Аналитические базы данных-специализированные РСУБД, оптимизированные для быстрой выборки данных из витрин. (колоночные)

Большие данные - это разнообразные данные, которые поступают с постоянно растущей скоростью и объем которых постоянно растет

В

В каких задачах не применяются регулярные выражения (детекция)

В объектно-ориентированном подходе (ООП) все сущности формализуются набором и методов (полей)

В формате данных JSON данные хранятся в виде (формата «ключ-заключение»)

В чем состоят основные отличия ETL от ELT процессов? (реализация, работа)

Возможно ли использование одновременно двух агрегирующих функций: SELECT MIN (price), MAX(price) FROM ORDERS; (Да, в результате мы получим минимальную и максимальную стоимости.)

Выберите верные основные задачи бизнес-аналитики (Изучение и формализация предметной области клиента,Оптимизация бизнес-процессов, Разработка характеристик IT продукта, Внедрение новых характеристик продукта, Расчет метрик качества принятия решений, Аналитическая отчетность и визуализация)

Выберите верные свойства, присущие распределенной системе обработки и хранения данных: (Совместное использование ресурсов; Совместная обработка данных; Обнаружение отказов; Масштабируемость по программным)

Выберите верные стадии ETL процесса, расположите их в верном порядке следования. Запишите число без пробелов, состоящее из цифр последовательности (экспорт, трансформация, загрузка)

Выберите верные стадии ETL процесса, расположите их в верном порядке следования.

Выберите верные утверждения об аналитических базах данных: (Основаны на системах массивных параллельных вычислений (МРР) и базах данных, поддерживающих такой режим работы, Предназначены такие БД для обработки данных на стадии работы непосредственно с витринами больших данных)

Выберите верные утверждения, относящиеся к понятию неструктурированных данных (не имеет, имеет большое, хранят в озерах)

Выберите задачи, решаемые с помощью машинного и глубокого обучения с приемлемым качеством (все ответы)

Выберите классические задачи обучения с учителем (классификация, регрессия)

Выберите корректно составленный запрос с функцией. GROUP BY (Orders GROUP BY seller\_id)

Выберите корректный пример использования функции. CONCAT (SELECT CONCAT (index'," ", 'city') FROM)

Выберите неверное утверждение о кросс-таблице (выполняет операцию)

Выберите правильный пример использования функции округления ROUND (SELECT id, ROUND (price \* discount, 2)

Выберите пример правильно составленного запроса с использованием агрегирующей функции SUM (SELECT SUM(price) FROM Orders;)

Д

Для каких типов данных НЕ имеет смысл выбор показателя в качестве группы при проведении агрегации данных (группировка)? (Вещественный тип данных непрерывного признака)

Для каких типов данных НЕ имеет смысл выбор показателя в качестве группы при проведении агрегации данных (группировка)? Вещественный тип данных непрерывного признака

Для чего необходимо ведение журналов транзакций транзакционными системами РСУБД? (журнализация, поддержка, восстановление)

З

Запишите число без пробелов, состоящее из цифр последовательности (экспорт, трансформация, загрузка)

К

Какая функция из перечисленных НЕ является агрегационной (использующимися при агрегации, группировке)? (корень числа)

Какие возможности предоставляет система управления базами данных? (все перечисленное)

Какие задачи решает агрегация (группировка) данных? (Подсчет агрегационных функций (статистик) по подвыборкам показателей по группам совпадающих дискретных признаков.)

Какие задачи решает аналитика данных? (получение, описание)

Какие особенности реляционных баз данных (все перечисленное)

Какие существуют основные свойства шкал измерений? (идентифицируемость, величина, абсолютный ноль, равенство инеравалов)

Какие факторы ограничивают кратное повышение производительности в решениях на базе распределенных архитектур? (все перечисленное)

Какими свойствами должно обладать хранилище данных? (все ответы)

Какое ключевое слово SOL запроса позволяет определить максимальное количество строк в выходной таблице данных? (LIMIT)

Какое ключевое слово SQL запроса позволяет определить условие фильтрации выборки данных? (WHERE)

Какое ключевое слово SQL запроса позволяет упорядочить строки в соответствии со значениями в выбранных полях?(ORDER BY)

Какое сравнение значений в полях таблицы при проведении процедуры фильтрации данных не является возможным? (заранее определенному типу данных)

Какой протокол используется для передачи данных между распределенными приложениями на Linux-серверах? (SSH)

Какой тип данных используется для хранения URL-адресов? (TEXT, VARCHAR)

Какой тип данных используется для хранения видеофайлов? (BLOB)

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)

М

Модель данных, предполагающая наличие ориентированной связи между объектами, от родительского объекта к дочерним называется (иерархической)

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

О

Основными функциями СУБД являются (Управление х2, поддержка языков, журнализация)

Отметьте верные примеры неструктурированных данных(Видеофайлы, Файлы изображений, Файлы, хр. текст. инф)

Отметьте верные стадии очистки данных (обработка х2, удаление, восстановление, верификация)

Отметьте основные принципы организации данных в витринах данных (секционирование, изоляция)

Отметьте основные принципы организации данных в витринах данных (Секционирование данных по тематикам (отделам), Изоляция витрин данных на уровне предоставления прав доступа на чтение определенному кругу ролей)

П

Процесс загрузки данных в долговременное хранилище данных заключается в (переносе данных из промежуточных таблиц, образованных на стадии предобработки данных, в структуры хранилища данных)

Р

Расположите стадии проведения аналитики данных в правильном порядке (25 14 67 38)

С

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, архитектура, репликация)

Совокупность программных и языковых средств для создания, удаления, изменения и любых других манипуляций с данными в БД (система управления базами данных)

Сопоставьте компоненты конвейера обработки данных с их функциями (2415)

Сортировка записей по нескольким полям табличных данных производится в соответствии с правилом:(Лексикографическим правилом в соответствии с выбранным пользователем порядком старшинства полей таблицы)

Среди шкал выберите те, в которых мерой среднего может служить медиана (Порядковая шкала (ординальная, ранговая), Интервальная шкала (разностей), Шкала отношений (абсолютная)

Среди шкал выберите те, в которых мерой среднего может служить среднее арифметическое Шкала отношений (абсолютная)

Среди шкал выберите те, на которых невозможно ввести отношение порядка: (Шкала отношений; Порядковая)

Среди шкал выберите те, с которыми невозможны арифметические операции (интервальная шкала, шкала отношений)

У

Упорядочивание записей в таблице по строковой переменной осуществляется по правилу (посимвольное)

Ч

Чем характеризуются архитектуры главный/подчиненный в исполнении серверных решений? (Все перечисленное)

Что такое SQL (Язык запросов к базам данных)

Что такое первичный ключ в реляционной модели данных? (Уникальный идентификатор записи в таблице)

Что такое реляционная модель данных? (Модель данных, основанная на связях между таблицами)

Что такое таблица в реляционной модели данных? (Структура, хранящая данные в виде строк и столбцов)

Ш

Шаг по оси измерения имеет фиксированный шаг. (квантование)

... развитие подразумевает собой уменьшение расходов для платформ и/или продуктов по средствам, времени и качеству решения задачи. (Экстенсивное)

... таблицы в обработке и хранении данных это перечень полей и их типов данных, которыми представлена таблица данных (Структура)

… система - это совокупность компьютеров, которые обмениваются данными и синхронизируются в общей сети, образуя «единый компьютер» для конечного пользователя(распределенная)

… определяет как и каким образом данные будут располагаться в БД и как к ним будет предоставлен доступ (схема данных)

… данные являются значениями признака, общее число которых конечной или бесконечно, но может быть подсчитано при помощи натуральных чисел. (Дискретные)

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

(Трансформация данных)

... данных - это процесс дополнения сырых данных той информацией, которая в исходном виде в них отсутствует, но необхолима пля качественного анализа. (обогащение)

... этап удаления нерелевантных значений показателей или записей данных с нетипичными значениями (очистка данных)

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

... данных - разбиение диапазона значений непрерывной или дискретной величины на конечное число интервалов.

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

... это цифровая система хранения, которая выполняет объединение и согласование больших объемов данных из разных источников. (хранилище данных)

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

... 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.

(Неструктурированные)

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

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

... называют данные, которые содержат ошибки, пропуски, дубликаты, несогласованность в формате или некорректную информацию (грязными)

... это система аналитической обработки данных. Она предназначена для подготовки отчетов, построения прогностических сценариев и выполнения статистических расчетов. (OLAP)

... 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 (Схема)

... данных - процесс получения обобщенных статистик для некоторой большой выборки табличных данных с целью получить важную информацию по уникальным группам категорий объектов (группировка)

... данных - процесс измерения показателя в дискретные моменты времени с заранее заданной частотой дискретизации, так чтобы перевести непрерывную или аналоговую информацию в цифровой вид (дискретизация)

- группа операторов определения, манипуляции данных, переводящих базу данных из одного согласованного состояния в другое согласованное состояние. (Транзакция)

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

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

Для манипуляций над данными в реляционных СУБД используют декларативный язык запросов (SOL)