



Федеральное агентство по рыболовству
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Калининградский государственный технический университет»
(ФГБОУ ВО «КГТУ»)

УТВЕРЖДАЮ
Начальник УРОПС

Фонд оценочных средств
(приложение к рабочей программе модуля)
«ETL-СИСТЕМЫ И БАЗЫ ДАННЫХ»

основной профессиональной образовательной программы магистратуры
по направлению подготовки

09.04.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА

ИНСТИТУТ
РАЗРАБОТЧИК

цифровых технологий
кафедра прикладной математики и информационных технологий

1 РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Таблица 1 – Планируемые результаты обучения по дисциплине, соотнесенные с установленными индикаторами достижения компетенций

Код и наименование компетенции	Индикаторы достижения компетенции	Дисциплина	Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции
ПК-4: Способен разрабатывать и оптимизировать средства сбора и хранения данных для решения профессиональных задач	ПК-4.1: Знает методы и средства сбора и хранения данных для решения профессиональных задач с использованием ETL-систем	ETL-системы и базы данных	<p><u>Знать:</u></p> <ul style="list-style-type: none"> - архитектуры и модели данных; - архитектуры и модели баз и хранилищ данных; - технологии и программное обеспечение систем хранения и ETL-обработки информации; - технологии, методы и инструментальные средства ETL-обработки обработки данных; <p><u>Уметь:</u></p> <ul style="list-style-type: none"> - пользоваться методами и инструментами получения, хранения, передачи, ETL-обработки данных; <p><u>Владеть:</u></p> <ul style="list-style-type: none"> - навыками разработки предложений по развитию и совершенствованию системы получения, хранения, передачи и ETL-обработки данных.

2 ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ СРЕДСТВ И КРИТЕРИИ ОЦЕНИВАНИЯ

2.1 К оценочным средствам текущего контроля успеваемости относятся:

- тестовые задания открытого и закрытого типов.

2.2 К оценочным средствам для промежуточной аттестации относятся:

- экзаменационные задания по дисциплине, представленные в виде тестовых заданий закрытого и открытого типов.

2.3 Критерии оценки результатов освоения дисциплины

Универсальная система оценивания результатов обучения включает в себя системы оценок: 1) «отлично», «хорошо», «удовлетворительно», «неудовлетворительно»; 2) «зачтено», «не зачтено»; 3) 100 – балльную/процентную систему и правило перевода оценок в пятибалльную систему (табл. 2).

Таблица 2 – Система оценок и критерии выставления оценки

Система оценок Критерий	2	3	4	5
	0-40%	41-60%	61-80 %	81-100 %
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
	«не зачтено»	«зачтено»		
1 Системность и полнота знаний в отношении изучаемых объектов	Обладает частичными и разрозненными знаниями, которые не может научно- корректно связывать между собой (только некоторые из которых может связывать между собой)	Обладает минимальным набором знаний, необходимым для системного взгляда на изучаемый объект	Обладает набором знаний, достаточным для системного взгляда на изучаемый объект	Обладает полной знаний и системным взглядом на изучаемый объект
2 Работа с информацией	Не в состоянии находить необходимую информацию, либо в состоянии находить отдельные фрагменты информации в рамках поставленной задачи	Может найти необходимую информацию в рамках поставленной задачи	Может найти, интерпретировать и систематизировать необходимую информацию в рамках поставленной задачи	Может найти, систематизировать необходимую информацию, а также выявить новые, дополнительные источники информации в рамках поставленной задачи
3.Научное осмысление изучаемого явления, процесса, объекта	Не может делать научно корректных выводов из имеющихся у него сведений, в состоянии проанализировать только некоторые из имеющихся у него сведений	В состоянии осуществлять научно корректный анализ предоставленной информации	В состоянии осуществлять систематический и научно корректный анализ предоставленной информации, вовлекает в исследование новые релевантные задачи данные	В состоянии осуществлять систематический и научно-корректный анализ предоставленной информации, вовлекает в исследование новые релевантные поставленной задаче данные, предлагает новые ракурсы поставленной задачи
4. Освоение стандартных алгоритмов решения профессиональных задач	В состоянии решать только фрагменты поставленной задачи в соответствии с заданным алгоритмом, не освоил	В состоянии решать поставленные задачи в соответствии с заданным алгоритмом	В состоянии решать поставленные задачи в соответствии с заданным алгоритмом, понимает	Не только владеет алгоритмом и понимает его основы, но и предлагает новые решения в рамках поставленной задачи

Система оценок Критерий	2	3	4	5
	0-40%	41-60%	61-80 %	81-100 %
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
	«не зачтено»	«зачтено»		
	предложенный алгоритм, допускает ошибки		основы предложенного алгоритма	

2.4 Критерии оценки курсовой работы

Оценка курсовой работы осуществляется в два этапа. Первый этап – после проверки работы, второй этап – после ее защиты. Для защиты курсовой работы студент должен подготовить пояснительную записку, а также представить устное сообщение с возможным представлением электронной презентации (по желанию) в течение 5–7 мин и ответить на вопросы. При защите студент получает оценку «отлично», «хорошо», «удовлетворительно», «неудовлетворительно». Оценки выставляются согласно системе оценок и критериям их выставления, указанной в таблице 3.

Таблица 3 – Критерии оценивания

Показатель оценивания	Критерий оценивания	Оценка
Теоретическая и практическая ценность КР	Работа обладает новизной, имеет определенную теоретическую или практическую ценность	5
	Отдельные положения работы могут быть новыми и значимыми в теоретическом или практическом плане	4
	Работа представляет собой изложение известных фактов и не содержит рекомендации по их практическому использованию	3
	Полученные результаты или решение задачи не являются верными	2
Содержание работы	Содержание полностью соответствует заявленной теме. Тема раскрыта полностью. Работа отличается логичностью. Выводы обоснованы	5
	Содержание работы соответствует заявленной теме, однако она раскрыта недостаточно обстоятельно. Работа выстроена достаточно логично	4
	Содержание работы не полностью соответствует заявленной теме, либо тема раскрыта недостаточно полно.	3
	Содержание работы не раскрывает заявленную тему.	2
Использование источников	Общее количество используемых источников 10 и более, включая литературу на иностранных языках. Используется литература последних лет издания. Внутритекстовые ссылки и библиография оформлены в соответствии с ГОСТ	5
	Общее количество используемых источников не соответствует норме. Имеются погрешности в оформлении библиографического аппарата	4

Показатель оценивания	Критерий оценивания	Оценка
	Количество используемых источников недостаточно или отсутствуют источники по теме работы. Используется литература давних лет издания. Имеются серьезные ошибки в библиографическом оформлении источников	3
	Изучено малое количество литературы. Нарушены правила внутритекстового цитирования, список литературы оформлен не в соответствии с действующим ГОСТ	2
Качество защиты	Студент демонстрирует хорошее знание вопроса, кратко и точно излагает свои мысли, умело ведет дискуссию.	5
	Студент владеет проблематикой и в целом правильно излагает свои мысли, однако ему не всегда удается аргументировать свою точку зрения при ответе на вопросы	4
	Студент затрудняется в кратком и четком изложении результатов своей работы.	3
	Студент плохо разбирается в теории вопроса. Не может изложить результаты своей работы.	2

3 ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

ПК-4: Способен разрабатывать и оптимизировать средства сбора и хранения данных для решения профессиональных задач.

Индикатор ПК-4.1: Знает методы и средства сбора и хранения данных для решения профессиональных задач с использованием ETL-систем.

Тестовые задания открытого типа:

1. Технология для хранения сложной структурированной и неструктурированной информации в компьютерных системах, обеспечивающая хранение фактов об окружающем мире, а механизм вывода которой использует различные формы логики для получения новых знаний - это база _____.

Регистр не важен.

Ответ: знаний

2. _____ данные - это произвольные данные, которые явно или неявно связаны с определенными датами или промежутками времени.

Введите характеристику в именительном падеже мн. числа, регистр не важен.

Ответ: темпоральные

3. Коммерческая фирма желает установить структуру своих клиентов с точки зрения вклада в доход фирмы. Требуется выполнить задачу анализа данных: _____

Определите тип задачи в именительном падеже, регистр не важен.

Ответ: кластеризация

4.

A	B	C	D	E	F
Стр_№	Разрешение	Погода	Температура	Влажность	Ветер
1	Да	Солнечно	25	85	False
2	Да	Солнечно	20	90	True
3	Нет	Пасмурно	15	78	False
4	Нет	Дождливо	28	96	False
5	Да	Солнечно	32	80	True

Без предварительной обработки классификационными могут быть указанные в таблице признаки: _____.

Введите буквенные заголовки колонок (без пробелов, через запятую, регистр не важен).

Ответ: B, C,F

5. Клиент покупает билет на самолет через интернет. В момент покупки стоит задача определить вероятность дополнительных покупок (отель, страховка, туристический тур) и предложить наиболее вероятные пользователю. Требуется выполнить задачу анализа данных: _____.

Определите тип задачи в именительном падеже, регистр не важен.

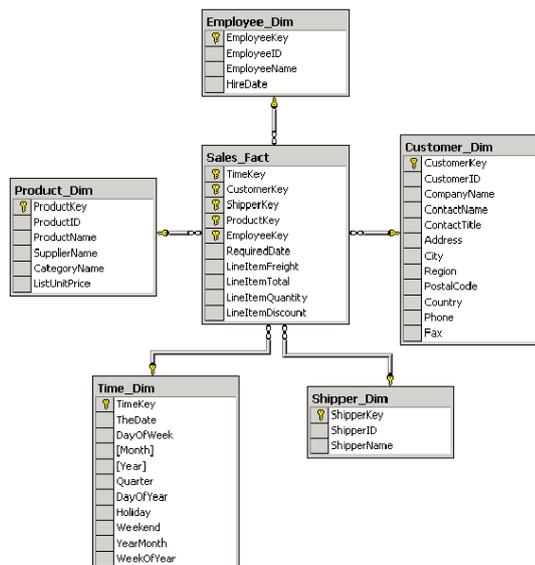
Ответ: прогнозирование

6 Шкала «очень плохо»-«плохо»-«средне»-«хорошо»-«очень хорошо» относится к типу: _____.

Введите тип шкалы в именительном падеже, регистр не важен.

Ответ: порядковая

7.



На рисунке изображена схема многомерной модели данных: _____.

Определите название схемы в именительном падеже, регистр не важен.

Ответ: звезда

Тестовые задания открытого типа (на дополнение):

1. Расшифруйте аббревиатуру ETL _____

Правильный ответ: Extract, Transform, Load

2. Дополните определение: «Кластерный анализ» это _____ . Приведите пример.

Правильный ответ: статистический метод выделения из множества элементов групп схожих между собой элементов

3. Какие термины могут употребляться вместо термина «кластерный анализ»? (не менее четырех вариантов) _____

Правильный ответ: кластеризация, самообучение, обучение без учителя, типология, классификация, автоматическая классификация, таксономический анализ, анализ образов

Тестовые задания открытого типа (с развернутым ответом):

1. Провести инфологическое проектирование базы данных компании, которая продает авиабилеты. База данных необходима для обслуживания пассажиров и менеджеров. БД содержит информацию о: пассажирах, купивших билеты, направлениях полетов, о рейсах и наличии билетов на рейс.

Ограничения: по одному направлению существует несколько рейсов. Рейсы прямые, то есть без пересадок. Билеты покупают на свободные места. Сущность «Самолет» не вводится, вместо нее работает сущность «Рейс». В билете должно быть указано название аэропорта.

Назначение базы данных.

Выдавать информацию о рейсах, следующих в аэропорт;

Выдавать список пассажиров, следующих данным рейсом;

Выдавать информацию о купленных билетах;

Показывать количество проданных билетов на рейс.

Характеристики объектов уже выделены. Ими являются: номер билета, название аэропорта, рейс, дата и время отправления, дата продажи, Ф.И.О. пассажира, ряд, место, сведения о паспорте пассажира, название направления.

1. Выполните распределение характеристик по сущностям и предложите ER-диаграмму со связями. Диаграмму выполните в виде «Название объекта» - список характеристик. Виды связи между объектами – письменно.

2. Какие поля необходимо добавить для создания схемы в реляционной базе данных в каждой сущности?

3. Нужны ли дополнительные таблицы для создания схемы БД (на примере реляционной СУБД Access)?

4. Представьте схему базы данных, указав, каким образом соединяются объекты. Примечание. Ответ представить в форме: *Сущность: (характеристика 1, характеристика 2...)*

Правильные ответы:

1. **Сущности:** «Аэропорт» (Название аэропорта); Пассажир (Ф.И.О., паспорт, пол); Рейс (макс. к-во билетов, Аэропорт отпр., Аэропорт приб., дата отпр., время отпр. к-во своб. билетов) Билет (Ф.И.О.,рейс,)

Связи между объектами: Пассажир- Билет (1:1); Рейс-Билет (1:M); Аэропорт-Рейс (M:M)

2. **Дополнительные поля:** В сущностях Рейс и Билет – «Код билета»
В сущностях Пассажир и Билет – «Код билета»; В сущностях Рейс и Аэропорт - «Код рейса» и «Код аэропорта» соответственно.

3. Нужна служебная таблица с двумя полями «Код аэропорта», «Код рейса»

4. Пассажир - Билет (1:1) по полю «Код билета» ; Рейс - Билет (1:M) по полю «Код билета»; Аэропорт-Рейс (M:M) «Код рейса» и «Код аэропорта» через служебную таблицу.

Критерии оценивания:

«зачтено» - если ответы на каждый пункт описаны правильно;

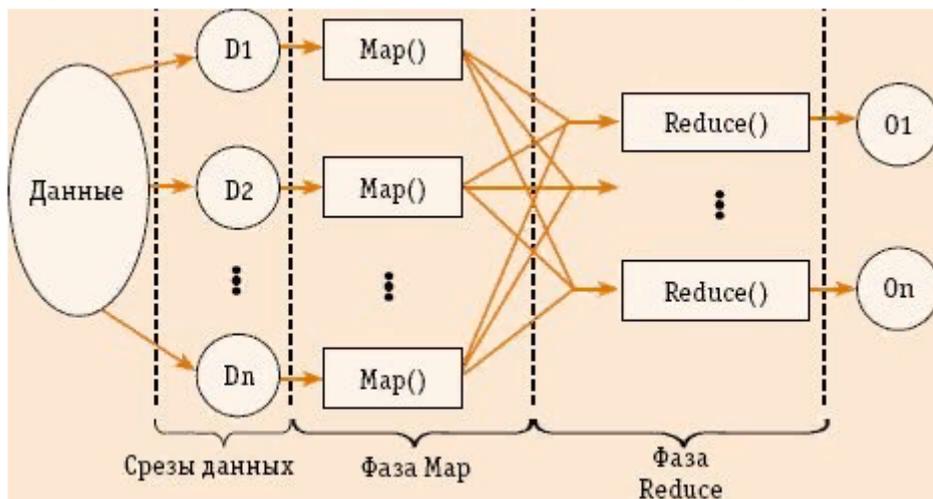
«зачтено» - в ответе правильно выделены сущности, правильно описаны связи между объектами, введены дополнительные поля, нет указаний как выполнить связь M:M;

«зачтено» - произведено правильное распределение полей в соответствии с сущностями, правильно определены типы связей, но неверно определены дополнительные поля, необходимые для создания физической модели

«зачтено» - правильно выполнен только первый пункт задания

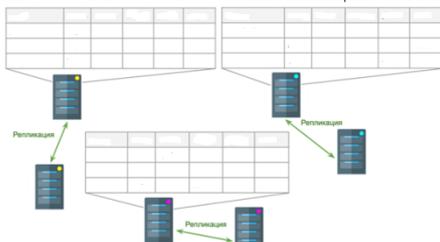
«незачтено» - ни один пункт задания не выполнен полностью.

2. Для базы данных компании, которая продает авиабилеты и содержит информацию о: пассажирах, купивших билеты, направлениях полетов, о рейсах и наличии билетов на рейс, разъяснить упрощенную схему MapReduce, представленную на рисунке

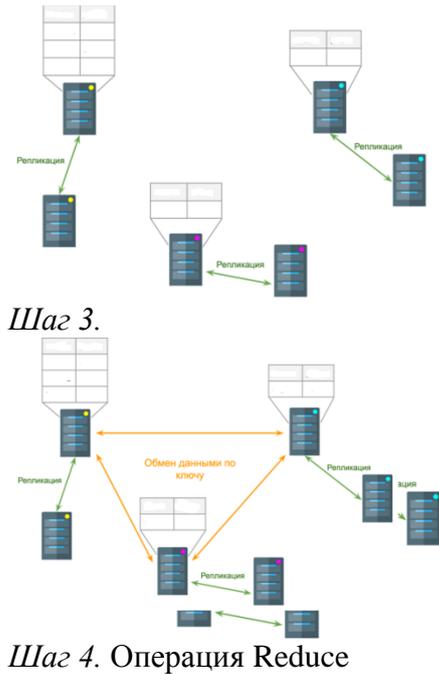


Компания нуждается в сведениях о наиболее популярных направлениях, которые приносят наибольшее количество прибыли. Охарактеризуйте каждый шаг

Шаг 1. Поясните с помощью схемы и определите, зачем «репликация»



Шаг 2.



Правильные ответы:
Операция Map:

Шаг 1. Специализированная программа HDFS (Hadoop Distributed File System) разделяет всю информацию по разным узлам сети. Репликация обеспечивает отказоустойчивость.

Шаг 2. Отфильтровываются нужные строки и выбираются колонки

Шаг 3. Определяем ключ и узлы собирают строчки с одинаковым ключом на одном узле

Операция Reduce:

Шаг 4. Для собранных строчек вычисляется сумма.

Критерии оценивания:

«зачтено» - последовательность и смысл каждого шага правильно охарактеризованы. Вместо слова «узел» может быть использовано слово «нод». Приведены примеры.

«зачтено» - последовательность и смысл каждого шага правильно охарактеризованы. Вместо слова «узел» может быть использовано слово «нод». Примеров нет.

«зачтено» - рассмотрен каждый шаг, общий смысл сохранен, но упущены важные детали

«зачтено» - указаны и рассмотрены не все пункты.

«зачтено» - дано общее определение *MapReduce*.

«незачтено» - не приведен ни один из вышеперечисленных вариантов правильных ответов

3. Для базы данных компании, которая продает авиабилеты и содержит информацию о: пассажирах, купивших билеты, направлениях полетов, о рейсах и наличии билетов на рейс, для организации СППР и принятия верных управленческих решений, было решено создать хранилище данных с целью получения информации об особенностях клиентов, а также выделения наиболее перспективных направлений и рейсов.

Вопросы и задания к кейсу:

Организация такого хранилища требует выбора измерений для анализа. Определите, какие измерения (не менее 4-х) вы бы ввели для анализа. Сформулируйте по одному из запросов аналитика, которые возможны с использованием таких измерений.

- a.
- b.
- c.
- d.

Правильный ответ:

- a. – дата вылета («Какие даты являются наиболее востребованными»);**
- b. – пол пассажира («Кто чаще летает женщины или мужчины по данному направлению?»);**
- c. –название аэропорта («Какое направление и в какие даты не пользуется популярностью?»)**
- d. . номер рейса («На какие рейсы остаются билеты?»)**

Критерии оценивания:

«зачтено» - выделены правильные измерения и сформулированы вопросы, которые соответствуют требованиям компании, сформулированным в вопросе;

«зачтено» - выделены правильные измерения, но формулировки вопросов не соответствуют требованиям компании;

«зачтено» - выделены правильные измерения, но нет формулировок гипотез аналитиков;

«зачтено» - указаны и рассмотрены менее 3-х измерений;

«зачтено» - указаны и рассмотрены менее 2-х измерений.

«незачтено» - не указано ни одного измерения

Тестовые задания закрытого типа (с одним вариантом ответа):

1. Направление информационных технологий, охватывающее всю область проблем, связанных с извлечением знаний из массивов данных, называется:

- a. интеллектуальный анализ данных**
- б. базы данных и знаний
- в. статистический анализ данных
- г. машинное обучение

2. Целью технологии MapReduce является обеспечение:

- a. параллельной обработки больших наборов данных в вычислительном кластере**
- б. хранения больших объемов, структурированных и неструктурированных данных
- в. выполнения сложного анализа данных
- г. обработка данных в режиме реального времени

3. **НЕ** используют принцип MapReduce технологии СУБД:

- a. Hadoop
- б. Cassandra

в. HDInsight

г. **Redis**

4. Преимущество колоночно-ориентированных СУБД в том, что они позволяют:

а. выполнять более сложные SQL-запросы по сравнению с реляционными СУБД

б. динамически дополнять содержание записей новыми полями

в. иметь более гибкие возможности аналитики

г. эффективно делать межколоночные сравнения

5. Результатом проектирования логической модели хранилища данных является:

а. физическая база данных

б. информационная модель данных

в. план тестирования хранилища данных

г. нормализованная схема отношений хранилища данных

д. физическая структура хранилища данных

е. спецификации модулей

6. Модель "сущность-связь" (Entity-Relationship, ER) для предметной области по программному обеспечению включает в себя следующие 5 сущностей:

а. "Software": содержит основную информацию о программном обеспечении, такую как название, версия, дата выпуска, цена и т.д.

б. "Developers": содержит информацию о разработчиках, такую как имя, контактная информация и т.д.

в. "Operating_systems": содержит информацию о операционных системах, с которыми совместимо программное обеспечение.

г. "Software_downloads": содержит информацию о загрузках программного обеспечения, такую как дата, местоположение и количество загрузок.

д. "Software_reviews": содержит информацию о отзывах пользователей о конкретном программном обеспечении, такую как оценка, комментарий и т.д.

Сущности "Software" и "Developers" связаны как:

а. один-к-одному, т. к. один программный продукт разрабатывается только одним разработчиком, и один разработчик может создать только один продукт

б. один-ко-многим, т. к. один программный продукт могут создавать несколько разработчиков

в. многие-ко-многим, т. к. один программный продукт могут создавать несколько разработчиков, и один разработчик может работать над несколькими продуктами.

г. многие-к-одному, т. к. несколько программных продуктов могут создаваться одним разработчиком

7. Функция Reduce в технологии MapReduce предназначена для:

а. обработки данных в режиме реального времени

б. обработки данных в режиме реального времени

в. сортировки данных

г. консолидации результатов на основе промежуточных данных

8. В контексте физического проектирования реляционных баз данных вертикальное разбиение таблицы – это процесс:

- а. перемещения некоторых колонок таблицы в другую новую таблицу, которая имеет тот же первичный ключ, что и исходная таблица
- б. создания двух независимых таблиц из одной таблицы
- в. декомпозиции таблицы на две или более таблиц с целью устранения частичной зависимости не ключевых колонок от составного первичного ключа
- г. создания независимых таблиц посредством намеренного дублирования колонок одной таблицы в другой

Тестовые задания закрытого типа (на последовательность и соответствие):

1. Укажите правильное соответствие между элементами логической модели и их определениями.

Элементы модели		Определения	
1	Отношение (связь)	A	– это характер участия сущности в связи.
2	Уникальный идентификатор связи	B	– это соотношение между числом сущностей, участвующих в образовании связи.
3	Степень (мощность) связи	C	– это сущность, которая не может присутствовать в базе данных, пока не существует связанного с ней экземпляра другой сущности.
4	Класс принадлежности сущности	D	– это сущность, которая порождает непересекающиеся подмножества экземпляров (категории сущности), связанных с сущностью через значения своего атрибута.
5	Рефлексивное отношение	E	– это категория порождающей сущности (сущности- супер-типа).
6	Слабая сущность	F	– представляет собой соединение (взаимоотношение) между двумя или более сущностями.
7	Супертип	G	– это супертип с порожденными им подтипами.
8	Подтип	H	– это атрибут для однозначной идентификации отношения.
9	Составная сущность	J	– это отношение, связывающее сущность саму с собой.

Введите последовательность букв в порядке расположения элементов модели в таблице, без пробелов, через запятую.

Правильный ответ: F, H, B, A, J, C, D, E, G

2. Укажите главную задачу на каждом этапе бизнес - модели создания хранилища данных (ХД).

Этапы		Главная задача этапа	
1	Формулирование требований	A	проектирование и разработка аналитических отчетов на спроектированной структуре данных

2	Создание вычислительной среды для хранилища данных	В	формирование программной среды, в которой пользователи будут извлекать данные из (ХД) и просматривать predetermined отчеты
3	Расширение и реализация концептуальной модели данных	С	добиться оптимальной производительности ETL-процессов, производства отчетов и их доставки конечному пользователю
4	Определение процедур извлечения, преобразования и загрузки данных	D	разработка логической и физической моделей данных для ХД
5	Проектирование аналитических отчетов	E	идентификация и определение процедур извлечения, очистки (фильтрации), преобразования и загрузки данных
6	Разработка приложений хранилища данных	F	идентификация требований заказчика ХД и оформление их в виде документа "Каталог требований"
7	Настройка производительности	G	убедиться, что ХД готово к эксплуатации. Как правило, проверка качества выполняется отдельной группой специалистов, не входящих в состав команды разработчиков
8	Проверка качества	H	создание информационно-вычислительной среды, в которой будет разрабатываться ХД

Введите последовательность букв в порядке расположения этапов в таблице, без пробелов, через запятую.

Правильный ответ: F, H, D, E, A, B, C, G

3. *Вопрос:* Процесс преобразования данных состоит из последовательности подпроцессов. Укажите соответствие между названием подпроцесса и его описанием.

Этапы		Главная задача этапа	
1	Очистка	A	Разделение данных на столбцы
2	Фильтрация	B	данные собираются и хранятся в едином формате.
3	Объединение	C	бизнес-правила применяются к данным и проверяются на достоверность, и если они обнаруживаются неверными, они возвращаются обратно в источник
4	Сортировка	D	отклонение данных, в которых отсутствует какое-либо значение по умолчанию или заданный формат
5	Дедупликация	E	удаление или удаление ошибок или несоответствий данных для улучшения качества данных
6	Обобщение	F	связывание релевантных данных из нескольких источников
7	Вывод	G	упорядочение записей по некоторому признаку
8	Проверка данных	H	выбор соответствующих строк или столбцов
9	Разбиение	I	поиск и удаление повторяющихся записей

Введите последовательность пар «Число» - «Буква»

Правильный ответ: 1-E; 2-H; 3-F; 4-G; 5- I; 6-B; 7-C; 8-D; 9-A

4. *Вопрос:* В базе данных PostgreSQL в результате выполнения операторов получаются результаты:

Оператор		Результат	
1	INSERT INTO software (name, version, release_date, price, company, language, type, platform, description) VALUES ('Windows 10', '1909', '2019-11-12', '199.99', 'Microsoft', 'English', 'Operating System', 'Windows', 'The latest version of the Microsoft Windows operating system.')	A	Q1
2	INSERT INTO software_downloads (software_id, download_date, download_location, download_count) VALUES (1, '2022-01-01', 'https://example.com/software1', 200), добавит в таблицу запись (связанную с записью таблицы software), в которой значение внешнего ключа равно ____.	B	1909
3	В базе данных PostgreSQL таблица, созданная запросом: CREATE TABLE Q1 (software_id SERIAL PRIMARY KEY, name VARCHAR(255) NOT NULL, version VARCHAR(255) NOT NULL, release_date DATE NOT NULL, price NUMERIC(10,2) NOT NULL, company VARCHAR(255) NOT NULL, language VARCHAR(255) NOT NULL, type VARCHAR(255) NOT NULL, platform VARCHAR(255) NOT NULL, description VARCHAR(255) NOT NULL); получит имя ____	C	1

Правильный ответ: 1-В; 2-С; 3-А

5. Расставьте номера по порядку выполнения действий в методике анализа больших данных:

Название шага методики анализа больших данных	
1	Систематизация данных
2	Эксплуатация модели
3	Проверка полученной модели
4	Сбор экспериментальных данных
5	Поиск модели, объясняющей имеющиеся данные
6	Получение новых экспериментальных данных на основе построенной модели

Введите последовательность цифр в порядке следования шагов при анализе больших данных без пробелов, через запятую.

Правильный ответ: 4, 1, 5, 3, 2, 6

Критерии оценивания:

Указания по оцениванию	Баллы
Элементы ответа: 4,1,5,3,2,6	
Указания по оцениванию:	
Последовательность указана верно	4
В ответе более 4 элементов стоят на своих местах	3
В ответе 2 элемента стоят на своих местах	2
В ответе только 1 элемент стоит на своем месте	1
В ответе все варианты указаны неверно	0

6. Укажите последовательность шагов для развертывания модели больших данных.

Наименование шагов	
1	Хранение данных , то есть хранение их в базах данных HDFS, NoSQL или HBase.
2	Обработка данных , то есть обработка с помощью фреймворков Hadoop, Spark, MapReduce или Flink и Pig
3	Прием данных , то есть извлечение данных из нескольких источников данных. Этот процесс включает сбор данных из нескольких источников, таких как сайты социальных сетей, корпоративное программное обеспечение и файлы журналов.

Введите последовательность цифр в порядке обработки, без пробелов, через запятую.

Правильный ответ: 3, 1, 2

Критерии оценивания:

Указания по оцениванию	Баллы
Элементы ответа: 3,1,2	
Указания по оцениванию:	
Последовательность указана верно	3
В ответе только 2 элемента стоят на своих местах	1
В ответе все варианты указаны неверно	0

4 ТИПОВЫЕ ЗАДАНИЯ НА КОНТРОЛЬНУЮ РАБОТУ, КУРСОВУЮ РАБОТУ/КУРСОВОЙ ПРОЕКТ

4.1 Курсовая работа (КР) направлена на закрепление полученных теоретических знаний и приобретение умений и навыков в области разработки баз данных и элементов ETL-системы.

Курсовая работа может выполняться индивидуально, либо в составе проектной команды. В случае командной разработки студенты готовят общую пояснительную записку с указанием (распределением) ролей участников команды.

Примерное задание и темы курсовой работы ниже.

Задание:

В ходе выполнения курсовой работы в зависимости от темы проведите:

1. Анализ предметной области и разработку концептуальной модели базы данных.

2. Разработка логической и физической модели база данных, ее реализация в подходящей СУБД.

3. Анализ структуры данных, подлежащих хранению в БД.

4. Разработка механизмов предподготовки, консолидации и загрузки данных из различных источников данных в разработанную БД.

Типовые темы курсовой работы:

1. Разработка базы данных и ETL-процессов информационной системы «Название ИС».

2. Разработка ETL-процессов информационной системы на базе 1С: Предприятие.

К защите курсовой работы допускается обучающийся, выполнивший работу по утвержденной теме.

5 СВЕДЕНИЯ О ФОНДЕ ОЦЕНОЧНЫХ СРЕДСТВ И ЕГО СОГЛАСОВАНИИ

Фонд оценочных средств для аттестации по дисциплине «ETL-системы и базы данных» представляет собой компонент основной профессиональной образовательной программы магистратуры по направлению подготовки 09.04.01 Информатика и вычислительная техника.

Фонд оценочных средств рассмотрен и одобрен на заседании методической комиссии института цифровых технологий (протокол № 2 от 26.04.2022 г.).

Фонд оценочных средств актуализирован. Изменения, дополнения рассмотрены и одобрены на заседании методической комиссии института цифровых технологий (протокол № 3 от 24.03.2023 г.).

Директор института



А.Б. Тристанов