

Документ подписан Министерством науки и высшего образования Российской Федерации
Информация о владельце:
ФИО: Макаренко Елена Николаевна
Должность: Ректор
Дата подписания: 19.08.2024 11:37:50
Уникальный программный ключ:
c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

УТВЕРЖДАЮ
Директор Института магистратуры
Иванова Е.А.
«03» июня 2024г.

**Рабочая программа дисциплины
Технологии анализа больших данных**

Направление 09.04.03 Прикладная информатика
магистерская программа 09.04.03.01 "Информационные системы и технологии в
бизнесе"

Для набора 2024 года

Квалификация
магистр

КАФЕДРА Информационных систем и прикладной информатики**Распределение часов дисциплины по курсам**

Курс Вид занятий	2		Итого	
	УП	РП		
Лекции	4	4	4	4
Практические	8	8	8	8
Итого ауд.	12	12	12	12
Контактная работа	12	12	12	12
Сам. работа	164	164	164	164
Часы на контроль	4	4	4	4
Итого	180	180	180	180

ОСНОВАНИЕ

Учебный план утвержден учёным советом вуза от 26.03.2024 протокол № 13.

Программу составил(и): к.э.н., доцент, Аручиди Н.А.

Зав. кафедрой: д.э.н., доцент Щербаков С.М.

Методическим советом направления: д.э.н., доцент, Щербаков С.М.

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	формирование у обучающихся способности осуществлять обработку больших объемов данных для решения профессиональных задач, эффективно применять методы, технологии и инструментальные средства анализа больших данных в профессиональной деятельности.
-----	--

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

ПК-4:Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика

ПК-6:Способен управлять этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации

В результате освоения дисциплины обучающийся должен:

Знать:
основы аналитики больших данных (соотнесено с индикатором ПК-4.1) этапы жизненного цикла методологической и технологической инфраструктуры (соотнесено с индикатором ПК-6.1)
Уметь:
создавать комплексные системы на основе аналитики больших данных (соотнесено с индикатором ПК-4.2) управлять получением, хранением, передачей, обработкой больших данных (соотнесено с индикатором ПК-6.2)
Владеть:
навыками руководства проектами по построению комплексных систем на основе аналитики больших данных в различных сферах (соотнесено с индикатором ПК-4.3) навыками управления качеством больших данных (соотнесено с индикатором ПК-6.3)

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Раздел 1. Основы построения и использования систем больших данных

№	Наименование темы / Вид занятия	Семе стр	Часов	Компетен- ции	Литература
1.1	Тема 1. «Основы систем больших данных». Понятие больших данных. Особенности сбора, хранения, обработки и анализа больших массивов данных. Источники больших данных. Базовые характеристики и метрики качества больших данных. Измерение и оценка качества больших данных. Методы и технологии управления качеством больших данных. Методы оценки рисков использования больших данных. Инженер по качеству данных, дата стьюард. Data Steward. Использование больших данных в науке, бизнесе, государственном управлении. / Лек /	2	2	ПК-4,ПК-6	Л1.1, Л1.2, Л2.1, Л2.3, Л2.2
1.2	Тема 2. «Методы работы с распределенными информационными системами. Управление качеством данных». Использование фреймворка в распределенной среде. Реализации фреймворка. Состав и возможности программного комплекса. Языки поисковых запросов. Управление качеством данных. Выполнение заданий с использованием Yandex Data Lens, ClickHouse. / Пр /	2	4	ПК-4,ПК-6	Л1.1, Л1.2, Л2.1, Л2.3, Л2.2
1.3	Тема 1. «Основы систем больших данных». Понятие больших данных. Особенности сбора, хранения, обработки и анализа больших массивов данных. Источники больших данных. Базовые характеристики и метрики качества больших данных. Измерение и оценка качества больших данных. Методы и технологии управления качеством больших данных. Методы оценки рисков использования больших данных. Инженер по качеству данных, дата стьюард. Data Steward. Использование больших данных в науке, бизнесе, государственном управлении. / Ср /	2	28	ПК-4,ПК-6	Л1.1, Л1.2, Л2.1, Л2.3, Л2.2
1.4	Тема 2. «Методы работы с распределенными информационными системами. Управление качеством данных». Использование фреймворка в распределенной среде. Реализации фреймворка. Состав и возможности программного комплекса. Языки поисковых запросов. Управление качеством данных. / Ср /	2	36	ПК-4,ПК-6	Л1.1, Л1.2, Л2.1, Л2.3, Л2.2

Раздел 2. Разработка и использование приложений на основе распределенных баз данных

№	Наименование темы / Вид занятия	Семе стр	Часов	Компетен- ции	Литература
---	---------------------------------	-------------	-------	------------------	------------

2.1	Тема 3. «Базы данных NoSQL». Варианты построения распределенных баз данных, репликация, фрагментация. Согласованность. CAP-теорема. Классы NoSQL баз данных. Примеры СУБД NoSQL. Семейства столбцов. Графовые СУБД. Neo4j. / Лек /	2	2	ПК-4,ПК-6	Л1.1, Л1.2, Л2.1, Л2.3, Л2.2
2.2	Тема 3. «Базы данных NoSQL». Варианты построения распределенных баз данных, репликация, фрагментация. Согласованность. CAP-теорема. Классы NoSQL баз данных. Примеры СУБД NoSQL. Семейства столбцов. Графовые СУБД. Neo4j. Выполнение заданий с использованием Yandex Data Lens, ClickHouse. / Пр /	2	2	ПК-4,ПК-6	Л1.1, Л1.2, Л2.1, Л2.3, Л2.2
2.3	Тема 4. «Документно-ориентированные распределенные СУБД». Понятие агрегата. Современные документо-ориентированные СУБД. Запросы к СУБД на языке JSON. Использование фреймворка в документо-ориентированных СУБД. Выполнение заданий с использованием Yandex Data Lens, ClickHouse. / Пр /	2	2	ПК-4,ПК-6	Л1.1, Л1.2, Л2.1, Л2.3, Л2.2
2.4	Тема 3. «Базы данных NoSQL». Варианты построения распределенных баз данных, репликация, фрагментация. Согласованность. CAP-теорема. Классы NoSQL баз данных. Примеры СУБД NoSQL. Семейства столбцов. Графовые СУБД. Neo4j. / Ср /	2	46	ПК-4,ПК-6	Л1.1, Л1.2, Л2.1, Л2.3, Л2.2
2.5	Тема 4. «Документно-ориентированные распределенные СУБД». Понятие агрегата. Современные документо-ориентированные СУБД. Запросы к СУБД на языке JSON. Использование фреймворка в документо-ориентированных СУБД. / Ср /	2	54	ПК-4,ПК-6	Л1.1, Л1.2, Л2.1, Л2.3, Л2.2
2.6	/ Зачёт /	2	4	ПК-4,ПК-6	Л1.1, Л1.2, Л2.1, Л2.3, Л2.2

4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Структура и содержание фонда оценочных средств для проведения текущей и промежуточной аттестации представлены в Приложении 1 к рабочей программе дисциплины.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л1.1	Жуковский О. И.	Информационные технологии и анализ данных: учебное пособие	Томск: Эль Контент, 2014	https://biblioclub.ru/index.php?page=book&id=480500 неограниченный доступ для зарегистрированных пользователей
Л1.2	Барсегян А., Куприянов М., Степаненко В., Холод И.	Технологии анализа данных: Data Mining, Text Mining, Visual Mining, OLAP. 2 изд.	Санкт-Петербург: БХВ-Петербург, 2008	https://ibooks.ru/reading.php?short=1&productid=3351 56 неограниченный доступ для зарегистрированных пользователей

5.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л2.1	Емельянов А. А.	Прикладная информатика: журнал	Москва: Синергия ПРЕСС, 2010	https://biblioclub.ru/index.php?page=book&id=120321 неограниченный доступ для зарегистрированных пользователей
Л2.2	Федин, Ф. О., Федин, Ф. Ф.	Анализ данных. Часть 2. Инструменты Data Mining: учебное пособие	Москва: Московский городской педагогический университет, 2012	https://www.iprbookshop.ru/26445.html неограниченный доступ для зарегистрированных пользователей

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л2.3	Пальмов, С. В.	Интеллектуальный анализ данных: учебное пособие	Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017	https://www.iprbookshop.ru/75376.html неограниченный доступ для зарегистрированных пользователей

5.3 Профессиональные базы данных и информационные справочные системы

Национальная электронная библиотека (НЭБ) - <https://rusneb.ru/>

ИСС «КонсультантПлюс»

ИСС «Гарант» <http://www.internet.garant.ru/>

5.4. Перечень программного обеспечения

Yandex Data Lens

ClickHouse

5.5. Учебно-методические материалы для студентов с ограниченными возможностями здоровья

При необходимости по заявлению обучающегося с ограниченными возможностями здоровья учебно-методические материалы предоставляются в формах, адаптированных к ограничениям здоровья и восприятия информации. Для лиц с нарушениями зрения: в форме аудиофайла; в печатной форме увеличенным шрифтом. Для лиц с нарушениями слуха: в форме электронного документа; в печатной форме. Для лиц с нарушениями опорно-двигательного аппарата: в форме электронного документа; в печатной форме.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Помещения для всех видов работ, предусмотренных учебным планом, укомплектованы необходимой специализированной учебной мебелью и техническими средствами обучения:

- столы, стулья;

- персональный компьютер / ноутбук (переносной);

- проектор, экран / интерактивная доска.

Практические занятия проводятся в компьютерных классах, рабочие места в которых оборудованы необходимыми лицензионными и/или свободно распространяемыми программными средствами и выходом в Интернет.

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Методические указания по освоению дисциплины представлены в Приложении 2 к рабочей программе дисциплины.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

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

1.1. Показатели и критерии оценивания компетенций:

ЗУН, составляющие компетенцию	Показатели оценивания	Критерии оценивания	Средства оценивания
ПК-4. Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика			
З. основы аналитики больших данных	знает методологию и принципы руководства проектами по созданию, поддержке и использованию комплексных систем на основе аналитики больших данных	полнота и содержательность ответа умение приводить примеры	ТЗ – тестовые задания (1-12), 3 – вопросы к зачету (1-19)
У. создавать комплексные системы на основе аналитики больших данных	решает задачи по руководству коллективной проектной деятельностью для создания, поддержки и использования комплексных систем на основе аналитики больших данных	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (1-2)
В. навыками руководства проектами по построению комплексных систем на основе аналитики больших данных в различных сферах	выявляет области деловой деятельности, которые потенциально могут получить отдачу от аналитики	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (1-2)
ПК-6: Способен управлять этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации			
З. технологии, методы и инструментальные средства обработки больших данных	знает технологии, методы и инструментальные средства обработки больших данных	полнота и содержательность ответа умение приводить примеры	ТЗ – тестовые задания (1-12), 3 – вопросы к зачету (10-18)
У. использовать методы и инструменты получения, хранения, передачи, обработки больших данных	пользуется методами и инструментами получения, хранения, передачи, обработки больших данных	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (3)
В. формализованного описания алгоритмов решения задач анализа данных	проектирует архитектуру информационных систем на основе нереляционных баз данных и распределенных систем хранения	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (3)

1.2 Шкалы оценивания:

Текущий контроль успеваемости и промежуточная аттестация осуществляется в рамках накопительной балльно-рейтинговой системы в 100-балльной шкале.

50-100 баллов (зачет);

0-49 баллов (незачет).

2. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Вопросы к зачету

1. Понятие Больших данных.
2. Особенности сбора, хранения, обработки и анализа больших массивов данных.
3. Источники больших данных.
4. Использование больших данных в науке, бизнесе, государственном управлении.
5. Использование фреймворка в распределенной среде.
6. Реализации фреймворка.
7. Языки поисковых запросов.
8. Управление качеством данных.

9. Варианты построения распределенных баз данных, репликация, фрагментация. Согласованность.
10. CAP-теорема.
11. Классы NoSQL баз данных.
12. Примеры СУБД NoSQL.
13. Семейства столбцов.
14. Графовые СУБД. Neo4j
15. Понятие агрегата.
16. Современные документо-ориентированные СУБД.
17. Запросы к СУБД.
18. Использование фреймворка в документо-ориентированных СУБД.

Зачетное задание включает два вопроса – один теоретический вопрос и одно практико-ориентированное задание из числа приведенных ниже практических заданий.

Критерии оценивания:

- 50-100 баллов («зачет») – изложенный материал фактически верен, наличие глубоких исчерпывающих знаний в объеме пройденной программы дисциплины в соответствии с поставленными программой курса целями и задачами обучения; правильные, уверенные действия по применению полученных знаний на практике, грамотное и логически стройное изложение материала при ответе, усвоение основной и знакомство с дополнительной литературой; наличие твердых и достаточно полных знаний в объеме пройденной программы дисциплины в соответствии с целями обучения, правильные действия по применению знаний на практике, четкое изложение материала, допускаются отдельные логические и стилистические погрешности, обучающийся усвоил основную литературу, рекомендованную в рабочей программе дисциплины; наличие твердых знаний в объеме пройденного курса в соответствии с целями обучения, изложение ответов с отдельными ошибками, уверенно исправленными после дополнительных вопросов; правильные в целом действия по применению знаний на практике;

- 0-49 баллов («незачет») – ответы не связаны с вопросами, наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неумение применять знания на практике, неуверенность и неточность ответов на дополнительные и наводящие вопросы.

Тестовые задания

1. На основе какого языка был создан язык программирования R?
 - a. C
 - b. S
 - c. Java
 - d. Python
2. Выберите верный ответ
 - a. большие данные – это обработка или хранение более 1 Тб информации
 - b. проблема больших данных – это такая проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна
 - c. большие данные – это огромная PR-акция крупных вендоров и не более того
 - d. большие данные – это явление, когда цифровые данные наиболее полно представляют изучаемый объект
3. Чего не могут стандартные реляционные базы (MySQL, Oracle, MS SQL)?
 - a. хранить данные с изменяющейся структурой
 - b. хранить данные на нескольких физических серверах
 - c. хранить большие объемы (терабайты+)
 - d. обрабатывать запросы большого количества клиентов (миллионы)
4. Для чего аналитику необходима "песочница"?
 - a. для высокопроизводительной аналитики за счёт использования оперативной памяти и inDB операций
 - b. для хранения всех полученных от заказчика данных
 - c. для построения отчетов о результатах анализа
 - d. для снижения затрат, связанных с репликацией данных
5. Назовите преимущества языка программирования R, как инструмента data mining, перед другими языками
 - a. быстрдействие

- b. большое кол-во библиотек алгоритмов ИАД и статистических пакетов
- c. удобная работа с данными
- d. работа с большими данными

6. Выберите неверное высказывание:

- a. большие объёмы данных приводят к слабой их структуризации, поэтому появляется такое разнообразие данных
- b. увеличившаяся производительность телекоммуникационных каналов привела к росту объёмов передаваемой информации
- c. удешевление систем хранения на единицу информации привело к росту рынка больших данных

7. Выберите технологии анализа больших данных:

- a. Бионический анализ
- b. Краудсорсинг
- c. Машинное обучение и нейронные сети
- d. Разработка математических алгоритмов
- e. Data Mining
- f. Проектирование аналитических систем

8. Машинное обучение — ...

- a. специализированный программный решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных
- b. эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов
- c. набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.
- d. подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных

9. Обучающая выборка — ...

- a. эта группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов
- b. набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат
- c. выявление в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности

10. Big Data или большие данные — это ...

- a. структурированные массивы данных большого объема. Их обрабатывают при помощи специальных автоматизированных инструментов, чтобы использовать для статистики, анализа, прогнозов и принятия решений.
- b. неструктурированные массивы данных большого объема. Их обрабатывают при помощи специальных автоматизированных инструментов, чтобы использовать для статистики, анализа, прогнозов и принятия решений.
- c. это крупные массивы разнообразной информации и стек специальных технологий для работы с ней.
- d. Все приведенные определения верны.

11. Через какие три v часто дают определение больших данных?

- a. volume (пер. с англ. объем)
- b. visibility (пер. с англ. видимость)
- c. vacant (пер. с англ. свободный)
- d. velocity (пер. с англ. скорость)
- e. variety (пер. с англ. многообразие)

12. Укажите, что из нижеперечисленного является платформой для хранения и обработки больших данных:

- a. NoSQL
- b. MapReduce
- c. Datawrapper
- d. HPC Systems

Критерии оценивания:

Из имеющегося банка тестов формируется тестовое задание, содержащее 10 тестовых вопросов для одного обучающегося. Каждый тестовый вопрос содержит 3-4 варианта ответов, один или несколько из которых – верные.

Правильный ответ на один тестовый вопрос – 1 балл, неправильный – 0 баллов.

Максимальное количество баллов за тестовые задания – 10.

Практические задания

Практическое задание №1

Настройка облачной среды и рабочей области машинного обучения. Создание рабочей области машинного обучения. Отправка существующих данных. Создание нового эксперимента. Обучение и анализ моделей. Создать эксперимент и обучить модель на примере набора данных. Создание кластеров. Библиотека машинного обучения. Программная платформа для написания заданий, обрабатывающих большие объемы данных. Входные данные разбиваются на независимые блоки, которые затем обрабатываются параллельно на узлах кластера.

Практическое задание №2

Графовая СУБД. Используя графовую СУБД, разработать базу данных модельной социальной сети. Узел социальной сети имеет следующие атрибуты: ФИО, пол, возраст, город, список групп по интересам. Семантика связи между узлами – дружеские отношения соответствующих персон («А является другом Б»). Разработать и протестировать запросы на выборку данных из созданной графовой базы данных. Разработать и протестировать запросы на обновление свойств узлов графа и выполнить с их помощью визуализацию графа.

Практическое задание №3

Документ-ориентированная СУБД. Разработать базу данных, предназначенную для хранения логов веб-сервера. Лог включает в себя следующие поля: адрес ресурса (URL), IP-адрес пользовательского компьютера, отметка времени начала просмотра ресурса, длительность просмотра ресурса. Разработать и протестировать запросы на выборку данных из созданных коллекций. Разработать и протестировать функции фреймворка для анализа посещаемости ресурсов web-сервера.

Критерии оценивания (для каждого задания):

25-30 б. – задание выполнено верно;

17-24 б.– при выполнении задания были допущены неточности, не влияющие на результат;

9-16 б. – при выполнении задания были допущены ошибки;

1-8 б. – при выполнении задания были допущены существенные ошибки.

Максимальное количество баллов за все практические задания – 90 (3 задания по 30 баллов).

3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Процедуры оценивания включают в себя текущий контроль и промежуточную аттестацию.

Текущий контроль успеваемости проводится с использованием оценочных средств, представленных в п. 2 данного приложения. Результаты текущего контроля доводятся до сведения студентов до промежуточной аттестации.

Промежуточная аттестация проводится в форме зачета.

Зачет проводится по расписанию промежуточной аттестации. Количество вопросов в зачетном задании – 2 (один теоретический вопрос и одно практико-ориентированное задание). Объявление результатов производится в день зачета. Результаты аттестации заносятся в зачетную ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику сессии, должны ликвидировать задолженность в установленном порядке.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Учебным планом предусмотрены следующие виды занятий:

- лекции;
- практические занятия.

В ходе лекционных занятий рассматриваются основные теоретические вопросы, даются рекомендации для самостоятельной работы и подготовке к практическим занятиям.

В ходе практических занятий углубляются и закрепляются знания студентов по ряду рассмотренных на лекциях вопросов, развиваются навыки практической работы.

При подготовке к практическим занятиям каждый студент должен:

- изучить рекомендованную учебную литературу;
- изучить конспекты лекций;
- подготовить ответы на все вопросы по изучаемой теме.

В процессе подготовки к практическим занятиям студенты могут воспользоваться консультациями преподавателя.

Вопросы, не рассмотренные на лекциях и практических занятиях, должны быть изучены студентами в ходе самостоятельной работы. Контроль самостоятельной работы студентов над учебной программой курса осуществляется в ходе занятий методом выполнения тестовых и практических заданий. В ходе самостоятельной работы каждый студент обязан прочитать основную и по возможности дополнительную литературу по изучаемой теме, дополнить конспекты лекций недостающим материалом, выписками из рекомендованных первоисточников, выделить непонятные термины, найти их значение в энциклопедических словарях.

Студент должен готовиться к предстоящему практическому занятию по всем обозначенным в рабочей программе дисциплины вопросам.

Для подготовки к занятиям, текущему контролю и промежуточной аттестации студенты могут воспользоваться электронно-библиотечными системами. Также обучающиеся могут взять на дом необходимую литературу на абонементе университетской библиотеки или воспользоваться читальными залами.