

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ В.Ф. УТКИНА»**

Кафедра «Вычислительная и прикладная математика»

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ
«Основы работы с большими данными»**

Направление подготовки
09.03.04 «Программная инженерия»

Направленность (профиль) подготовки
«Программное обеспечение систем искусственного интеллекта»

Уровень подготовки – бакалавриат

Квалификация выпускника – бакалавр

Форма обучения – очная

Срок обучения – 4 года

Рязань 2023 г.

1. ОБЩИЕ ПОЛОЖЕНИЯ

Оценочные материалы – это совокупность учебно-методических материалов и процедур, предназначенных для оценки качества освоения обучающимися данной дисциплины как части основной образовательной программы.

Цель – оценить соответствие знаний, умений и уровня приобретенных компетенций, обучающихся целям и требованиям основной образовательной программы в ходе проведения текущего контроля и промежуточной аттестации.

Основная задача – обеспечить оценку уровня сформированности общекультурных и профессиональных компетенций и индикаторов их достижения, приобретаемых обучающимися в соответствии с этими требованиями.

Контроль знаний обучающихся проводится в форме текущего контроля и промежуточной аттестации.

Текущий контроль успеваемости и промежуточная аттестация проводятся с целью определения степени усвоения учебного материала, своевременного выявления и устранения недостатков в подготовке обучающихся, организации работы обучающихся в ходе учебных занятий и оказания им индивидуальной помощи.

К контролю текущей успеваемости относятся проверка знаний, умений и навыков обучающихся на практических занятиях по результатам выполнения и защиты обучающимися индивидуальных заданий, по результатам выполнения контрольных работ и тестов, по результатам проверки качества конспектов лекций и иных материалов.

В качестве оценочных средств на протяжении семестра используется устные и письменные ответы студентов на индивидуальные вопросы, письменное тестирование по теоретическим разделам курса, реферат. Дополнительным средством оценки знаний и умений студентов является отчет о выполнении практических заданий и его защита.

По итогам курса обучающиеся сдают экзамен. Форма проведения – устный ответ с письменным подкреплением по утвержденным билетам, сформулированным с учетом содержания дисциплины. В билет для экзамена включается два теоретических вопроса. В процессе подготовки к устному ответу студент должен составить в письменном виде план ответа.

2. Перечень компетенций с указанием этапов их формирования

При освоении дисциплины формируются следующие компетенции: ПК-8 (индикаторы ПК-8.1, ПК-8.2), ПК-9 (индикаторы ПК-9.1, ПК-9.2).

Указанные компетенции формируются в соответствии со следующими этапами:

- формирование и развитие теоретических знаний, предусмотренных указанными компетенциями (лекционные занятия, самостоятельная работа студентов);
- приобретение и развитие практических умений, предусмотренных компетенциями (практические занятия, самостоятельная работа студентов);
- закрепление теоретических знаний, умений и практических навыков, предусмотренных компетенциями, в ходе решения конкретных задач на занятиях, выполнения индивидуальных заданий на практических занятиях и их защиты, а также в процессе сдачи экзамена.

3 Показатели и критерии оценивания компетенций (результатов) на различных этапах их формирования, описание шкал оценивания

Сформированность каждой компетенции в рамках освоения данной дисциплины оценивается по трехуровневой шкале:

- пороговый уровень является обязательным для всех обучающихся по завершении освоения дисциплины;

- продвинутый уровень характеризуется превышением минимальных характеристик сформированности компетенций по завершении освоения дисциплины;
- эталонный уровень характеризуется максимально возможной выраженностью компетенций и является важным качественным ориентиром для самосовершенствования.

При достаточном качестве освоения более 80% приведенных знаний, умений и навыков преподаватель оценивает освоение данной компетенции в рамках настоящей дисциплины на эталонном уровне, при освоении более 60% приведенных знаний, умений и навыков – на продвинутом, при освоении более 40% приведенных знаний, умений и навыков – на пороговом уровне. При освоении менее 40% приведенных знаний, умений и навыков компетенция в рамках настоящей дисциплины считается неосвоенной.

Уровень сформированности каждой компетенции на различных этапах ее формирования в процессе освоения данной дисциплины оценивается в ходе текущего контроля успеваемости и представлено различными видами оценочных средств.

В качестве шкалы оценивания принимается 100-бальная система с выделением (градацией) оценок в соответствии с Положением о текущем контроле успеваемости и промежуточной аттестации студентов МГТУ им. Н.Э. Баумана:

Рейтинг	Оценка на экзамене
85 – 100	отлично
71 – 84	хорошо
60 – 70	удовлетворительно
0 – 59	неудовлетворительно

Показатели достижения планируемых результатов обучения и критерии их оценивания на разных уровнях формирования компетенций приведены в таблице 1.

Таблица 1. Показатели достижения индикаторов компетенции

1	2	3	4
Компетенция: код по ФГОС 3++, формули- ровка	Индикаторы	Этап	Наименование оце- ночного средства
ПК-8 (09.03.04/02 Программно- алгоритмиче- ское обеспече- ние систем ис- кусственного интеллекта) Способен осу- ществлять сбор и подготовку данных для си- стем искус- ственного ин- теллекта	ПК-8.1 Осуществляет поиск дан- ных в открытых источниках, спе- циализированных библиотеках и репозиториях ЗНАТЬ - виды представления данных, мето- ды поиска и парсинга данных - уровни представления данных (ODS, DDL, семантический слой, мо- дель данных) УМЕТЬ - отделять достоверные источники данных от сомнительных, осуществ- лять критический отбор данных, про- верять их на целостность и непроти- воречивость ПК-8.2 Выполняет подготовку и разметку структурированных и не- структурированных данных для машинного обучения ЗНАТЬ	1	Рубежные контроли. Лабораторные работы, практические занятия.

1	2	3	4
	- методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных УМЕТЬ - использовать методы и технологии массово параллельной обработки и анализа данных		
ПК-9 (09.03.04/02 Программно-алгоритмическое обеспечение систем искусственного интеллекта) Способен разрабатывать системы анализа больших данных	ПК-9.1 Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учетом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных ЗНАТЬ - общедоступные репозитории и специализированные библиотеки, содержащие наборы больших данных - принципы работы экосистемы Hadoop, фреймворка SPARK - устройство интерфейсов между реляционными SQL-хранилищами данных и нереляционными NoSQL-хранилищами данных - предметно-ориентированные языки УМЕТЬ - настраивать и оптимизировать конфигурацию программного и аппаратного обеспечения с целью интеграции больших данных - разрабатывать программное обеспечение для очистки и валидации наборов больших данных - выполнять потоковую обработку данных (data streaming, event processing) - использовать шины данных (Apache Kafka) - использовать языки запросов, в том числе нереляционных, для поддержки различных типов данных (например, XML, RDF, JSON, мультимедиа) и операций с большими данными (например, матричные операции) ПК-9.2 Разрабатывает программные компоненты обработки, удаленной, распределенной и объединенной аналитики, использования результатов анализа, описания и	1	Рубежные контроли. Лабораторные работы, практические занятия.

1	2	3	4
	управления качеством и достоверностью больших данных ЗНАТЬ - принципы и методы анализа больших данных, включая спецификации и стандартизацию метаданных - устройство и принципы работы систем обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, ETL) - архитектуру и принципы работы промышленных решений, созданных на основе ИИ - методы и технологии машинного обучения на больших данных УМЕТЬ - разрабатывать программное обеспечение для анализа больших данных - разрабатывать программные и технические средства визуализации больших данных и результатов их анализа - использовать системы обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, ETL процессы и инструменты) - использовать технологии науки о данных и больших данных в разработке для решения практических задач промышленности - описывать и управлять качеством и достоверностью больших данных		

Критерии оценки результатов обучения для различных видов контрольных мероприятий приведены в таблице:

Критерии оценивания результатов рубежного контроля №1

От 20 до 23 баллов - студент выполнил задание полностью правильно; логично, четко и ясно излагает ответы на поставленные вопросы; ответ носит самостоятельный характер.

От 18 до 19 баллов - при выполнении задания студент допустил отдельные неточности (несущественные ошибки); ответ отличается меньшей обстоятельностью, глубиной, обоснованностью и полнотой; однако допущенные ошибки исправляются самим студентом после дополнительных вопросов.

От 14 до 17 баллов - при выполнении задания студент допустил неточности и существенные ошибки; при аргументации ответа студент не применяет теоретические знания для объяснения допущенных ошибок, в целом ответ отличается низким уровнем самостоятельности.

От 0 до 13 баллов - студент не выполнил задание; в ответе на вопросы студент допускает ошибки в определении основных понятий; беспорядочно и неуверенно излагает материал.

Критерии оценивания результатов рубежного контроля №2

От 20 до 23 баллов - студент выполнил задание полностью правильно; логично, четко и

ясно излагает ответы на поставленные вопросы; ответ носит самостоятельный характер. <i>От 18 до 19 баллов</i> - при выполнении задания студент допустил отдельные неточности (несущественные ошибки); ответ отличается меньшей обстоятельностью, глубиной, обоснованностью и полнотой; однако допущенные ошибки исправляются самим студентом после дополнительных вопросов. <i>От 14 до 17 баллов</i> - при выполнении задания студент допустил неточности и существенные ошибки; при аргументации ответа студент не применяет теоретические знания для объяснения допущенных ошибок, в целом ответ отличается низким уровнем самостоятельности. <i>От 0 до 13 баллов</i> - студент не выполнил задание; в ответе на вопросы студент допускает ошибки в определении основных понятий; беспорядочно и неуверенно излагает материал.
Критерии оценивания результатов рубежного контроля №3 <i>От 20 до 24 баллов</i> - студент выполнил задание полностью правильно; логично, четко и ясно излагает ответы на поставленные вопросы; ответ носит самостоятельный характер. <i>От 18 до 19 баллов</i> - при выполнении задания студент допустил отдельные неточности (несущественные ошибки); ответ отличается меньшей обстоятельностью, глубиной, обоснованностью и полнотой; однако допущенные ошибки исправляются самим студентом после дополнительных вопросов. <i>От 14 до 17 баллов</i> - при выполнении задания студент допустил неточности и существенные ошибки; при аргументации ответа студент не применяет теоретические знания для объяснения допущенных ошибок, в целом ответ отличается низким уровнем самостоятельности. <i>От 0 до 13 баллов</i> - студент не выполнил задание; в ответе на вопросы студент допускает ошибки в определении основных понятий; беспорядочно и неуверенно излагает материал.
Критерии оценивания на лабораторных работах и практических занятиях <i>От 3 до 5 баллов</i> - студентом лабораторная работа и практическое занятие выполнены полностью или с незначительными ошибками. <i>От 0 до 2 баллов</i> - ошибки или отсутствие выполненной работы.
Критерии оценивания ответов на экзамене (для ликвидации академической задолженности, устранения академической разницы): <i>От 85 до 100 баллов:</i> студент глубоко и полно владеет содержанием учебного материала и понятийным аппаратом; умеет связывать теорию с практикой, иллюстрировать примерами, фактами, данными научных исследований; осуществляет межпредметные связи, предложения, выводы; логично, четко и ясно излагает ответы на поставленные вопросы; умеет обосновывать свои суждения и профессионально-личностную позицию по излагаемому вопросу; ответ носит самостоятельный характер. <i>От 71 до 84 баллов:</i> ответ студента соответствует указанным выше критериям, но в содержании имеют место отдельные неточности (несущественные ошибки) при изложении теоретического и практического материала; ответ отличается меньшей обстоятельностью, глубиной, обоснованностью и полнотой; однако допущенные ошибки исправляются самим аспирантом после дополнительных вопросов экзаменатора. <i>От 60 до 70 баллов:</i> студент обнаруживает знание и понимание основных положений учебного материала, но излагает его неполно, непоследовательно, допускает неточности и существенные ошибки в определении понятий, формулировке положений; не применяет теоретические знания для объяснения эмпирических фактов и явлений, не обосновывает свои суждения; имеет место нарушение логики изложения; в целом ответ отличается низким уровнем самостоятельности, не содержит собственной профессионально-личностной позиции. <i>От 0 до 59 баллов:</i> студент имеет разрозненные, бессистемные знания; не умеет выделять главное и второстепенное; в ответе допускаются ошибки в определении понятий,

формулировке теоретических положений, искажающие их смысл; аспирант не ориентируется в нормативно-концептуальных, программно-методических, исследовательских материалах, беспорядочно и неуверенно излагает материал; не умеет соединять теоретические положения с практикой; не умеет применять знания для объяснения эмпирических фактов, не устанавливает межпредметные связи.

Использование показателей и критериев оценивания компетенций на различных этапах их формирования совместно со шкалой балльно-рейтинговой системы позволяет формировать результаты обучения по модулям.

Оценка результатов обучения

Неделя	Номер и название модуля	Формы контроля	Баллы (мин/макс)
1 семестр			
6	1. Рабочее пространство	Лабораторные работы	6/10
		Рубежный контроль	14/23
		ИТОГО	20/33
10	2. Базовые элементы обработки Больших данных	Лабораторные работы	6/10
		Рубежный контроль	14/23
		ИТОГО	20/33
15	3. Продвинутое элементы обработки Больших данных	Лабораторные работы	6/10
		Рубежный контроль	14/24
		ИТОГО	20/34
		ИТОГО за семестр	60/100

4. Типовые контрольные задания или иные материалы

ФОС по дисциплине содержит следующие оценочные средства, позволяющие оценить знания, умения и уровень приобретенных компетенций при текущем контроле и промежуточной аттестации, разбитые по модулям дисциплины:

- перечень вопросов к экзамену.
- макеты билетов к экзамену.

Средства для оценки различных уровней формирования компетенций по категориям «знать», «уметь», «владеть» обеспечивают реализацию основных принципов контроля, таких, как объективность и независимость, практико-ориентированность, междисциплинарность.

С учетом этого, контрольные вопросы (задания, задачи,) входящие в ФОС, для различных категорий и уровней освоения компетенций имеют следующий вид:

Уровень ЗНАТЬ

Дескрипторы	Пример задания из оценочного средства
виды представления данных, методы поиска и парсинга данных	1. Дать определение данных и их представления в базах данных. Описать различия методов и особенности поиска в реляционных и нереляционных базах данных. 2. Определение Большие данные. Классификация Больших данных. Требования. Опишите подходы

	обработки Больших данных, их ключевые особенности.
уровни представления данных (ODS, DDL, семантический слой, модель данных)	1. Архитектура HBase. Модель данных, формат хранения. Масштабируемость. Принцип работы регионов.
методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных	1. Масштабирование в Реляционных БД, плюсы и минусы. Решение проблемы с помощью NoSQL. Свойства NoSQL.
общедоступные репозитории и специализированные библиотеки, содержащие наборы больших данных	1. Особенности и требования Hadoop. Опишите главные компоненты Hadoop и их свойства.
принципы работы экосистемы Hadoop, фреймворка SPARK	1. Особенности и требования Hadoop. Опишите главные компоненты Hadoop и их свойства. 2. Определение MapReduce, взаимосвязь с Hadoop. Фазы MapReduce. Опишите принцип работы, входные и выходные данные. Каким образом происходит передача данных между компонентами? Различия между MapReduce и Apache Spark.
устройство интерфейсов между реляционными SQL-хранилищами данных и нереляционными NoSQL-хранилищами данных	1. Методы хранения данных. Перечислите положительные и отрицательные стороны каждого из них. Реляционная модель БД, определение, требования, плюсы и минусы. 2. Масштабирование в Реляционных БД, плюсы и минусы. Решение проблемы с помощью NoSQL. Свойства NoSQL.
предметно-ориентированные языки	1. Методы хранения данных. Перечислите положительные и отрицательные стороны каждого из них. Реляционная модель БД, определение, требования, плюсы и минусы. 2. Масштабирование в Реляционных БД, плюсы и минусы. Решение проблемы с помощью NoSQL. Свойства NoSQL.
принципы и методы анализа больших данных, включая спецификации и стандартизацию метаданных	1. Определение Большие данные. Классификация Больших данных. Требования. Опишите подходы обработки Больших данных, их ключевые особенности. 2. Особенности и требования Hadoop. Опишите главные компоненты Hadoop и их свойства.
устройство и принципы работы систем обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, ETL)	1. Методы хранения данных. Перечислите положительные и отрицательные стороны каждого из них. Реляционная модель БД, определение, требования, плюсы и минусы. 2. Масштабирование в Реляционных БД, плюсы и минусы. Решение проблемы с помощью NoSQL.

	Свойства NoSQL.
архитектуру и принципы работы промышленных решений, созданных на основе ИИ	1. Архитектура HBase. Модель данных, формат хранения. Масштабируемость. Принцип работы регионов. 2. HBase Master и Data Storage, запись и чтение данных. Особенности удаления данных, механизмы объединения. Какие существуют запросы данных из HBase? Приведите примеры.
методы и технологии машинного обучения на больших данных	1. Архитектура HBase. Модель данных, формат хранения. Масштабируемость. Принцип работы регионов. 2. HBase Master и Data Storage, запись и чтение данных. Особенности удаления данных, механизмы объединения. Какие существуют запросы данных из HBase? Приведите примеры.

Уровень УМЕТЬ

Дескрипторы	Пример задания из оценочного средства
отделять достоверные источники данных от сомнительных, осуществлять критических отбор данных, проверять их на целостность и непротиворечивость	1. Масштабирование в Реляционных БД, плюсы и минусы. Решение проблемы с помощью NoSQL. Свойства NoSQL.
использовать методы и технологии массово параллельной обработки и анализа данных	1. Определение MapReduce, взаимосвязь с Hadoop. Фазы MapReduce. Опишите принцип работы, входные и выходные данные. Каким образом происходит передача данных между компонентами? Различия между MapReduce и Apache Spark.
настраивать и оптимизировать конфигурацию программного и аппаратного обеспечения с целью интеграции больших данных	1. Масштабирование в Реляционных БД, плюсы и минусы. Решение проблемы с помощью NoSQL. Свойства NoSQL.
описывать и управлять качеством и достоверностью больших данных	1. Масштабирование в Реляционных БД, плюсы и минусы. Решение проблемы с помощью NoSQL. Свойства NoSQL.
разрабатывать программное обеспечение для очистки и валидации наборов больших данных	1. Определение Pig, компоненты, возможности использования. Отличия от реляционных баз данных. Опишите режимы выполнения, а также механизмы исполнения. 2. Типы данных и операторы в Pig Latin. Опишите сценарий чтения данных, а также операторы загрузки, сохранения данных. Операторы Pig Latin, назначение.
выполнять потоковую	1. Определение Hive, назначение, архитектуру.

обработку данных (data streaming, event processing)	ра. Компоненты Hive. 2. Партиционирование, особенности работы Hive. Использование бакетов. Приведите пример
использовать шины данных (Apache Kafka)	1. Определение MapReduce, взаимосвязь с Hadoop. Фазы MapReduce. Опишите принцип работы, входные и выходные данные. Каким образом происходит передача данных между компонентами? Различия между MapReduce и Apache Spark. 2. Особенности использования принципа без Reduce-фазы. Функциональность фреймворка MapReduce в Hadoop. Hadoop Streaming и его использование. Приведите пример.
использовать языки запросов, в том числе нереляционных, для поддержки различных типов данных (например, XML, RDF, JSON, мультимедиа) и операций с большими данными (например, матричные операции)	1. Методы хранения данных. Перечислите положительные и отрицательные стороны каждого из них. Реляционная модель БД, определение, требования, плюсы и минусы. 2. Определение Hive, назначение, архитектура. Компоненты Hive.
разрабатывать программное обеспечение для анализа больших данных	1. Определение Большие данные. Классификация Больших данных. Требования. Опишите подходы обработки Больших данных, их ключевые особенности. 2. Особенности и требования Hadoop. Опишите главные компоненты Hadoop и их свойства.
разрабатывать программные и технические средства визуализации больших данных и результатов их анализа	1. Определение Hive, назначение, архитектура. Компоненты Hive. 2. Партиционирование, особенности работы Hive. Использование бакетов. Приведите пример
использовать системы обработки и анализа больших массивов данных (SQL, NoSQL, Hadoop, ETL процессы и инструменты)	1. Масштабирование в Реляционных БД, плюсы и минусы. Решение проблемы с помощью NoSQL. Свойства NoSQL.
использовать технологии науки о данных и больших данных в разработке для решения практических задач промышленности	1. Типы данных и операторы в Pig Latin. Опишите сценарий чтения данных, а также операторы загрузки, сохранения данных. Операторы Pig Latin, назначение.

Комплект билетов к рубежному контролю № 1

Билет № 1

1. Определение Большие данные. Классификация Больших данных. Требования. Опишите подходы обработки Больших данных, их ключевые особенности.
2. Особенности и требования Hadoop. Опишите главные компоненты Hadoop и их свойства.

Билет № 2

1. HDFS. Особенности и отличия от NFS. Архитектура. Опишите схематично работу блочного хранения данных.
2. Опишите сценарий работы с файлами. Блок в HDFS, особенности характеристики. Репликация блоков.

Билет № 3

1. Опишите принципы работы с файлами (чтение, запись) с использованием HDFS. Опишите особенности работы с памятью NameNode.
2. Виды доступа к HDFS. Опишите режимы работы, их сценарии и ключевые компоненты.

Комплект билетов к рубежному контролю № 2

Билет № 1

1. Методы хранения данных. Перечислите положительные и отрицательные стороны каждого из них. Реляционная модель БД, определение, требования, плюсы и минусы.
2. Масштабирование в Реляционных БД, плюсы и минусы. Решение проблемы с помощью NoSQL. Свойства NoSQL.

Билет № 2

1. Модели согласованности данных. Приведите примеры, характеризующие каждую из них. В чем отличие BASE и ACID подхода.
2. Особенности HBase. Применение HBase. В каких случаях следует использовать? Ответ поясните. Строение ячеек, группировка, версионность.

Билет № 3

1. Архитектура HBase. Модель данных, формат хранения. Масштабируемость. Принцип работы регионов.
2. HBase Master и Data Storage, запись и чтение данных. Особенности удаления данных, механизмы объединения. Какие существуют запросы данных из HBase? Приведите примеры.

Комплект билетов к рубежному контролю № 3

Билет № 1

1. Определение MapReduce, взаимосвязь с Hadoop. Фазы MapReduce. Опишите принцип работы, входные и выходные данные. Каким образом происходит передача данных между компонентами? Различия между MapReduce и Apache Spark.

2. Особенности использования принципа без Reduce-фазы. Функциональность фреймворка MapReduce в Hadoop. Hadoop Streaming и его использование. Приведите пример.

Билет № 2

1. Определение PIG, компоненты, возможности использования. Отличия от реляционных баз данных. Опишите режимы выполнения, а также механизмы исполнения.

2. Типы данных и операторы в Pig Latin. Опишите сценарий чтения данных, а также операторы загрузки, сохранения данных. Операторы Pig Latin, назначение.

Билет № 3

1. Определение HIVE, назначение, архитектура. Компоненты HIVE.

2. Партиционирование, особенности работы HIVE. Использование бакетов. Приведите пример

Перечень лабораторных работ

Лабораторная работа 1.1-1.3 Настройка рабочей среды. Hadoop.

Цель работы: Знакомство с рабочей средой, установка необходимых компонентов. Знакомство с Hadoop. Практическое применение полученных знаний.

Задание:

В лабораторной работе необходимо установить необходимые компоненты для Hadoop, настроить пользователя в операционной системе, настроить подключение через зашифрованный канал, установить и развернуть Apache Hadoop. Настроить Apache Hadoop в псевдо-распределенном режиме работы. Настроить HDFS через конфигурационные файлы.

Задание по варианту.

Практическое занятие 1.1-1.3 Знакомство с HDFS

Цель работы: Знакомство с HDFS, особенностями работы. Практическое применение полученных знаний.

Задание:

В практическом занятии необходимо изучить команды HDFS. Необходимо загрузить данные в HDFS, получить данные от HDFS, разархивировать и прочитать файлы, а так же проверить файлы HDFS на отсутствие проблем. Так же необходимо научиться переходить в режим администратора, безопасный режим.

Задание по варианту.

Практическое занятие 2.1-2.2 Знакомство с HBase

Цель работы: Знакомство с HBase, установка необходимых компонентов. Практическое применение полученных знаний.

Задание:

В практическом занятии необходимо установить необходимые компоненты для HBase, установить и настроить HBase с получением доступа через подключение по локальной сети.

Задание по варианту.

Лабораторная работа 2.1-2.2 HBase. Shell

Цель работы: Знакомство с HBase, особенностями обработки данных. Практическое применение полученных знаний.

Задание:

В лабораторной работе необходимо изучить команды HBase. Необходимо получить информацию о HBase с помощью общих команд. Необходимо изучить работу команд определения данных, а также их манипулирования.

Задание по варианту.

Практическое занятие 3.1-3.2 MapReduce.

Цель работы: Знакомство с MapReduce. Практическое применение полученных знаний.

Задание:

Реализовать пример передачи данных в парадигме MapReduce. Запустить в локальном режиме. Запустить в Hadoop.

Задание по варианту.

Практическое занятие 3.3 Знакомство с Pig

Цель работы: Знакомство с Pig, установка необходимых компонентов. Практическое применение полученных знаний.

Задание:

Практическое занятие делится на две части.

Установить необходимые компоненты для Pig, установить и настроить Pig в системе в начале в локальном режиме, потом в Hadoop. Изучить возможности системы сохранять информацию, манипулировать информацией через команды пользователя.

Задание по варианту.

Лабораторная работа 3.1-3.3 Apache Hive

Цель работы: Знакомство с Apache Hive, установка необходимых компонентов. Практическое применение полученных знаний.

Задание:

В лабораторной работе необходимо установить необходимые компоненты для Apache Hive, установить и настроить Apache Hive в системе с Hadoop. Установить и запустить базу данных Derby. Изучить возможности системы сохранять информацию, манипулировать информацией через команды пользователя.

Задание по варианту.

Макет экзаменационного билета

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №1

по дисциплине «Основы работы с большими данными»

1. Алгоритм обратного распространения ошибки. Применение.
2. Принцип работы сети LSTM. Достоинства и недостатки этой сети.

Билет рассмотрен и утвержден на заседании кафедры «__» _____ 23 г.

Перечень вопросов к экзамену (для ликвидации академической задолженности или перезачета или для повышения балльной оценки)

1. Определение Большие данные. Классификация Больших данных. Требования. Опишите подходы обработки Больших данных, их ключевые особенности.
2. Особенности и требования Hadoop. Опишите главные компоненты Hadoop и их свойства.
3. HDFS. Особенности и отличия от NFS. Архитектура. Опишите схематично работу блочного хранения данных.
4. Опишите сценарий работы с файлами. Блок в HDFS, особенности характеристики. Репликация блоков.
5. Опишите принципы работы с файлами (чтение, запись) с использованием HDFS. Опишите особенности работы с памятью NameNode.
6. Виды доступа к HDFS. Опишите режимы работы, их сценарии и ключевые компоненты.
7. Методы хранения данных. Перечислите положительные и отрицательные стороны каждого из них. Реляционная модель БД, определение, требования, плюсы и минусы.
8. Масштабирование в Реляционных БД, плюсы и минусы. Решение проблемы с помощью NoSQL. Свойства NoSQL.
9. Модели согласованности данных. Приведите примеры, характеризующие каждую из них. В чем отличие BASE и ACID подхода.
10. Особенности HBase. Применение HBase. В каких случаях следует использовать? Ответ поясните. Строение ячеек, группировка, версионность.
11. Архитектура HBase. Модель данных, формат хранения. Масштабируемость. Принцип работы регионов.
12. HBase Master и Data Storage, запись и чтение данных. Особенности удаления данных, механизмы объединения. Какие существуют запросы данных из HBase?
13. Определение MapReduce, взаимосвязь с Hadoop. Фазы MapReduce. Опишите принцип работы, входные и выходные данные. Каким образом происходит передача данных между компонентами? Различия между MapReduce и Apache Spark.
14. Особенности использования принципа без Reduce-фазы. Функциональность фреймворка MapReduce в Hadoop. Hadoop Streaming и его использование. Приведите пример.
15. Определение Pig, компоненты, возможности использования. Отличия от реляционных баз данных. Опишите режимы выполнения, а также механизмы исполнения.
16. Типы данных и операторы в Pig Latin. Опишите сценарий чтения данных, а также операторы загрузки, сохранения данных. Операторы Pig Latin, назначение.
17. Определение Hive, назначение, архитектура. Компоненты Hive.
18. Партиционирование, особенности работы Hive. Использование бакетов. Приведите пример

Процедуры оценивания знаний, умений, навыков, формы и организация текущего контроля и промежуточной аттестации обучающихся

Текущий контроль и промежуточная аттестация студентов в университете ведется в соответствии с Положением о текущем контроле успеваемости и промежуточной аттестации студентов МГТУ им. Н.Э. Баумана.

Текущий контроль успеваемости

Дисциплина делится на 3 модуля. Каждый модуль учебной дисциплины включает в себя изучение законченного раздела, части дисциплины.

Основными видами контроля знаний, умений и навыков в течение каждого модуля учебной дисциплины являются рубежные контроли и лабораторные работы.

Текущий контроль по модулю учебной дисциплины осуществляется по графику учебного процесса. Сроки контрольных мероприятий (КМ) и сроки подведения итогов по модулям учебной дисциплины отображаются в рабочих учебных планах на семестр (отрезках). Студент должен выполнить все контрольные мероприятия, предусмотренные в модуле учебной дисциплины к указанному сроку, после чего преподаватель проставляет балльные оценки, набранные студентом по результатам текущего контроля модуля учебной дисциплины в ЭУ.

Контрольное мероприятие считается выполненным, если за него студент получил оценку в баллах, не ниже минимальной оценки, установленной программой дисциплины по данному мероприятию.

Студенты, не сдавшие контрольное мероприятие в установленный срок, продолжают работать над ним в соответствии с порядком, принятым кафедрой.

Промежуточная аттестация

Формой промежуточной аттестации по дисциплине является зачет.

Зачет

В рамках рейтинговой системы контроля успеваемости студентов, зачет по дисциплине формируется набором в течение семестра, предусмотренной в программе дисциплины, суммы баллов, при выполнении им всех контрольных мероприятий.

Дифференцированный зачет

Зачеты по курсовому проекту проходят в форме дифференцированного зачета с проставлением в зачетной ведомости оценок «отлично», «хорошо», «удовлетворительно» и «неудовлетворительно». Зачет по курсовому проекту проставляется по результатам защиты студентами курсового проекта перед комиссией, назначенной кафедрой.

Оценивание дисциплины ведется в соответствии с Положением о текущем контроле успеваемости и промежуточной аттестации студентов МГТУ им. Н.Э. Баумана.

Методика оценки по рейтингу

Студент, выполнивший все предусмотренные учебным планом задания и сдавший все контрольные мероприятия, получает итоговую оценку по дисциплине за семестр в соответствии со шкалой:

Рейтинг	Оценка на экзамене
85 – 100	отлично
71 – 84	хорошо
60 – 70	удовлетворительно
0 – 59	неудовлетворительно

Рейтинг студента по дисциплине за семестр определяется как сумма баллов, полученных им за все модули учебной дисциплины, и баллов за промежуточную аттестацию. Максимальное количество баллов за дисциплину в семестре устанавливается равным 100.