

شماره دانشجو: 99243100

اسم و اسم خانوادگی: عبدالقدیر فرتاش

تمرین سری 6

جواب مسئله (1)
بخش الف

Decomposition Join-Lossless چیست؟

فرآیندی است که در آن یک جدول به جداول کوچکتر تقسیم می‌شود، به طوری 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 گم شدن 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 وابستگی‌های تابعی (Closure of a Set of Functional Dependencies) F^+ : مجموعه‌ای از تمام وابستگی‌های تابعی است که از مجموعه اولیه F قابل استنتاج است. به عبارت دیگر، اگر یک وابستگی تابعی $X \rightarrow Y$ 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 $X \rightarrow Y$ وجود دارد، می‌توانیم از این وابستگی برای ساده‌سازی پرس‌جو استفاده کنیم.

بخش ج)

چرا BCNF همیشه تمام وابستگی‌ها را حفظ نمی‌کند؟

شکل نرمال (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 مسئله 2)

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

Step 1: compute $(\{\alpha\} - A)^+$ using the dependencies in F

Step 2: check that $(\{\alpha\} - A)^+$ contains A ; if it does, A is extraneous

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

1. $AB \rightarrow C$ امکان حذف B یا A وجود ندارد.
2. $C \rightarrow A$ نمی‌توان حذف کرد
3. $BC \rightarrow D$ امکان حذف B یا C وجود ندارد.
4. $ACD \rightarrow B$ $\{CDEGB.. \} = (CD)^+$ پارامتر A زاید است امکان حذف وجود دارد. پس حذف می‌کنیم.
5. $BE \rightarrow C$ امکان حذف وجود ندارد
6. $CE \rightarrow G$ امکان حذف وجود ندارد.
7. $CE \rightarrow A$ امکان حذف این رابطه وجود دارد چون $A \rightarrow C$ را میدهد
8. $CG \rightarrow B$ وجود دارد
9. بقیه حذف شان وجود ندارد.

نتیجه نهایی:

$AB \rightarrow C$
 $C \rightarrow A$
 $BC \rightarrow D$
 $CD \rightarrow B$
 $BE \rightarrow C$
 $CE \rightarrow G$
 $CG \rightarrow D$
 $D \rightarrow E$
 $D \rightarrow G$

برای یافتن کلیدهای کاندید در Attribute closure را حساب می‌کنیم

$$A^+ = \{A\} \times \quad D^+ = \{D, E, G\} \times$$

$$B^+ = \{B\} \times \quad E^+ = \{E\} \times$$

$$C^+ = \{A, C\} \times \quad G^+ = \{G\} \times$$

همان طریقی که می‌بینید هیچ یکی از Attribute نمی‌تواند کلید کاندید باشد

$$AB^+ = \{A, B\}$$

$$AB^+ = \{A, B, C\} \Rightarrow AB \rightarrow C$$

* C.K کلید کاندید

$$AB^+ = \{A, B, C, D\} \quad BC \rightarrow D$$

$$AB^+ = \{A, B, C, D, E, G\} \Rightarrow D \rightarrow EG$$

$$AB^+ \supseteq R \Rightarrow AB \text{ is C.K.} \checkmark$$

$$AC^+ = \{A, C\} \times$$

$$BC^+ = \{B, C, D, E, G, A\} \quad BC \rightarrow C.K. \checkmark$$

$$AD^+ = \{A, D, E, G\} \times$$

$$BD^+ = \{B, D, E, G, C, A\} \quad BD \text{ is C.K.} \checkmark$$

$$AE^+ = \{A, E\} \times$$

$$BE^+ = \{B, E, C, A, G, D\} \quad BE \text{ is C.K.} \checkmark$$

$$AG^+ = \{A, G\} \times$$

$$BG^+ = \{B, G\} \times$$

$$CD^+ = \{C, D, A, B, E, G\} \Rightarrow CD \rightarrow C.K. \checkmark$$

$$DE^+ = \{D, E, G\} \times$$

$$CE^+ = \{C, E, G, A, B, D\} \quad CE \text{ is C.K.} \checkmark$$

$$DG^+ = \{D, G\} \times$$

$$CG^+ = \{C, G, B, D, E, A\} \quad CG^+ \text{ is C.K.} \checkmark$$

$$EG^+ = \{E, G\} \times$$

$$ABX \rightarrow X \in \{C, D, E, G\} \rightarrow S.K. \quad ABX \rightarrow \text{یک سوپر کلید}$$

چون AB کلید کاندید است

$$ACD \rightarrow CD \rightarrow \text{کلید کاندید است} \rightarrow ACD \text{ یک سوپر کلید است}$$

$$\{CG^+, CD, CE, AB, BC, BD, BE\} \text{ کلیدها کاندید است}$$

جواب مسئله 3)

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 decomposition of R into R_1 and R_2 is lossless join if at least one of the following dependencies is in F^+ :
 - $R_1 \cap R_2 \rightarrow R_1$
 - $R_1 \cap R_2 \rightarrow R_2$

$$R_1(A, B, C), \quad R_2(A, D, E)$$

$$R_1 \cap R_2 = \{A\} \text{ and } A \rightarrow ABC$$

$$A \rightarrow BC, \quad A \rightarrow B, \quad A \rightarrow C$$

$$A \rightarrow B, \quad B \rightarrow D, \quad A \rightarrow D$$

A برای رابطه R_1 سوپرکلید است و چون $E \rightarrow A$ را میدهد و $A \rightarrow B, A \rightarrow C, A \rightarrow D$ پس با استفاده از E میتواند به همه ستون ها دسترسی پیدا کرده پس E یک کلید کاندید است.

ستون A بین دو رابطه $R_1(A, D, E)$ و $R_2(A, B, C)$ مشترک است. از آنجایی که A یک سوپرکلید برای کل رابطه R است، پس نتیجه می گیریم که این تجزیه **lossless** خواهد بود.

جواب مسئله 4)

برای تمام حالات FD را بررسی میکنیم.

$A \rightarrow B$	okey
$A \rightarrow C$	not okey
$B \rightarrow A$	not okey
$B \rightarrow C$	not okey
$C \rightarrow A$	not okey
$C \rightarrow B$	okey
$AB \rightarrow C$	not okey
$AC \rightarrow B$	okey
$BC \rightarrow A$	not okey

final answer:

$$A \rightarrow B, \quad C \rightarrow B, \quad AC \rightarrow B$$

جواب مسأله 5)

بخش الف)

جدول اصلی:

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

مرحله 1: نرمال سازی به NF1

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

- ✓ جدول ارائه شده اتمیک است، زیرا هیچ ستونی شامل مجموعه ای از مقادیر یا لیستی از داده ها نیست
- ✓ این جدول در 1NF است.

مرحله 2: نرمال سازی به NF2

برای 2NF باید شرایط زیر برقرار باشد:

1. جدول در 1NF باشد.
2. تمام ستون های غیر کلیدی، به کلید اصلی وابستگی کامل داشته باشند (نه وابستگی جزئی).

کلید اصلی پیشنهادی:

برای این جدول، کلید اصلی ترکیبی می تواند (EMP_ID, EMP_DEPT) باشد، زیرا این ترکیب تمام ردیف ها را یکتا می کند.
بررسی وابستگی ها:

- 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 NF2:

EMP_DETAILS

EMP_ID	EMP_COUNTRY
264	India
364	UK

EMP_DEPARTMENT

EMP_ID	EMP_DEPT
264	Designing
264	Testing
364	Stores
364	Developing

DEPT_DETAILS

EMP_DEPT	DEPT_TYPE	EMP_DEPT_NO
Designing	D394	283
Testing	D394	300
Stores	D283	232
Developing	D283	549

مرحله 3: نرمال سازی به 3NF

برای 3NF باید:

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

بررسی وابستگی ها:

- در جدول EMP_DEPT_NO ، DEPT_DETAILS به DEPT_TYPE وابسته است. این وابستگی باید حذف شود.

تفکیک جدول برای 3NF:

1. تفکیک اطلاعات دپارتمان:

EMP_DEPT_NUMBERS

EMP_DEPT	EMP_DEPT_NO
Designing	283
Testing	300
Stores	232
Developing	549

DEPARTMENT_TYPES

DEPT_TYPE	EMP_DEPT_NO
D394	283
D394	300
D283	232
D283	549

اکنون جدول در 3NF است

مرحله 4: نرمال سازی به BCNF

برای BCNF باید:

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

بررسی وابستگی ها:

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

نتایج نهایی جداول نرمال شده:

EMP_DETAILS

EMP_ID	EMP_COUNTRY
264	India
364	UK

EMP_DEPARTMENT

EMP_ID	EMP_DEPT
264	Designing
264	Testing
364	Stores
364	Developing

EMP_DEPT_NUMBERS

EMP_DEPT	EMP_DEPT_NO
Designing	283
Testing	300
Stores	232
Developing	549

DEPARTMENT_TYPES

DEPT_TYPE	EMP_DEPT_NO
D394	283
D394	300
D283	232
D283	549

بخش ب)

جدول رابطه ها

Student ID	Course ID	Instructor	Instructor Department
101	CS101	Prof. Smith	Computer Science
102	MATH201	Prof. Johnson	Mathematics
101	MATH201	Prof. Johnson	Mathematics

مرحله 1: شناسایی کلید اصلی

- **کلید اصلی:** ترکیبی از ستون های 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 شده ای 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.

Student ID	Course ID	Instructor	Instructor Department
101	101	Prof. Smith	Computer Science
102	201	Prof. Johnson	Mathematics
101	201	Prof. Johnson	Mathematics

- 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:

Students_Courses

Student ID	Course ID	Instructor
101	101	Prof. Smith
102	201	Prof. Johnson
101	201	Prof. Johnson

- 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:

- $AB \rightarrow C$
- $AB \rightarrow D$
- $D \rightarrow C$
- $DE \rightarrow B$
- $DEH \rightarrow A$
- $DEH \rightarrow B$
- $AC \rightarrow D$
- $AC \rightarrow D$

از $DEH \rightarrow AB$ می‌توان نتیجه گرفت که DEH یک 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 جزئی).

برای این جدول، کلید اصلی (DEH) است.

بررسی وابستگی‌ها:

- $D \rightarrow C$ یعنی C وابسته است به D، به جز از کلید اصلی وابسته است نه به همه کلید اصلی.
- $DE \rightarrow B$ یعنی B وابسته است به DE، به جز از کلید اصلی وابسته است نه به همه کلید اصلی.

پس رابطه‌های زیر داریم:

$$\begin{aligned} R_1(DEH, A, G); \quad & DEH \rightarrow A, \quad DEH \rightarrow G \\ R_2(DE, B); \quad & DE \rightarrow B \\ R_3(D, C); \quad & D \rightarrow C \end{aligned}$$

اکنون جدول در 2NF است

مرحله 3: نرمال‌سازی به 3NF

1. جدول در 2NF باشد.
 2. هیچ وابستگی متعددی (Transitive Dependency) بین ستون‌ها وجود نداشته باشد
- برای 3NF، باید تمام ستون‌های غیرکلیدی به کلید اصلی وابستگی مستقیم داشته باشند (نه از طریق وابستگی متعددی).

وابستگی متعددی $AB \rightarrow CD$:

- ستون‌های C و D از طریق AB تعیین می‌شوند، اما این ترکیب وابستگی‌های متعددی ایجاد می‌کند. باید تجزیه شود.
- $$AB \rightarrow C, \quad AB \rightarrow D, \quad D \rightarrow C$$

$$R_1(DEH, A, G), \quad R_2(DE, B), \quad R_3(D, C), \quad R_4(AB, D)$$