

به نام خدا

## تکلیف اول درس پایگاه داده‌ها ۱

ترم مهر ۹۹

فرمت ارسال : کل پاسخ ها را در قالب یک فایل pdf روی سامانه قرار دهید. نام فایل با شماره دانشجویی شروع شود و سپس فامیل (انگلیسی). مثال : 9511113-Irani.pdf  
نحوه ارسال روی سامانه : وارد سامانه [yekta.iut.ac.ir](http://yekta.iut.ac.ir) شوید و فایل پاسخنامه خود را در قسمت تکلیف اول قرار دهید.  
توجه: ارسال مستقیم با ایمیل و پیام‌رسان به استاد یا همکاران درس نادیده گرفته خواهد شد.

مهلت ارسال پاسخ: جمعه ۱۱ مهر ۹۹، ساعت ۱۳:۰۰

(تحويل صرفاً از طریق سامانه انجام می‌شود و در پایان ترم نیز نمرات از روی سامانه به گلستان منتقل می‌شود و ارسال دستی قابل ورود نیست)

۱) روی مفاهیم بورس تهران، سجام، کارگزاریهای بورس، کد بورسی تحقیق کنید و گزارش دهید که (الف) هر کدام چه پایگاه داده ای شامل چه اطلاعاتی از کاربران و شرکتها را دارند. برای هر پایگاه لیستی از موجودیتها و خصوصیات اصلی هر موجودیت را در گزارش بیاورید. (ب) این پایگاهها چه ارتباطاتی با یکدیگر دارند؟

۲) روی دو مورد از مشاغل مطرح در جدول ۱.۳ از فصل اول کتاب BK2 به انتخاب خودتان جستجو کنید و گزارش دهید که (الف) هر کدام از این دو شغل مرتبط با پایگاه به چه مهارتهایی نیاز دارند؟ (ب) میانگین حقوق آنها در کشورهای مختلف چه حدود است؟ راهنمایی: میتوانید از سایتهایی مانند LinkedIn و همچنین سایتهای مختلف job finder استفاده کنید.

۳) مدل سازی به چه معناست؟ فرض کنید میخواهید کاربران سامانه یکپارچه بهداشت (سیب) را مدلسازی کنید، چه مواردی برای مدلسازی این افراد لازم است؟  
راهنمایی: روی مفهوم پرونده الکترونیک سلامت (Electronic Health Record-EHR) و ساختار اطلاعاتی آن جستجو کنید و مشخص نمایید مهمترین موارد اطلاعاتی که درباره سلامت یک فرد در این نوع سامانه ها ثبت می شود شامل چه مواردی است.

۴) در فصل اول 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.

۵) پنج مورد از کارکردهای مهم Database Administrator را نام ببرید.

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؟

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؟

ج) اگر بخواهید از این file یک لیست تهیه کنید که دربردارنده‌ی last name, area code, city,

state, یا zip code باشد چگونه فایل فوق را تغییر می‌دهید؟

د) چه نوع افزونگی داده (data redundancy) در فایل فوق وجود دارد؟ چگونه این افزونگیها منجر به ناهنجاری (anomaly) می‌شود؟

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. Sattlemeir	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 افزونگی داده را در فایل فوق بررسی کنید و موارد را در صورت وجود بنویسید.

ب) با توجه به ساختار فیلدهای EMP\_PHONE و EMP\_NAME چه تغییر یا تغییراتی را پیشنهاد می‌دهید که بتوان اطلاعات داخل آنها را برای پاسخ به نیازها بصورت بهتر ذخیره 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.

$$\prod_{c.course\_id, c.title} (p_s (\prod_{course\_id}(section) - \prod_{course\_id}(\sigma_{year=2018}(section)))) \bowtie_{s.course\_id=c.course\_id} p_c (course))$$

	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	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

	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

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.

۱۰) دو دلیل برای اینکه چرا مقادیر NULL باید در پایگاه داده مورد توجه قرار گیرند را بیان کنید.

۱۱) در جدول زیر، هیچ دو استادی نام یکسانی ندارند. آیا از این گزاره می توان نتیجه گرفت که name میتواند یک super key یا primary key برای جدول instructor باشد؟

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

۱۲) با توجه به ساختار Banking Database که در زیر آمده است ، query های زیر را به زبان algebra relation 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)
```

Figure 2.15 Banking database

الف) primary key را برای هر جدول مشخص کنید.

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

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

د) تمام loan\_number ها برای وام های بیشتر از \$۱۰,۰۰۰ را پیدا کنید.

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

- a. Entity Supertypes
- b. Entity Subtypes
- c. Subtype Discriminator
- d. Entity Clustering

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

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

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