

تمرین اول

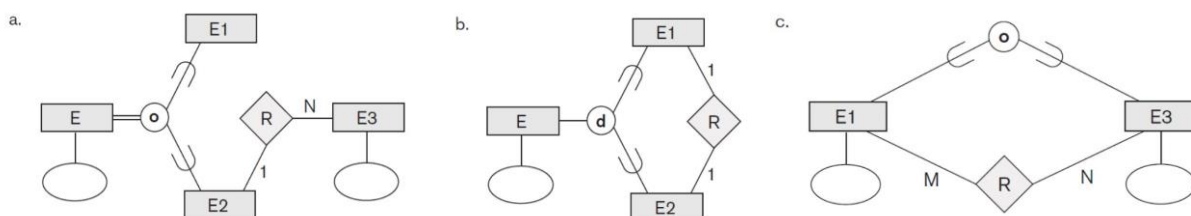
۱. داده‌های ساختمان، نیم‌ساختمند و ناساختمند را تعریف کنید و برای هر کدام مثال بزنید.
۲. دلایل استفاده از سلسله مراتب حافظه‌ها چیست؟ بیان کنید که چه داده‌ای و برای چه مدتی، در کدامیک از مراتب سلسله مذکور قرار می‌گیرد؟
۳. چرا در خط مشی فایلینگ، خطر بروز ناسازگاری در داده‌ها وجود دارد؟ دو گونه ناسازگاری را شرح دهید. مزایای استفاده از سیستم مدیریت پایگاه داده نسبت به سیستم فایلینگ چیست؟
۴. نسل‌های مختلف فناوریهای ذخیره و بازیابی اطلاعات را نامبرده و هر کدام را به اختصار توضیح داده و بگویید هر کدام چه ویژگی‌های جدیدی ارائه دادند.
۵. انواع مدل‌های داده‌ای پایگاه داده را نام برده و هر کدام را خلاصه‌وار شرح دهید. مدل رابطه‌ای جزو کدامیک از مدل‌های داده‌ای است؟ دیگر مدل‌های هم‌سطح مدل رابطه‌ای را نام ببرید، برای هر کدام مثالی ارائه کرده و با یکدیگر مقایسه کنید.
۶. پایگاه داده درون حافظه‌ای چیست؟ چه معایب و مزایایی دارد؟ برای نگهداری دائمی داده‌ها در این نوع پایگاه داده چه راه‌حلهایی پیشنهاد شده است؟ سه نمونه از این نوع پایگاه داده‌ها را نام ببرید.
۷. چگونه می‌توان هر مجموعه موجودیت ضعیف را به موجودیت قوی تبدیل کرد؟ با این حال آیا لازم است از موجودیت‌های ضعیف استفاده کنیم؟ پاسخ خود را با ذکر مثال شرح دهید.
۸. تفاوت‌های ER و EER چیست؟ آیا EER نقطه ضعیفی هم دارد؟
۹. در مدل‌سازی داده‌ها گاهی در دام می‌افتیم. سه دام رایج عبارتند از: حلقه‌ای، چند شاخه‌ای و شکاف. شرایط رخ دادن هر کدام به همراه یک مثال با نمایش مدل ER هر کدام آن‌ها را توضیح دهید و چگونگی رفع آن را نیز ذکر کنید.
۱۰. مدل EER پایگاه داده دانشگاه که نیازمندی‌های زیر را دارد را طراحی کنید:
 - پایگاه داده اطلاعات سه دسته از افراد را نگهداری می‌کند: کارمند، فارغ‌التحصیلان و دانشجویان. یک فرد می‌تواند به یک، دو یا هر سه دسته تعلق داشته باشد. اطلاعات هر شخص شامل نام، کد ملی، جنسیت، آدرس و تاریخ تولدش در پایگاه داده ذخیره می‌شود.
 - هر کارمند حقوق مشخصی دارد. سه دسته کارمند وجود دارد: اساتید، پرسنل و دانشجویار(دانشجویار نوعی از دانشجو است که مسئولیت‌های بیشتری از یک دانشجوی معمولی دارد). هر کارمند به یکی از این سه دسته تعلق دارد. اطلاعات هر فارغ‌التحصیل اعم از مدرک یا مدرک‌های دریافت شده توسط او در دانشگاه، سال فارغ‌التحصیلی و نام دانشکده‌اش ذخیره می‌شود. (هر دانشجو به یک دانشکده تعلق دارد)

- اساتید، دارای درجه و پرسنل دارای منصب کاری هستند. دانشجویارها به دو گروه تحقیق یار و تدریس یار تقسیم می شوند. عنوان پروژه تحقیقاتی تحقیق یاران و عنوان درس تدریس یاران در پایگاه داده ذخیره می شود.
- دانشجویان نیز به دسته کارشناسی، کارشناسی ارشد و دکترا تقسیم می شوند.

۱۱. مجموعه های موجودیت و صفتهای زیر را در نظر بگیرید. مشخص کنید که کدام یک از ارتباطات ذکر شده در ستون های جدول میان سطرهای ستون اول و آخر (از راست) برقرار است. (با تیک یا ضربدر در جدولی مشابه مشخص کنید)

| مجموعه موجودیت (مبدأ) | تعمیمی است از... | تخصیصی است از... | صفتهای دارد به نام... | ارتباط دارد با... | مجموعه موجودیت یا صفت (مقصد) |
|-----------------------|------------------|------------------|-----------------------|-------------------|------------------------------|
| دختر | | | | | شخص |
| دختر | | | | | مادر |
| دانشجو | | | | | شخص |
| دانشجو | | | | | شماره دانشجویی |
| دانشگاه | | | | | دانشجو |
| دانشگاه | | | | | کلاس درس |
| حیوان | | | | | اسب |
| اسب | | | | | سن |
| کارمند | | | | | شماره شناسایی |
| مبل | | | | | صندلی |
| صندلی | | | | | وزن |
| انسان | | | | | زن |
| سرباز | | | | | شخص |
| جنگجو | | | | | شخص |

۱۲. کدام مدل یا مدل های EER زیر دارای خطا هستند و چرا؟ مفروضات خود را به صورت شفاف بیان کنید.



۱۳. مدل ER پایگاه داده یک بیمارستان با نیازمندی‌های زیر را طراحی کنید: (چندی ارتباط‌ها و شناسه موجودیت‌ها را نیز مشخص نمایید)

- بیمارستان اطلاعات بیمار که شامل نام، سن، جنسیت، تاریخ تولد و آدرس او می‌شود را نگهداری می‌کند.
- بیمارستان هر بیمار را با شناسه منحصر به فردی متمایز می‌کند و برای او یک سابقه پذیرش ایجاد می‌کند. سابقه پذیرش شامل شماره پذیرش، تاریخ پذیرش و تاریخ مرخص شدن است.
- بیمارستان برای هر بیمار، یک دکتر را در نظر می‌گیرد. اطلاعات دکتر شامل نام، شناسه، زمینه تخصص و سنوات سابقه نزد بیمارستان ذخیره می‌شود. هر دکتر می‌تواند چند بیمار را زیر نظر داشته باشد.
- هر بیمار به یک بخش سپرده می‌شود. هر بخش با شماره بخش، نام، نوع بخش (درمانی، جراحی و ...) شناسایی می‌شود. هر بخش چندین تخت دارد. هر تخت دارای شماره و نوع (مثلاً، برقی یا ساده) است.
- اگر بیماری نیاز به جراحی داشته باشد، بیمارستان برای او برنامه‌ریزی‌های لازم را انجام می‌دهد. فرم عمل شامل اطلاعاتی از قبیل شماره عمل، تاریخ، ساعت، شناسه بیمار و شناسه دکتر است. چندین دکتر می‌توانند در یک عمل شرکت داشته باشند.

۱۴. از ما خواسته شده پایگاه داده‌ای برای یک شرکت معامله املاک طراحی کنیم که شعبه‌های زیادی سرتاسر کشور دارد. نیازمندی‌های آن در زیر ذکر شده است. مدل EER آن را طراحی کنید. (چندی ارتباط‌ها و شناسه موجودیت‌ها را نیز مشخص نمایید)

- هر شعبه دارای یک شماره شعبه منحصر به فرد است. همچنین تعدادی کارمند دارد. هر شعبه لیستی از خانه‌های موجود برای خرید یا اجاره دارد.
- هر خانه دارای یک شماره خانه منحصر به فرد، آدرس، قیمت اجاره، قیمت فروش، وضعیت و شماره شعبه است. "وضعیت" امکان خرید یا اجاره خانه را مشخص می‌کند. شماره شعبه نیز تعیین می‌کند که کدام شعبه می‌تواند این خانه را اجاره یا به فروش برساند.
- هر مشتری یک کد ملی، نام، شماره قرارداد و شماره خانه (که خرید یا اجاره کرده) دارد.
- هر قرارداد یک شماره قرارداد منحصر به فرد، نوع (اجاره یا خرید)، تاریخ، شماره شعبه و کد ملی مشتری است.
- اگر یک قرارداد از نوع اجاره باشد، مبلغ پیش، اجاره خانه ماهیانه و طول مدت قرارداد خواهد داشت و اگر از نوع خرید باشد، مبلغ فروش.
- طول مدت قرارداد می‌تواند شش ماه، یک سال یا دو سال باشد. مشتری همچنین می‌تواند قرارداد خود را تمدید نماید.
- هر تمدید قرارداد شامل یک شماره تمدید منحصر به فرد، شماره قرارداد اصلی و درصد اضافه شده به اجاره خانه (با توجه به تورم!!) است.
- مشتری می‌تواند قرارداد خود را فسخ کند. هر درخواست فسخ، یک شماره درخواست منحصر به فرد، کد ملی مشتری و هزینه فسخ دارد.

۱۵. (امتیازی) کنترل هم‌زمانی (Concurrency) در پایگاه داده های توزیع شده فعالیت است که دسترسی های همزمان و موازی به یک پایگاه داده را هماهنگ و تعدیل می کند. این کار به کاربران اجازه می دهد تا با حفظ ذهنیت هر کاربر مبنی بر اینکه به صورت انفرادی در حال کار بر روی یک سیستم اختصاصی است، به پایگاه داده دسترسی پیدا کنند. سختی اصلی در رسیدن به این هدف این است که جلوی تداخل بروزرسانی های انجام شده توسط یک فرد بر روی پایگاه داده را با بروزرسانی های انجام شده توسط فرد دیگر را بگیریم. توضیح دهید این کار در حال حاضر چگونه انجام می شود و چند نمونه از پایگاه داده هایی که این روش را پشتیبانی می کنند را نام ببرید.

نکات مهم انجام تمرین

- مستند تکلیف باید در دو نسخه Word و 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.
- در صورت بروز ابهام در مورد سوالات، می توانید از طریق آدرس bagheri.ramtin@gmail.com مکاتبه نمایید.
- از به کار بردن فینگلیش در پرسیدن سوالات از طریق ایمیل خودداری نمایید.
- هرگونه سوال و ابهام در مورد تمرینات حداکثر تا ۲۴ ساعت قبل از مهلت تکلیف پرسیده شود.
- 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، نمره ای برای پاسخ شما در پی نخواهد داشت.
- در صورت کشف تقلب، بر اساس مقررات آموزشی برخورد خواهد شد.
- پاسخها فقط می توانند به زبان فارسی باشند.
- حداکثر تعداد صفحات بایستی ۱۵ صفحه باشند.
- در صورت نقض هر کدام از نکات گفته شده، نمره کسر می گردد.

موفق باشید

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 پرترافیک 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 جویهای SQL خود را از رابط وب این سیستم روی پایگاه داده اعمال کنند و خروجی را برای ما بفرستند.

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

Stack Overflow چیست؟

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 سؤالات 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 جویها 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 جویهایی که قبلاً کاربران روی پایگاه داده اعمال کرده‌اند و ذخیره کرده‌اند را می‌بینید. شما نیز میتوانید پرس و جویهای جالب خودتان را ذخیره کنید و در اختیار سایر کاربران قرار دهید.

برای نوشتن یک پرس و جوی 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 جوی خود را به زبان SQL نوشته و روی دکمه Run Query کلیک کنید تا آن را روی پایگاه داده اجرا کنید. نتیجه پرس و جوی در قسمت زیرین صفحه به نمایش در می‌آید. برای نمونه، من یک پرس و جوی 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

edit description

 stackoverflow
Q&A for programmers

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

Database Schema

Posts

Users

Comments

Badges

CloseAsOffTopicReasonTypes

CloseReasonTypes

FlagTypes

Revisions

593737

593720

593719

permalink

hide sidebar >>

Run Query


Cancel

Options: ☐ Text-only results ☐ Include execution planSwitch sites: 

search by name or url

Results

Messages

 Download CSV

9189

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 بخوایم جدول PostTags را حذف کنیم کدام جدول های دیگر نیز حذف می شوند؟

۱ پرس و جوی 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 و نیز ضمیمه پاسخ نامه تان کنید. (در موارد زیر خروجی هر پرس و جوی یک جدول 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 را. (فیلد های نامرتب را نادیده بگیرید.)
- ۴. فرض کنید می خواهیم کاربرانی را که «به درد نمی خورند!!» 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 هایی 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

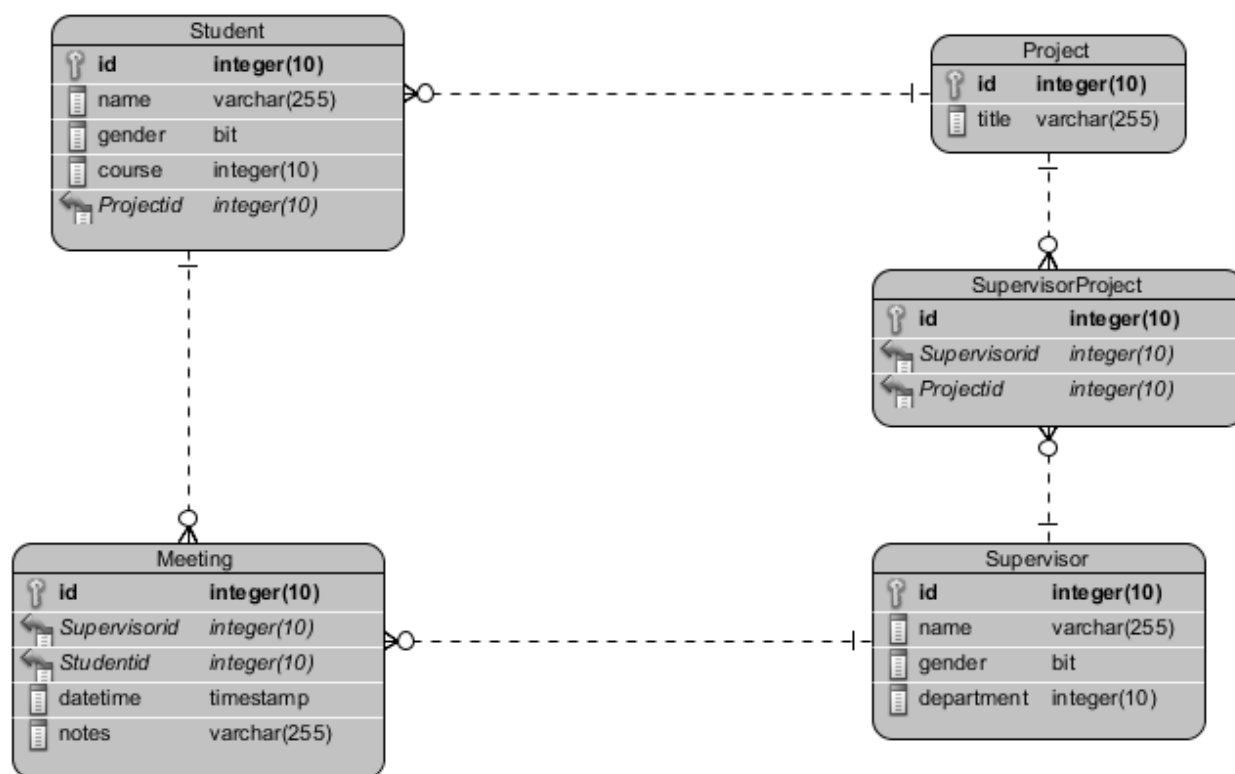
۱. تفاوت مشخصه unique و primary key چیست؟ آیا کلیدهای خارجی نیز باید یکتا باشند؟

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 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

تمرین سوم

۱. شکل زیر یک فرم پزشکی از بیمارستانی را نشان می‌دهد. با توجه به اطلاعات و مفاهیم بیان شده در شکل:

- تمام موجودیت‌های ممکن و صفات آن‌ها را بیان کنید. (در صورتی که صفت خاصی نیاز است و در شکل بیان نشده، آن را در نظر بگیرید).
- پس از شناسایی موجودیت‌ها و رابطه‌ها، انواع مختلف کلیدها را در رابطه‌های به‌دست آمده بیان کنید. (super-key, primary key, candidate key, alternate key و foreign key)

| | | | | | | | |
|----------------------|-----------------|----------------------|-----------|--------------------------|----------------------|---------------|-------------|
| فرم پزشکی بیمارستان | | | | | | | |
| شماره بیمار: ۹۵۰۰۳۴ | | | | | | | |
| نام بیمار: علی احمدی | | سن: ۵۵ | | نام پزشک معالج: رضا حسنی | | شماره بخش: ۱۰ | |
| | | نام بخش: قلب | | شماره تخت: ۵ | | | |
| شماره دارو | نام | توصیف | میزان دوز | روش استفاده | میزان استفاده روزانه | تاریخ شروع | تاریخ پایان |
| ۲۲۳ | مورفین | مسکن | 10mg/ml | تزریق | ۱۰ | ۱۳۹۵/۱/۲۰ | ۱۳۹۵/۱/۲۵ |
| ۱۰۵ | قرص نیتروگانتین | پایین آورنده فشارخون | 0.5mg/ml | خوراکی | ۲۰ | ۱۳۹۵/۱/۲۰ | ۱۳۹۵/۱/۳۰ |
| ۱۲۴۵ | قرص زیرزبانی | کاهنده بار قلب | 10mg/ml | خوراکی | ۵ | ۱۳۹۵/۱/۲۶ | ۱۳۹۵/۱/۳۰ |

۲. شمای زیر بیان‌کننده‌ی موجودیت‌ها و روابط موجود در یک شرکت است:

Employee(fname , lname, Ssn , Bdate , Address , Gender , salary , Super_ssn , Dno)

Department(Dname , Dno , manager_ssn , manager_start_date)

Dept_locations(Dno,Dlocations)

Project(Pname , Pnumber , Plocation , Dnum)

Works_on (Employee_ssn , Pnumber , hours)

Dependent(Essn , Dependent_name , Gender , Bdate , Relationship)

در ادامه هر کدام از موجودیت ها بیان می شوند:

هر کارمند (employee) دارای صفاتی از قبیل نام، نام خانوادگی، شماره کارمندی، تاریخ تولد، آدرس، جنسیت، حقوق، شماره کارمندی که مافوق وی است (می تواند تهی باشد) و شماره دپارتمانی است که وی در آن کار می کند.

هر دپارتمان (Department) دارای صفاتی از قبیل نام دپارتمان، شماره، شماره مدیر دپارتمان و تاریخ آغاز مدیریت آن فرد بر دپارتمان است.

Dept_locations بیانگر محل های آن دپارتمان است که دارای صفات شماره دپارتمان و محل های آن ها است.

شرکت دارای تعدادی پروژه (project) است که دارای صفات نام، شماره، محل و دپارتمان انجام دهنده ی آن هست.

هر فرد بر روی پروژه ای کار می کند که با Works_on نشان داده می شود و دارای صفات شماره کارمند، شماره پروژه و تعداد ساعت های کار بر روی پروژه است.

همچنین هر کارمند دارای تعداد فرد وابسته (خانواده وی) است (Dependent) که این موجودیت دارای صفات شماره کارمند، نام فرد وابسته، جنسیت، تاریخ تولید و ارتباط با کارمند می باشد.

با توجه به توضیحات فوق به موارد زیر پاسخ دهید:

- گراف ارجاع را رسم کنید.
- View های زیر را بیان کنید:
 - i. View ای دربرگیرنده نام دپارتمان، نام مدیر وی و میزان حقوق مدیر آن برای تمام دپارتمان ها
 - ii. View ای شامل اطلاعات نام پروژه، دپارتمانی که بر روی آن کار می کند، تعداد کارمندی که بر روی آن کار می کنند و مجموع ساعات کار افراد آن پروژه برای تمام پروژه ها.
 - iii. View ای شامل نام پروژه هایی که بیش از یک کارمند برای روی آن کار می کنند به همراه نام دپارتمانی که بر روی آن کار می کند.
- View زیر را در نظر بگیرید:

```
Create View DEPT_SUMMARY(D,C, Total_s,Average_s)
As Select Dno , COUNT(*) , SUM(salary) , AVG(salary)
FROM Employee
GROUP BY Dno;
```

کویری ها زیر را در نظر بگیرد. بررسی کنید که آیا قابل اجرا هستند؟ دلیل خود را به صورت مختصر توضیح دهید:

(نمونه ای از مقادیر جدول Employee در زیر آمده است.)

```
i.  SELECT      D,C
     FROM        DEPT_SUMMARY
     WHERE       Total_s > 100000;
```

ii. DELETE FROM DEPT_SUMMARY
WHERE C > 4;

| Fname | Lname | Ssn | Bdate | Address | Gender | Salary | Super_ssn | Dno |
|----------|-----------|-----|------------|---------|--------|---------|-----------|-----|
| Reza | Kamali | 1 | 1360/12/23 | Tehran | M | 1000000 | 10 | 5 |
| Hasan | Moradi | 10 | 1355/1/25 | Tehran | M | 1500000 | 15 | 5 |
| Zahra | Askari | 5 | 1365/7/7 | Tehran | F | 1000000 | 20 | 4 |
| Ali | Ahmadi | 20 | 1345/4/20 | Tehran | M | 2000000 | NULL | 4 |
| Fatemeh | Alavi | 12 | 1359/5/31 | Tehran | F | 1200000 | 10 | 5 |
| Mohammad | Mohammadi | 15 | 1340/6/13 | Tehran | M | 2500000 | NULL | 5 |
| Vali | Zareh | 13 | 1368/2/2 | Tehran | M | 1000000 | 20 | 4 |
| Maryam | Donyavi | 16 | 1365/5/20 | Tehran | F | 1500000 | NULL | 1 |

۳. شمای زیر را که در مورد یک هتل است در نظر بگیرید:

Hotel(hotelNo , hotelName , city)

Room (roomNo , hotelNo , type , price)

Booking(hotelNo , guestNo , dateFrom , dateTo , roomNo)

Guest(guestNo , guestName , guestAddress)

برای موقعیت‌های زیر، رهانا(Trigger) های مربوطه را بیان کنید:

- i. یک مهمان نمی‌تواند دو رزرواسیون که با یکدیگر همپوشانی دارند را انجام دهد.
- ii. برای اتاقی که رزرو شده است، نمی‌تواند رزرواسیون دیگری در بازه‌ای که اتاق رزرو شده انجام داد.
- iii. قیمت اتاق‌های دوفترة بایستی بیشتر از قیمت گران‌ترین اتاق یک‌فتره باشد.
- iv. فرض کنید امکان آن وجود داشته باشد که بتوان رکورد مربوط به یک رزرواسیون را حذف کرد. بدین منظور لازم است تا قبل از حذف یک رکورد از جدول، آن اطلاعات در یک فایل بانام "deleted-booking-log" ذخیره شود.
- v. در هنگام درج یک رکورد رزرواسیون جدید، فیلد dateTo به‌جای گرفتن مقدار پیش‌فرض، مقدار فیلد dateFrom را بگیرد.
- vi. امکان ایجاد یک رزرواسیون بین ساعت‌های ۲۳ تا ۲۳:۱۵ وجود ندارد.

۴. شمای زیر را در نظر بگیرید:

Student(snum:integer , sname:string, major:string , level:string, age: integer)

Class(name:string , meets_at:time , room:string, fid:integer)

Enrolled(snum:integer , cname:string)

Faculty(fid:integer , fname:string, deptid: integer)

برای موقعیت‌های زیر، اظهار (Assertion) مربوطه را بنویسید.

- i. برای هر کلاس حداقل ۵ و حداکثر ۳۰ دانش‌آموز ثبت‌نام می‌کنند.
- ii. هر عضو faculty باید حداقل ۲ درس را تدریس کند.
- iii. دپارتمان با بیشترین تعداد faculty، باید کمتر از ۲ برابر تعداد faculty های دپارتمان با کمترین تعداد faculty، عضو faculty داشته باشد.
- iv. تعداد دروس مختلفی که در رشته‌های علوم کامپیوتر (CS) ارائه (enroll) شده‌اند بیشتر از تعداد دروسی مختلفی است که در رشته‌های ریاضی (Math) ارائه شده‌اند.
- v. دو کلاس نمی‌توانند در یک‌زمان یکسان، در اتاق یکسان برگزار شوند.

۵. شمای زیر را در نظر بگیرید:

Employee(person_name , street, city)
 Works(person_name , company_name, salary)
 Company(company_name , city)
 Manages(person_name , manager_name)

برای موارد زیر، عبارت حساب رابطه‌ای معادل آن را بیان کنید:

- a. نام و آدرس خیابان و شهر اقامت تمام کارمندانی که برای شرکت "ایرانیان" کار می‌کنند و بیش از ۱۰۰ هزار دلار درآمد دارند.
- b. نام تمام کارمندانی که در همان شهری زندگی می‌کنند که شرکت آن‌ها در آن شهر قرار دارد.
- c. نام تمام کارمندانی که در همان شهر و خیابانی زندگی می‌کنند که مدیران آن‌ها در آن زندگی می‌کنند.
- d. نام تمام کارمندانی که در شرکت "ایرانیان" کار نمی‌کنند.
- e. نام تمام کارمندانی که بیش از تمام کارمندان شرکت "ایرانیان" حقوق دریافت می‌کنند.

۶. شمای مربوط به هتل در سؤال ۳ را در نظر بگیرید. بیان کنید که عبارات زیر چه چیزی را محاسبه می‌کنند:

- a. $\pi_{hotelno}(\sigma_{price>50}(Room))$
- b. $\pi_{hotelname}(hotel \bowtie_{Hotel.hotelno=Room.hotelno} (\sigma_{price>50}(Room)))$
- c. $\pi_{guestname,hotelno}(Booking \bowtie_{Booking.guestNo=Guest.guestNo} Guest) \div \pi_{hotelno}(\sigma_{city="tehran"}(Hotel))$

۷. فرض کنید $R = (A, B)$ و $S = (A, C)$ و در نظر بگیرید $r(R)$ و $s(S)$ بیانگر رابطه‌ها باشند. برای هر کدام از موارد زیر عبارت جبر رابطه‌ای معادل را بیان کنید:

- a. $\{ \langle a \rangle \mid \exists b (\langle a, b \rangle \in r \wedge b = 17) \}$
- b. $\{ \langle a \rangle \mid \exists b (\langle a, b \rangle \in r) \vee \forall c (\exists d (\langle d, c \rangle \in s) \Rightarrow \langle a, c \rangle \in s) \}$

۸. انواع قواعد جامعیت را بیان کرده و هر کدام را به اختصار توضیح دهید.

نکات مهم انجام تمرین

- مستند تکلیف باید در دو نسخه Word و PDF تحویل داده شود. کلیه محتویات تکلیف (مستند، کد منبع، خروجی نرم افزارها و غیره) بایستی در قالب یک فایل فشرده بانام DB-942-HW#-StudentNumber (# شماره تمرین است) به آدرس DB.94952@gmail.com ارسال گردد. عنوان نامه بایستی همانند نام فایل باشد.
- در صورت بروز ابهام در مورد سوالات، می توانید از طریق آدرس farzad.niazmand@gmail.com مکاتبه نمایید.
- از به کار بردن فینگلیش در پرسیدن سوالات از طریق ایمیل خودداری نمایید.
- هرگونه سوال و ابهام در مورد تمرینات حداکثر تا ۲۴ ساعت قبل از مهلت تکلیف پرسیده شود.
- تاخیر در ارسال تکلیف، مشمول کسر نمره خواهد بود. هرروز 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، نمره ای برای پاسخ شما در پی نخواهد داشت.
- در صورت کشف تقلب، بر اساس مقررات آموزشی برخورد خواهد شد.
- پاسخها فقط می توانند به زبان فارسی باشند.
- حداکثر تعداد صفحات بایستی ۱۵ صفحه باشند.
- در صورت نقض هرکدام از نکات گفته شده، نمره کسر می گردد.

موفق باشید

تمرین چهارم

۱. (امتیازی) فرم نرمال DKNF را با یک مثال توضیح دهید.

۲. با فرض داشتن شمای رابطه $R(A, B, C, D)$ آیا وابستگی چندمقداری $BC \rightarrow A$ منطقاً وابستگی‌های $A \rightarrow B$ و $A \rightarrow C$ را نتیجه می‌دهد؟ اگر جوابتان مثبت است اثبات کنید، در غیر اینصورت مثال نقضی ارائه کنید.

۳. فرض کنید شمای رابطه $R(A, B, C, D, E)$ را به دو رابطه زیر تجزیه کردیم و وابستگی‌های تابعی زیر را نیز داریم:

$R_1(A, B, C)$

$R_2(A, D, E)$

$F = \{A \rightarrow BC, CD \rightarrow E, B \rightarrow D, E \rightarrow A\}$

الف) با ذکر دلیل نشان دهید که این تجزیه، یک تجزیه بی‌کاست (lossless) است.

ب) آیا این تجزیه حافظ وابستگی (dependency preserving) است؟ (با ذکر دلیل)

۴. با ذکر توضیحات در هر گام، شمای زیر را تا سطح 4NF نرمال کنید:

Books (acs_no, isbn, title, author, publisher)

Users (userid, name, deptid, deptname)

acs_no \rightarrow isbn

isbn \rightarrow title

isbn \rightarrow publisher

isbn \rightarrow author

userid \rightarrow name

userid \rightarrow deptid

deptid \rightarrow deptname

۵. رابطه نشان داده شده در زیر لیست بیمارستان‌هایی را نمایش می‌دهد که نیاز به مواردی دارند که توسط تامین‌کنندگان برای بیمارستان تهیه می‌شوند. هر زمان که بیمارستانی (h) نیاز به مورد مشخصی (i) داشته باشد و تامین‌کننده‌ای (s) را تامین می‌کند و s در حال حاضر حداقل یک فقره از i را برای h تهیه کرده است، s، i را برای h تامین می‌کند. در این مثال فرض کنید هر مورد با نامش به صورت منحصر به فرد شناخته می‌شود.

| hospitalName | itemDescription | supplierNo |
|-----------------|------------------|------------|
| Western General | Antiseptic Wipes | S1 |
| Western General | Paper Towels | S2 |
| Yorkhill | Antiseptic Wipes | S2 |
| Western General | Antiseptic Wipes | S2 |

الف) توضیح دهید چرا رابطه بالا در 5NF نیست؟

ب) گام‌های نرمال‌سازی رابطه به 5NF را با توضیحات بیان کنید.

۶. شمای رابطه $R(A, B, C, D, E, F)$ وابستگی‌های تابعی زیر را برآورده می‌کند:

AB \rightarrow C
C \rightarrow A
BC \rightarrow D
ACD \rightarrow B
BE \rightarrow C
CE \rightarrow FA
CF \rightarrow BD
D \rightarrow EF

الف) هم‌ارز کاهش ناپذیر را برای این مجموعه از FD ها پیدا کنید.

ب) کلید(های) کاندید را مشخص کنید

ج) گراف وابستگی‌های تابعی مذکور را رسم نمایید

د) بستار $\{A, C\}^+$ را تحت این مجموعه از وابستگی‌های تابعی محاسبه کنید.

۷. متغیر رابطه‌ای NADDR با صفات NAME (منحصرفرد) STATE, CITY, STREET و ZIP مشخص می‌شود. برای هر کد پستی، فقط یک شهر و یک استان وجود دارد. برای هر خیابان، شهر و استان، فقط یک کد پستی وجود دارد.

الف) یک مجموعه کاهش ناپذیر از وابستگی‌های تابعی برای این متغیر رابطه‌ای ارائه کنید. همچنین کلید(های) کاندید را مشخص نمایید.

ب) آیا متغیر رابطه‌ای مذکور به شکل BCNF است؟ در صورت وجود طراحی بهتری را برای این مورد ارائه کنید.

۸. فرض کنید شمای رابطه $R(A, B, C, D, E, F, G)$ و مجموعه وابستگی‌های تابعی زیر را داریم.

BCD \rightarrow A
BC \rightarrow E
A \rightarrow F
F \rightarrow G
C \rightarrow D
A \rightarrow G
A \rightarrow B

الف) رابطه R را تا سطح 3NF تجزیه کنید. تمام گام‌های طی شده را به همراه توضیحات بیان کنید.

ب) آیا این تجزیه در BCNF است؟ دلیل خود را ذکر کنید.

۹. فرض کنید شمای رابطه $R(A, B, C, D, E)$ به سه رابطه $R_1(A, B, C)$ و $R_2(B, C, D)$ و $R_3(A, C, E)$ تجزیه شده است. با داشتن مجموعه وابستگی‌های تابعی $F = \{A \rightarrow D, D \rightarrow E, B \rightarrow E\}$ بی‌کاست بودن تجزیه را مورد بررسی قرار دهید. (راهنمایی: از آزمون Chase استفاده کنید)

نکات مهم انجام تمرین

- مستند تکلیف باید در دو نسخه Word و 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.
- در صورت دست‌نویس بودن تمرین، توجه داشته باشد که وجود ابهام و خوانا نبودن، کسر نمره را در پی خواهد داشت.
- در صورت بروز ابهام در مورد سوالات، می‌توانید از طریق آدرس bagheri.ramtin@gmail.com مکاتبه نمایید.
- از به کار بردن فینگلیش در پرسیدن سوالات از طریق ایمیل خودداری نمایید.
- هرگونه سوال و ابهام در مورد تمرینات حداکثر تا ۲۴ ساعت قبل از مهلت تکلیف پرسیده شود.
- تاخیر در ارسال تکلیف، مشمول کسر نمره خواهد بود. هر روز تاخیر مشمول 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. کپی‌برداری 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