

ORACLE BIG DATA: A MUST-HAVE TECHNOLOGY IN 2017



N. USPENSKIY
BigData Specialist, ORACLE Representative in
Russia, Kazakhstan, Mongolia, Russia

ORACLE, Russia
E-mail: nikita.uspenskiy@oracle.com

Abstract. The speaker will share information major Big Data market trends and most popular use cases in CIS region. The presentation will also cover Oracle Big Data portfolio overview and recommendations for Big Data projects implementation.

Oracle Big Data
Технологии, без которых не обойтись в 2017 году

Успенский Никита
Руководитель направления Big Data (Россия, Казахстан, Монголия)
Mob.: +7 916 074 2340
E-mail: nikita.uspenskiy@oracle.com

Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Рынок Big Data

4 года непрерывного технологического прогресса

Июнь 2012: Big Data Appliance X2, Big Data Connectors

Сентябрь 2016: On-Premise (Big Data Appliance X6, Big Data Connectors, Big Data SQL, Big Data Spatial and Graph, Big Data Discovery, ODI for Big Data, Golden Gate for Big Data); Cloud (Big Data Cloud Service, Big Data SQL Cloud Service, Big Data Discovery Cloud Service, Big Data Preparation Cloud Service, ...)

Forrester Wave: Big Data Hadoop-Optimized Systems

- Oracle – №1 среди 7 вендоров
- Oracle – №1 по 3/3 категорий:
 - Продукт
 - Стратегия
 - Присутствие на рынке

Current Strategy	Product	Strategy	Market Presence
Current Strategy	Product	Strategy	Market Presence
Oracle	90%	90%	90%
IBM	80%	80%	80%
EMC	70%	70%	70%
HP	60%	60%	60%
Microsoft	50%	50%	50%
NetScout	40%	40%	40%
Veritas	30%	30%	30%

Увеличение вовлеченности в игровой процесс с помощью аналитики больших данных

WARGAMING.NET
LET'S BATTLE

- Управление и аналитика до 300 миллионов событий в день
- Повышение предпочтений и сегментация игроков
- Быстрое устранение проблем с игровыми сценариями
- На Oracle Big Data Appliance и Oracle Database Appliance развернута Oracle Advanced Analytics и Oracle R Advanced Analytics for Hadoop


62%
Увеличение выручки в одном из регионов за счет улучшения отклика на потребности игроков

BTБ24

«Наша ИТ-стратегия опирается в том числе на трансформацию ИТ-инфраструктуры на базе технологических инноваций. Интегрированный стек продуктов Oracle отличается гибкостью бизнес-банка, а выбор Oracle Big Data Appliance снижает общую стоимость владения, дает не только экономию по стоимости владения данными, но и открывает перспективные для нас превращения в реально работающий анти.»

Сергей Русанов, член правления, директор департамента Банковских и информационных технологий Банка BTБ24

Пресс-релиз: [ссылка](#)
Статья на TADVISER: [ссылка](#)



ORACLE

Альфа Банк

«В начале мы хотели увидеть, действительно ли анализ больших данных приносит те плоды, о которых красиво рассказывают менеджеры по продажам. Мир становится другим, объем данных растет. Анализ информации становится источником дополнительного дохода для бизнеса. Мы в начале пути. У нас есть большие планы по использованию внешнего и внутреннего данных. Важны не просто технологии, а важен комплекс – сочетание технологий, людей, которые это делают и конкретные цели бизнеса»

Ирина Елистратова, главный директор BI



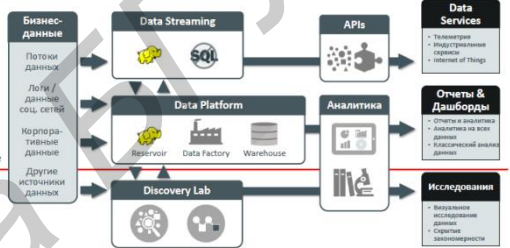
Описание проекта на сайте Snews: [ссылка](#)
Статья на сайте Банковского обозрения: [ссылка](#)

ORACLE

Архитектура ИТ

ORACLE

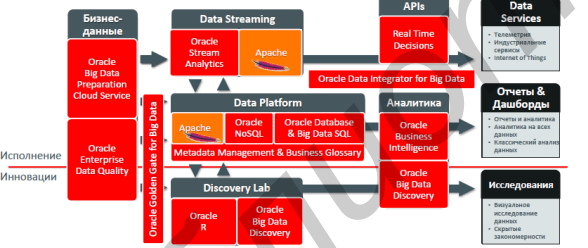
Data Management Platform: общая архитектура



Исполнение
Иновации

ORACLE

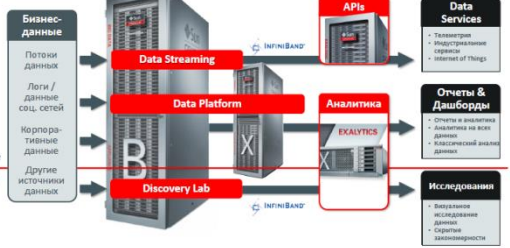
Data Management Platform: компоненты Oracle обеспечивают выполнение всех основных задач Big Data



Исполнение
Иновации

ORACLE

Data Management Platform: интегрированные программно-аппаратные комплексы




Исполнение
Иновации

ORACLE

Преимущества Oracle Big Data

ORACLE

1. Простота доступа к данным



ORACLE

Новые реалии в мире управления данными

Появление и развитие новых компонентов управления данными

Появление и развитие новых средств программирования

ORACLE

Big Data SQL

SQL

Hadoop: Логи, Профили клиентов

NoSQL

Relational: Транзакции

Информационная безопасность:
ВСЕ данные в рамках единых общих политик обеспечения безопасности

ORACLE

Единый интерфейс обращения к данным

Oracle Big Data SQL - нет необходимости переписывать существующие приложения

Предобработка запроса на узлах кластера

Часть данных

SQL

Быстрый возврат выборочных данных

Предобработка запроса на серверах хранения Exadata

ORACLE

Технологии увеличения скорости Big Data SQL

Сокращение ввода-вывода

User Query: 100 TB

Partition Pruning: 10 TB

Storage Indexing: 1 TB

Predicate Pushdown: 100 GB

ORACLE

Технологии Big Data SQL Security

- Единые модели безопасности для различных сред хранения данных
- Дополнительные опции по безопасности, такие как «data redaction» могут быть применены к объединенным источникам данных
- Oracle обеспечивает безопасность поверх Hadoop

ORACLE

Интеграция данных между традиционными реляционными СУБД и миром Big Data

SQL Query in place with Big Data SQL

Oracle Table Access for Hadoop

Oracle Loader for Hadoop (Big Data connectors)

Copy to BDA

Резервуар данных

Хранилище данных

ORACLE

Oracle Big Data

Открытая аналитическая платформа

ORACLE
BIG DATA APPLIANCE
OPTIMIZED

ORACLE

2. Исследование НОВЫХ ВОЗМОЖНОСТЕЙ

ORACLE

Из данных в Hadoop не так легко извлечь **пользу**

- Существующие аналитические средства не подходят
 - Требуется много усилий для подготовки данных
 - Ручная обработка наборов данных
 - Очистка данных и их подготовка сильно зависят от ETL инструментов
 - Аналитические запросы должны быть определены заранее
- Нет единого инструмента для анализа
 - Несогласованность данных и визуализации
 - Нет единого «источника правды»
- Нужен полный функционал
- Сложные компоненты Hadoop
 - Pig, Oozie, Sqoop, Hive, Spark
- Сложно найти специалистов
 - Языки программирования (например, Map Reduce, Python, Scala)
 - Мат. статистика и машинное обучение
 - Интерфейсы с командной строкой

Аналитика больших данных требует новых подходов

Интуитивный, интерактивный и наглядный пользовательский интерфейс

понятный **всем**, для быстрого поиска, изучения, трансформации и анализа данных в Hadoop.

а также для **общего пользования** результатами

Big Data Discovery. Визуальный интерфейс к Hadoop

Легко находите нужные наборы данных

- Доступ к интерактивному **каталогу** данных, расположенных в Hadoop
- Удобный **поиск** и **навигация** в наборах данных
- Аннотации, краткие описания и **рекомендации** к наборам данных
- **Загрузка** собственных данных
- Автоматическое **обогащение** данных
- Навигация по собственным и коллективным **проектам**

Изучайте данные и раскрывайте их потенциал

- Понимание **структуры** данных. Визуализация атрибутов по типам
- **Сортировка атрибутов по степени их важности** на основе алгоритмов машинного обучения
- Распределение данных, **качество данных** и выбросы
- Понимание **корреляции** между атрибутами
- **Оценка целесообразности** дальнейшей работы с набором данных

Трансформируйте и обогащайте данные

- Интуитивный интерфейс для **трансформации** данных
- Расширяемая **библиотека функций** (замена значений, конвертация типов, группировки, слияния и т.д.)
- **Предварительный просмотр** результатов, отмена, подтверждение и повторное выполнение трансформаций
- Тестирование на небольшом наборе данных в оперативной памяти, применение на полном наборе данных в **Hadoop**

Исследуйте данные для получения новых знаний

- **Объединение** различных данных для более глубокого анализа
- Фильтрация данных с помощью мощного **поисковика** и интуитивной **навигации с подсказками**
- Интерфейс «**Drag & drop**» для создания интерактивных визуальных дашбордов
- **Публикация** в Hadoop отчетов на основе смешанных наборов данных

Делитесь результатами с Вашими коллегами

- **Делитесь и взаимодействуйте** с Вашей командой
 - Делитесь проектами, закладками, снапшотами и т.д.
- **Публикуйте** данные в Hadoop
 - Трансформации и обогащение могут быть применены к исходным наборам данных
 - Публикация подготовленных наборов данных в HDFS
- **Используйте** результаты для работы с другими инструментами
 - Публикуйте данные в Hadoop в формате, оптимизированном для аналитических инструментов (например, ORAAI)
 - Используйте стандартные компоненты Hadoop (например, Pig, Hive, Impala, Python, т.д.)

Oracle Big Data Discovery отлично работает со всех **экосистемой Big Data**

3. Развертывание, сопровождение и TCO

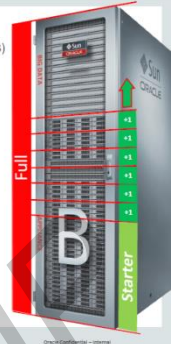


ORACLE

Copyright © 2013, Oracle and/or its affiliates. All rights reserved. | Oracle Confidential – internal

BIG DATA APPLIANCE – аппаратная платформа

- МОЩНЫЙ** До 792 ядер с 4.6TB RAM (расширяется до 13.8TB)
До 1728 TB дискового пространства
40 Gb/s InfiniBand
- МОДУЛЬНЫЙ** Минимум 6 узлов и расширение по 96TB до 20+ систем в кластере
- ОПТИМИЗИРОВАННЫЙ** Оптимизации для Linux
Сконфигурированная Java
Сетевые оптимизации
Оптимизации для Hadoop
Единая точка обновления и патчирования
- БЕЗОПАСНЫЙ** Kerberos аутентификация
Авторизация Apache Sentry
Аудит Oracle Audit Vault
Шифрование данных
Шифрование трафика
- УПРАВЛЯЕМЫЙ** Централизованное управление через OEM 12c



ORACLE

Copyright © 2013, Oracle and/or its affiliates. All rights reserved. | Oracle Confidential – internal

BIG DATA APPLIANCE – программное обеспечение

ORACLE

- Oracle Linux 5 или Oracle Linux 6
- Oracle Java – JDK 8
- Oracle R Distribution
- Oracle Big Data Appliance Enterprise Manager Plug-In
- Big Data SQL (опционально, лицензируется отдельно)
- Big Data Discovery (опционально, лицензируется отдельно)
- Oracle NoSQL Database Enterprise Edition (опционально, лицензируется отдельно)

cloudera

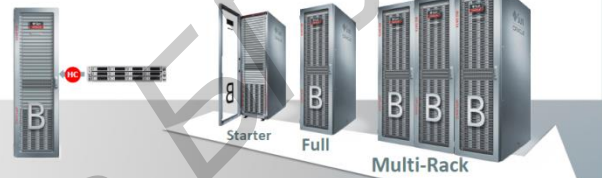
- Cloudera Enterprise 5 – Data Hub Edition
- Cloudera’s Distribution Including Apache Hadoop (CDH) with support for YARN and MR2
- Cloudera Impala
- HBase (as well as support for Accumulo)
- Cloudera Search
- Apache Spark
- Cloudera Manager including:
 - Cloudera Back-up and Disaster Recovery (BDR)
 - Cloudera Navigator



ORACLE

Copyright © 2013, Oracle and/or its affiliates. All rights reserved. | Oracle Confidential – internal

Эластичное масштабирование от Starter Rack to Multi-Rack



- Начальная конфигурация включает 6 BDA узлов и все свитчи
- Можно добавлять по одному BDA узлу
- Системы предыдущих поколений можно расширять новыми узлами

ORACLE

Copyright © 2013, Oracle and/or its affiliates. All rights reserved. | Oracle Confidential – internal

Что вам НЕ придется делать с Big Data Appliance (1/2)

- Подбирать и оптимизировать компоненты – серверы, диски, количество дисков, процессоры, сети, память и т.п.
- Заключать отдельный договор о поддержке с Cloudera
- Собирать кластер
- Настраивать сетевые коммутаторы
- Инсталлировать операционную систему на каждом узле и
- Отслеживать и устанавливать оптимальные версии драйверов и прошивок для каждого компонента
- Настраивать операционную систему для оптимальной производительности (у нас же очень много данных!)
- Настраивать Java



ORACLE

Copyright © 2013, Oracle and/or its affiliates. All rights reserved. | Oracle Confidential – internal

Что вам НЕ придется делать с Big Data Appliance (2/2)

- Инсталлировать дополнительное ПО от Cloudera
- Тестировать работоспособность и производительность каждого узла кластера
- Заниматься самостоятельно трудоемкой процедурой многоуровневого апгрейда и патчирования BIOS, OS, Java, Hadoop и т.п.
- И просто следить за тем, что нужно что-то проапгрейдить
- Изучать как это все сделать без остановки и прерывания работы пользователей
- Заниматься дизайном переконфигурации кластера при его расширении
- И т.д. и т.п.



ORACLE

Copyright © 2013, Oracle and/or its affiliates. All rights reserved. | Oracle Confidential – internal

TCO и развитие системы

- Экономия на лицензировании:
 - По кол-ву дисков (например, Oracle Big Data SQL):
 - Больше размер дисков → меньше лицензий для нужного объема данных
- Экономия на тех. поддержке ПО:
 - 12% от стоимости Oracle Big Data Appliance, тех. поддержка Cloudera учтена
 - Напрямую покупать дороже (цена на узел \$3000-\$7000, узлы могут быть «слабее»)
- Западные заказчики переходят на Enterprise-решения
- Рынок развивается в сторону конвергентных систем, а не подхода «сделай сам»

ORACLE

Copyright © 2013, Oracle and/or its affiliates. All rights reserved. | Oracle Confidential – internal


Техническая поддержка

- Premier Support: 24x7, русский язык в рабочие часы с 9:00 до 18:00
- Покрывает весь «стеck» программно-аппаратного комплекса:
 - Аппаратные компоненты
 - Системное ПО (OS, Java и т.д.)
 - Платформенное ПО Cloudera
- Единый интерфейс обращения – через Oracle
- 12% от стоимости решения в год

ORACLE

Copyright © 2013, Oracle and/or its affiliates. All rights reserved. | Oracle Confidential – internal

4. Производительность



Copyright © 2015, Oracle and/or its affiliates. All rights reserved. | 40

Oracle Big Data Appliance

В **2X** быстрее самосборного кластера¹




Oracle Big Data Appliance is a pre-configured, turnkey solution for big data processing. It is based on Oracle Exalytics In-Memory Machine and Oracle Big Data Appliance. It is designed to be easy to deploy and use, and it provides a high level of performance and scalability. It is also easy to integrate with existing applications and data sources.

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. | 41

1DMP (CleverDATA)

Тестирование партнерского решения в РФ

• Результаты тестирования:

- Обработка более 101,5 тысяч (в 1,45 раз больше изначально ожидаемого числа) запросов в секунду всего на одном узле Oracle Big Data Appliance. Время отклика на запрос в 99% случаев не превысило 1,17 миллисекунды (в 1,7 раза лучше ожидаемого).
- Время классификации Интернет-страниц и построение пользовательских профилей на 6 узлах Oracle Big Data Appliance составило 11 минут 17 секунд и оказалось в 5,43 раза лучше ожидаемых результатов.

<http://cleverdata.ru/1dmp-na-oracle-big-data-appliance/>

«Число клиентов Биржи Данных растет, а это значит, что актуальными становятся вопросы увеличения производительности и масштабируемости решений. Мы решили эти задачи за счет партнерства платформы на Oracle Big Data Appliance, что позволяет платформе 1DMP работать на объектах, производящих собственные решения по управлению большими Данными, масштабы и за короткое время классифицировать большие объемы информации, формировать профили клиентов для целей маркетингового взаимодействия. Компания Oracle Big Data Appliance позволил нам обеспечить короткое минимальное время отклика, что является критичным при работе с Данными»

Денис Афанасьев, генеральный директор CleverDATA



Copyright © 2015, Oracle and/or its affiliates. All rights reserved. | 42

IQPLATFORM® (Айкумен ИБС)

Тестирование партнерского решения в РФ

• Результаты тестирования:

- Скорость поиска в архиве выросла в 9 раз – с 40 000 до 366 600 документов в секунду

«Ускорение обработки данных с применением программно-аппаратных возможностей Oracle Exalytics, Oracle Exalytics и Oracle Big Data Appliance позволяет нашим клиентам наиболее эффективно обрабатывать большие массивы накопленной информации, используя быстрый онлайн-доступ к необходимым данным для их комплексного анализа прямо на входном потоке. Таким образом, открывается возможность решения нового класса исследовательских задач, требующих обработки данных в режиме реального времени»

Дмитрий Часовский, руководитель проектного отдела «Айкумен ИБС»

http://www.imen.ru/imen-oracle_exastark



Copyright © 2015, Oracle and/or its affiliates. All rights reserved. | 43

ForSMedia (ФОРС – ЦР)

Тестирование партнерского решения в РФ


• Результаты тестирования:

- Скорость поиска в архиве выросла в 9 раз – с 40 000 до 366 600 документов в секунду

«Стать впечатляющие результаты тестирования ForS Media на Oracle Exalytics In-Memory Machine и Oracle Big Data Appliance свидетельствуют о несомненных преимуществах этих оптимизированных программно-аппаратных комплексов перед традиционными, что открывает широкие перспективы для продвижения наших решений с использованием технологий Больших Данных среди компаний финансового, телекоммуникационного сектора, в ритейле и других индустриях»


Алексей Галосов, президент ФОРС

<http://www.fors.ru/pressroom/news/1892/>



Copyright © 2015, Oracle and/or its affiliates. All rights reserved. | 44

Примеры проектов



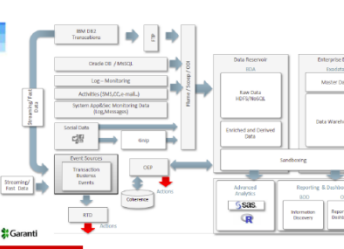
Copyright © 2015, Oracle and/or its affiliates. All rights reserved. | 45

Oracle Big Data: референсные заказчики




Copyright © 2015, Oracle and/or its affiliates. All rights reserved. | 46

Garanti Bank



7 сценариев применения

- Alert System
- Next Best Offer
- Customer SWA
- Data Pool
- Data Mining
- Urban Analytics
- Customer DNA



Copyright © 2015, Oracle and/or its affiliates. All rights reserved. | 47

