

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Воронежский государственный технический университет»

УТВЕРЖДАЮ

И.о. декана факультета информационных
технологий и компьютерной безопасности



_____/ А.В. Бредихин

19.03.2024

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
«Технологии обработки больших данных»

Направление подготовки 09.03.02 Информационные системы и технологии

Профиль Информационные системы и технологии цифровизации

Квалификация выпускника бакалавр

Нормативный период обучения 4 года

Форма обучения очная

Год начала подготовки 2022

Автор программы

С.Ю. Белецкая

И.о. заведующего кафедрой
систем

автоматизированного
проектирования и
информационных систем

П.Ю. Гусев

Руководитель ОПОП

Д.В. Иванов

Воронеж 2024

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

1.1. Цели дисциплины

Целью дисциплины является освоение моделей, методов, технологий и инструментов хранения и обработки больших данных

1.2. Задачи освоения дисциплины

- ознакомление студентов с основными классами задач обработки больших данных;
- изучение методов и технологий интеллектуального анализа больших объёмов данных;
- освоение технологий подготовки, хранения и распределённой обработки больших данных;
- приобретение навыков использования технологий и инструментов Big Data при решении практических задач.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Технологии обработки больших данных» относится к дисциплинам части, формируемой участниками образовательных отношений блока Б1.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Процесс изучения дисциплины «Технологии обработки больших данных» направлен на формирование следующих компетенций:

ПК-5 - Способен администрировать сетевую инфраструктуру и программное обеспечение в инфокоммуникационной среде организации

Компетенция	Результаты обучения, характеризующие сформированность компетенции
ПК-5	Знать методы обработки и анализа больших данных, технологии распределённой обработки данных
	Уметь использовать технологии хранения, обработки и анализа больших данных при решении профессиональных задач
	Владеть инструментами Big Data и навыками их практического применения в прикладных областях

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины «Технологии обработки больших данных» составляет 4 з.е.

Распределение трудоемкости дисциплины по видам занятий
очная форма обучения

Виды учебной работы	Всего часов	Семестры
		7
Аудиторные занятия (всего)	72	72
В том числе:		
Лекции	36	36
Лабораторные работы (ЛР)	36	36
Самостоятельная работа	72	72
Курсовой проект		
Часы на контроль		
Виды промежуточной аттестации Зачёт с оценкой	+	+
Общая трудоемкость: академические часы	144	144
зач.ед.	4	4

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1 Содержание разделов дисциплины и распределение трудоемкости по видам занятий

очная форма обучения

№ п/п	Наименование темы	Содержание раздела	Лекц	Лаб. зан.	СРС	Всего, час
1	Основы Big Data	Понятие больших данных. Особенности сбора, хранения, обработки и анализа больших массивов данных. Источники больших данных. Характеристики больших данных. Жизненный цикл больших данных. Этапы обработки больших данных. Задачи, решаемые при обработке больших данных. Принципы управления большими данными. Требования к системам хранения и обработки больших данных. Использование больших данных в информационных системах	4	4	10	18
2	Аналитика больших данных	Основные задачи анализа больших данных в информационных системах. Классификация современных методов анализа данных. Разведочный анализ больших данных. Сбор данных из разных источников. Предварительный анализ, обработка и очистка данных. Особенности технологии Data Mining. Задачи Data Mining и методы их решения. Этапы интеллектуального анализа данных. Использование методов Data Mining для	16	16	30	62

		анализа больших данных. Решение задач классификации, кластеризации, прогнозирования на основе больших данных. Методы машинного обучения, их классификация и особенности. Использование методов Machine learning для анализа больших данных. Нейросетевые технологии обработки информации.. Задачи, решаемые с использованием искусственных нейронных сетей. Классификация искусственных нейронных сетей. Подходы к построению и обучению искусственных нейронных сетей. Этапы нейросетевого анализа данных. Практическое использование нейросетевых технологий для решения задач обработки и анализа больших данных. Технологии визуализации и трансформации данных. Многомерный анализ данных. Современные инструментальные средства и системы анализа больших данных. Использование библиотек Python для решения аналитических задач обработки данных.				
3	Технологии и инструменты распределённой обработки больших данных	Базовые архитектуры распределённой обработки данных. Параллельные и облачные вычисления. Модель распределённых вычислений MapReduce, её назначение, особенности, области применения. Принцип работы MapReduce. Этапы обработки данных на основе MapReduce, базовые функции вычислительной модели. Платформа распределённой обработки больших данных Hadoop. Возможности, основы построения и архитектура Hadoop. Элементы кластера Hadoop. Основные компоненты Hadoop. Основы Yarn. Реализация и использование фреймворка MapReduce в распределённой среде. Распределённая файловая система HDFS, архитектура, принципы функционирования и реализация. Языки поисковых запросов для Hadoop. Состав экосистемы Hadoop, основные проекты и технологии. Apache Spark, его особенности, архитектура, принципы функционирования. Организация распределённой обработки слабоструктурированных данных на основе Apache Spark. Распределённая потоковая платформа Apache Kafka, возможности, архитектура, принципы функционирования. Организация потоковой передачи данных на основе Apache Kafka. Основные принципы работы Apache Hive. Использование современных инструментов BigData при проектировании и эксплуатации информационных систем.	8	8	16	32
4	Технологии хранения больших данных. Организация хранилищ данных на основе нереляционных СУБД	Технологии хранения больших данных на основе нереляционных СУБД. Основные характеристики NoSQL СУБД. Особенности нереляционных СУБД. Ключевые отличия между реляционными и нереляционными СУБД. Модели данных в нереляционных	8	8	16	32

		СУБД. Формулировка CAP теоремы и её следствий. Классификация NoSQL СУБД. Базы данных “ключ-значение”. Ассоциативные массивы, кэш-память и алгоритмы вытеснения. Принципы построения баз данных “ключ -значение. Прикладное использование. Примеры. Принципы организации СУБД Redis. Документо ориентированные базы данных. Принципы организации документо-ориентированных баз данных. Области использования. Пример ДОСУБД MongoDB. Программные интерфейсы для работы с MongoDB. Принципы построения колоночных СУБД. Модели представления данных, организация запросов. Примеры колоночных СУБД – Hbase, Cassandra. Графовые СУБД, их особенности и области применения. Способы организации и хранения аннотированных графов. Характеристики графовых баз данных. Графовая СУБД Neo4j. Программные интерфейсы для работы с Neo4j. Использование нереляционных СУБД при проектировании и эксплуатации информационных систем.				
Итого			36	36	72	144

5.2 Перечень лабораторных работ

- 1 Разведочный анализ данных с использованием библиотек Python.
2. Решение задач кластеризации и классификации больших данных в Python.
3. Решение задач прогнозирования с использованием библиотек Python.
- 4-5. Нейросетевой анализ данных в Python.
6. Платформа Hadoop и ее компоненты.
7. Распределенные вычисления на платформе Apache Spark
- 8-9. Изучение возможностей и работа с документо-ориентированной СУБД MongoDB.

6. ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ ПРОЕКТОВ (РАБОТ) И КОНТРОЛЬНЫХ РАБОТ

В соответствии с учебным планом освоение дисциплины не предусматривает выполнение курсового проекта (работы) или контрольной работы.

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

7.1.1 Этап текущего контроля

Результаты текущего контроля знаний и межсессионной аттестации оцениваются по следующей системе:

«аттестован»;

«не аттестован».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Аттестован	Не аттестован
ПК-5	Знать методы обработки и анализа больших данных, технологии распределённой обработки данных	Знание основных классов задач обработки и анализа данных, технологии интеллектуального анализа данных, методов Data Mining. Ответы на теоретические вопросы при защите лабораторных работ	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	Уметь использовать технологии хранения, обработки и анализа больших данных при решении профессиональных задач	Умение использовать технологии распределённой обработки больших данных, разрабатывать базы данных на основе NoSQL СУБД, применять на практике аналитические методы обработки данных. Выполнение лабораторных работ	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	Владеть инструментами Big Data и навыками их практического применения в прикладных областях	Владение навыками использования инструментальных средств Big Data при решении практических задач. Выполнение лабораторных работ	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах

7.1.2 Этап промежуточного контроля знаний

Результаты промежуточного контроля знаний оцениваются в 7 семестре для очной формы обучения в форме зачёта с оценкой по четырехбальной системе:

«отлично»;

«хорошо»;

«удовлетворительно»;

«неудовлетворительно»

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Отлично	Хорошо	Удовл.	Неудовл.
ПК-5	Знать методы обработки и анализа больших данных, технологии распределённой обработки данных	Тест	Выполнение теста на 90- 100%	Выполнение теста на 80-90%	Выполнение теста на 70-80%	В тесте менее 70% правильных ответов
	Уметь использовать технологии хранения, обработки и анализа больших данных при решении профессиональных задач	Решение стандартных практических задач	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
	Владеть инструментами Big Data и навыками их практического применения в прикладных областях	Решение прикладных задач в конкретной предметной области	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены

7.2 Примерный перечень оценочных средств (типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности)

7.2.1 Примерный перечень заданий для подготовки к тестированию

1 Какими свойствами, как правило, обладают большие данные? (возможно несколько ответов)

- а) структурированность;
- б) неструктурированность;**
- в) объем в терабайтах и петабайтах;**
- г) объем в мегабайтах и гигабайтах.

2 Перечислите четыре основных характеристики Big Data:

- а) Virtualization, Volume, Variability, Vehicle
- б) Variety, Velocity, Volume, Value**
- в) Verification, Volume, Velocity, Visualization
- г) Video, Value, Variety, Volume

3 Какой метод применяется для обнаружения скрытых знаний в больших объемах данных?

- а) **Data Mining;**
- б) дискриминантный анализ;
- в) OLAP;
- г) дисперсионный анализ.

4 Какой из алгоритмов относится к алгоритмам кластеризации?

- 1 Алгоритм Дейкстры
- 2 k-means – алгоритм**
- 3 Генетический алгоритм
- 4 Алгоритм Форда

5 В чем заключаются отличия задач классификации и кластеризации данных?

- 1 Отличий нет, эти понятия являются синонимами;
- 2 При кластеризации заранее известны группы, к которым должен быть отнесен объект;
- 3 При классификации заранее известны группы, к которым должен быть отнесен объект;**
- 4 При классификации заранее не известны группы, к которым должен быть отнесен объект.

6 Hadoop – это...

- а) **набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах**
- б) распределённая СУБД, позволяющая обрабатывать большие данные
- в) язык выполнения заданий в парадигме MapReduce
- г) распределённая файловая система, предназначенная для хранения файлов большого объёма

7 Какие компоненты по умолчанию не входят в состав проекта Apache Hadoop?

- а) HDFS
- б) Spark**
- в) MapReduce
- г) YARN

8. Какие модели хранения данных поддерживают базы данных NoSQL? (возможно несколько ответов)

- а) столбчатые;**
- б) ключ-значение;**
- в) графовые;**
- г) реляционные.

9 Какая из баз данных работает на основе стека Hadoop?

- а) HBase**
- б) Dynamo
- в Bigtable
- г) Cassandra

10 Какая из перечисленных NoSQL СУБД относится к документоориентированному типу?

- 1) HBase
- 2) MongoDB**
- 3) Redis
- 4) Cassandra

11. HDFS – это...

- а) СУБД
- б) среда программирования
- в) распределённая файловая система**
- г) система анализа данных

12 Какая из перечисленных NoSQL СУБД относится к колоночному типу?

- 1) Neo4j
- 2) MongoDB
- 3) Redis
- 4) Cassandra**

7.2.2 Примерный перечень заданий для решения стандартных задач

1 Какая из библиотек Python предназначена для работы с нейронными сетями?

- а) Pandas
- б) Keras**
- в) NumPy
- г) Statsmodels

2 Какая из библиотек Python может быть использована для решения задач кластерного анализа?

- а) Scikit Learn**
- б) Seaborn

- в) NumPy
- г) Pandas

3 Количество нейронов в слоях многослойного персептрона

- 1 Должны быть кратно двум
- 2 **Может быть произвольным**
- 3 Должно быть одинаковым
- 4 Может быть нулевым

4 Как происходит обучение нейронной сети?

- а) эксперты настраивают нейронную сеть
- б) сеть запускается на обучающем множестве, и незадействованные нейроны удаляются
- в) **сеть запускается на обучающем множестве, и подстраиваются весовые значения**
- г) сеть запускается на обучающем множестве, и добавляются или удаляются

5 Выберите верные утверждения о модели данных в HBase (возможно несколько ответов)

- 1) **внутри ячеек таблицы поддерживается хранение нескольких версий данных**
- 2) версионность данных не поддерживается на уровне стандартных средств
- 3) количество колонок созданной таблицы по умолчанию является наперед заданным
- 4) **колонки семейства можно идентифицировать по их общему префиксу**

6 Глубина вложенности документов в MobgoDB:

- 1) **не ограничена**
- 2) ограничена 2-мя уровнями вложенности
- 3) ограничена 1-им уровнем вложенности
- 4) ограничена 3-мя уровнями вложенности

7. Какая базовая структура данных стоит в основе модели MapReduce?

- а) **пара (ключ, значение)**
- б) бинарное дерево
- в) массив произвольной длины
- г) функция высшего порядка

8 На каких узлах в Hadoop MapReduce выполняются map- и reduce-задачи?

- 1) master nodes
- 2) job submission nodes
- 3) slave nodes**
- 4) namenodes

9. Каков размер блока HDFS по умолчанию (для релиза 2.6.0 и выше)?

- 1) 512 байт
- 2) 128 Мбайт**
- 3) 256 Мбайт
- 4) 64 Мбайт

10. Спарк-приложения выполняют вычисления

- 1) в кэше
- 2) в оперативной памяти**
- 3) на жёстком диске
- 4) с использованием внешнего хранилища состояний, например, сторонней СУБД или файловой системы

11. Выберите технологию потоковой обработки событий в режиме реального времени

- a) Spark Streaming
- б) Apache Kafka**
- в) Apache Hadoop
- г) MapReduce

12. Формат Parquet считается

- a) неструктурированным
- б) полуструктурированным
- в) строковым
- г) колоночным (столбцовым)**

13. Что является основной абстракцией для Spark SQL?

- a) набор RDD
- б) отсортированный список
- в) датафрейм**
- г) серия датафрейма

14 Какое количество фоновых процессов менеджера ресурсов и менеджера узлов работает при стандартной конфигурации кластера с YARN архитектурой?

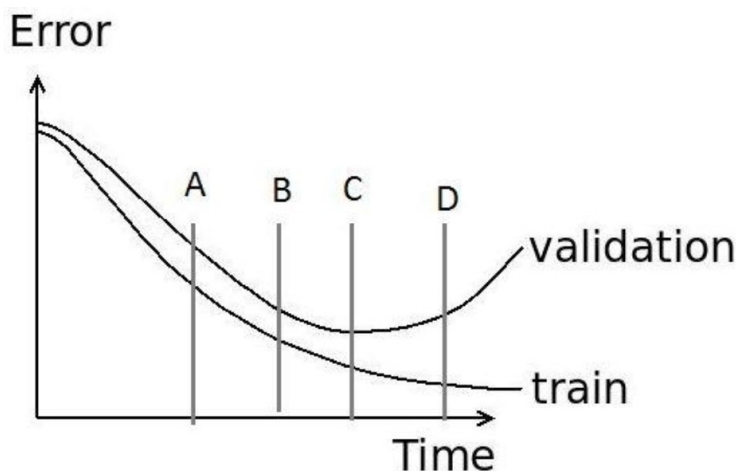
- 1) Один процесс менеджера ресурсов и один или более процессов менеджера узлов (по одному на каждом узле кластера)**
- 2) Один процесс менеджера ресурсов и один процесс менеджера узлов

3) Один или более процессов менеджера ресурсов (по одному на каждом узле кластера) и один процесс менеджера узлов

4) Один или более процессов менеджера ресурсов (по одному на каждом узле кластера) и один или более процессов менеджера узлов (по одному на каждом узле кластера)

7.2.3 Примерный перечень заданий для решения прикладных задач

1 При обучении нейронной сети в задаче распознавания образов был построен следующий график ошибки обучения и ошибки валидации модели. Какой момент времени лучше всего подходит для раннего останова?



- a) A
- б) B
- в) C**
- г) D

2. Опишите работу функции map в модели MapReduce

- 1) функция map возвращает некоторую функцию с аргументом в виде данного списка
- 2) функция map объединяет поэлементно два и более списка, возвращая список пар элементов
- 3) функция map осуществляет поиск данного элемента в списке-аргументе
- 4) функция map применяет данную функцию к каждому элементу списка, возвращая список результатов**

3. Опишите структуру нейронной сети, созданной с помощью программного кода:

```
model = Sequential([Dense(64, input_dim=100), activation('relu'),  
Dense(10), activation('softmax'),])
```

1) MLP 100-64-10, функция активации скрытого слоя relu, функция активации выходного слоя softmax

2) MLP 100-10-64, функция активации скрытого слоя relu, функция активации выходного слоя softmax

3) MLP 100-64-10, функция активации скрытого слоя softmax, функция активации выходного слоя relu

4) MLP 100-10-64, функция активации скрытого слоя relu, функция активации выходного слоя softmax

4 Как следует отредактировать запрос `db.teachers.update({name:'Василий'}, {$inc: {salary:5000}})`, чтобы обновить все найденные документы?

1) `db.teachers.update({name:'Василий'}, {$inc: {salary:5000}},false,false,true)`

2) `db.teachers.update({name:'Василий'}, {$inc: {salary:5000}},false)`

3) `db.teachers.update({name:'Василий'}, {$inc: {salary:5000}},false,true)`

4) `db.teachers.update({name:'Василий'}, {$inc: {salary:5000}},true)`

5. Результатом выполнения операции `db.unicorns.find().skip(5)` в MongoDB будет:

а) возврат первых 5 документов

б) возврат последних 5 документов

в) возврат всех документов, содержащих цифру 5

г) пропуск первых 5 документов коллекции и возврат всех остальных документов

6. Как следует отредактировать запрос `db.teachers.update({name:'Василий'}, {$inc: {salary:5000}})`, чтобы разрешить вставку документа, если его не существует в коллекции?

1) `db.teachers.update({name:'Василий'}, {$inc: {salary:5000}},false,false,true)`

2) `db.teachers.update({name:'Василий'}, {$inc: {salary:5000}},false)`

3) `db.teachers.update({name:'Василий'}, {$inc: {salary:5000}},false,true)`

4) `db.teachers.update({name:'Василий'}, {$inc: {salary:5000}},true)`

7 Для чего предназначена утилита `distcp`?

1) Для параллельного копирования больших объёмов данных из локальной файловой системы в HDFS

2) Для параллельного копирования образа файловой системы между основным узлом имён и резервным

3) Для параллельного копирования больших объёмов данных между кластерами Hadoop

4) Ни один из указанных вариантов

8 Для чего предназначена команда `hdfs dfsadmin -refreshNodes`?

- 1) Проверяет доступность всех узлов данных кластера и выдаёт список существующих, но недоступных по каким-либо причинам узлов
- 2) Вызывает повторное чтение конфигурационных файлов на всех узлах кластера (NameNode и DataNode)
- 3) Обновляет узел имён (NameNode) новым набором разрешённых узлов данных**
- 4) Повторно запускает фоновые процессы на всех узлах кластера

9. Для агрегации данных в случаях отношения «один-ко-многим» или «многие-ко-многим» классические СУРБД применяют конструкцию JOIN. Для таких случаев в MongoDB используется

- а) аналог конструкции JOIN
- б) только вложенные документы
- в) только технология DBRef
- г) вложенные документы и технология DBRef**

10 Результатом выполнения команды `db.source.copyTo(target)` будет

- 1) копирование всех документов из одной коллекции в другую**
- 2) копирование всех полей из одного документа в другой
- 3) копирование определенных полей документа
- 4) копирование всех коллекций

7.2.4 Примерный перечень вопросов для подготовки к зачету

1 Понятие больших данных. Основные задачи обработки больших данных в информационных системах.

2 Характеристики больших данных.

3 Жизненный цикл больших данных. Этапы обработки больших данных.

4 Принципы управления большими данными. Требования к системам хранения и обработки больших данных.

5. Основные задачи анализа данных в интеллектуальных информационных системах.

7. Источники больших данных. Технологии сбора данных из разных источников.

8. Классификация современных методов анализа данных. Понятие о технологии Data Mining.

9. Этапы интеллектуального анализа данных.

10. Первичный разведочный анализ данных.

11. Решение задач кластеризации на основе больших данных.

12. Решение задач классификации на основе больших данных.

13. Решение задач прогнозирования на основе больших данных.
14. Нейросетевые технологии анализа данных.
15. Инструментарий анализа больших данных
16. Требования к распределённым и системам обработки больших данных.
- 17 Модель распределённых вычислений MapReduce, её назначение, особенности, области применения. Этапы обработки данных на основе MapReduce, базовые функции вычислительной модели.
18. Платформа распределённой обработки больших данных Hadoop. Возможности, основы построения и архитектура Hadoop. Основные компоненты Hadoop.
19. Распределённая файловая система HDFS, архитектура, принципы функционирования и реализация. Языки поисковых запросов для Hadoop.
20. Состав экосистемы Hadoop, основные проекты и технологии.
21. Apache Spark, его особенности, архитектура, принципы функционирования. Организация распределённой обработки слабоструктурированных данных на основе Apache Spark.
22. Распределённая потоковая платформа Apache Kafka, возможности, архитектура, принципы функционирования. Организация потоковой передачи данных на основе Apache Kafka.
23. Основные принципы работы Apache Hive. Выполнение запросов, агрегирование и анализ данных.
24. Базы данных NoSQL. Особенности, классификация
- 25 Возможности NoSQL СУБД по обеспечению целостности, доступности скорости обработки информации. CAP-теорема.
- 26 Документо-ориентированные СУБД. Принципы построения, области применения, примеры.
- 27 Возможности СУБД MongoDB. Модели данных. Организация запросов.
28. Базы данных “Ключ-значение”. Принципы построения, примеры.
29. Колоночные СУБД. Принципы построения, примеры.
30. Графовые СУБД. Принципы построения, области применения, примеры.

7.2.5 Примерный перечень заданий для подготовки к экзамену

Экзамен не предусмотрен учебным планом

7.2.6. Методика выставления оценки при проведении промежуточной аттестации

Зачёт с оценкой в 7 семестре проводится по тест-билетам, каждый из которых содержит 10 вопросов и задачу. Каждый правильный ответ на вопрос в тесте оценивается 1 баллом, задача оценивается в 10 баллов (5 баллов верное решение и 5 баллов за верный ответ). Максимальное количество набранных баллов – 20.

1. Оценка «Неудовлетворительно» ставится в случае, если студент набрал менее 6 баллов.
 2. Оценка «Удовлетворительно» ставится в случае, если студент набрал от 6 до 10 баллов
 3. Оценка «Хорошо» ставится в случае, если студент набрал от 11 до 15 баллов.
 4. Оценка «Отлично» ставится, если студент набрал от 16 до 20 баллов
1. Зачёт ставится в случае, если студент набрал более 10 баллов.
 2. Незачёт ставится в случае, если студент набрал менее 10 баллов

7.2.7 Паспорт оценочных материалов

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции	Наименование оценочного средства
1	Основы Big Data	ПК-5	Тест, защита лабораторных работ
2	Аналитика больших данных	ПК-5	Тест, защита лабораторных работ
3	Технологии и инструменты распределённой обработки больших данных	ПК-5	Тест, защита лабораторных работ
4	Технологии хранения больших данных. Организация хранилищ данных на основе нереляционных СУБД	ПК-5	Тест, защита лабораторных работ

7.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Тестирование осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных тест-заданий на бумажном носителе. Время тестирования 30 мин. Затем осуществляется проверка теста экзаменатором и выставляется оценка согласно методики выставления оценки при проведении промежуточной аттестации.

Решение стандартных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

Решение прикладных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

8 УЧЕБНО МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ)

8.1 Перечень учебной литературы, необходимой для освоения дисциплины

1 Черемухин, А. Д. Большие данные [Электронный ресурс]: учебное пособие / А. Д. Черемухин. — Москва : Ай Пи Ар Медиа, 2023. — 782 с. — Режим доступа: <https://www.iprbookshop.ru/129721>

2 Чубукова, И. А. Data Mining : учебное пособие / И. А. Чубукова. — 4-е изд. — Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2024. — 469 с. — Режим доступа <https://www.iprbookshop.ru/133907>.

3. Пальмов С.В. Интеллектуальный анализ данных [Электронный ресурс]: учеб. пособие. – Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017. – 127 с. – Режим доступа <http://www.iprbookshop.ru/75376>

4. Протодяконов, А. В. Алгоритмы Data Science и их практическая реализация на Python [Электронный ресурс]: учебное пособие / А. В. Протодяконов, П. А. Пылов, В. Е. Садовников. — Москва, Вологда : Инфра-Инженерия, 2022. — 392 с. — Режим доступа: <https://www.iprbookshop.ru/124000>

5 Лэм, Ч. Hadoop в действии [Электронный ресурс]/ Ч. Лэм. — Москва : ДМК Пресс, 2019. — 424 с. — Режим доступа <https://www.iprbookshop.ru/124537>

8.2. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень лицензионного программного обеспечения, ресурсов информационно-телекоммуникационной сети «Интернет», современных профессиональных баз данных и информационных справочных систем:

Программное обеспечение

Hadoop (Hortonworks Data platform)

Apache Spark

Apache Kafka

MongoDB

HBase

Redis

IntelliJ Idea community (java)

OpenJDK 11

Python

Ресурсы информационно-телекоммуникационной сети «Интернет»

<http://www.edu.ru/>

Информационные справочные системы

<http://window.edu.ru>

<https://wiki.cchgeu.ru/>

Современные профессиональные базы данных

<https://habr.com/ru/>

<https://sources.ru/>

<https://proglib.io/>

**9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ
ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА**

Для проведения обучения по дисциплине используются:

Компьютерный класс

Учебная аудитория для проведения учебных занятий, включающая:

- рабочее место преподавателя (стол, стул);
- рабочие места обучающихся (столы, стулья)
- персональные компьютеры с установленным ПО, подключенные к сети Интернет (12 шт.);
- принтер;
- доска магнитно-маркерная поворотная
- оборудование для лекционных демонстраций и проекционная аппаратура.

Помещение для самостоятельной работы. Читальный зал с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду.

**10. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО
ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)**

По дисциплине «Технологии обработки больших данных» читаются лекции, проводятся лабораторные работы.

Основой изучения дисциплины являются лекции, на которых излагаются наиболее существенные и трудные вопросы, а также вопросы, не нашедшие отражения в учебной литературе.

Лабораторные работы выполняются на лабораторном оборудовании в соответствии с методиками, приведенными в указаниях к выполнению работ.

Вид учебных	Деятельность студента
-------------	-----------------------

занятий	
Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; помечать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначение вопросов, терминов, материала, которые вызывают трудности, поиск ответов в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на лекции или на практическом занятии.
Лабораторная работа	Лабораторные работы позволяют научиться применять теоретические знания, полученные на лекции при решении конкретных задач. Чтобы наиболее рационально и полно использовать все возможности лабораторных для подготовки к ним необходимо: следует разобрать лекцию по соответствующей теме, ознакомиться с соответствующим разделом учебника, проработать дополнительную литературу и источники, решить задачи и выполнить другие письменные задания.
Самостоятельная работа	Самостоятельная работа студентов способствует глубокому усвоения учебного материала и развитию навыков самообразования. Самостоятельная работа предполагает следующие составляющие: - работа с текстами: учебниками, справочниками, дополнительной литературой, а также проработка конспектов лекций; - выполнение домашних заданий и расчетов; - работа над темами для самостоятельного изучения; - участие в работе студенческих научных конференций, олимпиад; - подготовка к промежуточной аттестации.
Подготовка к промежуточной аттестации	Готовиться к промежуточной аттестации следует систематически, в течение всего семестра. Интенсивная подготовка должна начаться не позднее, чем за месяц-полтора до промежуточной аттестации. Данные перед зачётом дни эффективнее всего использовать для повторения и систематизации материала.

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

№ п/п	Перечень вносимых изменений	Дата внесения изменений	Подпись заведующего кафедрой, ответственной за реализацию ОПОП
----------	-----------------------------	----------------------------	--