

Курс «Хранилища данных»

Тема: ПО Oracle для построения
ХД

Барабанчиков
Игорь Витальевич

Типовые блоки BI-систем

Основные возможности BI-систем развиваются по четырем основным направлениям:

- **хранение данных,**
- **интеграция данных,**
- **анализ данных**
- **представление данных.**

Хранение данных

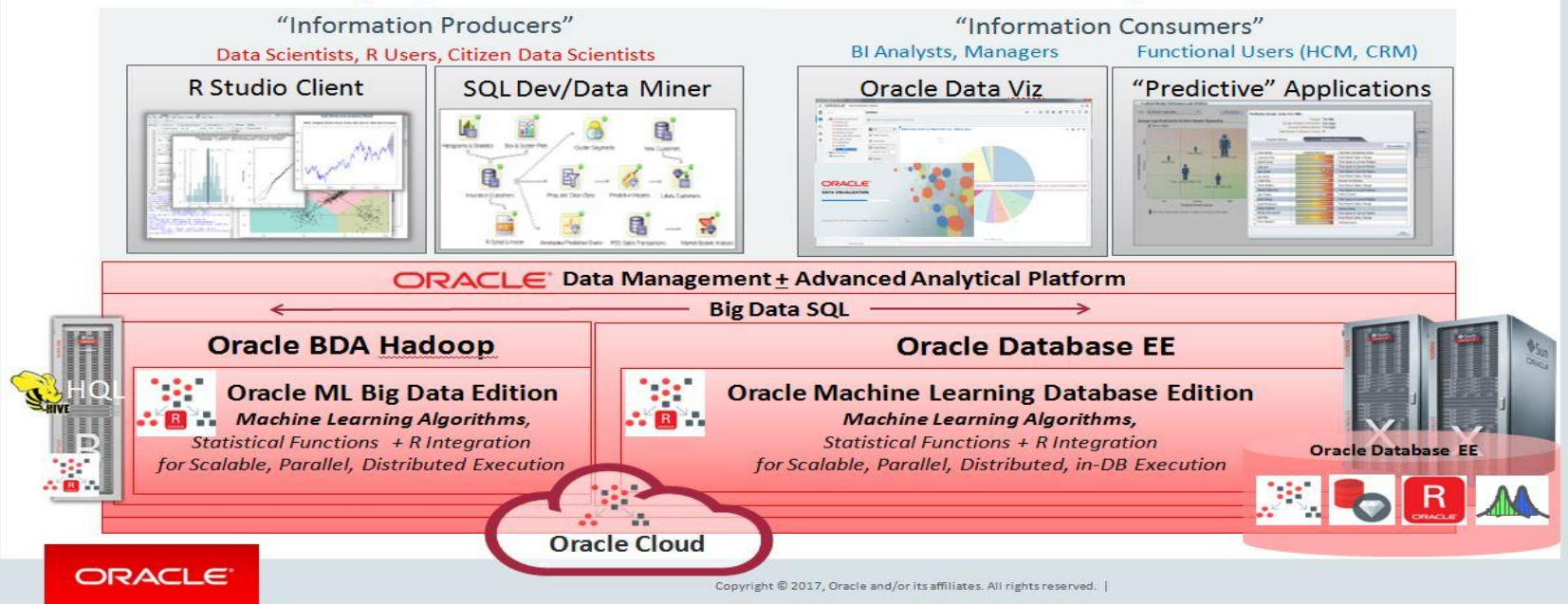
- Oracle Database – реляционная СУБД
- **Позволяет реализовать разные варианты серверов OLAP:**
 - Реляционный
 - Многомерный с помощью опции **Oracle OLAP**.
- Поддерживает опции:
 - **Oracle Partitioning**: предоставляет возможность секционирования таблиц и индексов

Oracle Advanced Analytics

Включает в себя два основных компонента:

- [Oracle Data Mining](#) — интеллектуальный анализ данных
- [Oracle R Enterprise](#) — Oracle внедрил поддержку языка R

Oracle's Machine Learning and Advanced Analytics Platforms Machine Learning Algorithms Embedded in the Data Management Platforms



Oracle Advanced Analytics

- Ключевым принципом опции Advanced Analytics Option, относящимся как к data mining, так и к R, является **приближение обработки к данным.**
- Вместо того, чтобы извлекать данные из базы, а также из файлов или других источников, а затем загружать их в сервер статистической обработки, **можно выполнять R-скрипты и алгоритмы data mining непосредственно в базе данных, что позволяет легко применять модели и классифицировать данные в режиме реального времени**

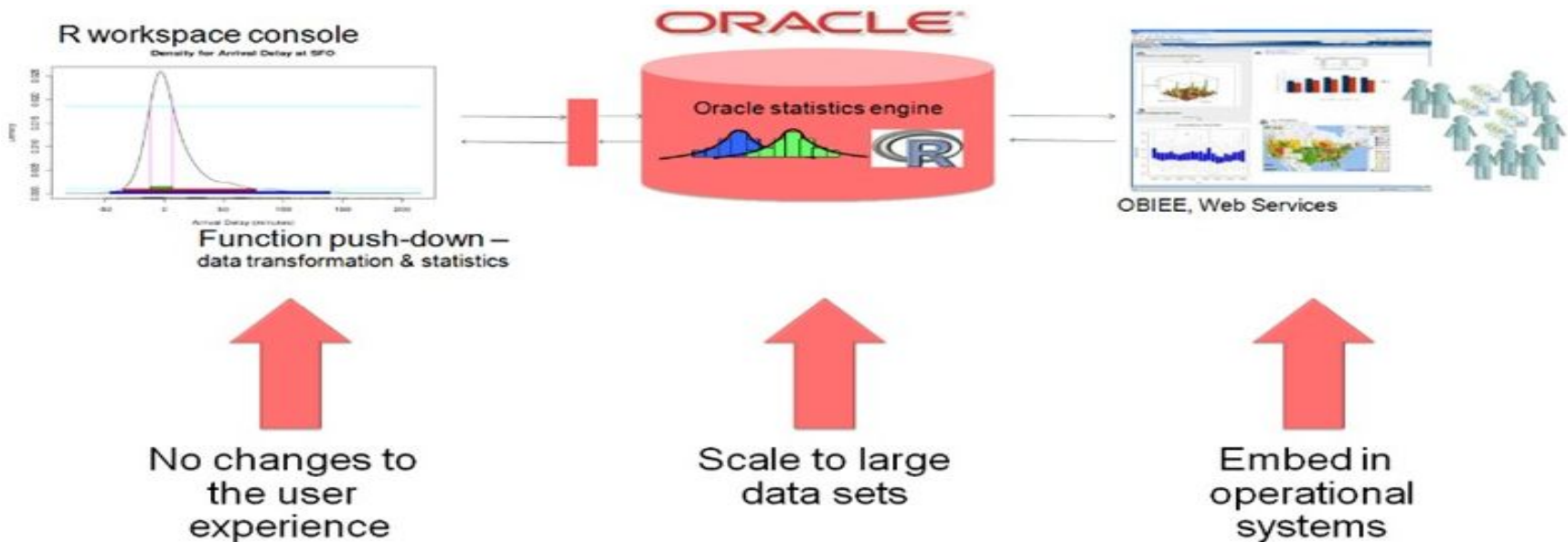
Oracle R Enterprise

Oracle R Enterprise представляет собой базовый R-пакет, расширенный для работы внутри базы данных Oracle следующими (лицензируемыми) элементами:

- R-пакеты для расширения стандартных пакетов, входящих в состав open-source R;
- Библиотеки базы данных для соединения с Oracle и запуска R-скриптов в базе данных;
- SQL-расширения для вызова R-функций

Преимущества

- **Не надо переучиваться** (есть опыт работы с R)
- **Масштабирование обработки** на большие объемы данных.
- **Возможность встраивания** интеллектуальных алгоритмов в свои системы.



Oracle's Adv. Analytics Machine Learning Algorithms



Classification

- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Explicit Semantic Analysis
- Gaussian Mixture Models



Clustering

- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)



Anomaly Detection

- One-Class Support Vector Machine (SVM)



Regression

- Generalized Linear Model
- Support Vector Machine (SVM)
- Random Forest
- Linear Model
- Stepwise Linear regression
- LASSO



Association Rules

- A priori



Attribute Importance

- Minimum Description Length
- Principal Component Analysis (PCA)
- Unsupervised Pair-wise KL Divergence



Predictive Queries

Statistical Functions

- Basic statistics: median, stdev, t-test, F-test, Pearson's, Chi-sq, Anova, etc.



Algorithm Support for Text

- Algorithms support text type
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA) for document similarity



Feature Extraction

- Principal Component Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)

Time Series

- Single Exponential Smoothing
- Double Exponential Smoothing

Open Source ML Algorithms

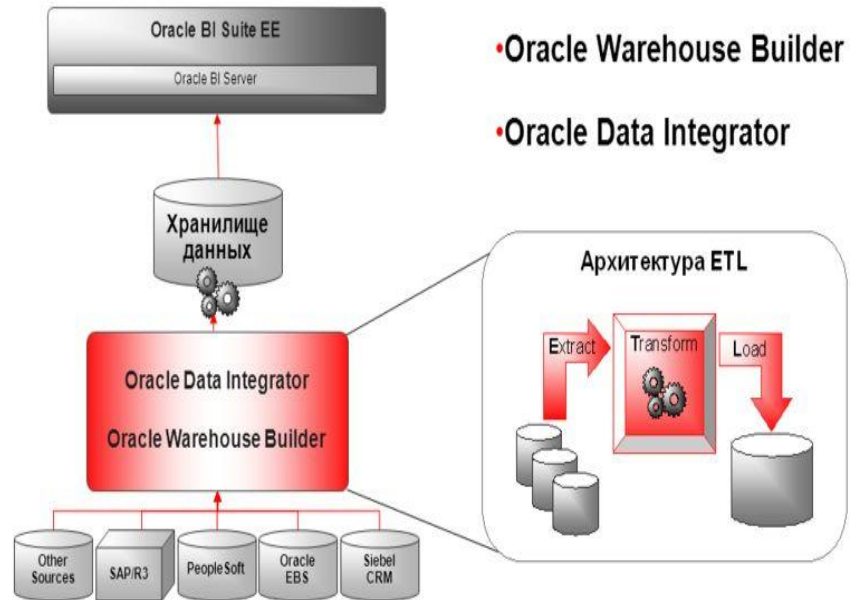
- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLlib algorithm integration



Инструменты ETL

- Oracle Warehouse Builder
- Oracle Data Integrator
- Oracle Golden Gate

■ ETL – средства (Extract – Load- Transform)

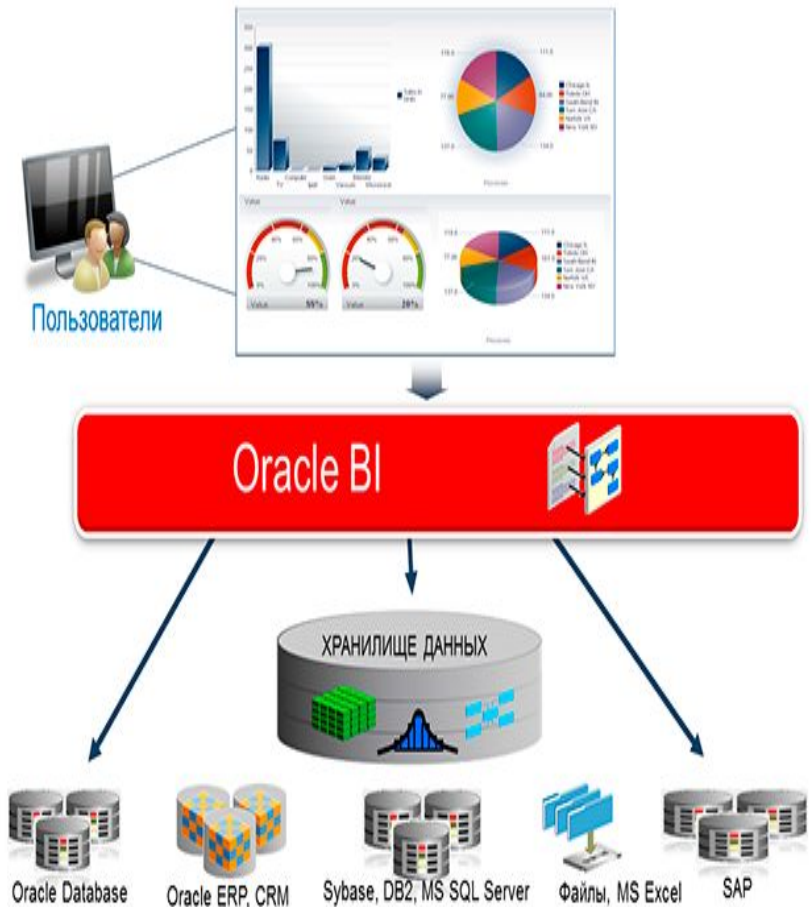


Oracle BI EE

Имеет 3-х уровневую архитектуру:

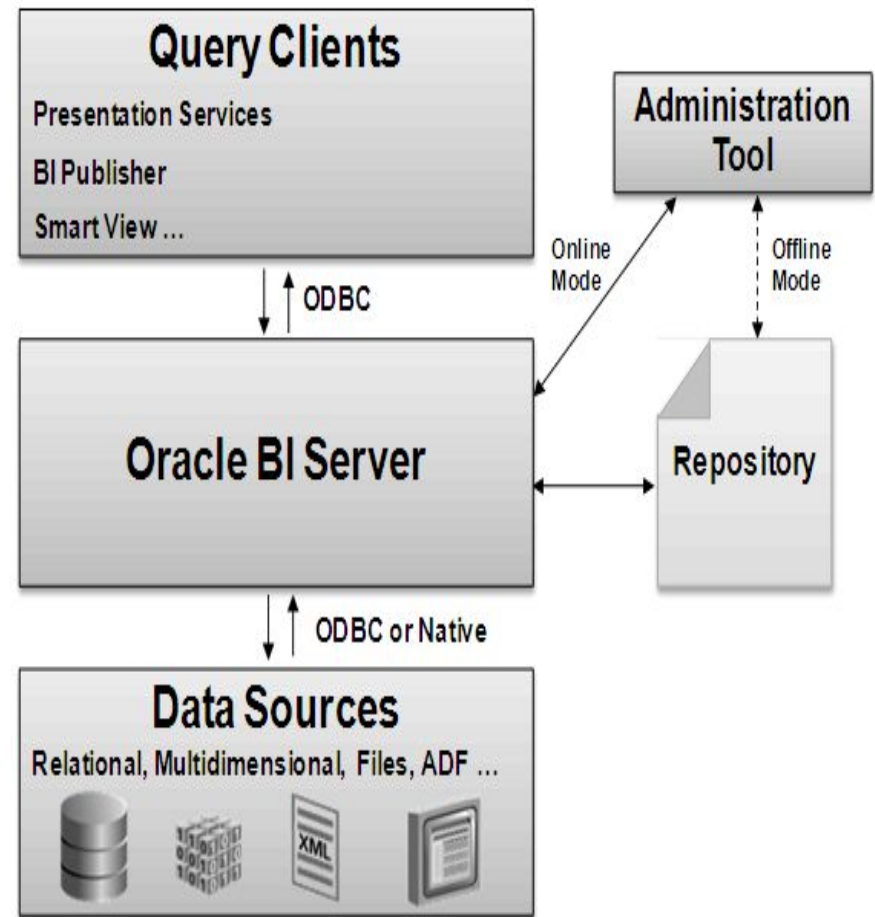
- **Источники данных**
- **Сервер Oracle BI**
- **Тонкий клиент**

Сервер может делать сложные вычисления, которые нельзя выполнить в источнике данных.



Oracle BI – виртуальное ХД

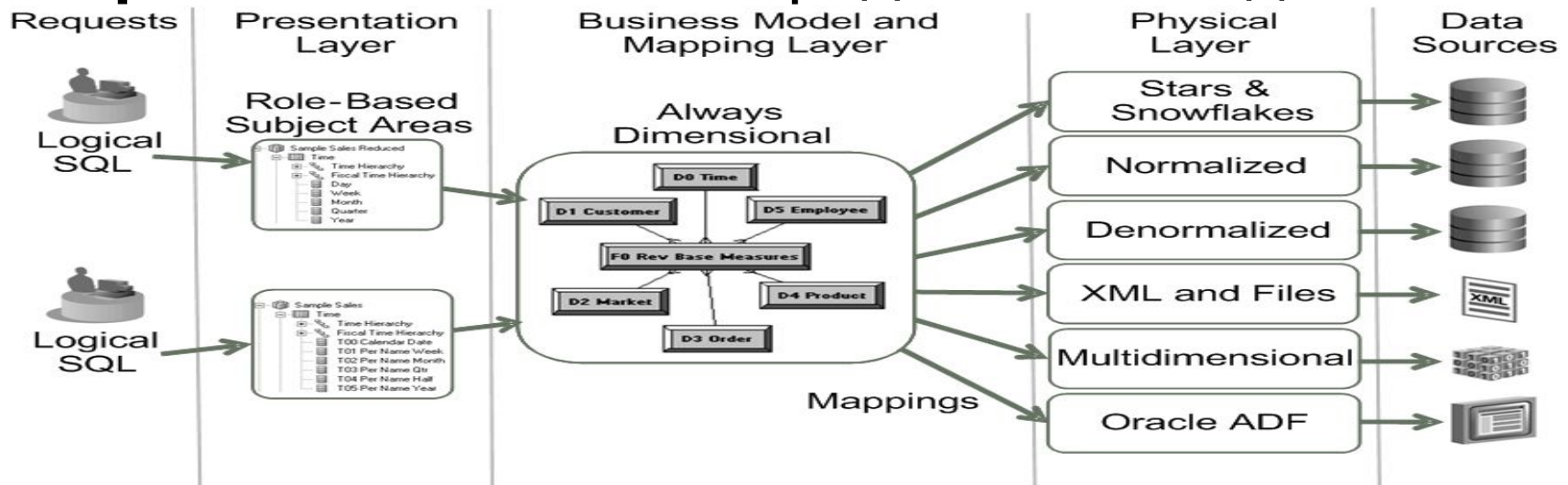
- **Репозиторий содержит метаданные,** которые позволяют отобразить источники данных на многомерный куб.
- Разработка репозитория Oracle BI выполняется с помощью программы **Administration Tool**.



Репозиторий Oracle BI

Состоит из 3-х уровней:

- **Физический** – соединение с источниками данных
- **Бизнес-модели** – отображение физических источников на многомерный куб
- **Презентационный** – представление данных



Oracle BI Answers

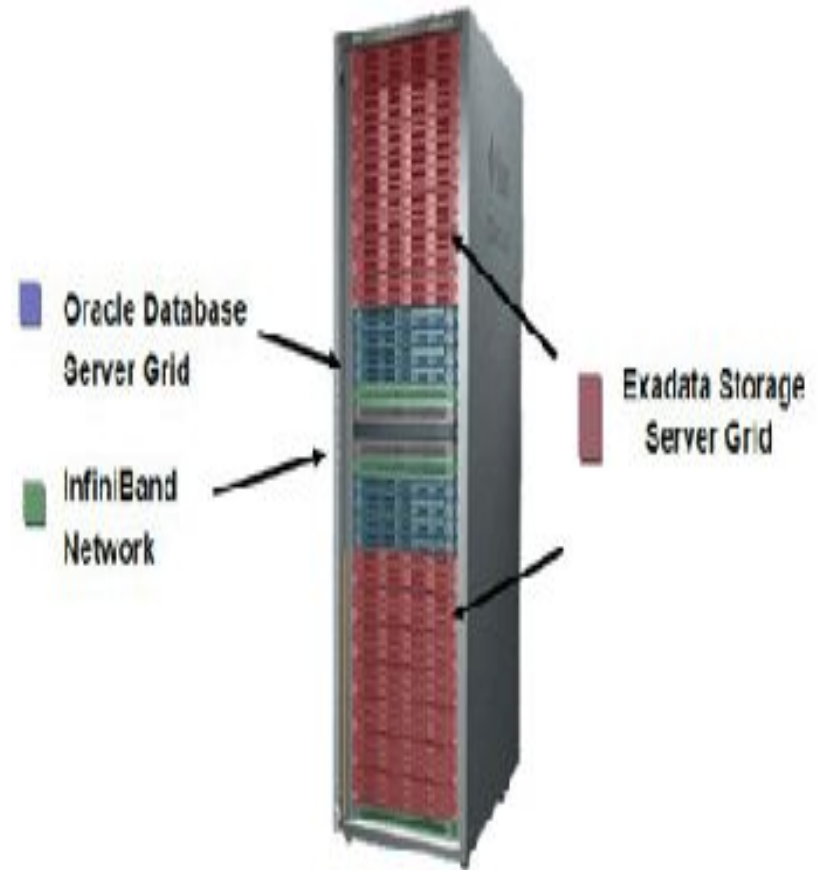
- Работает в браузере
- Имеет простой графический интерфейс.
- Имеет графические средства визуализации данных.
- Позволяет выполнять ad-hoc запросы
- Позволяет создавать информационные панели.



Oracle Exadata

- Программно-аппаратный комплекс, оптимизированный для работы с большими БД.
- Состоит из:
 - серверов БД
 - системы хранения
 - сетевого оборудования
 - специализированного ПО

ПО



Oracle Exalytics

Oracle Exalytics Business Intelligence Machine



- Программно-аппаратный комплекс для бизнес-анализа
- Экстремальная производительность
- Неограниченные возможности визуализации и анализа

Аналитическая платформа

Oracle

Oracle предоставляет все необходимые инструменты и технологии для построения корпоративной аналитической системы

Аналитическая платформа Oracle

ИНСТРУМЕНТЫ БИЗНЕС-АНАЛИЗА и ОТЧЁТНОСТИ

Oracle BI Suite EE

Oracle RTD

ХРАНИЛИЩА И ВИТРИНЫ ДАННЫХ

Oracle
Warehouse Builder

Oracle
Data Integrator

ORACLE DATABASE

Oracle Partitioning

Oracle OLAP

Oracle Data Mining

ORACLE
ESSBASE

АППАРАТНО-ПРОГРАММНАЯ ПЛАТФОРМА РЕАЛИЗАЦИИ ОЧЕНЬ БОЛЬШИХ БД

HP Oracle Exadata Storage Server

ORACLE

ИТОГИ

Продукты Oracle идеально подходят для строительства крупных, центрально-управляемых глобальных решений с широким набором BI-инструментов, продвинутой аналитикой на базе одной платформы.

Корпоративная аналитическая платформа *Oracle Business Intelligence*



ORACLE