



1920

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ**

филиал Федерального государственного бюджетного образовательного
учреждения высшего образования

«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

в г. Новороссийске

Кафедра информатики и математики

УТВЕРЖДАЮ

Проректор по работе с филиалами

ФГОУ ВО «Кубанский
государственный университет»

А.А.Евдокимов



01.07 2019 г.

РАБОЧАЯ ПРОГРАММА

Б1.О.36 УГЛУБЛЕННЫЙ АНАЛИЗ ДАННЫХ И BIG DATA

Направление подготовки: 01.03.02 Прикладная математика и информатика

Направленность (профиль): Математическое и информационное обеспечение
экономической деятельности

Программа подготовки: академическая

Форма обучения: очная

Квалификация (степень) выпускника: Бакалавр

Краснодар 2019

Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 01.03.02 Прикладная математика и информатика, утвержденного приказом Министерства образования и науки Российской Федерации № 9 от 10 января 2018 года.

Программу составил(и):

И.Г.Рзун , доцент канд.физ.-мат.наук



С.В. Дьяченко доцент канд.физ.-мат.наук



Рабочая программа дисциплины Углубленный анализ данных и Big Data обсуждена и утверждена на заседании кафедры Информатики и математики протокол № 11 от 20.06. 2019 г.

Заведующий кафедрой (выпускающей) Рзун И.Г.



Рабочая программа одобрена на заседании учебно-методической комиссии филиала УГС 01.00.00 «Математика и механика»
21.06.2019 г. протокол № 10

Председатель УМК



С.В. Дьяченко

Рецензенты:

Кунина М.К. Директор по развитию ООО «АЙТИ БИЗНЕС ЮГ»

Адамович А.Е. Директор ООО «Финам - Новороссийск»

Содержание рабочей программы дисциплины

1. Цели и задачи изучения дисциплины.....	4
1.1 Цель освоения дисциплины.....	4
1.2 Задачи дисциплины.....	4
1.3 Место дисциплины в структуре образовательной программы.....	4
1.4 Перечень планируемых результатов обучения по дисциплине	4
2. Структура и содержание дисциплины	6
2.1 Распределение трудоёмкости дисциплины по видам работ.....	6
2.2 Структура дисциплины:.....	7
2.3 Содержание разделов дисциплины:	8
2.3.1 Занятия лекционного типа	8
2.3.2 Занятия практического типа	11
2.3.3 Лабораторные занятия	11
2.3.4 Примерная тематика курсовых работ (проектов).....	13
2.4 Перечень учебно-методического обеспечения для самостоятельной работы.....	13
3. Образовательные технологии	14
4. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации	15
4.1 Фонд оценочных средств для проведения текущей аттестации.....	15
4.2 Фонд оценочных средств для проведения промежуточной аттестации	22
5. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	24
5.1 Основная литература:	24
5.2 Дополнительная литература:	Ошибка! Закладка не определена.
5.3. Периодические издания:	24
7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	25
7. Методические указания для обучающихся по освоению дисциплины	25
8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (при необходимости)	29
8.1 Перечень информационных технологий.....	29
8.2 Перечень необходимого программного обеспечения.....	29
8.3 Перечень информационных справочных систем:	29
9. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине	29

1. Цели и задачи изучения дисциплины

1.1 Цель освоения дисциплины

Целью дисциплины является изучение основ современных баз данных в объеме, необходимом для самостоятельной работы с базами данных и для освоения дисциплин, связанных с анализом, проектированием, разработкой и сопровождением корпоративных информационных систем.

1.2 Задачи дисциплины

Развитие навыков системного подхода к информационным системам, освоение основных моделей данных (реляционной, иерархической, объектно-реляционной и реляционной) и их отображений, изучение языков предназначенных для работы с реляционными, иерархическими и объектными базами данных, понимание проблематики хранилищ данных, представление о направлениях развития баз данных.

1.3 Место дисциплины в структуре образовательной программы

Дисциплина «Углубленный анализ данных и Big Data» относится к основной части учебного плана. Данный курс наиболее тесно связан с курсами:

- Дискретная математика;
- Основы программирования;
- Анализ, проектирование и разработка БД;
- Администрирование БД;
- Объектно - ориентированное программирование;
- Формализмы представления знаний.

1.4 Перечень планируемых результатов обучения по дисциплине

Изучение данной учебной дисциплины направлено на формирование у обучающихся профессиональных компетенций: ОПК-1; ПК-1

№ п.п.	Индекс компетен ции	Содержание компетенции (или её части)	В результате изучения учебной дисциплины обучающиеся должны		
			знать	уметь	владеть
	ОПК-1	Способен решать актуальные и значимые задачи прикладной математики и информатики	- понятие информации; - основные положения теории информации и кодирования; - общую характеристику процессов сбора, передачи, обработки и накопления информации; - технические и программные средства реализации информационных процессов; - современное состояние и направления развития	- работать в качестве пользователя персонального компьютера; - самостоятельно использовать внешние носители информации для обмена данными между машинами; - создавать резервные копии и архивы данных и программ; - работать с программными средствами общего назначения, соответствующими современным требованиям мирового рынка; - использовать	- навыками подготовки сложных иллюстрированных текстовых документов с использованием MS Word; - навыками решения расчетных экономических задач с применением MS Excel; - навыками создания и обработки реляционных баз данных средствами MS Access; - навыками

№ п.п.	Индекс компетен ции	Содержание компетенции (или её части)	В результате изучения учебной дисциплины обучающиеся должны		
			знать	уметь	владеть
			вычислительной техники и программных средств; - закономерности протекания информационных процессов в системах обработки информации; - принципы использования современных информационных технологий и инструментальных средств для решения различных задач в своей профессиональной деятельности; - основы защиты информации и сведений, составляющих государственную тайну; - методы обеспечения информационной безопасности экономического субъекта.	информационные системы и средства вычислительной техники в решении задач сбора, передачи, хранения и обработки экономической информации; - формулировать требования и принимать обоснованные решения по выбору аппаратно-программных средств для рационального решения задач, связанных с получением и преобразованием информации; - использовать в профессиональной деятельности сетевые средства поиска и обмена информацией.	подготовки электронных презентаций с использованием MS PowerPoint. - методами решения экономических задач с помощью специализированных программных продуктов; - навыками автоматизации решения экономических задач; - технологиями работы в локальных и глобальных информационных сетях; - приемами антивирусной защиты; - навыками работы с программами автоматизации бухгалтерского учета.
	ПК-1	Способен применять фундаментальные знания, полученные в области математических и (или) естественных наук, и использовать их в профессиональ	современный уровень развития прикладной математики и информационных технологий; источники данных о современных научных исследованиях.	проводить научные исследования с использованием новейших математических и информационных достижений, собирать, обрабатывать данные современных научных исследований,	информацией о перспективах развития современных математических теорий и информационных технологий, навыками участия в работе научных семинаров,

№ п.п.	Индекс компетен ции	Содержание компетенции (или её части)	В результате изучения учебной дисциплины обучающиеся должны		
			знать	уметь	владеть
		ной деятельности		необходимые для формирования выводов по соответствующим научным, профессиональным проблемам, использовать современные достижения в своей профессиональной деятельности, изучать новые научные результаты, научную литературу и научно- исследовательские проекты в соответствии с профилем объекта профессиональной деятельности, исследовать и разрабатывать математические модели, алгоритмы, методы, программное обеспечение, инструментальные средства по тематике проводимых научно- исследовательских проектов, составлять научные обзоры, рефераты и библиографии по тематике проводимых исследований.	научно- тематических конференций, симпозиумов; навыками подготовки научных и научно- технических публикаций.

2. Структура и содержание дисциплины

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 4 зач.ед. (144 часов), их распределение по видам работ представлено в таблице (для студентов ОФО).

Вид учебной работы	Всего часов	Семестры (часы)
		8
Контактная работа, в том числе:	42,3	42,3

Аудиторные занятия (всего):		42	42
Занятия лекционного типа		14	14
Лабораторные занятия		28	28
Занятия семинарского типа (семинары, практические занятия)			
Иная контактная работа:			
Контроль самостоятельной работы (КСР)			
Промежуточная аттестация (ИКР)		0,3	0,3
Самостоятельная работа, в том числе:		48	48
Курсовая работа			
Проработка учебного (теоретического) материала		26	26
Выполнение индивидуальных заданий		22	22
Реферат			
Подготовка к текущему контролю			
Контроль: экзамен		53,7	53,7
Подготовка к экзамену			
Общая трудоемкость	час.	144	144
	в том числе контактная работа	66,5	42,3
	зач. ед	4	4

Курсовые не предусмотрены

2.2 Структура дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы дисциплины, изучаемые в 8 семестре (для студентов ОФО)

№ разд ела	Наименование разделов	Количество часов						
		Всего	Контактная работа				Контр оль	Самост оятельн ая работа
			Л	ЛР	КСР	ИК Р		
1	2	3	4	5	6	7	8	9
1	Data science в мире больших данных	14	2	4				8
2	Процесс data science	14	2	4				8
3	Машинное обучение	14	2	4				8
4	Работа с большими данными на одном компьютере	16	2	6				8
5	Распределение хранения и обработки данных в инфраструктурах	16	4	4				8
6	Глубокий анализ текста	16	2	6				8
	Итого по дисциплине :	90	14	28				48
	Промежуточная аттестация (ИКР)	0,3				0,3		
	<i>Контроль</i>	53,7					53,7	
	<i>Всего:</i>	144	14	28		0,3	53,7	48

2.3 Содержание разделов дисциплины:

Раздел 1. Область применения data science и больших данных и их преимущества. Грани данных. Структурированные данные. Неструктурированные данные. Данные на естественном языке. Машинные данные. Графовые, или сетевые, данные. Аудио, видео и графика. Поточные данные. Распределенные файловые системы. Инфраструктура распределенного программирования. Инфраструктура интеграции данных. Инфраструктуры машинного обучения. Базы данных NoSQL. Инструменты планирования. Инструменты сравнительного анализа. Развертывание системы. Программирование служб. Безопасность. ОПК-1; ПК-1

Раздел 2. Обзор процесса data science. Определение целей исследования и создание проектного задания. Создание проектного задания. Сбор данных. Данные, хранимые в компании. Проверка качества данных. Очистка, интеграция и преобразование данных. Исправление ошибок. Комбинирование данных из разных источников.

Преобразование данных ОПК-1; ПК-1

Раздел 3. Применение машинного обучения в data science. Применение машинного обучения в процессе data science. Инструменты Python, используемые в машинном обучении. Процесс моделирования. Создание новых показателей и выбор модели. Тренировка модели. Проверка адекватности модели. Прогнозирование новых наблюдений. Типы машинного обучения. Контролируемое обучение. Неконтролируемое обучение. Частично контролируемое обучение ОПК-1; ПК-1

Раздел 4. Проблемы при работе с большими объемами данных. Общие методы обработки больших объемов данных. Правильный выбор алгоритма. Правильный выбор структуры данных. Правильный выбор инструментов. Общие рекомендации для программистов при работе с большими наборами данных. Не повторяйте уже выполненную работу. Используйте все возможности оборудования. Экономьте вычислительные ресурсы. Пример1: Прогнозирование вредоносных URL-адресов. Определение цели исследования. Сбор данных URL. Исследование данных. Построение модели. ОПК-1; ПК-1

Раздел 5. Распределение хранения и обработки данных в инфраструктурах. Hadoop: инфраструктура для хранения и обработки больших объемов данных. Spark: замена MapReduce с повышенной производительностью. Учебный пример: Оценка риска при кредитовании. Этап 1: Цель исследования. Этап 2: Сбор данных. Этап 3: Подготовка данных. Этап 4: Исследование данных и Этап 6: построение отчета ОПК-1; ПК-1

Раздел 6. Глубокий анализ текста в реальном мире. Методы глубокого анализа текста. Набор слов. Выделение основы и лемматизация. Классификатор на базе дерева принятия решений. Учебный пример: классификация сообщений Reddit. NLTK. Обзор процесса data science и этап 1: назначение цели исследования. Этап 2: Сбор данных. Этап 3: Подготовка данных. Этап 4: Исследование данных. Этап 3 (повторно): Подготовка данных (адаптированная) Этап 5: Анализ данных. Этап 6: Отображение и автоматизация. ОПК-1; ПК-1

2.3.1 Занятия лекционного типа

№	Наименование раздела	Содержание раздела	Форма текущего контроля
1	2	3	4
1	Data science в мире больших данных	Область применения data science и больших данных и их преимущества. Грани данных. Структурированные данные. Неструктурированные данные. Данные на естественном языке. Машинные данные. Графовые, или сетевые, данные. Аудио, видео	Вопросы для устного опроса

		и графика. Поточковые данные . Распределенные файловые системы. Инфраструктура распределенного программирования. Инфраструктура интеграции данных. Инфраструктуры машинного обучения. Базы данных NoSQL. Инструменты планирования. Инструменты сравнительного анализа. Развертывание системы. Программирование служб. Безопасность	
2	Процесс data science	Обзор процесса data science. Определение целей исследования и создание проектного задания. Создание проектного задания. Сбор данных. Данные, хранимые в компании. Проверка качества данных. Очистка, интеграция и преобразование данных. Исправление ошибок. Комбинирование данных из разных источников. Преобразование данных	Вопросы для устного опроса
3	Машинное обучение	Применение машинного обучения в data science. Применение машинного обучения в процессе data science. Инструменты Python, используемые в машинном обучении. Процесс моделирования. Создание новых показателей и выбор модели. Тренировка модели. Проверка адекватности модели. Прогнозирование новых наблюдений. Типы машинного обучения. Контролируемое обучение. Неконтролируемое обучение. Частично контролируемое обучение	Вопросы для устного опроса
4	Работа с большими данными на одном компьютере	Проблемы при работе с большими объемами данных. Общие методы обработки больших объемов данных. Правильный выбор алгоритма. Правильный выбор структуры данных. Правильный выбор инструментов. Общие рекомендации для программистов при работе с большими наборами данных. Не повторяйте уже выполненную работу. Используйте все возможности оборудования. Экономьте вычислительные ресурсы. Пример1: Прогнозирование вредоносных URL-адресов. Определение цели исследования. Сбор	Вопросы для устного опроса

		данных URL. Исследование данных. Построение модели.	
5	Распределение хранения и обработки данных в инфраструктурах	Распределение хранения и обработки данных в инфраструктурах. Hadoop: инфраструктура для хранения и обработки больших объемов данных. Spark: замена MapReduce с повышенной производительностью. Учебный пример: Оценка риска при кредитовании. Этап 1: Цель исследования. Этап 2: Сбор данных. Этап 3: Подготовка данных. Этап 4: Исследование данных и Этап 6: построение отчета	Вопросы для устного опроса
6	Глубокий анализ текста	Глубокий анализ текста в реальном мире. Методы глубокого анализа текста. Набор слов. Выделение основы и лемматизация. Классификатор на базе дерева принятия решений. Учебный пример: классификация сообщений Reddit. NLTK. Обзор процесса data science и этап 1: назначение цели исследования. Этап 2: Сбор данных. Этап 3: Подготовка данных. Этап 4: Исследование данных. Этап 3 (повторно): Подготовка данных (адаптированная) Этап 5: Анализ данных. Этап 6: Отображение и автоматизация.	Вопросы для устного опроса

Вопросы для устного опроса

1. Область применения data science и больших данных и их преимущества. Грани данных.
2. Структурированные данные. Неструктурированные данные. Данные на естественном языке.
3. Машинные данные. Графовые, или сетевые, данные. Аудио, видео и графика. Поточковые данные .
4. Распределенные файловые системы. Инфраструктура распределенного программирования. Инфраструктура интеграции данных. Инфраструктуры машинного обучения.
5. Базы данных NoSQL. Инструменты планирования. Инструменты сравнительного анализа. Развертывание системы. Программирование служб. Безопасность
6. Обзор процесса data science. Определение целей исследования и создание проектного задания.
7. Создание проектного задания. Сбор данных. Данные, хранимые в компании. Проверка качества данных.
8. Очистка, интеграция и преобразование данных. Исправление ошибок. Комбинирование данных из разных источников. Преобразование данных
9. Применение машинного обучения в data science. Применение машинного обучения в процессе data science.
10. Инструменты Python, используемые в машинном обучении. Процесс моделирования.
11. Создание новых показателей и выбор модели. Тренировка модели. Проверка адекватности модели. Прогнозирование новых наблюдений.

12. Типы машинного обучения. Контролируемое обучение. Неконтролируемое обучение. Частично контролируемое обучение
13. Проблемы при работе с большими объемами данных. Общие методы обработки больших объемов данных.
14. Правильный выбор алгоритма. Правильный выбор структуры данных. Правильный выбор инструментов.
15. Общие рекомендации для программистов при работе с большими наборами данных.
16. Пример1: Прогнозирование вредоносных URL-адресов. Определение цели исследования. Сбор данных URL. Исследование данных. Построение модели.
17. Распределение хранения и обработки данных в инфраструктурах. Hadoop: инфраструктура для хранения и обработки больших объемов данных.
18. Spark: замена MapReduce с повышенной производительностью.
19. Учебный пример: Оценка риска при кредитовании. Этап 1: Цель исследования. Этап 2: Сбор данных. Этап 3: Подготовка данных. Этап 4: Исследование данных и Этап 6: построение отчета
20. Глубокий анализ текста в реальном мире. Методы глубокого анализа текста.
21. Набор слов. Выделение основы и лемматизация. Классификатор на базе дерева принятия решений.
22. Учебный пример: классификация сообщений Reddit. NLTK. Обзор процесса data science и этап 1: назначение цели исследования. Этап 2: Сбор данных. Этап 3: Подготовка данных.
23. Этап 4: Исследование данных. Этап 3 (повторно): Подготовка данных (адаптированная) Этап 5: Анализ данных. Этап 6: Отображение и автоматизация.

2.3.2 Занятия практического типа

Практические занятия - не предусмотрены

2.3.3 Лабораторные занятия

№	Наименование раздела	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
1	Data science в мире больших данных	Область применения data science и больших данных и их преимущества. Грани данных. Структурированные данные. Неструктурированные данные. Данные на естественном языке. Машинные данные. Графовые, или сетевые, данные. Аудио, видео и графика. Поточковые данные. Распределенные файловые системы. Инфраструктура распределенного программирования. Инфраструктура интеграции данных. Инфраструктуры машинного обучения. Базы данных NoSQL. Инструменты планирования. Инструменты сравнительного анализа. Развертывание системы. Программирование служб. Безопасность	лабораторная

2	Процесс data science	Обзор процесса data science. Определение целей исследования и создание проектного задания. Создание проектного задания. Сбор данных. Данные, хранимые в компании. Проверка качества данных. Очистка, интеграция и преобразование данных. Исправление ошибок. Комбинирование данных из разных источников. Преобразование данных	лабораторная
3	Машинное обучение	Применение машинного обучения в data science. Применение машинного обучения в процессе data science. Инструменты Python, используемые в машинном обучении. Процесс моделирования. Создание новых показателей и выбор модели. Тренировка модели. Проверка адекватности модели. Прогнозирование новых наблюдений. Типы машинного обучения. Контролируемое обучение. Неконтролируемое обучение. Частично контролируемое обучение	лабораторная
4	Работа с большими данными на одном компьютере	Проблемы при работе с большими объемами данных. Общие методы обработки больших объемов данных. Правильный выбор алгоритма. Правильный выбор структуры данных. Правильный выбор инструментов. Общие рекомендации для программистов при работе с большими наборами данных. Не повторяйте уже выполненную работу. Используйте все возможности оборудования. Экономьте вычислительные ресурсы. Пример1: Прогнозирование вредоносных URL-адресов. Определение цели исследования. Сбор данных URL. Исследование данных. Построение модели.	лабораторная
5	Распределение хранения и обработки данных в инфраструктурах	Распределение хранения и обработки данных в инфраструктурах. Hadoop: инфраструктура для хранения	лабораторная

		и обработки больших объемов данных. Spark: замена MapReduce с повышенной производительностью. Учебный пример: Оценка риска при кредитовании. Этап 1: Цель исследования. Этап 2: Сбор данных. Этап 3: Подготовка данных. Этап 4: Исследование данных и Этап 6: построение отчета	
6	Глубокий анализ текста	Глубокий анализ текста в реальном мире. Методы глубокого анализа текста. Набор слов. Выделение основы и лемматизация. Классификатор на базе дерева принятия решений. Учебный пример: классификация сообщений Reddit. NLTK. Обзор процесса data science и этап 1: назначение цели исследования. Этап 2: Сбор данных. Этап 3: Подготовка данных. Этап 4: Исследование данных. Этап 3 (повторно): Подготовка данных (адаптированная) Этап 5: Анализ данных. Этап 6: Отображение и автоматизация.	лабораторная

2.3.4 Примерная тематика курсовых работ (проектов)

Курсовая работа не предусмотрена

2.4 Перечень учебно-методического обеспечения для самостоятельной работы

№	Наименование раздела	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1.	Проработка учебного (теоретического) материала	Базы данных. Практическое применение СУБД SQL и NoSQL-типа для применения проектирования информационных систем: Учебное пособие / Мартишин С.А., Симонов В.Л., Храпченко М.В. - М.: ИД ФОРУМ, НИЦ ИНФРА-М, 2017. [Электронный ресурс] - URL: http://znanium.com/catalog.php?bookinfo=556449 , 05.10.2017. Советов, Б. Я. Базы данных : учебник для прикладного бакалавриата / Б. Я. Советов, В. В. Цехановский, В. Д. Чертовской. — 2-е изд. — М. : Издательство Юрайт, 2017. [Электронный ресурс] - URL: https://www.biblio-online.ru/viewer/502697C3-F440-4628-B9B8-28E18BCB4337#/, 05.10.2017.
2.	Выполнение	Базы данных. Практическое применение СУБД SQL и

индивидуальных заданий	NoSQL-типа для применения проектирования информационных систем: Учебное пособие / Мартишин С.А., Симонов В.Л., Храпченко М.В. - М.:ИД ФОРУМ, НИЦ ИНФРА-М, 2017. [Электронный ресурс] - URL:http://znanium.com/catalog.php?bookinfo=556449, 05.10.2017. Советов, Б. Я. Базы данных : учебник для прикладного бакалавриата / Б. Я. Советов, В. В. Цехановский, В. Д. Чертовской. — 2-е изд. — М. : Издательство Юрайт, 2017. [Электронный ресурс] - URL:https://www.biblio-online.ru/viewer/502697C3-F440-4628-B9B8-28E18BCB4337#/, 05.10.2017.
------------------------	---

При изучении дисциплины «Углубленный анализ данных и Big Data» обязательными являются следующие формы самостоятельной работы:

- разбор теоретического материала по пособиям, конспектам лекций;
- самостоятельное изучение указанных теоретических вопросов;
- решение задач по темам занятий;
- выполнение домашней контрольной работы;
- подготовка к экзамену

Согласно письма Министерства образования и науки РФ № МОН-25486 от 21.06.2017г «О разработке адаптированных образовательных программ» -Разработка адаптивной программы необходима в случае наличия в образовательной организации хотя бы одного обучающегося с ограниченными возможностями здоровья.

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,
- в форме аудиофайла,
- в печатной форме на языке Брайля.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,
- в форме аудиофайла.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. Образовательные технологии

С точки зрения применяемых методов используются как традиционные информационно-объяснительные лекции, так и интерактивная подача материала с мультимедийной системой. Компьютерные технологии в данном случае обеспечивают возможность разнопланового отображения алгоритмов и демонстрационного материала. Такое сочетание позволяет оптимально использовать отведенное время и раскрывать логику и содержание дисциплины.

Лекции представляют собой систематические обзоры основных аспектов дисциплины.

Лабораторные занятия позволяют научиться применять теоретические знания при решении и исследовании конкретных задач. Лабораторные занятия проводятся в компьютерных классах, при этом практикуется работа в группах. Подход разбора конкретных ситуаций широко используется как преподавателем, так и студентами при проведении анализа результатов самостоятельной работы. Это обусловлено тем, что в процессе исследования часто встречаются задачи, для которых единых подходов не существует. Каждая конкретная задача при своем исследовании имеет множество подходов, а это требует разбора и оценки целой совокупности конкретных ситуаций.

При освоении дисциплины используются следующие сочетания видов учебной работы с методами и формами активизации познавательной деятельности бакалавров для достижения запланированных результатов обучения и формирования компетенций.

В соответствии с требованиями ФГОС ВО по направлению подготовки бакалавров программа по дисциплине «Углубленный анализ данных и Big Data» предусматривает использование в учебном процессе следующих образовательных технологий: чтение лекций с использованием мультимедийных технологий; разбор конкретных ситуаций.

Компьютерные технологии позволяют проводить сравнительный анализ научных исследований по данной проблеме, являясь средством разнопланового отображения алгоритмов и демонстрационного материала.

Подход разбора конкретных ситуаций широко используется как преподавателем, так и бакалаврами во время лекций и анализа результатов самостоятельной работы. Это обусловлено тем, что в процессе работы с базами данных часто встречаются задачи, для которых единых подходов не существует. Каждая конкретная задача при своем исследовании имеет множество подходов, а это требует разбора и оценки целой совокупности конкретных ситуаций. Особенно этот подход широко используется при определении адекватности математической модели, результатам компьютерных экспериментов.

Цель *лекции* – обзор понятий баз данных.

Цель *лабораторного занятия* – научиться применять теоретические знания при решении и исследовании конкретных задач.

Темы, задания и вопросы для самостоятельной работы призваны сформировать навыки поиска информации, умения самостоятельно расширять и углублять знания, полученные в ходе лекционных и практических занятий.

4. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

4.1 Фонд оценочных средств для проведения текущей аттестации

Учебная деятельность проходит в соответствии с графиком учебного процесса. Процесс самостоятельной работы контролируется во время аудиторных занятий и индивидуальных консультаций. Самостоятельная работа студентов проводится в форме изучения отдельных теоретических вопросов по предлагаемой литературе и решения индивидуальных задач повышенной сложности.

Фонд оценочных средств дисциплины состоит из средств текущего контроля (см. примерные варианты контрольных работ, индивидуальных заданий, задач и вопросов) и итоговой аттестации (экзамена).

В качестве оценочных средств, используемых для текущего контроля успеваемости, предлагается перечень вопросов, которые прорабатываются в процессе освоения курса. Данный перечень охватывает все основные разделы курса, включая знания, получаемые во время самостоятельной работы.

Оценка успеваемости осуществляется по результатам: самостоятельного

выполнения лабораторных работ, устного опроса при сдаче выполненных самостоятельных заданий, ответов на экзамене.

Аттестация по учебной дисциплине проводится в виде экзамена. Экзаменационный билет содержит два теоретических вопроса и одну задачу. Студент готовит ответы на билет в письменной форме в течение установленного времени. Далее экзамен протекает в форме собеседования.

Примерное содержание самостоятельных и лабораторных работ

В качестве методического обеспечения для аудиторной и внеаудиторной самостоятельной работы приводятся варианты заданий, позволяющие студентам ориентироваться в необходимом уровне требований при подготовке к занятиям.

1. Пример варианта РАР (включено в учебный план дисциплины).

Разработать модель предметной области (полная атрибутивная модель в нотации IDEF в системе ERWin) и структуру представления информации (таблицы, поля, тип данных и связи) для информационной поддержки деятельности предприятия, которое занимается прокладкой компьютерных сетей и разработкой программных комплексов для организаций.

Набор данных приведен в таблице:

№	Поле	Описание
1	EmployeeID	Идентификатор работника
2	EmployeeName	ФИО работника
3	Address	Домашний адрес
4	District	Район города
5	Experience	Опыт работы по специальности
6	Year	Год рождения
7	Language	Базовый язык программирования
8	Base	Образование (вуз)
9	Comment	Примечания
10	Salary	Зарплата
11	Bonus	Премия
12	GrossSalary	Полная зарплата
13	Exempt	Льготы
14	Picture	Фото работника
15	ProjectID	Идентификатор проекта
16	ProjectName	Название проекта
17	ProjectStart	Дата начала проекта
18	ProjectStop	Дата окончания проекта
19	Chief	Руководитель проекта
20	Customer	Заказчик проекта
21	Cost	Стоимость разработки
22	Phone	Телефон заказчика
23	Bank	Банк заказчика
24	Account	Номер счета в банке
25	INN	ИНН заказчика
26	AddressCust	Адрес заказчика
27	FioWorker	Ответственный от заказчика
28	PhoneWorker	Телефон ответственного
29	BonusAll	Премия, %, при досрочном выполнении
30	EmployeeStart	Начало участия работника в проекте

При разработке структуры представления информации учесть приведенные описания деятельности компании:

- Одновременно находится в разработке несколько проектов.
- Один разработчик может участвовать в нескольких проектах одновременно, но зарплата его от этого не зависит.
- Одна организация может заказать несколько разработок.
- Стоимость каждого проекта оговаривается отдельно.
- При досрочном выполнении работы заказчик перечисляет предприятию определенный, заранее оговоренный процент премии.

2. Пример типового индивидуального задания (операция эквивалентного соединения отношений).

Даны два отношения:

Отношение R

R.a1 R.a2

X 1

X 2

Y 1

Y 3

Y 4

Отношение S

S.b1 S.b2

1 p

2 g

3 p

5 h

Выполнить Левое внешнее соединение и показать результат.

3. Пример типового индивидуального задания (создание таблицы с рекурсивными связями и запрос на выбор информации).

Написать скрипт на языке SQL для создания таблицы (Люди) с рекурсивными связями. Таблица содержит информацию о людях, между которыми имеются родственные связи (КодЧеловека, Фамилия, КодСупруга, КодМатери, КодОтца). КодСупруга, КодМатери, КодОтца – внешние ключи.

Используя созданную таблицу «Люди» с рекурсивной связью, написать запрос на языке SQL для получения в результате соответствия между человеком, его супругом/супругой (вывести только людей, которые имеют супруга/супругу).

4. Пример типового индивидуального задания (создание таблиц со связями и ограничениями).

База данных содержит таблицы:

Товар (Кодтовара, Названиетовара, Цена)

Клиент (Кодклиента, Фамилия, Датарождения, Город)

Заказ (Кодзаказа, Кодтовара, Кодклиента, Количествотовара, ДатаЗаказа).

Написать скрипт на языке SQL для создания таблиц и связей базы данных. В скрипте предусмотреть:

- Выбор соответствующих типов данных для полей таблиц;
- Создание первичных ключей типа «Счетчик» для таблиц Товар и Клиент;
- Задание значения по умолчанию для поля ДатаЗаказа, равное текущей дате;

- Задание составного первичного ключа таблицы Заказ по полям Кодтовара, Кодзаказа;
- Задать проверку вводимых значений поля Количествотовара в диапазоне от 0 до 50;
- Задать условие удаления в таблице Заказ товара, который удаляется из таблицы Товар.

5. Пример типового индивидуального задания (запрос с использованием соединения таблиц).

База данных содержит таблицы:

Товар (Кодтовара, Названиетовара, Цена)

Поставщик (Кодпоставщика, Фамилия, Датарождения, Город)

Наскладе (Кодтовара, Кодпоставщика, Количествотовара, ДатаПоставки).

Написать запрос на языке SQL для вывода информации по Поставщикам, которые никогда не делали поставок на склад. Вывести Фамилия.

6. Пример типового индивидуального задания (запрос с использованием соединения таблиц).

База данных содержит таблицы:

Товар (Кодтовара, Названиетовара, Цена)

Клиент (Кодклиента, Фамилия, Датарождения, Город)

Заказ (Кодтовара, Кодклиента, Количествотовара, ДатаЗаказа).

Написать запрос на языке SQL для вывода информации по заказам, произведенным после задаваемой даты, без использования в запросе ключевого слова join. Вывести Названиетовара, Цена, Фамилия, Город, Количествотовара, ДатаЗаказа.

7. Пример типового индивидуального задания (создание и использование представлений).

Создать модифицируемое представление на основе запроса. С помощью оператора INSERT INTO добавить в представление записи. Проверить выполнение оператора INSERT INTO при использовании параметра WITH CHECK OPTION. Для изменения параметра WITH CHECK OPTION использовать оператор изменения представления ALTER VIEW. Выбрать поля из модифицированного представления.

8. Пример типового индивидуального задания (разработка структуры хранилища данных).

Имеется информация для анализа Продаж, которая включает информационные компоненты «Время», «Категория товара», «Товар», «Регион», «Продавец», «Покупатель», «Сумма», «Количество».

1. Разработайте структуру хранилища данных для этой информации (измерения, 2-3 атрибута для таблиц измерений, показатели).

2. Для двух измерений предложите свой вариант иерархии измерения (не менее трех уровней).

3. Нарисуйте логическую структуру хранилища данных, используя каноническую схему «Снежинка» для измерений с предложенными иерархиями (п. 2).

9. Пример типового индивидуального задания (оптимизация выполнения запроса).

Провести анализ влияния селективности на план выполнения запроса. Получить два различных плана выполнения запроса для условия, обеспечивающего высокую и низкую селективность.

10. Пример типового индивидуального задания (транзакции и управление параллельной работой).

Провести исследование функционирования уровня изоляции READ UNCOMMITTED.

11. Пример типового индивидуального задания (Data Mining).

Построить дерево решений для анализа покупателей заданного товара (файл исходных данных «Покупатели»). Варьировать используемые методы построения дерева решений и параметр разбиения узла. Провести интерпретацию полученных результатов.

12. Пример типового индивидуального задания (Data Mining).

Для одного из вариантов модели (дерево решений для анализа покупателей заданного товара) проверить правильность прогнозирования для заданного набора строк данных (взять подмножество исходных данных). Использовать меню «Использование модели». Провести оценку точности модели.

Самостоятельное изучение разделов дисциплины

Самостоятельная работа студентов является неотъемлемой частью процесса подготовки. Под самостоятельной работой понимается часть учебной планируемой работы, которая выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия.

Самостоятельная работа направлена на усвоение системы научных и профессиональных знаний, формирования умений и навыков, приобретение опыта самостоятельной творческой деятельности. Самостоятельная работа помогает формировать культуру мышления, расширять интеллектуальный потенциал магистрантов.

Виды самостоятельной работы по дисциплине:

- 1) по целям:
 - подготовка к лекциям,
 - к практическим занятиям,
- 2) по характеру работы:
 - изучение литературы,
 - конспекта лекций;
 - поиск литературы в библиотеке;

работа с обучающими и контролирующими программами.

Самостоятельная работа

Целью самостоятельной работы студента является углубление знаний, полученных в результате аудиторных занятий. Вырабатываются навыки самостоятельной работы. Закрепляются опыт и знания полученные во время лабораторных занятий.

Для контроля знаний периодически проводятся аудиторные самостоятельные работы.

Вопросы для самоподготовки:

1. Область применения data science и больших данных и их преимущества. Грани данных.
2. Структурированные данные. Неструктурированные данные. Данные на естественном языке.
3. Машинные данные. Графовые, или сетевые, данные. Аудио, видео и графика. Поточковые данные .
4. Распределенные файловые системы. Инфраструктура распределенного программирования. Инфраструктура интеграции данных. Инфраструктуры машинного обучения.
5. Базы данных NoSQL. Инструменты планирования. Инструменты сравнительного анализа. Развертывание системы. Программирование служб. Безопасность

6. Обзор процесса data science. Определение целей исследования и создание проектного задания.
7. Создание проектного задания. Сбор данных. Данные, хранимые в компании. Проверка качества данных.
8. Очистка, интеграция и преобразование данных. Исправление ошибок. Комбинирование данных из разных источников. Преобразование данных
9. Применение машинного обучения в data science. Применение машинного обучения в процессе data science.
10. Инструменты Python, используемые в машинном обучении. Процесс моделирования.
11. Создание новых показателей и выбор модели. Тренировка модели. Проверка адекватности модели. Прогнозирование новых наблюдений.
12. Типы машинного обучения. Контролируемое обучение. Неконтролируемое обучение. Частично контролируемое обучение
13. Проблемы при работе с большими объемами данных. Общие методы обработки больших объемов данных.
14. Правильный выбор алгоритма. Правильный выбор структуры данных. Правильный выбор инструментов.
15. Общие рекомендации для программистов при работе с большими наборами данных.
16. Пример1: Прогнозирование вредоносных URL-адресов. Определение цели исследования. Сбор данных URL. Исследование данных. Построение модели.
17. Распределение хранения и обработки данных в инфраструктурах. Hadoop: инфраструктура для хранения и обработки больших объемов данных.
18. Spark: замена MapReduce с повышенной производительностью.
19. Учебный пример: Оценка риска при кредитовании. Этап 1: Цель исследования. Этап 2: Сбор данных. Этап 3: Подготовка данных. Этап 4: Исследование данных и Этап 6: построение отчета
20. Глубокий анализ текста в реальном мире. Методы глубокого анализа текста.
21. Набор слов. Выделение основы и лемматизация. Классификатор на базе дерева принятия решений.
22. Учебный пример: классификация сообщений Reddit. NLTK. Обзор процесса data science и этап 1: назначение цели исследования. Этап 2: Сбор данных. Этап 3: Подготовка данных.
23. Этап 4: Исследование данных. Этап 3 (повторно): Подготовка данных (адаптированная) Этап 5: Анализ данных. Этап 6: Отображение и автоматизация.

Пример варианта структуры данных прикладной области (Банк):

Таблицы:

- Сотрудники (Код сотрудника, ФИО, Возраст, Пол, Адрес, Телефон, Паспортные данные, Код должности).
- Должности (Код должности, Наименование должности, Оклад, Обязанности, Требования).
- Вклады (Код вклада, Наименование вклада, Минимальный срок вклада, Минимальная сумма вклада, Код валюты, Процентная ставка, Дополнительные условия).
- Валюта (Код валюты, Наименование, Обменный курс).
- Вкладчики (ФИО вкладчика, Адрес, Телефон, Паспортные данные, Дата вклада, Дата возврата, Код вклада, Сумма вклада, Сумма возврата, Отметка о возврате вклада, Код сотрудника).

Примеры индивидуальных заданий:

1. Создать в БД необходимые таблицы согласно варианту с помощью команд CREATE TABLE, определить типы таблиц (родительская или подчиненная), типы полей и их размеры, поля типа Primary key и Foreign key. При создании таблиц использовать

- свойство `IDENTITY`, задать вычисляемый столбец, заполнение значением по умолчанию, принадлежность значений от до, невозможность удаления строки из родительской таблицы.
2. Используя оператор `INSERT` заполнить таблицы записями. Создать варианты использования оператора `INSERT` (использовать или не использовать список столбцов, признаки `NULL` и `NOT NULL` для столбцов, заполнение по умолчанию, столбцы типа «счетчик», вычисляемые столбцы).
 3. Выбор нескольких полей таблицы (таблиц). Задать условия отбора для полей с числовыми значениями, с данными типа «дата» и с символьными данными, объединив их соответствующими логическими операциями. Задание сортировки по одному из полей. Использовать вычисляемое поле (арифметическое или строковое выражения). Для вычисляемого поля задать имя.
 4. Выбор нескольких полей таблицы (таблиц). Выбор записей с использованием агрегатных функций для числовых или символьных полей с заданием группировки (при необходимости изменить записи в таблицах для получения групп).
 5. Выбор нескольких полей таблицы (таблиц). Выбор записей с использованием агрегатных функций для числовых или символьных полей с заданием группировки и с использованием условия обработки сформированных групп (`HAVING`) (записи в таблицах должны обеспечить получение групп).
 6. Выбор нескольких полей таблиц. Выполнить условие соединения трех таблиц с использованием и без использования оператора `JOIN`. В запросе использовать краткое обозначение таблиц (синонимы или псевдонимы).
 7. Выбор нескольких полей двух таблиц. Выполнить условие левого или правого внешнего соединения.
 8. Выбор нескольких полей таблицы (таблиц) с использованием подзапроса, возвращающего единственное значение. Выполнить запрос для данных, обеспечивающих корректное выполнение запроса, для данных, когда подзапрос возвращает несколько значений, для данных, когда скалярный подзапрос не возвращает ни одного значения.
 9. Добавить в таблицу записи, используя выбранные записи по условию из другой таблицы.
 10. Создать модифицируемое представление. С помощью оператора `INSERT INTO` добавить в представление записи. Проверить выполнение оператора `INSERT INTO` при использовании параметра `WITH CHECK OPTION`. Для изменения параметра `WITH CHECK OPTION` использовать оператор изменения представления `ALTER VIEW`. Выбрать поля из модифицированного представления.
 11. На базе имеющихся запросов (код SQL запросов нужно изменить таким образом, чтобы в них можно было передавать значения полей, по которым осуществляется поиск) создать:
 - Процедуру без параметров.
 - Процедуру с входными параметрами
 - Процедуру с входными параметрами и выходным параметромНаписать примеры обращений к процедурам.
 12. Провести анализ влияния селективности на план выполнения.
 13. Для последовательности запросов на изменение данных задать именованную транзакцию и две точки сохранения в ней. Выполнить откат к одной из точек сохранения. Произвести фиксацию транзакции. Состояние данных после фиксации транзакции должно соответствовать начальному состоянию.
 14. Анализ данных методами Data Mining в SQL Server с использованием клиентского средства Excel. Провести подготовку исходных данных (файл исходных данных

Клиенты), создать структуру анализа данных. Для созданной структуры анализа данных создать модели, используя Алгоритм кластеризации. Создать варианты модели, используя альтернативные методы кластеризации, задав разбиение на 2 кластера. Сравнить характеристики полученных кластеров, определить параметры, которые сильно влияют на состав кластеров. Создать варианты моделей, удалив слабо влияющие параметры. Сравнить характеристики полученных кластеров.

4.2 Фонд оценочных средств для проведения промежуточной аттестации

Фонд оценочных средств дисциплины (ФОС) состоит из средств входного контроля знаний по дисциплине, текущего контроля выполнения заданий и средств для промежуточной аттестации:

1. контрольные вопросы;
2. коллоквиум;
3. лабораторные занятия;

Эти средства содержат перечень:

- вопросов, ответы на которые дают возможность студенту продемонстрировать, а преподавателю оценить степень усвоения теоретических и фактических знаний на уровне знакомства;
- заданий, позволяющих оценить приобретенные студентами практические умения на репродуктивном уровне.

Оценка успеваемости специалистов осуществляется по результатам:

- самостоятельного выполнения лабораторной работы,
- взаимного рецензирования работ друг друга,
- анализа подготовленных рефератов,
- устного опроса при сдаче выполненных индивидуальных заданий, защите отчетов по лабораторным работам для выявления знания и понимания теоретического материала дисциплины.

Перечень вопросов к экзамену

1. Область применения data science и больших данных и их преимущества. Грани данных.
2. Структурированные данные. Неструктурированные данные. Данные на естественном языке.
3. Машинные данные. Графовые, или сетевые, данные. Аудио, видео и графика. Поточные данные.
4. Распределенные файловые системы. Инфраструктура распределенного программирования. Инфраструктура интеграции данных. Инфраструктуры машинного обучения.
5. Базы данных NoSQL. Инструменты планирования. Инструменты сравнительного анализа. Развертывание системы. Программирование служб. Безопасность
6. Обзор процесса data science. Определение целей исследования и создание проектного задания.
7. Создание проектного задания. Сбор данных. Данные, хранимые в компании. Проверка качества данных.
8. Очистка, интеграция и преобразование данных. Исправление ошибок. Комбинирование данных из разных источников. Преобразование данных
9. Применение машинного обучения в data science. Применение машинного обучения в процессе data science.
10. Инструменты Python, используемые в машинном обучении. Процесс моделирования.
11. Создание новых показателей и выбор модели. Тренировка модели. Проверка адекватности модели. Прогнозирование новых наблюдений.

12. Типы машинного обучения. Контролируемое обучение. Неконтролируемое обучение. Частично контролируемое обучение
13. Проблемы при работе с большими объемами данных. Общие методы обработки больших объемов данных.
14. Правильный выбор алгоритма. Правильный выбор структуры данных. Правильный выбор инструментов.
15. Общие рекомендации для программистов при работе с большими наборами данных.
16. Пример1: Прогнозирование вредоносных URL-адресов. Определение цели исследования. Сбор данных URL. Исследование данных. Построение модели.
17. Распределение хранения и обработки данных в инфраструктурах. Hadoop: инфраструктура для хранения и обработки больших объемов данных.
18. Spark: замена MapReduce с повышенной производительностью.
19. Учебный пример: Оценка риска при кредитовании. Этап 1: Цель исследования. Этап 2: Сбор данных. Этап 3: Подготовка данных. Этап 4: Исследование данных и Этап 6: построение отчета
20. Глубокий анализ текста в реальном мире. Методы глубокого анализа текста.
21. Набор слов. Выделение основы и лемматизация. Классификатор на базе дерева принятия решений.
22. Учебный пример: классификация сообщений Reddit. NLTK. Обзор процесса data science и этап 1: назначение цели исследования. Этап 2: Сбор данных. Этап 3: Подготовка данных.
23. Этап 4: Исследование данных. Этап 3 (повторно): Подготовка данных (адаптированная) Этап 5: Анализ данных. Этап 6: Отображение и автоматизация.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

- при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;
- при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;
- при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

5. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

5.1 Основная литература:

1. Анализ данных : учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2019. — 490 с. — (Бакалавр. Академический курс). — ISBN 978-5-534-00616-2. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://biblio-online.ru/bcode/432178> (дата обращения: 12.09.2019).
2. Статистика [Электронный ресурс]: учебник и практикум для академического бакалавриата / В. С. Мхитарян [и др.] ; под ред. В. С. Мхитаряна. — М. : Издательство Юрайт, 2017. — 464 с. - URL: <https://biblio-online.ru/viewer/8B223896-5381-4624-B8AB-F179B8E4C027#page/1>
3. Статистика [Текст] : учебник и практикум для академического бакалавриата : учебник для студентов вузов, обучающихся по экономическим направлениям и специальностям : учебник для студентов вузов, обучающихся по направлению "Статистика" и другим экономическим специальностям / [В. С. Мхитарян и др.] ; под ред. В. С. Мхитаряна ; Нац. исслед. ун-т "Высшая школа экономики". - Москва : Юрайт, 2016. - 464 с. **5**
4. Васильев, А. А. Теория вероятностей и математическая статистика : учебник и практикум для академического бакалавриата / А. А. Васильев. — 2-е изд., испр. и доп. — М. : Издательство Юрайт, 2017. — 253 с. — (Серия : Бакалавр. Академический курс). ЭБС. URL: <https://biblio-online.ru/viewer/3F13A609-9D28-44A2-A070-1A025A293A4F#page/1>
5. Анализ данных : учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под ред. В. С. Мхитаряна. — М. : Издательство Юрайт, 2017. — 490 с. — (Серия : Бакалавр. Академический курс). — ISBN 978-5-534-00616-2. ЭБС. URL <https://biblio-online.ru/viewer/CC38E97A-CCE5-4470-90F1-3B6D35ACC0B4#page/1>
6. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — М. : Издательство Юрайт, 2017. — 174 с. — (Серия : Авторский учебник). — ISBN 978-5-534-03762-3. ЭБС. URL <https://www.biblio-online.ru/viewer/46A41F93-BC46-401C-A30E-27C0FB60B9DE#page/1>

5.2 Дополнительная литература:

1. Гобарева, Яна Львовна. Бизнес-аналитика средствами Excel [Текст] : учебное пособие для студентов, обучающихся по направлению "Экономика" (квалификация (степень) - бакалавр) / Я. Л. Гобарева, О. Ю. Городецкая, А. В. Золотарюк ; Финансовый ун-т при Правительстве Рос. Федерации. - 2-е изд., испр. и доп. - Москва : Вузовский учебник : ИНФРА-М, 2015. - 335 с.
2. Коваленко, Анна Владимировна (КубГУ). Математические основы финансово-экономического анализа [Текст] : учебное пособие . Ч. 1 : Многомерный статистический анализ / А. В. Коваленко, М. Х. Уртенев, У. А. Узденов . - Москва : Academia, 2010. - 303 с.
3. Коваленко, Анна Владимировна (КубГУ). Многомерный статистический анализ предприятий [Текст] : монография / А. В. Коваленко, М. Х. Уртенев, У. А. Узденов. - Москва : [ACADEMIA], 2009. - 240 с.
4. Халафян, Алексан Альбертович (КубГУ). STATISTICA 6. Статистический анализ данных [Текст] : учебное пособие для студентов вузов / А. А. Халафян. - [3-е изд.]. - М. : [БИНОМ-Пресс], 2008. - 508 с.

5.3. Периодические издания:

1. "Алгебра и логика" / Институт математики им.Соболева СО РАН /Периодичность – 6 раз в год

1. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

№	Наименование электронного ресурса	Ссылка на электронный адрес
1.	Электронный каталог Научной библиотеки КубГУ	https://www.kubsu.ru/
2.	Электронная библиотечная система «BOOK.ru» ООО «КноРус медиа»	https://www.book.ru
3.	Электронная библиотечная система "Университетская библиотека ONLINE"	www.biblioclub.ru
4.	Электронная библиотечная система «ZNANIUM.COM» ООО «ЗНАНИУМ»	www.znanium.com
5.	Электронная библиотечная система издательства "Лань"	http://e.lanbook.com/
6.	Электронная библиотечная система "Юрайт"	http://www.biblio-online.ru

1. Российское образование, федеральный портал [Официальный сайт] — URL: <http://www.edu.ru>

2. Образовательный портал «Учеба» [Официальный сайт] URL: <http://www.ucheba.com/>

3. Портал «Российское образование» [Официальный сайт] URL: <http://www.edu.ru/>

4. Единое окно доступа к образовательным ресурсам «Единое окно» [Официальный сайт] URL: <http://window.edu.ru/>

5. Федеральная университетская компьютерная сеть России [Официальный сайт] URL: <http://www.runnet.ru/>

6. Служба тематических толковых словарей [Официальный сайт] URL: <http://www.glossary.ru/>

7. Образовательный портал [Официальный сайт] URL: «Академик» <http://dic.academic.ru/>

8. Web of Sciense (архив с 2002 года) рефераты [Официальный сайт] URL: <http://webofknowledge.com>.

9. Лекториум “(Минобрнауки РФ) единая Интернет-библиотека лекций [Официальный сайт] URL <http://www.lektorium.tv/>

10. Электронный архив документов КубГУ полнотекстов [Официальный сайт] URL: <http://docspace.kubsu.ru>

10.Электронная библиотечная система "Юрайт". URL:<http://www.biblio-online.ru>

7. Методические указания для обучающихся по освоению дисциплины

Согласно письма Министерства образования и науки РФ № МОН-25486 от 21.06.2017г «О разработке адаптированных образовательных программ» -Разработка адаптивной программы необходима в случае наличия в образовательной организации хотя бы одного обучающегося с ограниченными возможностями здоровья.

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

Система обучения основывается на рациональном сочетании нескольких видов учебных занятий (в первую очередь, лекций и практических (лабораторных) занятий), работа на которых обладает определенной спецификой.

Подготовка к лекциям.

Знакомство с дисциплиной происходит уже на первой лекции, где от требуется не просто внимание, но и самостоятельное оформление конспекта. Конспектирование лекций – сложный вид аудиторной работы, предполагающий интенсивную умственную деятельность студента. Конспект является полезным тогда, когда записано самое существенное. Не надо стремиться записать дословно всю лекцию. Такое «конспектирование» приносит больше вреда, чем пользы. Целесообразно вначале понять основную мысль, излагаемую лектором, а затем записать ее. Желательно запись осуществлять на одной странице листа или оставляя поля, на которых позднее, при самостоятельной работе с конспектом, можно сделать дополнительные записи, отметить непонятные места.

Конспект лекции лучше подразделять на пункты, соблюдая красную строку. Этому в большой степени будут способствовать вопросы плана лекции, предложенные преподавателям. Следует обращать внимание на акценты, выводы, которые делает лектор, отмечая наиболее важные моменты в лекционном материале замечаниями «важно», «хорошо запомнить» и т.п. Можно делать это и с помощью разноцветных маркеров или ручек, подчеркивая термины и определения.

Работая над конспектом лекций, Вам всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, кропотливая работа с лекционным материалом позволит глубоко овладеть теоретическим материалом.

Подготовка к практическим (лабораторным) занятиям.

Подготовку к каждому практическому занятию необходимо начать с ознакомления с планом практического занятия, который отражает содержание предложенной темы. Тщательное продумывание и изучение вопросов плана основывается на проработке текущего материала лекции, а затем изучения обязательной и дополнительной литературы, рекомендованной к данной теме. Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса.

Подготовка к лабораторным занятиям и практикумам носит различный характер, как по содержанию, так и по сложности исполнения. Проведение прямых и косвенных измерений предполагает детальное знание измерительных приборов, их возможностей, умение вносить своевременные поправки для получения более точных результатов. Многие лабораторные занятия требуют большой исследовательской работы, изучения дополнительной научной литературы.

В процессе подготовки к практическим занятиям, необходимо обратить особое внимание на самостоятельное изучение рекомендованной литературы. При всей полноте конспектирования лекции в ней невозможно изложить весь материал. Поэтому самостоятельная работа с учебниками, учебными пособиями, научной, справочной литературой, материалами периодических изданий и Интернета является наиболее эффективным методом получения дополнительных знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала.

Защита лабораторных работ должна происходить, как правило, в часы, отведенные на лабораторные занятия. Студент может быть допущен к следующей лабораторной работе только в том случае, если у него не защищено не более двух предыдущих работ.

Рекомендации по работе с литературой.

Работу с литературой целесообразно начать с изучения общих работ по теме, а также учебников и учебных пособий. Далее рекомендуется перейти к анализу монографий и статей, рассматривающих отдельные аспекты проблем, изучаемых в рамках курса, а также официальных материалов и неопубликованных документов

(научно-исследовательские работы, диссертации), в которых могут содержаться основные вопросы изучаемой проблемы.

Работу с источниками надо начинать с ознакомительного чтения, т.е. просмотреть текст, выделяя его структурные единицы. При ознакомительном чтении закладками отмечаются те страницы, которые требуют более внимательного изучения.

В зависимости от результатов ознакомительного чтения выбирается дальнейший способ работы с источником. Если для разрешения поставленной задачи требуется изучение некоторых фрагментов текста, то используется метод выборочного чтения. Если в книге нет подробного оглавления, следует обратить внимание ученика на предметные и именные указатели.

Избранные фрагменты или весь текст (если он целиком имеет отношение к теме) требуют вдумчивого, неторопливого чтения с «мысленной проработкой» материала. Такое чтение предполагает выделение: 1) главного в тексте; 2) основных аргументов; 3) выводов. Особое внимание следует обратить на то, вытекает тезис из аргументов или нет.

Необходимо также проанализировать, какие из утверждений автора носят проблематичный, гипотетический характер, и уловить скрытые вопросы.

Понятно, что умение таким образом работать с текстом приходит далеко не сразу. Наилучший способ научиться выделять главное в тексте, улавливать проблематичный характер утверждений, давать оценку авторской позиции – это сравнительное чтение, в ходе которого Вы знакомитесь с различными мнениями по одному и тому же вопросу, сравниваете весомость и доказательность аргументов сторон и делаете вывод о наибольшей убедительности той или иной позиции.

Если в литературе встречаются разные точки зрения по тому или иному вопросу из-за сложности прошедших событий и правовых явлений, нельзя их отвергать, не разобравшись. При наличии расхождений между авторами необходимо найти рациональное зерно у каждого из них, что позволит глубже усвоить предмет изучения и более критично оценивать изучаемые вопросы. Знакомясь с особыми позициями авторов, нужно определять их схожие суждения, аргументы, выводы, а затем сравнивать их между собой и применять из них ту, которая более убедительна.

Следующим этапом работы с литературными источниками является создание конспектов, фиксирующих основные тезисы и аргументы..

Таким образом, при работе с источниками и литературой важно уметь:

- сопоставлять, сравнивать, классифицировать, группировать, систематизировать информацию в соответствии с определенной учебной задачей;
- обобщать полученную информацию, оценивать прослушанное и прочитанное;
- фиксировать основное содержание сообщений; формулировать, устно и письменно, основную идею сообщения; составлять план, формулировать тезисы;
- готовить и презентовать развернутые сообщения типа доклада;
- работать в разных режимах (индивидуально, в паре, в группе), взаимодействуя друг с другом;
- пользоваться реферативными и справочными материалами;
- контролировать свои действия и действия своих товарищей, объективно оценивать свои действия;
- обращаться за помощью, дополнительными разъяснениями к преподавателю, другим студентам;
- пользоваться лингвистической или контекстуальной догадкой, словарями различного характера, различного рода подсказками, опорами в тексте (ключевые слова, структура текста, предваряющая информация и др.);
- использовать при говорении и письме перифраз, синонимичные средства, слова-описания общих понятий, разъяснения, примеры, толкования, «словотворчество»;

- повторять или перефразировать реплику собеседника в подтверждении понимания его высказывания или вопроса;
- обратиться за помощью к собеседнику (уточнить вопрос, переспросить и др.);
- использовать мимику, жесты (вообще и в тех случаях, когда языковых средств не хватает для выражения тех или иных коммуникативных намерений).

Подготовка к промежуточной аттестации.

При подготовке к промежуточной аттестации целесообразно:

- внимательно изучить перечень вопросов и определить, в каких источниках находятся сведения, необходимые для ответа на них;
- внимательно прочитать рекомендованную литературу;
- составить краткие конспекты ответов (планы ответов).

Лекционные занятия проводятся в соответствии с тематическим ланом, при изложении материала рекомендуется использовать презентации в среде PowerPoint и фрагменты печатных материалов по теме лекции.

В ходе интерактивных занятий следует проводить разбор конкретных примеров, максимально приближенных к реальным данным, соответствующих экономической и финансовой информации.

Основное внимание при проведении практических занятий следует уделять развитию навыков формирования рациональных схем данных предметной области, реализации этих схем в среде современных СУБД с использованием языка SQL, формирования сложных содержательных запросов по выбору данных, использования методов и алгоритмов анализа данных.

При этом задача состоит в обучении профессиональным навыкам разработки и использования баз данных и анализа данных в среде современных СУБД.

Проведение практических занятий осуществляется в компьютерных классах и включает в себя реализацию всех этапов создания и использования баз данных в среде СУБД MS SQL Server (создание базы данных, реализация схемы базы данных, включая все таблицы, связи, ограничения, индексы, ввод и редактирование данных, создание и отладка сложных логических прикладных запросов, анализ и оптимизация выполнения сформированных запросов). Также использования методов анализа данных (подготовки исходных данных, создания структур для анализа, использования альтернативных алгоритмов обработки для созданных структур). Следует обратить внимание, что примеры данных в таблицах должны обеспечить получение корректных и полных результатов запросов.

Поскольку большая часть учебного времени отводится на самостоятельное изучение дисциплины, рекомендуется уделить особое внимание организации и планированию самостоятельной работы, раскрыв существующие возможности созданных в университете корпоративных образовательных ресурсов (электронная библиотека, компьютерные обучающие программы, электронные учебные ресурсы, учебно-методические комплексы (УМК), облачные сервисы).

Практические занятия в компьютерных классах позволяют студентам сформировать навыки работы с современными СУБД, аналитическими системами на их базе и CASE – системами для решения прикладных экономических задач.

Методика проведения занятий заключается в совместном решении студентами учебной группы под руководством преподавателя типовых задач по изучаемым темам дисциплины, которые далее выполняются на вариантах индивидуальных данных. Итогом таких занятий является самостоятельное решение студентами задачи на реальных данных.

Внедрение активных и интерактивных элементов в проведение занятий по дисциплине может осуществляться разными методами: семинар с групповым обсуждением, опрос, компьютерный эксперимент и др.

Интерактивная форма проведения занятий способствует формированию профессиональных компетенций для успешного освоения основных дисциплин блока программы. Реализация интерактивной формы обеспечивается базой данных прикладной предметной области, коллективной работой над решениями задач, отсутствием единственного решения, единой целью в поиске решения. Конечная цель - выработать у студентов умение реализовывать и оценивать альтернативные варианты различных аспектов функционирования современных СУБД и аналитических систем.

Занятие посвящено изучению транзакций и их использования для организации эффективной многопользовательской работы с базой данных. Цель занятия – показать многовариантность организации уровней изоляций данных при параллельной работе многих пользователей и необходимость выработки разумных компромиссов между степенью блокировки данных и поддержки их целостности для повышения эффективности работы многопользовательской СУБД.

1. Исходными данными для реализации занятия являются база данных прикладной предметной области, созданная на предшествующих практических занятиях.
2. Студенты получают индивидуальные задания по реализации различных уровней изоляции (READ UNCOMMITTED, READ COMMITTED, REPEATABLE READ, SERIALIZABLE).
3. Студенты разрабатывают сценарии моделирования параллельной работы пользователей, организацию заданных уровней изоляции, запросы на считывание и редактирование информации в базах данных, технологию проверки промежуточных и финальных значений данных при реализации транзакций с различными уровнями изоляции и использования системных средств СУБД для отслеживания текущих состояний сервера и эффективности его работы.
4. Для каждого варианта готовится отчет (выступление) с описанием реализации решений пункта 3. Основное внимание уделяется оценке эффективности вариантов.
5. Студенты докладывают результаты реализации своих решений. Общее обсуждение и дискуссия проводится как по реализации решений, так и по сравнительной оценке эффективности работы многопользовательской СУБД. Сравнительный анализ помогает понять взаимосвязь степени блокировки данных и поддержки их целостности и их влияние на эффективность работы многопользовательской СУБД.

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (при необходимости)

8.1 Перечень информационных технологий.

- Компьютерное тестирование по итогам изучения разделов дисциплины.
- Проверка домашних заданий и консультирование посредством электронной почты.
- Использование программного обеспечения при проведении лабораторных занятий.

8.2 Перечень необходимого программного обеспечения.

- ОС Windows XP, Архиватор WinRAR, Браузер Internet Explorer. Пакет программ Microsoft Office 2003, 2007, Oracle Database 11g Express Edition бесплатная для разработки, развертывания;

8.3 Перечень информационных справочных систем:

Информационных справочных систем по этому предмету не предусмотрено.

9. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине .

№	Наименование специальных помещений и	Номера аудиторий / кабинетов
---	--------------------------------------	------------------------------

	помещений для самостоятельной работы	
1.	учебные аудитории для проведения занятий лекционного типа	501,502,503,505,506,507,508, 509, 510,513,514
2.	учебные аудитории для проведения занятий семинарского типа	501,502,503,505,506,507,508, 509, 510,513,514
3.	Компьютерные классы с выходом в Интернет	503,509,510
4.	учебные аудитории для выполнения научно – исследовательской работы (курсового проектирования)	Кабинет курсового проектирования (выполнения курсовых работ) - № 503 Оборудование: мультимедийный проектор, экран, персональные компьютеры, учебная мебель, доска учебная, выход в Интернет, учебно-наглядные пособия (тематические иллюстрации), принтер, презентации на электронном носителе, сплит-система
5.	учебные аудитории для самостоятельной работы, с рабочими местами, оснащенными компьютерной техникой с подключением к сети «Интернет» и обеспечением неограниченного доступа в электронную информационно-образовательную среду организации для каждого обучающегося, в соответствии с объемом изучаемых дисциплин	Кабинет для самостоятельной работы - № 504 Оборудование: персональные компьютеры, учебная мебель, доска учебная, выход в Интернет, МФУ (многофункциональное устройство)
6.	Исследовательские лаборатории (центров), оснащенные лабораторным оборудованием	Компьютерный класс № 510 : мультимедийный проектор, экран, персональные компьютеры, учебная мебель, доска учебная, выход в Интернет, наглядные пособия. Сетевое оборудование CISCO (маршрутизаторы, коммутаторы, 19-ти дюймовый сетевой шкаф) сплит-система, стенд «Архитектура ПЭВМ»
7.	Кабинет групповых и индивидуальных консультаций	№508 Оборудование: персональный компьютер, учебная мебель, доска учебная, учебно-наглядные пособия (тематические иллюстрации), сканер, доска магнитно-маркерная, стеллажи с учебной и периодической литературой

8.	Помещение для хранения и профилактического обслуживания учебного оборудования	Помещение № 511, Помещение № 516, Помещение № 517, Помещение № 518
9.	Помещение для проведения текущей и промежуточной аттестации	501,502,503,505,506,507,508, 509, 510,513,514

Согласно письма Министерства образования и науки РФ № МОН-25486 от 21.06.2017г «О разработке адаптированных образовательных программ» -Разработка адаптивной программы необходима в случае наличия в образовательной организации хотя бы одного обучающегося с ограниченными возможностями здоровья

Для обучающихся из числа инвалидов обучение проводится организацией с учетом особенностей их психофизического развития, их индивидуальных возможностей и состояния здоровья (далее - индивидуальные особенности).

При проведении обучения инвалидов обеспечивается соблюдение следующих общих требований:

- проведение обучения для инвалидов в одной аудитории совместно с обучающимися, не имеющими ограниченных возможностей здоровья, если это не создает трудностей для обучающихся;

- присутствие в аудитории ассистента (ассистентов), оказывающего обучающимся инвалидам необходимую техническую помощь с учетом их индивидуальных особенностей;

- пользование необходимыми обучающимся инвалидам техническими средствами с учетом их индивидуальных особенностей;

- обеспечение возможности беспрепятственного доступа обучающихся инвалидов в аудитории, туалетные и другие помещения, а также их пребывания в указанных помещениях;

В зависимости от индивидуальных особенностей обучающихся с ограниченными возможностями здоровья, организация обеспечивает выполнение следующих требований при проведении занятий:

а) для слепых:

- на компьютере со специализированным программным обеспечением для слепых, либо надиктовываются ассистенту;

б) для слабовидящих:

- задания и иные материалы оформляются увеличенным шрифтом;
- обеспечивается индивидуальное равномерное освещение не менее 300 люкс;
- при необходимости обучающимся предоставляется увеличивающее устройство, допускается использование увеличивающих устройств, имеющихся у обучающихся;

в) для глухих и слабослышащих, с тяжелыми нарушениями речи:

- обеспечивается наличие звукоусиливающей аппаратуры коллективного пользования, при необходимости обучающимся предоставляется звукоусиливающая аппаратура индивидуального пользования;

г) для лиц с нарушениями опорно-двигательного аппарата (тяжелыми нарушениями двигательных функций верхних конечностей или отсутствием верхних конечностей):

- письменные задания выполняются обучающимися на компьютере со специализированным программным обеспечением или надиктовываются ассистенту;

Обучающийся инвалид при поступлении подает письменное заявление о необходимости создания для него специальных условий при проведении обучения с указанием особенностей его психофизического развития, индивидуальных возможностей и состояния здоровья (далее - индивидуальные особенности). К заявлению прилагаются

документы, подтверждающие наличие у обучающегося индивидуальных особенностей (при отсутствии указанных документов в организации).