13. Этапы проектирования БД: логическая модель.

Погическим проектированием баз данных формально можно считать процесс создания модели мира самой базы данных, без той системы, которая будет позволять ей работать, и прочих физических деталей. Точность и полнота играют в этом процессе ключевую роль. Одним из главных преимуществ этого этапа является то, что всегда можно взять черновой проект, отложить его в сторону и начать все заново или просто внести желаемые поправки. Гораздо легче менять те или иные детали на этапе проектирования, чем иметь дело с проблемами уже реализованной производственной базы данных, которая плохо спроектирована.

Этап логического проектирования иногда делят на концептуальную и логическую часть, но это отличие чисто номинально. Концептуальное проектирование базы данных обычно предшествует этапу логического проектирования и подразумевает моделирование информации без использования какой-либо базовой модели данных. Что касается этапа логического проектирования, то он уже предполагает явное применение конкретной модели данных, например, реляционной, и фокусирование внимания на логических отношениях, которые были определены на этапе концептуального проектирования. частности, логическое проектирование заключается в концептуальном моделировании базы данных и гарантии того, что данные в таблицах проходят проверку на целостность и не являются избыточными. Для удовлетворения этих требований реализуются принципы нормализации данных.

структурированы, т.к можно установить соответствие между единицами знаний.

 Наличие ситуационных связей. Связи описывают множество текущих ситуаций объекта. Данные трудно поддаются анализу. Из структуры и состава знаний по ситуации возможно построение процедур анализа знаний.

14. Этапы проектирования БД: физическая модель.

Физическое проектирование является третьим и последним этапом создания проекта базы данных. при выполнении которого проектировщик принимает решения о способах реализации разрабатываемой базы данных. Во время предыдущего этапа проектирования была определена логическая структура базы данных (которая описывает отношения и ограничения в рассматриваемой прикладной области). Хотя эта структура не зависит от конкретной целевой СУБД, она создается с учетом выбранной модели хранения данных, например реляционной, сетевой или иерархической. Однако, приступая к физическому проектированию базы данных, прежде всего необходимо выбрать конкретную целевую СУБД. Поэтому физическое проектирование неразрывно связано с конкретной СУБД. Между логическим и физическим проектированием существует постоянная обратная связь, так как решения, принимаемые на этапе физического проектирования с целью повышения производительности системы, способны повлиять на структуру логической модели данных.

Как правило, основной целью физического проектирования базы данных является описание способа физической реализации логического проекта базы данных.

Этапы методологии физического проектирования баз данных:

- Перенос глобальной логической модели данных в среду целевой СУБД.
- . Проектирование основных отношений.
- Разработка способов получения производных данных.
- Реализация ограничений предметной области.
- Проектирование физического представления базы данных.
- Анализ транзакций.
- Выбор файловой структуры.
- Определение индексов.
- Определение требований к дисковой памяти.
- Проектирование пользовательских представлений.
- Разработка механизмов защиты.
- Обоснование необходимости введения контролируемой избыточности.
- Текущий контроль настройка операционной системы

15. Дайте определения понятиям: Знания и База знаний. Назовите отличия данных от знаний.

База знаний — база данных, содержащая правила вывода и информацию о человеческом опыте и знаниях в некоторой предметной области. В самообучающихся системах база знаний также содержит информацию, являющуюся результатом решения предыдущих задач. Знания—факты, сообщения об окружающей среде, процедуры и правила манипулирования фактами, а также информация о том, когда и как следует применять эти процедуры и правила.

В целом, знания — это проверенный практикой результат познания действительности, вид информации, которая отображает знания человек, специалиста в предметной области.

Знания различаются: есть декларативные (факты) и процедурные (правила). Декларативные – это знания об определенных явлениях, событиях, свойствах объектов. Процедурные – это знания о действиях, которые нужно предпринять для достижения какой-либо цели.

Отличия:

- 1) Интерпретация. Хранимые дынные могут быть интерпретированы только человеком или программой. Данные не несут информации. Знания содержат как данные, так и их описание (правила интерпретации).
- 2) Наличие связей классификации. Данные не имеют эффективного описания связей между различными типами данных. Знания

16. Модели представления знаний: Продукционная модель. Привести пример.

Продукционная модель знания — модель, основанная на правилах, позволяет представить знание в виде предложений типа «Если (условие), то (действие)».

Продукционная модель — фрагменты Семантической сети, основанные на временных отношениях между состояниями объектов.

Недостатки: большая база и длительность работы, необходимость выполнять проверки при моделировании БЗ.

Преимущества: удобство модификации Б3, хорошо интерпретировать результат и получить все промежуточные выводы.

ЕСЛИ

«двигатель не заводится»

И

«стартер двигателя не работает»

TO

«неполадки в системе электропитания стартера»

Модели представления знаний: Сеть Фреймов. Привести пример.

Фреймовая модель представляет собой систематизированную психологическую модель памяти человека и его сознания.

Фрейм (англ. frame – рамка, каркас) – структура данных для представления некоторого концептуального объекта. Информация, относящаяся к фрейму, содержится в составляющих его слотах.

Слот (англ. slot – щель, прорезь) может быть терминальным (листом иерархии) или представлять собой фрейм нижнего уровня.



18. Модели представления знаний: Семантическая сеть. Привести пример.

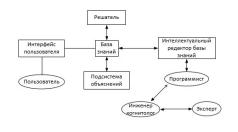
Семантическая сеть — ориентированный граф, вершины которого есть понятия, а дуги отношения между ними.

<u>Преимущества:</u> - простота вывода - наглядность <u>Недостатки:</u> пропадает наглядность при увеличении числа сущностей.



19. Что такое экспертная система. Блоксхема экспертной системы.

Экспертная система — это сложные программные комплексы, аккумулирующие знания специалистов в конкретных предметных областях и тиражирующие эти знания для консультаций менее квал-х пользователей.



20. Определение экспертной системы. Технология разработки экспертной системы (схема этапов и стадии).

Экспертная система — это сложные программные комплексы, аккумулирующие знания специалистов в конкретных предметных областях и тиражирующие эти знания для консультаций менее квал-х пользователей.

Стадии: 1)Демонстрационный прототип

- 2)Исследовательский прототип
- 3)Действующий прототип
- 4)Промышленная ЭС
- 5)Коммерческая ЭС



21. Моделирование: определение, цели, виды.

Модели́рование — замещение исследуемого объекта его условным образом. Основная задача — познание свойств оригинала через исследование модели.

Виды моделей:

- **1) Физические модели.** Реальное воплощение физических свойств оригинала.
- 2)Математические модели. Представляют собой формализованное описание объекта или системы с помощью некоторого языка, например: совокупности математических соотношений и формул.
- 3)Имитационная модель. Сложная система предоставляется в виде алгоритма или логических действий (блоков), в которых также используется аппарат математического моделирования, статистики и теории вероятности.

Цели:

- 1) Изучение объекта-позновательная цели Механизмов, свойств и характеристик
- Управление объектами и системами с целью выработки оптимальных управленческих воздействий и характеристик системы.

22. Моделирование: свойства, этапы разработки модели, погрешности моделирования.

Модели́рование — замещение исследуемого объекта его условным образом. Основная задача познание свойств оригинала через исследование модели.

Свойства моделей: 1) Адекватность – это степень соответствия модели исследуемому объекту.

2) Простота. Чем больше кол-во свойств объекта описывает модель, тем более она сложна. Но не всегда чем более сложна модель, тем более выше ее адекватность. Нужно стремиться найти более простую модель для достижения требуемых результатов.

Этапы разработки: 1)Определение цели.

- 2) Разработка концептуальной модели
- 3)Формализация модели 4) Прог-я реалиазция
- 5)Планирование модельных экспериментов
- 6)Реализация плана экспериментов
- 7) Анализ и интерпретация результатов

Погрешности моделирования:

Ошибка: error = x*x_u

Абсолютная погрешность: $e = |x-x_{ij}|$

Относительная ошибка: $\Delta = e/x_{ij}$

23. Имитационное моделирование: определение, свойства, этапы создания, требования к моделям.

 Имитационное
 моделирование
 —
 метод, описывающие

 позволяющий
 строить модели, описывающие
 описывающие

 процессы
 так, как они проходили бы в действительности.
 в

Свойства: Процесс функционирования сложной системы представляется в виде определенного алгоритма, то есть логических действий, которые и реализуются на компьютере. По результатам реализации могут быть сделаны те или иные выводы относительно исходного процесса. Им. моделирование предоставляет возможность преодолеть аналитические трудности и найти ответ на поставленные вопросы о поведении системы.

Этапы создания им модели:

- 1) Формулируют вопросы о поведении сложной системы.
- 2)Декомпозиция системы 3) Формулируют гипотезы 4)Задание системного времени 5)Задание свойств системы 6)Случайные параметры заменяются машинной реализацией

Требования к моделям: 1) Полнота 2)Гибкость

- 3)Скорость разработки 4)Структура модели
- 5)Информационное обеспечение 6)Требования к ТС и ПО.

24 моделирование случайных процессов с использованием случайных чисел. Структура случайного процесса.

Под статистическим моделированием понимается воспроизведение с помощью ЭВМ функционирования вероятностной модели назвадариготающиетом, чтобы научиться воспроизводить с помощью ЭВМ поведение таких моделей, например:

- с помощью специальных методов и средств вырабатывать программы реализации случайных чисел:
- с помощью этих чисел получать реализацию случайных величин или случайных процессов с более сложными законами распределения;
- с помощью полученных реализации вычислять значения величин, характеризующих модель, и производить обработку результатов экспериментов;

Устанавливать связь алгоритмов моделирования с алгоритмами решения задач вычислительной математики с помощью метода Монте-Карло и строить так называемые «фиктивные» модели, т.е. модели, не имеющие связи с объектом моделирования, но удобные в вычислительном отношении и позволяющие вычислять нужные нам характеристики объекта.

Моделирование случайных процессов строится на основе базовых распределений случайных величин.

Структура случайного процесса: Структуру случайного процесса можно установить по корреляционной

функции или по известной плотности распределения. В зависимости от типа законов распределения можно выделить нормальные, равномерные, релеевские, пуассоновские и другие случайные процессы.

25.Основные характеристики случайных величин.

Математическое ожидание: число, вокруг которого сосредоточены значения случайной величины.

$$M[X] = m_x = \sum_{i=1}^{N} x_i \cdot P_i$$

Р_і=вероятность этого значения

Х_і=возможное значение случайной величины

Дисперсия: определяет рассеяние случайной величины около ее мат. ожидания

$$D[X] = \sum_{i=1}^{N} (x_i - m_x)^2 \cdot P_i$$

Среднеквадратичное отношение:

$$\sigma_x^{\bullet} = \sqrt{D_x^{\bullet}}$$

Коэффициент корреляции: характеризует степень линейной связи между величинами

$$r_{yy}^{*} = \frac{\sum_{i=1}^{n} (x_{i} - m_{x}^{*})(y_{i} - m_{y}^{*})}{(n-1)\sigma_{-}^{*}\sigma_{-}^{*}}$$

26. Генерация псевдослучайных чисел: равномерное распределение, нормальное распределение.

Для метода статистического моделирования на ЭВМ большая доля машинного времени расходуется на действия со случайными числами. Результаты статистического моделирования существенно зависят качества исходных последовательностей случайных чисел. Поэтому наличие простых и способов экономичных формирования последовательностей случайных чисел требуемого качества во многом определяет возможности практического использования машинного моделирования систем.

Равномерное распределение:



Метод серединных квадратов.

- 1.Выбирается число разрядов n (целое)
- 2.Задается 2n разрядное число $x_0=a_1, a_2... a_{2n}$, меньше
- 3. Текущее число возводится в квадрат: $(x_0)^2 = b_1, b_2...$
- 4.Определяется новое значение искомого числа $x_1=b_{n+1}, b_{n+2} \dots b_{3n}$, путем выделения 2n средних разрядов из квадрата исходного числа.
- $5.x_0=x_1$, переход к пункту 3.

Рекурсивный метод.

Основан формуле $Z_{i} = a_{0} \cdot Z_{i-2} + a_{1} \cdot Z_{i-1}$ $x_i = Z_i \mod M$

где a₀, a₁ – целые положительные коэффициенты.

М – целое положительное число, модуль

Z_i – рассчитываемые целые положительные числа

Х_і – рассчитываемые квазиравномерные числа

Нормальное распределение

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



Алгоритм имитации.

Основан на: сумма независимых случайных величин имеет асимптотически гауссовское распределение

$$x = m_x + \sigma_x \cdot \sqrt{\frac{12}{n}} \cdot \left(\sum_{i=1}^n R_i - \frac{n}{2} \right)$$

R_і – случайные квазиравномерные числа

n – положительное число

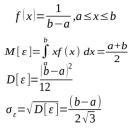
27. Генерация псевдослучайных чисел: равномерное распределение, экспоненциальное распределение.

Для метода статистического моделирования на ЭВМ большая доля машинного времени расходуется на действия со случайными числами. Результаты статистического моделирования существенно зависят последовательностей качества исходных случайных чисел. Поэтому наличие простых и способов экономичных формирования последовательностей случайных чисел требуемого качества во многом определяет возможности практического использования машинного моделирования систем.

Равномерное распределение:

Параметры





Метод серединных квадратов.

- 1.Выбирается число разрядов n (целое)
- 2.Задается 2n разрядное число $x_0=a_1, a_2... a_{2n}$, меньше
- 3. Текущее число возводится в квадрат: $(x_0)^2 = b_1, b_2...$
- 4.Определяется новое значение искомого числа $x_1 = b_{n+1}, b_{n+2} \dots b_{3n},$ путем выделения 2n средних разрядов из квадрата исходного числа
- $5.x_0=x_1$, переход к пункту 3.

Рекурсивный метод.

Основан

формуле

$$Z_i = a_0 \cdot Z_{i-2} + a_1 \cdot Z_{i-1}$$

 $x_i = Z_i \mod M$

где а₀, а₁ – целые положительные коэффициенты.

М – целое положительное число, модуль

- Z_i рассчитываемые целые положительные числа
- Хі рассчитываемые квазиравномерные числа

Экспоненциальное распределение

Алгоритм имитации:

$$x_i = -\frac{1}{\lambda} \ln(1 - R_i)$$

R_і –квазиравномерные случайные числа.

28. Метод случайного поиска и его применение для решения задач.

Случайными или стохастическими методами поиска называют методы, в которых для определения оптимума целевой функции или направления движения к оптимуму сознательно используются случайные числа. Необходимым условием применения данного метода является получение случайных чисел с заданными вероятностными характеристиками. Эта задача далеко не простая. Поэтому на практике почти всегда вместо случайных чисел используются так называемые псевдослучайные (ПСЧ). Доказано, что если ПСЧ удовлетворяют определенным требованиям, то они с достаточной для практики точностью могут заменить случайные числа.

Под поиском понимается процесс отыскания такого значения критерия оптимизации (целевой функции), которое близко к оптимальному и в то же время удовлетворяет всем ограничениям.

Различают 3 вида случайного писка:

- 1.Ненапиравленный случайный поиск.
- 2. Направленный случайный поиск
- 3.Направленный поиск с самообучением (если изменение параметра привело к улучшению, то изменяем параметр в том же направлении)

Алгоритм случайного поиска:

1.Задаем случайную точку х₀, объявляем ее текущей и вычисляем в ней значение целевой функции Q(x₀).

- 2.Текущей точке придаем приращение Δx , вычисляется значение целевой функции $Q' = Q(x_i + \Delta x)$ 3. Если значение целевой функции улучшилось (Q`<Q), то данную точку делаем текущей.
- 4. Проверка условия останова, если выполняется, тогда поиск закончен, иначе п2.

Случайный поиск используется при:

- 1.Отсутствии или не желательном использовании аналитических и численных методов
- возможность создания модели функционирования системы (процесса)
- 3.Наличие большого числа случайных факторов в исследуемой системе(процессе).

29. Задача принятия решения. Постановка задачи, виды задач, типовые задачи.

Постановка задачи. Требуется каждой альтернативе дать приоритет(рейтинг) получить рейтинг альтернатив.

Типовая задача принятия решения:

- 1) Рассматриваются несколько вариантов решения
- 2) Задан критерий, по которому определяется в какой мере то или иное решение является подходящим
- 3) Известны условия, в которых решается проблема и причины, влияющие на выбор того или иного решения.

Виды задач:

- 1) Задача выбора (выбрать несколько вариантов из группы возможных)
- 2) Задача распределения ресурсов (множество вариантов)

Типовые задачи:

- 1) Анализ рисков. (кредиты, страхование)
- 2) Распределение ресурсов
- 3) Планирование от достигнутого

- 4) Планирование (анализ перспективных стратегий развития)
- 5) Анализ эффективность-стоимость(выбор направления деятельности с макс отдачей при мин потерях)
- 6) Поиск существенных факторов (при решении масштабных проблем и проблем стратегического планирования)
- 7) Диагностика
- 8) Построение зависимостей

30. Задача принятия решения. Сопутствующие задачи и проблемы.

Задачи принятия решения

Постановка задачи. Требуется каждой альтернативе дать приоритет(рейтинг) => получить рейтинг альтернатив.

Типовая задача принятия решения:

- 1) Рассматриваются несколько вариантов
- 2) Задан критерий, по которому определяется в какой мере то или иное решение является подходящим
- 3) Известны условия, в которых решается проблема и причины, влияющие на выбор того или иного решения.

Сопутствующие задачи.

- 1)Выявить наиболее неясные и противоречивые этапы создания модели для поддержки принятия решения
- 2)Разбить большую задачу о принятии решения на ряд малых задач.
- 3)Оценить и минимизировать противоречивость использующихся для определения приоритетов рассматриваемых решений (очистка данных)

4)Установить предварительно условия, при которых по найденному рейтингу приоритетов возможных решений можно сделать выбор лучшего совпадения.

- 5)Проверить факторы и варианты решений на избыточность параметров.
- 6)Оценить устойчивость результатов, полученных в результате применения метода.

Сопутствующие проблемы.

- 1)Полный набор решений и факторов неизвестен.
- 2)Большой набор альтернатив и факторов, влияющих на их отбор.
- 3)В реальных задачах влияния различных факторов на выбор оптимального решения сложны и запутаны
- 4)Нет точной количественной информации, необходимой для решения задачи (проблема добычи данных)
- 5)Имеющиеся данные противоречивы
- 6)Нет четкой и универсальной методики составления рейтинга рассматриваемых решений.

31. Методы поддержки принятия решений: информационный поиск. Основные определения, виды, методы поиска и оценки эффективности.

Информационный поиск - процесс поиска неструктурированной документальной информации и наука об этом поиске.

Этапы поиска:

- 1) определение (уточнение) информационной потребности и формулировка информационного запроса;
- 2) установление совокупности возможных держателей информационных массивов (источников);
- 3) извлечение информации из выявленных информационных массивов:
- 4) ознакомление с полученной информацией и оценка результатов поиска.

Различают поиск: полнотекстовый - по всему содержимому документа: по метаданным - по неким атрибутам документа, поддерживаемым системой (название документа, дата создания, размер, автор и т.д.); по изображению - по содержанию изображения.

Информационный ПОИСК осуществляется следующими методами: адресный - процесс поиска документов по чисто формальным признакам, указанным в запросе; семантический - процесс поиска документов содержанию; ПО ИХ документальный - процесс поиска в хранилище информационно-поисковой системы первичных

документов или в базе данных вторичных документов, соответствующих запросу пользователя; фактографический - процесс поиска фактов, соответствующих информационному запросу.

Оценка эффективности.

Точность – отношение числа релевантных документов, найденных ИП, к общему числу найденных документов.

Полнота - отношение числа найденных релевантных документов, к общему числу релевантных документов в базе.

Выпадение – характеризует вероятность нахождения нерелевантного ресурса и определяется, как отношение числа найденных нерелевантных документов к общему числу нерелевантных документов в базе

Mepa (F)

F = 2*(точность) * (полнота) / (точность + полнота).

32. Методы поддержки принятия решений: интеллектуальный анализ данных.

Интеллектуальный анализ данных (datamining)

- это выявление скрытых закономерностей или взаимосвязей между переменными в больших массивах необработанных данных

Такой анализ предусматривает решение задач классификации, моделирования и прогнозирования, кластеризации, сокращения описания, ассоциации. анализа отклонений, визуализации и др.



Цель DM - поиск в больших объемах данных неочевидных, объективных и полезных на практике закономерностей.

Задачи DM – 1)Классификация(отнесение входного объекта, события к одному из заранее известных 2)Кластеризация(разделение входных классов) событий, объектов на группы (кластеры) по степени похожести друг на друга) 3)Сокращение описания(для визуализации данных, лаконизма моделей, упрощения счета, сжатия собираемой и хранимой информации) повторяющихся 5)Прогнозирование

7)Визуализация

Этапы технологии (подготовительные)

- 1)Формирование гипотезы 2)Сбор данных
- 3)Подготовка данных(фильтрация)

Этапы технологии (обучение)

4)Выбор модели 5)Подбор параметров модели и алгоритма обучения 6)Обучение модели(автоматический поиск остальных параметров модели)

4)Ассоциация (поиск

последовательностей)

6)Анализ отклонений

Этапы технологии (анализ результатов)

- 7)Анализ качества обучения, если плохо то п4, п5.
- 8)Анализ выявленных закономерностей знаний, если плохо то п1, п4, п5.

Этапы технологии(анализ результатов)

Проверка адекватности полученных моделей на тестовых данных.

33. Методы поддержки принятия решений: извлечение знаний из баз данных (KDD).

KDD. Извлечение знаний в базах данных - процесс обнаружения полезных знаний в базах данных. Знания в виде: закономерностей, правил прогнозов, связей и

Последовательность операций KDD:

- 1) Консолидация данных процесс их извлечения из различных источников (online-систем, СУБД, файлов, Интернета и т.д.) и загрузка в централизованное хранилище данных.
- 2) Подготовка анализируемых выборок данных (в том числе, обучающих), загрузка их из хранилища данных или других источников в аналитическое приложение.
- 3) Очистка данных от факторов, мешающих их корректному анализу (шум. аномальные значений. дубликаты, противоречия, пропуски).
- 4) Трансформация оптимизация данных для решения определенной задачи. Обычно на данном этапе выполняется исключение незначащих факторов, снижения размерности входных нормализация, обогащение.
- 5) Анализ данных применение методов и технологий Data Mining: построение и обучение моделей (например, нейронных сетей), решение задач классификации, регрессии, кластеризации, прогнозирования и т.д.

6) Интерпретация и визуализация результатов анализа, их применение в бизнес-приложениях.

Характеристика KDD

Knowledge Discovery in Databases не задает набор методов обработки или пригодные для анализа алгоритмы, он определяет последовательность действий, которую необходимо выполнить для того, чтобы из исходных данных получить знания. Данный подход универсальный и не зависит от предметной области.

34. Методы поддержки принятия решений: рассуждения на основе прецедентов. Основные методы извлечения прецедентов.

Прецедент - случай, имевший место ранее и служащий примером или оправданием для последующих случаев подобного рода.

Рассуждение на основе прецедентов. Вывод на основе прецедентов является подходом, позволяющим решить новую задачу, используя или адаптируя решение уже известной задачи.

Преимущества:

- 1) Возможность напрямую использовать опыт, накопленный системой без интенсивного привлечения эксперта в предметной области.
- 2) Возможность сокращения времени поиска решения поставленной задачи за счет использования уже имеющегося решения для подобной задачи.
- 3) Отсутствует необходимость полного рассмотрения знаний о конкретной предметной области.

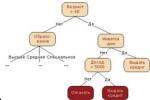
Недостатки:

- 1) При описании прецедентов обычно ограничиваются поверхностными знаниями о предметной области.
- 2) Большое количество прецедентов может привести к снижению производительности системы.
- 3) Выбор критериев для индексации и сравнения прецедентов является сложной задачей.

4) Невозможность получения решения задач, для которых нет прецедентов или степень их сходства меньше заданного порогового значения.

Основные методы извлечения прецедентов:

- 1) Метод ближайшего соседа. С целью определения степени сходства на множестве параметров, используемых для описания прецедентов и текущей ситуации, вводится определенная метрика. Преимущества: простота реализации и универсальность. Недостатки: сложность выбора метрики и требовательность к ресурсам.
- 2) Метод извлечения прецедентов на основе



деревьев решений.

3) Метод извлечения с учетом применимости прецедентов. Сохранения прецедентов вместе с комментариями по их применению позволяет сделать поиск решения более эффективным, заранее отбрасывая часть заведомо неперспективных прецедентов.