

به نام خدا

بررسی و تحلیل دیتاست‌های دیجی کالا

بخش اول

نام درس: پایگاه داده-بهار ۹۸

استاد درس: دکتر امین غیبی

گردآورنده: امیرعلی کابلی

شماره دانشجویی: ۹۶۱۳۰۲۰

در بخش اول این پروژه باید با توجه به دیتاست مربوط به لیست محصولات اطلاعات کالاهای خواسته شده را استخراج و دیتابیزی به فرم 5NF طراحی کنیم و اطلاعات بدست آمده را در داخل این دیتابیس ذخیره سازی کنیم

در دیتاست مختص به لیست محصولات هر کالا توسط مشخصاتش ظاهر شده که این مشخصات به شرح زیر هستند:

id, product_title_fa, product_title_en, url_code, title_alt, category_title_fa, category_keywords, brand_name_fa, brand_name_en, product_attributes

در میان این مشخصات محتویات title_alt حاوی چند مقدار است و product_attributes هم حاوی چندین ویژگی از کالا به صورت key-value هست که در طراحی دیتابیس باید برای هر یک کدام از این ویژگی ها یک جدول مجزا در نظر گرفت

پس به ازای هر یک از کالاها ۳ جدول باید در نظر گرفته شود که اولی برای نگهداری داده های آن نوع کالا و جدول دوم برای نگهداری title_alt های مختص به دسته کالا و جدول سوم هم برای نگهداری ویژگی های مربوطه به هر دسته خواهد بود

در طراحی این جداول برای جدول اول ستون های زیر را در نظر می گیریم:

product_id (primary key), product_title_fa, product_title_en, url_code, category_title_fa, category_keywords, brand_name_fa, brand_name_en

برای جدول دوم که نگهدارنده ویژگی title_alt کالاهاست ستون های زیر را در نظر می گیریم:

id (primary key), productid (foreign key (product_id)), alt_title

و برای جدول سوم که حاوی ویژگی های مختلف هر کالا است ستون های زیر را در نظر می گیریم:

productid (primary key)(foreign key (product_id)), ستون های مختص به ویژگی های آن کالا

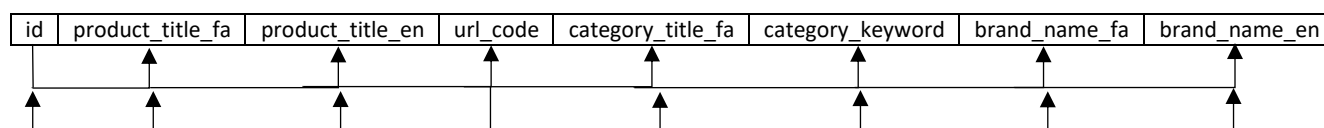
حال ادعا می کنیم دیتابیس طراحی شده به فرم 5NF قرار دارد

در وهله اول می توان گفت به فرمت 1NF قرار دارد چون به ازای هر یک از ستون هایی که در آن ها ویژگی های چندمقداری یا چندبخشی وجود داشت مقادیر موجود در آن ستون را به صورت یک جدول درآوردیم و مقادیر آن

ستون را در جدول جدید نگهداری می‌کنیم پس هیچ‌یک از ستون‌ها دارای ویژگی‌های چندمقداری یا چندبخشی نیستند

سپس می‌توان گفت هر یک از این جداول هم به صورت 2NF قرار دارند چون کلید همه جداول تک عضوی هست پس نمی‌توان ستونی از کلید را حذف کرد تا باز هم بقیه ویژگی‌ها به آن وابسته بمانند

و طراحی انجام شده به فرمت 3NF است چون FDهای جداول alt_title ها و information ها همه ستون‌ها تنها وابسته به کلید جدول یعنی id یا productid هستند و روی جداول اصلی FDها به فرم زیر است



و چون به ازای رابطه متعددی که بین attributeها برقرار شود attributeهای میانی candidate key هستند پس می‌توان گفت که طراحی گفته شده شرایط 3NF را داراست

و چون طراحی این دیتابیس به فرمت 3NF قرار دارد و کلید همه جداول هم تک عضوی است می‌توان گفت این دیتابیس به فرم 5NF هم هست

برای بخش دوم باید روی دیتاست مربوط به تاریخچه خرید مشتریان و بر روی فیلد تاریخ سفارشات ایندکس مناسب قرار می‌دادیم و تحلیل می‌کردیم که کوئری‌های که روی این دیتابیس زده می‌شود آیا این ایندکس را مورد استفاده قرار می‌دهند یا خیر

هرگاه از دستور explain برای تحلیل و آنالیز کوئری‌ها استفاده می‌کنیم خروجی این دستور شامل چند بخش است:

۱. id: شناسه کوئری را مشخص می‌کند

۲. select_type: نوع کوئری را مشخص می‌کند که مقدار می‌تواند یکی از simple یا primary یا union یا dependent union باشد ... باشد

۳. table: نام جدولی که روی آن کوئری زدیم را نشان می‌دهد

۴. partition: نام پارتیشن‌های هریک از جداول را نشان می‌دهد

۵. type: چگونی پیوستن هریک از جدول را نشان می‌دهد

۶. possible_keys: کلیه ایندکس‌هایی که امکان استفاده از آن‌ها فراهم است

۷. key: ایندکس انتخاب شده را نشان می‌دهد

۸. key_len: طول کلیدی که mysql تصمیم می‌گیرد استفاده کند را نشان می‌دهد

۹. rows: تعداد رکوردهایی را نشان می‌دهد که بررسی می‌شوند

۱۰. extra: اطلاعات اضافه برای مشخص کردن داده‌ها را نشان می‌دهد

کوئری اول:

```
select orderid,cityname
from maintable
where orderdate between '2014-07-10' and '2014-07-12';
```

	orderid	cityname
1	1301337	تهران
2	1301155	تربت حیدریه
3	1301550	تهران
4	1301625	اهواز
5	1301604	تهران
6	1301867	تهران
7	1301993	تهران
8	1302288	تهران
9	1302046	قشم
10	1302565	اسدیه

خروجی:

خروجی explain:

id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
1	SIMPLE	maintable	<null>	range	dateindex	dateindex	6	<null>	10	100	Using index condition

با توجه به خروجی بالا می‌توان گفت این جدول از ایندکس قرار داده شده روی تاریخ برای یافتن اطلاعات موردنظر استفاده می‌کند

کوئری دوم:

```
select cityname
from maintable
where orderdate between '2013-00-00' and '2015-00-00'
group by cityname
having sum(quantity)>100;
```

خروجی:

	cityname
1	تهران
2	اصفهان
3	کرج

خروجی explain:

id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
1	SIMPLE	maintable	<null>	range	dateindex	dateindex	6	<null>	4568	100	Using index condition; Using MRR; Using temporary

که باز هم با توجه به خروجی بالا می توان گفت از ایندکس استفاده شده است

کوئری سوم:

```
select orderid
from maintable
where orderdate>'2016-00-00' and quantity>100;
```

خروجی:

	orderid
1	4152468
2	11842273

خروجی explain:

id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
1	SIMPLE	maintable	<null>	ALL	dateindex	<null>	<null>	<null>	199375	16.66	Using where

در این کوئری می‌توان دید که از ایندکس استفاده نشده است

کوئری چهارم:

```
select orderid
from maintable
where orderdate between '2017-01-23 00:00:00' and '2017-01-23 01:00:00' or
       orderdate between '2017-01-23 23:00:00' and '2017-01-23 23:59:59';
```

خروجی:

	orderid
1	6145677
2	6146910
3	6146539
4	6145888
5	6155231
6	6157087
7	6155352
8	6155365
9	6156857
10	6155439
11	6155450

خروجی explain:

id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
1	SIMPLE	maintable	<null>	range	dateindex	dateindex	6	<null>	11	100	Using index condition

با توجه به خروجی بالا می توان گفت که این کوئری هم از ایندکس موردنظر استفاده کرده است

کوئری پنجم:

```
select cityname,sum(quantity)
from maintable
where orderdate<'2015-00-00'
group by cityname
having sum(quantity)>50
order by sum(quantity) desc;
```

	cityname	`sum(quantity)`
1	تهران	2617
2	کرج	154
3	اصفهان	110
4	رشت	79
5	مشهد	79
6	اهواز	72
7	یزد	57
8	شیراز	53

خروجی:

خروجی explain:

id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
1	SIMPLE	maintable	<null>	range	dateindex	dateindex	6	<null>	4568	100	Using index condition; Using MRR; Using temporary;

و با توجه به خروجی بالا می‌توان گفت در این کوئری هم از ایندکس قرار داده شده روی تاریخ سفارشات استفاده شده است