# Teoretiska frågor

1. Beskriv kort hur en relationsdatabas fungerar.

En relationdatabas organiserar data i tabeller med rader och kolumner.

varje tabell representerar en specifik typ av information, och relationer mellan tabeller etableras genom användning av nycklar.

SQL används för att hantera och manipulera data, och databasen strävar efter

normalisering för att optimera organisationen och bevara dataintegritet.

1. Vad menas med ”CRUD” flödet?

CRUD står för Create, Read, Update, Delete.

Det är en förkortning som används för att beskriva de grundläggande operationerna som kan utföras på data

i en databas eller annan lagrinsstruktur.

Create betyder att skapa en ny data och lägga till den i databasen.

Read innebär att hämmta eller läsa data från databasen.

Update innebär att ändra eller uppdatera befintlig data i databasen.

Delete vilket är sista steget i CRUD och det innebär att radera eller ta bort data från databasen.

1. Beskriv kort vad en ”left join” och ”inner join” är. Varför använder man det?

bådde inner och left join är typer av SQL- joins som används för att kombinera rader från två eller flera

tabeller baserat på en gemensam kolumn.

en innerjoin returnerar endast de rader där det finns matchande värden i båda tabellerna.

om det inte finns någon matchning i tabellen så exkluderas den raden från resultatet.

leftjoin returnerar alla raden från den första tabellen och matchande rader från den högra tabellen.

Om det inte finns någon matchning från högra tabellen blir resultatet i vänstra tabellen NULL-värden.

1. Beskriv kort vad indexering i SQL innebär.

indexering i SQL är en metod för att förkorta och snabba upp sökningar i en databas

genom att skapa en strukturerad lista över unika värden i en eller flera kolumner. Det gör snabbare åtkomst

till specifika rader och förbättrar sökprestanda. men den borda användas noggrant så att du undviker överflödig

påverkan på lagringsutrymme och prestanda vid datauppdateringar.

1. Beskriv kort vad en vy i SQL är.

en vy i SQL är en virtuell tabell som baseras på resultatet av en SQL- fråga.

Den representerar ett sätt att abstrahera och organisera data från en eller flera tabeller.

användare kan interagera med en vy på samma sätt som med en vanlig tabell trots att den inte lagrar data själv.

1. Beskriv kort vad en lagrad procedur i SQL är.

en lagrad procedur i SQL är en fördefinierad uppsättning SQL- instruktioner som lagras i databasen

och kan återanvändas vid behov. Dessa procedurer utför vanligtvis en specifik uppgift eller sekvens av uppgifter.

fördelarna med lagrad procedurer inkluderar återanvändbarhet, ökad prestanda genom minskat nätverkstrafik och möjlighet

att centralisera databaslogik.

================= Svar till teorifrågor ska ta max 1 sida (men detta är flexibel=================

=================Teorifrågorna ingår inte i den 2-5 sidor och räknas separat==================

==========================Börja rapporten på en ny sida==========================

# 2. Programmeringsuppgift och Rapport

1) AdventureWorks är ett fiktivt företag och en databas som används av Microsoft som en exempeldatabas för inlärnings- och teständamål. AdventureWorks2022 är en exempeldatabas som representerar en fiktiv cykel- och sportbutik vid namn Adventure Works Cycles. Databasen innehåller flera tabeller som omfattar olika aspekter av företagets verksamhet, inklusive anställda, produkter, kunder, försäljning, och produktion. Den används ofta i Microsofts dokumentation, handledningar och utbildningsmaterial för att visa olika funktioner och kapaciteter i SQL Server. I den ingår exempeldata som representerar information för det här fiktiva företaget som heter Adventure Works Cycles. Några av de huvudsakliga tabellerna och informationen som kan ingå i AdventureWorks-databasen är:

* Person:
  + Innehåller information om personer, såsom anställda och kunder. Tabellen kan inkludera namn, adresser, och kontaktuppgifter.
* HumanResources:
  + Kan innehålla information om anställda, inklusive anställningsstatus, löneinformation och relaterade uppgifter.
* Production:
  + Relaterad till produktionsinformation och kan innehålla detaljer om produkter, produktkategorier, och tillverkningsprocesser.
* Sales:
  + Innehåller information om försäljning, kunder, order och relaterade data. Det kan även finnas data om försäljningsrepresentanter och försäljningstransaktioner.

Frågeställningar som kan besvaras med databasen med användning av SQL är exempelvis information om produkter, kunder och försäljning. Här är ett exempel på visualisering av en datasett:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

2)Jag valde att göra ett konfidensintervall av average UnitPrice, alltså medelvärdet av priser per unit. Konfidensintervallet jag fick fram var mellan 460.. och 469. Detta tolkas då som att priset på en vara från tabellen ligger mellan 460…. Och 469….

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

3)Denna kod som användes i denna uppgift avser att sammanställa information om de anställda på detta fiktiva företag. Informationen i de statistiska analyserna, alltså tabellerna ovan, är en grundläggande bas för vidare analyser. Med hjälp av de samlade tabellerna över de anställda kan vi senare ta reda på flera frågeställningar, som exempelvis hur många anställda som jobbar inom varje departement (alltså hur många som har samma jobbtitel), eller hur många olika persontyper finns. Med hjälp av den sammanfattade datan i tabellerna kan vi senare analysera datan och visualisera den. Ett exempel på ett sätt att analysera datan är konfidensintervallet som gjordes i fråga 2. I tabellerna vi får fram kan vi bland annat se de olika jobbtitlarna samt arbetarnas business entity id-nummer. Vi kan också se att de anställdas pay frequency är lika med 2, alltså verkar de ha fått 2 utbetalningar av sin lön än så länge. En rekommendation skulle kunna vara att visualisera datan i tabellerna i form av diagram, för att enklare tolka den. Genom att analysera jobbtitlarna och ID-numren kan man dra slutsatser om företagets hierarki och den övergripande organisatoriska strukturen. Det kan ge insikter om ledningsnivåer, mellanchefer och operativa roller. Antalet löneutbetalningar per anställd kan användas som en indikator på personalomsättning.

1. Executive summary:

Denna Python-kod skapar en anslutning till AdventureWorks2022- databasen via SQLAlchemy och ODBC-driver. En SQL-query kombinerar och filtrerar data från flera tabeller, väljer specifika kolumner och använder Pandas för att skapa och presentera resultaten i en läsbar tabell. En andra del använder Pandas för att utforska och visa data från olika tabeller. Ytterligare en del fokuserar på en specifik tabell (SalesOrderDetail) och visar de första fem raderna. Slutligen beräknas ett 95% konfidensintervall för pris per enhet, vilket ger en statistisk uppskattning av osäkerheten kring medelvärdet baserat på insamlad data från tabellen. I det här exemplet presenteras informationen av ett fiktivt företags anställda personer i en tabell. Detta är andvändbart när vi exempelvis vill se organisatoriska strukturen i företaget eller personalomsättningen. Vidare beräkningar, analyser och visualiseringar av dessa tabeller kan sedan göras. Detta är andvändbart för ett företag som ett sätt att kontrollera sina resurser och implementera nödvändiga förändringar för att öka effektiviteten av arbetet.

1. Muntlig presentation spelades in. Skickades söndag den 7 januari 2024.

Utvärdering:

==Rapporten ska vara 2-5 sidor. Det går bra att skriva mer, speciellt om man använder grafer== =================MEN skriv koncist och fundera på vad du vill lyfta fram================

============================Börja reflektionen på en ny sida============================

=========================Längden av reflektionen är ungefär 1 sida=======================

# Reflektion på eget arbete

1. Utmaningar du haft under arbetet samt hur du hanterat dem.
2. Vilket betyg du anser att du skall ha och varför.
3. Tips du hade ”gett till dig själv” i början av kursen nu när du slutfört den.