1. Язык PL/SQL. Типы данных(простые и RECORD). Неявное объявление переменных.

Язык PL/SQL. Типы данных (простые и RECORD):

Простые типы данных в PL/SQL включают целые числа, числа с плавающей точкой, символы, строки и т.д. Они используются для хранения значений.

Тип данных RECORD позволяет создавать пользовательские структуры данных, объединяя несколько полей в одном объекте.

2. Язык PL/SQL. Хранимые процедуры и функции.

Хранимые процедуры и функции в PL/SQL представляют собой блоки кода, которые могут быть вызваны и выполняться в базе данных.

Процедуры выполняют действия, а функции возвращают значения.

Они используются для повторного использования кода, улучшения производительности и обеспечения безопасности данных.

3. Язык PL/SQL. Курсоры, команды для работы с курсорами. Атрибуты курсоров. Неявные курсоры и их атрибуты.

Неявные курсоры и их атрибуты:

\*Курсоры в PL/SQL используются для обработки результатов запросов к базе данных.

\*Команды для работы с курсорами включают открытие, закрытие, получение данных и перемещение по результатам запроса.

\*Атрибуты курсоров предоставляют информацию о состоянии и результате работы курсора.

\*Неявные курсоры автоматически создаются при выполнении запросов SELECT, а их атрибуты могут быть использованы для доступа к данным.

4. Язык PL/SQL. Пакеты. Схема на запись и схема на чтение.

Пакеты в PL/SQL позволяют группировать связанные процедуры, функции, типы данных и другие объекты в одной единице.

Схема на запись определяет, какие объекты пакета могут изменять данные в базе данных.

Схема на чтение определяет, какие объекты пакета могут только читать данные из базы данных.

Пакеты используются для организации кода, повышения модульности и обеспечения безопасности данных.

5.Язык PL/SQL. Механизм обработки исключений в PL/SQL. Пользовательские исключения.

Механизм обработки исключений в PL/SQL позволяет обрабатывать ошибки и исключительные ситуации в коде.

Пользовательские исключения - это исключения, созданные и определенные пользователем для обработки конкретных ситуаций.

В блоке кода PL/SQL можно определить обработчики исключений, которые выполняют определенные действия при возникновении ошибки.

6. Язык PL/SQL. Триггеры. Виды триггеров. Псевдозаписи и триггерные предикаты.

Триггеры в PL/SQL - это блоки кода, которые автоматически выполняются при определенных событиях, таких как вставка, обновление или удаление данных.

Виды триггеров включают триггеры до и после события, триггеры для строки или столбца, триггеры на уровне таблицы или базы данных.

Псевдозаписи в триггерах представляют старые и новые значения строк, которые вызвали триггер.

Триггерные предикаты - это условия, определяющие, когда триггер должен сработать.

7.Язык PL/SQL. Триггеры. Мутирующие таблицы.

Мутирующие таблицы в PL/SQL - это таблицы, на которые выполняется операция внутри триггера.

При обработке триггера на мутирующую таблицу возникает ошибка мутирования таблицы.

Чтение или модификация данных в той же таблице, на которую наложен триггер, требует специальной обработки, чтобы избежать ошибки мутирования таблицы.

8. Основные понятия и определения интеграции данных. Методы интеграции данных. Схема на запись и схема на чтение.

Схема на чтение определяет, какие объекты и операции позволяют получать доступ к данным в интегрированной системе без возможности их изменения.

В схеме на чтение предоставляются различные представления данных для анализа, отчетности и принятия решений.

Методы интеграции данных могут включать синхронную и асинхронную репликацию, использование ETL-процессов, применение виртуальных таблиц и другие подходы.

Цель интеграции данных - обеспечить консистентность, точность и доступность данных из разных источников для целей анализа и принятия решений.

9. Характеристики OLAP и OLTP систем. Сравнительный анализ OLAP и OLTP. Схема на запись и схема на чтение.

OLAP (Online Analytical Processing) - это технология обработки данных, предназначенная для анализа больших объемов данных и поддержки принятия стратегических решений.

OLTP (Online Transaction Processing) - это технология обработки транзакций в реальном времени, используемая для выполнения операций в рамках бизнес-процессов.

OLAP-системы ориентированы на сложные аналитические запросы и могут работать с большими объемами данных.

OLTP-системы ориентированы на операции ввода-вывода, обновление и получение небольших объемов данных.

Схема на запись (write schema) в OLAP и OLTP системах определяет структуру хранения данных и правила изменения данных.

Схема на чтение (read schema) в OLAP и OLTP системах определяет представления и доступ к данным для аналитических запросов и операций чтения.

10. Понятие строковых и колоночных СУБД. Сравнительный анализ строковых и колоночных СУБД.

Строковая СУБД (строковая база данных) организует данные по строкам, где каждая строка представляет собой запись, содержащую значения всех полей.

Колоночная СУБД (колоночная база данных) организует данные по столбцам, где каждый столбец представляет собой отдельное поле, содержащее значения для всех записей.

Строковые СУБД обычно лучше подходят для операций чтения и записи целых записей, а колоночные СУБД эффективнее при аналитических запросах, где требуется обработка большого объема данных.

Строковые СУБД обеспечивают более быстрый доступ к отдельным записям, а колоночные СУБД - более эффективное сжатие данных и операции агрегации.

11. Методологии построения архитектуры хранилища данных. Модель Инмона.

Модель Инмона (Inmon) предлагает построение хранилища данных с централизованным подходом.

В модели Инмона создается центральное хранилище данных (Enterprise Data Warehouse), которое является источником для всех аналитических и отчетных запросов.

Центральное хранилище данных строится на основе нормализованной модели данных, где данные разделены на отдельные таблицы для обеспечения целостности и удобства обновления.

В модели Инмона акцент делается на интеграции и качестве данных, а также на однозначном источнике правды.

12.Методологии построения архитектуры хранилища данных. Модель Кимбалла.

Каждая звездная схема содержит фактовую таблицу, которая хранит числовые данные, и связанные с ней измерительные таблицы, которые содержат описательные атрибуты.

Модель Кимбалла ориентирована на быстрый доступ к данным и удовлетворение аналитических запросов.

В модели Кимбалла используются понятия измерений, фактов и иерархий для организации данных.

Модель Кимбалла облегчает понимание данных и упрощает процесс анализа и отчетности.

Важным аспектом модели Кимбалла является процесс ETL (извлечение, преобразование и загрузка), который позволяет переносить данные из операционных систем в хранилище данных.

13. Определение транзакции. ACID-свойства. Команды управления транзакциями. Ошибки параллельной работы транзакций.

Транзакция представляет собой логическую единицу работы с базой данных, которая состоит из одной или нескольких операций.

Транзакция должна быть атомарной, целостной, изолированной и устойчивой (обладать свойствами ACID).

ACID-свойства транзакций:

Атомарность: Транзакция считается атомарной, если все ее операции либо выполняются полностью, либо не выполняются вообще.

Целостность: Транзакция должна сохранять целостность данных. После успешного завершения транзакции, база данных должна оставаться в согласованном состоянии.

Изолированность: Каждая транзакция должна работать изолированно от других транзакций, чтобы предотвратить взаимное влияние операций.

Устойчивость: Успешно завершенная транзакция должна сохранять свои изменения в базе данных даже в случае сбоя системы.

Команды управления транзакциями:

COMMIT: Завершает текущую транзакцию и сохраняет все изменения в базе данных.

ROLLBACK: Отменяет текущую транзакцию и возвращает базу данных к состоянию перед началом транзакции.

SAVEPOINT: Устанавливает точку сохранения внутри текущей транзакции, чтобы можно было откатиться к этой точке при необходимости.

Ошибки параллельной работы транзакций:

Потерянное обновление (Lost Update): Когда две или более транзакции пытаются обновить одни и те же данные, и результатом является потеря одного из обновлений.

Грязное чтение (Dirty Read): Когда одна транзакция считывает данные, которые были изменены другой транзакцией, но позже отменены (откатлены).

Неповторяемое чтение (Non-repeatable Read): Когда одна транзакция выполняет повторное чтение данных и видит, что они были изменены другой транзакцией.

Фантомное чтение (Phantom Read): Когда одна транзакция выполняет чтение данных и видит новые записи, которых ранее не было.

14. Основные этапы ETL-процесса. Понятие «плохих данных». Инструменты для визуализации данных.

Трансформация (Transformation): Преобразование и очистка данных, включая фильтрацию, агрегацию, объединение и преобразование форматов данных. Цель этого этапа - обеспечить соответствие данных требованиям и стандартам целевой системы.

Загрузка (Loading): Загрузка преобразованных данных в целевую систему, такую как хранилище данных или целевая база данных. Этот этап может включать создание таблиц, индексов и других структур данных.

Валидация (Validation): Проверка данных на соответствие заданным правилам и бизнес-правилам. Валидация может включать проверку целостности данных, проверку на дубликаты, проверку форматов и т. д.

Обновление (Update): Обновление целевой системы с учетом новых данных, полученных в результате ETL-процесса. Обновление может быть выполнено по расписанию или при наступлении определенных событий.

Понятие "плохих данных":

Под "плохими данными" понимаются данные, которые содержат ошибки, отсутствующую или несогласованную информацию, аномалии или несоответствия форматам и стандартам.

Плохие данные могут быть причиной некорректных аналитических результатов, ошибок в принятии решений и негативного влияния на работу системы.

Инструменты для визуализации данных:

Инструменты для визуализации данных позволяют представить данные в графическом виде, чтобы облегчить понимание и анализ информации.

Некоторые популярные инструменты для визуализации данных включают Tableau, Power BI, QlikView, Google Data Studio и др.

Визуализация данных может включать диаграммы, графики, схемы, интерактивные панели управления и другие элементы, которые помогают пользователю получить представление о данных и выявить закономерности и тренды.