

**Рабочая программа дисциплины (модуля) «Аналитика больших данных»,
включающая оценочные и методические материалы**

1. Требования к результатам обучения по дисциплине (модулю)

1.1. Перечень компетенций, формируемых дисциплиной (модулем) в процессе освоения образовательной программы

Группа компетенций	Категория компетенций	Коды и содержание компетенций
Профессиональные	-	ПК-2. Создает и использует программные средства обработки больших объемов данных, в том числе с использованием методов машинного обучения и искусственного интеллекта, применяет результаты обработки в решении промышленных и бизнес-задач

1.2. Компетенции и индикаторы их достижения, формируемых дисциплиной (модулем) в процессе освоения образовательной программы

Код компетенции	Код индикатора компетенции	Содержание индикатора компетенции
ПК-2	ПК-2.1	Владеет инструментами и средствами работы с большими данными
	ПК-2.2	Выполняет анализ больших данных на основе требований
	ПК-2.3	Создает и применяет программные системы на основе результатов обработки больших данных для решения промышленных и бизнес-задач

1.3. Результаты обучения по дисциплине (модулю)

машинного обучения, предназначенных для анализа данных и построения предсказательных моделей.

В результате изучения дисциплины (модуля) обучающийся должен

знать:

- теорию базы данных;
- основные структуры и модели данных;
- методы обработки данных.

уметь:

- идентифицировать класс разрабатываемой системы управления базами данными (СУБД) в зависимости от выполняемых ею задач и аппаратных средств, определенных в техническом задании на разработку СУБД;
- анализировать, оптимизировать и документировать проектные процессы, а также поддерживать их на всем протяжении жизненного цикла разработки.

владеть:

- навыками разработки структуры управления базами данными в целом и ее отдельных компонентов;
- способностью формирования плана разработки, управления рисками.

2. Объем, структура и содержание дисциплины (модуля)

2.1. Объем дисциплины (модуля)

<i>Виды учебной работы</i>	<i>Формы обучения</i>
	<i>Очная</i>
Общая трудоемкость: зачетные единицы/часы	4/144
Контактная работа:	84
Лекции	36
Лабораторные работы	24
Практические занятия, семинары	24
Промежуточная аттестация: экзамен	36
Самостоятельная работа (СР)	24

2.2. Темы (разделы) дисциплины (модуля) с указанием отведенного на них количества часов по формам образовательной деятельности

Очная форма обучения

№ п/п		Наименование тем (разделов)	Виды учебной работы (в часах)						СР
			Контактная работа						
			Занятия лекционного типа		Занятия семинарского типа				
			Л	Иные	ПЗ	С	ЛР	Иные	
1.	Хранение и обработка больших наборов данных в Apache Hive.	18	0	12	0	12	0	12	
2.	Анализ больших наборов данных в Apache Hive.	18	0	12	0	12	0	12	

2.3. Содержание дисциплины (модуля), структурированное по темам (разделам) и видам работ

Содержание лекционного курса

№ п/п	Наименование тем (разделов)	Содержание лекционного курса
1.	Хранение и обработка больших наборов данных в Apache Hive.	Введение в анализ больших данных. Методы описательной статистики Свойства подхода к анализу данных data mining. Этапы анализа больших данных согласно стандарту CRISP-DM Cross Industry Standard Process for Data Mining Особенности архитектуры кластера Apache Hadoop, распределенная файловая система Hadoop HDFS. Хранилище данных Apache Hive. Применение модели распределённых вычислений MapReduce при обработке данных в Apache Hive .
2.	Анализ больших наборов данных в Apache Hive.	Метод корреляционного анализа данных. Оценка параметров модели Метод дисперсионного анализа данных. Оценка параметров модели Метод линейного регрессионного анализа данных. Оценка параметров модели Метод ковариационного анализа данных. Оценка параметров модели

Содержание занятий семинарского типа

№ п/п	Наименование тем (разделов)	Тип	Содержание занятий семинарского типа
1.	Хранение и обработка больших наборов данных в Apache Hive.	ЛР	Создание и заполнение «простой» таблицы в Apache Hive Использование строковых функции в Apache Hive Использование числовых в Apache Hive Использование функций для работы с датами в Apache Hive
		ПР	Установка и настройка программного обеспечения («виртуальный» шаблон Hadoop хранилище Apache Hive, Hue и др.) Проектирование и генерация датасета, загрузка в Apache Hive Проектирование и создание разделов (partitions) в таблице Apache Hive. Проектирование и создание кластеров (buckets) в таблице Apache Hive
2.	Анализ больших наборов данных в Apache Hive.	ЛР	Использование функций преобразования типов в Apache Использование функции задания условия в Apache Hive Использование групповых функции в Apache Hive

			Использование оконных в Apache Hive
		ПР	Проектирование «сложной» таблицы и применение функций для работы с массивами, картами и структурами Использование функций преобразования типов в запросах на языке HiveQL Проектирование датасета, создание «управляемой» таблицы применение аналитических функций анализа данных. Проектирование датасета, создание «управляемой» таблицы и применение функций для регрессионного анализа

Содержание самостоятельной работы

№ п/п	Наименование тем (разделов)	Содержание самостоятельной работы
1.	Хранение и обработка больших наборов данных в Apache Hive.	Работа с лекционным материалом, поиск и обзор литературы и электронных источников информации по индивидуально заданной проблеме курса; Работа в электронном курсе (изучение теоретического материала, выполнение тестов и контрольных работ, разработка программ); Выполнение домашних заданий; Подготовка к лабораторным работам; Разработка программ для решения нестандартных задач; Анализ научных публикаций по заранее определенной преподавателем теме.
2.	Анализ больших наборов данных в Apache Hive.	Работа с лекционным материалом, поиск и обзор литературы и электронных источников информации по индивидуально заданной проблеме курса; Работа в электронном курсе (изучение теоретического материала, выполнение тестов и контрольных работ, разработка программ); Выполнение домашних заданий; Подготовка к лабораторным работам; Разработка программ для решения нестандартных задач; Анализ научных публикаций по заранее определенной преподавателем теме.

3. Оценочные материалы для проведения текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине (модулю)

По дисциплине (модулю) предусмотрены следующие виды контроля качества освоения:

- текущий контроль успеваемости;
- промежуточная аттестация обучающихся по дисциплине (модулю).

3.1.1. Оценочные материалы для проведения текущей аттестации по дисциплине (модулю)

№ п/п	Контролируемые темы (разделы)	Наименование оценочного средства
1.	Хранение и обработка больших наборов данных в Apache Hive.	Контрольная работа
2.	Анализ больших наборов данных в Apache Hive.	Контрольная работа

3.1.1. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности в процессе текущего контроля успеваемости

Контрольная работа

Контрольная работа №1:

Что такое большие данные и какие проблемы они представляют для анализа?

Объясните различия между MapReduce и Spark.

Что такое ETL-процессы и почему они важны для аналитики больших данных?

Объясните основные этапы процесса машинного обучения.

Какие характеристики делают Hadoop распределенной файловой системой?

Контрольная работа №2:

Каким образом можно оптимизировать запросы в распределенных системах хранения данных?

Как обеспечивается безопасность данных при их обработке и хранении?

Расскажите о принципах управления доступом к данным в распределенных системах.

Какие методы шифрования используются в аналитике больших данных?

Как предотвратить утечку конфиденциальных данных при анализе больших объемов информации?

Какие принципы лежат в основе визуализации данных?

3.1.2. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности в ходе текущего контроля успеваемости

Контрольная работа

Оценивается не только глубина знаний поставленных вопросов, но и умение изложить письменно.

Критерии оценивания: последовательность, полнота, логичность изложения, анализ различных точек зрения, самостоятельное обобщение материала. Изложение материала без фактических ошибок.

Оценка «отлично» ставится в случае, когда соблюдены все критерии.

Оценка «хорошо» ставится, если обучающийся твердо знает материал, грамотно и по существу излагает его, знает практическую базу, но допускает несущественные погрешности.

Оценка «удовлетворительно» ставится, если обучающийся освоил только основной материал, но не знает отдельных деталей, допускает неточности, недостаточно правильные формулировки, нарушает последовательность в изложении материала, затрудняется с ответами, показывает отсутствие должной связи между анализом, аргументацией и выводами.

Оценка «неудовлетворительно» ставится, если обучающийся не отвечает на поставленные вопросы.

3.2. Оценочные материалы для проведения промежуточной аттестации

3.2.1. Критерии оценки результатов обучения по дисциплине (модулю)

Шкала оценивания	Результаты обучения	Показатели оценивания результатов обучения
ОТЛИЧНО	Знает:	- обучающийся глубоко и всесторонне усвоил материал, уверенно, логично, последовательно и грамотно его излагает, опираясь на знания основной и дополнительной литературы, - на основе системных научных знаний делает квалифицированные выводы и обобщения, свободно оперирует категориями и понятиями.
	Умеет:	- обучающийся умеет самостоятельно и правильно решать учебно-профессиональные задачи или задания, уверенно, логично, последовательно и аргументировано излагать свое решение, используя научные понятия, ссылаясь на нормативную базу.
	Владеет:	- обучающийся владеет рациональными методами (с использованием рациональных методик) решения сложных профессиональных задач, представленных деловыми играми, кейсами и т.д.; При решении продемонстрировал навыки - выделения главного, - связкой теоретических положений с требованиями руководящих документов, - изложения мыслей в логической последовательности, - самостоятельного анализа факты, событий, явлений, процессов в их взаимосвязи и диалектическом развитии.
ХОРОШО	Знает:	- обучающийся твердо усвоил материал, достаточно грамотно его излагает, опираясь на знания основной и дополнительной литературы,

		- затрудняется в формулировании квалифицированных выводов и обобщений, оперирует категориями и понятиями, но не всегда правильно их верифицирует.
	Умеет:	- обучающийся умеет самостоятельно и в основном правильно решать учебно-профессиональные задачи или задания, уверенно, логично, последовательно и аргументировано излагать свое решение, не в полной мере используя научные понятия и ссылки на нормативную базу.
	Владеет:	- обучающийся в целом владеет рациональными методами решения сложных профессиональных задач, представленных деловыми играми, кейсами и т.д.; При решении смог продемонстрировать достаточность, но не глубинность навыков, - выделения главного, - изложения мыслей в логической последовательности, - связи теоретических положений с требованиями руководящих документов, - самостоятельного анализа факты, событий, явлений, процессов в их взаимосвязи и диалектическом развитии.
УДОВЛЕТВОРИТЕЛЬНО	Знает:	- обучающийся ориентируется в материале, однако затрудняется в его изложении; - показывает недостаточность знаний основной и дополнительной литературы; - слабо аргументирует научные положения; - практически не способен сформулировать выводы и обобщения; - частично владеет системой понятий.
	Умеет:	- обучающийся в основном умеет решить учебно-профессиональную задачу или задание, но допускает ошибки, слабо аргументирует свое решение, недостаточно использует научные понятия и руководящие документы.
	Владеет:	- обучающийся владеет некоторыми рациональными методами решения сложных профессиональных задач, представленных деловыми играми, кейсами и т.д.; При решении продемонстрировал недостаточность навыков - выделения главного, - изложения мыслей в логической последовательности, - связи теоретических положений с требованиями руководящих документов, - самостоятельного анализа факты, событий, явлений, процессов в их взаимосвязи и диалектическом развитии.
НЕУДОВЛЕТВОРИТЕЛЬНО	Знает:	- обучающийся не усвоил значительной части материала; - не может аргументировать научные положения; - не формулирует квалифицированных выводов и обобщений; - не владеет системой понятий.
	Умеет:	обучающийся не показал умение решать учебно-профессиональную задачу или задание.
	Владеет:	не выполнены требования, предъявляемые к навыкам, оцениваемым «удовлетворительно».

3.2.2. Контрольные задания и/или иные материалы для проведения промежуточной аттестации

Список вопросов для устных ответов

1. Какие основные характеристики определяют большие данные, и почему они представляют вызовы для традиционных методов анализа?
2. Объясните концепцию "жизненного цикла данных" и опишите основные этапы этого цикла.
3. В чем заключается различие между понятиями "структурированные данные" и "неструктурированные данные"? Приведите примеры каждого типа данных.
4. Какие технологии и инструменты широко используются в аналитике больших данных, и какие задачи они решают?

5. Что такое ETL-процессы, и почему они важны для успешной работы с данными в контексте аналитики больших данных?
6. Расскажите о принципах работы алгоритмов MapReduce. В каких сценариях они могут быть полезны?
7. Какие языки программирования широко используются в анализе больших данных? Объясните, почему выбор языка программирования важен.
8. Какие методы оптимизации запросов могут быть применены при работе с большими объемами данных? Приведите примеры.
9. Что такое NoSQL базы данных, и какие типы NoSQL баз существуют? Когда целесообразно использовать NoSQL вместо реляционных баз данных?
10. Какие принципы лежат в основе работы с данными в распределенных системах хранения, например, Apache Hadoop?
11. Каким образом можно обеспечить безопасность данных в аналитике больших данных? Укажите основные принципы и методы.
12. Расскажите о принципах работы машинного обучения. Какие этапы включают в себя процессы обучения модели?
13. Каким образом визуализация данных может улучшить понимание информации? Приведите примеры сценариев использования визуализации.
14. Какие вызовы могут возникнуть при работе с неструктурированными данными, и как их можно преодолеть?
15. В чем заключается роль аналитики больших данных в принятии бизнес-решений? Какие преимущества она предоставляет организации?
16. Каким образом можно избежать переобучения модели машинного обучения? Приведите методы и стратегии.
17. Какие особенности работы с данными в реальном времени важны для аналитики больших данных? Укажите примеры приложений реального времени.
18. Как влияет шардирование данных на производительность NoSQL баз данных? Как выбрать подходящий метод шардирования?
19. В чем суть принципа "делегированного управления данными" в контексте аналитики больших данных?
20. Какие этические и законные вопросы могут возникнуть при сборе, хранении и использовании больших данных? Как организация может обеспечить соблюдение норм и стандартов в этой области?

3.2.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков в ходе промежуточной аттестации

Процедура оценивания знаний (устный ответ)

Предел длительности	10 минут
Предлагаемое количество заданий	2
Последовательность выборки вопросов из каждого раздела	Случайная
Критерии оценки	<ul style="list-style-type: none"> - требуемый объем и структура - изложение материала без фактических ошибок - логика изложения - использование соответствующей терминологии - стиль речи и культура речи - подбор примеров из научной литературы и практики
«5» если	требования к ответу выполнены в полном объеме
«4» если	в целом выполнены требования к ответу, однако есть небольшие неточности в изложении некоторых вопросов
«3» если	требования выполнены частично – не выдержан объем, есть фактические ошибки, нарушена логика изложения, недостаточно используется соответствующая терминология

4. Учебно-методическое и материально-техническое обеспечение дисциплины (модуля)

4.1. Электронные учебные издания

1. Макшанов, А. В. Технологии интеллектуального анализа данных: учебное пособие / А. В. Макшанов, А. Е. Журавлев. — 2-е изд., стер. — Санкт-Петербург: Лань, 2022. — 212 с. — ISBN 978-5-8114-4493-9. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/206711> (дата обращения: 05.12.2023). — Режим доступа: для авториз. пользователей.
2. Радченко, И. А. Технологии и инфраструктура Big Data: учебное пособие / И. А. Радченко, И. Н. Николаев. — Санкт-Петербург: НИУ ИТМО, 2018. — 52 с. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/136430> (дата обращения: 05.12.2023). — Режим доступа: для авториз. пользователей.

.2. Электронные образовательные ресурсы

1. Электронная библиотечная система «ЭБС ЮРАЙТ» Biblio-online.ru (ЭБС «Юрайт») [Электронный ресурс]. — URL: <https://urait.ru/>.
2. Электронно-библиотечная система ZNANIUM [Электронный ресурс]. — URL: <https://znanium.com/>.
3. Электронная библиотечная система «Консультант студента» [Электронный ресурс]. — URL: <https://www.studentlibrary.ru/>.
4. e-Library.ru: Научная электронная библиотека [Электронный ресурс]. — URL: <http://elibrary.ru/>.
5. Научная электронная библиотека «КиберЛенинка» [Электронный ресурс]. — URL: <http://cyberleninka.ru/>.
6. Информационная система «Единое окно доступа к образовательным ресурсам» [Электронный ресурс]. — URL: <http://window.edu.ru/>.
7. Федеральный центр информационно-образовательных ресурсов [Электронный ресурс]. — URL: <http://fcior.edu.ru/>.

4.3. Современные профессиональные базы данных и информационные справочные системы

Обучающимся обеспечен доступ (удаленный доступ) к ниже следующим современным профессиональным базам данных и информационным справочным системам:

1. Словари и энциклопедии на Академике [Электронный ресурс]. — URL: <http://dic.academic.ru>.
2. Система информационно-правового обеспечения «Гарант» [Электронный ресурс]. — <http://www.garant.ru/>.
3. База данных Института философии РАН: Философские ресурсы: Текстовые ресурсы: <https://iphras.ru/page52248384.htm>.

4.4. Комплект лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства

1. Лицензионное программное обеспечение: операционная система Microsoft Windows, пакет офисных приложений Microsoft Office.
2. Свободно распространяемое программное обеспечение: свободные пакеты офисных приложений Apache Open Office, LibreOffice.
3. Программное обеспечение отечественного производства: справочно-правовая система «Гарант» (Электронный периодический справочник «Система ГАРАНТ»), образовательная платформа ЮРАЙТ (Электронная библиотечная система «ЭБС ЮРАЙТ» Biblio-online.ru (ЭБС «Юрайт»)), электронно-библиотечная система ZNANIUM, электронная библиотечная система «Консультант студента».

4.5. Оборудование и технические средства обучения

Для реализации дисциплины (модуля) используются учебные аудитории для проведения учебных занятий, которые оснащены оборудованием и техническими средствами обучения, и помещения для самостоятельной работы обучающихся, которые оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и

обеспечены доступом в электронную информационно-образовательную среду РХТУ им. Д.И. Менделеева. Допускается замена оборудования его виртуальными аналогами.

Наименование учебных аудиторий для проведения учебных занятий и помещений для самостоятельной работы*	Оснащенность учебных аудиторий для проведения учебных занятий и помещений для самостоятельной работы оборудованием и техническими средствами обучения
Учебные аудитории для проведения учебных занятий	Учебная аудитория укомплектована специализированной мебелью, отвечающей всем установленным нормам и требованиям, оборудованием и техническими средствами обучения (мобильное мультимедийное оборудование).
Помещение для самостоятельной работы	Помещение оснащено компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду РХТУ им. Д.И. Менделеева и к ЭБС.

* Номер конкретной аудитории указан в приказе об аудиторном фонде, расписании учебных занятий и расписании промежуточной аттестации.