



به نام زیبای هستی بخش
دانشگاه صنعتی شریف - دانشکده کامپیوتر، طراحی پایگاه داده‌ها
مهدی آخی

مدت پاسخگویی: ۱۲۰ دقیقه

تعداد سوالات: ۵ سوال

تاریخ آزمون: ۱۴۰۳/۳/۲۱

شماره دانشجویی:

نام و نام خانوادگی:

سوال ۱. نرمال‌سازی:

فرض کنید یک رابطه‌ی R بدین شکل داریم: $R(A, B, C, D, E)$ که دارای وابستگی‌های (FD) زیر است:

$$AB \rightarrow C, BC \rightarrow D, CD \rightarrow E, DE \rightarrow A$$

الف) از بین گزینه‌های زیر کدام‌ها کلید کاندید کمینه هستند؟

$\{AB, BC, CD, ABD, BDE, BCE\}$

ب) رابطه را تا سطح BCNF نرمال‌سازی کنید. برای هر سطح دلیل نرمال بودن یا نبودن را شرح دهید (فرض کنید رابطه نرمال سطح اول هست).

سوالات صحیح غلط (یک خط توضیح بنویسید):

پ) هر رابطه‌ای که دارای دو ویژگی است، BCNF هست.

ت) به خاطر این که Denormalization باعث افزایش تکرار داده‌ها می‌شود، هیچوقت نباید از آن استفاده کرد.

سوال ۲. بهینه‌سازی: روابط و پرسوجوی SQL زیر را در نظر بگیرید:

Department(DID, Name, Building_Num, Telephone, FID)

Student(SID, DID, Enroll_Year, Nationality)

Finance(FID, Budget, Expenses, ...)

SELECT D.Name, F.Budget

FROM Student S, Department D, Finance F

WHERE S.DID = D.DID

AND D.FID = F.FID AND (D.Building_Num > 5 AND D.Building_Num <= 10)

AND (S.Enroll_Year = 1395 OR S.Enroll_Year = 1393);

یک سری آمار در مورد داده‌ها داریم:

- Building_Num یک عدد بین ۱ تا ۲۰ است.
- هر Building تعداد یکسانی دپارتمان (دانشکده) دارد.
- ثبت‌نام دانشجویان Student.Enroll_Year دامنه‌ای از سال ۱۳۹۱ تا ۱۳۹۶ است، تعداد دانشجویان هر سال:

1391	1392	1393	1394	1395	1396
5000	1400	2000	3000	7000	1600

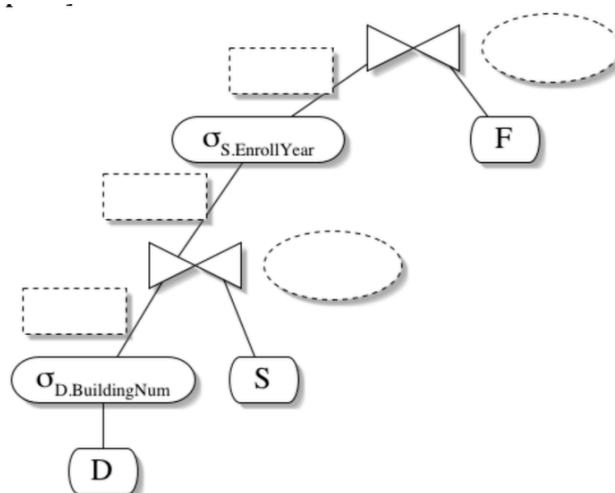
تعداد تاپل‌ها برای هر رابطه به صورت زیر است:

Student: 20000

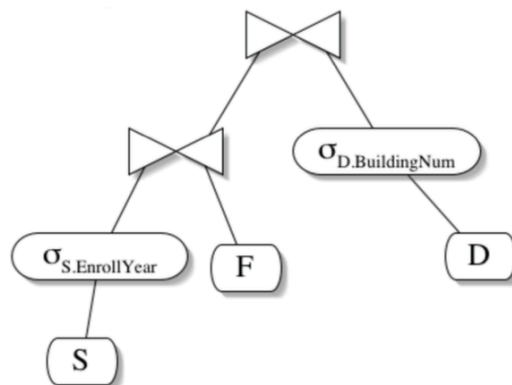
Department: 100

Finance: 100

الف) طرح و نقشه پرسوجوی ذیل را در نظر بگیرید. تعداد روابط ایجاد شده در هر بخش این پرسوجو را در داخل مستطیل‌های خطچین وارد کنید. (با بیضی‌ها کاری نداشته باشید).



ب) به نظر شما ممکن است یک بهینه‌ساز (مثل چیزی که داخل کلاس توضیح دادیم) پیشنهادی مانند شکل زیر برای بهینه‌سازی پرسوجوی یاد شده بدهد؟ توضیح دهید



سوال ۳. تراکنش و همزمانی: Bob و Anne در یک شرکت کار می‌کنند و هر دو یک حساب مشترک دارند. امروز (از شانس بد شما) به صورت تصادفی به صورت همزمان رفتن بانک که کارهای بانکیشون رو انجام بدن. فعالیت بانکی اونها به صورت زیر است.

الف) نمودار تقدم (Precedence) را برای تراکنش‌ها رسم کنید.

ب) آیا زمان‌بندی ارائه شده در شکل بخش الف سوال conflict serializable است؟ اگر بله لطفا یک زمان‌بندی معادل ارائه دهید.

پ) فرض کنید که دیتابیس بانک از الگوریتم 2PL شرح داده شده در کلاس برای پیاده‌سازی توافقی در تراکنش‌ها استفاده کرده است. فرض کنید که درخواست‌های Bob و Anne به صورتی که در قسمت الف نمایش داده شده است، انجام می‌شوند. الگوریتم را روی این توالی انجام تراکنش اجرا کنید و توضیح دهید که چه اتفاقی خواهد افتاد! (در بانک اگر نتواند قفل را برای یک عملیات خاص به دست آورد، فرآیند انجام درخواست فرد تا زمان به دست آوردن قفل متوقف می‌شود)

T_Bob	
Description	Action
Looks at checking balance	R(C)
Withdraw \$100 from checking	W(C)
	commit

T_Anne	
Description	Action
Looks at savings balance	R(S)
Looks at checking balance	R(C)
Transfer \$200 from checking to savings	W(C)
	W(S)
	commit

سوال ۴. ایندکس: دستورات SQL زیر را که به صورت دائمی در سیستم اجرا می‌شوند در نظر بگیرید:

1. SELECT DISTINCT name, address FROM Repayment WHERE borrower_id = ?;
2. SELECT * FROM Repayment WHERE borrower_id = ? AND repayment_date > ?;
3. SELECT borrower_id, loan_amount FROM Repayment WHERE loan_amount BETWEEN ? AND ?;

این دستورات بر روی رابطه‌ی زیر ساخته شده‌اند:

Repayment(borrower_id, name, address, loan_amount, request_date, repayment_date, request_amount)

به نظر شما ایندکس اصلی (Primary) کدام است؟ ایندکس‌هایی که نیاز است تا سرعت این کوئری‌ها افزایش یابد را بنویسید.

سوال ۵. RAID: به نظر شما کدام سطح RAID برای آرکایو کردن و بکاپ گرفتن مناسب است؟ این تناسب را در مقایسه با وضعیتی در نظر بگیرید که Replication داده‌ها صرفاً برای reliability در لحظه وجود دارد نه برای نگهداری مطمئن داده‌ها در طولانی مدت. اگر طراحی دیگری در نظر دارید بنویسید.