

データベース概説 (第5回)

2008.05.26

宇陀則彦

1

リレーショナルデータベースの設計

- スキーマ設計
 - 設計がよくないと、更新時に異常が起きる。
- リレーションの正規化
 - 異常が起きないようにリレーションを分解する。
 - **One Fact One Placeの原則** (複数の事実を一つのリレーションに入れると不整合が起きやすくなる。)
 - キーと関数従属性を元に形式化

2

重複≠冗長

名前	Tel
荒井	03-1111-2222
飯塚	0298-33-4444
上田	026-99-0000

電話番号を複数持っている場合は、リレーションにどのように格納するか。

2つ以上のデータを入れることは許されない。

名前	Tel
荒井	03-1111-2222, 03-1111-2233
飯塚	0298-33-4444, 090-555-66666
上田	026-99-0000

3

重複≠冗長

名前	電話番号
荒井	03-1111-2222
荒井	03-1111-2233
飯塚	0298-33-4444
飯塚	0298-77-8888
上田	026-99-0000

名前は重複しているが、冗長ではない。

一つの事実が一つのタブルに書かれているから。

例えば、一番上の荒井を消すと「荒井が03-1111-2222を持つ」という事実が見失われてしまう。

どの値を消しても一つの事実が見失われる。
→冗長ではない。

4

重複≠冗長

連絡先

名前	電話番号	住所
荒井	03-1111-2222	東京
荒井	03-1111-2233	東京
飯塚	0298-33-4444	茨城
飯塚	0298-77-8888	茨城
上田	026-99-0000	長野

住所の「東京」が重複。
一つの「東京」を消しても残り一つから荒井の住所は判明

↓
「東京」の重複は冗長

「荒井の住所は東京」という一つの事実が2つのタブルに書かれている。

荒井が千葉へ引っ越ししたら、2箇所の「東京」の両方を書き換えなければならない
更新時に余計な手間。

5

リレーションを「分解」:
射影を使って二つに分ける。

名前	電話番号	住所
荒井	03-1111-2222	東京
荒井	03-1111-2233	東京
飯塚	0298-33-4444	茨城
飯塚	0298-77-8888	茨城
上田	026-99-0000	長野

名前と電番で射影

名前と住所で射影

名前	電話番号
荒井	03-1111-2222
荒井	03-1111-2223
飯塚	0298-33-4444
飯塚	0298-77-8888
上田	026-99-0000

名前	住所
荒井	東京
飯塚	茨城
上田	長野

「住所」の冗長な重複が消えた!

リレーションの設計

- 更新時異常
 - タップル挿入時異常
 - タップル削除時異常
 - タップル修正時異常
- 関数従属性
 - 公理系
 - 閉包
- 多値従属性

7

更新時異常の例

顧客名	商品名	数量	単価	金額
丸八	テレビ	3	19800	59400
亀屋	テレビ	10	19800	198000
花商	洗濯機	5	55800	279000
杉金	ビデオ	1	29800	29800

- **タップル挿入時異常**
 - 電子レンジを ¥32000- で扱うことにした
 - 顧客はまだ決まっていない。(主キー制約に違反)

8

更新時異常の例

顧客名	商品名	数量	単価	金額
丸八	テレビ	3	19800	59400
亀屋	テレビ	10	19800	198000
花商	洗濯機	5	55800	279000
杉金	ビデオ	1	29800	29800

- **タップル削除時異常**
 - 花商から洗濯機の注文が取り消された
 - 洗濯機を扱っている事実と単価のデータが消滅

9

更新時異常の例

顧客名	商品名	数量	単価	金額
丸八	テレビ	3	19800	59400
亀屋	テレビ	10	19800	198000
花商	洗濯機	5	55800	279000
杉金	ビデオ	1	29800	29800

- **タップル修正時異常**
 - テレビの単価を ¥15800- に値下げした
 - リレーション中の単価と金額をタップル毎に修正
 - 花商が注文を間違えていた(本当はテレビだった。)
 - 洗濯機のデータが消滅

10

分解

- 更新時異常が起きないように分解する。
- よい分解と悪い分解がある。
 - よい分解
 - 自然結合したら元のリレーションに戻る。
 - **「情報無損失分解」**
 - 悪い分解
 - 自然結合しても元のリレーションに戻らない。

11

情報無損失分解の例

取引

顧客名	商品名	数量	単価	金額
丸八	テレビ	3	19800	59400
亀屋	テレビ	10	19800	198000
花商	洗濯機	5	55800	279000
杉金	ビデオ	1	29800	29800

伝票

顧客名	商品名	数量	金額
丸八	テレビ	3	59400
亀屋	テレビ	10	198000
花商	洗濯機	5	279000
杉金	ビデオ	1	29800

商品

商品名	単価
テレビ	19800
洗濯機	55800
ビデオ	29800

12

情報無損失分解の例

伝票				商品	
顧客名	商品名	数量	金額	商品名	単価
丸八	テレビ	3	59400	テレビ	19800
亀屋	テレビ	10	198000	洗濯機	55800
花商	洗濯機	5	279000	ビデオ	29800
杉金	ビデオ	1	29800		

取引				
顧客名	商品名	数量	単価	金額
丸八	テレビ	3	19800	59400
亀屋	テレビ	10	19800	198000
花商	洗濯機	5	55800	279000
杉金	ビデオ	1	29800	29800

13

情報無損失分解ではない例

顧客名	商品名	数量	単価	金額
丸八	テレビ	3	19800	59400
亀屋	テレビ	10	19800	198000
花商	洗濯機	5	55800	279000
杉金	ビデオ	1	29800	29800

顧客名	商品名
丸八	テレビ
亀屋	テレビ
花商	洗濯機
杉金	ビデオ

商品名	数量	単価	金額
テレビ	3	19800	59400
テレビ	10	19800	198000
洗濯機	5	55800	279000
ビデオ	1	29800	29800

14

情報無損失分解ではない例

顧客名	商品名
丸八	テレビ
亀屋	テレビ
花商	洗濯機
杉金	ビデオ

商品名	数量	単価	金額
テレビ	3	19800	59400
テレビ	10	19800	198000
洗濯機	5	55800	279000
ビデオ	1	29800	29800

顧客名	商品名	数量	単価	金額
丸八	テレビ	3	19800	59400
丸八	テレビ	10	19800	198000
亀屋	テレビ	3	19800	59400
亀屋	テレビ	10	19800	198000
花商	洗濯機	5	55800	279000
杉金	ビデオ	1	29800	29800

これは？

15

分解の定義

- リレーションスキーマ $RS = \{A_1, \dots, A_n\}$ に対して, RS_i $RS(1 \leq i \leq m)$ なるリレーションスキーマの集合 $\rho = \{RS_1, \dots, RS_m\}$ のうち,

$$RS_1 \dots RS_m = RS$$

となるものをRSの分解(decomposition)と呼ぶ

- このような ρ を求めることをRSを分解するという

16

情報無損失分解の定義

- RSに関する関数従属性集合FとRSの分解 $\rho = \{RS_1, \dots, RS_m\}$ が与えられたとき, F を満足するRSの全てのインスタンスRについて

$$R = \pi_{RS_1}(R) \bowtie \dots \bowtie \pi_{RS_m}(R)$$

が成立するとき,

分解 ρ はRSの情報無損失分解である

という。

17

情報無損失で分解するには？

関数従属性をうまく使って分解すれば
情報無損失分解できる。

18

関数従属性 (functional dependency)

リレーション R の属性集合 X, Y に
関数従属性 $X \rightarrow Y$ が存在する

$$(\forall t, t' \in R)(t[X] = t'[X] \Rightarrow t[Y] = t'[Y])$$

「属性 X の値が決まると属性 Y の値もひとつに決まる」を
「 Y は X に関数的に従属する」と言う。

X を主体にして言うと「 X は Y を関数的に決定する」となる。

記号: $X \rightarrow Y$ X を決定項、 Y を従属項という。

(注意: X や Y が一つの属性でなく、属性の組の場合もある。)

19

関数従属性の例

「営業」のリレーションスキーマ

{商品番号, 顧客番号, 社員番号, 販売価格}

においては,

- {商品番号, 顧客番号} \rightarrow {販売価格}
- {顧客番号} \rightarrow {社員番号} などが成立

({} は省略することがある。)

20

関数従属性の例

学籍番号	科目	得点	評価	入学年
99101	情報科学	60	可	1999
2000209	経営学原論	90	優	2000
2001124	情報科学	75	可	2001
2001195	情報科学	85	良	2001
2001204	経済学概論	50	不可	2001

{学籍番号, 科目} \rightarrow {得点}

{科目, 得点} \rightarrow {評価}

{学籍番号} \rightarrow {入学年}

21

関数従属性と情報無損失分解

R1

A	B	C	D
a1	b1	f(a1)	d1
a1	b2	f(a1)	d2
a2	b1	f(a2)	d3
a2	b2	f(a2)	d4

R2

A	C
a1	f(a1)
a2	f(a2)

R3

A	B	D
a1	b1	d1
a1	b2	d2
a2	b1	d3
a2	b2	d4

リレーション $R1(A, B, C, D)$ で関数従属性 $A \rightarrow C$ があるとする。
すると、**CはAの値により決まるので $c=f(a)$ と書いてよい。**

AとCの関数従属性は、R2に保存され、A, B, Dの関係はR3に残る。
→自然結合すれば、元のR1ができる。

大雑把に言えば、関数従属性を保存すればよい。²²

情報無損失分解と関数従属性の関係 (詳細)

- $\rho = \{RS_1, RS_2\}$ をリレーションスキーマ RS の分解とする。

- F を RS に関する関数従属性集合とする

このとき、分解 ρ が情報無損失分解となる十分条件は、次の通りである。

$RS_1 \cap RS_2 \rightarrow RS_1 - RS_2$, もしくは、
 $RS_1 \cap RS_2 \rightarrow RS_2 - RS_1$ が成立

23

情報無損失分解の例

- リレーションスキーマ $RS = \{\text{商品番号, 顧客番号, 社員番号, 販売価格}\}$

- $F = \{\text{商品番号, 顧客番号} \rightarrow \text{販売価格}, \text{顧客番号} \rightarrow \text{社員番号}\}$

とすると,

{商品番号, 顧客番号, 販売価格}, {顧客番号, 社員番号} は情報無損失分解だが、

{商品番号, 顧客番号}, {商品番号, 社員番号, 販売価格} は情報無損失分解ではない。

24

分解

- 更新時異常が起きないように分解する。
- よい分解と悪い分解がある。
 - よい分解
 - 自然結合したら元のリレーションに戻る。
 - 「情報無損失分解」
 - 悪い分解
 - 自然結合しても元のリレーションに戻らない。

25

正規化の理論

- 更新時異常が起きないリレーションとは？
- 情報無損失分解で分解するには？
 - リレーションの正規化
 - 正規化に基づく分解は情報無損失分解になる。
 - 更新異常を起こさない。
 - 高次の正規形(どこまでやるか)
 - 第一正規形、第二正規形、第三正規形
 - ボイス・コッド正規形、第四正規形、第五正規形

26

第一正規形(1st Normal Form)

ドメインはシンプルでなければならない。

シンプルとは、ドメインが他のドメインの直積になっていたり、ドメインのべき集合であってはならないということ。

27

第一正規形ではないリレーション

支店名	所在	商品	数量	単価	金額	合計	取引日		
							年	月	日
みなと	横浜	S T	3 2	1,000 950	3,000 1,900	4,900	2004	09	24
おえど	東京	S U V	1 10 5	1,000 1,200 1,800	1,000 12,000 9,000	22,000	2004	09	28
はかた	福岡	T	8	950	7,600	7,600	2004	10	18
なにわ	大阪	U	25	1,200	30,000	30,000	2004	11	06

28

第一正規形における更新時異常

顧客名	商品名	数量	単価	金額
丸八	テレビ	3	19800	59400
亀屋	テレビ	10	19800	198000
花商	洗濯機	5	55800	279000
杉金	ビデオ	1	29800	29800

- タプル挿入時異常
 - 電子レンジを ¥32000- で扱うことにした
 - 顧客はまだ決まっていない。(主キー制約に違反)

29

第一正規形における更新時異常

顧客名	商品名	数量	単価	金額
丸八	テレビ	3	19800	59400
亀屋	テレビ	10	19800	198000
花商	洗濯機	5	55800	279000
杉金	ビデオ	1	29800	29800

- タプル削除時異常
 - 花商から洗濯機の注文が取り消された
 - 洗濯機を扱っている事実と単価のデータが消滅

30

第一正規形における更新時異常の例

顧客名	商品名	数量	単価	金額
丸八	テレビ	3	19800	59400
亀屋	テレビ	10	19800	198000
花商	洗濯機	5	55800	279000
杉金	ビデオ	1	29800	29800

- **タプル修正時異常**
 - テレビの単価を¥15800-に値下げした
 - リレーション中の単価と金額をタプル毎に修正
 - 花商が注文を間違えていた(本当はテレビだった。)
 - 洗濯機のデータが消滅

31

第二正規形(2nd Normal Form)

R は第一正規形である。

R のすべての非キー属性は R の各候補キーに完全関数従属している。

32

完全関数従属性

- $X \rightarrow Y$ をリレーション R の関数従属性とする。
- X の真部分集合 X' に対して $X' \rightarrow Y$ が成立しない。

{学籍番号, 科目} \rightarrow {得点}
 {科目, 得点} \rightarrow {評価}

33

第二正規形に分解

顧客名	商品名	数量	単価	金額
丸八	テレビ	3	19800	59400
亀屋	テレビ	10	19800	198000
花商	洗濯機	5	55800	279000
杉金	ビデオ	1	29800	29800

- ドメインが単純 (第一正規形)
- 候補キー {顧客名, 商品名}
 - {顧客名, 商品名} \rightarrow 数量 (完全関数従属)
 - {顧客名, 商品名} \rightarrow 金額 (完全関数従属)
 - {顧客名, 商品名} \rightarrow 単価
 - 商品名 \rightarrow 単価 (完全関数従属)

34

第二正規形に分解

顧客名	商品名	数量	金額
丸八	テレビ	3	59400
亀屋	テレビ	10	198000
花商	洗濯機	5	279000
杉金	ビデオ	1	29800

商品名	単価
テレビ	19800
洗濯機	55800
ビデオ	29800

- 候補キー {顧客名, 商品名}
 - {顧客名, 商品名} \rightarrow 数量 (完全関数従属)
 - {顧客名, 商品名} \rightarrow 金額 (完全関数従属)
- 候補キー {商品名}
 - 商品名 \rightarrow 単価 (完全関数従属)

35

第二正規形に分解

顧客名	商品名	数量	金額
丸八	テレビ	3	59400
亀屋	テレビ	10	198000
花商	洗濯機	5	279000
杉金	ビデオ	1	29800

商品名	単価
テレビ	19800
洗濯機	55800
ビデオ	29800

- **タプル挿入時異常**
 電子レンジを¥32000-で扱うことにした。顧客はまだ決まっていない。(主キー制約に違反)
- **タプル修正時異常**
 - テレビの単価を¥15800-に値下げした
 - リレーション中の単価と金額をタプル毎に修正
 - 花商が注文を間違えていた(本当はテレビだった。)
 洗濯機のデータが消滅
- **タプル削除時異常**
 花商から洗濯機の注文が取り消された。洗濯機を扱っている事実と単価のデータが消滅

第二正規形

ID	社員名	給与	所属	勤務地
65	鈴木	50	K55	神奈川
58	濱田	40	K41	東京
23	宇田	60	K41	東京
34	佐藤	40	K55	神奈川
21	石井	40	K58	静岡

- 候補キー {ID}
 - ID → 社員名
 - ID → 給与
 - ID → 所属
 - ID → 勤務地

非キー属性は候補キーに完全関数従属 → 第二正規形

37

第二正規形

ID	社員名	給与	所属	勤務地
65	鈴木	50	K55	神奈川
58	濱田	40	K41	東京
23	宇田	60	K41	東京
34	佐藤	40	K55	神奈川
21	石井	40	K58	静岡

- タプル挿入時異常
 - 新しい部門 K45 (所在地: 千葉)
 - 主キー制約に抵触

38

第二正規形

ID	社員名	給与	所属	勤務地
65	鈴木	50	K55	神奈川
58	濱田	40	K41	東京
23	宇田	60	K41	東京
34	佐藤	40	K55	神奈川
21	石井	40	K58	静岡

- タプル削除時異常
 - 石井さんが退職
 - K58 (所在地: 静岡) が消滅

39

第二正規形

ID	社員名	給与	所属	勤務地
65	鈴木	50	K55	神奈川
58	濱田	40	K41	東京
23	宇田	60	K41	東京
34	佐藤	40	K55	神奈川
21	石井	40	K58	静岡

- タプル修正時異常
 - K41が埼玉に引っ越し
 - 所属がK41のタプルの勤務地をすべて埼玉に書き換え
 - 石井さんがK55に配置換え
 - K58 (所在地: 静岡) が消滅

40

第二正規形

ID	社員名	給与	所属	勤務地
65	鈴木	50	K55	神奈川
58	濱田	40	K41	東京
23	宇田	60	K41	東京
34	佐藤	40	K55	神奈川
21	石井	40	K58	静岡

- 候補キー {ID}
 - ID → 社員名
 - ID → 給与
 - ID → 所属
 - ID → 勤務地

非キー属性は候補キーに完全関数従属 → 第二正規形

$(ID \rightarrow 所属) \wedge (所属 \rightarrow 勤務地) \Rightarrow ID \rightarrow 勤務地$

42

第三正規形(3rd Normal Form)

R は第二正規形である
 R のすべての非キー属性は、
 R のいかなる候補キーにも
 推移的に従属しない

第三正規形

ID	社員名	給与	所属	勤務地
65	鈴木	50	K55	神奈川
58	濱田	40	K41	東京
23	宇田	60	K41	東京
34	佐藤	40	K55	神奈川
21	石井	40	K58	静岡

ID	社員名	給与	所属
65	鈴木	50	K55
58	濱田	40	K41
23	宇田	60	K41
34	佐藤	40	K55
21	石井	40	K58

所属	勤務地
K55	神奈川
K41	東京
K55	神奈川
K58	静岡

43

第三正規形

ID	社員名	給与	所属	勤務地
65	鈴木	50	K55	神奈川
58	濱田	40	K41	東京
23	宇田	60	K41	東京
34	佐藤	40	K55	神奈川
21	石井	40	K58	静岡

第二正規形の更新時異常が解消されたか確認してみよう。

ID	社員名	給与	所属
65	鈴木	50	K55
58	濱田	40	K41
23	宇田	60	K41
34	佐藤	40	K55
21	石井	40	K58

所属	勤務地
K55	神奈川
K41	東京
K55	神奈川
K58	静岡

44

まとめ：正規形の考え方

- リレーションをうまく分解することにより、不具合をなくす。
- 分解したとしても、自然結合をおこなうことにより分解前の情報を再現できる(無損失結合分解)
- 正規形は、「従属性」を用いることにより行なわれる。

45

正規化の手順

- 属性値が単純な値であること。(第一正規形)
 - データに繰り返しが無い。
 - 入れ子になっていない。
- 関数従属性を洗い出す。
- 部分従属を除く。(第二正規形)
- 推移従属を除く。(第三正規形)

46

正規化の手順(例)

履修科目(科目名、担当名、単位数、教科書名、価格)

47

正規化の手順(例)

履修科目(科目名、担当名、単位数、教科書名、価格)

これを第2正規形にするには、
部分従属「科目名→単位数」を分解で他へ移し、除く。

科目(科目名、単位数)

履修科目(科目名、担当名、教科書名、価格)

48

正規化の手順(例)

科目(科目名、単位数)
履修科目(科目名、担当名、教科書名、価格)

推移従属「教科書名→価格」を分解で移す。

科目(科目名、単位数)
科目担当(科目名、担当名、教科書名)
教科書(教科書名、価格)

49

正規化の手順(練習)

問1 第三正規形になるように分解せよ。
支店(支店名、課名、所在地、支代電番、課長社番)

問2 第三正規形になるように分解せよ。
社員(社員番号、社員名、入社日、所属支店、所在地、支代電番)

50