

Kunskapskontroll i SQL-DS23

Student: Muhammad Mahmudur Rahman

Kurs: SQL

Kurslärare: Antonio Prgomet och Márk Mészáros

Utbildning: Data Scientist

Yrkeshögskola: EC Utbildning

Datum: 2024-01-05

Teoretiska Frågor

1. Beskriv kort hur en relationsdatabas fungerar.
2. Vad menas med "CRUD" flödet?
3. Beskriv kort vad en "left join" och "inner join" är. Varför använder man det?
4. Beskriv kort vad indexering i SQL innebär.
5. Beskriv kort vad en vy i SQL är. 6. Beskriv kort vad en lagrad procedur i SQL är.
6. Beskriv kort vad en lagrad procedur i SQL är.

1. Beskriv kort hur en relationsdatabas fungerar.

Svar:

En relationsdatabas är en samling information som organiserar data i fördefinierade relationer där data lagras i en eller flera tabeller eller relationer av kolumner och rader, vilket gör det enkelt att se och förstå hur olika datastrukturer relaterar till varandra. Tabellerna är sammankopplade med unika ID:n eller nycklar. När en användare behöver komma åt specifik information kan de använda en nyckel för att komma åt alla tabeller med data som har förutbestämts att vara relaterade till den nyckeln.

Anta att jag jobbar på ett företag som säljer cyklar online. Min organisation använder en relationsdatabas för att hantera data relaterad till frakt, kundinformation, inventering och webbplatstrafik. Jag har en nyckel till den här databasen som kommer åt alla tabeller relaterade till frakt, och jag måste ta reda på om du har tillräckligt med lager för att skicka ut förra veckans beställningar.

Eftersom relationsdatabasen känner igen att det finns ett samband mellan fraktinformation och lager, kan jag använda min nyckel för att komma åt lagernummer och fraktförfrågningar för att jämföra data. Under denna begäran behöver jag inte komma åt någon information om webbplatstrafik eftersom min nyckel endast kommer åt tabellerna med data som är relaterade till frakt.

2. Vad menas med "CRUD" flödet?

Svar:

CRUD är ett tillvägagångssätt i mjukvaruapplikationer som beskriver fyra grundläggande funktioner: Create, Read, Update, Delete. Alla operationer i relationsdatabasapplikationer översatta till en vanlig HTTP-metod, SQL-sats eller DDS-funktion kan representeras av vilken bokstav som helst i akronymen. De fyra CRUD-operationerna kan utföra olika aktiviteter på specifika data som valts i databasen.

CREATE

CREATE-operationen lägger till en ny post i en databas. I RDBMS hänvisas till en databastabellrad som en post, medan kolumner kallas attribut eller fält. CREATE-operationen lägger till en eller flera nya poster med distinkta fältvärden i en tabell.

READ

READ returnerar poster från en databastabell baserat på vissa sökkriterier. READ-operationen kan returnera alla poster och några eller alla fält.

UPDATE

UPDATE används för att ändra befintliga poster i databasen. Det kan till exempel vara adressändring i en kunddatabas eller prisändring i en produkt-databas.

DELETE

DELETE-operationer tillåter användaren att ta bort poster från databasen. En hård radering tar bort posten helt, medan en mjuk radering flaggar posten men lämnar den på plats, till exempel, i löneadministrationen där anställningsregister måste föras även efter att en anställd har lämnat företaget.

3. Beskriv kort vad en "left join" och "inner join" är. Varför använder man det?

Svar:

LEFT JOIN

Kommandot LEFT JOIN returnerar alla rader från den vänstra tabellen (tabell 1) och de matchande raderna från den högra tabellen (tabell 2). Resultatet är NULL från höger sida, om det inte finns någon matchning.

LEFT JOIN används när vi behöver alla poster från "vänster tabell", oavsett om de har par i "höger tabell" eller inte.

INNER JOIN

INNER JOIN kombinerar poster från två tabeller när det finns matchande värden i ett fält som är gemensamt för båda tabellerna.

INNER JOIN används när vi endast vill returnera poster med par på båda sidor.

4. Beskriv kort vad indexering i SQL innebär.

Svar:

SQL-index är speciella uppslagstabeller som används för att påskynda processen för datahämtning. De håller pekare som hänvisar till data som lagras i en databas, vilket gör det lättare att hitta de nödvändiga dataposterna i en databastabell. För att förstå

bättre kan vi säga att SQL-index fungerar på samma sätt som indexet för en bok eller en journal.

5. Beskriv kort vad en vy i SQL är. 6. Beskriv kort vad en lagrad procedur i SQL är.

Svar:

I SQL är en vy en virtuell tabell baserad på resultatuppsättningen av en SQL-sats.

En vy innehåller rader och kolumner, precis som en riktig tabell. Fälten i en vy är fält från en eller flera riktiga tabeller i databasen.

Vi kan lägga till SQL-satser och funktioner i en vy och presentera data som om data kom från en enda tabell.

En vy skapas med CREATE VIEW-satsen.

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

Svar:

En lagrad procedur är en förberedd SQL-kod som vi kan spara, så koden kan återanvändas om och om igen.

Så om vi har en SQL-fråga som skrivs om och om igen, sparar vi den som en lagrad procedur och anropar den för att köra den.

Vi kan också skicka parametrar till en lagrad procedur, så att den lagrade proceduren kan agera utifrån de parametervärden som skickas.

Rapport

När du är klar med koden skall du göra en skriftlig rapport som innehåller följande:

1. En presentation på fråga 1 ovan där du gjorde "en deskriptiv sammanfattning över databasen AdventureWorks2022".
2. Presentera resultatet av din statistiska analys och göra ett konfidensintervall för något som du själv bestämmer. Hur tolkar du resultaten?
3. Analysera datan och resultaten du tagit fram. Vilka slutsatser och rekommendationer kan du ge?
4. "Executive summary". Cirka 10 meningar som sammanfattar det viktigaste (detta brukar t.ex. chefer vara intresserade av).
5. Skriv vilket datum du gjorde din muntliga presentation av koden.

1. En presentation på fråga 1 ovan där du gjorde "en deskriptiv sammanfattning över databasen AdventureWorks2022".

Svar:

Adventure Works Cycles, det fiktiva företaget som AdventureWorks exempeldatabaser bygger på, är ett stort, multinationellt tillverkningsföretag. Företaget tillverkar och säljer metall- och kompositcyklar till de nordamerikanska, europeiska och asiatiska kommersiella marknaderna.

AdventureWorks 2022 Sample Database innehåller information om en fiktiv cykeltillverkare, Adventure Works Cycles, och omfattar olika aspekter av dess verksamhet, inklusive försäljning, produkter, kunder och mer.

Nyckeltabeller och datakategorier

Faktatabeller

Internetförsäljningsdata: Databasen innehåller försäljningsrelaterad information, inklusive order, orderdetaljer och försäljningsområden.

Budget: Data om budgetar och ekonomiskt resultat.

Måtttabeller

Produktdata: Information om produkter, kategorier och beskrivningar.

Kunddata: Kundinformation, inklusive namn, adresser och kontaktinformation.

Kalenderdata: Information om beskrivningar av år, kvartal, månad och dag.

Innan jag påbörjade analysen av Database AdventureWorks2022 behövde det gå igenom några grundläggande steg för att ansluta databasen i pandas dataframe som installerade SQLAlchemy, importerade nödvändiga bibliotek, skapade engine och kopplade ihop engine med databasen.

Efter att ha kopplat till databasen kollade jag först översikten över databasen som till exempel inspektera databasen, hur många scheman det finns i databasen, tabeller under varje schema så att jag kan göra en plan för vad jag ska göra.

Efter att ha gjort de mycket grundläggande stegen i databasen började jag analysera databasen. Det analyserades flera scheman och tabeller för att få en helhetsbild av databaserna så att en sammanfattning kan göras och lämna rapporten till chefen för företaget där jag nystartat mitt jobb som dataanalytiker.

Jag analyserade främst Försäljningsschema och tabeller under Försäljningsschemat. Förutom försäljningsscheman analyserade jag även produktion, inköp, personal, säljare och personscheman.

Eftersom jag är ny i detta yrke var det ganska intressant och spännande att analysera databasen och resultaten var också tillfredsställande.

2. Statistisk analys

Ett konfidensintervall visar sannolikheten att en parameter hamnar mellan ett par värden runt medelvärdet. Här använder jag 95% konfidensintervall som definierar ett värdeintervall som kan användas 95% säkert innehåller populationsmedelvärdet.

Här har jag använt formeln $\bar{x} \pm t^*(s/\sqrt{n})$ där \bar{x} =Samplemedelvärde, t =Önskad konfidensnivå och frihetsgrader $(n-1)$, s =Samplestandardavvikelse och n =Samplestorlek .

Från min beräkning fann jag,

Genomsnittligt listpris, $\bar{x} = 1586.737$

Exempel på standardavvikelse, $s = 996,268721$

Provstorlek, $n = 97$

Frihetsgrader, $n-1 = 97-1 = 96$

I boken Tabeller och Formeler för Statistiska Beräkningar, från Tabell 4. t-fördelningen, finns värde för fördelningen för 96 frihetsgrader som är $t=1,980$

Så, konfidensintervall för genomsnittligt listpris = $1586.737 \pm 1.980*(996.268.721 / \sqrt{97}) = 1586.737 \pm 200.288$ där Övregräns = 1787.025 och Nedregräns = 1386.449

Det bedöms att genomsnittligt listpris ligger i intervallet 1787.025 till 1386.449 prisenhet.

3. Analysera databasen och fynden

Nordamerika, Europa och Stillahavsområdet sålde mest respektive enligt YTD och förra årets försäljning, medan Kanada, Australien och Storbritannien sålde mest efter länder.

Försäljningsanalysen:

Sales schema	Findings
Total order	274914
Average order quantity	2.2660797744751355
Total Unit Price	56423747.6147
Average unit price	465.0934956741429
Subtotal	1.098464e+08
Percentage of tax	2.582642e+05
Average Sales Per Person	223
Total freight paid	3183430.2518
Number of customers	19820
Number of discount given	16
Discount given in percentage	3.5199999999999996
Number of sales store	701

Analys av produktionsschema:

Production schema	Findings
AverageListPrice	437.4042
Total number of products	504
Four kinds of products	Bikes, Components, Clothing and Accessories
Standard cost of products	171535.0024
Average cost of products	434.2658288607595

Analys av inköpsschema:

Purchasing schema	Findings
Total number of vendors	104

Analys av personalschema:

Human Resources schema	Findings
Number of departments	16
Number of job titles	67
Number of employees	290
Total salary per week	210983.776

Analys av säljare schema:

Salesperson schema	Findings
Total sale YTD	36277591.903400004
Total sale last year	23685963.624399997
Total bonus	48610.0

Top selling products:

Product Category	Total Quantity Sold	Total Sales
Clothing	18627	479780.78
Bikes	7526	4601049.01
Accessories	6586	143811.43
Components	1080	198050.52

Top 10 vendors:

Vendor Name	Total Amount Ordered
Superior Bicycles	5.034267e+06
Professional Athletic Consultants	3.379946e+06
Chicago City Saddles	3.347165e+06
Jackson Authority	2.821334e+06
Vision Cycles, Inc.	2.777685e+06
Sport Fan Co.	2.675889e+06
Proseware, Inc.	2.593901e+06
Greenwood Athletic Company	2.472770e+06
Crowley Sport	2.472770e+06
Mitchell Sports	2.424284e+06

Top 10 cities by customer:

Cities	Customer
London	420
Paris	386
Burien	212
Harmoni	212
Bellingham	210
Beaverton	210
Chula Vista	206
Berkeley	200
Burlingame	198
Bellflower	194

Top 10 cities by revenue:

Cities	Revenue
Toronto	5.034270e+06
Paris	3.071177e+06
Seattle	2.233808e+06
Burnaby	1.767308e+06
Richmond	1.503464e+06
Montreal	1.419264e+06
Garland	1.274667e+06
Memphis	1.267787e+06
Nashua	1.246122e+06
Bellflower	1.244107e+06

4. Executive summary:

Efter att ha analyserat databasen AdventureWorks2022 kan jag komma fram till en sammanfattning av vilken typ av produkt företaget säljer, försäljning YTD och försäljning förra året, i vilken del av världens produkter säljs, avdelning och anställda på företaget, säljare, kunder och leverantörer.

Totala värden:

Totalt antal produkter: 504
Totalt antal leverantörer: 104
Total beställning: 274914
Totalt enhetspris: 56423747.6147
Totalt antal kunder: 19820
Total försäljning butiker: 701
Totalt antal anställda: 290
Totallön per vecka: 210983.776
Total försäljning YTD: 36277591.903400004
Total försäljning förra året: 23685963.624399997

Genomsnittliga värden:

Genomsnittlig orderkvantitet: 2,2660797744751355
Genomsnittligt enhetspris: 465,0934956741429
Genomsnittlig försäljning per person: 223
Genomsnittligt listpris: 437,4042
Genomsnittlig kostnad för produkter: 434,2658288607595

Toppvärden:

Topleverantör: Superior Cyklar har beställt belopp 5.034267e+06
Topp stad efter kund: London med 420 kunder
Topp stad efter intäkter: Toronto med 5,034270e+06 intäkt

Statistisk analys gjordes genomsnittligt listpris = 1586.737 med 95% konfidensintervall och bedömningen är att genomsnittligt listpris ligger i intervallet 1787.025 till 1386.449 prishet.

Efter att ha analyserat databasen AdventureWorks2022 kommer mina resultat som är totala, genomsnittliga och toppvärden och dessutom statistisk analys för genomsnittligt listpris att ge chefen en **helhetsbild av databasen** och hjälpa honom för **vidare analys, planering och beslutsprocess**.

5. Muntlig presentation av koden: Den 4 december, torsdag kl. 11:40-12:00 i Microsoft Teams.

Efter att du är klar skall du skriva en kort redogörelse i slutet av rapporten för:

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.

1. Svar:

När jag fick utmaningar under tiden jag gjorde mitt projekt försöker jag alltid lösa detta själv. För när ett problem uppstår gillar jag att lösa själv och om jag kan lösa det själv gör det mig extra självförtroende som inte bara är viktigt för studierna utan också viktigt för livet. Om jag misslyckas med att lösa själv, frågar jag alltid mina lärare eller diskuterar med mina klasskamrater.

2. Svar:

Jag hoppas starkt på att få "Väl godkänd (VG)" betyget baserat på följande faktorer: Jag började denna kurs SQL med noll kunskap om kodning. När jag började kursen, det kändes som att jag ramlade ner mitt i havet och jag kunde inte simma och jag ville överleva! Att delta i alla klasslektioner, gruppdiskussioner, hårt arbete (studera 40 timmar/vecka), prata med kurslärare och klasskamrater, titta på YouTube-handledningar, gå igenom onlineresurser som är LinkedIn Learning, DataCamp, Udemy, PluralSight hjälpte mig mycket! Således lärde jag mig simma och nu har jag kommit till stranden av havet från mitten av havet genom att simma! Det känns att ingenting är omöjligt! Nu tycker jag om SQL-kodning. Trivs är väldigt viktigt som våra duktiga kurslärare kunde ge oss, de gjorde kursen lätt och intressant för oss :)

3. Svar:

I början av kursen sa jag till mig att gå på alla klassföreläsningar, gruppdiskussioner, hårt arbete (40 timmar/vecka) och träna mycket (kodning) som hjälpte mig mycket och nu känner jag att jag gjorde rätt och jag kommer att fortsätta denna strategi för kommande kurser.