



Первый проректор-проректор по учебной работе

09 2026 г.

Б1.В.02.12 «Инженерия данных»

Самара 2026

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы	4
2. Место дисциплины (модуля) в структуре образовательной программы	5
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся	5
4. Содержание дисциплины (модуля), структурированное по темам (разделам), с указанием отведенного на них количества академических часов и видов учебных занятий	6
4.1 Содержание лекционных занятий	6
4.2 Содержание лабораторных занятий	7
4.3 Содержание практических занятий	8
4.4. Содержание самостоятельной работы	8
5. Перечень учебной литературы и учебно-методического обеспечения по дисциплине (модулю)	9
6. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения	9
7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», профессиональных баз данных, информационно-справочных систем	10
8. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)	10
9. Методические материалы	10
10. Фонд оценочных средств по дисциплине (модулю)	12

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Наименование категории (группы) компетенций	Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Результаты обучения (знать, уметь, владеть, соотнесенные с индикаторами достижения компетенции)
Профессиональные компетенции			
Не предусмотрено	ПК-2 Способность проектировать и разрабатывать программные средства работы с базами данных современными инструментальными средствами и технологиями	ПК-2.1 Знать: средства СУБД для выявления проблем производительности при выполнении и повышении пропускной способности баз данных	Знать архитектуры платформ данных, средства контроля выполнения запросов, потоков загрузки и пропускной способности хранилищ.
		ПК-2.2 Уметь: осуществлять поиск, хранение, обработку и анализ информации и баз данных, представление их в требуемом формате	Уметь извлекать, очищать, преобразовывать, хранить, анализировать и представлять данные в пакетных и потоковых конвейерах.
		ПК-2.3 Владеть: проектированием баз данных, поддержанием схемы БД в соответствии с изменениями в требованиях и предметной области	Владеть навыками проектирования хранилищ, моделей и контрактов данных, а также управляемого изменения схем и конвейеров.
Универсальные компетенции			
Системное и критическое мышление	УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Знать: основы системного и критического анализа информации	Знать методы системного анализа источников, требований, качества, происхождения, ограничений и рисков данных.
		УК-1.2 Уметь: разрабатывать стратегию действий, принимать конкретные решения для ее реализации	Уметь разрабатывать стратегию построения конвейера, сравнивать варианты архитектуры и принимать обоснованные решения.
		УК-1.3 Владеть: методиками постановки цели, определения способов ее достижения, разработки стратегий действий	Владеть методиками постановки цели, декомпозиции потока данных, выбора критериев и планирования реализации и проверки решения.

2. Место дисциплины (модуля) в структуре образовательной программы

Место дисциплины (модуля) в структуре образовательной программы: **часть, формируемая участниками образовательных отношений**

Код компетенции	Предшествующие дисциплины	Параллельно осваиваемые дисциплины	Последующие дисциплины
ПК-2	Анализ данных и машинное обучение; Большие данные; Визуализация данных; Информационная бизнес-аналитика; Нейронные сети; Основы работы с NoSQL-базами данных	Компьютерное зрение; Подготовка к процедуре защиты и защита выпускной квалификационной работы; Технологии интернета вещей	
УК-1	Анализ данных и машинное обучение; Большие данные; Введение в цифровую экономику; Визуализация данных; Методология разработки DevOps; Нейронные сети; Объектно-ориентированное программирование; Основы работы с NoSQL-базами данных; Основы экономического анализа; Производственная практика: научно-исследовательская работа; Развертывание и жизненный цикл программного обеспечения; Разработка и принятие управленческих решений; Стратегическое планирование; Учебная практика: научно-исследовательская работа (получение первичных навыков научно-исследовательской работы); Учебная практика: проектная практика; Финансовый менеджмент; Цифровая трансформация бизнеса; Экономика предприятия; Экономическое развитие и инновационная экономика; Языки программирования	Компьютерное зрение; Моделирование бизнес-процессов; Подготовка к процедуре защиты и защита выпускной квалификационной работы; Производственная практика: преддипломная практика; Технологии интернета вещей	

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Вид учебной работы	Всего часов / часов в электронной форме	8 семестр часов / часов в электронной форме
Аудиторная контактная работа (всего), в том числе:	32	32
Лабораторные работы	16	16

Лекции	16	16
Внеаудиторная контактная работа, КСР	3	3
Самостоятельная работа (всего), в том числе:	17	17
подготовка к зачету	4	4
подготовка к лабораторным работам	13	13
Контроль	20	20
Итого: час	72	72
Итого: з.е.	2	2

4. Содержание дисциплины (модуля), структурированное по темам (разделам), с указанием отведенного на них количества академических часов и видов учебных занятий

№ раздела	Наименование раздела дисциплины	Виды учебной нагрузки и их трудоемкость, часы				
		ЛЗ	ЛР	ПЗ	СРС	Всего часов
1	Архитектура и жизненный цикл данных	4	4	0	4	12
2	Загрузка, интеграция и качество данных	4	4	0	4	12
3	Хранилища и преобразование данных	4	4	0	4	12
4	Оркестрация, наблюдаемость и управление данными	4	4	0	5	13
	КСР	0	0	0	0	3
	Контроль	0	0	0	0	20
	Итого	16	16	0	17	72

4.1 Содержание лекционных занятий

№ занятия	Наименование раздела	Тема лекции	Содержание лекции (перечень дидактических единиц: рассматриваемых подтем, вопросов)	Количество часов / часов в электронной форме
8 семестр				
1	Архитектура и жизненный цикл данных	Инженерия данных и архитектура платформ данных	Роль инженера данных. Жизненный цикл данных. Источники, транспорт, обработка, хранение и потребление. Централизованные и распределенные архитектуры.	2
2	Архитектура и жизненный цикл данных	Источники, форматы и контракты данных	Базы данных, файлы, API и события. Структурированные и слабоструктурированные форматы. Схемы, контракты, метаданные и происхождение данных.	2

3	Загрузка, интеграция и качество данных	Пакетная и потоковая загрузка данных	Полная и инкрементальная загрузка. Захват изменений. Очереди и журналы событий. Гарантии доставки, повторная обработка и идемпотентность.	2
4	Загрузка, интеграция и качество данных	Профилирование, очистка и контроль качества данных	Полнота, точность, согласованность, уникальность и своевременность. Профилирование, правила проверки, очистка, карантин и отчетность.	2
5	Хранилища и преобразование данных	ETL, ELT и моделирование хранилищ данных	Извлечение, преобразование и загрузка. Нормализованные и размерные модели. Факты, измерения, историчность и витрины данных.	2
6	Хранилища и преобразование данных	Озера данных, lakehouse и форматы хранения	Зоны озера данных. Колоночные форматы и секционирование. Каталоги, транзакционные таблицы, эволюция схемы и управление файлами.	2
7	Оркестрация, наблюдаемость и управление данными	Оркестрация и наблюдаемость конвейеров данных	Графы задач, зависимости, расписания и повторные запуски. Метрики, журналы, происхождение, соглашения об уровне сервиса и обработка сбоев.	2
8	Оркестрация, наблюдаемость и управление данными	Производительность, безопасность и управление данными	Планы запросов, индексы, секционирование и масштабирование. Доступ, маскирование, каталогизация, качество, владельцы и управляемое изменение.	2
Итого за семестр:				16
Итого:				16

4.2 Содержание лабораторных занятий

№ занятия	Наименование раздела	Тема лабораторного занятия	Содержание лабораторного занятия (перечень дидактических единиц: рассматриваемых подтем, вопросов)	Количество часов / часов в электронной форме
8 семестр				
1	Архитектура и жизненный цикл данных	Профилирование источника и контракт данных	Исследовать структуру и качество исходного набора, определить поля, типы, ключи и критические ограничения.	2
2	Архитектура и жизненный цикл данных	Профилирование источника и контракт данных	Сформировать контракт и правила эволюции схемы, реализовать автоматическую проверку корректных и ошибочных данных.	2
3	Загрузка, интеграция и качество данных	Конвейер пакетной и потоковой загрузки	Реализовать полную и инкрементальную загрузку, зафиксировать контрольную точку и сведения о происхождении.	2
4	Загрузка, интеграция и качество данных	Конвейер пакетной и потоковой загрузки	Добавить повторную обработку, идемпотентность и карантин ошибочных записей, проверить сбой и восстановление.	2

5	Хранилища и преобразование данных	Хранилище и преобразование данных	Спроектировать модель фактов и измерений, реализовать загрузку очищенных данных и основные преобразования.	2
6	Хранилища и преобразование данных	Хранилище и преобразование данных	Создать витрину и агрегаты, добавить секционирование или индексы, сравнить производительность типовых запросов.	2
7	Оркестрация, наблюдаемость и управление данными	Оркестрация и контроль качества конвейера	Описать граф задач с зависимостями и расписанием, настроить проверки качества, журналы и повторные запуски.	2
8	Оркестрация, наблюдаемость и управление данными	Оркестрация и контроль качества конвейера	Добавить мониторинг и оповещение, воспроизвести отказ, восстановить конвейер и оформить техническую документацию.	2
Итого за семестр:				16
Итого:				16

4.3 Содержание практических занятий

Учебные занятия не реализуются.

4.4. Содержание самостоятельной работы

Наименование раздела	Вид самостоятельной работы	Содержание самостоятельной работы (перечень дидактических единиц: рассматриваемых подтем, вопросов)	Количество часов
8 семестр			
Архитектура и жизненный цикл данных	Подготовка к лекционным и лабораторным занятиям	Повторение архитектур, жизненного цикла, источников, форматов, контрактов, метаданных и происхождения данных.	3
Архитектура и жизненный цикл данных	Подготовка к зачету	Повторение теоретических вопросов и типовых задач по разделу	1
Загрузка, интеграция и качество данных	Подготовка к лекционным и лабораторным занятиям	Повторение пакетной и потоковой загрузки, гарантий доставки, идемпотентности, профилирования и контроля качества.	3
Загрузка, интеграция и качество данных	Подготовка к зачету	Повторение теоретических вопросов и типовых задач по разделу	1
Хранилища и преобразование данных	Подготовка к лекционным и лабораторным занятиям	Повторение ETL, ELT, размерного моделирования, озер данных, lakehouse, форматов хранения и эволюции схемы.	3
Хранилища и преобразование данных	Подготовка к зачету	Повторение теоретических вопросов и типовых задач по разделу	1

Оркестрация, наблюдаемость и управление данными	Подготовка к лекционным и лабораторным занятиям	Повторение оркестрации, наблюдаемости, производительности, безопасности, каталогизации и управления данными.	4
Оркестрация, наблюдаемость и управление данными	Подготовка к зачету	Повторение теоретических вопросов и типовых задач по разделу	1
Итого за семестр:			17
Итого:			17

5. Перечень учебной литературы и учебно-методического обеспечения по дисциплине (модулю)

№ п/п	Библиографическое описание	Ресурс НТБ СамГТУ (ЭБС СамГТУ, IPRbooks и т.д.)
Основная литература		
1	Архипенков, С. Хранилища данных : От концепции до внедрения / С.Архипенков, Д.Голубев, О.Максименко.- М., ДИАЛОГ-МИФИ, 2002.- 528 с.	Книжный фонд
2	Круз, Р. Структуры данных и проектирование программ / Пер.с 3-го англ.изд. К.Г.Финогенова.- М., БИНОМ.Лаб.знаний, 2008.- 765 с.	Книжный фонд
Дополнительная литература		
3	Большие данные: учебное пособие / Черемухин А.Д., Ай Пи Ар Медиа: 2023.- Режим доступа: https://elib.samgtu.ru/getinfo?uid=els_samgtu iprbooks 129721	Электронный ресурс
4	Разработка веб-приложений на платформе Spring: Учебно-методическое пособие / Суханов В.И., Издательство Уральского университета, ред. Тимошенко С.И.: 2023.- Режим доступа: https://elib.samgtu.ru/getinfo?uid=els_samgtu iprbooks 157380	Электронный ресурс

Доступ обучающихся к ЭР НТБ СамГТУ (elib.samgtu.ru) осуществляется посредством электронной информационной образовательной среды университета и сайта НТБ СамГТУ по логину и паролю.

6. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения

При проведении лекционных занятий используется мультимедийное оборудование.

Организовано взаимодействие обучающегося и преподавателя с использованием электронной ин-формационной образовательной среды университета.

№ п/п	Наименование	Производитель	Способ распространения
1	PostgreSQL	PostgreSQL Global Development Group (Зарубежный)	Свободно распространяемое

2	Apache Airflow	Apache Software Foundation (Зарубежный)	Свободно распространяемое
3	Apache Kafka	Apache Software Foundation (Зарубежный)	Свободно распространяемое
4	Python	Python Software Foundation (Зарубежный)	Свободно распространяемое

7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», профессиональных баз данных, информационно-справочных систем

№ п/п	Наименование	Краткое описание	Режим доступа
1	Библиотека компьютерной литературы	http://it.eup.ru/	Ресурсы открытого доступа
2	Книги. Программирование [Электронный ресурс]. - Microsoft Project 2010: The Missing Manual	books.tr200.ru/v.php	Ресурсы открытого доступа
3	Научная электронная библиотека «Киберленинка»	https://cyberleninka.ru	Ресурсы открытого доступа
4	Библиотека учебно-методической литературы системы "Единое окно"	http://window.edu.ru/	Ресурсы открытого доступа

8. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Лекционные занятия

Аудитория, оснащенная презентационной техникой: проектор, экран, компьютер или ноутбук. Аудитория оснащенная учебной мебелью.

Лабораторные занятия

Компьютерный класс с доступом в локальную сеть и Интернет, оснащенный компьютерами, редактором программного кода и средствами демонстрации учебных материалов.

Самостоятельная работа

Компьютерный класс с доступом в Интернет и электронную информационно-образовательную среду СамГТУ. Читальный зал НТБ СамГТУ. Ресурсы УИТ СамГТУ.

9. Методические материалы

Методические рекомендации при работе на лекции

До лекции студент должен просмотреть учебно-методическую и научную литературу по теме лекции с тем, чтобы иметь представление о проблемах, которые будут разбираться в лекции.

Перед началом лекции обучающимся сообщается тема лекции, план, вопросы, подлежащие рассмотрению, доводятся основные литературные источники. Весь учебный материал, сообщаемый преподавателем, должен не просто прослушиваться. Он должен быть активно воспринят, т.е. услышан, осмыслен, понят, зафиксирован на бумаге и закреплён в памяти. Приступая к слушанию нового учебного материала, полезно мысленно установить его связь с ранее изученным. Следя за техникой чтения лекции (акцент на существенном, повышение тона, изменение ритма, пауза и т.п.), необходимо вслед за

преподавателем уметь выделять основные категории, законы и определять их содержание, проблемы, предполагать их возможные решения, доказательства и выводы. Осуществляя такую работу, можно значительно облегчить себе понимание учебного материала, его конспектирование и дальнейшее изучение.

Конспектирование лекции позволяет обработать, систематизировать и лучше сохранить полученную информацию с тем, чтобы в будущем можно было восстановить в памяти основные, содержательные моменты. Типичная ошибка, совершаемая обучающимся, дословное конспектирование речи преподавателя. Как правило, при записи «слово в слово» не остается времени на обдумывание, анализ и синтез информации. Отбирая нужную информацию, главные мысли, проблемы, решения и выводы, необходимо сокращать текст, строить его таким образом, чтобы потом можно было легко в нем разобраться. Желательно оставить в рабочих конспектах поля, на которых можно будет делать пометки из рекомендованной литературы, дополняющие материал прослушанной лекции, а также подчеркивающие особую важность тех или иных теоретических положений. С окончанием лекции работа над конспектом не может считаться завершенной. Нужно еще восстановить отдельные места, проверить, все ли понятно, уточнить что-то на консультации и т.п. с тем, чтобы конспект мог быть использован в процессе подготовки к практическим занятиям, зачету, экзамену. Конспект лекции – незаменимый учебный документ, необходимый для самостоятельной работы.

Методические рекомендации при работе на лабораторном занятии

Проведение лабораторной работы делится на две условные части: теоретическую и практическую.

Необходимыми структурными элементами занятия являются проведение лабораторной работы, проверка усвоенного материала, включающая обсуждение теоретических основ выполняемой работы.

Перед лабораторной работой, как правило, проводится технико-теоретический инструктаж по использованию необходимого оборудования. Преподаватель корректирует деятельность обучающегося в процессе выполнения работы (при необходимости). После завершения лабораторной работы подводятся итоги, обсуждаются результаты деятельности.

Возможны следующие формы организации лабораторных работ: фронтальная, групповая и индивидуальная. При фронтальной форме выполняется одна и та же работа (при этом возможны различные варианты заданий). При групповой форме работа выполняется группой (командой). При индивидуальной форме обучающимися выполняются индивидуальные работы.

По каждой лабораторной работе имеются методические указания по их выполнению, включающие необходимый теоретический и практический материал, содержащие элементы и последовательную инструкцию по проведению выбранной работы, индивидуальные варианты заданий, требования и форму отчетности по данной работе.

Методические рекомендации по выполнению самостоятельной работы

Организация самостоятельной работы обучающихся ориентируется на активные методы овладения знаниями, развитие творческих способностей, переход от поточного к индивидуализированному обучению с учетом потребностей и возможностей обучающегося.

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. Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса.

Самостоятельная работа реализуется:

- непосредственно в процессе аудиторных занятий;
- на лекциях, практических занятиях;
- в контакте с преподавателем вне рамок расписания;
- на консультациях по учебным вопросам, в ходе творческих контактов, при ликвидации задолженностей, при выполнении индивидуальных заданий и т.д.;
- в библиотеке, дома, на кафедре при выполнении обучающимся учебных и практических задач.

Эффективным средством осуществления обучающимся самостоятельной работы является электронная информационно-образовательная среда университета, которая обеспечивает доступ к учебным планам, рабочим программам дисциплин (модулей), практик, к изданиям электронных библиотечных систем.

10. Фонд оценочных средств по дисциплине (модулю)

Фонд оценочных средств представлен в приложении № 1.

Приложение 1 к рабочей программе дисциплины
Б1.В.02.12 «Инженерия данных»

**Фонд оценочных средств
по дисциплине
Б1.В.02.12 «Инженерия данных»**

Код и направление подготовки (специальность)	02.03.03 Математическое обеспечение и администрирование информационных систем
Направленность (профиль)	Анализ данных в бизнес-аналитике
Квалификация	Бакалавр
Форма обучения	Очная
Год начала подготовки	2026
Институт / факультет	Институт инженерно-экономического и гуманитарного образования
Выпускающая кафедра	кафедра "Высшая математика"
Кафедра-разработчик	кафедра "Высшая математика"
Объем дисциплины, ч. / з.е.	72 / 2
Форма контроля (промежуточная аттестация)	Зачет

**Перечень планируемых результатов обучения по дисциплине (модулю),
соотнесенных с планируемыми результатами освоения образовательной
программы**

Наименование категории (группы) компетенций	Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Результаты обучения (знать, уметь, владеть, соотнесенные с индикаторами достижения компетенции)
Профессиональные компетенции			
Не предусмотрено	ПК-2 Способность проектировать и разрабатывать программные средства работы с базами данных современными инструментальными средствами и технологиями	ПК-2.1 Знать: средства СУБД для выявления проблем производительности при выполнении и повышении пропускной способности баз данных	Знать архитектуры платформ данных, средства контроля выполнения запросов, потоков загрузки и пропускной способности хранилищ.
		ПК-2.2 Уметь: осуществлять поиск, хранение, обработку и анализ информации и баз данных, представление их в требуемом формате	Уметь извлекать, очищать, преобразовывать, хранить, анализировать и представлять данные в пакетных и потоковых конвейерах.
		ПК-2.3 Владеть: проектированием баз данных, поддержанием схемы БД в соответствии с изменениями в требованиях и предметной области	Владеть навыками проектирования хранилищ, моделей и контрактов данных, а также управляемого изменения схем и конвейеров.
Универсальные компетенции			
Системное и критическое мышление	УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Знать: основы системного и критического анализа информации	Знать методы системного анализа источников, требований, качества, происхождения, ограничений и рисков данных.
		УК-1.2 Уметь: разрабатывать стратегию действий, принимать конкретные решения для ее реализации	Уметь разрабатывать стратегию построения конвейера, сравнивать варианты архитектуры и принимать обоснованные решения.
		УК-1.3 Владеть: методиками постановки цели, определения способов ее достижения, разработки стратегий действий	Владеть методиками постановки цели, декомпозиции потока данных, выбора критериев и планирования реализации и проверки решения.

Матрица соответствия оценочных средств запланированным результатам обучения

Код индикатора достижения компетенции	Результаты обучения	Оценочные средства	Текущий контроль успеваемости	Промежуточная аттестация
Архитектура и жизненный цикл данных				
ПК-2.1 Знать: средства СУБД для выявления проблем производительности при выполнении и повышении пропускной способности баз данных	Знать архитектуры платформ данных, средства контроля выполнения запросов, потоков загрузки и пропускной способности хранилищ.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
ПК-2.2 Уметь: осуществлять поиск, хранение, обработку и анализ информации и баз данных, представление их в требуемом формате	Уметь извлекать, очищать, преобразовывать, хранить, анализировать и представлять данные в пакетных и потоковых конвейерах.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
ПК-2.3 Владеть: проектированием баз данных, поддержанием схемы БД в соответствии с изменениями в требованиях и предметной области	Владеть навыками проектирования хранилищ, моделей и контрактов данных, а также управляемого изменения схем и конвейеров.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
УК-1.1 Знать: основы системного и критического анализа информации	Знать методы системного анализа источников, требований, качества, происхождения, ограничений и рисков данных.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
УК-1.2 Уметь: разрабатывать стратегию действий, принимать конкретные решения для ее реализации	Уметь разрабатывать стратегию построения конвейера, сравнивать варианты архитектуры и принимать обоснованные решения.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
УК-1.3 Владеть: методиками постановки цели, определения способов ее достижения, разработки стратегий действий	Владеть методиками постановки цели, декомпозиции потока данных, выбора критериев и планирования реализации и проверки решения.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
Загрузка, интеграция и качество данных				
ПК-2.1 Знать: средства СУБД для выявления проблем производительности при выполнении и повышении пропускной способности баз данных	Знать архитектуры платформ данных, средства контроля выполнения запросов, потоков загрузки и пропускной способности хранилищ.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да

ПК-2.2 Уметь: осуществлять поиск, хранение, обработку и анализ информации и баз данных, представление их в требуемом формате	Уметь извлекать, очищать, преобразовывать, хранить, анализировать и представлять данные в пакетных и потоковых конвейерах.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
ПК-2.3 Владеть: проектированием баз данных, поддержанием схемы БД в соответствии с изменениями в требованиях и предметной области	Владеть навыками проектирования хранилищ, моделей и контрактов данных, а также управляемого изменения схем и конвейеров.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
УК-1.1 Знать: основы системного и критического анализа информации	Знать методы системного анализа источников, требований, качества, происхождения, ограничений и рисков данных.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
УК-1.2 Уметь: разрабатывать стратегию действий, принимать конкретные решения для ее реализации	Уметь разрабатывать стратегию построения конвейера, сравнивать варианты архитектуры и принимать обоснованные решения.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
УК-1.3 Владеть: методиками постановки цели, определения способов ее достижения, разработки стратегий действий	Владеть методиками постановки цели, декомпозиции потока данных, выбора критериев и планирования реализации и проверки решения.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
Хранилища и преобразование данных				
ПК-2.1 Знать: средства СУБД для выявления проблем производительности при выполнении и повышении пропускной способности баз данных	Знать архитектуры платформ данных, средства контроля выполнения запросов, потоков загрузки и пропускной способности хранилищ.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
ПК-2.2 Уметь: осуществлять поиск, хранение, обработку и анализ информации и баз данных, представление их в требуемом формате	Уметь извлекать, очищать, преобразовывать, хранить, анализировать и представлять данные в пакетных и потоковых конвейерах.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
ПК-2.3 Владеть: проектированием баз данных, поддержанием схемы БД в соответствии с изменениями в требованиях и предметной области	Владеть навыками проектирования хранилищ, моделей и контрактов данных, а также управляемого изменения схем и конвейеров.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
УК-1.1 Знать: основы системного и критического анализа информации	Знать методы системного анализа источников, требований, качества, происхождения, ограничений и рисков данных.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да

УК-1.2 Уметь: разрабатывать стратегию действий, принимать конкретные решения для ее реализации	Уметь разрабатывать стратегию построения конвейера, сравнивать варианты архитектуры и принимать обоснованные решения.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
УК-1.3 Владеть: методиками постановки цели, определения способов ее достижения, разработки стратегий действий	Владеть методиками постановки цели, декомпозиции потока данных, выбора критериев и планирования реализации и проверки решения.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
Оркестрация, наблюдаемость и управление данными				
ПК-2.1 Знать: средства СУБД для выявления проблем производительности при выполнении и повышении пропускной способности баз данных	Знать архитектуры платформ данных, средства контроля выполнения запросов, потоков загрузки и пропускной способности хранилищ.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
ПК-2.2 Уметь: осуществлять поиск, хранение, обработку и анализ информации и баз данных, представление их в требуемом формате	Уметь извлекать, очищать, преобразовывать, хранить, анализировать и представлять данные в пакетных и потоковых конвейерах.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
ПК-2.3 Владеть: проектированием баз данных, поддержанием схемы БД в соответствии с изменениями в требованиях и предметной области	Владеть навыками проектирования хранилищ, моделей и контрактов данных, а также управляемого изменения схем и конвейеров.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
УК-1.1 Знать: основы системного и критического анализа информации	Знать методы системного анализа источников, требований, качества, происхождения, ограничений и рисков данных.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
УК-1.2 Уметь: разрабатывать стратегию действий, принимать конкретные решения для ее реализации	Уметь разрабатывать стратегию построения конвейера, сравнивать варианты архитектуры и принимать обоснованные решения.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да
УК-1.3 Владеть: методиками постановки цели, определения способов ее достижения, разработки стратегий действий	Владеть методиками постановки цели, декомпозиции потока данных, выбора критериев и планирования реализации и проверки решения.	Подготовка к ЛЗ и ЛР, подготовка к зачету	Да	Да

Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующие процесс формирования компетенций в ходе освоения образовательной программы

Формы текущего контроля успеваемости Текущий контроль осуществляется по результатам выполнения и защиты лабораторных работ, контрольных заданий и тестирования по материалам дисциплины.

Пример контрольной работы по разделу «Архитектура и жизненный цикл данных»

1. Определить источники, потребителей и этапы жизненного цикла данных.
2. Выбрать архитектуру платформы по заданным требованиям.
3. Описать схему и семантику исходного набора.
4. Разработать контракт данных и правила его изменения.
5. Определить метаданные и происхождение данных.

Пример контрольной работы по разделу «Загрузка, интеграция и качество данных»

1. Разработать полную и инкрементальную загрузку.
2. Обеспечить идемпотентность повторного запуска.
3. Сформулировать измеримые правила качества.
4. Определить карантин и обработку ошибочных записей.
5. Разработать сценарий восстановления после сбоя.

Пример контрольной работы по разделу «Хранилища и преобразование данных»

1. Спроектировать таблицу фактов и измерения.
2. Определить ETL- или ELT-последовательность.
3. Разработать витрину и типовые агрегаты.
4. Выбрать формат и секционирование файлов.
5. Подготовить управляемое изменение схемы.

Пример контрольной работы по разделу «Оркестрация, наблюдаемость и управление данными»

1. Построить граф задач и зависимостей.
2. Определить расписание и правила повторного запуска.
3. Выбрать метрики, журналы и оповещения конвейера.
4. Проанализировать план запроса и предложить улучшение.
5. Определить владельца, доступ и правила каталогизации набора.

Семестр 8 Форма промежуточной аттестации — зачет. Зачет проводится в письменно-устной форме. Билет включает два теоретических вопроса и одно практическое задание. Результаты выполненных и защищенных лабораторных работ учитываются при промежуточной аттестации.

Федеральное государственное бюджетное образовательное учреждение высшего образования
Самарский государственный технический университет
ЗАЧЕТНЫЙ БИЛЕТ № 1
по дисциплине Б1.В.02.13 «Инженерия данных»
Направление подготовки 02.03.03 Математическое обеспечение и администрирование информационных систем
1. Роль инженера данных и жизненный цикл данных.
2. Очистка, карантин и отчетность по качеству.
3. Практическое задание. Выполнить профилирование набора и сформировать контракт данных.

Вопросы к зачету (8 семестр)

1. Роль инженера данных и жизненный цикл данных.
2. Компоненты и архитектуры платформ данных.
3. Источники данных: базы, файлы, API и события.
4. Структурированные и слабоструктурированные форматы.
5. Схемы, контракты, метаданные и происхождение данных.
6. Полная и инкрементальная пакетная загрузка.
7. Захват изменений в источнике.
8. Потокковая передача и журналы событий.
9. Гарантии доставки, повторная обработка и идемпотентность.
10. Измерения качества и профилирование данных.
11. Очистка, карантин и отчетность по качеству.
12. ETL, ELT и проектирование преобразований.
13. Нормализованные и размерные модели данных.
14. Факты, измерения, историчность и витрины.
15. Озера данных, зоны и колоночные форматы.
16. Lakehouse, каталоги и эволюция схемы.
17. Оркестрация, графы задач и повторные запуски.
18. Наблюдаемость и соглашения об уровне сервиса данных.
19. Производительность запросов и конвейеров.
20. Безопасность, каталогизация и управление данными.

Типовые практические задания к зачету

1. Выполнить профилирование набора и сформировать контракт данных.
2. Разработать схему полной и инкрементальной загрузки.
3. Предусмотреть идемпотентность и восстановление после сбоя.
4. Сформулировать правила качества и обработку ошибочных записей.
5. Спроектировать модель фактов и измерений.
6. Разработать ETL- или ELT-преобразование и витрину данных.
7. Построить граф задач с проверками, повторными запусками и оповещением.
8. Проанализировать производительность запроса и предложить улучшения.

Тесты по компетенциям
по дисциплине Б1.В.02.13 «Инженерия данных»
(шифр и наименование дисциплины)

№ задания	Содержание задания	Ключ к ответу	Наименование темы РПД	Семестр								
ПК-2 Способность проектировать и разрабатывать программные средства работы с базами данных современными инструментальными средствами и технологиями												
1	Какой компонент координирует зависимости и расписание задач? 1. Оркестратор. 2. Текстовый редактор. 3. DNS-сервер. 4. Архиватор.	1.	Оркестрация и наблюдаемость конвейеров данных	8								
2	Согласованное описание полей, типов и правил набора называется контрактом _____.	данных.	Источники, форматы и контракты данных	8								
3	Установите соответствие между компонентом и назначением.	<table><tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr><tr><td>2</td><td>4</td><td>3</td><td>1</td></tr></table>	А	Б	В	Г	2	4	3	1	Инженерия данных и архитектура платформ данных	8
	А		Б	В	Г							
	2		4	3	1							
	А. Источник.		1. Представление данных для конкретной аналитической задачи.									
	Б. Транспорт.		2. Место возникновения исходных данных.									
В. Хранилище.	3. Долговременное размещение подготовленных наборов.											
Г. Витрина.	4. Передача данных между компонентами.											
Запишите выбранные цифры под соответствующими буквами в виде таблицы:												
<table><tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr><tr><td></td><td></td><td></td><td></td></tr></table>		А	Б	В	Г							
А	Б	В	Г									
4	Что должен содержать контракт данных? 1. Схему. 2. Семантику полей. 3. Правила качества. 4. Цвет панели.	1, 2, 3.	Источники, форматы и контракты данных	8								
5	Установите последовательность пакетной обработки: А. Загрузить проверенный набор в целевую таблицу. Б. Извлечь данные из источника. В. Выполнить преобразования. Г. Проверить схему и качество.	Б, Г, В, А.	Пакетная и потоковая загрузка данных	8								
6	Зачем конвейеру идемпотентность?	Повторный запуск с теми же входными данными не должен создавать дубликаты или искажать результат, что обеспечивает безопасное восстановление после сбоя.	Пакетная и потоковая загрузка данных	8								
7	Что обычно хранит измеримые события предметной области? 1. Таблица фактов. 2. Таблица ролей. 3. Файл конфигурации. 4. Журнал сборки.	1.	ETL, ELT и моделирование хранилищ данных	8								
8	Источник добавил обязательное поле. Как безопасно изменить конвейер?	Обновить контракт и версию схемы, проверить совместимость, адаптировать прием и преобразование, выполнить тесты на старых и новых данных и контролировать переход.	Источники, форматы и контракты данных	8								
9	Какие свойства относятся к качеству данных? 1. Полнота. 2. Уникальность. 3. Своевременность. 4. Цвет файла.	1, 2, 3.	Профилирование, очистка и контроль качества данных	8								
10	Изолированная область для записей, не прошедших проверку, называется _____.	карантином.	Профилирование, очистка и контроль качества данных	8								
11	Какой формат эффективен для аналитического чтения отдельных столбцов? 1. Колоночный. 2. Растровый. 3. Аудиоформат. 4. Архив без схемы.	1.	Озера данных, lakehouse и форматы хранения	8								

№ задания	Содержание задания	Ключ к ответу	Наименование темы РПД	Семестр								
12	Установите соответствие между зоной озера и содержанием.	<table><tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr><tr><td>2</td><td>3</td><td>1</td><td>4</td></tr></table>	А	Б	В	Г	2	3	1	4	Озера данных, lakehouse и форматы хранения	8
	А		Б	В	Г							
	2		3	1	4							
	А. Сырая зона.		1. Подготовленные наборы с подтвержденным качеством.									
	Б. Очищенная зона.		2. Исходные данные без содержательных преобразований.									
В. Доверенная зона.	3. Данные после стандартизации и очистки.											
Г. Архивная зона.	4. Редко используемая история длительного хранения.											
	Запишите выбранные цифры под соответствующими буквами в виде таблицы:											
	<table><tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	А	Б	В	Г							
А	Б	В	Г									
13	Чем ETL отличается от ELT?	В ETL преобразование выполняется до загрузки в целевое хранилище, а в ELT данные сначала загружаются и преобразуются средствами целевой платформы.	ETL, ELT и моделирование хранилищ данных	8								
14	Установите последовательность восстановления конвейера: А. Повторно запустить с контрольной точки. Б. Определить последнюю успешную операцию. В. Проверить целевой результат. Г. Устранить причину сбоя.	Б, Г, А, В.	Оркестрация и наблюдаемость конвейеров данных	8								
15	Что помогает ускорить аналитический запрос? 1. Подходящий индекс. 2. Секционирование. 3. Предварительный агрегат. 4. Случайное дублирование.	1, 2, 3.	Производительность, безопасность и управление данными	8								
16	После повторного запуска появились дубликаты. Предложите исправление.	Ввести устойчивый ключ записи и операции слияния, хранить контрольную точку, сделать запись идемпотентной и проверить повторный запуск на одинаковом наборе.	Пакетная и потоковая загрузка данных	8								
17	Что описывает путь данных от источника до результата? 1. Происхождение данных. 2. Цветовая схема. 3. Имя сервера. 4. Размер шрифта.	1.	Источники, форматы и контракты данных	8								
18	Структура зависимостей задач без циклов называется направленным ациклическим _____.	графом.	Оркестрация и наблюдаемость конвейеров данных	8								
19	Зачем секционировать крупную таблицу по времени?	Это сокращает объем чтения для запросов по интервалу, упрощает управление историей и может ускорить загрузку и удаление периодов.	Производительность, безопасность и управление данными	8								
20	Загрузка укладывается в срок, но данные неполны. Какие проверки добавить?	Контролировать число записей, обязательные поля, уникальность ключей, диапазоны и сверку итогов с источником, сохранять отчет и блокировать публикацию неполного набора.	Профилирование, очистка и контроль качества данных	8								
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач												

№ задания	Содержание задания	Ключ к ответу	Наименование темы РПД	Семестр																								
21	С чего начинается проектирование платформы данных? 1. С определения цели, потребителей и требований. 2. С выбора случайного инструмента. 3. С удаления метаданных. 4. С создания всех таблиц.	1.	Инженерия данных и архитектура платформ данных	8																								
22	Разделение потока данных на связанные этапы называется _____.	декомпозицией.	Инженерия данных и архитектура платформ данных	8																								
23	Установите соответствие между критерием и способом проверки. <table><tr><td>А. Полнота.</td><td>1. Сравнение связанных значений и правил.</td></tr><tr><td>Б. Уникальность.</td><td>2. Подсчет отсутствующих обязательных значений.</td></tr><tr><td>В. Своевременность.</td><td>3. Контроль задержки появления данных.</td></tr><tr><td>Г. Согласованность.</td><td>4. Поиск повторяющихся ключей.</td></tr></table> Запишите выбранные цифры под соответствующими буквами в виде таблицы: <table><tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	А. Полнота.	1. Сравнение связанных значений и правил.	Б. Уникальность.	2. Подсчет отсутствующих обязательных значений.	В. Своевременность.	3. Контроль задержки появления данных.	Г. Согласованность.	4. Поиск повторяющихся ключей.	А	Б	В	Г					<table><tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr><tr><td>2</td><td>4</td><td>3</td><td>1</td></tr></table>	А	Б	В	Г	2	4	3	1	Профилирование, очистка и контроль качества данных	8
А. Полнота.	1. Сравнение связанных значений и правил.																											
Б. Уникальность.	2. Подсчет отсутствующих обязательных значений.																											
В. Своевременность.	3. Контроль задержки появления данных.																											
Г. Согласованность.	4. Поиск повторяющихся ключей.																											
А	Б	В	Г																									
А	Б	В	Г																									
2	4	3	1																									
24	Какие сведения нужны для выбора архитектуры? 1. Объем и скорость данных. 2. Требования потребителей. 3. Ограничения и риски. 4. Только популярность инструмента.	1, 2, 3.	Инженерия данных и архитектура платформ данных	8																								
25	Установите последовательность разработки конвейера: А. Проверить решение по критериям. Б. Выбрать архитектуру и компоненты. В. Сформулировать требования. Г. Реализовать минимальный поток.	В, Б, Г, А.	Инженерия данных и архитектура платформ данных	8																								
26	Почему инструмент нельзя выбирать только по максимальной производительности?	Необходимо учитывать надежность, стоимость, компетенции команды, совместимость, сопровождение, безопасность и реальные требования к нагрузке.	Инженерия данных и архитектура платформ данных	8																								
27	Что делает сравнение хранилищ обоснованным? 1. Одинаковая нагрузка и единые критерии. 2. Разные наборы без описания. 3. Название продукта. 4. Цвет интерфейса.	1.	Производительность, безопасность и управление данными	8																								
28	Потребители требуют свежие данные, но потоковая архитектура дорога. Как принять решение?	Уточнить допустимую задержку, оценить пакетную микрозагрузку и потоковый вариант по стоимости и рискам, создать прототип и выбрать минимально достаточное решение.	Пакетная и потоковая загрузка данных	8																								
29	Что фиксируется при принятии риска качества? 1. Причина. 2. Владелец. 3. Срок и компенсирующая мера. 4. Отказ от наблюдения.	1, 2, 3.	Производительность, безопасность и управление данными	8																								
30	Проверяемое предположение о причине задержки конвейера называется _____.	гипотезой.	Оркестрация и наблюдаемость конвейеров данных	8																								
31	Что подтверждает полезность витрины? 1. Проверка по запросам и показателям потребителя. 2. Число столбцов. 3. Имя таблицы. 4. Размер документации.	1.	ETL, ELT и моделирование хранилищ данных	8																								

№ задания	Содержание задания	Ключ к ответу	Наименование темы РПД	Семестр								
32	Установите соответствие между этапом анализа и результатом.	<table><tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr><tr><td>2</td><td>3</td><td>4</td><td>1</td></tr></table>	А	Б	В	Г	2	3	4	1	Инженерия данных и архитектура платформ данных	8
	А		Б	В	Г							
	2		3	4	1							
	А. Постановка цели.		1. Данные о соответствии критериям.									
Б. Анализ источников.	2. Ожидаемый результат и потребители.											
В. Сравнение архитектур.	3. Характеристики, ограничения и качество входа.											
	Г. Проверка прототипа.	4. Обоснованный выбор компонентов.										
	Запишите выбранные цифры под соответствующими буквами в виде таблицы:											
	<table><tr><td>А</td><td>Б</td><td>В</td><td>Г</td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	А	Б	В	Г							
А	Б	В	Г									
33	Зачем проверять восстановление конвейера после сбоя?	Проверка подтверждает корректность контрольных точек, повторных запусков, идемпотентности и целевого результата, а не только работу без ошибок.	Оркестрация и наблюдаемость конвейеров данных	8								
34	Установите последовательность анализа сбоя: А. Проверить гипотезу и исправление. Б. Зафиксировать симптомы и время. В. Оформить вывод и предупреждающую меру. Г. Локализовать проблемную задачу и данные.	Б, Г, А, В.	Оркестрация и наблюдаемость конвейеров данных	8								
35	Какие данные нужны для оценки производительности? 1. Объем входа. 2. Время стадий. 3. Использование ресурсов. 4. Только среднее без распределения.	1, 2, 3.	Производительность, безопасность и управление данными	8								
36	Два подразделения по-разному трактуют одно поле. Как действовать?	Согласовать владельца и определение, зафиксировать семантику в контракте и каталоге, определить преобразование и проверить потребителей перед изменением.	Источники, форматы и контракты данных	8								
37	Что снижает риск миграции схемы? 1. Версионирование, совместимость и поэтапная проверка. 2. Изменение без уведомления. 3. Удаление истории. 4. Отказ от тестов.	1.	Озера данных, lakehouse и форматы хранения	8								
38	Ответственное лицо за определение и качество набора называется _____ данных.	владельцем.	Производительность, безопасность и управление данными	8								
39	Почему архитектурное решение необходимо документировать?	Документация фиксирует требования, ограничения, альтернативы и причины выбора, обеспечивая проверку, сопровождение и изменение платформы.	Инженерия данных и архитектура платформ данных	8								
40	Проверка качества увеличила задержку, но почти не находит ошибок. Как оценить ее необходимость?	Сопоставить стоимость и задержку с выявляемыми рисками, проверить правило на истории, изменить частоту или место выполнения либо удалить неэффективную проверку.	Профилирование, очистка и контроль качества данных	8								

Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующие процесс формирования компетенций

Оценка достижения запланированных результатов обучения производится по итогам отдельных видов текущего контроля и промежуточной аттестации.

Шкала оценивания

«Зачтено» выставляется, если сформированность заявленных знаний, умений и навыков обучающегося составляет 60% и более и обязательные части задания выполнены положительно.

«Не зачтено» выставляется, если результат составляет менее 60%, выявлены существенные пробелы в знаниях и обучающийся не может выполнить типовое практическое задание.

Ответы и решения обучающихся оцениваются по следующим общим критериям: распознавание проблемы; определение значимой информации; анализ проблемы; аргументированность; использование подходящей стратегии; правильность решения; обоснованность выводов; творческий подход; общая грамотность.

Текущий контроль

Результат тестирования определяется как доля набранных баллов от максимально возможного количества баллов. Положительным считается результат не менее 60%. Практическое или ситуационное задание оценивается с учетом правильности решения, полноты анализа, самостоятельности, аргументированности и качества выводов.

Промежуточная аттестация

Зачет в 8-м семестре проводится по вопросам и практическим заданиям дисциплины. Билет включает два теоретических вопроса и одно практическое задание. При проведении зачета могут учитываться положительные результаты текущего контроля.

Уровень освоения	Результат зачета
60–100%	Зачтено
Менее 60%	Не зачтено

Обучающиеся выполняют задания в сроки, установленные преподавателем. Итоговый результат определяется по совокупности результатов выполнения теоретической и практической частей зачета.