МИНОБРНАУКИ РОССИИ

Федеральное государственное автономное образовательное

учреждение высшего образования «Южный федеральный университет»

(ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ)

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**Технологии Big Data**

Ростов-на-Дону – Таганрог 2020

Содержание

[I. Цели и задачи освоения дисциплины 4](#_Toc44274470)

[II. Место дисциплины в структуре образовательной программы 4](#_Toc44274471)

[III. Требования к результатам освоения дисциплины 5](#_Toc44274472)

[IV. Содержание и структура дисциплины 7](#_Toc44274473)

[4.1. Содержание дисциплины, структурированное по темам 7](#_Toc44274474)

[4.2. План внеаудиторной самостоятельной работы 8](#_Toc44274475)

[4.3. Содержание учебного материала 10](#_Toc44274476)

[V. Образовательные технологии 11](#_Toc44274477)

[VI. Учебно-методическое обеспечение дисциплины 12](#_Toc44274478)

[6.1. Основная литература 12](#_Toc44274479)

[6.2. Дополнительная литература 12](#_Toc44274480)

[6.3. Периодические издания 12](#_Toc44274481)

[6.4. Перечень ресурсов сети Интернет 12](#_Toc44274482)

[VII. Материально-техническое обеспечение дисциплины 13](#_Toc44274483)

[VIII. Методические указания для обучающихся по освоению дисциплины 13](#_Toc44274484)

[IX. Учебная карта дисциплины 15](#_Toc44274485)

[X. Фонд оценочных средств 16](#_Toc44274486)

[10.1. Паспорт фонда оценочных средств 16](#_Toc44274487)

[10.2. Контрольная работа № 1 (тестирование) 16](#_Toc44274488)

[10.3. Практические задания №№ 1-8 (выполнение, защита результатов) 19](#_Toc44274489)

[10.4. Лабораторные работы №№ 1–10 (выполнение, подготовка отчёта, защита отчёта) 20](#_Toc44274490)

[10.5. Экзаменационные вопросы и билеты 23](#_Toc44274491)

# Цели и задачи освоения дисциплины

Цели освоения дисциплины:

* формирование у студентов профессиональные компетенции в области разработки и использования систем обработки и анализа больших массивов данных. Данная цель соотносится с целью образовательной программы в части изучения технологий разработки специализированных программных систем, отвечающих за обработку больших данных.

Задачи освоения дисциплины:

* изучение специфики анализа и обработки больших массивов данных;
* овладение основами разработки моделей, методов, алгоритмов и программно-аппаратных платформ для реализации проектов Big Data;
* освоение базовых решений в области построения архитектур больших данных;
* приобретение навыков применения существующих решений в области обработки больших данных.

# Место дисциплины в структуре образовательной программы

Дисциплина относится к модулю общеинститутских дисциплин образовательной программы.

Данная дисциплина опирается на базовые знания, умения и навыки, формируемые при получении предшествующего уровня образования.

Знания, умения и навыки, формируемые данной дисциплиной, потребуются при освоении следующих дисциплин и практик образовательной программы:

* «Исследовательский проект»;
* для прохождения производственной и преддипломной практик, подготовки и защиты выпускной квалификационной работы.

# Требования к результатам освоения дисциплины

Освоение дисциплины направлено на формирование следующих компетенций в соответствии с образовательным стандартом и образовательной программой:

Перечень планируемых результатов обучения по дисциплине, соотнесённых с индикаторами достижения компетенций

| Компетенция | Индикаторы достижения компетенции | Результаты обучения |
| --- | --- | --- |
| ОПК-6. Способен разрабатывать компоненты программно-аппаратных комплексов обработки информации и автоматизированного проектирования и адаптировать подобные зарубежные комплексы к нуждам отечественных предприятий | ОПК-6.1. Анализирует техническое задание, разрабатывает компоненты программно-аппаратных комплексов для решения задач обработки информации и автоматизированного проектирования | Знания:   * *основных тенденций развития интеллектуальных технологий в области обработки больших данных;* * *методологических основ интеллектуального анализа больших данных.* |
| Умения:   * *применять методы интеллектуального анализа больших данных для решения профессиональных задач;* * *осуществлять разработку оригинальных алгоритмов и программных средств Big Data в условиях информационной неопределенности.* |
| Навыки:   * *использования методов интеллектуального анализа больших данных при разработке алгоритмов и программных средств Big Data для решения профессиональных задач в условиях информационной неопределенности.* |
| ОПК-6.2. Анализирует функциональные требования к программно-аппаратному обеспечению с учетом национальных стандартов обработки информации и автоматизированного проектирования | Знания:   * *национальных стандартов обработки информации и автоматизированного проектирования;* * *основных функциональных требований* *к программно-аппаратному обеспечению с учетом национальных стандартов обработки информации и автоматизированного проектирования.* |
| Умения:   * *применять на практике сведения из национальных стандартов обработки информации и автоматизированного проектирования;* * *осуществлять разработку оригинальных алгоритмов и программных средств на основе применения функциональных требований к программно-аппаратному обеспечению с учетом национальных стандартов.* |
| Навыки:   * *использования функциональных требований к программно-аппаратному обеспечению с учетом национальных стандартов.* |
| ОПК-6.3. Конфигурирует программно-аппаратные комплексы обработки информации и автоматизированного проектирования | Знания:   * *методов разработки программно-аппаратных комплексов в технологиях Big Data;* * *эффективных решений мировых лидеров IT- индустрии в области создания конфигурируемых программно-аппаратных комплексов для решения задач анализа и обработки больших данных.* |
| Умения:   * *применять на практике различные варианты типовых конфигураций программно-аппаратных комплексов;* * *осуществлять разработку конфигураций программно-аппаратных комплексов средств обработки больших данных.* |
| Навыки:   * *использования конфигурируемых программно-аппаратных комплексов для решения задач анализа и обработки больших данных.* |
| ОПК-6.4. Адаптирует зарубежные комплексы обработки информации и автоматизированного проектирования в соответствии с национальными стандартами и нуждами отечественных предприятий | Знания:   * *зарубежных комплексов обработки информации;* * *нужд отечественных предприятий в средствах* *обработки информации.* |
| Умения:   * *адаптировать зарубежные комплексы обработки информации* *в соответствии с национальными стандартами и нуждами отечественных предприятий.* |
| Навыки:   * *использования адаптированных зарубежных комплексов обработки информации в соответствии с национальными стандартами и нуждами отечественных предприятий* |

# Содержание и структура дисциплины

Трудоёмкость дисциплины составляет 6 зачётных единиц, 216 часов,

в том числе 1 зачётная единица, 36 часов на экзамен

Форма промежуточной аттестации: экзамен

## Содержание дисциплины, структурированное по темам

| № п/п | Темы дисциплины | Семестр | Виды учебной работы и их трудоёмкость, часы  (в том числе с использованием онлайн-курсов) | | | | Наименования оценочных средств |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Контактная работа | | | Самостоя­тельная работа |
| Лекции | Практические занятия | Лабораторные занятия |
| **Модуль 1. Технологии Big Data** | | | | | | | |
| 1 | Специфика анализа и обработки больших массивов данных с учетом национальных информационных стандартов | 1 | 2 | 2 | 2 | 14 | * практические занятия №№ 1,2 (выполнение, защита результатов) |
| 2 | Основы разработки моделей, методов, алгоритмов и программно-аппаратных платформ для реализации проектов Big Data | 1 | 2 | 2 | 2 | 14 | * лабораторная работа № 1 (выполнение, подготовка отчёта, защита отчёта); * практическое занятие № 3 (выполнение, защита результатов) |
| 3 | Модель программирования Map Reduce и программная платформа Hadoop | 1 | 4 | 4 | 4 | 28 | * практические занятия №№ 4,5 (выполнение, защита результатов) |
| 4 | Базовые решения в области построения архитектур больших данных | 1 | 2 | 2 | 2 | 14 | * лабораторная работа № 2 (выполнение, подготовка отчёта, защита отчёта); * практическое занятие № 6 (выполнение, защита результатов) |
| 5 | Состав и предназначение основных утилит программной платформы в технологиях Big Data | 1 | 4 | 4 | 4 | 28 | * практические занятия №№ 7,8 (выполнение, защита результатов) |
| 6 | Эффективные решения мировых лидеров IT- индустрии в области Big Data | 1 | 4 | 4 | 4 | 28 | * лабораторная работа № 3 (выполнение, подготовка отчёта, защита отчёта); * контрольная работа № 1 (тестирование) |
| Промежуточная аттестация | | 1 | – | – | – | 36 | Экзаменационные вопросы и билеты |
| **Итого часов** | | **1** | **18** | **18** | **18** | **162** | **–** |

## План внеаудиторной самостоятельной работы

| № п/п | Темы дисциплины | Семестр | Вид самостоятельной работы | Сроки выполнения (нед.) | Затраты времени (часы) | Учебно-методическое обеспечение |
| --- | --- | --- | --- | --- | --- | --- |
| **Модуль 1. Технологии Big Data** | | | | | | |
| 1 | Специфика анализа и обработки больших массивов данных с учетом национальных информационных стандартов | 1 | * проработка и повторение материала лекционных занятий; * подготовка к практическим занятиям | 1–2 | 14 | [1], [3], [5] |
| 2 | Основы разработки моделей, методов, алгоритмов и программно-аппаратных платформ для реализации проектов Big Data | 1 | * проработка и повторение материала лекционных занятий; * подготовка к лабораторным работам, подготовка отчётов о выполнении лабораторных работ, подготовка к защите отчётов о выполнении лабораторных работ; * подготовка к практическому занятию | 3–4 | 14 | [2], [5]–[6] |
| 3 | Модель программирования Map Reduce и программная платформа Hadoop | 1 | * проработка и повторение материала лекционных занятий; * подготовка к практическим занятиям | 5–8 | 28 | [4], [6] |
| 4 | Базовые решения в области построения архитектур больших данных | 1 | * проработка и повторение материала лекционных занятий; * подготовка к лабораторным работам, подготовка отчётов о выполнении лабораторных работ, подготовка к защите отчётов о выполнении лабораторных работ; * подготовка к практическому занятию | 9–10 | 14 | [2], [6] |
| 5 | Состав и предназначение основных утилит программной платформы в технологиях Big Data | 1 | * проработка и повторение материала лекционных занятий; * подготовка к практическим занятиям | 11–14 | 28 | [4]–[6] |
| 6 | Эффективные решения мировых лидеров IT- индустрии в области Big Data | 1 | * подготовка к лабораторным работам, подготовка отчётов о выполнении лабораторных работ, подготовка к защите отчётов о выполнении лабораторных работ; * проработка и повторение материала лекционных занятий; * подготовка к контрольной работе | 15–18 | 28 | [4]–[6] |
| **Подготовка к экзамену** | | | | | **36** | **[1]–[6]** |
| **Общая трудоёмкость самостоятельной работы по дисциплине** | | | | | **162** | **–** |

## Содержание учебного материала

**Модуль 1. Технологии Big Data**

**Специфика анализа и обработки больших массивов данных с учетом национальных информационных стандартов**: Альтернативная ценность данных. Датификация: генезис, характеристики, технологии. Дата-форсайты, прогнозное моделирование. Информация и коммуникация в современной инфосфере. Датификация информации и коммуникации: новейшие модели. Интернет и формирование новой модели профессиональной коммуникации.

**Основы разработки моделей, методов, алгоритмов и программно-аппаратных платформ для реализации проектов Big Data**: Методология исследования и использование интернета в профессиональной социальной коммуникации и Big Data. Предпосылки формирования больших массивов данных. Специфика современных «больших данных». Big Data и кибербезопасность. Конфиденциальность. Деанонимизация данных как проблема.

**Модель программирования Map Reduce и программная платформа Hadoop**: Big Data: к вопросу определения понятия. Характеристики, особенности Big Data. Data Science. Инструменты, алгоритмы, оборудование для реализации проектов Big Data. Технологический цикл применения Big Data. Категоризация массивов данных.

**Базовые решения в области построения архитектур больших данных**: Специфика Data Mining. Основы языка R (надстройка Rhadoop). Map Reduce, Hadoop. Особенности анализа Big Data. Концепции и инструментарий. Возможности использования больших данных. Трансляция данных. Переход к решениям на основе больших данных. Явные и скрытые корреляции. Особенности первичного, вторичного использования больших данных.

**Состав и предназначение основных утилит программной платформы в технологиях Big Data**: Файловая система HDFS. Утилиты HADOOP. База данных HBase. Модель данных HBase. Архитектура HBase. Основные процессы HBase. СУБД Hive. Партицирование и бакетирование в Hive. СУБД Impala. Процедурный язык Apache Pig. Библиотека Mahout. Инструмент кластерных вычислений Spark. Утилита Tez. Планировщик Oozie. Web Интерфейс Hue. Утилита загрузки Sqoop.

**Эффективные решения мировых лидеров IT- индустрии в области Big Data**: Модуль Yarn. Oracle Big Data Appliance. Oracle Exadata. Oracle Exalytics. SAP HANA Data Platform for Big Data. Технологии Fujitsu для BigData.

Перечень тем практических занятий

| №  п/п | Тема практического занятия | Количество часов |
| --- | --- | --- |
| **Модуль 1. Технологии Big Data** | | |
| 1 | **Высоконагруженные приложения.** Шардирование.  Реплицирование. Рост нагрузки и падение серверов. Серверный стек технологий | 2 |
| 2 | **Java в BigData.** Collection Framework. JVM internals. Работа с файлами. Как работает загрузка классов. Пулы соединений/потоков | 2 |
| 3 | **HDFS. «**Демоны» HDFS. Консольные команды. Java API. | 2 |
| 4 | **NoSQL.** Key-value хранилища. Документоориентированные БД. Таблицы в оперативной памяти. Cassandra. CQL. Hbase. | 2 |
| 5 | **MapReduce.** Основная идея. Mapper, Reducer и Combiner. Запуск MR приложения. | 4 |
| 6 | **Hive.** SQL синтаксис для выполнения MR-заданий. Создание таблиц и их заполнение. Техники оптимизации. | 2 |
| 7 | **Основы Scala.** Особенности Scala. Сборка и запуск проекта. Подключение зависимостей в sbt. | 2 |
| 8 | **Spark.** Быстрые вычисления в памяти. Dataset API. Азы машинного обучения. Использование MLlib. | 2 |
| **Всего часов** | | **18** |

Перечень лабораторных работ

| №  п/п | Название лабораторной работы | Количество часов |
| --- | --- | --- |
| **Модуль 1. Технологии Big Data** | | |
| 1 | Лабораторная работа № 1 **Технологический цикл реализации дата-проекта** | 6 |
| 2 | Лабораторная работа № 2 **Практикум-апробация дата-проекта** | 6 |
| 3 | Лабораторная работа № 3 **Презентация дата-проекта** | 6 |
| **Всего часов** | | **18** |

# Образовательные технологии

Эффективность обучения повышает использование современных технических средств, технологий и методик изучения предмета. В организации процесса обучения по дисциплине используются как традиционные, характерные лекционно-семинарской форме обучения, так и инновационные (интерактивные, имитационные, проектные) технологии.

Для достижения поставленных целей преподавания дисциплины реализуются следующие методы и формы активизации познавательной и научной деятельности:

1) изучение теоретического материала дисциплины с использованием компьютерных технологий;

2) самостоятельное изучение теоретического материала дисциплины с использованием Интернет технологий, информационных баз, методических разработок, специальной учебной и научной литературы;

3) дискуссия как способ закрепления теоретического материала и формирования четко осознаваемой собственной точки зрения;

4) IT-методы как способ обучения эффективному оперированию информацией и ее обработки;

5) опережающая СР как форма углубленного изучения и закрепления знаний, а также развития практических умений, заключающейся в работе студентов с лекционным материалом, поиск и анализ литературы и электронных источников информации по заданной проблеме, выполнении домашних заданий, изучении тем, вынесенных на самостоятельную проработку, подготовка к рубежному контролю по модулю;

6) индивидуальный подход к обучению;

7) проблемное обучение как способ развития самостоятельности в решении возникающих в процессе обучения и профессиональной деятельности задач;

8) кейс-метод можно представить как сложную систему, в которую интегрированы различные методы познания. В него входят моделирование, системный анализ, проблемный метод, мысленный эксперимент, методы описания, классификации, которые выполняют в кейс-методе свои функции.

9) интерактивные образовательные технологии, используемые в аудиторных занятиях.

Лабораторный практикум по курсу включает элементы интерактивного проблемно-ориентированного подхода к обучению за счет фокусирования внимания студентов на анализе и разрешении конкретных задач, когда важно не только решить проблему проектирования интегрированных информационных систем, но корректно ее поставить и сформулировать.

В результате выполнения заданий в приведенной серии лабораторных занятий студенты будут обладать знаниями и умениями создания и реализации дата-проектов.

Прикладные задачи выбираются студентами или выдаются преподавателем. Поощряется решение творческих, нетривиальных задач. Предлагается создание групп по 2-3 человека. Каждая группа в конце курса представит проект, реализующий решение прикладной задачи.

Наряду с традиционными образовательными технологиями, для реализации дисциплины могут использоваться технологии электронного обучения и дистанционные образовательные технологий в электронной информационно-образовательной среде Южного федерального университета. Лекционные занятия и другие формы контактной работы обучающихся с преподавателем могут проводиться с использованием платформ Microsoft Teams, Cisco, Moodle (BigBlueButton) и др., что позволяет обеспечить онлайн и офлайн взаимодействие преподавателя с обучающимися в рамках дисциплины.

Основными методами текущего контроля являются электронный учёт и контроль учебных достижений студентов (использование средств сервиса балльно-рейтинговой системы; ведение электронного журнала успеваемости, проведение электронного тестирования и применение других средств контроля с использованием системы электронного обучения).

# Учебно-методическое обеспечение дисциплины

## Основная литература

1. Гергель В. П. Технологии построения и использования кластерных систем: курс – http://biblioclub.ru/index.php?page=book&id=233768.

2. Короткова О. А. Характеристика программных средств АРМ и перспективы их развития – http://biblioclub.ru/index.php?page=book&id= 142937.

3. Галатенко В. А. Мобильное программирование приложений реального времени в стандарте POSIX – http://biblioclub.ru/index.php?page= book&id=233766.

## Дополнительная литература

4. Демидова Л. А., Кираковский В. В., Пылькин А. Н. Принятие решений в условиях неопределенности – http://biblioclub.ru/index.php?page= book&id=253180.

5. Подколзин А. С. Компьютерное моделирование логических процессов. Архитектура и языки решателя задач – http://biblioclub.ru/ index.php?page=book&id=68419.

6. Володин К. И., Гульницкий Л. Л., Пожариский И. Ф., Чуркин В. П., Янбухтин Т. К. Автоматизированная система научно-технической информации - разработка и эксплуатация – http://biblioclub.ru/ index.php?page=book&id=220229.

## Периодические издания

* Научный журнал «Вестник компьютерных и информационных технологий» (<http://www.vkit.ru/>);
* Научный журнал «Информационные системы и технологии» (<http://oreluniver.ru/science/journal/isit>);
* Научный журнал «Информационные технологии» (<http://novtex.ru/IT/>);
* Научный журнал «Известия ЮФУ. Технические науки» (<http://izv-tn.tti.sfedu.ru/index.php/izv_tn>).

## Перечень ресурсов сети Интернет

* Портал «Гуманитарное образование» http://www.humanities.edu.ru/
* Федеральный портал «Российское образование» http://www.edu.ru/
* Федеральное хранилище «Единая коллекция цифровых образовательных ресурсов» http://school-collection.edu.ru/

# Материально-техническое обеспечение дисциплины

При реализации дисциплины используются следующие помещения, оборудование и программное обеспечение:

* Лекционные занятия проводятся в аудиториях, оснащенных интерактивной доской типа SmartBoard (или иным аналогичным по функциональному назначению оборудованием с возможностью подключения ноутбука и удаленным Wi-Fi доступом);
* Практические и лабораторные занятия проводятся в аудиториях, оснащенных персональными компьютерами с предустановленным лицензионным программным обеспечением Microsoft Windows, Microsoft Office PowerPoint, которое позволяет работать с видео-аудио материалами, создавать и демонстрировать презентации, с выходом в сеть Интернет;
* Интегрированная среда разработки Eclipse (GPL - <http://www.eclipse.org/legal/eplfaq.php>);
* OC Linux CentOS7 (GPL - <http://www.gnu.org/licenses/gpl.html>);
* Авторские презентации.

# Методические указания для обучающихся по освоению дисциплины

В ходе организации обучения по учебной дисциплине «Технологии BIGDATA» применяются метод проблемного обучения, применяются метод анализа первоисточников – неадаптированных философских текстов, методы логических заданий, обучающей игры, дискуссии, метод тестирования.

Исходя из значительного объема учебного материала, в преподавании курса «Технологии BIGDATA» широко применяется проблемный метод чтения лекций. Лекционный курс содержит преимущественно теоретический материал, отражающий современное состояние научных концепций по данной тематике и подкрепленный разъяснениями и комментариями на конкретных примерах. В процессе лекционного занятия студенты слушают преподавателя, задают вопросы, часть информации конспектируют. При этом активно используются компьютерная, проекционная техника и презентации, концентрирующих внимание слушателей на ключевых моментах лекционного материала и ориентирующие на последовательное изложение материала при разборе конкретных ситуаций проблемного характера.

Проведение лекционных, практических и лабораторных занятий осуществляется с постановкой проблемных вопросов, допускающих возникновение дискуссий, что предполагает активное включение студентов в образовательный процесс.

В организации процесса обучения по дисциплине «Технологии BIGDATA» используются как традиционные, характерные лекционно-семинарской форме обучения, так и инновационные (интерактивные, имитационные, проектные) технологии. Используемые технологии обеспечивают:

– формирование компетенций, осознанное усвоение знаний, качественное освоение умений их применять и формирование заинтересованного отношения к изучаемым объектам в единстве;

– продуктивность познавательной деятельности, научный поиск, создание субъективно и объективно новых знаний или других продуктов;

– ориентацию на студентов, стимулирование их активности, самостоятельности, инициативы и ответственности;

– контекстный характер обучения, то есть привязку к реальным профессиональным задачам;

– вовлеченность студентов в выполняемую деятельность, возможность проявить и развить свой интеллектуальный, творческий, личностный, деловой потенциал.

Лекционная часть курса включает следующие компоненты системы знаний учебной дисциплины: понятийный аппарат (тезаурус курса), теоретические утверждения, разъяснения и комментарии; междисциплинарные точки зрения на тенденции развития компьютерных технологии; описание рассматриваемых разделов; ретроспективный и перспективный взгляды на изучаемую проблематику. Технология обучения предусматривает систематическое обновление содержания лекционной части курса с использованием принципа «бенчмаркинга» (ориентация на лучшие отечественные и зарубежные аналоги учебной дисциплины); использование балльно-рейтинговой системы для оценки достижения каждым слушателем курса ожидаемых результатов (задач) программы; дополнение рейтинговой системы элементами тестирования; активного «контекстного» обучения, когда мотивация к усвоению знаний достигается путем выстраивания отношений между конкретными знаниями по программе курса и сферами их возможного применения в области профессиональной деятельности, а студенты имеют возможность ассоциировать свой собственный опыт с предметом изучаемого курса. Лекционная часть курса содержит фундаментальные и прикладные научные результаты в области теории, методов и принципов компьютерных технологий.

Самостоятельная работа (СР) направлена на повышение качества обучения, углубление и закрепление знаний студента, развитие аналитических навыков по проблематике учебной дисциплины, активизацию учебно-познавательной деятельности студентов и снижение аудиторной нагрузки. Баллы, полученные по СР студентом, обязательно учитываются при итоговой аттестации по курсу. Формы контроля СР включают устную беседу по теме с преподавателем. Технология обучения предусматривает выработку навыков презентации результатов выполняемых работ и создание условий для командной работы над комплексной темой с распределением функций и ответственности между членами коллектива.

# Учебная карта дисциплины

Курс 1, семестр 1, очная форма обучения

|  |  |  |  |
| --- | --- | --- | --- |
| №  п/п | Виды контрольных мероприятий  (наименования оценочных средств) | Количество баллов | |
| Текущий контроль | Рубежный контроль |
| **Модуль 1. Технологии Big Data** | | | |
| 1 | Практические занятия №№ 1-8 (выполнение, защита результатов) | 16  (8 работ × 2 балла) |  |
| 2 | Лабораторные работы №№ 1–3 (выполнение, подготовка отчёта, защита отчётов) | 24 (3 работы × 8 баллов) | – |
| 3 | Контрольная работа № 1 (тестирование) | – | 20 |
| **Всего** | | **40** | **20** |
| Бонусные баллы | | **до 10 баллов**  За участие в научных конференциях, опубликование тезисов и статей неделе академической мобильности (5 баллов); академическая активность в ходе изучения дисциплины (5 баллов). | |
| **Промежуточная аттестация  в форме экзамена** | | **40 баллов**  Экзамен считается сданным при получении не менее 22 баллов, для допуска к экзамену необходимо набрать не менее 38 баллов по сумме текущего и рубежного контроля.  Оценка по дисциплине выставляется по сумме баллов за текущий контроль, рубежный контроль и экзамен:   * 85–100 баллов – оценка «отлично»; * 71–84 балла – оценка «хорошо»; * 60–70 баллов – оценка «удовлетворительно»; * менее 60 баллов – оценка «неудовлетворительно» | |

# Фонд оценочных средств

## Паспорт фонда оценочных средств

| № п/п | Индикатор достижения компетенции | Наименование оценочного средства |
| --- | --- | --- |
| 1 | ОПК-6.1. Анализирует техническое задание, разрабатывает компоненты программно-аппаратных комплексов для решения задач обработки информации и автоматизированного проектирования | * контрольная работа № 1 (тестирование); * практические занятия № 1-8 (выполнение, защита результатов); * экзаменационные вопросы и билеты |
| 2 | ОПК-6.2. Анализирует функциональные требования к программно-аппаратному обеспечению с учетом национальных стандартов обработки информации и автоматизированного проектирования | * лабораторные работы № 1-3 (выполнение, подготовка отчёта, защита отчёта); * экзаменационные вопросы и билеты |
| 3 | ОПК-6.3. Конфигурирует программно-аппаратные комплексы обработки информации и автоматизированного проектирования | * лабораторные работы № 1-3 (выполнение, подготовка отчёта, защита отчёта); * экзаменационные вопросы и билеты |
| 4 | ОПК-6.4. Адаптирует зарубежные комплексы обработки информации и автоматизированного проектирования в соответствии с национальными стандартами и нуждами отечественных предприятий | * контрольная работа № 1 (тестирование); * практические занятия № 1-8 (выполнение, защита результатов); * экзаменационные вопросы и билеты |

## Контрольная работа № 1 (тестирование)

**База тестовых заданий и ключи**

1. Data Mining – это процесс обнаружения в сырых данных знаний, необходимых для:  
   расчета целевых показателей; **принятия решений в различных сферах человеческой деятельности**; определения области допустимых значений.
2. Атрибут – это:  
   мера оценки; значение переменной; **свойство, характеризующее объект**
3. В процессе работы Data Mining программы пользователь может получить такие результаты:  
   **только статистически достоверные результаты**; полученные на основе методов математического анализа; расчетные показатели процесса.
4. Данные могут быть получены в результате:  
   **наблюдений; экспериментов;** расчетов.
5. Данные представляют собой:  
   **факты; наблюдения;** меры оценки.
6. Данные – это …  
   **необработанный материал, предоставляемый поставщиками данных и используемый потребителями для формирования информации на основе данных**; расчетные характеристики процесса; результаты обработки имеющихся параметров.
7. Два основных типа переменных в статистике:  
   **качественные и количественные**; расчетные и априорные; простые и сложные.
8. Для какой шкалы применимы только такие операции как равно и не равно?  
   порядковая шкала; регрессивная шкала; **номинальная шкала**; прогрессивная шкала.
9. Для какой шкалы применимы только такие операции как равно, не равно, больше, меньше?  
   **порядковая шкала**; регрессивная шкала; номинальная шкала; прогрессивная шкала
10. Задачей классификации можно назвать предсказание...  
    **категориальной зависимой переменной, основываясь на выборке непрерывных и/или категориальных переменных**; категориальной независимой переменной, основываясь на выборке непрерывных и/или категориальных переменных.
11. Задачей регрессии можно назвать предсказание...  
    **числовой зависимой переменной, основываясь на выборке непрерывных и/или категориальных переменных**; текстовой зависимой переменной, основываясь на выборке непрерывных и/или категориальных переменных; числовой независимой переменной, основываясь на выборке непрерывных и/или категориальных переменных
12. Задачи классификации решаются следующими алгоритмами:
    * **нейронные сети**
    * **линейной регрессии**
    * Дейкстры
    * Отжига
13. Задачу классификации нельзя решить с помощью...  
    **алгоритма Apriori**; нейронной сетью; алгоритма линейной регрессии
14. Закономерности, найденные в процессе использования технологии Data Mining должны обладать такими свойствами:
    * **быть практически полезными**
    * **быть объективными**
    * **быть неочевидными**
    * быть точными
15. Изначальная предопределенность классов является характеристикой задачи …  
    кластеризации; **классификации**; интеграции
16. Инструменты Data Mining  
    работают на основе точных методов; **могут самостоятельно строить гипотезы о взаимосвязях в данных**; позволяют найти глобальный оптимум
17. Интервальная шкала – это шкала  
    **разности между значениями которой могут быть вычислены, однако их отношения не имеют смысла**; суммы значений которые могут быть вычислены, однако их отношения не имеют смысла; произведения значений которые могут быть вычислены, однако их отношения не имеют смысла
18. К какой категории данных относится вес измеряемых объектов:  
    **непрерывным данным**; категориальным данным; основным данным
19. К классу описательных задач Data Mining относятся такие задачи:
    * **кластеризация**
    * **поиск ассоциативных правил**
    * размещение
    * компоновка
20. К классу прогнозирующих задач Data Mining относятся такие задачи:
    * размещение
    * компоновка
    * **регрессия**
    * **классификация**
21. Классификация относится к стратегии:  
    **обучения с учителем**; обучение без учителя; обучение учителя
22. Классификация относится к:
    * **управляемому обучению**
    * неуправляемому обучению
    * **контролируемому обучению**
    * обучению по правилам
23. Классификация — это ...  
    отнесение объектов к одному из заранее известных кластеров; **отнесение объектов к одному из заранее известных классов**; отнесение объектов к одной из заранее известных предметных областей
24. Метод деревьев решений применяется для решения задач …  
    кластеризации; идентификации; интеграции; **классификации**
25. Множество примеров, используемое для конструирования модели, называется...  
    **обучающим множеством**; тестовым множеством; эталонным множеством; предметным множеством
26. Множество примеров, используемое для проверки работы сконструированной модели, называется...  
    обучающим множеством; **тестовым множеством**; эталонным множеством; предметным множеством
27. Назовите факторы, обусловившие возникновение и развитие Data Mining:
    * **совершенствование алгоритмов обработки информации**
    * **накопление большого количества ретроспективных данных**
    * **совершенствование технологий хранения и записи данных**
    * **совершенствование аппаратного и программного обеспечения**
    * появление методов математического анализа
28. Номинальная шкала – это шкала,  
    **содержащая только категории, которые не могут упорядочиваться**; содержащая только категории, которые могут упорядочиваться; содержащая только категории, которые могут интегрироваться
29. Объект описывается как …  
    **набор атрибутов**; набор переменных; набор правил; набор характеристик
30. Объектом не является:  
    **переменная**; набор переменных; набор правил
31. Основная характеристика задачи бинарной классификации:  
    **зависимая переменная может принимать только два значения**; зависимая переменная может принимать только три значения; зависимая переменная может принимать только четыре значения
32. Оцените правильность утверждения: "Извлечение полезных сведений невозможно без хорошего понимания сути данных"  
    **утверждение верно**; утверждение неверно
33. Оцените правильность утверждения:"Data Mining может заменить аналитика"
    * **технология всего лишь дает аналитику инструмент для облегчения и улучшения его работы**
    * технология не может дать ответы на те вопросы, которые не были заданы
34. Подготовка данных в процессе Data Mining является:  
    **существенным этапом работы**; подготовительным этапом работы; заключительным этапом работы
35. Порядковая шкала – это шкала, содержащая  
    **категории, которые могут упорядочиваться**; описание предметной области; описание области; описание области допустимых значений

**Спецификация теста**

Данные тестовые задания предназначены для использования в качестве средства рубежного контроля учебных достижений магистрантов по курсу «Технологии Big Data».

Материалы тестовых заданий предусматривают необходимый минимум проверки знаний по дисциплине, а также степени овладения студентами знаниями в области разработки программно-аппаратных. В тесте представлено по **10** вопросов, вес каждого вопроса – **2** балла.

## Практические задания №№ 1-8 (выполнение, защита результатов)

***Цель проведения*:**

Практические задания по дисциплине «Технологии Big Data» необходимы для формирования у обучающихся навыков применения средств и инструментов разработки систем анализа и обработки больших слабоструктурированных данных. В результате выполнения заданий в приведенной серии практических занятий студенты будут обладать знаниями и умениями быстрого создания простых и прототипирования сложных информационных приложений для решения задач Big Data. Для проведения практических занятий используется открытое программное обеспечение. Также организуется постоянный доступ к Интернет. Прикладные задачи выбираются студентами или выдаются преподавателем. Поощряется решение творческих, нетривиальных задач. Предлагается создание групп по 3-4 человека.

| №  задания | Тема | Кол-во часов |
| --- | --- | --- |
| 1 | **Высоконагруженные приложения.** Шардирование.  Реплицирование. Рост нагрузки и падение серверов. Серверный стек технологий | 2 |
| 2 | **Java в Big Data.** Collection Framework. JVM internals. Работа с файлами. Как работает загрузка классов. Пулы соединений/потоков | 2 |
| 3 | **HDFS. «**Демоны» HDFS. Консольные команды. Java API. | 2 |
| 4 | **NoSQL.** Key-value хранилища. Документоориентированные БД. Таблицы в оперативной памяти. Cassandra. CQL. Hbase. | 2 |
| 5 | **MapReduce.** Основная идея. Mapper, Reducer и Combiner. Запуск MR приложения. | 4 |
| 6 | **Hive.** SQL синтаксис для выполнения MR-заданий. Создание таблиц и их заполнение. Техники оптимизации. | 2 |
| 7 | **Основы Scala.** Особенности Scala. Сборка и запуск проекта. Подключение зависимостей в sbt. | 2 |
| 8 | **Spark.** Быстрые вычисления в памяти. Dataset API. Азы машинного обучения. Использование MLlib. | 2 |
| ***Всего часов*** | | 18 |

***Требования к содержанию и оформлению отчётов*:**

Отчет по каждому практическому заданию должен содержать титульный лист с указанием названия задания, фамилий выполнивших студентов, номера группы (подгруппы), фамилию преподавателя; цель практического задания; ход выполнения задания с представлением копий изображений экранов монитора, демонстрирующих основные этапы проведенной работы. Оценивание каждого практического задания проводится во время демонстрации и пояснения полученных результатов.

***Критерии оценивания*:**

Согласно учебной карте дисциплины на практические задания отводится по **18** баллов. Дисциплина содержит **8** практических задания. Каждое практическое задание оценивается максимум в **4** балла и минимум в **2**.

Баллы выставляются обучающимся после демонстрации результатов выполненного задания согласно следующим критериям (критерии представлены из расчета максимальной оценки – **2** балла, расчет критериев для максимальной оценки в **4** балла производится пропорционально представленной):

**2 балла** – обучающийся принимал активное участие в выполнении задания, все поставленные задачи решены самостоятельно и в полном объеме, даваемые пояснения к результатам верны и позволяют высоко оценить уровень полученных навыков;

**1,5 балла** – обучающийся принимал участие в выполнении задания, большинство поставленных задач решены самостоятельно и в полном объеме, даваемые пояснения к результатам в большинстве своем верны и позволяют считать уровень полученных навыков удовлетворительным;

**1 балл** – обучающийся принимал участие в выполнении задания, большинство поставленных задач решены самостоятельно, но не в полном объеме, даваемые пояснения к результатам в большинстве своем верны и позволяют считать уровень полученных навыков достаточным;

**0,5 балла** – обучающийся принимал некоторое участие в выполнении задания, не все поставленные задачи были решены, либо большинство из них решены частично, даваемые пояснения к результатам фрагментарны и не всегда верны, что не позволяет оценить уровень полученных навыков на достаточном уровне;

**0 баллов** – обучающийся не принимал участие в выполнении задания, либо даваемые им пояснения к результатам принципиально неверны.

## Лабораторные работы №№ 1–10 (выполнение, подготовка отчёта, защита отчёта)

| № | Тема | Кол-во часов |
| --- | --- | --- |
| 1 | Лабораторная работа № 1 **Технологический цикл реализации дата-проекта** *(Методология реализации дата-проекта. Типология данных для различных типов проектов. Технологии, методы формирования креативного задания для коммуниканта, программиста, веб-дизайнера. Разработка аналитического кейса. Формирование коммуникантом целеполагания для креативной команды. Распределение функций. Разработка творческого и технического задания. Поиск данных. Выбор аналитического инструментария. Этапы реализации проекта. Типология визуализации, выбор оптимального решения. Корректировка результатов каждого этапа, оптимизация промежуточных результатов. Апробация проекта)* | 4 |
| 2 | Лабораторная работа № 2 **Практикум-апробация дата-проекта в формате хакатона** *(Основные понятия, методология, методы работы профессионального коммуниканта в формате хакатона. Формирование креативной команды. Распределение функций. Разработка творческого и технического задания. Поиск данных. Выбор аналитического инструментария. Этапы реализации проекта. Корректировка результатов каждого этапа, оптимизация промежуточных результатов. Апробация проекта. Выявление и акцентуация особенностей интеграции коммуникации в дата-проекте)* | 6 |
| 3 | Лабораторная работа № 3 **Презентация дата-проекта и обсуждение проблемных моментов** *(Типология дата-проектов. Особенности презентации дата проектов в политике. Специфика дата-проектов в бизнесе. Особенности, проблемные моменты проектов в сфере больших данных. Типология презентаций дата-проектов)* | 6 |

Каждая из лабораторных работ включает в свой состав:

- проведение исследований и экспериментов с использованием аппаратно-программной платформы, проведение обработки (на аудиторных занятиях);

- обработку полученных результатов исследований и экспериментов и составление отчёта о выполнении лабораторной работы, подготовку к защите отчётов о выполнении лабораторных работ (в рамках внеаудиторной самостоятельной работы);

- защиту отчётов о выполнении лабораторных работ (на аудиторных занятиях).

**Требования к содержанию и оформлению отчётов о выполнении лабораторных работ:**

- отчёт должен содержать титульный лист с указанием названия работы, фамилий выполнивших студентов, номера группы (подгруппы, бригады), фамилии преподавателя; цель лабораторной работы; перечень используемых в лабораторной работе программных и аппаратных средств и инструментов; ход работы со всеми полученными результатами исследований и экспериментов в текстовом или графическом виде; обработку результатов исследований и экспериментов в соответствии с требованиями задания на выполнение лабораторной работы, выводы по лабораторной работе;

- отчёт должен быть оформлен в соответствии с требованиями к оформлению текстовых документов (ГОСТ 7.32-2001, ГОСТ 2.105-95).

**Критерии оценивания лабораторных работ**

Согласно учебной карте дисциплины на лабораторные работы отводится **30** баллов. Каждая лабораторная работа оценивается максимум в **10** баллов. При оценивании лабораторной работы принимается во внимание активное участие в выполнении лабораторной работы; соблюдение требований к содержанию и оформлению отчёта о выполнении лабораторной работы, ответы на вопросы в процессе защиты отчёта о выполнении лабораторной работы.

Условием допуска к защите отчёта о выполнении лабораторной работы является факт выполнения исследований и экспериментов лабораторной работы (в рамках аудиторных лабораторных занятий). Обучающийся, не выполнявший лабораторную работу, получает по ней **0** (ноль) баллов.

Оценивание каждой лабораторной работы проводится во время процедуры защиты отчёта о выполнении работы. Оценивание каждой лабораторной работы (в пределах **8** баллов) осуществляется по следующим элементам оценивания:

до **2** баллов – оценивание выполнения работы и пояснений обучающимся хода выполнения работы;

до **4** баллов – оценивание отчёта о выполнении лабораторной работы и пояснений обучающимся содержимого отчёта о выполнении лабораторной работы;

до **2** баллов – оценивание ответа обучающегося на вопросы, связанные с принципами проведения соответствующих видов исследований и экспериментов.

Допускается каждый элемент оценивания лабораторной работы оценивать дробным числом баллов; после суммирования баллов по всем элементам оценивания набранная сумма баллов округляется до целого значения.

**Критерии оценивания выполнения работы и пояснений обучающимся хода выполнения работы, а также ответа обучающегося на вопросы:**

**2 балла** – обучающийся принимал активное участие в выполнении работы, может дать пояснения хода выполнения работы (как во время выполнения, так и на последующих занятиях при защите отчёта о выполнении работы);

**1,25 балла** – обучающийся принимал участие в выполнении работы, может дать частичные пояснения хода выполнения работы (как во время выполнения, так и на последующих занятиях при защите отчёта о выполнении работы); ошибки и неточности в ответах самостоятельно исправляются обучающимся в ходе ответов на дополнительные вопросы;

**0,75 балла** – обучающийся принимал участие в выполнении работы, может дать частичные пояснения хода выполнения работы (как во время выполнения, так и на последующих занятиях при защите отчёта о выполнении работы); обучающийся испытывает затруднения в самостоятельном исправлении ошибок и неточности в ответах;

**0 баллов** – обучающийся не может дать даже частичные пояснения хода выполнения работы, не смотря на участие в выполнении работы (присутствии на лабораторном занятии).

**Критерии оценивания отчёта о выполнении лабораторной работы и пояснений обучающимся содержимого отчёта о выполнении лабораторной работы:**

**4 балла** – отчёт о выполнении работы содержит все требуемые разделы; отчёт соответствует варианту выполнения экспериментальных исследований; отчёт оформлен в соответствии с требованиями к оформлению текстовых документов (ГОСТ 7.32-2001, ГОСТ 2.105-95); отчёт содержит все требуемые результаты; обучающийся может дать полные пояснения любого места отчёта о выполнении лабораторной работы, в том числе использованных процедур струк­турного и объектно-ориентированного анализа и полученных результатов;

**2,1-3,9 балла** – отчёт о выполнении работы содержит все требуемые разделы; отчёт соответствует варианту выполнения экспериментальных исследований; отчёт оформлен в соответствии с требованиями к оформлению текстовых документов; отчёт содержит все требуемые результаты; обучающийся может дать частичные пояснения отдельных фрагментов отчёта о выполнении лабораторной работы, в том числе использованных процедур струк­турного и объектно-ориентированного анализа и полученных результатов; ошибки и неточности в ответах самостоятельно исправляются обучающимся в ходе ответов на дополнительные вопросы;

**0,1-2,0 баллов** – отчёт о выполнении работы содержит основные требуемые разделы (в том числе ход работы и обработку результатов); отчёт соответствует варианту выполнения экспериментальных исследований; отчёт в целом оформлен в соответствии с требованиями к оформлению текстовых документов без грубых нарушений требований; отчёт содержит большинство требуемых результатов; обучающийся может дать частичные пояснения отдельных фрагментов отчёта о выполнении лабораторной работы, в том числе использованных процедур струк­турного и объектно-ориентированного анализа и полученных результатов; обучающийся испытывает затруднения в самостоятельном исправлении ошибок и неточности в ответах;

**0 баллов** – либо отчёт о выполнении работы не содержит все требуемые разделы; либо отчёт не соответствует варианту выполнения экспериментальных исследований; либо отчёт оформлен с грубыми нарушениями требований к оформлению текстовых документов; либо отчёт не содержит большинства требуемых результатов; либо отчёт не содержит большинства требуемых графических нотаций струк­турного и объектно-ориентированного анализа; либо обучающийся не может дать даже частичные пояснения отдельных фрагментов отчёта о выполнении лабораторной работы, в том числе использованных процедур струк­турного и объектно-ориентированного анализа и полученных результатов.

## Экзаменационные вопросы и билеты

Экзамен проводится в классической устной форме. Каждый билет содержит два теоретических вопроса.

***Вопросы к экзамену***:

1. Какой характер роста данных в технологиях BigData?
2. Развитие каких технологий и устройств привело к появлению термина BigData?
3. Что является основным глобальным источником роста объемов данных?
4. Каковы причины актуальности технологий BigData?
5. Какие этапы включает в себя парадигма BigData?
6. Какие факторы тормозят внедрение технологий BigData?
7. Что представляют собой процедуры ETL?
8. Какой основной недостаток полной автоматизации аналитических процедур?
9. Опишите основные характеристики BigData?
10. Опишите основные методы, используемые для решения задач BigData?
11. Подходы NoSQL.
12. Технология Hadoop.
13. Парадигма MapReduce.
14. Файловая система HDFS.
15. Утилиты Hadoop.
16. Состав модуля YARN.
17. Принципы работы с BigData.
18. Платформа Oracle для BigData.
19. Кластер Oracle Big data Appliance.
20. Кластер Oracle Exadata.
21. Кластер Oracle Exalytics.
22. Java Hotspot VM.
23. Oracle Data Mining Algorithms.
24. Oracle Endeca Information Discovery.
25. SAP HANA Data Platform для Big Data.
26. Технологии OLTP BI и OLAP в хранилищах SAP HANA.
27. Задачи аналитики в SAP HANA.
28. Аналитика на уровне СУБД в SAP HANA.
29. Интеграционная модель SAP HANA.
30. Типовые задачи для SAP ESP.
31. Принципы построения модели данных в HBase.
32. Архитектура HBase.
33. Опишите основные процессы HBase.
34. Архитектура Hive.
35. Основные объекты Hive.
36. Партицирование и бакетирование в Hive.
37. Организация обработки запросов в Impala.
38. Опишите основные функции Impala.
39. Опишите операторы Pig Latin.
40. Основные части скрипта на Pig Latin.
41. Опишите состав библиотеки Mahout.
42. Опишите концепцию Spark.
43. Какие типы задач выполняет Oozie?
44. Опишите состав потока работ Oozie.
45. Архитектура Yarn.

**Критерии оценки:**

**(35-40 баллов)** – выставляются на экзамене при полном самостоятельном верном ответе на вопросы билета: даны определения, проанализированы различные точки зрения, концептуальные основы рассматриваемой проблемы, приведены примеры. Или если при ответе были допущены неточности, но студент отвечает на любой вопрос, предложенный экзаменатором. Материал излагается логично и последовательно. Студент свободно владеет понятийным аппаратом, умеет его использовать.

**(29-34 балла)** – выставляются на экзамене при достаточно полном ответе на вопросы билета. В ответе проанализированы не все точки зрения на рассматриваемую проблему. Студент владеет понятийным аппаратом, но при его использовании допускает неточности, испытывает небольшие затруднения при обобщении теоретического материала и формулировке выводов. Допускаются некоторые неточности при изложении фактического материала. Во время беседы задавались наводящие вопросы, на которые студент давал незамедлительные ответы.

**(22-28 баллов)** – выставляются на экзамене при неполном ответе на вопросы билета. Студент в основном знает содержание понятий, но допускает ошибки в их использовании, испытывает значительные трудности при обобщении теоретического материала и в формулировке выводов, не умеет доказательно обосновать свои суждения. Предпочитает отвечать на вопросы воспроизводящего характера и путается при ответах на видоизмененные вопросы. Материал в значительной степени излагается бессистемно и с нарушением логических связей.

Экзамен считается несданным **(0 баллов)**, если студент имеет разрозненные, бессистемные знания, не умеет выделять главное и второстепенное, допускает ошибки в определении понятий, искажает их смысл, беспорядочно и неуверенно излагает материал, не может применять знания для решения практических задач. Или же демонстрирует полное незнание и непонимание учебного материала или отказывается отвечать.

**Форма экзаменационного билета**

 МИНОБРНАУКИ РОССИИ

Федеральное государственное автономное образовательное

учреждение высшего образования

«ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

**ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №** \_\_\_

По дисциплине **Технологии Big Data**

Структурное подразделение **Институт компьютерных технологий и информационной безопасности**

Направление **09.04.01 Информатика и вычислительная техника**

1 Вопрос……………………………………………………………………………

2 Вопрос……………………………………………………………………………

Составитель        \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_И.О.Фамилия(подпись)

«\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_20     г.

Председатель УМС ИКТИБ

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_И.О.Фамилия(подпись)

«\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_20     г.