

# **Изграждане на интерфейс за предаване на големи данни от система Hadoop към система Oracle ERP система (E-Business Suite)**

Настоящата разработка е направена в резултат на извършване на неикономическа дейност – фундаментални научни изследвания за ефективно сътрудничество, в резултат от дейност по договора между УНСС и „РИЛА СОЛЮШЪНС“ ЕАД за работа в научната област ИЗПОЛЗВАНЕ НА ГОЛЕМИ ДАННИ В ERP СИСТЕМИ, свързани с изпълнение на проект „Дигитализация на икономиката в среда на Големи данни (ДИГД)“ BG05M2OP001-1.002-0002-C02.

## **Съдържание**

Въведение .....	3
Цел на разработката .....	3
Описание на структурата на разработката .....	3
1. Анализ на възможностите за обмен на данни .....	4
1.1. Видове данни касаещи бизнеса .....	4
1.2. Hadoop като система за обработка на големи данни .....	6
1.3. Oracle ERP като система за обработка на бизнес данни .....	8
2. Проектиране и разработване на интерфейс за големи данни .....	11
3. Тестване на разработения интерфейс .....	14
Източници .....	14

## **Фигури**

<i>Фигура 1: Организационна структура на предприятие .....</i>	<i>6</i>
<i>Фигура 2: Архитектура на Cloudera/Hadoop .....</i>	<i>8</i>
<i>Фигура 3: Трислойна архитектура на Oracle E-Business Suite .....</i>	<i>10</i>
<i>Фигура 4: Базирана на екранни форми десктоп архитектура .....</i>	<i>11</i>
<i>Фигура 5: Интерфейс между Oracle EBS и Hadoop .....</i>	<i>13</i>

## **Въведение**

### **Цел на разработката**

В тази разработка (документ) са представени резултатите от работата по анализ на възможността за изграждане на интерфейс за предаване на големи данни от технологичната система Hadoop към бизнес приложението Oracle ERP (с използване на продукта Oracle E-Business Suite R12) с цел проучване на потенциала за използването на големи данни в различни функционални области.

### **Описание на структурата на разработката**

В **точка първа** са разгледани видовете данни в контекста на обработката им в системите за управление на ресурсите в бизнеса. Анализирани са възможностите на системата за обработка на големи данни Hadoop. Анализирани са възможностите на бизнес приложението Oracle E-Business Suite за моделиране на процесите в бизнеса.

В **точка втора** е описано проектирането и разработването на интерфейс за обмен на данни между двете системи с цел използването им за обработки в различни бизнес процеси.

В **точка втора** са представени резултатите от проведените тестове.

Разработката завършва със списък на използваните информационни източници.

# **1. Анализ на възможностите за обмен на данни**

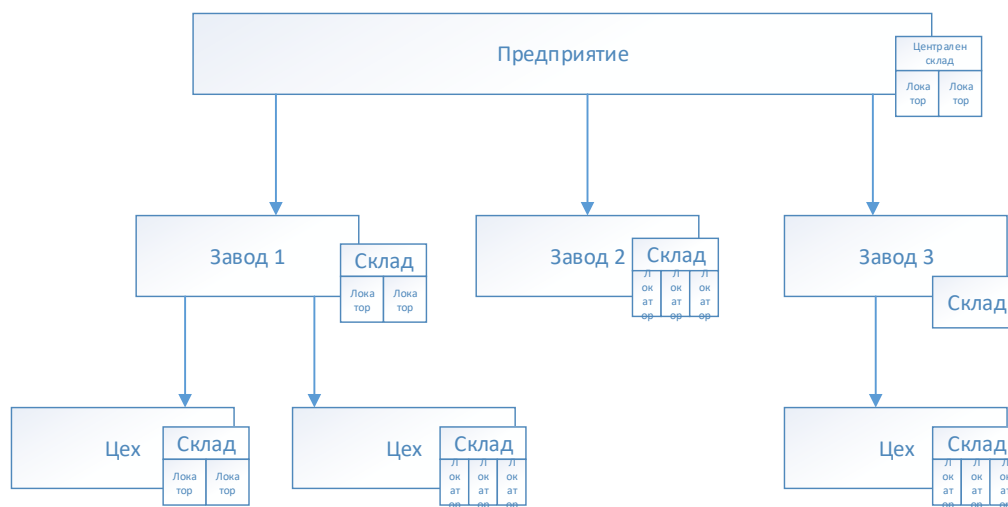
## **1.1. Видове данни касаещи бизнеса**

Системите за управление на бизнеса, и в частност системите за управление на ресурсите на предприятието, обикновено обработват голямо по обем и разнообразие количество структуриране данни. За да се изгради правилно системата тези данни се разпределят в следните видове:

1. Структурни данни – описват структурата и взаимовръзките между отделните организационни звена на предприятието, като заводи, отдели, факултети, департаменти, складове, локатори, производствени единици, военни поделения и други.
2. Параметрични данни – относително постоянни данни, които се залагат като номенклатури в системата, променят се или се допълват относително рядко, като например списък на персонала, номенклатура на доставчици и клиенти, номенклатура на материали, продукцията, стоки и услуги, рецептурници, локации и други.
3. Оперативни данни – текущи данни, постъпващи през ежедневния документооборот или генерирани при обработката им, като данни от поръчки, фактури за покупки или продажби и други.

Предмет на нашето изследване са оперативните данни, които се генерират и обработват в големи обеми при някои обработки, и в определени ситуации придобиват характеристиките на Големи данни.

На фигура 1 е представен пример за приложение на структурни данни за едно предприятие. Структурата на складовото стопанство в системата се формира от складове, подскладове и локатори.



**Фигура 1: Организационна структура на предприятие**

Възможно е по изключение да се обработват и данни, които да се използват за допълване на информацията в номенклатурите.

Системите за съхранение и обработка на големи данни могат да поддържат структурирани, полуструктурирани и неструктурирани данни. За целите на бизнеса всички данни са важни, по тази причина правим анализ кой от данните могат да се обменят между двата вида системи така, че обработката им да има прагматична стойност.

Когато данните се предават от системата за големи данни към системата за дигитализация на бизнеса, трябва да има технологична възможност те да се обработват. За целта за всеки конкретен бизнес процес трябва да анализира и да се намери подходящ софтуер, който да допълва/завършва обработката.

Например, данните да се визуализират, да се структурират или да се анализират със средствата на изкуствения интелект.

## 1.2. Hadoop като система за обработка на големи данни

В съвременното общество генерирането и обработката на данни нараства с огромни темпове, което наложи създаването на системи за управление на големи обеми от данни, известни като Big Data. По тази причина бяха създадени системи за обработка на големи данни, които в повечето случаи са съчетани и с прилагане на методи от изкуствения интелект.

Big Data е организирано масивното количество данни, които не могат да се съхраняват, обработват и анализират, използвайки традиционни начини.

Основните елементи на Big Data са:

- Обем - голямо количество данни, генерирани всяка секунда.
- Скорост - Скоростта, с която се генерират, събират и анализират данните.
- Сортиране - различните видове данни могат да се подреждат по различни признаци, като структурирани, полуструктурирани, неструктурирани.
- Стойност - възможността да превърнете данните в полезна информация за бизнеса.
- Достоверност - верификация по отношение на качеството и точността.

Системата за обработка на големи данни Hadoop служи за съхранение и организиране с цел използването им в бизнес приложения.

Hadoop е рамка, която използва разпределено съхранение и паралелна обработка за съхранение и управление на големи данни. Това е софтуерът, който най-много се използва от анализаторите на данни за боравене с големи информационни масиви [1].

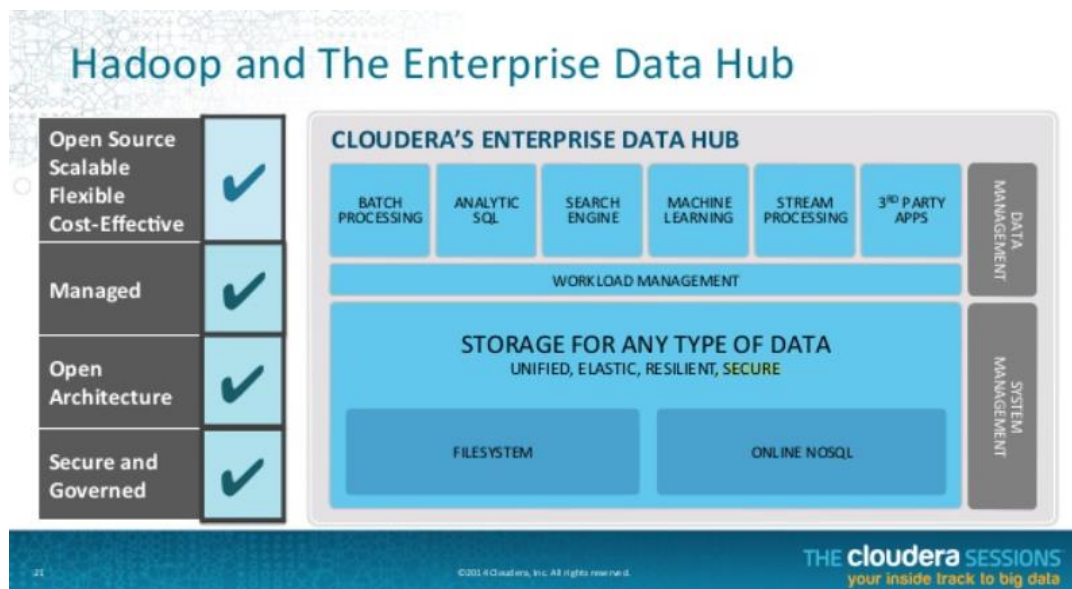
Hadoop е популярен инструмент за обработка на големи информационни масиви, използван от много компании по целия свят, като British Airways, Uber, The Bank of Scotland, Netflix, The National Security Agency (NSA), Expedia, Twitter и други.

Компоненти на Hadoop са [2]:

1. Hadoop разпределена файлова система (HDFS) е единицата за съхранение.
2. Hadoop MapReduce е процесорната единица.
3. Hadoop YARN (Yet Another Resource Negotiator) служи за управление на ресурси.

Същевременно Hadoop не може да се използва шаблонно, а изисква специфични знания за настройки и програмиране за постигане на желаните резултати. Тоест за всяко конкретно приложение на системата е необходимо да се направи анализ, проектиране и изграждане (програмиране) на конкретна ИКТ система.

В нашата разработка използваме версията Cloudera/Hadoop, чиято архитектура е представена на фигура 2.



*Фигура 2: Архитектура на Cloudera/Hadoop*

Избрахме да работим с Cloudera/Hadoop, защото е широко използвана технология Big Data за съхранение, обработка и анализ на големи масиви от данни, която е гъвкава за използване с други софтуерни приложения, включително с Oracle E-Business Suite.

### 1.3. Oracle ERP като система за обработка на бизнес данни

Софтуерната система Oracle ERP (E-Business Suite - EBS) е една от многото Oracle приложения.

Oracle E-Business Suite Architecture [3] е рамка за многостепенни, разпределени изчислителни технологии, която поддържа продуктите на Oracle E-Business Suite. В този модел различни сървъри или услуги се разпределят между три нива.

Сървър е процес или група от процеси, които работят на една машина и предоставя определена функционалност. Например веб услугите обработват HTTP заявки, а услугите за форми (Forms Services) обработват заявки за дейности, свързани с екранните форми на Oracle.

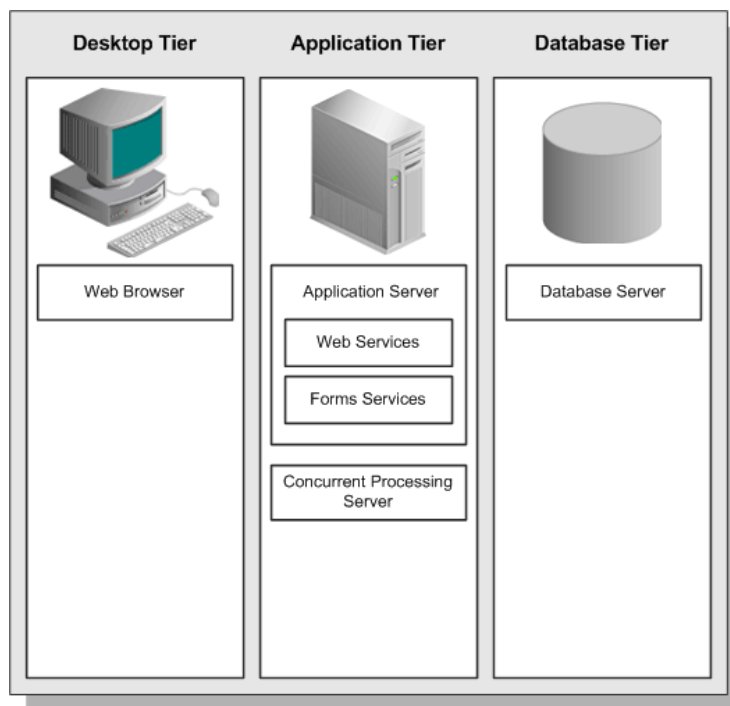
Слой (Tier) е логическо групиране на услугите, потенциално разпространено в повече от една физическа машина.

Трислойната архитектура на Oracle E-Business Suite е съставена от:

- Database Tier - слой на базата данни, което поддържа и управлява базата данни на Oracle;
- Application Tier - слой на приложението, което поддържа и управлява различните компоненти на Oracle E-Business Suite, и
- Desktop Tier - десктоп слой, който осигурява потребителския интерфейс чрез компонент на добавка към стандартен веб браузър.

Трислойната архитектура на Oracle E-Business Suite е представена на фигура 3.





*Фигура 3: Трислойна архитектура на Oracle E-Business Suite*

Машина може да бъде посочена като възел, особено в контекста на група компютри, които работят в тясно сътрудничество в клъстер. Всяко конфигурация може да се състои от един или повече възли и всеки възел може потенциално да побере повече от един слой.

Например базата данни може да се разположи на същия възел като един или повече компоненти на слоя на приложението.

Централизиране на софтуера Oracle E-Business Suite на слоевете на приложението елиминира необходимостта от инсталиране и поддръжка на приложен софтуер на всеки настолен клиентски компютър, а също така дава възможност на Oracle E-Business Suite да се мащабира добре с нарастващо натоварване.

Разширяване на тази концепция по-нататък и една от ключовите ползи от използването на този модел на споделена Tier файлова система е необходимостта да се поддържа само едно копие на съответния код на Oracle E-Business Suite, вместо копие за всяко приложение на tier машина.

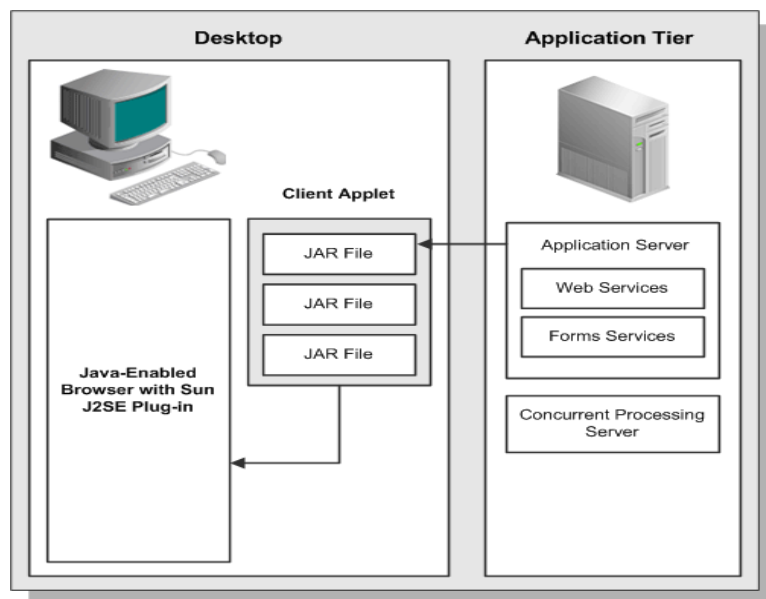
Слоеве на базата данни увеличават възможността за използването на Oracle Real Application Клъстери (Oracle RAC), където множество възли поддържат

един екземпляр на база данни, за да се даде по-голям капацитет и мащабируемост.

Връзката между слоя на приложението и десктоп слоя може да работи успешно през Wide Area Network (WAN). Това е така, защото слоя на работния плот и приложението обменят минимално количество информация, например само стойности на полета, които са се променили. Това е от голяма полза при глобални операции с потребители в различни географски точки, изискващи по-малко мрежов трафик и намалява телекомуникационните разходи, както и подобрява времето за реакция.

Десктоп слоя подsigурява работата на клиентският интерфейс, който се предоставя чрез HTML за HTML-базирани приложения, и чрез аplet Java в Уеб браузър за традиционните приложения, базирани на екранни форми.

На фигура 4 е представена базираната на екранни форми десктоп архитектура.



**Фигура 4: Базирана на форми десктоп архитектура**

За целите на нашия проект избираме за работим с Oracle E-Business Suite R12.

## **2. Проектиране и разработване на интерфейс за големи данни**

В рамките на проекта проектирахме и разработихме интерфейс за предаване на големи данни от Cloudera/Hadoop към Oracle E-Business Suite R12 и обратно.

С цел разширяване на функционалните възможности на системата за управление на ресурсите на предприятието Oracle EBS анализирахме технологичните характеристики [6] за обмен на данни между двете системи, проектирахме, разработихме и тествахме интерфейс, който може да предава данни в двете посоки.

Разработеният интерфейс позволява Големи данни, обработвани и съхранявани в Cloudera/Hadoop система да бъдат извлечени и използвани в Oracle EBS системата и оттам в цялостното функциониране на ERP системата и обратно.

Разработеният интерфейс може да се раздели на два вида. Интерфейс за обработка на структурирани данни и интерфейс за зареждане на неструктурирани данни от Hadoop системата.

За изграждането на интерфейса, обработващ структурирани данни, се използва междинна Oracle схема, специално създадена за целта.

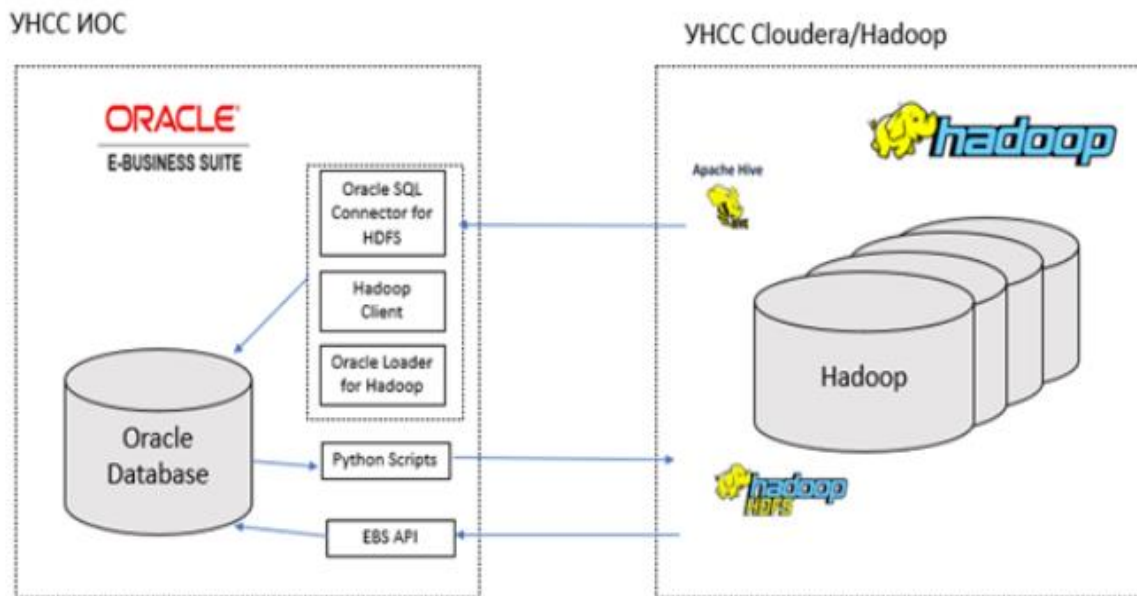
Инсталираната Oracle Big Data ни дава възможност да извличаме данни от Hive таблици, използвайки специално създадени таблици в Oracle. Работата с тези таблици е идентична със стандартните Oracle таблици, с тази разлика че данните се извличат в момента на обръщане към таблицата.

С помощта на Oracle SQL Developer, можем да търсим, филтрираме, сортираме и обработваме данните по начин, по който ни е необходим [6].

В Oracle EBS използваме стандартните интерфейсни таблици, които ни позволяват данни външни за системата да бъдат автоматично въведени и използвани.

Създаваме процес, в който свързваме таблица от междинната схема с интерфейсна таблица на Oracle EBS и последната стъпка е стандартен процес, който въвежда и валидира данните в таблиците на Oracle EBS системата.

На фигура 5 е показана принципна схема на разработката.



**Фигура 5: Интерфейс между Oracle EBS и Hadoop**

Изграденият интерфейс, обработващ структурирани данни е двустранен. Освен извличане на данни от Cloudera/Hadoop система е изградена и възможността да се извличат данни от Oracle EBS системата и да се записват в Hadoop.

Реализирано е с помощта на Oracle Big Data включващ Oracle Copy to Hadoop инструмента. Автоматизирането на процеса е реализирано с помощта на Python скриптове.

Поради ограничението на Oracle Big Data SQL за работа с неструктурирани данни [4], интерфейсът, който беше изграден, е полуавтоматичен. От Hadoop с помощта на Hadoop Get командата, копираме файловете от HDFS в локалната файлова система.

Стандартните интерфейсни таблици с помощта на които се импортват данни в Oracle EBS системата не дават възможност за работа с неструктурирани данни и

за това вместо тях се използва API - FND\_WEBATTCH.add\_attachment. С негова помощ се зареждат файлове в базата от данни използвайки Generic File Manager Access Utility (FNDGFU). В случая данните (файловете) се въвеждат в FND\_ATTACHED\_DOCUMENTS таблицата.

За да се използва тази функционалност са необходимо допълнителни настройки в Oracle EBS системата, които да позволят в конкретната форма да се използва функционалността, която позволява да се прикачат файлове [5]. Добавя се нова Категория Документи или се модифицира вече съществуваща, за да съхраняват файловете (в случая създадохме нова MEDICAL\_ASSESSMENTS).

Следваща стъпка е да се свържат тези категории с модула, който обработва данните (Document Entities) и създаване на Attachment Functions за данните, които искаме да се зареждат.

### 3. Тестване на разработения интерфейс

Разработеният интерфейс може да се използва за предаване на данни от всички широко използвани формати.

Проведените тестове с предаване на данни от оперативни документи, като поръчки и фактури, преминаха успешно.

Проведени бяха тестове с различни формати за образи и видео, които също преминаха успешно.

Тестовите показаха, че след като бъдат прехвърлени данните от едната система в другата, е необходимо те да се свържат с определен бизнес процес, за да може да се извлече прагматичната стойност на тези данни.

#### Източници

- 1) *What Is Hadoop? Components of Hadoop and How Does It Work* ([simplilearn.com](http://simplilearn.com))
- 2) *Hadoop Ecosystem and Their Components - A Complete Tutorial - DataFlair* ([data-flair.training](http://data-flair.training))
- 3) *Oracle E-Business Suite Concepts, 122oacg.pdf* ([oracle.com](http://oracle.com))
- 4) *Fusion Middleware Integrating Big Data with Oracle Data Integrator*,  
<https://docs.oracle.com/middleware/12212/odi/odi-big-data/intro.htm#ODIBD109>
- 5) *Oracle Big Data: Interactive Quick Reference*,  
[https://www.oracle.com/webfolder/technetwork/tutorials/tutorial/bda/BigDataPoster/OUT\\_PUT\\_poster/bigdata.html](https://www.oracle.com/webfolder/technetwork/tutorials/tutorial/bda/BigDataPoster/OUT_PUT_poster/bigdata.html)
- 6) *Oracle Workflow Developer's Guide, 122wfdg.pdf* ([oracle.com](http://oracle.com))