

موسسه بابان

انتشارات بابان و انتشارات راهیان ارشد

درس و کنکور ارشد

پایگاه داده‌ها

(حل تشریحی سوالات کنکور ۱۴۰۰)

ویژه‌ی داوطلبان کنکور کارشناسی ارشد مهندسی کامپیوتر و IT

براساس کتب مرجع

راما کریشنان، آبراهام سیلبرشاتز و رامز المصری

ارسطو خلیلی فر

سؤالات کنکور کارشناسی ارشد سال ۱۴۰۰

۱- هم‌ارزی‌های جبر رابطه‌ای زیر را در نظر بگیرید. این هم‌ارزی‌ها ممکن است همواره درست باشند، در بعضی شرایط درست باشند، یا همواره نادرست باشند. در این عبارت‌ها، R یک رابطه (Relation)، σ ها شرط‌هایی بر روی R و π ها زیر مجموعه‌هایی از صفت‌های R هستند. کدام هم‌ارزی همواره درست است؟

(مهندس کامپیوتر-دولتی-۱۴۰۰)

$$\begin{array}{ll} \Pi_{a1}(\Pi_{a2}(R)) \equiv \Pi_{a2}(\Pi_{a1}(R)) & (۲) \quad \sigma_{c1}(\sigma_{c2}(R)) \equiv \sigma_{c2}(\sigma_{c1}(R)) \quad (۱) \\ \Pi_{a1}(\Pi_{a2}(R)) \equiv \Pi_{a1}(R) & (۴) \quad \Pi_{a1}(\sigma_{c1}(R)) \equiv \sigma_{c1}(\Pi_{a1}(R)) \quad (۳) \end{array}$$

۲- رابطه $R(A,B,C,D)$ و این وابستگی‌های تابعی را در نظر بگیرید، کدام گزینه در مورد رابطه R درست است؟

(مهندس کامپیوتر-دولتی-۱۴۰۰)

$$B \rightarrow C, CD \rightarrow B$$

(۲) R در BCNF است.

(۱) R در 2NF نیست.

(۴) R در 3NF است، اما در BCNF نیست.

(۳) R در 2NF است، اما در 3NF نیست.

۳- رابطه $R(A,B,C,D,E)$ و این وابستگی‌های تابعی را در نظر بگیرید، تعداد کلیدهای کاندید R چند تاست؟

(مهندس کامپیوتر-دولتی-۱۴۰۰)

$$AB \rightarrow CDE, E \rightarrow BC$$

(۴) 4

(۳) 3

(۲) 2

(۱) 1

(مهندس کامپیوتر-دولتی-۱۴۰۰)

۴- این شمای پایگاه داده را در نظر بگیرید:

Student(sid, sname, age)

Course(cid, cname, credits)

Takes(sid, cid, grade)

می‌خواهیم sid دانشجویانی را پیدا کنیم که هم در درس Database و هم در درس Math ثبت نام کرده‌اند. کدام پرس و جوی SQL برای این منظور مناسب است؟

I. SELECT T1.sid
FROM Course C1, Takes T1
WHERE C1.cid = T1.cid AND C1.cname = 'Database'

INTERSECT

SELECT T2.sid
FROM Course C2, Takes T2
WHERE C2.cid = T2.cid AND C2.cname = 'Math'

II. SELECT T1.sid
FROM Course C1, Takes T1
WHERE C1.cid = T1.cid AND C1.cname = 'Database'
AND T1.sid IN (SELECT T2.sid
FROM Course C2, Takes T2
WHERE C2.cid = T2.cid AND C2.cname = 'Math')

```

III. SELECT T1.sid
FROM Course C1 , Takes T1
WHERE C1.cid = T1.cid AND C1.cname = 'Database'
AND EXISTS (SELECT *
FROM Course C2 , Takes T2
WHERE C2.cid = T2.cid AND C2.cname = 'Math'
AND C2.sid = C1.sid)

```

(۱) فقط I (۲) فقط II (۳) فقط I و II (۴) I و II و III

۵- شمای رابطه‌ای زیر، پایگاه داده موسسات آموزش هنر است.

در این پایگاه داده اسامی هنرجویانی که در هر موسسه عضو هستند ذخیره شده است.

جدول هنرهای مورد علاقه، نام رشته‌های هنری مورد علاقه هر هنرجو را نشان می‌دهد.

جدول دوره‌های هنری نشان می‌دهد در هر موسسه چه رشته‌های هنری ارائه می‌شود.

(مهندسی کامپیوتر-دولتی-۱۴۰۰)

Student (SID , Name)

Institute (IID , IName , IAddress)

Membership (SID, IID)

Favirate Field (SID , Field)

Offered Field (IID , Field)

کدام جبر رابطه‌ای لیست تمام هنرجوها را می‌دهد که فقط در موسسه‌هایی عضوند که هیچ رشته هنری خارج از علاقه‌مندی آنها را ارائه نمی‌دهد؟

(۱) $\Pi_{SID}(\text{Favirate Field} \bowtie \text{Membership} \bowtie \text{Offered Field})$

(۲) $\Pi_{SID}(\text{Membership}) - \Pi_{SID}(\text{Favirate Field} \bowtie \text{Membership} \bowtie \text{Offered Field})$

(۳) $\Pi_{SID}(\text{Membership}) - \Pi_{SID}(\text{Membership} - \Pi_{SID, IID}(\text{Favirate Field} \bowtie \text{Offered Field}))$

(۴) $\Pi_{SID}(\text{Membership}) - \Pi_{SID}(\Pi_{SID, Field}(\text{Membership} \bowtie \text{Offered Field}) - \text{Favirate Field})$

(مهندسی کامپیوتر-دولتی-۱۴۰۰)

۶- حاصل تجزیه رابطه زیر بر اساس 3NF چند رابطه خواهد بود؟

$R(A, B, C, D, E)$

$A \rightarrow B, C$

$B, C \rightarrow A, D$

$D \rightarrow E$

(۱) رابطه 1 (۲) رابطه 2 (۳) رابطه 3 (۴) رابطه 4

پاسخ سؤالات کنکور کارشناسی ارشد سال ۱۴۰۰

۱- گزینه (۱) صحیح است.

$$\sigma_{c1}(\sigma_{c2}(R)) \equiv \sigma_{c2}(\sigma_{c1}(R)) \quad (۱)$$

گزینه اول درست است، زیرا ترتیب اعمال شرط‌های $c1$ و $c2$ تأثیری در نتیجه نهایی ندارد و هر دو طرف عبارت هم‌ارزی معادل $\sigma_{c1} \& c2(R)$ می‌باشد.

$$\Pi_{a1}(\Pi_{a2}(R)) \equiv \Pi_{a2}(\Pi_{a1}(R)) \quad (۲)$$

گزینه دوم نادرست است، زیرا عملگر پرتو (Π) همواره خاصیت جابه‌جایی ندارد. توجه کنید که عبارت $\Pi_{a1}(\Pi_{a2}(R))$ فقط در صورتی قابل انجام است که ستون‌های $a1$ زیر مجموعه ستون‌های $a2$ باشد یعنی $a1 \subseteq a2$ باشد که در آن صورت لزوماً $a2 \subseteq a1$ نیست (مگر در حالت استثنایی که $a1 = a2$ باشد)، بنابراین عبارت $\Pi_{a1}(\Pi_{a2}(R))$ در حالت کلی و در همه شرایط قابل انجام نیست و همچنین همواره معادل عبارت $\Pi_{a2}(\Pi_{a1}(R))$ هم نیست.

$$\Pi_{a1}(\sigma_{c1}(R)) \equiv \sigma_{c1}(\Pi_{a1}(R)) \quad (۳)$$

گزینه سوم نادرست است، زیرا دو عملگر σ و Π به صورت مشروط دارای خاصیت جابه‌جایی هستند. به طور کلی اگر R یک رابطه، L زیر مجموعه‌ای از ستون‌ها و θ مجموعه‌ای از شروط بر روی سطرها باشد، آنگاه تساوی زیر زمانی برقرار است که ستون‌های عملگر σ زیر مجموعه ستون‌های عملگر Π باشد. یعنی $\sigma \subseteq \Pi$

$$\Pi_L(\sigma_\theta(R)) = \sigma_\theta(\Pi_L(R))$$

$$\Pi_{a1}(\Pi_{a2}(R)) \equiv \Pi_{a1}(R) \quad (۴)$$

گزینه چهارم نادرست است، زیرا $\Pi_{a1}(\Pi_{a2}(R)) \equiv \Pi_{a1}(R)$ فقط در شرایطی هم‌ارز است که ستون‌های $a1$ زیر مجموعه ستون‌های $a2$ باشد یعنی $a1 \subseteq a2$ باشد.

۲- گزینه (۴) صحیح است.

با توجه به وابستگی‌های مطرح شده برای رابطه $R(A,B,C,D)$ داریم:

$$B \rightarrow C$$

$$CD \rightarrow B$$

$$ABCD - BC = AD$$

بنابر رابطه فوق صفات AD حتماً باید عضو کلید کاندید باشند. بستار صفات AD به صورت زیر است:

$$\{A,D\}^+ = \{A,D\}$$

براساس بستار فوق، صفات AD ، فقط ستون‌های AD را تولید می‌کند، پس صفات AD فقط عضو کلید کاندید می‌باشد و کلید کاندید نمی‌باشد.

قانون سوم ارسطو

هرگاه عضو کلید کاندید، حاصل از تفاضل قانون اول، برخی از ستون‌ها را تولید کند، بدین معنی است که، جدول موردنظر، چندین کلید کاندید دارد، که این عضو کلید کاندید، در بین تمامی کلیدهای کاندید، به طور مشترک قرار دارد، بنابراین صفات دیگری نیز، باید عضو کلید کاندید را همراهی کنند تا کلید کاندید ایجاد گردد.

همچنین مطابق این قانون، صفاتی که توسط عضو کلید کاندید، قابل دسترسی هستند، در کلید کاندید جایگاهی نخواهند داشت.

براساس بستار فوق، صفات AD به عنوان عضو کلید کاندید همه ستون‌ها را تولید نمی‌کند، بنابراین مطابق قانون سوم ارسطو باید صفاتی در کنار صفت AD قرار گیرد تا کلید کاندید تولید گردد. این صفات کناری از صفات باقی مانده انتخاب می‌گردند، صفات باقی مانده عبارتند از B و C بنابراین با ترکیب صفات B یا C با صفت AD کلیدهای کاندید تولید می‌گردند.

$$\{ADB\}^+ = \{A, D, B, C\}$$

$$\{ADC\}^+ = \{A, D, C, B\}$$

پس ترکیبات صفات ADB و ADC کلیدهای کاندید جدول R هستند.

توجه: دقت کنید که هیچ‌گاه، کلید کاندید نباید عضو زائد داشته باشد. بنابراین رابطه داده شده در مجموع دارای دو کلید کاندید است. که مطابق قانون سوم ارسطو، عضو کلید کاندید AD، در بین هر دو کلید کاندید به طور مشترک قرار دارد.

حال یک‌بار دیگر وابستگی‌های مطرح شده برای رابطه $R(A, B, C, D)$ را در نظر بگیرید:

وابستگی معکوس $B \rightarrow C$ عضو کلید کاندید C

وابستگی معکوس $CD \rightarrow B$ عضو کلید کاندید B

در وابستگی‌های فوق، وابستگی بخشی وجود ندارد. بنابراین جدول مربوطه در نرمال فرم دوم هم قرار دارد.

در وابستگی‌های فوق، وابستگی انتقالی وجود ندارد. بنابراین جدول مربوطه در نرمال فرم سوم هم قرار دارد.

به طور کلی می‌توان شروط قرار داشتن یک جدول در نرمال فرم BCNF را به صورت زیر بیان کرد:

- جدول باید در نرمال فرم سوم باشد.

- جدول باید فاقد وابستگی معکوس باشد.

وابستگی معکوس: وابستگی یک عضو کلید کاندید به عضو یک کلید کاندید دیگر یا مؤلفه غیرکلیدی را، وابستگی معکوس می‌نامند.

عضو کلید کاندید \rightarrow عضو کلید کاندید

عضو کلید کاندید \rightarrow غیرکلید

در وابستگی‌های فوق، وابستگی معکوس وجود دارد. بنابراین جدول مربوطه در نرمال فرم BCNF قرار ندارد.

در یک نگاه دیگر برای نرمال فرم BCNF می‌توان گفت، جدولی در نرمال فرم BCNF قرار دارد که همه‌ی شروع‌های وابستگی‌ها، ابرکلید باشد. به بیان دیگر هرگاه سمت چپ همه وابستگی‌ها، ابرکلید باشد، آن‌گاه آن جدول در BCNF قرار دارد که در وابستگی‌های فوق این چنین نیست. پس BCNF هم نیست. در وابستگی‌های فوق، سمت چپ وابستگی‌های اول و دوم ابرکلید نیست. بنابراین این جدول به دلیل نقض شرایط مربوطه، در نرمال فرم BCNF قرار ندارد. در نتیجه گزینه چهارم پاسخ سوال است.

۳- گزینه (۲) صحیح است.

با توجه به وابستگی‌های مطرح شده برای رابطه $R(A, B, C, D, E)$ داریم:

$$AB \rightarrow CDE$$

$$E \rightarrow BC$$

$$ABCD E - BCDE = A$$

بنابر رابطه فوق صفت A حتماً باید عضو کلید کاندید باشد. بستار صفات A به صورت زیر است:

$$\{A\}^+ = \{A\}$$

براساس بستار فوق، صفت A، فقط ستون A را تولید می‌کند، پس صفت A فقط عضو کلید کاندید می‌باشد و کلید کاندید نمی‌باشد.

قانون سوم ارسطو

هرگاه عضو کلید کاندید، حاصل از تفاضل قانون اول (روش اول یا دوم)، برخی از ستون‌ها را تولید کند، بدین معنی است که، جدول موردنظر، چندین کلید کاندید دارد، که این عضو کلید کاندید، در بین تمامی کلیدهای کاندید، به طور مشترک قرار دارد، بنابراین صفات دیگری نیز، باید عضو کلید کاندید را همراهی کنند تا کلید کاندید ایجاد گردد.

همچنین مطابق این قانون، صفاتی که توسط عضو کلید کاندید، قابل دسترسی هستند، در کلید کاندید جایگاهی نخواهند داشت.

براساس بستار فوق، صفت A به عنوان عضو کلید کاندید همه ستون‌ها را تولید نمی‌کند، بنابراین مطابق قانون سوم ارسطو باید صفاتی در کنار صفت A قرار گیرد تا کلید کاندید تولید گردد. این صفات کناری از صفات باقی مانده انتخاب می‌گردند، صفات باقی مانده عبارتند از B و C و D و E البته از این مجموعه صفات C و D را هم کنار می‌گذاریم، زیرا کمکی در تولید صفات دیگر نمی‌کنند. چون در سمت چپ هیچ یک از وابستگی‌های تابعی نیامده‌اند. بنابراین با ترکیب صفات B یا E با صفت A کلیدهای کاندید تولید می‌گردند.

$$\{AB\}^+ = \{A, B, C, D, E\}$$

$$\{AE\}^+ = \{A, E, B, C, D\}$$

پس ترکیبات صفات AB و AE کلیدهای کاندید جدول R هستند.

بستار صفات AB به صورت زیر است:

$$\{AB\}^+ = \{A, B, C, D, E\}$$

براساس بستار فوق، صفات AB، همه ستون‌ها را تولید می‌کنند، پس صفات AB، کلید کاندید است. همچنین از آن‌جا که مطابق وابستگی‌ها $E \rightarrow B$ ، پس می‌توان ترکیب دو خصیصه (A, E) را هم کلید کاندید دیگری برای این جدول تلقی کرد. چون وقتی (A, B) کلید کاندید است و همه ستون‌ها را تولید می‌کند، پس (A, E) هم کلید کاندید است و همه ستون‌ها را تولید می‌کند، زیرا در نهایت طبق وابستگی $E \rightarrow B$ ، صفت E، صفت B را می‌دهد و (A, E) به (A, B) تبدیل می‌گردد. بستار صفات (A, E) به صورت زیر است:

$$\{AE\}^+ = \{A, E, B, C, D\}$$

براساس بستار فوق، صفات (A, E) ، همه ستون‌ها را تولید می‌کنند، پس صفات (A, E) کلید کاندید است. **توجه:** دقت کنید که هیچ‌گاه، کلید کاندید نباید عضو زائد داشته باشد. بنابراین رابطه داده شده، در مجموع، دارای دو کلید کاندید است که مطابق قانون سوم ارسطو، عضو کلید کاندید A در بین هر دو کلید کاندید به طور مشترک قرار دارد.

۴- گزینه (۳) صحیح است.

سه جدول Student، Course و Takes با مقادیر زیر را در نظر بگیرید:

Sid	Sname	age	Sid	Cid	Grade	Cid	CName	Credits
s1	sn1		s1	c1	10	c1	Database	1
s2	sn2		s1	c2	12	c2	Math	2
s3	sn3		s2	c1	14	Course C1		
Student			s2	c2	16			
			s3	c1	18			
			Takes T1					

مطابق پرس و جوی مطرح شده در گزاره I، داریم:

```
I. SELECT T1.sid
FROM Course C1, Takes T1
WHERE C1.cid = T1.cid AND C1.cname = 'Database'
```

INTERSECT

```
SELECT T2.sid
FROM Course C2, Takes T2
WHERE C2.cid = T2.cid AND C2.cname = 'Math'
```

با توجه به جداول فوق، خروجی بخش اول پرس و جوی گزاره‌ی I پس از انجام عملگر ضرب دکارتی و اعمال شرط به صورت زیر است:

Sid	Cid	Grade	Cid	CName	Credits
s1	c1	10	c1	Database	1
s2	c1	14	c1	Database	1
s3	c1	18	c1	Database	1

و در نهایت پس از اعمال دستور `SELECT T1.sid` خروجی نهایی بخش اول پرس و جوی گزاره‌ی I به صورت زیر است:

Sid
s1
s2
s3

با توجه به جداول فوق، خروجی بخش دوم پرس و جوی گزاره‌ی I پس از انجام عملگر ضرب دکارتی و اعمال شرط به صورت زیر است:

Sid	Cid	Grade	Cid	CName	Credits
s1	c2	12	c2	Math	2
s2	c2	16	c2	Math	2

و در نهایت پس از اعمال دستور `SELECT T2.sid` خروجی نهایی بخش دوم پرس و جوی گزاره‌ی I به صورت زیر است:

Sid
s1
s2

در نهایت پس از انجام عملگر `INTERSECT` خروجی پرس و جو به صورت زیر خواهد بود:

Sid	INTERSECT	Sid	=	Sid
s1		s1		s1
s2		s2		s2
s3				

بنابراین گزاره‌ی I مطابق پرس و جوی مطرح شده در صورت سوال است. یعنی «شماره دانشجویی، دانشجویانی که هم در درس Database و هم در درس Math ثبت نام کرده‌اند را استخراج می‌کند.»
مطابق پرس و جوی مطرح شده در گزاره II، داریم:

```
II. SELECT T1.sid
FROM Course C1, Takes T1
WHERE C1.cid = T1.cid AND C1.cname = 'Database'
AND T1.sid IN (SELECT T2.sid
FROM Course C2, Takes T2
WHERE C2.cid = T2.cid AND C2.cname = 'Math')
```

با توجه به وجود دستور `in`، زیر پرس و جوی داخلی فوق یک `Normal Subquery` است، یعنی ابتدا زیر پرس و جوی داخلی یک بار و برای همیشه اجرا می‌گردد، سپس پرس و جوی خارجی به ازای حرکت در هر یک از سطرها، از مقادیر زیر پرس و جوی داخلی استفاده می‌کند.

در پرس و جوی فوق به ازای حرکت در هر سطر از جدول حاصل از ضرب دکارتی `Course C1, Takes T1` و شرط اعمال شده در پرس و جوی خارجی مجموعه جلوی `in` بررسی می‌گردد که آیا `Sid` داخل این

مجموعه قرار دارد یا خیر. اگر قرار داشت سطر مورد نظر از جدول مذکور در پرس و جوی خارجی در خروجی نمایش داده می شود.

توسط دستور where در فرم زیر:

```
SELECT T1.sid
FROM Course C1, Takes T1
WHERE C1.cid = T1.cid AND C1.cname = 'Database'
AND T1.sid IN (...)
```

برای هر سطر از جدول مذکور کل مجموعه ی جلوی in که حاصل یک مقایسه می باشد، بررسی می گردد، اگر sid موجود در هر سطر، داخل sidهای مجموعه جلوی in بود، آنگاه شرط جلوی where که همان in است، TRUE می گردد و سطر مورد نظر از جدول مذکور انتخاب می گردد و این رویه برای تک تک سطرهای جدول مذکور، تا به انتهای جدول ادامه پیدا می کند. به بیان دیگر این پرس و جو شماره دانشجویانی از جدول مذکور را می دهد که در پرائتز مقابل in قرار دارند. به عبارت دیگر پرس و جوی گزاره ی II شماره دانشجویی، دانشجویانی که هم در درس Database و هم در درس Math ثبت نام کرده اند را استخراج می کند.

با توجه به جداول فوق، خروجی زیر پرس و جوی داخلی پس از انجام عملگر ضرب دکارتی و اعمال شرط به صورت زیر است:

Sid	Cid	Grade	Cid	CName	Credits
s1	c2	12	c2	Math	2
s2	c2	16	c2	Math	2

و در نهایت پس از اعمال دستور SELECT T2.sid خروجی به صورت زیر است:

Sid
s1
s2

بنابراین در ادامه پرس و جوی زیر را خواهیم داشت:

```
SELECT T1.sid
FROM Course C1, Takes T1
WHERE C1.cid = T1.cid AND C1.cname = 'Database'
AND T1.sid IN (s1,s2)
```

همانطور که گفتیم برای هر سطر از جدول مذکور کل مجموعه ی جلوی in که حاصل یک مقایسه می باشد، بررسی می گردد، اگر Sid موجود در هر سطر، داخل Sidهای مجموعه جلوی in بود، آنگاه شرط جلوی where که همان in است، TRUE می گردد و سطر مورد نظر از جدول مذکور انتخاب می گردد و این رویه برای تک تک سطرهای جدول، تا به انتهای جدول ادامه پیدا می کند. به بیان دیگر این پرس و جو شماره دانشجویانی از جدول مذکور را می دهد که در پرائتز مقابل in قرار دارند. با توجه به جداول فوق، خروجی جدول مذکور به صورت زیر است:

Sid	Cid	Grade	Cid	CName	Credits
s1	c1	10	c1	Database	1
s2	c1	14	c1	Database	1
s3	c1	18	c1	Database	1

خروجی نهایی پرس و جوی فوق پس از انجام عملگر in به ازای هر سطر جدول مذکور به صورت زیر است:

Sid	Cid	Grade	Cid	Cname	Credit	IN	Sid	=
s1	c1	10	c1	Database	1		s1	
s2	c1	14	c1	Database	1		s2	
s3	c1	18	c1	Database	1			

Sid	Cid	Grade	Cid	Cname	Credit
s1	c1	10	c1	Database	1
s2	c1	14	c1	Database	1

و در نهایت پس از اعمال دستور SELECT T1.sid به صورت زیر است:

Sid
s1
s2

بنابراین گزاره‌ی II مطابق پرس و جوی مطرح شده در صورت سوال است. یعنی «شماره دانشجویی، دانشجویانی که هم در درس Database و هم در درس Math ثبت نام کرده‌اند را استخراج می‌کند».

مطابق پرس و جوی مطرح شده در گزاره III، داریم:

```

III. SELECT T1.sid
FROM Course C1 , Takes T1
WHERE C1.cid = T1.cid AND C1.cname = 'Database'
AND EXISTS (SELECT *
FROM Course C2 , Takes T2
WHERE C2.cid = T2.cid AND C2.cname = 'Math'
AND C2.sid = C1.sid)

```

که البته پرس و جوی مطرح شده در گزاره III خطای نحوی دارد زیرا در $C2.sid = C1.sid$ وجود C1 و C2 به معنی جدول Course است که اصلاً در جدول Course ستون sid وجود ندارد، که فرم اصلاح شده آن به صورت زیر است:

```

III. SELECT T1.sid
FROM Course C1 , Takes T1
WHERE C1.cid = T1.cid AND C1.cname = 'Database'
AND EXISTS (SELECT *
FROM Course C2 , Takes T2
WHERE C2.cid = T2.cid AND C2.cname = 'Math'
AND T2.sid = T1.sid)

```

توجه: در فرم اصلاح شده‌ی گزاره III، شرط اتصال $T2.sid = T1.sid$ جایگزین شرط اتصال $C2.sid = C1.sid$ شده است.

توجه: گزاره III به فرم اصلاح نشده، خطای نحوی دارد و از سوی کامپایلر اجرا نمی‌شود، بنابراین گزاره III مطابق پرس و جوی مطرح شده در صورت سوال نیست.

با توجه به وجود دستور Exists، زیر پرس وجوی داخلی فوق یک Correlated Subquery است، یعنی به ازای حرکت در هریک از سطرها پرس وجوی خارجی، یکبار به طور کامل از ابتدا تا انتها زیر پرس وجوی داخلی اجرا و بر اساس شرطی که زیر پرس وجوی داخلی را به پرس وجوی خارجی متصل می‌کند، بررسی انجام می‌شود. مانند دو حلقه تو در تو، که به ازای هربار اجرای حلقه خارجی، یکبار به طور کامل حلقه داخلی اجرا می‌گردد.

توسط دستور where در فرم زیر:

```
SELECT T1.sid
FROM Course C1, Takes T1
WHERE C1.cid = T1.cid AND C1.cname = 'Database'
AND EXISTS (...)
```

برای هر سطر از جدول حاصل از ضرب دکارتی Course C1, Takes T1 و شرط اعمال شده شرط جلوی Exists که حاصل یک مقایسه می‌باشد، محاسبه می‌گردد، اگر غیرتهی بود، شرط جلوی where که همان Exists است، TRUE می‌گردد و سطر مورد نظر از جدول مذکور انتخاب می‌گردد و این رویه برای تک تک سطرها جدول مذکور، تا به انتهای جدول ادامه پیدا می‌کند. به بیان دیگر این پرس وجو شماره دانشجویانی را می‌دهد که پُرانتز مقابل Exists برای آن‌ها غیرتهی است. این پُرانتز هنگامی غیرتهی می‌شود که حاصل مقایسه بیان شده در این پُرانتز غیرتهی شود. حاصل این مقایسه در صورتی غیرتهی می‌شود که شماره دانشجویی که درس Database ثبت نام کرده است، درس Math هم ثبت نام کرده باشد. به عبارت دیگر فرم اصلاح شده پرس وجوی گزاره‌ی III شماره دانشجویی، دانشجویانی که هم در درس Database و هم در درس Math ثبت نام کرده‌اند را استخراج می‌کند. که مطابق پرس و جوی مطرح شده در صورت سوال است.

عبارت موجود در جلوی دستور Exists که به صورت زیر است:

```
SELECT *
FROM Course C2, Takes T2
WHERE C2.cid = T2.cid AND C2.cname = 'Math'
AND T2.sid = T1.sid)
```

توسط دستور $T1.Sid = T2.Sid$ به محیط خارج یعنی جدول حاصل از ضرب دکارتی Course C1, Takes T1 و شرط اعمال شده متصل می‌گردد. حال به ازای حرکت در هر سطر از جدول مذکور، یک بار به طور کامل سطرها حاصل از ضرب دکارتی دو جدول Course C2, Takes T2 از ابتدا تا انتها با توجه به شرط اتصال بررسی می‌گردد.

ابتدا برای سطر اول از جدول حاصل از ضرب دکارتی Course C1, Takes T1 با توجه به شرط

اتصال $T1.Sid = T2.Sid$ داریم:
S1 S1

Sid	Cid	Grade	Cid	Cname	Credit	EXISTS
s1	c1	10	c1	Database	1	
s2	c1	14	c1	Database	1	
s3	c1	18	c1	Database	1	

Sid	Cid	Grade	Cid	CName	Credits
s1	c2	12	c2	Math	2
s2	c2	16	c2	Math	2

جلوی Exists برابر غیرتهی گردید، بنابراین شرط where در پشت exists، TRUE می‌گردد. بنابراین سطر اول از جدول حاصل از ضرب دکارتی Course C1, Takes T1 در خروجی نمایش داده می‌شود. در ادامه برای سطر دوم از جدول حاصل از ضرب دکارتی Course C1, Takes T1 با توجه به شرط اتصال $T1.Sid = T2.Sid$ داریم:

S2 S2

Sid	Cid	Grade	Cid	Cname	Credit	EXISTS
s1	c1	10	c1	Database	1	
s2	c1	14	c1	Database	1	
s3	c1	18	c1	Database	1	

Sid	Cid	Grade	Cid	CName	Credits
s1	c2	12	c2	Math	2
s2	c2	16	c2	Math	2

جلوی Exists برابر غیرتهی گردید، بنابراین شرط where در پشت exists، TRUE می‌گردد. بنابراین سطر دوم از جدول حاصل از ضرب دکارتی Course C1, Takes T1 در خروجی نمایش داده می‌شود.

در ادامه برای سطر سوم از جدول حاصل از ضرب دکارتی Course C1, Takes T1 با توجه به شرط اتصال $T1.Sid = T2.Sid$ داریم:

S3 ×

Sid	Cid	Grade	Cid	Cname	Credit	EXISTS
s1	c1	10	c1	Database	1	
s2	c1	14	c1	Database	1	
s3	c1	18	c1	Database	1	

Sid	Cid	Grade	Cid	CName	Credits
s1	c2	12	c2	Math	2
s2	c2	16	c2	Math	2

جلوی Exists برابر تھی گردید، بنابراین شرط where در پشت exists، FALSE می‌گردد. بنابراین سطر سوم از جدول حاصل از ضرب دکارتی Course C1, Takes T1 در خروجی نمایش داده نمی‌شود. بنابراین خروجی پرس‌وجو به صورت زیر خواهد بود:

Sid	Cid	Grade	Cid	Cname	Credit	EXISTS
s1	c1	10	c1	Database	1	
s2	c1	14	c1	Database	1	

و در نهایت پس از اعمال دستور SELECT T1.sid خروجی نهایی به صورت زیر است:

Sid
s1
s2

بنابراین فرم اصلاح شده گزاره‌ی III مطابق پرس و جوی مطرح شده در صورت سوال است. یعنی «شماره دانشجویی، دانشجویانی که هم در درس Database و هم در درس Math ثبت نام کرده‌اند را استخراج می‌کند.»

توجه: دقت کنید که فرم اصلاح نشده گزاره III خطای نحوی و کامپایلری دارد و مطابق پرس‌وجوی مطرح شده در صورت سوال نیست؛ البته فرم اصلاح شده آنرا بررسی کردیم.

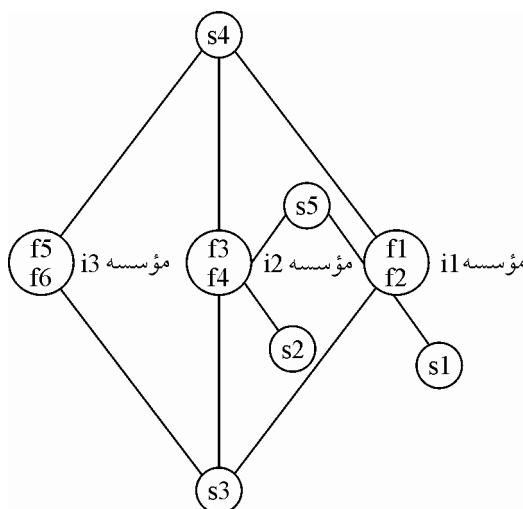
توجه: سازمان سنجش آموزش کشور در کلید اولیه خود ابتدا گزینه چهارم را به عنوان پاسخ اعلام نمود، سپس در کلید نهایی نظر خود را عوض کرد و گزینه سوم را به عنوان پاسخ اعلام کرد، که عمل درستی را انجام داده است.

۵- گزینه (۴) صحیح است.

سه جدول Faviorate Field, Offered Field و Membership با مقادیر زیر را در نظر بگیرید:

SID	Field	SID	IID	IID	Field
s1	f1	s1	i1	i1	f1
s1	f2	s2	i2	i1	f2
s2	f3	s3	i3	i2	f3
s2	f4	s3	i1	i2	f4
s3	f5	s3	i2	i3	f5
s3	f6	s4	i3	i3	f6
s3	f1	s4	i1	Offered Field	
s3	f3	s4	i2		
s4	f5	s5	i1	Membership	
s4	f1	s5	i2		
s4	f3				
s5	f1				
s5	f2				
s5	f3				
s5	f4				
Favorate Field					

شکل زیر روابط بین سه جدول فوق را نمایش می‌دهد:



مطابق پرس و جوی مطرح شده در گزینه چهارم، داریم:

$$\Pi_{SID}(\text{Membership}) - \Pi_{SID}(\Pi_{SID, \text{Field}}(\text{Membership} \bowtie \text{Offered Field}) - \text{Favirate Field})$$

ابتدا توسط عبارت $\Pi_{SID}(\text{Membership})$ در سمت چپ عملگر تفاضل شماره کلیه هنجویان از جدول Membership به صورت زیر استخراج می‌گردد:

SID
s1
s2
s3
s4
s5

توجه: در جبر رابطه‌ای، سطرهای تکراری در خروجی قرار نمی‌گیرند.

همچنین براساس بخش سمت راست عملگر تفاضل داریم:

در عبارت $\Pi_{SID}(\Pi_{SID, \text{Field}}(\text{Membership} \bowtie \text{Offered Field}) - \text{Favirate Field})$ ، الحاق طبیعی پرائتز داخلی یعنی $\Pi_{SID, \text{Field}}(\text{Membership} \bowtie \text{Offered Field})$ به صورت زیر انجام می‌شود:

SID	IID	Feild
s1	i1	f1
s1	i1	f2
s2	i2	f3
s2	i2	f4
s3	i3	f5
s3	i3	f6
s3	i1	f1
s3	i1	f2
s3	i2	f3
s3	i2	f4
s4	i3	f5
s4	i3	f6
s4	i1	f1
s4	i1	f2
s4	i2	f3
s4	i2	f4
s5	i1	f1
s5	i1	f2
s5	i2	f3
s5	i2	f4

سپس در ادامه، توسط عملگر $\Pi_{SID, Field}$ ، شماره هر هنرجو در کنار تمام فیلدهای هنری هر موسسه‌ای که هنرجو در آن حضور دارد به صورت زیر استخراج می‌گردد:

SID	Feild
s1	f1
s1	f2
s2	f3
s2	f4
s3	f5
s3	f6
s3	f1
s3	f2
s3	f3
s3	f4
s4	f5
s4	f6
s4	f1
s4	f2
s4	f3
s4	f4
s5	f1
s5	f2
s5	f3
s5	f4

سپس در ادامه با انجام عملگر تفاضل، خروجی به صورت زیر استخراج می‌گردد:

SID	Feild	-	SID	Feild	=	SID	Feild
s1	f1		s1	f1		s3	f2
s1	f2		s1	f2		s3	f4
s2	f3		s2	f3		s4	f6
s2	f4		s2	f4		s4	f2
s3	f5		s3	f5		s4	f4
s3	f6		s3	f6			
s3	f1		s3	f1			
s3	f2		s3	f3			
s3	f3		s4	f5			
s3	f4		s4	f1			
s4	f5		s4	f3			
s4	f6		s5	f1			
s4	f1		s5	f2			
s4	f2		s5	f3			
s4	f3		s5	f4			
s4	f4						
s5	f1						
s5	f2						
s5	f3						
s5	f4						

Favirate Field

سپس در ادامه، توسط عملگر Π_{SID} ، خروجی به صورت زیر استخراج می‌گردد:

SID
s3
s4

در انتها، با انجام عملگر تفاضل نهایی، شماره هنجویان فوق، از مجموعه شماره کلیه هنجویان کنار گذاشته می‌شود. آنچه باقی می‌ماند، شماره یا لیست تمام هنجویانی است که فقط در موسسه‌هایی عضوند که هیچ رشته هنری خارج از علاقه‌مندی آنها را ارائه نمی‌دهد. دقت کنید که بر اساس صورت سوال، مهم نیست که یک هنرجو فقط عضو یک موسسه باشد یا نباشد، اما اگر عضو هر موسسه‌ای بود باید در تمام دوره‌های هنری موسسه شرکت کرده باشد. بنابراین واضح است که پاسخ گزینه چهارم است.

خروجی نهایی پرس و جوی مطرح شده به صورت زیر است:

SID	-	SID	=	SID
s1		s3		s1
s2		s4		s2
s3				s5
s4				
s5				

توجه: در سایر گزینه‌ها اطلاعات نامرتبط و نامعتبر استخراج می‌گردد.

۶- گزینه (۲) صحیح است.

با توجه به وابستگی‌های مطرح شده برای رابطه $R(A, B, C, D, E)$ داریم:

$$A \rightarrow B, C$$

$$B, C \rightarrow A, D$$

$$D \rightarrow E$$

$$ABCDE - ABCDE = \text{تهی}$$

قانون چهارم ارسطو

هرگاه عضو کلید کاندید، حاصل از تفاضل قانون اول، تهی گردد، بدین معنی است که، جدول فوق چندین کلید کاندید دارد، که هیچ عضو کلید کاندید مشترکی، بین تمامی کلیدهای کاندید آن وجود ندارد. بنابراین باید کلید کاندید با بررسی دقیق بر روی مجموعه وابستگی کشف گردد.

$$\{A\}^+ = \{A, B, C, D, E\}$$

براساس بستر فوق، صفت A، همه ستون‌ها را بدون عضو زائد تولید می‌کند، پس صفت A، کلید کاندید می‌باشد.

حال کلید کاندید A را در نظر بگیرید، از آن‌جا که مطابق وابستگی $BC \rightarrow A$ ، صفات BC، صفت A را می‌دهد. می‌توان به جای صفت A در کلید کاندید A، صفات BC را قرار داد. چون مجدداً مطابق وابستگی $BC \rightarrow A$ ، صفت A را می‌دهد که منجر به ایجاد کلید کاندید BC می‌گردد. مطابق الگوی زیر:

$$A \Rightarrow BC$$

↑

$$BC$$

$$\{BC\}^+ = \{B, C, A, D, E\}$$

براساس بستر فوق، صفات BC، همه ستون‌ها را بدون عضو زائد تولید می‌کند، پس صفات BC کلید کاندید می‌باشد.

بررسی مجموعه وابستگی‌های مطرح شده نشان داد که (A) و (B, C) به عنوان کلیدهای کاندید برای رابطه R محسوب می‌شوند.

توجه: همان‌طور که مشاهده می‌شود، مطابق قانون چهارم ارسطو، هیچ عضو کلید کاندید مشترکی، بین

تمامی کلیدهای کاندید فوق وجود ندارد.

واضح است که جدول مطرح شده در فرم اول نرمال قرار دارد.

به طور کلی می توان شروط قرار داشتن یک جدول در نرمال فرم دوم را به صورت زیر بیان کرد:

• جدول باید در نرمال فرم اول باشد.

• جدول باید فاقد وابستگی بخشی باشد.

وابستگی بخشی: وابستگی یک مؤلفه غیرکلیدی، به جزئی از کلید کاندید را وابستگی بخشی می نامند.

مؤلفه غیرکلید: هر صفتی که عضو هیچ کلید کاندیدی نباشد، به عنوان مؤلفه غیرکلیدی نامیده می شود.

مؤلفه جزء کلید کاندید: هر صفتی که عضو حداقل یک کلید کاندید باشد، به عنوان مؤلفه جزء کلید نامیده می شود.

وابستگی های مطرح شده را در نظر بگیرید:

کلید کاندید $A \longrightarrow BC$ وابستگی کامل

کلید کاندید $BC \longrightarrow A$ وابستگی کامل

غیرکلید $BC \longrightarrow D$ وابستگی کامل

غیرکلید $D \longrightarrow E$ وابستگی انتقالی

در وابستگی های فوق، وابستگی بخشی وجود ندارد. بنابراین جدول مربوطه در نرمال فرم دوم هم قرار دارد.

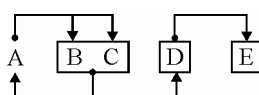
به طور کلی می توان شروط قرار داشتن یک جدول در نرمال فرم سوم را به صورت زیر بیان کرد:

• جدول باید در نرمال فرم دوم باشد.

• جدول باید فاقد وابستگی انتقالی باشد.

وابستگی انتقالی: وابستگی یک مؤلفه غیرکلیدی به یک مؤلفه غیرکلیدی دیگر را وابستگی انتقالی می نامند.

در جدول زیر:

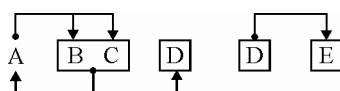


وابستگی تابعی زیر:

$D \longrightarrow E$

به عنوان وابستگی انتقالی محسوب می گردد. بنابراین جدول R در نرمال فرم سوم قرار ندارد.

بنابراین برای قرار دادن جدول R، در نرمال فرم سوم، باید وابستگی های انتقالی از جدول R خارج گردند.



بنابراین جدول R به دو جدول کوچکتر تجزیه می گردد.

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

با توجه به کلیدهای کاندید در دو جدول R_1 و R_2 می‌توان دریافت که این دو جدول فاقد وابستگی بخشی و انتقالی هستند و بنابراین در نرمال فرم سوم قرار دارند. همچنین از آن جا که سمت چپ تمام وابستگی‌های دو جدول R_1 و R_2 ، ابرکلید است، بنابراین دو جدول R_1 و R_2 در سطح BCNF نیز قرار دارند. به این ترتیب با تجزیه جدول R به دو جدول R_1 و R_2 می‌توان جدول R را در نرمال فرم سوم و سطح BCNF قرار داد.

تست‌های فصل هشتم: وابستگی تابعی

۸۱- در رابطه $R = (A, B, C, D, E, F)$ وابستگی‌های زیر برقرار است: (مهندسی IT - دولتی ۱۴۰۰)

$A \rightarrow F$

$B \rightarrow C$

$ABC \rightarrow E$

در این رابطه چند سوپر کلید وجود دارد؟

۱ (۱)

۳ (۲)

۶ (۳)

۸ (۴)

عنوان کتاب: پایگاه داده‌ها

مؤلف: ارسطو خلیلی فر

ناشر: انتشارات راهیان ارشد

آدرس سایت گروه بابان: khalilifar.ir

پاسخ تست‌های فصل هشتم: وابستگی تابعی

۸۱- گزینه (۴) صحیح است.

به طور کلی کلید کاندید باید دو شرط زیر را داشته باشد:

۱- ابرکلید باشد (خاصیت کلیدی داشته باشد) یعنی همه خصیصه‌ها را تولید کند.

۲- عضو زائد نداشته باشد.

به طور کلی عضو کلید کاندید از روابط زیر به دست می‌آید:

قانون اول ارسطو

روش اول:

اجتماع تمام خصیصه‌های سمت راست وابستگی‌های غیر بدیهی - تمام خصیصه‌های جدول = عضو کلید کاندید

روش دوم:

$$\text{عضو کلید کاندید} = R - \bigcup_{i=1}^n [x_i \text{ (چپ)} - y_i \text{ (راست)}]$$

توجه: عبارت $[x_i \text{ (چپ)} - y_i \text{ (راست)}]$ به طور مستقل بر روی تک تک وابستگی‌ها انجام می‌گردد.

$$A \rightarrow BC \Rightarrow BC - A = BC$$

مثال:

توجه: استفاده از روش اول مستلزم گام ابتدایی حذف وابستگی‌های بدیهی است، اگر در حذف وابستگی‌های بدیهی دچار خطا می‌شوید، از روش دوم استفاده نمایید.

با توجه به وابستگی‌های مطرح شده برای رابطه $R(A,B,C,D,E,F)$ داریم:

$$A \rightarrow F$$

$$B \rightarrow C$$

$$ABC \rightarrow E$$

$$ABCDEF - CEF = ABD$$

بنابر رابطه فوق صفات ABD حتماً باید عضو کلید کاندید باشد. بستر صفات ABD به صورت زیر است:

$$\{ABD\}^+ = \{A,B,D,F,C,E\}$$

براساس بستر فوق، صفات ABD ، همه ستون‌ها را تولید می‌کند، پس صفات ABD کلید کاندید می‌باشد.

قانون دوم ارسطو

هرگاه عضو کلید کاندید، حاصل از تفاضل قانون اول (روش اول یا دوم) همه ستون‌ها را تولید

کند، آن عضو کلید کاندید، تنها کلید کاندید جدول خواهد بود.
ساخت ابرکلید: هیچ یا ترکیبی از صفات باقیمانده + کلید کاندید $ABD =$ ابرکلید

A B D +	CEF	ترکیبات صفات باقی مانده	ابرکلیدها
	0 0 0	تهی	A B D
	0 0 1	F	A B D F
	0 1 0	E	A B D E
	0 1 1	E F	A B D E F
	1 0 0	C	A B D C
	1 0 1	C F	A B D C F
	1 1 0	C E	A B D C E
	1 1 1	C E F	A B D C E F

توجه: واضح است که 8 ابرکلید، ایجاد می گردد. (2^3)

بنابراین حاصل جمع ابرکلیدها 8 عدد خواهد بود، به صورت زیر:

ابرکلیدهای **ABD** ، **ABDF** ، **ABDE** ، **ABDEF** ، **ABDC** ، **ABDCF** ، **ABDCE** و **ABDCEF** به همین سادگی.

تست‌های فصل سوم: مدل رابطه‌ای

(مهندسی IT - دولتی ۱۴۰۰)

۸۲- کدام مورد در خصوص رابطه درست است؟

- ۱) رابطه‌ای نرمال است که هیچ یک از صفات ساده‌اش چند مقداری نباشند.
- ۲) کلید کاندید رابطه می‌تواند کاهش‌پذیر باشد.
- ۳) تاپل‌های یک رابطه نظم مکانی دارند.
- ۴) رابطه تاپل تکراری ندارد.

عنوان کتاب: پایگاه داده‌ها

مؤلف: ارسطو خلیلی فر

ناشر: انتشارات راهیان ارشد

آدرس سایت گروه بابان: khalilifar.ir

پاسخ تست‌های فصل هشتم: وابستگی تابعی

۸۲- گزینه (۴) صحیح است.

ویژگی‌های رابطه در مدل رابطه‌ای

۱- در رابطه، تاپل تکراری وجود ندارد. زیرا رابطه یک مجموعه است و مجموعه در ریاضیات طبق تعریف تاپل تکراری ندارد. بنابراین در جدول هم رکوردهای تکراری نداریم.

۲- تاپل‌ها در رابطه نظم مکانی ندارند. زیرا مجموعه نظم ندارد. بنابراین ترتیب رکوردها در جدول اهمیت ندارد.

۳- ترتیب صفات خاصه در یک تاپل مهم نیست. البته به شرطی که ترتیب صفات خاصه در تمام تاپل‌های موجود در یک رابطه با یکدیگر سازگار باشند. بنابراین ترتیب فیلدها در یک جدول نیز مهم نیست، نظم ندارد و می‌توان آنرا جلیه‌جا کرد. البته به شرطی که ترتیب فیلدها در تمام رکوردهای موجود در یک جدول با یکدیگر سازگار باشند.

برای مثال رابطه R_3 به عنوان یک زیر مجموعه دلخواه از حاصلضرب دکارتی سه دامنه D_1 ، D_2 و D_3 را در نظر بگیرید:

$$R_3 = \{(a, x, 1), (a, x, 2), (b, x, 1), (b, x, 2)\}$$

اگر ترتیب تاپل‌های رابطه R_7 به صورت زیر باشد:

$$R_7 = \{(x, 1, a), (x, 2, a), (x, 1, b), (x, 2, b)\}$$

رابطه R_7 معادل رابطه R_3 است، زیرا ترتیب صفات خاصه در تمام تاپل‌های موجود در رابطه R_7 با یکدیگر سازگار است. ($R_3 = R_7$)

اگر ترتیب تاپل‌های رابطه R_8 به صورت زیر باشد:

$$R_8 = \{(x, 1, a), (2, x, a), (1, b, x), (x, 2, b)\}$$

رابطه R_8 معادل رابطه R_3 نیست، زیرا ترتیب صفات خاصه در تمام تاپل‌های موجود در رابطه R_8 با یکدیگر سازگار نیست. ($R_3 \neq R_8$)

۴- همه‌ی مقادیر صفات خاصه در تاپل‌ها تجزیه‌ناپذیرند. به عبارت دیگر در مدل رابطه‌ای نمی‌توان فیلد مرکبی که از فیلدهای ساده تشکیل شده‌است، تعریف نمود. یعنی فیلدهای موجود در جدول مدل رابطه‌ای نباید مرکب مثل تاریخ تولد و آدرس باشد و نباید چندمقداری مثل مدرک تحصیلی باشد. به عبارت دیگر هیچ رابطه‌ای نمی‌تواند شامل تاپل‌های تو در تو، تاپل تو تاپل یا جدول تو جدول باشد، به بیان دیگر مولفه‌های هریک از تاپل‌های رابطه نباید مرکب و چندمقداری باشد، به عبارت ساده‌تر هر مولفه یک تاپل نمی‌تواند خودش یک تاپل باشد.

توجه: بهترین راه نمایش و پیاده‌سازی رابطه استفاده از جدول است، بنابراین جدول معادل رابطه در نظر گرفته می‌شود.

گزینه اول نادرست است. زیرا، اگر هیچ‌یک از صفات ساده یک رابطه چندمقداری و مرکب نباشد، این مربوط به فقط و فقط نرمال سطح اول (1nf) است. در حالی که نرمال‌سازی تا سطح سوم (3nf) اجباری است.

گزینه دوم نادرست است. زیرا، ابرکلیدی که عضو زائد نداشته باشد، کلید کاندید است، به عبارت دیگر ابرکلید کمینه را کلید کاندید می‌گویند. منظور از ابرکلید کمینه، ابرکلیدی نیست که کمترین تعداد صفت را داشته باشد، بلکه منظور ابرکلیدی است که صفت زائد نداشته باشد. بنابراین کلید کاندید یک رابطه خودش به طور ذاتی کمینه است و نمی‌تواند کاهش‌پذیر باشد.

گزینه سوم نادرست است. زیرا، تاپل‌ها در رابطه نظم مکانی ندارند. زیرا مجموعه نظم ندارد. بنابراین ترتیب رکوردها در جدول اهمیت ندارد.

گزینه چهارم درست است. زیرا، در رابطه، تاپل تکراری وجود ندارد. زیرا رابطه یک مجموعه است و مجموعه در ریاضیات طبق تعریف تاپل تکراری ندارد. بنابراین در جدول هم رکوردهای تکراری نداریم.

تست‌های فصل هشتم: وابستگی تابعی

۸۳- فرض کنید در یک جدول $N \geq 3$ خصیصه داریم. در این جدول دو کلید کاندید وجود دارد. یک کلید کاندید دارای دو خصیصه است و یک کلید کاندید دیگر دارای یک خصیصه. بین خصیصه‌های این دو کلید همپوشانی وجود ندارد. تعداد کل سوپر کلیدها در این جدول کدام است؟

(مهندسی IT - دولتی ۱۴۰۰)

(۱) $3 \times 2^{N-2}$

(۲) $3 \times 2^{N-3}$

(۳) $5 \times 2^{N-3}$

(۴) $5 \times 2^{N-2}$

عنوان کتاب: پایگاه داده‌ها

مؤلف: ارسطو خلیلی‌فر

ناشر: انتشارات راهیان ارشد

آدرس سایت گروه بابان: khalilifar.ir

پاسخ تست‌های فصل هشتم: وابستگی تابعی

۸۳- گزینه (۳) صحیح است.

برای درک سوال و راه حل آن به مثال زیر دقت نمایید:

مثال: با توجه به رابطه $R(A,B,C,D,E)$ و مجموعه وابستگی‌های تابعی زیر، کدام مورد نادرست است؟

$$F = \{A \rightarrow BC, CD \rightarrow E, B \rightarrow D, E \rightarrow A\}$$

(۱) $\{E\}$ و $\{C,B\}$ هر دو کلید کاندید هستند. (۲) $\{B\}$ و $\{C,D\}$ هر دو کلید کاندید هستند.

(۳) $\{E\}$ و $\{C,D\}$ هر دو کلید کاندید هستند. (۴) $\{A\}$ و $\{E\}$ هر دو کلید کاندید هستند.

پاسخ: گزینه (۲) صحیح است.

به طور کلی کلید کاندید باید دو شرط زیر را داشته باشد:

۱- ابرکلید باشد (خاصیت کلیدی داشته باشد) یعنی همه خصیصه‌ها را تولید کند.

۲- عضو زائد نداشته باشد.

به طور کلی عضو کلید کاندید از روابط زیر به دست می‌آید:

قانون اول ارسطو

روش اول:

اجتماع تمام خصیصه‌های سمت راست وابستگی‌های غیر بدیهی - تمام خصیصه‌های جدول = عضو کلید کاندید

روش دوم:

$$\text{عضو کلید کاندید} = R - \bigcup_{i=1}^n [(x_i \text{ (چپ)}) - (y_i \text{ (راست)})]$$

توجه: عبارت $[(x_i \text{ (چپ)}) - (y_i \text{ (راست)})]$ به طور مستقل بر روی تک تک وابستگی‌ها انجام می‌گردد.

$$A \rightarrow BC \Rightarrow BC - A = BC$$

مثال:

توجه: استفاده از روش اول مستلزم گام ابتدایی حذف وابستگی‌های بدیهی است، اگر در حذف

وابستگی‌های بدیهی دچار خطا می‌شوید، از روش دوم استفاده نمایید.

با توجه به وابستگی‌های مطرح شده برای رابطه $R(A,B,C,D,E)$ داریم:

$$A \rightarrow BC$$

$$CD \rightarrow E$$

$$B \rightarrow D$$

$$E \rightarrow A$$

$$ABCDE - ABCDE = \text{تهی}$$

قانون چهارم ارسطو

هرگاه عضو کلید کاندید، حاصل از تفاضل قانون اول (روش اول یا دوم)، تهی گردد، بدین معنی است که، جدول فوق چندین کلید کاندید دارد، که هیچ عضو کلید کاندید مشترکی، بین تمامی کلیدهای کاندید آن وجود ندارد. بنابراین باید کلید کاندید با بررسی دقیق بر روی مجموعه وابستگی کشف گردد.

$$\{B\}^+ = \{B, D\}$$

براساس بستار فوق، صفت B، فقط ستون‌های B و D را تولید می‌کند، پس صفت B، کلید کاندید نمی‌باشد. بنابراین گزینه دوم پاسخ سوال خواهد بود.

$$\{A\}^+ = \{A, B, C, D, E\}$$

براساس بستار فوق، صفت A، همه ستون‌ها را بدون عضو زائد تولید می‌کند، پس صفت A، کلید کاندید می‌باشد.

در ادامه به شکل بازگشتی جهت کشف مابقی کلیدهای کاندید داریم:

صفت E، ستون A را تولید می‌کند. پس صفت E کلید کاندید است، به صورت زیر:

$$\{E\}^+ = \{E, A, B, C, D\}$$

براساس بستار فوق، صفت E، همه ستون‌ها را بدون عضو زائد تولید می‌کند، پس صفت E، کلید کاندید می‌باشد.

صفات CD، ستون E را تولید می‌کند. پس صفات CD کلید کاندید است، به صورت زیر:

$$\{CD\}^+ = \{C, D, E, A, B\}$$

براساس بستار فوق، صفات CD، همه ستون‌ها را بدون عضو زائد تولید می‌کند، پس صفات CD، کلید کاندید می‌باشد.

صفت B، ستون D را تولید می‌کند. پس صفات CB کلید کاندید است، به صورت زیر:

$$\{CB\}^+ = \{C, B, D, E, A\}$$

براساس بستار فوق، صفات CB، همه ستون‌ها را بدون عضو زائد تولید می‌کند، پس صفات CB، کلید کاندید می‌باشد.

توجه: همان‌طور که مشاهده می‌شود، مطابق قانون چهارم ارسطو، هیچ عضو کلید کاندید مشترکی، بین تمامی کلیدهای کاندید فوق وجود ندارد.

توجه: پس صورت سوال مطرح شده در شرایط قانون چهارم ارسطو قرار دارد، زیرا اول اینکه بیشتر از یک کلید کاندید دارد و دوم اینکه هیچ عضو کلید کاندید مشترکی، بین تمامی کلیدهای کاندید آن وجود ندارد.

توجه: فرض کنید کلید کاندید تک خصیصه برابر A و کلید کاندید دو خصیصه برابر BC باشد.

توجه: در یک جدول با $N \geq 3$ خصیصه داریم:

ترکیب اول: هیچ یا ترکیبی از صفات باقی مانده + کلید کاندید A = ابرکلید

A +	BCDEF...	ترکیبات صفات باقی مانده	ابرکلیدها
	00000...	تهی	A

	11111...	BCDEF...	ABCDEF...

توجه: واضح است که در این حالت (2^{n-1}) ابرکلید، ایجاد می‌گردد.

ترکیب دوم: هیچ یا ترکیبی از صفات باقیمانده + کلید کاندید BC = ابرکلید

B C +	ADEF...	ترکیبات صفات باقی مانده	ابرکلیدها
	0000...	تهی	B C

	1111...	ADEF...	BCADEF...

توجه: واضح است که در این حالت (2^{n-2}) ابرکلید، ایجاد می‌گردد.

بنابراین حاصل جمع ابرکلیدها $\left[(2^{n-1}) + (2^{n-2}) \right]$ عدد خواهد بود، که از این مجموعه حاصل،

ابرکلیدهای زیر در دو مجموعه فوق تکراری هستند:

ترکیب تکراری‌ها: هیچ یا ترکیبی از صفات باقیمانده + ابرکلید ABC = ابرکلید

توجه: ترکیب دو کلید کاندید یعنی کلید کاندید A و کلید کاندید BC باهم می‌شود، ابرکلید.

A B C +	DEF...	ترکیبات صفات باقی مانده	ابرکلیدها
	000...	تهی	A B C

	111...	DEF...	ABCDEF...

توجه: واضح است که (2^{n-3}) ابرکلید تکراری، میان دو مجموعه ترکیب اول و دوم وجود دارد.

توجه: بنابراین با کنار گذاشتن ابرکلیدهای تکراری، در نهایت $5 \times (2^{n-3})$ ابرکلید خواهیم داشت،

به صورت زیر:

$$\left[(2^{n-1}) + (2^{n-2}) \right] - \left[(2^{n-3}) \right] = 2^{n-3} (2^2 + 2^1 - 1) = 5 \times (2^{n-3})$$

تست‌های فصل ششم: SQL دستورات DML

۸۴- رابطه $R = (A, B, C, D, E, F)$ را در نظر بگیرید. اگر مجموعه وابستگی‌های تابعی F روی R برقرار باشد، کدام یک از تجزیه‌های زیر دارای گمشدگی (lossless) است؟

(مهندسی IT- دولتی ۹۹)

$F = \{A \rightarrow BC, CD \rightarrow E, B \rightarrow D, E \rightarrow A\}$

(۱) $R_1(A, B, C), R_2(C, D, E)$

(۲) $R_1(A, B, C), R_2(A, D, E)$

(۳) $R_1(E, B, C), R_2(E, D, A)$

(۴) $R_1(A, B, C), R_2(B, C, D, E)$

عنوان کتاب: پایگاه داده‌ها

مؤلف: ارسطو خلیلی‌فر

ناشر: انتشارات راهیان ارشد

آدرس سایت گروه بابان: khalilifar.ir

پاسخ تست‌های فصل نهم: نرمال‌سازی

۸۴- گزینه (۱) صحیح است.

صورت سوال ابهام و خطا دارد، کلمه فارسی «گمشدگی» معادل کلمه انگلیسی «lossless» نیست، معنی درست و فارسی کلمه «lossless» عبارت «بدون گمشدگی» است. و دیگر اینکه صفت F موجود در جدول R در هیچ یک از گزینه‌ها نیامده است. بنابراین سوال قابل حل نیست.

توجه: سازمان سنجش آموزش کشور، در کلید اولیه و نهایی خود، گزینه اول را به عنوان پاسخ اعلام کرده بود. که کار درستی نبوده است و سوال باید حذف می‌شد که نشد.

شرایط تجزیه مطلوب به صورت زیر است:

۱- شرط لازم (Nonloss join): دو جدول الحاق‌پذیر باشند (صفت مشترک در دو جدول، حداقل در یکی کلید کاندید باشد) این شرط، مهم‌ترین شرط نرمال‌سازی است، زیرا شرط لازم است.

۲- شرط کافی (حفظ وابستگی‌های تابعی): وابستگی‌های تابعی جدول پایه قابل استنتاج از روی وابستگی‌های تابعی جداول حاصل از تجزیه باشد یا در آن‌ها موجود باشد.

توجه: در تجزیه بدون گمشدگی (Lossless)، صفت مشترک در یکی از روابط باید کلید کاندید باشد. همچنین هنگامی که در فرآیند نرمال‌سازی یک جدول تجزیه می‌شود، این تجزیه باید به گونه‌ای انجام شود که امکان بازسازی محتوایی و ساختاری جدول اولیه بر اساس جداول کوچکتر فراهم باشد، به این صورت که با الحاق جداول کوچکتر به یکدیگر دقیقاً جدول اولیه (پایه) ایجاد گردد (نه سطر یا ستونی از جدول اولیه از دست برود و نه سطر یا ستونی به جدول اولیه اضافه شود). چنین تجزیه‌ای را تجزیه بدون گمشدگی (Lossless) می‌گویند.

اگر صورت سوال را به فرم زیر اصلاح کنیم، یعنی صفت F از جدول R حذف شود و معنی «lossless» را بدون گمشدگی معنی کنیم:

رابطه $R = (A, B, C, D, E)$ را در نظر بگیرید. اگر مجموعه وابستگی‌های تابعی F روی R برقرار باشد، کدام یک از تجزیه‌های زیر بدون گمشدگی (lossless) است؟

آنگاه بررسی گزینه‌ها بر اساس وابستگی‌های مطرح شده به صورت زیر خواهد بود:

$A \rightarrow BC$
 $CD \rightarrow E$
 $B \rightarrow D$
 $E \rightarrow A$

گزینه اول:

$$R_1(A, B, C)$$

$$\{C\}^+ = \{C, \text{Stop}\}$$

$$R_2(C, D, E)$$

$$\{C\}^+ = \{C, \text{Stop}\}$$

در گزینه اول، ستون مشترک C در $R_1(A, B, C)$ یا $R_2(C, D, E)$ کلید کاندید نیست. بنابراین «دارای گمشدگی» است.

گزینه دوم:

$$R_1(A, B, C)$$

$$A \rightarrow BC$$

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

$$R_2(A, D, E)$$

$$\{A\}^+ = \{A, \text{Stop}\}$$

در گزینه دوم، ستون مشترک A در $R_1(A, B, C)$ یا $R_2(A, D, E)$ در جدول R_1 کلید کاندید است. بنابراین «بدون گمشدگی» (Lossless) است.

گزینه سوم:

$$R_1(E, B, C)$$

$$\{E\}^+ = \{E, \text{Stop}\}$$

$$R_2(E, D, A)$$

$$E \rightarrow A$$

$$\{E\}^+ = \{E, A, \text{Stop}\}$$

در گزینه سوم، ستون مشترک E در $R_1(E, B, C)$ یا $R_2(E, D, A)$ کلید کاندید نیست. بنابراین «دارای گمشدگی» است.

گزینه چهارم:

$R_1(A, B, C)$

$\{BC\}^+ = \{B, C, \text{Stop}\}$

$R_2(B, C, D, E)$

$CD \rightarrow E$

$B \rightarrow D$

$\{BC\}^+ = \{B, C, D, E\}$

در گزینه دوم، ستون مشترک BC در $R_1(A, B, C)$ یا $R_2(B, C, D, E)$ در جدول R_2 کلید کاندید است. بنابراین «بدون گمشدگی» (Lossless) است.

تست‌های فصل چهارم: جبر رابطه‌ای

۸۵- رابطه $R = (a, b, c)$ و عبارات جبر رابطه‌ای زیر را در نظر بگیرید: (مهندسی IT - دولتی ۱۴۰۰)

$$Q_1 : \rho(S(a, b_1, c_1), R)$$

$$Q_2 : \Pi_{b,c}(\sigma_{b=c} R)$$

$$\rho(T, S \bowtie R)$$

$$\Pi_{b_1,c}(\sigma_{b_1=c} T)$$

- (۱) Q_1 و Q_2 پاسخ‌های یکسان تولید می‌کنند.
(۲) Q_1 و Q_2 پاسخ‌های متفاوت تولید می‌کنند.
(۳) پاسخ Q_2 زیر مجموعه‌ای از پاسخ Q_1 است.
(۴) پاسخ Q_1 زیر مجموعه‌ای از پاسخ Q_2 است.

عنوان کتاب: پایگاه داده‌ها

مؤلف: ارسطو خلیلی‌فر

ناشر: انتشارات راهیان ارشد

آدرس سایت گروه بابان: khalilifar.ir

پاسخ تست‌های فصل چهارم: جبر رابطه‌ای

۸۵- گزینه (۳) صحیح است.

جدول $R(a,b,c)$ را با مقادیر زیر در نظر بگیرید:

a	b	c
1	2	2
3	4	4
5	6	7
5	7	6

جدول R

بخش اول از پرس و جوی Q_1 :

$$Q_1 : \rho(S(a, b_1, c_1), R)$$

عملگر ρ جهت نام‌گذاری مجدد یک جدول یا نام‌گذاری مجدد ستون‌های یک جدول به طور موقت مورد استفاده قرار می‌گیرد. محدوده اعتبار عملگر ρ در همان پرس و جوی مربوطه است، و پس از پایان پرس و جو نام‌گذاری جدید دیگر اعتبار ندارد. این عملگر فقط یک نام‌گذاری مجدد برای یک جدول است و نه ذخیره‌سازی مجدد جدول مورد نظر.

فرم کلی عملگر ρ برای نام‌گذاری مجدد یک جدول به طور موقت به صورت زیر است:

$$\rho_E(R)$$

فرم کلی عملگر ρ برای نام‌گذاری مجدد ستون‌های یک جدول به طور موقت به صورت زیر است:

$$\rho_{E(L)}(R)$$

در بخش E نام‌گذاری مجدد برای رابطه R مشخص می‌شود. و در بخش L نام‌گذاری مجدد برای ستون‌های رابطه R مشخص می‌شود.

در بخش R نام جدول یا جداول به عنوان مکان یا محل پرس و جو مشخص می‌گردد، این مکان جستجو می‌تواند خروجی یک پرس و جوی دیگر یعنی یک عبارت جبر رابطه‌ای دیگر باشد.

$$Q_1 : \rho(S(a, b_1, c_1), R)$$

در بخش اول پرس و جوی Q_1 ، جدول R به جدول S با ستون‌های a, b_1, c_1 نام‌گذاری مجدد می‌شود، به صورت زیر:

a	b ₁	c ₁
1	2	2
3	4	4
5	6	7
5	7	6

جدول S

در بخش دوم از پرس و جوی Q₁:

$$\rho(T, S \bowtie R)$$

عملگر الحاق طبیعی با نماد \bowtie یا Join نشان داده می‌شود. عملگر \bowtie جهت الحاق سطرهای پیوندپذیر مابین دو جدول مورد استفاده قرار می‌گیرد. سطرهایی از دو جدول که مقدارشان در ستون یا ستون‌های مشترک مساوی است به عنوان سطرهای پیوندپذیر نامیده می‌شوند. در جبر رابطه‌ای الحاق طبیعی هر دو رابطه دلخواه امکان‌پذیر نیست و شرط خاصی برای الحاق طبیعی دارد. داشتن ستون یا ستون‌های مشترک مابین دو جدول شرط انجام عملگر الحاق طبیعی است. اگر دو جدولی که با یکدیگر الحاق طبیعی می‌شوند دارای ستون یا ستون‌های مشترک باشند، آنگاه سطرهایی از دو جدول در خروجی قرار می‌گیرد که مقادیرشان در ستون یا ستون‌های مشترک برابر، یکسان و مساوی باشد. همچنین ستون یا ستون‌های مشترک فقط یکبار در خروجی ظاهر می‌شوند. اما اگر دو جدولی که با یکدیگر الحاق طبیعی می‌شوند دارای ستون یا ستون‌های مشترک نباشند، آنگاه عملگر الحاق طبیعی، دقیقاً مانند عملگر ضرب دکارتی رفتار خواهد کرد.

فرم کلی عملگر \bowtie به صورت زیر است:

$$R_3 = R_1 \bowtie R_2$$

در بخش R₁ و R₂ نام جدول یا جداول یا خروجی یک پرس و جو یعنی یک عبارت جبر رابطه‌ای دیگر می‌تواند قرار گیرد.

$$\rho(T, S \bowtie R)$$

در بخش دوم پرس و جوی Q₁، جدول S و R الحاق طبیعی می‌شود و خروجی به جدول T با ستون‌های سابق نام‌گذاری مجدد می‌شود، به صورت زیر:

a	b ₁	c ₁	⋈	a	b	c	=	a	b ₁	c ₁	b	c
1	2	2		1	2	2		1	2	2	2	2
3	4	4		3	4	4		3	4	4	4	4
5	6	7		5	6	7		5	6	7	6	7
5	7	6		5	7	6		5	6	7	7	6
جدول S				جدول R				5	7	6	6	7
								5	7	6	7	6
								جدول T				

توجه: ستون مشترک در الحاق طبیعی فوق فقط و فقط ستون a است. چون قبلا ستون b به b₁ و ستون c به c₁ نام گذاری مجدد شده است.

در بخش سوم از پرس و جوی Q₁:

$$\Pi_{b_1, c}(\sigma_{b_1=c} T)$$

عملگر σ جهت انتخاب سطر در یک جدول مورد استفاده قرار می گیرد.

فرم کلی عملگر σ به صورت زیر است:

$$R_2 = \sigma_{\theta}(R_1)$$

در بخش θ شرط انتخاب سطر به عنوان ملاک انتخاب سطر مشخص می گردد. همچنین در این بخش علائم ریاضی $=, >, <, \leq, \geq, \neq, \neg, \vee, \wedge$ می تواند مورد استفاده قرار گیرد.

در بخش R₁ نام جدول یا جداول به عنوان مکان یا محل پرس و جو مشخص می گردد، این مکان جستجو می تواند خروجی یک پرس و جوی دیگر یعنی یک عبارت جبر رابطه ای دیگر باشد.

عملگر Π جهت انتخاب ستون در یک جدول مورد استفاده قرار می گیرد.

فرم کلی عملگر Π به صورت زیر است:

$$R_2 = \Pi_L(R_1)$$

در بخش L مدل انتخاب ستون به عنوان ستون های مورد نیاز مشخص می گردد.

در بخش R₁ نام جدول یا جداول به عنوان مکان یا محل پرس و جو مشخص می گردد، این مکان جستجو می تواند خروجی یک پرس و جوی دیگر یعنی یک عبارت جبر رابطه ای دیگر باشد.

$$\Pi_{b_1, c}(\sigma_{b_1=c} T)$$

در بخش سوم پرس و جوی Q₁، ابتدا سطرهایی از جدول T بر اساس عملگر و شرط انتخاب سطر $\sigma_{b_1=c}$ انتخاب می شود، به صورت زیر:

a	b ₁	c ₁	b	c
1	2	2	2	2
3	4	4	4	4
5	6	7	7	6
5	7	6	6	7

سپس بر اساس عملگر و مدل انتخاب ستون $\Pi_{b_1,c}$ ، خروجی نهایی پرس و جوی Q_1 مشخص می‌شود، به صورت زیر:

b ₁	c
2	2
4	4
6	6
7	7

پرس و جوی Q_1 :

$$Q_2 : \Pi_{b,c}(\sigma_{b=c} R)$$

در پرس و جوی Q_2 ، ابتدا سطرهایی از جدول R بر اساس عملگر و شرط انتخاب سطر $\sigma_{b=c}$ انتخاب می‌شود، به صورت زیر:

a	b	c
1	2	2
3	4	4

سپس بر اساس عملگر و مدل انتخاب ستون $\Pi_{b,c}$ ، خروجی نهایی پرس و جوی Q_2 مشخص می‌شود، به صورت زیر:

b	c
2	2
4	4

واضح است که مابین پرس و جوی Q_1 و Q_2 رابطه $Q_2 \subseteq Q_1$ برقرار است، به صورت زیر:

b	c	\subseteq	b ₁	c
2	2		2	2
4	4		4	4
			6	6
			7	7
Q_2			Q_1	

شرط لازم برای زیرمجموعه بودن رعایت شروط سازگاری است، در جبر رابطه‌ای دو شرط به عنوان شروط سازگاری مطرح است:

شرط اول: تعداد ستون‌های دو جدول یکسان باشد، به عبارت دیگر دو رابطه (جدول) هم درجه باشند.

شرط دوم: نوع یا دامنه ستون‌های متناظر در دو جدول یکسان باشد. در نام‌گذاری مجدد، نوع ستون‌ها عوض نشده‌است، بنابراین نوع یا دامنه ستون‌های متناظر در دو جدول یکسان است.

اگر بخواهیم دو شرط فوق را در یک جمله بیان کنیم، اینطور خواهد بود، شروط سازگاری یعنی تیتراها در دو رابطه (جدول) یکسان باشد، به عبارت دیگر هم‌تیترا باشند.

و شرط کافی برای زیر مجموعه بودن، قرار داشتن کلیه سطرهای رابطه سمت چپ در رابطه سمت راست است.

بنابراین پرواضح است که گزینه سوم پاسخ سوال است.

تست‌های فصل نهم: نرمال‌سازی

۸۶- رابطه $R = (A, B, C, D, E, F, G)$ را با وابستگی‌های زیر در نظر بگیرید:

(مهندسی IT - دولتی ۱۴۰۰)

$AF \rightarrow BE$

$FC \rightarrow DE$

$F \rightarrow CD$

$D \rightarrow E$

$C \rightarrow A$

حاصل تجزیه 3NF این رابطه چند رابطه خواهد بود؟

۲ (۱)

۳ (۲)

۴ (۳)

۵ (۴)

عنوان کتاب: پایگاه داده‌ها

مؤلف: ارسطو خلیلی‌فر

ناشر: انتشارات راهیان ارشد

آدرس سایت گروه بابان: khalilifar.ir

پاسخ تست‌های فصل نهم: نرمال‌سازی

۸۶- گزینه (۲ و ۳) صحیح است.

به طور کلی کلید کاندید باید دو شرط زیر را داشته باشد:

۱- ابرکلید باشد (خاصیت کلیدی داشته باشد) یعنی همه خصیصه‌ها را تولید کند.

۲- عضو زائد نداشته باشد.

به طور کلی عضو کلید کاندید از روابط زیر بدست می‌آید:

قانون اول ارسطو

روش اول:

اجتماع تمام خصیصه‌های سمت راست وابستگی‌های غیربدیهی - تمام خصیصه‌های جدول = عضو کلید کاندید

روش دوم:

$$\text{عضو کلید کاندید} = R - \bigcup_{i=1}^n [x_i - (y_i \text{ (چپ) (راست)})]$$

با توجه به وابستگی‌های مطرح شده برای رابطه $R(A, B, C, D, E, F, G)$ داریم:

$AF \rightarrow BE$

$FC \rightarrow DE$

$F \rightarrow CD$

$D \rightarrow E$

$C \rightarrow A$

$$ABCDEF - ABCDE = FG$$

بنابر رابطه فوق صفات FG حتماً باید عضو کلید کاندید باشد. بستار صفات FG به صورت زیر است:

$$\{FG\}^+ = \{F, G, C, D, E, A, B\}$$

براساس بستار فوق، صفات FG ، همه ستون‌ها را تولید می‌کند، پس صفات FG کلید کاندید می‌باشد.

قانون دوم ارسطو

هرگاه عضو کلید کاندید، حاصل از تفاضل قانون اول (روش اول یا دوم)، همه ستون‌ها را تولید کند، آن عضو کلید کاندید، تنها کلید کاندید جدول خواهد بود.

جدولی در سطح نرمال فرم اول قرار دارد که صفت چند مقداری و مرکب نداشته باشد. از آنجا که در جدول R ، صفت چند مقداری و مرکب وجود ندارد، بنابراین این جدول در نرمال فرم اول قرار دارد.

با توجه به وابستگی‌های مطرح شده برای رابطه $R(A, B, C, D, E, F, G)$ داریم:

$AF \rightarrow BE$

$FC \rightarrow DE$

$F \rightarrow CD$

$D \rightarrow E$

$C \rightarrow A$

توجه: به طور کلی در نرمال‌سازی پس از کشف کلید کاندید، جهت کشف وابستگی بخشی و انتقالی و کنترل بهتر وابستگی‌ها می‌توان وابستگی‌ها را بهینه و کمینه کرد، اما با عدم حذف وابستگی‌های اصلی و عدم درج وابستگی‌ها فرعی. در اغلب موارد خود وابستگی‌ها گویای وابستگی بخشی و انتقالی هستند و نیاز به کمینه‌سازی وابستگی‌ها هم نیست. مهمترین عنصر نرمال‌سازی صحیح، کشف دقیق و درست کلید کاندید توسط قوانین چهارگانه است.

توجه: در وابستگی $AF \rightarrow BE$ از وابستگی‌های $F \rightarrow A$ $\left\{ \begin{array}{l} F \rightarrow C \\ C \rightarrow A \end{array} \right.$ ستون F می‌دهد A را، پس درست چپ وابستگی $AF \rightarrow BE$ ستون A زائد است، چون F می‌تواند خود A را تولید کند، پس به جای وابستگی $AF \rightarrow BE$ می‌توان وابستگی کمینه $F \rightarrow BE$ را جایگزین کرد.

توجه: در وابستگی $FC \rightarrow DE$ از وابستگی $F \rightarrow CD$ ستون F می‌دهد C را، پس درست چپ وابستگی $FC \rightarrow DE$ ستون C زائد است، چون F می‌تواند خود C را تولید کند، پس به جای وابستگی $FC \rightarrow DE$ می‌توان وابستگی کمینه $F \rightarrow DE$ را جایگزین کرد.

پس از حذف وابستگی‌های فرعی و قابل تولید از سایر وابستگی‌ها، وابستگی‌های کمینه به صورت زیر است:

$F \rightarrow B$

$F \rightarrow C$

$F \rightarrow D$

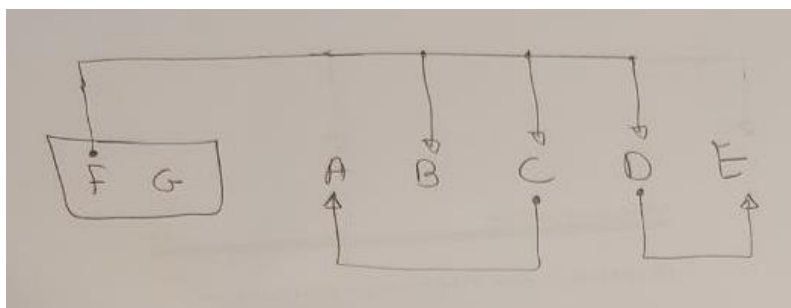
$D \rightarrow E$

$C \rightarrow A$

توجه: بر اساس وابستگی‌ها فوق واضح است که صفت F همه ستون‌ها را به طور مستقیم و غیرمستقیم بجز صفت G را تولید می‌کند، بستار صفت F به صورت زیر است:

$$\{F\}^+ = \{F, C, D, E, A, B\}$$

نمودار وابستگی‌های جدول R به صورت زیر است:



به طور کلی می‌توان شروط قرار داشتن یک جدول در نرمال فرم دوم را به صورت زیر بیان کرد:

- جدول باید در نرمال فرم اول باشد.
- جدول باید فاقد وابستگی بخشی باشد.

وابستگی بخشی: وابستگی یک مؤلفه غیرکلیدی، به جزئی از کلید کاندید را وابستگی بخشی می‌نامند. با توجه به این که ترکیب صفات FG کلید کاندید رابطه R می‌باشد، وابستگی‌های تابعی زیر:

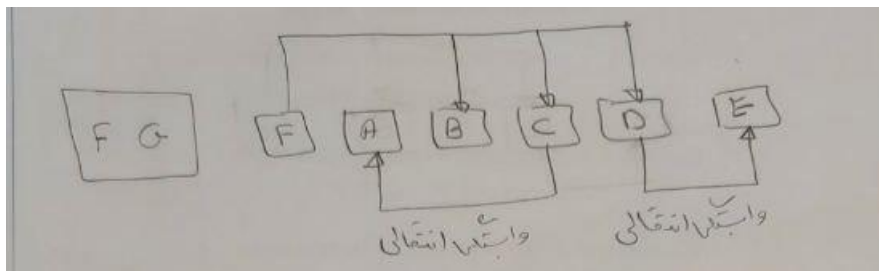
$F \rightarrow B$

$F \rightarrow C$

$F \rightarrow D$

به عنوان وابستگی بخشی محسوب می‌گردند. بنابراین جدول R در نرمال فرم دوم قرار ندارد.

بنابراین برای قرار دادن جدول R، در نرمال فرم دوم، باید وابستگی‌های بخشی از جدول R خارج گردند.



بنابراین جدول R به دو جدول کوچکتر تجزیه می‌گردد:

$R_1(F, G)$ $R_2(F, A, B, C, D, E)$

با توجه به کلید کاندید FG در جدول R_1 می‌توان دریافت که این جدول فاقد وابستگی بخشی و انتقالی است، بنابراین در نرمال فرم سوم قرار دارد. همچنین از آن جا که سمت چپ تمام وابستگی‌های جدول R_1 ، ابرکلید است، بنابراین جدول R_1 در سطح BCNF نیز قرار دارد. از نگاه دیگر می‌توان گفت جدول R_1 تمام کلید است پس BCNF است، از منظری دیگر نیز می‌توان گفت هر جدول با 2 ستون همواره و تحت هر شرایطی قطعاً BCNF است. مطابق خواسته سوال راه حل را تا سطح 3NF برای جدول R_2 ادامه می‌دهیم.

به طور کلی می‌توان شروط قرار داشتن یک جدول در نرمال فرم سوم را به صورت زیر بیان کرد:

- جدول باید در نرمال فرم دوم باشد.
- جدول باید فاقد وابستگی انتقالی باشد.

وابستگی انتقالی: وابستگی یک مؤلفه غیرکلیدی به یک مؤلفه غیرکلیدی دیگر را وابستگی انتقالی می‌نامند.

در جدول R_2 وابستگی تابعی زیر:

$D \rightarrow E$

$C \rightarrow A$

به عنوان وابستگی انتقالی محسوب می‌گردد. بنابراین جدول R_2 در نرمال فرم سوم قرار ندارد. بنابراین برای قرار دادن جدول R_2 ، در نرمال فرم سوم، باید وابستگی‌های انتقالی از جدول R_2 خارج گردند.



بنابراین جدول R_2 به سه جدول کوچکتر تجزیه می‌گردد.

$R_{21}(F, B, C, D)$ $R_{22}(C, A)$ $R_{23}(D, E)$

با توجه به کلیدهای کاندید در سه جدول R_{21} و R_{22} و R_{23} می‌توان دریافت که این سه جدول فاقد وابستگی بخشی و انتقالی هستند و بنابراین در نرمال فرم سوم قرار دارند. همچنین از آن جا که سمت چپ تمام وابستگی‌های سه جدول R_{21} و R_{22} و R_{23} ، ابرکلید است، بنابراین سه جدول R_{21} و R_{22} و R_{23} در سطح BCNF نیز قرار دارند. به این ترتیب با تجزیه جدول R به چهار جدول R_1 ، R_{21} ، R_{22} و R_{23} می‌توان جدول R را در نرمال فرم سوم و سطح BCNF قرار داد.

توجه: سازمان سنجش آموزش کشور در کلید اولیه خود ابتدا گزینه دوم را به عنوان پاسخ اعلام نمود، سپس در کلید نهایی نظر خود را عوض کرد و گزینه دوم و سوم را به عنوان پاسخ اعلام کرد، که عمل نادرستی را انجام داده است. چون فقط و فقط همان گزینه سوم درست و پاسخ سوال است.