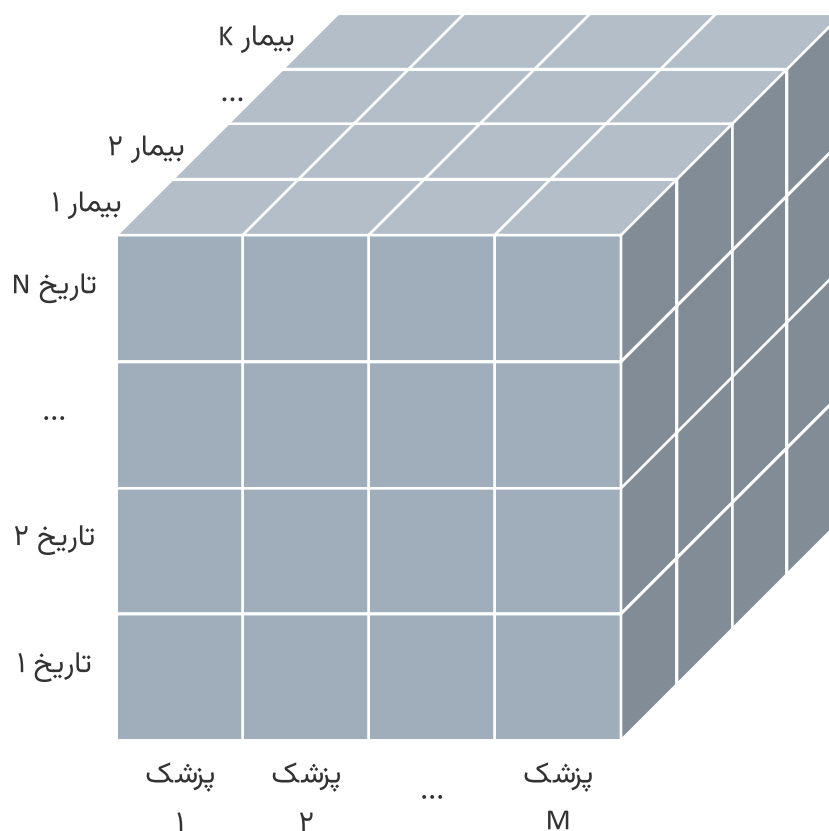


تکلیف فصل سوم درس داده‌کاوی

مصطفی سبیلو (۴۰۱۰۸۲۵۴)

مهدی محمدی (۴۰۱۱۲۰۹۴)

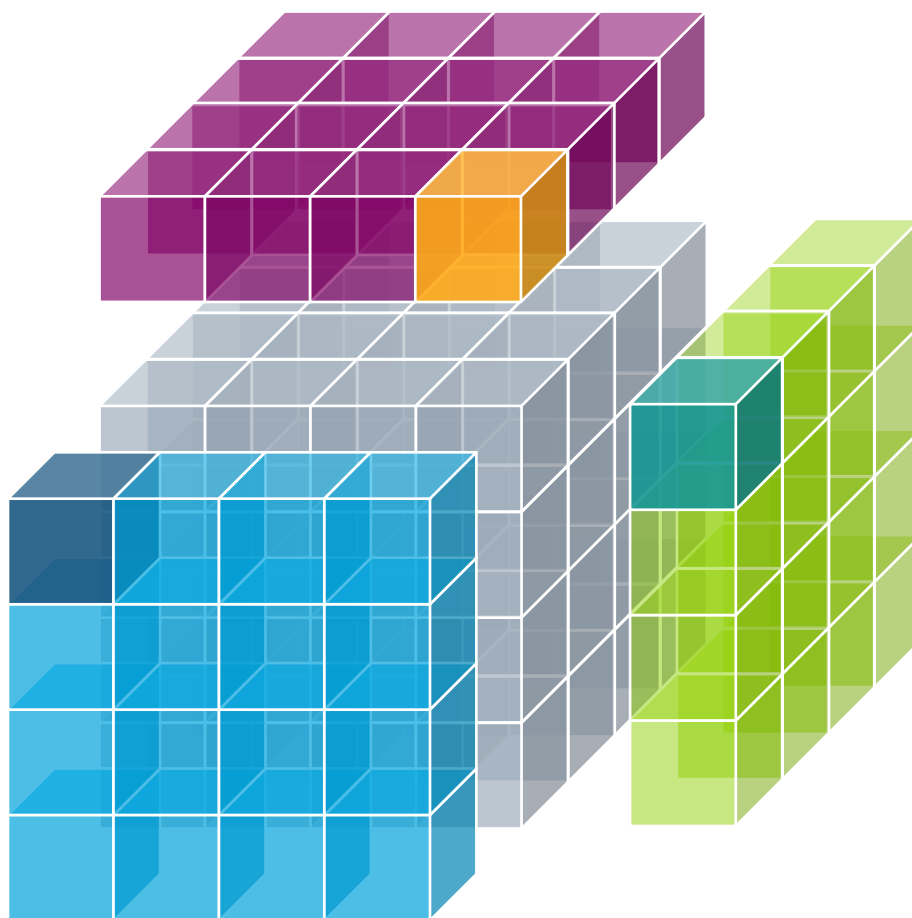
سؤال ۱) یک انباره داده پزشکی با سه بُعد زمان (سال، ماه، روز)، پزشک و بیمار را در نظر بگیرید. بر مبنای این سه بُعد دو فقره داده شامل تعداد مراجعات و مبلغ ویزیت ذخیره شده است. الف) یک نمایش مکعبی مناسب به همراه لایه‌های حاصل جمع برای این انباره داده ارائه نمایید.



مکعب اصلی سه‌بُعدی ترسیمی با توجه به موجودیت‌های پزشک، بیمار و تاریخ.

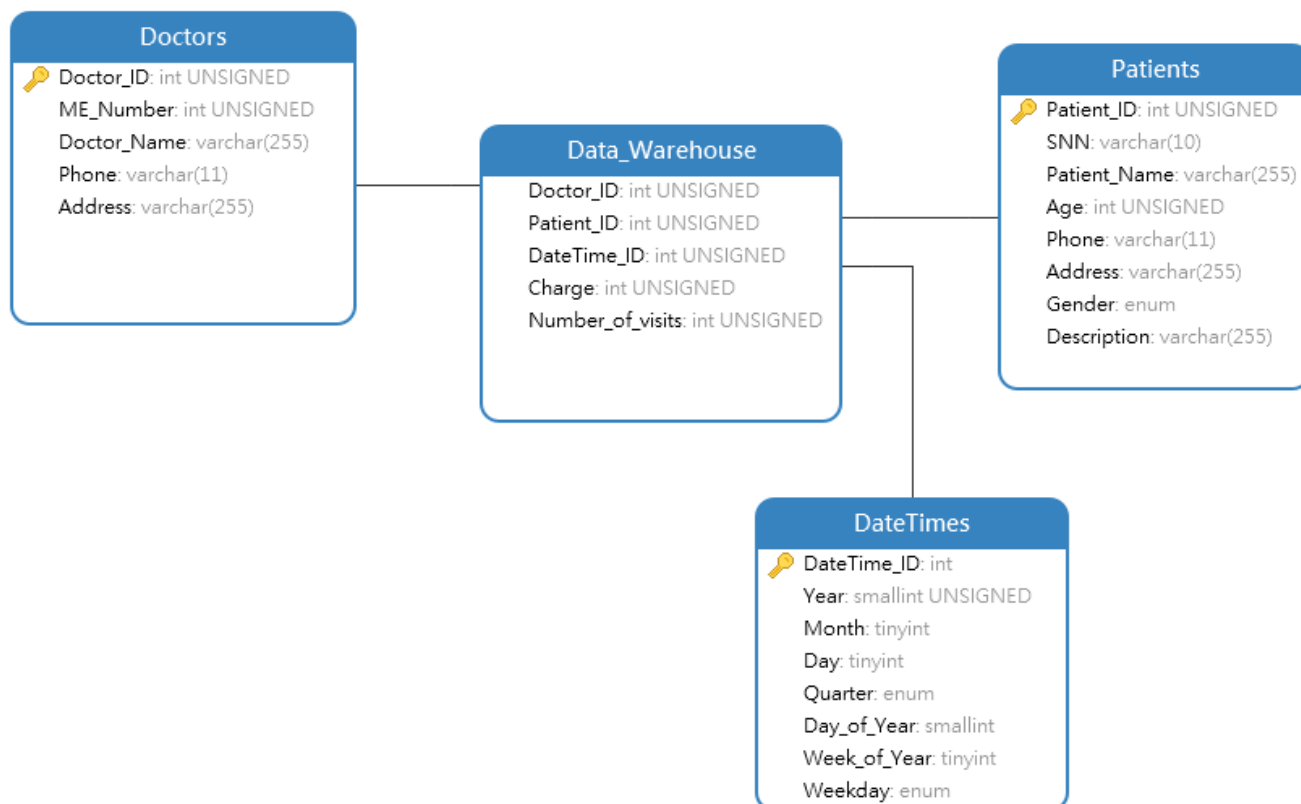
مشخصات هر یک لایه‌های تجمیعی مکعب.

۱	مکعب یک‌دریک در یک	مجموع حق ویزیت و مجموع دفعات مراجعه یک بیمار خاص در یک زمان خاص به همه پزشک‌ها
۲	هر سطر	مجموع حق ویزیت و مجموع دفعات مراجعه همه بیمارها به همه پزشک‌ها در یک زمان خاص
۳	هر ستون	مجموع حق ویزیت و مجموع دفعات مراجعه یک بیمار خاص در همه زمان‌ها به همه پزشک‌ها
۴	مکعب یک‌دریک در یک	مجموع حق ویزیت و مجموع دفعات مراجعه همه بیمارها به یک پزشک خاص در یک زمان خاص
۵	هر سطر	مجموع حق ویزیت و مجموع دفعات مراجعه همه بیمارها به همه پزشک‌ها در یک زمان خاص
۶	هر ستون	مجموع حق ویزیت و مجموع دفعات مراجعه همه بیمارها در همه زمان‌ها به یک پزشک خاص
۷	مکعب یک‌دریک در یک	مجموع حق ویزیت و مجموع دفعات مراجعه یک بیمار خاص به یک پزشک خاص در همه زمان‌ها
۸	هر سطر	مجموع حق ویزیت و مجموع دفعات مراجعه یک بیمار خاص به همه پزشک‌ها در همه زمان‌ها
۹	هر ستون	مجموع حق ویزیت و مجموع دفعات مراجعه همه بیمارها به یک پزشک خاص در همه زمان‌ها



مکعب اصلی سه‌بُعدی ترسیمی همراه با لایه‌های تجمیعی با توجه به موجودیت‌های پزشک، بیمار و تاریخ.

ب) با فرض تعدادی فیلد داده مناسب برای پزشک، بیمار و زمان، یک نمایش در قالب شمای ستاره‌ای برای این انبار داده ارائه کنید.



شمای ستاره‌ای انبار داده پزشک، بیمار و زمان.

پ) چگونه می‌توان در نمایش مکعبی میزان ویزیت پرداخت‌شده برای هر پزشک در یک سال مشخص را نشان داد.

۱. برحسب زمان از روز به سال Roll-Up می‌کنیم.

۲. برحسب سال موردنظر مثلاً ۱۴۰۱ روی مکعب برش می‌زنیم.

۳. از روی تکتک بیماران یک Roll-Up بر روی تمامی بیماران انجام می‌پذیرد.

ت) با فرض ذخیره داده‌ها در جداول رابطه‌ای، معادل مورد خواسته‌شده در بند ب را در قالب یک دستور SQL ارائه نمایید.

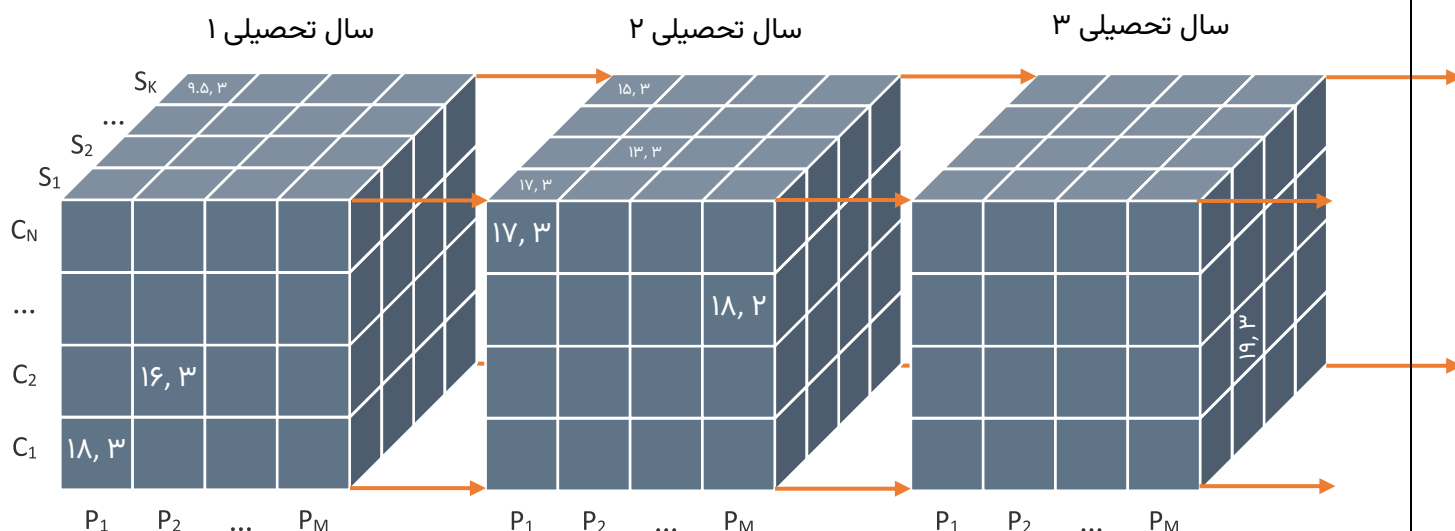
برای این کار ابتدا جدول Fee شامل فیلدهای Year، Month، Day، Doctor_ID، Hospital، Patient Number_of_visits، Charge ایجاد کرده و پرس و جوی (query) زیر را اجرا می‌کنیم.

```

SELECT Doctor, SUM (Charge)
FROM Fee
WHERE Year = 1401
GROUP BY Doctor_ID
  
```

سؤال ۲) یک انباره داده آموزش دانشگاه با چهار بُعد دانشجو، درس، ترم تحصیلی و مدرس را در نظر بگیرید. بر مبنای این چهار بُعد دو فقره داده شامل تعداد واحد و نمره درس ذخیره شده است.

- دانشجو: S
- درس: C
- مدرس: P



هایپرکیوب چهاربُعدی اصلی ترسیمی با توجه به موجودیت‌های استاد، دانشجو، درس و سال تحصیلی.

در شکل بالا نمونه‌هایی از داده‌ها بر روی مکعب‌ها درج گردیده است که تفسیر آن‌ها در زیر آمده است:

دانشجوی ۱: در سال تحصیلی ۱ درس ۳ واحدی ۱ را با استاد ۱ اخذ نموده و نمره ۱۸ کسب کرده است.

دانشجوی ۱: در سال تحصیلی ۱ درس ۳ واحدی ۲ را با استاد ۲ اخذ نموده و نمره ۱۶ کسب کرده است.

دانشجوی ۱: در سال تحصیلی ۲ درس ۳ واحدی N را با استاد ۱ اخذ نموده و نمره ۱۷ کسب کرده است.

دانشجوی ۱: در سال تحصیلی ۲ درس ۲ واحدی ... را با استاد M اخذ نموده و نمره ۱۸ کسب کرده است.

دانشجوی ۲: در سال تحصیلی ۲ درس ۳ واحدی N را با استاد ۲ اخذ نموده و نمره ۱۳ کسب کرده است.

دانشجوی ۲: در سال تحصیلی ۳ درس ۳ واحدی ۲ را با استاد M اخذ نموده و نمره ۱۹ کسب کرده است.

دانشجوی K: در سال تحصیلی ۱ درس ۳ واحدی N را با استاد ۱ اخذ نموده و نمره ۹.۵ کسب کرده است.

دانشجوی K: در سال تحصیلی ۲ درس ۳ واحدی N را با استاد ۱ اخذ نموده و نمره ۱۳ کسب کرده است.

الف) با فرض تعدادی فیلد داده مناسب برای هر بُعد، یک نمایش در قالب شمای برف دانه برای این انبار داده ارائه کنید.



شمای برف دانه انبار داده استاد، دانشجو، درس و سال تحصیلی.

ب) بر مبنای نمایش مکعبی نمایش‌های Roll-Up و Drill-Down ممکن را در این انبار داده نشان دهید.

مثال ۱. دانشجوی ۱ چه درس‌هایی را در نیمسال اول سال تحصیلی ۱ با استاد ۱ گذرانده است؟

در سال تحصیلی ۱، صفحه مربوط به دانشجوی ۱ برش و صفحه مربوط به استاد ۱ برش و پس‌از آن Drill-Down به نیمسال تحصیلی.

مثال ۲. اساتید دپارتمان ریاضی که درس ۱ را در سال تحصیلی ۲ ارائه کرده‌اند؟

انتخاب سال تحصیلی ۲، Roll-Up از شناسه اساتید به دپارتمان، برش دپارتمان ریاضی، پس‌از آن برش درس ۱.

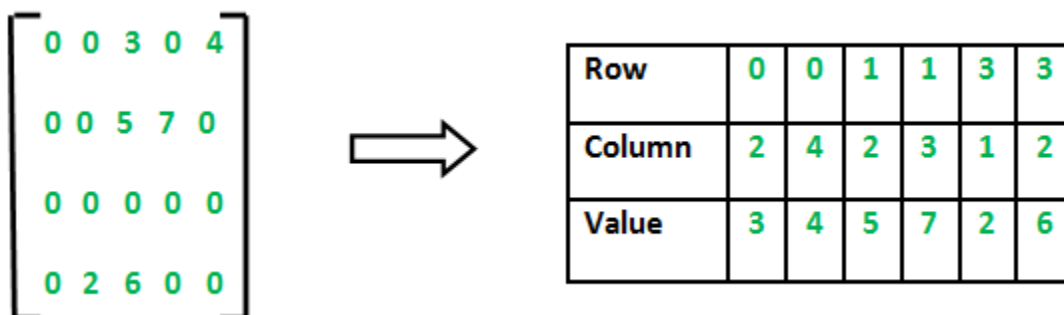
این‌ها دو نمونه از مثال‌های برش، Drill-Down و Roll-Up هستند که بیان شدند. به نحو مشابه می‌توان چنین عملیاتی را بر روی Hyper-Cube‌ها انجام داد.

۳) یک مشکل در نمایش‌های مکعبی، خلوت بودن آن‌ها به دلیل خالی بودن تعداد زیادی از خانه‌های حاوی داده است. بر مبنای دانش خود، دو ساختمان داده یکی بر مبنای آرایه‌ها و دیگری بر مبنای لیست پیوندی برای ذخیره داده‌ها در حافظه اصلی ارائه نمایید. به‌عنوان مثال می‌توانید از یکی از موارد ارائه‌شده در دو سؤال بالا استفاده نمایید.

روش ۱. استفاده از آرایه‌ها:

ارائه دوبعدی برای نشان دادن یک ماتریس پراکنده استفاده می‌شود که در آن سه ردیف با نام‌های زیر وجود دارد:

- ردیف: شاخص ردیف، جایی که عنصر غیر صفر در آن قرار دارد.
- ستون: نمایه ستون که عنصر غیر صفر در آن قرار دارد.
- مقدار: مقدار عنصر غیر صفر واقع در شاخص (ردیف، ستون).

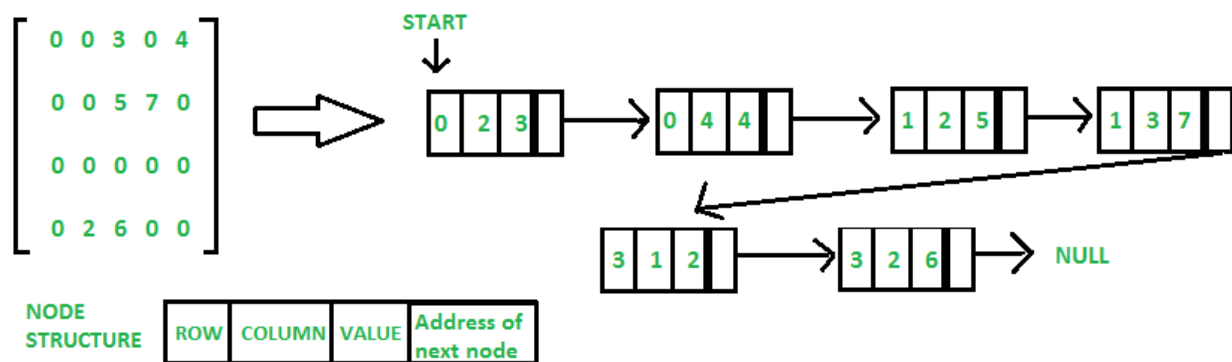


نمونه‌ای از نحوه استفاده از آرایه‌ها جهت ذخیره‌سازی ماتریس‌های خلوت.

روش ۲. استفاده از لیست‌های پیوندی:

در لیست پیوندی، هر گره دارای چهار فیلد است. این چهار فیلد به‌صورت زیر تعریف می‌شوند:

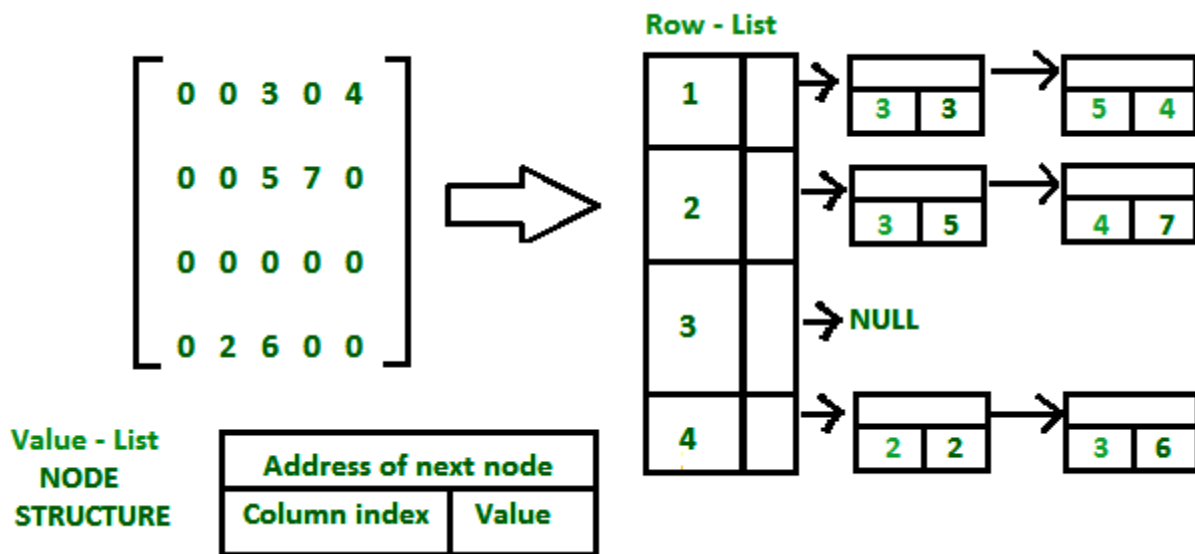
- ردیف: شاخص ردیف، جایی که عنصر غیر صفر در آن قرار دارد.
- ستون: نمایه ستون که عنصر غیر صفر در آن قرار دارد.
- مقدار: مقدار عنصر غیر صفر واقع در شاخص (ردیف، ستون).
- گره بعدی: آدرس گره بعدی.



نمونه‌ای از نحوه استفاده از لیست‌ها جهت ذخیره‌سازی ماتریس‌های خلوت.

روش ۳. فهرست فهرست‌ها (LIL):

یکی از نمایش‌های ممکن ماتریس پراکنده، فهرست فهرست‌ها (LIL) است. جایی که از یک لیست برای نشان دادن ردیف‌ها استفاده می‌شود و هر ردیف شامل لیست سه‌گانه است: نمایه ستون، مقدار (عنصر غیر صفر) و فیلد آدرس، برای عناصر غیر صفر. برای بهترین عملکرد، هر دو لیست باید به ترتیب کلیدهای صعودی ذخیره شوند.



نمونه‌ای از نحوه استفاده از فهرست فهرست‌ها جهت ذخیره‌سازی ماتریس‌های خلوت.