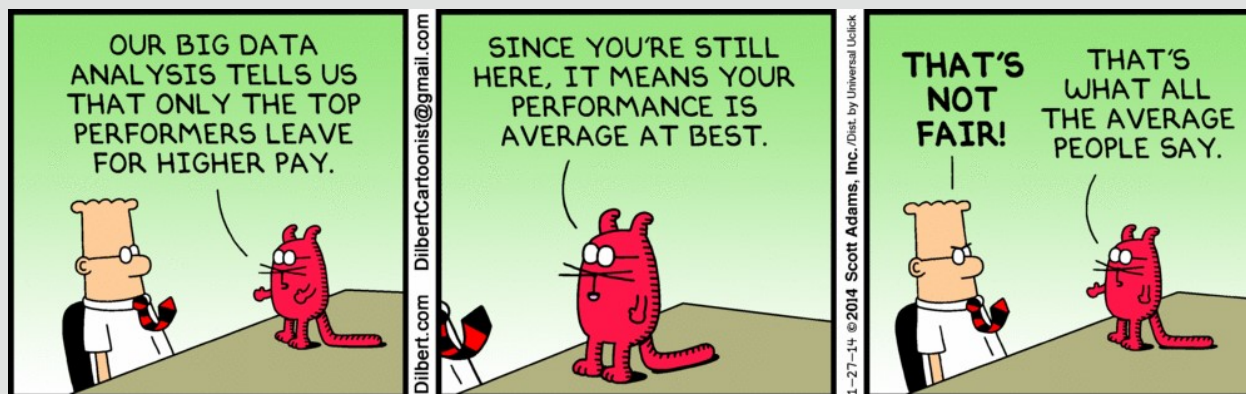




ORACLE

Oracle Big Data

Обзор Технологий Oracle BigData



Роман Рахамимов
Технический Консультант
Roman.Rakhamimov@oracle.com

Safe Harbor

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, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

Statements in this presentation relating to Oracle's future plans, expectations, beliefs, intentions and prospects are "forward-looking statements" and are subject to material risks and uncertainties. A detailed discussion of these factors and other risks that affect our business is contained in Oracle's Securities and Exchange Commission (SEC) filings, including our most recent reports on Form 10-K and Form 10-Q under the heading "Risk Factors." These filings are available on the SEC's website or on Oracle's website at <http://www.oracle.com/investor>. All information in this presentation is current as of September 2019 and Oracle undertakes no duty to update any statement in light of new information or future events.

A DAY IN DATA

The exponential growth of data is undisputed, but the numbers behind this explosion - fuelled by internet of things and the use of connected devices - are hard to comprehend, particularly when looked at in the context of one day

500m

tweets are sent every day
Twitter



4PB

of data created by Facebook, including

350m photos
100m hours of video watch time
Facebook Research

294bn

billion emails are sent
Radicati Group

320bn

emails to be sent each day by 2021

306bn

emails to be sent each day by 2020

3.9bn

people use emails

4TB

of data produced by a connected car
Intel

ACCUMULATED DIGITAL UNIVERSE OF DATA

4.4ZB

44ZB

Perf.

2013

2020

DEMYSTIFYING DATA UNITS

From the more familiar 'bit' or 'megabyte', larger units of measurement are more frequently being used to explain the masses of data

Unit	Value	Size
b bit	0 or 1	1/8 of a byte
B byte	8 bits	1 byte
KB kilobyte	1,000 bytes	1,000 bytes
MB megabyte	1,000 ² bytes	1,000,000 bytes
GB gigabyte	1,000 ³ bytes	1,000,000,000 bytes
TB terabyte	1,000 ⁴ bytes	1,000,000,000,000 bytes
PB petabyte	1,000 ⁵ bytes	1,000,000,000,000,000 bytes
EB exabyte	1,000 ⁶ bytes	1,000,000,000,000,000,000 bytes
ZB zettabyte	1,000 ⁷ bytes	1,000,000,000,000,000,000,000 bytes
YB yottabyte	1,000 ⁸ bytes	1,000,000,000,000,000,000,000,000 bytes

*A lowercase "b" is used as an abbreviation for bits, while an uppercase "B" represents bytes.

65bn

messages sent over WhatsApp and two billion minutes of voice and video calls made
Facebook



Searches made a day

5bn

Searches made a day from Google

3.5bn

Smart Insights



463EB

of data will be created every day by 2025
BBC

95m

photos and videos are shared on Instagram
Instagram Business

Instagram Business

28PB

to be generated from wearable devices by 2020
Statista



RACONTEUR



Для чего нужны данные?

1 Привлечение, рост и сохранение клиентов

- Персонализация клиентского опыта
- Привлечение и сохранение новых клиентов
- Повышение доходности клиента

2 Оптимизация затрат и сокращение мошенничества

- Повышение эффективности инфраструктуры
- Выявление и предотвращение мошенничества
- Безопасность

3 Глубокое понимание и анализ бизнеса

- Использование и анализ всех имеющихся данных
- Оптимизация аналитической работы
- Новые виды анализа

4 Повышение эффективности и выполнение требований

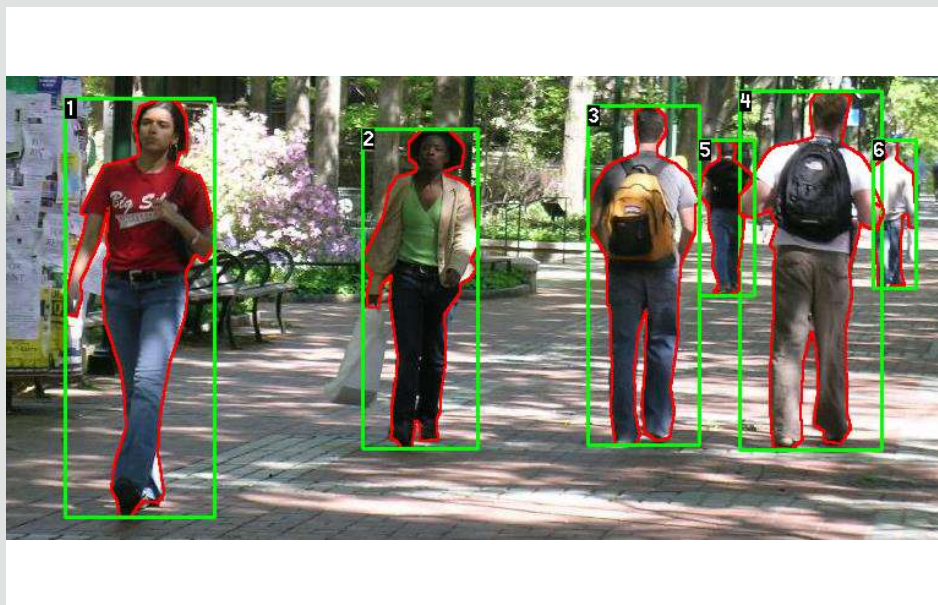
- Финансовая и операционная эффективность
- Оценка и управление рисками
- Соответствие требованиям регуляторов

5 Создание новых моделей бизнеса

- Продукты и услуги на основе данных
- Новые виды партнерства
- Быстрый ввод услуг и оценка эффекта



Пример использования



Идентификация объектов на видео



Мониторинг массовых мероприятий

Еще Пример: Предиктивной аналитика для управления электроцентробежными насосами



Цель

- Использование инструментов Big Data для выявления причин сбоев автоматического перезапуска насосов после аварийного отключения электропитания

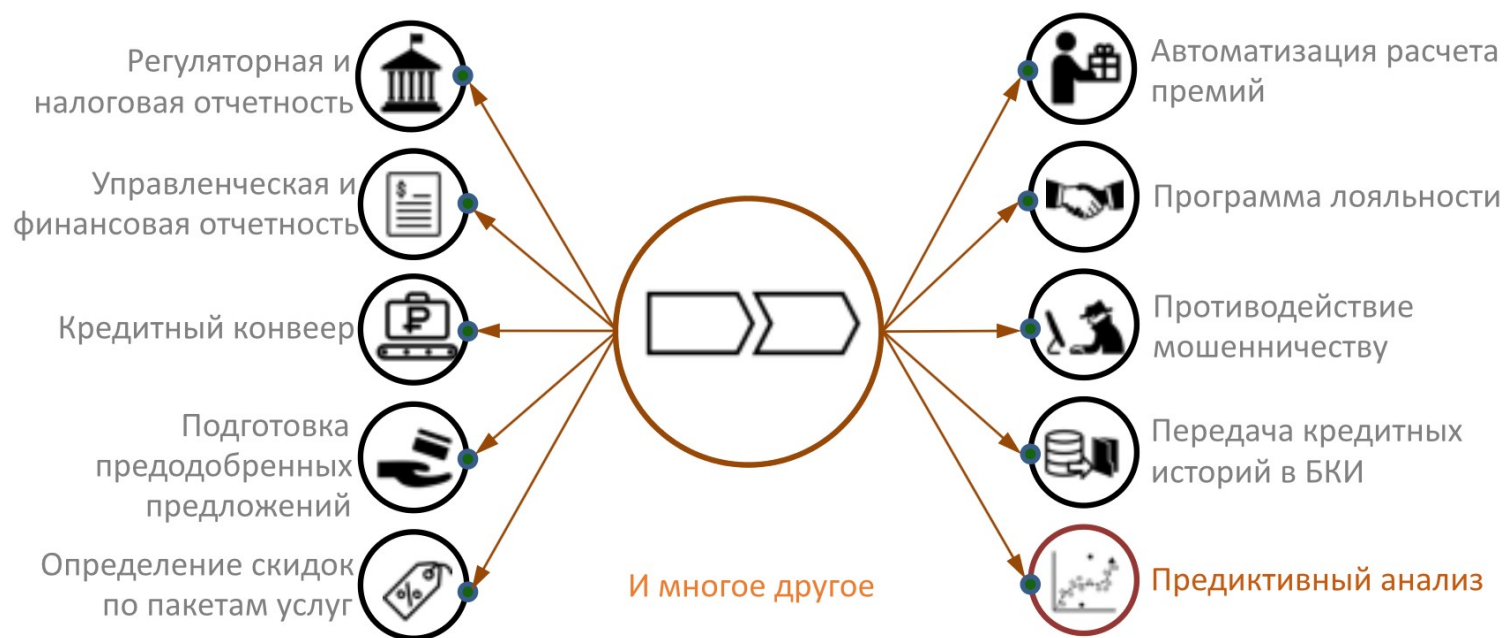
Результат

- Созданы визуализированные модели цепочек событий, относящиеся к самозапуску насосов, а также карты вероятностного распределения причинно-следственных связей
- Применение инструментов Big Data позволило не только сформировать и проверить набор различных гипотез о причинах сбоев в автозапуске, но и получить информацию о ранее неизвестных взаимосвязях в работе насосного оборудования

На Примере Финансовой Организации



Поддерживаемые бизнес-процессы



Что Такое Большие Данные?

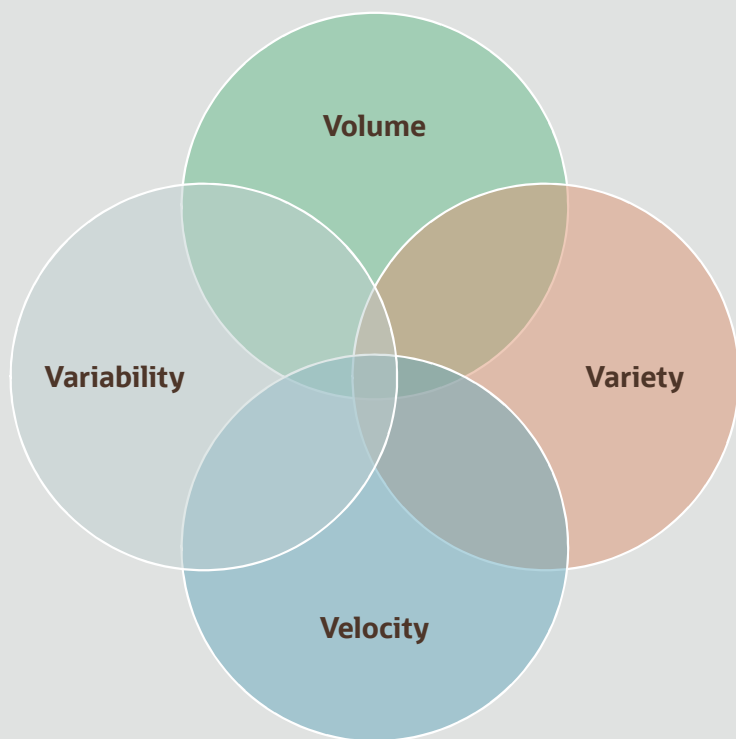
- **Феномен**, связанный с появлением технологических возможностей анализировать огромные массивы данных
- Множество технологий, порой конкурирующих

Хранение	HDFS, HBase, NoSQL, RDBMS, Object Store ...
Обработка (движок)	MapReduce, Tez, Hive, Pig, Spark, Storm, Impala, Mahout ...
Потоковая обработка	Spark Streaming, Kafka ...
Доступ к данным	SQL, BigData SQL, HUE, flume, Alluxio ...
Управление	YARN, Mesos, Ambari, Cloudera Manager ...

- Часть задач можно решать в обычных СУБД, но это может быть дороже
- Не всегда «неструктурированные данные»



Характеристики Больших данных



Volume (объем)

- к 2025 году ожидается 165 Зетабайт данных (165 триллионов ГБ)

Variety (разнообразие)

- СУБД, текст, аудио, видео, социальные сети, ...

Velocity (скорость)

- цифровая цивилизация генерирует массу данных

Variability (непостоянство)

- неоднородность форматов, качество данных

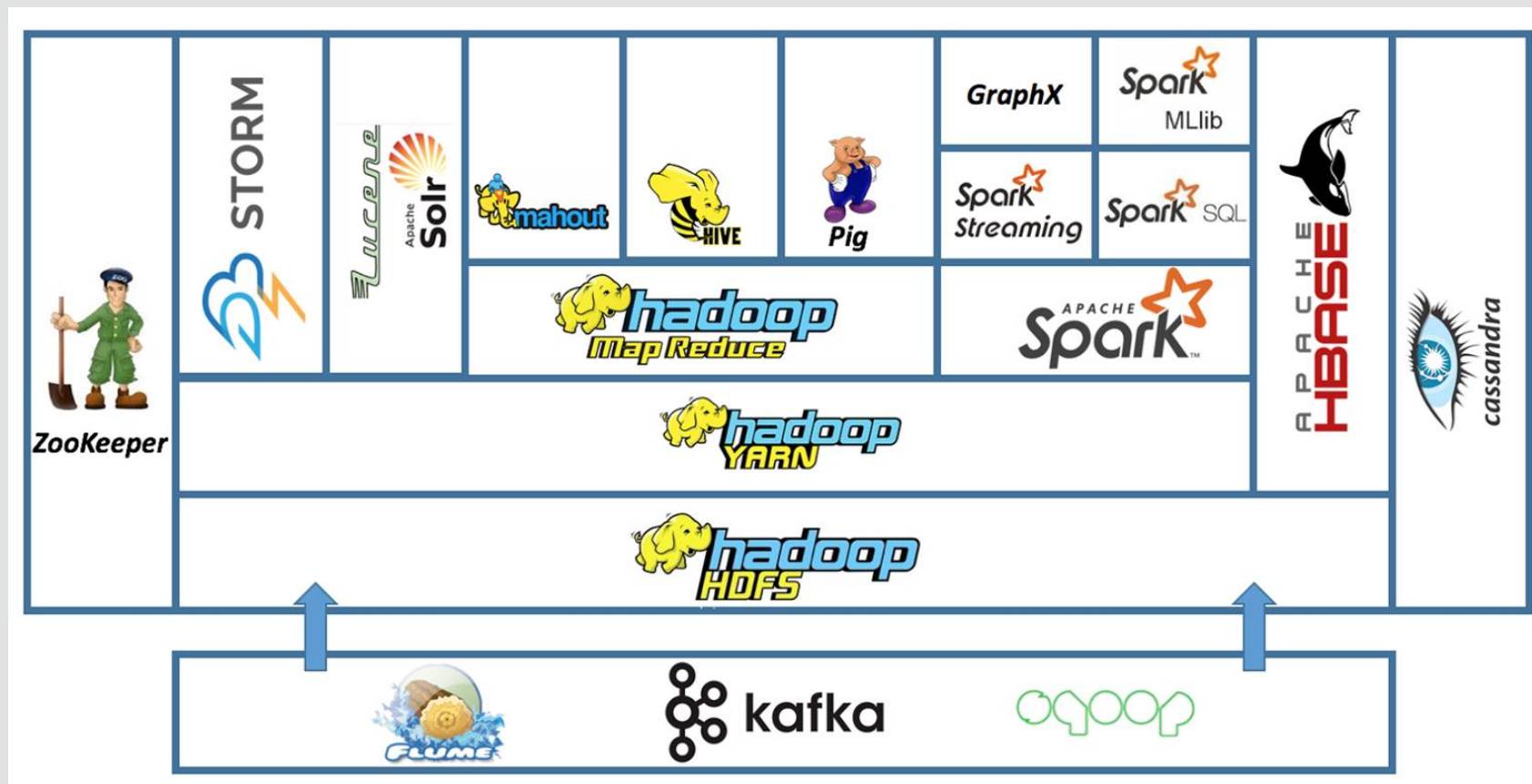
Big Data – другими словами

Данные,
которых
много
настолько
что:

- их **дорого** хранить и обрабатывать, используя традиционные технологии
- это **отражается на производительности** существующих систем
- это **ограничивает возможности** накопления данных для более расширенного анализа



Разнородные Инструменты





Видение Oracle

Go Deeper with Oracle's Big Data Products



Data Integration

Big data comes from a number of sources. Weaving all of that data into a comprehensive and manageable fabric is a process in itself. The result is a unified view of



Big Data Management

Managing big data sources, processes, and output is a tricky balancing act, especially when compliance and other factors come

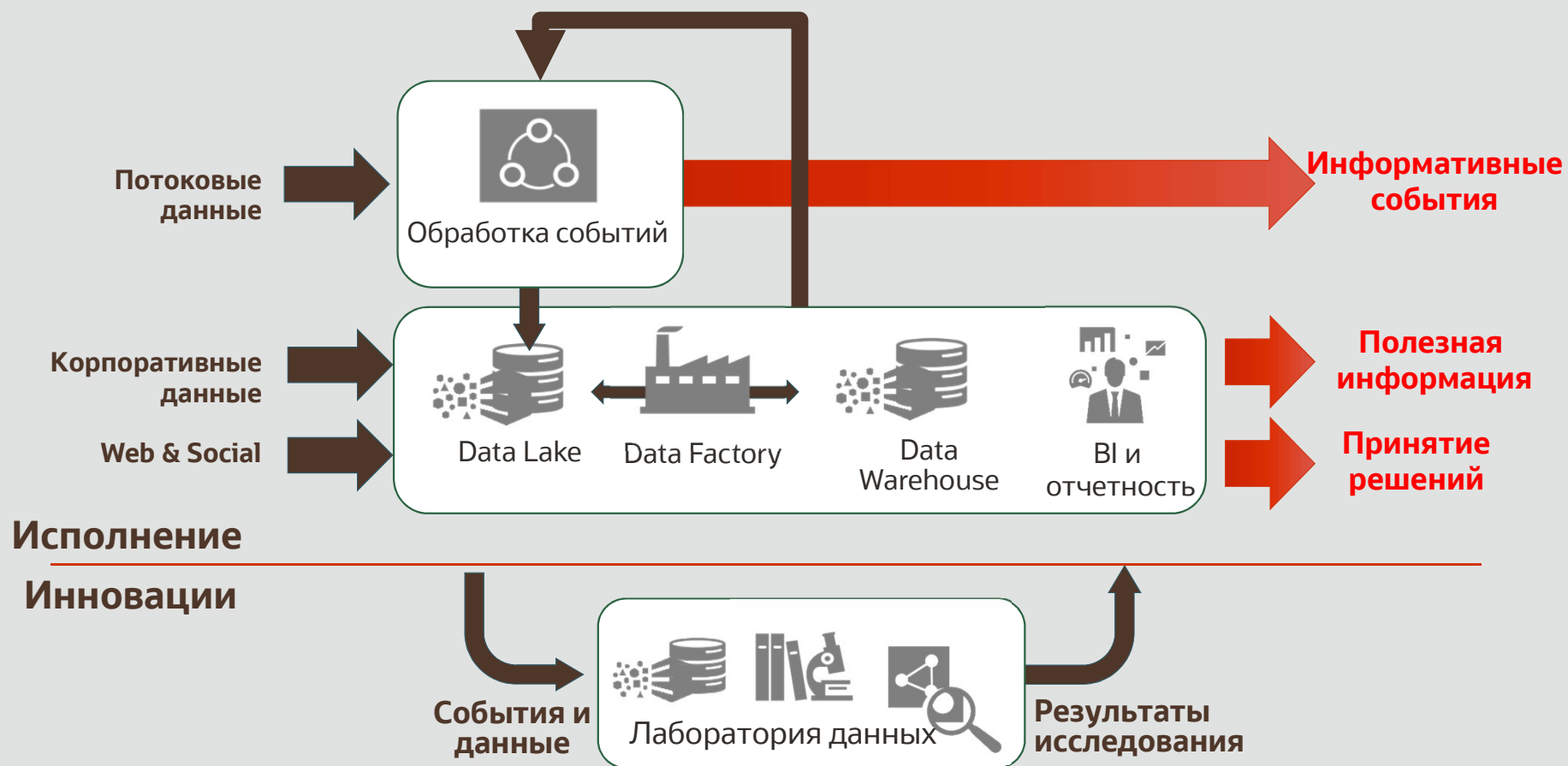


Data Science

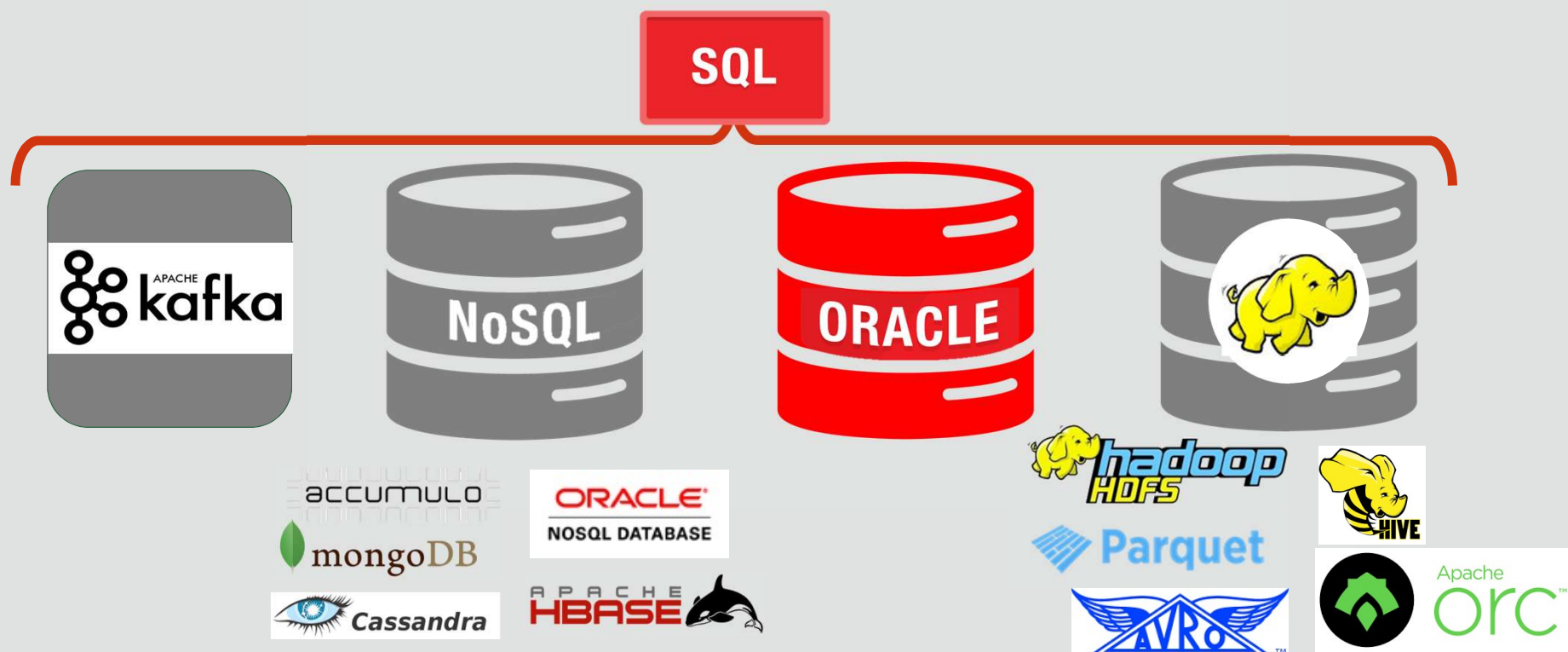
Through data science, numbers aren't just numbers anymore. They tell a story about your organization's past, present, and future. Data science is all about discovery.



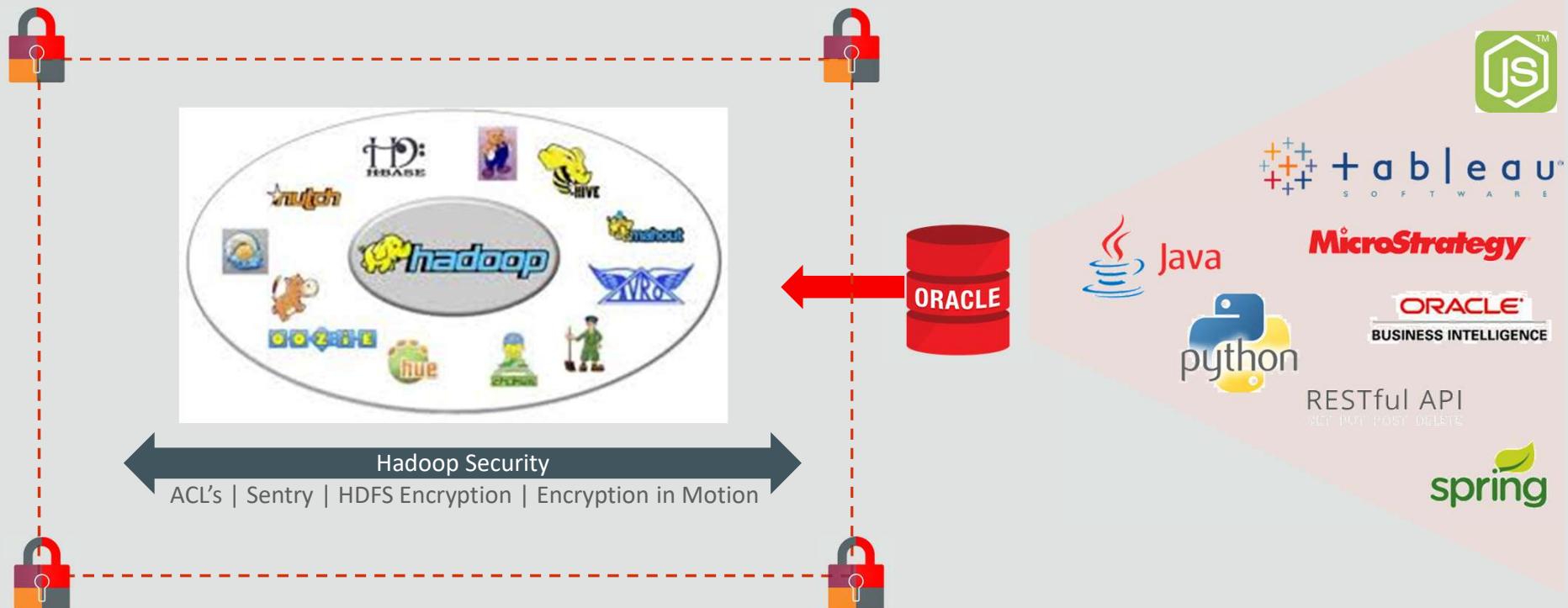
Архитектура Oracle BigData



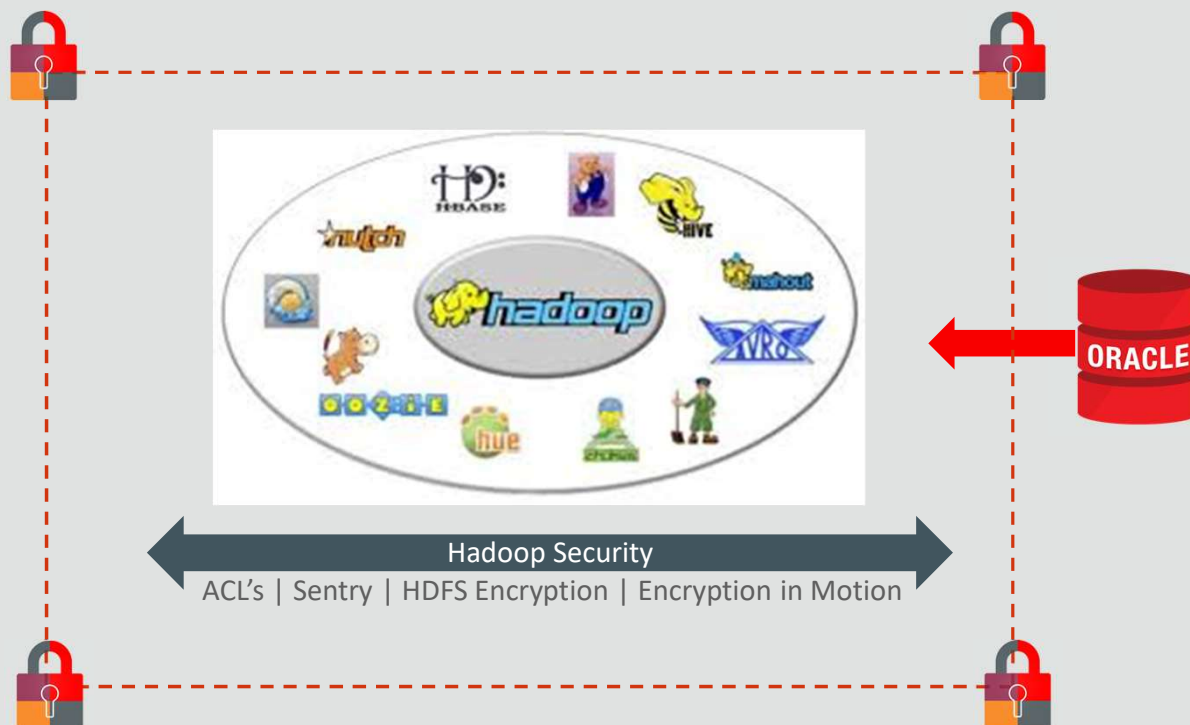
Технология Big Data SQL/Cloud SQL



Технологии Big Data SQL Security



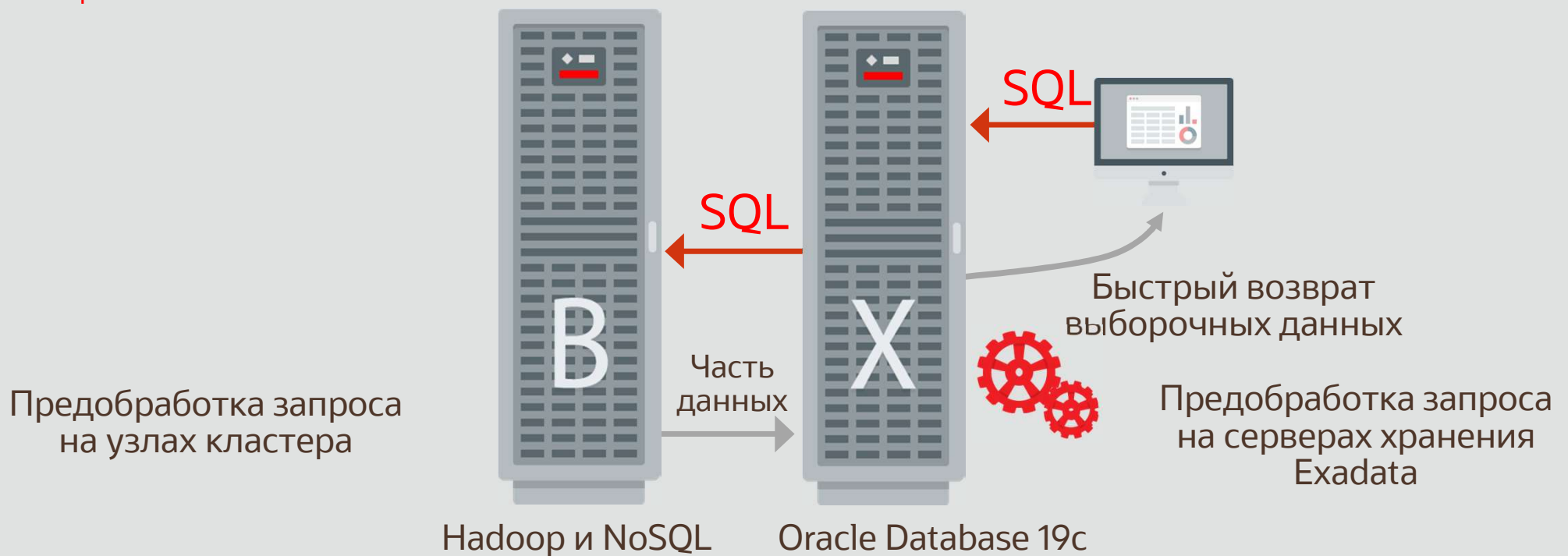
Технологии Big Data SQL Security



- 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

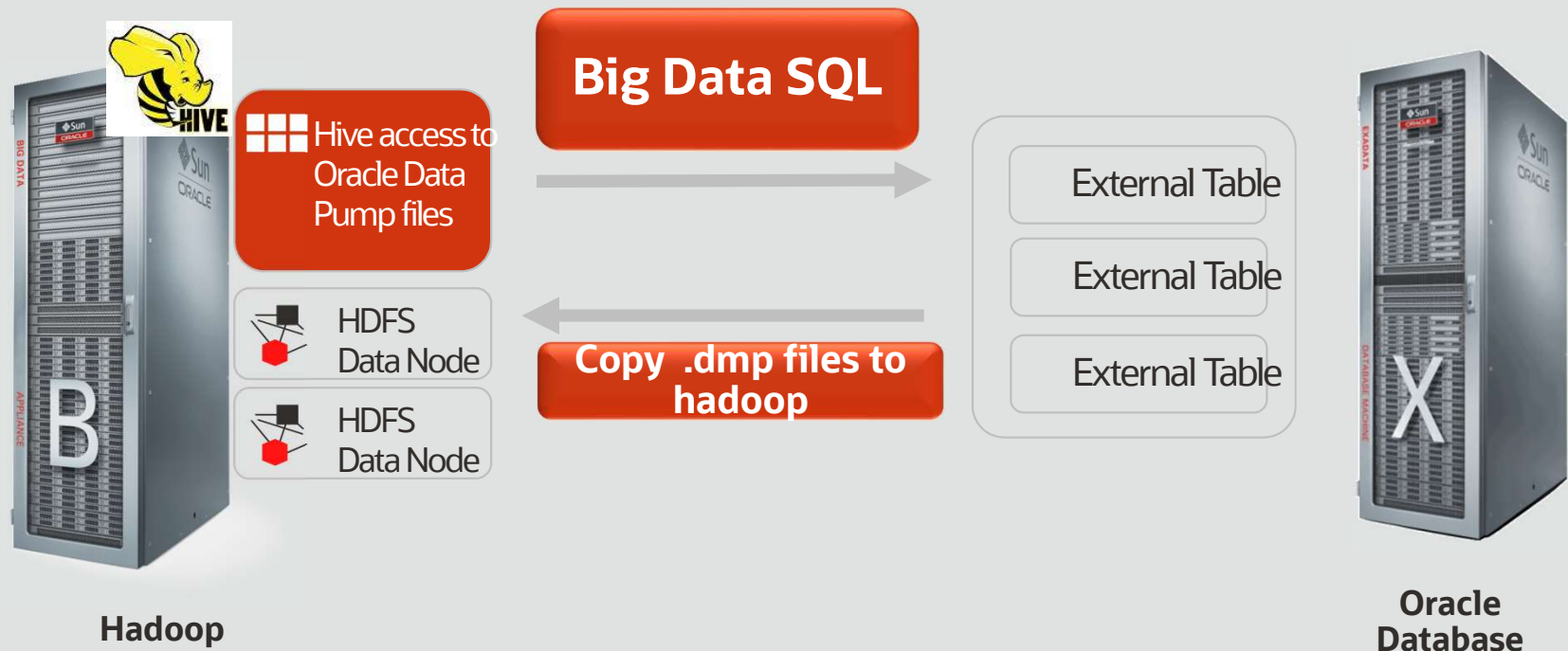
Единый интерфейс обращения к данным

Oracle Big Data SQL - нет необходимости переписывать существующие приложения



Oracle Big Data SQL

Copy to Hadoop



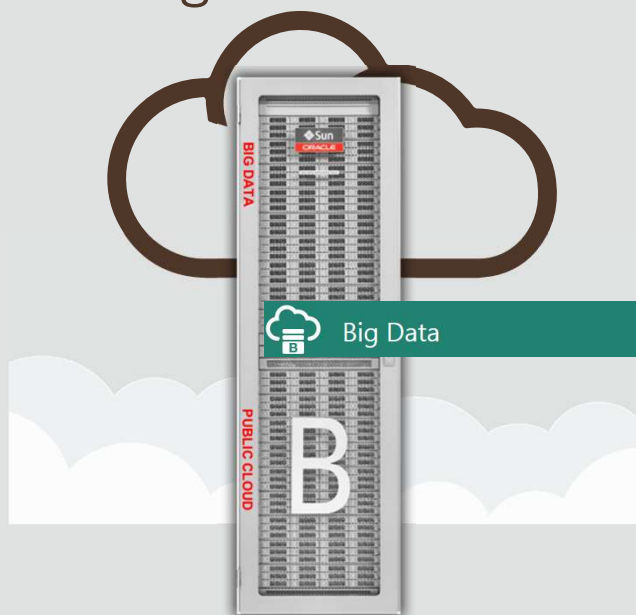
Варианты развертывания

Big Data Appliance X8



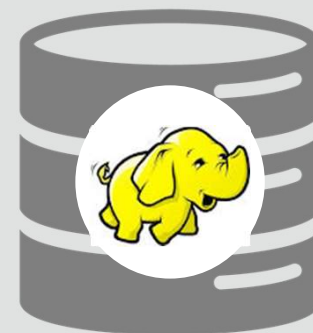
On-Premises

Big Data Service



Big Data Service в
облаке Oracle

Non Oracle BigData



BIG DATA **APPLIANCE** – аппаратная платформа

МОЩНЫЙ

Каждый узел:

48 ядер с **256 GB RAM** (расширяется до 1,5 TB)
120 TB дискового пространства

МОДУЛЬНЫЙ

Минимум 6 узлов и расширение по **120 TB**
до 20+ систем в кластере

ОПТИМИЗИРОВАННЫЙ

Оптимизации для **Linux**
Сконфигурированная **Java**
Сетевые оптимизации **40 Gb/s Infiniband**
Оптимизации для **Hadoop**
Единая точка обновления и патчирования

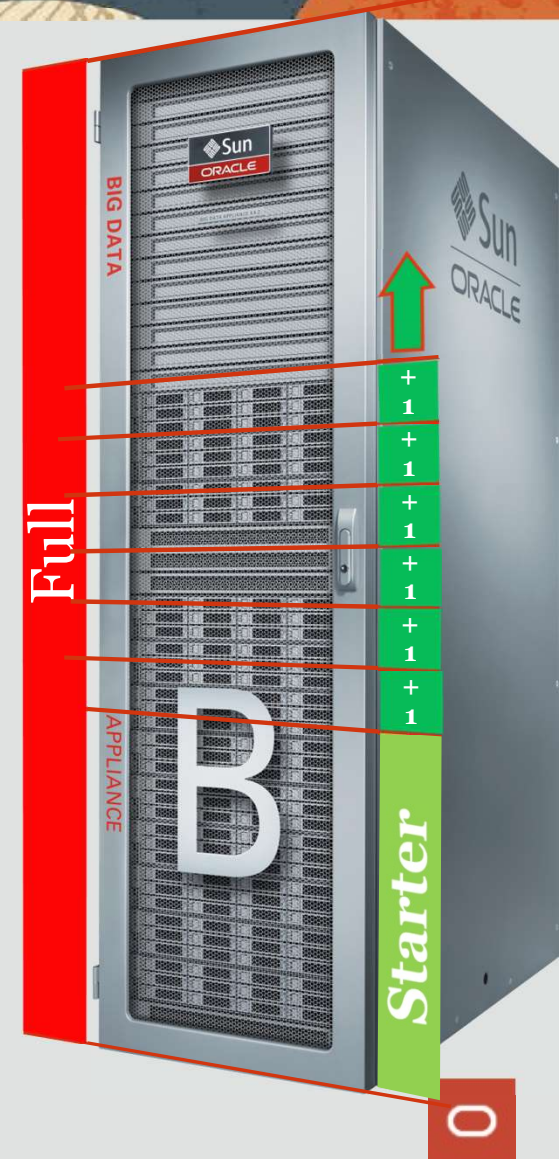
БЕЗОПАСНЫЙ

Kerberos аутентификация
Авторизация Apache Sentry
Аудит Oracle Audit Vault
Шифрование данных
Шифрование трафика



УПРАВЛЯЕМЫЙ

Централизованное управление через OEM 12c



BIG DATA **APPLIANCE** – программное обеспечение

ORACLE®

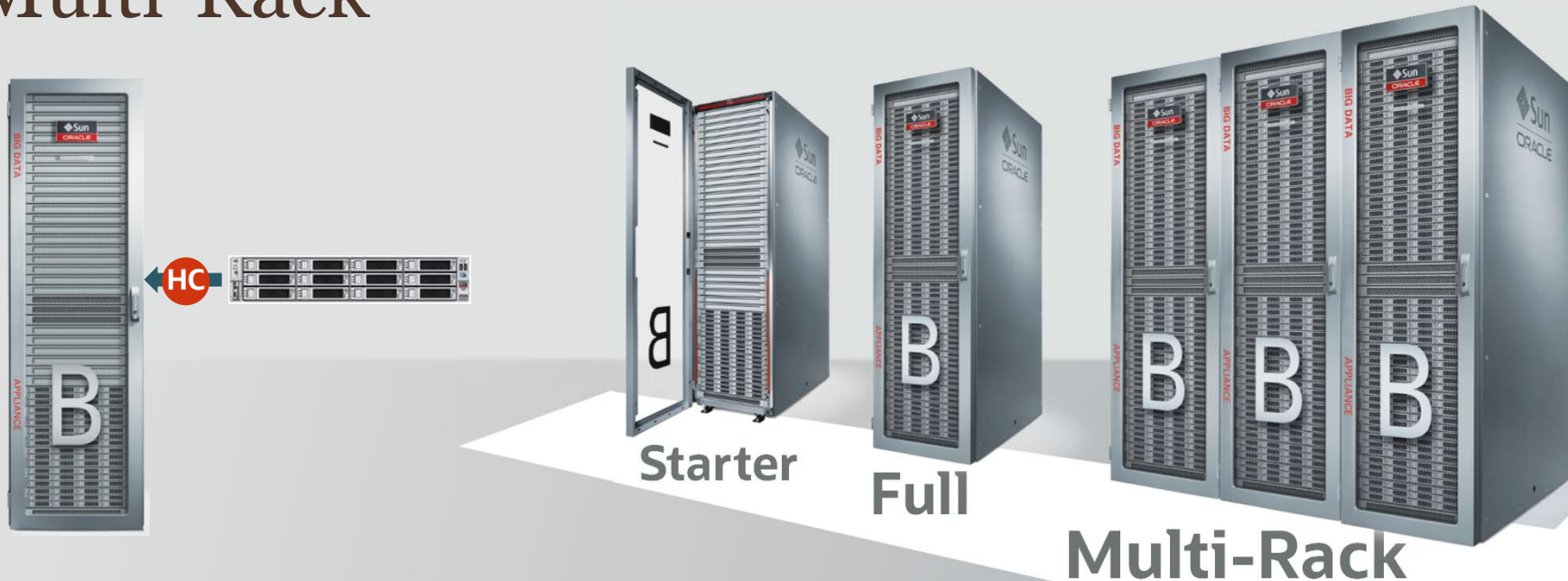
- Oracle Linux 5 или Oracle Linux 6
- Oracle Java – JDK 8
- Oracle R Distribution
- Oracle Big Data Appliance Enterprise Manager Plug-In
- **Big Data SQL (опционально, лицензируется отдельно)**
- **Oracle NoSQL Database Enterprise Edition (опционально, лицензируется отдельно)**

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



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

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
- 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



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 многуровневого 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 перебалансировки кластера при его расширении
- И т.д. и т.п.



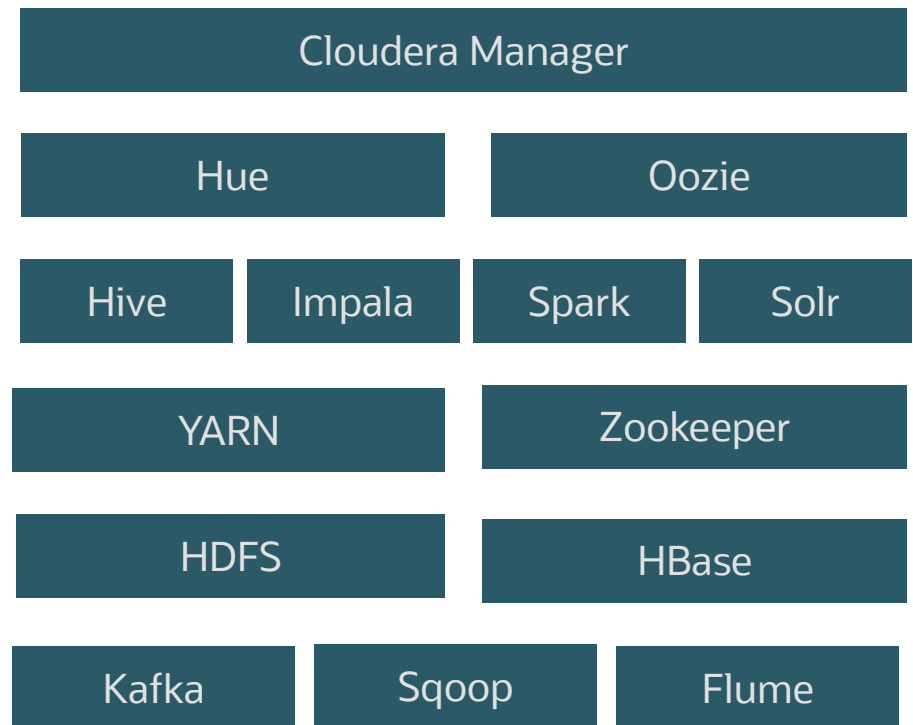
Самая Полная Hadoop Платформа

Использует Cloudera Enterprise Data Hub

- Вертикальная интеграция
- Spark, Kafka, Impala, Hive, Solr, и т.д.
- Функции Безопасности и Управления автоматически устанавливаются автоматически deployed

Выберите свою версию Cloudera

- Соответствуйте вашему текущему разворачиванию
- Разверните новые версии



Sentry Authorization, Navigator Audit

Big Data Service 1

Провиженинг и Эластичность

- Провиженинг полностью настроенного кластера Hadoop и Spark по требованию. Возможность расширения в зависимости от нагрузки.
- Расширение Hadoop и Spark кластера по требованию, добавление дополнительных ОСРЦ для кластера.



Big Data Service 2

Комплексный софт

- Big Data Connectors: Oracle R Advanced Analytics for Hadoop – Разрабатывайте модели на R и запускайте их в параллели на Hadoop и ускоряйте используя Spark
- Smart Data Movement:
 - Oracle Data Integrator – Трансформируйте и обогащайте данные внутри вашего big data кластера
 - Oracle SQL Connector for Hadoop and Oracle Loader for Hadoop – Интеграция данных в режиме online или offline
- Big Data Spatial and Graph: Обработка гео-данных и аналитика графов



Аналитика в Oracle Big Data

Oracle Machine Learning

Scalable, distributed analytics
for Big Data

Statistics & Advanced Matrix
Computation

Regression Models

Classification Models



Oracle Spatial Analysis

Proximity and containment
analysis

Location Data Enrichment

Vector and raster data
preparation



Oracle Graph Analysis

Property Graphs

Social Media relationships

Churn analysis

Cyber-Security

Object recognition



Правильный Кластер для Любой Нагрузки

Development



Test



Data Science



Data Lakes



VM Standard
1-24 OCPUs,
15-320 GB RAM
Up to 1 PB Block



**VM
Dense IO**
8-24 OCPUs,
120-320 GB RAM,
6.4-25.6TB NVMe



**Bare Metal
Standard**
52 OCPUs,
768 GB RAM
Up to 1 PB
Block



**Bare Metal HPC
Dense IO**
36 OCPUs (3.7 GHz),
384 GB RAM
6.7TB NVMe



**Bare Metal
Dense IO**
52 OCPUs,
768 GB RAM,
51.2 TB NVMe

Производительность

Гибкость



Примерная стоимость

▲ Oracle Big Data Service - Compute - Standard - OCPU Per Hour (B91128) / 72

Ocpu Per Hour

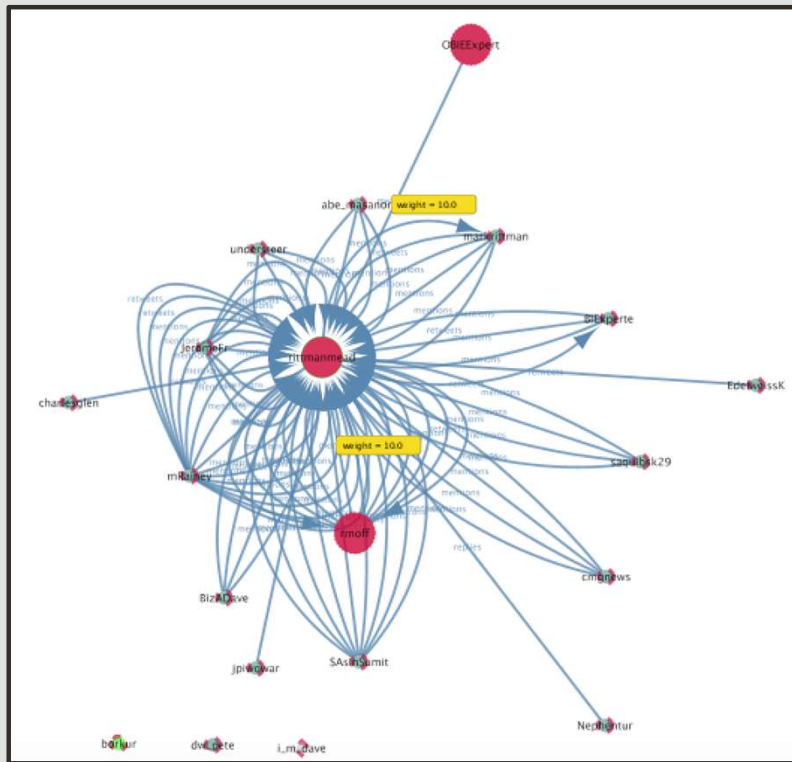
▲ Oracle Big Data Service - Compute - Dense I/O - OCPU Per Hour (B91129) / 72

Ocpu Per Hour



Graph Analytics

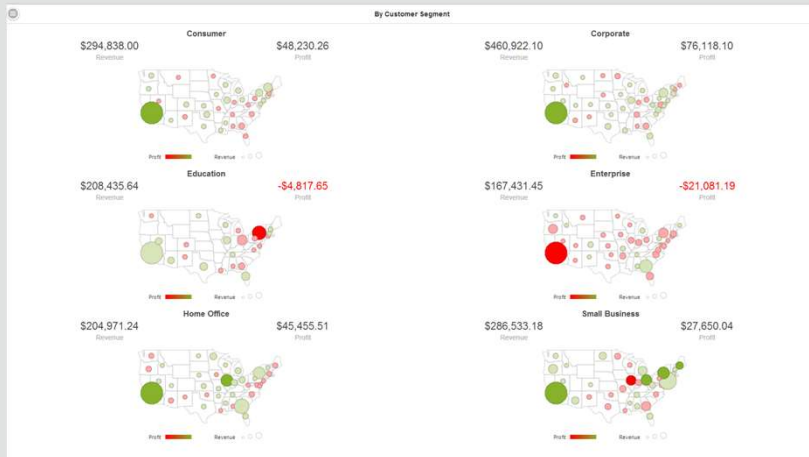
Oracle Big Data Spatial and Graph



- Highly scalable, extreme performance graph analytics and query
 - In-memory analytics
- 50+ pre-built analytics algorithms
- Powerful graph query language PGQL
 - SQL like language for specifying graph patterns
- Rich UI
 - Notebook, CLI, Java API, Graph visualization
- **Enterprise capabilities** – built on Oracle infrastructure
 - Manageability, fine-grained security, high availability, integration, and more

Spatial Analytics and Preparation

Oracle Big Data Spatial and Graph



- Supports both spatial processing and spatial analytics
 - Rich set (~50) of spatial operators and functions
 - Both Vector and Raster services
 - Includes map visualization
- Application-centric approach vs. GIS-centric
 - Works on any data that includes location info, datatype and file format
- Geo-Enrichment services including global geographic hierarchy
- Commercial, supported software

Machine Learning

Oracle Machine Learning



- Use standard R interface to computing statistics, data preparation, joins and view creation over Hive data
- Up to 10x plus faster model building than comparable open source offerings
 - Analyze billions of rows of data using Spark in the presence of limited memory
- 20+ integrated algorithms
 - Custom Spark and Map Reduce implementations
 - Wrapped Apache Spark MLlib algorithms
- Write custom Map Reduce jobs using R and CRAN packages
- Commercial, supported software with Oracle R Advanced Analytics for Hadoop

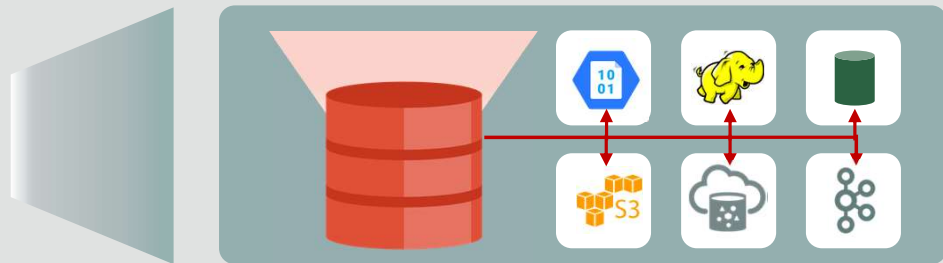
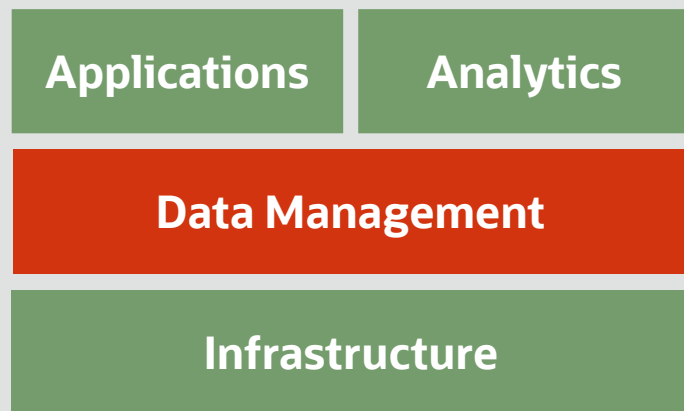
ORACLE

Oracle Cloud Infrastructure Data Catalog

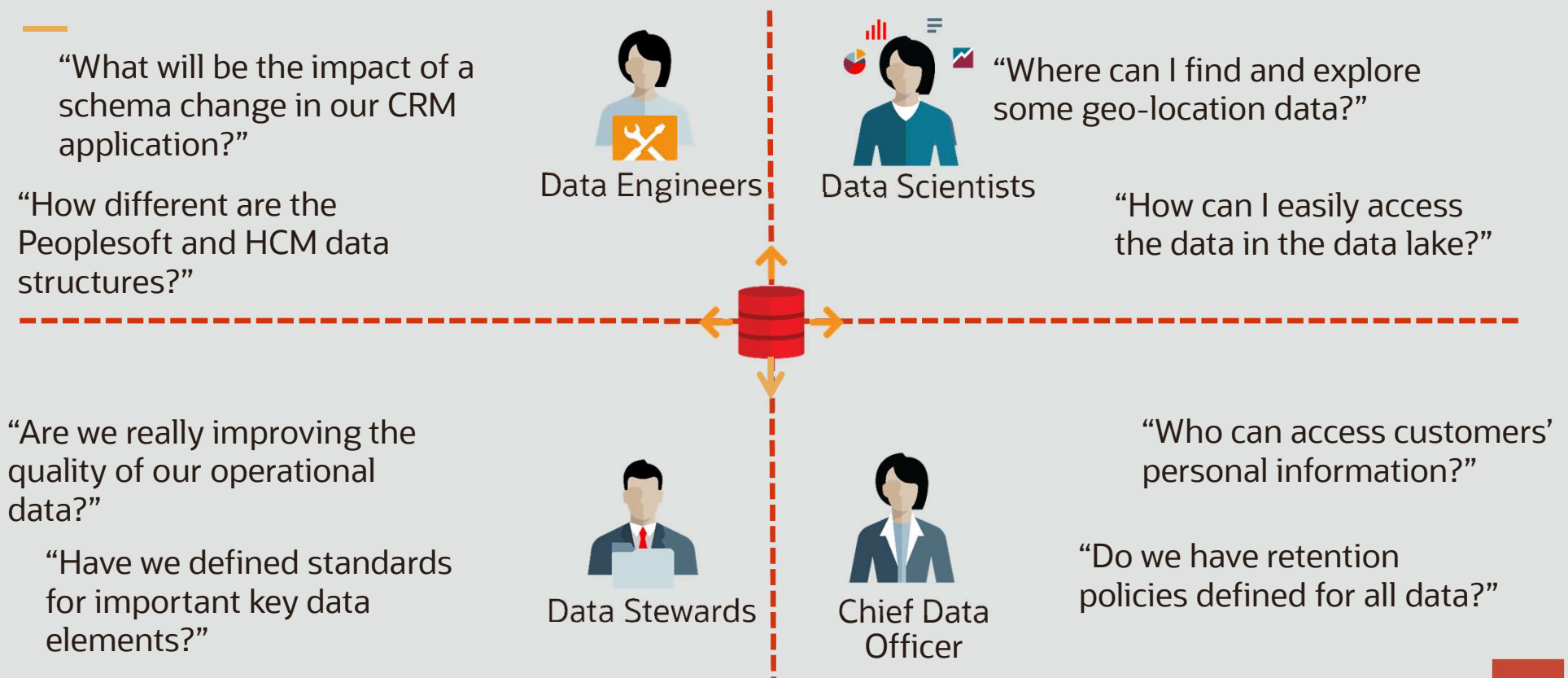
Copyright © 2019 Oracle and/or its affiliates.



Oracle Cloud

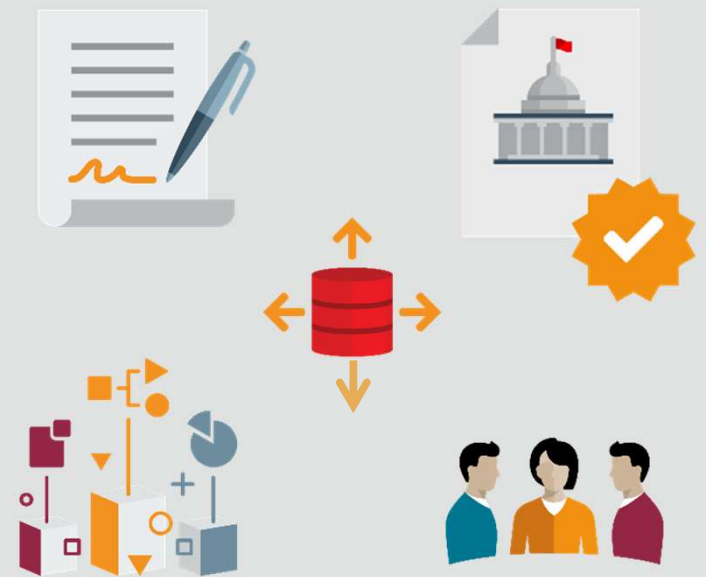


Finding trusted data is still a challenge...

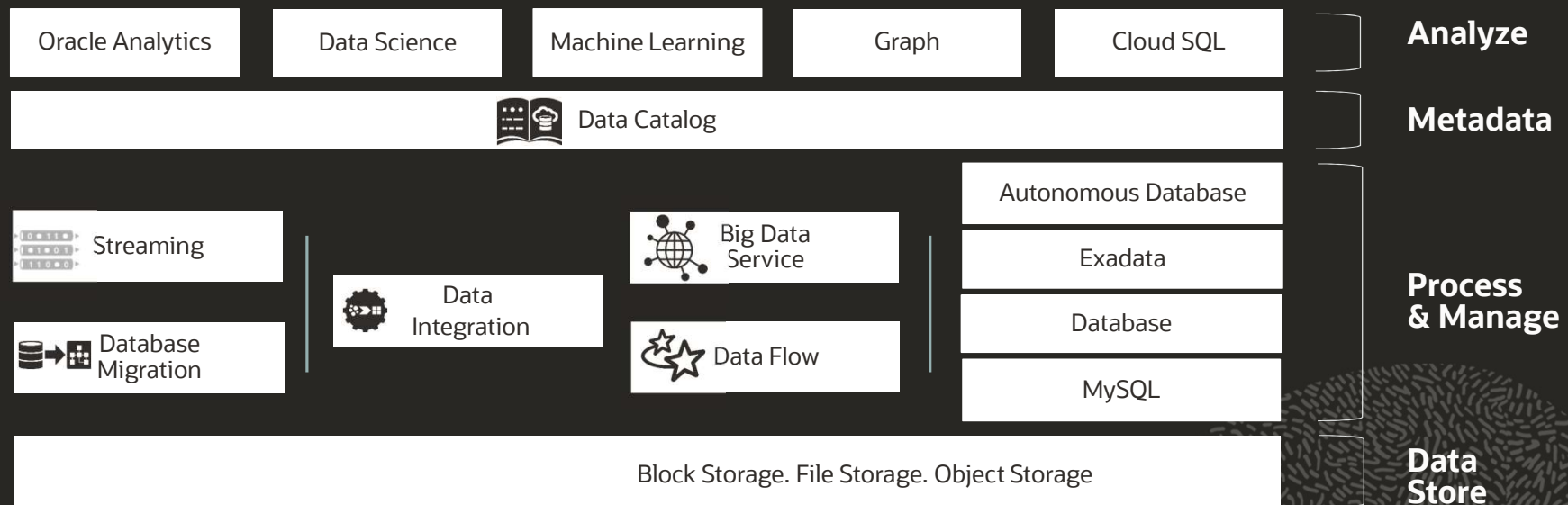


Complexity, Regulations: Governance

- What data do I have?
- Where did it come from?
- How is it moving in the organization?
- Who can access it? Who do I share it with?
- Can I trust it? What can I do with it?
- Is it sensitive? Is it Protected?



Data & AI Services Platform Vision



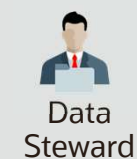


OCI Data Catalog: Find Trusted Data

Self-service discovery

Foundation for data governance

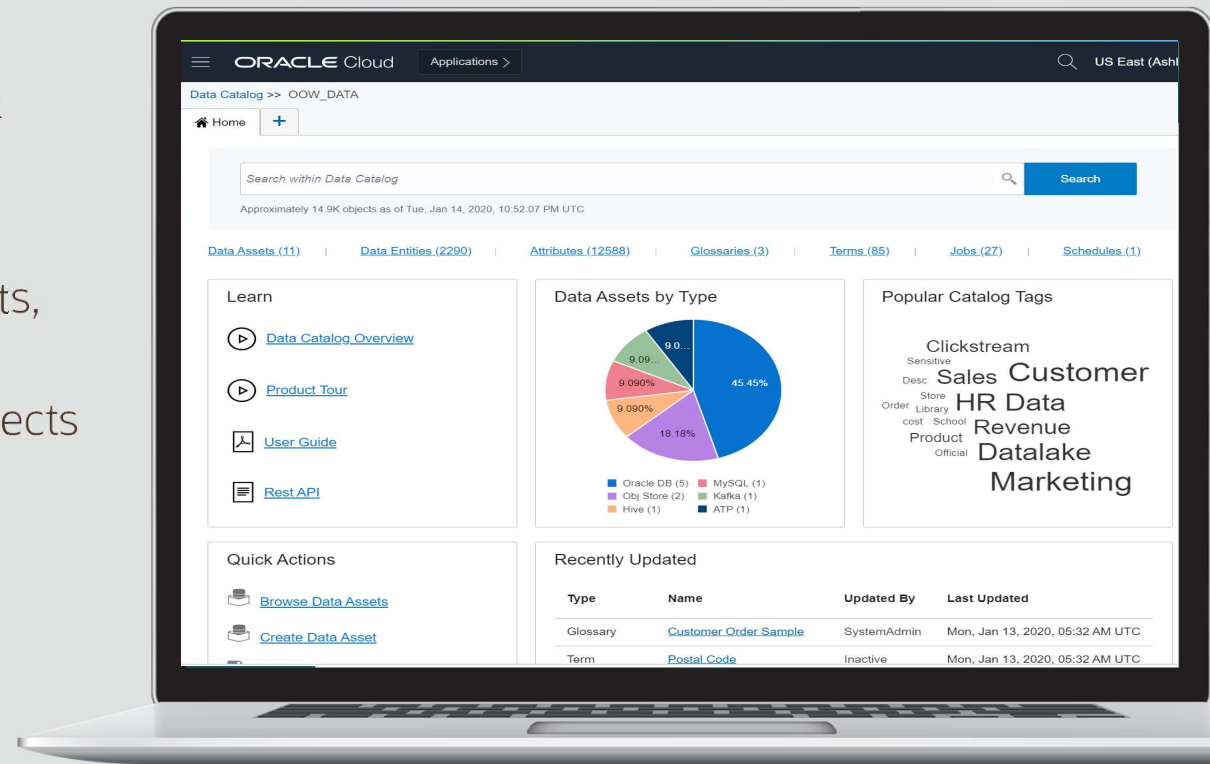
Enterprise-wide data intelligence





Data Catalog: Single Collaborative Environment

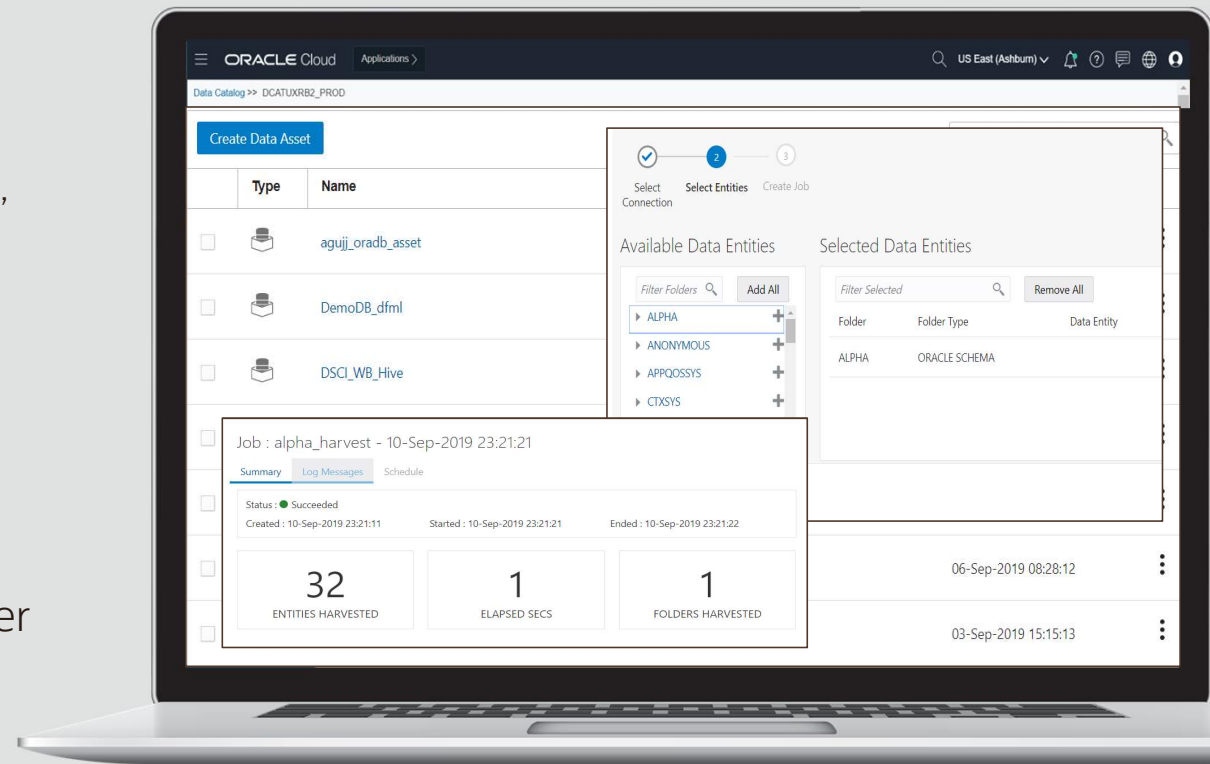
- Home Page with helpful shortcuts & operational stats
- Search and Browse
- Quick Actions to manage Data Assets, Glossaries, Jobs and Schedules
- Popular Tags, Recently Updated Objects





Data Catalog: discover all your data

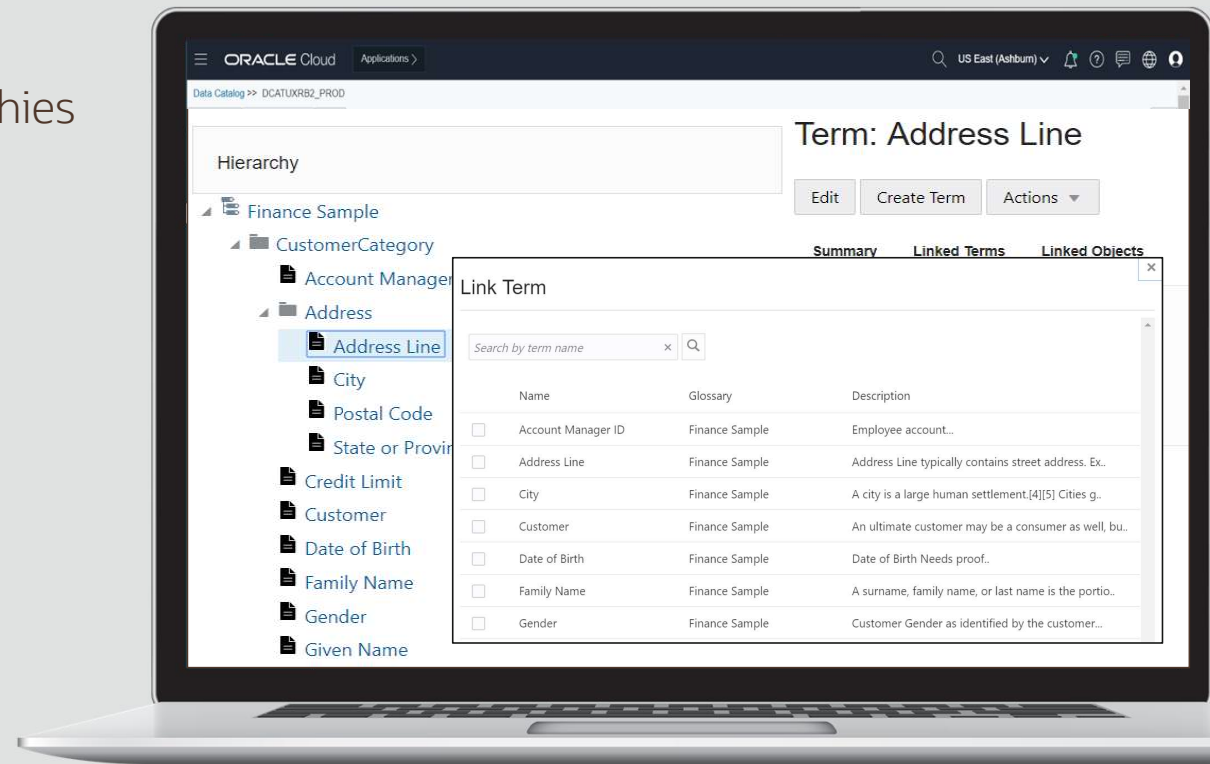
- Catalogs data from multiple sources:
 - Object Storage (CSV, ORC, Avro, Parquet, JSON, XLSX)
 - Oracle Database, ATP, ADW
 - MySQL
 - Apache Hive, Apache Kafka
- On demand and schedule based
- REST API support to integrate with other applications





Data Catalog: Business Glossary and Enrichment

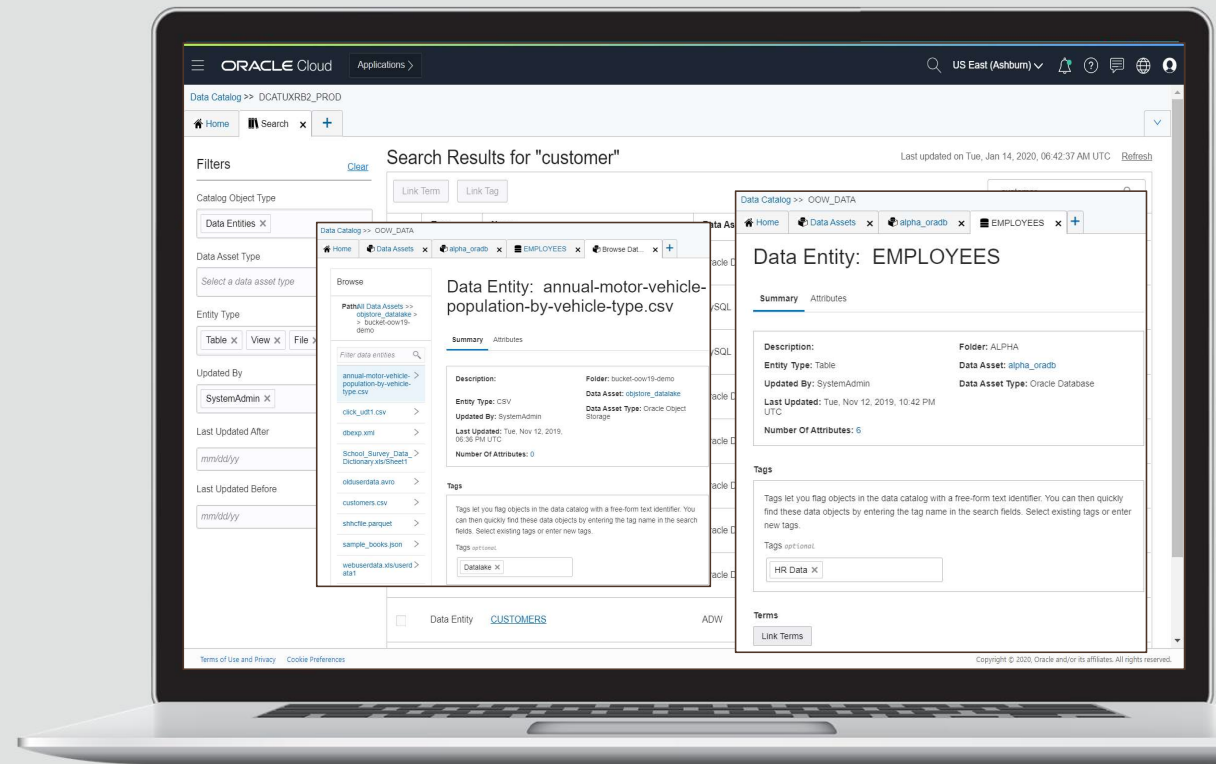
- Manage Business glossaries with hierarchies as simple taxonomies or full glossaries
- Define Business Terms and Categories
- Free form Tags for user annotations
- Assets links to business terms and tags





Data Catalog: Search & Exploration

- Search Data based on technical names, terms, tags
- View details of various objects
- Browse Catalog based on Data Assets



Why OCI Data Catalog

- Fully managed with no additional cost
- Business glossary
- Easy application integration



The Data Asset Inventory is the core of the Data Catalog

Create an inventory of data assets to enable:

- Relevant data to be found quickly and used in a variety of ways

 - For example, as a Data Scientist 'shopping for data', a Data Engineer migrating data, etc.

- New data assets to be registered quickly and easily

 - To allow Data Scientists to work in sandboxes but still use the Catalog as a central connection point

- Enterprise reuse of prepared data

 - Make sure that effort spent preparing data is reusable

- Knowledge sharing

 - To ensure that knowledge about data assets is captured and shared across the organization

- Governance of data assets

ORACLE

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Restricted For Internal Use Only

Classification of data assets enables data ownership, access control and



Data Catalog Functionality (18.4.3 – MVP)

Discover (Harvest)

- Metadata harvesting from:
 - Object Store (CSV, XLS, Parquet, JSON)
 - Oracle DB
 - MySQL
 - Hive
 - Kafka
- Schedule based crawlers for automatic harvesting*

Organize

- Metadata repository and persistence
- Asset Tagging
 - Free form / Glossary based
- Extensible Metadata Model
- REST API
- Event driven metadata integration

Enrich

- Business glossaries with hierarchies usable as simple taxonomies or fuller glossaries
- Optional data sampling to aid understanding

Find and Trace

- Web based UI
 - Browse
 - Search
 - Manage Glossaries
 - Trace Data Lineage
 - Register Assets
- Lineage graphs for SPARK and Hive
- Audit logging

Functionality exposed through its AuthZ and REST APIs

Detailed Scope of first release: Data Catalog for Big Data

Connectivity: Oracle Big Data Cloud, Oracle Object Store, HIVE, Oracle DB, MySQL DB, Kafka, ADWC, Autonomous Transaction Processing

File Type Support: CSV, Parquet, ORC, Simple JSON

Object Support in Harvesting: Database Tables and Views, File Structures including multi-file

Data Entity Info from Harvesting: Number of rows, Size, Attributes and Types, Data Sample

Data Entity Enrichment Options: Tags, Links to Glossary Terms, Owner, Permissions

Search: Find Objects by name, tag or linked term. Refine by Date, Object Type etc.

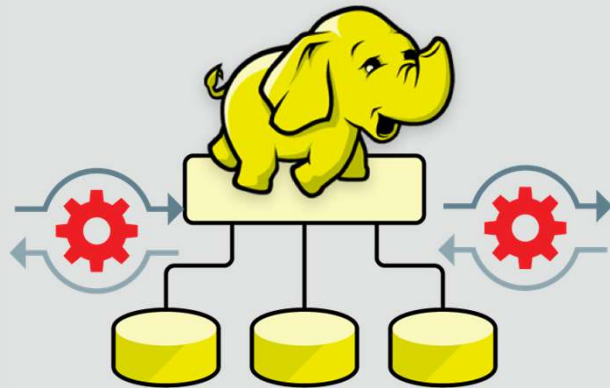
Glossary: Provides a framework for 'governed tags' with formal definitions, in a hierarchy. Simple term status. Linkage to data entities and data entity attributes. Browse for data entities using the glossary. Search using glossary links. Import and Export support.

Data Lineage: Entity-level lineage for data entities in Spark applications or HIVE statements on BDC

Audit and Access Control: Privilege control over key functions – harvesting, viewing data, modifying glossary, querying audit logs. Audit logs accessible by API only.

Data Connectivity

Oracle Big Data Connectors



- Connect Hadoop and Spark directly to data in Oracle Database
 - Parallel Access to Oracle Tables
 - Adhere to all Security in Oracle for Access
- Enable fast ETL by leveraging the Hadoop compute infrastructure through OLH
- Enable a wide variety of Oracle Database version with OSCH
- All Supported with Oracle Data Integrator EE

ORACLE

Истории Успеха

Copyright © 2019 Oracle and/or its affiliates.



ВТБ24

«Наша ИТ-стратегия опирается в том числе на трансформацию ИТ-инфраструктуры на базе технологических инноваций.

Интегрированный стек продуктов Oracle отвечает потребностям бизнеса банка, а выбор Oracle Big Data Appliance, снижая общую стоимость хранения, дает не только экономию по стоимости владения данными, но и открывает перспективы для их превращения в реально работающий актив.»

Сергей Русанов,
член правления, директор департамента банковских
и информационных технологий банка ВТБ24

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



Wargaming.net



«В день мы накапливаем в наших системах хранения данных около 300 млрд телеметрических событий – и это только от World of Tanks на ПК. Добавить к этому поток данных от других игр и платформ – и становится понятно, какой океан информации захлестывает Wargaming.»

Крейг Фрайар (Craig Fryar), руководитель подразделения бизнес-аналитики компании Wargaming

История успеха на Oracle.com: [ссылка](#)

Блог на Oracle.com: [ссылка](#)

Статья на TAdviser: [ссылка](#)



Ссылки на Youtube

- Применение озер данных в компании Megafon - <https://www.youtube.com/watch?v=Aexlo-D6Czc>
- Oracle Helps CERN Use Big Data to Explore the Universe - https://www.youtube.com/watch?v=u8w0e_kelX8
- Wargaming Big Data Journey - <https://www.youtube.com/watch?v=bt5FGlDhdYw>



ORACLE

Сколько Стоит?

Copyright © 2019 Oracle and/or its affiliates.



Oracle Big Data Service

▲ Oracle Big Data Service - Compute - Standard - OCPU Per Hour (B91128) / 72

Ocpu Per Hour

▲ Oracle Big Data Service - Compute - Dense I/O - OCPU Per Hour (B91129) / 72

Ocpu Per Hour



Oracle Big Data Appliance & Software

Big Data Appliance

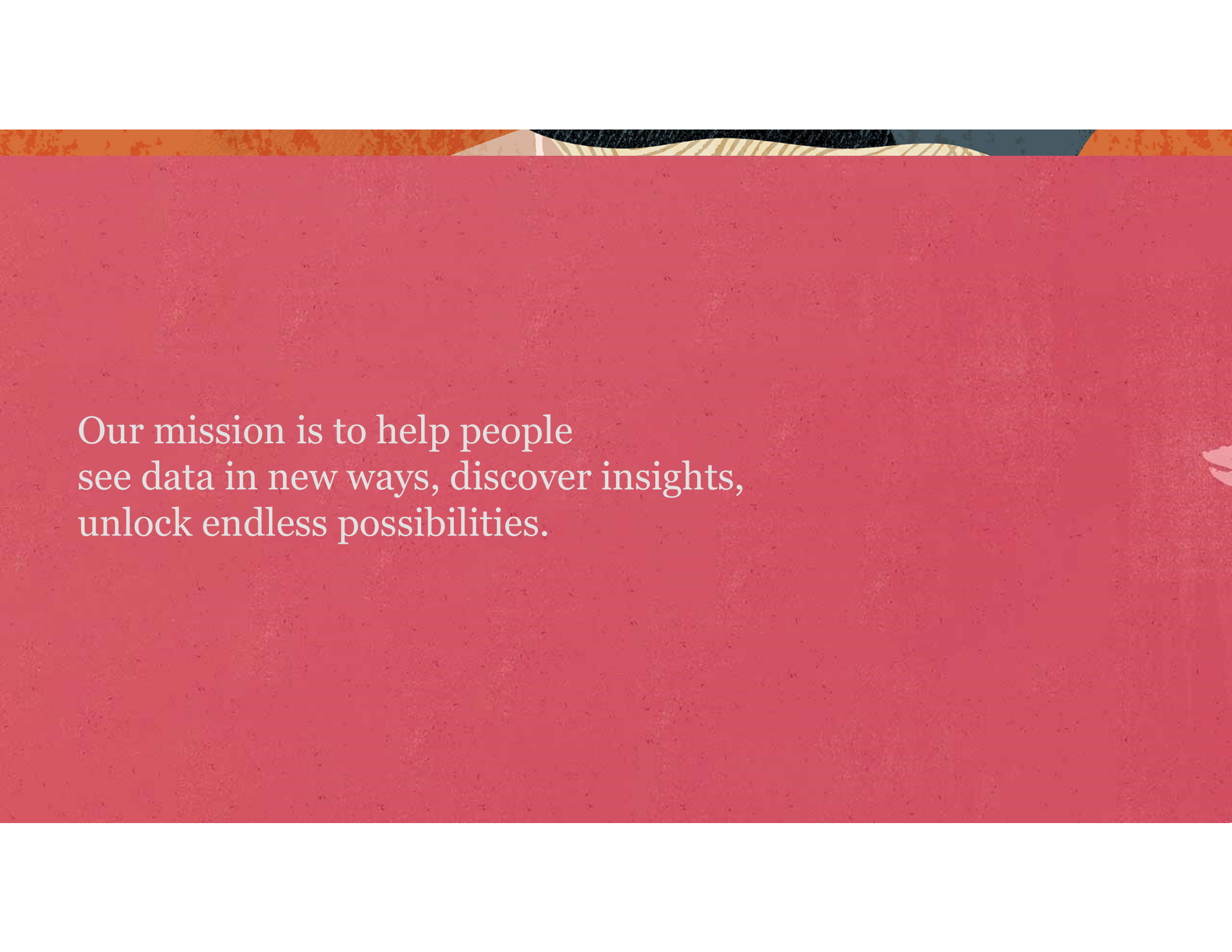
Big Data
Appliance Price

Big Data Appliance

Big Data Appliance X8-2 High Capacity (HC) Node Plus InfiniBand Infrastructure

38,500



The image features a large, solid red area with a fine, grainy texture that occupies the lower two-thirds of the frame. Above this red area is a horizontal band containing abstract, geometric shapes in shades of orange, yellow, and black. The text is positioned on the left side of the red field.

Our mission is to help people
see data in new ways, discover insights,
unlock endless possibilities.



ORACLE