|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  | Приложение 1 | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | **ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ** | | | | | | | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | **БОЛЬШИЕ ДАННЫЕ** | | | | | | | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | ***Назначение оценочных материалов*** | | | | | | | |  |  |
| Фонд оценочных материалов (ФОМ) создается в соответствии с требованиями ФГОС ВО для аттестацииобучающихся на соответствие их учебных достижений поэтапным требованиям основной профессиональнойобразовательной программе (ОПОП) при проведении входного и текущего оценивания, а также промежуточнойаттестации обучающихся. ФОС является составной частью нормативно-методического обеспечения системы оценкикачества освоения ОПОП ВО, входит в состав ОПОП.  Фонд оценочных материалов – комплект методических материалов, нормирующих процедуры оцениваниярезультатов обучения, т.е. установления соответствия учебных достижений запланированным результатам обучения итребованиям образовательных программ, рабочих программ модулей (дисциплин).  Фонд оценочных материалов сформирован на основе ключевых принципов оценивания:  - валидности: объекты оценки должны соответствовать поставленным целям обучения;  - надежности: использование единообразных стандартов и критериев для оценивания достижений;  - объективности: разные студенты должны иметь равные возможности добиться успеха.  Основными параметрами и свойствами ФОМ являются:  • предметная направленность (соответствие предмету изучения конкретной учебной дисциплины);  • содержание (состав и взаимосвязь структурных единиц, образующих содержание теоретической и практическойсоставляющих учебной дисциплины);  • объем (количественный состав оценочных средств, входящих в ФОМ);  • качество оценочных средств и ФОМ в целом, обеспечивающее получение объективных и достоверныхрезультатов при проведении контроля с различными целями.  Целью ФОМ является проверка сформированности у студентов компетенций: | | | | | | | | | | | |
| **УК-1** - Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач | | | | | | | | | | | |
| **ОПК-4** - Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности | | | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | ***Карта компетенций*** | | | | | | | |  |  |
| Контролируемые компетенции | | | | | Планируемый результат обучения | | | | | | |
|
| УК-1 - Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач | | | | | Знать: методы сбора, способы хранения и обработки информации. | | | | | | |
| Уметь: применять методы обработки данных | | | | | | |
| Владеть: методами интерпретации данных | | | | | | |
| ОПК-4 - Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности | | | | | Знать: технологии сбора, хранения и обработки больших объемов данных. | | | | | | |
| Уметь: выполнять обработку и визуализацию данных. | | | | | | |
| Владеть: методами обработки данных. | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | ***Показатели оценивания планируемых результатов обучения*** | | | | | | | |  |  |  |
| Элементы компетенций | | | Показатели оценивания | | | | Средства оценивания | | | | |
|
| УК-1  Знать | | | Знать: методы сбора, способы хранения и обработки информации. | | | | Подготовка к аудиторным занятиям  Зачет | | | | |
| УК-1  Уметь | | | Уметь: применять методы обработки данных | | | | Выполнение практических заданий  Подготовка к аудиторным занятиям  Зачет | | | | |
| УК-1  Владеть | | | Владеть: методами интерпретации данных | | | | Выполнение практических заданий  Подготовка к аудиторным занятиям  Зачет | | | | |
| ОПК-4  Знать | | | Знать: технологии сбора, хранения и обработки больших объемов данных. | | | | Подготовка к аудиторным занятиям  Зачет | | | | |
| ОПК-4  Уметь | | | Уметь: выполнять обработку и визуализацию данных. | | | | Выполнение практических заданий  Подготовка к аудиторным занятиям  Зачет | | | | |
| ОПК-4  Владеть | | | Владеть: методами обработки данных. | | | | Выполнение практических заданий  Подготовка к аудиторным занятиям  Зачет | | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | ***Матрица компетентностных задач по дисциплине*** | | | | | | | |  |  |  |
| Контролируемые блоки (темы) дисциплины | | | | Контролируемые компетенции (или их части) | | Оценочные средства | | | | | |
|
| Большие данные | | | | УК-1, ОПК-4 | | Практические задания  Вопросы для самостоятельного контроля знаний студентов | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | ***Оценочные средства*** | | | | | | | |  |  |  |
|  | ***Текущий контроль*** | | | | | | | |  |  |  |
| Целью текущего контроля знаний является установление подробной, реальной картины студенческих достиженийи успешности усвоения ими учебной программы на данный момент времени. В условиях рейтинговой системы контролярезультаты текущего оценивания студента используются как показатель его текущего рейтинга.  Текущий контроль успеваемости осуществляется в течение семестра, в ходе повседневной учебной работы поиндивидуальной инициативе преподавателя. Данный вид контроля стимулирует у студентов стремление ксистематической самостоятельной работе по изучению дисциплины. | | | | | | | | | | | |
|
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | **Описание видов практических занятий, предусмотренных РПД** | | | | | | | |  |  |  |
| **Выполнение** **практических** **заданий** | | | | | | | | | | | |
| Практические задания выдаются студентам с целью применения полученных знаний на практике под руководством преподавателя. Практические задания могут быть представлены в виде решения задач, проблемных заданий, тренингов и иных видах, направленных на получение практических знаний. | | | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | **Описание видов самостоятельной работы, предусмотренных РПД** | | | | | | | |  |  |  |
| **Подготовка** **к** **аудиторным** **занятиям** | | | | | | | | | | | |
| Подготовка к аудиторным занятиям состоит из изучения материала по соответствующей теме и ответов на вопросы для самоконтроля. Проверка уровня подготовки студентов к занятиям может проводится устным опросом, тестом, контрольной работой или иными видами текущего контроля. | | | | | | | | | | | |

.

Задания для текущего контроля

УК-1: Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач

Знать - методы сбора, способы хранения и обработки информации.

1. Среди шкал выберите те, с которыми невозможны арифметические операции:

Номинальная шкала (категориальная, наименований)

Порядковая шкала (ординальная, ранговая)

\*Интервальная шкала (разностей)

\*Шкала отношений (абсолютная)

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 50% правильных ответов.

1. Среди шкал выберите те, на которых невозможно ввести отношение порядка:

\*Номинальная шкала (категориальная, наименований)

Порядковая шкала (ординальная, ранговая)

Интервальная шкала (разностей)

Шкала отношений (абсолютная)

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Среди шкал выберите те, в которых мерой среднего может служить среднее арифметическое:

Номинальная шкала (категориальная, наименований)

Порядковая шкала (ординальная, ранговая)

\*Интервальная шкала (разностей)

\*Шкала отношений (абсолютная)

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 50% правильных ответов.

1. Среди шкал выберите те, в которых мерой среднего может служить медиана:

Номинальная шкала (категориальная, наименований)

\*Порядковая шкала (ординальная, ранговая)

\*Интервальная шкала (разностей)

\*Шкала отношений (абсолютная)

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

1. Что такое реляционная модель данных?

Модель данных, основанная на файловой структуре

\*Модель данных, основанная на связях между таблицами

Модель данных, основанная на иерархической структуре

Модель данных, основанная на объектах

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Совокупность программных и языковых средств для создания, удаления, изменения и любых других манипуляций с данными в БД:

\*система управления базами данных

реляционная база данных

утилита для формирования запросов к базе данных

визуализатор связей и отношений в базе данных

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Что такое таблица в реляционной модели данных?

Множество свойств, описывающих объекты

Список связей между объектами

\*Структура, хранящая данные в виде строк и столбцов

Файл, содержащий информацию о данных

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какие особенности реляционных баз данных?

Использование структуры типа таблицы с записями

Возможность связывания таблиц

Удобство в использовании и администрировании

\*Все перечисленные выше

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какие принципы лежат в основе реляционной модели данных?

\*Принципы ООП

Принципы структурного программирования

Принципы представления данных и их связей

Принципы функционального программирования

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Основными функциями СУБД являются:

\*Управление данными во внешней памяти

\*Управление буферами оперативной памяти

Автоматическое формирование запросов к БД на основе конструктора запросов

\*Поддержка языков базы данных

Поддержка удобного интерфейса работы с БД

\*Журнализация и резервное копирование базы данных

Визуализация связей таблиц и отношений

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 75% правильных ответов.

УК-1: Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач

Уметь - применять методы обработки данных

1. Отметьте верные задачи ОБРАБОТКИ и АНАЛИЗА больших данных фото и видеоформата:

\*Сегментация

\*Оптическое распознавание символоов

\*Обнаружение краев

\*Сопоставление шаблонов

Время, отводимое на выполнение задания (2 минуты).

Критерий оценивания: 75% правильных ответов.

1. С помощью каких ключевых технологий и архитектурных решений достигается высокая скорость записи и чтения больших файлов в серверных решениях на базе кластерной архитектуры?

\*Распределенная система обработки данных

\*Архитектура главный/подчиненный

Последовательная обработка данных

\*Репликация блоков данных

Использование операционной системы на базе ядра Linux

Время, отводимое на выполнение задания (2 минуты).

Критерий оценивания: 66% правильных ответов.

1. Чем характеризуются архитектуры главный/подчиненный в исполнении серверных решений?

\*Большое число объединенных стоек с несколькими серверными узлами

\*Один главный и несколько резервных узлов для управления стойками

\*Наличие большого числа точек записи в серверных узлах на стойках

\*Централизованная балансировка нагрузки по чтению, записи и обработки данных

Время, отводимое на выполнение задания (2 минуты).

Критерий оценивания: 75% правильных ответов.

1. Какие факторы ограничивают кратное повышение производительности в решениях на базе распределенных архитектур?

\*соединение исполнителей в сети

\*программная балансировка нагрузки на исполнителях

качество написания последовательной части программного кода

\*синхронизация исполнителей

\*помехи при передаче данных

формат передаваемой информации по каналу связи

Время, отводимое на выполнение задания (2 минуты).

Критерий оценивания: 75% правильных ответов.

1. Выберите верные свойства, присущие распределенной системе обработки и хранения данных:

\*совместное использование ресурсов

низкая скорость чтения

\*совместная обработка данных

проприеритарность

\*масштабируемость по программным и аппаратным ресурсам

\*обнаружение отказов

закрытость

Время, отводимое на выполнение задания (2 минуты).

Критерий оценивания: 75% правильных ответов.

1. В чем состоит существенная особенность временных рядов как данных табличного типа?

\*Содержат зависимые упорядоченные во времени последовательности значений

Временные промежутки между измерениями непостоянны

\*Содержат время в качестве отдельного показателя

\*Показывают развитие показателя одного конкретного объекта во времени

Данные показателей (столбцы) временных рядов (если их несколько) не зависят друг от друга

Время, отводимое на выполнение задания (2 минуты).

Критерий оценивания: 66% правильных ответов.

1. В формате данных JSON данные хранятся в виде:

Строк с разделителями столбцов, отличными от разделителя дробной части вещественных чисел

\*Коллекции данных формата "ключ-значение", где доступ к значениям полей данных осуществляется по ключу

В виде таблицы, с разлинованными строками и столбцами в виде отдельных объектов документа

В виде неструктурированной последовательности битов, расшифровывающейся в определенном формате как целостный объект

Время, отводимое на выполнение задания (3 минуты).

Критерий оценивания: 100% правильных ответов.

1. Отметьте верные примеры неструктурированных данных:

\*Файлы изображений

Таблицы данных транзакций

\*Файлы, хранящие текстовую информацию на естественном языке

Таблицы реляционной базы данных

\*Видеофайлы

Файлы формата JSON с перечнем объектов с одинаковым набором ключей

Время, отводимое на выполнение задания (2 минуты).

Критерий оценивания: 66% правильных ответов.

1. Выберите верные утверждения, относящиеся к понятию неструктурированных данных:

Организованная, типизированная информация, относящаяся к одной сущности

\*Не имеет предопределенной организации и имеет множество форм

Франят в долговременном хранилище данных, реляционных базах данных

\*Хранят в озёрах данных, файловых базы данных

Имеет несколько предопределенных форматов

\*Имеет большое количество различных форматов данных

Время, отводимое на выполнение задания (2 минуты).

Критерий оценивания: 66% правильных ответов.

1. Какие существуют основные свойства шкал измерений?

\*Идентифицируемость

\*Величина

Дискретизируемость

\*Равенство интервалов

Равенство значений

\*Абсолютный ноль

Квантуемость

Время, отводимое на выполнение задания (2 минуты).

Критерий оценивания: 75% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_\_\_\_\_ данные  — данные, которые не соответствуют заранее определённой модели данных, и, как правило, представлены в форме текста с датами, цифрами, фактами, расположенными в нём в произвольной форме.

Правильный ответ: Неструктурированные, неструктурированные

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_\_ данные являются значениями признака, общее число которых конечной или бесконечно, но может быть подсчитано при помощи натруральных чисел.

Правильный ответ: Дискретные, дискретные

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуски в определении

\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_ – это совокупность данных, хранящихся и упорядоченных в соответствии с определенной структурой.

Правильный ответ: База данных, база данных, БД

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуски в определении

\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_ определяет как и каким образом данные будут располагаться в БД и как к ним будет предоставляться доступ.

Правильный ответ: Модель данных, модель данных

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

Модель данных, предполагающая наличие ориентированной связи между объектами, от родительского объекта к дочерним называется \_\_\_\_\_\_\_\_\_\_

Правильный ответ: иерархической, иерархическая

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

Модель данных, предполагающая наличие связей между таблицами данных (отношений), основанных на значениях ключей в полях структурированных табличных объектов называется \_\_\_\_\_\_\_\_\_\_

Правильный ответ: реляционной

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

Для манипуляций над данными в реляционных СУБД используют декларативный язык запросов \_\_\_\_\_

Правильный ответ: SQL, sql, Sql, structured query language

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_ базы данных – список таблиц, их атрибутов, типов данных, ограничений, ключей и связей между таблицами, необходимый для корректной организации хранения данных в памяти вычислительного устройства и доступа к данным извне, как на запись, так и на чтение.

Правильный ответ: схема

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

Доступ к данным в реляционных СУБД классически осуществляется с помощью \_\_\_, подъязыка SQL

Правильный ответ: DML, dml, data manipulation language

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_ — группа операторов определения, манипуляции данных, переводящих базу данных из одного согласованного состояния в другое согласованное состояние.

Правильный ответ: Транзакция, транзакция

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

УК-1: Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач

Владеть - методами интерпретации данных

1. Анализ основных свойств данных, нахождение в них общих закономерностей, распределений и аномалий, построение начальных моделей на основе визуализаций называют:

\*разведочный АД

гибридный АД

кросс АД

независимый АД

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Операция выборки строк (объектов, наблюдений) из таблицы данных в соответствии с логическим правилом сравнения значений выбранного атрибута с определенным значением

называется:

\*фильтрацией

сортировкой

кластеризацией

сегментацией

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Несколько условий с логическими связками «И», «ИЛИ», «НЕ» использующихся для отбора нескольких значений из разных столбцов называется:

\*сложный фильтр

\*составной фильтр

независимый фильтр

колоночный фильтр

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Получение отсутствующей информации путём обработки уже имеющихся данных называется \_\_\_ обогащение данных:

\*локальное

составное

независимое

внешнее

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Получение отсутствующей информации путём получения из иных информационных систем называется \_\_\_ обогащение данных:

локальное

составное

независимое

\*внешнее

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Процесс изучения данных с целью достижения понимания их структуры, содержимого и оценки качества называется \_\_\_ данных:

сегментация

кластеризация

обогащение

\*профайлинг

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Значение, отличающееся от остальных наблюдений настолько, что его формирование предполагается под влиянием иных механизмов и факторов, чем у большинства других данных в наборе (до 3 σ) называется:

экстремальное значение

пик

импульс

\*выброс

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Значение, отличающееся от остальных наблюдений от 3 σ до 5 σ называется:

\*экстремальное значение

пик

импульс

выброс

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Уникальный атрибут внешней таблицы, создающий связь с данной по совпадающим значениям в столбце называется \_\_ ключ:

\*внешний

\*вторичный

первичный

независимый

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Список таблиц, их атрибутов, типов данных, ограничений, ключей и связей между таблицами, необходимый для корректной организации хранения данных в памяти ВУ и доступа к данным извне, как на запись, так и на чтение называется \_\_ БД:

\*схема

состав

сегмент

кластер

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 1000 записей. Какое максимальное количество записей с пропусками может быть в ней, чтобы их можно было проигнорировать на этапе профайлинга в рамках стратегии борьбы с пропусками: \_\_.

Правильный ответ: 10

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 5000 записей. Какое минимальное количество записей с пропусками на этапе профайлинга должно быть в ней, чтобы необходимо было отказаться от последующего анализа этого набора данных в рамках стратегии борьбы с пропусками: \_\_.

Правильный ответ: 2500

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 5000 записей. Задается фильтрация: фильтру A не удовлетворяют 200 записей, фильтру B – 300, фильтру C – 400. Какое максимальное количество записей может отображаться в БД после применения всех фильтраций: \_\_.

Правильный ответ: 4600

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 500 записей. Задается фильтрация: фильтру A не удовлетворяют 20 записей, фильтру B – 30, фильтру C – 40. Какое минимальное количество записей может отображаться в БД после применения всех фильтраций: \_\_.

Правильный ответ: 410

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Структура БД содержит 3 поля X, Y, Z. БД содержит 10 записей. Известно, что поля в БД содержат 9 пропусков. Какое максимальное количество чистых данных (записей без пропусков) возможно в такой БД: \_\_\_.

Правильный ответ: 7

Время выполнения: 3 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Структура БД содержит 3 поля X, Y, Z. БД содержит 10 записей. Известно, что поля в БД содержат 9 пропусков. Какое минимальное количество чистых данных (записей без пропусков) возможно в такой БД: \_\_\_.

Правильный ответ: 1

Время выполнения: 3 минуты.

Критерий оценивания: верный ответ соответствует 100%

1. Некоторая БД содержит 2 поля X и Y. X может быть целым числом от 0 до 10, а Y – пол и может принимать значения «М» или «Ж». Какое максимальное количество чистых данных (без дубликатов) может содержать такая БД: \_\_\_\_\_\_\_\_.

Правильный ответ: 22

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Какой объем памяти отводится для хранения вещественного числа в формате с плавающей запятой, если на мантиссу отводится 10 бит, а на порядок числа 5 бит: \_\_\_\_\_\_\_\_.

Правильный ответ: 16

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Некоторая задача решается исполнителем за время Т. 90% задачи можно решать параллельно. Найти во сколько раз уменьшится время решения задачи, если число исполнителей увеличится на 8, а накладными расходами на коммуникацию можно пренебречь: \_\_\_\_\_\_\_\_.

Правильный ответ: 5

Время выполнения: 3 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Непрерывная величина зависит от времени и характеризует вес курицы до 5 кг. Указать число уровней квантования при переводе такой информации из непрерывного вида в дискретный при помощи линейного квантования с интервалом 1 кг:\_\_\_\_\_\_\_\_.

Правильный ответ: 6

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

ОПК-4: Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

Знать - технологии сбора, хранения и обработки больших объемов данных.

1. Какой протокол используется для передачи данных между распределенными приложениями на Linux-серверах?

HTTP

SMTP

FTP

\*SSH

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какая директория на Linux-серверах используется для хранения системных журналов?

\*/var/logs

/logs

/usr/share/logs

/etc/logs

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какой тип данных используется для хранения URL-адресов?

\*VARCHAR

BOOLEAN

INT

\*TEXT

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какой тип данных используется для хранения изображений?

\*BLOB

BOOLEAN

INT

VARCHAR

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какой тип данных используется для хранения аудиофайлов?

AUDIO

INT

BOOLEAN

\*BLOB

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какой тип данных используется для хранения видеофайлов?

VIDEO

BOOLEAN

INT

\*BLOB

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Отметьте верные задачи ОБРАБОТКИ/ПРЕДОБРАБОТКИ неструктурированных данных, рассматривающиеся в обработке текстовой информации:

Сегментация текста

\*Структуризация текстовых данных

Поиск смысла в тексте

\*Векторизация текста

Анализ тональности текста

\*Генерация признаков на основе текста

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

1. Отметьте верные задачи АНАЛИЗА неструктурированных данных, рассматривающиеся в анализе текстовой информации:

\*Классификация отзывов

Стемминг, лемматизация

\*Кластеризация документов

\*Автодополнение поисковых запросов

Очистка текста от ошибок

\*Поиск именованных сущностей, ключевых слов

\*Сегментация текста

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 80% правильных ответов.

1. Отметьте верные задачи ОБРАБОТКИ и АНАЛИЗА больших данных аудиоформата:

\*очистка от шумов

дискретизация аналогового сигнала

\*выявление споттера (слово-триггер обычно для голосовых ассистентов)

\*разделение дорожек на речь и фоновый шум

\*выявление акцента / языка

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 75% правильных ответов.

1. Отметьте верные задачи ОБРАБОТКИ и АНАЛИЗА больших данных фото и видеоформата:

\*Сегментация

\*Оптическое распознавание символоов

\*Обнаружение краев

\*Сопоставление шаблонов

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 75% правильных ответов.

ОПК-4: Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

Уметь - выполнять обработку и визуализацию данных.

1. Какие возможности предоставляет система управления базами данных?

Создание и управление базами данных

Хранение и обработка данных

Работа с данными из разных приложений

\*Все перечисленные выше

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Что такое SQL?

Язык программирования

\*Язык запросов к базам данных

Язык маршрутизации пакетов в сети

Язык разметки текста

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Какие операции можно выполнять с таблицами в реляционной модели данных?

Выборка данных

Вставка данных

Удаление данных

\*Все перечисленные выше

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Какие связи между таблицами могут быть реализованы в реляционной модели данных?

Один-ко-многим (1:N)

Многие-ко-многим (N:M)

Один-к-одному (1:1)

\*Все перечисленные выше

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Что такое первичный ключ в реляционной модели данных?

\*Уникальный идентификатор записи в таблице

Связь между разными таблицами

Индекс для ускорения поиска данных

Поле, содержащее данные в текстовом виде

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Для чего необходимо ведение журналов транзакций транзакционными системами РСУБД?

\*Поддержки целостности данных

Контроля выполнения запросов

\*Журналирования запросов

Автоматического выполнения похожих запросов

\*Восстановления реляционной БД

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 66%.

1. На каких стадиях обработки данных применяюися методы трансофрмации данных?

Извлечение данных из внешних источников

\*Стадия очистки данных

\*Стадия подготовки данных перед загрузкой данных в хранилище

Загрузка данных в долговременное хранилище данных

\*Оптимизация данных для проведения моделирования

\*Организация витрин данных

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 75%.

1. Упорядочивание записей в таблице по строковой переменной осуществляется по правилу:

\*Посимвольное лексикографическое правило в соотвествии с положением символа в кодовой таблице символов

По правилу Левенштейна от пустой строки

По правилу Левенштейна от строки "9999"

Лексикографическое правило в соотвествии с положением слова в словаре

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Сортировка записей по нескольким полям табличных данных производится в соответсвии с правилом:

\*Лексикографическим правилом в соотвествии с выбранным пользователем порядком старшинства полей таблицы

Лексикографическим правилом в соотвествии с порядком старшинства полей, определенным по естественному правилу приведения типов

Один единственным выбранным полем из перечня полей

Сортировка табличных данных по нескольким полям не производится

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Какие возможны сравнения значений в полях таблицы при проведении процедуры фильтрации данных?

\*Очередное значение поля находится среди заранее определенного списка значений

\*Очередное значение поля находится в интервале значений

Очередное значение поля принадлежит заранее определенному типу данных

\*Очередное значение поля содержится в полуинтервале значений

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 66%.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_ данных – разбиение диапазона значений непрерывной или дискретной величины на конечное число интервалов. Шаг по оси измерения имеет фиксированный шаг.

Правильный ответ: квантование, Квантование

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_ данные – дополнительное определение для табличных данных с большим количеством категориальных атрибутов (столбцов).

Правильный ответ: Многомерные, многомерные

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_ данных — это процесс дополнения сырых данных той информацией, которая в исходном виде в них отсутствует, но необходима для качественного анализа.

Правильный ответ: обогащение, Обогащение

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуски в определении

Существует два вида обогащения данных, а именно \_\_\_\_\_\_\_\_\_ обогащение данных - получение отсутствующей информации путём обработки уже имеющихся данных, а также \_\_\_\_\_\_\_ обогащение данных - получение отсутствующей информации из иных информационных систем.

Правильный ответ: локальное, внешнее

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуски в определении

Стадия \_\_\_\_\_\_\_\_ данных решает задачу преобразования данных, полученных на стадии загрузки из внешних источников, в необходимую для задач анализа форму.

Правильный ответ: предобработки, Предобработки

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуски в определении

\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_ — этап удаления нерелевантных значений показателей или записей данных с нетипичными значениями.

Правильный ответ: Очистка данных, очистка данных, очистка, Очистка

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуски в определении

\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_ — это цифровая система хранения, которая выполняет объединение и согласование больших объемов данных из разных источников.

Правильный ответ: хранилище данных, Хранилище данных

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуски в определении

\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_ — это часть хранилища данных, секционированная для отделов или направлений бизнеса (например, продажи, маркетинг или финансы).

Правильный ответ: Витрина данных, витрина данных, витрина, Витрина

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Выберите верные стадии ETL процесса, расположите их в верном порядке следования. Запишите число без пробелов, состоящее из цифр последовательности.

1. Загрузка данных в хранилище

2. Организация витрин данных

3. Экспорт данных из внешних источников

4. Трансформация, предобработка данных

5. Визуализация данных

6. Определение метрик качества данных

Правильный ответ: 341

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Сопоставьте компоненты конвейера обработки данных с их функциями.

А. Сортировка

Б. Фильтрация

В. Слияние

Г. Группировка

1. Объединение таблиц на основе правил соединения по совпадающим значениям в выбранных полях с использованием алгебры множеств

2. Упорядочивание записей в таблице данных в соответствии с порядком значений выбранного поля или полей в лексикографическом порядке по возрастанию или убыванию значений

3. Обучение классификатора дискретных данных на основе известных данных предикторов

4. Выборка данных на основе логических выражений сравнений над значениями в полях таблицы данных

5. Вычисление агрегационных функций над подвыборками данных, сформированных на основе совпадения значений в выбранном поле групп

6. Объединение записей по совпадающей струкутре полей нескольких таблиц

В качестве ответа напишите четырехзначное число, содержащее последовательность цифр верных ответов в порядке их сопоставления с первыми буквами алфавита. Пример: если верное сопоставление обозначено как: A-5, Б-1, В-3, Г-6, то в ответе необходимо записать 5136.

Правильный ответ: 2415

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

ОПК-4: Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

Владеть - методами обработки данных.

1. Анализ основных свойств данных, нахождение в них общих закономерностей, распределений и аномалий, построение начальных моделей на основе визуализаций называют:

\*разведочный АД

гибридный АД

кросс АД

независимый АД

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Операция выборки строк (объектов, наблюдений) из таблицы данных в соответствии с логическим правилом сравнения значений выбранного атрибута с определенным значением

называется:

\*фильтрацией

сортировкой

кластеризацией

сегментацией

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Несколько условий с логическими связками «И», «ИЛИ», «НЕ» использующихся для отбора нескольких значений из разных столбцов называется:

\*сложный фильтр

\*составной фильтр

независимый фильтр

колоночный фильтр

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Получение отсутствующей информации путём обработки уже имеющихся данных называется \_\_\_ обогащение данных:

\*локальное

составное

независимое

внешнее

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Получение отсутствующей информации путём получения из иных информационных систем называется \_\_\_ обогащение данных:

локальное

составное

независимое

\*внешнее

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Процесс изучения данных с целью достижения понимания их структуры, содержимого и оценки качества называется \_\_\_ данных:

сегментация

кластеризация

обогащение

\*профайлинг

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Значение, отличающееся от остальных наблюдений настолько, что его формирование предполагается под влиянием иных механизмов и факторов, чем у большинства других данных в наборе (до 3 σ) называется:

экстремальное значение

пик

импульс

\*выброс

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Значение, отличающееся от остальных наблюдений от 3 σ до 5 σ называется:

\*экстремальное значение

пик

импульс

выброс

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Уникальный атрибут внешней таблицы, создающий связь с данной по совпадающим значениям в столбце называется \_\_ ключ:

\*внешний

\*вторичный

первичный

независимый

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Список таблиц, их атрибутов, типов данных, ограничений, ключей и связей между таблицами, необходимый для корректной организации хранения данных в памяти ВУ и доступа к данным извне, как на запись, так и на чтение называется \_\_ БД:

\*схема

состав

сегмент

кластер

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 1000 записей. Какое максимальное количество записей с пропусками может быть в ней, чтобы их можно было проигнорировать на этапе профайлинга в рамках стратегии борьбы с пропусками: \_\_.

Правильный ответ: 10

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 5000 записей. Какое минимальное количество записей с пропусками на этапе профайлинга должно быть в ней, чтобы необходимо было отказаться от последующего анализа этого набора данных в рамках стратегии борьбы с пропусками: \_\_.

Правильный ответ: 2500

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 5000 записей. Задается фильтрация: фильтру A не удовлетворяют 200 записей, фильтру B – 300, фильтру C – 400. Какое максимальное количество записей может отображаться в БД после применения всех фильтраций: \_\_.

Правильный ответ: 4600

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 500 записей. Задается фильтрация: фильтру A не удовлетворяют 20 записей, фильтру B – 30, фильтру C – 40. Какое минимальное количество записей может отображаться в БД после применения всех фильтраций: \_\_.

Правильный ответ: 410

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Структура БД содержит 3 поля X, Y, Z. БД содержит 10 записей. Известно, что поля в БД содержат 9 пропусков. Какое максимальное количество чистых данных (записей без пропусков) возможно в такой БД: \_\_\_.

Правильный ответ: 7

Время выполнения: 3 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Структура БД содержит 3 поля X, Y, Z. БД содержит 10 записей. Известно, что поля в БД содержат 9 пропусков. Какое минимальное количество чистых данных (записей без пропусков) возможно в такой БД: \_\_\_.

Правильный ответ: 1

Время выполнения: 3 минуты.

Критерий оценивания: верный ответ соответствует 100%

1. Некоторая БД содержит 2 поля X и Y. X может быть целым числом от 0 до 10, а Y – пол и может принимать значения «М» или «Ж». Какое максимальное количество чистых данных (без дубликатов) может содержать такая БД: \_\_\_\_\_\_\_\_.

Правильный ответ: 22

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Какой объем памяти отводится для хранения вещественного числа в формате с плавающей запятой, если на мантиссу отводится 10 бит, а на порядок числа 5 бит: \_\_\_\_\_\_\_\_.

Правильный ответ: 16

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Некоторая задача решается исполнителем за время Т. 90% задачи можно решать параллельно. Найти во сколько раз уменьшится время решения задачи, если число исполнителей увеличится на 8, а накладными расходами на коммуникацию можно пренебречь: \_\_\_\_\_\_\_\_.

Правильный ответ: 5

Время выполнения: 3 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Непрерывная величина зависит от времени и характеризует вес курицы до 5 кг. Указать число уровней квантования при переводе такой информации из непрерывного вида в дискретный при помощи линейного квантования с интервалом 1 кг:\_\_\_\_\_\_\_\_.

Правильный ответ: 6

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

Промежуточная аттестация

УК-1: Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач

Знать - методы сбора, способы хранения и обработки информации.

1. С помощью каких ключевых технологий и архитектурных решений достигается высокая скорость записи и чтения больших файлов в серверных решениях на базе кластерной архитектуры?

\*Распределенная система обработки данных

\*Архитектура главный/подчиненный

Последовательная обработка данных

\*Репликация блоков данных

Использование операционной системы на базе ядра Linux

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

1. Чем характеризуются архитектуры главный/подчиненный в исполнении серверных решений?

\*Большое число объединенных стоек с несколькими серверными узлами

\*Один главный и несколько резервных узлов для управления стойками

\*Наличие большого числа точек записи в серверных узлах на стойках

\*Централизованная балансировка нагрузки по чтению, записи и обработки данных

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

Какие факторы ограничивают кратное повышение производительности в решениях на базе распределенных архитектур?

\*соединение исполнителей в сети

\*программная балансировка нагрузки на исполнителях

качество написания последовательной части программного кода

\*синхронизация исполнителей

\*помехи при передаче данных

формат передаваемой информации по каналу связи

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 75% правильных ответов.

Выберите верные свойства, присущие распределенной системе обработки и хранения данных:

\*совместное использование ресурсов

низкая скорость чтения

\*совместная обработка данных

проприеритарность

\*масштабируемость по программным и аппаратным ресурсам

\*обнаружение отказов

Закрытость

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 75% правильных ответов.

1. В чем состоит существенная особенность временных рядов как данных табличного типа?

\*Содержат зависимые упорядоченные во времени последовательности значений

Временные промежутки между измерениями непостоянны

\*Содержат время в качестве отдельного показателя

\*Показывают развитие показателя одного конкретного объекта во времени

Данные показателей (столбцы) временных рядов (если их несколько) не зависят друг от друга

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

1. В формате данных JSON данные хранятся в виде:

Строк с разделителями столбцов, отличными от разделителя дробной части вещественных чисел

\*Коллекции данных формата "ключ-значение", где доступ к значениям полей данных осуществляется по ключу

В виде таблицы, с разлинованными строками и столбцами в виде отдельных объектов документа

В виде неструктурированной последовательности битов, расшифровывающейся в определенном формате как целостный объект

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

Отметьте верные примеры неструктурированных данных:

\*Файлы изображений

Таблицы данных транзакций

\*Файлы, хранящие текстовую информацию на естественном языке

Таблицы реляционной базы данных

\*Видеофайлы

Файлы формата JSON с перечнем объектов с одинаковым набором ключей

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

1. Выберите верные утверждения, относящиеся к понятию неструктурированных данных:

Организованная, типизированная информация, относящаяся к одной сущности

\*Не имеет предопределенной организации и имеет множество форм

Франят в долговременном хранилище данных, реляционных базах данных

\*Хранят в озёрах данных, файловых базы данных

Имеет несколько предопределенных форматов

\*Имеет большое количество различных форматов данных

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

1. Какие существуют основные свойства шкал измерений?

\*Идентифицируемость

\*Величина

Дискретизируемость

\*Равенство интервалов

Равенство значений

\*Абсолютный ноль

Квантуемость

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 75% правильных ответов.

Среди шкал выберите те, с которыми невозможны арифметические операции:

Номинальная шкала (категориальная, наименований)

Порядковая шкала (ординальная, ранговая)

\*Интервальная шкала (разностей)

\*Шкала отношений (абсолютная)

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 50% правильных ответов.

1. Среди шкал выберите те, на которых невозможно ввести отношение порядка:

\*Номинальная шкала (категориальная, наименований)

Порядковая шкала (ординальная, ранговая)

Интервальная шкала (разностей)

Шкала отношений (абсолютная)

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Среди шкал выберите те, в которых мерой среднего может служить среднее арифметическое:

Номинальная шкала (категориальная, наименований)

Порядковая шкала (ординальная, ранговая)

\*Интервальная шкала (разностей)

\*Шкала отношений (абсолютная)

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 50% правильных ответов.

1. Среди шкал выберите те, в которых мерой среднего может служить медиана:

Номинальная шкала (категориальная, наименований)

\*Порядковая шкала (ординальная, ранговая)

\*Интервальная шкала (разностей)

\*Шкала отношений (абсолютная)

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

Что такое реляционная модель данных?

Модель данных, основанная на файловой структуре

\*Модель данных, основанная на связях между таблицами

Модель данных, основанная на иерархической структуре

Модель данных, основанная на объектах

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Совокупность программных и языковых средств для создания, удаления, изменения и любых других манипуляций с данными в БД:

\*система управления базами данных

реляционная база данных

утилита для формирования запросов к базе данных

визуализатор связей и отношений в базе данных

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Что такое таблица в реляционной модели данных?

Множество свойств, описывающих объекты

Список связей между объектами

\*Структура, хранящая данные в виде строк и столбцов

Файл, содержащий информацию о данных

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какие особенности реляционных баз данных?

Использование структуры типа таблицы с записями

Возможность связывания таблиц

Удобство в использовании и администрировании

\*Все перечисленные выше

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какие принципы лежат в основе реляционной модели данных?

\*Принципы ООП

Принципы структурного программирования

Принципы представления данных и их связей

Принципы функционального программирования

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какой вид операции JOIN в SQL соответствует декартовому произведению множеств?  
   \*FULL  
   RIGHT  
   LEFT  
   OUTER  
   INNER

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Единица измерения информации это?  
   Ватт  
   \*бит  
   ампер  
   вольт  
   кулон

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Графовые базы данных основаны на:  
   \*Графах  
   Герцогах  
   Матрицах  
   СЛАУ

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Что из перечисленного является ключевым словом в языке SQL?  
   \*SELECT  
   COS  
   \*WHERE  
   \*JOIN  
   WELCOME  
   EXIT

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Что из перечисленного является СУБД?  
   \*Cassandra  
   \*MongoDB  
   \*Postgres  
   Mongol  
   Hawaii  
   CyberDB

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Наличие первичного ключа в таблице:  
   Обязательно  
   \*Не обязательно  
   Зависит от СУБД  
   Зависит от типа СУБД

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Наличие вторичного ключа в таблице:  
   Обязательно  
   \*Не обязательно  
   Зависит от СУБД  
   Зависит от типа СУБД

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

УК-1: Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач

Уметь - применять методы обработки данных

1. Заполните пропуск в определении

\_\_\_\_\_\_\_\_\_\_\_\_ система — это совокупность компьютеров, которые обмениваются данными и синхронизируются в общей сети, образуя «единый компьютер» для конечного пользователя.

Правильный ответ: Распределенная, распределенная

Время, отводимое на выполнение задания (3 минуты).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_\_\_\_\_\_ масштабирование означает добавление дополнительных узлов в систему (или удаление узлов из) системы, например добавление нового компьютера в распределенное программное приложение.

Правильный ответ: Горизонтальное, горизонтальное

Время, отводимое на выполнение задания (3 минуты).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_\_\_\_\_\_ масштабирование означает добавление ресурсов (или удаление ресурсов) к одному узлу, как правило, с добавлением ЦП, памяти или хранилища к одному компьютеру.

Правильный ответ: Вертикальное, вертикальное

Время, отводимое на выполнение задания (3 минуты).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_\_\_\_\_\_ развитие подразумевает собой уменьшение расходов для платформ и/или продуктов по средствам, времени и качеству решения задачи.

Правильный ответ: Экстенсивное, экстенсивное

Время, отводимое на выполнение задания (3 минуты).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_\_\_\_\_\_ развитие подразумевает собой балансировку нагрузки системы, увеличение количества ресурсов для обработки данных, репликация баз данных, мониторинг и автоматическое отслеживание производительности компонентов системы.

Правильный ответ: Эффективное, эффективное

Время, отводимое на выполнение задания (3 минуты).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

В объектно-ориентированном подходе (ООП) все сущности формализуются набором \_\_\_\_\_\_\_ и методов.

Правильный ответ: полей, параметров

Время, отводимое на выполнение задания (3 минуты).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_\_ таблицы в обработке и хранении данных это перечень полей и их типов данных, которыми представлена таблица данных.

Правильный ответ: структура, Структура

Время, отводимое на выполнение задания (3 минуты).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_\_\_\_\_ данные  — данные, которые не соответствуют заранее определённой модели данных, и, как правило, представлены в форме текста с датами, цифрами, фактами, расположенными в нём в произвольной форме.

Правильный ответ: Неструктурированные, неструктурированные

Время, отводимое на выполнение задания (3 минуты).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_\_ данные являются значениями признака, общее число которых конечной или бесконечно, но может быть подсчитано при помощи натуральных чисел.

Правильный ответ: Дискретные, дискретные

Время, отводимое на выполнение задания (3 минуты).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуски в определении

\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_ – это совокупность данных, хранящихся и упорядоченных в соответствии с определённой структурой.

Правильный ответ: База данных, база данных, БД

Время, отводимое на выполнение задания (3 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какой протокол используется для передачи данных между распределенными приложениями на Linux-серверах?

HTTP

SMTP

FTP

\*SSH

Время, отводимое на выполнение задания (2 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какая директория на Linux-серверах используется для хранения системных журналов?

\*/var/logs

/logs

/usr/share/logs

/etc/logs

Время, отводимое на выполнение задания (1 минута).

Критерий оценивания: 100% правильных ответов.

1. Какой тип данных используется для хранения URL-адресов?

\*VARCHAR

BOOLEAN

INT

\*TEXT

Время, отводимое на выполнение задания (1 минута).

Критерий оценивания: 50% правильных ответов.

1. Какой тип данных используется для хранения изображений?

\*BLOB

BOOLEAN

INT

VARCHAR

Время, отводимое на выполнение задания (1 минута).

Критерий оценивания: 100% правильных ответов.

1. Какой тип данных используется для хранения аудиофайлов?

AUDIO

INT

BOOLEAN

\*BLOB

Время, отводимое на выполнение задания (1 минута).

Критерий оценивания: 100% правильных ответов.

1. Какой тип данных используется для хранения видеофайлов?

VIDEO

BOOLEAN

INT

\*BLOB

Время, отводимое на выполнение задания (1 минута).

Критерий оценивания: 100% правильных ответов.

1. Отметьте верные задачи ОБРАБОТКИ/ПРЕДОБРАБОТКИ неструктурированных данных, рассматривающиеся в обработке текстовой информации:

Сегментация текста

\*Структуризация текстовых данных

Поиск смысла в тексте

\*Векторизация текста

Анализ тональности текста

\*Генерация признаков на основе текста

Время, отводимое на выполнение задания (1 минута).

Критерий оценивания: 66% правильных ответов.

1. Отметьте верные задачи АНАЛИЗА неструктурированных данных, рассматривающиеся в анализе текстовой информации:

\*Классификация отзывов

Стемминг, лемматизация

\*Кластеризация документов

\*Автодополнение поисковых запросов

Очистка текста от ошибок

\*Поиск именованных сущностей, ключевых слов

\*Сегментация текста

Время, отводимое на выполнение задания (1 минута).

Критерий оценивания: 75% правильных ответов.

1. Отметьте верные задачи ОБРАБОТКИ и АНАЛИЗА больших данных аудиоформата:

\*очистка от шумов

дискретизация аналогового сигнала

\*выявление споттера (слово-триггер обычно для голосовых ассистентов)

\*разделение дорожек на речь и фоновый шум

\*выявление акцента / языка

Время, отводимое на выполнение задания (2 минуты).

Критерий оценивания: 75% правильных ответов.

1. Отметьте верные задачи ОБРАБОТКИ и АНАЛИЗА больших данных фото и видеоформата:

\*Сегментация

\*Оптическое распознавание символоов

\*Обнаружение краев

\*Сопоставление шаблонов

Время, отводимое на выполнение задания (3 минуты).

Критерий оценивания: 75% правильных ответов.

УК-1: Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач

Владеть - методами интерпретации данных

1. Как влияет на качество перевода непрерывной (аналоговой) информации в дискретную частота дискретизации - ν.

\* Увеличение ν – улучшает качество перевода

Увеличение ν – ухудшает качество перевода

Уменьшение ν – улучшает качество перевода

Изменение ν – не влияет на качество перевода

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Как изменится период (интервал) между точками оценки точных значений аналогового сигнала при переводе непрерывной информации в дискретную при увеличении частоты дискретизации ν в 2 раза.

Т увеличится в 2 раза

\* Т уменьшится в 2 раза

Т увеличится в 4 раза

Т уменьшится в 8 раз

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 2000 записей. Какое максимальное количество записей с пропусками может быть в ней, чтобы их можно было проигнорировать на этапе профайлинга в рамках стратегии борьбы с пропусками:

10

\*20

30

50

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 10000 записей. Какое минимальное количество записей с пропусками на этапе профайлинга должно быть в ней, чтобы необходимо было отказаться от последующего анализа этого набора данных в рамках стратегии борьбы с пропусками:

1000

2000

3000

\*5000

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 5000 записей. Задается фильтрация: фильтру A не удовлетворяют 100 записей, фильтру B – 200, фильтру C – 300. Какое максимальное количество записей может отображаться в БД после применения всех фильтраций: \_\_.

4100

4600

\*4700

4900

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 500 записей. Задается фильтрация: фильтру A не удовлетворяют 10 записей, фильтру B – 20, фильтру C – 30. Какое минимальное количество записей может отображаться в БД после применения всех фильтраций: \_\_.

410

\*440

470

490

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Единица обработки информации при конвейерной обработке информации единообразной структуры - это:

сегмент

\*контейнер

кластер

байт

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Отличие каппа от лямбда архитектуры обработки данных следующее:

Данные хранятся в разных хранилищах

\* слияние направлений обработки данных: потоковой и пакетной.

\* данные хранятся в одном хранилище данных.

Запросы проводятся к разным ХД

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Витрины данных, объединяющие источники первичных данных из существующего хранилища данных и других внешних источников данных:

первичные

\* гибридные

зависимые

независимые

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В случае, если данные извлекаются из внутренних или внешних источников и загружаются в независимый репозиторий, то витрины данных, не связанные с хранилищем данных называются:

первичные

гибридные

зависимые

\*независимые

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Сколько бит информации содержит слово из 5 символов кодировки ASCII \_\_\_\_\_\_\_\_.

Правильный ответ: 40

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Сколько бит информации содержит слово из 3 символов кодировки ASCII7: \_\_\_\_\_\_\_.

Правильный ответ: 21

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_ данных – процесс измерения показателя в дискретные моменты времени с заранее заданной частотой дискретизации, так чтобы перевести непрерывную или аналоговую информацию в цифровой вид.

Правильный ответ: дискретизация, Дискретизация

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_ данных – разбиение диапазона значений непрерывной или дискретной величины на конечное число интервалов. Шаг по оси измерения имеет фиксированный шаг.

Правильный ответ: квантование, Квантование

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_ данные – дополнительное определение для табличных данных с большим количеством категориальных атрибутов (столбцов).

Правильный ответ: Многомерные, многомерные

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_ данных — это процесс дополнения сырых данных той информацией, которая в исходном виде в них отсутствует, но необходима для качественного анализа.

Правильный ответ: обогащение, Обогащение

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуски в определении

Существует два вида обогащения данных, а именно \_\_\_\_\_\_\_\_\_ обогащение данных - получение отсутствующей информации путём обработки уже имеющихся данных, а также \_\_\_\_\_\_\_ обогащение данных - получение отсутствующей информации из иных информационных систем.

Правильный ответ: локальное, внешнее

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуски в определении

Стадия \_\_\_\_\_\_\_\_ данных решает задачу преобразования данных, полученных на стадии загрузки из внешних источников, в необходимую для задач анализа форму.

Правильный ответ: предобработки, Предобработки

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуски в определении

\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_ — это цифровая система хранения, которая выполняет объединение и согласование больших объемов данных из разных источников.

Правильный ответ: хранилище данных, Хранилище данных

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуски в определении

\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_ — это часть хранилища данных, секционированная для отделов или направлений бизнеса (например, продажи, маркетинг или финансы).

Правильный ответ: Витрина данных, витрина данных, витрина, Витрина

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

ОПК-4: Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

Знать - технологии сбора, хранения и обработки больших объемов данных.

1. С помощью каких ключевых технологий и архитектурных решений достигается высокая скорость записи и чтения больших файлов в серверных решениях на базе кластерной архитектуры?

\*Распределенная система обработки данных

\*Архитектура главный/подчиненный

Последовательная обработка данных

\*Репликация блоков данных

Использование операционной системы на базе ядра Linux

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

1. Чем характеризуются архитектуры главный/подчиненный в исполнении серверных решений?

\*Большое число объединенных стоек с несколькими серверными узлами

\*Один главный и несколько резервных узлов для управления стойками

\*Наличие большого числа точек записи в серверных узлах на стойках

\*Централизованная балансировка нагрузки по чтению, записи и обработки данных

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

Какие факторы ограничивают кратное повышение производительности в решениях на базе распределенных архитектур?

\*соединение исполнителей в сети

\*программная балансировка нагрузки на исполнителях

качество написания последовательной части программного кода

\*синхронизация исполнителей

\*помехи при передаче данных

формат передаваемой информации по каналу связи

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 75% правильных ответов.

Выберите верные свойства, присущие распределенной системе обработки и хранения данных:

\*совместное использование ресурсов

низкая скорость чтения

\*совместная обработка данных

проприеритарность

\*масштабируемость по программным и аппаратным ресурсам

\*обнаружение отказов

Закрытость

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 75% правильных ответов.

1. В чем состоит существенная особенность временных рядов как данных табличного типа?

\*Содержат зависимые упорядоченные во времени последовательности значений

Временные промежутки между измерениями непостоянны

\*Содержат время в качестве отдельного показателя

\*Показывают развитие показателя одного конкретного объекта во времени

Данные показателей (столбцы) временных рядов (если их несколько) не зависят друг от друга

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

1. В формате данных JSON данные хранятся в виде:

Строк с разделителями столбцов, отличными от разделителя дробной части вещественных чисел

\*Коллекции данных формата "ключ-значение", где доступ к значениям полей данных осуществляется по ключу

В виде таблицы, с разлинованными строками и столбцами в виде отдельных объектов документа

В виде неструктурированной последовательности битов, расшифровывающейся в определенном формате как целостный объект

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

Отметьте верные примеры неструктурированных данных:

\*Файлы изображений

Таблицы данных транзакций

\*Файлы, хранящие текстовую информацию на естественном языке

Таблицы реляционной базы данных

\*Видеофайлы

Файлы формата JSON с перечнем объектов с одинаковым набором ключей

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

1. Выберите верные утверждения, относящиеся к понятию неструктурированных данных:

Организованная, типизированная информация, относящаяся к одной сущности

\*Не имеет предопределенной организации и имеет множество форм

Франят в долговременном хранилище данных, реляционных базах данных

\*Хранят в озёрах данных, файловых базы данных

Имеет несколько предопределенных форматов

\*Имеет большое количество различных форматов данных

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

1. Какие существуют основные свойства шкал измерений?

\*Идентифицируемость

\*Величина

Дискретизируемость

\*Равенство интервалов

Равенство значений

\*Абсолютный ноль

Квантуемость

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 75% правильных ответов.

Среди шкал выберите те, с которыми невозможны арифметические операции:

Номинальная шкала (категориальная, наименований)

Порядковая шкала (ординальная, ранговая)

\*Интервальная шкала (разностей)

\*Шкала отношений (абсолютная)

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 50% правильных ответов.

1. Среди шкал выберите те, на которых невозможно ввести отношение порядка:

\*Номинальная шкала (категориальная, наименований)

Порядковая шкала (ординальная, ранговая)

Интервальная шкала (разностей)

Шкала отношений (абсолютная)

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Среди шкал выберите те, в которых мерой среднего может служить среднее арифметическое:

Номинальная шкала (категориальная, наименований)

Порядковая шкала (ординальная, ранговая)

\*Интервальная шкала (разностей)

\*Шкала отношений (абсолютная)

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 50% правильных ответов.

1. Среди шкал выберите те, в которых мерой среднего может служить медиана:

Номинальная шкала (категориальная, наименований)

\*Порядковая шкала (ординальная, ранговая)

\*Интервальная шкала (разностей)

\*Шкала отношений (абсолютная)

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 66% правильных ответов.

Что такое реляционная модель данных?

Модель данных, основанная на файловой структуре

\*Модель данных, основанная на связях между таблицами

Модель данных, основанная на иерархической структуре

Модель данных, основанная на объектах

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Совокупность программных и языковых средств для создания, удаления, изменения и любых других манипуляций с данными в БД:

\*система управления базами данных

реляционная база данных

утилита для формирования запросов к базе данных

визуализатор связей и отношений в базе данных

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Что такое таблица в реляционной модели данных?

Множество свойств, описывающих объекты

Список связей между объектами

\*Структура, хранящая данные в виде строк и столбцов

Файл, содержащий информацию о данных

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какие особенности реляционных баз данных?

Использование структуры типа таблицы с записями

Возможность связывания таблиц

Удобство в использовании и администрировании

\*Все перечисленные выше

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какие принципы лежат в основе реляционной модели данных?

\*Принципы ООП

Принципы структурного программирования

Принципы представления данных и их связей

Принципы функционального программирования

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Какой вид операции JOIN в SQL соответствует декартовому произведению множеств?  
   \*FULL  
   RIGHT  
   LEFT  
   OUTER  
   INNER

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Единица измерения информации это?  
   Ватт  
   \*бит  
   ампер  
   вольт  
   кулон

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Графовые базы данных основаны на:  
   \*Графах  
   Герцогах  
   Матрицах  
   СЛАУ

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Что из перечисленного является ключевым словом в языке SQL?  
   \*SELECT  
   COS  
   \*WHERE  
   \*JOIN  
   WELCOME  
   EXIT

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Что из перечисленного является СУБД?  
   \*Cassandra  
   \*MongoDB  
   \*Postgres  
   Mongol  
   Hawaii  
   CyberDB

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Наличие первичного ключа в таблице:  
   Обязательно  
   \*Не обязательно  
   Зависит от СУБД  
   Зависит от типа СУБД

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

1. Наличие вторичного ключа в таблице:  
   Обязательно  
   \*Не обязательно  
   Зависит от СУБД  
   Зависит от типа СУБД

Время, отводимое на выполнение задания (1 минуты).

Критерий оценивания: 100% правильных ответов.

ОПК-4: Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

Уметь - выполнять обработку и визуализацию данных.

1. Заполните пропуск в определении

Для манипуляций над данными в реляционных СУБД используют декларативный язык запросов \_\_\_\_\_

Правильный ответ: SQL, sql, Sql, structured query language

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуск в определении

\_\_\_\_\_ базы данных – список таблиц, их атрибутов, типов данных, ограничений, ключей и связей между таблицами, необходимый для корректной организации хранения данных в памяти вычислительного устройства и доступа к данным извне, как на запись, так и на чтение.

Правильный ответ: схема

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуск в определении

Доступ к данным в реляционных СУБД классически осуществляется с помощью \_\_\_, подъязыка SQL

Правильный ответ: DML, dml, data manipulation language

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуск в определении

\_\_\_\_\_\_ — группа операторов определения, манипуляции данных, переводящих базу данных из одного согласованного состояния в другое согласованное состояние.

Правильный ответ: Транзакция, транзакция

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуск в определении

\_\_\_\_\_\_ называют данные, которые содержат ошибки, пропуски, дубликаты, несогласованность в формате или некорректную информацию.

Правильный ответ: Грязными, грязными, грязные

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуски в определении

\_\_\_\_\_\_ \_\_\_\_\_\_\_\_ - это выполнение различных преобразований данных с целью их подготовки к анализу или моделированию.

Правильный ответ: Трансформация данных, трансформация данных

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуск в определении

\_\_\_\_\_\_ табличных данных – преобразование, упорядочивающее набор объектов (строк) или наблюдений в связи с правилом упорядочивания по выбранным атрибутам.

Правильный ответ: Сортировка, сортировка

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуск в определении

\_\_\_\_\_\_ данных — операция выборки строк (объектов) или наблюдений из таблицы данных в соответствии с логическим правилом сравнения значений выбранного атрибута с определенным значением.

Правильный ответ: Фильтрация, фильтрация

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_ данных – процесс получения обобщенных статистик для некоторой большой выборки табличных данных с целью получить важную информацию по уникальным группам категорий объектов.

Правильный ответ: группировка, Группировка, Агрегация, агрегация

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_ данных – процесс измерения показателя в дискретные моменты времени с заранее заданной частотой дискретизации, так чтобы перевести непрерывную или аналоговую информацию в цифровой вид.

Правильный ответ: дискретизация, Дискретизация

Время выполнения: 5 минут.

Критерий оценивания: верный ответ соответствует 100%.

1. Корректным ли является сравнение значений вычисленного вещественного поля на точное равенство с заранее заданным пользователем вещественным значением для проведения фильтрации?

\*Нет, в общем смысле сравнение не является корректным, так как из-за ошибок округления при вычислении значений столбца результат может быть неконтролируемым

Корректным, в случае, если вычисление запланировано должно арифметически равняться заданному значению

Нет, в общем смысле сравнение не является корректным, так как такое сравнение не определено

Корректно всегда

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. В каких задачах применяются регулярные выражения?

Детекция объектов на изображении

\*Замена строк в текстовых данных

\*Парсинг текстовых данных из сырой текстовой информации веб-страниц

Классификация текстовой информации

\*Валидация текстовых данных на корректность ввода

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 66%.

1. Какие задачи решает агрегация (группировка) данных?

Упорядочивание записей (показателей) по значениям в группах дискретных признаков

Выборка данных (показателей) по совпадению значений в группах дискретных признаков с заранее заданным значением

Изменение структуры таблицы и выбор нового перечня столбцов на основе уникальных значений в группах дискретных признаков

\*Подсчет агрегационных функций (статистик) по подвыборкам показателей по группам совпадающих дискретных признаков

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Для каких типов данных имеет смысл выбор показателя в качестве группы при проведении агрегации данных (группировка)?

\*Строковый тип данных категориального признака

Вещественный тип данных непрерывного признака

\*Целочисленный тип данных дискретного признака

\*Дата/время начала месяца

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 66%.

1. Какие функции из перечисленных являются агрегационными (использующимися при агрегации, группировке)?

\*Среднее арифметическое

Модуль числа

\*Медиана

\*Сумма

\*Максимум

Корень числа

\*Последнее значение

Последний символ в строке

Экспонента

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 80%.

1. Выберите верные утверждения о кросс-таблице:

\*Кросс-таблица позволяет вычислять агрегационные функции по числовым значениям в записях на пересечении двух совпадающих значений категориального признака

\*Вычисления агрегации в кросс-таблице всегда происходят по меньшим по размеру числовым векторам по сравнению с базовой одномерной агрегацией

Кросс-таблица выполняет операцию слияния нескольких таблиц и последующую агрегацию значений в ней по всем категориальным признакам

\*Кросс-таблица это частный случай OLAP-куба для двух измерений категориальных признаков

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 66%.

1. Какой вид соединения (слияния) таблиц удаляет все записи с ключами не принаджежащими сразу двум выбранным таблицам в результирующей таблице?

Левое соединение

Правое соединение

Полное соединение

\*Внутреннее соединение

Перекрестное соединение

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Какой(-ие) вид(-ы) соединения(-ий) (слияния) таблиц дополняет(-ют) поля новыми полями присоединяемой таблицы, оставляя исходное число записей основной таблицы в результирующей?

\*Левое соединение

\*Правое соединение

Полное соединение

Внутреннее соединение

Перекрёстное соединение

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 50%.

1. Из каких стадий состоит классический конвейер обработки больших данных?

\*Сбор и извлечение данных из внешних источников.

\*Предобработка данных

Настройка параметров дерева решений

\*Структуризация данных.

Консолидация данных

\*Загрузка данных в долговременное хранилище данных.

\*Организация витрин данных.

Обучение глубокой нейронной модели

\*Построение аналитических отчётностей

Деструктуризация данных

\*Построение моделей над данными

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 80%.

1. Отметьте верные стадии очистки данных:

\*Обработка пропусков

\*Удаление дубликатов и противоречий

Автоматическая настройка струкутры табличных данных

\*Обработка выбросов (нетипичных значений)

Консолидация данных

\*Восстановление структуры данных

Проведение статистических тестов

\*Верификация целостности данных

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 80%.

ОПК-4: Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

Владеть - методами обработки данных.

1. Как влияет на качество перевода непрерывной (аналоговой) информации в дискретную частота дискретизации - ν.

\* Увеличение ν – улучшает качество перевода

Увеличение ν – ухудшает качество перевода

Уменьшение ν – улучшает качество перевода

Изменение ν – не влияет на качество перевода

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Как изменится период (интервал) между точками оценки точных значений аналогового сигнала при переводе непрерывной информации в дискретную при увеличении частоты дискретизации ν в 2 раза.

Т увеличится в 2 раза

\* Т уменьшится в 2 раза

Т увеличится в 4 раза

Т уменьшится в 8 раз

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 2000 записей. Какое максимальное количество записей с пропусками может быть в ней, чтобы их можно было проигнорировать на этапе профайлинга в рамках стратегии борьбы с пропусками:

10

\*20

30

50

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 10000 записей. Какое минимальное количество записей с пропусками на этапе профайлинга должно быть в ней, чтобы необходимо было отказаться от последующего анализа этого набора данных в рамках стратегии борьбы с пропусками:

1000

2000

3000

\*5000

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 5000 записей. Задается фильтрация: фильтру A не удовлетворяют 100 записей, фильтру B – 200, фильтру C – 300. Какое максимальное количество записей может отображаться в БД после применения всех фильтраций: \_\_.

4100

4600

\*4700

4900

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В некоторой БД 500 записей. Задается фильтрация: фильтру A не удовлетворяют 10 записей, фильтру B – 20, фильтру C – 30. Какое минимальное количество записей может отображаться в БД после применения всех фильтраций: \_\_.

410

\*440

470

490

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Единица обработки информации при конвейерной обработке информации единообразной структуры - это:

сегмент

\*контейнер

кластер

байт

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Отличие каппа от лямбда архитектуры обработки данных следующее:

Данные хранятся в разных хранилищах

\* слияние направлений обработки данных: потоковой и пакетной.

\* данные хранятся в одном хранилище данных.

Запросы проводятся к разным ХД

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Витрины данных, объединяющие источники первичных данных из существующего хранилища данных и других внешних источников данных:

первичные

\* гибридные

зависимые

независимые

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. В случае, если данные извлекаются из внутренних или внешних источников и загружаются в независимый репозиторий, то витрины данных, не связанные с хранилищем данных называются:

первичные

гибридные

зависимые

\*независимые

Время выполнения: 2 минуты.

Критерий оценивания: верный ответ соответствует 100%.

1. Сколько бит информации содержит слово из 5 символов кодировки ASCII \_\_\_\_\_\_\_\_.

Правильный ответ: 40

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Сколько бит информации содержит слово из 3 символов кодировки ASCII7: \_\_\_\_\_\_\_.

Правильный ответ: 21

Время выполнения: 1 минута.

Критерий оценивания: верный ответ соответствует 100%.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_ данных – процесс измерения показателя в дискретные моменты времени с заранее заданной частотой дискретизации, так чтобы перевести непрерывную или аналоговую информацию в цифровой вид.

Правильный ответ: дискретизация, Дискретизация

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_ данных – разбиение диапазона значений непрерывной или дискретной величины на конечное число интервалов. Шаг по оси измерения имеет фиксированный шаг.

Правильный ответ: квантование, Квантование

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_ данные – дополнительное определение для табличных данных с большим количеством категориальных атрибутов (столбцов).

Правильный ответ: Многомерные, многомерные

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуск в определении

\_\_\_\_\_\_\_ данных — это процесс дополнения сырых данных той информацией, которая в исходном виде в них отсутствует, но необходима для качественного анализа.

Правильный ответ: обогащение, Обогащение

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуски в определении

Существует два вида обогащения данных, а именно \_\_\_\_\_\_\_\_\_ обогащение данных - получение отсутствующей информации путём обработки уже имеющихся данных, а также \_\_\_\_\_\_\_ обогащение данных - получение отсутствующей информации из иных информационных систем.

Правильный ответ: локальное, внешнее

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуски в определении

Стадия \_\_\_\_\_\_\_\_ данных решает задачу преобразования данных, полученных на стадии загрузки из внешних источников, в необходимую для задач анализа форму.

Правильный ответ: предобработки, Предобработки

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуски в определении

\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_ — это цифровая система хранения, которая выполняет объединение и согласование больших объемов данных из разных источников.

Правильный ответ: хранилище данных, Хранилище данных

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.

1. Заполните пропуски в определении

\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_ — это часть хранилища данных, секционированная для отделов или направлений бизнеса (например, продажи, маркетинг или финансы).

Правильный ответ: Витрина данных, витрина данных, витрина, Витрина

Время, отводимое на выполнение задания (5 минут).

Критерий оценивания: 100% правильных ответов.