



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ
(национальный исследовательский университет)»

Выпускная квалификационная работа по теме:
«Информационная система с Web-интерфейсом для курса дистанционного обучения
по дисциплине "Хранение и обработка больших данных"»

Научный руководитель:

Д.ф.-м.н., проф. каф. 810Б, Ревизников Дмитрий Леонидович

Дипломник: группа 8О-203М, Журба Н.Д.

Москва

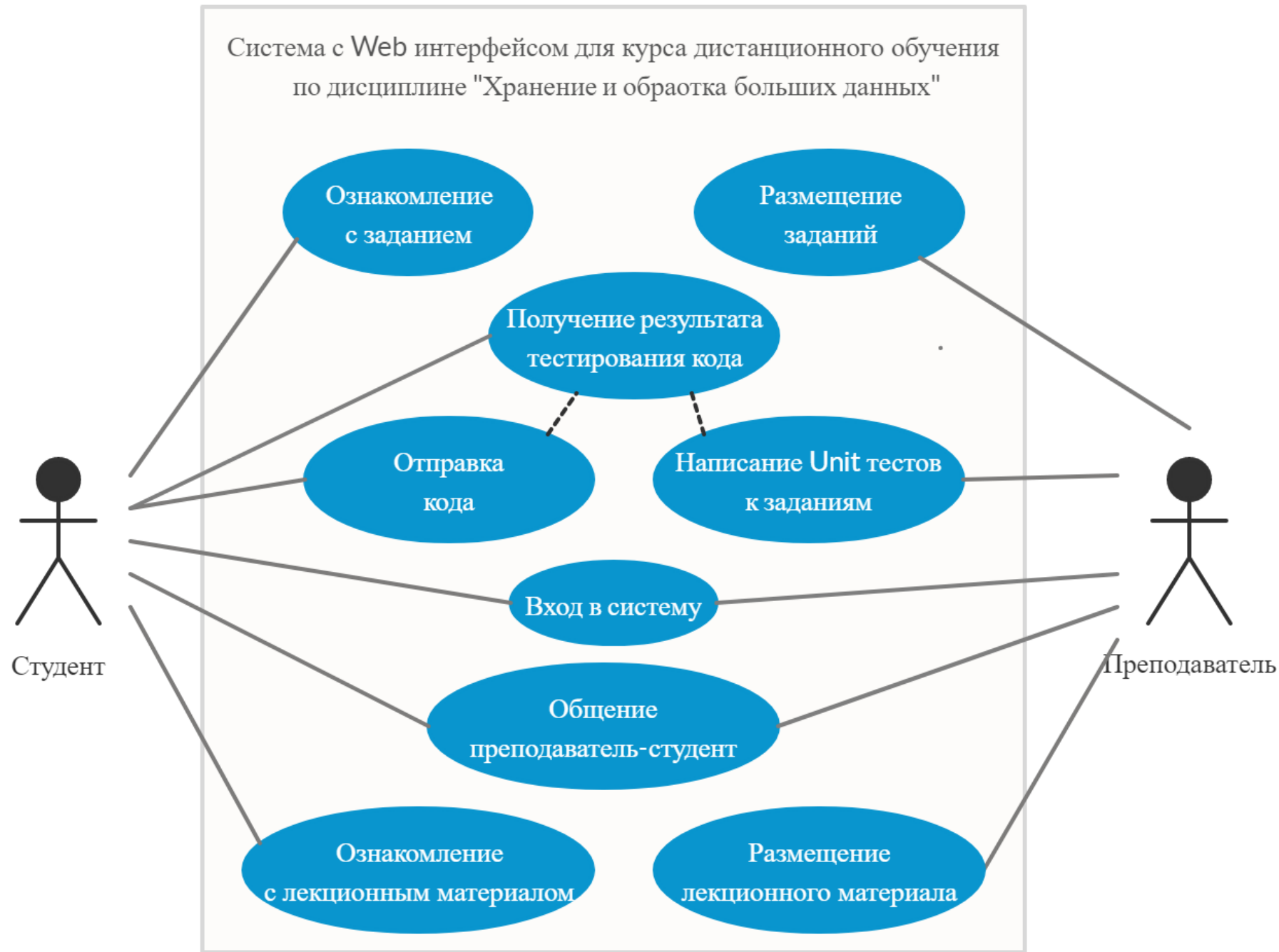
2020 г.

Цель и задачи работы.

Цель - реализовать Информационную систему с Web-интерфейсом для курса дистанционного обучения по дисциплине "Хранение и обработка больших данных"

Задачи:

1. Провести анализ предметной области дисциплины «Хранение и обработка больших данных» и применимость к ней существующих доступных систем онлайн-обучения (LMS)
2. Разработать архитектуру информационной системы с автоматизированной проверкой практических заданий дисциплины
3. Реализовать прототип информационной системы с автоматизированной проверкой задач по анализу больших данных с помощью Apache Spark



Логика работы системы проверки задания

4



Анализ существующих решений

5



Основной функционал существующих решений:

- Доставка обучающего контента
- Коммуникация с преподавателем
- Учет учебного прогресса

Основной недостаток, относительно предметной области данной работы:

- Невозможность интерактивной автоматической проверки кода с целью получения практических навыков

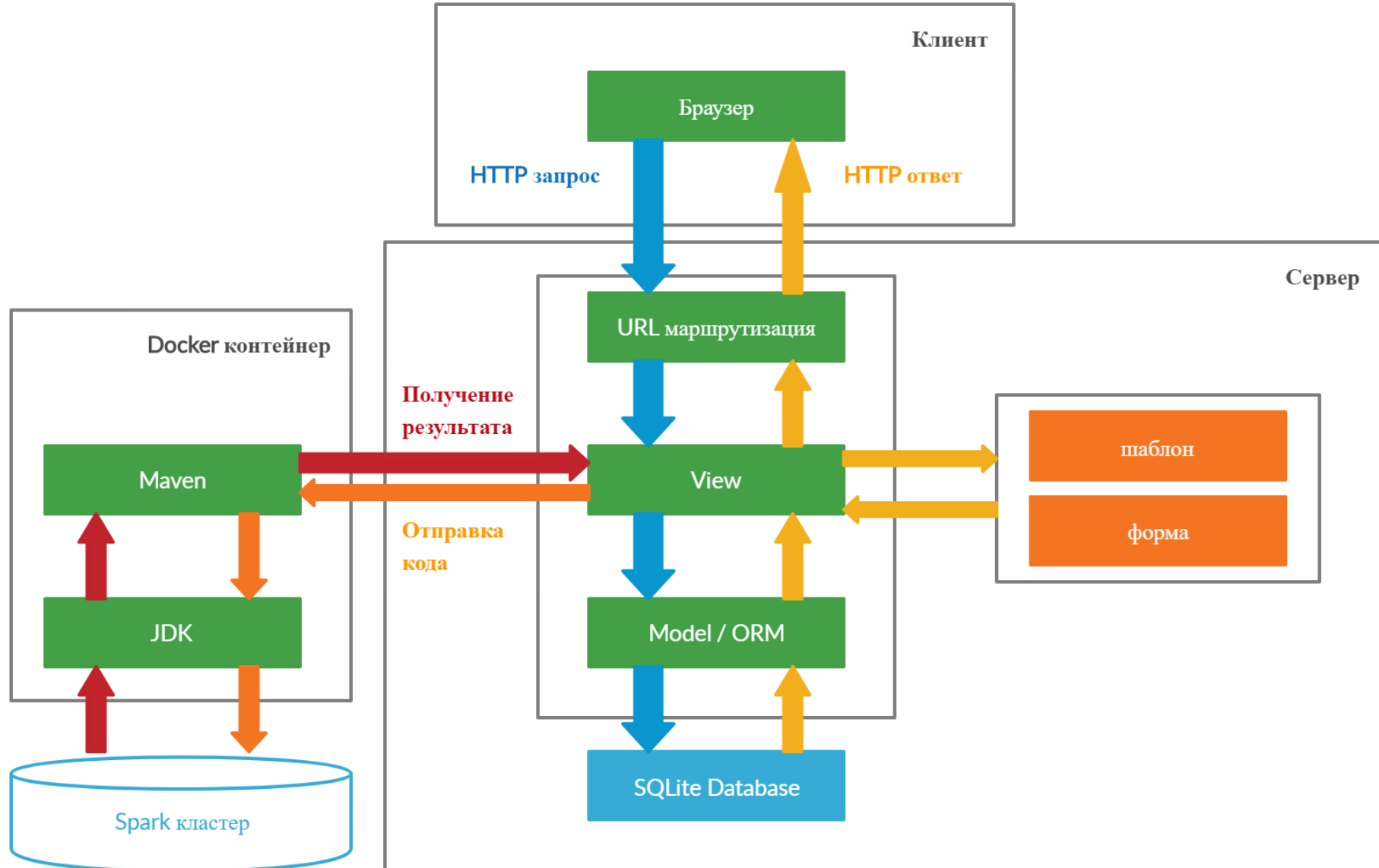
Технологии используемые для реализации поставленных задач.

6

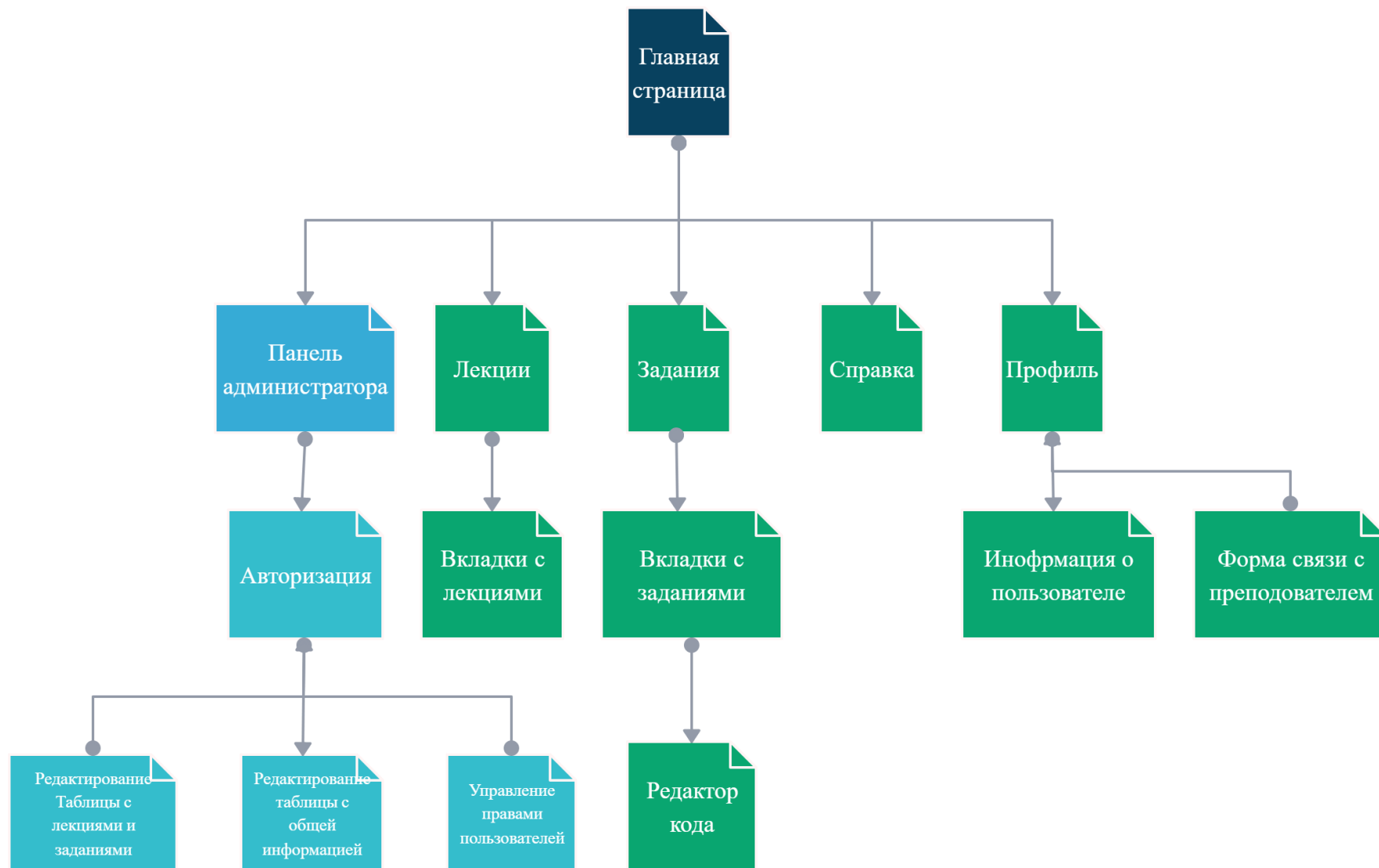
- HTML;
- CSS;
- Python;
- Фреймворк Django;
- СУБД SQLite;
- Jinja;
- Bootstrap.
- Maven
- Apache Spark
- Docker



Архитектура системы



Карта сайта





Scala

Scala — мультипарадигмальный язык программирования, спроектированный кратким и типобезопасным для простого и быстрого создания компонентного программного обеспечения, сочетающий возможности функционального и объектно-ориентированного программирования.

[Узнать больше](#)

Контакты преподавателя

Тел: +7 999 999 99 99

Email: prepod@gmail.com

Telegram: [@prepod](https://t.me/prepod)

Полезные ссылки

[Ссылка 1](#)

[Ссылка 2](#)

[Ссылка 3](#)

О сайте

[Информация о сайте](#)

Напишите класс, реализующий подсчёт слов в тексте:

Класс принимает как аргумент "file"

```

1 private static void wordcount(String fileName) {
2     SparkConf sparkConf = new SparkConf().setMaster("local"
3     ).setAppName("3D word counter");
4     JavaSparkContext sparkContext = new JavaSparkContext
5     (sparkConf);
6     JavaRDD<String> inputFile = sparkContext.textFile(fileName
7     );
8     JavaRDD<String> wordsFromFile = inputFile.flatMap(content
9     -> Arrays.asList(content.split(" ")));
10
11     JavaPairRDD countData = wordsFromFile.mapToPair(t -> new
12     Tuple2(t, 1)).reduceByKey((x, y) -> (int) x + (int) y);
13     countData.saveAsTextFile("countData");
14 }

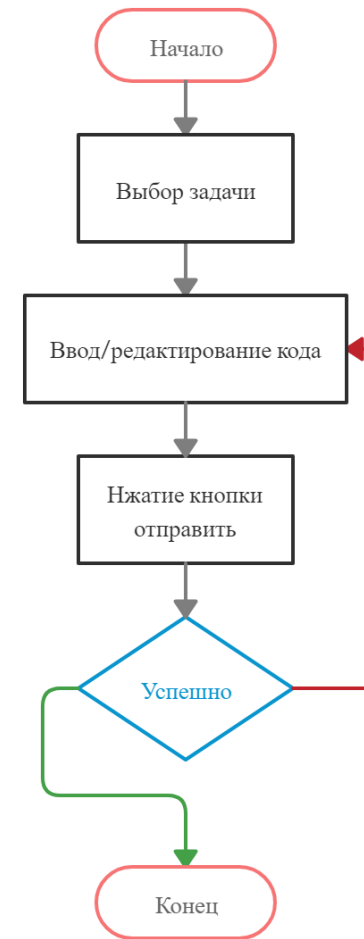
```

Отправить

```

[INFO] Scanning for projects...
[INFO]
[INFO] -----< yshensh:spark-word-count >-----
[INFO] Building spark-word-count 0.1.0-SNAPSHOT
[INFO] -----[ jar ]-----
[INFO]
[INFO] --- maven-resources-plugin:2.6:resources (default-resources) @ spark-word-count ---
[INFO] skip non existing resourceDirectory C:\Users\SKYNET\Desktop\spark-word-count-master\src\main\resources
[INFO]
[INFO] --- maven-compiler-plugin:3.1:compile (default-compile) @ spark-word-count ---
[INFO] Nothing to compile - all classes are up to date
[INFO]
[INFO] --- scala-maven-plugin:3.2.1:compile (default) @ spark-word-count ---
[INFO] C:\Users\SKYNET\Desktop\spark-word-count-master\src\main\scala:-1: info: compiling
[INFO] Compiling 1 source files to C:\Users\SKYNET\Desktop\spark-word-count-master\target\classes at 1592056725031
[INFO] prepare-compile in 0 s
[INFO] compile in 3 s
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 7.612 s
[INFO] Finished at: 2020-06-13T16:58:48+03:00
[INFO] -----
Результат:
(hello,1)
(test,1)
(world,1)

```



Контакты преподавателя

Тел: +7 999 999 99 99
 Email: prepod@gmail.com
 Telegram: @prepod

Полезные ссылки

Ссылка 1
 Ссылка 2
 Ссылка 3

О сайте



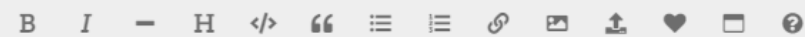
Информация о сайте

Изменить Лекция

11
История

Тема

Лекция 1

 Editor Preview

```
1 https://www.youtube.com/watch?v=_e0SBbpFPpA
```

```
2
```

```
3 Apache Spark (от англ. spark – искра, вспышка) – фреймворк с открытым исходным кодом для реализации распределённой обработки неструктурированных и слабоструктурированных данных, входящий в экосистему проектов Hadoop. В отличие от классического обработчика из ядра Hadoop, реализующего двухуровневую концепцию MapReduce с хранением промежуточных данных на накопителях, Spark работает в парадигме резидентных вычислений (англ. in-memory computing) – обрабатывает данные в оперативной памяти, благодаря чему позволяет получать значительный выигрыш в скорости работы для некоторых классов задач, в частности, возможность многократного доступа к загруженным в память пользовательским данным делает библиотеку привлекательной для алгоритмов машинного обучения.
```

Файл

Выберите файл | Файл не выбран

Удалить

Сохранить и продолжить редактирование

Сохранить и добавить другой объект

Сохранить

Изменить Лекция

12ория

Тема

Лекция 1

✎ Editor

👁 Preview



Apache Spark (от англ. spark — искра, вспышка) — фреймворк с открытым исходным кодом для реализации распределённой обработки неструктурированных и слабоструктурированных данных, входящий в экосистему проектов Hadoop. В отличие от классического обработчика из ядра Hadoop, реализующего двухуровневую концепцию MapReduce с хранением промежуточных данных на накопителях, Spark работает в парадигме резидентных вычислений (англ. in-memory computing) — обрабатывает данные в оперативной памяти, благодаря чему позволяет получать значительный выигрыш в скорости работы для некоторых классов задач, в частности, возможность многократного доступа к загруженным в память пользовательским данным делает библиотеку привлекательной для алгоритмов машинного обучения.

Файл

Выберите файл

Файл не выбран

Активация Windows

Чтобы активировать Windows, перейдите в раздел
"Параметры".

Удалить

Сохранить и продолжить редактирование

Сохранить и добавить другой объект

Сохранить

Профиль



Фамилия Имя
Отчество

Студент

Группа: номер

Информация

[Задать вопрос Преподавателю](#)

История профиля

Зарегистрирован:	12-06-2020
Последняя активность:	12-06-2020 / 09:11
Группа:	номер группы
Направление:	направление
Сданных работ:	●●●● 4/5
email:	email@email.com

Контакты преподавателя

Тел: +7 999 999 99 99

Email: prepod@gmail.com

Telegram: [@prepod](#)

Полезные ссылки

[Ссылка 1](#)


[Ссылка 2](#)

[Ссылка 3](#)

О сайте

[Информация о сайте](#)

Профиль



Фамилия Имя
Отчество

Студент
Группа: номер

Информация [Задать вопрос Преподавателю](#)

Текст Вашего сообщения

Текст сообщения будет отправлен преподавателю

[Отправить](#)

Контакты преподавателя

Тел: +7 999 999 99 99

Email: prepod@gmail.com

Telegram: [@prepod](#)

Полезные ссылки

[Ссылка 1](#)

[Ссылка 2](#)

[Ссылка 3](#)

О сайте

[Информация о сайте](#)

Выводы

В ходе выполнения работы были выполнены следующие задачи:

- Проведен анализ предметной области дисциплины «Хранение и обработка больших данных» и применимость к ней существующих доступных систем онлайн-обучения (LMS)
- Разработана архитектура информационной системы с автоматизированной проверкой практических заданий дисциплины
- Реализован прототип информационной системы с автоматизированной проверкой задач по анализу больших данных с помощью Apache Spark