

**اسم واسم خانوادگی: عبدالقدیر فرتاش شماره دانشجویی: 99243100**

**تمرین سری 6**

**جواب مسئاله1)   
بخش الف)**

### ****Decomposition Join-Lossless چیست؟****

فرآیندی است که در آن یک جدول به جداول کوچک‌تر تقسیم می‌شود، به‌طوری که با **Join** کردن این جداول، جدول اصلی بدون از دست دادن داده‌ها بازسازی شود.

### ****چرا در طراحی پایگاه داده مهم است؟****

1. **حفظ داده‌ها:** مانع از گم‌شدن یا نادرست بازسازی شدن اطلاعات می‌شود.
2. **کاهش افزونگی:** داده‌های تکراری حذف می‌شوند و سازگاری حفظ می‌شود.
3. **بهبود عملکرد:** جداول کوچک‌تر، کوئری‌ها را سریع‌تر و کارآمدتر می‌کنند.
4. **حفظ وابستگی‌ها:** قوانین و محدودیت‌های رابطه اصلی باقی می‌مانند.

**بخش ب)**

**F+ چیست؟**

* **مجموعه بسته وابستگی های تابعی:** (Closure of a Set of Functional Dependencies) F+ مجموعه ای از تمام وابستگی های تابعی است که از مجموعه اولیه F قابل استنتاج است. به عبارت دیگر، اگر یک وابستگی تابعی X→Y در F+ وجود دارد، به این معنی است که می‌توان این وابستگی را از وابستگی‌های موجود در F با استفاده از قوانین ارفاق (Armstrong’s axioms) استنباط کرد.

**دو کاربرد مهم F+:**

1. **نرمال‌سازی پایگاه داده:** با استفاده از F+ می‌توانیم وابستگی‌های تابعی را شناسایی کنیم که باعث نقض نرمال‌سازی می‌شوند. این به ما کمک می‌کند تا جداول را به گونه‌ای طراحی کنیم که از تکرار داده‌ها و ناهنجاری‌ها جلوگیری شود.
2. **بهینه‌سازی پرس‌وجوها:** با دانستن F+ می‌توانیم پرس‌وجوهایی را بهینه کنیم. برای مثال، اگر بدانیم که یک وابستگی تابعی X→Y وجود دارد، می‌توانیم از این وابستگی برای ساده‌سازی پرس‌جو استفاده کنیم.

**بخش ج)**

## چرا BCNF همیشه تمام وابستگی‌ها را حفظ نمی‌کند؟

شکل نرمال BCNF (Boyce-Codd Normal Form) به گونه‌ای طراحی شده است که وابستگی‌های تابعی را حذف کند تا افزونگی داده‌ها کاهش یابد و یکپارچگی داده‌ها بهبود یابد. اما در برخی موارد، ممکن است تمام وابستگی‌ها حفظ نشوند. دلایل اصلی این موضوع عبارتند از:

* **وابستگی‌های چندگانه**: اگر در یک رابطه وابستگی‌های چندگانه وجود داشته باشد، ممکن است نتوانیم تمام وابستگی‌ها را حفظ کنیم.
* **کلیدهای ترکیبی**: در برخی موارد، کلیدهای ترکیبی ممکن است باعث شوند که نتوانیم تمام وابستگی‌ها را حفظ کنیم.
* **پیچیدگی ساختار داده‌ها**: در برخی موارد، ساختار داده‌ها به گونه‌ای است که حفظ تمام وابستگی‌ها ممکن است بسیار پیچیده و زمان‌بر باشد..

**جواب مسئاله2)**

**بخش الف)**

#### وابستگی‌های اولیه:

AB → C

C → A

BC → D

ACD → B

BE → C

CE → GA

CG → BD

D → EG

#### مرحله 1: حذف اتصالات اضافی از سمت راست **(RHS)**:

**نتیجه بعد از این مرحله:**

AB → C

C → A

BC → D

ACD → B

BE → C

CE → G

CE → A

CG → B

CG → D

D → E

D → G

#### مرحله 2: حذف اتصالات اضافی از سمت چپ **LHS**

#### 



هر FD را بررسی می‌کنیم تا ببینیم آیا می‌توان برخی از مقادیر سمت چپ را حذف کرد:

1. AB → C امکان حذف B یا A وجود ندارد.
2. C → Aنمی‌توان حذف کرد
3. BC → D امکان حذف B یا C وجود ندارد.
4. (CD)+ = {CDEG**B**..} ACD → Bپارامتر A زاید است امکان حذف وجود دارد. پس حذف می کنیم.
5. BE→C امکان حذف وجود ندارد
6. CE→G امکان حذف وجود ندارد.
7. CE→A امکان حذف این رابطه وجود دارد چونA C→ را میدهد
8. CG→B وجود دارد
9. بقیه حذف شان وجود ندارد.

**نتیجه نهایی:**

AB → C

C → A

BC → D

CD → B

BE → C

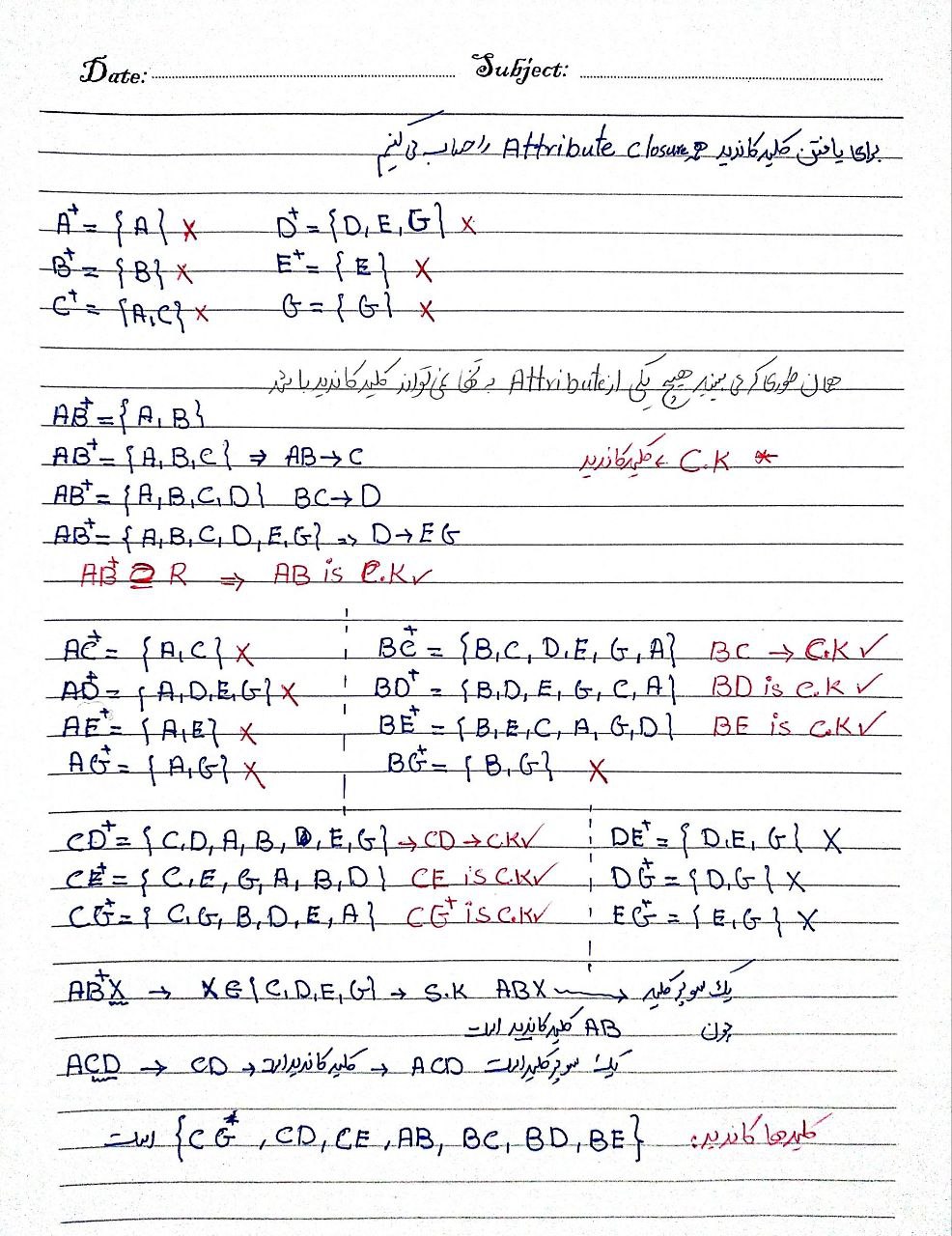
CE → G

CG → D

D → E

D → G

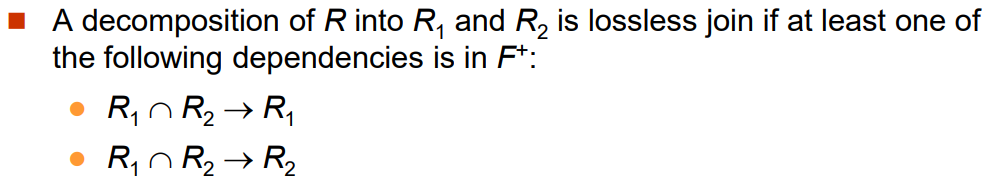
**بخش ب)**



**جواب مسئاله3)**

**شرط Lossless Join :** برای اطمینان از **lossless decomposition**، باید بخش مشترک بین روابط تجزیه شده یک **سوپرکلید** برای حداقل یکی از روابط باشد.

به عبارت دیگر:

* ستون مشترک باید بتواند تمام اطلاعات هر رابطه را بازسازی کند.  
  

A برای رابطه R1 سوپر کلید است و چون را میدهد و پس با استفاده از E میتواند به همه ستون ها دسترسی پیدا کرده پس E یک کلید کاندید است.

ستون A بین دو رابطه R1(A,D,E) وR2(A,B,C) مشترک است. از آنجایی کهA یک سوپرکلید برای کل رابطه R است، پس نتیجه می گیریم که این تجزیه **lossless** خواهد بود.

**جواب مسئاله4)**

برای تمام حالات FD را برسی میکنیم.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| EMP\_ID | EMP\_COUNTRY | EMP\_DEPT | DEPT\_TYPE | EMP\_DEPT\_NO |
| 264 | India | Designing | D394 | 283 |
| 264 | India | Testing | D394 | 300 |
| 364 | UK | Stores | D283 | 232 |
| 364 | UK | Developing | D283 | 549 |

**جواب مسئاله5)**

**بخش الف)**

جدول اصلی:

### ****مرحله 1: نرمال‌سازی به 1NF****

برای تبدیل به **1NF** باید مطمئن شویم که تمام مقادیر جدول اتمیک (Atomic) باشند.

* جدول ارائه‌شده اتمیک است، زیرا هیچ ستونی شامل مجموعه‌ای از مقادیر یا لیستی از داده‌ها نیست
* این جدول در **1NF** است.

### ****مرحله 2: نرمال‌سازی به 2NF****

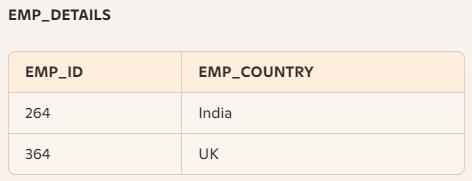
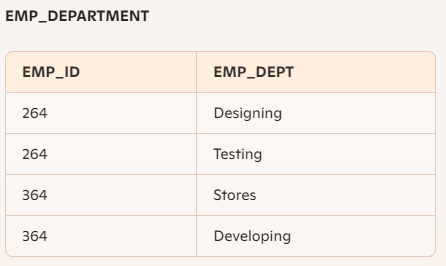
برای **2NF** باید شرایط زیر برقرار باشد:

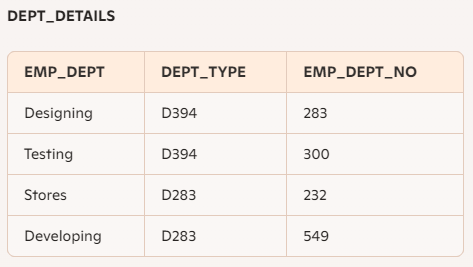
1. جدول در **1NF** باشد.
2. تمام ستون‌های غیر کلیدی، به کلید اصلی وابستگی کامل داشته باشند (نه وابستگی جزیی).

#### کلید اصلی پیشنهادی: برای این جدول، کلید اصلی ترکیبی می‌تواند **(EMP\_ID, EMP\_DEPT)** باشد، زیرا این ترکیب تمام ردیف‌ها را یکتا می‌کند.

#### بررسی وابستگی‌ها:

* **EMP\_COUNTRY** فقط به **EMP\_ID** وابسته است، نه ترکیب کامل کلید.
* **DEPT\_TYPE** و **EMP\_DEPT\_NO** فقط به **EMP\_DEPT** وابسته هستند.

**تفکیک جدول برای 2NF:**

****

### ****مرحله 3: نرمال‌سازی به3NF****

برای **3NF** باید:

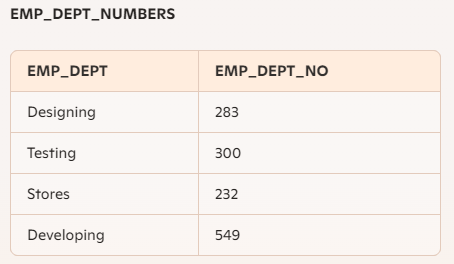
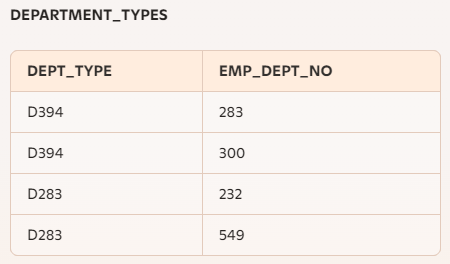
1. جدول در **2NF** باشد.
2. هیچ وابستگی متعددی (Transitive Dependency) بین ستون‌ها وجود نداشته باشد.

#### بررسی وابستگی‌ها:

* در جدول **DEPT\_DETAILS**، **EMP\_DEPT\_NO** به **DEPT\_TYPE** وابسته است. این وابستگی باید حذف شود.

تفکیک جدول برای NF3:

1. تفکیک اطلاعات دپارتمان:



اکنون جدول در **3NF** است

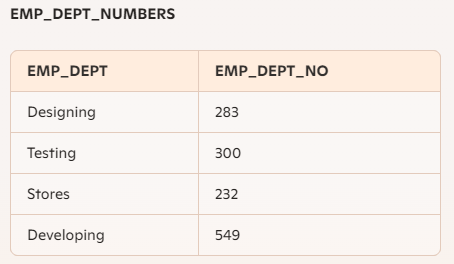
### ****مرحله 4: نرمال‌سازی به BCNF****

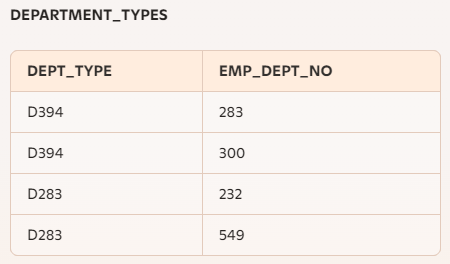
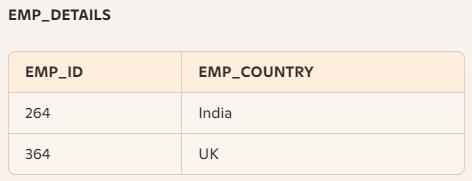
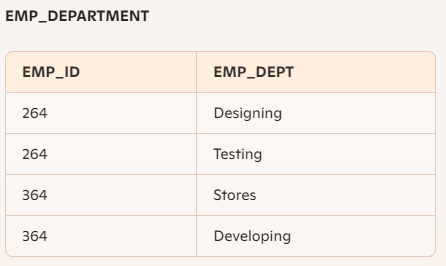
برای **BCNF** باید:

1. جدول در **3NF** باشد.
2. هر وابستگی تابعی در جدول، شرط **کلید بودن** داشته باشد.

#### بررسی وابستگی‌ها:

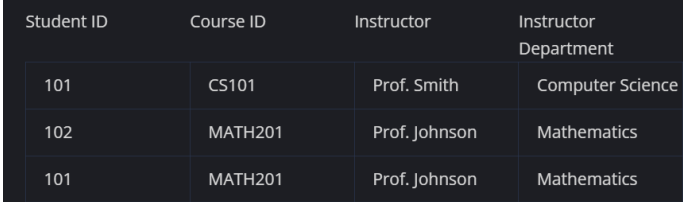
* در جدول **وابستگی نداریم پس اکنون جدول در BCNF است.**

**نتایج نهایی جداول نرمال شده:**

****

**بخش ب)**

**جدول رابطه ها)**

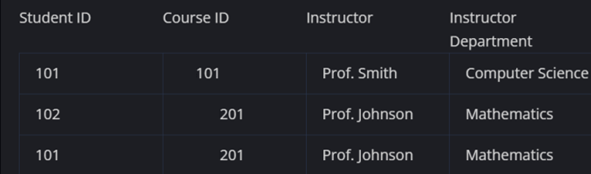
****

**مرحله 1: شناسایی کلید اصلی**

* **کلید اصلی:** ترکیبی از ستون‌های Student ID و Course ID است. زیرا برای شناسایی منحصر به فرد هر ردیف، به هر دو ستون نیاز داریم.

**مرحله 2: بررسی فرم نرمال اول (1NF)**

* **تکرار مقادیر:** در این جدول، مقادیر تکرار شده‌ای وجود ندارد.
* **اتمی بودن مقادیر:** اگر به ستون Course ID نگاه کنیم مقادیر اتمیک نیست چون از دو بخش تشکیل شده است یکی اسم دانشکده و دیگری کد درس پس باید این تجزیه شود چون اسم دانشکده ها داریم نیاز نیست CS, MATH را در جدول داشته باشیم.
* **هیچ ستونی حاوی چندین مقدار نیست:** هر سلول تنها یک مقدار را نگه می‌دارد.



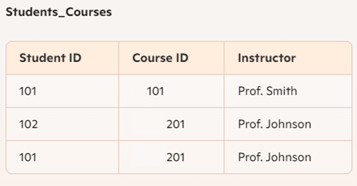
* **نتیجه:** حالا جدول در فرم نرمال اول(1NF) قرار دارد.

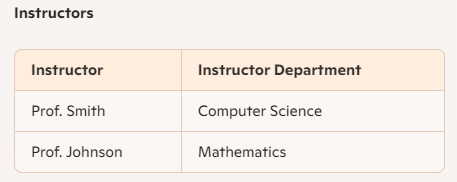
**مرحله 3: بررسی فرم نرمال دوم (2NF)**

* **وابستگی جزئی:** ستون‌های Instructor و Instructor Department فقط به بخشی از کلید اصلی (Instructor) وابسته هستند و به کلید اصلی کامل وابسته نیستند.
* **نتیجه:** جدول در فرم نرمال دوم قرار ندارد.

**تبدیل به فرم نرمال دوم (2NF)**

برای تبدیل جدول به فرم نرمال دوم، جدول را به دو جدول تقسیم می‌کنیم:

* **جدول اول (Students\_Courses):**
  + Student ID
  + Course ID
  + Instructor
* **جدول دوم (Instructors):**
  + Instructor
  + Instructor Department

**مرحله 5: بررسی فرم نرمال سوم (3NF)**

* **وابستگی گذرا:** در جدول Instructors وابستگی گذرایی وجود ندارد. همه ستون‌ها به کلید اصلی وابسته هستند.
* **نتیجه:** هر دو جدول در فرم نرمال سوم قرار دارند.

**نتیجه نهایی:**

با تقسیم جدول اصلی به دو جدول Students\_Courses و Instructors، جدول‌ها به فرم نرمال سوم رسیده‌اند.

**بخش ج)**

**مرحله 1: تبدیل به شکل 1NF و شناسایی کلیدهای کاندید)** ابتدا باید تابعی بستگی را به شکل 1NF تبدیل کرده و کلیدهای کاندید برای رابطه را شناسایی کنیم. پس داریم:

* AB→C
* AB→D
* D→C
* DE→B
* DEH→A
* DEH→B
* AC→D
* AC→D

از DEH→AB می‌توان نتیجه گرفت که DEHیک کلید کاندید است زیرا تمام ویژگی‌های دیگر را تعیین می‌کند.

### ****مرحله 2: نرمال‌سازی به 2NF)**** برای ****2NF**** باید شرایط زیر برقرار باشد:

1. جدول در **1NF** باشد.
2. تمام ستون‌های غیر کلیدی، به کلید اصلی وابستگی کامل داشته باشند (نه وابستگی جزیی).

برای این جدول، کلید اصلی**(**DEH**)**است.

#### بررسی وابستگی‌ها:

* D→C یعنی Cوابسته است بهD ، به جز از کلید اصلی وابسته است نه به همه کلید اصلی.
* DE→B یعنی Bوابسته است بهDE ، به جز از کلید اصلی وابسته است نه به همه کلید اصلی.

پس رابطه های زیر داریم:

اکنون جدول در**2NF**  است

### ****مرحله 3: نرمال‌سازی به3NF**** :

1. جدول در **2NF** باشد.
2. هیچ وابستگی متعددی (Transitive Dependency) بین ستون‌ها وجود نداشته باشد

برای NF3، باید تمام ستون‌های غیرکلیدی به کلید اصلی وابستگی مستقیم داشته باشند (نه از طریق وابستگی متعددی).

#### وابستگی متعددی: AB→CD

* ستون‌های C و D از طریق ABتعیین می‌شوند، اما این ترکیب وابستگی‌های متعددی ایجاد می‌کند. باید تجزیه شود.