

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Андрей Драгомирович Хлутков
Должность: директор
Дата подписания: 20.05.2026 14:35:48
Уникальный программный ключ:
880f7c07c583b07b775f6604a630281b13ca9fd2

Приложение 4
к образовательной программе

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.ДЭ.02.02 «Большие данные в бизнес-аналитике»
(индекс, наименование дисциплины в соответствии с учебным планом)

38.04.05 «Бизнес-информатика»
(код, наименование направления подготовки/специальности)

«Аналитическое обеспечение информационной безопасности»

(наименование образовательной программы)

очная
(форма обучения)

Год набора 2026

Город
Санкт-Петербург, 2026 г.

Автор–составитель:

Доктор военных наук профессор, заведующий кафедрой бизнес-информатики
Наумов Владимир Николаевич

Заведующий кафедрой бизнес-информатики

Доктор военных наук, профессор Наумов Владимир Николаевич

РПД Б1.В.ДЭ.02.02 «Большие данные в бизнес-аналитике» одобрена
протоколом заседания кафедры бизнес-информатики № 6 от 26.03.2026 г.

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы
2. Объем и место дисциплины в структуре образовательной программы
3. Содержание и структура дисциплины
4. Материалы текущего контроля успеваемости обучающихся
5. Оценочные материалы промежуточной аттестации по дисциплине
6. Методические материалы для освоения дисциплины

Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет"

7.1. Основная литература

7.2. Дополнительная литература

7.3. Нормативные правовые документы и иная правовая информация

7.4. Интернет-ресурсы

7.5. Иные источники

Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

1.Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы

1.1. Дисциплина Б1.В.ДЭ.02.02 «Большие данные в бизнес-аналитике» обеспечивает овладение следующими компетенциями.

Таблица 1.1

ОТФ/ТФ и реквизиты ПС (при наличии)**	Код компетенци и **	Наименова ние Компетенци и **	Код индикатора достижения компетенци й **	Наименова ние индикатора достижения компетенци й **	Образовательный результат **
08.037 Бизнес-аналитик Е Управление бизнес-анализом ТД.3. Сбор информации, анализ, оценка эффективности проводимого бизнес-анализа в организации	ПКс-3	Способен управлять работами по сопровождению и проектами по созданию (модификации) ИС, автоматизирующих задачи организационного управления и бизнес-процессы	ПКс-3.1	Управляет данными при решении задач бизнес-аналитики	<p>ПКс 3.1. Знает</p> <ul style="list-style-type: none"> – Возможность и использования свободно распространяемого программного обеспечения в организации в объеме, необходимом для целей бизнес-анализа – Перспективные и существующие цифровые технологии и цифровые возможности для бизнеса в контексте предметной области и специфики деятельности организации <p>ПКс-3.1 Умеет</p> <ul style="list-style-type: none"> – Пользоваться системами анализа и визуализации данных; – Применять IT-инструменты (приложения и платформы) для обеспечения работ по бизнес-анализу

					– Отбирать, применять и адаптировать соответствующие методы, инструменты и техники анализа бизнес-ситуации и предметной области, включая методы анализа данных
--	--	--	--	--	--

** Дисциплина может формировать компетенцию полностью или частично.*

***Должно соответствовать Приложению 1 к образовательной программе*

2.Объем и место дисциплины в структуре образовательной программы

Объем дисциплины

Общий объем дисциплины Б1.В.ДЭ.02.02 «Большие данные в бизнес-аналитике» - 4 зачетные единицы – 144 акад.час; объем академических часов, выделенных на контактную работу обучающихся с преподавателем - 39 акад час, из них 12 акад.час – лекции, 16 час – практические занятия, 9 часов – контактная работа на аттестацию в период экзаменационных сессий, 2 акад.час – консультация к экзамену, 18 акад.час – самостоятельная работа при подготовке к экзамену и 87 акад. час. выделяется на самостоятельную работу обучающихся.

Дисциплина Б1.В.ДЭ.02.02 «Большие данные в бизнес-аналитике» относится к элективным дисциплинам части образовательной программы формируемой участниками образовательных отношений, преподается на 1 курсе в 1 семестре, когда обучаемые получают знания и формируют навыки на таких дисциплинах как:

Б1.В.02 «Математические методы статистической обработки и анализа данных», Б1.В.05 «Методы бизнес-аналитики» и является основой для подготовки к магистерским диссертациям.

3.Содержание и структура дисциплины

Таблица 3

3.1. Структура дисциплины Очная форма обучения

№ п/п	Наименование тем и (или) разделов	ВСЕГО	Объем дисциплины, ак.час											Форма текущего контроля успеваемости, промежуточной аттестации	
			Контактная работа обучающихся с преподавателем по видам учебных занятий							Самостоятельная работа					
			Период теоретического обучения					Период промежуточной аттестации (сессия)			СРк р	СРэ к	СР		
			Занятия лекционного типа		Занятия семинарского типа		И К	КС Р	К Э	Каттэ к					Контроль
Л	ВЛ	ЛР	ПЗ												
Тема 1	Основные понятия	32	4			4								24	Т, ПКЗ

	концепции. Введение в экосистему больших данных														
Тема 2	Методы предобработки и данных	28	2			4								22	ПКЗ, Т
Тема 3	Базы данных	29	4			4								21	ПКЗ, Т
Тема 4	Хранилища данных и аналитические платформы	26	2			4								20	ПКЗ, Т
Промежуточная аттестация		29	29						2	9				18	
Итого		144	12	0	0	16	0	0	2	9	0	0	18	87	

Используемые сокращения:

Л – лекции - занятия, предусматривающие преимущественную передачу учебной информации обучающимся педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях,).

ВЛ – видео лекции.

ЛР – лабораторные работы.

ПЗ – практические занятия (за исключением лабораторных работ).

ИК – индивидуальные консультации.

КСР – контроль самостоятельной работы

КЭ – консультации перед экзаменом

Каттэк – контактная работа на аттестацию в период экзаменационных сессий
Контроль - контактная работа на аттестацию в период экзаменационных сессий для заочной формы обучения
СРкр – самостоятельная работа на подготовку курсовой работы/ курсового проекта.
СРэк – самостоятельная работа на подготовку к экзамену.
СР – самостоятельная работа в семестре на подготовку к учебным занятиям.
Т – тестирование.
ПКЗ – практические контрольные задания.
ПИЗ – профессионально-исследовательские задания.

3.2 Содержание дисциплины

Тема 1. Основные понятия и концепции. Введение в экосистему больших данных

Понятия данные, знания. Типы данных и форматы хранения. Проблемы управления данными. Принципы управления данными. Стратегия управления данными. Жизненный цикл данных. Модели управления данными. Проблемы управления данными. Архитектура данных. Управление корпоративными требованиями в рамках проектов. Моделирование и проектирование данных. Виды моделируемых данных. Компоненты модели данных. Хранение и операции с данными.

Понятие и характеристики больших данных. Архитектурные паттерны. Lambda и Карра-архитектуры; микросервисы для обработки данных; роль message brokers (Kafka, RabbitMQ). Инструментальный ландшафт экосистемы больших данных. Обзор стека: Hadoop, Spark, Airflow, dbt, Snowflake, ClickHouse; критерии выбора.

ETL и ELT-процессы. Схема конвейера

Тема 2. Методы предобработки данных

Предобработка и очистка данных. Обработка пропусков, выбросов, дубликатов; нормализация, кодирование категориальных признаков. Агрегация и трансформация. Группировки, оконные функции, pivot/unpivot; работа с временными рядами. Методы выборки и стратификации. Репрезентативная выборка из больших массивов; стратификация по бизнес-сегментам. Оценка качества данных. Метрики полноты, точности, согласованности; профилирование данных.

Тема 3. Базы данных

Реляционные СУБД для аналитики. Использование метода ER диаграмм при проектировании БД. Выделение информационных объектов. Обязательные и необязательные сущности. Арность связи. Диаграммы ER экземпляров. Диаграммы ER типов или ER диаграммы. Правила формирования отношений 1:1. Правила формирования отношений 1:M. Правила формирования отношений M:M. Правила формирования отношений при арности больше 2. Этапы проектирования базы данных связи.

Логическая модель. Уровни: определения, сущностей, полный атрибутивный. Виды связей. Идентифицирующая и не идентифицирующая связи 1:M. Физическая модель.

Клиент-серверная архитектура. Файлы базы данных.

Создание базы данных с помощью SQL Server Management Studio. Типы данных. Создание таблиц. Диаграмма БД. Изменение макета таблиц. Работа с данными таблицы.

Создание, модификация и удаление базы данных операторами T-SQL. Спецификация файла. Создание базы данных с использованием вторичной файловой группы. Удаление базы данных.

Перенос базы данных с одного компьютера на другой. Отсоединение и присоединение БД. Резервное копирование БД и восстановление БД. Манипулирование с данными. Запросы. Организация запросов.

NoSQL-системы: типы и применение. Document (MongoDB), Key-Value (Redis), Column-family (Cassandra), Graph (Neo4j).

NewSQL и гибридные решения. Векторные базы данных для AI-аналитики

Тема 4. Хранилища данных и аналитические платформы

Концепция Data Warehouse. Отчетность vs аналитика; OLAP vs OLTP; история развития (Inmon vs Kimball). Моделирование данных для аналитики. Схемы «звезда», «снежинка», «галлактика»; slowly changing dimensions (SCD). Современные облачные DWH. Data Lake и Lake House. Различия, преимущества, риски; форматы таблиц (Delta Lake, Iceberg, Hudi). Витрины данных и self-service BI. Сравнительный анализ BI-платформ. Круг Громова. Примеры BI-платформ. Круг Громова.

4. Типы оценочных материалов, показатели и критерии оценивания

Оценочные материалы по дисциплине Б1.В.ДЭ.02.02 «Большие данные в бизнес-аналитике» входят в состав оценочных материалов по образовательной программе. Совокупность оценочных материалов по всем дисциплинам образовательной программы составляет фонд оценочных средств (далее – ФОС). ФОС используется при проведении текущего контроля успеваемости и промежуточной аттестации обучающихся с целью оценивания достижения обучающимися планируемых результатов обучения.

4.2. ФОС разработан как комплекс проверочных заданий различного типа и уровня сложности, включает критерии и шкалы оценивания, а также «ключи» правильных ответов. ФОС формируется как отдельный документ и хранится в электронном виде, доступ к ФОС предоставлен ограниченному кругу лиц.

4.3. Для самостоятельной работы обучающихся при подготовке к текущему контролю успеваемости и промежуточной аттестации в рабочих программах дисциплин размещены типовые проверочные задания, которые можно условно разделить на задания закрытого, комбинированного и открытого типов.

Задания закрытого типа — это тестовые задания, в которых каждый вопрос сопровождается готовыми вариантами ответов, из которых необходимо выбрать один или несколько правильных.

Задания комбинированного типа – это тестовые задания, в которых каждый вопрос сопровождается готовыми вариантами ответов, из которых необходимо выбрать один или несколько правильных и обосновать свой выбор.

Задания открытого типа — это задания, в которых на каждый вопрос должен быть предложен развернутый обоснованный ответ.

В зависимости от типа задания рекомендованы определенная последовательность выполнения и система оценивания выполнения заданий.

4.4. Типы заданий, сценарии выполнения, критерии оценивания

ТИП ЗАДАНИЯ	ИНСТРУКЦИЯ	СЦЕНАРИИ ВЫПОЛНЕНИЯ	КРИТЕРИИ ОЦЕНИВАНИЯ
Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных	Прочитайте текст, выберите правильный ответ	<ol style="list-style-type: none"> 1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В). 	Ответ считается верным, если правильно указана цифра или буква
Задание закрытого типа на установление соответствия	Прочитайте текст и установите соответствие	<ol style="list-style-type: none"> 1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидаются пары элементов. 2. Внимательно прочитать оба списка: список 1 – вопросы, утверждения, факты, понятия и т.д.; список 2 – утверждения, свойства объектов и т.д. 3. Сопоставить элементы списка 1 с элементами списка 2, сформировать пары элементов. 4. Записать попарно буквы и цифры (в зависимости от задания) вариантов ответа (например, А1 или Б4). 	Ответ считается верным, если правильно указаны цифры или буквы
Задание закрытого типа с выбором нескольких правильных ответов из нескольких вариантов предложенных	Прочитайте текст, выберите правильные ответы	<ol style="list-style-type: none"> 1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать несколько правильных ответов. 4. Записать только номера (или буквы) выбранного варианта ответа (например, 1 4 или А Г). 	Ответ считается верным, если правильно установлены все соответствия (позиции из одного столбца верно сопоставлены с позициями другого)
Задание закрытого типа на установление последовательности	Прочитайте текст и установите последовательность	<ol style="list-style-type: none"> 1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается последовательность элементов. 	Ответ считается верным, если правильно указана вся последовательность цифр

		<p>2. Внимательно прочитать предложенные варианты ответа.</p> <p>3. Построить верную последовательность из предложенных элементов.</p> <p>4. Записать буквы/цифры (в зависимости от задания) вариантов ответа в нужной последовательности (например, БВА или 135).</p>	
<p>Задание комбинированного типа с выбором одного правильного ответа из предложенных и обоснованием выбора</p>	<p>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа</p>	<p>1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.</p> <p>2. Внимательно прочитать предложенные варианты ответа.</p> <p>3. Выбрать один верный ответ.</p> <p>4. Записать только номер (или букву) выбранного варианта ответа.</p> <p>5. Записать аргументы, обосновывающие выбор ответа (например, 4 текст обоснования).</p>	<p>Ответ считается верным, если правильно указана цифра или буква и приведены корректные аргументы, используемые при выборе ответа</p>
<p>Задание открытого типа с развернутым ответом</p>	<p>Прочитайте текст и запишите развернутый обоснованный ответ</p>	<p>1. Внимательно прочитать текст задания и понять суть вопроса.</p> <p>2. Продумать логику и полноту ответа.</p> <p>3. Записать ответ, используя четкие компактные формулировки.</p> <p>4. В случае расчетной задачи, записать решение и ответ</p>	<p>Ответ считается верным:</p> <p>1. Отсутствие фактических ошибок.</p> <p>2. Раскрытие объема используемых понятий (полнота ответа).</p> <p>3. Обоснованность ответа (наличие аргументов).</p> <p>4. Логическая последовательность излагаемого материала.</p>

4.5. Общая шкала оценивания результатов текущего контроля успеваемости и промежуточной аттестации обучающихся с применением БРС

Итоговая балльная оценка	Традиционная система	Бинарная система	ECTS	
			Для традиционной системы	Для бинарной системы
95-100	Отлично	Зачтено	A	P/ Passed
85-94			B	P/ Passed
75-84	Хорошо		C	P/ Passed
65-74			D	P/ Passed
55-64	Удовлетворительно		E	P/ Passed
0-54	Неудовлетворительно	Не зачтено	F	F/Failed

Соотношение баллов за текущий контроль успеваемости и промежуточную аттестацию, а также повторную промежуточную аттестацию:

Максимальная сумма баллов за текущий контроль успеваемости	Максимальная сумма баллов за промежуточную аттестацию	Максимальная итоговая балльная оценка	Максимальная сумма баллов за повторную промежуточную аттестацию
60 баллов	40 баллов	100 баллов	100баллов

3. Формы аттестации, типовые оценочные материалы для текущего контроля успеваемости обучающихся, критерии и шкалы оценивания по контрольным точкам

5.1. В ходе реализации дисциплины используются следующие формы текущего контроля успеваемости обучающихся (в том числе, задания к контрольным точкам):

Т – тестирование, ПКЗ – практические контрольные задания.

Тема 1. Основные понятия и концепции. Введение в экосистему больших данных

Тестовые задания:

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.
2. Внимательно прочитать предложенные вариант-ты ответа.
3. Выбрать один верный ответ.

4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или B).

Тест 1. Какая из перечисленных характеристик НЕ входит в классическую 5V-модель больших данных?

- a) Volume (объём)
- b) Velocity (скорость)
- c) Value (ценность)
- d) Viability (жизнеспособность)
- e) Variety (разнообразиие)

Правильный ответ: d) Viability (жизнеспособность)

Тест 2. Какой объём данных традиционно считается пороговым для отнесения к «большим данным»?

- a) От 1 ГБ
- b) От 100 ГБ
- c) От 1 ТБ
- d) Не существует фиксированного порога — зависит от контекста и возможностей организации
- e) От 1 ПБ

Правильный ответ: d) Не существует фиксированного порога — зависит от контекста и возможностей организации

Обоснование: «Большие данные» определяются не абсолютным объёмом, а соотношением объёма данных и возможностей организации по их обработке традиционными инструментами.

Тест 3. Какой тип данных относится к «полуструктурированным»?

- a) Реляционные таблицы в SQL-базе данных
- b) Текстовые документы Word
- c) JSON, XML файлы
- d) Видеофайлы
- e) Аудиозаписи

Правильный ответ: c) JSON, XML файлы

Обоснование: Полуструктурированные данные имеют некоторую организацию (теги, ключи), но не соответствуют строгой схеме реляционных таблиц. JSON и XML — классические примеры.

Практическое контрольное задание по теме 1. Аналитическая платформа для ритейл-сети "SmartRetail"

Бизнес-цели:

- Прогнозирование спроса с точностью $\geq 85\%$
- Персонализация предложений для 2 млн+ клиентов
- Мониторинг эффективности маркетинговых кампаний в реальном времени
- Снижение логистических издержек за счёт оптимизации запасов

Источники данных:

- Транзакционные данные из 1С (структурированные, ~50 ГБ/день)
- Данные с кассовых терминалов (потокковые, ~1000 событий/сек)
- Логи мобильного приложения (полуструктурированные JSON, ~20 ГБ/день)
- Отзывы клиентов из соцсетей и чатов (неструктурированный текст)
- Данные с IoT-датчиков в магазинах (температура, трафик, видеоаналитика)

Текущие проблемы:

- Данные разрознены, нет единой версии правды
- Отчёты формируются с задержкой 24–48 часов
- Невозможно оперативно реагировать на аномалии (например, всплеск возвратов)
- Отсутствие инструментов для предиктивной аналитики

Тема 2. Методы предобработки данных

Тестовые задания:

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.
2. Внимательно прочитать предложенные варианты ответов.
3. Выбрать один верный ответ.
4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или B)..

Тест 1. Какой механизм пропусков является **НАИБОЛЕЕ** проблематичным для анализа и требует специализированных методов?
A) MCAR (Missing Completely At Random)
B) MAR (Missing At Random)
C) MNAR (Missing Not At Random) — вероятность пропуска зависит от самого ненаблюдаемого значения

D) Все механизмы одинаково просты в обработке

Правильный ответ: C

Обоснование: При MNAR пропуск систематически связан с ненаблюдаемым значением (напр., люди с высоким доходом реже указывают зарплату). Это ведёт к смещению оценок; требуются selection models, sensitivity analysis или экспертные допущения.

Тест 2. В чём основное преимущество множественного импутирования (MICE) перед простым удалением строк с пропусками?

- A) Оно быстрее выполняется
 - B) Оно сохраняет размер выборки и учитывает неопределённость импутации через создание нескольких датасетов
 - C) Оно автоматически исправляет все выбросы
 - D) Оно не требует проверки допущений
- Правильный ответ: B

Обоснование: MICE генерирует m импутированных датасетов, анализирует каждый и объединяет результаты (Rubin's rules). Это сохраняет статистическую мощьность и корректно оценивает стандартные ошибки, в отличие от listwise deletion.

Тест 3. Какой метод обнаружения выбросов наиболее устойчив к многомерным аномалиям?

- A) Правило $1.5 \times IQR$ для одномерных боксплотов
 - B) $Z\text{-score} > |3|$ для нормальных распределений
 - C) Isolation Forest — алгоритм, изолирующий аномалии через случайные разбиения пространства
 - D) Визуальный осмотр рассеянных диаграмм
- Правильный ответ: C

Обоснование: Isolation Forest эффективно выявляет многомерные выбросы, не требуя предположений о распределении. Он основан на том, что аномалии «изолируются» быстрее при случайных разбиениях, чем нормальные наблюдения.

Практическое контрольное задание по теме 2

Тема: «Комплексный EDA датасета опроса о восприятии ИИ»

Загрузить датасет (предоставляется: ~500 наблюдений, 25 переменных, включая пропуски).

Выполнить разведочный анализ:

Описать структуру данных, типы переменных, долю пропусков.

Визуализировать распределения ключевых переменных (гистограммы, boxplot).

Построить корреляционную матрицу с выделением значимых связей.

Применить mice для множественного импутирования ($m=5$), сравнить результаты с полным удалением пропусков.

Выявить и обосновать обработку выбросов (мин. 2 метода).

Подготовить отчёт в R Markdown с интерпретацией выводов для менеджера.

Тема 3. Базы данных

Тестовые задания:

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитать предложенные вариант-ты ответа.
3. Выбрать один верный ответ.
4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.
2. Внимательно прочитать предложенные вариант-ты ответа.
3. Выбрать один верный ответ.
4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).

Тест 1. Что такое «партиционирование» таблицы в контексте аналитических баз данных?

- a) Создание резервной копии таблицы
- b) Разделение таблицы на логические части (партиции) для ускорения запросов и упрощения управления данными
- c) Шифрование чувствительных данных в таблице
- d) Репликация таблицы на несколько серверов для отказоустойчивости

Тест 2. Какой тип NoSQL-базы данных наиболее подходит для хранения иерархических данных с гибкой схемой (например, профили пользователей с произвольными атрибутами)?

- a) Key-Value (Redis)
- b) Document-oriented (MongoDB)
- c) Column-family (Cassandra)
- d) Graph-oriented (Neo4j)

Правильный ответ: b) Document-oriented (MongoDB)

Обоснование: Документные БД хранят данные в форматах JSON/BSON, поддерживают вложенные структуры и гибкую схему, что идеально для профилей с переменным набором атрибутов. Key-Value слишком прост, Column-family оптимален для аналитики по колонкам, Graph — для связей.

Тест 3. Указать какой вид связи представлен на рисунке



- a) Многие ко многим
- b) Идентифицирующая связь один ко многим
- c) Неидентифицирующая связь один ко многим

Тест 4. Для какой бизнес-задачи наиболее целесообразно использовать графовую базу данных (например, Neo4j)?

- a) Хранение исторических продаж для отчётности
- b) Анализ социальных связей и рекомендательные системы («друзья друзей», «похожие пользователи»)
- c) Кэширование сессий пользователей
- d) Полнотекстовый поиск по документам

Правильный ответ: b) Анализ социальных связей и рекомендательные системы («друзья друзей», «похожие пользователи»)

Обоснование: Графовые БД оптимизированы для обхода связей (edges) между сущностями (nodes), что делает их незаменимыми для задач, где важны отношения: фрод-детекция, рекомендации, анализ влияний.

Практическое Контрольное задание 3. Проектирование структуры БД

Цель работы: разработка структуры БД. Исследовать предметную область и выделить сущности. Продумать определение для каждой сущности. Определить для сущностей атрибуты. Выделить ключевые атрибуты. Проанализировать связи между сущностями. Определить вид связи. Дать каждой связи имя. Построить диаграмму ER –типа с учетом всех сущностей и их связей. Сформировать набор предварительных отношений с указанием предполагаемого первичного ключа для каждого отношения и использованием диаграмм ER типа. Добавить не ключевые атрибуты в отношения. Привести предварительные отношения к нормальной форме. При необходимости пересмотреть ER диаграммы. После преобразования ER диаграмм осуществить повторное выполнение предыдущих этапов проектирования.

Тема 4. Хранилища данных и аналитические платформы

Тест по теме 4

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.
2. Внимательно прочитать предложенные варианты-ты ответа.
3. Выбрать один верный ответ.
4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или B).

Тест 1. В чём основное различие между системами OLTP и OLAP?

- a) OLTP использует реляционные базы данных, а OLAP — только NoSQL
- b) OLTP оптимизирован для транзакционных операций (вставка, обновление), а OLAP — для аналитических запросов (агрегация, отчётность)
- c) OLTP работает только в облаке, а OLAP — только локально
- d) OLTP поддерживает только структурированные данные, а OLAP — только неструктурированные

Правильный ответ: b) OLTP оптимизирован для транзакционных операций (вставка, обновление), а OLAP — для аналитических запросов (агрегация, отчётность)

Обоснование: OLTP-системы (например, операционные БД интернет-магазина) обеспечивают высокую скорость коротких транзакций. OLAP-системы (хранилища данных) оптимизированы для выполнения сложных аналитических запросов к большим объёмам исторических данных.

Тест 2. Какая схема моделирования данных наиболее характерна для хранилищ данных (Data Warehouse)?

- a) Третья нормальная форма (3NF)
- b) Схема «звезда» (Star Schema)
- c) Иерархическая модель
- d) Сетевая модель

Правильный ответ: b) Схема «звезда» (Star Schema)

Обоснование: Схема «звезда» (фактовая таблица + связанные таблицы измерений) оптимизирована для аналитических запросов: упрощает JOIN, ускоряет агрегации, интуитивно понятна бизнес-пользователям. 3NF характерна для операционных систем (нормализация для избежания аномалий).

Тест 3. Что такое «медленно меняющееся измерение» (Slowly Changing Dimension, SCD) типа 2?

- a) Измерение, которое никогда не изменяется
- b) Измерение, где при изменении атрибута создаётся новая запись с версионированием (сохраняется история)
- c) Измерение, где изменения перезаписывают старое значение без сохранения истории
- d) Измерение, которое обновляется только раз в год

Правильный ответ: b) Измерение, где при изменении атрибута создаётся новая запись с версионированием (сохраняется история)

Обоснование: SCD Type 2 сохраняет историю изменений (например, смена региона клиента): добавляется новая строка с новыми значениями, датой начала/окончания действия, флагом актуальности. Критично для корректной ретроспективной аналитики.

Практическое задание по теме 4. Разработка аналитического дашборда в Yandex DataLens

Маркетинговый департамент компании TechStore (интернет-магазин электроники) запрашивает единую панель мониторинга для оценки эффективности рекламных каналов.

Бизнес-вопросы:

- Какой канал привлечения (Google Ads, Яндекс.Директ, соцсети, email) даёт наибольшую конверсию?
- Как меняется САС (Customer Acquisition Cost) по каналам в динамике?

- Какие товарные категории наиболее маржинальны по каждому каналу?
- Где наблюдаются аномалии в расходах или конверсиях?

Доступные данные (предоставляются в формате CSV/ClickHouse):

- campaigns.csv: id_кампании, канал, дата_старта, дата_окончания, бюджет
- ads_spend.csv: дата, id_кампании, расходы, показы, клики
- orders.csv: id_заказа, дата, id_клиента, сумма, категория_товара, id_кампании
- customers.csv: id_клиента, регион, сегмент (новый/повторный)

Технические условия:

- Данные содержат ~100 000 строк, период: 12 месяцев
- Требуется обеспечить возможность фильтрации по дате, каналу, региону
- Дашборд должен быть доступен для просмотра руководству (только чтение)

Задание 1. Подготовка и подключение данных.

Задание 2. Проектирование и разработка дашборда

Задание 3. Реализация дашборда в Yandex DataLens

5.2. Типовые оценочные материалы для текущего контроля успеваемости обучающихся (вне контрольных точек) приведены в п.6.2.

5.3. Один или несколько тематических блоков дисциплины завершаются контрольной точкой (далее – КТ). Текущий контроль успеваемости по дисциплине предусматривает не менее 2 (двух) и не более 10 (десяти) КТ в течение периода освоения дисциплины.

Максимальное количество баллов за любой тип работ в рамках КТ составляет 100 (сто) баллов.

Распределение весовых коэффициентов по КТ в рамках текущего контроля успеваемости по дисциплине и формулы расчета:

Наименование контрольной точки	Максимальное количество баллов за работу в рамках КТ, которое может набрать студент	Коэффициент веса контрольной точки	Результат контрольной точки, участвующий в формировании итоговой балльной оценки по дисциплине (отражается в журнале БРС в СДО)
КТ - 1	100	0,15	15
КТ - 2	100	0,15	15
КТ-3	100	0,15	15
КЕ-4	100	0,15	15

Итого:	x	0,6	60
--------	---	-----	----

Формула расчета результата контрольной точки:

Результат контрольной точки = Количество баллов за работу в рамках КТ X Коэффициент веса контрольной точки.

5.4 Формы текущего контроля успеваемости обучающихся в рамках КТ и типовые оценочные материалы:

КТ-1

Тема 1.

Тестирование.

Практическое-контрольное задание (ПКЗ).

Тема 2.

Тестирование.

Практическое-контрольное задание (ПКЗ).

Тема 3

Тестирование.

Практическое-контрольное задание (ПКЗ).

Тема 4.

Тестирование

Практическое-контрольное задание (ПКЗ).

Для каждой формы текущего контроля успеваемости обучающихся в рамках КТ определены критерии оценивания результатов выполнения задания.

1. Критерии оценивания тестирования за каждый тест в КТ1:

Критерии оценки	Диапазон баллов	Описание критерия
<i>Количество правильных ответов</i>	<i>0</i>	<i>Количество правильных ответов менее 55%</i>
	<i>25</i>	<i>Количество правильных ответов от 55% до 64%</i>
	<i>50</i>	<i>Количество правильных ответов от 65% до 74%</i>
	<i>75</i>	<i>Количество правильных ответов от 75% до 84%</i>
	<i>100</i>	<i>Количество правильных ответов от 85% до 100%</i>
Итого максимально:	100	

2. Критерии оценивания ПКЗ:

Критерии оценки	Диапазон баллов	Описание критерия
<i>Содержание и раскрытие выбранных понятий</i>	<i>41-70</i>	<i>Детальное, последовательное описание всех понятий на примере выбранной системы</i>
	<i>21-40</i>	<i>Поверхностное описание без привязки к выбранной системе</i>
	<i>0-20</i>	<i>Понятия раскрыты минимально или не раскрыты вовсе</i>
<i>Количество выполненных заданий</i>	<i>30</i>	<i>Количество выполненных заданий от 85% до 100%</i>
	<i>15</i>	<i>Количество выполненных заданий от 55% до 84%</i>
	<i>0</i>	<i>Количество выполненных заданий менее 55%</i>
Итого максимально:	100	

6. Формы промежуточной аттестации, критерии и шкала оценивания, типовые оценочные материалы по дисциплине

6.1. Промежуточная аттестация проводится в форме экзамена

Зачет проводится путем выполнения итогового теста в ЭО/ДОТ, где могут быть использованы следующие формы: устно в ДОТ - в форме обоснованных ответов на задания различного типа; письменно в СДО - в форме письменного решения заданий различного типа; тестирование в СДО.

6.2. Типовые оценочные материалы промежуточной аттестации.

Типовые оценочные материалы промежуточной аттестации

Вопросы к экзамену по дисциплине «Большие данные в бизнес-аналитике»

1. Проблемы управления данными. Принципы управления данными. Стратегия управления данными. Жизненный цикл данных.
2. Модели управления данными. Архитектура данных.
3. Моделирование и проектирование данных. Виды моделируемых данных. Компоненты модели данных. Хранение и операции с данными.
4. Понятие и характеристики больших данных.
5. Архитектурные паттерны. Lambda и Карпа-архитектуры; микросервисы для обработки данных; роль message brokers (Kafka, RabbitMQ).
6. Инструментальный ландшафт экосистемы больших данных. Обзор стека: Hadoop, Spark, Airflow, dbt, Snowflake, ClickHouse; критерии выбора.
7. ETL и ELT-процессы. Схема конвейера.

8. Предобработка и очистка данных.
9. Обработка пропусков, выбросов, дубликатов; нормализация, кодирование категориальных признаков.
10. Агрегация и трансформация. Группировки, оконные функции, pivot/unpivot; работа с временными рядами.
11. Методы выборки и стратификации. Репрезентативная выборка из больших массивов; стратификация по бизнес-сегментам.
12. Оценка качества данных. Метрики полноты, точности, согласованности; профилирование данных.
13. Реляционные СУБД для аналитики. Использование метода ER диаграмм при проектировании БД.
14. Диаграммы ER экземпляров. Диаграммы ER типов или ER диаграммы. Правила формирования отношений 1:1. Правила формирования отношений 1:M. Правила формирования отношений M:M.
15. Логическая модель. Уровни: определения, сущностей, полный атрибутивный. Виды связей. Идентифицирующая и не идентифицирующая связи 1:M. Физическая модель.
16. Клиент-серверная архитектура. Файлы базы данных.
17. NoSQL-системы: типы и применение. Document (MongoDB), Key-Value (Redis), Column-family (Cassandra), Graph (Neo4j).
18. NewSQL и гибридные решения.
19. Векторные базы данных для AI-аналитики
20. Концепция Data Warehouse. Отчетность vs аналитика; OLAP vs OLTP; история развития (Inmon vs Kimball).
21. Моделирование данных для аналитики. Схемы «звезда», «снежинка», «галлактика»; slowly changing dimensions (SCD).
22. Современные облачные DWH. Data Lake и Lake House. Различия, преимущества, риски; форматы таблиц (Delta Lake, Iceberg, Hudi).
23. Витрины данных и self-service BI.
24. Круг Громова. Сравнительный анализ BI-платформ.
25. Примеры BI-платформ.

Типовые проверочные задания для самоподготовки обучающегося к промежуточной аттестации:

ТИП ЗАДАНИЯ	СЦЕНАРИИ ВЫПОЛНЕНИЯ	ТИПОВЫЕ ЗАДАНИЯ
Задание закрытого типа с выбором одного правильного ответа из нескольких предложенных	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или B).	Какой формат хранения данных оптимизирован для аналитических запросов в экосистеме Hadoop/Spark? a) CSV b) JSON c) Parquet d) XML e) TXT
		Что такое Apache Airflow? a) Инструмент для визуализации данных b) Платформа для оркестрации и планирования

		workflows (ETL-конвейеров) с) Распределённая база данных d) Инструмент машинного обучения а) е) Фреймворк для веб-разработки												
Задание закрытого типа на установление соответствия	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидаются пары элементов. 2. Внимательно прочитать оба списка: список 1 – вопросы, утверждения, факты, понятия и т.д.; список 2 – утверждения, свойства объектов и т.д. 3. Сопоставить элементы списка 1 с элементами списка 2, сформировать пары элементов. 4. Записать попарно буквы и цифры (в зависимости от задания) вариантов ответа (например, А1 или Б4).	<p>Характеристика (5V) Выберите правильное описание</p> <p>Volume Выберите правильное описание А) Данные поступают со скоростью 10 000 событий/сек с IoT-датчиков, требуется обработка в реальном времени</p> <p>Velocity Б) В датасете 30% пропущенных значений в поле «доход», что снижает доверие к прогнозам</p> <p>Variety В) Компания хранит 500 ТБ исторических транзакций за 5 лет для ретроспективного анализа Г) Данные включают структурированные таблицы, JSON-логи, тексты отзывов и видео с камер наблюдения</p> <p>Veracity Д) После внедрения аналитической платформы компания снизила логистические издержки на 15%</p> <p>Value</p> <p>2. Выберите правильное описание архитектурных паттернов</p> <table border="1"> <thead> <tr> <th>Архитектурный подход</th> <th>Выберите правильное описание</th> </tr> </thead> <tbody> <tr> <td>Lambda-архитектура</td> <td>А) Все данные обрабатываются как единый поток; исторические пересчёты выполняются повторным пропуском потока</td> </tr> <tr> <td>Kappa-архитектура</td> <td>Б) Данные хранятся в «сыром» виде без предварительной схемы; структура накладывается при чтении (schema-on-read)</td> </tr> <tr> <td>Data Lake</td> <td>В) Сочетает batch-слой для точной обработки и speed-слой для низких задержек; результаты объединяются в serving-слое</td> </tr> <tr> <td>Data Warehouse</td> <td>Г) Централизованное хранилище со строгой схемой (schema-on-write), оптимизированное для отчётности и BI</td> </tr> <tr> <td>Lake House</td> <td>Д) Объединяет гибкость Data Lake и управление данными DWH; поддерживает ACID-транзакции поверх файловых форматов</td> </tr> </tbody> </table>	Архитектурный подход	Выберите правильное описание	Lambda-архитектура	А) Все данные обрабатываются как единый поток; исторические пересчёты выполняются повторным пропуском потока	Kappa-архитектура	Б) Данные хранятся в «сыром» виде без предварительной схемы; структура накладывается при чтении (schema-on-read)	Data Lake	В) Сочетает batch-слой для точной обработки и speed-слой для низких задержек; результаты объединяются в serving-слое	Data Warehouse	Г) Централизованное хранилище со строгой схемой (schema-on-write), оптимизированное для отчётности и BI	Lake House	Д) Объединяет гибкость Data Lake и управление данными DWH; поддерживает ACID-транзакции поверх файловых форматов
Архитектурный подход	Выберите правильное описание													
Lambda-архитектура	А) Все данные обрабатываются как единый поток; исторические пересчёты выполняются повторным пропуском потока													
Kappa-архитектура	Б) Данные хранятся в «сыром» виде без предварительной схемы; структура накладывается при чтении (schema-on-read)													
Data Lake	В) Сочетает batch-слой для точной обработки и speed-слой для низких задержек; результаты объединяются в serving-слое													
Data Warehouse	Г) Централизованное хранилище со строгой схемой (schema-on-write), оптимизированное для отчётности и BI													
Lake House	Д) Объединяет гибкость Data Lake и управление данными DWH; поддерживает ACID-транзакции поверх файловых форматов													
Задание закрытого типа с выбором нескольких правильных ответов из нескольких вариантов предложенных	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа.	Какие характеристики из модели 5V непосредственно влияют на выбор архитектуры обработки данных? А) Volume (объём) Б) Velocity (скорость) В) Variety (разнообразие) Г) Veracity (достоверность)												

	<p>3. Выбрать несколько правильных ответов. 4. Записать только номера (или буквы) выбранного варианта ответа (например, 1 4 или А Г).</p>	<p>Д) Value (ценность)</p> <p>Какие сценарии являются типичными обоснованиями для внедрения решений класса Big Data? А) Необходимость обработки данных объёмом более 10 ТБ с ростом >20% в год Б) Требуется консолидация данных из 10+ разнородных источников (структурированных и неструктурированных) В) Бизнесу нужны отчёты с задержкой не более 5 минут от события Г) Компания хочет заменить Excel на более современный инструмент визуализации Д) Требуется обеспечение строгой согласованности данных (strong consistency) для всех операций</p>																												
<p>Задание закрытого типа на установление последовательности</p>	<p>1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается последовательность элементов. 2. Внимательно прочитать предложенные варианты ответа. 3. Построить верную последовательность из предложенных элементов. 4. Записать буквы/цифры (в зависимости от задания) вариантов ответа в нужной последовательности (например, БВА или 135).</p>	<p>1. Определите этапы процесса оценки PLS</p> <table border="1" data-bbox="890 719 1481 1413"> <thead> <tr> <th>Буква</th> <th>Этап</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>Оценка измерительной модели: расчёт loadings, CR, AVE, HTMT</td> </tr> <tr> <td>B</td> <td>Спецификация концептуальной рамки и формулировка гипотез</td> </tr> <tr> <td>C</td> <td>Интерпретация результатов и формирование бизнес-рекомендаций (IPMA, executive summary)</td> </tr> <tr> <td>D</td> <td>Тестирование гипотез с помощью бутстрэппинга (≥ 5000 итераций), расчёт медиации/модерации</td> </tr> <tr> <td>E</td> <td>Оценка структурной модели: коэффициенты путей, R^2, f^2, Q^2, SRMR</td> </tr> <tr> <td>F</td> <td>Подготовка данных: обработка пропусков, выбросов, проверка мультиколлинеарности</td> </tr> </tbody> </table> <p>2. Определите последовательность этапов EDA</p> <table border="1" data-bbox="890 1447 1481 2080"> <thead> <tr> <th>Буква</th> <th>Этап</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>Оценка измерительной модели: расчёт loadings, CR, AVE, HTMT</td> </tr> <tr> <td>B</td> <td>Спецификация концептуальной рамки и формулировка гипотез</td> </tr> <tr> <td>C</td> <td>Интерпретация результатов и формирование бизнес-рекомендаций (IPMA, executive summary)</td> </tr> <tr> <td>D</td> <td>Тестирование гипотез с помощью бутстрэппинга (≥ 5000 итераций), расчёт медиации/модерации</td> </tr> <tr> <td>E</td> <td>Оценка структурной модели: коэффициенты путей, R^2, f^2, Q^2, SRMR</td> </tr> <tr> <td>F</td> <td>Подготовка данных: обработка пропусков,</td> </tr> </tbody> </table>	Буква	Этап	A	Оценка измерительной модели: расчёт loadings, CR, AVE, HTMT	B	Спецификация концептуальной рамки и формулировка гипотез	C	Интерпретация результатов и формирование бизнес-рекомендаций (IPMA, executive summary)	D	Тестирование гипотез с помощью бутстрэппинга (≥ 5000 итераций), расчёт медиации/модерации	E	Оценка структурной модели: коэффициенты путей, R^2 , f^2 , Q^2 , SRMR	F	Подготовка данных: обработка пропусков, выбросов, проверка мультиколлинеарности	Буква	Этап	A	Оценка измерительной модели: расчёт loadings, CR, AVE, HTMT	B	Спецификация концептуальной рамки и формулировка гипотез	C	Интерпретация результатов и формирование бизнес-рекомендаций (IPMA, executive summary)	D	Тестирование гипотез с помощью бутстрэппинга (≥ 5000 итераций), расчёт медиации/модерации	E	Оценка структурной модели: коэффициенты путей, R^2 , f^2 , Q^2 , SRMR	F	Подготовка данных: обработка пропусков,
Буква	Этап																													
A	Оценка измерительной модели: расчёт loadings, CR, AVE, HTMT																													
B	Спецификация концептуальной рамки и формулировка гипотез																													
C	Интерпретация результатов и формирование бизнес-рекомендаций (IPMA, executive summary)																													
D	Тестирование гипотез с помощью бутстрэппинга (≥ 5000 итераций), расчёт медиации/модерации																													
E	Оценка структурной модели: коэффициенты путей, R^2 , f^2 , Q^2 , SRMR																													
F	Подготовка данных: обработка пропусков, выбросов, проверка мультиколлинеарности																													
Буква	Этап																													
A	Оценка измерительной модели: расчёт loadings, CR, AVE, HTMT																													
B	Спецификация концептуальной рамки и формулировка гипотез																													
C	Интерпретация результатов и формирование бизнес-рекомендаций (IPMA, executive summary)																													
D	Тестирование гипотез с помощью бутстрэппинга (≥ 5000 итераций), расчёт медиации/модерации																													
E	Оценка структурной модели: коэффициенты путей, R^2 , f^2 , Q^2 , SRMR																													
F	Подготовка данных: обработка пропусков,																													

		выбросов, проверка мультиколлинеарности														
<p>Задание комбинированного типа с выбором одного правильного ответа из предложенных и обоснованием выбора</p>	<p>1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа. 5. Записать аргументы, обосновывающие выбор ответа (например, 4 текст обоснования).</p>	<p>Компания внедряет аналитическую платформу на базе облачного хранилища. Необходимо выстроить корректную последовательность этапов разработки конвейера данных</p> <table border="1" data-bbox="887 342 1441 987"> <thead> <tr> <th>Буква</th> <th>Этап процесса</th> </tr> </thead> <tbody> <tr> <td>А</td> <td>Разработка и тестирование трансформаций (очистка, агрегация, обогащение)</td> </tr> <tr> <td>Б</td> <td>Формализация бизнес-требований и определение ключевых метрик (KPI)</td> </tr> <tr> <td>В</td> <td>Загрузка сырых данных в промежуточную зону (Staging/Raw layer)</td> </tr> <tr> <td>Г</td> <td>Публикация витрин данных и подключение BI-инструментов</td> </tr> <tr> <td>Д</td> <td>Подключение к источникам и извлечение данных</td> </tr> <tr> <td>Е</td> <td>Валидация качества данных и настройка мониторинга (алерты, SLA)</td> </tr> </tbody> </table>	Буква	Этап процесса	А	Разработка и тестирование трансформаций (очистка, агрегация, обогащение)	Б	Формализация бизнес-требований и определение ключевых метрик (KPI)	В	Загрузка сырых данных в промежуточную зону (Staging/Raw layer)	Г	Публикация витрин данных и подключение BI-инструментов	Д	Подключение к источникам и извлечение данных	Е	Валидация качества данных и настройка мониторинга (алерты, SLA)
Буква	Этап процесса															
А	Разработка и тестирование трансформаций (очистка, агрегация, обогащение)															
Б	Формализация бизнес-требований и определение ключевых метрик (KPI)															
В	Загрузка сырых данных в промежуточную зону (Staging/Raw layer)															
Г	Публикация витрин данных и подключение BI-инструментов															
Д	Подключение к источникам и извлечение данных															
Е	Валидация качества данных и настройка мониторинга (алерты, SLA)															
<p>Задание открытого типа с развернутым ответом</p>	<p>1. Внимательно прочитать текст задания и понять суть вопроса. 2. Продумать логику и полноту ответа. 3. Записать ответ, используя четкие компактные формулировки. 4. В случае расчетной задачи, записать решение и ответ</p>	<p>2. Аналитик разрабатывает аналитическую модель для отчёта по продажам. Установите корректную последовательность шагов проектирования.</p> <p>А) Определение фактовой таблицы и выбор гранулярности (уровня детализации) Б) Согласование бизнес-гlossария и метрик со стейкхолдерами В) Создание физической модели (DDL) и настройка индексов/партиций Г) Проектирование таблиц измерений и связей с фактами Д) Настройка стратегии историзации (SCD Type 2/3) для измерений Е) Тестирование аналитических запросов и оптимизация производительности</p> <p>Какие 3–4 конкретных бизнес-вопроса должна решать платформа в первую очередь? Какие KPI изменятся и как будет измеряться успех проекта через 6 месяцев?</p>														

6.3. Критерии и шкала оценивания на основе БРС.

Критерии и балльная шкала определены в БРС. В moodle оценка выставляется по результатам одной попытки. На попытку дается 60 минут. В тест включены 30 вопросов из разных мет учебной дисциплины.

Максимальное число баллов на зачет – 40. Из них тридцать баллов за тест и 10 баллов за активность на занятиях слушателей или наличие сертификата по одному из курсов, посвященных предсказательной аналитике.

6.4. Описание дополнительных материалов и оборудования, необходимых для выполнения проверочных заданий (*при необходимости*).

Для решения задач открытого типа (кейсов, ПКЗ), тестовых заданий студенту разрешается использование калькулятора; программ для работы с электронными таблицами для обработки, анализа и визуализации данных. Для построения интеллект-карты и моделей в различных нотациях студенту можно использовать любой соответствующий онлайн-инструмент.

7. Методические материалы по освоению дисциплины (модуля)

Для закрепления изученного материала даны вопросы по каждой теме дисциплины, на которые следует самостоятельно найти ответы, основные материалы размещены в moodle.

Важной составной частью учебного процесса в вузе являются практические занятия. Практические занятия проводятся главным образом по дисциплинам, требующим закрепления навыков решения задач, и помогают студентам глубже усвоить учебный материал, приобрести умения применять принципы системного подхода к решению разнообразных задач, определять и оценивать ресурсы и существующие ограничения разного рода проектов.

Необходимо помнить, что на лекции обычно рассматривается не весь материал, а только его часть. Остальная его часть восполняется в процессе самостоятельной работы. В связи с этим работа с рекомендованной литературой обязательна.

После изучения базовых тем курса проводится текущий контроль знаний студентов в виде опроса или письменного тестирования. Типовые тесты и задания по темам дисциплины приведены в специальном разделе данной рабочей программы.

8. Учебная литература и ресурсы информационно-телекоммуникационной сети Интернет

8.1. Основная литература

1. Алламмор Джей, Гроотендорст Маартен. Большие языковые модели на практике. Понимание языка и генерация текстов. – Астана, Алист, 2026. -416 с.

2. Анализ данных : учебник для вузов / под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2026. — 448 с. — (Высшее образование). — ISBN 978-5-534-19964-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/583032> (дата обращения: 01.05.2026).

1. Афанасьев, Владимир Николаевич. Анализ временных рядов и прогнозирование. - Саратов: Ай Пи Ар Медиа. – 310 с. Текст : электронный. - URL: <http://www.iprbookshop.ru/90196.html> (дата обращения: 12.11.2020). - Режим доступа: для авторизир. Пользователей

2. Макшанов А.В., Журавлев А.Е., Тындыкарь Л.Н. Большие данные. Big Data: учебник для вузов. — 3-е изд., стер. — СПб.: Лань, 2023. — 188 с. — ISBN 978-5-507-46866-9

3. Миркин, Б. Г. Базовые методы анализа данных : учебник и практикум для вузов / Б. Г. Миркин. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2026. — 297 с. — (Высшее образование). — ISBN 978-5-534-19709-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/583143> (дата обращения: 01.05.2026).

4. Митина О.А. Бизнес-аналитика. Введение в обработку и анализ данных: учебник для вузов. — М.: Юрайт, 2026. — 172 с. — ISBN 978-5-534-21811-4

8.2. Дополнительная литература

1. Балдин К.В. Эконометрика : учеб. пособие для студентов вузов, обучающихся по специальностям "Финансы и кредит", "Бухгалтерский учет, анализ и аудит" / К.В. Балдин, О.Ф. Быстров, М.М. Соколов. - М. : ЮНИТИ-ДАНА, 2004. - 254с.

2. Гинько А.Ю. Анализ и визуализация данных в Yandex DataLens. Подробное руководство: от новичка до эксперта. — М.: ДМК Пресс, 2022. — 356 с. — ISBN 978-5-93700-171-9.

3. Коллективная монография. Большие данные и цифровая революция: как анализ данных меняет мир / под ред. М.А. Гайипова. — Ташкент: Fan va texnologiyalar, 2025. — 312 с.

4. Курносов Ю.В. Бизнес-аналитика: инструменты, методы, практика. — М.: ДМК Пресс, 2023. — 420 с.

8.3 Нормативные правовые документы и иная правовая информация

Не используются

8.4 Интернет-ресурсы

Обучающимся обеспечен доступ к материалам курса в СДО Академии <http://lms.ranepa.ru>, а также через сайт научной библиотеки к следующим подписным электронным ресурсам:

Русскоязычные ресурсы

- Электронные учебники электронно-библиотечной системы (ЭБС) «Айбукс»
- Электронные учебники электронно-библиотечной системы (ЭБС) «Юрайт»
- Электронные учебники электронно-библиотечной системы (ЭБС) «Лань»
- Электронные учебники электронно-библиотечной системы (ЭБС) «ZNANIUM.COM»
- Электронные учебники электронно-библиотечной системы (ЭБС) «BOOK.RU»
- Электронные учебники электронно-библиотечной системы (ЭБС) «IPR SMART»

Облачные ресурсы курсов и репозиториев

- Документация: Apache Airflow, dbt, Spark, Kafka (официальные сайты)
- Онлайн-курсы: Stepik «Инженерия данных», Coursera «Data Engineering with Google Cloud»
- netology.ru
- Платформы для практики: Kaggle Datasets, Google BigQuery Sandbox, Yandex Cloud Trial
- Профессиональные сообщества: Data Engineering Russia, Analytics Community, Telegram-чаты по Big Data

9. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

№ п/п	Наименование
1.	Специализированные залы для проведения лекций, оснащенные персональным компьютером/ноутбуком и мультимедийным проектором
2.	Аудитории и компьютерные классы, оборудованные посадочными местами и персональными компьютерами с выходом в Интернет для проведения практических занятий
3.	«МТС Линк» — российская платформа для онлайн-коммуникаций и совместной работы команд ; «Яндекс Телемост» — сервис для видеоконференций от Яндекса; Я-мессенджер
4.	Технические средства обучения: персональные компьютеры; программные средства, обеспечивающие просмотр видеофайлов в форматах AVI, MPEG-4, DivX, RMVB, WMV; программы для работы с электронными таблицами для обработки, анализа и визуализации данных; соответствующие онлайн-инструменты для построения интеллект-карты и моделей в различных нотациях
5.	Научная библиотека (в т.ч. электронные информационные ресурсы научной библиотеки)

6.	СДО Академии https://lms.ranepa.ru/
7.	R, Rstudio, Anaconda Navigator, VS code, JASP, jamovi, excel