



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

УТВЕРЖДАЮ
И.о.директора института

Фонд оценочных средств
(приложение к рабочей программе дисциплины)
«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, ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ СРЕДСТВ И КРИТЕРИИ ОЦЕНИВАНИЯ

1.1 Результаты освоения дисциплины

Таблица 1 – Планируемые результаты обучения по дисциплине, соотнесенные с установленными компетенциями

Код и наименование компетенции	Дисциплина	Результаты обучения (владения, умения и знания), соотнесенные с компетенциями
ОПК-3: Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями ОПК-4: Способен применять на практике новые научные принципы и методы исследований	ETL-системы и базы данных	<u>Знать:</u> - 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; <u>Уметь:</u> - пользоваться методами и инструментами получения, хранения, передачи, ETL-обработки данных; <u>Владеть:</u> - навыками разработки предложений по развитию и совершенствованию системы получения, хранения, передачи и ETL-обработки данных.

1.2 К оценочным средствам текущего контроля успеваемости относятся:

- тестовые задания открытого и закрытого типов.

К оценочным средствам для промежуточной аттестации в форме экзамена (первый семестр) относятся:

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

1.3 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 полнотой 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	

1.4 Оценивание тестовых заданий закрытого типа осуществляется по системе зачтено/ не зачтено («зачтено» – 41-100% правильных ответов; «не зачтено» – менее 40 % правильных ответов) или пятибалльной системе (оценка «неудовлетворительно» - менее 40 % правильных ответов; оценка «удовлетворительно» - от 41 до 60 % правильных ответов; оценка «хорошо» - от 61 до 80% правильных ответов; оценка «отлично» - от 81 до 100 % правильных ответов).

Тестовые задания открытого типа оцениваются по системе «зачтено/ не зачтено». Оценивается верность ответа по существу вопроса, при этом не учитывается порядок слов в словосочетании, верность окончаний, падежи.

2 ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

ОПК-3: Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями

Тестовые задания закрытого типа:

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

1. интеллектуальный анализ данных
2. базы данных и знаний
3. статистический анализ данных
4. машинное обучение

2. Целью технологии MapReduce является обеспечение:

1. параллельной обработки больших наборов данных в вычислительном кластере
2. хранения больших объемов структурированных и неструктурированных данных
3. выполнения сложного анализа данных
4. обработка данных в режиме реального времени

3. НЕ используют принцип MapReduce технологии СУБД:

1. Hadoop
2. Cassandra
3. HDInsight
4. Redis

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

1. выполнять более сложные SQL-запросы по сравнению с реляционными СУБД
- 2. динамически дополнять содержание записей новыми полями**
3. иметь более гибкие возможности аналитики
4. эффективно делать межколоночные сравнения

5. 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. Регистр не важен.

Ответ: F,H,B,A,J,C,D,E,G

6. Главная задача 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	B	формирование программной среды, в которой пользователи будут извлекать данные из (ХД) и просматривать предопределенные отчеты
3	Создание вычислительной среды для хранилища данных	C	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. Регистр не важен.

Ответ: F,H,D,E,A,B,C,G

7. Технология преобразования данных состоит из ряда процедур. Соответствие между названием процедуры и ее функцией:

Этапы		Главная задача этапа	
1	Очистка	A	разделение данных на столбцы
2	Фильтрация	B	приведение данных к некоторой общей шкале без потери информации о различии диапазонов
3	Объединение	C	предобработка с целью приведению данных к единому формату и представлению, наиболее удобному для использования определенного вида обработки
4	Сортировка	D	процесс проверки информации на соответствие определённым критериям и правилам до её использования в дальнейшей обработке
5	Дедупликация	E	удаление ошибок или несоответствий данных для улучшения их качества
6	Нормализация	F	связывание релевантных данных из нескольких источников
7	Стандартизация	G	упорядочение записей по некоторому критерию
8	Валидация данных	H	выбор строк или столбцов в соответствии с критериями
9	Разбиение	I	поиск и удаление повторяющихся записей

Введите последовательность пар «Число» - «Буква»

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

8. Правильный порядок выполнения действий в методике 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.

Ответ: 4, 1, 5, 3, 2, 6

Критерии оценивания: последовательность указана верно – 5, в ответе более 4 элементов стоят на своих местах – 4, в ответе 2 элемента стоят на своих местах 3, в ответе только 1 элемент стоит на своем месте – 2, в ответе все варианты указаны неверно – 0.

Тестовые задания открытого типа:

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

Регистр не важен.

Ответ: знаний

10. _____ 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

11. Расшифровка аббревиатуры ETL _____

Ответ: Extract, Transform, Load

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

Ответ: статистический метод выделения из множества элементов групп схожих между собой элементов

13. Вместо термина «кластерный анализ» могут употребляться термины: _____ (не менее четырех вариантов).

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

14. _____ данных – это предметно-ориентированная база данных, содержащая данные по одному из направлений деятельности, информация в которой хранится оптимизированно с точки зрения решения конкретных задач.

Введите название в именительном падеже, регистр не важен.

Ответ: витрина

15. Хранилище всех типов необработанных данных в оригинальном формате, которые специалисты по обработке и анализу данных могут использовать для различных проектов – это _____ данных.

Введите название в именительном падеже, регистр не важен.

Ответ: озеро

16. Компонентом программных решений класса Business Intelligence является _____ - технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу.

Вставьте пропущенное слово (аббревиатура, англ.) Допускаются иные формулировки ответа, не искажающие его смысла.

Ответ: OLAP

17. Совокупность элементов данных и спецификаций, содержащих описание данных информационной системы и процессов их обработки – это _____ .

Введите название в именительном падеже, регистр не важен..

Ответ: метаданные

18. Ключ в измерении хранилища данных, **НЕ** имеющий смыслового значения, используемый для связей – _____ ключ.

Ответ: суррогатный

19. Язык запросов, используемый для работы с многомерными структурами данных (в контексте OLAP баз данных) : _____ (аббр., англ.)

Ответ: MDX

20. Охарактеризуйте технологию CDC (Change Data Capture).

Ответ: CDC позволяет отслеживать изменения в данных, происходящие в базе данных и передавать их в целевые системы в реальном времени или близком к нему. Отслеживаются и фиксируются только изменения, происходящие в базе данных (добавление, обновление, удаление записей), в результате загружается не весь объем данных, что значительно повышает эффективность ETL-процессов.

ОПК-4: Способен применять на практике новые научные принципы и методы исследований

Тестовые задания закрытого типа:

21. Результатом проектирования логической модели хранилища данных является:

1. физическая база данных
2. информационная модель данных
3. план тестирования хранилища данных

4. нормализованная схема отношений хранилища данных

5. физическая структура хранилища данных

6. спецификации модулей

22.

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, в качестве классификационных, могут быть выбраны признаки, указанные в таблице: _____ .

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

Ответ: B,C,F

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

1. "Software": содержит основную информацию о программном обеспечении, такую как название, версия, дата выпуска, цена и т.д.
2. "Developers": содержит информацию о разработчиках, такую как имя, контактная информация и т.д.
3. "Operating_systems": содержит информацию о операционных системах, с которыми совместимо программное обеспечение.
4. "Software_downloads": содержит информацию о загрузках программного обеспечения, такую как дата, местоположение и количество загрузок.
5. "Software_reviews": содержит информацию об отзывах пользователей о конкретном программном обеспечении, такую как оценка, комментарий и т.д.

Между сущностями "Software" и "Developers" связь определяется как:

1. один-к-одному, т. к. один программный продукт разрабатывается только одним разработчиком, и один разработчик может создать только один продукт
2. один-ко-многим, т. к. один программный продукт могут создавать несколько разработчиков
3. **многие-ко-многим, т. к. один программный продукт могут создавать несколько разработчиков, и один разработчик может работать над несколькими продуктами.**
4. многие-к-одному, т. к. несколько программных продуктов могут создаваться одним разработчиком

24. Функция Reduce в технологии MapReduce предназначена для:

1. обработки данных в режиме реального времени

2. обработки данных в режиме реального времени
3. сортировки данных
4. консолидации результатов на основе промежуточных данных

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

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

Тестовые задания открытого типа:

26. Коммерческая фирма желает установить типизацию своих клиентов с точки зрения вклада в доход фирмы. Выполняется задача анализа данных: _____

Укажите тип задачи в именительном падеже, регистр не важен.

Ответ: кластеризация

27. 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: _____

Укажите тип задачи в именительном падеже, регистр не важен.

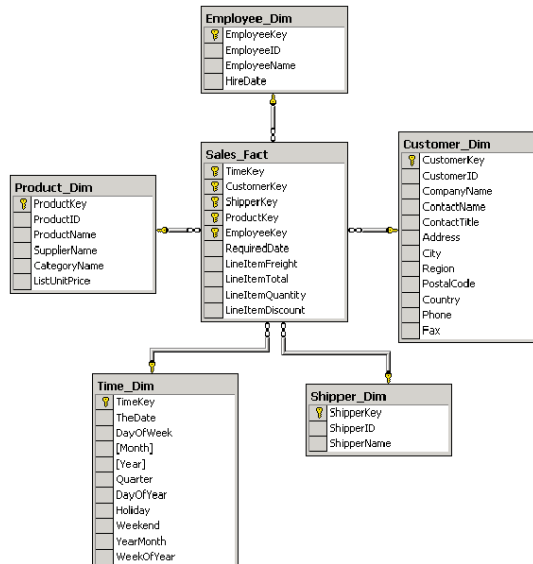
Ответ: прогнозирование

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

Введите тип шкалы в именительном падеже, регистр не важен.

Ответ: порядковая

29.



На рисунке изображена схема многомерной модели данных: _____ .

Укажите название схемы в именительном падеже, регистр не важен.

Ответ: звезда

30. 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...)

Ответ:

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) 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;

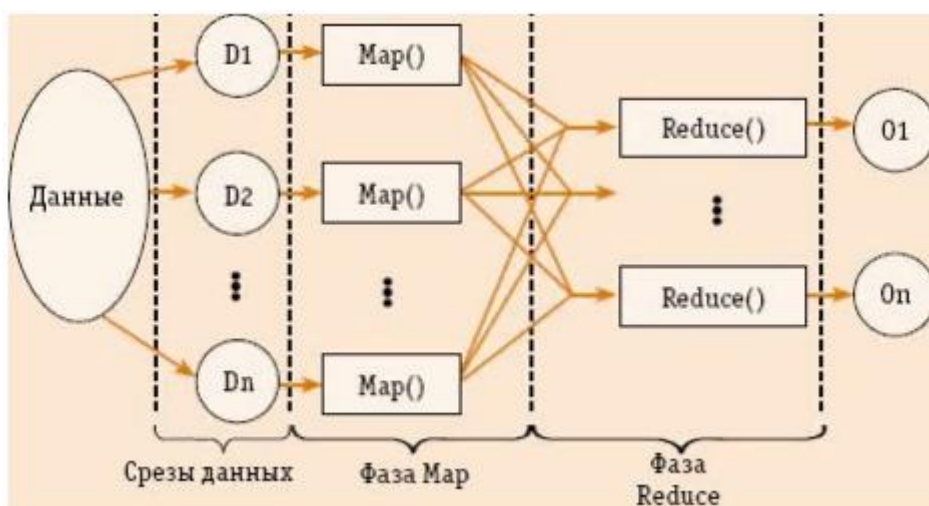
«зачтено» - в ответе правильно выделены сущности, правильно описаны связи между объектами, введены дополнительные поля, нет указаний как выполнить связь M:M;

«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

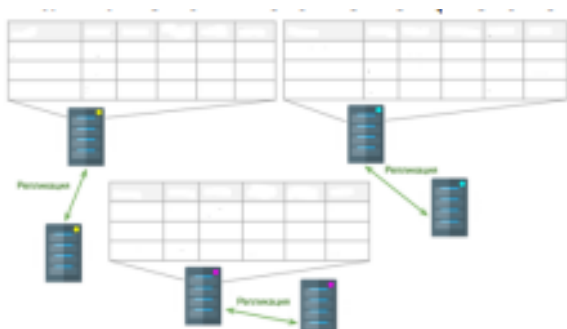
«не зачтено» - ни один пункт задания не выполнен полностью.

31. 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 максимальную прибыль. Охарактеризуйте каждый шаг.

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

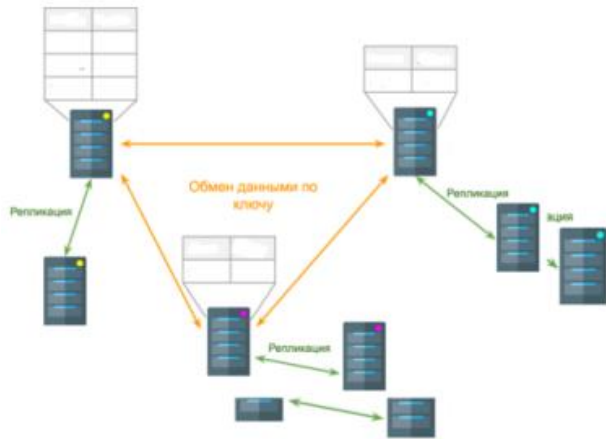


Шаг 2.



+

Шаг 3.



Шаг 4. Операция Reduce

Ответ:

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.

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

32. 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.

а.

- 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.

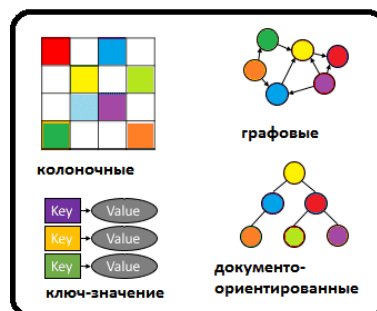
«не зачтено» - не указано ни одного измерения

33. В схеме отношения ПРЕПОДАВАТЕЛЬ_ПРЕДМЕТ (Шифр, Предмет, Фамилия, Должность, Оклад, Часы) ключом (ключами) отношения является: _____ .

Введите название поля (полей - без пробелов, через запятую, регистр не важен).

Ответ: Шифр, Предмет

34.

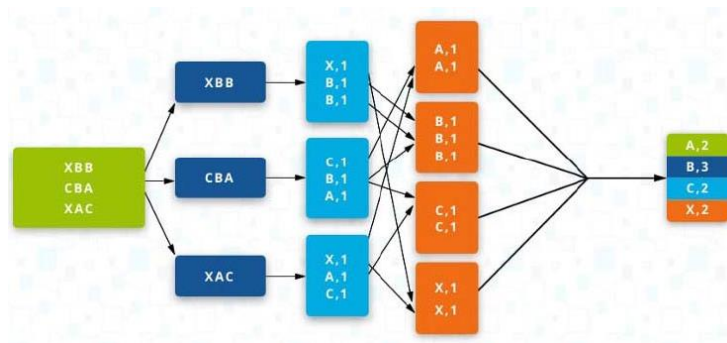


Для работы со всеми представленными на рисунке типами хранилищ данных используют _____-СУБД.

Укажите вид (аббревиатура, англ.). Допускаются иные формулировки ответа, не искажающие его смысла.

Ответ: NoSQL.

35. На рисунке:



изображена схема работы алгоритма _____.

Введите название (англ.) Регистр важен. Допускаются иные формулировки ответа, не искажающие его смысла.

Ответ: MapReduce

36. 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 _____.

Введите имя таблицы, регистр важен.

Ответ: Q1

37. В базе данных PostgreSQL в результате выполнения оператора:

```
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.')
```

в добавляемой записи поле «version» примет значение _____.

Ответ: 1909

38. В PostgreSQL в результате выполнения запроса создана таблица:

```
CREATE TABLE software_downloads (
  download_id SERIAL PRIMARY KEY,
  software_id INTEGER REFERENCES software(software_id) NOT NULL,
  download_date DATE NOT NULL,
  download_location VARCHAR(255) NOT NULL,
```


download_count INTEGER NOT NULL
);

Запрос:

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 _____.

Введите число.

Ответ: 1

Таблица 3

software_id	name	version	release_date	price	company	language	type	platform
1	Windows 10	1909	2019-11-12	199.99	Microsoft	English	Operating System	Windows
2	Office 365	2019	2018-09-24	69.99	Microsoft	English	Productivity Suite	Windows, Mac, iOS, Android
3	Adobe Photoshop	2021	2020-10-20	20.99	Adobe Systems	English	Image Editing Software	Windows, Mac
4	MacOS	Catalina	2019-10-07	0.0	Apple	English	Operating System	Mac
5	Visual Studio Code	1.58	2021-03-17	0.0	Microsoft	English	Source Code Editor	Windows, Mac, Linux
6	VSCode	1.58	2021-03-17	0.0	Microsoft	English	Source Code Editor	Windows, Mac, Linux
7	Google Chrome	89.0	2021-03-02	0.0	Google	English	Web Browser	Windows, Mac, Linux, iOS, Android
8	Mozilla Firefox	87.0	2021-03-09	0.0	Mozilla Foundation	English	Web Browser	Windows, Mac, Linux
9	Ubuntu	20.04	2020-04-23	0.0	Canonical	English	Operating System	Windows, Mac, Linux
10	Spotify	8.6.0	2021-01-25	0.0	Spotify AB	English	Music Streaming Service	Windows, Mac, Linux, iOS, Android
11	Git	2.29.2	2021-03-29	0.0	Git Project	English	Version Control System	Windows, Mac, Linux
12	Docker	19.03.15	2021-02-22	0.0	Docker Inc.	English	Containerization Platform	Windows, Mac, Linux
13	Java	15	2020-09-15	0.0	Oracle	English	Programming Language	Windows, Mac, Linux
14	Python	3.10	2021-03-15	0.0	Python Software Foundation	English	Programming Language	Windows, Mac, Linux
15	SQL Server	2019	2019-11-04	0.0	Microsoft	English	Relational Database Management System	Windows

software_id	name	version	release_date	price	company	language	type	platform
16	MySQL	8.0.24	2021-03-15	0.0	Oracle	English	Relational Database Management System	Windows, Mac, Linux
17	PostgreSQL	15.1	2021-03-29	0.0	PostgreSQL Global Development Group	English	Relational Database Management System	Windows, Mac, Linux

39. Из таблицы 3 после выполнения запроса:

`SELECT * FROM software WHERE type <> 'Operating System' AND company='Microsoft'`

будет извлечено количество записей: _____

Введите число.

Ответ: 4

40. В таблицу 3 столбец license_type (до 50 символов) добавится запросом: _____

Ответ: ALTER TABLE software ADD COLUMN license_type VARCHAR(50)

41. Запрос к таблице 3 для подсчета количества бесплатных программ на платформе Linux:

Ответ: SELECT COUNT(*) FROM software WHERE price = 0.0 AND platform LIKE '%Linux%'

42. Запрос к таблице 3 для подсчета общего дохода от платного ПО:

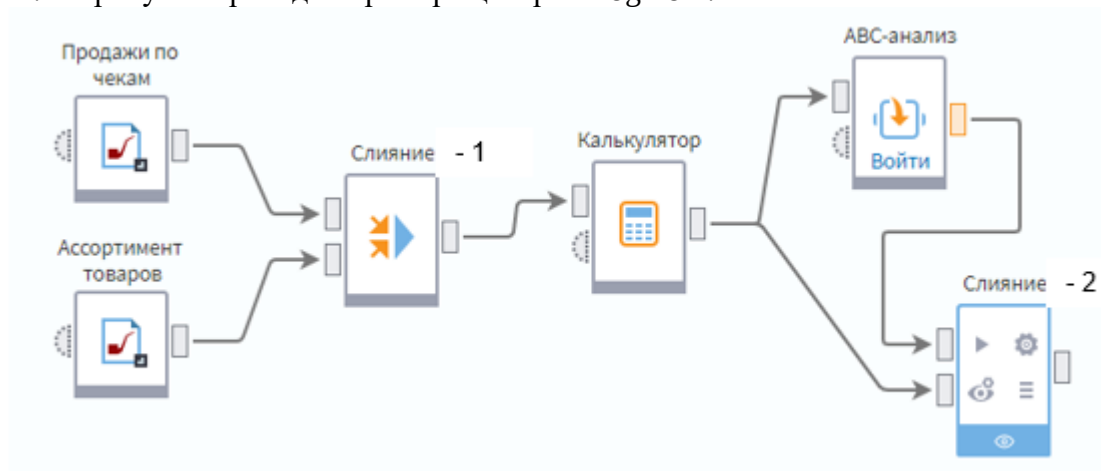
Ответ: SELECT SUM(price) FROM software WHERE price > 0

43. В технологии MapReduce обработку и преобразование данных в промежуточную форму выполняет функция _____ .

Регистр важен.

Ответ: Map

44. На рисунке приведен пример сценария Loginom.



Узлы обозначены буквами:

A	B	C	D	E	F
---	---	---	---	---	---

Продажи по чекам	Ассортимент товаров	Слияние - 1	Калькулятор	ABC -анализ	Слияние - 2
------------------	---------------------	-------------	-------------	-------------	-------------

В понятиях ETL-процесса за извлечение данных из внешних источников отвечает:

_____ .
Введите обозначения узла (узлов - без пробелов, через запятую, регистр не важен).

Ответ: A,B

45. На рисунке представлен фрагмент окна настройки подключения базы данных PostgreSQL в системе Loginom:

PostgreSQL (параметры подключения)

Метка: PostgreSQL

Строка подключения: [Тестировать](#)

Логин: Пароль:

Настройки SSL:

В Строке подключения указывают количество параметров: обязательных - _____, опциональных - _____ .

Введите два числа через запятую, без пробелов.

Ответ: 1,2

3 ТИПОВЫЕ ЗАДАНИЯ НА КОНТРОЛЬНУЮ РАБОТУ, КУРСОВУЮ РАБОТУ/ КУРСОВОЙ ПРОЕКТ, РАСЧЕТНО-ГРАФИЧЕСКУЮ РАБОТУ

Учебным планом предусмотрено выполнение курсовой работы (КР).

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

КР 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 КР 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 разнородных 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.

4 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.

Преподаватель-разработчик – к.п.н. Розен Н.Б.,

Фонд оценочных средств рассмотрен и одобрен кафедрой прикладной математики и информационных технологий.

И.о. заведующего кафедрой



И.Г. Булан

Фонд оценочных средств рассмотрен и одобрен методической комиссией института цифровых технологий (протокол №3 от 29.04.2025 г).

Председатель методической комиссии



О.С. Витренко