

## طراحی پایگاه داده ها

دانشکده مهندسی کامپیوتر

مریم رضایی  
بهار ۱۴۰۴



### تمرین پنجم - بخش تئوری

تاریخ انتشار: ۱۱ 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 تماماً توسط خود دانشجو انجام شود. حتماً 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: ۲۱ خرداد ۱۴۰۴

### سوالات تئوری (۵۰ نمره)

پرسش ۱ (۱۵ نمره) صحیح یا غلط بودن موارد زیر را تعیین کنید. موارد غلط را تصحیح کنید.

(آ) Independent Data Marts معمولاً داده های خود را از Enterprise Data Warehouse دریافت می کنند.

(ب) Independent Data Marts معمولاً هزینه ی عملیاتی پایین تری نسبت به Enterprise Data Warehouse دارند.

(ج) در EDW، داده ها پس از ذخیره سازی همچنان تغییر می کنند و به روزرسانی می شوند.

(د) Data Mart نسخه ای کوچک تر از Data Warehouse است که معمولاً برای یک بخش خاص سازمان مانند مالی یا بازاریابی استفاده می شود.

(ه) هدف اصلی یک انبار داده (Data Warehouse) تحقیق روی داده های مختلف است.

(و) یک Data Mart برای بهینه سازی عملکرد در کاربردهای مشخص و قابل پیش بینی طراحی شده است.

(ز) پایگاه های داده NoSQL برای سیستم های تحلیلی و گزارش گیری بهتر از SQL هستند.

(ح) پایگاه داده های NoSQL معمولاً برای پردازش تراکش های بانکی و مالی توصیه می شوند.

(ط) Replication به این معنی است که داده ها فقط روی یک سرور ذخیره می شوند تا از تکرار جلوگیری شود.

(ی) NoSQL برای تحلیل داده های بلادرنگ (Real-Time Analytics) مناسب است.

(ک) پایگاه داده های توزیع شده معمولاً از نظر تحمل خطا (Fault Tolerance) عملکرد بهتری نسبت به پایگاه داده های متمرکز دارند.

(ل) توزیع داده ها بین چندین سرور می تواند منجر به مشکلات سازگاری داده (Data Consistency) شود.

(م) در پایگاه داده های توزیع شده، دسترسی به داده ها همیشه سریع تر از پایگاه داده های متمرکز است.

پرسش ۲ (۱۰ نمره) سوالات زیر را درباره معماری های Data Warehouse پاسخ دهید.

(آ) چه مشکلاتی ممکن است هنگام استفاده از Independent Data Marts ایجاد شود؟

(ب) انواع معماری Data Warehouse را نام ببرید و هر کدام را توضیح دهید.

پرسش ۳ (۱۵ نمره) با توجه به مفاهیم انبار داده به سوالات زیر پاسخ دهید :

(آ) انواع داده های موجود در انبار داده ها چیست؟ هر کدام را به اختصار توضیح دهید.

(ب) مزیت وجود و استفاده از data mart ها چیست؟

(ج) مفهوم OLAP را به اختصار توضیح دهید و سه مورد از مزیت های OLAP نسبت به OLTP را بیان کنید.

(د) انواع OLAP cube ها را نام ببرید و هر کدام را به اختصار توضیح دهید.

(ه) با توجه به جدول فروش زیر مشخص کنید که هرکدام از پرسمان های SQL زیر مربوط به کدام یک از عملیات های OLAP می باشد و برای هرکدام به اختصار دلیل خود را توضیح دهید؟ (عملیات ها را از بین Roll-up، Drill-down، Slice، Dice و Pivot انتخاب کنید)  
جدول فروش شامل ستون های id، region، city، product، sales\_amount می باشد.  
مورد ۱:

```
SELECT region, SUM(sales_amount) AS total_sales
FROM sales
GROUP BY region;
```

مورد ۲:

```
SELECT region, city, product, sales_amount
FROM sales
WHERE (region IN ('North', 'South'))
AND (product = 'Laptop')
AND (sales_amount > 4000);
```

مورد ۳:

```
SELECT region, city, sales_amount
FROM sales
WHERE product = 'Laptop';
```

(و) انواع شماهای موحود در انبار داده ها را نام ببرید و هرکدام را به اختصار توضیح دهید.

پرسش ۴ (۱۰ نمره) با توجه به پایگاه داده های nosql به سوالات زیر پاسخ دهید :

(آ) توزیع بار در پایگاه داده های توزیع شده به چه معناست و در nosql از چه تکنیکی برای توزیع بار استفاده می شود؟

(ب) تئوری CAP در پایگاه داده های توزیع شده را توضیح دهید.

(ج) فرض کنید یک پایگاه داده سندگرا مانند MongoDB را در اختیار دارید و یک مجموعه به نام products در این پایگاه داده تعریف کرده ایم. چند نمونه از سند های این مجموعه مانند زیر است (داده ها فرضی هستند)

```
{
  "id" : 1,
  "Name" : "Laptop",
  "category" : "Electronics",
  "price" : 1200,
  "Stock" : 15
}

{
  "id" : 2,
  "Name" : "TV",
  "category" : "Electronics",
  "price" : 1500,
  "Stock" : 10
}
```

با توجه به این ساختار برای موارد زیر پرسمان مناسب طراحی کنید:

مورد ۱: نام کالاهایی که قیمت آن ها بیشتر از ۱۰۰۰ واحد می باشد.

مورد ۲: مجموع مقدار موجودی (stock) برای هر دسته بندی (category) کدام است؟