

به نام خدا



دانشگاه تهران

دانشکده فنی

دانشکده مهندسی برق و کامپیوتر



درس داده کاوی

تمرین ۲

اسفند ماه ۱۴۰۱

❖ فهرست

سوال ۱ ۳
سوال ۲ ۴
سوال ۳ ۵
سوال ۴ ۶
سوال ۵ ۷
ملاحظات (حتما مطالعه شود) ۸

سوال ۱

فرض کنید که یک data warehouse از سه بعد زمان، پزشک و بیمار و نیز دو معیار تعداد و هزینه‌ی ویزیت تشکیل شده است.

الف) شمای ستاره‌ای را برای این data warehouse ترسیم کنید.

ب) با شروع از cuboid پایه $[day, doctor, patient]$ ، چه اعمال OLAP ای باید انجام شود تا مجموع درآمد هر پزشک در سال ۲۰۲۲ حاصل شود؟

ج) با شروع از cuboid پایه $[day, doctor, patient]$ ، از چه اعمال OLAP ای استفاده می‌کنید تا مجموع هزینه‌ای که هر بیمار بابت ویزیت شدن توسط دندان‌پزشک‌ها در سال ۲۰۲۲ پرداخت کرده است، را به دست آورید؟

سوال ۲

در ارتباط با محاسبه‌ی معیارها در یک data cube، هر کدام از توابع زیر به کدام یک از سه دسته‌ی Algebraic، Distributive و Holistic تعلق دارد؟ به طور کامل توضیح دهید.

- Sum of the largest 30 values
- Mode, if the data is guaranteed to be binary
- Median
- Variance

راهنمایی: فرمول محاسبه‌ی واریانس به صورت زیر است که در آن، \bar{x}_i میانگین x_i ها است.

$$variance = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_i)^2$$

سوال ۳

فرض کنید که cuboid پایه‌ی یک data cube دارای دو سلول زیر است که در آن‌ها برای هر i ، $a_i \neq b_i$ است.

$$(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}): 1$$

$$(\underline{a_1}, b_2, \underline{a_3}, b_4, b_5, b_6, b_7, b_8, \underline{a_9}, b_{10}): 1$$

الف) چند cuboid در این data cube وجود دارد؟

ب) چند سلول بسته‌ی غیرتهی در این data cube موجود است؟

ج) این data cube، چند سلول aggregate غیرتهی دارد؟

د) این data cube، چند سلول aggregate بسته‌ی غیرتهی دارد؟

ه) اگر شرط $\text{minimum support} = 2$ را در نظر بگیریم، تعداد سلول‌های aggregate غیرتهی در iceberg cube متناظر چقدر خواهد بود؟

سوال ۴

یک cuboid پایه با سه بعد A، B و C را در نظر بگیرید که کاردینالیتی این ابعاد به ترتیب 10000، 100 و 1000 است و فرض کنید که هر بعد به صورت مساوی به ۱۰ قسمت برای chunking تقسیم می‌شود. با توجه به مراحل محاسبه‌ی data cube کامل با استفاده از الگوریتم Multi-Way Array Aggregation، به سؤالات زیر پاسخ دهید و برای هر بخش، راه‌حل خود را به طور کامل شرح دهید.

الف) ترتیب بهینه‌ی پیمایش chunk ها در cuboid پایه به منظور ساختن cuboid های دوبعدی را بیان کنید.

ب) اگر هر سلول cube، یک معیار را در داخل ۴ بایت ذخیره کند، چه مقدار فضا در حافظه‌ی اصلی برای ساختن cuboid های دو بعدی مورد نیاز است؟ محاسبات خود را بر مبنای ترتیبی که برای پیمایش chunk ها در قسمت الف به دست آورید، انجام دهید.

ج) هر کدام از cuboid های یک بعدی A، B و C از روی کدام یک از cuboid های دوبعدی ساخته می‌شود؟

سوال ۵

هدف ما در این بخش، محاسبه‌ی یک iceberg cube با استفاده از الگوریتم BUC است. ورودی الگوریتم، یک مجموعه داده با ۳ بعد A، B، C و معیار Count است که tuple های زیر را شامل می‌شود.

A	B	C	Count
a1	b1	c1	1
a1	b2	c1	1
a1	b3	c2	1
a2	b4	c2	1
a2	b3	c3	1

الف) ترتیب پردازش ابعاد را به گونه‌ای تعیین کنید که الگوریتم BUC بهترین کارایی را داشته باشد. دلایل خود را برای انتخاب این ترتیب در گزارش بنویسید.

ب) با توجه به ترتیبی که برای بررسی ابعاد در بخش الف مشخص کردید، الگوریتم BUC را روی مجموعه داده‌ی فوق اجرا کنید و iceberg cube را با شرط $\text{minimum support} = 2$ محاسبه نمایید.

توجه: پیش از حل این سؤال، بهتر است که شبه‌کد الگوریتم BUC را از کتاب مرجع مطالعه فرمایید.

ملاحظات (حتما مطالعه شود)

- تمامی نتایج شما باید در یک فایل فشرده با عنوان DM_HW2_StudentID تحویل داده شود.
- خوانایی و دقت بررسی‌ها در گزارش نهایی از اهمیت ویژه‌ای برخوردار است. به تمرین‌هایی که به صورت کاغذی تحویل داده شوند یا به صورت عکس در سایت بارگذاری شوند، ترتیب اثری داده نخواهد شد.
 - مهلت تحویل تمرین به هیچ عنوان تمدید نخواهد شد. مجموعاً ۱۴ روز برای تمامی تمرین‌ها و پروژه‌ی درس به عنوان Grace day در نظر گرفته می‌شود و پس از پایان مجموعاً ۱۴ روز، برای هر تمرینی که پس از زمان اختصاص یافته ارسال شود روزی ۱۵ درصد از نمره آن تمرین کسر خواهد شد.
 - توجه کنید این تمرین باید به صورت تک نفره انجام شود و پاسخ‌های ارائه شده باید نتیجه فعالیت فرد نویسنده باشد (همفکری و به اتفاق هم نوشتن تمرین نیز ممنوع است). در صورت مشاهده تقلب به همه افراد مشارکت کننده، نمره تمرین صفر و به استاد نیز گزارش می‌گردد.
 - در صورت بروز هرگونه مشکل با ایمیل زیر در ارتباط باشید:

[mailto: mahsan.a.a@gmail.com](mailto:mahsan.a.a@gmail.com)

مهلت تحویل بدون جریمه: ۱۴۰۲ / ۱ / ۱۸