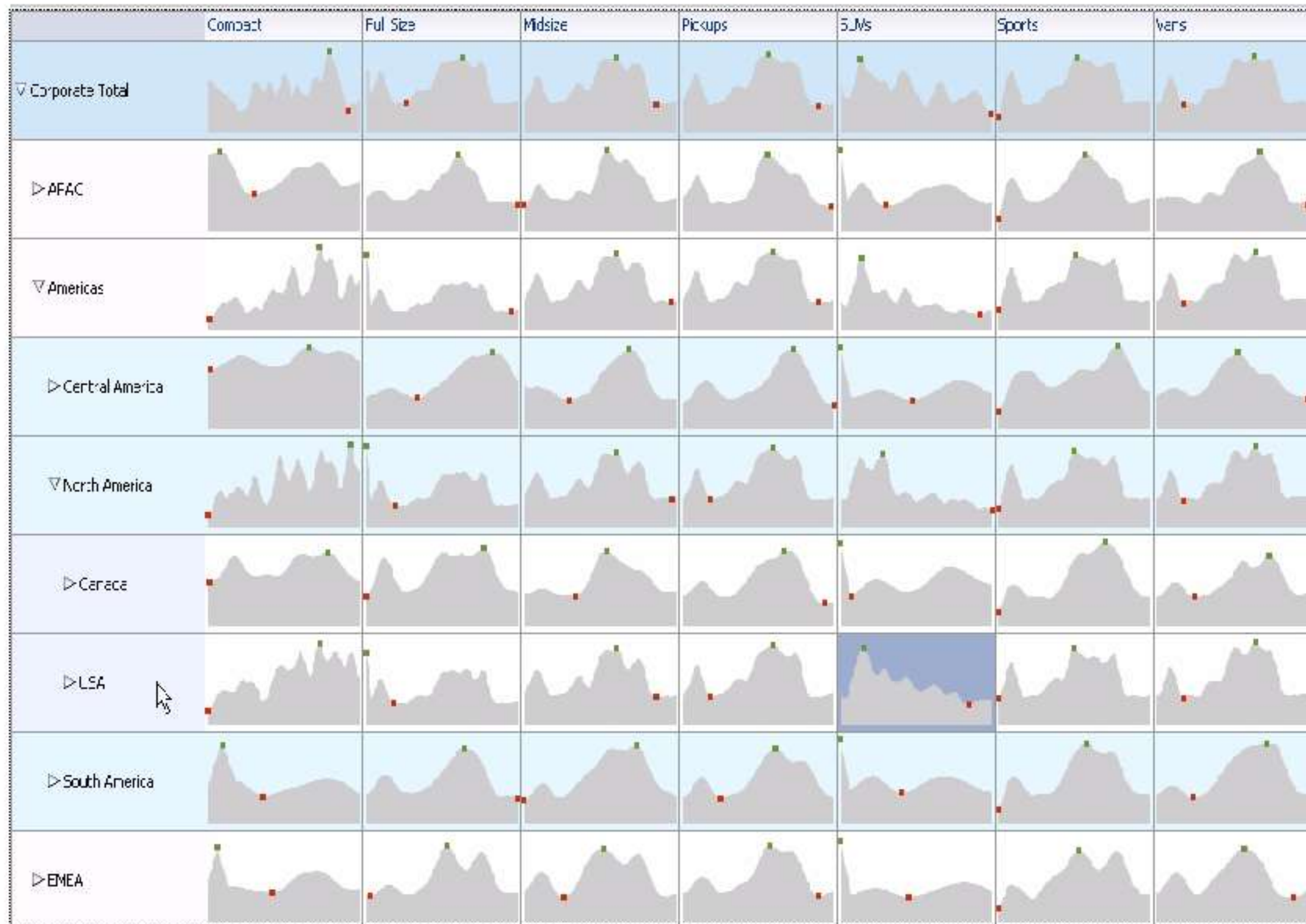
Financing Option

- ☒ (All Column Values)
- ☐ Lease

> Trend Data by Vehicle Type

> Trend Data by Registration Type

☒ Trend Data by Region



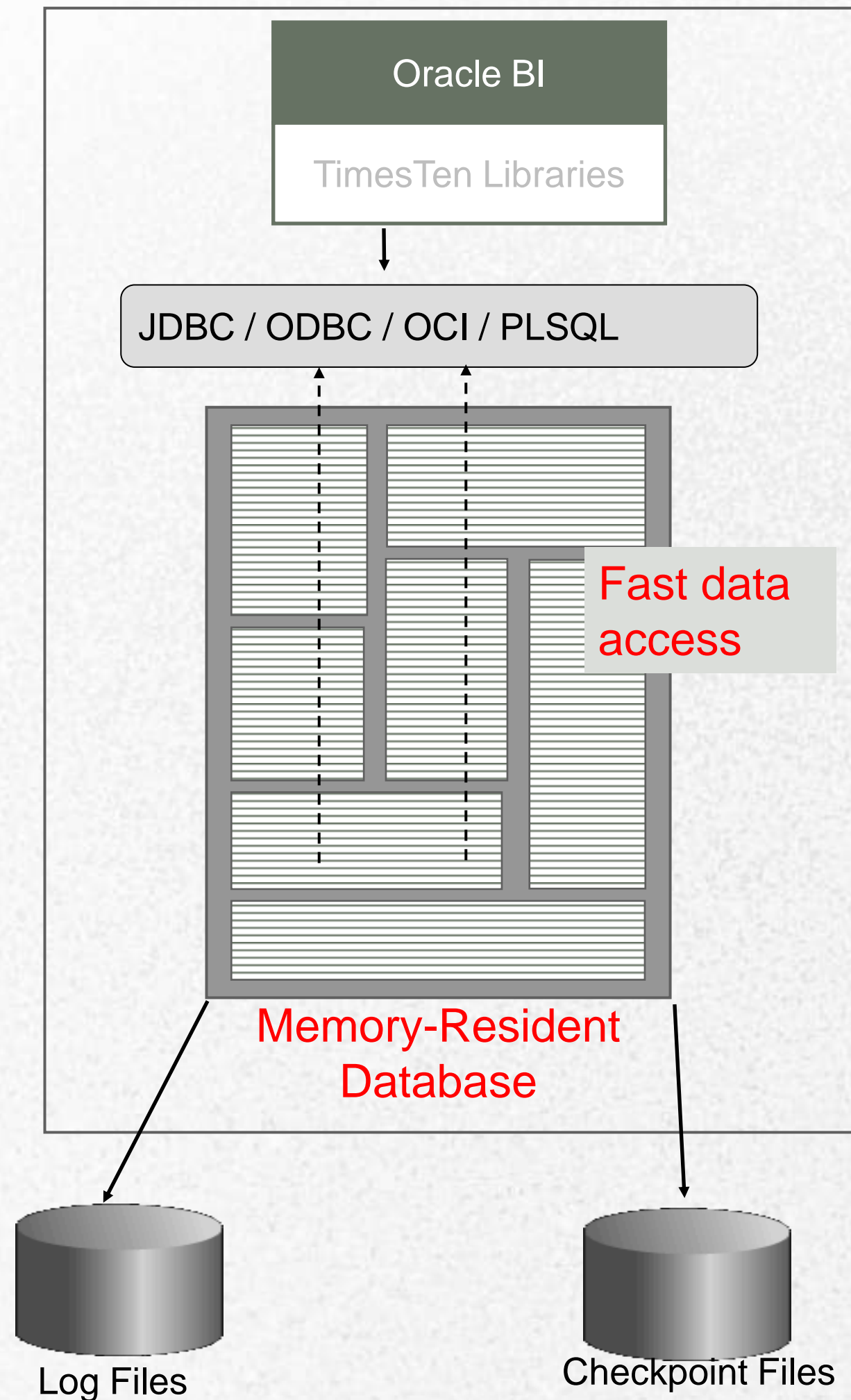
Oracle Essbase

- Универсальный **OLAP-сервер** для хранения, обработки и представления информации
- Высокая производительность
- Моделирование сложной аналитики
- Основа продукта Hyperion Planning
- Essbase на Exalytics оптимизирован для повышения производительности работы Hyperion Planning и других EPM-приложений



Times Ten for Exalytics

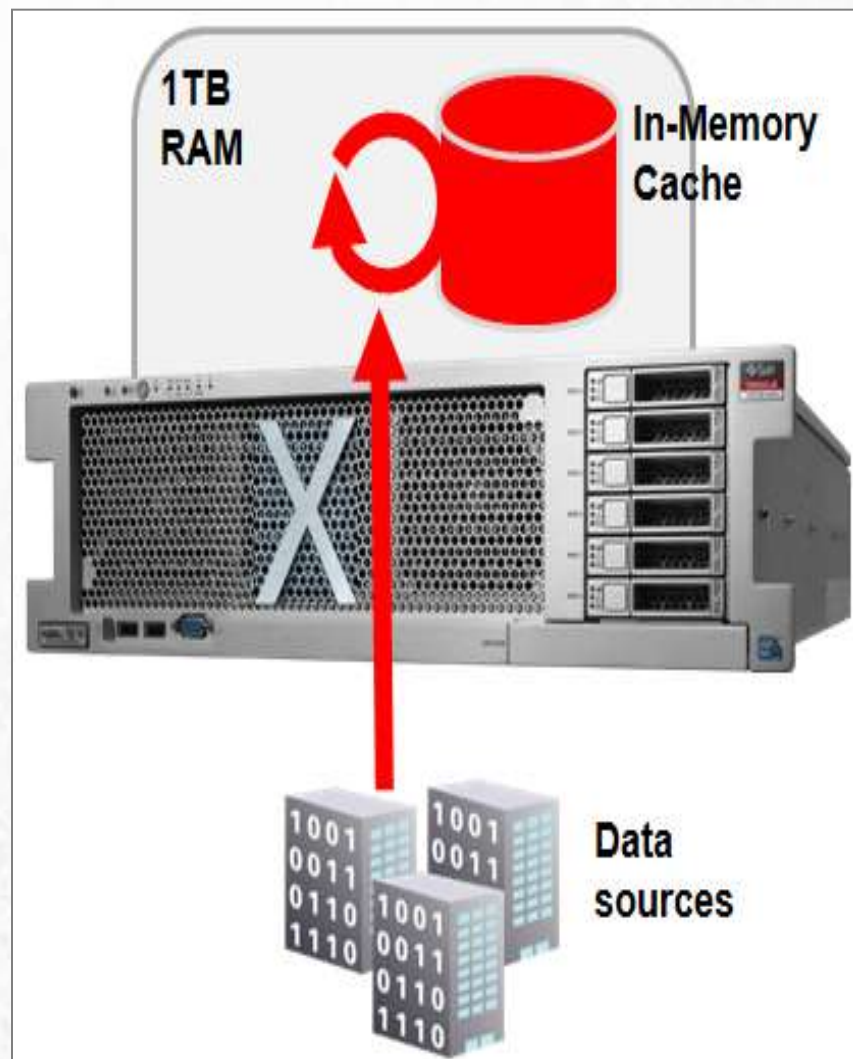
На основе Oracle TimesTen In-Memory Database



- 100% In-memory RDBMS
 - База данных в оперативной памяти
- Высокая производительность
 - Снижение времени отклика
 - Высокая пропускная способность
- Сохранение в дисковой памяти
 - Транзакции и контрольные точки копируются для постоянного хранения
- Колоночная компрессия
 - Сжатие от 5 до 10 раз
 - Аналитические алгоритмы работают непосредственно с компрессированными данными
- Аналитические функции
 - Эффективное выполнение аналитических функций
 - Разгрузка BI-сервера

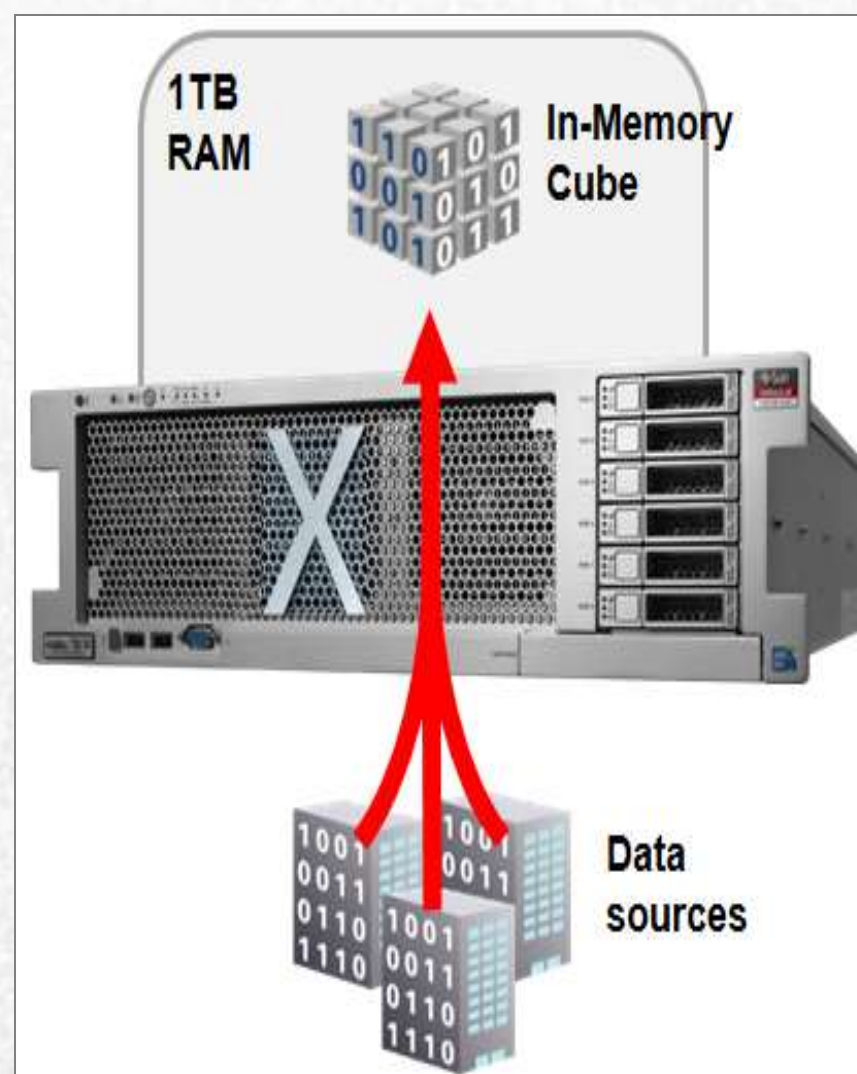
Адаптивные in-memory акселераторы

Кэширование данных в оперативной памяти



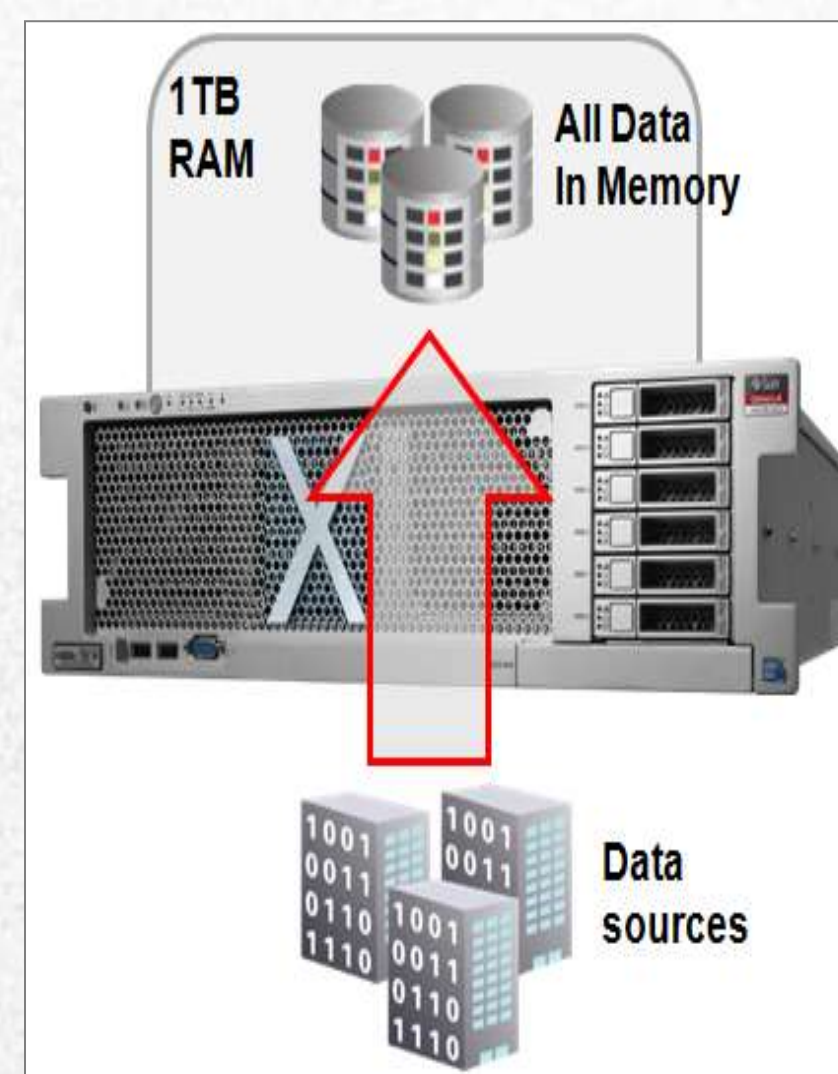
In-Memory Intelligent Result Cache

Хранение в оперативной памяти результатов выполненных запросов



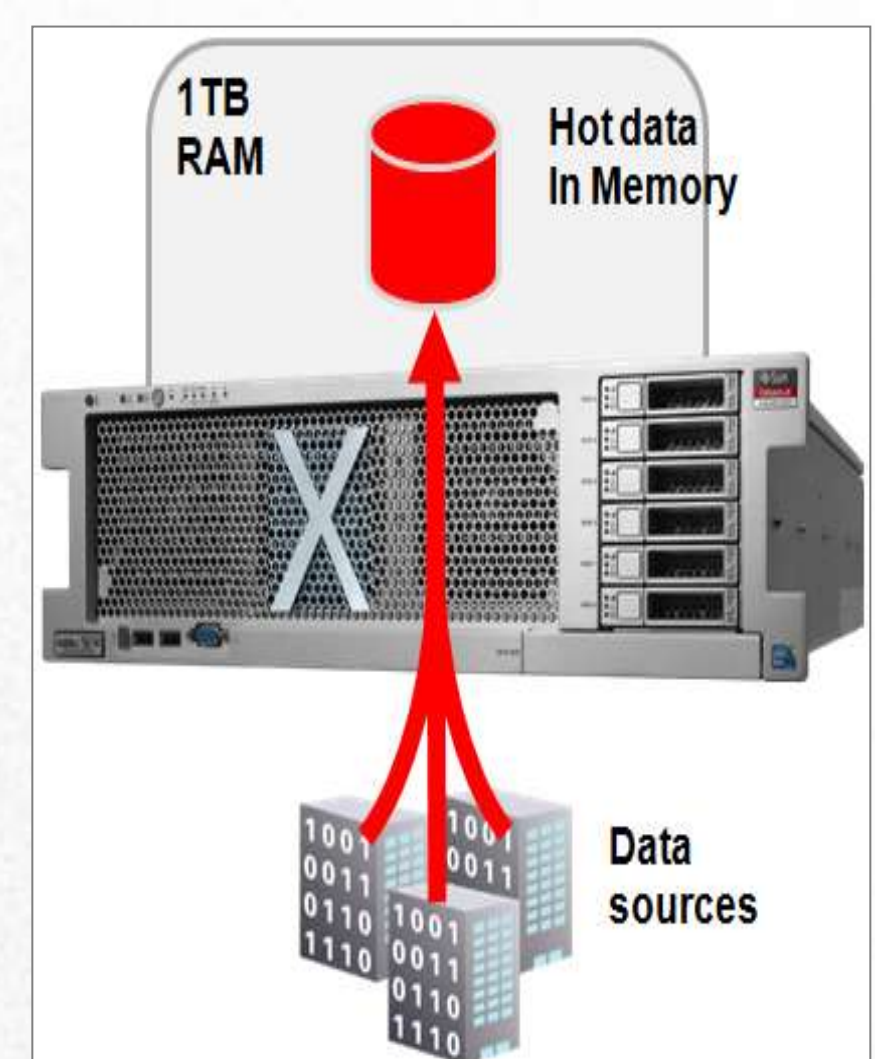
In-Memory Essbase Cubes

Кэширование в оперативной памяти Essbase-кубов
Нет в 1-ом релизе



In-Memory Data Warehouse

Хранение в оперативной памяти всего хранилища данных



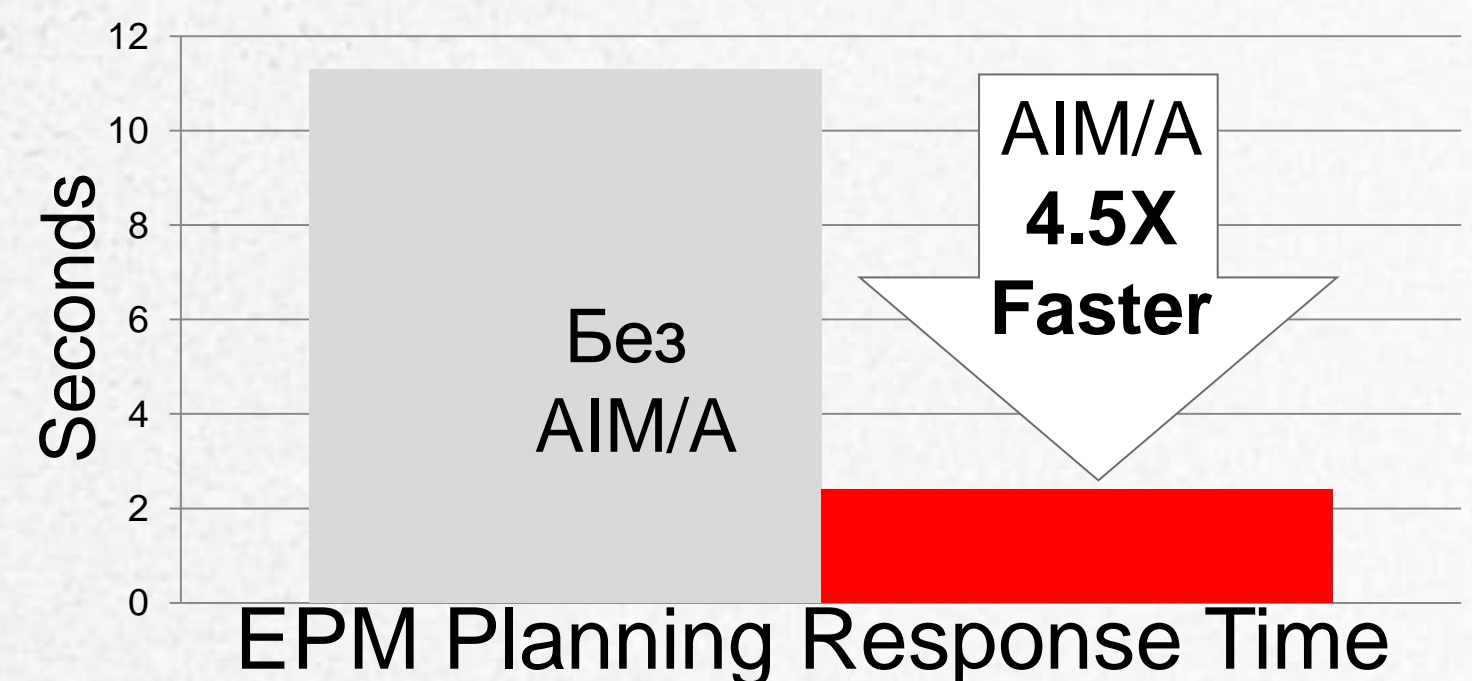
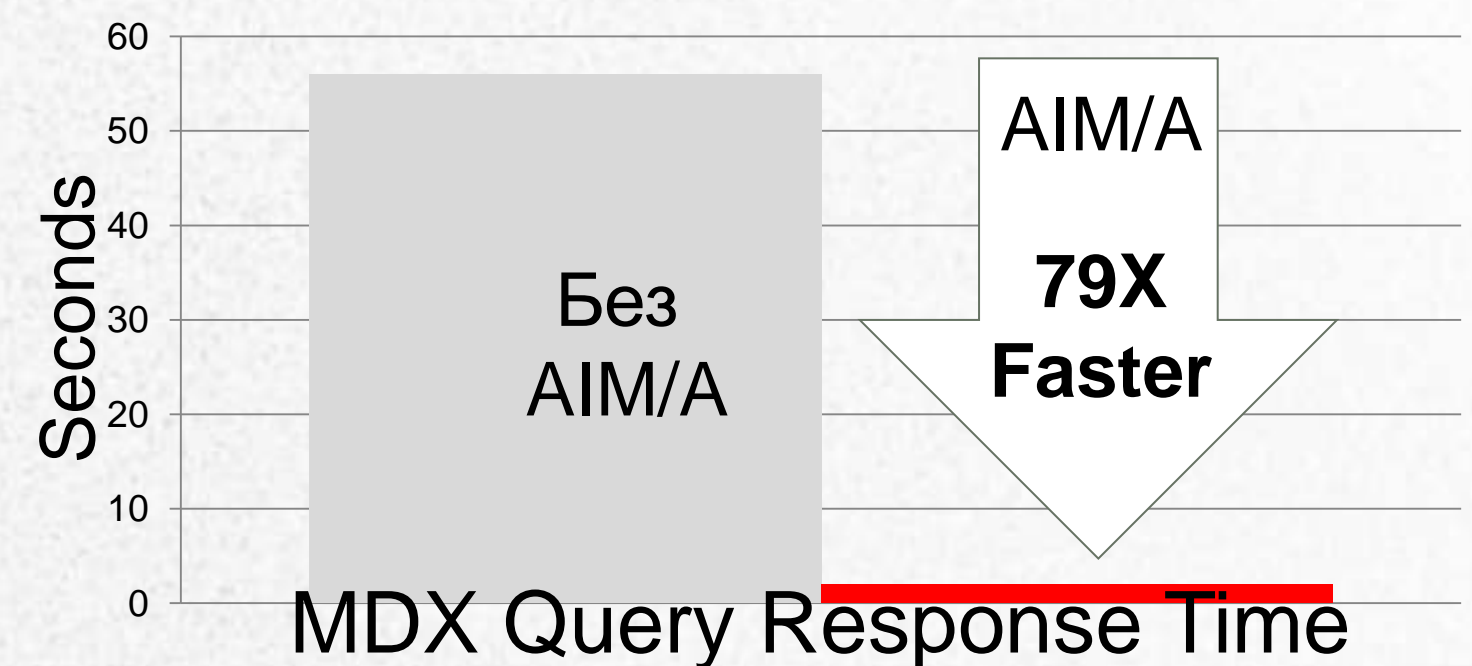
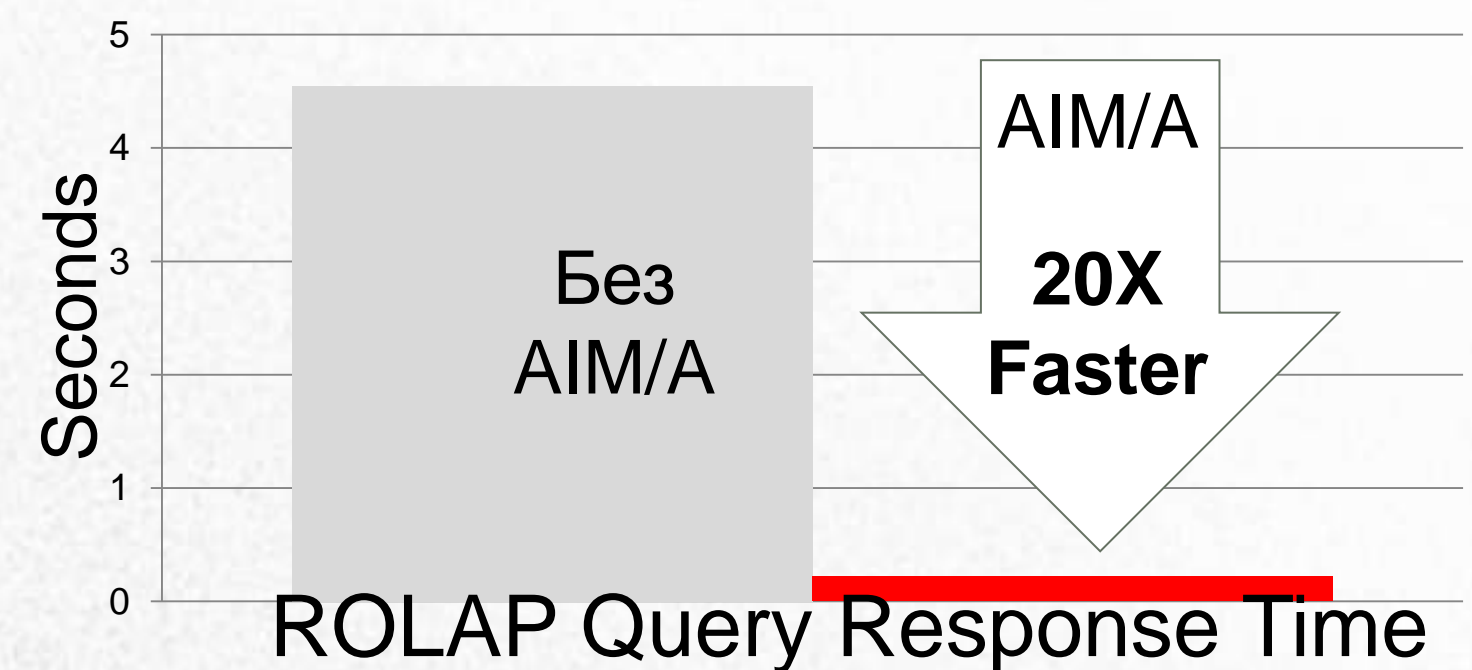
In-Memory Adaptive Data Marts

Кэширование в TimesTen
Автоматическое формирование витрины на основе статистики запросов

Что дают in-memory акселераторы

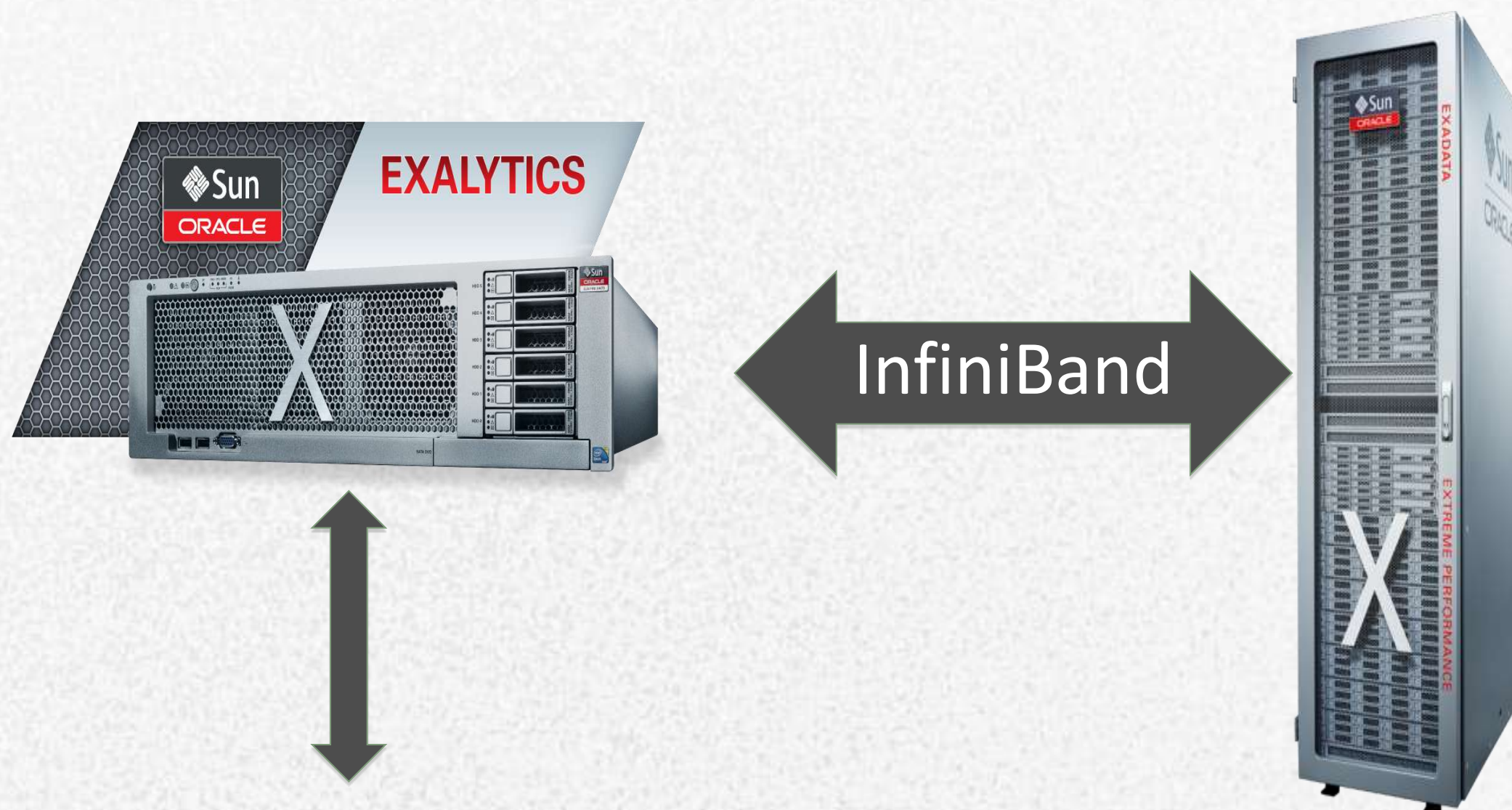
Повышение производительности

- **In-Memory Analytics: ROLAP**
 - **20X** снижение времени отклика
 - **50,000** пользователей на одном комплексе
- **In-Memory Analytics: MOLAP**
 - **79X** снижение времени отклика при чтении
 - **16X** снижение времени отклика при записи
- **In-Memory Analytics: EPM Planning**
 - **4.5X** снижение времени отклика
 - **10,000** пользователей на одном комплексе



Бизнес-анализ и источники данных

Аналитика в оперативной памяти



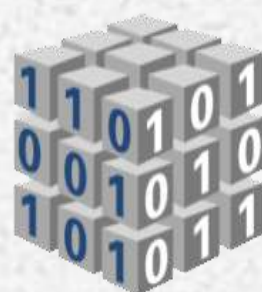
Экстремальная производительность хранилища данных

Любые источники данных

OLTP & ODS Data Warehouse Data Mart



OLAP Sources



Packaged Applications (Oracle, SAP, Others)



Unstructured & Semi-Structured



Excel XML/Office

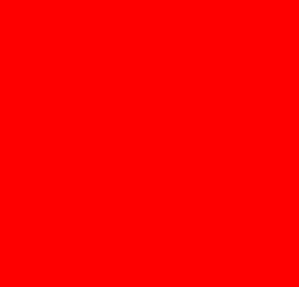


Business Process



План

- Платформа Oracle для Больших Данных
- Аналитическая in-memory машина Oracle Exalytics
- **От бизнес-анализа к исследованию данных – Oracle Endeca Information Discovery**
- Встроенная аналитика для статистических исследований – Oracle R Enterprise



**Oracle Endeca Information Discovery –
платформа для исследования
структурированных,
слабо-структурированных и
неструктурированных данных с
меняющейся структурой в условиях
нечетких критериев поиска**

Новые требования в анализу данных

Аналитика Больших Данных

**Большое
разнообразие
данных**



**Не все можно описать
традиционными
реляционным и
многомерными моделями
данных**

**Большая
скорость
изменения**



**Нельзя ориентироваться
на заранее
установленную модель
данных**

**Больше
непредвиденных
запросов**



**Неизвестно, какую
информацию
можно запрашивать**

Что такое Data Discovery?



- Инструменты обеспечивают высокую степень удобства (usability), гибкости управления процессом моделирования и создания контента
- Высокая степень интерактивности и расширенные возможности визуализации интерфейса, основанного на in-memory архитектурах и быстрым развертыванием
- Возможность пользователям исследовать данные самостоятельно без предварительного обучения
- Примеры: QlikTech QlikView, Tableau, Tibco Spotfire, Microsoft PowerPivot, [Oracle Endeca Information Discovery](#) (previously Endeca Latitude)

**Source: Gartner, June 17, 2011, "Emerging Technology Analysis: Visualization-Based Data Discovery Tools"



- Быстрая разработка и ускоренная реакция на быстро меняющиеся бизнес-требования
- Инкрементальное внедрение, быстрое прототипирование
- Данные, метаданные и приложения (отчеты, дэшборды) не разделяются («one and the same»)
- Легкость и простота создания новых приложений и изменения существующих

**Source: Forrester, "April 22, 2010, "Agile BI Out of the Box," Boris Evelson

От бизнес-анализа к исследованию данных

Когда возникает необходимость нового подхода к анализу данных

- Пользователи не знают **какие вопросы** они будут формулировать:
 - Огромное число различных параметров и характеристик и их состав заранее неизвестен
 - Быстрые изменения информационных источников
 - Невозможно каждый раз обращаться к ИТ специалистам для изменения модели данных
- ИТ-специалисты не знают, **какая модель** данных будет использоваться:
 - Разнообразие данных и наличие большого числа внешних информационных источников
 - Неструктурированные данные
 - Схемы исходных данных часто меняются

Endeca

Историческая справка

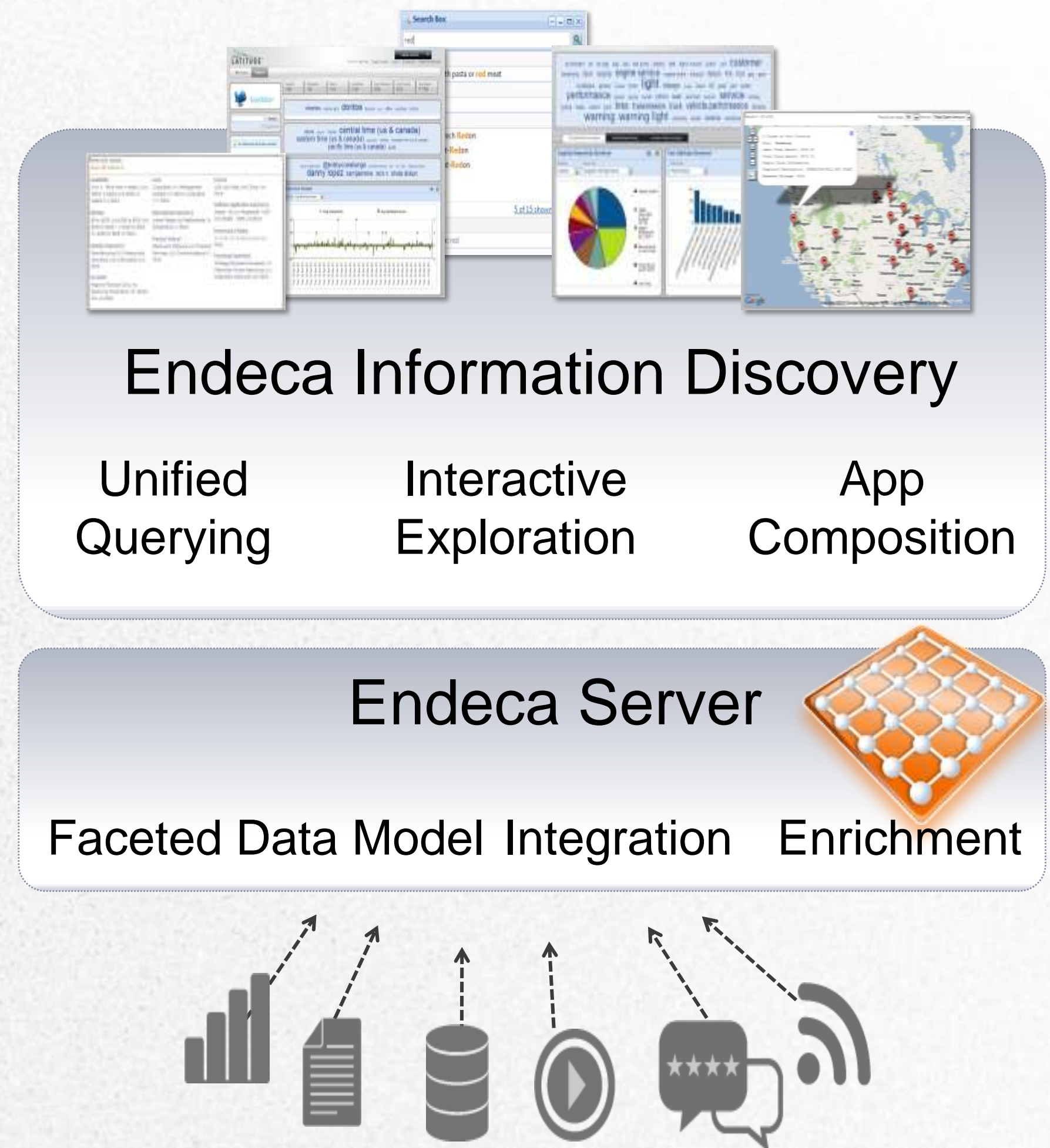
- Основана в Кембридже, МА в 1999
- Более 600 клиентов
- 33% of the Fortune 100
- Анилиз неструктурированной информации, Большие данные
- От систем поиска (для электронной коммерции) к «BI beyond the data warehouse»
- *Entdecken* (немецкий) = to discover, открытие



Oracle Endeca Information Discovery

Платформа для исследования информации

- **Endeca Server**
 - Поисково-аналитическая база данных
- **Information Integration Suite**
 - Загрузка данных в Endeca Server
- **Endeca Studio**
 - Быстрая компонентная разработка приложений для исследования данных



Особенности интерфейса пользователей

- Удобство и простота использования
 - На основе 10-летнего опыта работы в области разработки поисковых систем для электронной коммерции
- Поиск + Фасетная навигация + Визуальный анализ
 - Поиск и выбор атрибутов в стиле вэб сайтов
- Интерактивные исследования
 - Без заранее определенного сценария



Фасетный поиск, фасетная навигация

- Технологии доступа к информации, организованной на основе системы фасетной классификации
- Поиск путем уточнений, навигация по независимым параметрам
- Модель информационного поиска – набор характеристик
- Теоретические основы фасетной классификации -- Ш. Р. Ранганатан («Классификация двоеточием», 1933)

The screenshot shows a search interface for "Digital cameras". It includes a "Refine your results" section with facets for Manufacturer, Resolution, Zoom range, and More. A "Regular search results list" is shown at the bottom with a "COMPARE SELECTED" button. Callouts explain the concepts of facets, constraints, and breadcrumb trails.

Digital cameras

Refine your results

Manufacturer

- Canon USA (5)
- Sony (2)
- Nikon (2)
- Olympus (6)
- Pentax (2)

Resolution

- 6 megapixels (3)
- 8 megapixels and up (14)

Zoom range

- 3X to 4X (11)
- 8X to 12X (1)

More

- LCD size
- Image stabilizer
- Flash memory
- Still image format
- Maximum ISO

[See all >](#)

you selected: \$400 - \$500 SLR remove all

17 results

Regular search results list

1 2 next

Show 10 results per page Sort by: Review date

COMPARE SELECTED

 **Canon EOS Rebel XS (silver, with 18-55mm lens)** \$459 to \$699 at 15 stores

Callouts:

- Manufacturer is a **facet**, a way of categorizing the results
- Canon, Sony, and Nikon are **constraints**, or facet values
- The **facet count** or constraint count shows how many results match each value
- The **breadcrumb** trail shows what constraints have already been applied and allows for their removal

Разработка приложений

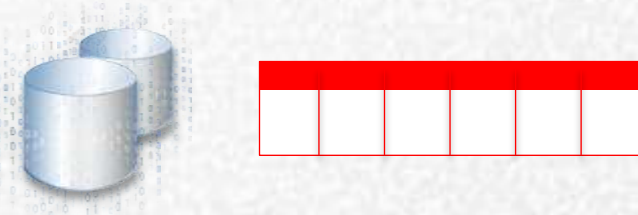
Источники данных

Загрузка в Endeca Server (без модели)

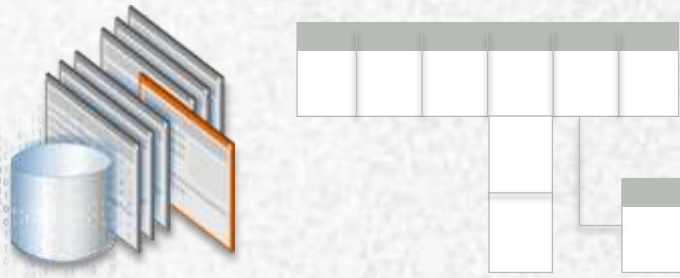
Drag-and-drop инструменты создания приложения

Интерактивный поиск, навигация и анализ

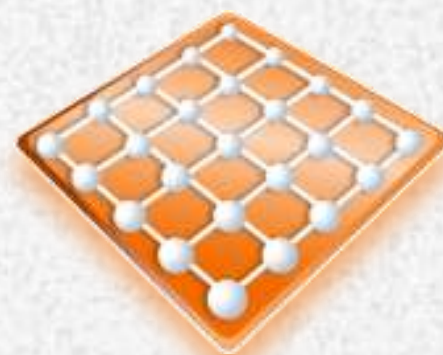
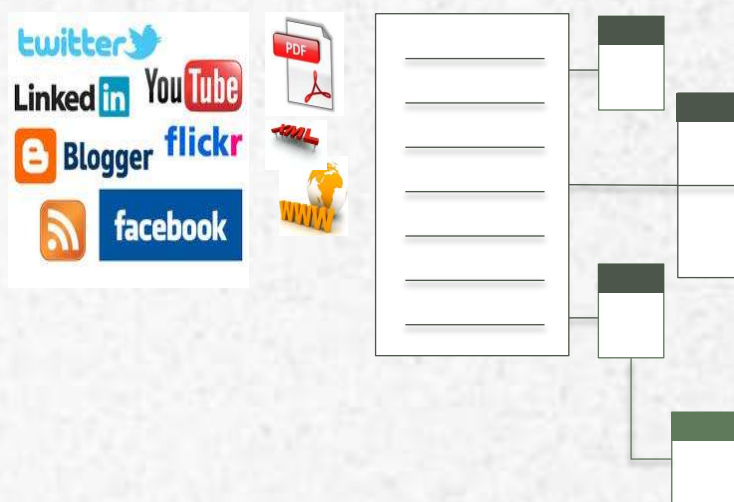
Структурированные



Слабоструктурированные



Неструктурированные



Фасетная модель данных

Позволяет быстро загружать и изменять данные из любого источника

- Набор записей, каждая из которых имеет собственную «структуру»
 - Многозначные поля
 - Неструктурированные поля (тексты)
- Один из видов «Key Value» модели
- Модель состоит из
 - Записей
 - Атрибутов
- Каждая запись – это набор пар (атрибут, значение)
- Нет никакой разбиения на таблицы
- Нет понятия схемы данных

- TxnID = 12324
- ProductID = 506
- Category = Mountain Bike
- Amount = \$499.99
- Suspension = Fox 32 F-Series
- FrameType = Aluminium
- Saddle = Bontrager SSR
- Mountain Accessories = Fork and shock sag meter
- Mountain Accessories = Water Bottle
- Review = A great bike for off road. Smooth ride over the bumps
- ReviewSentiment = Positive
- ReviewTerm = Great
- ReviewTerm = Off Road
- ReviewTerm = Smooth
- ReviewTerm = Bumps

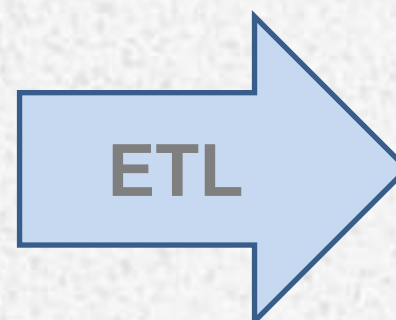
- TxnID = 12325
- ProductID = 507
- Category = Road Bike
- Amount = \$1399.49
- Weight = 20lb.
- FrameType = Composite
- Saddle = Bontrager Race
- Review = Disappointing for the price. The frame feels heavier than I expected.
- ReviewSentiment = Negative
- ReviewTerm = Disappointing
- ReviewTerm = Price
- ReviewTerm = Heavier

Загрузка данных : Структурированные источники

- Структурированные данные загружаются в фасетную модель естественным образом
 - Каждый кортеж становится записью
 - Каждый столбец становится атрибутом

Transaction			
TxnID	ProductID	Category	Amount
12324	506	Mountain Bike	499
12325	507	Road Bike	1399

Реляционная таблица



- TxnID = 12324
- ProductID = 506
- Category = Mountain Bike
- Amount = \$499.99
- Suspension = Fox 32 F-Series
- FrameType = Aluminium
- Saddle = Bontrager SSR
- Mountain Accessories = Fork and shock sag meter
- Mountain Accessories = Water Bottle
- Review = A great bike for off road. Smooth ride over the bumps
- ReviewSentiment = Positive
- ReviewTerm = Great
- ReviewTerm = Off Road
- ReviewTerm = Smooth
- ReviewTerm = Bumps

- TxnID = 12325
- ProductID = 507
- Category = Road Bike
- Amount = \$1399.49
- Weight = 20lb.
- FrameType = Composite
- Saddle = Bontrager Race
- Review = Disappointing for the price. The frame feels heavier than I expected.
- ReviewSentiment = Negative
- ReviewTerm = Disappointing
- ReviewTerm = Price
- ReviewTerm = Heavier

Загрузка данных : Слабо-структурированные источники

- Слабоструктурированные данные из различных источников -- промышленные системы, XML источники, внешние приложения. Загружаются в виде пар атрибут-значение
- Гетерогенные структуры записей

```
<ProductData ProductID="506">
  <Attribute Key="Suspension">Fox 32 F-
Series</Attribute>
  <Attribute Key="FrameType">Aluminium</Attribute>
  <Attribute Key="Saddle">Bontrager SSR</Attribute>
  <Attribute Key="Mountain Accessories">
    Fork and shock sag meter</Attribute>
  <Attribute Key="Mountain Accessories">
    Water Bottle</Attribute>
</ProductData>

<ProductData ProductID="507">
  <Attribute Key="Weight">20lb.</Attribute>
  <Attribute Key="FrameType">Composite</Attribute>
  <Attribute Key="Saddle">Bontrager
Race</Attribute>
</ProductData>
```

ETL

- TxnID = 12324
- ProductID = 506
- Category = Mountain Bike
- Amount = \$499.99
- **Suspension = Fox 32 F-Series**
- **FrameType = Aluminium**
- **Saddle = Bontrager SSR**
- **Mountain Accessories = Fork and shock sag meter**
- **Mountain Accessories = Water Bottle**
- Review = A great bike for off road. Smooth ride over the bumps
- ReviewSentiment = Positive
- ReviewTerm = Great
- ReviewTerm = Off Road
- ReviewTerm = Smooth
- ReviewTerm = Bumps

- TxnID = 12325
- ProductID = 507
- Category = Road Bike
- Amount = \$1399.49
- **Weight = 20lb.**
- **FrameType = Composite**
- **Saddle = Bontrager Race**
- Review = Disappointing for the price. The frame feels heavier than I expected.
- ReviewSentiment = Negative
- ReviewTerm = Disappointing
- ReviewTerm = Price
- ReviewTerm = Heavier



Загрузка данных : *Неструктурированные источники*

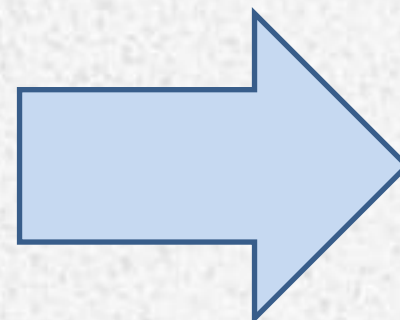
- Неструктурированные данные «присоединяются» к остальным данным на основе значения ключа
- В определенных случаях неструктурированные данные могут храниться в виде независимых записей
- Текстовые документы, RSS, Twitter, Facebook, форумы, etc..

Review: #1301
Product: 506

A great bike for off road. Smooth ride over the bumps

Review: #1327
Product: 507

Disappointing for the price. The frame feels heavier than I expected.



- TxnID = 12324
- ProductID = 506
- Category = Mountain Bike
- Amount = \$499.99
- Suspension = Fox 32 F-Series
- FrameType = Aluminium
- Saddle = Bontrager SSR
- Mountain Accessories = Fork and shock sag meter
- Mountain Accessories = Water Bottle
- **Review = A great bike for off road. Smooth ride over the bumps**
- ReviewSentiment = Positive
- ReviewTerm = Great
- ReviewTerm = Off Road
- ReviewTerm = Smooth
- ReviewTerm = Bumps

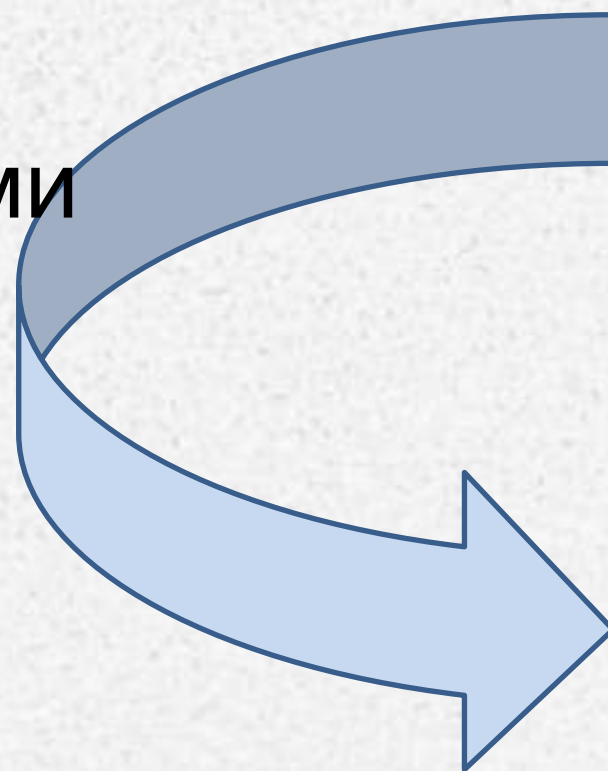
- TxnID = 12325
- ProductID = 507
- Category = Road Bike
- Amount = \$1399.49
- Weight = 20lb.
- FrameType = Composite
- Saddle = Bontrager Race
- **Review = Disappointing for the price. The frame feels heavier than I expected.**
- ReviewSentiment = Negative
- ReviewTerm = Disappointing
- ReviewTerm = Price
- ReviewTerm = Heavier



Обогащение (Enrichment)

Добавление в модель новой информации, которых «не было» в источнике

- Использование дополнительных пакетов работы с текстами, пространственными данными и др.
- Примеры:
 - Извлечение объектов по именам
 - Анализ эмоциональной окраски текстов
 - Совмещение с геопространственными
 - Term extraction
 - Geospatial matching



- TxnID = 12324
- ProductID = 506
- Category = Mountain Bike
- Amount = \$499.99
- Suspension = Fox 32 F-Series
- FrameType = Aluminium
- Saddle = Bontrager SSR
- Mountain Accessories = Fork and shock sag meter
- Mountain Accessories = Water Bottle
- Review = A great bike for off road. Smooth ride over the bumps
- **ReviewSentiment = Positive**
- **ReviewTerm = Great**
- **ReviewTerm = Off Road**
- **ReviewTerm = Smooth**
- **ReviewTerm = Bumps**

- TxnID = 12325
- ProductID = 507
- Category = Road Bike
- Amount = \$1399.49
- Weight = 20lb.
- FrameType = Composite
- Saddle = Bontrager Race
- Review = Disappointing for the price. The frame feels heavier than I expected.
- **ReviewSentiment = Negative**
- **ReviewTerm = Disappointing**
- **ReviewTerm = Price**
- **ReviewTerm = Heavier**



Welcome

Staffing

Utilization

Refine your results:

Availability

More than 4 weeks (1500) Within 2 weeks (816) Now (739) More

Bill Rate

\$0 to \$100 (627) \$100 to \$200 (789) \$200 to \$300 (401) \$300 to \$400 (423) \$400 to \$500 (195) More

Industry Experience

Manufacturing (2038) Finance and Insurance (1567) Information (1338) More

Accounts

Cisco Systems, Inc. (151) Intel Corporation (127) Xerox Corporation (111) More

Level

Consultant (716) Senior Consultant (517) Management Analyst (502) More

International Experience

United States (2376) Netherlands (188) Switzerland (117) More

Practice Vertical

Electronics Software (754) Financial Services (366) Communications (263) More

Country

USA (1466) China (358) Germany (87) More

Software Application Experience

Siebel - BI (862) Peoplesoft - CRM (818) Siebel - CRM (786) More

Performance Rating

1 (461) 2 (560) 3 (853) 4 (363) 5 (198) More

Functional Experience

Enterprise Human Resources (710) Integration IT Strategy (705) Infrastructure Services Management (514) More

Search

☐ Search Within

Geography Search:

miles from

Submit

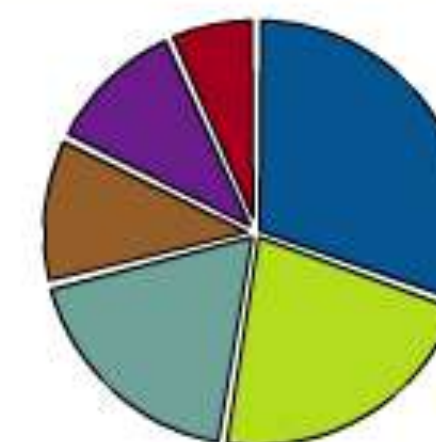
Averages

Metrics:

Billing Rate

Group By:

Level



- Senior Manager 30%
- Manager 22%
- Senior Consultant 19%
- Consultant 11%
- Senior Management Analyst 11%
- Management Analyst 7%

Consultants

2,435

Average Performance Rating

2.70

Avg Billing Rate

\$199.40

Actions

Choose a column set:

Position

Location

Resume

<input type="checkbox"/>	Employee Name	Billing Rate	Level	Performance	Practice Area	Practice Vertical
<input type="checkbox"/>	Chesick, Harry	437	Manager	1	North America	Electronics Software
<input type="checkbox"/>	Fleitas, Cleo	361	Senior Consultant	1	North America	Electronics Software
<input type="checkbox"/>	Gantzler, Harvey	111	Management Analyst	5	North America	Electronics Software
<input type="checkbox"/>	Dibenedict, Ruthe	110	Management Analyst	5	North America	Financial Services
<input type="checkbox"/>	Cote, Lissette	373	Senior Consultant	1	North America	Aerospace & Defense
<input type="checkbox"/>	Mabrey, Ilana	280	Senior Consultant	5	EMEA	Content
<input type="checkbox"/>	Zwahlen, Lee	265	Consultant	3	North America	Electronics Software
<input type="checkbox"/>	Adelmann, Gus	350	Senior Consultant	2	North America	Financial Services
<input type="checkbox"/>	Baldiviez, Becki	148	Senior Management Ar	4	North America	Electronics Software
<input type="checkbox"/>	Suryanarayana, Mia	134	Management Analyst	2	North America	Automotive



Welcome

Staffing

Utilization

Refine your results:

[Clear All Filters](#)

Availability

Now More than 4 weeks (1500)
Within 4 weeks (935) Within 2
weeks (816) More

Bill Rate

\$0 to \$100 (259) \$100 to \$200 (224)
\$200 to \$300 (114) \$300 to \$400
(97) \$400 to \$500 (45) More

Industry Experience

Manufacturing (552) Finance and
Insurance (448) Information (244)
More

Accounts

Regions Financial Corp. (63)
Wachovia Corporation (58) Yahoo!
Inc. (54) More

Level

Consultant (247) Management
Analyst (157) Senior Consultant
(151) More

International Experience

United States (695) Netherlands (76)
Switzerland (37) More

Practice Vertical

Electronics Software (222) Financial
Services (128) Communications (87)
More

Country

Select refinements, or search for a particular value:

☐ Austria (14) X
☐ Belgium (8) X
☐ China (126) X
☐ Denmark (6) X
☐ Germany (25) X
☐ India (168) X
☐ Netherlands (3) X
☐ Spain (31) X
☐ Switzerland (5) X
☐ USA (355) X
Select All Deselect All
Submit Selections

Search

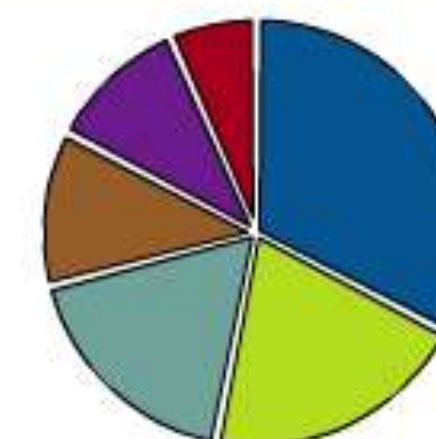
☐ Search Within

miles from

Submit

Group By:

Level



■ Senior Manager 33%
■ Manager 20%
■ Senior Consultant 18%
■ Senior Management Analyst 12%
■ Consultant 11%
■ Management Analyst 7%

Consultants
739

Average Performance Rating
2.90

Avg Billing Rate
\$173.72

Actions

Choose a column set: Position Location Resume

<input type="checkbox"/>	Employee Name	Billing Rate	Level	Performance	Practice Area	Practice Vertical
<input type="checkbox"/>	Gantzler, Harvey	111	Management Analyst	5	North America	Electronics Software
<input type="checkbox"/>	Dibenedict, Ruthe	110	Management Analyst	5	North America	Financial Services
<input type="checkbox"/>	Cote, Lissette	373	Senior Consultant	1	North America	Aerospace & Defense
<input type="checkbox"/>	Mabrey, Ilana	280	Senior Consultant	5	EMEA	Content
<input type="checkbox"/>	Zwahlen, Lee	265	Consultant	3	North America	Electronics Software
<input type="checkbox"/>	Adelmann, Gus	350	Senior Consultant	2	North America	Financial Services
<input type="checkbox"/>	Karapetyan, Charissa	267	Consultant	3	North America	Retail Wholesale
<input type="checkbox"/>	Vormwald, Evan	309	Senior Consultant	4	North America	Financial Services
<input type="checkbox"/>	Ingrahm, Cameron	446	Senior Manager	2	North America	Financial Services
<input type="checkbox"/>	Minges, Wiley	144	Management Analyst	1	North America	Retail Wholesale



Welcome

Staffing

Utilization

Refine your results:

[Clear All Filters](#)

Availability

Now More than 4 weeks (379)
Within 4 weeks (223) Within 2
weeks (177) More

Bill Rate

\$0 to \$100 **\$100 to \$200**
\$200 to \$300 (83) \$300 to \$400 (84)
\$400 to \$500 (34) More

Industry Experience

Manufacturing (133) Finance and
Insurance (101) Information (72)
More

Accounts

Jabil Circuit, Inc. (29) Regions
Financial Corp. (23) Scientific-
Atlanta, Inc. (17) More

State

Virginia (21) Tennessee (19)
California (16) More

Level

Management Analyst (85) Senior
Management Analyst (69) More

International Experience

United States (154) Japan (8)
Netherlands (3) *United States* More

Practice Vertical

Electronics Software (51) Financial
Services (27) Retail Wholesale (22)
More

Country

USA India (168) China (126)
Spain (11) More

Software Application Experience

Siebel - BI (82) Siebel - CRM (48)
Cognos (43) More

Performance Rating

1 (28) 2 (31) 3 (44) 4 (25) 5 (26) More

Functional Experience

Strategy/Business Processes (66)
Enterprise Human Resources (46)
CRM Architecture/Integration (41)
More

Search

☐ Search Within

Geography Search:

500

miles from Detroit

Submit

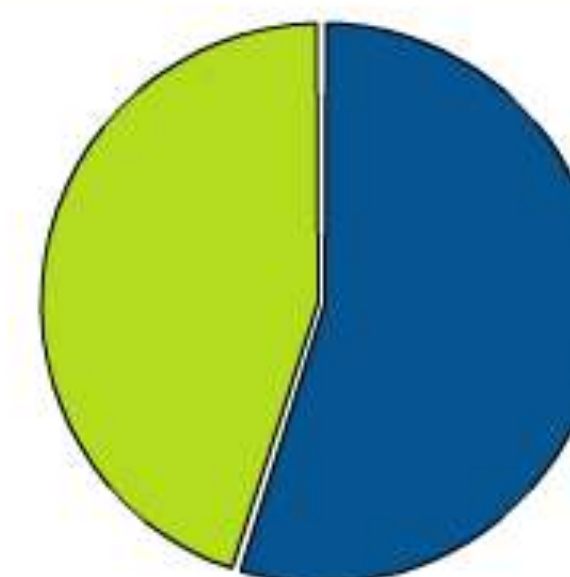
Averages

Metrics:

Billing Rate

Group By:

Level



■ Senior Management Analyst 55%

■ Management Analyst 45%

Consultants

154

Average Performance Rating

2.94

Avg Billing Rate

\$140.48

Actions

Choose a column set: Position Location Resume

	Employee Name	Billing Rate	Country	Region	State	City
<input type="checkbox"/>	Gantzler, Harvey	111	USA	Mid-Atlantic	Maryland	Lexington Park
<input type="checkbox"/>	Dibenedict, Ruthe	110	USA	New England	New Hampshire	Nashua
<input type="checkbox"/>	Minges, Wiley	144	USA	West	California	Mountain View
<input type="checkbox"/>	Lorent, Sebrina	126	USA	Mid-Atlantic	Maryland	Rockville
<input type="checkbox"/>	Crete, Shawna	164	USA	Mid-Atlantic	Pennsylvania	Radnor
<input type="checkbox"/>	Momeni, Harley	126	USA	South	North Carolina	Raleigh
<input type="checkbox"/>	Feist, Casimira	134	USA	South	Tennessee	Dallas
<input type="checkbox"/>	Bough, Levi	157	USA	Mid-Atlantic	Maryland	Rockville
<input type="checkbox"/>	Toa, Patience	165	USA	West	California	Sacramento
<input type="checkbox"/>	Broder, Vanessa	156	USA	West	California	San Diego



Welcome

Staffing

Utilization

Refine your results:

Availability

More than 4 weeks (1500) Within 2 weeks (816) Now (739) More

Bill Rate

\$0 to \$100 (627) \$100 to \$200 (789) \$200 to \$300 (401) \$300 to \$400 (423) \$400 to \$500 (195) More

Industry Experience

Manufacturing (2038) Finance and Insurance (1567) Information (1338) More

Accounts

Cisco Systems, Inc. (151) Intel Corporation (127) Xerox Corporation (111) More

Level

Consultant (716) Senior Consultant (517) Management Analyst (502) More

International Experience

United States (2376) Netherlands (188) Switzerland (117) More

Practice Vertical

Electronics Software (754) Financial Services (366) Communications (263) More

Country

USA (1466) China (358) Germany (87) More

Software Application Experience

Siebel - BI (862) Peoplesoft - CRM (818) Siebel - CRM (786) More

Performance Rating

1 (461) 2 (560) 3 (853) 4 (363) 5 (198) More

Functional Experience

Enterprise Human Resources (710) Integration IT Strategy (705) Infrastructure Services Management (514) More

Search

☐ Search Within

Geography Search:

miles from

Submit

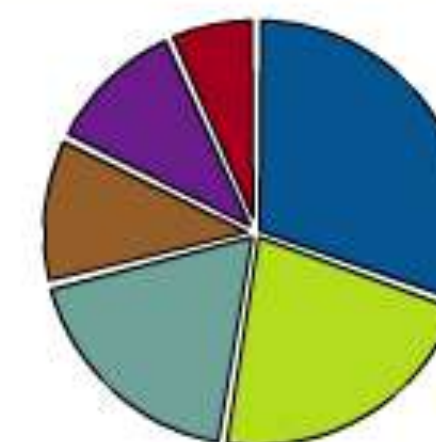
Averages

Metrics:

Billing Rate

Group By:

Level



- Senior Manager 30%
- Manager 22%
- Senior Consultant 19%
- Consultant 11%
- Senior Management Analyst 11%
- Management Analyst 7%

Consultants

2,435

Average Performance Rating

2.70

Avg Billing Rate

\$199.40

Actions

Choose a column set: Position Location Resume

<input type="checkbox"/>	Employee Name	Billing Rate	Level	Performance	Practice Area	Practice Vertical
<input type="checkbox"/>	Chesick, Harry	437	Manager	1	North America	Electronics Software
<input type="checkbox"/>	Fleitas, Cleo	361	Senior Consultant	1	North America	Electronics Software
<input type="checkbox"/>	Gantzler, Harvey	111	Management Analyst	5	North America	Electronics Software
<input type="checkbox"/>	Dibenedict, Ruthe	110	Management Analyst	5	North America	Financial Services
<input type="checkbox"/>	Cote, Lissette	373	Senior Consultant	1	North America	Aerospace & Defense
<input type="checkbox"/>	Mabrey, Ilana	280	Senior Consultant	5	EMEA	Content
<input type="checkbox"/>	Zwahlen, Lee	265	Consultant	3	North America	Electronics Software
<input type="checkbox"/>	Adelmann, Gus	350	Senior Consultant	2	North America	Financial Services
<input type="checkbox"/>	Baldiviez, Becki	148	Senior Management Ar	4	North America	Electronics Software
<input type="checkbox"/>	Suryanarayana, Mia	134	Management Analyst	2	North America	Automotive



Welcome

Staffing

Utilization

Refine your results:

[Clear All Filters](#)

Range Filters:

Office Geocode : 500.0 miles from (42,-83)

Availability

Now More than 4 weeks (182)
Within 4 weeks (88) Within 2 weeks
(70) More

Bill Rate

\$0 to \$100 **\$100 to \$200**
\$200 to \$300 (33) \$300 to \$400 (38)
\$400 to \$500 (12) More

Industry Experience

Manufacturing (57) Finance and
Insurance (46) Information (29)
More

Accounts

PNC Financial Services (12) Jabil
Circuit, Inc. (11) Host Marriott
Corporation (10) More

State

Virginia (16) Maryland (15) Ohio (9)
More

Level

Management Analyst (31) Senior
Management Analyst (28) More

International Experience

United States (59) Japan (6)
Netherlands (2) *United States* More

Practice Vertical

Electronics Software (18) Financial
Services (12) Retail Wholesale (8)
More

Software Application Experience

Siebel - BI (31) Siebel - CRM (22)
Microsoft (17) More

Performance Rating

1 (12) 2 (15) 3 (16) 4 (7) 5 (9) More

Functional Experience

Strategy/Business Processes (27)
Integration Enterprise Program
Management (17) Infrastructure
Project Lifecycle (15) More

French

Search

☒ Search Within

Geography Search:

miles from

Submit

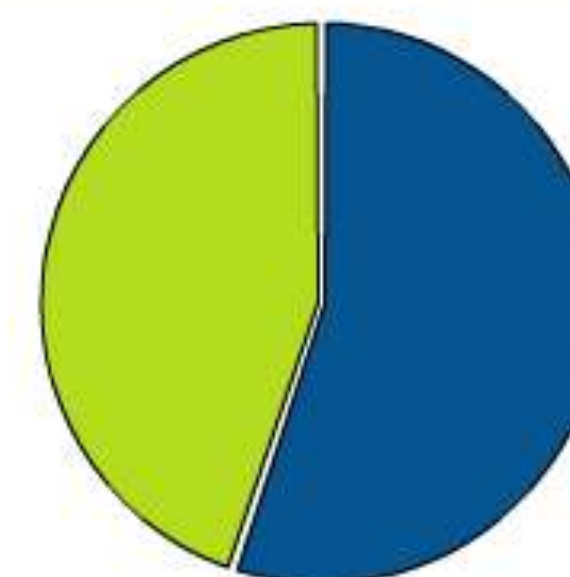
Averages

Metrics:

Billing Rate

Group By:

Level



■ Senior Management Analyst 55%

■ Management Analyst 45%

Consultants

59

Average Performance Rating

2.76

Avg Billing Rate

\$142.76

Actions

Choose a column set: Position Location Resume

	Employee Name	Billing Rate	Country	Region	State	City
<input type="checkbox"/>	Gantzler, Harvey	111	USA	Mid-Atlantic	Maryland	Lexington Park
<input type="checkbox"/>	Lorent, Sebrina	126	USA	Mid-Atlantic	Maryland	Rockville
<input type="checkbox"/>	Crete, Shawna	164	USA	Mid-Atlantic	Pennsylvania	Radnor
<input type="checkbox"/>	Bough, Levi	157	USA	Mid-Atlantic	Maryland	Rockville
<input type="checkbox"/>	Klepchick, Aaron	126	USA	Mid-Atlantic	Pennsylvania	Harrisburg
<input type="checkbox"/>	Clardy, Angel	157	USA	Mid-Atlantic	Pennsylvania	Radnor
<input type="checkbox"/>	Huizar, Ervin	118	USA	South	Virginia	Richmond
<input type="checkbox"/>	Thronson, Kieth	111	USA	South	Tennessee	Nashville
<input type="checkbox"/>	Ravelec, Alanna	143	USA	Midwest	Ohio	Cleveland

Refine your results:

[Clear All Filters](#)[Search for 'French'](#)

Range Filters:

[Office Geocode : 500.0 miles from \(42,-83\)](#)

Availability

Now More than 4 weeks (28)
Within 2 weeks (15) Within 4 weeks (15) More

Bill Rate

\$0 to \$100 **\$100 to \$200**
\$200 to \$300 (3) \$300 to \$400 (4)
\$400 to \$500 (3) More

Industry Experience

Manufacturing (13) Finance and Insurance (13) Information (8)
Finance and Insurance *Manufacturing*
More

Accounts

RadioShack Corporation (4) PNC
Financial Services (3)
Amazon.com, Inc. (3) More

State

Maryland (4) Virginia (3) Illinois (2)
More

Level

Management Analyst (7) Senior
Management Analyst (6) More

International Experience

United States (13) Netherlands (1)
Japan (1) *United States* More

Practice Vertical

Financial Services (4) Electronics
Software (3) Communications (2)
More

Software Application Experience

Peoplesoft - CRM (8) Vignette (6)
Siebel - BI (5) More

Performance Rating

1 (4) 2 (2) 3 (2) 4 (2) 5 (3) More

Functional Experience

Strategy/Business Processes (7)
Integration Quality Management
(5) Infrastructure (4) More

Search

☐ Search Within

Geography Search:

miles from

Submit

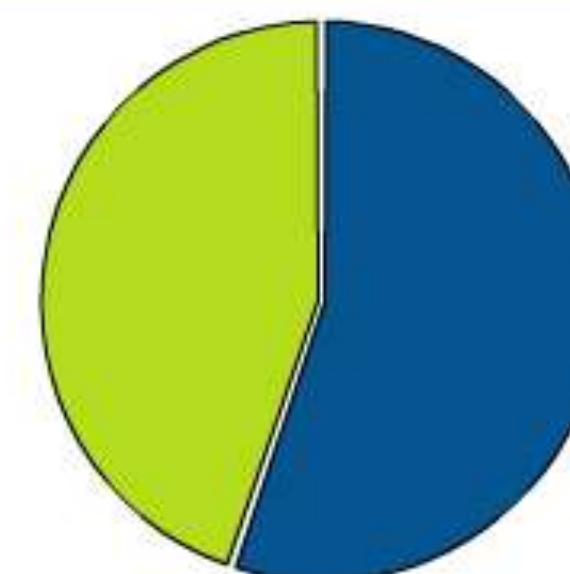
Averages

Metrics:

Billing Rate

Group By:

Level



■ Senior Management Analyst 55%
■ Management Analyst 45%

Consultants
13

Average Performance Rating
2.85

Avg Billing Rate
\$141.92

Actions

Choose a column set: Position Location Resume

<input type="checkbox"/>	Employee Name	Billing Rate	Resume Link	Resume Text Snippet
<input type="checkbox"/>	Basu, Londa	111	Londa_Basu_482.doc	...for mapping and documenting a global printing c
<input type="checkbox"/>	Thronson, Kieth	111	Kieth_Thronson_235.doc	...Committee, European Logistics Association (ELA
<input type="checkbox"/>	Bough, Levi	157	Levi_Bough_555.doc	...and spoken English, Spanish, Portuguese, and l
<input type="checkbox"/>	Dierking, Sanjuanita	140	Sanjuanita_Dierking_1517.doc	...Low/Medium Window ...and spoken English, Spanish, Portuguese, and Persian, as well as having working knowledge of French. She has excellent problem solving and analytical skills and the ability to work independently and...
<input type="checkbox"/>	Clavin, Jarrett	134	Jarrett_Clavin_1.doc	...of Calgary. Minor: Ma
<input type="checkbox"/>	Scarritt, Jade	119	Jade_Scarritt_617.doc	...Low/Medium Windows 95 High Windows NT Med
<input type="checkbox"/>	Polizio, Hoyt	144	Hoyt_Polizio_460.doc	...Team Leadership, Research and Analysis skills.
<input type="checkbox"/>	Darty, Miyoko	144	Miyoko_Darty_373.doc	...Professional qualifications Associate of the Roy
<input type="checkbox"/>	Hielscher, Liberty	173	Liberty_Hielscher_1592.doc	...holds a Bachelor of Commerce in Finance from t
<input type="checkbox"/>	Kosh, Xiao	148	Xiao_Kosh_1119.doc	...directing sales & marketing, and collections. LA

Полная аналитическая платформа Oracle



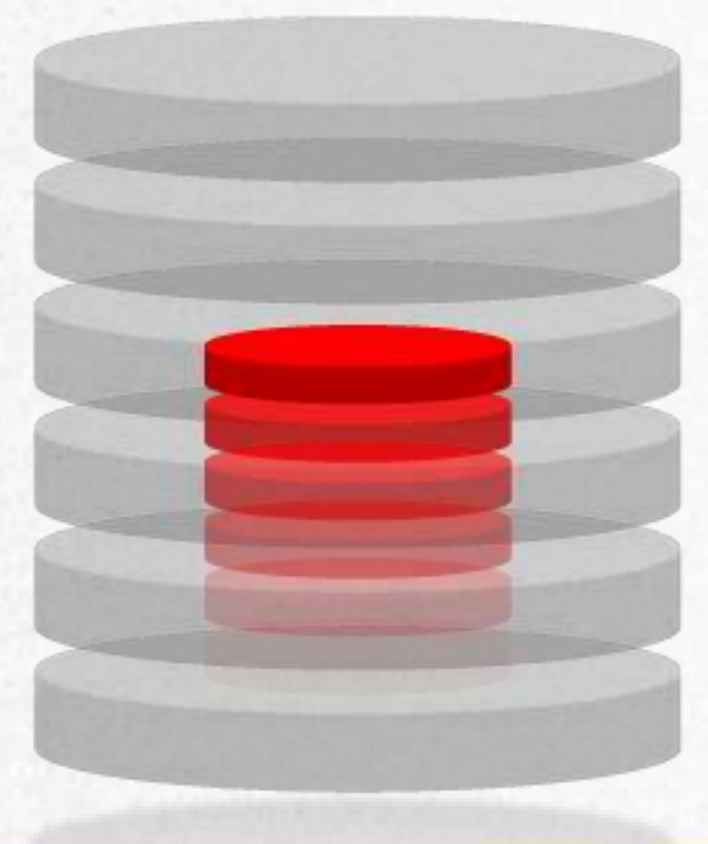
План

- Платформа Oracle для Больших Данных
- Аналитическая in-memory машина Oracle Exalytics
- От бизнес-анализа к исследованию данных – Oracle Endeca Information Discovery
- **Встроенная аналитика для статистических исследований – Oracle R Enterprise**

Oracle Advanced Analytics

Статистические исследования и data mining

Новая опция для СУБД Oracle Database 11g, объединяет среду статистических исследований ***Oracle R Enterprise*** и ***Oracle Data Mining***

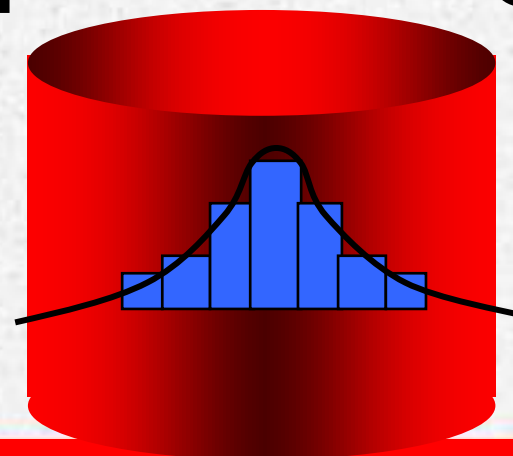


Статистика & Data Mining в Oracle

Встроенная в базу данных аналитика

Аналитические и статистические функции

- Функции ранжирования
- Агрегирование
- Сравнение с предыдущими периодами
- Линейная регрессия
- Корреляции
- Базовая статистика
- Проверка гипотез
- Подбор распределений



Oracle Data Mining

- Встроенные в базу данных процедуры автоматического выявления закономерностей в больших массивах данных
- API для разработки приложений, встраивания data mining в существующие приложения и системы
- Инфраструктура вместо готовой инструментальной среды
- Oracle Data Miner



ORACLE®

Новое решение Oracle R Enterprise

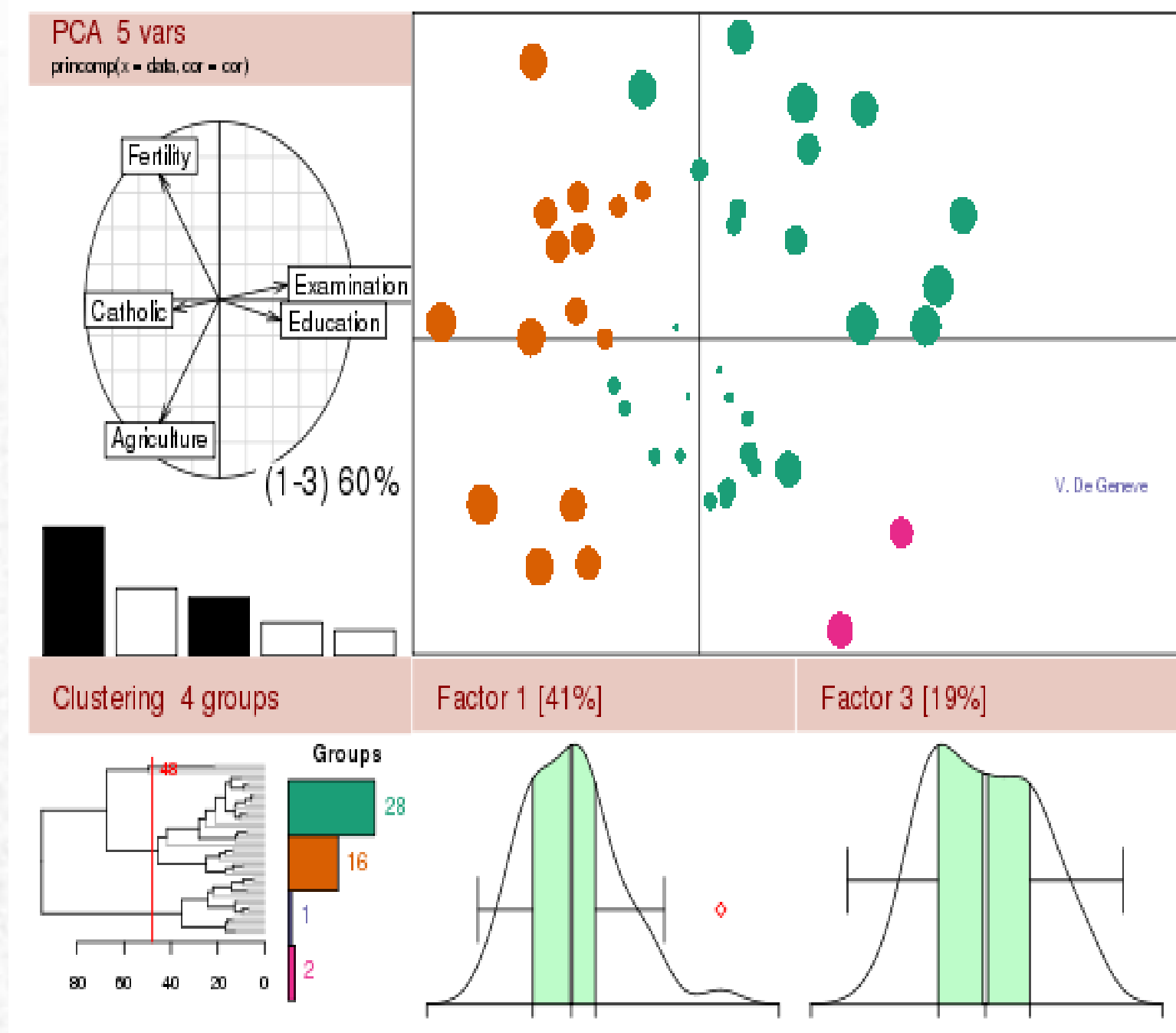


ORACLE® **R** Enterprise
Open Source



ORACLE®

Проект R для статистических вычислений



- Язык для статистических исследований и работы с графикой (Росс Айхэк, Роберт Джентельмен, Оклендский ун-т, 1997)
- Open source проект, R Foundation
- Широкий спектр различных функций (временные ряды, прогнозирование, классификация, кластеризация и др)
- Важное отличительное преимущество – простые средства построения самых сложных графиков и диаграмм
- Возможность расширения, технология разработки дополнительных пакетов участниками проекта

Open Source

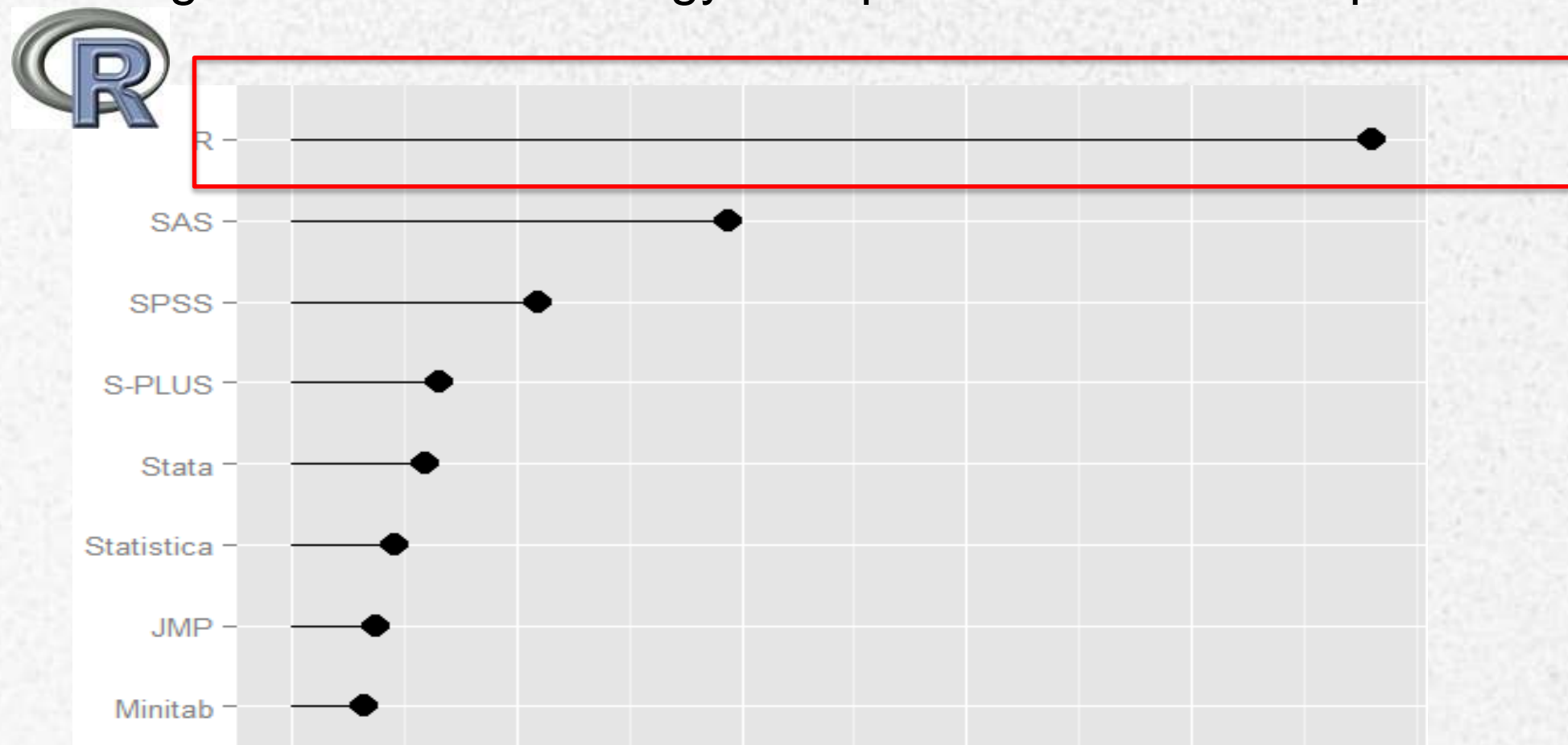


Частично благодаря появлению концепции *Big Data*, бизнес-анализ(BI) остается быстро растущим рынком Одновременно с ростом рынка BI постоянно увеличиваются инвестиции в предиктивную аналитику; **R является не только хорошим готовым инструментом, но и идеальной средой для исследований в области углубленной аналитики.** R ориентирован на расширения и интегрируется с инструментами бизнес-анализа , обогащая отчеты глубокой аналитикой.

“Hype Cycle for Analytic Applications, 2011, 30 August 2011

Gartner

<http://www.gartner.com/technology/core/products/research/topics/businessIntelligence.jsp>



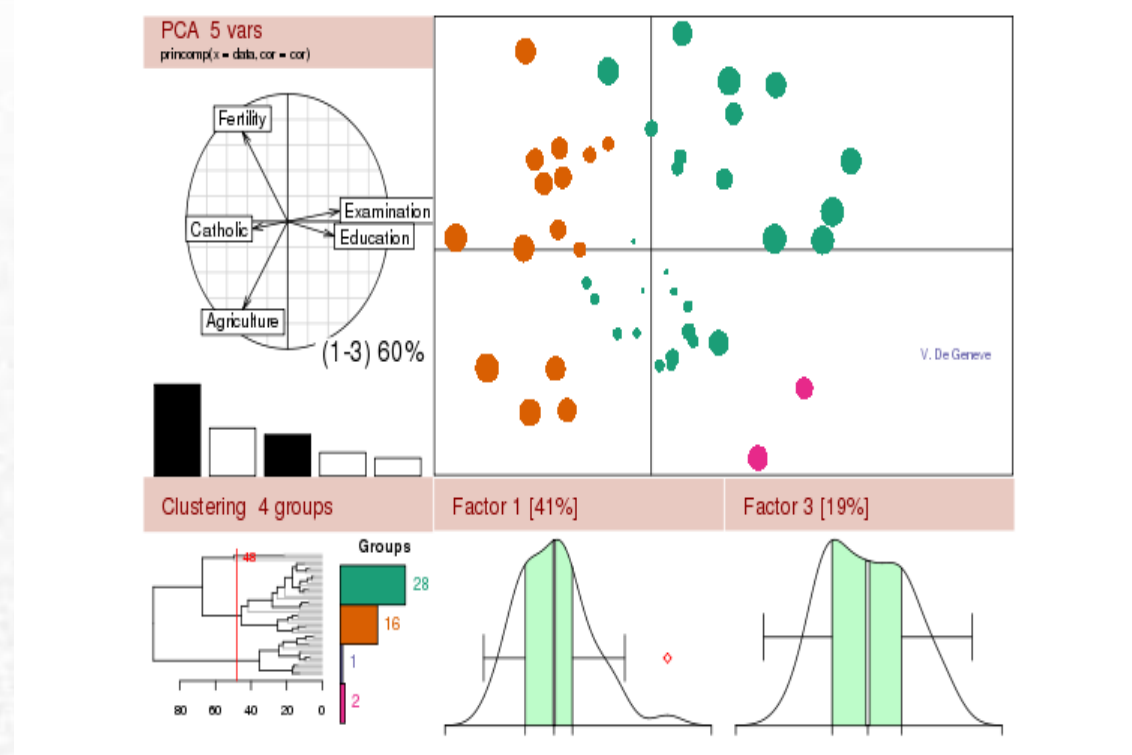
Кол-во f web site линков, которые указывают на основной сайт инструментальной среды March 19, 2011.

<http://www.r4stats.com/popularity>

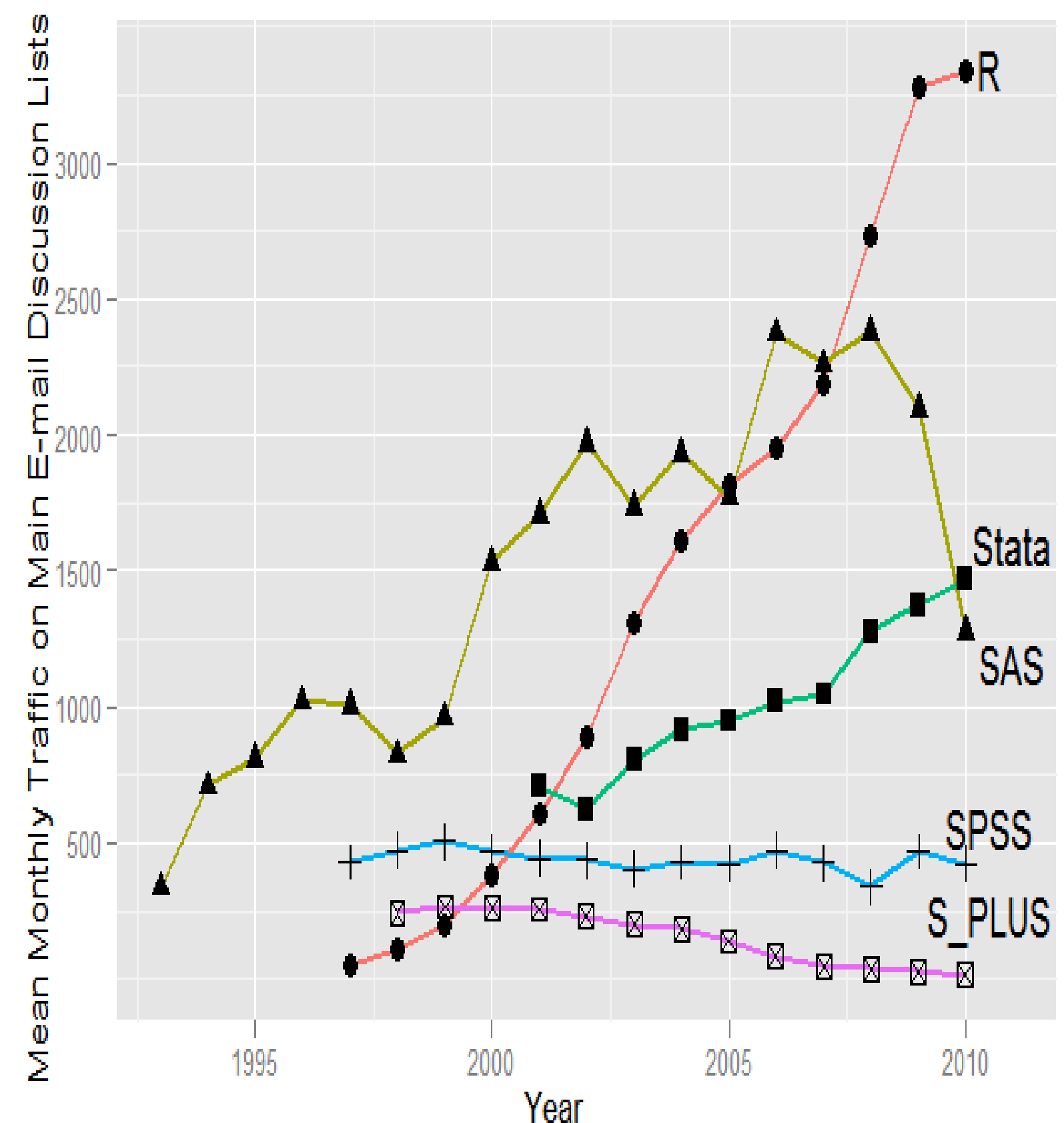
ORACLE

Растущая популярность проекта R

The R Project for Statistical Computing



- Благодаря быстрому развитию и большому интересу со стороны специалистов R заслужил репутацию нового стандарта в области статистического программного обеспечения

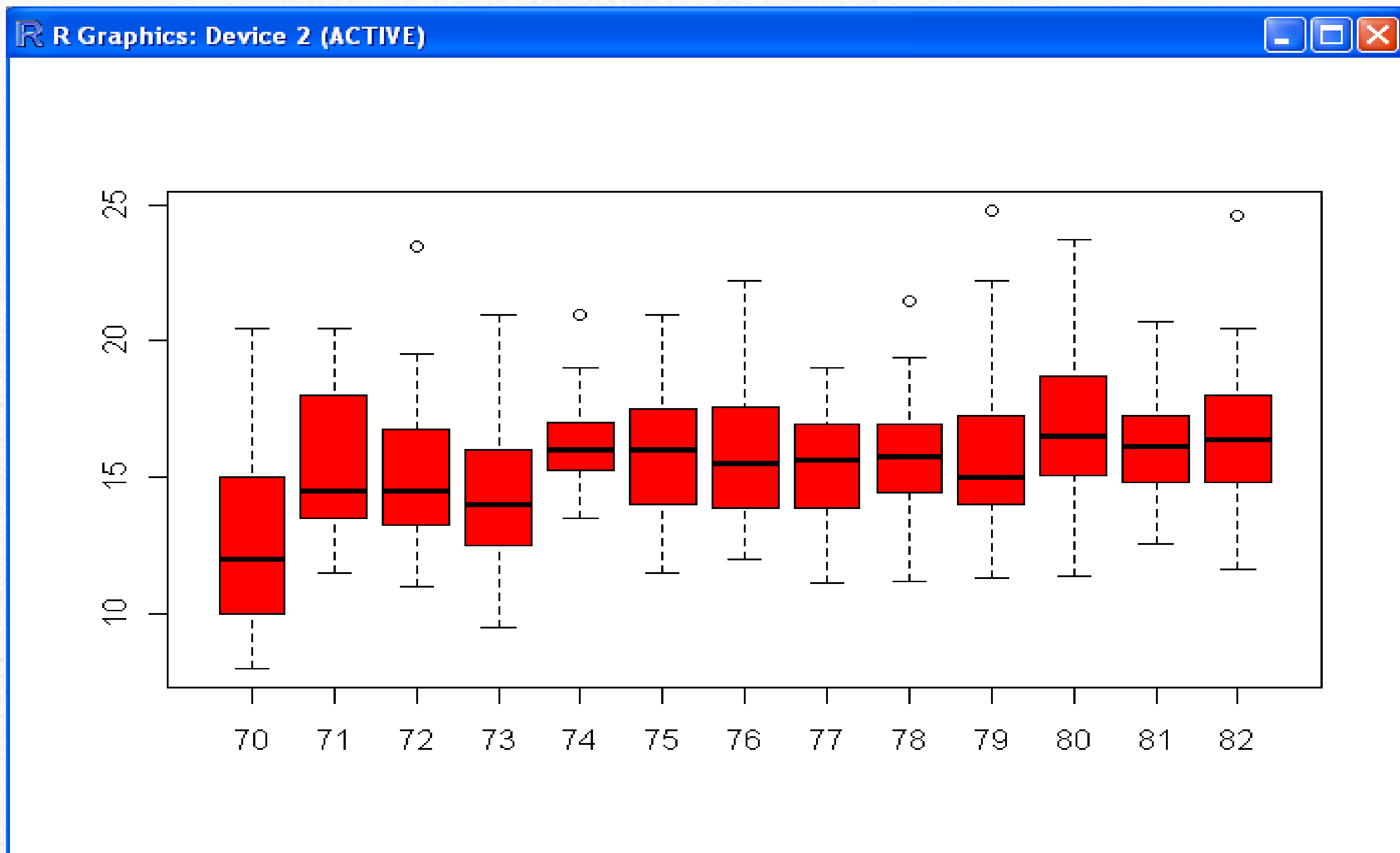


Трудно оценить точно, сколько людей используют R, по оценкам специалистов примерно 250,000 людей работают с этой средой регулярно

“Data Analysts Captivated by R’s Power”, New York Times, Jan 6, 2009

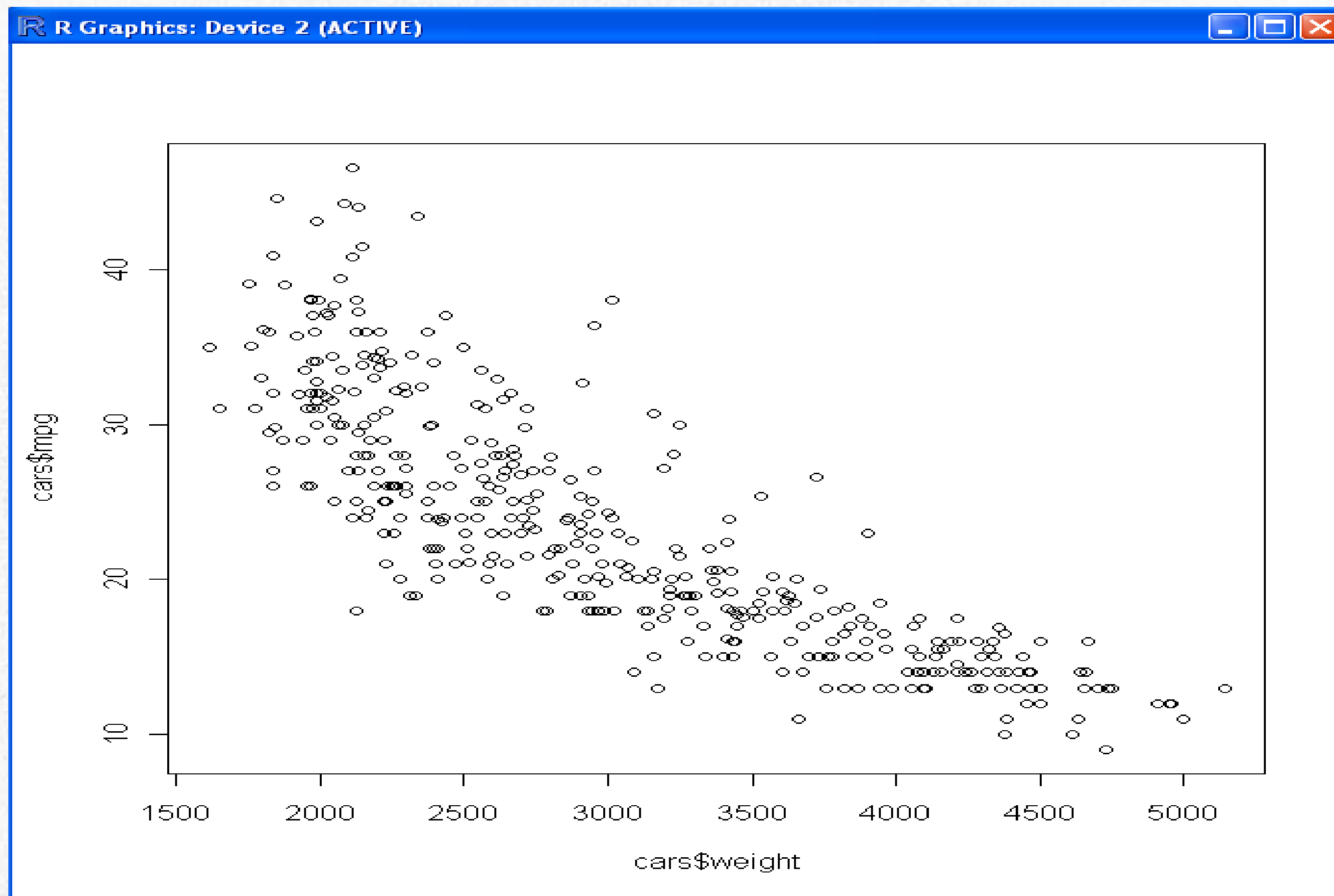
R Graphics

```
R> boxplot(split(cars$acceleration,  
cars$model.year), col = "red")
```



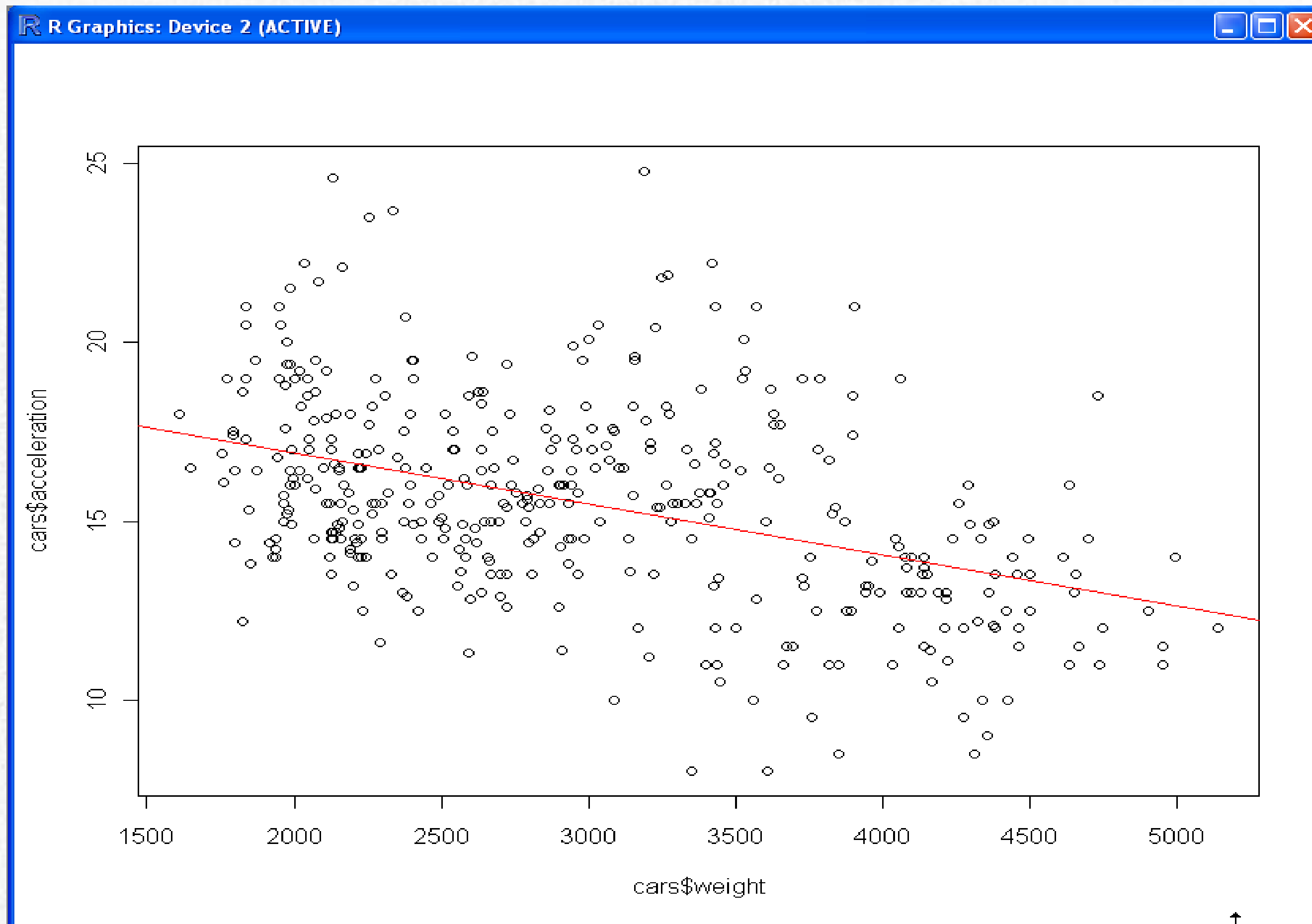
R Graphics

```
R> plot(cars$weight, cars$mpg)
```

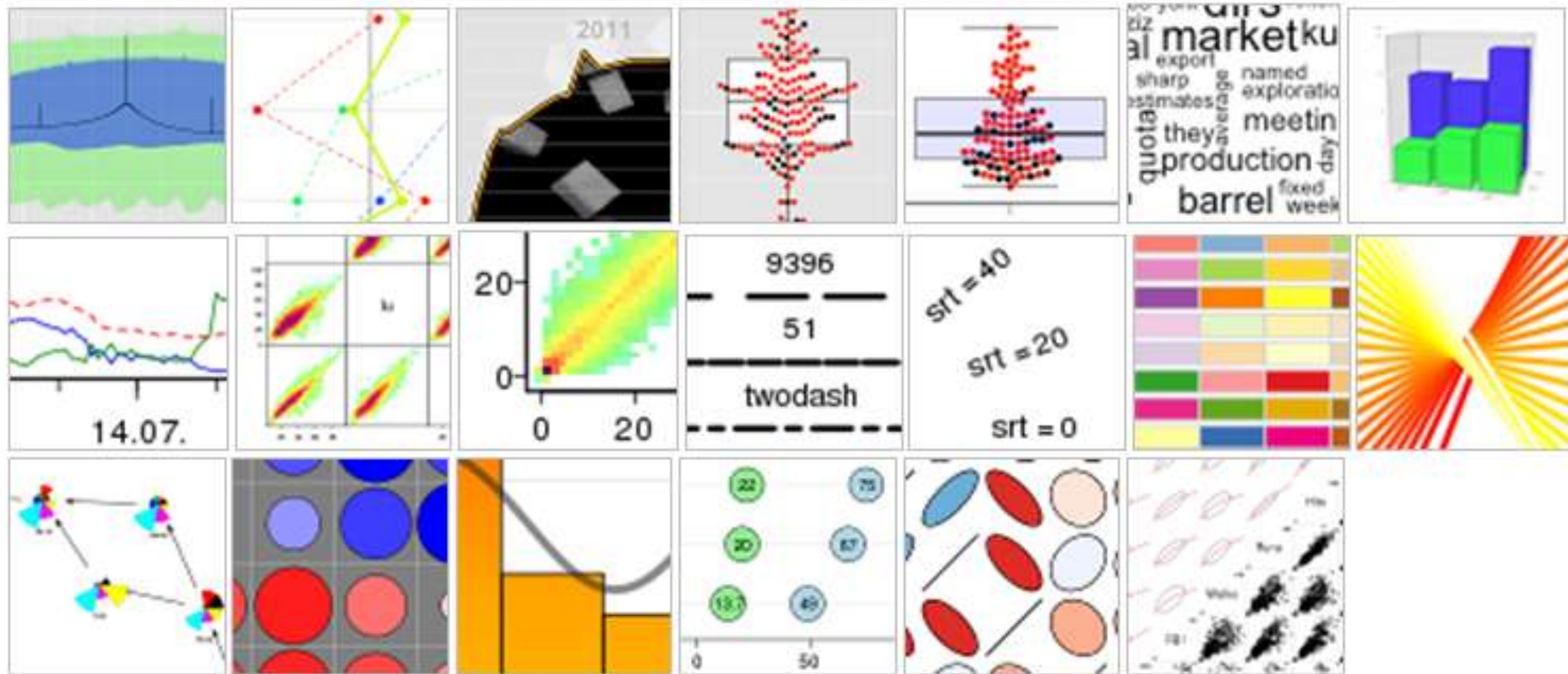


R Graphics

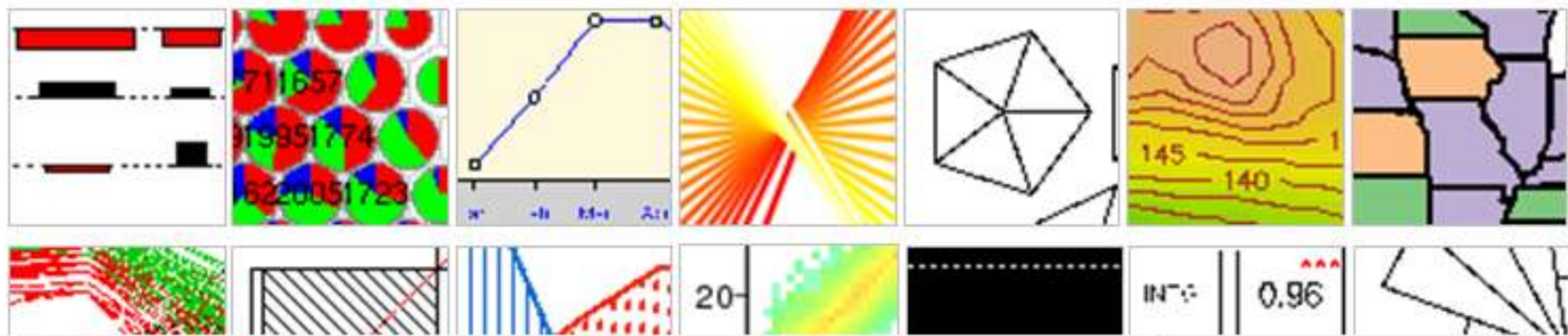
```
R> abline(coef(lm(acceleration ~ weight, cars)), col = "red")
```



Графики и диаграммы в R



» Random entries



Линейное моделирование

```
R Console

R> anova(lm.D9 <- lm(weight ~ group))
Analysis of Variance Table

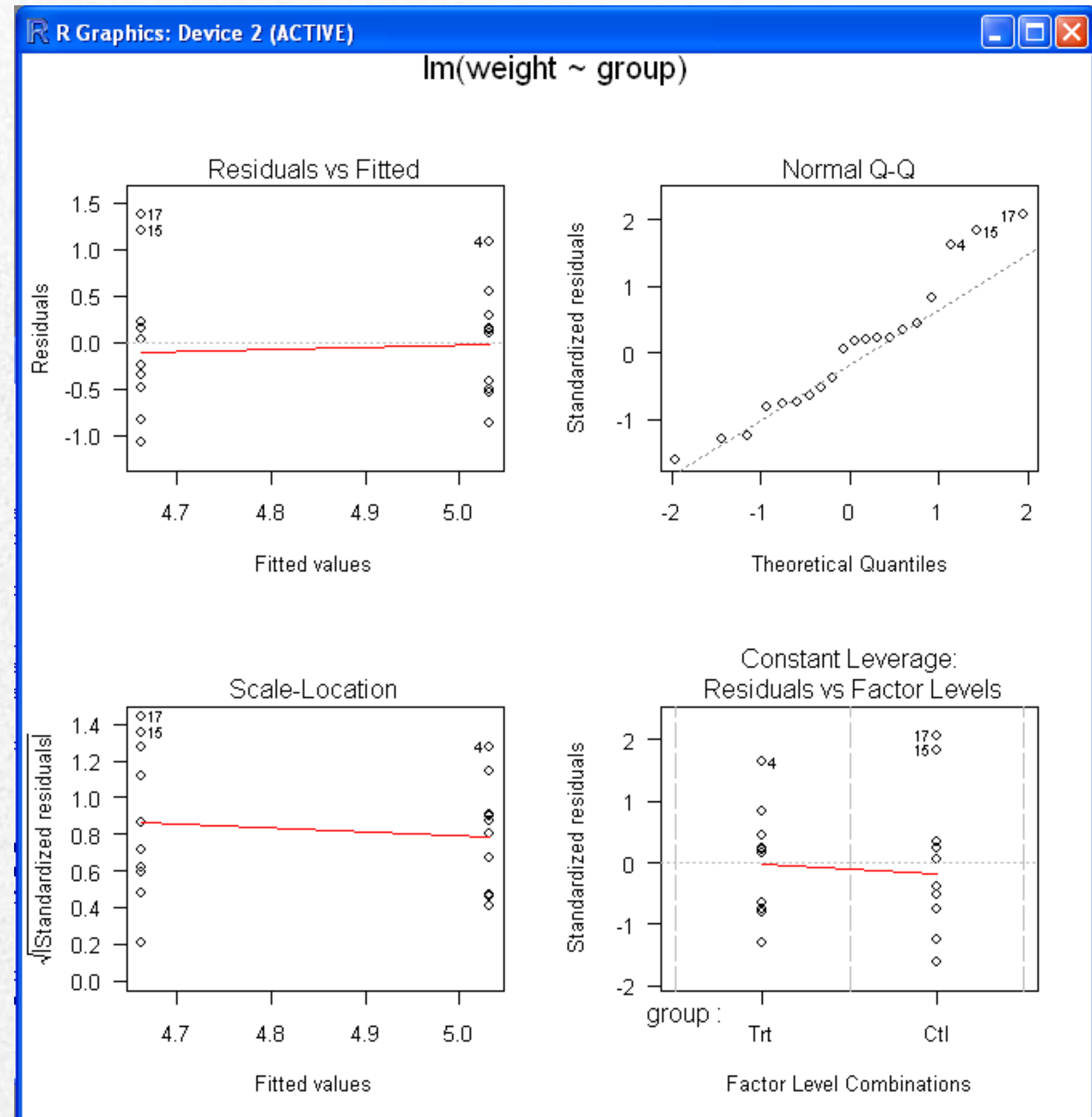
Response: weight
          Df Sum Sq Mean Sq F value Pr(>F)
group       1  0.6882  0.68820   1.4191  0.249
Residuals  18  8.7293  0.48496
R> summary(lm.D90 <- lm(weight ~ group - 1))# omitting intercept

Call:
lm(formula = weight ~ group - 1)

Residuals:
    Min       1Q   Median       3Q      Max
-1.0710 -0.4938  0.0685  0.2462  1.3690

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
groupCtl     5.0320     0.2202   22.85 9.55e-15 ***
groupTrt     4.6610     0.2202   21.16 3.62e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6964 on 18 degrees of freedom
Multiple R-squared:  0.9818,    Adjusted R-squared:  0.9798
F-statistic: 485.1 on 2 and 18 DF,  p-value: < 2.2e-16
```



ARIMA --прогнозирование

```
year200801 <- ONTIME_S[(ONTIME_S$YEAR==2008)& (ONTIME_S$MONTH==1),]
```

```
y <- ore.pull(year200801)
```

```
gc()
```

```
delays <- tapply(y$ARRDELAY, y$DAYOFMONTH, mean, n
```

```
delays <- ts(delays, start=1, end=31, frequency=1)
```

```
# Create a Kalman filter with the first 5 delays and predict the
```

```
preds <- c()
```

```
ses <- c()
```

```
# 1 step predictions
```

```
for (i in 5:length(delays))
```

```
{
```

```
  fit <- arima(delays[1:i], c(1,2,1))
```

```
  # predict 1 step into the future.
```

```
  pred <- predict(fit)
```

```
  preds <- c(preds, pred$pred)
```

```
  ses <- c(ses, pred$se)
```

```
}
```

```
plot(5:length(delays), preds, type='l', col='green',
```

```
  ylim=range(c(preds+2*ses, preds-2*ses)), xlab="Day of month",
```

```
  ylab="Predicted average delay (in minutes)",
```

```
  main="Average delays by day for January 2008")
```

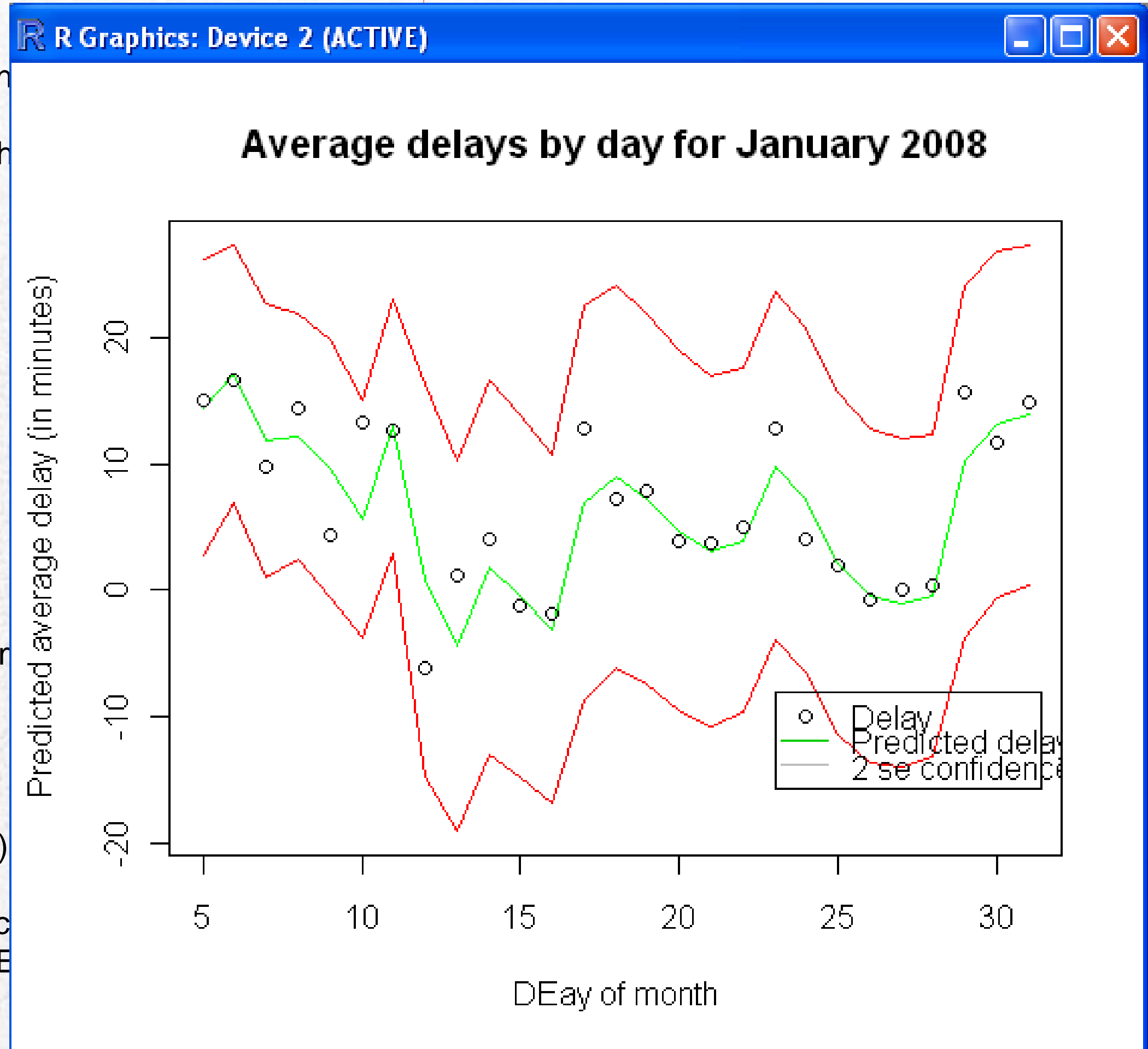
```
lines(5:length(delays), preds+2*ses, col='red')
```

```
lines(5:length(delays), preds-2*ses, col='red')
```

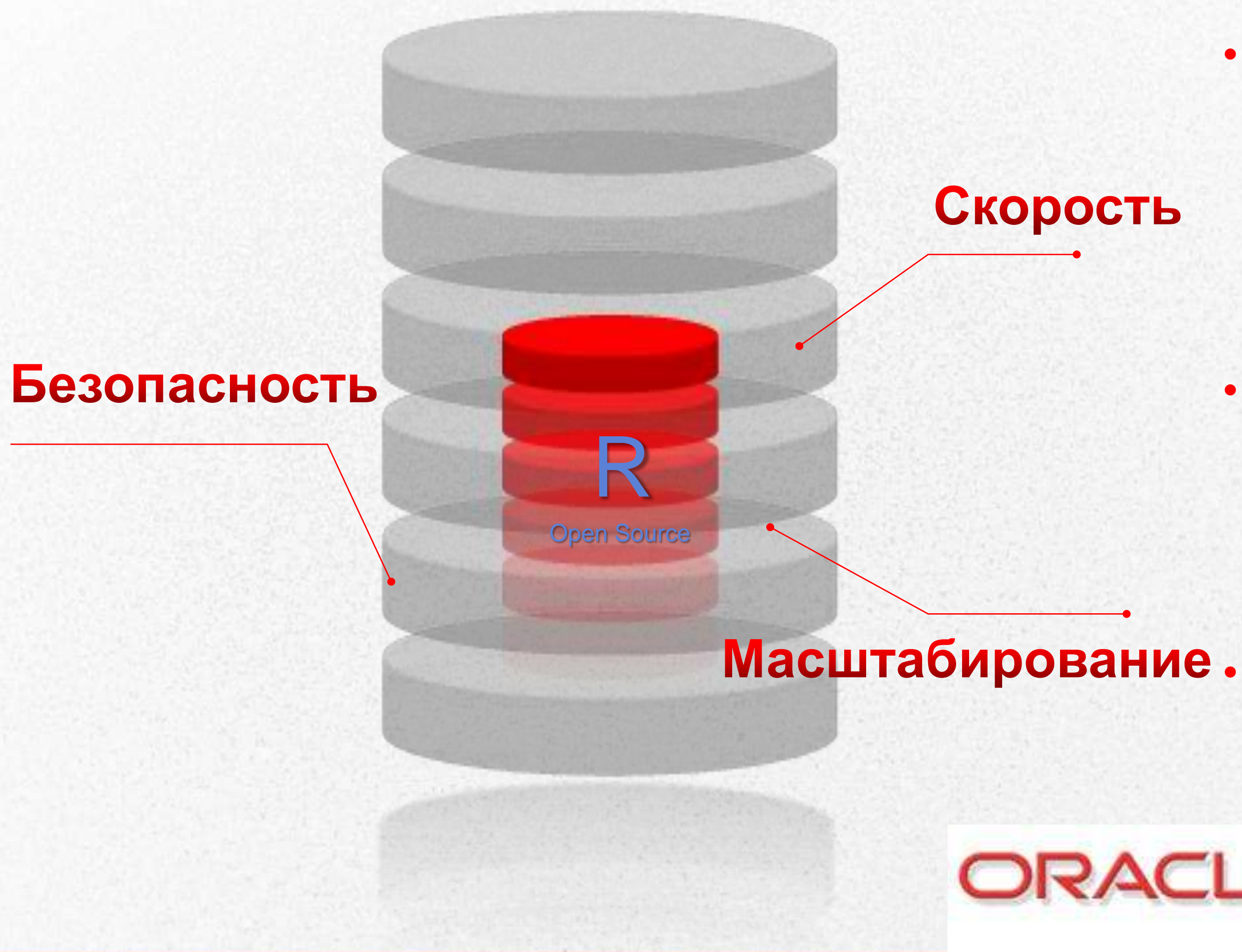
```
points(5:length(delays), as.vector(delays[5:length(delays)]))
```

```
legend( 23, -8, c("Delay", "Predicted delay", "2 se confidence
```

```
  col=c(1, 3, 8), lty=c(0, 1, 1), pch=c(1, -1, -1), merge=TRUE
```



Oracle R Enterprise



- R «встраивается» в Oracle database
- Данные сохраняются и статистические вычисления выполняются в базе данных
- 100% совместимость с R интерфейсом и клиентскими приложениями
- Дополняет Oracle Data Mining

ORACLE® **R** Enterprise
Open Source



Спасибо за внимание!

