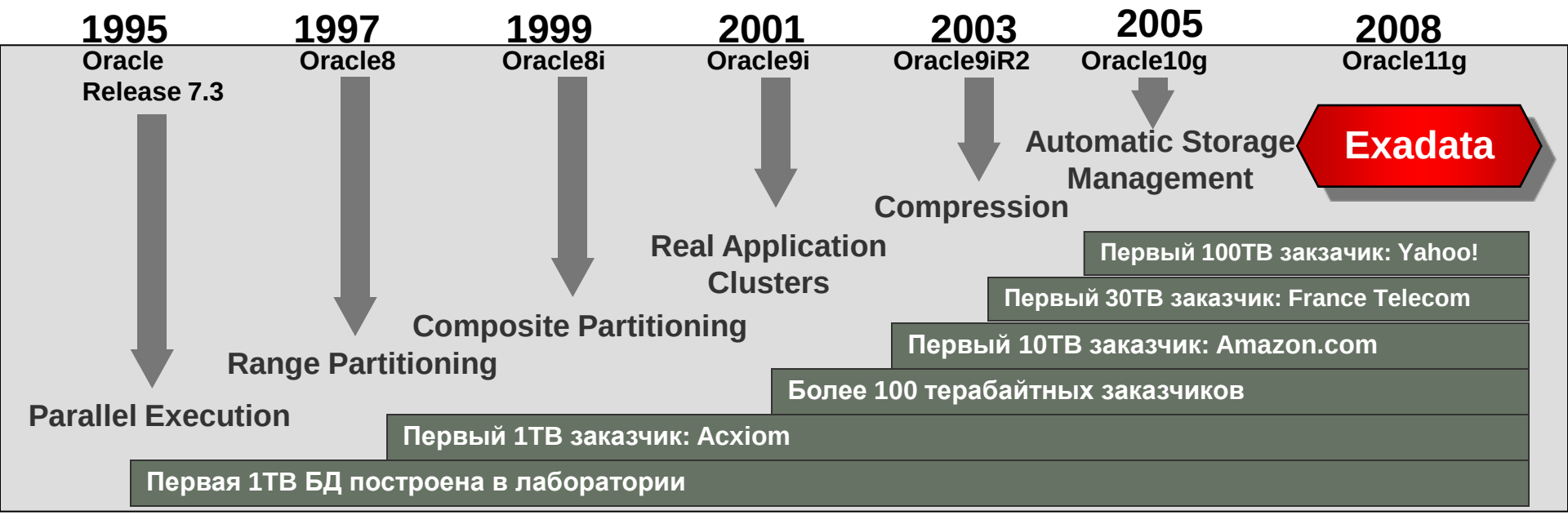
ORACLE

Oracle BIG DATA Appliance и другие технологии для обработки Больших Данных

Андрей Пивоваров
Менеджер по технологическому консалтингу

Поддержка технологиями Oracle Database сверхбольших хранилищ данных



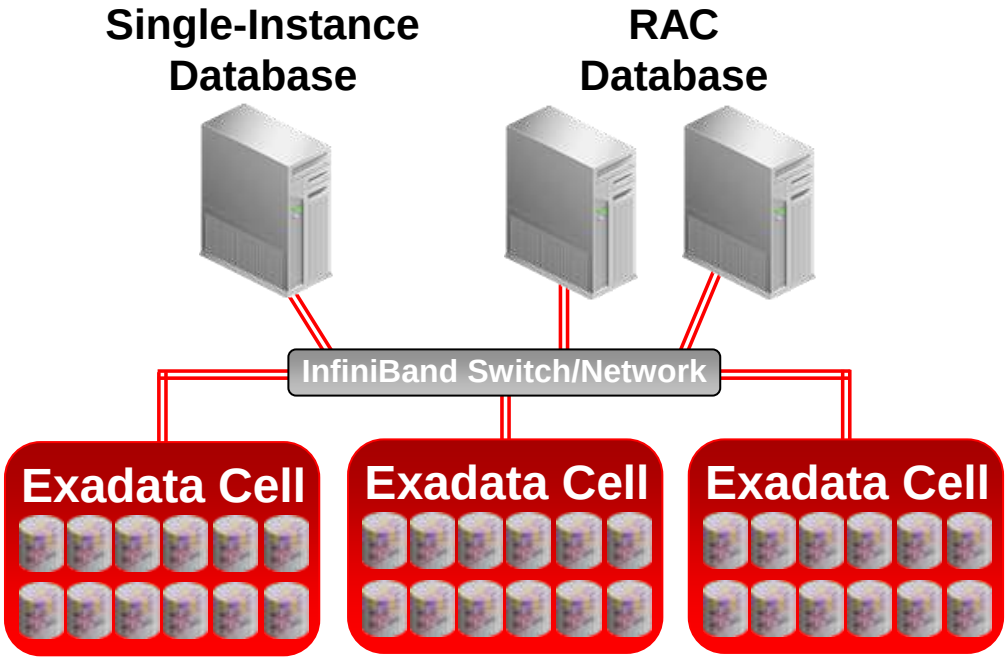
Что такое Oracle Exadata?

- Стратегическое аппаратно-программное решение Oracle для
  - OLTP
  - Хранилищ данных
  - Смешанных нагрузок
  - Консолидации приложений на базе Oracle Database
- Построено на основе:
  - Oracle Database
    - Т.е. все приложения, работающие на Oracle, могут работать на Exadata
  - Oracle Hardware (ex-Sun)

Exadata на аппаратном уровне

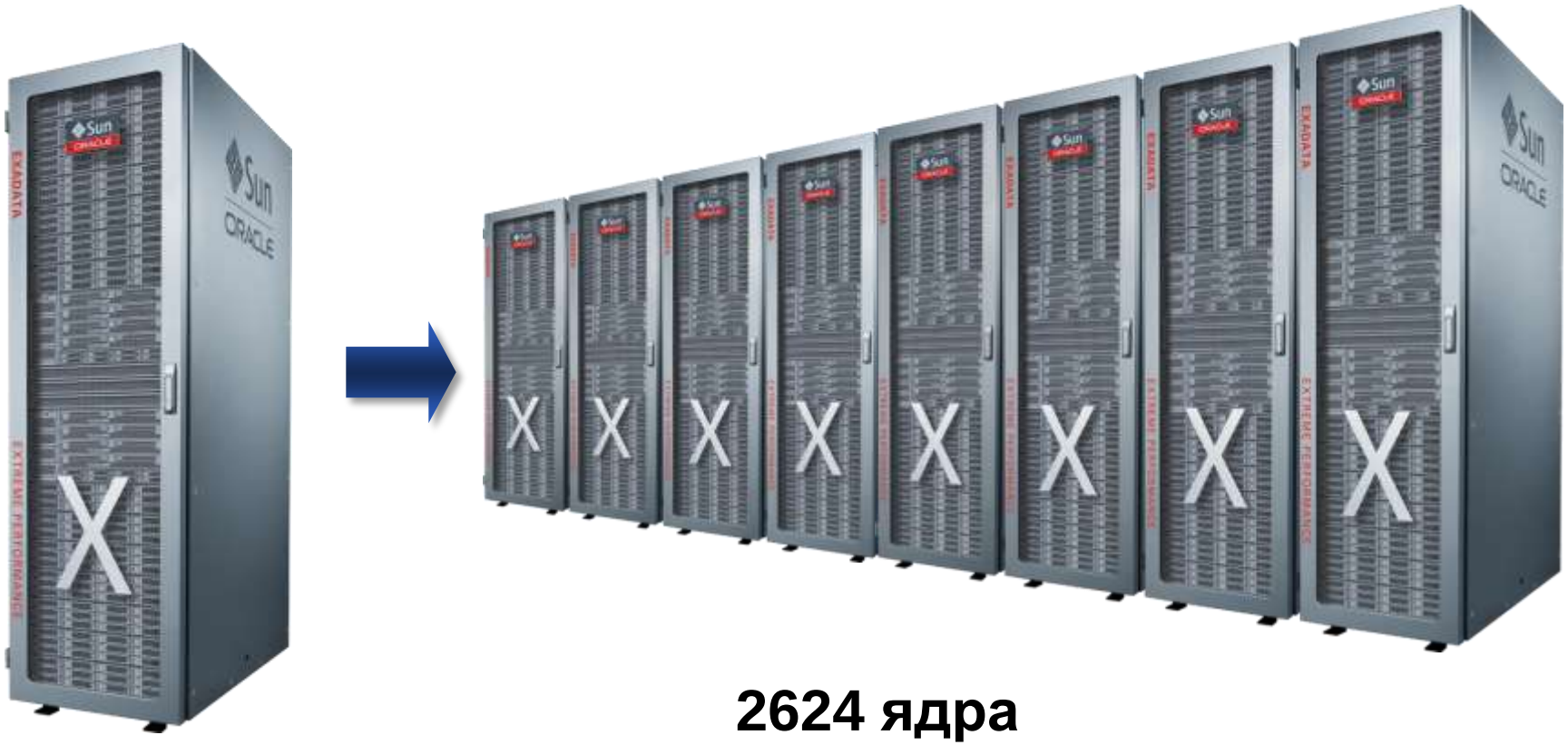
- Быстрая дисковая подсистема
- Использование 40Gb/s Infiniband
- Использование FLASH карт (до 5.3TB)
- Много RAM (до 4TB)
- Много процессорных ядер (до 160+168)

Конфигурация системы с Exadata



- Каждая ячейка Exadata – самостоятельный сервер с установленными дисками и ПО Exadata
- Данные «размазаны» между многими ячейками Exadata
- Нет ограничения на количество ячеек в системе
- Ячейки работают в режиме MPP

Масштабируемость до 8 шкафов



2624 ядра

4 петабайта несжатых данных

Big Data

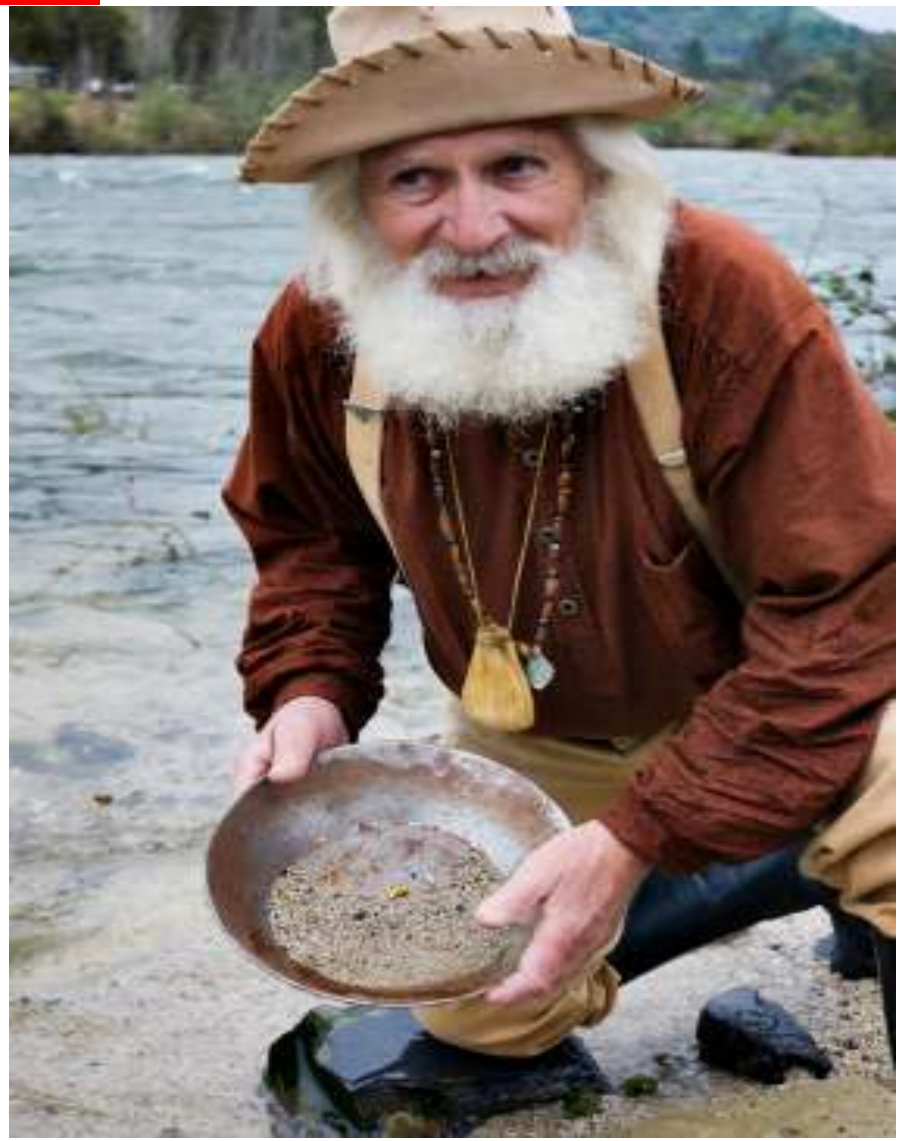
Что такое Big Data?

Что такое Big Data?

- Данные, которые могут очень быстро накапливаться, при этом, обычно (но не всегда) информационная плотность их низкая.
  - Логи, данные телеметрии, датчиков, полуструктурированные данные и неструктурированные данные, записи в социальных сетях, вебсайты и т.д.

- Данные, которые хранить очень дорого
  - Важно! Часто компании держат в хранилище данные только за последние несколько месяцев или год не потому, что им больше не нужно, а потому, что это дорого

Особенности обработки больших данных



- Большие объемы данных нужно хранить желательно дешевле, чем в традиционных СУБД.
- Могут не использоваться многие возможности РСУБД
- Для того, чтобы найти крупицу ценной информации, нужно переработать огромный объем данных
- При этом экстремальная производительность может быть не нужна

Общие принципы построения Big Data систем

- Построены из большого количества (до десятков тысяч) узлов, на основе относительно дешевого оборудования
- Каждый узел является сервером и хранения и обработки данных
- Обработка данных ведется в массивно-параллельном режиме
  - MapReduce
- Данные хранятся в нескольких копиях (обычно в трех) и отказ узла или двух не ведет к потере данных
- Система практически неограниченно масштабируется

Современные технологии обработки Big Data

NoSQL DB

- Not Only SQL – СУБД, часто построенные по принципу «ключ-значение»
- Быстрая запись и выборка по ключу

MapReduce

- Фреймворк для распределенных вычислений и обработки данных на тысячах узлах
- Можно использовать через SQL-подобные инструменты

Hadoop

- Лидирующая реализация MapReduce (проект Apache)
- Масштабируемая пакетная обработка
- Большое количество существующих наработок

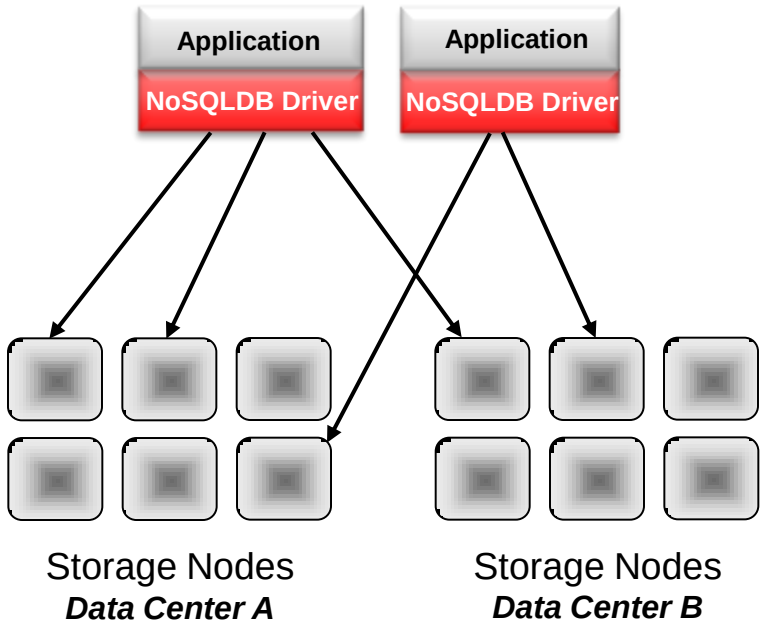
HDFS

- Hadoop Distributed File System
- Для построения дешевых, распределенных, масштабируемых хранилищ

Новый продукт - Oracle NoSQL DB

Распределенная, масштабируемая key-value база данных

- Простая модель данных
  - Пара Key-value с подходом major+sub-key
  - Операции read/insert/update/delete
- Масштабируемость
  - Динамическое партиционирование и перераспределение
  - Оптимизированный доступ к данным
- Высокая доступность
  - Одна или более реплик
  - Катастрофоустойчивость засчет разнесения реплик
  - Устойчивость к отказу мастера
  - Нет одной точки отказа
- Прозрачная балансировка нагрузки
  - Чтение с мастера или реплики
  - Драйвер знает о сетевой топологии и временах задержки



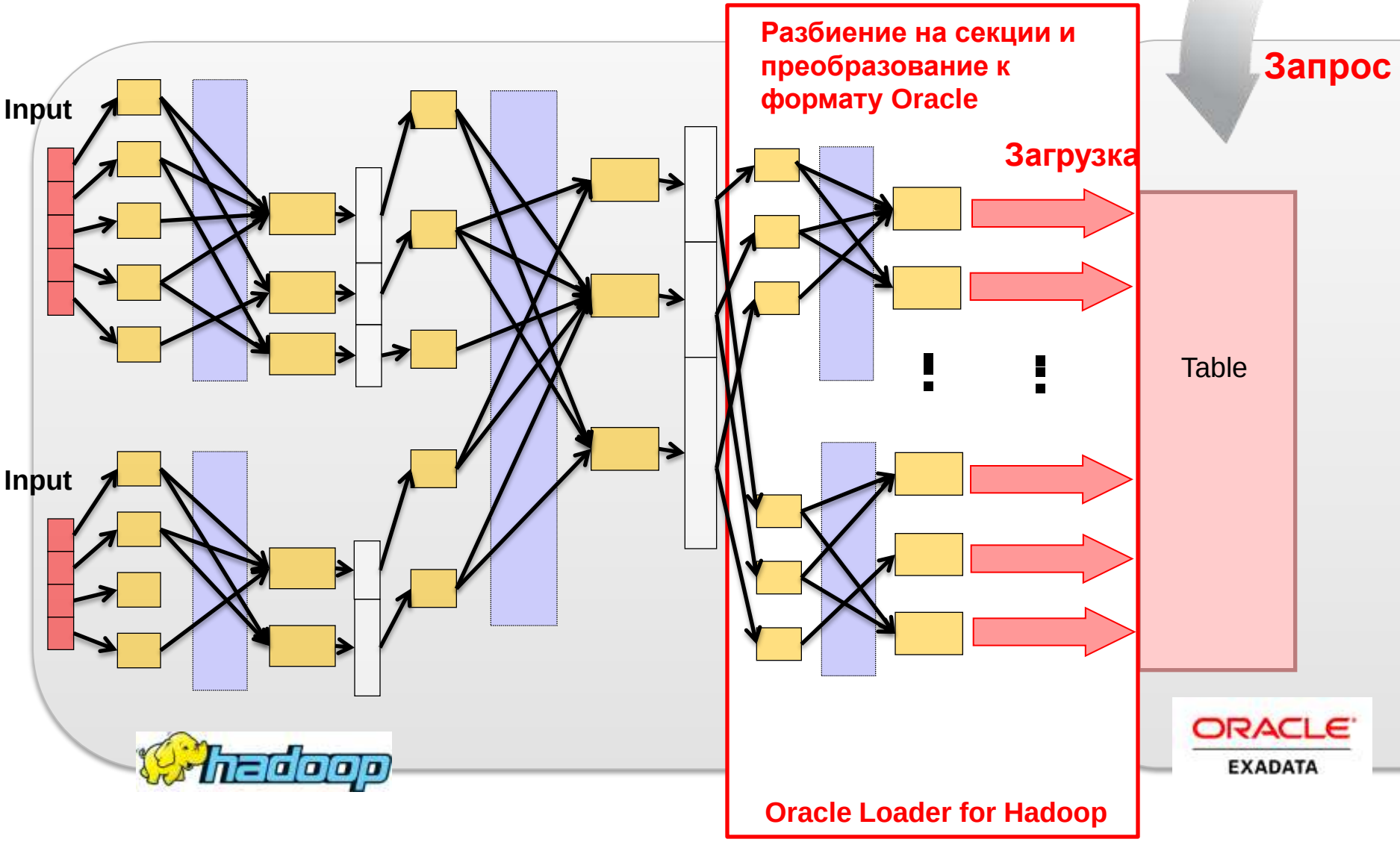
Что такое Hadoop?

- Apache Hadoop - это распределенная вычислительная архитектура:
  - Open source (проект Apache Software Foundation)
  - Включает в себя распределенную файловую систему HDFS
  - Служит для пакетной обработки и ETL
  - Обрабатывает данные в массивно-параллельном режиме (MapReduce)
  - Работает на очень больших кластерах (от сотен до тысяч узлов) на дешевом «железе»
  - Автоматически обрабатывает отказ узлов, и перераспределение данных
  - Используется во многих известных проектах
    - Yahoo – более 10000 узлов на Linux, для обработки поиска
    - Кроме этого – Apple, Twitter, LinkedIn, Amazon, Last.fm и др.
    - Facebook – более 30PB на Hadoop

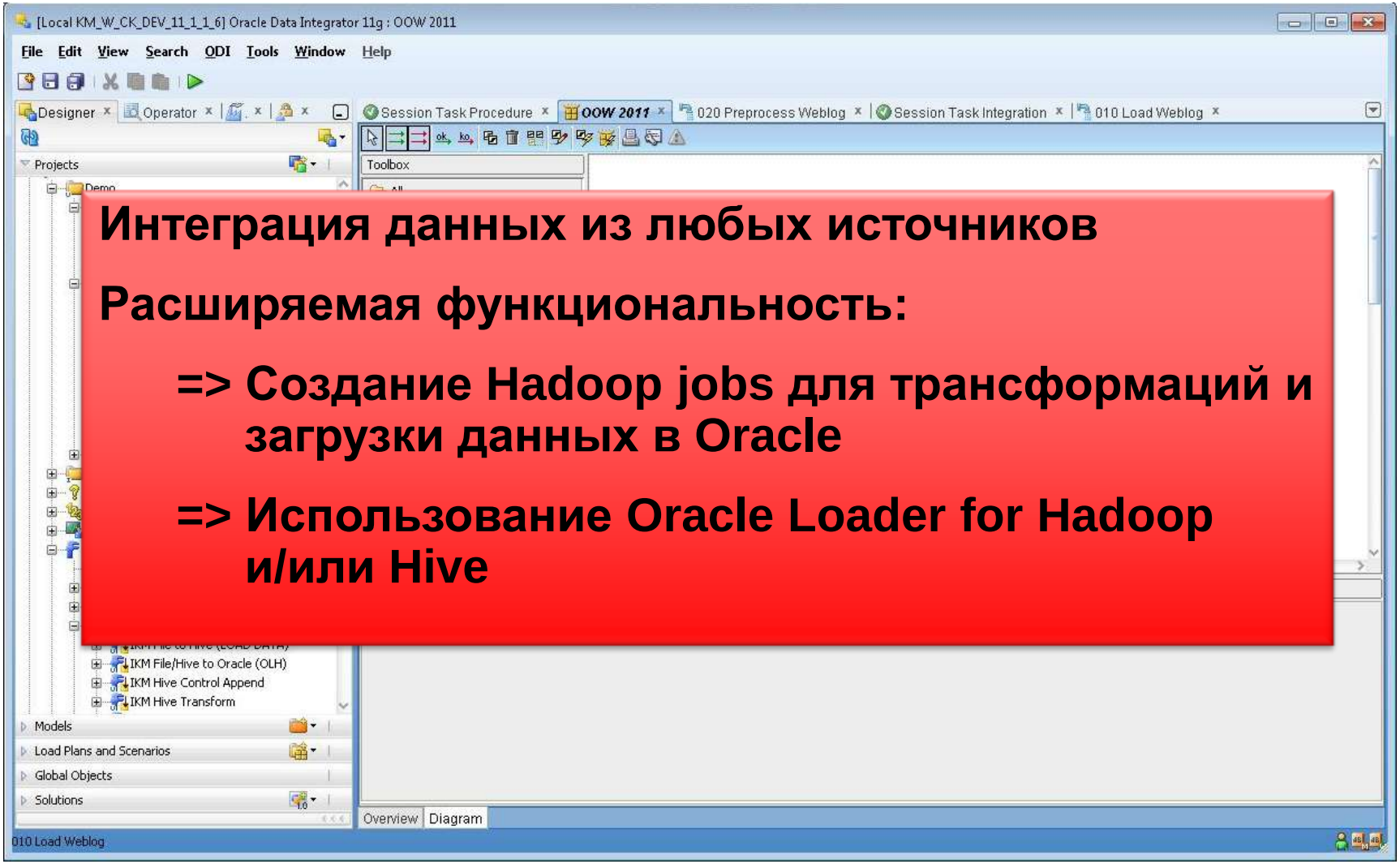
Проекты, использующие Hadoop

- Apache Hive
  - Инфраструктура, реализующая хранилище данных на Hadoop
  - Разработана в Facebook
  - Есть SQL-подобный язык HiveQL
  - Не замена Oracle DB
- Apache HBase
  - Нереляционная СУБД, позволяющая хранить и обрабатывать огромные объемы разреженных данных
  - Данные хранятся в структурах: индекс строки, индекс колонки, временная метка
  - Используется компрессия
  - Рассчитана на хранения петабайтов данных

Oracle Loader for Hadoop



Oracle Data Integrator



Интеграция данных из любых источников

Расширяемая функциональность:

=> Создание Hadoop jobs для трансформаций и загрузки данных в Oracle

=> Использование Oracle Loader for Hadoop и/или Hive

Big Data

Что такое Big Data Appliance?

Предпосылки для Big Data Appliance

- Oracle NoSQL DB, Hadoop, доступны к скачиванию и использованию
- Однако, даже несмотря на то, что Hadoop – Open Source, настройка и конфигурирование кластера из десятков узлов требует высокой квалификации

- Для того, чтобы помочь заказчикам использовать преимущества работы с Big Data, Oracle создает оптимизированный комплекс Big Data Appliance

Oracle Big Data Appliance Hardware

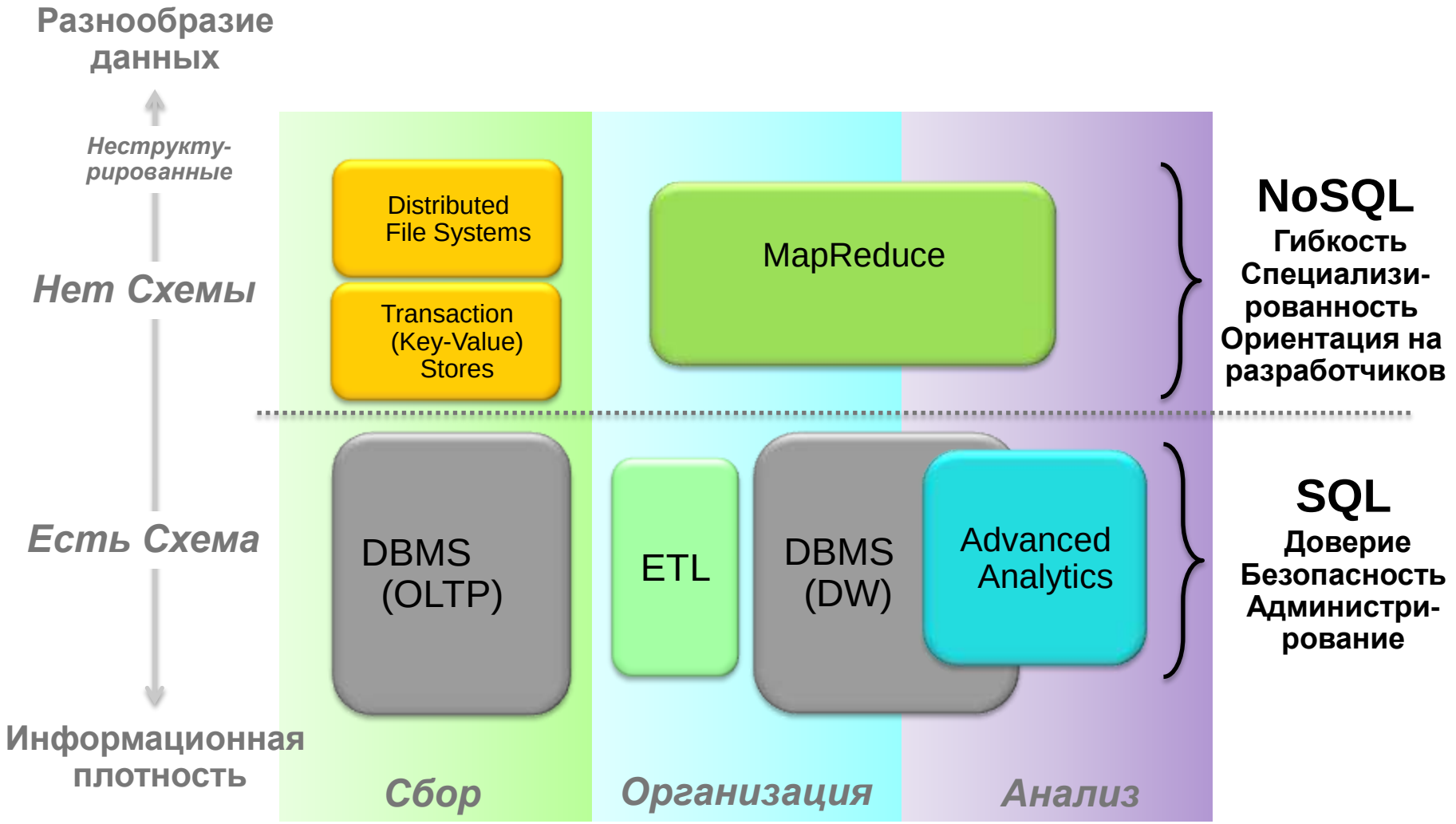
- 18 Sun X4270 M2 Servers
  - 48 GB memory per node = 864 GB memory
  - 12 Intel cores per node = 216 cores
  - 36 TB storage per node = 648 TB storage
- 40 Gb p/sec InfiniBand
- 10 Gb p/sec Ethernet

Oracle Big Data Appliance Software

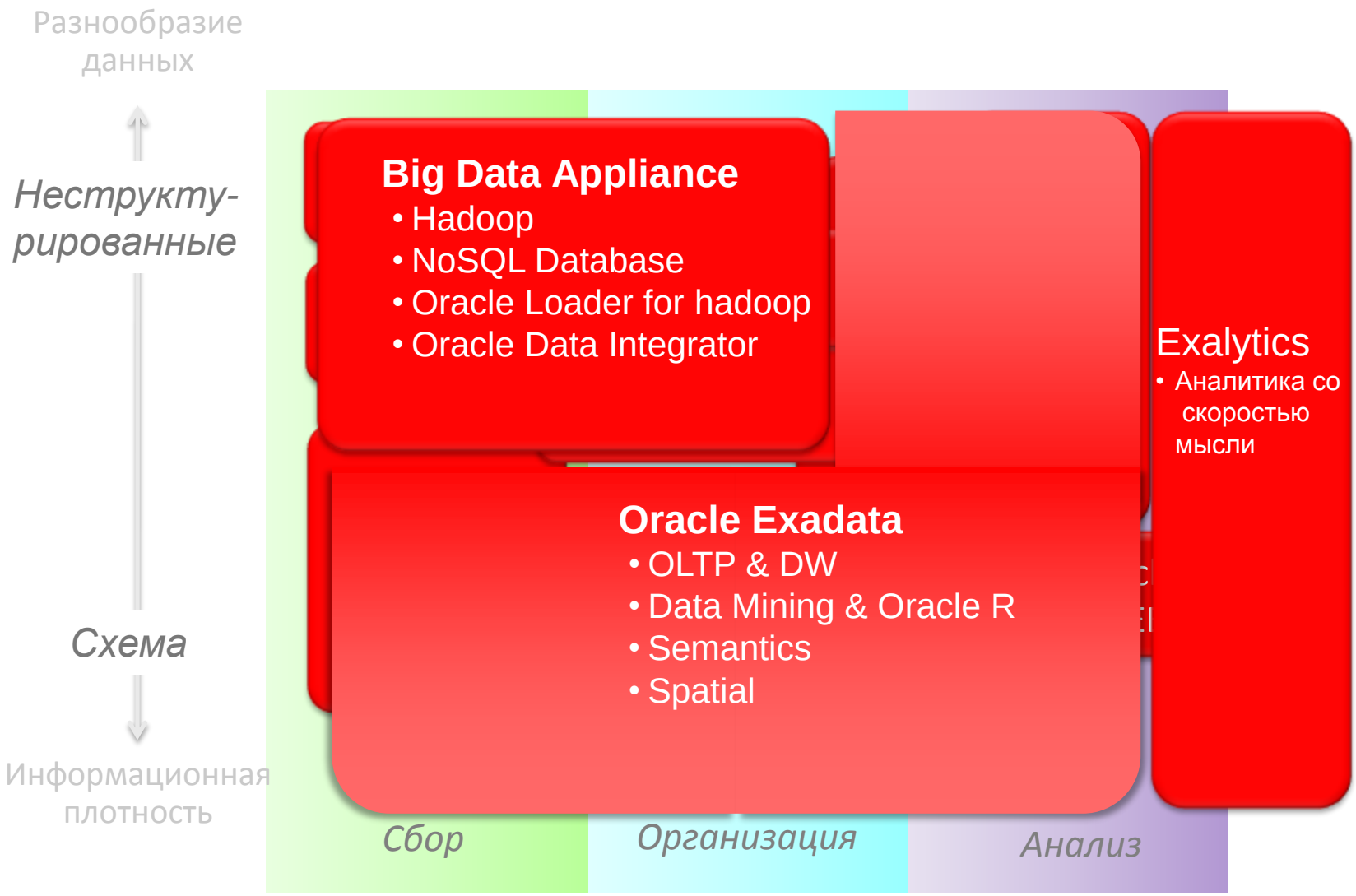
- Oracle Linux 5.6
- Java Hotspot VM
- Cloudera Hadoop Distribution

  Hadoop Core, HDFS, Hive, HBase, Zookeeper, Oozie, Mahout, Sqoop, Administration Tools
- R Distribution
- Oracle Loader for Hadoop
- Oracle NoSQL Database
- Oracle Adapters for Hadoop:
  - Oracle R Connector for Hadoop
  - Oracle SQL to HDFS Connector
  - Oracle Data Integrator Application Adapter for Hadoop
  - Oracle Loader for Hadoop

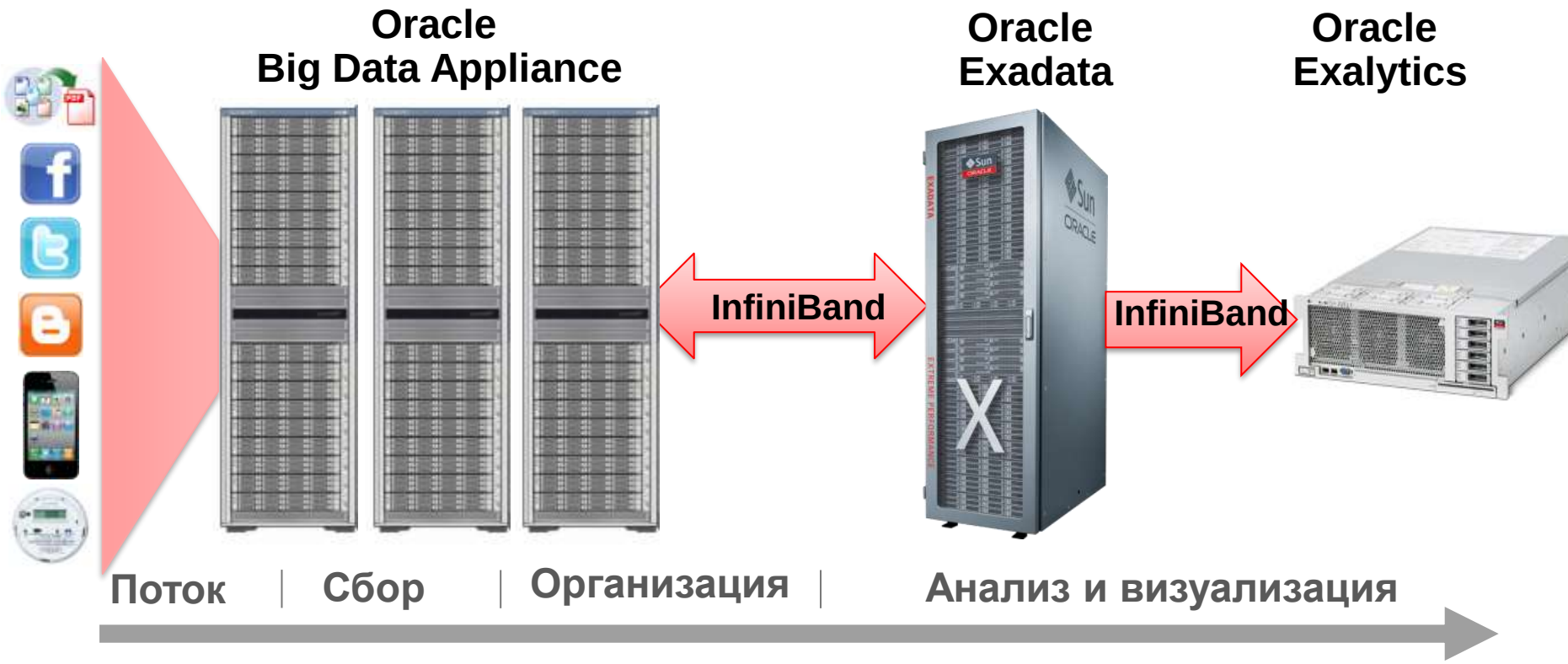
Подходы к обработке данных



Решения Oracle



Возможная архитектура



Типовые примеры использования Big Data

- Социальные сети(LinkedIn, Facebook, Digg, Google+, etc.)
- Персонализация (Amazon, Ebay, Yahoo, etc.)
- Обслуживание в веб (Apple, Cisco, AT&T, HP, Motorola, Nokia)
  - Обслуживание клиентов
  - Отслеживание устройств
- Банки и финансы (JP Morgan, Wells Fargo)
  - Выявление мошенничеств
- Поиск по документам (Thomson Reuters, exLibris)
- Безопасность
  - Анализ логов, видео, аудио
- Наука
  - Геофизика (Halliburton)
  - Биология и медицина

Big Data

Пример анализа данных в Twitter

Что такое Twitter?

- Сервис микроблогов
  - Каждое сообщение не может быть длиннее 140 символов (как в SMS)
- Большинство сообщений – публичные
- Поддерживается поиск по сообщениям в реальном времени
- Пользователи часто описывают свой опыт по работе с какими-то организациями в Twitter
  - Часто эмоционально

twitter @delta Have an account? Sign in

Results for @delta

Tweets · Top ▾ Refine results »

rdsanchezjr Ruben D. Sanchez Jr
Nice!...:-) @Delta Air Lines posts Q4 profit of $425 million sbne.ws/r/9Tb9
1 minute ago

donalderyan Donald Ryan
I never get tired of the "Your Upgrade is Now Confirmed" emails. Thanks @Delta
1 minute ago

BiscoffCookies Biscoff
A great mention of Biscoff cookies as a favorite snack served on @Delta Air Lines. jaunted.com/story/2012/1/1...
7 minutes ago

mnicolewilliams Nicole Williams
Starting to fly lots and just crossed @delta off the list of airlines I'll use. $50 for an earlier flight that has open seats?! No thanks.
10 minutes ago

andriiwarana Andrés Arana
Where do the bags go after they pass through those black rubber flaps at the airport? @Delta give us the answer: youtu.be/ocbxS5aWUSo?hd...
11 minutes ago

mollyoehmichen Molly Oehmichen
Got my first "Your Upgrade is Now Confirmed" email from @Delta for #hollanding next week. Let the plane-drinking begin!
17 minutes ago

NixFred Fred Nix

Follow "@delta" and more on Twitter

Twitter delivers instant updates on what's happening around the world. Sign up today and follow your interests!

Sign up »

People results for @delta · view all

Delta Delta Follow
We're listening to your feedback and posting news, ti...

delta_goodrem Delta Goodrem Follow
I'm Fascinated by life the universe and everything so...

DeltaAssist Delta Assist Follow
We're listening around the clock, 7 days a week. We t...

TriDelta Delta Delta Delta Follow
Let us steadfastly love one another.

Popular images & videos

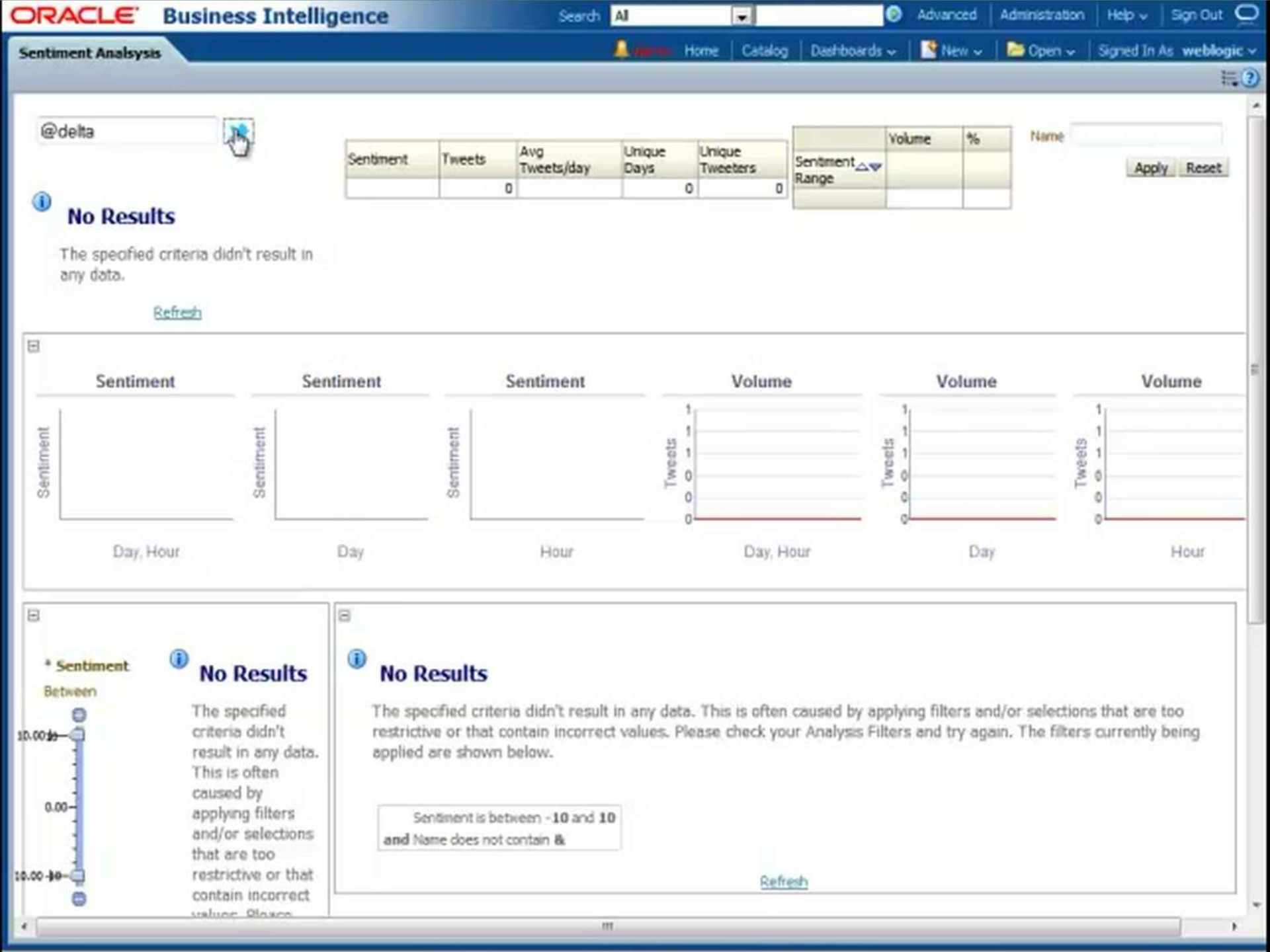
These results include media shared by people you don't follow.

Display media

Trends:

Oeste 2 x 3 São Paulo
#orgullosodesermadridista
#HighSchoolMemories

Done Internet | Protected Mode: On 100%


ORACLE Business Intelligence
Search All
Advanced
Administration
Help
Sign Out
Sentiment Analysis
Home
Catalog
Dashboards
New
Open
Signed In As weblogic
@delta
Sentiment
Tweets
Avg Tweets/day
Unique Days
Unique Tweeters
0
0
0
Sentiment Range
Volume
%
Name
Apply
Reset
No Results
The specified criteria didn't result in any data.
Refresh
Sentiment
Volume
Day, Hour
Day
Hour
* Sentiment
Between
10.00
0.00
-10.00
No Results
The specified criteria didn't result in any data. This is often caused by applying filters and/or selections that are too restrictive or that contain incorrect values.
No Results
The specified criteria didn't result in any data. This is often caused by applying filters and/or selections that are too restrictive or that contain incorrect values. Please check your Analysis Filters and try again. The filters currently being applied are shown below.
Sentiment is between -10 and 10
and Name does not contain &
Refresh

telligence

Search All

Advanced | Adm

Alerts! Home | Catalog | Dashboards | New

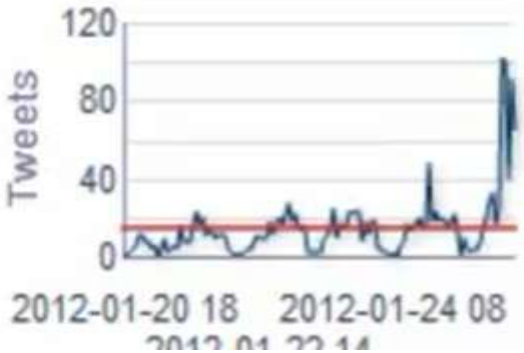
Sentiment	Tweets	Avg Tweets/day	Unique Days	Unique Tweeters
0.1	1,786	297	6	1,112

	Volume	%
Sentiment Range		
Pos. Sentiment	548	30.7%
Neutral Sentiment	807	45.2%
Neg. Sentiment	431	24.1%

Sentiment

Sentiment

Volume



Volume

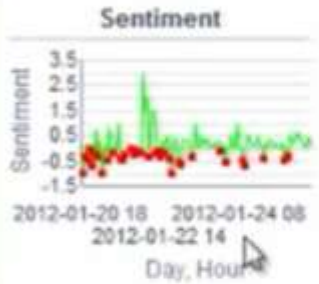
Current Search Term: **@delta**

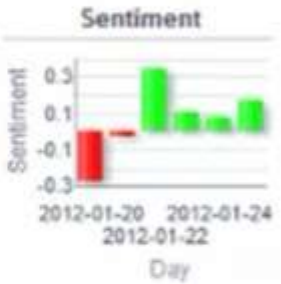
Sentiment	Tweets	Avg Tweets/day	Unique Days	Unique Tweeters
0.1	1,786	297	6	1,112

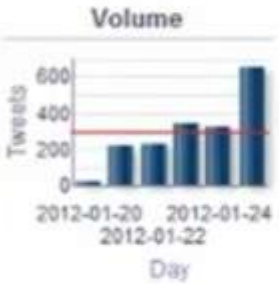
Sentiment Range	Volume	%
Pos. Sentiment	548	30.7%
Neutral Sentiment	807	45.2%
Neg. Sentiment	431	24.1%

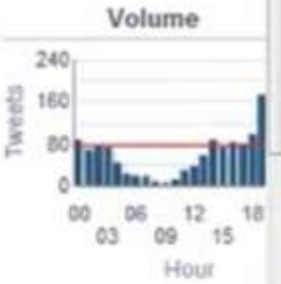
Name

Apply Reset









Top Tweeters

* Sentiment

Between

10.00 0.00 -10.00

Apply Reset

Name	Sentiment	Tweets
Y Aerospace	0.1	40
Pam A	-0.1	18
Mike Vizdos	0.1	12
US Air News	-0.2	12
tiffany helmly	-0.1	12
Kyle Luttrell	-1.0	11
Leif Skartland	-0.1	10
Fred Nix	0.3	9
Keith Flatley	-0.4	9
Ayuz M Mayangsari	0.0	8

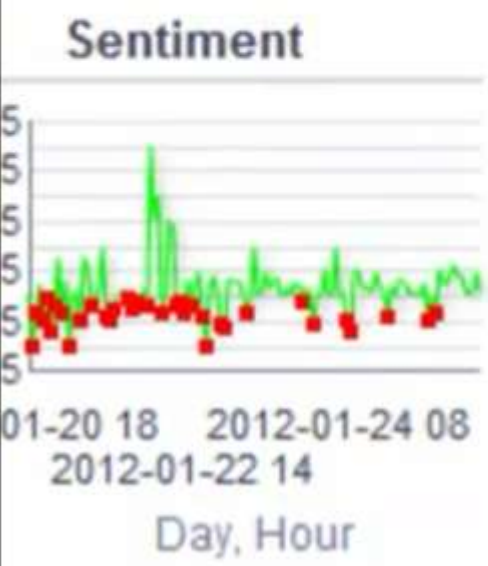
Detail all Tweet

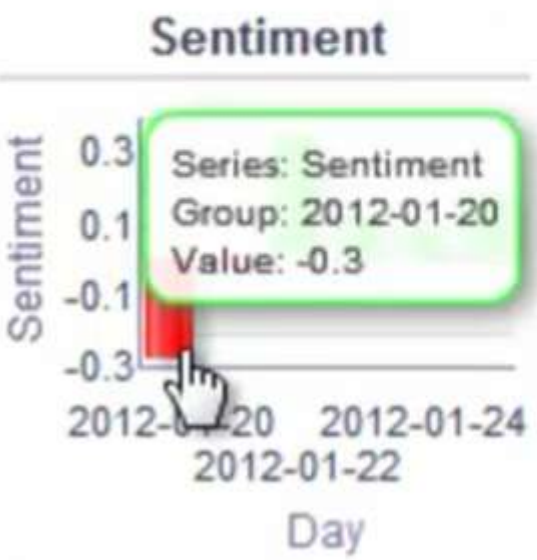
Name	Date	Sentiment	Tweet
BLatta	2012-01-25 23:29:33	2	>RT @BiscoffCookies: A great mention of Biscoff cookies as a favorite snack served on @Delta Air Lines. http://t.co/KdsPzpYq
John Maden	2012-01-25 23:28:20	-1	>Mrs. @katearoni? RT @donalderyan: I never get tired of the "Your Upgrade is Now Confirmed" emails. Thanks @Delta
Ruben D. Sanchez Jr	2012-01-25 23:28:06	1	>Nice!...:-) @Delta Air Lines posts Q4 profit of $425 million http://t.co/sB4CGXaP
Donald Ryan	2012-01-25 23:27:48	-1	>I never get tired of the "Your Upgrade is Now Confirmed" emails. Thanks @Delta
Ruben D. Sanchez Jr	2012-01-25 23:27:36	0	>@Delta Air Lines posts Q4 profit of $425 million

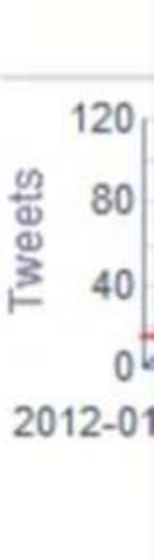
ORACLE

Current Search Term: **@delta**

0.1	1,786	297	6







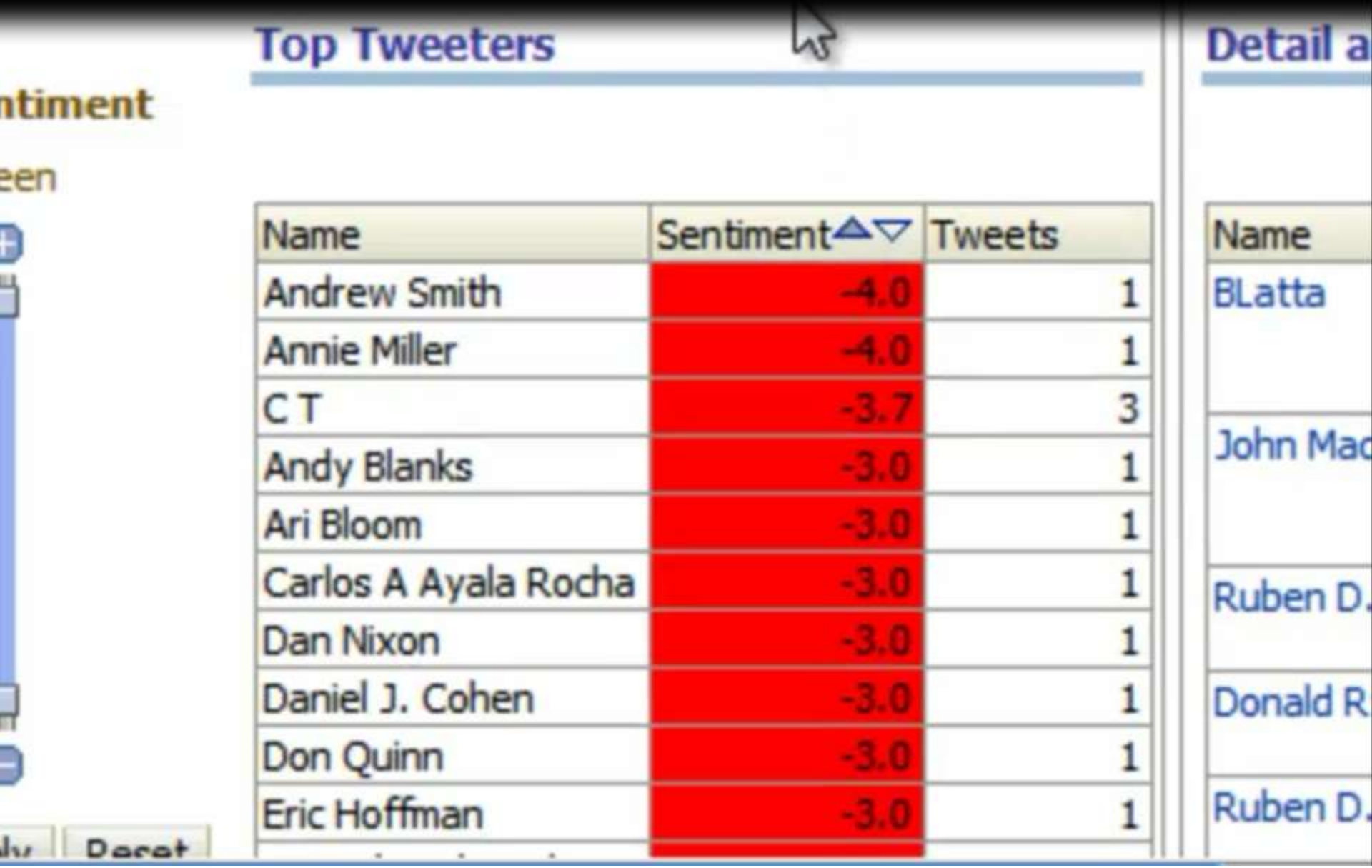
ORACLE

Top Tweeters

Name	Sentiment	Tweets
Y Aerospace	0.1	40
Pam A	-0.3	18
Mike Vizdos	0.1	12
US Air News	-0.2	12
tiffany helmly	-0.7	12
Kyle Luttrell	-1.0	11
Leif Skartland	-0.5	10
Fred Nix	0.3	9
Keith Flatley	-0.4	9
Ayuz M Mayangsari	0.0	8

Detail al

Name
BLatta
John Mad
Ruben D.
Donald Ry
Ruben D.


Top Tweeters
Detail a
ntiment
een
Name
Sentiment
Tweets
Andrew Smith -4.0 1
Annie Miller -4.0 1
C T -3.7 3
Andy Blanks -3.0 1
Ari Bloom -3.0 1
Carlos A Ayala Rocha -3.0 1
Dan Nixon -3.0 1
Daniel J. Cohen -3.0 1
Don Quinn -3.0 1
Eric Hoffman -3.0 1
Name
BLatta
John Mac
Ruben D.
Donald R
Ruben D.
Reset

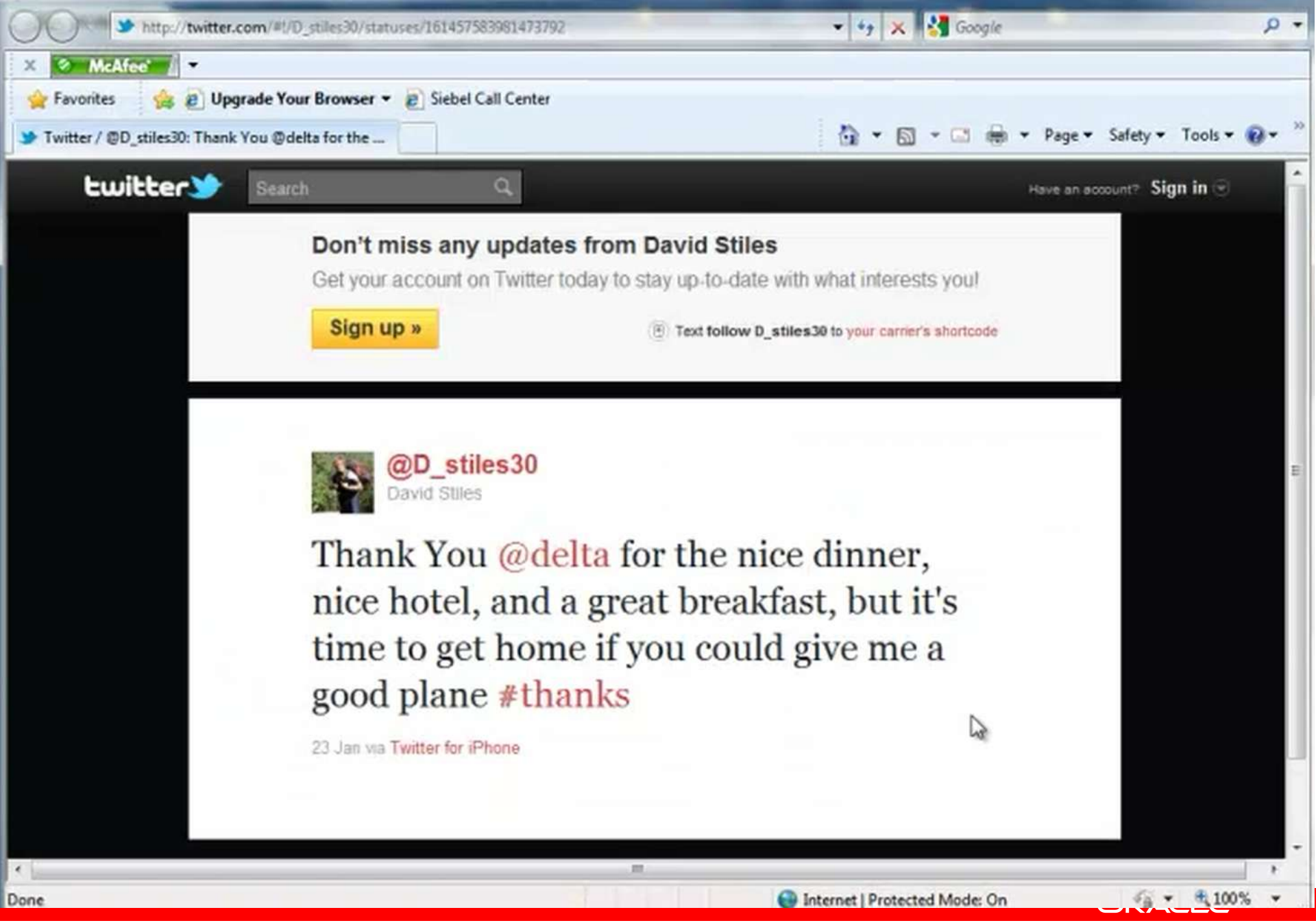
Detail all Tweet

Name	Date	Sentiment	Tweet
BLatta	2012-01-25 23:29:33	2	>RT @BiscoffCookies: A great mention of Biscoff cookies as a favorite snack served on @Delta Air Lines. http://t.co/KdsPzpYq
John Maden	2012-01-25 23:28:20	-1	>Mrs. @katearoni? RT @donalderyan: I never get tired of the "Your Upgrade is Now Confirmed" emails. Thanks @Delta
Ruben D. Sanchez Jr	2012-01-25 23:28:06	1	>Nice!...:-) @Delta Air Lines posts Q4 profit of $425 million http://t.co/sB4CGXaP
Donald Ryan	2012-01-25 23:27:48	-1	>I never get tired of the "Your Upgrade is Now Confirmed" emails. Thanks @Delta
Ruben D. Sanchez Jr	2012-01-25 23:27:36	0	>@Delta Air Lines posts Q4 profit of $425 million

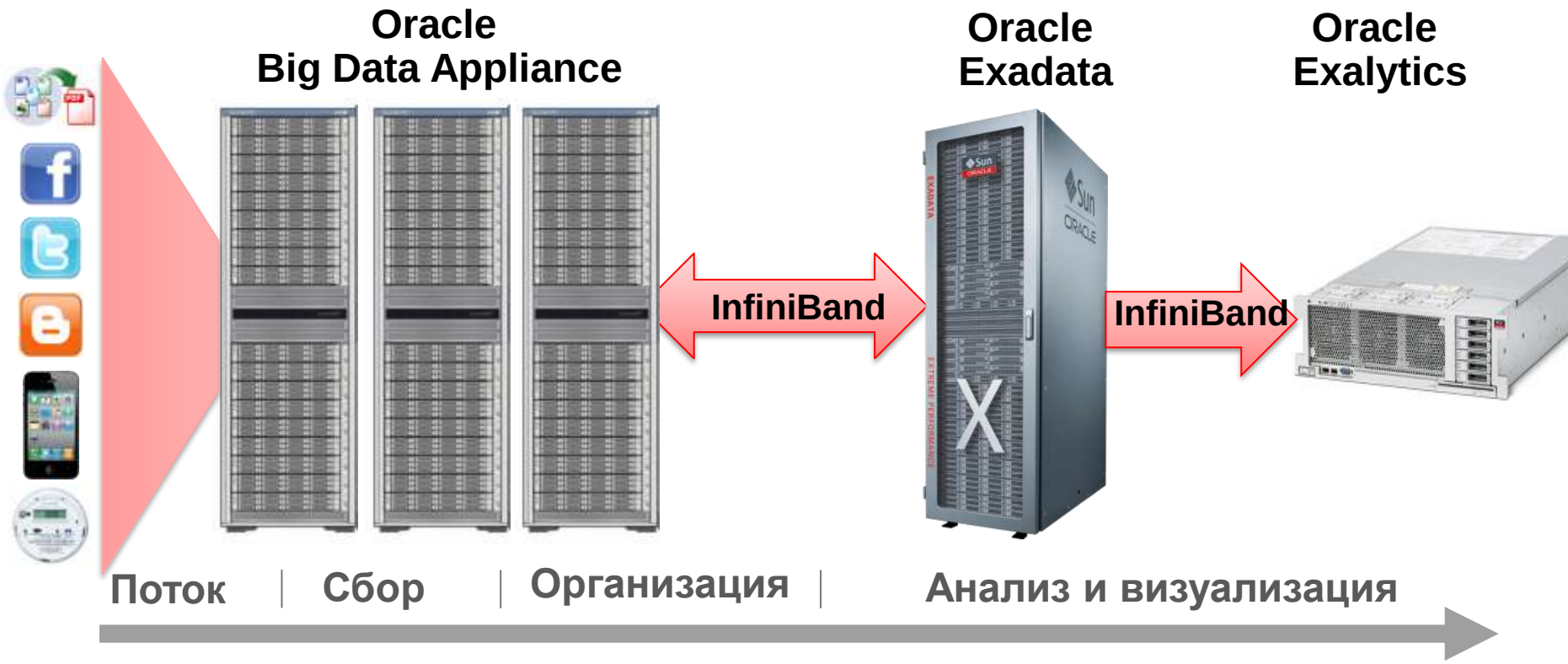
ORACLE

Detail all Tweet

Name	Date	Sentiment	Tweet
David Stiles	2012-01-23 14:37:42	5	>Thank You @delta for the nice dinner, nice hotel, and a great breakfast, but it's time to get home if you could give me a good plane #thanks
Morten Wang	2012-01-25 12:31:01	5	>Would like to take the opportunity to thank @Delta customer service for taking great care of me! #yay #awesome
Amy Turek	2012-01-25 19:53:49	4	>@Delta!! The terminals at JFK & SFO are SO cool plus you get a free bag :) RT @mint: #Minters, what is your favorite airline?
Angela Natividad	2012-01-23 02:02:49	4	>RT @katyzack: Good start to the week. @AKQA honored on @Creativitymag's A-List for standout work w/ @Heineken, @Xbox, @Audi @Delta


http://twitter.com/#!/D_stiles30/statuses/161457583981473792
Google
McAfee
Favorites
Upgrade Your Browser
Siebel Call Center
Twitter / @D_stiles30: Thank You @delta for the ...
Page
Safety
Tools
twitter
Search
Have an account? Sign in
Don't miss any updates from David Stiles
Get your account on Twitter today to stay up-to-date with what interests you!
Sign up »
Text follow D_stiles30 to your carrier's shortcode
@D_stiles30
David Stiles
Thank You @delta for the nice dinner, nice hotel, and a great breakfast, but it's time to get home if you could give me a good plane #thanks
23 Jan via Twitter for iPhone
Done
Internet | Protected Mode: On
100%

Возможная архитектура



Если есть вопросы

Andrey.Pivovarov@oracle.com

http://OracleBI.RU

http://www.oracle.com/bigdata

Questions