ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Иркутский государственный университет путей сообщения»
(ФГБОУ ВО ИрГУПС)

УТВЕРЖДЕНА
приказом ректора
от «02» июня 2023 г. № 424-1

Б1.В.ДВ.11.01 Базы данных Big Data

рабочая программа дисциплины

Специальность/направление подготовки – 27.03.02 Управление качеством
Специализация/профиль – Управление качеством в производственно-технологических системах
Квалификация выпускника – Бакалавр
Форма и срок обучения – очная форма 4 года
Кафедра-разработчик программы – Информационные системы и защита информации

Общая трудоемкость в з.е. – 2
Часов по учебному плану (УП) – 72
В том числе в форме практической подготовки (ПП) – 6
(очная)

Формы промежуточной аттестации
очная форма обучения:
зачет 8 семестр

Очная форма обучения **Распределение часов дисциплины по семестрам**

Семестр	8	**Итого**
Вид занятий	Часов по УП	**Часов по УП**
Аудиторная контактная работа по видам учебных занятий/ в т.ч. в форме ПП*	24/6	**24/6**
– лекции	12	**12**
– практические (семинарские)		
– лабораторные	12/6	**12/6**
Самостоятельная работа	48	**48**
Итого	72/6	**72/6**

* В форме ПП – в форме практической подготовки.

ИРКУТСК

Электронный документ выгружен из ЕИС ФГБОУ ВО ИрГУПС и соответствует оригиналу
Подписант ФГБОУ ВО ИрГУПС Трофимов Ю.А.
00920FD815CE68F8C4CA795540563D259C с 07.02.2024 05:46 по 02.05.2025 05:46 GMT+03:00
Подпись соответствует файлу документа

Рабочая программа дисциплины разработана в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по направлению подготовки 27.03.02 Управление качеством, утвержденным Приказом Министерства науки и высшего образования Российской Федерации от 31.07.2020 № 869.

Программу составил(и):
К.э.н., доцент, заведующий кафедрой, Т. К. Кириллова

Рабочая программа рассмотрена и одобрена для использования в учебном процессе на заседании кафедры «Информационные системы и защита информации», протокол от «2» июня 2023 г. № 12

Зав. кафедрой, к. э. н, доцент Т.К. Кириллова

СОГЛАСОВАНО

Кафедра «Управление качеством и инженерная графика», протокол от «2» июня 2023 г. № 9

Зав. кафедрой, к. т. н., доцент Е.Д. Молчанова

1 ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ	
1.1 Цели дисциплины	
1	изучение обучающимися современного состояния теории Big Data
2	приобретение знаний о технологиях подготовки, хранения, обработки и анализа больших данных
3	применение статистических и математических методов для анализа больших объемов информации
1.2 Задачи дисциплины	
1	изучить современное состояние теории Big Data
2	дать представление обучающимся об основных методах хранения и анализа больших данных в том числе в сфере управления качеством
3	сформировать представления о возможностях и ограничениях основных методов обработки, анализа больших данных
1.3 Цель воспитания и задачи воспитательной работы в рамках дисциплины	
Профессионально-трудовое воспитание обучающихся	
Цель профессионально-трудового воспитания – формирование у обучающихся осознанной профессиональной ориентации, понимания общественного смысла труда и значимости его для себя лично, ответственного, сознательного и творческого отношения к будущей деятельности, профессиональной этики, способности предвидеть изменения, которые могут возникнуть в профессиональной деятельности, и умению работать в изменённых, вновь созданных условиях труда. Цель достигается по мере решения в единстве следующих задач: – формирование сознательного отношения к выбранной профессии; – воспитание чести, гордости, любви к профессии, сознательного отношения к профессиональному долгу, понимаемому как личная ответственность и обязанность; – формирование психологии профессионала; – формирование профессиональной культуры, этики профессионального общения; – формирование социальной компетентности и другие задачи, связанные с имиджем профессии и авторитетом транспортной отрасли	

2 МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП	
Блок/часть ОПОП	Блок 1. Дисциплины / Часть, формируемая участниками образовательных отношений
2.1 Дисциплины и практики, на которых основывается изучение данной дисциплины	
1	Б1.О.23 Основы обеспечения качества
2	Б1.О.25 Метрология
3	Б1.О.40 История управления качеством
4	Б1.О.44 Статистические методы в управлении качеством
5	Б1.В.ДВ.03.01 Взаимоотношения с потребителями в системе менеджмента качества
6	Б1.В.ДВ.04.01 Аудит системы менеджмента качества
7	Б1.В.ДВ.05.01 Управление процессами
8	Б1.В.ДВ.06.01 Подтверждение соответствия
9	Б1.В.ДВ.07.01 Средства и методы управления качеством
10	Б1.В.ДВ.09.01 Документирование в системе менеджмента качества
11	Б2.О.03(П) Производственная - технологическая (производственно-технологическая) практика
2.2 Дисциплины и практики, для которых изучение данной дисциплины необходимо как предшествующее	
1	Б3.02(Д) Защита выпускной квалификационной работы

3 ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ТРЕБОВАНИЯМИ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения
ПК-1 Способен осуществлять работу по управлению качеством продукции (работ, услуг)	ПК-1.1 Анализирует причины снижения качества продукции (работ, услуг) и разработка предложений по их устранению	Знать: основные понятия и терминологию в области технологий обработки и анализа больших данных; способы оценки снижения качества продукции; основное программное обеспечение работы по анализу больших наборов данных
		Уметь: разрабатывать и анализировать концептуальные и теоретические модели прикладных задач анализа больших данных; использовать и применять углубленные знания в области управления качеством; создавать алгоритмы анализа и обработки большого объема данных с применением моделей Data Mining
		Владеть: навыками создания и решения моделей,

		необходимых в сфере управления; программными инструментами для работы с большими наборами данных; методикой прогнозирования

4 СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Код	Наименование разделов, тем и видов работ	Очная форма					*Код индикатора достижения компетенции
		Семестр	Часы				
			Лек	Пр	Лаб	СР	
1.0	**Раздел 1. Технологии хранения и обработки больших данных.**						
1.1	Тема 1. Современные подходы к обработке и хранению	8	2		2	8	ПК-1.1
1.2	Тема 2. Извлечение и визуализация данных	8	2		2	8	ПК-1.1
1.3	Тема 3. Подготовка данных к анализу. Методика извлечения знаний	8	2		2/2	8	ПК-1.1
2.0	**Раздел 2. Информационное и программное обеспечение Big Data.**						
2.1	Тема 4. Алгоритмы параллельной обработки данных	8	2		2/2	8	ПК-1.1
2.2	Тема 5. Аналитические платформы: классификация и особенности применения	8	2		2/2	8	ПК-1.1
2.3	Тема 6. Понятие сценария и узла обработки	8	2		2	8	ПК-1.1
	Форма промежуточной аттестации – зачет	8					ПК-1.1
	Итого часов (без учёта часов на промежуточную аттестацию)		12		12/6	48	

5 ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

Фонд оценочных средств для проведения текущего контроля успеваемости и промежуточной аттестации по дисциплине оформлен в виде приложения № 1 к рабочей программе дисциплины и размещен в электронной информационно-образовательной среде Университета, доступной обучающемуся через его личный кабинет

6 УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1 Учебная литература

6.1.1 Основная литература

	Библиографическое описание	Кол-во экз. в библиотеке/ онлайн
6.1.1.1	Базы данных : методические указания / составители А. Р. Мангушева. — Казань : КНИТУ, 2017. — 16 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/138352 (дата обращения: 14.05.2024).	Онлайн
6.1.1.2	Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 3-е изд., стер. — Санкт-Петербург : Лань, 2023. — 188 с. — ISBN 978-5-507-46866-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/322664 (дата обращения: 14.05.2024).	Онлайн
6.1.1.3	Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/182452 (дата обращения: 14.05.2024).	Онлайн
6.1.1.4	Берикашвили, В. Ш. Статистическая обработка данных, планирование эксперимента и случайные процессы : учебное пособие для бакалавриата и магистратуры / В. Ш. Берикашвили, С. П. Оськин. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2019. — 164 с. — (Бакалавр и магистр. Академический курс). — ISBN 978-5-534-09216-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/427449 (дата обращения: 14.05.2024)	Онлайн

6.1.2 Дополнительная литература

	Библиографическое описание	Кол-во экз. в библиотеке/ онлайн
6.1.2.1	Бутаков, Н. А. Обработка больших данных с Apache Spark : учебно-методическое пособие : [16+] / Н. А. Бутаков, М. В. Петров, Д. Насонов. – Санкт-Петербург : Университет ИТМО, 2019. – 52 с. : ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=566771 (дата обращения: 14.05.2024).	Онлайн
6.1.2.2	Информационное обеспечение и базы данных : учебное пособие / составитель А. Ф. Похилько. — Ульяновск : УлГТУ, 2019. — 127 с. — ISBN 978-5-9795-1964-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/165031 (дата обращения: 14.05.2024).	Онлайн
6.1.3 Учебно-методические разработки (в т. ч. для самостоятельной работы обучающихся)		
	Библиографическое описание	Кол-во экз. в библиотеке/ онлайн
6.1.3.1	Кириллова, Т.К. Методические указания по изучению дисциплины Б1.В.ДВ.11.01 Базы данных Big Data по направлению подготовки 27.03.02 Управление качеством, профиль Управление качеством в производственно-технологических системах / Т.К.Кириллова; ИрГУПС. – Иркутск : ИрГУПС, 2023. – 11 с. - Текст: электронный. - URL: https://www.irgups.ru/eis/for_site/umkd_files/mu_7740_1492_2023_1_signed.pdf	Онлайн
6.2 Ресурсы информационно-телекоммуникационной сети «Интернет»		
6.2.1	Научная электронная библиотека eLIBRARY.RU — https://elibrary.ru/	
6.2.2	Электронная библиотека Учебно-методического центра по образованию на железнодорожном транспорте «ЭБ УМЦ ЖДТ» — https://umczdt.ru/books/	
6.2.3	Электронно-библиотечная система «Издательство Лань», https://e.lanbook.com/	
6.2.4	Электронно-библиотечная система «Образовательная платформа ЮРАЙТ», https://urait.ru/	
6.2.5	Электронно-библиотечная система «Университетская библиотека онлайн», https://biblioclub.ru/	
6.3 Программное обеспечение и информационные справочные системы		
6.3.1 Базовое программное обеспечение		
6.3.1.1	Microsoft Windows Professional 10, государственный контракт от 20.07.2021 № 0334100010021000013-01	
6.3.1.2	Microsoft Office Russian 2010, государственный контракт от 20.07.2021 № 0334100010021000013-01	
6.3.1.3	FoxitReader, свободно распространяемое программное обеспечение http://free-software.com.ua/pdf-viewer/foxit-reader/	
6.3.1.4	Adobe Acrobat Reader DC свободно распространяемое программное обеспечение https://get.adobe.com/ru/reader/enterprise/	
6.3.1.5	Яндекс. Браузер. Прикладное программное обеспечение общего назначения, Офисные приложения, лицензия – свободно распространяемое программное обеспечение по лицензии BSD License	
6.3.2 Специализированное программное обеспечение		
6.3.2.1	Packet Tracer УЧ. ПРОЦ. Универсальная общественная лицензия GNU, http://www.packettracernetwork.com/	
6.3.2.2	PuTTY свободно распространяемый клиент для различных протоколов удалённого доступа УЧ. ПРОЦ. http://www.putty.org/	
6.3.3 Информационные справочные системы		
6.3.3.1	Не предусмотрены	
6.4 Правовые и нормативные документы		
6.4.1	Не предусмотрены	

7 ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ УЧЕБНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ

1	Корпуса А, Б, В, Г, Д, Е ИрГУПС находятся по адресу г. Иркутск, ул. Чернышевского, д. 15; корпус Л ИрГУПС находится – по адресу г. Иркутск, ул. Лермонтова, д.80
2	Учебная аудитория Д-822 для проведения лекционных и практических занятий, лабораторных работ, групповых и индивидуальных консультаций, курсового проектирования (выполнения курсовых работ), текущего контроля и промежуточной аттестации Основное оборудование: специализированная мебель, мультимедиапроектор, экран, (ноутбук переносной).
3	Лаборатория Д-508 «Информационные системы и сетевые технологии». «Сети и системы передачи информации» для проведения лекционных и практических занятий, лабораторных работ, групповых и индивидуальных консультаций, курсового проектирования (выполнения курсовых работ), текущего

	контроля и промежуточной аттестации. Основное оборудование: специализированная мебель, мультимедиапроектор переносной), экран (переносной), компьютер.
4	Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой, подключенной к информационно-телекоммуникационной сети «Интернет», и обеспечены доступом в электронную информационно-образовательную среду ИрГУПС. Помещения для самостоятельной работы обучающихся: – читальные залы; – учебные залы вычислительной техники А-401, А-509, А-513, А-516, Д-501, Д-503, Д-505, Д-507; – помещения для хранения и профилактического обслуживания учебного оборудования – А-521

8 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Вид учебной деятельности	Организация учебной деятельности обучающегося
Лекция	Лекция (от латинского «lection» – чтение) – вид аудиторных учебных занятий. Лекция: закладывает основы научных знаний в систематизированной, последовательной, обобщенной форме; раскрывает состояние и перспективы развития соответствующей области науки и техники; концентрирует внимание обучающихся на наиболее сложных, узловых вопросах; стимулирует познавательную активность обучающихся. Во время лекционных занятий обучающийся должен уметь сконцентрировать внимание на изучаемых проблемах и включить в работу все виды памяти: словесную, образную и моторно-двигательную. Для этого весь материал, излагаемый преподавателем, обучающемуся необходимо конспектировать. На полях конспекта следует помечать вопросы, выделенные обучающимся для консультации с преподавателем. Выводы, полученные в виде формул, рекомендуется в конспекте подчеркивать или обводить рамкой, чтобы лучше запоминались. Полезно составить краткий справочник, содержащий определения важнейших понятий лекции. К каждому занятию следует разобрать материал предыдущей лекции. Изучая материал по учебнику или конспекту лекций, следует переходить к следующему вопросу только в том случае, когда хорошо усвоен предыдущий вопрос. Ряд вопросов дисциплины может быть вынесен на самостоятельное изучение. Такое задание требует оперативного выполнения. В конспекте лекций необходимо оставить место для освещения упомянутых вопросов. Обозначить вопросы, термины, материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удается разобраться в материале, то необходимо сформулировать вопрос и задать преподавателю на консультации, на практическом занятии
Лабораторная работа	Основной целью лабораторных работ является теоретическое обоснование, наглядное и/или экспериментальное подтверждение и/или проверка существенных теоретических положений (законов, закономерностей) анализ существующих методик и методов их реализации и т.д. Они занимают преимущественное место при изучении дисциплин обязательной части и части, формируемой участниками образовательных отношений Блока 1. Исходя из цели, содержанием лабораторных работ могут быть: - экспериментальная проверка формул, методик расчета; - проведение натурных измерений свойств, рабочих параметров, режимов работы при помощи лабораторного оборудования и/или стендов и макетов; - ознакомление, анализ и теоретические выкладки по устройству, принципу действия и способам обслуживания аппаратов, деталей машин, механизмов, процессов, протекающих в них при этом и т.д.; - наглядная графическая интерпретация чертежей, схем, объемных поверхностей и т.д., воспроизводимых с помощью специализированного программного обеспечения; - имитационное моделирование процессов, протекающих в сложных химических, физических, механических, электрических и пр. объектах; - наглядное представление о работе персонала конкретной организации или подразделения ОАО «РЖД» посредством моделирования штатных и внештатных ситуаций в виртуальных специализированных АРМ (автоматизированных рабочих мест); - установление и подтверждение закономерностей (путем сравнения проведенного эксперимента и рассчитанных значений) и т.д.; - ознакомление с методиками проведения экспериментов, наглядным устройством стенд-макетов и пр.; - установление свойств веществ, их качественных и количественных характеристик; - анализ различных характеристик процессов, в том числе производственных и иных процессов; - расчет параметров различных явлений и процессов, смоделировать которые не

	возможно в реальных условиях (например, чрезвычайные ситуации и пр.); - наблюдение развития явлений, процессов и др. Допускается иное содержание лабораторных работ, если это будет способствовать реализации целей и задач дисциплины и формированию соответствующих компетенций. По характеру выполняемых лабораторных работ возможны: - ознакомительные работы, используемые для закрепления изученного теоретического материалы; - аналитические работы, используемые для получения новой информации на основе формализованных методов; - творческие работы, ориентированные на самостоятельный выбор подходов решения задач. Прежде, чем приступить к лабораторным занятиям, обучающимся необходимо повторить теоретический материал по теме работы. Каждая лабораторная работа оснащена методическими указаниями, разработанными преподавателями, ведущими дисциплину
Самостоятельная работа	Обучение по дисциплине «Базы данных Big Data» предусматривает активную самостоятельную работу обучающегося. В разделе 4 рабочей программы, который называется «Структура и содержание дисциплины», все часы самостоятельной работы расписаны по темам и вопросам, а также указана необходимая учебная литература: обучающийся изучает учебный материал, разбирает примеры и решает разноуровневые задачи в рамках выполнения как общих домашних заданий, так и индивидуальных домашних заданий (ИДЗ) и других видов работ, предусмотренных рабочей программой дисциплины. При выполнении домашних заданий обучающемуся следует обратиться к задачам, решенным на предыдущих практических занятиях, решенным домашним работам, а также к примерам, приводимым лектором. Если этого будет недостаточно для выполнения всей работы можно дополнительно воспользоваться учебными пособиями, приведенными в разделе 6.1 «Учебная литература». Если, несмотря на изученный материал, задание выполнить не удается, то в обязательном порядке необходимо посетить консультацию преподавателя, ведущего практические занятия, и/или консультацию лектора. Домашние задания, индивидуальные домашние задания и другие работы, предусмотренные рабочей программой дисциплины должны быть выполнены обучающимся в установленные преподавателем сроки в соответствии с требованиями к оформлению текстовой и графической документации, сформулированным в Положении «Требования к оформлению текстовой и графической документации. Нормоконтроль»
Комплекс учебно-методических материалов по всем видам учебной деятельности, предусмотренным рабочей программой дисциплины (модуля), размещен в электронной информационно-образовательной среде ИрГУПС, доступной обучающемуся через его личный кабинет	

Приложение № 1 к рабочей программе

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

для проведения текущего контроля успеваемости и промежуточной аттестации

1. Общие положения

Фонд оценочных средств (ФОС) является составной частью нормативно-методического обеспечения системы оценки качества освоения обучающимися образовательной программы.

Фонд оценочных средств предназначен для использования обучающимися, преподавателями, администрацией ИрГУПС, а также сторонними образовательными организациями для оценивания качества освоения образовательной программы и уровня сформированности компетенций у обучающихся.

Задачами ФОС являются:

– оценка достижений обучающихся в процессе изучения дисциплины;

– обеспечение соответствия результатов обучения задачам будущей профессиональной деятельности через совершенствование традиционных и внедрение инновационных методов обучения в образовательный процесс;

– самоподготовка и самоконтроль обучающихся в процессе обучения.

Фонд оценочных средств сформирован на основе ключевых принципов оценивания: валидность, надежность, объективность, эффективность.

Для оценки уровня сформированности компетенций используется трехуровневая система:

– минимальный уровень освоения, обязательный для всех обучающихся по завершению освоения образовательной программы; дает общее представление о виде деятельности, основных закономерностях функционирования объектов профессиональной деятельности, методов и алгоритмов решения практических задач;

– базовый уровень освоения, превышение минимальных характеристик сформированности компетенций; позволяет решать типовые задачи, принимать профессиональные и управленческие решения по известным алгоритмам, правилам и методикам;

– высокий уровень освоения, максимально возможная выраженность характеристик компетенций; предполагает готовность решать практические задачи повышенной сложности, нетиповые задачи, принимать профессиональные и управленческие решения в условиях неполной определенности, при недостаточном документальном, нормативном и методическом обеспечении.

2. Перечень компетенций, в формировании которых участвует дисциплина.
Программа контрольно-оценочных мероприятий.
Показатели оценивания компетенций, критерии оценки

Дисциплина «Базы данных Big Data» участвует в формировании компетенций:

ПК-1. Способен осуществлять работу по управлению качеством продукции (работ, услуг)

Программа контрольно-оценочных мероприятий очная форма обучения

№	Наименование контрольно-оценочного мероприятия	Объект контроля	Код индикатора достижения компетенции	Наименование оценочного средства (форма проведения*)
8 семестр				
1.0	**Раздел 1. Технологии хранения и обработки больших данных**			
1.1	Текущий контроль	Тема 1. Современные подходы к обработке и хранению	ПК-1.1	Лабораторная работа (письменно/устно) Собеседование (устно)
1.2	Текущий контроль	Тема 2. Извлечение и визуализация данных	ПК-1.1	Лабораторная работа (письменно/устно)
1.3	Текущий контроль	Тема 3. Подготовка данных к анализу. Методика извлечения знаний	ПК-1.1	Доклад (устно) В рамках ПП**: Лабораторная работа (письменно/устно)
2.0	**Раздел 2. Информационное и программное обеспечение Big Data**			
2.1	Текущий контроль	Тема 3. Алгоритмы параллельной обработки данных	ПК-1.1	Доклад (устно) В рамках ПП**: Лабораторная работа (письменно/устно)
2.2	Текущий контроль	Тема 4. Аналитические платформы: классификация и особенности применения	ПК-1.1	Доклад (устно) В рамках ПП**: Лабораторная работа (письменно/устно)
2.3	Текущий контроль	Тема 5. Понятие сценария и узла обработки	ПК-1.1	Лабораторная работа (письменно/устно)
	Промежуточная аттестация	Раздел 1. Технологии хранения и обработки больших данных. Раздел 2. Информационное и программное обеспечение Big Data.	ПК-1.1	Зачет (собеседование) Зачет - тестирование (компьютерные технологии)

*Форма проведения контрольно-оценочного мероприятия: устно, письменно, компьютерные технологии.

**ПП – практическая подготовка

Описание показателей и критериев оценивания компетенций.
Описание шкал оценивания

Контроль качества освоения дисциплины включает в себя текущий контроль успеваемости и промежуточную аттестацию. Текущий контроль успеваемости и промежуточная аттестация обучающихся проводятся в целях установления соответствия достижений обучающихся поэтапным требованиям образовательной программы к результатам обучения и формирования компетенций.

Текущий контроль успеваемости – основной вид систематической проверки знаний, умений, навыков обучающихся. Задача текущего контроля – оперативное и регулярное управление учебной деятельностью обучающихся на основе обратной связи и корректировки. Результаты оценивания учитываются в виде средней оценки при проведении промежуточной аттестации.

Для оценивания результатов обучения используется четырехбалльная шкала: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно» и двухбалльная шкала: «зачтено», «не зачтено».

Перечень оценочных средств, используемых для оценивания компетенций, а также краткая характеристика этих средств приведены в таблице.

Текущий контроль

№	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в ФОС
1	Собеседование	Средство контроля на практическом занятии, организованное как специальная беседа преподавателя с обучающимся на темы, связанные с изучаемой дисциплиной, и рассчитанное на выяснение объема знаний обучающегося по определенному разделу, теме, проблеме и т.п. Может быть использовано для оценки знаний обучающихся	Вопросы для собеседования по темам/разделам дисциплины
2	Доклад	Продукт самостоятельной работы обучающегося, представляющий собой публичное выступление по представлению полученных результатов решения определенной учебно-практической, учебно-исследовательской или научной темы. Может быть использовано для оценки знаний, умений, навыков и (или) опыта деятельности обучающихся	Темы докладов
3	Лабораторная работа	Средство, позволяющее оценить умение обучающегося письменно/устно излагать суть поставленной задачи, самостоятельно применять стандартные методы решения поставленной задачи с использованием имеющейся лабораторной базы, проводить анализ полученного результата работы. Может быть использовано для оценки умений, навыков и (или) опыта деятельности обучающихся	Образец задания для выполнения лабораторной работы и примерный перечень вопросов для ее защиты

Промежуточная аттестация

№	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в ФОС
1	Зачет	Средство, позволяющее оценить знания, умения, навыков и (или) опыта деятельности обучающегося по дисциплине. Может быть использовано для оценки знаний, умений, навыков и (или) опыта деятельности обучающихся	Перечень теоретических вопросов и практических заданий к зачету
2	Тест – промежуточная аттестация в форме зачета	Система автоматизированного контроля освоения компетенций (части компетенций) обучающимся по дисциплине (модулю) с использованием информационно-коммуникационных технологий. Может быть использовано для оценки знаний, умений, навыков и (или) опыта деятельности обучающихся	Фонд тестовых заданий

Критерии и шкалы оценивания компетенций в результате изучения дисциплины при проведении промежуточной аттестации в форме зачета. Шкала оценивания уровня освоения компетенций

Шкала оценивания	Критерии оценивания	Уровень освоения компетенции
«зачтено»	Обучающийся правильно ответил на теоретические вопросы. Показал отличные знания в рамках учебного материала. Правильно выполнил практические задания. Показал отличные умения и владения навыками	Высокий

	применения полученных знаний и умений при решении задач в рамках учебного материала. Ответил на все дополнительные вопросы	
	Обучающийся с небольшими неточностями ответил на теоретические вопросы. Показал хорошие знания в рамках учебного материала. С небольшими неточностями выполнил практические задания. Показал хорошие умения и владения навыками применения полученных знаний и умений при решении задач в рамках учебного материала. Ответил на большинство дополнительных вопросов	Базовый
	Обучающийся с существенными неточностями ответил на теоретические вопросы. Показал удовлетворительные знания в рамках учебного материала. С существенными неточностями выполнил практические задания. Показал удовлетворительные умения и владения навыками применения полученных знаний и умений при решении задач в рамках учебного материала. Допустил много неточностей при ответе на дополнительные вопросы	Минимальный
«не зачтено»	Обучающийся при ответе на теоретические вопросы и при выполнении практических заданий продемонстрировал недостаточный уровень знаний и умений при решении задач в рамках учебного материала. При ответах на дополнительные вопросы было допущено множество неправильных ответов	Компетенция не сформирована

Тест – промежуточная аттестация в форме зачета

Шкала оценивания	Критерии оценивания
«зачтено»	Обучающийся верно ответил на 70 % и более тестовых заданий при прохождении тестирования
«не зачтено»	Обучающийся верно ответил на 69 % и менее тестовых заданий при прохождении тестирования

Критерии и шкалы оценивания результатов обучения при проведении текущего контроля успеваемости

Собеседование

Шкалы оценивания		Критерии оценивания
«отлично»	«зачтено»	Глубокое и прочное усвоение программного материала. Полные, последовательные, грамотные и логически излагаемые ответы при видоизменении задания. Обучающийся свободно справляется с поставленными задачами, может обосновать принятые решения, демонстрирует владение разносторонними навыками и приемами выполнения практических работ
«хорошо»		Знание программного материала, грамотное изложение, без существенных неточностей в ответе на вопрос, правильное применение теоретических знаний, владение необходимыми навыками при выполнении практических задач
«удовлетворительно»		Обучающийся демонстрирует усвоение основного материала, при ответе допускаются неточности, при ответе недостаточно правильные формулировки, нарушение последовательности в изложении программного материала, затруднения в выполнении практических заданий Слабое знание программного материала, при ответе возникают ошибки, затруднения при выполнении практических работ
«неудовлетворительно»	«не зачтено»	Не было попытки выполнить задание

Доклад

Шкалы оценивания		Критерии оценивания
«отлично»	«зачтено»	Доклад создан с использованием компьютерных технологий

		(презентация PowerPoint, Flash–презентация, видео-презентация и др.) Использованы дополнительные источники информации. Содержание заданной темы раскрыто в полном объеме. Отражена структура доклада (вступление, основная часть, заключение, присутствуют выводы и примеры). Оформление работы. Оригинальность выполнения (работа сделана самостоятельно, представлена впервые)
«хорошо»		Доклад создан с использованием компьютерных технологий (презентация PowerPoint, Flash–презентация, видео-презентация и др.) Содержание доклада включает в себя информацию из основных источников (методическое пособие), дополнительные источники информации не использовались. Содержание заданной темы раскрыто не в полном объеме. Структура доклада сохранена (вступление, основная часть, заключение, присутствуют выводы и примеры)
«удовлетворительно»		Доклад сделан устно, без использования компьютерных технологий. Содержание доклада ограничено информацией только из методического пособия. Содержание заданной темы раскрыто не в полном объеме. Отсутствуют выводы и примеры. Оригинальность выполнения низкая
«неудовлетворительно»	«не зачтено»	Доклад сделан устно, без использования компьютерных технологий и других наглядных материалов. Содержание ограничено информацией только из методического пособия. Заданная тема доклада не раскрыта, основная мысль доклада не передана

Лабораторная работа

Шкалы оценивания		Критерии оценивания
«отлично»	«зачтено»	Лабораторная работа выполнена в обозначенный преподавателем срок, письменный отчет без замечаний. Лабораторная работа выполнена обучающимся в полном объеме с соблюдением необходимой последовательности. Обучающийся работал полностью самостоятельно; показал необходимые для проведения работы теоретические знания, практические умения и навыки. Работа (отчет) оформлена аккуратно, в наиболее оптимальной для фиксации результатов форме
«хорошо»		Лабораторная работа выполнена в обозначенный преподавателем срок, письменный отчет с небольшими недочетами. Лабораторная работа выполнена обучающимся в полном объеме и самостоятельно. Допущены отклонения от необходимой последовательности выполнения, не влияющие на правильность конечного результата. Работа показывает знание обучающимся основного теоретического материала и овладение умениями, необходимыми для самостоятельного выполнения работы. Допущены неточности и небрежность в оформлении результатов работы (отчета)
«удовлетворительно»		Лабораторная работа выполнена с задержкой, письменный отчет с недочетами. Лабораторная работа выполняется и оформляется обучающимся при посторонней помощи. На выполнение работы затрачивается много времени. Обучающийся показывает знания теоретического материала, но испытывает затруднение при самостоятельной работе с источниками знаний или приборами
«неудовлетворительно»	«не зачтено»	Лабораторная работа не выполнена, письменный отчет не представлен. Результаты, полученные обучающимся, не позволяют сделать правильных выводов и полностью расходятся с поставленной целью. Показывается плохое знание теоретического материала и отсутствие необходимых умений. Лабораторная работа не выполнена, у учащегося отсутствуют необходимые для проведения работы теоретические знания, практические умения и навыки

3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности

3.1 Типовые контрольные задания для проведения собеседования

Контрольные варианты заданий выложены в электронной информационно-образовательной среде ИрГУПС, доступной обучающемуся через его личный кабинет.

Ниже приведен образец типовых вариантов заданий для проведения собеседований.

Образец типового варианта вопросов для проведения собеседования
«Тема 1. Современные подходы к обработке и хранению»

1. Инструментальные средства по машинному обучению
2. Проверка адекватности моделей бинарного выбора. Селекция моделей на основе информационных критериев.
3. Модели множественного выбора с неупорядоченными альтернативами.
4. Модель наивного байесовского классификатора.
5. Алгоритмы дерева принятия решения для решения задач классификатора

3.2 Типовые контрольные темы для написания докладов

Контрольные варианты заданий выложены в электронной информационно-образовательной среде ИрГУПС, доступной обучающемуся через его личный кабинет.

Ниже приведен образец типовых вариантов тем для написания докладов.

Образец тем докладов
«Тема 3. Подготовка данных к анализу. Методика извлечения знаний»

1. Модель данных и информационная архитектура предприятия.
2. ETL: автоматизация подготовки данных.
3. Инкрементальная и автоматическая выгрузка. Пайплайн.
4. Хранилище данных с несколькими источниками.
5. Документно-ориентированная СУБД MongoDB.
6. Использование неструктурированных данных в качестве источника
7. Колоночно-аналитическая СУБД Vertic

Образец тем докладов
«Тема 4. Алгоритмы параллельной обработки данных»

1. Алгоритмы случайного леса для решения задач классификации
2. Ансамбль алгоритмов randomforest.
3. Классификации данных методом опорных векторов.
4. Алгоритм линейный классификации Вапника.
5. Метод релевантных векторов.
6. Метод градиентного бустинга.
7. Ресемплинг как метод бустрап-анализа.
8. Алгоритм сэмплирования SMOTE

Образец тем докладов
«Тема 5. Аналитические платформы: классификация и особенности применения»

1. Apache Airflow как инструмент разработки, планирования и мониторинга рабочих процессов.
2. Озеро данных (Data Lake).
3. Hadoop для масштабируемой, распределенной обработки больших наборов данных.
4. Модель распределенных вычислений MapReduce.
5. Потоковая обработка данных. Kafka.
6. Облачные технологии для масштабирования систем. Яндекс. Облако.

3.3 Типовые задания для выполнения лабораторной работы и примерный перечень вопросов для ее защиты

Контрольные варианты заданий выложены в электронной информационно-образовательной среде ИрГУПС, доступной обучающемуся через его личный кабинет.

Ниже приведен образец типовых вариантов заданий для выполнения лабораторных

работ и примерный перечень вопросов для их защиты.

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Тема 1. Современные подходы к обработке и хранению»

В таблице представлен образец датасета, представляющий собой фрагмент лог-файла, фиксирующего события запуска (started) и остановки (как успешной (finished), так и неуспешной (failed)) системных служб некоторой UNIX-системы.

Дата	Время	Служба	Статус
2022-10-03	13:25	sshd	started
2022-12-12	10:01	ftpd	finished
2022-12-12	11:45	ftpd	failed
2022-12-12	12:14	ftpd	failed
2022-12-12	09:23	sshd	failed
…			
2020-12-08	00:04	httpd	failed

а) Составьте алгоритм MapReduce, позволяющий на основе данных, представленных выше, решить следующую задачу:

Определите количество дат, за которые служба sshd завершила некорректно свою работу большее количество раз, чем служба ftpd .

Оформите его в виде псевдокода, показав реализацию основных операций над данными.

б) Определите классы – параметры шаблонных классов Mapper и Reducer. Напишите прототипы функций map(…) и reduce(…).

в) Требуется ли в данной задаче для повышения производительности создавать специальную реализацию класса Combiner? Если да, то каким образом? Если нет, то почему? Поясните ответы.

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Тема 2. Извлечение и визуализация данных»

1. Установите соединение с кластером и откройте удаленный рабочий стол. Номер учетной записи выберите в соответствии с номером рабочего компьютера в аудитории. Логин и пароль к учетной записи получите у преподавателя.

2. Исследуйте веб-интерфейс Cloudera Manager, доступный по адресу http://localhost:7180, позволяющий управлять Hadoop-кластером и получать информацию о системе в удобном графическом представлении. Проанализируйте представленную информацию.

3. Создайте в домашней директории папку, в которой будут располагаться Ваши файлы. В имя директории желательно включить фамилии пользователей:

$ mkdir ~/ivanov_directory

В созданной директории создайте еще одну папку:

$ mkdir ~/ivanov_directory/WordCount

5. Скопируйте файл с исходным кодом WordCount.java и набор текстовых файлов в созданную директорию WordCount с помощью программы WinSCP.

6. Узнайте объем свободного пространства в файловой системе HDFS и выведите содержимое Вашей домашней папки в ней:

$ hadoop fs -df -h

$ hadoop fs -ls /user/user0

7. Создайте в системе HDFS папку, аналогичную созданной ранее:

$ hadoop fs -mkdir -p /user/user[0-9]/ivanov_directory/WordCount

8. Аналогичным образом создайте в системе HDFS директорию, необходимую для хранения исследуемых текстовых файлов:

$ hadoop fs -mkdir /user/user[0-9]/ivanov_directory/WordCount/input

9. Скопируйте полученные у преподавателя текстовые файлы в папку input. Проверьте их наличие после копирования:

$ hadoop fs -copyFromLocal ~/ivanov_directory/WordCount/*.txt /user/user[0-9]/ivanov_directory/WordCount/input

$ hadoop fs -ls /user/user[0-9]/ivanov_directory/WordCount/input

10. Установите требуемые для работы переменные окружения:

$ export JAVA_HOME=/usr/java/jdk1.7.0_67-cloudera/

$ export HADOOP_CLASSPATH=$HADOOP_CLASSPATH:/ $JAVA_HOME/lib/tools.jar

11. Перейдите в директорию WordCount. Скомпилируйте файл с исходным кодом WordCount.java, который представляет программу, определяющую количество повторений слов в наборе текстовых файлов.

Для этого в командной строке выполните следующую команду:

$ hadoop com.sun.tools.javac.Main WordCount.java

12. Соберите Java-пакет из всех полученных в результате компиляции файлов:

$ $JAVA_HOME/bin/jar cf wc.jar WordCount*.class

13. Запустите программу для подсчета количества слов в системе Hadoop:

$ hadoop jar wc.jar WordCount /user/user[0-9]/ivanov_directory/ WordCount/input /user/user[0-9]/ivanov_directory/WordCount/output

Если система сообщает о том, что директория вывода уже создана, необходимо удалить ее командой:

$ hadoop fs -rm -r /user/user0/ivanov_directory/WordCount/output

и запустить программу еще раз.

14. По окончанию работы программы проверьте статус ее завершения. Если выполнение программы произошло с ошибками, исправьте их и запустите расчет вновь.

15. Добейтесь успешного завершения программы. Зафиксируйте и проанализируйте консольный вывод приложения. Просмотрите информацию, предоставляемую Cloudera Manager о завершившейся задаче.

16. Проверьте содержимое директории output. Отметьте формат файлов, находящихся в ней.

$ hadoop fs -ls /user/user[0-9]/ivanov_directory/WordCount/output

17. Выведите содержимое всех файлов в директории output в файл result.txt в текущей директории.

$ hadoop fs -getmerge /user/user0/ivanov_directory/WordCount/output/ result.txt

18. Просмотрите содержание полученного файла с помощью текстового редактора и проанализируйте его содержимое. Фрагмент файла поместите в протокол.

19. Доработайте программу по подсчету числа вхождения слов в текстовые файлы, добавив в нее еще один MapReduce этап. Данный этап должен принимать на вход результат подсчета количества вхождений слов в тексты, осуществленный программой WordCount, и выводить слова с характеристикой встречаемости их в тексте (согласно таблице 1):

Таблица 1. Соотношение между количеством слов и характеристикой встречаемости

Количество вхождений слова в тексты	Количество вхождений слова в тексты
1 – 10	Очень редко
10 – 50	Редко
50 – 200	Часто
> 200	Очень часто

Проанализируйте результаты, выводимые модифицированной программой. Доработанный исходный код программы, результаты ее работы, скрины информационных окон Cloudera Manager поместите в протокол. Сделайте выводы.

Вопросы:

1. Понятие больших данных (Big Data). Проблемы, связанные с обработкой больших данных традиционными инструментами. Необходимость обработки больших данных.

2. Понятие технологии MapReduce, основные принципы, лежащие в ее основе. Характеристики этапов map и reduce. Примеры задач MapReduce.
3. Hadoop как распределенная система обработки больших данных. Стек технологий Hadoop. Масштабируемость и производительность Hadoop. Принцип определения доступности узлов. Понятие Heartbeat.
4. Понятие HDFS. Архитектура HDFS. NameNode и DataNode: понятие и функции. Обеспечение надежности хранения данных. Понятие репликации. Характеристики и ключевые особенности HDFS как файловой системы.
5. Процесс чтения и записи данных в HDFS.

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Тема 3. Подготовка данных к анализу. Методика извлечения знаний»

Изучение технологии Apache Spar

Установите соединение с кластером и откройте удаленный рабочий стол. Номер учетной записи выберите в соответствии с номером рабочего компьютера в аудитории.

2. Создайте в домашней директории папку, в которой будутрасполагаться Ваши файлы. В имя директории желательно включить фамилии пользователей. В этой директории создайте еще одну папку, в которой будет храниться Spark-проект.

3. Скопируйте файл с исходным кодом WordCount.java и набор текстовых файлов в созданную директорию SparkWordCount с помощью программы WinSCP.

4. Создайте аналогичную папку в системе HDFS:

$ hadoop fs -mkdir -p ~/ivanov_directory/SparkWordCount

5. Создайте в системе HDFS директорию, необходимую для хранения исследуемых текстовых файлов:

$ hadoop fs -mkdir ivanov_directory/SparkWordCount/input

6. Скопируйте полученные у преподавателя текстовые файлы в папку input. Проверьте их наличие после копирования:

$ hadoop fs -copyFromLocal ~/ivanov_directory/SparkWordCount/
*.txt /user/user[0-9]/ivanov_directory/SparkWordCount/input
$ hadoop fs -ls /user/user[0-9]/ivanov_directory/SparkWordCount/
input

7. Установите необходимые для работы программы переменные
окружения:

$ export JAVA_HOME=/usr/java/jdk1.7.0_67-cloudera/
$ export HADOOP_CLASSPATH=$(echo /cloudera/parcels/CDH/
jars/*.jar | tr ' ' ':'):.

8. Перейдите в директорию SparkWordCount. Скомпилируйте файл с исходным кодом WordCount.java, который представляет программу, определяющую количество повторений слов в наборе текстовых файлов.

Для этого в командной строке выполните следующую команду:

$ $JAVA_HOME/bin/javac -classpath $HADOOP_CLASSPATH -d .
WordCount.java

9. Соберите Java-пакет из всех полученных в результате компиляции файлов:

$ $JAVA_HOME/bin/jar cf wc.jar WordCount*.class

10. Запустите программу для подсчета количества слов в Spark:

$ spark-submit --class WordCount --master yarn --deploy-mode
cluster --executor-memory 512m wc.jar ivanov_directory/
SparkWordCount/input ivanov_directory/SparkWordCount/output

11. Дождитесь окончания работы задачи. Обратите внимание на ее статус завершения. Представьте скрин вывода. Определите id-задачи в системе yarn (строка вида application_1454960730847_0050). Получите лог работы задачи и результаты ее работы и перенаправьте вывод в локальный файл следующей командой:

$ yarn logs -applicationId application_1454960730847_0050 >

~/ivanov_directory/SparkWordCount/SparkWordCount.log

12. По окончанию работы программы проверьте статус ее завершения. Если выполнение программы произошло с ошибками, исправьте их и запустите расчет вновь.

13. Добейтесь успешного завершения программы. Зафиксируйте и проанализируйте консольный вывод приложения. Просмотрите информацию о Spark, предоставляемую Cloudera Manager. На вкладке «History server Web UI» найдите изображение графа DAG, соответствующего выполнившемуся заданию (см. рисунок 2), и поместите его в протокол.

14. Проверьте содержимое директории output. Выведите содержимое всех файлов в директории output в файл result.txt в текущей директории. $ hadoop fs -getmerge ivanov_directory/SparkWordCount/output/ result.txt

15. Просмотрите содержание полученного файла с помощью текстового редактора и проанализируйте его содержимое. Фрагмент файла поместите в протокол.

16. Доработайте программу по подсчету числа вхождения слов в текстовые файлы, добавив в нее требуемое количество MapReduce-фаз, определяющих количество слов с заданной характеристикой встречаемости в текстах (см. таблицу 1).

17. Проанализируйте результаты, выводимые модифицированной программой. Доработанный исходный код программы, результаты ее работы, скрины информационных окон Cloudera Manager поместите в протокол. Сделайте выводы.

Вопросы:

1. Фреймворк обработки больших данных Spark, его назначение, функции и отличия от Hadoop MapReduce.
2. Понятие устойчивого распределенного набора данных (RDD). Понятие раздела RDD (partition). Способы создания RDD. Трансформации (transformations) и действия (actions). Кэширование (cache) данных в Spark.
3. RDD и PairRDD: понятие, назначение. Основные трансформации (transformations) и действия (actions) над ними. Понятие кортежа (tuple2), связь между JavaPairRDD и JavaRDD посредством кортежа.
4. Реализация концепции MapReduce в фреймворке Spark. Функции map, flatMap, mapValues, reduce, reduceByKey.
5. Архитектура среды Spark времени выполнения. Модели запуска Spark-приложений (YARN, Mesos, Standalone).
6. Понятие драйвера (driver) и исполнителей (executors). Понятия задания (job), этапа (stage) и задачи (task). Модель ленивых вычислений (lazy) и ее применение в Spark.
7. Понятие ориентированного ациклического графа (Directed Acyclic Graph).

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Тема 4. Алгоритмы параллельной обработки данных»

1. Реализуйте последовательную обработку элементов вектора, например, умножение элементов вектора на число. Число элементов вектора задается параметром *N*.

2. Реализуйте многопоточную обработку элементов вектора, используя разделение вектора на равное число элементов. Число потоков задается параметром *M*.

3. Выполните анализ эффективности многопоточной обработки при разных параметрах (10, 100, 1000, 100000) и *M* (2, 3, 4, 5, 10). Результаты представьте в табличной форме.

4. Выполните анализ эффективности при усложнении обработки каждого элемента вектора.

5. Исследуйте эффективность разделения по диапазону при неравномерной вычислительной сложности обработки элементов вектора.

6. Исследуйте эффективность параллелизма при круговом разделении элементов вектора. Сравните с эффективностью разделения по диапазону.

Вопросы:

1. Почему эффект от распараллеливания наблюдается только при большем числе элементов?

2. Почему увеличение сложности обработки повышает эффективность многопоточной обработки?

3. Какое число потоков является оптимальным для конкретной вычислительной системы?

4. Почему неравномерность загрузки потоков приводит к снижению эффективности многопоточной обработки?

5. Какие другие варианты декомпозиции позволяют увеличить равномерность загрузки потоков?

6. В какой ситуации круговая декомпозиция не обеспечивает равномерную загрузку потоков?

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Тема 5. Аналитические платформы: классификация и особенности применения»

Аналитическая платформа (АП) «Deductor» применима для решения большого спектра задач, таких как создание аналитической отчетности, прогнозирование, поиск закономерностей и пр. Можно сказать, что данная система применима в задачах, где требуется консолидация и отображение данных различными способами, построение моделей и последующее применение полученных моделей к новым данным.

Задание

1. Опишите назначение и возможности АП «Deductor».
2. Запустите программу «Deductor Studio Academic»,ознакомьтесь с назначением кнопок и контекстным меню главного окна программы.
3. Воспользуйтесь мастером импорта данных (импортируйте любой файл, например из C:\Program Files\BaseGroup\ Deductor\ Samples\ *.txt).
4. Ознакомьтесь с доступными способами обработки данных.
5. Изучите возможности мастера визуализации и экспорта. Какие параметры доступны для мастера экспорта данных?
6. Создайте отчет.

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Тема 6. Понятие сценария и узла обработки»

1. Создайте новый проект и сохраните его под именем stud_name.

2. Заполните свойства проекта.

3. Просмотрите файл проекта через любой текстовый редактор.

4. Сделайте видимой вкладку Подключения.

5. Создайте текстовый файл с данными, содержащими буквенные, цифровые форматы, а также формат даты, в виде матрицы, и сохраните его в текстовом редакторе на рабочем диске под именем тест2.txt.

6. Создайте новый проект и сохраните его под именем stud2_name.

7. Проведите импорт подготовленного текстового файла в созданный проект, настройте параметры импорта таким образом, чтобы данные файла отображались корректно.

8. Метку узла переименуйте посредством контекстного меню в Проба импорта. В комментарии к узлу впишите: Текстовый файл тест2.txt с разделителями.

9. Добавьте к узлу узел Настройка набора данных и задайте следующие метки к столбцам:

Поле1, Поле2, Поле3 и далее по количеству столбцов.

10. Экспортируйте набор данных в текстовый файл с настройками, предлагаемыми по умолчанию, сохраняя файл с названием пробаэкспорт.txt на рабочем диске.

11. Импортируйте только что экспортированный файл в Deductor, задав в поле Мастера импорта путь сохраненного файла.

12. Присоедините к новому узлу импорта (путем копирования) предыдущую ветвь, начиная с узла Настройка набора данных.

13. Между экспортом и настройкой набора данных вставьте еще один узел настройки,

в котором измените тип столбца Поле 3 на логический.

14. Удалите только что вставленный узел.

15. Сохраните проект.

Сформируйте отчет по проделанной работе, сохраняя каждый момент работы с платформой путем копирования и сохранения экрана и переноса картинок в текстовый редактор. Отчет по выполненной работе должен содержать:

1. Титульный лист.
2. Постановку задачи и цель работы.
3. Данные в виде твердой копии подготовленного текстового файла тест2.txt.
4. Распечатку табличного и графического материала, сохраненного в файл по мере работы в аналитической платформе.
5. Замечания о ходе выполнения работы.
6. Выводы о работе аналитической платформы Deductor

Вопросы:

1. Из каких частей состоит Deductor?
2. Какие варианты поставки Deductor существуют?
3. Чем отличается версия Professional от Academic?
4. Имеются ли ограничения по количеству обрабатываемых записей в версии Deductor Academic?
5. Сколько категорий пользователей Deductor можно выделить?

3.4 Типовые контрольные задания для проведения тестирования

Фонд тестовых заданий по дисциплине содержит тестовые задания, распределенные по разделам и темам, с указанием их количества и типа.

Структура фонда тестовых заданий по дисциплине

Индикатор достижения компетенции	Тема в соответствии с РПД	Характеристика ТЗ	Количество тестовых заданий, типы ТЗ
ПК-1.1	Тема 1. Современные подходы к обработке и хранению	Знание	2 – ОТЗ 2 – ЗТЗ
		Умение	2 – ОТЗ 3 – ЗТЗ
ПК-1.1	Тема 2. Извлечение и визуализация данных	Знание	4 – ОТЗ 4 – ЗТЗ
		Умение	2 – ОТЗ 2 – ЗТЗ
		Навык и (или) опыт деятельности/ действие	1 – ОТЗ 1 – ЗТЗ
ПК-1.1	Тема 3. Подготовка данных к анализу. Методика извлечения знаний	Знание	3 – ОТЗ 2 – ЗТЗ
		Умение	1 – ОТЗ 1 – ЗТЗ
ПК-1.1	Тема 3. Алгоритмы параллельной обработки данных	Знание	2 – ОТЗ 2 – ЗТЗ
		Умение	2 – ОТЗ 2 – ЗТЗ
		Навык и (или) опыт деятельности/ действие	1 – ОТЗ 1 – ЗТЗ
ПК-1.1	Тема 4. Аналитические платформы: классификация и особенности применения	Знание	4 – ОТЗ 4 – ЗТЗ
		Умение	2 – ОТЗ 2 – ЗТЗ
		Навык и (или) опыт деятельности/	2 – ОТЗ 2 – ЗТЗ

		действие	
ПК-1.1	Тема 5. Понятие сценария и узла обработки	Знание	5 – ОТЗ 5 – ЗТЗ
		Итого	60

Полный комплект ФТЗ хранится в электронной информационно-образовательной среде ИрГУПС и обучающийся имеет возможность ознакомиться с демонстрационным вариантом ФТЗ.

Ниже приведен образец типового варианта итогового теста, предусмотренного рабочей программой дисциплины.

1. Суммарная оперативная память IBM Watson составляет порядка:
А) 100 гигабайт
Б) 5000 терабайт
В) 10 зетабайт
Г) 15 терабайт

2. Кто ввел термин Большие данные?
А) Клиффорд Линч
Б) Алан Тьюринг
В) Бьерн Страуструп
Г) Дональд Кнут

3. Какие данные занимают больше мировой памяти относительно остальных?
А) Structured Data
Б) Unstructured Data
В) Semi-Structured Data
Г) Quasi-Structured Data

4. Языком, на котором был разработан RabbitMQ, является:
Ответ: Erlang

5. Сопоставьте сущность и определение

А	Data Warehouse (Хранилище данных)	Хранилище, в котором хранится огромное количество необработанных данных в собственном формате, включая структурированные, полуструктурированные и неструктурированные данные. Структура данных и требования не определены, пока данные не потребуются. Обеспечивают большую гибкость, чем более жесткое хранилище данных
Б	Data Mart (Витрина данных)	База данных, оптимизированная для анализа реляционных данных, поступающих из транзакционных систем и бизнес-приложений. Структура данных и схема определены заранее для оптимизации для быстрых запросов SQL, результаты которых обычно используются для оперативной отчетности и анализа. Данные очищаются, обогащаются и трансформируются, чтобы они могли выступать в качестве «единого источника истины», которому пользователи могут доверять
В	Data Lake (Озеро данных)	Подмножество хранилища данных, обычно для определенной бизнес-функции (маркетинг, продажи, складские запасы, финансы и т.п.)

Ответ: А=Б, Б=В, В=А

6. Какая компания создала технологию MapReduce?
А) Google

Б) Yahoo
В) EMC
Г) Oracle

7. В каком году впервые был введен термин Большие данные?
Ответ: 2008

8. Данные текстовых файлов с определенными паттернами для их обработки (например, XML) являются:
Ответ: полуструктурированными

9. Что означает термин «Big Data» в информационных технологиях?
А) Комплексный набор методов для создания файлов большого объѐма
Б) **Комплексный набор методов обработки структурированных и неструктурированных данных колоссальных объемов**.
В) Файлы с большим количеством данных.
Г) Представление времени, дня, месяца и года в качестве значения количества миллисекунд, прошедших с начала нашей эры.

10. Какие из перечисленных пунктов являются достоинствами MapReduce?
А) Оптимальная производительность
Б) Эффективное применение в маленьких кластерах с небольшим объемом данных
В) Масштабируемость
Г) Отказоустойчивость

11. Сколько уровней имеет лямбда-архитектура?
Ответ: 3

12. Точка, в которой математическая функция стремится к бесконечности или имеет какие-либо иные нерегулярности поведения
Ответ: Open source

13. Функция округления до единиц вверх в языке «R»:
Ответ: Ceiling(x)

14. Что из этого не является типом визуализации?
А) График
Б) Текст
В) Круговая диаграмма
Г) Гистограмма

15. **Инструмент для кластерных вычислений**
Ответ: Spark

16. Что такое Oozie?
А) Распределенный координационный сервис
Б) Нереляционная распределенная база данных
В) Язык управления потоком данных и исполнительная среда для анализа больших объемов данных
Г) Сервис для записи и планировки заданий Hadoop

17. Чему примерно равен объем всей существующей на земле информации (в байтах)?
Ответ: 10^21

18. Какой язык программирования из перечисленных является наиболее важным для аналитика?

Ответ: R

3.5 Перечень теоретических вопросов к зачету

(для оценки знаний)

Раздел 1 «Технологии хранения и обработки больших данных»

1. Технологии Big Data: дать определение для Big Data, назначение
2. Big Data, история появления и основные принципы Big Data.
3. Достоинства и недостатки Big Data.
4. Технологии управления знаниями, визуализации знаний и интеллектуальные карты.
5. Данные, информация, знания, модели. Наука о данных.
6. Эволюционное развитие архитектур и данных.
7. Критерии больших данных. Источники больших данных.
8. Возможные этапы работы с большими данными.
9. Примеры и истории успеха работы с большими данными: торговля, финансы, кадры.
10. Обзор подходов к работе с данными: от языка простых запросов до методов анализа больших данных.

Раздел 2 «Информационное и программное обеспечение Big Data»

Интеллектуальный анализ данных: краткий обзор подходов.

12. Генетические алгоритмы.
13. Деревья принятия решений.
14. Визуализация больших данных.
15. Специфика хранения и обработки больших данных.
16. Парадигма MapReduce
17. Дать определение понятиям, назначение технологии, привести примеры программного обеспечения для визуализации знаний и построения интеллектуальных карт.
18. Файловая система HDFS.
19. Особенности хранилищ данных NoSQL.
20. Архитектура высоконагруженных систем.
21. Логические связки, таблицы истинности.
22.СУБД PostgreSQL.
23. Модель данных и информационная архитектура предприятия.
24. Процесс миграции на новую схему базы данных
25. ETL: автоматизация подготовки данных

3.6 Перечень типовых простых практических заданий к зачету

(для оценки умений)

1. Вывести наименование товара, наименование бренда и цену, для товаров, выпущенных в 2018 году. Упорядочить по убыванию цены.

2. Вывести имя, фамилию клиента, его телефон и дату заказа для всех клиентов. Упорядочить по фамилии клиента. Есть ли клиенты, не сделавшие ни одного заказа?

3. Вывести наименование товара, название склада и количество товара на данном складе. Оставить все товары, но для товаров, которых никогда не было ни на одном складе должны указываться значения null, -- как для названия склада, так и для количества. Если товар на складе был, но остаток нулевой, то должны выводиться название склада и ноль. Упорядочить по двум столбцам 1) название склада и 2) наименование товара.

4. Вывести наименование товара, название склада и количество товара на данном складе, в том числе вывести все товары и склады, для которых остаток нулевой. Упорядочить по двум столбцам 1) название склада и 2) наименование товара.

5. Вывести наименование товара, цену товара, номер заказа для заказа 55. Упорядочить по убыванию цены товара.

6. Вывести в формате "Имя Фамилия" для каждого сотрудника его начальника

(manager) и его номер телефона. Упорядочить по "имя фамилия" начальника. Для самого главного начальника (над которым нет начальников) в соответствующих столбцах вывести null.

7. Вывести наименование товара, его цену, год выпуска и остаток на складе для товаров, которых на складе осталось в 2 раза меньше, чем средний остаток по трем брендам, для которых остатки самые большие.

8. Вывести информацию о номере заказа, дате заказа, id сотрудника, который принял заказ, но только для заказов, которые были оформлены сотрудниками, для которых менеджером является Fabiola Jackson. Вывести только первые 10 заказов, отсортировав по возрастанию даты заказа.

9. Вывести наименование бренда, сумму остатков на всех складах по товарам данного бренда. Но только для трех топ-брендов, для которых сумма максимальна.

10. Вывести средний остаток по трем брендам, для которых сумма остатков на всех складах максимальна

3.7 Перечень типовых практических заданий к зачету

(для оценки навыков и (или) опыта деятельности)

1. Реализовать дополнительные операции над таблицами (объединение таблиц, нахождение общей части нескольких таблиц, исключение общих записей и др.).
2. Реализовать методы балансировки деревьев поиска и выполнить вычислительные эксперименты по оценке эффективности операций с таблицей при балансировке, сравнить работу с таблицей при наличии и в отсутствии методов балансировки.
3. Реализовать дополнительный набор операций для обработки результатов успеваемости.
4. Реализовать дополнительный набор графических форм вывода результатов обработки.
5. Реализовать расширенное представление таблиц (наличие нескольких ключей, поиск записей с использованием нескольких ключей, поиск записей по значениям из тела записей и т.д.).
6. Выполнить замену разработанных программ для реализации таблиц (без изменения программ обработки результатов экзаменационной сессии) на средства работы с таблицами в библиотеке STL. Оценить трудоемкость выполнения такой замены (общий размер изменяемого программного кода).
7. Дан набор данных заданной структуры и программа SAS Data step, производящая определенную обработку и вычисления с использованием данного набора. Перепишите эту программу на SAS DS2 с использованием параллельных нитей и созданием пользовательского пакета, чтобы результат обработки сохранился тем же, но код мог выполняться в параллельной среде.
8. Дан набор заданной структуры, постройте модель прогнозирования отклика с использованием процедуры impstat с алгоритмом random forest с заданным числом деревьев. Примените полученную модель к тестовому набору данных той же структуры, визуализируйте полученный график Lift. Постройте на том же наборе модель с использование высокопроизводительной версии метода GLM. Примените к тестовому набору. Сравните результаты GLM и Random Forest по AUC.
9. Дан текстовый корпус документов, лежащих в указанной директории. Создайте в SAS Tex Miner проект, который: выберет файлы с расширением pdf; осуществит парсинг набора с определением частей речи и сохранением в признаковом пространстве только существительных и глаголов; осуществит фильтрацию документов и признаков с использованием заданной схемой определения весов лексем (например, на основе tf-idf); выделит заданное количество ключевых тематик по методу SVD. В ответе укажите топ 5 ключевых слов во второй выявленной тематике. Какой документ имеет наибольший вес в это тематике?

4. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности

В таблице приведены описания процедур проведения контрольно-оценочных мероприятий и процедур оценивания результатов обучения с помощью оценочных средств в соответствии с рабочей программой дисциплины.

Наименование оценочного средства	Описания процедуры проведения контрольно-оценочного мероприятия и процедуры оценивания результатов обучения
Собеседование	Собеседование, предусмотренное рабочей программой дисциплины, проводится на практическом занятии. Преподаватель на практическом занятии, предшествующем занятию проведения контроля, доводит до обучающихся тему, вопросы для подготовки к собеседованию. Результаты собеседования преподаватель доводит до обучающихся сразу после завершения собеседования
Доклад	Защита докладов, предусмотренных рабочей программой дисциплины, проводится во время практических занятий. Преподаватель на практическом занятии, предшествующем занятию проведения контроля, доводит до обучающихся: тему докладов и требования, предъявляемые к их выполнению и защите
Лабораторная работа	Защита лабораторных работ проводится во время лабораторных занятий. Во время проведения защиты лабораторной работы пользоваться учебниками, справочниками, конспектами лекций, тетрадями не разрешено. Преподаватель на лабораторной работе, предшествующей занятию проведения защиты лабораторной работы, доводит до обучающихся: номер защищаемой лабораторной работы, время на защиту лабораторной работы. Преподаватель информирует обучающихся о результатах защиты лабораторной работы сразу после ее контрольно-оценочного мероприятия

Для организации и проведения промежуточной аттестации составляются типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы.

Перечень теоретических вопросов и типовые практические задания разного уровня сложности для проведения промежуточной аттестации обучающиеся получают в начале семестра через электронную информационно-образовательную среду ИрГУПС (личный кабинет обучающегося).

Описание процедур проведения промежуточной аттестации в форме зачета и оценивания результатов обучения

При проведении промежуточной аттестации в форме зачета преподаватель может воспользоваться результатами текущего контроля успеваемости в течение семестра. С целью использования результатов текущего контроля успеваемости, преподаватель подсчитывает среднюю оценку уровня сформированности компетенций обучающегося (сумма оценок, полученных обучающимся, делится на число оценок).

Шкала и критерии оценивания уровня сформированности компетенций в результате изучения дисциплины при проведении промежуточной аттестации в форме зачета по результатам текущего контроля (без дополнительного аттестационного испытания)

Средняя оценка уровня сформированности компетенций по результатам текущего контроля	Шкала оценивания
Оценка не менее 3,0 и нет ни одной неудовлетворительной оценки по текущему контролю	«зачтено»
Оценка менее 3,0 или получена хотя бы одна неудовлетворительная оценка по текущему контролю	«не зачтено»

Если оценка уровня сформированности компетенций обучающегося не соответствует критериям получения зачета без дополнительного аттестационного испытания, то

промежуточная аттестация проводится в форме собеседования по перечню теоретических вопросов и типовых практических задач или в форме компьютерного тестирования.

Промежуточная аттестация в форме зачета с проведением аттестационного испытания проходит на последнем занятии по дисциплине.

При проведении промежуточной аттестации в форме компьютерного тестирования вариант тестового задания формируется из фонда тестовых заданий по дисциплине случайным образом, но с условием: 50 % заданий должны быть заданиями открытого типа и 50 % заданий – закрытого типа.