Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Максимов Алексей Борисович
Должность: директор департамента по образовательной политике
Дата подписания: 23.05.2024 11:29:32
Уникальный программный ключ:
8db180d1a3f02ac9e60521a5672742735c18b1d6

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение
высшего образования
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
Факультет информационных технологий

УТВЕРЖДЕНО
Декан факультета
Информационных технологий
_____/ Д.Г. Демидов /
«15» февраля 2024 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Хранилища данных»

Направление подготовки/специальность
09.03.03 Прикладная информатика

Профиль/специализация
Большие и открытые данные

Квалификация
бакалавр

Формы обучения
очная

Москва, 2024 г.

Разработчик(и):

К.т.н., доцент. / В.С. Ноздрин /

Согласовано:
Заведующий кафедрой «Прикладной информатики»,
К.э.н, доцент / С.В. Суворов /

Содержание

Оглавление

1 Цели, задачи и планируемые результаты обучения по дисциплине 4
2 Место дисциплины в структуре образовательной программы 5
3 Структура и содержание дисциплины 5
3.1 Виды учебной работы и трудоемкость 5
3.2 Тематический план изучения дисциплины 6
3.3 Содержание дисциплины 7
3.4 Тематика семинарских/практических и лабораторных занятий 7
3.5 Тематика курсовых проектов (курсовых работ) 8
4 Учебно-методическое и информационное обеспечение 8
4.1 Нормативные документы и ГОСТы 8
4.2 Основная литература 8
4.3 Дополнительная литература 8
4.4 Электронные образовательные ресурсы 8
4.5 Лицензионное и свободно распространяемое программное обеспечение 8
4.6 Современные профессиональные базы данных и информационные справочные системы 8
5 Материально-техническое обеспечение 9
5.1 Требования к оборудованию и помещению для занятий 9
5.2 Требования к программному обеспечению 9
6 Методические рекомендации 9
6.1 Методические рекомендации для преподавателя по организации обучения 9
6.2 Методические указания для обучающихся по освоению дисциплины 9
7 Фонд оценочных средств 10
7.1 Методы контроля и оценивания результатов обучения 10
7.2 Шкала и критерии оценивания результатов обучения 10
7.3 Оценочные средства 12

1 Цели, задачи и планируемые результаты обучения по дисциплине

Цель дисциплины:

- расширенное формирование у студентов представления о принципах и методах машинного обучения;
- знакомство студентов с современными методами работы с большими данными;
- закрепление получаемых в семестре знаний и навыков на практике;
- формирование взаимосвязей, получаемых в семестре знаний и навыков с изученными ранее и изучаемых параллельно с данной дисциплиной;
- подготовка студентов к деятельности в соответствии с квалификационной характеристикой бакалавра.

Задачи дисциплины:

- освоение методологии обработки больших данных;
- использование компьютерных технологий реализации методов машинного обучения;
- изучение и освоение теоретического материала, как в процессе контактной, так и в ходе самостоятельной работы;
- выполнение предоставленных практических заданий различных форм, как в процессе контактной, так и в ходе самостоятельной работы;
- самостоятельная работа над тематикой дисциплины для формирования компетенций основной образовательной программы (далее, ООП).

Планируемые результаты обучения

- знакомство студентов с современными методами работы с большими данными;
- закрепление получаемых в семестре знаний и навыков на практике;
- формирование взаимосвязей, получаемых в семестре знаний и навыков с изученными ранее и изучаемых параллельно с данной дисциплиной;
- подготовка студентов к деятельности в соответствии с квалификационной характеристикой бакалавра.

Обучение по дисциплине «Хранилища данных» направлено на формирование у обучающихся следующих компетенций:

Код и наименование компетенций	**Индикаторы достижения компетенции**
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.	**Знать**: Принципы сбора, отбора и обобщения информации. Методики системного подхода для решения профессиональных задач. **Уметь**: Анализировать и систематизировать разнородные данные.

	Оценивать эффективность процедур анализа проблем и принятия решений в профессиональной деятельности. **Владеть**: Навыками научного поиска и практической работы с информационными источниками. Методами принятия решений.
УК-2. Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений.	**Знать**: Необходимые для осуществления профессиональной деятельности правовые нормы. Методологические основы принятия управленческого решения. **Уметь**: Анализировать альтернативные варианты решений для достижения намеченных результатов. Разрабатывать план, определять целевые этапы и основные направления работ. **Владеть**: Методиками разработки целей и задач проекта. Методами оценки продолжительности и стоимости проекта. Методами оценки потребности в ресурсах.

2 Место дисциплины в структуре образовательной программы

Дисциплина относится к части, формируемой участниками образовательных отношений блока Б1 «Дисциплины (модули) Большие открытые данные».

Она взаимосвязана логически и содержательно-методически со следующими дисциплинами ООП:

- Основы баз данных;
- Проектирование и разработка баз данных;
- Методы и технологии обработки больших данных.

3 Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет _5_ зачетных(е) единиц(ы) (_180_ часов).

3.1 Виды учебной работы и трудоемкость (по формам обучения)

3.1.1 Очная форма обучения

№ п/п	Вид учебной работы	Количество часов	Семестры	
			7	
1	**Аудиторные занятия**	**72**	72	
	В том числе:			
1.1	Лекции	18	18	
1.2	Семинарские/практические занятия			
1.3	Лабораторные занятия	54	54	
2	**Самостоятельная работа**	**108**		
	В том числе:			

2.1	СРС	108	108	
3	**Промежуточная аттестация**			
	Зачет/диф.зачет/экзамен	**Зачет**	Зачет	
	Итого:	**180**	180	

3.2 Тематический план изучения дисциплины (по формам обучения)

3.2.1 Очная форма обучения

№ п/п	Разделы/темы дисциплины	Трудоемкость, час					
		Всего	Аудиторная работа				Самостоятельная работа
			Лекции	Семинарские/ практические занятия	Лабораторные занятия	Практическая подготовка	
1	Раздел 1.						
1.1	Тема 1. Pig и Hive	20	2		6		12
1.2	Тема 2. Введение в Data Vault	20	2		6		12
1.3	Тема 3. Git и GitHub	20	2		6		12
1.4	Тема 4. Документные СУБД (MongoDb)	20	2		6		12
1.5	Тема 5. Графовые базы данных (Neo4j)	20	2		6		12
1.6	Тема 6. Обучение каскада хаара OpenCV Метод Виолы-Джонса	20	2		6		12
1.7	Тема 7. Многомерные хранилища данных OLAP	20	2		6		12
1.8	Тема 8. Архитектура OLAP проекта	20	2		6		12

1.9	Тема 9. Язык доступа к многомерным данным MDX	20	2		6		12
Итого		**180**	**18**		**54**		**108**

3.3 Содержание дисциплины

Раздел 1.

Тема 1. Pig и Hive

Основные возможности PIG. Компоненты PIG. Концепция HIVE.

Тема 2. Введение в Data Vault

Хабы, сателлиты, ссылки, преимущества и недостатки.

Тема 3. Git и GitHub

Система управления версиями, подготовка репозитория.

Тема 4. Документные СУБД (MongoDb)

Настройка, конфигурирование, функциональность, команды.

Тема 5. Графовые базы данных (Neo4j)

Сферы применения, особенности, язык запросов.

Тема 6. Обучение каскада хаара OpenCV Метод Виолы-Джонса

Основы компьютерного зрения. Распознавание образов. Обучение каскада Хаара.

Тема 7. Многомерные хранилища данных OLAP

Основные понятия. Основные требования. Как реализовать OLAP?

Тема 8. Архитектура OLAP проекта

Логическая архитектура. Компоненты таблицы фактов. Проектирование измерений. Измерения и иерархии.

Тема 9. Язык доступа к многомерным данным MDX

Основы языка, назначение, терминология MDX, MDX-выражения, функции и свойства.

3.4 Тематика семинарских/практических и лабораторных занятий

3.4.1 Семинарские/практические занятия

3.4.2 Лабораторные занятия

Лабораторная работа №1 Pig & Hive

Лабораторная работа №2 Реализация структуры Data Vault

Лабораторная работа №3. Git и GitHub

Лабораторная работа №4 MongoDb

Лабораторная работа №5 по Neo4j

Лабораторная работа №6 Организация хранилища данных для датасета по компьютерному зрению

Лабораторная работа №7 OLAP-кубы в Excel
Лабораторная работа №8 Моделирование хранилища данных
Лабораторная работа №9 OLAP на MS SQL

3.5 Тематика курсовых проектов (курсовых работ)

4 Учебно-методическое и информационное обеспечение

4.1 Нормативные документы и ГОСТы

1. Федеральный государственный образовательный стандарт высшего образования (уровень бакалавриата) по направлению подготовки 09.03.03 «Прикладная информатика», утвержденный приказом Министерства образования и науки Российской Федерации от 19.09.2017 №922 «Об утверждении федерального государственного стандарта высшего образования – бакалавриат по направлению подготовки 09.03.03 Прикладная информатика. https://fgos.ru/fgos/fgos-09-03-03-prikladnaya-informatika-922/

4.2 Основная литература

1. Дэви Силен, Арно Мейсман Основы Data Science и Big Data. Python и наука о данных. СПб, Питер, 2017. – 336 стр.
2. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям [Электронный ресурс], СПб, Питер, 2013. – 704 стр. Режим доступа: http://www.padaread.com/?book=52105

4.3 Дополнительная литература

1. Заенцев И.В. Нейронные сети: основные модели [Электронный ресурс], Воронеж, 1999, - 76 стр. Режим доступа: http://computersbooks.net/books/web-programmirovanie/zaencev-iv/1999/files/neyronnieseti1999.pdf

4.4 Электронные образовательные ресурсы

1. https://lms.mospolytech.ru/course/view.php?id=2241

4.5 Лицензионное и свободно распространяемое программное обеспечение

1. Операционная система, Windows 11 (или ниже) - Microsoft Open License
2. Офисные приложения, Microsoft Office 2013(или ниже) - Microsoft Open License

4.6 Современные профессиональные базы данных и информационные справочные системы

1.не предусмотрено

5 Материально-техническое обеспечение

5.1 Требования к оборудованию и помещению для занятий

Лабораторные работы и самостоятельная работа студентов должны проводиться в специализированной аудитории, оснащенной современной оргтехникой и персональными компьютерами с программным обеспечением в соответствии с тематикой изучаемого материала. Число рабочих мест в аудитории должно быть достаточным для обеспечения индивидуальной работы студентов. Рабочее место преподавателя должно быть оснащены современным компьютером с подключенным к нему проектором на настенный экран, или иным аналогичным по функциональному назначению оборудованием.

5.2 Требования к программному обеспечению

Для выполнения лабораторных работ и самостоятельной работы необходимо следующее программное обеспечение:

Microsoft Windows.

Веб-браузер, Chrome.

ПО, предоставленное преподавателем.

6 Методические рекомендации

6.1 Методические рекомендации для преподавателя по организации обучения

1. При подготовке к занятиям следует предварительно проработать материал занятия, предусмотрев его подачу точно в отведенное для этого время занятия. Следует подготовить необходимые материалы – теоретические сведения, задачи и др. При проведении занятия следует контролировать подачу материала и решение заданий с учетом учебного времени, отведенного для занятия.

2. При проверке работ и отчетов следует учитывать не только правильность выполнения заданий, но и оптимальность выбранных методов решения, правильность выполнения всех его шагов.

6.2 Методические указания для обучающихся по освоению дисциплины

Изучение дисциплины осуществляется в строгом соответствии с целевой установкой в тесной взаимосвязи учебным планом. Основой теоретической подготовки студентов являются аудиторные занятия, семинары и практики.

В процессе самостоятельной работы студенты закрепляют и углубляют знания, полученные во время аудиторных занятий, дорабатывают конспекты и записи, готовятся к промежуточной аттестации, а также самостоятельно изучают отдельные темы учебной программы.

На занятиях студентов, в том числе предполагающих практическую деятельность, осуществляется закрепление полученных, в том числе и в процессе самостоятельной работы, знаний. Особое внимание обращается на развитие умений и навыков установления связи положений теории с профессиональной деятельностью будущего специалиста.

Самостоятельная работа осуществляется индивидуально. Контроль самостоятельной работы организуется в двух формах:

самоконтроль и самооценка студента;

контроль со стороны преподавателей (текущий и промежуточный).

Текущий контроль осуществляется на аудиторных занятиях, промежуточный контроль осуществляется на зачете в письменной (устной) форме.

Критериями оценки результатов самостоятельной работы студента являются:

уровень освоения студентом учебного материала;

умения студента использовать теоретические знания при выполнении практических задач;

сформированность компетенций;

оформление материала в соответствии с требованиями..

7 Фонд оценочных средств

7.1 Методы контроля и оценивания результатов обучения

В процессе обучения используются следующие оценочные формы самостоятельной работы студентов, оценочные средства текущего контроля успеваемости и промежуточных аттестаций:

Лабораторные работы, зачет.

7.2 Шкала и критерии оценивания результатов обучения

Показателем оценивания компетенций на различных этапах их формирования является достижение обучающимися планируемых результатов обучения по дисциплине.

Показатель	**Критерии оценивания**			
	2	**3**	**4**	**5**
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.				
Знать: Принципы сбора, отбора и обобщения информации. Методики системного подхода для решения профессиональных задач. **Уметь**: Анализировать и систематизировать разнородные данные. Оценивать эффективность процедур анализа проблем и принятия решений в профессиональной деятельности. **Владеть**: Навыками научного поиска и практической	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие материалу дисциплины знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3).	Обучающийся демонстрирует неполное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся испытывает значительные затруднения при оперировании	Обучающийся демонстрирует частичное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	Обучающийся демонстрирует полное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Свободно оперирует приобретенными знаниями.

работы с информационными источниками. Методами принятия решений.		знаниями при их переносе на новые ситуации.		
УК-2. Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений.				
Знать: Необходимые для осуществления профессиональной деятельности правовые нормы. Методологические основы принятия управленческого решения. **Уметь**: Анализировать альтернативные варианты решений для достижения намеченных результатов. Разрабатывать план, определять целевые этапы и основные направления работ. **Владеть**: Методиками разработки целей и задач проекта. Методами оценки продолжительности и стоимости проекта. Методами оценки потребности в ресурсах.	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие материалу дисциплины знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3).	Обучающийся демонстрирует неполное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.	Обучающийся демонстрирует частичное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	Обучающийся демонстрирует полное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины «Знать» (см. п. 3). Свободно оперирует приобретенными знаниями.

Шкала оценивания результатов промежуточной аттестации и её описание:

Форма промежуточной аттестации: зачет.

Промежуточная аттестация обучающихся в форме зачёта проводится по результатам выполнения всех видов учебной работы, предусмотренных учебным планом по данной дисциплине (модулю), при этом учитываются результаты текущего контроля успеваемости в

течение семестра. Оценка степени достижения обучающимися планируемых результатов обучения по дисциплине (модулю) проводится преподавателем, ведущим занятия по дисциплине (модулю) методом экспертной оценки. По итогам промежуточной аттестации по дисциплине (модулю) выставляется оценка «зачтено» или «не зачтено».

Шкала оценивания	**Описание**
Зачтено	Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, умений, навыков приведенным в таблицах показателей, оперирует приобретенными знаниями, умениями, навыками, применяет их в ситуациях повышенной сложности. При этом могут быть допущены незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
Не зачтено	Не выполнен один или более видов учебной работы, предусмотренных учебным планом. Студент демонстрирует неполное соответствие знаний, умений, навыков приведенным в таблицах показателей, допускаются значительные ошибки, проявляется отсутствие знаний, умений, навыков по ряду показателей, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.

7.3 Оценочные средства

7.3.1 Текущий контроль

Подготовка и защита лабораторных работ

7.3.2 Промежуточная аттестация

Вопросы на зачет

1. Область применения хранилищ данных и больших данных и их преимущества
2. Виды данных. Структурированные данные
3. Виды данных. Неструктурированные данные
4. Виды данных. Данные на естественном языке
5. Виды данных. Машинные данные
6. Виды данных. Графовые, или сетевые, данные
7. Виды данных. Аудио, видео и графика
8. Виды данных. Потоковые данные
9. Понятие гиперкуба. OLAP.
10. ROLAP-системы
11. MOLAP-системы
12. HOLAP-системы
13. Распределенные файловые системы

14. Базы данных NoSQL
15. OLAP – системы
16. Операции в OLAP-системах
17. Реализация OLAP –систем
18. Файл MOLAP
19. Схема «звезда» в ROLAP
20. Схема «снежинка» в ROLAP
21. Задачи Data Mining
22. Основные методы в Data Mining
23. Область применения Data Mining
24. Логическая архитектура OLAP Server: объекты
25. Планирование OLAP куба
26. Проектирование измерений. Свойства атрибутов измерений
27. Проектирование измерений. Ключи атрибутов
28. Проектирование измерений. Медленно меняющиеся измерения
29. Проектирование измерений. Измерения с пропущенными данными
30. Меняющиеся измерения. Динамические измерения (calculated members)
31. Меняющиеся измерения. Медленно меняющиеся измерения (Slowly Changed Dimensions - SHD)
32. Меняющиеся измерения. Деформированные измерения (ragged dimensions)
33. Разработка физического уровня куба: Секции куба.
34. Разработка физического уровня куба: Тип хранения данных (Storage Mode) - ROLAP, MOLAP,HOLAP.
35. Разработка физического уровня куба: Упреждающее кеширование (Proactive Caching).
36. Разработка физического уровня куба: Удаленные (remote) секции.
37. Назначение MDX
38. Терминология MDX: термины, ось, элемент, многомерные координаты.
39. Терминология MDX: ячейка, кортеж, slice, dice, множество.
40. MDX: ссылки на объекты и уникальные имена
41. Предложение SELECT в MDX: SELECT: Полный кортеж
42. Предложение SELECT в MDX: SELECT: Неполный кортеж
43. Предложение SELECT в MDX: SELECT с ограничением WHERE
44. Выражения MDX: простые и сложные
45. Выражения MDX: Выражения измерений и иерархий. Выражения элементов.
46. Выражения MDX: Выражения наборов. Выражения кубов
47. MDX-функции элементов
48. MDX: Функции наборов
49. MDX: Функции наборов – навигация
50. MDX: Функции наборов – сортировка
51. MDX: Функции наборов – фильтр
52. MDX: Строковые функции. Методы преобразования
53. MDX: Логические функции
54. MDX: Числовые функции – агрегация
55. MDX: Числовые функции. Статистика
56. Внутренние свойства элементов, зависящие и не зависящие от контекста
57. MDX: Свойства ячеек
58. Вычисляемые элементы (Calculated Members):
59. MDX сценарии с заданной областью видимости (Scoped MDX Scripts)
60. Развертывание OLAP проекта

61. Оптимизация производительности в кубах
62. Управление отчетами. Доставка отчетов
63. Что такое Data Mining
64. Data Mining: для каких данных?
65. Функциональность Data Mining
66. Классификация систем Data Mining
67. Задачи Data Mining
68. Интеграция Data Mining с СУБД и OLAP
69. Основные моменты в Data Mining
70. Data Mining задачи
71. Data Mining объекты
72. Data Mining процесс разработки
73. Планирование проекта Data Mining
74. Подготовка данных для Data Mining
75. Выбор Data Mining алгоритма
76. Использование Data Mining в Business Intelligence