

به نام خدا

پاسخنامه تکلیف اول درس پایگاه داده ها ۱

ترم اول ۱۳۹۹ – ۱۴۰۰

۴- در فصل یکم درباره چندین فایده اصلی پایگاه داده صحبت شد. آیا برای استفاده از آن عیوبی نیز متصور است؟ اگر جوابتان مثبت است مختصراً شرح دهید.

راهنمایی: از متن کتابهای BK1, BK2 می توانید برای پاسخ به این سوال و برخی سوالات دیگر استفاده کنید.

پاسخ :

دو مورد از معایب پایگاه داده در زیر آمده است:

الف- نصب پایگاه داده نیازمند دانش، پول، مهارت و زمان بیشتر است.

ب- پیچیدگی پایگاه داده ممکن است باعث کارایی ضعیف شود.

۵- پنج مورد از کارکردهای مهم database administrator را نام ببرید.

پاسخ :

(۱) ایجاد scheme

(۲) تعریف ساختار حافظه و متودهای دسترسی

(۳) تغییر scheme و ساختار فیزیکی در صورت نیاز

(۴) اعطای مجوز برای دسترسی به داده ها

(۵) مشخص کردن integrity constraints

۶- با توجه به file structure زیر به سوالات پاسخ دهید.

PROJECT_CODE	PROJECT_MANAGER	MANAGER_PHONE	MANAGER_ADDRESS	PROJECT_BID_PRICE
21-5Z	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	16833460.00
25-2D	Jane D. Grant	615-898-9909	218 Clark Blvd., Nashville, TN 36362	12500000.00
25-5A	George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185	32512420.00
25-9T	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	21563234.00
27-4Q	George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185	10314545.00
29-2D	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	25559999.00
31-7P	William K. Moor	904-445-2719	216 Morton Rd., Stetson, FL 30155	56850000.00

الف) این فایل چند record دارد ؟ به ازای هر record چند field دارد؟

پاسخ :

فایل شامل ۷ رکورد است (21-5Z تا 31-7P) و هر رکورد شامل ۵ فیلد است. (PROJECT_CODE تا

PROJECT_BID_PRICE)

ب) اگر بخواهید روی رکوردهای فوق بر اساس شهر جستجو کنید با چه مشکلی روبه‌رو می‌شوید؟ چگونه این مشکل را با تغییر فایل فوق برطرف می‌کنید؟

پاسخ :

نام شهر ها در فیلد `MANAGER_ADDRESS` است و جدا کردن کاراکترهای آن در سطح `application` در بهترین حالت، سخت خواهد بود. نوشتن کوئری ها بسیار سخت خواهد شد و وقتی بخواهیم درون هر رشته را جستجو کنیم زمان انجام کوئری ها افزایش خواهد یافت. اگر حتما به لیست کردن شهرها نیاز داریم بهتر است شهر را به عنوان یک ستون جدا ذخیره کنیم.

ج) اگر بخواهید از این `file` یک لیست تهیه کنید که دربردارنده `last name, area code, city, state` یا `zip code` باشد چگونه فایل فوق را تغییر می‌دهید؟

پاسخ :

هر چه آدرس را به بخش های بیشتری تجزیه کنیم ، اطلاعات بیشتری را می توانیم استخراج کنیم. برای مثال تجزیه ی `MANAGER_ADDRESS` به `(MGR_STREET, MGR_CITY, MGR_STATE, MGR_ZIP)` امکان بازیابی هر یک از این اطلاعات را در موقع لزوم به ما می دهد.

د) چه `data redundancies` در فایل فوق وجود دارد؟ چگونه این `redundancies` منجر به ناهنجاری می‌شود؟

پاسخ :

توجه کنید که نام `Holly B. Parker` سه بار در جدول آمده است که بیانگر این است که او سه پروژه را مدیریت می کند. این تکرار ها بیانگر این است که بین مدیر و پروژه یک رابطه ی یک به چند برقرار است. یعنی یک پروژه فقط یک مدیر دارد ولی یک مدیر می تواند بیش از یک پروژه را مدیریت کند. این تکرار ها باعث شده آدرس و شماره ی این مدیر چند بار تکرار شود که این امر باعث می شود در صورتی که آدرس و شماره این مدیر عوض شود، همه فیلدهای این سه رکورد نیازمند تغییر باشد. اگر یکی از آنها تغییر نکند، اطلاعات متفاوت برای یک فرد داریم و در طول زمان ، `valid` بودن داده ها از بین می رود. همین موارد برای `George F. Dorts` نیز برقرار است.

۷- با توجه به file structure زیر به سوالات پاسخ دهید.

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CODE	JOB_CHG_HOUR	PROJ_HOURS	EMP_PHONE
1	Hurricane	101	John D. Newson	EE	85.00	13.3	653-234-3245
1	Hurricane	105	David F. Schwann	CT	60.00	16.2	653-234-1123
1	Hurricane	110	Anne R. Ramoras	CT	60.00	14.3	615-233-5568
2	Coast	101	John D. Newson	EE	85.00	19.8	653-234-3254
2	Coast	108	June H. Sattlemeier	EE	85.00	17.5	905-554-7812
3	Satellite	110	Anne R. Ramoras	CT	62.00	11.6	615-233-5568
3	Satellite	105	David F. Schwann	CT	26.00	23.4	653-234-1123
3	Satellite	123	Mary D. Chen	EE	85.00	19.1	615-233-5432
3	Satellite	112	Allecia R. Smith	BE	85.00	20.7	615-678-6879

الف (مشکل data redundancy را در فایل زیر بررسی کنید

پاسخ :

اگر JOB-CH-HOUR برای JOB_CODE = EE از ۸۵ به ۹۰ تغییر کند باید برای همه سطر ها این تغییر را اعمال کنیم. همچنین اگر June H. Sattlemeier از فایل حذف شود، همه اطلاعات دیگر فیلد های مربوط به رکورد او حذف می شود و به علت حذف PROJ_HOURS ، نهایتا پروژه ی coast از نظر پرداخت هزینه ها با مشکل مواجه می شود.

ب (با توجه به EMP_NAME و EMP_PHONE چه تغییر یا تغییراتی را پیشنهاد می دهید؟

پاسخ :

یک پیشنهاد خوب ، اتمیک کردن دیتا است . این یعنی تا حد ممکن اطلاعات را بشکنیم. مثلا EMP_NAME را به EMP_FNAME, EMP_INITIAL, EMP_LNAME بشکنیم. هم چنین می توان EMP_PHONE را به EMP_PHONE and EMP_AREACODE تجزیه کرد.

۸- با توجه به دو جدول داده شده (تصویر اول یک snapshot از جدول course و تصویر دوم مربوط به جدول section است) ، query زیر را مرحله به مرحله اجرا کنید (از داخلی ترین پرانتز) و جدول حاصل از اعمال هر مرحله را نشان دهید.

$$\Pi_{c.course_id, c.title} (\rho_s (\Pi_{course_id(section)} - \Pi_{course_id} (\sigma_{year=2018}(section))) \bowtie_{s.course_id=c.course_id} \rho_c (course))$$

	course_id [PK] character varying (8)	title character varying (50)	dept_name character varying (20)	credits numeric (2)
1	101	Diffusion and Phase Transfo...	Mech. Eng.	3
2	105	Image Processing	Astronomy	3
3	123	Differential Equations	Mech. Eng.	3
4	127	Thermodynamics	Geology	3
5	130	Differential Geometry	Physics	3
6	133	Antidisestablishmentarian...	Biology	4
7	137	Manufacturing	Finance	3
8	139	Number Theory	English	4
9	158	Elastic Structures	Cybernetics	3
10	169	Marine Mammals	Elec. Eng.	3

	course_id [PK] character varying (8)	sec_id [PK] character varying (8)	semester [PK] character varying (6)	year [PK] numeric (4)	building character varying (15)	room_number character varying (7)	time_slot_id character varying (4)
1	105	1	Fall		2009 Chandler	375	C
2	105	2	Fall		2002 Taylor	183	C
3	137	1	Spring		2002 Fairchild	145	I
4	158	1	Fall		2008 Whitman	434	F
5	158	2	Spring		2008 Taylor	812	D
6	169	1	Spring		2007 Gates	314	A
7	169	2	Fall		2002 Drown	757	L
8	192	1	Fall		2002 Polya	808	B
9	200	1	Spring		2007 Saucon	180	D
10	200	2	Fall		2002 Chandler	375	D

پاسخ :

از داخلی ترین پرانتز شروع می کنیم.

1. ابتدا $\sigma_{year=2018}(section)$: خروجی این عملیات تهی می شود زیرا در جدول section هیچ

رکوردی با سال 2018 وجود ندارد.

2. سپس $\Pi_{course_id}(\sigma_{year=2018}(section))$: از آنجایی که قسمت (1) خروجی نداشتیم ، در این

قسمت هم خروجی همچنان تهی است.

3. سپس $\Pi_{course_id}(section) - \Pi_{course_id}(\sigma_{year=2018}(section))$: خروجی این عملیات همان جدول

section است . پس :

	course_id [PK] character varying (8)
1	105
2	105
3	137
4	158
5	158
6	169
7	169
8	192
9	200
10	200

4. سپس $\rho_s (\Pi_{course_id}(section) - \Pi_{course_id}(\sigma_{year=2018}(section)))$: این صرفاً یک تغییر نام است

یعنی :

	course_id [PK] character varying (8)
1	105
2	105
3	137
4	158
5	158
6	169
7	169
8	192
9	200
10	200

Tabel S

5. سپس ρ_c (course) : این عمل هم تغییر نام است پس :

	course_id [PK] character varying (8)	title character varying (50)	dept_name character varying (20)	credits numeric (2)
1	101	Diffusion and Phase Transfo...	Mech. Eng.	3
2	105	Image Processing	Astronomy	3
3	123	Differential Equations	Mech. Eng.	3
4	127	Thermodynamics	Geology	3
5	130	Differential Geometry	Physics	3
6	133	Antidisestablishmentarianis...	Biology	4
7	137	Manufacturing	Finance	3
8	139	Number Theory	English	4
9	158	Elastic Structures	Cybernetics	3
10	169	Marine Mammals	Elec. Eng.	3

Table C

6. سپس ρ_c (course) $\bowtie_{s.course_id=c.course_id} (\sigma_{year=2018}(\Pi_{course_id}(\Pi_{course_id(section)})))$: ρ_s

این عملیات جوی دو جدول S و C است . پس :

	course_id character varying (8)	course_id character varying (8)	title character varying (50)	dept_name character varying (20)	credits numeric (2)
1	105	105	Image Processing	Astronomy	3
2	137	137	Manufacturing	Finance	3
3	158	158	Elastic Structures	Cybernetics	3
4	169	169	Marine Mammals	Elec. Eng.	3

7. در نهایت خروجی کل کوئری داده شده :

	course_id character varying (8)	title character varying (50)
1	105	Image Processing
2	137	Manufacturing
3	158	Elastic Structures
4	169	Marine Mammals

* توجه داشته باشید که در جبر رابطه ای ، سطور یکسان از نتیجه نهایی حذف می شوند.

۹- با توجه به جدول time_slot در پایگاه داده دانشگاه که در کلاس معرفی شده است توضیح دهید که چرا

time_slot_id و day و start_day به عنوان primary key در نظر گرفته شده اند.

پاسخ :

ویژگی های day و start_time بخشی از کلید اصلی هستند زیرا هر کلاس به احتمال زیاد در چندین

روز متفاوت و حتی ممکن است بیشتر از یک بار در روز برگزار شود.

۱۰- دو دلیل برای اینکه چرا مقادیر NULL باید در پایگاه داده معرفی شوند را بیان کنید.

پاسخ :

برای زمانی که اطلاعاتی درباره فیلد مربوطه نداریم ، وجود این نوع لازم است . مثلاً وقتی درباره حقوق یک استاد اطلاعاتی نداریم نباید برای آن مقدار صفر تخصیص دهیم زیرا حقوق او صفر نیست و ما فقط نمیدانیم مقدارش چقدر است پس باید چیزی مثل null داشته باشیم.
همچنین برای نشان دادن missing values هم از مقدار NULL استفاده می‌شود.

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

۱۱- در جدول رو به رو ، هیچ دو استادی نام یکسانی

ندارند. آیا از این گزاره می‌توان نتیجه گرفت که

name میتواند یک super key یا primary key

برای جدول instructor باشد ؟

پاسخ :

خیر ، برای این مثال خاص نام همه ی استادان

منحصر به فرد است وای در کل ممکن است این

شرایط برقرار نباشد.

۱۲- با توجه به Banking Database که در زیر آمده است کوئری‌های زیر را به زبان جبرخطی بنویسید :

branch(branch_name, branch_city, assets)

customer (customer_name, customer_street, customer_city)

loan (loan_number, branch_name, amount)

borrower (customer_name, loan_number)

account (account_number, branch_name, balance)

depositor (customer_name, account_number)

الف (primary key را برای هر جدول مشخص کنید.

پاسخ :

Tabel	Primary Key
branch	branch_name
customer	customer_name
loan	loan_number
borrower	loan_number
account	account_number
depositor	customer_name , account_number

چندین نکته :

۱- فرض بر این است که مشتری یا شعبه ای با نام همسان نخواهیم داشت (در طراحی اصولی این فرض غلط است و باید از فیلدی مانند ID در این مواقع استفاده شود ولی در این مسئله مجازیم به داشتن این فرض)

۲- هر وام توسط فقط و فقط یک نفر گرفته می شود ولی هر شخص می تواند چندین وام بگیرد.

۳- هر حساب می تواند چندین صاحب داشته باشد (حساب های مشترک) و هر شخص می تواند چندین حساب داشته باشد.

ب) با توجه به primary key هایی که در قسمت (الف) مشخص کردید ، foreign key های مناسب را مشخص کنید.

پاسخ :

Tabel	Foreign Key
branch	None
customer	None
loan	branch_name → branch
borrower	Customer_name → customer Loan_number → loan
account	branch_name → branch
depositor	Customer_name → customer Account_number → account

* کلید اصلی یک جدول می تواند کلید خارجی نیز باشد (رجوع به schema diagram دیتابیس دانشگاه و مثلاً جدول advisor)

ج) نام تمام شعبه هایی که در "Chicago" قرار دارند را پیدا کنید.

پاسخ :

$\Pi_{\text{branch_name}} (\sigma_{\text{branch_city}=\text{"Chicago"}} (\text{branch}))$

د) تمام loan_number ها برای وام های بیشتر از \$10,000 را پیدا کنید.

پاسخ :

$\Pi_{\text{loan_number}} (\sigma_{\text{amount}>10000} (\text{loan}))$

۱۳- درباره ی مفاهیم زیر در حوزه ی Advanced Data Modeling تحقیق کنید و برای هر کدام در حد چند خط تعریفی ارائه دهید.

1. Entity Supertypes
2. Entity Subtypes
3. Subtype Discriminator
4. Entity Clustering

پاسخ :

* موارد a و b :

entity supertype ، یک entity از نوع عمومی است ، که با یک یا چند entity subtype مرتبط است. با ذکر یک مثال این موضوع را بیشتر بررسی میکنیم . فرض کنید در یک سازمان کارمندان مختلفی استخدام میشوند که هر کدام مهارت های متفاوتی دارند (مثلا خلبان، مکانیک و قرار دادن این کارمندان در جدول employee منجر به تولید فیلدهایی با مقادیر null می شود . چرا که مثلا در فیلد مربوط به license ها ، همه کارمندان یک نوع گواهی ندارند و در رکورد آنها مجبور به استفاده از null میشویم. هم چنین جدا کردن این نوع کارمندان از یک دیگر و قرار دادنشان در جداول مختلف باعث میشود آنها بتوانند در ارتباطی که صرفا مخصوص خودشان است شرکت کنند . پس در این سازمان employee یک supertype بوده و mechanic و pilot در واقع subtype هستند .

* مورد c :

Subtype Discriminator به attribute ای گفته می شود که subtype و supertype از طریق آن با هم دیگر ارتباط دارند. مثلا در مثال قبل ، (EMP_TYPE) به عنوان Subtype Discriminator شناخته می شود.

* مورد d :

entity clustering یک entity type مجازی است که برای نمایش چند entity و روابط بین آنها در ERD استفاده میشود.

۱۴- شش مورد از ویژگی های مطلوب برای primary key را نام برده و هر کدام را مختصرا توضیح دهید.

پاسخ :

Unique values

Nonintelligent

No change over time

Preferably single-attribute

Preferably numeric

Security-compliant

برای توضیحات بیشتر هر مورد به جدول ۵.۳ صفحه ۱۷۸ از BK2 مراجعه کنید.

۱۵- در چه مواقعی از composite primary key استفاده می کنیم؟

پاسخ :

برای composite entity ها و weak entity ها. در (در کتاب BK2 صفحه ی ۱۷۷ ، مورد 3c-5) توضیحات کامل آورده شده است.

۱۶- Surrogate primary key را تعریف کرده و کاربرد آن را توضیح دهید.

پاسخ :

گاهی برای یک entity در دنیای واقعی PK وجود ندارد. surrogate key یک PK است که توسط طراح دیتابیس ایجاد میشود. برای توضیحات بیشتر به صفحه ی ۱۷۹ مورد 3d-5 مراجعه کنید.