

طراحی پایگاه داده ها

دانشکده مهندسی کامپیوتر

مریم رمضانی
بهار ۱۴۰۴



تمرین چهارم

تاریخ انتشار: ۱ خرداد ۱۴۰۴

۱. پرسش های خود درمورد این تمرین را در سامانه کوثر مطرح کنید.

۲. سیاست ارسال با تاخیر: شما در مجموع در طول نیم سال می توانید از ۹ روز تاخیر استفاده کنید. همچنین هر تمرین را می توانید تا حداکثر ۴ روز با تاخیر تحویل دهید. تاخیرها با مقیاس ساعت محاسبه شده و به بالا گرد می شوند.

۳. سیاست مشارکت دانشجویان در حل کردن تمرین: دانشجویان می توانند در حل تمرین برای رفع ابهام و یا به دست آوردن ایده های کلی با یکدیگر مشورت و همفکری کنند. این کار مورد تایید و تشویق تیم ارائه ای درس می باشد؛ چرا که هم فکری و کار گروهی می تواند موجب تقویت یادگیری شود. اما به دست آوردن جزئیات راه حل و نگارش پاسخ باید تماما توسط خود دانشجو انجام شود. حتما در انتهای پاسخ های ارسالی خود نام افرادی که با آنها همفکری کردید را ذکر کنید.

۴. این تمرین ۱ نمره از ۶ نمره کل تمرین را شامل می شود

تاریخ تحویل: ۱۰ خرداد ۱۴۰۴

سوالات تئوری (۱۰۰ نمره)

پرسش ۱ (۸ نمره) درستی یا نادرستی هر یک از گزاره های زیر را با ذکر دلیل مشخص یا برای آن ها مثال نقض بزنید.

(آ) کاهش تعداد جداول از اهداف عملیات نرمال سازی محسوب می شود.

(ب) نرمال سازی عموما به افزایش سرعت پرس و جو ها منجر می شود.

(ج) اگر جدول فقط یک ستون با مقادیر غیر تکراری داشته باشد، می توان گفت شرط لازم و کافی برای قرار داشتن آن در فرم نرمال پنجم برقرار است.

(د) وجود یک secondary index بر روی یک جدول موجب افزایش سرعت عملیات update در آن جدول می شود

(ه) اگر در یک رابطه فقط کلید اصلی به عنوان index تعریف شود، استفاده از یک bitmap index بهبود عملکرد خاصی در کوئری های select نخواهد داشت.

(و) تبدیل یک رابطه از فرم نرمال دوم به فرم نرمال سوم همیشه ساختار رابطه را ساده تر می کند.

پرسش ۲ (۲۰ نمره) رابطه ای $R(A, B, C, D, E)$ با وابستگی های تابعی $F = \{A \rightarrow C, A \rightarrow CE, B \rightarrow C, C \rightarrow D, E \rightarrow D, CD \rightarrow E\}$ را در نظر بگیرید و به پرسش های زیر پاسخ دهید. در هر گام تمامی مراحل و نوع قاعده ای استفاده شده را ذکر کنید.

(آ) پوشش استاندارد^۱ این مجموعه را بیابید.

(ب) تمامی کلیدهای کاندید این رابطه را مشخص کنید. تعداد ابرکلیدها چندتا است؟

(ج) بی کاست^۲ و حافظ وابستگی^۳ بودن هر یک از تجزیه های زیر را بررسی کنید.

$R2(B, C, D, E), R1(A, C)$ •

$R2(B, C, D, E), R1(A, B)$ •

$R2(C, D, E), R1(A, B, C)$ •

(د) آیا تجزیه ای این رابطه به روابط $R3(D, E)$ و $R2(C, E), R1(A, B, C)$ از دید بی کاست و حافظ وابستگی بودن تجزیه ی خوبی است؟

پرسش ۳ (۲۵ نمره) تصور کنید که وظیفه باز طراحی پایگاه داده برای سیستم مدیریت کتابخانه دانشگاه به شما محول شده است. این سیستم باید بتواند انواع مختلفی از امانت ها، فرمت های مختلف و پیگیری تراکنش های خاص کاربران را مدیریت کند. جدول اولیه ای که برای تجزیه و تحلیل و بهینه سازی به شما داده شده است به شرح زیر است:

Library(LoanID, MemberID, MemberName, MemberAddress, MemberType, ItemID, ItemTitle, ItemType, ItemFormat, AuthorID, AuthorName, DueDate, LoanDate, ReturnDate, LateFee)

وابستگی های عملکردی به شرح زیر است:

- LoanID \rightarrow MemberID, ItemID, LoanDate, DueDate, ReturnDate, LateFee
- MemberID \rightarrow MemberName, MemberAddress, MemberType

۱. پوشش استاندارد: canonical cover

۲. بی کاست: lossless

۳. حافظ وابستگی: dependency preserving

- ItemID \rightarrow ItemTitle, ItemType, ItemFormat, AuthorID
- AuthorID \rightarrow AuthorName
- MemberType, ItemType \rightarrow LateFee

(آ) انواع آنومالی‌ها (درج، به‌روزرسانی، حذف) که ممکن است با طرح فعلی رخ دهند را شناسایی کنید.

(ب) جدول Library را تا فرم نرمال BCNF نرمال‌سازی کنید. هر مرحله از فرآیند نرمال‌سازی را شامل ایجاد جدول‌های جدید و بازتوزیع صفات توضیح دهید.

(ج) ارزیابی کنید که طراحی چگونه تغییرات آینده را پشتیبانی می‌کند، مانند افزودن انواع جدید موارد یا تغییر سیاست‌های امانت بر اساس انواع اعضا.

پرسش ۴ (۲۰ نمره) با توجه به روابط زیر به سوالات گفته شده پاسخ دهید:

(آ) با در نظر گرفتن رابطه $R(A, B, C, D, E)$ و مجموعه وابستگی‌های تابعی زیر:

$$F = \{AB \rightarrow C, CD \rightarrow E, DE \rightarrow B\}$$

تمام ابرکلیدهای این رابطه را بیابید.

(ب) رابطه $R(A, B, C, D, E, F, G)$ با کلید اصلی BCD و با مجموعه وابستگی‌های تابعی زیر مفروض است:

$$FD = \{BCD \rightarrow E, BCD \rightarrow B, BCD \rightarrow F, BD \rightarrow A, F \rightarrow G, E \rightarrow F, C \rightarrow E\}$$

در صورت تبدیل آن به ۳NF چند رابطه ایجاد می‌شود؟ مراحل تجزیه روابط را به طور کامل توضیح دهید.

پرسش ۵ (۱۲ نمره)

(آ) دو نوع اصلی indexing در پایگاه داده را نام ببرید و تفاوت‌های کلیدی آن‌ها را توضیح دهید.

(ب) فرض کنید یک جدول کاربران شامل ویژگی‌های زیر است:

Description	Attribute
کلید اصلی	UserID
نام کاربری (یونیک)	Username
ایمیل کاربر (یونیک)	Email
سن کاربر	Age
شهر محل سکونت	City
تاریخ ثبت نام	RegistrationDate

اگر بخواهید این جدول را برای سه نوع کوئری زیر بهینه کنید، چه استراتژی ایندکس‌گذاری‌ای پیشنهاد می‌دهید؟ توضیح دهید که چرا انتخاب شما بهترین گزینه است.

- یافتن سریع کاربری با نام کاربری مشخص.
- پیدا کردن تمام کاربران ثبت‌نام شده در یک بازه زمانی خاص.
- جستجوی تمام کاربران یک شهر خاص بر اساس سن.

پرسش ۶ (۱۵ نمره) در مورد افراز^۴ جدول به پرسش‌های زیر پاسخ دهید.

(آ) برای هر کدام از نمونه‌های زیر با ذکر دلیل بیان کنید از بین روش‌های افراز hash، list، range و round robin کدام ترجیح داده می‌شود.

- یک فروشگاه اینترنتی بزرگ که محصولاتی از نوع‌های مختلفی دارد.
- سیستمی که لاگ‌های مختلف با نرخ بالا از منابع مختلف دریافت و ذخیره می‌کند.
- یک برنامه با تعداد زیادی کاربر که هر کدام از آن‌ها شناسه‌ای دارند.
- سیستم فروشی که تاریخ فروش اجناس را ذخیره می‌کند.

(ب) عملیات درج، حذف و به‌روزرسانی در یک جدول افراز شده چگونه انجام می‌شود؟ آیا اینکه از چه روشی برای افراز استفاده شده است تغییری در این روند ایجاد می‌کند؟ توضیح دهید.

(ج) روش‌های افراز hash، range و round robin را در هر کدام از شرایط زیر، با فرض موازی انجام شدن عملیات، مقایسه کنید.

- اسکن کل جدول
- پرس‌وجوی نقطه‌ای (برای مثال: سن=۲۰)
- پرس‌وجوی بازه‌ای (برای مثال: سن>۲۰، سن>۳۰)

^۴ افراز: partition