

1)

A) [newmin, newmax]

B) [(old_minmean)/stddev,(old_maxmean)/stddev]

به طور کلی دامنه برای تمام دیتاست از منفی بی نهایت تا مثبت بی نهایت است.

C) [-1,1]

محدوده باید همانند دیگر Z_score ها باشد.

D) [-1,1]

2)

A) $Z_score = x_i - \min / \max - \min$

$$Z_s1 = 200 - 200 / 800 = 0$$

$$Z_s2 = 0.125$$

$$Z_s3 = 0.25$$

$$Z_s4 = 0.5$$

$$Z_s5 = 1$$

B) $\mu = 500 \quad \sigma = 282.25 \quad Z_score = x - \mu / \sigma$

$$Z_s1 = -1.0$$

$$Z_s2 = -0.7$$

$$Z_s3 = -0.3$$

$$Z_s4 = 0.3$$

$$Z_s5 = 1.7$$

C)

$$z_{if} = \frac{x_{if} - m_f}{s_f}$$

$$S = (300 + 200 + 100 + 100 + 500) / 5 = 1200 / 5 = 240$$

$$Z_{s1} = -1.25$$

$$Z_{s2} = -0.83$$

$$Z_{s3} = -0.41$$

$$Z_{s4} = 0.41$$

$$Z_{s5} = 2.08$$

D)

$$v1=200 / 10^3 = 0.2$$

$$v2=300 / 10^3 = 0.3$$

$$v3=400 / 10^3 = 0.4$$

$$v4=600 / 10^3 = 0.6$$

$$v5=1000 / 10^4 = 0.1$$

3)

A)

Bin 1: 5 , 10 , 11 , 13

Bin 2: 15 , 35 , 50 , 55

Bin 3: 72 , 92 , 204 , 215

B)

$$\text{width} = (215 - 5) / 3 = 70$$

bin 1: 5 , 10 , 11 , 13 , 15 , 35 , 50 , 55 , 72

bin 2: 92

bin 3: 204 , 215

C)

bin 1: 5 , 10 , 11 , 13 , 15

bin 2: 35 , 50 , 55 , 72 , 92

bin 3: 204 , 215

A) Snowflake schema, fact constellation, starnet query model

شمای snowflake و fact constellation هر دو متفاوت از شمای star هستند، که شامل یک جدول fact و یک مجموعه از جدول های بعد میباشد. شمای snowflake شامل تعدادی جدول نرمالایز شده ی بعد میباشد در حالیکه fact constellation شامل یک مجموعه از جدول های واقعی میباشد که تعدادی از جدول های بعد را به اشتراک میگذارد. Starnet query model یک مدل تحقیق و جستجو است نه یک مدل شمایی، که شامل یک مجموعه از خط های شعاعی و ستاره ای است که از یک نقطه مرکزی ساطع میشوند. هر خط شعاعی یک بعد را توصیف میکند و هر نقطه حول خط، تراز بعد را مشخص میکند. هر مرحله که در اطراف نقطه مرکزی است کم شدن مراحل از سلسله مراتب بعد را نشان میدهد. Starnet query model همانطور که از نامش پیداست برای جستجو و تحقیق و دادن یک دید کلی از عملگرهای olap به کاربران استفاده میشود.

B)

data cleaning : تمیز کردن داده ها

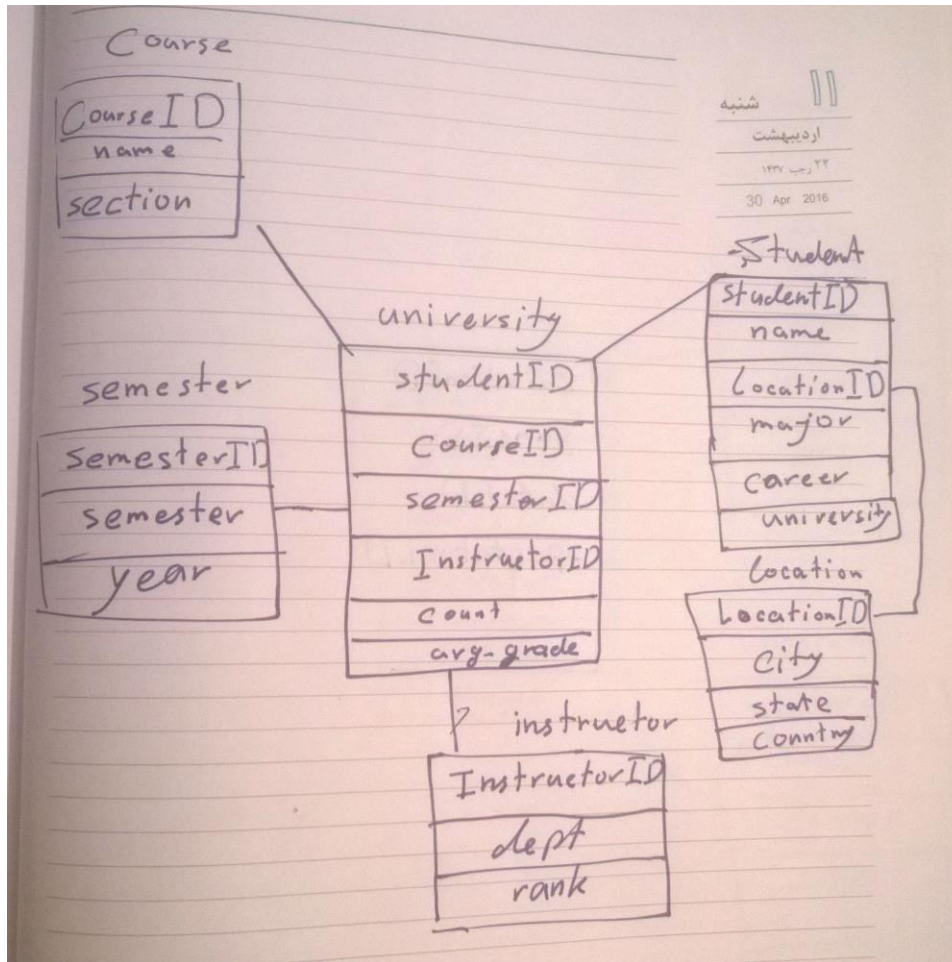
پروژه ی تشخیص خطاها در داده و تصحیح آنها تا جایی که ممکن است

Data transformation: انتقال داده ها

پروژه ی تبدیل داده از source های نا همگن به انبار داده ی یکپارچه است.

Refresh عملی است که باعث میشود اطلاعات موجود در انبار داده آپدیت شود.

5) A)



B)

Roll-up on course from courseID to section

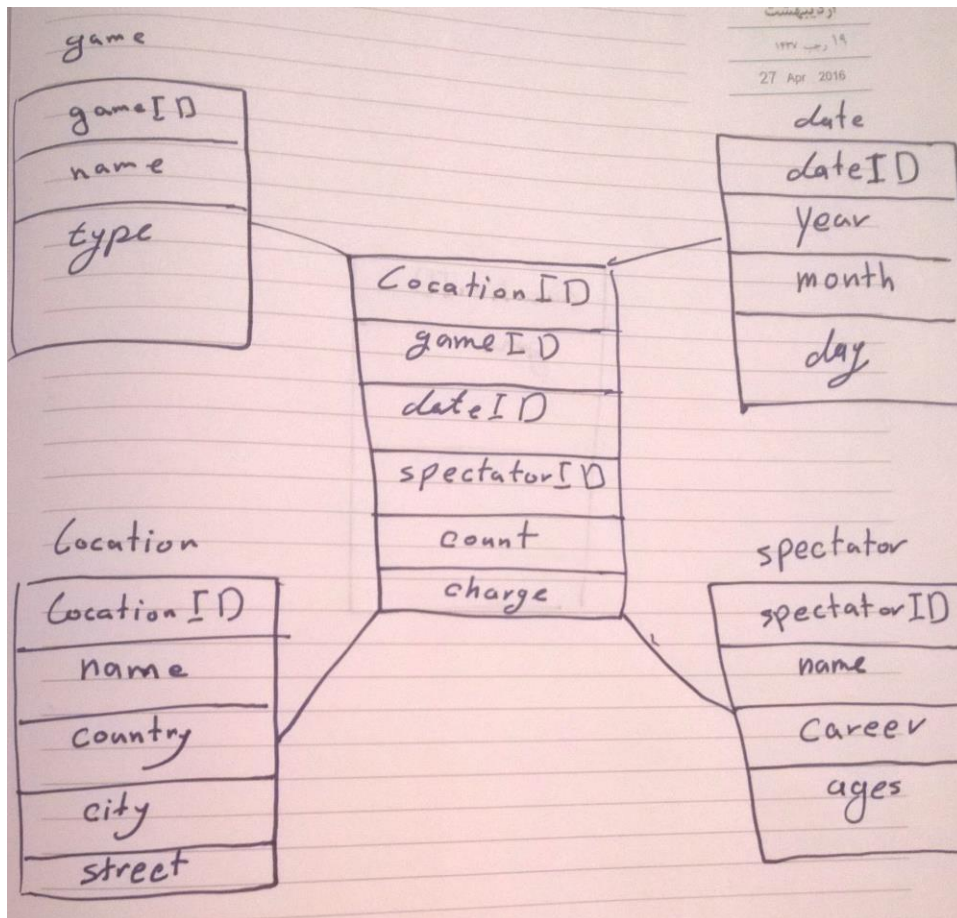
Roll-up on student from studentID to university

Dice on course, student with section=CS and university = Big University

Drill-down on student from university to name.

C) $5^4 = 625$ cuboids

6) A)



B)

Roll-up on date from dateID to year

Roll-up on game from gameID to all

Roll-up on location from locationID to name

Roll-up on spectator from spectatorID to career

Dice with career = students, name= GM_Place , and year=2010

C)

اگر در این مکعب، اگر location به عنوان bitmap index باشد. بعد مقایسه و پیوست و عملیات بر روی location به حساب بی‌تی کاهش پیدا کند که به صورت قابل ملاحظه ای زمان پردازش را کاهش می‌دهد.

علاوه بر این، رشته‌های طولانی نام محل را می‌توان با یک بیت نشان داد، که منجر به کاهش قابل توجهی در فضا و I/O می‌شود.

Bitmap index برای ابعاد با کاردینالیتی پایین سودمند است، ولی برای ابعاد با کاردینالیتی بالا مانند date در این مثال، می‌تواند بسیار طولانی باشد

7)

A) p^n

B) p

C) $(p + 1)^n$

D) $(2^n - 1) \times p + 1$