

**Рабочая программа дисциплины (модуля) «Технологии обработки информации»,
включая оценочные материалы**

1. Требования к результатам обучения по дисциплине (модулю)

1.1. Перечень компетенций, формируемых дисциплиной (модулем) в процессе освоения образовательной программы

Группа компетенций	Категория компетенций	Коды и содержание компетенций
Универсальные	-	-
Общепрофессиональные	-	ОПК-5. Способен устанавливать программное и аппаратное обеспечение для информационных и автоматизированных систем
	-	ОПК-7. Способен применять в практической деятельности основные концепции, принципы, теории и факты, связанные с информатикой
	-	ОПК-8. Способен осуществлять поиск, хранение, обработку и анализ информации из различных источников и баз данных, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий
Профессиональные	-	-

1.2. Компетенции и индикаторы их достижения, формируемых дисциплиной (модулем) в процессе освоения образовательной программы

Код компетенции	Код индикатора компетенции	Содержание индикатора компетенции
ОПК-5	ОПК-5.1	Использует современные программные средства для настройки и управления информационными и автоматизированными системами
ОПК-5	ОПК-5.2	Использует современные аппаратные средства для интеграции в информационные и автоматизированные системы
ОПК-5	ОПК-5.3	Применяет методы установки системного и прикладного программного обеспечения для информационных и автоматизированных систем
ОПК-7	ОПК-7.1	Анализирует практики использования основных концепций, принципов, теорий и фактов, связанных с информатикой, в профессиональной деятельности
ОПК-7	ОПК-7.2	Способен применять в практической деятельности основные концепции, принципы, теории и факты, связанные с информатикой.
ОПК-7	ОПК-7.3	Использует в практической деятельности основные концепции, принципы, теории и факты, связанные с информатикой.
ОПК-8	ОПК-8.1	Понимает методы поиска, хранения, обработки и анализа информации из различных источников и баз данных
ОПК-8	ОПК-8.2	Формирует решения задач поиска, хранения и анализа информации из различных источников и баз данных
ОПК-8	ОПК-8.3	Использует современные информационные, компьютерные и сетевые технологии для поиска, хранения и анализа информации из различных источников и баз данных

1.3. Результаты обучения по дисциплине (модулю)

Цель изучения дисциплины (модуля) – освоение обучающимися современных технологий для обработки и анализа информации и эффективных методов ее обработки с применением современных ЭВМ, а также формирование целостной системы знаний в области создания, накопления, обработки и использования информационных ресурсов.

В результате изучения дисциплины (модуля) обучающийся должен **знать:**

- методологию современных информационных технологий, используемых в анализе данных; методы, средства и алгоритмы для выполнения анализа данных по заданным фильтрам и критериям; принципы работы в аналитических платформах;
- уметь:**
- выбирать современные информационные технологии и программные средства, для анализа данных, осуществлять разработку моделей обработки информации в аналитической платформе;
- владеть:**
- навыками анализа и моделирования данных в программных средствах (в т.ч. отечественного производства), в частности в аналитической платформе.

2. Объем, структура и содержание дисциплины (модуля)

2.1. Объем дисциплины (модуля)

<i>Виды учебной работы</i>	<i>Формы обучения</i>
	<i>Очная</i>
Общая трудоемкость: зачетные единицы/часы	3/108
Контактная работа:	72
Лекции	36
Лабораторные работы	0
Практические занятия, семинары	36
Промежуточная аттестация: зачет	0
Самостоятельная работа (СР)	36

2.2. Темы (разделы) дисциплины (модуля) с указанием отведенного на них количества часов по формам образовательной деятельности

Очная форма обучения

№ п/п	Наименование тем (разделов)	Виды учебной работы (в часах)						СР
		Контактная работа						
		Занятия лекционного типа		Занятия семинарского типа				
		Л	Иные	ПЗ	С	ЛР	Иные	
1.	Базовые технологии обработки информации.	2	0	4	0	0	0	11
2.	Прикладные инструменты для анализа информации.	4	0	6	0	0	0	11
3.	Принцип организации обработки данных в системе Loginom, основанный на сценариях.	4	0	6	0	0	0	12
4.	Стандартизация и очистка информации.	2	0	6	0	0	0	11
5.	Регулярные выражения.	2	0	6	0	0	0	11
6.	Основы хранилищ и витрин данных. Многомерная модель данных.	4	0	6	0	0	0	11
7.	Многомерные и реляционные модели данных.	4	0	6	0	0	0	12
8.	Витрины данных.	2	0	6	0	0	0	11
9.	Загрузка данных в хранилище.	2	0	6	0	0	0	11
10.	Метод Data Vault.	2	0	6	0	0	0	11
11.	Программирование	4	0	6	0	0	0	11

	на языке Python для сбора и анализа данных.							
--	---------------------------------------------	--	--	--	--	--	--	--

Примечания:

Л – лекции, ПЗ – практические занятия, С – семинары, ЛР – лабораторные работы, СР – самостоятельная работа.

2.3. Содержание дисциплины (модуля), структурированное по темам (разделам) и видам работ

Содержание лекционного курса

№ п/п	Наименование тем (разделов)	Содержание лекционного курса
1.	Базовые технологии обработки информации.	Создание, модификация данных, контроль, безопасность и целостность данных, поиск информации, поддержка принятия решения, создание документов, сводок, отчетов, преобразование информации. Задачи обработки данных различных типов. Прикладные области обработки данных. Сжатие информации, алгоритмы сжатия.
2.	Прикладные инструменты для анализа информации.	Исторический обзор инструментов бизнес-аналитики для анализа информации. Аналитическая платформа Loginom и ее редакции. Использование структурного и объектно-ориентированного подхода к моделированию в Loginom.
3.	Принцип организации обработки данных в системе Loginom, основанный на сценариях.	Основы работы с Loginom: работа с пакетами, сценарии, компоненты и узлы, модификаторы доступа, настройки портов и автосинхронизация, компонент калькулятор, переменные и параметризация узлов, компоненты условие и замена. Компонент цикл, наследование и производные компоненты, импорт из промышленных источников данных.
4.	Стандартизация и очистка информации.	Причины загрязнения данных и виды ошибок. Методы очистки данных: использование словарей, и таблиц замены, анализ строк, регулярные выражения, частотный анализ, контрольные числа. Объединение различных методов очистки данных. Общий алгоритм очистки.
5.	Регулярные выражения.	История появления регулярных выражений. Диалекты и возможности регулярных выражений. Синтаксис регулярных выражений, языка Perl: метасимволы, квантификаторы, примеры использования.
6.	Основы хранилищ и витрин данных. Многомерная модель данных.	Основные положения концепции ХД (хранилищ данных). Определение и свойства ХД. Детализированные и агрегированные данные, метаданные, виды метаданных. Краткий обзор моделей ХД. Многомерное представление данных. Базовые понятия многомерной модели данных. Структура многомерного куба. Работа с измерениями.
7.	Многомерные и реляционные модели данных.	MOLAP и ROLAP. Агрегаты и нерегламентированные запросы. Пример MOLAP. Реляционные ХД ROLAP, схема «звезда» и «созвездие». Схема «созвездие». Сравнение MOLAP и ROLAP. Гибридная модель HOLAP.
8.	Витрины данных.	Концепция витрин данных. Проектирование витрины данных. Пример проектирование витрины данных.
9.	Загрузка данных в хранилище.	Загрузка данных в хранилище. Поддержка историчности. Медленно меняющиеся измерения. Неполная загрузка данных. Многопоточная организация процесса загрузки данных. Постзагрузочные операции.
10.	Метод Data Vault.	Недостатки существующих моделей ХД. Основные концепции и определение Data Vault. Эволюция от нормальных форм к Data Vault. Основные компоненты модели Data Vault: концентраторы, сссылки, спутники. Пример модели Data Vault. Обобщенная схема хранилища Data Vault. Преимущества и недостатки Data Vault.
11.	Программирование на языке Python для сбора и анализа данных.	Типы данных в Python: строки, списки, кортежи, словари, множества. Функции, их определение. Элементы функционального программирования: list comprehensions,

		лямбда-функции, рекурсия. Элементы объектно-ориентированного программирования.
--	--	--------------------------------------------------------------------------------

Содержание занятий семинарского типа

№ п/п	Наименование тем (разделов)	Тип	Содержание занятий семинарского типа
1.	Базовые технологии обработки информации.	ПЗ	Создание, модификация данных, контроль, безопасность и целостность данных, поиск информации, поддержка принятия решения, создание документов, сводок, отчетов, преобразование информации. Задачи обработки данных различных типов. Прикладные области обработки данных. Сжатие информации, алгоритмы сжатия.
2.	Прикладные инструменты для анализа информации.	ПЗ	Исторический обзор инструментов бизнес-аналитики для анализа информации. Аналитическая платформа Loginom и ее редакции. Использование структурного и объектно-ориентированного подхода к моделированию в Loginom.
3.	Принцип организации обработки данных в системе Loginom, основанный на сценариях.	ПЗ	Основы работы с Loginom: работа с пакетами, сценарии, компоненты и узлы, модификаторы доступа, настройки портов и автосинхронизация, компонент калькулятор, переменные и параметризация узлов, компоненты условие и замена. Компонент цикл, наследование и производные компоненты, импорт из промышленных источников данных.
4.	Стандартизация и очистка информации.	ПЗ	Причины загрязнения данных и виды ошибок. Методы очистки данных: использование словарей, и таблиц замены, анализ строк, регулярные выражения, частотный анализ, контрольные числа. Объединение различных методов очистки данных. Общий алгоритм очистки.
5.	Регулярные выражения.	ПЗ	История появления регулярных выражений. Диалекты и возможности регулярных выражений. Синтаксис регулярных выражений, языка Perl: метасимволы, квантификаторы, примеры использования.
6.	Основы хранилищ и витрин данных. Многомерная модель данных.	ПЗ	Основные положения концепции ХД (хранилищ данных). Определение и свойства ХД. Детализированные и агрегированные данные, метаданные, виды метаданных. Краткий обзор моделей ХД. Многомерное представление данных. Базовые понятия многомерной модели данных. Структура многомерного куба.
7.	Многомерные и реляционные модели данных.	ПЗ	MOLAP и ROLAP. Агрегаты и нерегламентированные запросы. Пример MOLAP. Реляционные ХД ROLAP, схема «звезда» и «созвездие». Схема «созвездие». Сравнение MOLAP и ROLAP. Гибридная модель HOLAP.
8.	Витрины данных.	ПЗ	Концепция витрин данных. Проектирование витрины данных. Пример проектирование витрины данных.
9.	Загрузка данных в хранилище.	ПЗ	Загрузка данных в хранилище. Поддержка историчности. Медленно меняющиеся измерения. Неполная загрузка данных. Многопоточная организация процесса загрузки данных.
10.	Метод Data Vault.	ПЗ	Недостатки существующих моделей ХД. Основные концепции и определение Data Vault. Эволюция от нормальных форм к Data Vault. Основные компоненты модели Data Vault: концентраторы, ссылки, спутники. Пример модели Data Vault. Обобщенная схема хранилища Data Vault.
11.	Программирование на языке Python для сбора и анализа данных.	ПЗ	Типы данных в Python: строки, списки, кортежи, словари, множества. Функции, их определение. Элементы функционального программирования: list comprehensions, лямбда-функции, рекурсия.

Содержание самостоятельной работы

№ п/п	Наименование тем (разделов)	Содержание самостоятельной работы
1.	Базовые технологии обработки информации.	Прикладные области обработки данных. Сжатие информации, алгоритмы сжатия.
2.	Прикладные инструменты для анализа информации.	Использование структурного и объектно-ориентированного подхода к моделированию в Loginom.
3.	Принцип организации обработки данных в системе Loginom, основанный на сценариях.	Компонент цикл, наследование и производные компоненты, импорт из промышленных источников данных.
4.	Стандартизация и очистка информации.	Общий алгоритм очистки.
5.	Регулярные выражения.	Синтаксис регулярных выражений, языка Perl: метасимволы, квантификаторы, примеры использования.
6.	Основы хранилищ и витрин данных. Многомерная модель данных.	Структура многомерного куба. Работа с измерениями.
7.	Многомерные и реляционные модели данных.	Сравнение MOLAP и ROLAP. Гибридная модель HOLAP.
8.	Витрины данных.	Пример проектирование витрины данных.
9.	Загрузка данных в хранилище.	Постзагрузочные операции.
10.	Метод Data Vault.	Обобщенная схема хранилища Data Vault. Преимущества и недостатки Data Vault.
11.	Программирование на языке Python для сбора и анализа данных.	Элементы объектно-ориентированного программирования.

3. Оценочные материалы для проведения текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине (модулю)

По дисциплине (модулю) предусмотрены следующие виды контроля качества освоения:

- текущий контроль успеваемости;
- промежуточная аттестация обучающихся по дисциплине (модулю).

3.1. Оценочные материалы для проведения текущей аттестации по дисциплине (модулю)

№ п/п	Контролируемые темы (разделы)	Наименование оценочного средства
1.	Базовые технологии обработки информации.	Устный опрос. Кейсы. Дискуссионные процедуры. Мини-тест
2.	Прикладные инструменты для анализа информации.	Устный опрос. Кейсы. Дискуссионные процедуры. Мини-тест
3.	Принцип организации обработки данных в системе Loginom, основанный на сценариях.	Устный опрос. Кейсы. Дискуссионные процедуры. Мини-тест
4.	Стандартизация и очистка информации.	Устный опрос. Кейсы. Дискуссионные процедуры. Мини-тест
5.	Регулярные выражения.	Устный опрос. Кейсы. Дискуссионные процедуры. Мини-тест
6.	Основы хранилищ и витрин данных. Многомерная модель данных.	Устный опрос. Кейсы. Дискуссионные процедуры. Мини-тест
7.	Многомерные и реляционные модели данных.	Устный опрос. Кейсы. Дискуссионные процедуры. Мини-тест
8.	Витрины данных.	Устный опрос. Кейсы. Дискуссионные процедуры. Мини-тест
9.	Загрузка данных в хранилище.	Устный опрос. Кейсы. Дискуссионные процедуры. Мини-тест
10.	Метод Data Vault.	Устный опрос. Кейсы. Дискуссионные процедуры. Мини-тест
11.	Программирование на языке Python для сбора и анализа данных.	Устный опрос. Кейсы. Дискуссионные процедуры. Мини-тест

3.1.1 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности в процессе текущего контроля успеваемости

Устный опрос. Кейсы (ситуации и задачи с заданными условиями). Дискуссионные процедуры (круглый стол, дискуссия, полемика, диспут, дебаты, мини-конференции)

Занятие № 1. Основы работы в Loginom

Вопросы для устного опроса:

1. Понятие пакета. Принципы работы с пакетами.
2. Модули, отчеты, ссылки.
3. Компоненты и узлы.
4. Сценарии обработки, правила разработки сценариев.
5. Группировка узлов при создании сценариев.
6. Запуск узла на выполнение и приостановка работы узла.
7. Использование переменных при создании сценариев.
8. Порты, настройка портов. Автосинхронизация.
9. Настройка входных портов. Конфигурирование порта.
10. Навигация по сценарию.
11. Модификаторы доступа, настройки модификаторов доступа для узлов.
12. Компоненты импорта в Loginom.
13. Компоненты фильтр строк и сортировка.
14. Обработка в фильтре пустых значений.
15. Типы полей в Loginom.
16. Операция синхронизации полей.
17. Обязательные и необязательные столбцы.
18. Проектирование без данных.
19. Компонент калькулятор.
20. Правила составления в компоненте калькулятор. Функции калькулятора.
21. Визуализаторы в Loginom. Настройка визуализаторов.
22. Переменные и параметризация узлов
23. Компоненты условие и замена.
24. Компонент цикл.
25. Подмодели.
26. Наследование и производные компоненты.
27. Импорт из промышленных источников данных.

Вопросы для групповой дискуссии

1. Какие существуют редакции платформы LOGINOM с локальным режимом запуска?
2. Сколько пакетов одновременно можно открыть в клиентском приложении Loginom Studio или Loginom Desktop?
3. Из каких групп объектов состоит пакет?
4. Как определить, что узел сценария активирован?
5. С какими видами объектов может работать каждый узел сценария?
6. Для чего используются порты для узла?
7. Как изображаются обязательные и необязательные порты, в чем состоит их отличие друг от друга?
8. Для чего применяются модификаторы доступа? Какие есть модификаторы доступа в Loginom?
9. Какие группы компонент импорта присутствуют в Loginom, и чем они отличаются?
10. Каким образом можно задать условие фильтрации в компоненте: фильтр?
11. Как обработать пустые значения в Loginom?
12. Что такое меппинг?
13. Как происходит меппинг?

14. Что означает проектирование без данных?
15. Что означает механизм «ленивых вычислений» используемый в компоненте: Калькулятор.
16. Для чего используется компонент Калькулятор?
17. При импорте текстового файла у части столбцов конвертация типов данных происходит с потерями. Что может помочь для исправления ситуации?
18. Какие типы данных совместимы между собой в Loginom?
19. Какие метаданные столбцов не принимаются во внимание алгоритмом автоматического связывания?
20. Имеется цело поле: Количество визитов с числом посещений клиентом магазина. Требуется преобразовать его в три значения: 1 визит; 2-4 визита; 5 и более. Каким компонентом можно это сделать?
21. Может ли в реальном сценарии присутствовать узел Подмодель, который не имеет входных портов?
22. Компонент Цикл. Какой тип цикла понадобится, если требуется обработать все четные строки набора данных?
23. Можно ли компонент Цикл настроить на компонент Узел-ссылка?

Кейсы (решение задач):

Задание 1.

1. Используйте файл «Товары.txt», который имеет следующие поля:
 - Артикул – артикул товара;
 - Наименование – наименование товара;
 - Дата продажи – дата последней продажи товара.
 - Тип клиента – поле принимает значения Тип 1 или Тип 2.
2. Создайте пакет в Loginom.
3. Импортируйте файл «Товары.txt».
4. Исключите из набора записи, в которых для товара отсутствует Артикул.
5. Рассчитайте сколько месяцев прошло от даты последней покупки каждого товара до 01.04.2018
6. Отсортируйте набор данных по количеству месяцев по убыванию.
7. Добавьте поле логического типа: Вывод из продажи и установите значение True, для товаров у которых от даты последней продажи до 01.04.2018 прошло более 10 месяцев.

Задание 2.

8. Используйте файл «Вероятности визитов.txt», который имеет следующие поля:
9. Код клиента – уникальный идентификатор клиента;
10. Отклик – поле принимает значения True или False;
11. Вероятность визита – поле принимает значения в диапазоне от 0 до 1;
12. Признак тестового множества – поле принимает значение True или False;
13. Тип клиента – поле принимает значения Тип 1 или Тип 2.
14. Разработайте сценарий который рассчитывает долгосрочную стоимость клиента CLV – Customer Lifetime Value по следующей формуле: $CLV = m(r / (1+I-r))$, где:
 - 15. m – ожидаемый годовой доход с одного клиента в рублях;
 - 16. r – коэффициент удержания клиента, от 0 до 1.
 - 17. I – годовая ставка дисконтирования в долях, от 0 до 1.
18. Рассчитайте для каждого клиента из файла «Вероятности визитов.txt» значение показателя CLV, если ставка дисконтирования равна 12%, а коэффициент удержания и ожидаемый годовой доход зависят от типа клиента:
 - 19. $r = 0,9$ и $m = 7200$ для клиентов с типом 1;
 - 20. $r = 0,7$ и $m = 9500$ для клиентов с типом 2;
21. Переделайте то же самое для другой формулы CLV с учетом ежегодного роста прибыли: $CLV = m(1+I - r / (1+g))$, где $g = 8\%$.

22. Отсортируйте по убыванию CLV.
23. При решении задач используйте подмодели, производные компоненты и наследование.

Занятие № 2. Регулярные выражения. Очистка регионов

Вопросы для устного опроса:

1. Очистка данных.
2. Виды очистки.
3. Правила составления регулярных выражений.
4. Очистка данных без использования регулярных выражений.
5. Функции калькулятора: RegExDomain, RegExMail, RegExMatch, RegExMatchCount, RegExMatchedExp, RegExMatchedNamedSubExp и другие.
6. Дополнительные функции калькулятора для работы со строками.

Вопросы для групповой дискуссии:

1. Что означает ошибка: *неправдоподобие данных*?
2. Словарь, состоящий из объектов, классификационных группировок, на которые они разбиты по степени сходства, и идентифицирующих их кодов, называется:
3. Чему равно расстояние Дамерау-Левенштейна между словами *бумага* и *уборка*?
4. Какие расстояния над строками учитываются при расчете расстояния Дамерау-Левенштейна?
5. От какого языка программирования ведут свою историю большинство диалектов *регулярных выражений*?
6. Что означает комбинация символов \.+ в регулярных выражениях Perl?
7. Выберите команд, для которой метасимвол \E не является символом конца ее действия:
8. Какой шаг типового алгоритма необходимо повторять несколько раз в процессе очистки?

Кейсы (решение задач):

Задание 1.

1. Создайте пакет Loginom и импортируйте в него данные из файла «organization_samples.lgd».
2. Разработайте регулярное выражение, которое может осуществить поиск и выделение из строки организационно-правовую форму: ООО, ОАО, ЗАО и ИП.
3. Изучите функции Калькулятора по работе со строками, полезными при очистке данных: Count, DamDevDist, Replace, Space, SubStr.
4. Скомбинируйте дополнительные функции таким образом, чтобы с их помощью выделить организационно-правовую форму или название организации, не используя регулярные выражения.

Занятие № 3. Проектирование хранилища данных типа ROLAP

Вопросы для устного опроса:

1. Основной принцип ROLAP.
2. Таблица фактов и таблицы измерений.
3. Схема «Звезда».
4. Схема «Снежинка».
5. Сравнение схем ROLAP.
6. Доступ к данным в ROLAP - системах.
7. Принципы Кодда для OLAP – систем.
8. Сравнение архитектур MOLAP и ROLAP.
9. Семантический слой хранилища данных.
10. Компонент: Метаданные.

Вопросы для групповой дискуссии:

1. Дайте определение хранилища данных.
2. В чем состоит идея принципа ROLAP?

3. Кто сформулировал основные принципы реляционной модели данных?
4. Какие принципы построения ROLAP, различают на логическом уровне?
5. Какие логические схемы архитектуры ХД ROLAP позволяют работать с иерархическими измерениями?
6. Какая схема архитектуры ROLAP более производительна?
7. Какой принцип лежит в основе логической структуры типа: «звезда»?
8. Как определяется плотность заполнения гиперкуба?
9. Значения чего располагаются в ячейках гиперкуба?
10. Как осуществляется доступ к данным в ROLAP - системах?
11. Перечислите типы данных СУБД Vertica.
12. Для чего в ХД используется семантический слой?
13. В чем состоит задача компонента: Метаданные

Кейсы (решение задач):

Задание. Проектирование ХД по схеме «Снежинка»:

1. Выполните проектирование хранилища данных типа «Снежинка», на основе данных о продажах фармацевтической сети.
2. Перед проектированием изучите структуру имеющихся данных и определите какие данные относятся к фактам, а какие к измерениям.
3. Определите таблицы фактов и измерений.
4. Постройте схему хранилища в виде ER- модели, используя программу DBeaver.
5. Создайте новую базу данных в колоночной СУБД Vertica и выполните подключение к ней в программе DBeaver.
6. Создайте таблицы хранилища данных в DBeaver и свяжите их.
7. Выполните загрузку данных, используя Loginom.
8. Экспортируйте данные в созданное хранилище.
9. Выполните настройку компонента Метаданные, и получите различные срезы хранилища данных: задание значений семантического слоя, настройка фильтров, получение данных.
10. Организуйте несколько SQL запросов к ХД.
11. Подготовиться к ответам на вопросы.

Занятие № 4. Проектирование хранилища данных типа Data Vault

Вопросы для устного опроса:

1. Основные концепции и определение Data Vault.
2. Основные компоненты модели Data Vault: концентраторы, сссылки, спутники.
3. Пример модели Data Vault.
4. Обобщенная схема хранилища Data Vault.
5. Преимущества и недостатки Data Vault.

Вопросы для групповой дискуссии:

1. Какие проблемы может решить Data Vault?
2. Какое количество вспомогательных алгоритмов может присутствовать в основном алгоритме?
3. Какой из компонентов модели Data Vault используется для представления бизнес-сущностей?
4. В основе идеи метода Data Vault лежит переход от предметно-ориентированного подхода к ...?
5. Сколько родительских таблиц может быть у спутника в модели Data Vault?
6. В каком порядке происходит загрузка элементов Data Vault в хранилище данных?
7. Для чего используется таблица ссылок в Data Vault?
8. Для чего используются спутники в Data Vault?

Кейсы (решение задач):

Задание. Проектирование ХД по схеме «Data Vault»:

1. Выполните проектирование хранилища данных типа «Data Vault», на основе

- данных о продажах фармацевтической сети.
2. Определите бизнес-ключи, каждый из которых будет представлять отдельная таблица-концентратор.
 3. Сформируйте таблицу ссылок.
 4. Сформируйте таблицы спутники.
 5. Постройте схему хранилища в виде ER- модели, используя программу DBeaver.
 6. Выполните проектирование хранилища данных по схеме Data Vault в программе DBeaver.
 7. Создайте новую базу данных в колоночной СУБД Vertica и выполните подключение к ней в программе DBeaver.
 8. Импортируйте данных в созданное хранилище данных, используя Loginom.
 9. Подготовиться к ответам на вопросы.

Занятие № 5. Программирование на языке Python для сбора и анализа данных.

Вопросы для устного опроса:

1. Переменные, операторы, управляющие конструкции, циклы.
2. Типы данных в Python: строки, списки, кортежи, словари, множества.
3. Функции, их определение.
4. Элементы функционального программирования.
5. Ввод-вывод. Работа с файлами различных форматов: CSV, XML, JSON и др.
6. Использование регулярных выражений для обработки текстов и извлечения данных.
7. Чтение и обработка (парсинг) веб-страниц (файлов HTML) с помощью BeautifulSoup, извлечение данных из веб-страниц.
8. Работа с внешними API различных сервисов.
9. Математические и статистические инструменты Python.

Вопросы для самоконтроля:

1. Каким образом реализуются в языке Python базовые алгоритмические конструкции?
2. Как реализуются различные типы данных в языке программирования Python?
3. В чем состоят особенности вызова функций в языке программирования Python?
4. В чем состоит идея функционального программирования?
5. Для чего используют лямбда-функции в Python?
6. Для чего используется библиотека Pandas?
7. Как реализуется парсинг веб-страниц в Python?

Кейсы (решение задач):

1. Преобразовать 1000 файлов из формата Excel в формат CSV за несколько секунд.
2. Взять с сайта ЦИК данные о результатах произвольных выборов по всем участковым избирательным комиссиям, найти участки, на которых кандидат-победитель набрал более 80% голосов, построить по ним диаграмму рассеяния «явка- результат», найти коэффициенты регрессии.
3. Скачать все статьи, опубликованные в некотором СМИ за последние три месяца; построить график зависимости частоты использования ряда слов-маркеров в комментариях к статьям в зависимости от времени публикации комментария.
4. С помощью API сайта Государственной Думы, найти всех действующих и бывших депутатов, имеющих учёную степень доктора наук.

Мини-тест

1. Как зовут математика, который впервые упомянул задачу нахождения расстояния между строками?
 1. Стивен Клини;
 2. Владимир Левенштейн;
 3. Уорен Мак Каллох;
 4. Фредерик Дамерау;

2. Какие строки полностью соответствуют *регулярному выражению*:
[a-я]{3,}\s?\d?\s[A-я]+
 1. п. Лесной;
 2. ул. Королева, 12;
 3. улица 1 мая;
 4. город Владивосток;
 5. Проспект Вернадского;
 6. морской вокзал;
 7. Москва;
 3. Что обозначают *мнимые метасимволы*?
 1. позицию символов в строке;
 2. начало и конец строки;
 3. экранированный символ;
 4. определенный символ в строке;
 4. Какой компонент из списка не имеет входных портов:
 1. Текстовый файл (экспорт);
 2. Узел ссылка;
 3. Текстовый файл (импорт);
 5. Как в сценарии воспользоваться узлом, находящимся в другом модуле текущего пакета?
 1. использовать компонент Узел-ссылка;
 2. добавить ссылку на другой модуль через соответствующую команду в дереве пакетов;
 3. нет такой возможности;
 6. Какие метаданные столбцов не принимаются во внимание алгоритмом автоматического связывания?
 1. вид данных;
 2. тип данных;
 3. метка столбца;
 4. имя столбца;
 7. Сколько родительских таблиц может быть у спутника в модели Data Vault?
 1. любое количество;
 2. только одна;
 3. ни одной;
 8. Первым шагом при разработке логической структуры витрины данных является:
 1. выбор главной сущности;
 2. Исследование организационной структуры компании;
 3. Обеспечение администрирования;
 4. Проектирование интерфейса пользователя;
- Ключ:** 1-2; 2-(3,4,6); 3-1; 4-3; 5-1; 6-1; 7-2; 8-1

3.1.2. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности в ходе текущего контроля успеваемости

Устный ответ

Оценка знаний предполагает дифференцированный подход к обучающемуся, учет его индивидуальных способностей, степень усвоения и систематизации основных понятий и категорий по дисциплине. Кроме того, оценивается не только глубина знаний поставленных вопросов, но и умение использовать в ответе практический материал. Оценивается культура речи, владение навыками ораторского искусства.

Критерии оценивания: последовательность, полнота, логичность изложения, анализ различных точек зрения, самостоятельное обобщение материала, использование профессиональных терминов, культура речи, навыки ораторского искусства. Изложение материала без фактических ошибок.

Оценка «отлично» ставится в случае, когда материал излагается исчерпывающе,

последовательно, грамотно и логически стройно, при этом раскрываются не только основные понятия, но и анализируются точки зрения различных авторов. Обучающийся не затрудняется с ответом, соблюдает культуру речи.

Оценка «хорошо» ставится, если обучающийся твердо знает материал, грамотно и по существу излагает его, знает практическую базу, но при ответе на вопрос допускает несущественные погрешности.

Оценка «удовлетворительно» ставится, если обучающийся освоил только основной материал, но не знает отдельных деталей, допускает неточности, недостаточно правильные формулировки, нарушает последовательность в изложении материала, затрудняется с ответами, показывает отсутствие должной связи между анализом, аргументацией и выводами.

Оценка «неудовлетворительно» ставится, если обучающийся не отвечает на поставленные вопросы.

Кейсы (ситуации и задачи с заданными условиями)

Обучающийся должен уметь выделить основные положения из текста задачи, которые требуют анализа и служат условиями решения. Исходя из поставленного вопроса в задаче, попытаться максимально точно определить проблему и соответственно решить ее.

Задачи могут решаться устно и/или письменно. При решении задач также важно правильно сформулировать и записать вопросы, начиная с более общих и, кончая частными.

Критерии оценивания – оценка учитывает методы и средства, использованные при решении ситуационной, проблемной задачи.

Оценка «отлично» ставится в случае, когда обучающийся выполнил задание (решил задачу), используя в полном объеме теоретические знания и практические навыки, полученные в процессе обучения.

Оценка «хорошо» ставится, если обучающийся в целом выполнил все требования, но не совсем четко определяется опора на теоретические положения, изложенные в научной литературе по данному вопросу.

Оценка «удовлетворительно» ставится, если обучающийся показал положительные результаты в процессе решения задачи.

Оценка «неудовлетворительно» ставится, если обучающийся не выполнил все требования.

Дискуссионные процедуры

Круглый стол, дискуссия, полемика, диспут, дебаты, мини-конференции являются средствами, позволяющими включить обучающихся в процесс обсуждения спорного вопроса, проблемы и оценить их умение аргументировать собственную точку зрения. Задание дается заранее, определяется круг вопросов для обсуждения, группы участников этого обсуждения.

Дискуссионные процедуры могут быть использованы для того, чтобы студенты:

– лучше поняли усвояемый материал на фоне разнообразных позиций и мнений, не обязательно достигая общего мнения;

– смогли постичь смысл изучаемого материала, который иногда чувствуют интуитивно, но не могут высказать вербально, четко и ясно, или конструировать новый смысл, новую позицию;

– смогли согласовать свою позицию или действия относительно обсуждаемой проблемы.

Критерии оценивания – оцениваются действия всех участников группы. Понимание проблемы, высказывания и действия полностью соответствуют заданным целям. Соответствие реальной действительности решений, выработанных в ходе игры. Владение терминологией, демонстрация владения учебным материалом по теме игры, владение методами аргументации, умение работать в группе (умение слушать, конструктивно вести

беседу, убеждать, управлять временем, бесконфликтно общаться), достижение игровых целей, (соответствие роли – при ролевой игре). Ясность и стиль изложения.

Оценка *«отлично»* ставится в случае, когда все требования выполнены в полном объеме.

Оценка *«хорошо»* ставится, если обучающиеся в целом демонстрируют понимание проблемы, высказывания и действия полностью соответствуют заданным целям. Решения, выработанные в ходе игры, полностью соответствуют реальной действительности. Но некоторые объяснения не совсем аргументированы, нарушены нормы общения, нарушены временные рамки, нарушен стиль изложения.

Оценка *«удовлетворительно»* ставится, если обучающиеся в целом демонстрируют понимание проблемы, высказывания и действия в целом соответствуют заданным целям. Однако, решения, выработанные в ходе игры, не совсем соответствуют реальной действительности. Некоторые объяснения не совсем аргументированы, нарушены временные рамки, нарушен стиль изложения.

Оценка *«неудовлетворительно»* ставится, если обучающиеся не понимают проблему, их высказывания не соответствуют заданным целям.

Тестирование

Является одним из средств контроля знаний обучающихся по дисциплине (модулю).

Критерии оценивания – правильный ответ на вопрос

Оценка *«отлично»* ставится в случае, если правильно выполнено 90-100% заданий.

Оценка *«хорошо»* ставится, если правильно выполнено 70-89% заданий.

Оценка *«удовлетворительно»* ставится в случае, если правильно выполнено 50-69% заданий.

Оценка *«неудовлетворительно»* ставится, если правильно выполнено менее 50% заданий.

3.2. Оценочные материалы для проведения промежуточной аттестации

3.2.1. Критерии оценки результатов обучения по дисциплине (модулю)

Шкала оценивания	Результаты обучения	Показатели оценивания результатов обучения
ОТЛИЧНО	Знает:	- обучающийся глубоко и всесторонне усвоил материал, уверенно, логично, последовательно и грамотно его излагает, опираясь на знания основной и дополнительной литературы, - на основе системных научных знаний делает квалифицированные выводы и обобщения, свободно оперирует категориями и понятиями.
	Умеет:	- обучающийся умеет самостоятельно и правильно решать учебно-профессиональные задачи или задания, уверенно, логично, последовательно и аргументировано излагать свое решение, используя научные понятия, ссылаясь на нормативную базу.
	Владеет:	- обучающийся владеет рациональными методами (с использованием рациональных методик) решения сложных профессиональных задач, представленных деловыми играми, кейсами и т.д.; При решении продемонстрировал навыки - выделения главного, - связкой теоретических положений с требованиями руководящих документов, - изложения мыслей в логической последовательности, - самостоятельного анализа факты, событий, явлений, процессов в их взаимосвязи и диалектическом развитии.
ХОРОШО	Знает:	- обучающийся твердо усвоил материал, достаточно грамотно его излагает, опираясь на знания основной и дополнительной литературы, - затрудняется в формулировании квалифицированных выводов и обобщений, оперирует категориями и понятиями, но не всегда правильно их верифицирует.
	Умеет:	- обучающийся умеет самостоятельно и в основном правильно решать учебно-профессиональные задачи или задания, уверенно, логично, последовательно и аргументировано излагать свое решение, не в

		полной мере используя научные понятия и ссылки на нормативную базу.
	Владеет:	- обучающийся в целом владеет рациональными методами решения сложных профессиональных задач, представленных деловыми играми, кейсами и т.д.; При решении смог продемонстрировать достаточность, но не глубинность навыков, - выделения главного, - изложения мыслей в логической последовательности, - связи теоретических положений с требованиями руководящих документов, - самостоятельного анализа факты, событий, явлений, процессов в их взаимосвязи и диалектическом развитии.
УДОВЛЕТВОРИТЕЛЬНО	Знает:	- обучающийся ориентируется в материале, однако затрудняется в его изложении; - показывает недостаточность знаний основной и дополнительной литературы; - слабо аргументирует научные положения; - практически не способен сформулировать выводы и обобщения; - частично владеет системой понятий.
	Умеет:	- обучающийся в основном умеет решить учебно-профессиональную задачу или задание, но допускает ошибки, слабо аргументирует свое решение, недостаточно использует научные понятия и руководящие документы.
	Владеет:	- обучающийся владеет некоторыми рациональными методами решения сложных профессиональных задач, представленных деловыми играми, кейсами и т.д.; При решении продемонстрировал недостаточность навыков - выделения главного, - изложения мыслей в логической последовательности, - связи теоретических положений с требованиями руководящих документов, - самостоятельного анализа факты, событий, явлений, процессов в их взаимосвязи и диалектическом развитии.
НЕУДОВЛЕТВОРИТЕЛЬНО	Знает:	- обучающийся не усвоил значительной части материала; - не может аргументировать научные положения; - не формулирует квалифицированных выводов и обобщений; - не владеет системой понятий.
	Умеет:	обучающийся не показал умение решать учебно-профессиональную задачу или задание.
	Владеет:	не выполнены требования, предъявляемые к навыкам, оцениваемым «удовлетворительно».

3.2.2. Контрольные задания и/или иные материалы для проведения промежуточной аттестации

Список вопросов для устных ответов (варианты теста)

1. Какие существуют редакции платформы LOGINOM с локальным режимом запуска?
2. Сколько пакетов одновременно можно открыть в клиентском приложении Loginom Studio или Loginom Desktop?
3. Из каких групп объектов состоит пакет?
4. Как определить, что узел сценария активирован?
5. С какими видами объектов может работать каждый узел сценария?
6. Для чего используются порты для узла?
7. Как изображаются порты, работающие с набором переменных?
8. Как изображаются порты, работающие с набором входных данных?
9. Каким образом активируются порты переменных сценария?
10. Как осуществляется настройка портов?
11. Как изображаются обязательные и необязательные порты, в чем состоит их отличие друг от друга?

12. Для чего применяются модификаторы доступа? Какие есть модификаторы доступа в Loginom?
13. Какие группы компонент импорта присутствуют в Loginom, и чем они отличаются?
14. Каким образом можно задать условие фильтрации в компоненте: фильтр?
15. Как обработать пустые значения в Loginom?
16. Какие типы полей существуют в Loginom?
17. Что такое меппинг?
18. Как происходит меппинг?
19. Что означает проектирование без данных?
20. Что означает механизм «ленивых вычислений» используемый в компоненте: Калькулятор.
21. Для чего используется компонент Калькулятор?
22. Перечислите основные правила составления выражений в компоненте Калькулятор.
23. Какие категории функций присутствуют в калькуляторе?
24. Как осуществляется настройка визуализатора в Loginom?
25. Может ли узел быть активным, а один из его обязательных портов неконфигурированным?
26. Как в сценарии воспользоваться узлом, находящимся в другом модуле текущего пакета?
27. При импорте текстового файла у части столбцов конвертация типов данных происходит с потерями. Что может помочь для исправления ситуации?
28. Какие типы данных совместимы между собой в Loginom?
29. Какие метаданные столбцов не принимаются во внимание алгоритмом автоматического связывания?
30. Имеется цело поле: Количество визитов с числом посещений клиентом магазина. Требуется преобразовать его в три значения: 1 визит; 2-4 визита; 5 и более. Каким компонентом можно это сделать?
31. Может ли в реальном сценарии присутствовать узел Подмодель, который не имеет входных портов?
32. Компонент Цикл. Какой тип цикла понадобится, если требуется обработать все четные строки набора данных?
33. Можно ли компонент Цикл настроить на компонент Узел-ссылка?
34. Как визуально понять, что на полотне производный узел, а не базовый?
35. Можно ли разработать сценарий не имея данных?
36. Как реализуется механизм наследования?
37. Вы настроили узел N1 на компоненте Калькулятор с полем C, содержащим выражение $A+B$, где A и B – поля. Затем создали производный компонент N2 и поменяли в его экземпляре выражение на $A+B+1$. После этого в базовом узле, N1, поменяли выражение на $A+B+3$ и обновили конфигурацию производного узла. Какое выражение в итоге будет в узле N2.
38. Вы настроили узел N1 на компоненте Калькулятор с полем C, содержащим выражение $A+B$, где A и B – поля. Затем создали производный компонент N2 и поменяли в его экземпляре выражение на $A+B+1$. После этого в базовом узле, N1, поменяли выражение на $A+B+3$ и восстановили исходную конфигурацию производного узла. Какое выражение в итоге будет в узле N2.

Тексты проблемно-аналитических и (или) практических учебно-профессиональных задач

1. Используйте файл «Вероятности визитов.txt», который имеет следующие поля:
 - Код клиента – уникальный идентификатор клиента;
 - Отклик – поле принимает значения True или False;
 - Вероятность визита – поле принимает значения в диапазоне от 0 до 1;

- Признак тестового множества – поле принимает значение True или False;
 - Тип клиента – поле принимает значения Тип 1 или Тип 2.
2. Разработайте сценарий который рассчитывает долгосрочную стоимость клиента CLV – Customer Lifetime Value по следующей формуле: $CLV = m(r / (1+I-r))$, где:
 - m – ожидаемый годовой доход с одного клиента в рублях;
 - r - коэффициент удержания клиента, от 0 до 1.
 - I – годовая ставка дисконтирования в долях, от 0 до 1.
 3. Рассчитайте для каждого клиента из файла «Вероятности визитов.txt» значение показателя CLV, если ставка дисконтирования равна 12%, а коэффициент удержания и ожидаемый годовой доход зависят от типа клиента:
 - r = 0,9 и m = 7200 для клиентов с типом 1;
 - r = 0,7 и m = 9500 для клиентов с типом 2;
 4. Переделайте то же самое для другой формулы CLV с учетом ежегодного роста прибыли: $CLV = m(1+I - r / (1+g))$, где g = 8%.
 5. Отсортируйте по убыванию CLV.
 6. При решении задач используйте подмодели, производные компоненты и наследование.

3.2.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков в ходе промежуточной аттестации

Процедура оценивания знаний (устный ответ)

Предел длительности	10 минут
Предлагаемое количество заданий	2 вопроса
Последовательность выборки вопросов из каждого раздела	Случайная
Критерии оценки	<ul style="list-style-type: none"> - требуемый объем и структура - изложение материала без фактических ошибок - логика изложения - использование соответствующей терминологии - стиль речи и культура речи - подбор примеров их научной литературы и практики
«5» если	требования к ответу выполнены в полном объеме
«4» если	в целом выполнены требования к ответу, однако есть небольшие неточности в изложении некоторых вопросов
«3» если	требования выполнены частично – не выдержан объем, есть фактические ошибки, нарушена логика изложения, недостаточно используется соответствующая терминологии

Процедура оценивания умений и навыков (решение проблемно-аналитических и практических учебно-профессиональных задач)

Предлагаемое количество заданий	1
Последовательность выборки	Случайная
Критерии оценки:	<ul style="list-style-type: none"> - выделение и понимание проблемы - умение обобщать, сопоставлять различные точки зрения - полнота использования источников - наличие авторской позиции - соответствие ответа поставленному вопросу - использование социального опыта, материалов СМИ, статистических данных - логичность изложения - умение сделать квалифицированные выводы и обобщения с точки зрения решения профессиональных задач - умение привести пример - опора на теоретические положения - владение соответствующей терминологией
«5» если	требования к ответу выполнены в полном объеме
«4» если	в целом выполнены требования к ответу, однако есть небольшие неточности в изложении некоторых вопросов. Затрудняется в формулировании квалифицированных

	выводов и обобщений
«3» если	требования выполнены частично – пытается обосновать свою точку зрения, однако слабо аргументирует научные положения, практически не способен самостоятельно сформулировать выводы и обобщения, не видит связь с профессиональной деятельностью

4. Учебно-методическое и материально-техническое обеспечение дисциплины (модуля)

4.1. Электронные учебные издания

1. Дронова, Е. Н. Программное обеспечение ЭВМ и технологии обработки информации : учебное пособие / Е. Н. Дронова. — Барнаул : Алтайский государственный педагогический университет, 2018. — 138 с. — ISBN 978-5-88210-911-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/102760.html>. — Режим доступа: для авторизир. пользователей
2. Кандаурова, Н. В. Технологии обработки информации : учебное пособие / Н. В. Кандаурова, В. С. Чеканов. — Ставрополь : Северо-Кавказский федеральный университет, 2014. — 175 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/63145.html>. — Режим доступа: для авторизир. пользователей
3. Кучинский, В. Ф. Сетевые технологии обработки информации : учебное пособие / В. Ф. Кучинский. — Санкт-Петербург : Университет ИТМО, 2015. — 118 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/68119.html>. — Режим доступа: для авторизир. пользователей

4.2. Электронные образовательные ресурсы

1. Электронная библиотечная система «ЭБС ЮРАЙТ» Biblio-online.ru (ЭБС «Юрайт») [Электронный ресурс]. — URL: <https://urait.ru/>.
2. Электронно-библиотечная система ZNANIUM [Электронный ресурс]. — URL: <https://znanium.com/>.
3. Электронная библиотечная система «Консультант студента» [Электронный ресурс]. — URL: <https://www.studentlibrary.ru/>.
4. e-Library.ru: Научная электронная библиотека [Электронный ресурс]. — URL: <http://elibrary.ru/>.
5. Научная электронная библиотека «КиберЛенинка» [Электронный ресурс]. — URL: <http://cyberleninka.ru/>.
6. Информационная система «Единое окно доступа к образовательным ресурсам» [Электронный ресурс]. — URL: <http://window.edu.ru/>.
7. Федеральный центр информационно-образовательных ресурсов [Электронный ресурс]. — URL: <http://fcior.edu.ru/>.

4.3. Современные профессиональные базы данных и информационные справочные системы

Обучающимся обеспечен доступ (удаленный доступ) к ниже следующим современным профессиональным базам данных и информационным справочным системам:

1. Словари и энциклопедии на Академике [Электронный ресурс]. — URL: <http://dic.academic.ru>.
2. Система информационно-правового обеспечения «Гарант» [Электронный ресурс]. — URL: <http://ivo.garant.ru/>.

4.4. Комплект лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства

1. Лицензионное программное обеспечение: операционная система Microsoft Windows, пакет офисных приложений Microsoft Office.

2. Свободно распространяемое программное обеспечение: свободные пакеты офисных приложений Apache Open Office, LibreOffice.
3. Программное обеспечение отечественного производства: справочно-правовая система «Гарант» (Электронный периодический справочник «Система ГАРАНТ»), образовательная платформа ЮРАЙТ (Электронная библиотечная система «ЭБС ЮРАЙТ» Biblio-online.ru (ЭБС «Юрайт»)), электронно-библиотечная система ZNANIUM, электронная библиотечная система «Консультант студента».

4.5. Оборудование и технические средства обучения

Для реализации дисциплины (модуля) используются учебные аудитории для проведения учебных занятий, которые оснащены оборудованием и техническими средствами обучения, и помещения для самостоятельной работы обучающихся, которые оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечены доступом в электронную информационно-образовательную среду РХТУ им. Д.И. Менделеева. Допускается замена оборудования его виртуальными аналогами.

Наименование учебных аудиторий для проведения учебных занятий и помещений для самостоятельной работы*	Оснащенность учебных аудиторий для проведения учебных занятий и помещений для самостоятельной работы оборудованием и техническими средствами обучения
Учебные аудитории для проведения учебных занятий	Учебная аудитория укомплектована специализированной мебелью, отвечающей всем установленным нормам и требованиям, оборудованием и техническими средствами обучения (мобильное мультимедийное оборудование).
Помещение для самостоятельной работы	Помещение оснащено компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду РХТУ им. Д.И. Менделеева и к ЭБС.

* Номер конкретной аудитории указан в приказе об аудиторном фонде, расписании учебных занятий и расписании промежуточной аттестации.