

تمرین دوم

هدف:

در این تمرین قصد داریم با روند طراحی شمای پایگاه داده‌های رابطه‌ای و چگونگی تعامل با آن‌ها از طریق زبان SQL بیشتر آشنا شویم. بدین منظور از بروزترین پایگاه داده آرشیو شده یکی از پر ترافیک ترین وب سایت‌های دنیا یعنی Stack Overflow استفاده می‌کنیم. خوشبختانه این پایگاه داده از طریق رابط وب نیز قابل دسترسی است و دوستان می‌باید پرس و جوهای SQL خود را از رابط وب این سیستم روی پایگاه داده اعمال کنند و خروجی را برای ما بفرستند.

علاوه بر این، یک سری سوالات تئوری در مورد این درس مطرح می‌شود که بسیار ساده و کوتاه جواب‌اند و از جهت یادآوری مطرح میشوند.

چيست Stack Overflow؟

این سایت محلی برای پرسش و پاسخ برای برنامه‌نویسان است. افراد می‌توانند به صورت رایگان در این سایت عضو شوند و به پرسیدن و یا پاسخ دادن به سوالات مشغول شوند. برای اطلاعات بیشتر در مورد عملکرد این وبسایت به آدرس زیر مراجعه کنید.

<http://stackoverflow.com/tour>

پایگاه داده Stack Overflow

همان‌طور که در قسمت قبلی با هدف و عملکرد این وب سایت آشنا شدیم. بدیهی است که پشت این وب سایت باید یک پایگاه داده قدرتمند باشد. این پایگاه داده باید اطلاعاتی نظیر اطلاعات کاربران، اطلاعات پست‌های ارسالی (سوال‌ها و جواب‌ها) و سایر اطلاعات را در خود نگهداری کند. این پایگاه داده از نوع رابطه‌ای است و برای موجودیت‌های سیستم جدول‌هایی در نظر گرفته است به طوری که در حدود ۳۰ جدول به هم وابسته در خود جا داده است. در این تمرین برای سادگی فقط پرس و جوها را به چند جدول اصلی محدود می‌کنیم.

رابط وب این پایگاه داده چیست و چگونه باید از آن استفاده کرد؟

برای دسترسی به رابط وب این پایگاه داده دوستان می‌باید به آدرس زیر مراجعه کرده و از آن قسمت روی گزینه Stack Overflow کلیک کنند.

<https://data.stackexchange.com/>

پس از کلیک کردن، شما به صفحه‌ای مشابه تصویر صفحه بعد (شکل ۱) هدایت میشوید. در این صفحه در قسمت اصلی لیستی از پرس و جو‌هایی که قبلاً کاربران روی پایگاه داده اعمال کرده‌اند و ذخیره کرده‌اند را می‌بینید. شما نیز میتوانید پرس و جو‌های جالب خودتان را ذخیره کنید و در اختیار سایر کاربران قرار دهید.

برای نوشتن یک پرس و جو جدید هم می‌باید روی دکمه Compose Query در قسمت بالا سمت راست صفحه کلیک کنید. پس از کلیک کردن به صفحه‌ای مشابه تصویر ۲ هدایت می‌شوید. در این صفحه می‌توانید پرس و جوهای خود را به زبان SQL نوشته و روی دکمه Run Query کلیک کنید تا آن را روی پایگاه داده اجرا کنید. نتیجه پرس و جو هم در قسمت زیرین صفحه به نمایش در می‌آید. برای نمونه، من یک پرس و جو که تعداد کاربران با اعتبار بیش از 10k را نشان می‌دهد، نوشته و اجرا کرده‌ام که نتیجه آن عدد ۹۱۸۹ است. در سمت راست این صفحه هم شمای جداول این پایگاه داده قابل دسترسی است.

All Queries

featured recent favorite popular everything

☆ 99k
255 views

How many upvotes do I have for each tag?

sep 29 11 sam.saffron

37
queries

☆ 91k
184 views

How Unsung am I?

oct 4 11 Eric

stackoverflow
Q&A for programmers

☆ 70k
153 views

My Comment Score distribution

dec 22 11 sam.saffron

11m
questions
18m
answers

☆ 78k
121 views

What is my accepted answer percentage rate

oct 3 11 sam.saffron

47m
comments

شکل ۱: پرس و جویهای ارسالی از سوی کاربران روی پایگاه داده Stack Overflow

Enter a title for your query

stackoverflow
Q&A for programmers

edit description

```
1 SELECT COUNT([Id])
2 FROM [Users]
3 WHERE [reputation] > 10000
```

Database Schema	
Posts	
Users	
Comments	
Badges	
CloseAsOffTopicReasonTypes	i
CloseReasonTypes	i
FlagTypes	i
Revisions	
593737	
593720	
593719	

permalink

hide sidebar >>

Run Query Cancel Options: Text-only results Include execution plan

Switch sites: | search by name or url

Results	Messages
9189	

[Download CSV](#)

شکل ۲: صفحه مربوط به نوشتن پرس و جوی جدید

شمای پایگاه داده

برای دسترسی به اطلاعات کاملتری در مورد شمای این پایگاه داده، دوستان می توانند به آدرس زیر مراجعه کنند.

<http://meta.stackexchange.com/questions/2677/database-schema-documentation-for-the-public-data-dump-and-sede>

https://en.wikipedia.org/wiki/Academic_honor_code

دانشجویان عزیز می باید بیان زیر را به صورت دست خط نوشته و سپس آن را امضاء کنند. تمرین هایی که به این عبارت متعهد نیستند تصحیح نخواهد شد.

«من متعهد می شوم که در هنگام پاسخ به سوالات این تمرین فقط از یافته ها و تلاش های خودم استفاده کنم. و از دیدگاه اخلاقی رونویسی از تلاش دیگران برایم مذموم است.» - نام و نام خانوادگی. امضاء

سوالات بخش DDL

۱. فرض کنید همه جدول های سیستم طراحی شده است و از ما خواسته شده است تا دستور SQL مربوط به ایجاد جدول Comments را بنویسیم به طوری که موارد زیر هم لحاظ شده باشند. شما چگونه این کار را خواهید کرد؟ (برای دسترسی به ستون های مورد نیاز این جدول به شمای پایگاه داده مراجعه کنید)

- ستون score مقادیرش همواره بزرگ تر یا مساوی صفر است.
- کلیدهای خارجی به درستی اعمال شوند.

۲. اکنون ما جدول Comments را ایجاد کرده ایم. فرض کنید از ما خواسته شود تا یک ستون دیگر به این جدول اضافه کنیم که مخاطب یک کامنت را مشخص می کند. برای این کار باید یک ستون به نام CalleeUserId به جدول اضافه کنیم که ID کاربر پذیرنده کامنت را مشخص میکند. شما چگونه این کار را انجام خواهید داد؟ (البته توجه داشته باشید که همه کامنت ها ممکن است مخاطب نداشته باشند.)

۳. همان طور که می دانیم یک سوال می تواند چندین برچسب داشته باشد. با توجه به شمای ارائه شده از پایگاه داده، موجودیت برچسب، یک موجودیت ضعیف است یا قوی؟ نمودار ER ای رسم کنید که فقط مربوط به ارتباط سوالات و برچسب هایشان باشند. (برای سادگی در رسم نمودار، سایر ارجاعات را در نظر نگیرید.)

۱.۳. در عمل، طراحان این پایگاه داده از یک ستون از نوع رشته ای به نام Tags در جدول Posts استفاده کرده اند که برچسب های یک پست را به صورت رشته ای و به هم متصل شده در خود نگهداری میکند. به نظر شما معایب و مزایای این روش چیست؟

۲.۳. علاوه بر داشتن ستون Tags در جدول Posts یک جدول دیگر به نام PostTags وجود دارد که ارتباط بین پست ها و برچسب ها را مشخص میکند. به نظر شما علاوه بر داشتن ستون Tags در جدول Posts، دیگر چه نیازی به جدول PostTags است؟

۴. با توجه به اصول طراحی منطقی پایگاه داده ها و آنچه تاکنون از طراحی شمای یکی پایگاه داده آموختید، آیا شمای پایگاه داده Stack Overflow منطقی به نظر میرسد؟ آیا همه اصولی که تاکنون آموختید در آن برقرار است؟ چه اصولی نقض شده است؟ برای پاسخ به این سوال نکات موجود در آدرس زیر را نیز هم در نظر بگیرید.

<http://stackexchange.com/performance>

۵. همانطور که از شمای پایگاه داده مشخص است، جدول Posts ستون هایی دارد که کلید خارجی به خودش است. برای نمونه ستون های AcceptedAnswerId و ParentId به کلید اصلی همان جدول یعنی Posts اشاره دارند. به هنگام ساخت این جدول اگر در SQL مربوطه مشخص کنیم که ستون ParentId کلید خارجی به ستون Id جدول Posts است، به دلیل اینکه هنوز جدول Posts ساخته نشده است، ممکن است از سوی DBMS خطایی مبنی بر عدم حضور جدول Posts ظاهر شود. برای حل این مشکل چه راهکاری می اندیشید؟

۶. اگر در هنگام تعریف جدول ها برای حذف کردن سیاست Cascade انتخاب شده باشد. آنگاه به نظر شما اگر بخواییم جدول PostTags را حذف کنیم کدام جدول های دیگر نیز حذف می شوند؟

۱ پرس و جوهای این قسمت را قادر نخواهید روی پایگاه داده اعمال کنید. زیرا این کپی از پایگاه داده فقط خواندنی است.

سوالات بخش DML

۱. موارد زیر را به صورت پرس و جوی SQL بنویسید و روی پایگاه اجرا کرده و خروجی را به همراه تاریخ اجرا برای پرس و جوی و نتیجه ضمیمه پاسخ نامه تان کنید. (در موارد زیر خروجی هر پرس و جوی تنها یک عدد است.)

- ◆ تعداد کامنت هایی که در زیر سوالات قرار داده شده است. (نکته: سوال گونه خاصی از پست است)
- ◆ شناسه کاربری که بیشترین تعداد کامنت را گذاشته است.
- ◆ میانگین تعداد جواب ها برای یک سوال
- ◆ تعداد کاربرانی که بیش از ۱۰۰ پاسخ ارسال کرده اند..
- ◆ طول بزرگترین زنجیره ایجاد شده در جدول PostLinks که در طول زنجیره نوع لینک ها یکسان هستند. (آیا نوشتن این پرس و جوی امکان پذیر است؟)

۲. موارد زیر را به صورت پرس و جوی SQL بنویسید و روی پایگاه اجرا کرده و چند سطر ابتدایی آن را به همراه تاریخ اجرا برای پرس و جوی نیز ضمیمه پاسخ نامه تان کنید. (در موارد زیر خروجی هر پرس و جوی شامل چندین ستون است.)

- ◆ شناسه سوالاتی که بیش از ۵ پاسخ دارند به همراه میانگین امتیاز پاسخ هایشان (جدول خروجی دو ستون به نام های شناسه سوال و میانگین امتیاز خواهد داشت.)
- ◆ لیستی از شناسه کاربران و میزان اعتبار آنها به طوری که آنها در سیستم هیچ پرسشی مطرح نکرده اند و از طرف دیگر بیش از ۱۰ پاسخ ارسال کرده اند. این لیست باید بر حسب میزان اعتبار به صورت نزولی مرتب شده باشد.
- ◆ اگر برای هر فرد یک معیار «کیفیت پاسخگویی» در نظر بگیریم که به صورت میانگین امتیاز پاسخ هایش تعریف می شود، آنگاه لیستی از شناسه کاربری با کیفیت ترین پاسخ دهندگان تهیه کنید. (توجه کنید که این معیار فقط برای کسانی تعریف می شود که حداقل یک پاسخ در سیستم ارسال کرده اند.)
- ◆ لیستی شامل سه ستون تهیه کنید که شامل شناسه کاربری، تعداد نشان های طلایش و تعداد پست های ارسالی اش باشد.

۳. با توجه به ساختار کلیدهای اصلی و خارجی جدول Posts، فرض کنید می خواهیم یک پرسش به همراه یک پاسخ پذیرفته شده اش را به این جدول اضافه کنیم. برای اینکار گام به گام کارهایی که انجام می دهید را توصیف کنید به همراه عبارت SQL آنها. (فیلد های نامرتب را نادیده بگیرید.)

۴. فرض کنید می خواهیم کاربرانی را که «به درد نمیخورند!!» را از سیستم حذف کنیم به طوری که اطلاعات کاربری آنها که در جدول Posts قرار گرفته اند به همراه تمام پست ها و کامنت هایشان حذف شود. برای این کار گام به گام توضیح دهید که چه پرس و جو هایی را اجرا خواهید کرد. فرضیات :

- سیاست حذف کردن به صورت RESTRICT تعریف شده است.
- معیار «به درد نخوردن» را خودتان تعریف کنید. این معیار نباید فقط شامل ستون های جدول Users باشد نمونه های برای این معیار:
 - کاربرانی که میانگین امتیاز پاسخ هایشان منفی است.
 - کاربرانی که فقط سوال می پرسند و هیچ جوابی ارسال نمی کنند.
 - کاربرانی که هیچ نشانی ندارند.

سوالات تئوری

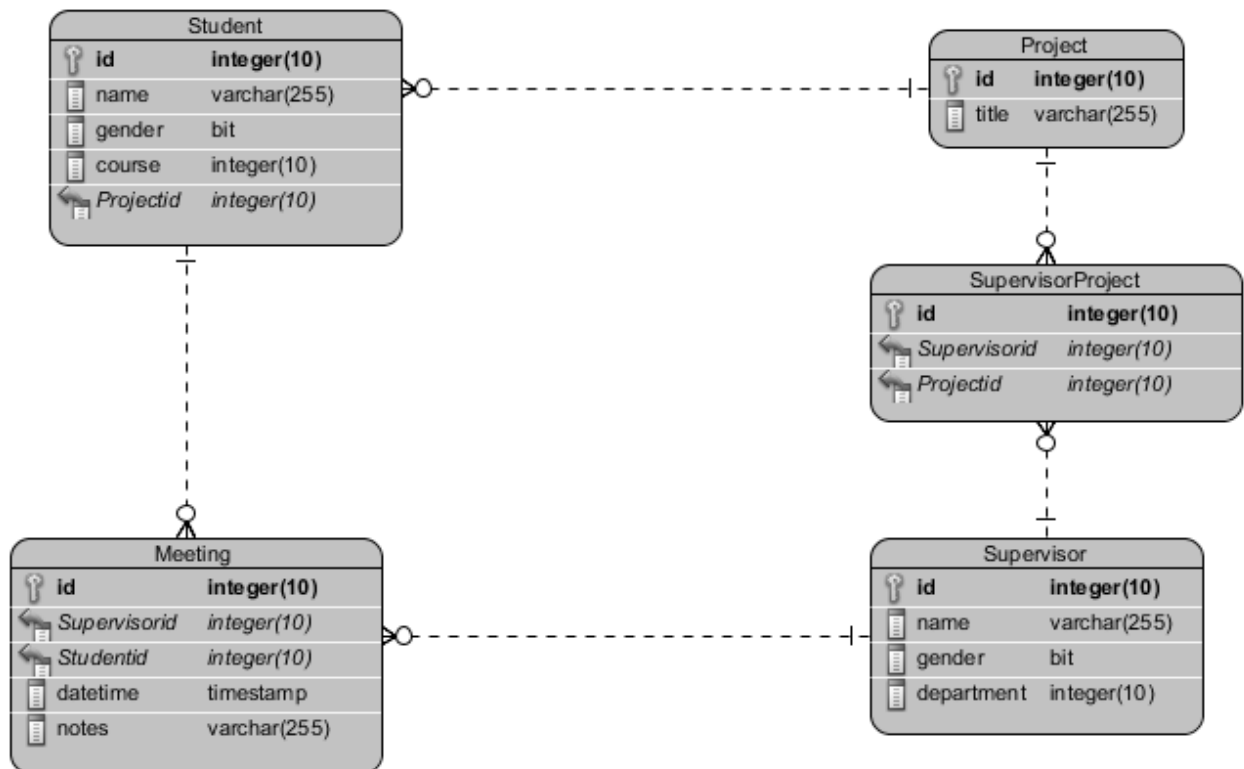
۱. تفاوت مشخصه unique و primary key چیست؟ آیا کلیدهای خارجی نیز باید یکتا باشند؟

۲. چه مواقعی لازم است تا کلید اصلی به صورت مجموعه ای از چند ستون تعریف شود؟

۳. تحقیق کنید که تفاوت موارد زیر در چیست؟

- INNER JOIN
- LEFT OUTER JOIN
- RIGHT OUTER JOIN
- FULL OUTER JOIN

۴. (اختیاری) نمودار ER سیستمی طراحی کنید که متناظر با شمای پایگاه داده رابطه ای زیر باشد.



شکل ۳: برگرفته از یکی از سوالات سایت Stack Overflow به همراه تغییراتی اندک

راهنمایی

اگر در مورد سوالات این تمرین سوالی و یا ابهامی هست، لطفاً با رایانامه زیر مکاتبه کنید.

abarkhordari@ce.sharif.edu

موفق باشید

علیرضا برخوردار

نکات مهم انجام تمرین

- مستند تکلیف باید به صورت PDF تحویل داده شود. کلیه محتویات تکلیف (مستند، کد منبع، خروجی نرم افزارها و غیره) بایستی در قالب یک فایل فشرده با نام DB-942-HW#-StudentNumber (شماره تمرین است) به آدرس DB.94952@gmail.com ارسال گردد. عنوان نامه بایستی همانند نام فایل باشد.
- تاخیر در ارسال تکلیف، مشمول کسر نمره خواهد بود. هر روز تاخیر مشمول ۰.۲۵٪ کسر نمره خواهد گردید.
- پاسخ هر سوال باید دقیق و متناسب با سوال باشد. از ذکر مطالب مبهم، نامرتبط و زائد خودداری نمایید.
- در صورت استفاده از منبع برای پاسخ به سوال، ذکر آدرس منبع الزامی است.
- پاسخها باید با کلمات خودتان بیان شوند. کپی برداری عینی از منابع، نمره‌ای برای پاسخ شما در پی نخواهد داشت.
- در صورت کشف تقلب، بر اساس مقررات آموزشی برخورد خواهد شد.
- پاسخها فقط می‌توانند به زبان فارسی باشند.
- حداکثر تعداد صفحات بایستی ۱۵ صفحه باشند.
- در صورت نقض هر کدام از نکات گفته شده، نمره کسر می‌گردد.

موفق باشید