

Компонент ОПОП 09.04.01 Информатика и вычислительная техника
Направленность (профиль) Компьютерный анализ и интерпретация данных.
Data Science.

Б1.О.13
шифр дисциплины

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Дисциплины
(модуля)

«Технология больших данных»

Разработчик:
Золотов О.В.
ФИО

ДОЦЕНТ
должность

канд. физ.-мат. наук
ученая степень,
звание

Утверждено на заседании кафедры
Информационных технологий
наименование кафедры

протокол № 6 от 01.02.2024 г.

Заведующий кафедрой ИТ



подпись

_____ Ляш О.И.
ФИО

1. Критерии и средства оценивания компетенций и индикаторов их достижения, формируемых дисциплиной (модулем)

Код и наименование компетенции	Код и наименование индикатора(ов) достижения компетенции	Результаты обучения по дисциплине (модулю)			Оценочные средства текущего контроля	Оценочные средства промежуточной аттестации
		<i>Знать</i>	<i>Уметь</i>	<i>Владеть</i>		
<p>ОПК-6. Способен разрабатывать компоненты программно-аппаратных комплексов обработки информации и автоматизированного проектирования</p>	<p>ИД-1_{опк-6} Понимает принципы работы программно-аппаратных комплексов обработки информации и автоматизированного проектирования</p> <p>ИД-2_{опк-6} Анализирует существующие подходы к разработке компонент программно-аппаратных комплексов обработки информации и автоматизированного проектирования</p> <p>ИД-3_{опк-6} Использует инструментальные средства разработки компонент программно-аппаратных комплексов</p>	<p>- принципы работы программно-аппаратных комплексов обработки информации и автоматизированного проектирования, ориентированных на технологии работы с большими данными;</p> <p>- специфику зарубежных комплексов обработки информации и автоматизированного проектирования, применяющих технологии работы с большими данными.</p>	<p>- анализировать существующие подходы к разработке компонент программно-аппаратных комплексов обработки информации и автоматизированного проектирования, ориентированных на технологии работы с большими данными;</p> <p>- анализировать цифровые потребности отечественных предприятий в области технологии работы с большими данными.</p>	<p>- инструментальными средствами разработки компонент программно-аппаратных комплексов, ориентированных на применение технологий работы с большими данными;</p> <p>- навыком адаптации комплексов обработки информации в соответствии с нуждами в сфере технологии работы с большими данными.</p>	<p>- комплект заданий для выполнения лабораторных работ;</p> <p>- типовые задания по вариантам для выполнения расчетно-графической работы.</p>	<p>Экзаменационные билеты</p> <p>Результаты текущего контроля</p>

<p>ОПК-7. Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий</p>	<p>ИД-1_{опк-7} Понимает специфику зарубежных комплексов обработки информации и автоматизированного проектирования</p> <p>ИД-2_{опк-7} Анализирует цифровые потребности отечественных предприятий</p> <p>ИД-3_{опк-7} Адаптирует комплексы обработки информации и автоматизированного проектирования к нуждам конкретного предприятия</p>					
--	--	--	--	--	--	--

2. Оценка уровня сформированности компетенций (индикаторов их достижения)

Показатели оценивания компетенций (индикаторов их достижения)	Шкала и критерии оценки уровня сформированности компетенций (индикаторов их достижения)			
	Ниже порогового («неудовлетворительно»)	Пороговый («удовлетворительно»)	Продвинутый («хорошо»)	Высокий («отлично»)
Полнота знаний	Уровень знаний ниже минимальных требований. Имели место грубые ошибки.	Минимально допустимый уровень знаний. Допущены не грубые ошибки.	Уровень знаний в объёме, соответствующем программе подготовки. Допущены некоторые погрешности.	Уровень знаний в объёме, соответствующем программе подготовки.
Наличие умений	При выполнении стандартных заданий не продемонстрированы основные умения. Имели место грубые ошибки.	Продемонстрированы основные умения. Выполнены типовые задания с не грубыми ошибками. Выполнены все задания, но не в полном объёме (отсутствуют пояснения, неполные выводы)	Продемонстрированы все основные умения. Выполнены все основные задания с некоторыми погрешностями. Выполнены все задания в полном объёме, но некоторые с недочетами.	Продемонстрированы все основные умения. Выполнены все основные и дополнительные задания без ошибок и погрешностей. Задания выполнены в полном объёме без недочетов.
Наличие навыков (владение опытом)	При выполнении стандартных заданий не продемонстрированы базовые навыки. Имели место грубые ошибки.	Имеется минимальный набор навыков для выполнения стандартных заданий с некоторыми недочетами.	Продемонстрированы базовые навыки при выполнении стандартных заданий с некоторыми недочетами.	Продемонстрированы все основные умения. Выполнены все основные и дополнительные задания без ошибок и погрешностей. Продемонстрирован творческий подход к решению нестандартных задач.
Характеристика сформированности компетенции	Компетенции фактически не сформированы. Имеющихся знаний, умений, навыков недостаточно для решения практических (профессиональных) задач. ИЛИ Зачетное количество баллов не набрано согласно установленному диапазону	Сформированность компетенций соответствует минимальным требованиям. Имеющихся знаний, умений, навыков в целом достаточно для решения практических (профессиональных) задач. ИЛИ Набрано зачетное количество баллов согласно установленному диапазону	Сформированность компетенций в целом соответствует требованиям. Имеющихся знаний, умений, навыков достаточно для решения стандартных профессиональных задач. ИЛИ Набрано зачетное количество баллов согласно установленному диапазону	Сформированность компетенций полностью соответствует требованиям. Имеющихся знаний, умений, навыков в полной мере достаточно для решения сложных, в том числе нестандартных, профессиональных задач. ИЛИ Набрано зачетное количество баллов согласно установленному диапазону

3. Критерии и шкала оценивания заданий текущего контроля

3.1 Критерии и шкала оценивания лабораторных работ

Перечень лабораторных работ, описание порядка выполнения и защиты работы, требования к результатам работы, структуре и содержанию отчета и т.п. представлены в методических материалах по освоению дисциплины (модуля) и в электронном курсе в ЭИОС МАУ.

Оценка/баллы	Критерии оценивания
<i>Отлично / 3</i>	Задание выполнено полностью и правильно. Отчет по лабораторной/практической работе подготовлен качественно в соответствии с требованиями. Полнота ответов на вопросы преподавателя при защите работы.
<i>Хорошо / 2.5</i>	Задание выполнено полностью, но нет достаточного обоснования или при верном решении допущена незначительная ошибка, не влияющая на правильную последовательность рассуждений. Все требования, предъявляемые к работе, выполнены.
<i>Удовлетворительно / 2</i>	Задания выполнены частично с ошибками. Демонстрирует средний уровень выполнения задания на лабораторную/практическую работу. Большинство требований, предъявляемых к заданию, выполнены.
<i>Неудовлетворительно / 0</i>	Задание выполнено со значительным количеством ошибок на низком уровне. Многие требования, предъявляемые к заданию, не выполнены. ИЛИ Задание не выполнено.

3.2 Критерии и шкала оценивания расчетно-графической работы

Перечень заданий, рекомендации по выполнению представлены в методических материалах по освоению дисциплины (модуля) и в электронном курсе в ЭИОС МАУ.

В ФОС включен типовой вариант задания на выполнение расчетно-графической работы:

В рамках выполнения РГР необходимо выполнить следующее задание.

1. Выполнить установку и настройку среды исполнения. Версия операционной системы, jvm и jdk, а также использование системы виртуализации / контейнеризации оставляются на усмотрение исполнителя с учетом совместимости с используемыми продуктами.
2. Установить в среду исполнения Apache Hadoop, выполнить настройку системных переменных. Продемонстрировать работу Apache Hadoop в standalone режиме на примере запуска любого примера из поставляемых с дистрибутивом Apache Hadoop.
3. Установить в среду исполнения Apache Pig, Apache Hive, Apache HBase, Apache Phoenix, Python-библиотеку mrjob, необходимые зависимости для Hadoop PIPES. Выполнить настройки, необходимые для выполнения следующего пункта.
4. Реализовать приложение подсчета частоты появления слов в тексте
 - на ЯВУ Java и запустить данное приложение как задачу Map-Reduce с помощью Apache Hadoop, работающего в псевдораспределенном режиме;
 - на ЯВУ Python3 с использованием API Hadoop STREAMING;

- на ЯВУ С++ с использованием API Hadoop PIPES;
- на ЯВУ Python3 с использованием библиотеки mrjob;
- на ЯВУ Apache Pig;
- с использованием Apache Hive;
- с использованием Apache HBase и Apache Phoenix.

Оценка/баллы	Критерии оценивания
Отлично / 34	Работа выполнена полностью, без ошибок (возможна одна неточность, описка, не являющаяся следствием непонимания материала).
Хорошо / 33	Работа выполнена полностью, но обоснования шагов решения недостаточны, допущена одна негрубая ошибка или два-три недочета, не влияющих на правильную последовательность рассуждений.
Удовлетворительно / 31	В работе допущено более одной грубой ошибки или более двух-трех недочетов, но обучающийся владеет обязательными умениями по проверяемой теме.
Неудовлетворительно / 0	В работе есть грубые ошибки и недочеты ИЛИ Контрольная работа не выполнена.

Критерии и шкала оценивания посещаемости занятий

Посещение занятий обучающимися определяется в процентном соотношении

Баллы	Критерии оценки
10	посещаемость 75 - 100 %
5	посещаемость 50 - 74 %
0	посещаемость менее 50 %

4. Критерии и шкала оценивания результатов обучения по дисциплине (модулю) при проведении промежуточной аттестации

Критерии и шкала оценивания результатов освоения дисциплины (модуля) с экзаменом

Для дисциплин (модулей), заканчивающихся экзаменом, результат промежуточной аттестации складывается из баллов, набранных в ходе текущего контроля и при проведении экзамена:

В ФОС включен список типовых вопросов и заданий к экзамену и типовой вариант экзаменационного билета:

Вопросы к экзамену:

1. Понятие больших данных (BigData).
2. Этические и юридические сложности работ в области больших данных
3. Существующие программные решения в области больших данных.
4. Основные компоненты Apache Hadoop. Apache Hadoop: Common
5. Основные компоненты Apache Hadoop. Apache Hadoop: HDFS
6. Основные компоненты Apache Hadoop. Apache Hadoop: YARN

7. Основные компоненты Apache Hadoop. Apache Hadoop: MapReduce
8. Интерфейс Hadoop Streaming. Назначение, особенности использования
9. Интерфейс Hadoop PIPES. Назначение, особенности использования
10. Организация Map-Reduce вычислений на базе Python-библиотеки MRJob
11. Основные возможности Apache Pig
12. Основные возможности Apache Hive и HiveQL
13. Основные возможности NoSQL СУБД Apache HBase
14. Основные возможности NoSQL СУБД Apache Phoenix
15. Основные возможности и особенности РСУБД ClickHouse

Типовые задания (задачи):

1. Опишите основные шаги установки и особенности настройки Apache Hadoop.
2. Опишите основные шаги установки и особенности настройки Apache Pig.
3. Опишите основные шаги установки и особенности настройки Apache Hive.
4. Опишите основные шаги установки и особенности настройки Apache HBase.
5. Опишите основные шаги установки и особенности настройки Apache Phoenix.
6. Опишите основные шаги установки и особенности настройки Python-библиотеки mrjob
7. Напишите программу на любом ЯВУ, демонстрирующую типовой сценарий использования Hadoop Streaming

Типовой вариант экзаменационного билета:

Федеральное государственное автономное образовательное учреждение
высшего образования
«МУРМАНСКИЙ АРКТИЧЕСКИЙ УНИВЕРСИТЕТ»
кафедра Информационных технологий
направление обучение 09.04.01 «Информатика и вычислительная техника»

Технология больших данных

БИЛЕТ № 1

Вопрос 1

Основные возможности распределенной файловой системы HDFS. Архитектура. Особенности, достоинства, недостатки.

Вопрос 2

Синтаксис HiveQL. Построение запросов.

Вопрос 3.

Задача.

Оценка	Критерии оценки ответа на экзамене
<i>Отлично</i>	Обучающийся глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, не затрудняется с ответом при видоизменении вопроса. Владеет специальной терминологией, демонстрирует общую эрудицию в предметной области, использует при ответе ссылки на материал специализированных источников, в том числе на Интернет-ресурсы.
<i>Хорошо</i>	Обучающийся твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, владеет специальной терминологией на достаточном уровне; могут возникнуть затруднения при ответе на уточняющие вопросы по рассматриваемой теме; в целом демонстрирует общую эрудицию в предметной области.

Удовлетворительно	Обучающийся имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, плохо владеет специальной терминологией, допускает существенные ошибки при ответе, недостаточно ориентируется в источниках специализированных знаний.
Неудовлетворительно	Обучающийся не знает значительной части программного материала, допускает существенные ошибки, нарушения логической последовательности в изложении программного материала, не владеет специальной терминологией, не ориентируется в источниках специализированных знаний. Нет ответа на поставленный вопрос.

Оценка, полученная на экзамене, переводится в баллы («5» - 20 баллов, «4» - 15 баллов, «3» - 10 баллов) и суммируется с баллами, набранными в ходе текущего контроля.

Итоговая оценка по дисциплине (модулю)	Суммарные баллы по дисциплине (модулю), в том числе	Критерии оценивания
Отлично	91 - 100	Выполнены все контрольные точки текущего контроля на высоком уровне. Экзамен сдан
Хорошо	81-90	Выполнены все контрольные точки текущего контроля. Экзамен сдан
Удовлетворительно	70- 80	Контрольные точки выполнены в неполном объеме. Экзамен сдан
Неудовлетворительно	69 и менее	Контрольные точки не выполнены или не сдан экзамен

5. Задания диагностической работы для оценки результатов обучения по дисциплине (модулю) в рамках внутренней и внешней независимой оценки качества образования

ФОС содержит задания для оценивания знаний, умений и навыков, демонстрирующих уровень сформированности компетенций и индикаторов их достижения в процессе освоения дисциплины (модуля).

Комплект заданий разработан таким образом, чтобы осуществить процедуру оценки каждой компетенции, формируемых дисциплиной (модулем), у обучающегося в письменной форме.

Содержание комплекта заданий включает тестовые задания.

Комплект заданий диагностической работы

ОПК-6. Способен разрабатывать компоненты программно-аппаратных комплексов обработки информации и автоматизированного проектирования	
1	<p><i>Считается, что термин BigData был введен</i></p> <p>А. Сергеем Брином (Sergey Brin) и Ларри Пейджем (Lawrence Edward «Larry» Page) в процессе разработки таких продуктов Google как BigTable, Hadoop и т.п.</p> <p>Б. Джеффом Безосом (Jeffrey Preston «Jeff» Bezos) в процессе разработки облачных сервисов Amazon</p> <p>В. Биллом Гейтсом (William Henry Gates III) в процессе разработки облачных сервисов Microsoft</p> <p>Г. Клиффордом Линчем (Clifford Lynch) в процессе обсуждения проблемы доступности преимущественно научных данных в долгосрочной перспективе</p> <p>Ответ: Г. Клиффордом Линчем (Clifford Lynch) в процессе обсуждения проблемы доступности преимущественно научных данных в долгосрочной перспективе</p>
2	<p><i>Основная роль в обеспечении обработки (хранения, преобразования, предоставления пользователю) в оригинальном понятии BigData отводится</i></p> <p>А. сохраняющим институтам / preservation institutions</p> <p>Б. облачной инфраструктуре / cloud infrastructure</p> <p>В. сетевой инфраструктуре / network infrastructure</p> <p>Г. квантовым компьютерам / quantum computers</p> <p>Ответ: А. сохраняющим институтам / preservation institutions</p>
3	<p><i>Используемые в современных работах и выступлениях определения BigData как правило</i></p> <p>А. соответствуют оригинальному понятию BigData</p> <p>Б. частично соответствуют оригинальному понятию BigData</p> <p>В. не соответствуют оригинальному понятию BigData</p> <p>Г. нет верного ответа</p> <p>Ответ: В. не соответствуют оригинальному понятию BigData</p>
4	<p><i>Проблемы организации обработки больших объемов данных</i></p> <p>А. не могут быть формализованы и математически описаны</p> <p>Б. являются уникальными для BigData</p> <p>В. аналогичны проблемам, анализировавшимся в 1970х-1980х гг. Эдгаром Коддом (Edgar F. Codd) при работе в IBM над реляционной СУБД System R (нынешний</p>

	<p>потомок – РСУБД IBM DB2) и языком запросов SEQUEL (в настоящее время известен как SQL) Г. нет верного ответа</p> <p>Ответ: В. аналогичны проблемам, анализировавшимся в 1970х-1980х гг. Эдгаром Коддом (Edgar F. Codd) при работе в IBM над реляционной СУБД System R (нынешний потомок – IBM DB2) и языком запросов SEQUEL (в настоящее время известен как SQL)</p>
5	<p><i>В современных работах и публикациях используемое определение BigData часто можно охарактеризовать как</i></p> <p>А. конкретное, точное, измеримое, непротиворечивое Б. отражающее аспекты конкретного решения или продукта В. общепринятое, т.е. устоявшееся, применяемое большинством сторон, каноническое Г. нет верного ответа</p> <p>Ответ: Б. отражающее аспекты конкретного решения или продукта</p>
6	<p><i>Типовые задачи при подготовке данных и обучении моделей включают</i></p> <p>А. учет (заполнение) пропусков Б. кодирование нечисловых признаков В. приведение данных к единому масштабу и стандартизация Г. всё перечисленное</p> <p>Ответ: Г. всё перечисленное</p>
7	<p><i>Среди всех существующих (мировых) данных наибольшую долю составляют</i></p> <p>А. структурированные данные Б. неструктурированные данные В. частично структурированные (полуструктурированные) данные Г. квазиструктурированные данные</p> <p>Ответ: Б. неструктурированные данные</p>
8	<p><i>Технологии BigData – это</i></p> <p>А. комплексный набор методов представления фактов, понятий или инструкций в форме, приемлемой для интерпретации или обработки Б. комплексный набор методов обработки структурированных данных колоссальных объемов. В. комплексный набор методов обработки структурированных и неструктурированных данных колоссальных объемов.</p>

	<p>Г. нет верного ответа</p> <p>Ответ: В. комплексный набор методов обработки структурированных и неструктурированных данных колоссальных объемов.</p>
9	<p><i>При работе с большими данными (BigData) в зависимости от их природы локальное законодательство может предъявлять специальные требования</i></p> <p>А. к физическому месту хранения данных Б. к трансграничной передаче данных В. к регистрации юридического лица, осуществляющего обработку – получение, хранение, преобразование, передачу – таких данных Г. всё перечисленное</p> <p>Ответ: Г. всё перечисленное</p>
10	<p><i>Представим, что Вы – международный стартап, зарегистрированный на территории РФ, который разрабатывает продукт, конкурирующий с профессиональной социальной сетью ученых ResearchGate. Зарегистрированный пользователь вашего продукта является гражданином(кой) страны Евросоюза (например, Румынии) и направил(а) вам официальный запрос на предоставление ему ВСЕХ данных, которые у Вас имеются в отношении него (нее). Должны ли Вы предоставить ему (ей) эту информацию?</i></p> <p>А. нет, не должны (мы находимся в юрисдикции РФ) Б. должны, но только те, что предусмотрены законодательством РФ В. должны, но только те, которые обязались предоставить ему по пользовательскому соглашению / публичной оферте Г. должны (согласно требований GDPR)</p> <p>Ответ: Г. должны (согласно требований GDPR)</p>
<p>ОПК-7. Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий</p>	
1	<p><i>Что из нижеперечисленного является компонентой Apache Hadoop?</i></p> <p>А. HDFS Б. YARN В. MapReduce Г. всё перечисленное</p> <p>Ответ: Г. всё перечисленное</p>

2	<p><i>HDFS по сравнению с другими файловыми системами обладает</i></p> <p>А. большим размер блока Б. примерно таким же размером блока В. меньшим размером блока Г. нет правильного ответа</p> <p>Ответ: А. большим размер блока</p>
3	<p><i>В HDFS зеркалирование и репликация осуществляется на уровне</i></p> <p>А. узла данных Б. клиента В. кластера Г. нет правильного ответа</p> <p>Ответ: В. кластера</p>
4	<p><i>В HDFS зеркалирование и репликация осуществляется на уровне</i></p> <p>А. узла данных Б. клиента В. кластера Г. нет правильного ответа</p> <p>Ответ: В. кластера</p>
5	<p><i>В HDFS проверка целостности данных является «обязанностью»</i></p> <p>А. узла данных Б. клиента В. кластера Г. нет правильного ответа</p> <p>Ответ: Б. клиента</p>
6	<p><i>HDFS оптимизирована для</i></p> <p>А. потокового считывания (передачи) данных Б. произвольного доступа (чтение, запись) к большому количеству файлов В. эффективного осуществления конкурентной записи с произвольным доступом Г. нет правильного ответа</p>

	<p>Ответ: А. потокового считывания (передачи) данных</p>
7	<p><i>В классическом HDFS для файлов НЕ поддерживается операция</i></p> <p>А. записи Б. чтения В. модификации (изменения) Г. репликации</p> <p>Ответ: В. модификации (изменения)</p>
8	<p><i>Выберите НЕ верное утверждение</i></p> <p>А. HBase – столбцовая (колоночная) СУБД Б. создатели HBase «вдохновлялись» Google BigTable В. HBase – реляционная СУБД Г. HBase – NoSQL СУБД</p> <p>Ответ: В. HBase – реляционная СУБД</p>
9	<p><i>Выберите верное утверждение</i></p> <p>А. Apache Phoenix изначально проектировалась для работы с HBase, поэтому хорошо учитывает архитектуру HBase и не вносит больших накладных расходов Б. Apache Hive изначально проектировалась для работы с HBase, поэтому хорошо учитывает архитектуру HBase и не вносит больших накладных расходов В. Apache Pig изначально проектировалась с учетом особенностей HBase, поэтому хорошо учитывает архитектуру HBase и не вносит больших накладных расходов Г. нет правильного ответа</p> <p>Ответ: А. Apache Phoenix изначально проектировалась для работы с HBase, поэтому хорошо учитывает архитектуру HBase и не вносит больших накладных расходов</p>
10	<p><i>Выберите НЕ верное утверждение</i></p> <p><i>СУБД ClickHouse</i></p> <p>А. рассчитана на OLAP-сценарий использования Б. рассчитана на OLTP-сценарий использования В. хранит данные по столбцам Г. поддерживает запросы на собственном подмножестве (диалекте) SQL</p>

	Ответ: Б. рассчитана на OLTP-сценарий использования
--	--

При оценивании тестовых заданий диагностической карты используются следующие критерии и шкала оценивания тестирования

Оценка/баллы	Критерии оценки
<i>Отлично</i>	90-100 % правильных ответов
<i>Хорошо</i>	70-89 % правильных ответов
<i>Удовлетворительно</i>	50-69 % правильных ответов
<i>Неудовлетворительно</i>	49% и меньше правильных ответов

Полученные баллы (за каждую компетенцию) равны целому (округленному до целого) проценту правильных ответов. Итоговый балл за выполнение диагностической работы равен сумме баллов, полученных за каждую компетенцию, умноженной на $1 / K$, где K – количество компетенций, проверяемых в диагностической работе.