

УДК 001.895

НОВЫЕ ПОДХОДЫ УПРАВЛЕНИЯ ДАННЫМИ



Успенский Никита

*Руководитель направления Big Data & Analytics
(Беларусь, Украина, Чехия) в корпорации Oracle*

Oracle Corporation, USA

E-mail: nikita.uspenskiy@oracle.com

Н. Успенский

Руководитель направления Big Data & Analytics (Беларусь, Украина, Чехия) в корпорации Oracle.

Аннотация. Представлены современные подходы управления данными, принятые в корпорации Oracle

Ключевые слова: архитектура управления, архитектура хранилища данных, корпоративное хранилище данных.

Современные подходы управления данными в корпорации Oracle

Концептуальная архитектура управления данными применяемая в корпорации Oracle, представляет собой следующую систему (рисунок 1).

1. Создание витрины данных.
2. Выбор источника данных (с использованием Sales CRM, Financial ERP, Social Data, IoT Data).
- 3 Бесшовная интеграция или управление онлайн-потокком данных. базируется на Data Integration Platform Cloud, а также на Event Hub Cloud.
4. Выгрузка данных или загрузка напрямую в хранилище данных.
5. Высокопроизводительная подготовка и обработка данных.
6. «Сырые» данные для продвинутой аналитики.
7. Исследование данных и самостоятельная аналитика.
8. Представление результатов (с использованием LOB, Analysts, Data Scientists, Developers).

Архитектура современного хранилища данных, применяемая в корпорации Oracle, представляет собой следующую систему (рисунок 2).

1. Средства разработки (Oracle SQL Developer).
2. Сервисы интеграции данных (Oracle Data Integration Platform Cloud, сторонние средства, развернутые в ОСI, сторонние средства, развернутые в ЦОДе).
3. Продвинутая аналитика (Oracle Analytics Cloud, сторонние средства, развернутые в ОСI, Сторонние средства, развернутые в ЦОДе).
4. Oracle Object Storage Cloud (Файлы и подготовка данных).
5. Autonomous Data Warehouse (Управление сервисом, Встроенные средства доступа).

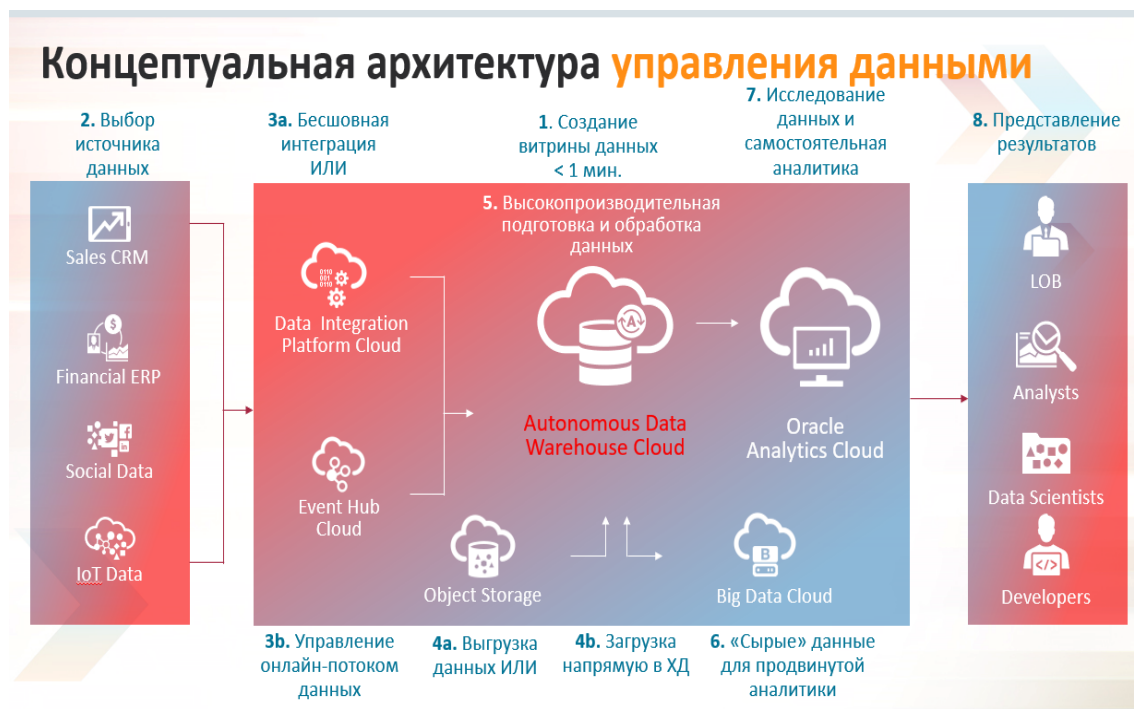


Рисунок 1. Концептуальная архитектура управления данными корпорации Oracle

Архитектура современного Хранилища Данных

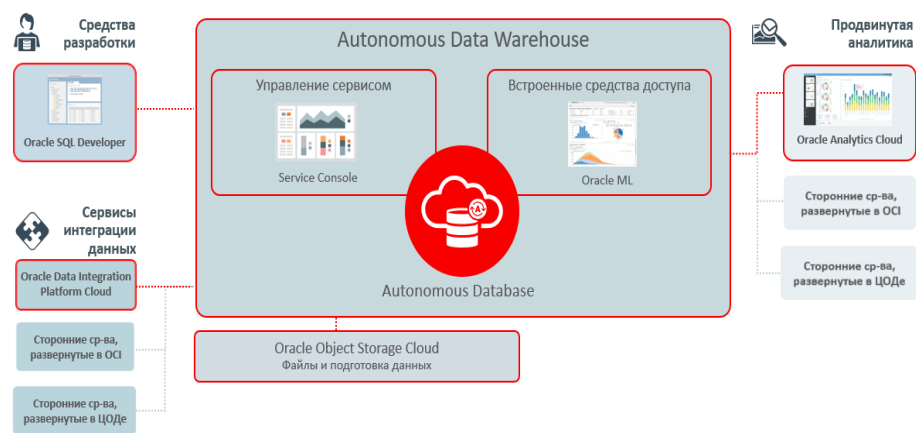


Рисунок 2. Архитектура современного хранилища данных в корпорации Oracle

Схема Analytics & Big Data представлена на рисунке 3.

1) Analytics Cloud характеризуется следующими параметрами: простая и продвинутая визуализация данных; дополнительные возможности машинного обучения: R, Python, интеграция с Apache Spark; современное мобильное приложение.

2) Autonomous Data Warehouse характеризуется следующими параметрами: витрины данных для бизнес-пользователей; быстрое развертывание высокопроизводительной аналитической платформы; минимизация рутинных операций.

3) Big Data Cloud характеризуется следующими параметрами: идеальная платформа для проектов «data science»; гибкий и доступный по цене Hadoop-кластер.

Analytics & Big Data



Рисунок 3. Схема Analytics & Big Data применяемая в корпорации Oracle

Autonomous Data Warehouse Cloud. Oracle Autonomous Data Warehouse Cloud (ADW) обеспечивает следующие сервисы.

1) Автоматизированное хранилище данных. Наиболее передовые средства автоматизации позволяют пользователю легко загрузить данные и сразу начать с ними работу.

2) На базе Exadata. С помощью технологий Oracle Database и Exadata обеспечиваются максимальная производительность и многопользовательский доступ к данным.

3) Обеспечивает любой размер. «Эластичность» сервиса позволяет мгновенно масштабировать хранилище данных, что приводит к значительному сокращению затрат.

Oracle ADW обеспечивает следующие преимущества: значительное увеличение производительности; снижение объема данных за счет компрессии; сокращение совокупной стоимости владения (TCO); гибкое масштабирование ресурсов; быстрое и простое развертывание; автоматизация рутинных операций; интерфейс исследования данных для бизнес-пользователей – Data Visualization Desktop.

Основные области применения ADW:

1) Витрины данных: аналитика для структурных подразделений и подготовка пользовательских отчетов; быстрый поиск бизнес-идей.

2) «Песочницы» для работы с данными: возможность сразу работать над алгоритмами, а не налаживать инфраструктуру; поддержка SQL, Python и R (через Data Visualization Desktop).

3) Срочные и временные аналитические задачи: меньше зависимости от службы ИТ; возможность отключить сервис ХД, когда проект завершен.

4) Обмен данными между компаниями: платформа для консолидированной аналитики и совместных проектов; доступ к данным для внешних консультантов и аналитиков.

Корпоративное хранилище данных обеспечивает: тысячи офисов и десятки тысяч сделок в день; быстрое развертывание новой БД; 8 минут вместо 2 недель ранее; время запуска проекта сократилось на 70%; каждый час, который ранее тратился на администрирование, теперь направлен на инновации

Гибкое масштабирование позволяет получить: быстрое увеличение мощности для обеспечения бизнес-задач при одновременной уменьшении стоимости; формирование отчетов из 1.2 млрд. строк за считанные секунды вместо 20 минут ранее; автоматическое патчирование устранило простой системы и обеспечило работу 24x7.

Быстрая загрузка данных обеспечивает: аналитический консалтинг для спортивных организаций; быстрое развертывание тестовых сред и сокращение времени загрузки данных в 4 раза; машинное обучение позволяет определить, когда спортивным фанатам стоит предложить продление билетов на сезон.

Схема Analytics Cloud представлена на рисунке 4.

Oracle Analytics Cloud

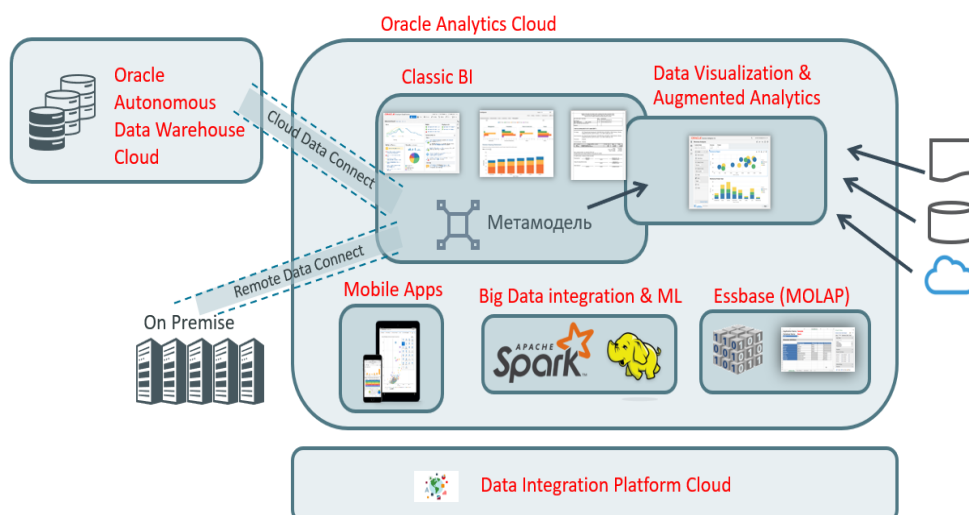


Рисунок 4. Схема Analytics Cloud

Разработано новое приложение (рисунок 5) для телефона, предлагающее нужную бизнес-аналитику в смартфоне.



Рисунок 5. Приложение Day by Day

Централизованная аналитическая платформа для аналитики и отчетности международной компании. Для крупнейшей бразильской железнодорожной транспортной компании (рисунок 6) необходимо выполнить задачу: использовать данные из разрозненных систем для предотвращения инцидентов, снижения затрат и количества внеплановых остановок. Внеплановые остановки обходятся \$100,000+ в час. Решение: Oracle Analytics Cloud и Big Data Cloud Service сохраняют и анализируют потоковые данные сенсоров. Результат:

предотвращены 4 инцидента, сэкономлено более \$400,000. Семь минут требуется, чтобы данные были доступны руководству железнодорожной транспортной компании. Запуск системы за 3 месяца.



Рисунок 6. Бразильская железнодорожная транспортная компания

Биометрические ID карты. Поставлена задача: обеспечить надежный доступ к информационным данным биометрических ID карт (рисунок 7) без расширения аппаратной инфраструктуры. Решение: полностью облачное хранилище данных с аналитикой на базе BI Cloud Service и Database Cloud Service. Результат: 70% - увеличение скорости обработки заявок страховщиков; 20% - уменьшение TCO (Total Cost of Ownership) с DBCS (Double Byte Character Set); 5% - быстрее запуск новых продуктов на рынок; на 30% удалось снизить время и ресурсы на поддержку новых поставщиков страховок.



Рисунок 7. Система биометрических ID карт

Oracle Big Data Cloud позволяет быстро, безопасно и недорого получить преимущества работы с большими объемами информации, которые усилены следующей гибкой и интегрированной инфраструктурой:

- 1) Elastic: масштабирование до сотен узлов; масштабирование Compute и Storage независимо друг от друга.
- 2) Managed: обработка данных и кода; обновление программного обеспечения одной кнопкой.
- 3) Highly Available: облачное хранилище как Data lake; нулевая потеря данных и нулевое время простоя.
- 4) Measurable Differentiation: интегрировано с Oracle Database, Data Analytics Stack и Oracle PaaS/SaaS; единая платформа для анализа Big Data и транзакционных данных.

На рисунке 8 представлены некоторые стоимостные характеристики.

Configuration Options	Pay As You Go	Monthly Flex
Oracle Big Data Cloud Service - Compute Edition	\$451	\$308
Big Data - Compute	\$431	\$287
Storage Classic	\$20	\$20

Summary Pricing:

Model	Price
Pay As You Go *	\$451 /mo
Monthly Flex *	\$308 /mo

Рисунок 8. Скриншот стоимостного сайта

Характеристики: 24x7; 2 OCPU; 30 GB RAM; 1 TB HDFS Storage; 1 TB Block Storage (persistent); Traffic (in-bound) – free; Traffic (out-bound first GB) – free.

Корпоративная аналитика розничной сети обладает следующими возможностями. Снижение TCO (Total Cost of Ownership): платежная платформа для лучшего обслуживания клиентов; миграция в облако; гибкие возможности роста с небольших нагрузок; развернуты Big Data Cloud, Oracle Analytics Cloud и Storage Cloud.

Консолидация данных розничной онлайн-сети обладает следующими возможностями. Снижение TCO (Total Cost of Ownership): разгрузка ETL (Extract, Transform, Load, извлечение, преобразование, загрузка) хранилища данных; поддержка маркетинговых кампаний по увеличению среднего чека; гибкие возможности роста с небольших нагрузок; сокращение затрат; уменьшение до 50% операционных издержек.

Заключение

В статье описаны общие направления развития продуктов корпорации Oracle. Содержание статьи предназначено исключительно для информационных целей и не может быть включено в состав каких-либо контрактов. Данная статья не предусматривает обязательств по предоставлению каких-либо материалов, программных кодов или функциональных возможностей и не предназначена для использования в качестве обоснования принятия решений о покупке продуктов корпорации Oracle. Разработка, выпуск, сроки реализации и стоимость любых описанных функций или возможностей продуктов Oracle могут измениться и осуществляются исключительно по усмотрению корпорации Oracle.

NEW DATA MANAGEMENT APPROACHES

N. USPENSKIY

*Head of Big Data & Analytics
(Belarus, Ukraine, Czech Republic)*

Oracle Corporation, USA

E-mail: nikita.uspenskiy@oracle.com

Abstract. Modern data management approaches adopted by Oracle Corporation are presented.

Keywords: management architecture, data warehouse architecture, corporate data storage.