

Документ подписан простой электронной подписью

Информация о владельце

ФИО: Степанов Павел Иванович

Должность: Руководитель НТИ НИЯУ МИФИ

Дата подписания: 27.02.2026 14:25:54

Уникальный программный ключ

8c65c591e26b2d8e46092774011010071215

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский ядерный университет «МИФИ»

Новоуральский технологический институт –

филиал федерального государственного автономного образовательного учреждения
высшего образования «Национальный исследовательский ядерный университет
«МИФИ»

УТВЕРЖДЕНА

Ученым советом НТИ НИЯУ МИФИ

Протокол №1 от 30.01.2024г.

Рабочая программа учебной дисциплины "Инструменты и технологии анализа больших данных"

Направление подготовки 09.03.01 – Информатика и вычислительная техника

Профиль подготовки Информационные технологии и бизнес-аналитика

Квалификация (степень) выпускника Академический бакалавр

Форма обучения Очно-заочная

	Очно-заочная форма обучения
Семестр	9
Трудоемкость, ЗЕТ	2 ЗЕТ
Трудоемкость, ч.	72 ч.
Аудиторные занятия, в т.ч.:	30 ч.
- лекции	10 ч.
- лабораторные работы	0 ч.
- практические работы	20 ч.
Самостоятельная работа	42 ч.
Контроль	0 ч.
Форма итогового контроля	зачет

Программу составил
ст. преподаватель кафедры АУ

Матвеев В.Е.

СОДЕРЖАНИЕ

1 ЦЕЛИ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ.....	4
2 МЕСТО УЧЕБНОЙ ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВПО	4
3 ФОРМИРУЕМЫЕ КОМПЕТЕНЦИИ И ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ	5
4 ВОСПИТАТЕЛЬНЫЙ ПОТЕНЦИАЛ ДИСЦИПЛИНЫ	6
5 СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ.....	8
5.1 Структура курса «Инструменты и технологии анализа больших данных».....	8
5.2 Содержание лекционных занятий (9-й семестр) – 10 часов	9
5.3 Темы практических занятий (9-й семестр) – 20 часов	9
5.4 Самостоятельная работа – 42 часа	9
6 ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ.....	11
7 ФОНД ОЦЕНОЧНЫХ СРЕДСТВ	12
8 УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ.....	14
9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ	15
Приложение 1. Перечень учебно-методического обеспечения самостоятельной работы студентов.	17
Приложение 2. Методические указания для студентов по освоению дисциплины.....	18
Приложение 3. Балльно-рейтинговая система оценки.....	19
Приложение 4. Фонд оценочных средств.	20

1 ЦЕЛИ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

Целью освоения данной учебной дисциплины является формирование у студентов системных знаний и практических навыков в области современных инструментов и технологий анализа больших данных, необходимых для эффективного сбора, хранения, обработки и анализа больших объемов разнородной информации с использованием современных программных платформ и архитектур, а также для принятия обоснованных управленческих и технических решений на основе полученных данных.

2 МЕСТО УЧЕБНОЙ ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВПО

В соответствии с кредитно-модульной системой подготовки бакалавров по направлению 09.03.01 «Информатика и вычислительная техника» учебная дисциплина «Инструменты и технологии анализа больших данных» имеет индекс Б1.В.01.04, т.е. входит в профессиональный модуль.

Для успешного освоения учебного курса необходимо знание разделов дисциплин: «Программирование», «Информационные технологии».

3 ФОРМИРУЕМЫЕ КОМПЕТЕНЦИИ И ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Универсальные компетенции выпускников и индикаторы их достижения:

Код и наименование общепрофессиональной компетенции	Код и наименование индикатора достижения общепрофессиональной компетенции
<p>УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.</p>	<p>З-УК-1 Знать: методики сбора и обработки информации; актуальные российские и зарубежные источники информации в сфере профессиональной деятельности; метод системного анализа</p> <p>У-УК-1 Уметь: применять методики поиска, сбора и обработки информации; осуществлять критический анализ и синтез информации, полученной из разных источников.</p> <p>В-УК-1 Владеть: методами поиска, сбора и обработки, критического анализа и синтеза информации; методикой системного подхода для решения поставленных задач.</p>

Профессиональные компетенции выпускников и индикаторы их достижения:

Код и наименование общепрофессиональной компетенции	Код и наименование индикатора достижения общепрофессиональной компетенции
<p>ПК-10.3 Способен создавать программный код в соответствии с техническим заданием (готовыми спецификациями).</p>	<p>З-ПК-10.3 Знать: основы программирования, современные объектно-ориентированные языки программирования, современные структурные языки программирования.</p> <p>У-ПК-10.3 Уметь: кодировать на языках программирования, тестировать результаты кодирования.</p> <p>В-ПК-10.3 Владеть: методами разработки кода ИС и баз данных ИС, верификации кода ИС и баз данных ИС относительно дизайна ИС и структуры баз данных ИС, устранения обнаруженных несоответствий.</p>

4 ВОСПИТАТЕЛЬНЫЙ ПОТЕНЦИАЛ ДИСЦИПЛИНЫ

Цели и задачи воспитания, воспитательный потенциал дисциплин:

Направления/цели воспитания	Задачи воспитания (код)	Воспитательный потенциал дисциплин
Профессиональное воспитание	В17 Формирование чувства личной ответственности за научно-технологическое развитие России, за результаты исследований и их последствий.	1.Использование воспитательного потенциала дисциплин профессионального модуля для формирования чувства личной ответственности за достижение лидерства России в ведущих научно-технических секторах и фундаментальных исследованиях, обеспечивающих ее экономическое развитие и внешнюю безопасность, посредством контекстного обучения, обсуждения социальной и практической значимости результатов научных исследований и технологических разработок. 2.Использование воспитательного потенциала дисциплин профессионального модуля для формирования социальной ответственности ученого за результаты исследований и их последствия, развития исследовательских качеств посредством выполнения учебно-исследовательских заданий, ориентированных на изучение и проверку научных фактов, критический анализ публикаций в профессиональной области, вовлечения в реальные междисциплинарные научно-исследовательские проекты.
	В18 Формирование ответственности за профессиональный выбор, профессиональное развитие и профессиональные решения.	1.Организация научно-практических конференций, круглых столов, встреч с выдающимися учеными и ведущими специалистами отраслей реального сектора экономики; научно-проектной деятельности. 2. Участие в подготовке публикаций в высокорейтинговых международных журналах.

<p>В23 формирование культуры информационной безопасности</p>	<p>Использование воспитательного потенциала дисциплин профессионального модуля для формирование базовых навыков информационной безопасности через изучение последствий халатного отношения к работе с информационными системами, базами данных (включая персональные данные), приемах и методах злоумышленников, потенциальном уроне пользователям.</p>
---	---

5 СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

5.1 Структура курса «Инструменты и технологии анализа больших данных»

Общая трудоемкость дисциплины составляет **2 зачетных единиц, 72 часа.**

№ п/п	Название темы/раздела учебной дисциплины	Виды учебной деятельности, включая самостоятельную ра- боту студентов и трудоемкость (в часах)				Текущий контроль (форма*, неделя)	Максимальный балл за раздел	Индикаторы освоения компетенции
		Лекции	Лабораторные занятия	Практические занятия	Самостоятельная работа			
9-й семестр								
1.	Понятие, функции и задачи "больших" данных	2	-	10	18	Пр1	25	3-УК-1 У-УК-1 В-УК-1 3-ПК-10.1 У-ПК-10.1 В-ПК-10.1
2.	Технологии анализа "больших" данных	4	-	10	18	Пр2	25	
3.	Перспективы развития сервисов "больших" данных	4	-	-	6	Тз	10	
Итого:		10	-	20	42		60	
Зачет						ОВ	40	

*Сокращение наименований форм текущего контроля и аттестации разделов:

Тз – тестовое задание; ОВ – Ответ на вопрос; Пр – Практическая работа.

5.2 Содержание лекционных занятий (9-й семестр) – 10 часов

Неделя	Раздел курса, № занятия	Темы лекционных занятий	Трудоемкость, час.
1 - 4	Раздел 1 Л1	Лекция 1. В данной лекции рассматривается определение понятия «Большие данные» (Big Data), основные характеристики и особенности данных, которые отличают их от традиционных наборов информации. Обсуждаются ключевые функции больших данных: сбор, хранение, обработка и аналитика. Особое внимание уделяется задачам, решаемым с помощью больших данных, включая интеграцию разнородных источников, повышение качества данных.	2
5 - 6	Раздел 2 Л2-3	Лекции 2-3. Лекция посвящена обзору современных технологий и инструментов, применяемых для анализа больших данных. Рассматриваются архитектуры распределенного хранения и обработки, такие как Hadoop и Spark, а также технологии потоковой обработки данных (Apache Kafka, Apache Flink).	4
7 - 8	Раздел 3 Л4-5	Лекция 4-5. В лекции рассматриваются тенденции и перспективы развития сервисов больших данных в будущем. Обсуждаются вызовы, связанные с постоянно растущими объемами и разнообразием данных, роль искусственного интеллекта и машинного обучения в автоматизации аналитики. Анализируется влияние облачных технологий, мультиоблачных и гибридных архитектур на построение современных сервисов.	4

5.3 Темы практических занятий (9-й семестр) – 20 часов

Неделя	Раздел курса, № занятия	Темы практических занятий Мероприятие по текущему аудиторному контролю знаний	Трудоемкость, час.
1 - 2	Раздел 1 Пр1	Практическая работа 1. Изучение работы MongoDB.	10
3-4	Раздел 2 Пр2	Практическая работа 2. Основы программирования на Python. NoSQL для обработки больших данных. Clouder Hadoop для обработки больших данных. Deductor Academic для анализа данных.	10

5.4 Самостоятельная работа – 42 часа

Самостоятельная работа студента по учебной дисциплине регламентируется «Положением об организации самостоятельной работы студентов в НТИ НИЯУ МИФИ».

№ п/п	Виды самостоятельной работы / разделы курса	Трудоемкость, час.

№ п/п	Виды самостоятельной работы / разделы курса	Трудоемкость, час.
1.	Изучение текущего материала по теме лекции. Подготовка к практической работе 1.	18
2.	Изучение текущего материала по теме лекции. Подготовка к практической работе 2.	18
3.	Изучение текущего материала по теме лекции. Подготовка к промежуточному тестированию (Тз).	6

Перечень учебно-методического обеспечения самостоятельной работы студентов приведен в Приложении 1.

Методические указания для студентов по освоению дисциплины приведены в Приложении 2.

6 ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

При реализации программы дисциплины «Инструменты и технологии анализа больших данных» используются различные образовательные технологии – аудиторные занятия проводятся в форме лекций и лабораторных (практических) занятий.

В процессе изучения дисциплины на лекциях, которые проводятся в специализированной аудитории, используется мультимедийный проектор и заранее подготовленный демонстрационный материал.

В начале каждого семестра все желающие студенты обеспечиваются электронными версиями методических пособий, имеющихся на кафедре, по изучаемому курсу для работы дома.

На сервере кафедры организован каталог со всеми методическими пособиями, разработанными на кафедре, для возможности постоянного студенческого доступа к ним с любого компьютера во время всех видов занятий.

Самостоятельная работа студентов подразумевает проработку лекционного материала с использованием рекомендуемой литературы (методических пособий по курсу) для подготовки к лабораторным и контрольным работам, контрольным тестам и зачету, а также выполнение контрольных домашних заданий и самостоятельное изучение ряда тем.

Для повышения уровня знаний студентов по курсу «Инструменты и технологии анализа больших данных» в течение семестра организуются консультации преподавателей (согласно графику консультаций кафедры АУ). Во время консультационных занятий:

- проводится объяснение непонятных для студентов разделов теоретического курса;
- разъясняются алгоритмы решения задач индивидуальных домашних заданий;
- принимаются задолженности по тестовым и контрольным работам и т.д.

Перечень учебно-методического обеспечения самостоятельной работы студентов приведен в Приложении 1.

Методические указания для студентов по освоению дисциплины приведены в Приложении 2.

Реализация компетентного подхода предусматривает широкое использование в учебном процессе активных и интерактивных форм проведения занятий, предполагающих активную обратную связь между преподавателем и студентами.

В процессе изучения дисциплины «Инструменты и технологии анализа больших данных» используются интерактивные формы обучения при проведении лабораторных (практических) занятий:

- выступление студентов с докладом по теме для самостоятельного изучения;
- защита домашнего контрольного задания;
- дискуссии;
- презентации.

7 ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Фонд оценочных средств по дисциплине обеспечивает проверку освоения планируемых результатов обучения (компетенций и их индикаторов) посредством мероприятий рубежного и промежуточного контроля по дисциплине. Связь между формируемыми компетенциями и формами контроля их освоения представлена в таблице:

№ п.п.	Контролируемые модули, разделы (темы) дисциплины	Результаты освоения ООП		Виды аттестации		Наименование оценочного средства
		Код контролируемой компетенции	Индикаторы освоения компетенции	Текущий контроль	Промежуточная аттестация	
1	Понятие, функции и задачи "больших" данных	УК-1 ПК-10.1	З-УК-1 У-УК-1 В-УК-1 З-ПК-10.1 У-ПК-10.1 В-ПК-10.1	Пр1	По итогам текущего контроля	Защита практической работы 1
2	Технологии анализа "больших" данных			Пр2	По итогам текущего контроля	Защита практической работы 2
3	Перспективы развития сервисов "больших" данных			Тз	По итогам текущего контроля	Прохождение тестового задания.

В целях повышения эффективности процесса обучения студентов и стимулирования их самостоятельной работы в течение семестра используется система контроля текущей успеваемости и достижения ПР УД, включающая:

- посещение лекций;
- выполнение лабораторных работ;
- посещение семинаров;
- выполнение практических контрольных работ (проверка практических навыков студента);
- выполнение контрольных тестов (программированный экспресс-опрос по теоретическому материалу);
- самостоятельное изучение ряда тем.

Для оценки достижений студента используется балльно-рейтинговая система (Приложение 3).

Для целей промежуточной аттестации используется фонд оценочных средств (ФОС) по дисциплине (Приложение 4).

Результаты каждого тестового задания оцениваются в баллах, на основании которых выставляется оценка.

Задание, по которому проводится тест, считается зачтенным, если по нему набрано не менее половины от максимального количества баллов.

К зачету в конце семестра студент допускается, если он сдал все лабораторные работы, выполнил все тестовые задания на положительные оценки, а также сдал все домашние контрольные задания.

На зачете студенту предлагается выполнить два теоретических вопроса и одно конкретное практическое задание на компьютере по различным темам курса.

Итоговая экзаменационная оценка по курсу выводится с учетом балла, полученного на экзамене, и баллов, полученных по указанным выше компонентам аттестации текущей работы студента в семестре. Шкала перевода баллов в традиционную систему оценок представлена в следующей таблице:

Оценка по 5 бальной шкале	Зачет	Сумма баллов по дисциплине	Оценка (ECTS)	Градация
5 (отлично)	Зачтено	90-100	A	Отлично
4 (хорошо)		85-89	B	Очень хорошо
		75-84	C	Хорошо
		70-74	D	Удовлетворительно
65-69				
3 (удовлетворительно)		60-64	E	Посредственно
2 (неудовлетворительно)	Не зачтено	Ниже 60	F	Неудовлетворительно

8 УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

8.1 Перечень литературы, необходимой для освоения дисциплины

1 Уорд, Б. Инновации SQL Server 2019. Использование технологий больших данных и машинного обучения / Боб Уорд ; пер. с англ. Н. Б. Желновой. - Москва : ДМК Пресс, 2020. - 408 с. - ISBN 978-5-97060-595-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1225362>.

2 Форман, Д. Много цифр. Анализ больших данных при помощи Excel / Форман Д.; Пер. с англ. Соколовой А. - Москва : Альпина Пабли., 2016. - 461 с. ISBN 978-5-9614-5032-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/551044>.

3 Адлер, Ю. П. Статистическое управление процессами. «Большие данные» : учебное пособие / Ю. П. Адлер, Е. А. Черных. - Москва : Изд. Дом МИСиС, 2016. - 52 с. - ISBN 978-5-87623-969-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1232190>

4 О'Нил, К. Data Science. Инсайдерская информация для новичков. Включая язык R : практическое руководство / К. О'Нил, Р. Шатт. - Санкт-Петербург : Питер, 2019. - 368 с. - (Серия «Библиотека программиста»). - ISBN 978-5-4461-0622-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1756149>.

8.2 Перечень ресурсов информационно-коммуникационной сети «Интернет», необходимых для освоения дисциплины

Наименование ресурса	Электронный адрес ресурса
1) ЭБС «Лань»	https://e.lanbook.com
2) Образовательная платформа Юрайт	https://urait.ru/bcode/468952
3) Образовательный портал НИЯУ МИФИ	https://online.mephi.ru/
4) Научная библиотека НИЯУ МИФИ	http://library.mephi.ru/

9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Учебная дисциплина обеспечена учебно-методической документацией и материалами. Её содержание представлено в локальной сети учебного заведения и находится в режиме свободного доступа для студентов. Доступ студентов для самостоятельной подготовки осуществляется через компьютеры библиотеки и компьютерных классов НТИ НИЯУ МИФИ.

Лекционные занятия:

1. комплект электронных презентаций/слайдов,
2. аудитория, оснащенная презентационной техникой (проектор, экран, ноутбук).

Лабораторные работы:

1. лаборатория 106 (Лаборатория локальных систем управления) и 102 (лаборатория микропроцессорных систем) оснащенные персональными компьютерами.

Практические занятия:

1. компьютерный класс,
2. презентационная техника (проектор, экран, ноутбук)
3. электронные тестовые задания, разработанные для данной дисциплины

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ

к рабочей программе по курсу
«Инструменты и технологии анализа больших данных»
для ООП ВПО 09.03.01

на 20___/20___ уч.год

В рабочую программу вносятся следующие изменения:

Рабочая программа пересмотрена и одобрена на заседании кафедры «__»_____20___ г.
Заведующий кафедрой АУ

на 20___/20___ уч.год

В рабочую программу вносятся следующие изменения:

Рабочая программа пересмотрена и одобрена на заседании кафедры «__»_____20___ г.
Заведующий кафедрой АУ

на 20___/20___ уч.год

В рабочую программу вносятся следующие изменения:

Рабочая программа пересмотрена и одобрена на заседании кафедры «__»_____20___ г.
Заведующий кафедрой АУ

Программа действительна

на 20___/20___ уч.год _____ (заведующий кафедрой АУ)

**ПРИЛОЖЕНИЕ 1. ПЕРЕЧЕНЬ УЧЕБНО-МЕТОДИЧЕСКОГО
ОБЕСПЕЧЕНИЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ.**

№	Литература	Год	Курс	Номер группы	Семестр	Кол-во студентов	Кол-во книг	Коэффициент книгообеспеченности
Основная литература								
1	Уорд, Б. Инновации SQL Server 2019. Использование технологий больших данных и машинного обучения / Боб Уорд ; пер. с англ. Н. Б. Желновой. - Москва : ДМК Пресс, 2020. - 408 с. - ISBN 978-5-97060-595-0. - Текст : электронный. - URL: https://znanium.com/catalog/product/1225362 .	2020	5	ИТ-54	9	14	14	1,0
2	Форман, Д. Много цифр. Анализ больших данных при помощи Excel / Форман Д.; Пер. с англ. Соколовой А. - Москва : Альпина Пабли., 2016. - 461 с. ISBN 978-5-9614-5032-3. - Текст : электронный. - URL: https://znanium.com/catalog/product/551044 .	2016	5	ИТ-54	9	14	14	1,0
Дополнительная литература								
1	Адлер, Ю. П. Статистическое управление процессами. «Большие данные» : учебное пособие / Ю. П. Адлер, Е. А. Черных. - Москва : Изд. Дом МИСиС, 2016. - 52 с. - ISBN 978-5-87623-969-3. - Текст : электронный. - URL: https://znanium.com/catalog/product/1232190	2016	5	ИТ-54	9	14	10	1,0
2	О'Нил, К. Data Science. Инсайдерская информация для новичков. Включая язык R : практическое руководство / К. О'Нил, Р. Шатт. - Санкт-Петербург : Питер, 2019. - 368 с. - (Серия «Библиотека программиста»). - ISBN 978-5-4461-0622-6. - Текст : электронный. - URL: https://znanium.com/catalog/product/1756149 .	2019	5	ИТ-54	9	14	10	1,0

ПРИЛОЖЕНИЕ 2. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ СТУДЕНТОВ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ.

- Стандарт организации СТО НТИ-2-2014. Требования к оформлению текстовой документации;
- Методические рекомендации по организации самостоятельной работы обучающихся НТИ НИЯУ МИФИ.

ПРИЛОЖЕНИЕ 3. БАЛЛЬНО-РЕЙТИНГОВАЯ СИСТЕМА ОЦЕНКИ.

Распределение баллов текущего рейтинга по видам деятельности студента направления подготовки 09.03.01 при изучении курса "Инструменты и технологии анализа больших данных" предоставлены в таблице 3.1.

Таблица 3.1 – Распределение баллов текущего рейтинга за семестр 9

№ п/п	Наименование раздела	Рубежный контроль	Максимальный балл
1	Понятие, функции и задачи "больших" данных	Пр1	25
2	Технологии анализа "больших" данных	Пр2	25
3	Перспективы развития сервисов "больших" данных	Тз	10
4	Зачет		40
ИТОГО			100

ПРИЛОЖЕНИЕ 4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ.

Для оценки результатов обучения в зависимости от оцениваемого средства используются следующие шкалы оценок:

Критерии оценок	Шкала оценок
1	2
Тест	
Полнота знаний теоретического контролируемого материала	При текущем контроле знаний количество баллов определяется количеством правильных ответов на вопросы теста: Тестовое задание 1 (Тз) – макс. 10 <i>баллов</i>
Зачет	
Полнота знаний теоретического контролируемого материала	При промежуточной аттестации количество баллов определяется качеством и полнотой ответа студента на предоставленный вопрос. Задание на зачет – макс. 30 баллов Задание на зачет – ответить на два вопроса из приведенного списка. За каждый вопрос – макс. 20 баллов: 15 баллов ставится за полный ответ на вопрос. 10 баллов ставится за достаточно полный ответ на вопрос с незначительными недочетами. 5 баллов ставится в случае неполного ответа на вопрос. 0 баллов ставится, если в беседе со студентом выясняется, что он не знает основных понятий и определений курса. В индивидуальном порядке по теме лекций могут быть заданы на зачете дополнительные вопросы (из перечня).

Материалы, необходимые для оценки результатов обучения

Перечень вопросов к зачету (9 семестр)

1. Что понимается под термином «Большие данные» (Big Data)?
2. Назовите три основных характеристики больших данных, известные как «3 V».
3. Какую роль выполняет система HDFS в экосистеме Hadoop?
4. В чем суть метода MapReduce?
5. Что такое Apache Spark и чем он отличается от Hadoop MapReduce?
6. Какие типы данных входят в категорию структурированных, полуструктурированных и неструктурированных?
7. Объясните принцип работы Apache Kafka.
8. Что такое ETL-процесс и зачем он нужен в аналитике больших данных?
9. Для чего используется язык запросов HiveQL?
10. Что представляет собой Data Lake и чем он отличается от Data Warehouse?
11. Какие задачи решают системы NoSQL и в чем их отличие от традиционных реляционных баз данных?
12. Что такое потоковая обработка данных и какие инструменты для нее применяются?
13. Опишите, что такое кластеризация данных и в чем ее цель.
14. В чем преимущество хранения данных в колоночных форматах, таких как Parquet или ORC?
15. Что такое Data Mining и какие задачи он помогает решать?
16. Какие основные этапы включает в себя процесс подготовки данных (data preprocessing)?
17. Объясните, что такое машинное обучение и какие типы алгоритмов применяются в анализе данных.
18. Как обеспечивается масштабируемость систем анализа больших данных?

19. Что такое временной ряд и где применяется его анализ?
20. В чем разница между OLTP и OLAP системами?
21. Как организуется распределенная обработка данных в рамках кластера?
22. Какие инструменты хорошо подходят для визуализации больших данных?
23. Что такое индексирование данных и какую роль он играет в аналитике?
24. Опишите назначение и особенности Apache Flink.
25. Что означает термин «шардинг» в контексте баз данных?
26. Расскажите о принципах работы и назначении Apache Airflow.
27. Почему обработка данных in-memory ускоряет анализ в приложениях типа Spark?
28. Какие форматы данных используются для обмена и хранения больших данных?
29. В чем суть и преимущества параллельной обработки данных?
30. Как обеспечивается безопасность и управление доступом в системах больших данных?

Тестовое задание 1
по дисциплине «Инструменты и технологии анализа больших данных»

1. Что такое «большие данные» (Big Data)?
 - a) Небольшой набор структурированных данных
 - b) Данные, превышающие возможности традиционных СУБД
 - c) Данные, которые хранятся только в облаке
 - d) Исключительно текстовые данные
2. Какое из следующих свойств относится к «трем V» больших данных?
 - a) Скорость, валидность, визуальность
 - b) Объём, разнообразие, скорость
 - c) Вероятность, величина, вариативность
 - d) Визуализация, валидация, верификация
3. Какой инструмент чаще всего используется для распределённой обработки больших данных?
 - a) Microsoft Excel
 - b) Hadoop
 - c) MySQL
 - d) PowerBI
4. Что такое MapReduce?
 - a) Язык программирования для анализа данных
 - b) Модель программирования для обработки больших данных
 - c) Система хранения данных
 - d) Визуализационный инструмент
5. Какой язык используется для запросов в базе данных Hadoop Hive?
 - a) SQL
 - b) Python
 - c) JavaScript
 - d) C++
6. Что из нижеперечисленного не является типом данных в больших данных?
 - a) Структурированные
 - b) Полуструктурированные
 - c) Неструктурированные
 - d) Унифицированные
7. Какая технология обеспечивает потоковую обработку данных в реальном времени?
 - a) Apache Spark Streaming
 - b) Hadoop MapReduce
 - c) Excel Power Query
 - d) Tableau
8. Что из перечисленного является одной из основных задач анализа больших данных?
 - a) Игнорирование шума в данных
 - b) Визуализация данных
 - c) Оптимизация запросов в СУБД
 - d) Ручное хранение данных
9. Как называется процесс преобразования сырых данных в удобный для анализа формат?
 - a) Хранение
 - b) Очистка данных
 - c) Сжатие
 - d) Архивирование
10. Какой инструмент используется для визуализации и построения отчетов по большим данным?
 - a) Microsoft Word
 - b) Tableau

- c) Hadoop
 - d) Git
11. Что такое ETL-процесс в контексте анализа данных?
 - a) Экспорт, трансфер, лимитирование
 - b) Извлечение, трансформация, загрузка
 - c) Электронная текстовая лингвистика
 - d) Анализ временных линий
 12. Какой фреймворк больше подходит для in-memory аналитики?
 - a) Hadoop MapReduce
 - b) Apache Spark
 - c) MySQL
 - d) Oracle DB
 13. Что из нижеперечисленного лучше всего описывает NoSQL базы данных?
 - a) Исключительно реляционные данные
 - b) Гибкая схема и поддержка несвязанных данных
 - c) Не поддерживают масштабирование
 - d) Поддержка только текстовых данных
 14. Какой язык программирования наиболее часто применяется для анализа больших данных?
 - a) C#
 - b) Python
 - c) HTML
 - d) Swift
 15. Что из перечисленного не является примером источника больших данных?
 - a) Социальные сети
 - b) Сенсоры IoT
 - c) Классические текстовые документы на диске пользователя
 - d) Веб-сайты и журналы серверов
 16. Какова основная задача кластеризации в анализе данных?
 - a) Разделение данных на группы по схожим характеристикам
 - b) Сортировка данных
 - c) Удаление повторяющихся данных
 - d) Сжатие данных
 17. Какой из следующих форматов данных часто применяется для хранения полуструктурированных данных?
 - a) CSV
 - b) JSON
 - c) TXT
 - d) XLS
 18. Что такое Data Lake?
 - a) Хранилище структурированных данных
 - b) Централизованное хранилище для всех типов данных
 - c) Система мониторинга данных
 - d) Архив старых данных
 19. Какая технология применяется для обеспечения масштабируемого хранения больших объемов данных?
 - a) FTP
 - b) HDFS (Hadoop Distributed File System)
 - c) SMTP
 - d) DNS
 20. Что из перечисленного является преимуществом использования облачных платформ для анализа больших данных?
 - a) Ограниченная пропускная способность
 - b) Масштабируемость и гибкость ресурсов

- c) Исключение безопасности данных
- d) Необходимость локального оборудования

Тестовое задание 2
по дисциплине «Инструменты и технологии анализа больших данных»

Какой компонент Hadoop отвечает за распределённое хранение данных?

- a) YARN
- b) HDFS
- c) MapReduce
- d) Spark

Какой формат файла является наиболее эффективным для хранения колоночных данных?

- a) CSV
- b) JSON
- c) Parquet
- d) TXT

Что такое Data Warehouse?

- a) Хранилище временных данных
- b) Специализированное хранилище для аналитических данных
- c) Хранилище неструктурированных данных
- d) Онлайн база данных для транзакций

Какой из инструментов предназначен для построения ETL-процессов?

- a) Apache NiFi
- b) Tableau
- c) Hadoop HDFS
- d) Microsoft Word

В чем основное отличие Apache Spark от Hadoop MapReduce?

- a) Spark работает только с текстовыми файлами
- b) Spark поддерживает in-memory вычисления, MapReduce — нет
- c) Spark — это облачная платформа, Hadoop — нет
- d) Spark не поддерживает распределённую обработку

Что из перечисленного не является платформой для анализа больших данных?

- a) Apache Flink
- b) Apache Kafka
- c) Microsoft Excel
- d) Google BigQuery

Для чего используется Apache Kafka?

- a) Для хранения данных
- b) Для обработки и передачи потоковых данных
- c) Для визуализации данных
- d) Для машинного обучения

Что такое машинное обучение (Machine Learning) в контексте больших данных?

- a) Метод хранения данных
- b) Метод создания моделей для прогнозирования на основе данных
- c) Язык программирования
- d) Файл формата для данных

Какой из перечисленных алгоритмов используется для классификации в анализе данных?

- a) K-means
- b) Linear Regression
- c) Decision Tree
- d) PCA

Что такое «шардинг» в контексте СУБД?

- a) Объединение нескольких баз данных

- b) Разделение базы данных на несколько частей для повышения производительности
- c) Резервное копирование данных
- d) Процесс визуализации данных

Что означает понятие «временной ряд» в анализе данных?

- a) Последовательность данных, собранных во времени
- b) Объем данных
- c) Множество данных без временных меток
- d) Тип базы данных

Какие данные можно охарактеризовать как неструктурированные?

- a) Табличные данные
- b) Текст, изображения, видео
- c) JSON и XML
- d) Данные из SQL баз

Что обозначает термин «Data Mining»?

- a) Процесс добычи полезной информации из больших массивов данных
- b) Хранение данных
- c) Архивирование данных
- d) Визуализация данных

Какой принцип лежит в основе работы распределённых систем хранения данных?

- a) Централизованное хранение
- b) Дублирование и репликация данных на нескольких узлах
- c) Автоматическое удаление данных
- d) Хранение только текстовых файлов

Какой тип анализа позволяет выявлять связи и шаблоны в больших объемах данных?

- a) Описательный анализ
- b) Диагностический анализ
- c) Предиктивный анализ
- d) Разведочный анализ

Для чего предназначен язык программирования R?

- a) Управление базами данных
- b) Статистический анализ и визуализация данных
- c) Веб-разработка
- d) Системное программирование

Какой формат файлов наиболее часто применяется для обмена табличными данными?

- a) PDF
- b) CSV
- c) MP3
- d) HTML

Какой термин описывает процесс преобразования данных из одного формата в другой?

- a) Трансформация
- b) Инициализация
- c) Кэширование
- d) Индексирование

Что из перечисленного является примером инструмента для автоматизации рабочих процессов с данными?

- a) Apache Airflow
- b) Adobe Photoshop
- c) Microsoft PowerPoint
- d) Notepad++

Какая из следующих технологий лучше всего подходит для анализа потоковых данных?

- a) Apache Hadoop
- b) Apache Storm
- c) MySQL
- d) SQLite