

تمرین اول درس اصول طراحی پایگاه داده

سوال اول

- A. مفهوم Data Integrity در پایگاه داده به معنی وجود دقت بالا، ساختارمند بودن و validity در دیتا و طراحی است. چهار نوع Data Integrity داریم:
- a. در Entity Integrity به معنای این است که دیتا نباید null و یا یکسان باشند که این امر به کمک primary key محقق میشود. برای مثال برای پایگاه داده کتاب دو ستون primary key و book_id داشته باشیم.
 - b. در Referential Integrity استفاده از foreign key در جدول دوم و متصل کردن به primary key در جدول اول کمک میکند در پایگاه داده کتاب ها، کتاب ها میتوانند از یک نویسنده و از یک ژانر باشند.
 - c. دو نوع دیگر از Data Integrity هم داریم: User-Defined Integrity و Domain Integrity.
- همین امر در system files هم وجود دارد به طوری که دیتا باید ساختارمند باشد تا در بهینه ترین زمان بتوان به تمام دیتا دسترسی داشت و دیتا غلط وارد سیستم نشود (بخش validation)
- به طور کلی داشتن Data Isolation در DBSM بهتر و کاهش Data Redundancy رو DBSM راحت تر است.
- B. دیتا structured در جدول می آید، مانند مشخصات دانشجویان (اسم، فامیلی، شماره تماس و شماره دانشجویی) در Excel. دیتا semi-structured را نمیتوان به فرم جدول نوشت اما یک ساختار خودش را دارد مانند HTML یا JSON. دیتا unstructured ساختار مشخصی ندارد مانند متن که نمیتوان به فرم مشخصی تبدیل کرد.
- C. این کلید واژه زمانی استفاده میشود که میخواهیم در صورت پاک شدن والد (parent)، فرزند (child) نیز از پایگاه داده پاک شود. این زمانی است که فرزند تماما به والد وابسته است و بدون والد خود معنی ندارد. با این کار حجم پایگاه داده به مرور زمان افزایش پیدا نمیکند زیرا دیتا بدون استفاده (redundant) اتوماتیک پاک میشوند. برای مثال پایگاه داده ساختمان ها و اتاق های آنها. در صورت حذف یک ساختمان نیازی به اتاق ها نیست.
- D. پنج قسمت اصلی DBMS عبارتند از: hardware, software, procedure, DAL یا همان Data Access Language زبانی است که به وسیله آن رکورد های پایگاه داده را تغییر میدهیم (برای مثال آپدیت یا پاک کردن) زبانی که برای این کار استفاده میشود SQL است که این امکان را به ما میدهد تا دیتابیس و جدول بسازیم، در آنها تغییرات بدیم و به آنها دیتا اضافه کنیم.
- E. سه ستون اصلی در اینستاگرام، جداول users, comments, posts, followings هستند. جدول comments به posts متصل است چرا که هر پست میتواند تعدادی کامنت داشته باشد (رابطه چند به یک). جدول followings به users متصل است چرا که کاربر ها میتوانند یکدیگر را فالو کنند (رابطه چند به چند). جدول users به posts متصل است چرا که هر کاربر میتواند تعدادی پست بگذارد (رابطه یک به چند).

سوال دوم

A.

a. $Id_diff \leftarrow \Pi_{user_id}(user) - \Pi_{user_id}(purchases)$ b. $b \leftarrow \Pi_{name}(\sigma_{user_id \text{ in } Id_diff}(user))$

Explanation: (a) Get IDs that exist but have no purchases, then (b) get names from IDs that exist in *Id_diff* column

B.

a. $males \leftarrow \Pi_{user_id}(\sigma_{gender=Male}(user))$ b. $females \leftarrow \Pi_{user_id}(\sigma_{gender=Female}(user))$ c. $suites \leftarrow \Pi_{product_id}(\sigma_{category="suit"}(product))$ d. $shoes \leftarrow \Pi_{product_id}(\sigma_{category="shoe"}(product))$ e. $male_suits \leftarrow \Pi_{user_id}(males) \cap \Pi_{user_id}(suites)$ f. $female_shoes \leftarrow \Pi_{user_id}(females) \cap \Pi_{user_id}(shoes)$ g. $all_intersected_users \leftarrow \Pi_{user_id}(females) \cup \Pi_{user_id}(shoes)$

h. Explanation: For easier understanding, all conditions are splitted and then grouped for the final query.

C.

a. $online_purchases_id \leftarrow \Pi_{product_id}(\sigma_{is_paid_online=True}(purchases))$ b. $prices \leftarrow \Pi_{price}(\sigma_{product_id \text{ in } \Pi_{product_id}(online_purchases_id)}(products))$ c. $\Pi_{avg(price)}(prices)$

Explanation: (a) Get all product IDs where payment method is online. (b) then get the process accordingly and (c) calculate the average.

D.

a. $offline_purchases_id \leftarrow \Pi_{product_id}(\sigma_{is_paid_online=False}(purchases))$ b. $names \leftarrow$
 $\Pi_{name}(\sigma_{product_id \text{ in } \Pi_{product_id}(offline_purchases_id) \text{ and created_at} > 1390}(products))$
c. $\Pi_{name}(names)$

Explanation: (a) get IDs for offline payments and (b) get the product names from the offline_purchases_id into names. (c) Then query.

توجه شود که خروجی ها جدول در نظر گرفته شده اند نه لیست، پس برای select های نهایی هم جدول فرض شده اند
(D.c or C.c)

سوال سوم

کوئری های مربوطه در فایل 3.sql ضمیمه شده داخل فایل زیپ قرار داده شده است. تمام کد ها با دیتای Mock ران گرفته شده و تست شده اند.

تمرین اول درس اصول طراحی پایگاه داده

سوال چهارم

کوئری های مربوطه در فایل 4.sql ضمیمه شده داخل فایل زیپ قرار داده شده است. تمام کد ها با دیتای Mock ران گرفته شده و تست شده اند.