



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 аудио 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-параллельна 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»



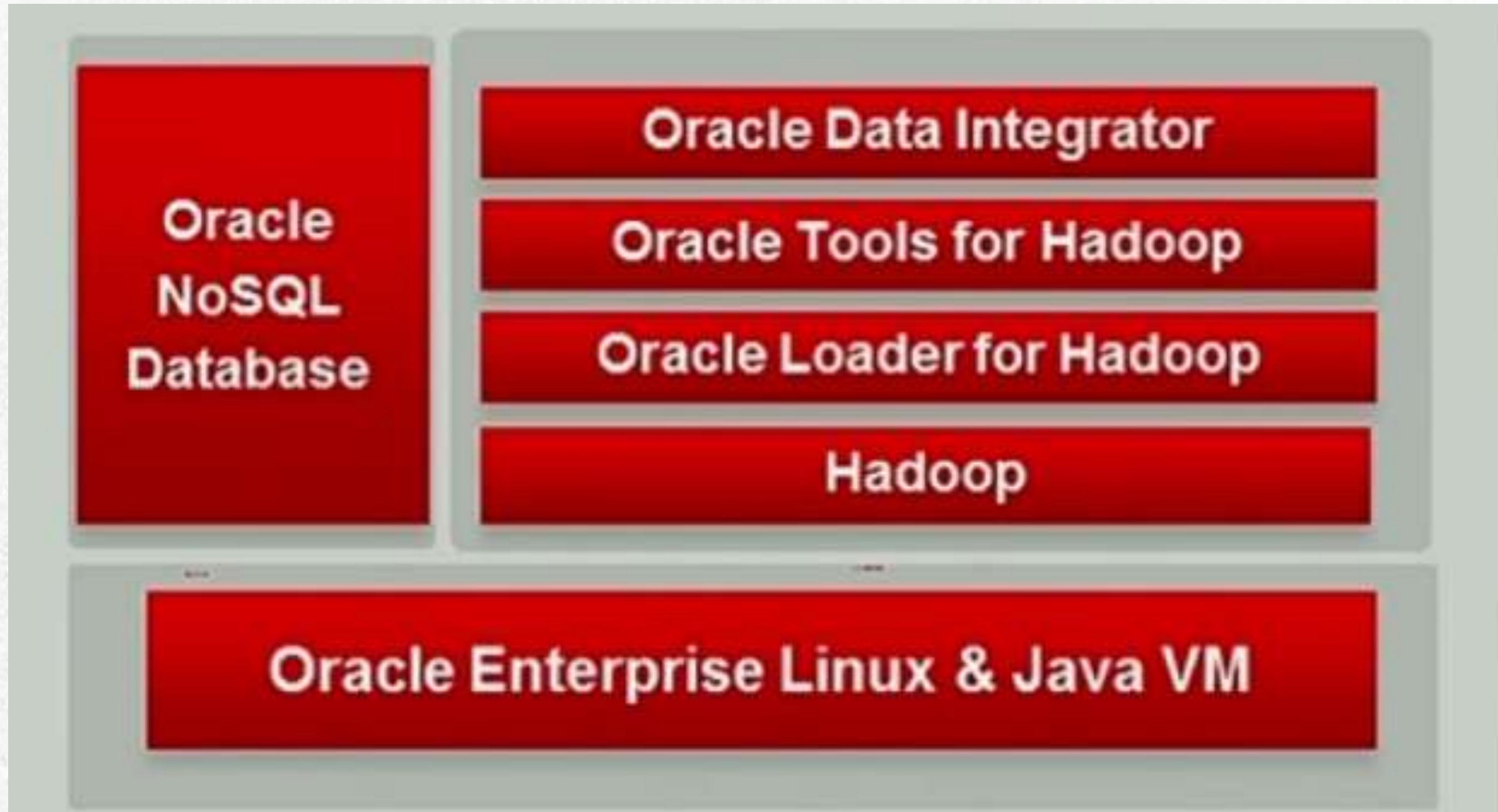
vvv
Velocity, Volume, Veracity
(скорость, объем,
многообразие)

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»

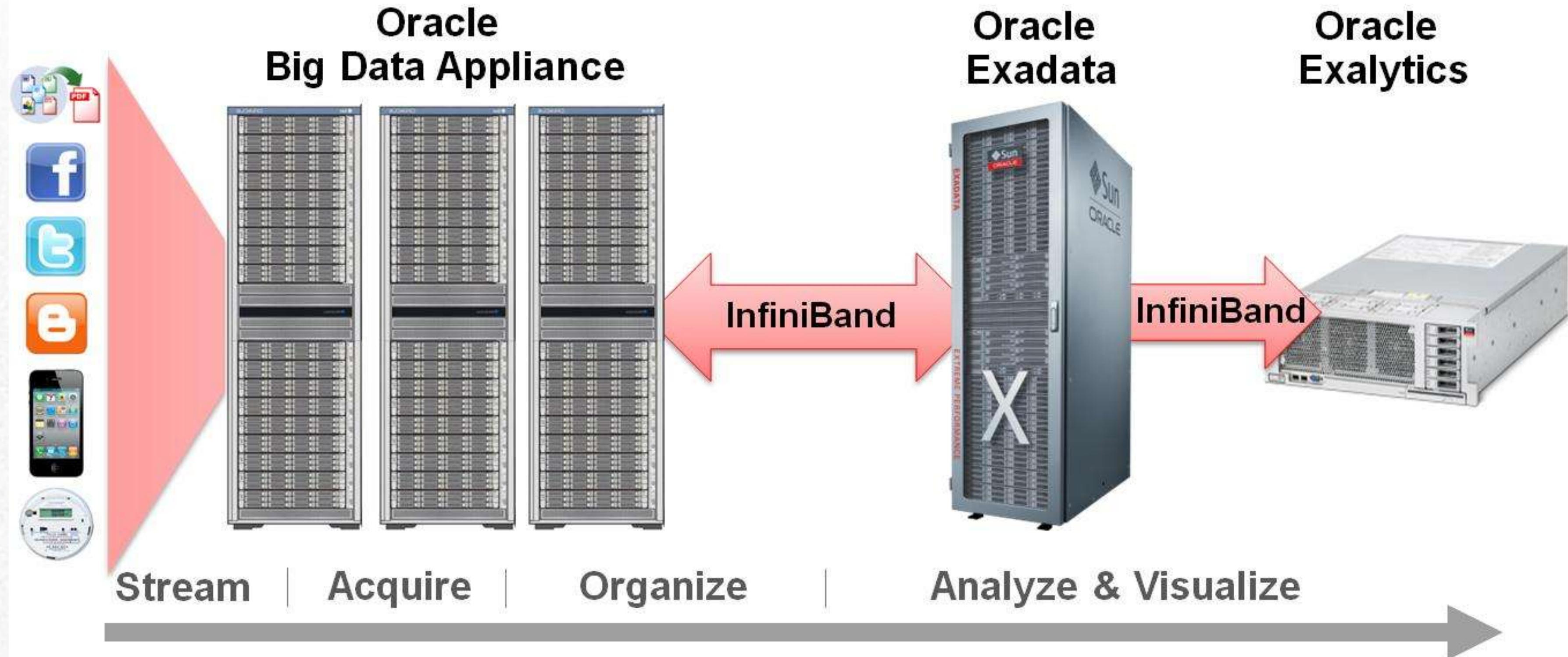


Аппаратное обеспечение

- 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



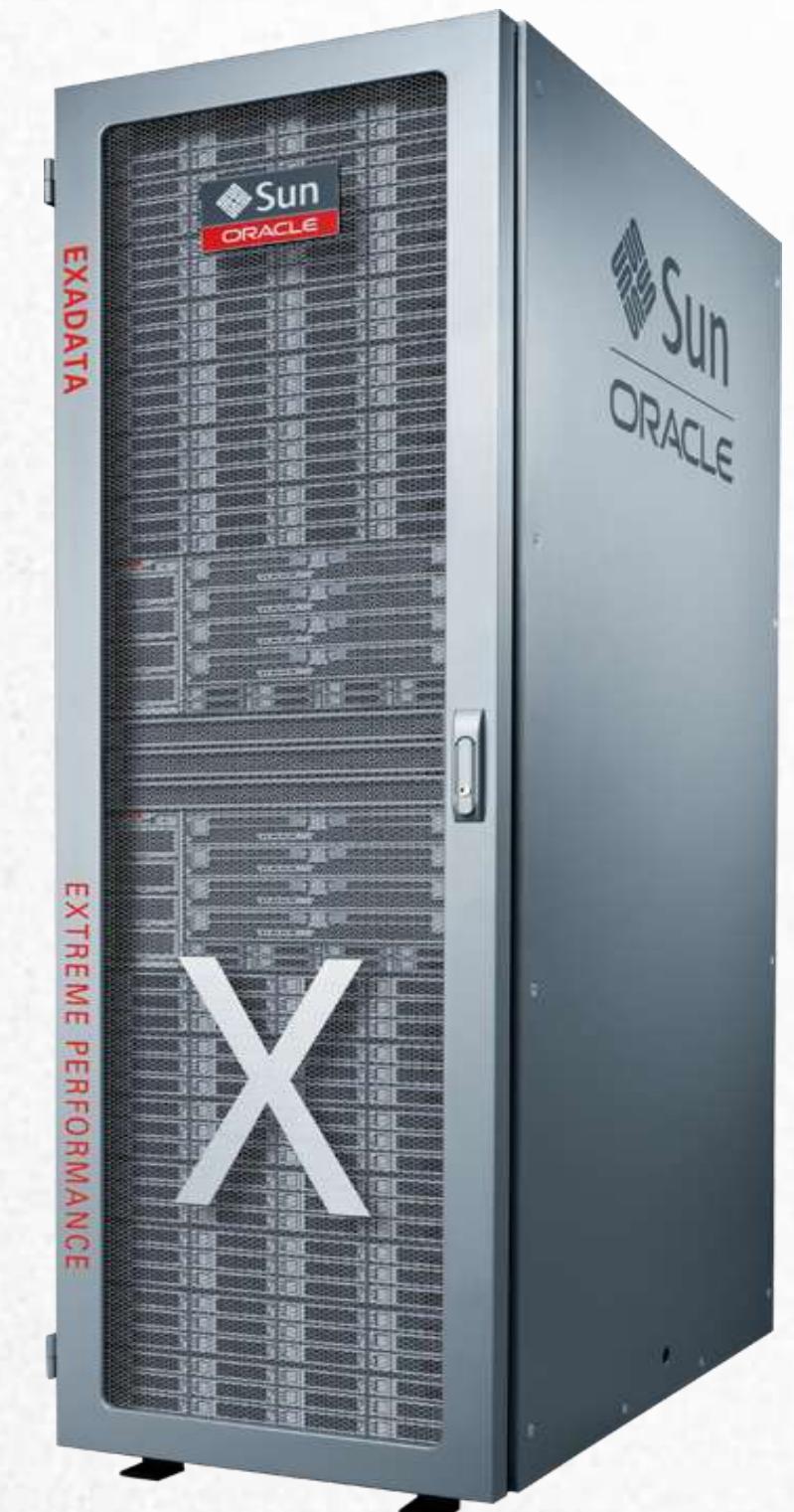
Общая архитектура Oracle для анализа Больших Данных



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

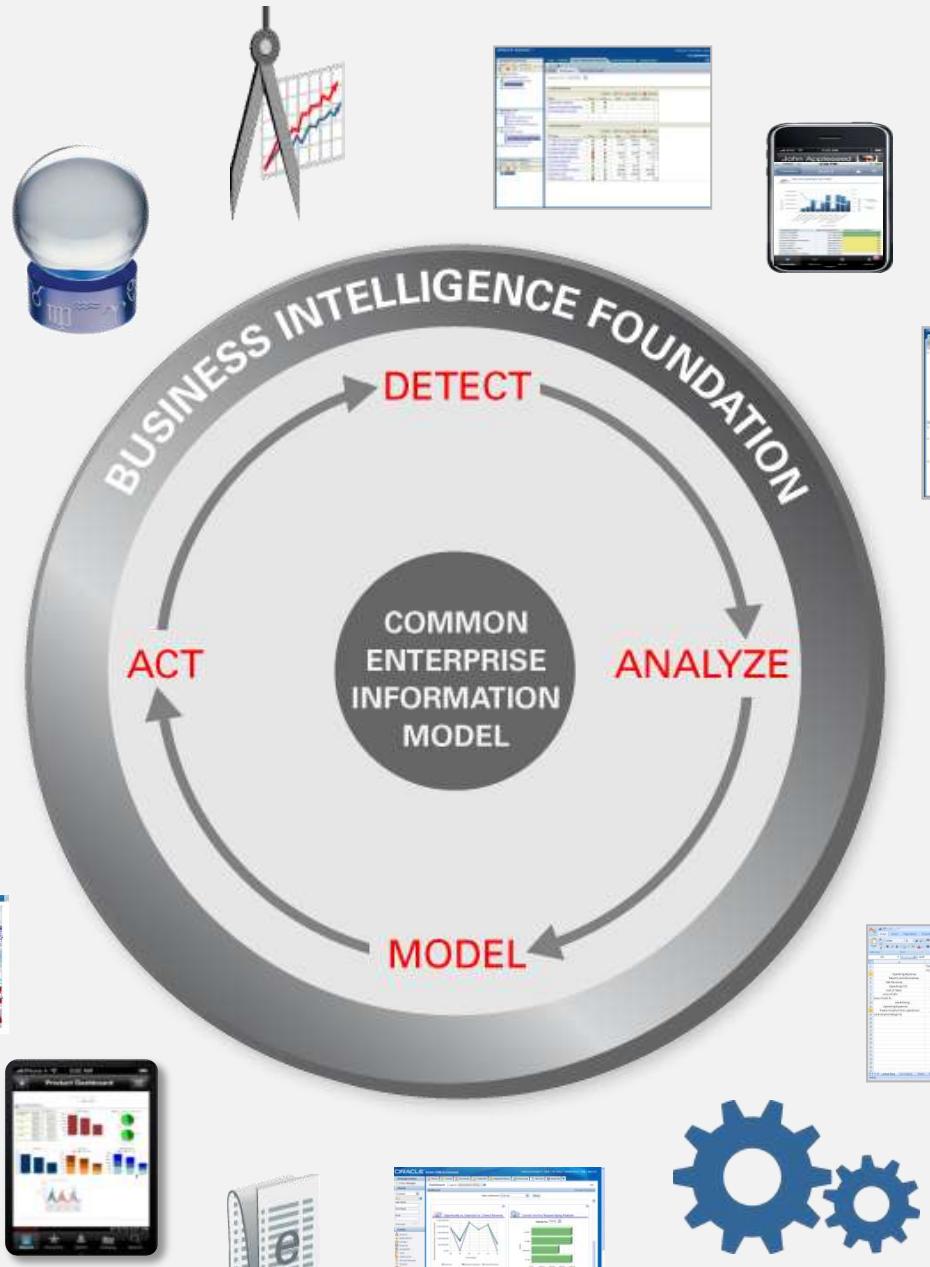
- Платформа Oracle для Больших Данных
- **Аналитическая in-memory машина Oracle Exalytics**
- От бизнес-анализа к исследованию данных – Oracle Endeca Information Discovery
- Встроенная аналитика для статистических исследований – Oracle R Enterprise

Аналитическая машина Oracle Exalytics

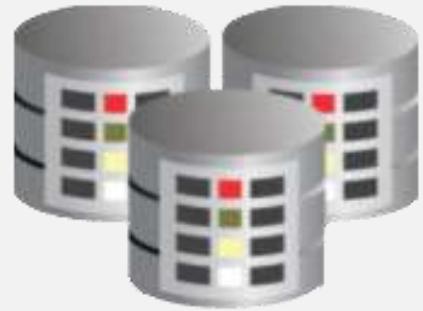
- Программно-аппаратный комплекс для бизнес-анализа
 - «Экстремальная производительность и неограниченная визуализация»
- Реализация концепции «аналитика в оперативной памяти» (in-memory analytics)



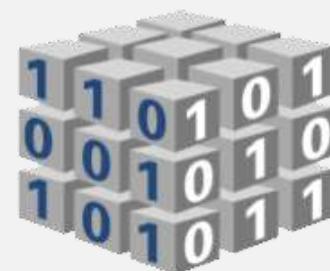
Oracle Exalytics – основные компоненты



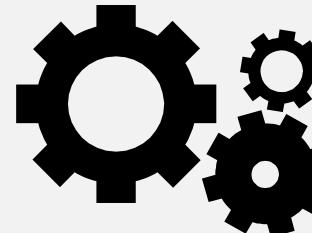
Oracle Business Intelligence Suite – специальная
редакция для Exalytics



TimesTen for Exalytics



Memory Optimized Essbase



Adaptive In-Memory Tools

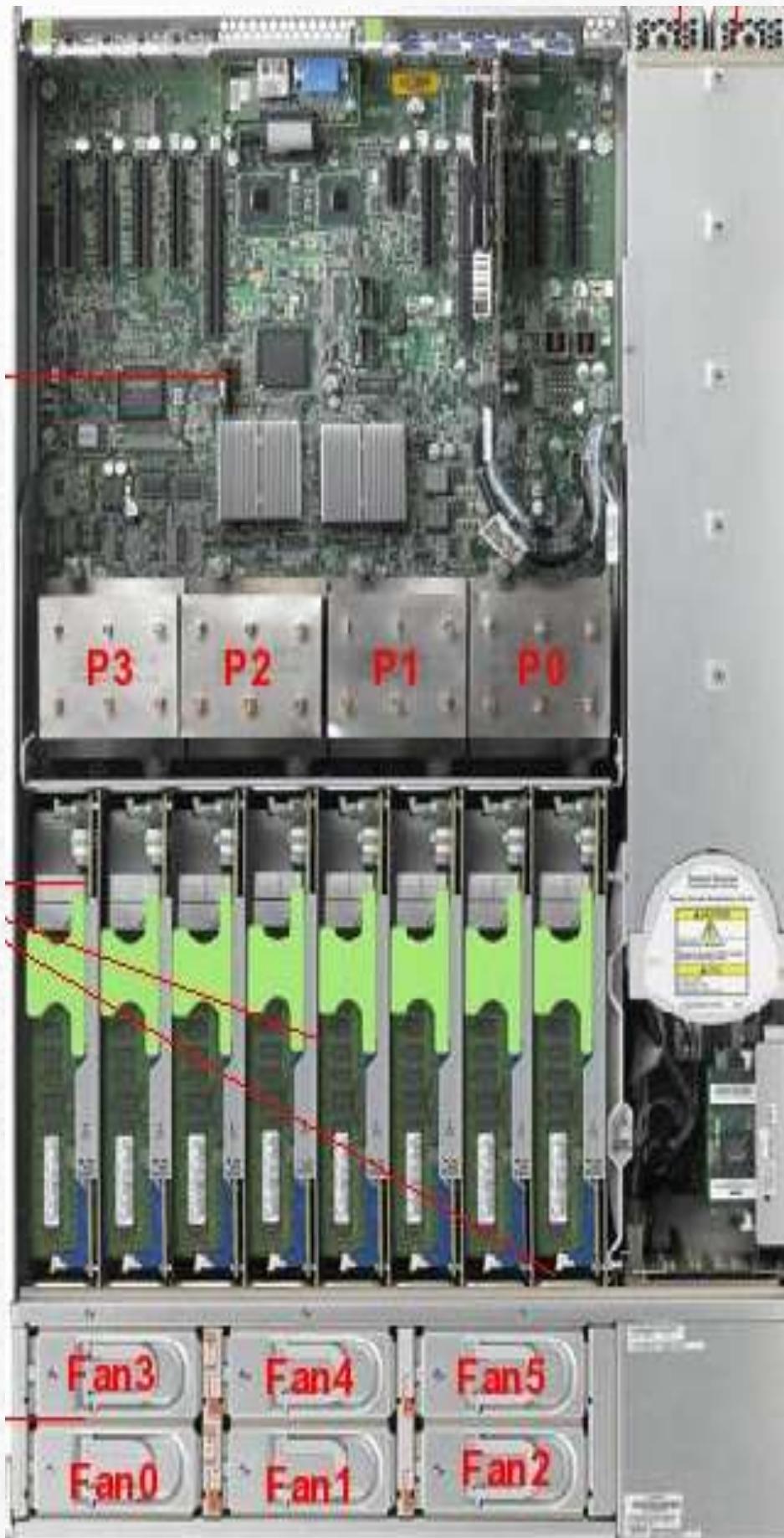
Программное
обеспечение для
In-Memory аналитики



1 TB RAM
40 Processing Cores
High Speed Networking

Аппаратный комплекс
для
In-Memory аналитики

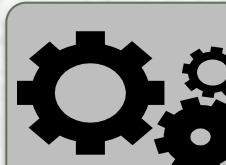
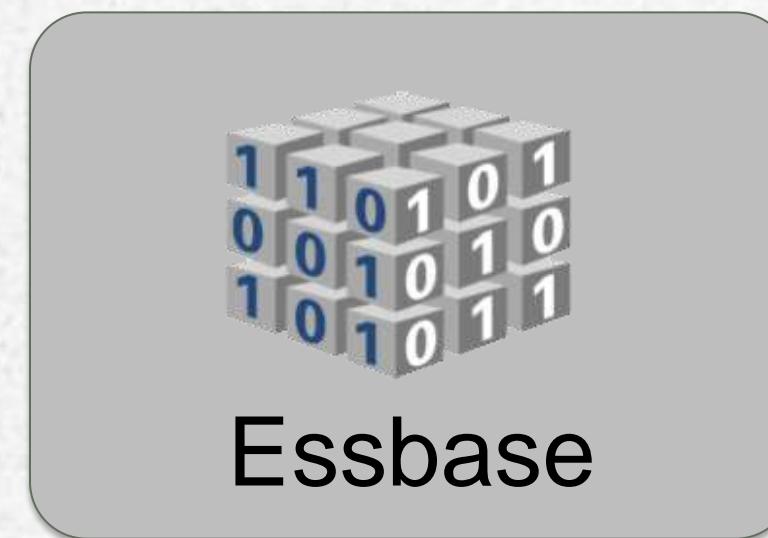
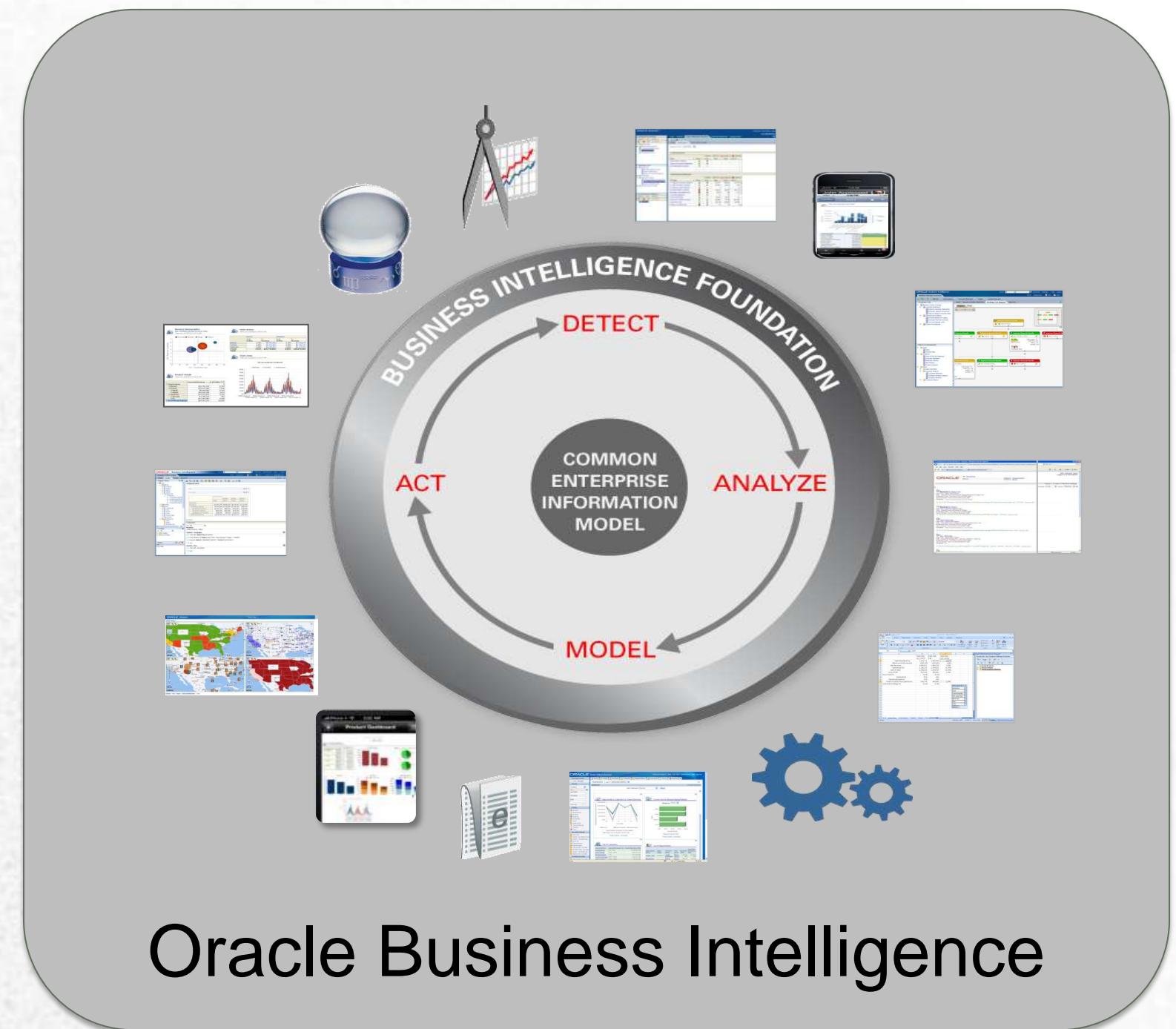
Аппаратное обеспечение



- **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



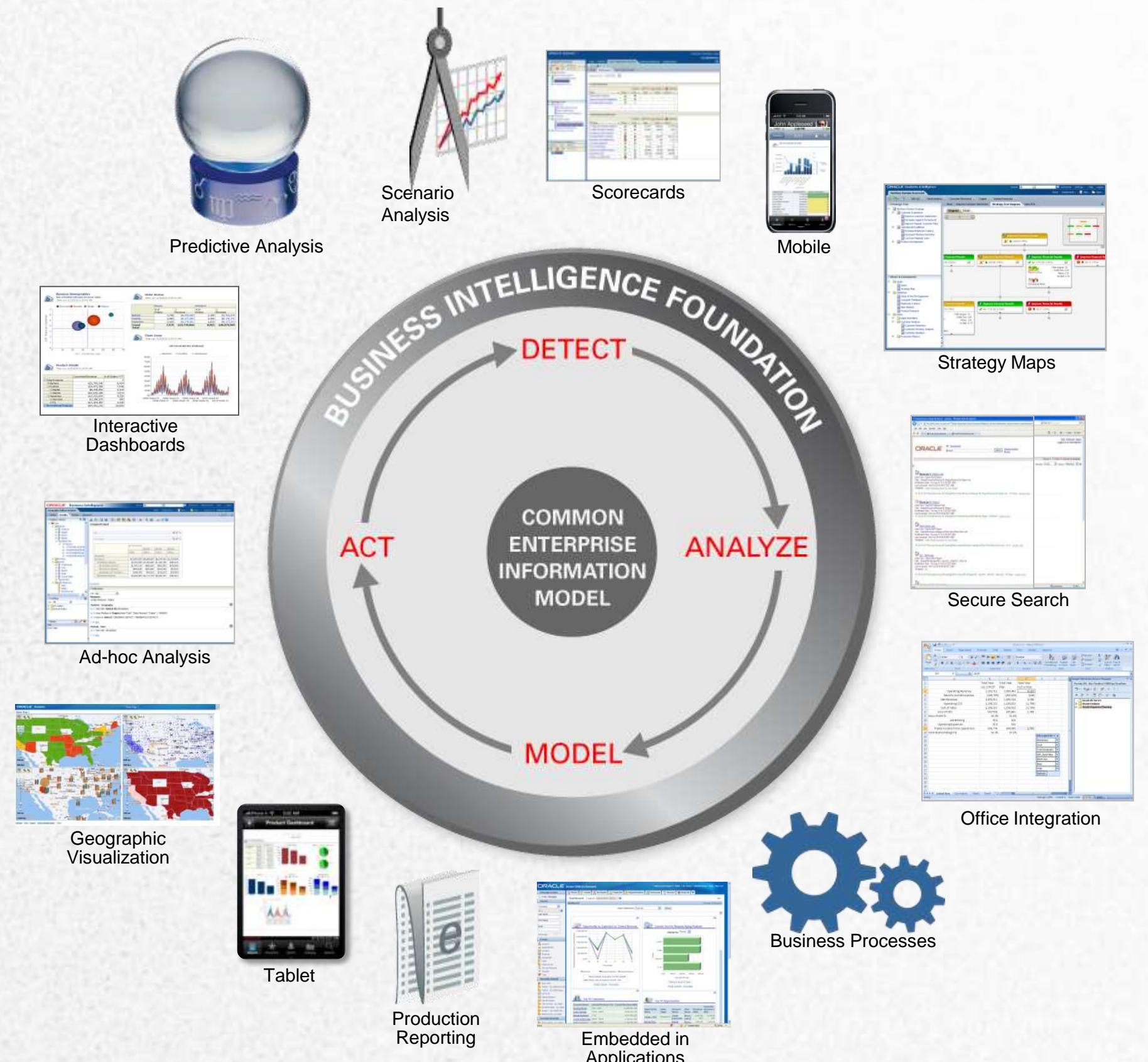
Адаптивные in-memory акселераторы

ORACLE®

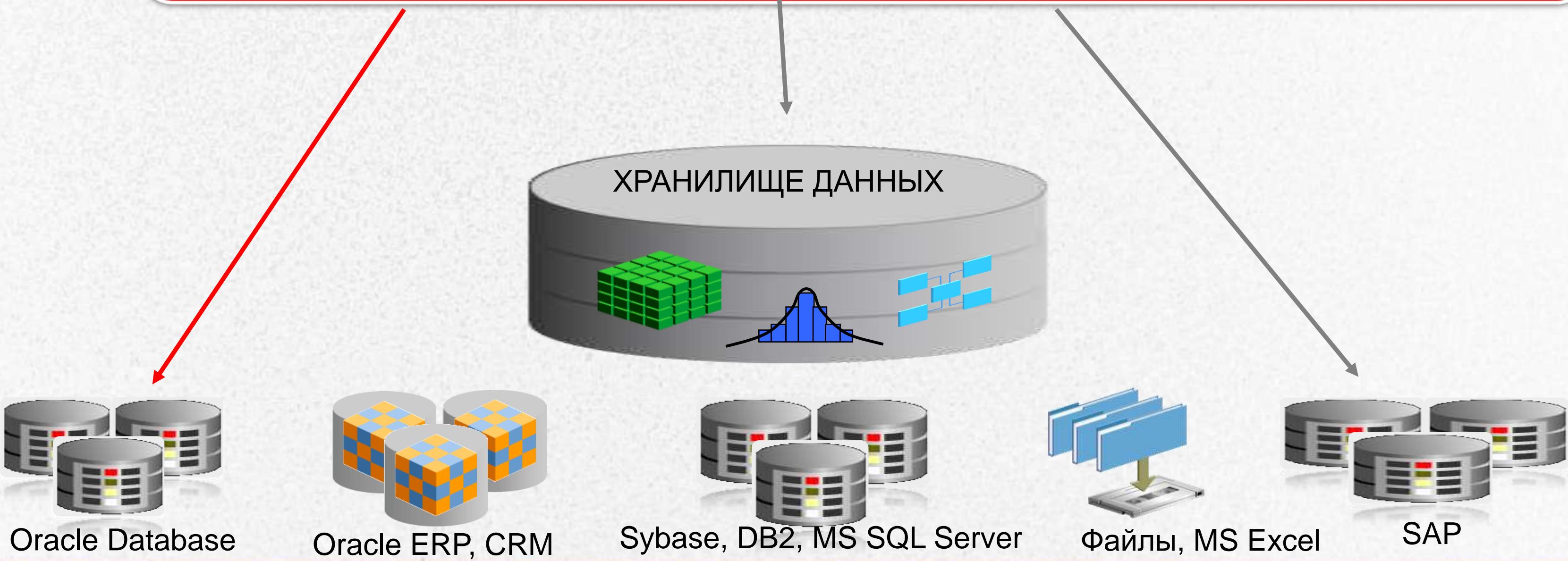
Oracle Business Intelligence

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®

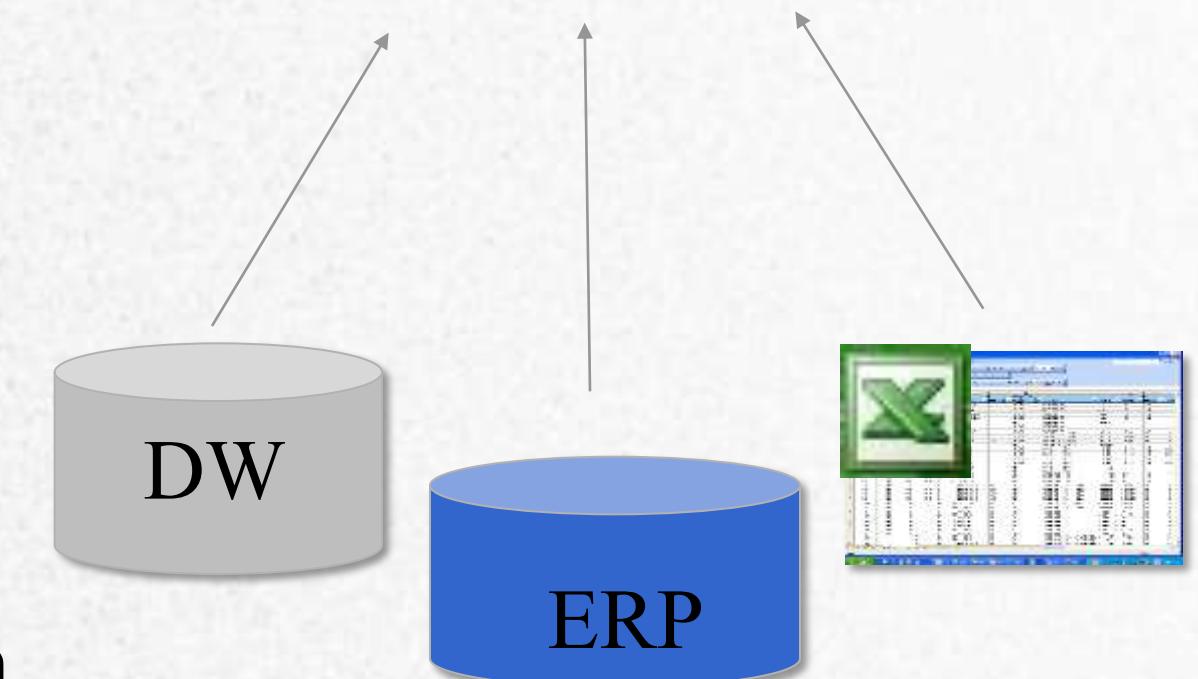
Аналитический сервер

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



Oracle
BI Server



Все виды бизнес-анализа на единой платформе

Не только отчеты

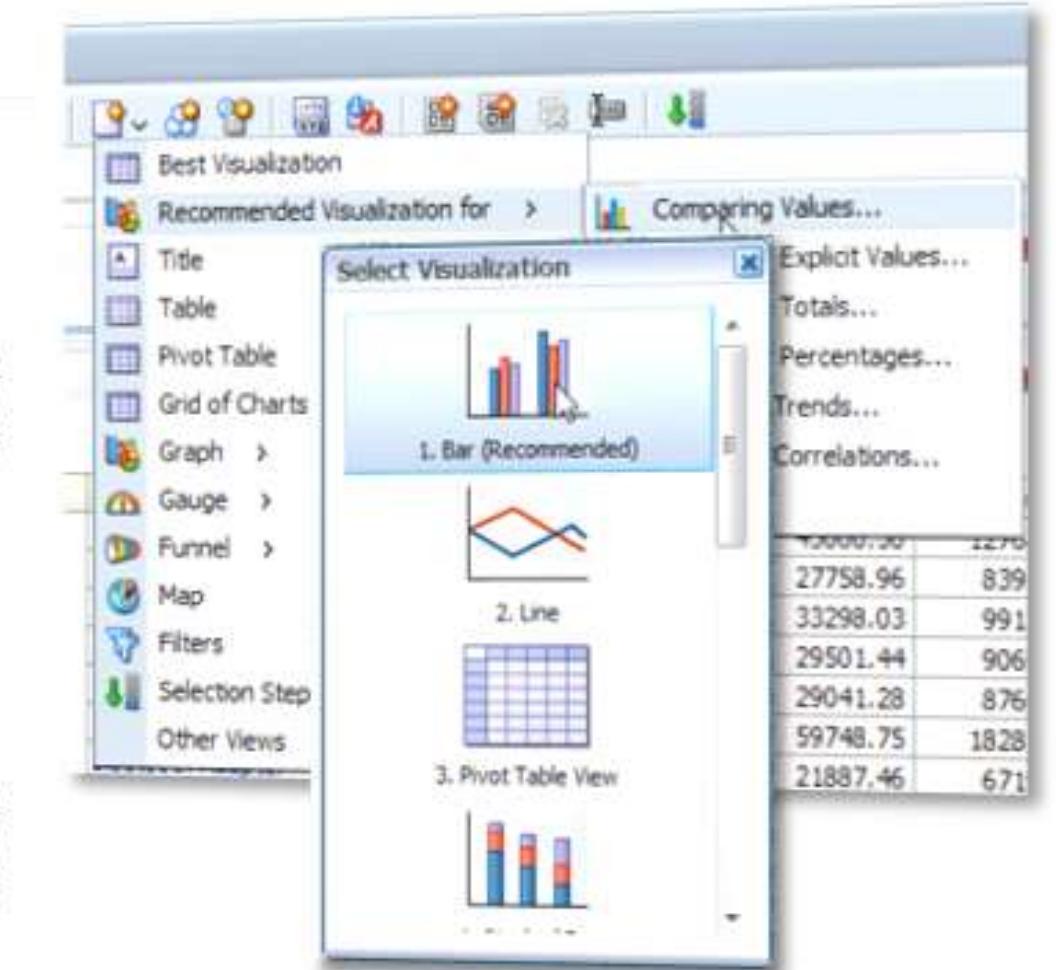
- Интерактивный анализ
- Пространственная аналитика
- Активный бизнес-анализ
- BI Collaboration
- Ключевые показатели и стратегическое управление
- Мобильная аналитика



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



Records: 911,520,524

Company

Vehicle Type

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

Fuel Technology

- Diesel
- Gasoline
- Hybrid
- Electric

* Sales Period

Between

2012

2011

2006

2001

2000

Financing Option

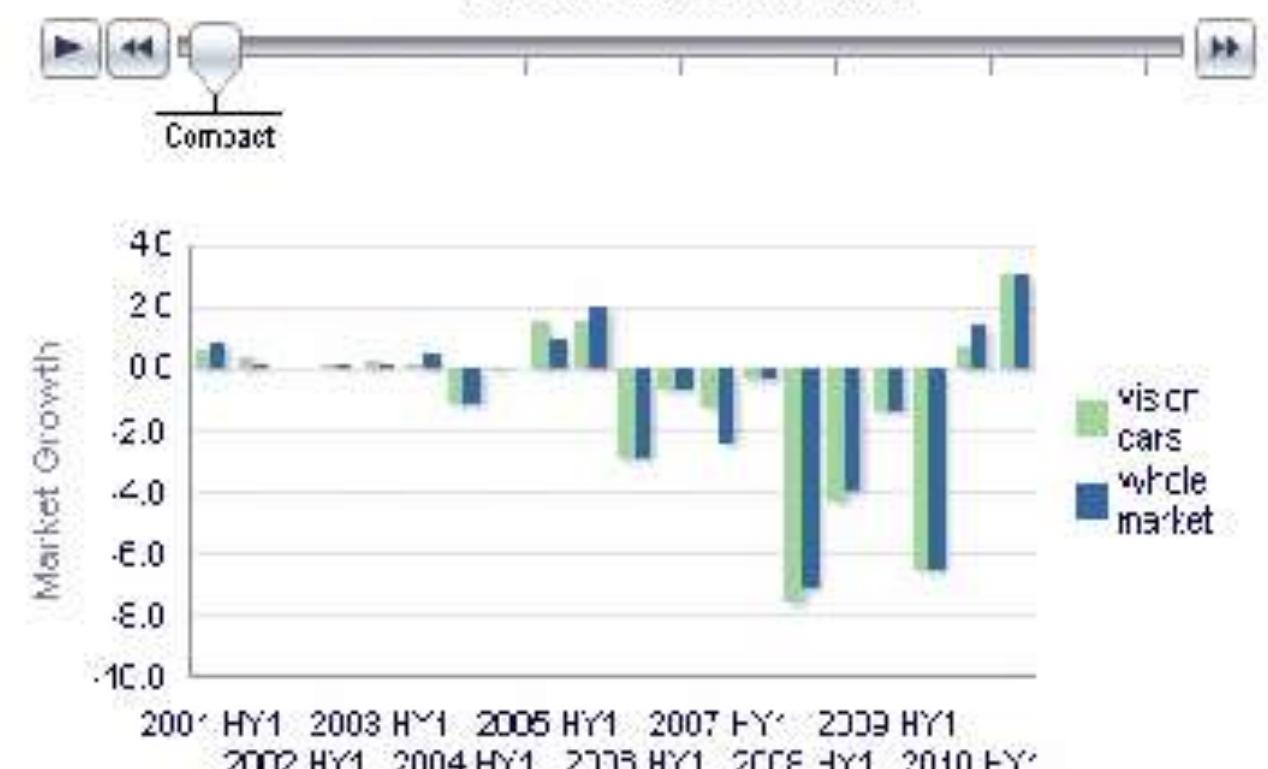
 (All Column Values) Lease

▼ Trend Data by Vehicle Type

Registrations Summary

	Registrations	Trend	Growth
Compact	121,395,717		
Full Size	62,703,324		
Midsize	105,633,147		
Pickups	74,386,170		
SUVs	135,399,076		
Sports	21,719,545		
Vans	42,174,097		
Total	563,915,164		

Market Growth Over Time

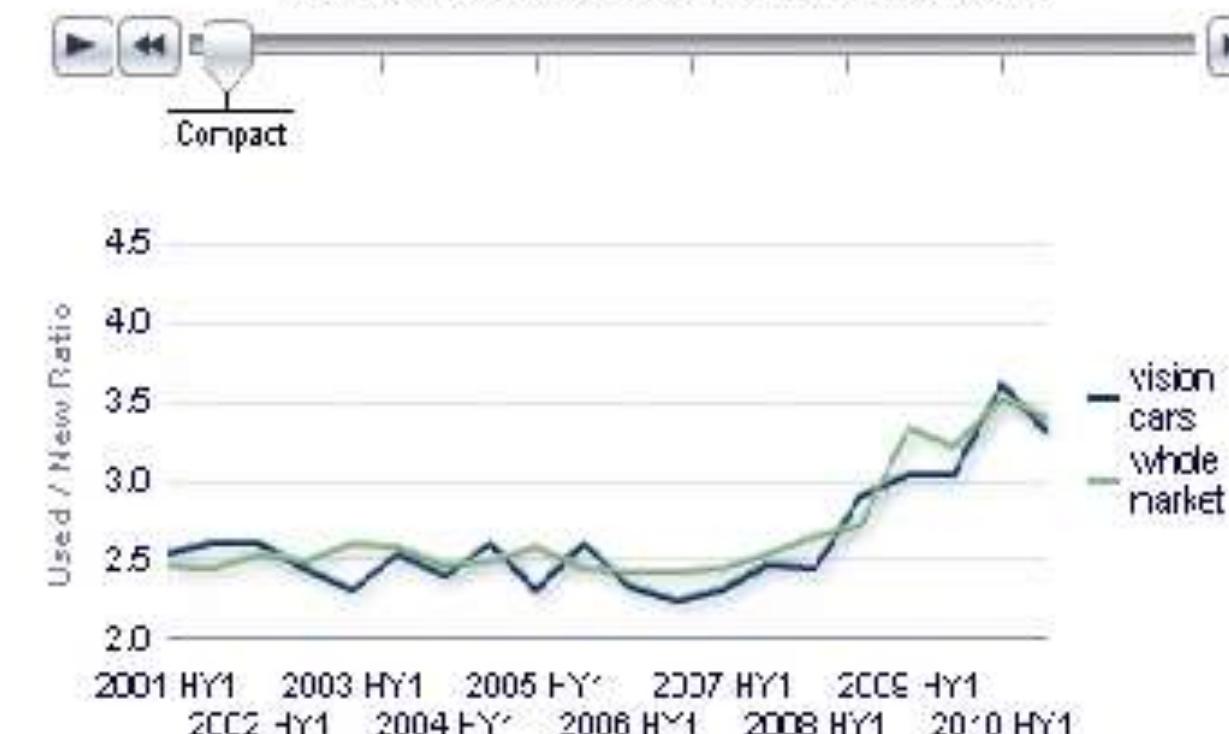


▼ Trend Data by Registration Type

New and Pre-Owned Registration Breakdown

	Registrations	Pre-Owned	New	Pre-Owned / New Ratio
Compact	121,395,717	86,056,037	35,340,677	2.64
Full Size	62,703,324	45,835,081	15,863,243	2.72
Midsize	105,633,147	76,690,237	23,941,913	2.65
Pickups	74,386,170	53,812,062	20,563,113	2.62
SUVs	135,399,076	98,446,765	37,440,314	2.63
Sports	21,719,545	15,797,465	5,922,180	2.67
Vans	42,174,097	30,362,638	11,810,462	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



Records: 41,305,900

> Trend Data by Vehicle Type

> Trend Data by Registration Type

> Trend Data by Region

Company

Mson Cars

Vehicle Type

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

Fuel Technology

- Diesel
- Gasoline
- Hybrid
- Electric

* Sales Period

Between

2012

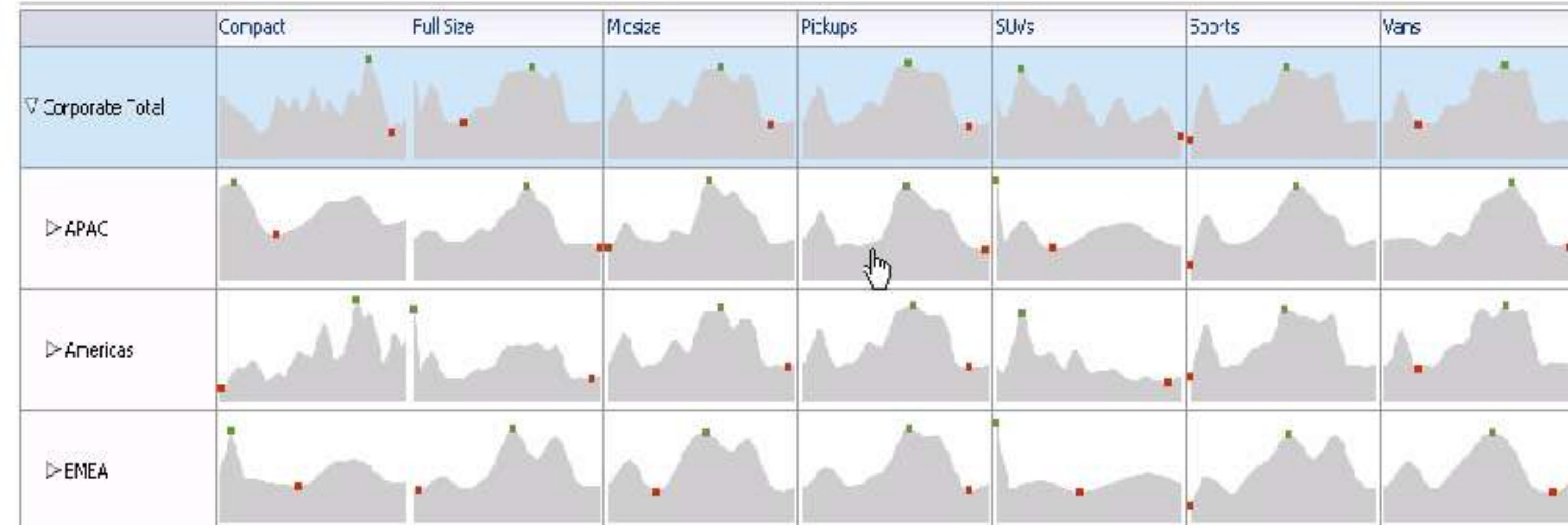
2011

2006

2001

2000

Financing Option

 (All Column Values) Lease

Records: 41,305,900

Trend Data by Vehicle Type

Trend Data by Registration Type

Trend Data by Region

Company

Mson Cars

Vehicle Type

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

Fuel Technology

- Diesel
- Gasoline
- Hybrid
- Electric

* Sales Period

Between

2012

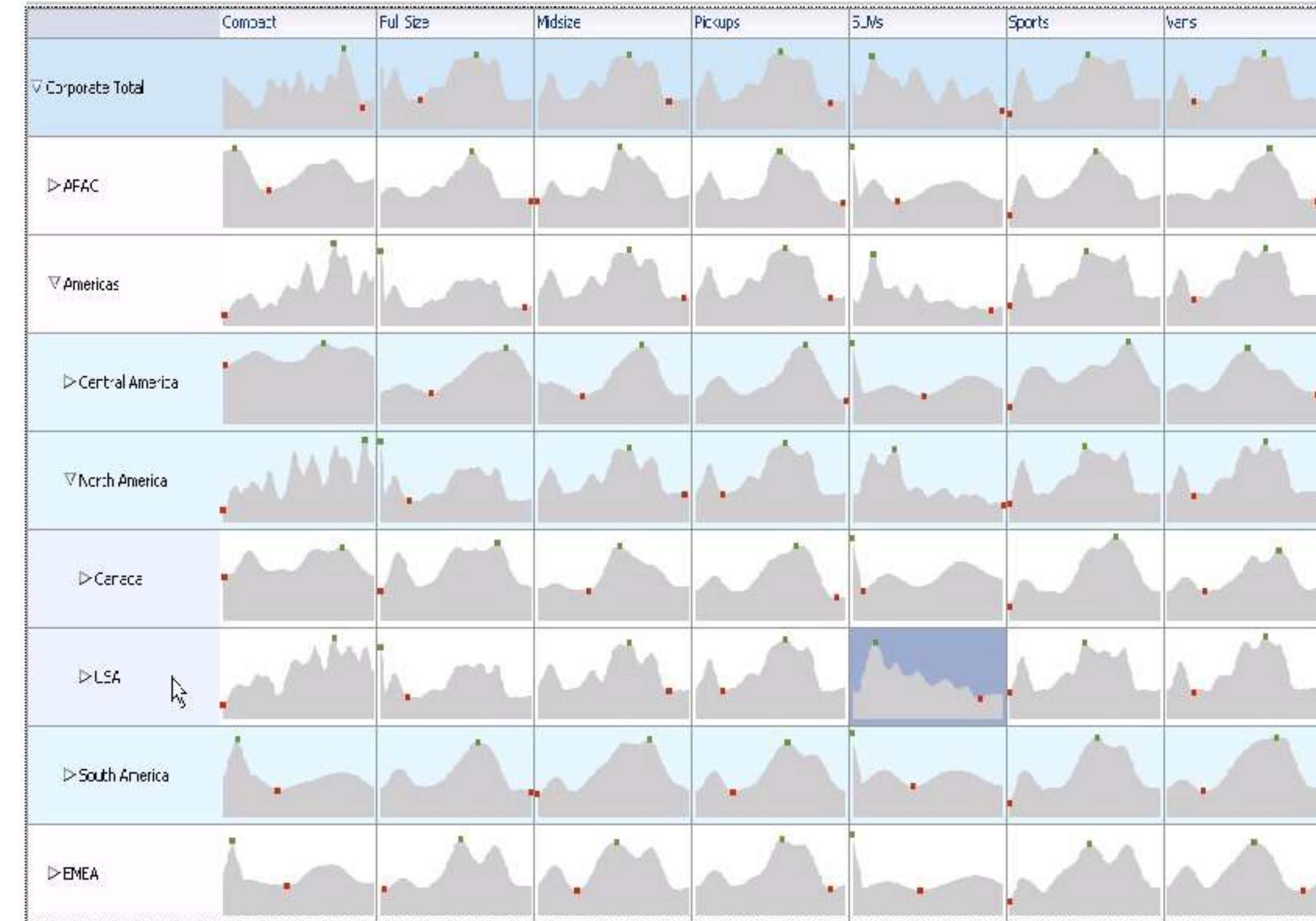
2011

2006

2001

2000

Financing Option

 (All Column Values) Lease

Oracle Essbase

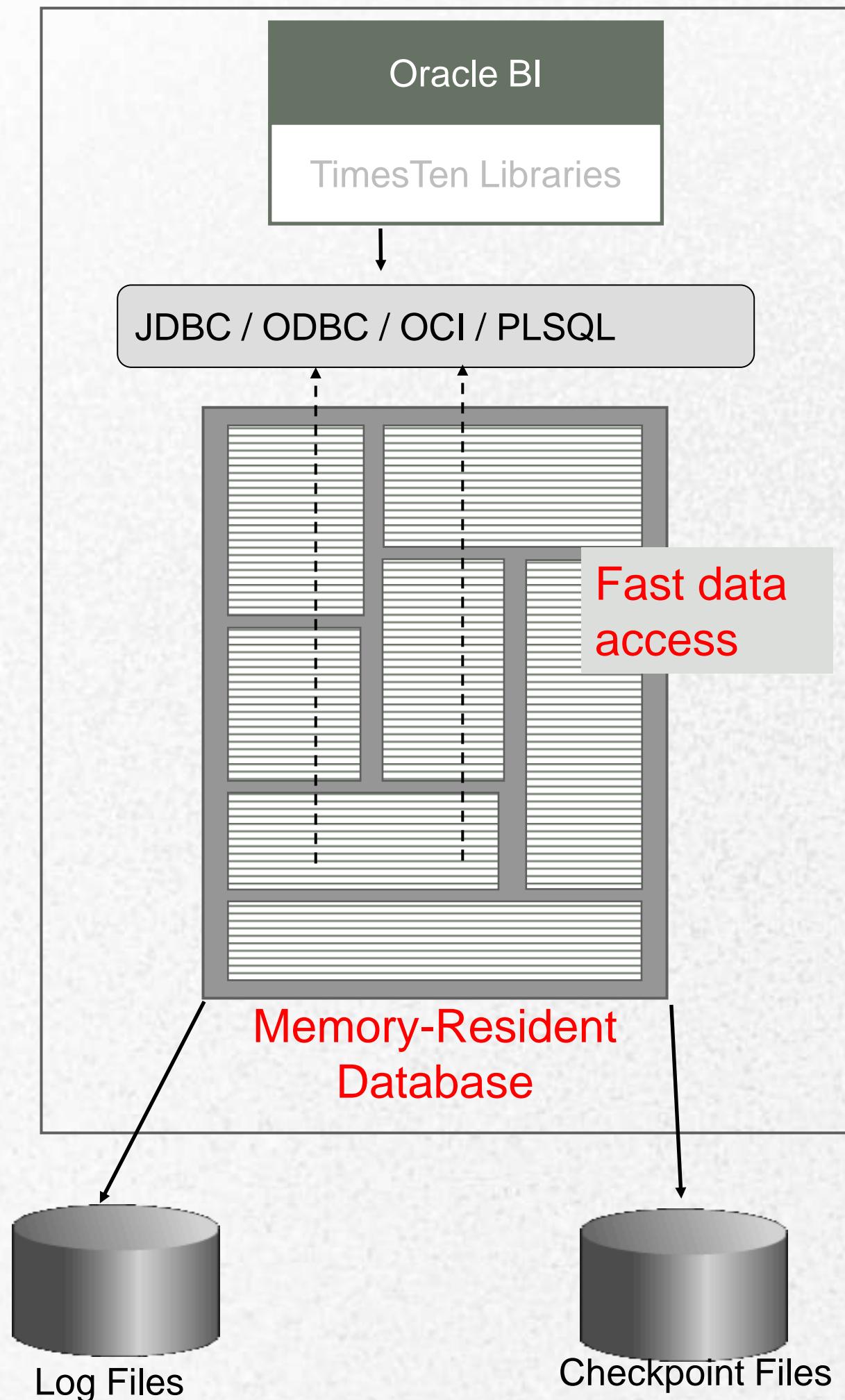
- Универсальный OLAP-сервер для хранения, обработки и представления информации
- Высокая производительность
- Моделирование сложной аналитики
- Основа продукта Hyperion Planning
- Essbase на Exalytics оптимизирован для повышения производительности работы Hyperion Planning и других EPM-приложений



ORACLE®

Times Ten for Exalytics

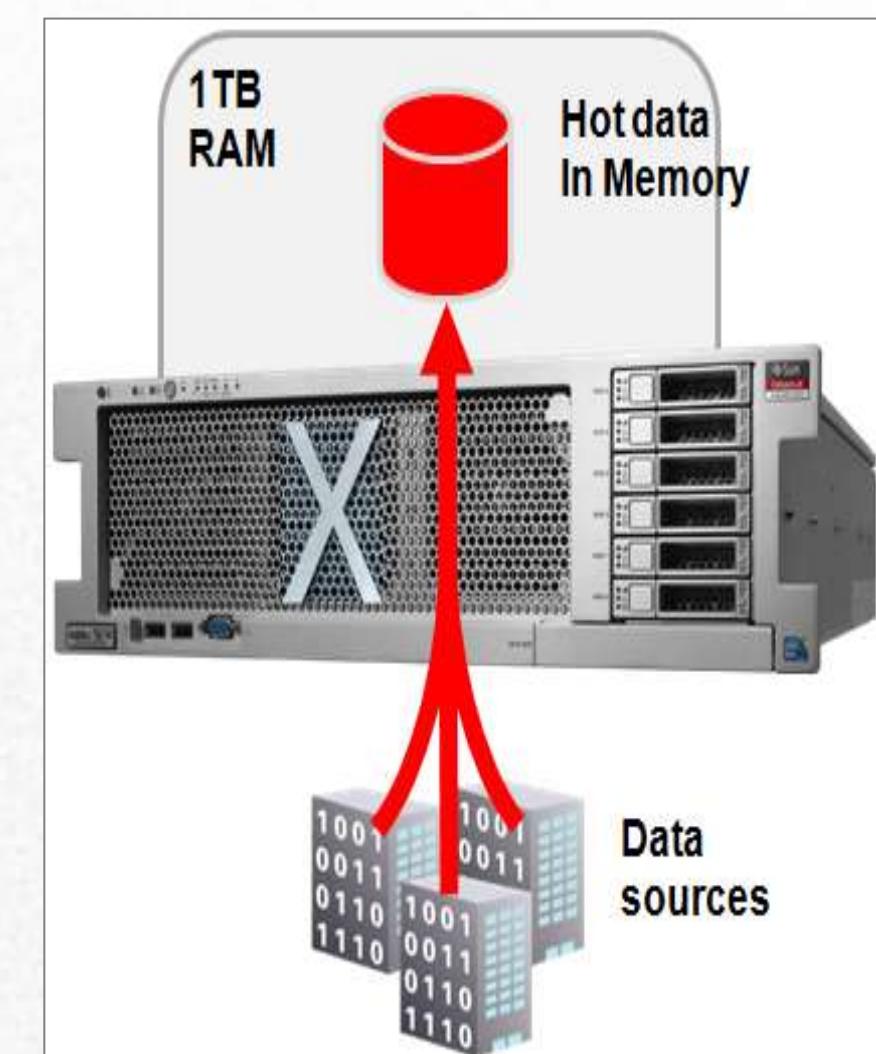
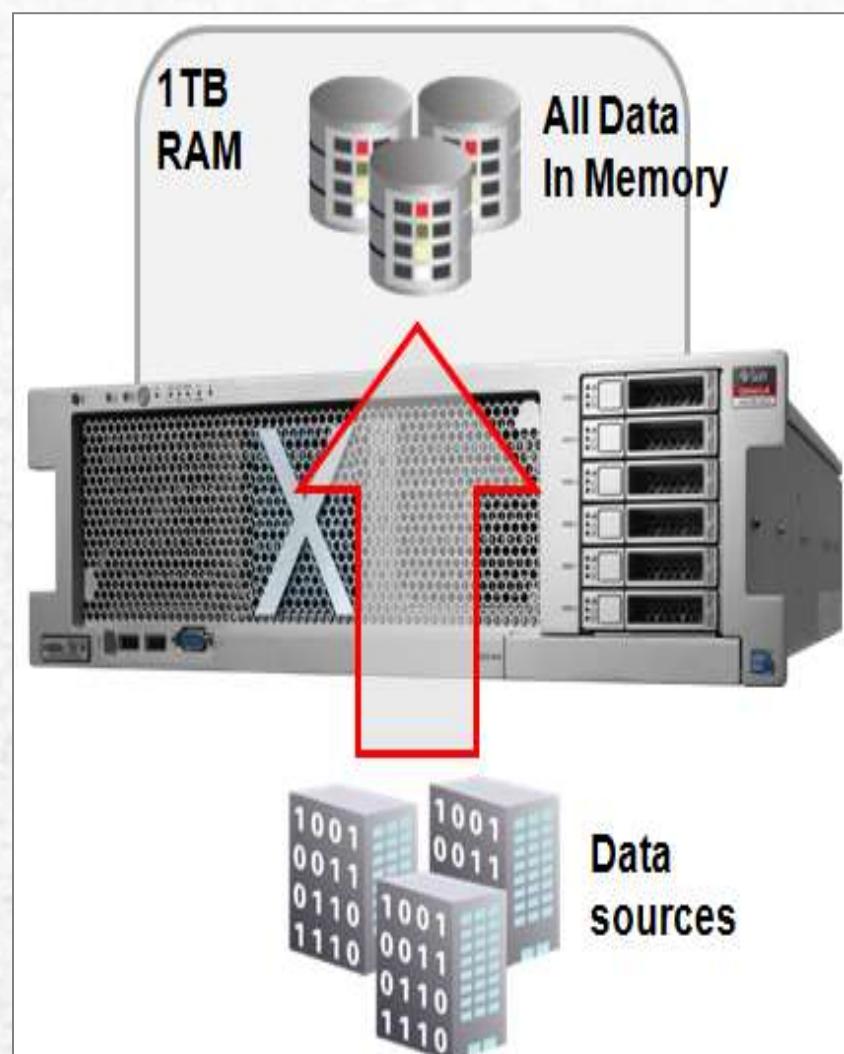
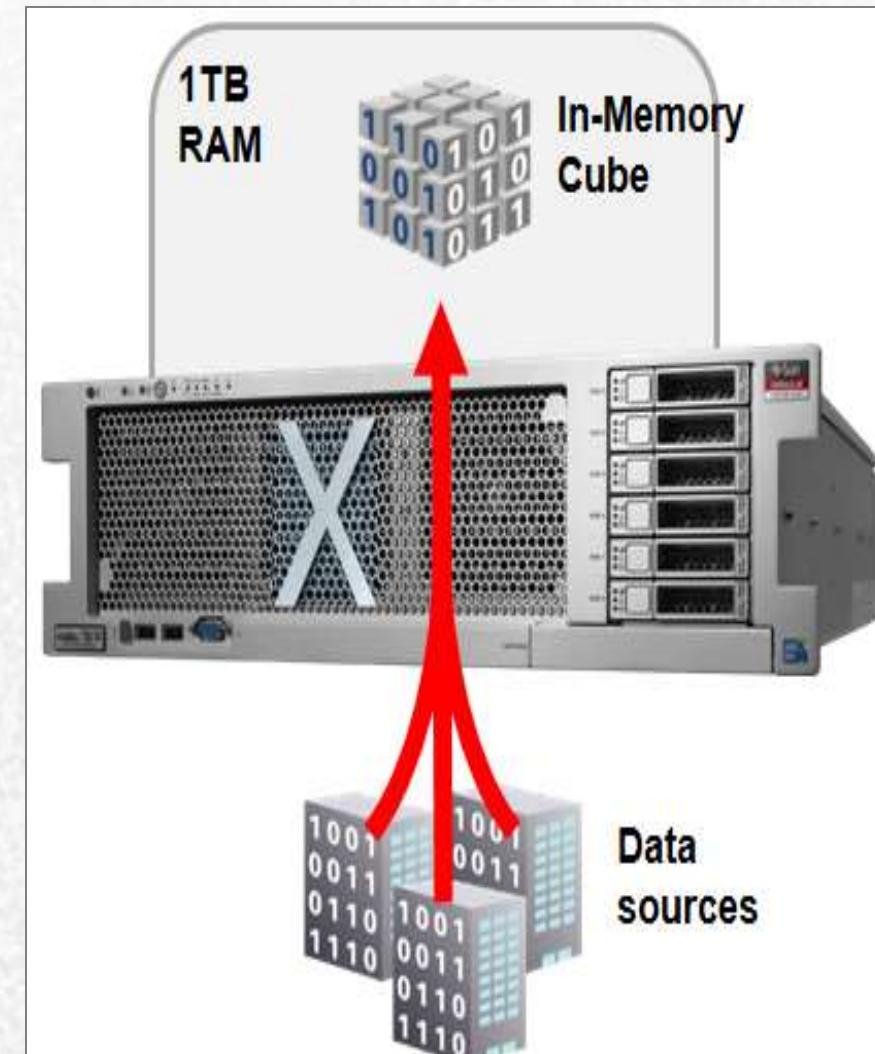
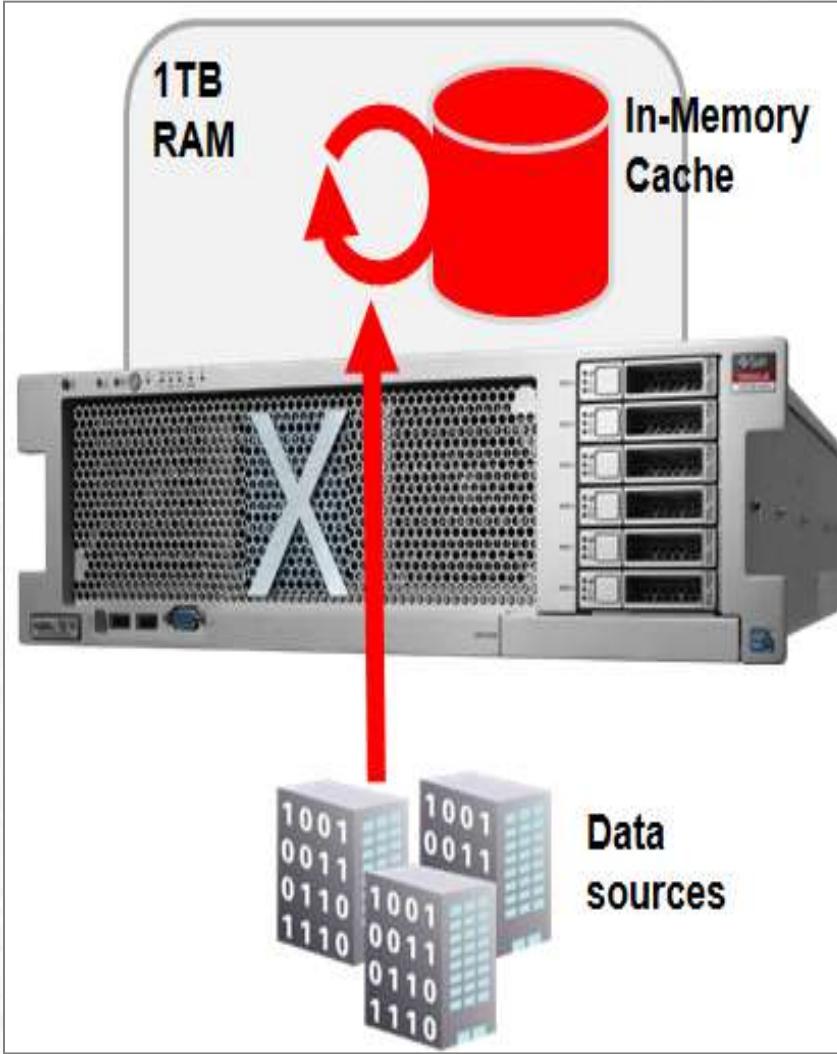
На основе *Oracle TimesTen In-Memory Database*



- 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

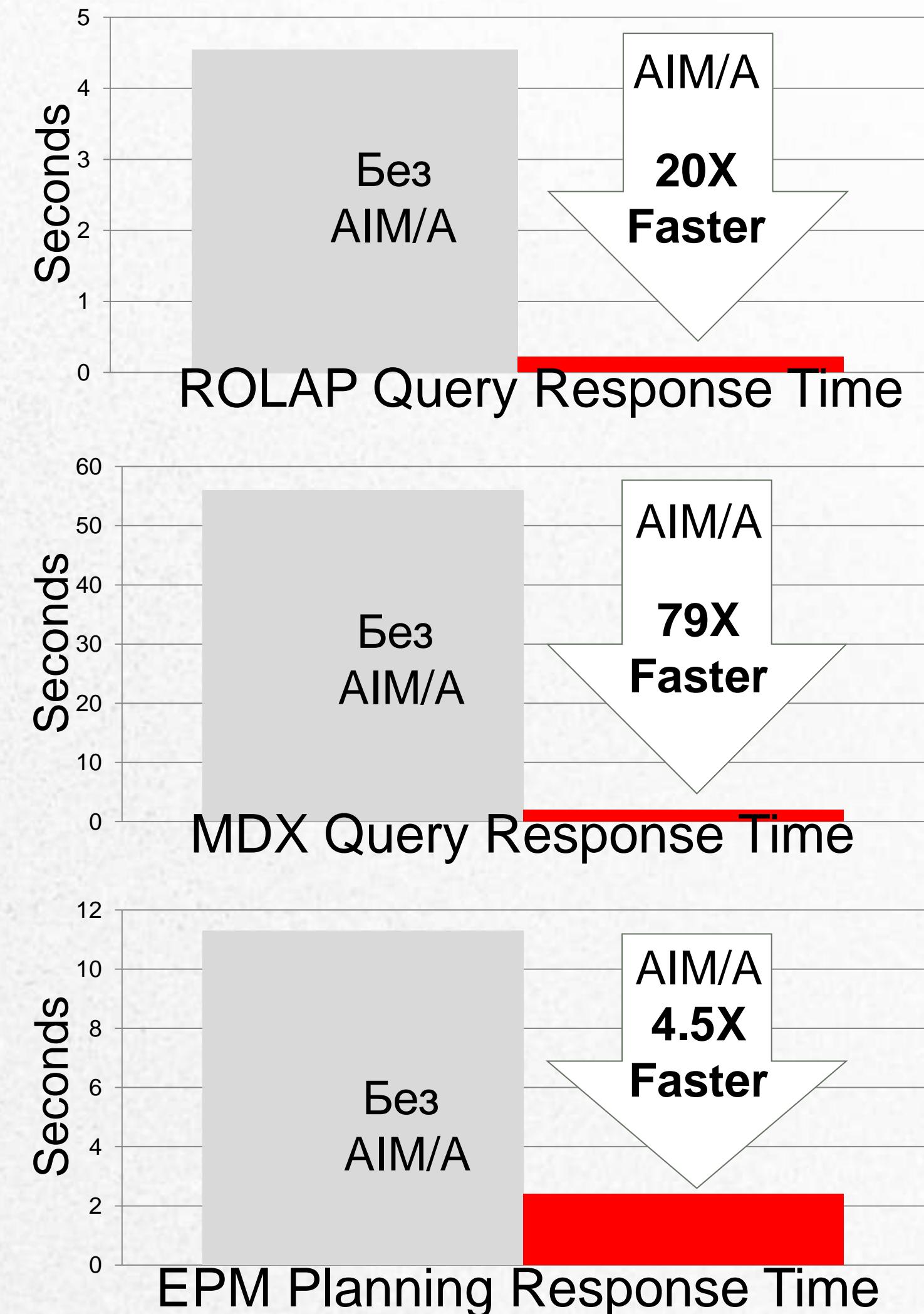
In-Memory Data Warehouse
Хранение в оперативной памяти всего хранилища данных

In-Memory Adaptive Data Marts
Кэширование в TimesTen
Автоматическое формирование витрины на основе статистики запросов

Что дают in-memory акселераторы

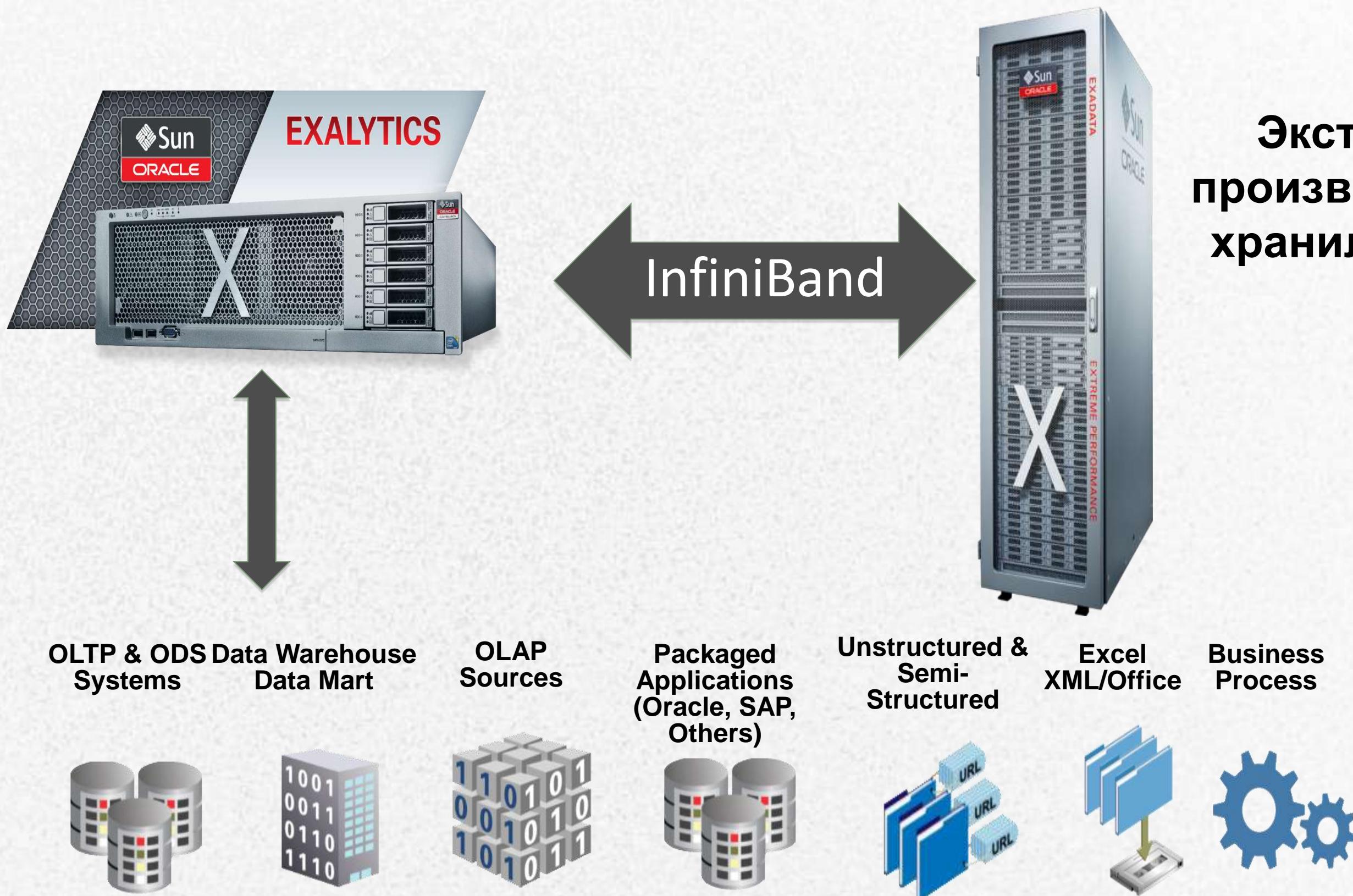
Повышение производительности

- **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

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 анализе 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, дашборды) 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

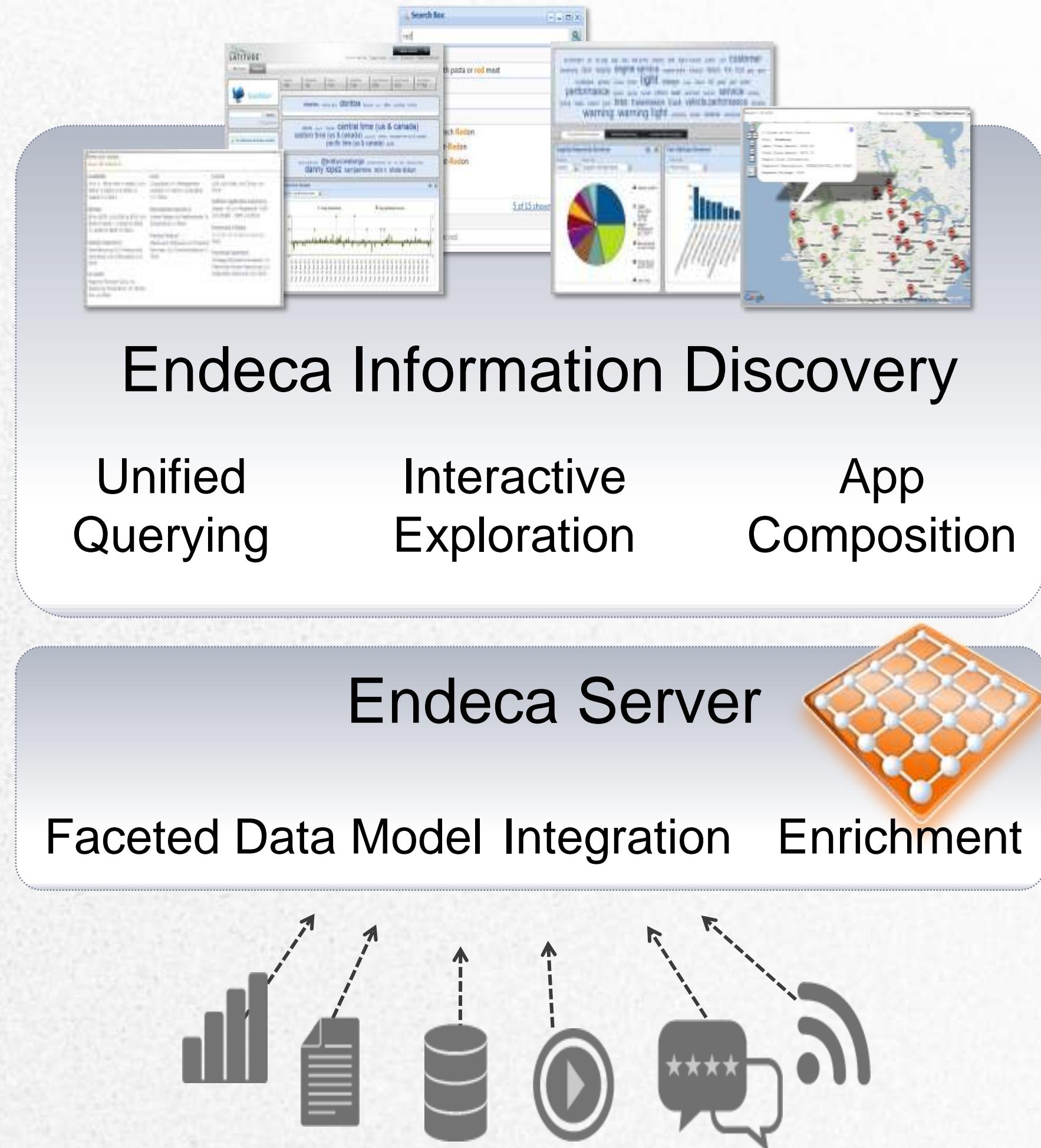
- Основана в Кембридже, МА в 1999
- Более 600 клиентов
- 33% of the Fortune 100
- Аналisis неструктурированной информации, Большие данные
- От систем поиска (для электронной коммерции) к «BI beyond the data warehouse»
- *Entdecken* (немецкий) = to discover, открытие



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

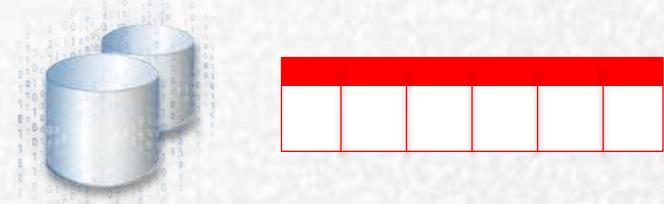
- Технологии доступа к информации, организованной на основе системы фасетной классификации
- Поиск путем уточнений, навигация по независимым параметрам
- Модель информационного поиска – набор характеристик
- Теоретические основы фасетной классификации --
Ш. Р. Ранганатан («Классификация двоеточием»,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

The screenshot illustrates a faceted search interface for digital cameras. At the top, the title 'Digital cameras' is displayed. Below it, a box explains that 'Manufacturer is a facet, a way of categorizing the results'. Another box states that 'Canon, Sony, and Nikon are constraints, or facet values'. A third box notes that 'The facet count or constraint count shows how many results match each value'. The main search area shows facets for 'Manufacturer' (with options for Canon USA, Sony, Nikon, Olympus, and Pentax), 'Resolution' (with options for 6 megapixels and up, 8 megapixels and up, 10 megapixels and up, 12 megapixels and up, and 14 megapixels and up), 'Zoom range' (with options for 3X to 4X and 6X to 12X), and 'More' (with options for LCD size, Image stabilizer, Flash memory, Still image format, and Maximum ISO). A 'Refine your results' button is located above the facets. Below the facets, a breadcrumb trail shows 'you selected: \$400 - \$500 SLR remove all'. The search results are displayed as a 'Regular search results list' showing 17 results, with page navigation buttons for 1, 2, and 'next'. The results list includes a camera icon, the product name 'Canon EOS Rebel XS (silver, with 18-55mm lens)', the price '\$459 to \$699', and the availability 'at 15 stores'. A 'COMPARE SELECTED' button is also present. The bottom of the interface features a 'Show 10 results per page' dropdown, a 'Sort by' dropdown set to 'Review date', and a 'SEARCH' button.

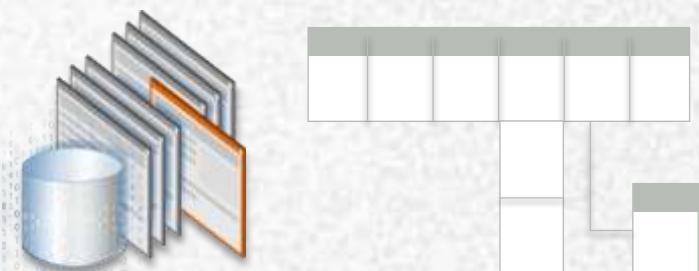
Разработка приложений

Источники данных

Структурированные



Слабоструктурированные



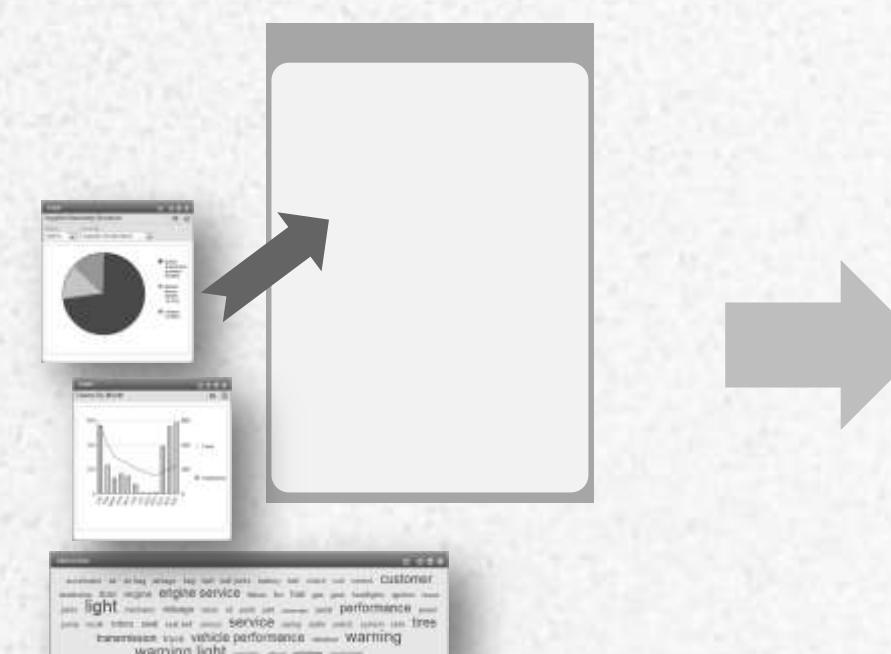
Неструктурированные



Загрузка в Endeca Server (без модели)

(без модели)

Drag-and-drop
инструменты
создания
приложения



Интерактивный
поиск, навигация
и анализ



Фасетная модель данных

Позволяет быстро загружать и изменять данные из любого источника

- Набор записей, каждая из которых имеет собственную «структурой»
 - Многозначные поля
 - Неструктурированные поля (тексты)
- Один из видов «Key Value» модели
- Модель состоит из
 - Записей
 - Атрибутов
- Каждая запись – это набор пар (атрибут, значение)
- Нет никакого 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

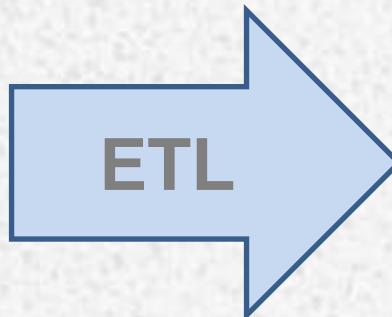
- TxnID = 12325
- ProductID = 507
- Category = Road Bike
- Amount = \$1399.49
- Weight = 20lb.
- FrameType = Composite
- Saddle = Bontrager Race
- Review = Disappointing for the price. The frame feels heavier than I expected.
- ReviewSentiment = Negative
- ReviewTerm = Disappointing
- ReviewTerm = Price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

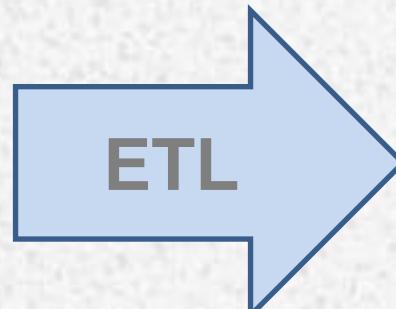
- TxnID = 12325
- ProductID = 507
- Category = Road Bike
- Amount = \$1399.49
- Weight = 20lb.
- FrameType = Composite
- Saddle = Bontrager Race
- Review = Disappointing for the price. The frame feels heavier than I expected.
- ReviewSentiment = Negative
- ReviewTerm = Disappointing
- ReviewTerm = Price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



- 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

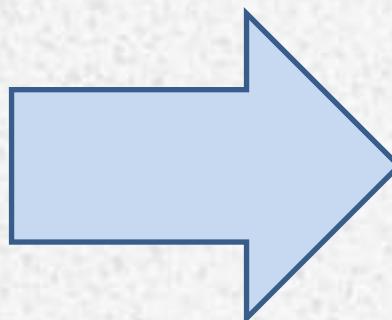
- Неструктурированные данные «присоединяются» к остальным данным на основе значения ключа
- В определенных случаях неструктурированные данные могут храниться в виде независимых записей
- Текстовые документы, RSS, Twitter, Facebook, форумы, etc..

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

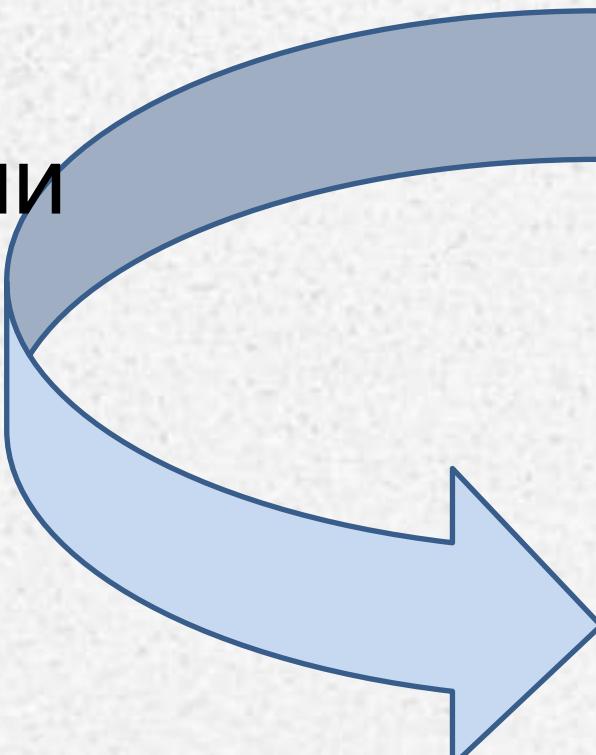
- TxnID = 12325
- ProductID = 507
- Category = Road Bike
- Amount = \$1399.49
- Weight = 20lb.
- FrameType = Composite
- Saddle = Bontrager Race
- Review = Disappointing for the price. The frame feels heavier than I expected.**
- ReviewSentiment = Negative
- ReviewTerm = Disappointing
- ReviewTerm =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

Staffing - endeca.com

ENDECA LATITUDE®

Welcome! Demo User | Toggle Controls | Logout | Demo Home | Create new bookmark

Welcome Staffing Utilization

Refine your results:

[Clear All Filters](#)

Availability	Level
Now More than 4 weeks (1500)	Consultant (247) Management
Within 4 weeks (935) Within 2 weeks (816) More	Analyst (157) Senior Consultant (151) More
Bill Rate	International Experience
\$0 to \$100 (259) \$100 to \$200 (224)	United States (695) Netherlands (76)
\$200 to \$300 (114) \$300 to \$400 (97) \$400 to \$500 (45) More	Switzerland (37) More
Industry Experience	Practice Vertical
Manufacturing (552) Finance and Insurance (448) Information (244) More	Electronics Software (222) Financial Services (128) Communications (87) More
Accounts	
Regions Financial Corp. (63) Wachovia Corporation (58) Yahoo! Inc. (54) More	

Consultants: 739

Average Performance Rating: 2.90

Avg Billing Rate: \$173.72

Actions ▾

Choose a column set: Position Location Resume

	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

Country

Select refinements, or search for a particular value:

Austria (14) X

Belgium (8) X

China (126) X

Denmark (8) X

Germany (25) X

India (188) X

Netherlands (3) X

Spain (31) X

Switzerland (5) X

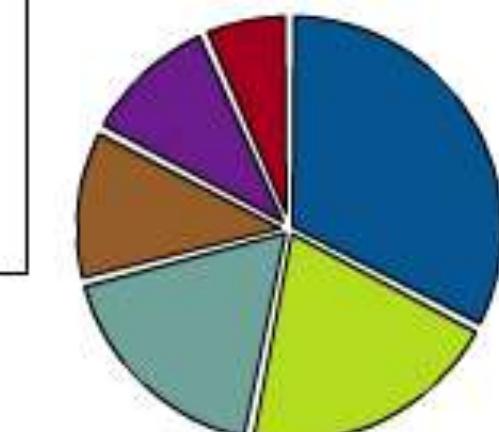
USA (355) X

Select All Deselect All

Submit Selections

Search: miles from

Group By:



Senior Manager 33%

Manager 20%

Senior Consultant 18%

Senior Management Analyst 12%

Consultant 11%

Management Analyst 7%

Staffing - endeca.com

ENDECA
LATITUDE®

Welcome! Demo User | Toggle Controls | Logout | Demo Home | Create new bookmark

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

State

Virginia (18) Maryland (15) Ohio (9)

More

Software Application Experience

Siebel - BI (31) Siebel - CRM (22)

Microsoft (17) More

Bill Rate

[\\$0 to \\$100](#) [\\$100 to \\$200](#)

\$200 to \$300 (33) \$300 to \$400 (36)

\$400 to \$500 (12) More

Level

Management Analyst (31) Senior

Management Analyst (28) More

Performance Rating

1 (12) 2 (15) 3 (18) 4 (7) 5 (9) More

Industry Experience

Manufacturing (57) Finance and

Insurance (48) Information (29)

More

International Experience

United States (59) Japan (6)

Netherlands (2) United States More

Functional Experience

Strategy/Business Processes (27)

Integration Enterprise Program

Management (17) Infrastructure

Project Lifecycle (15) More

Accounts

PNC Financial Services (12) Jabil

Circuit, Inc. (11) Host Marriott

Corporation (10) More

Practice Vertical

Electronics Software (18) Financial

Services (12) Retail Wholesale (8)

More

Consultants

59

Average Performance Rating

2.76

Avg Billing Rate

\$142.76

French

Search

Search Within

Geography Search:

miles from

Averages

Metrics:

Billing Rate

Group By:

Level

Senior Management Analyst 55%

Management Analyst 45%

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"/>	Thronsdson, Kieth	111	USA	South	Tennessee	Nashville
<input type="checkbox"/>	Pavelec, Alanna	143	USA	Midwest	Ohio	Cleveland

Staffing - endeca.com

Refine your results:
[Clear All Filters](#) [Search for 'French'](#)

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

Availability	State	Software Application Experience
Now More than 4 weeks (28)	Maryland (4) Virginia (3) Illinois (2)	Peoplesoft - CRM (8) Vignette (8)
Within 2 weeks (15) Within 4 weeks (15) More	More	Siebel - BI (5) More
Bill Rate	Level	Performance Rating
\$0 to \$100 \$100 to \$200	Management Analyst (7) Senior Management Analyst (8) More	1 (4) 2 (2) 3 (2) 4 (2) 5 (3) More
\$200 to \$300 (3) \$300 to \$400 (4)	International Experience	Functional Experience
\$400 to \$500 (3) More	United States (13) Netherlands (1)	Strategy/Business Processes (7)
Industry Experience	Japan (1) United States More	Integration Quality Management (5) Infrastructure (4) More
Manufacturing (13) Finance and Insurance (13) Information (6)	Practice Vertical	
Finance and Insurance Manufacturing More	Financial Services (4) Electronics Software (3) Communications (2) More	
Accounts		
RadioShack Corporation (4) PNC Financial Services (3)		
Amazon.com, Inc. (3) More		

Consultants 13 **Average Performance Rating** 2.85 **Avg Billing Rate** \$141.92

Geography Search
 miles from

Averages

Metrics: Billing Rate Group By: Level

Level	Percentage
Senior Management Analyst	55%
Management Analyst	45%

Actions ▾ Choose a column set: Position Location Resume

	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...
<input type="checkbox"/>	Throndson, Kieth	111	Kieth_Throndson_235.doc	...Committee, European Logistics Association (ELA)
<input type="checkbox"/>	Bough, Levi	157	Levi_Bough_555.doc	...and spoken English, Spanish, Portuguese, and I...
<input type="checkbox"/>	Dierking, Sanjuanita	140	Sanjuanita_Dierking_1517.doc	...Low/Medium Windows 95 High Windows NT Medi...
<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 Medi...
<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

BI Foundation Suite

Запросы,
интерактивные
отчеты, и анализ

OLAP анализ

Регламентная
отчетность

Data Discovery

- Интерактивные информационные панели, навигация и детализация

Business
Intelligence
Enterprise Edition

Essbase

BI Publisher

Endeca Information
Discovery

- Высоко-производительный генератор отчетов, высокое качество печати

- Фасетный поиск и навигация для структурированных данных и неструктурированной информации

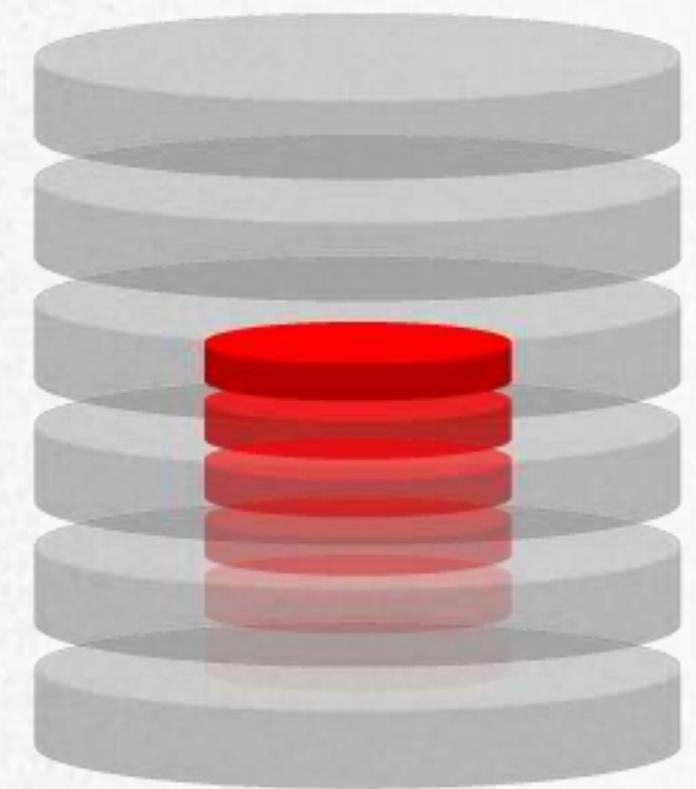
План

- 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

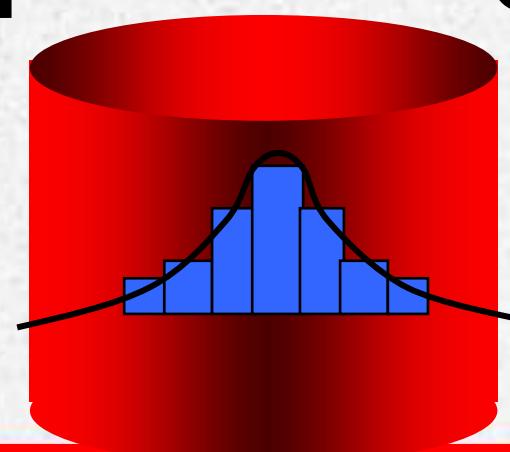
Встроенная в базу данных аналитика

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



Новое решение Oracle R Enterprise

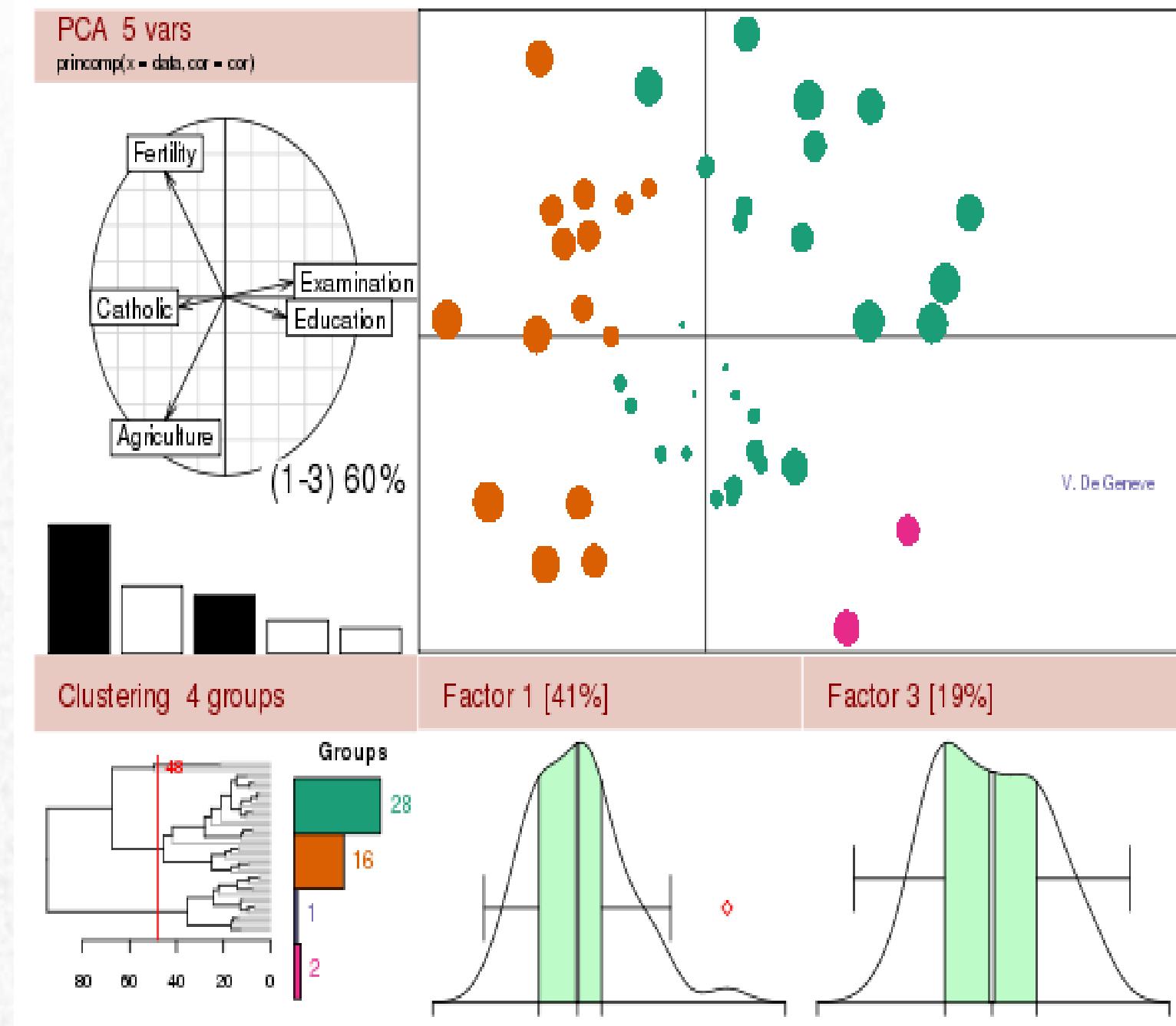
New

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

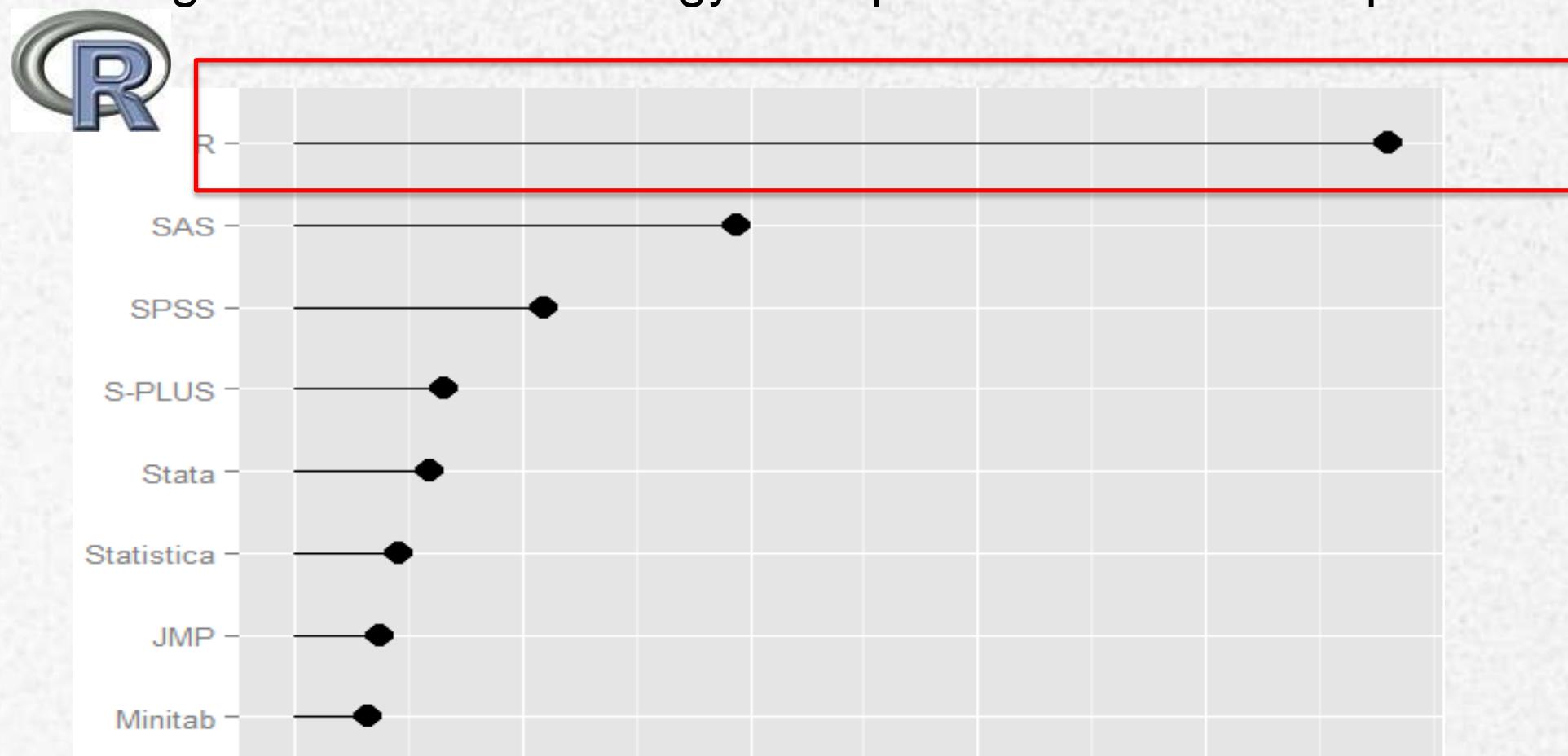


Частично благодаря появлению концепции *Big Data*, бизнес-анализ(BI) остается быстро растущим рынком Одновременно с ростом рынка BI постоянно увеличиваются инвестиции в предиктивную аналитику; **R является не только хорошим готовым инструментом, но и идеальной средой для исследований в области углубленной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

Gartner.

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

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

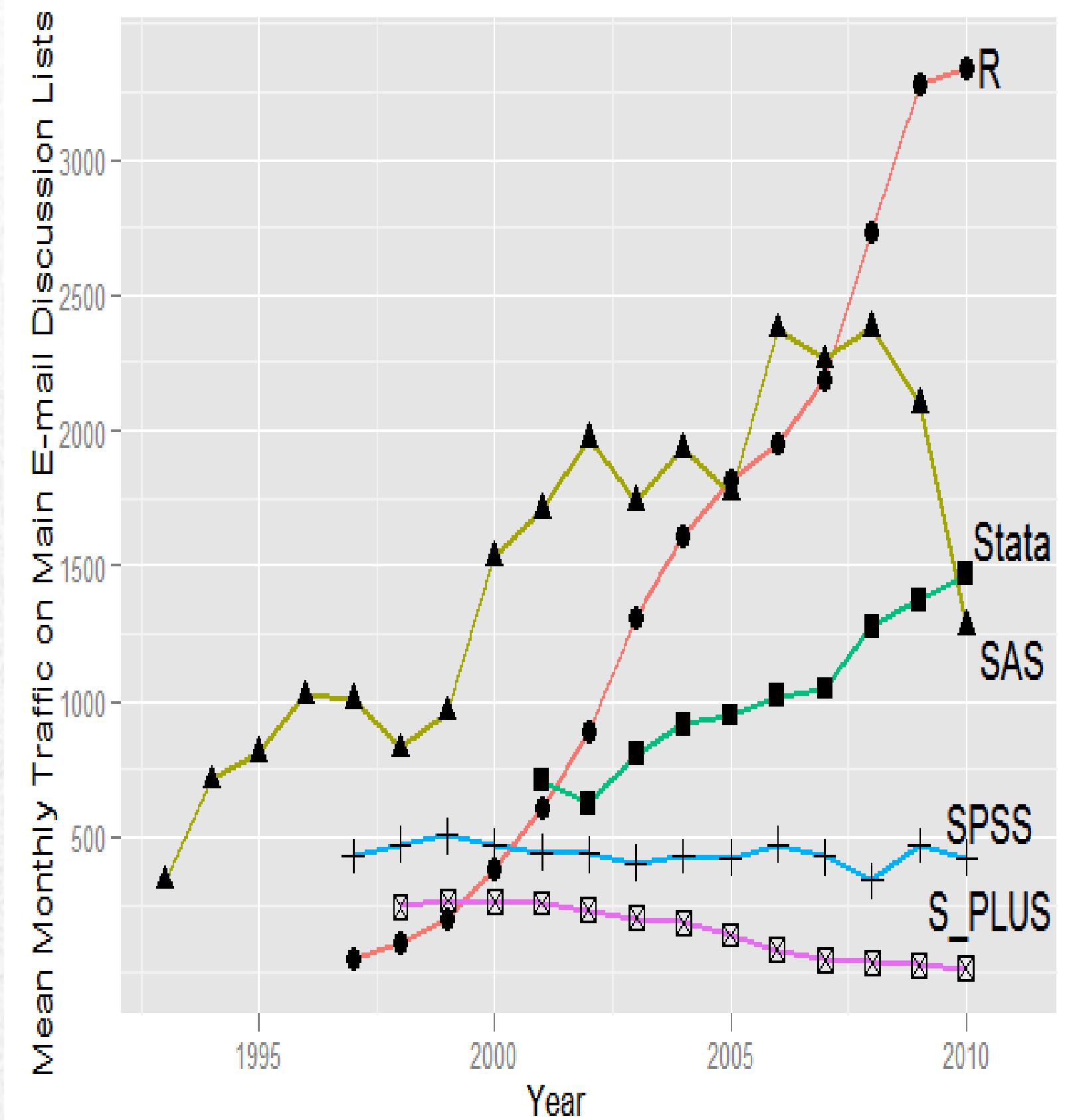
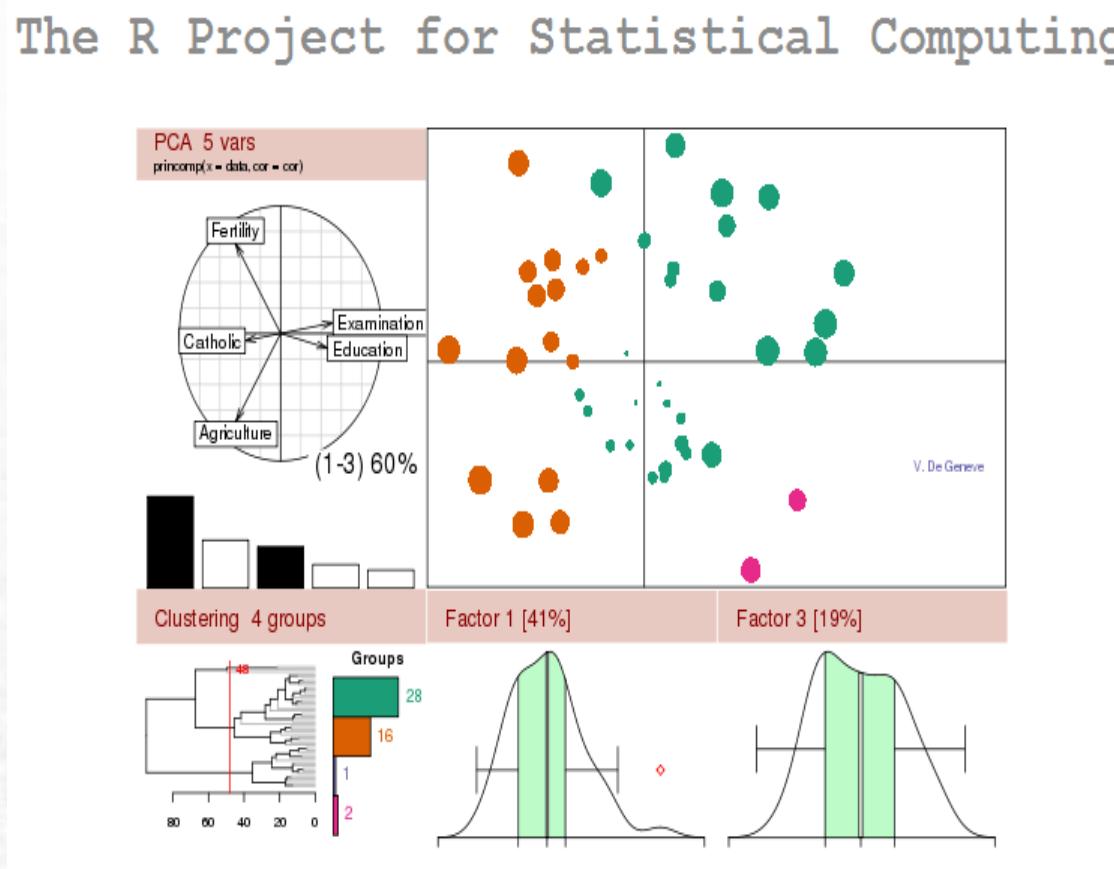


Кол-во f web site линков, которые указывают на основной сайт инструментальной среды March 19, 2011.

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

ORACLE

Растущая популярность проекта R



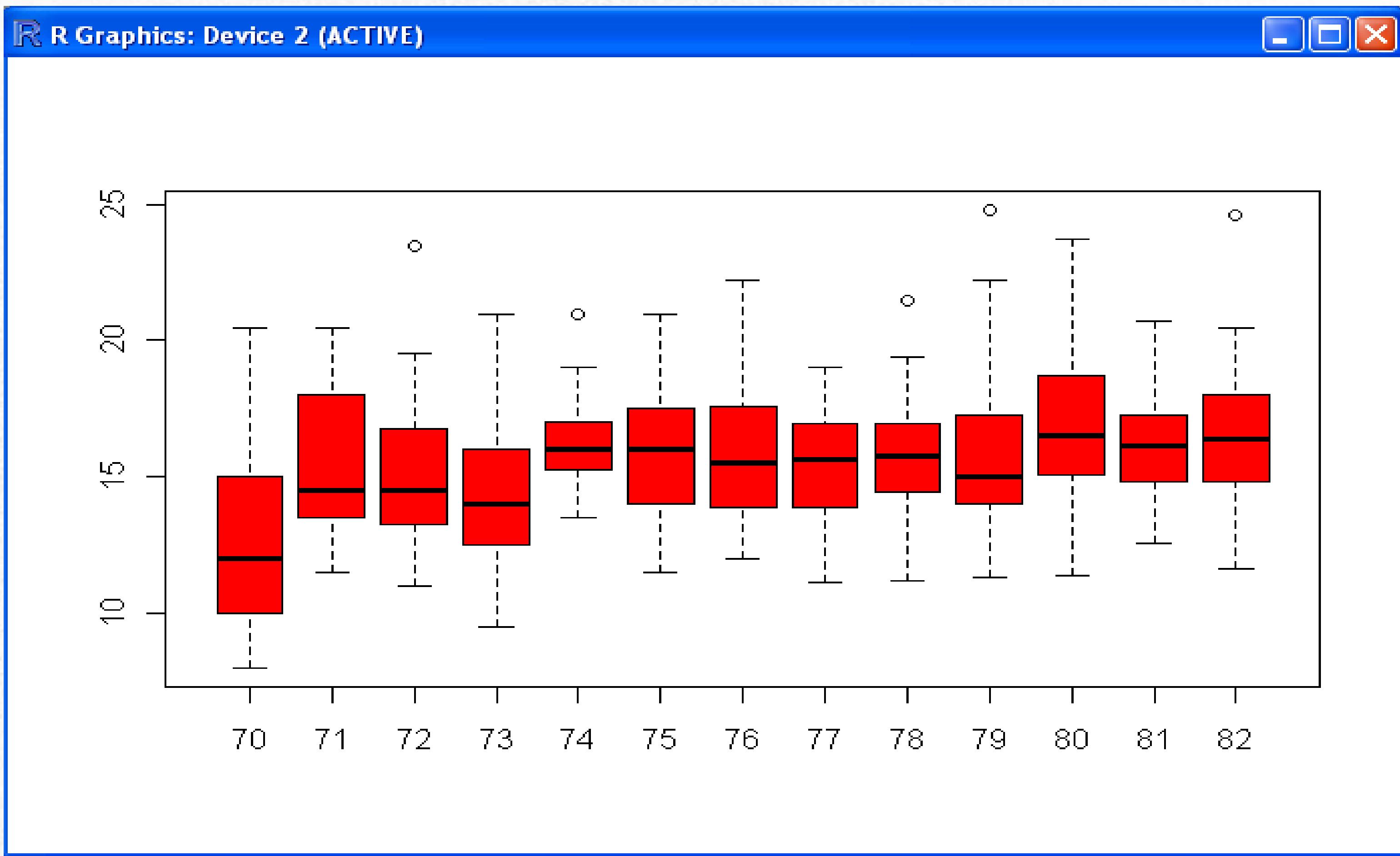
- Благодаря быстрому развитию и большому интересу со стороны специалистов R заслужил репутацию нового стандарта в области статистического программного обеспечения

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

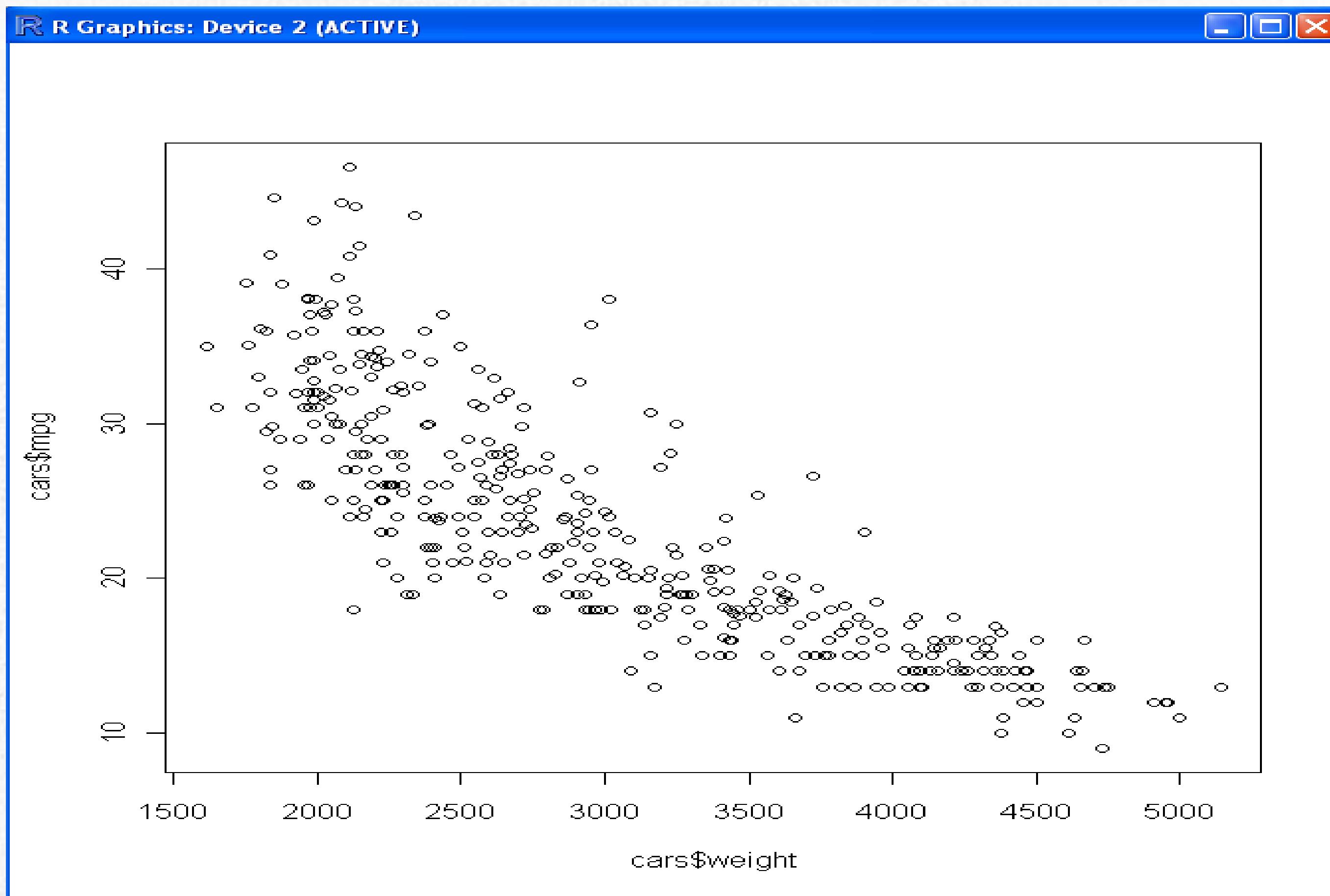
R Graphics

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



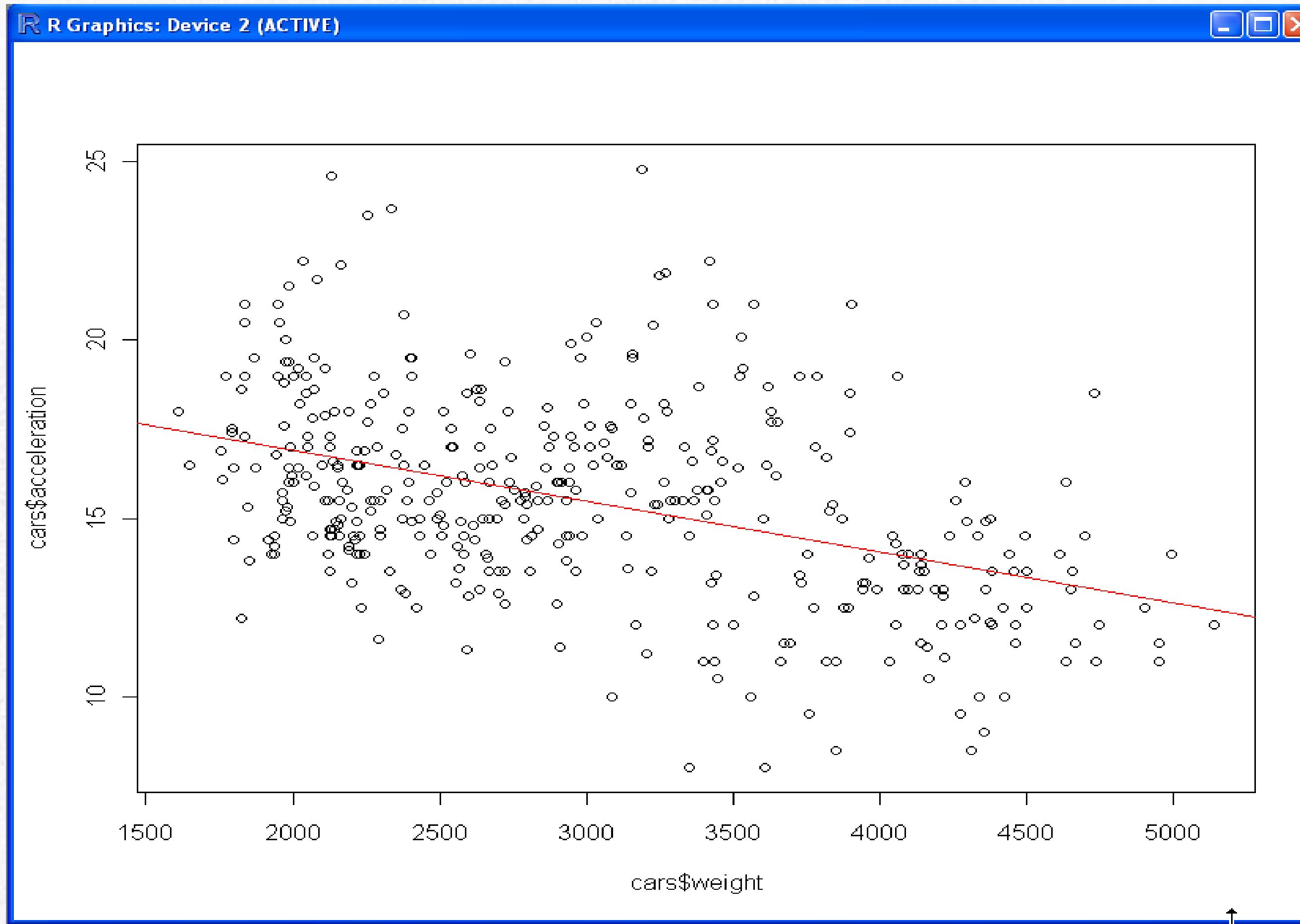
R Graphics

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

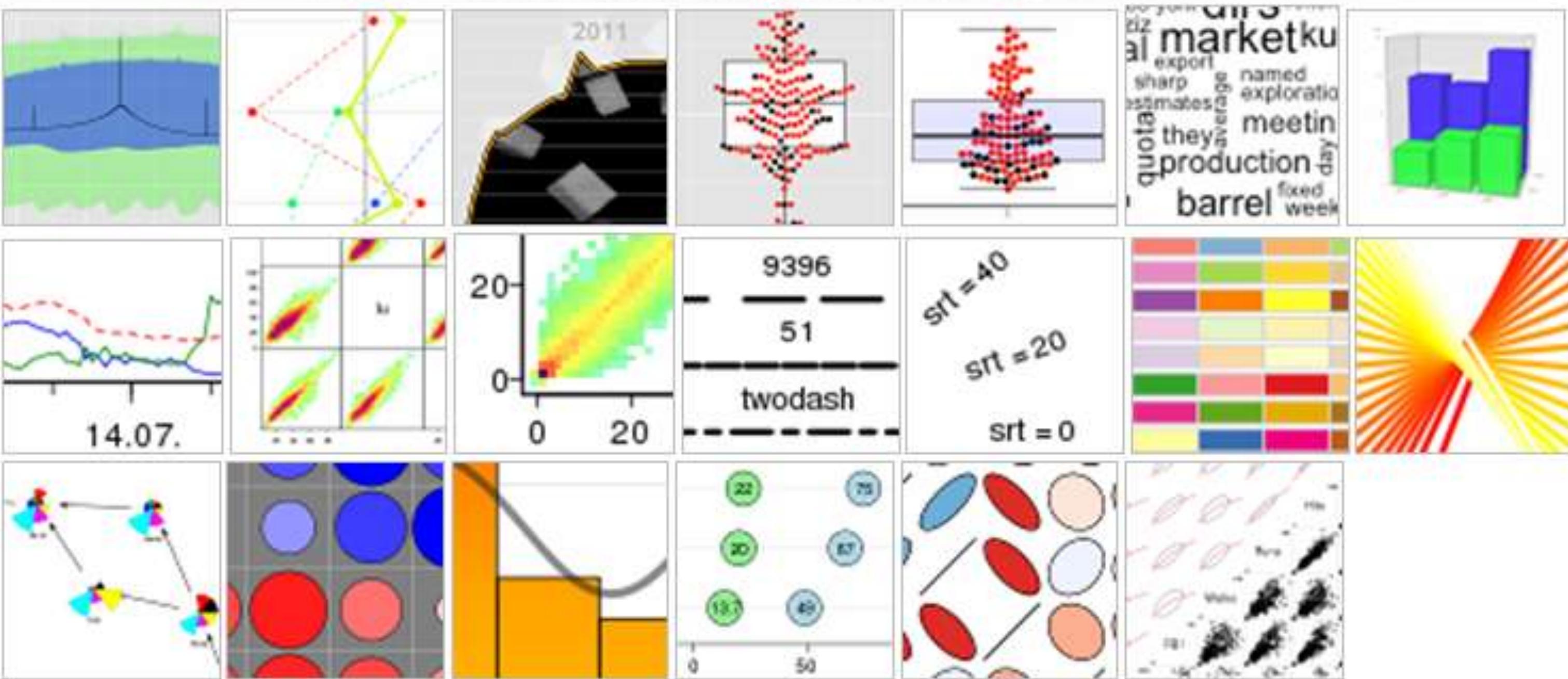


R Graphics

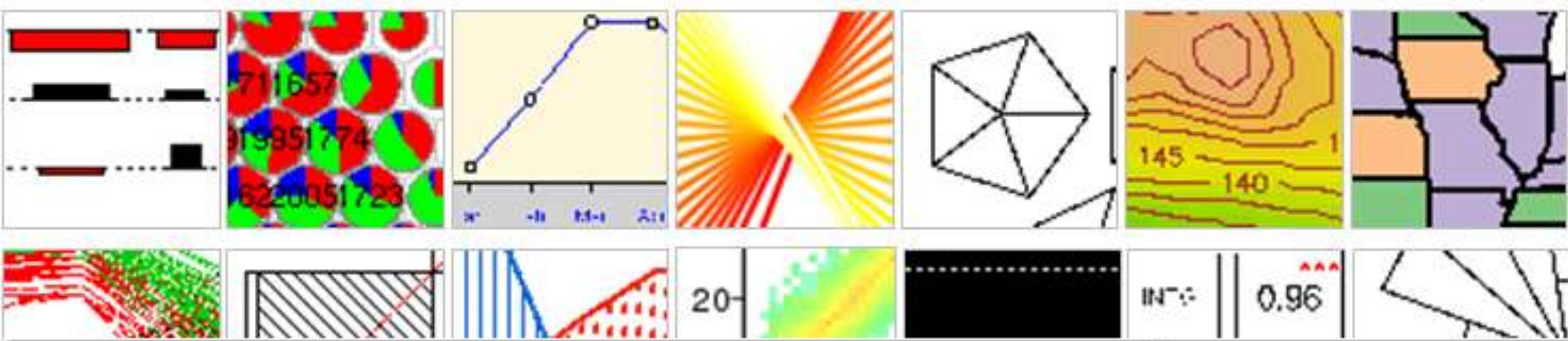
```
R> abline(coef(lm(acceleration ~ weight, cars)), col = "red")
```



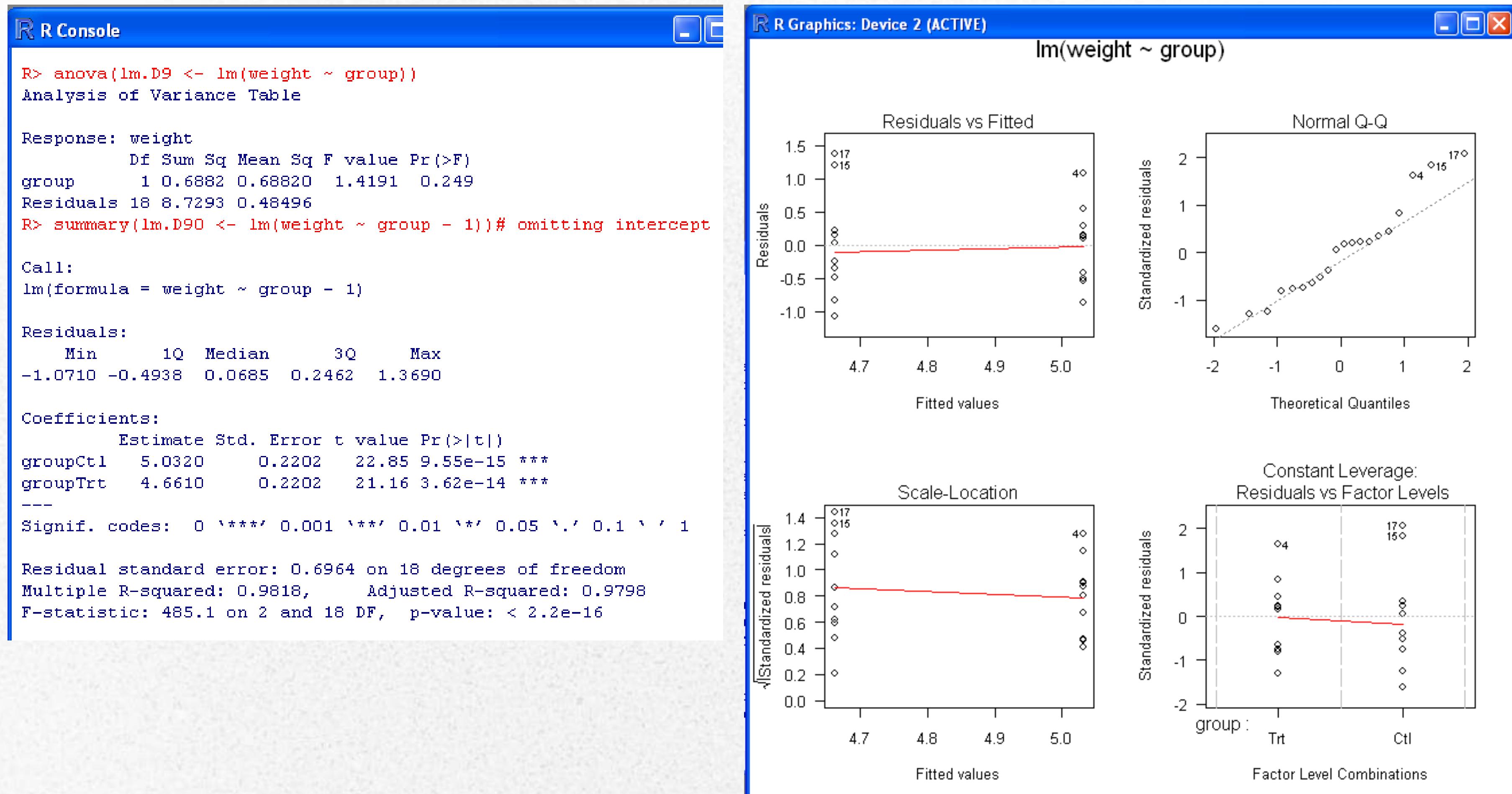
Графики и диаграммы в R



» Random entries

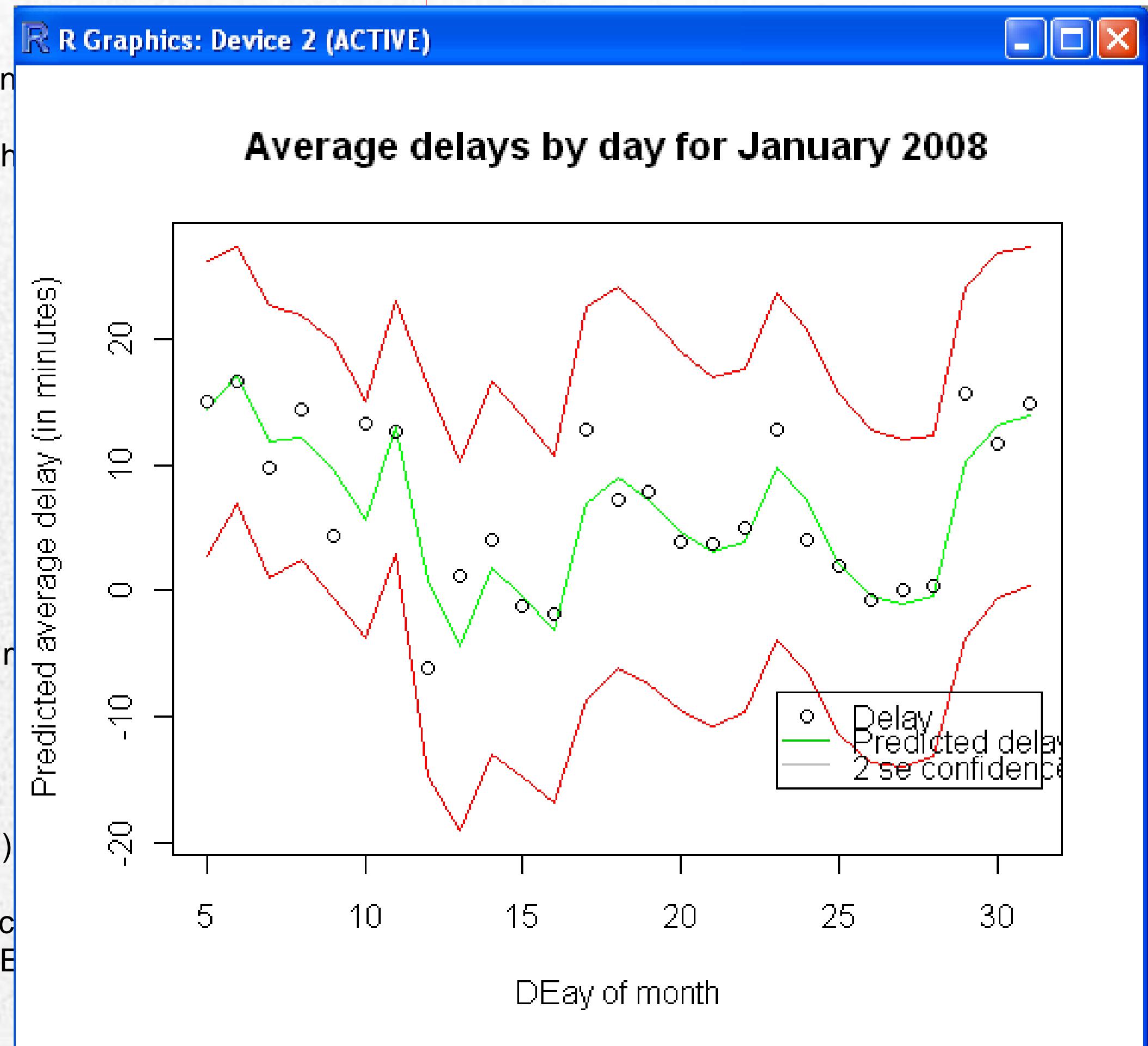


Линейное моделирование

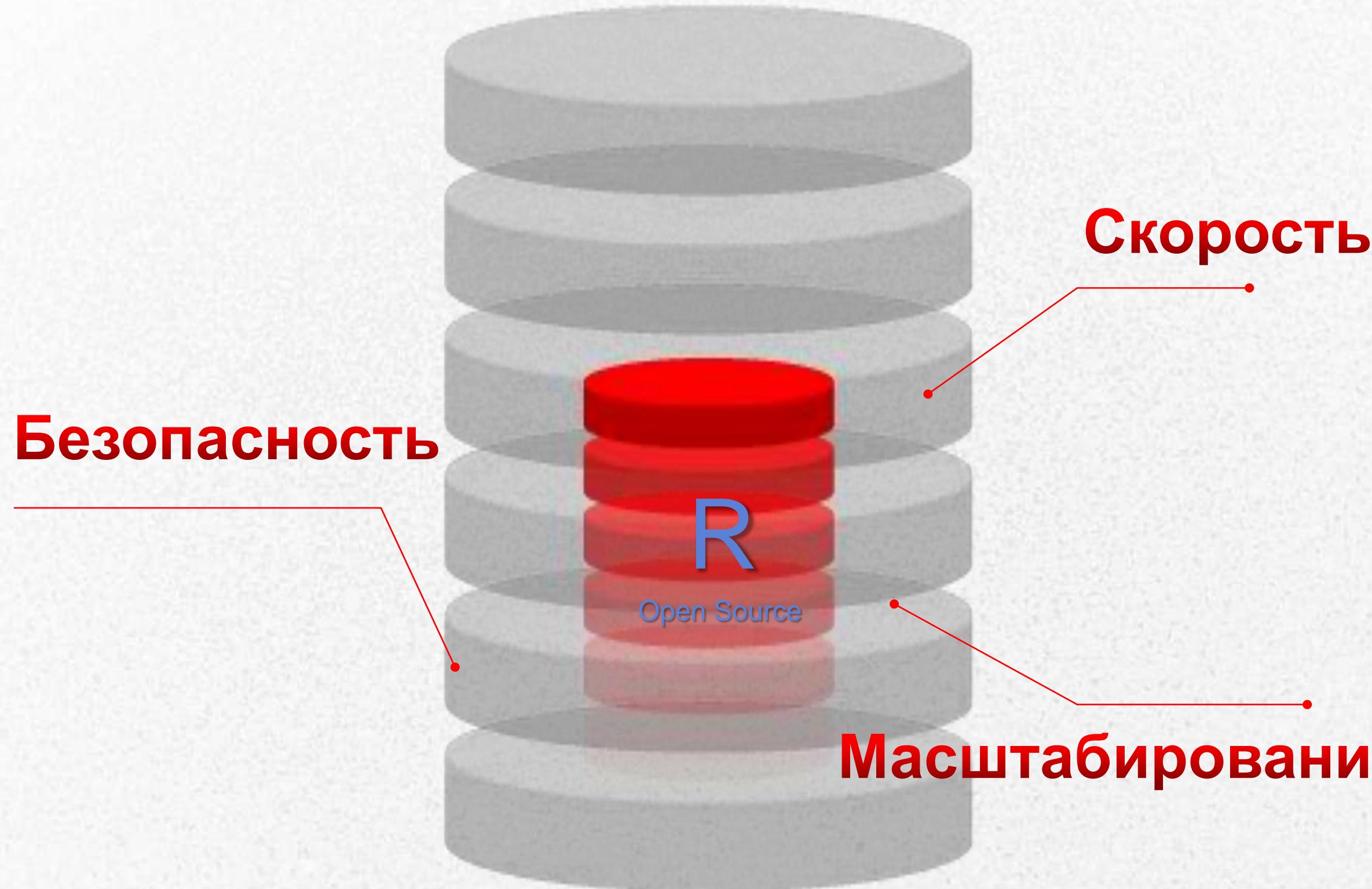


ARIMA --прогнозирование

```
year200801 <- ONTIME_S[(ONTIME_S$YEAR==2008)& (ONTIME_S$MONTH==1),]  
y <- ore.pull(year200801)  
gc()  
delays <- tapply(y$ARRDELAY, y$DAYOFMONTH, mean, na.rm=TRUE)  
delays <- ts(delays, start=1, end=31, frequency=1)  
# Create a Kalman filter with the first 5 delays and predict the rest  
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="DEay 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

ORACLE

Спасибо за внимание!

