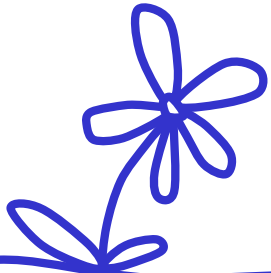


300秒 で 高次正規形 を
紐解け…なさそう

2010/09/03 おーがみ





自己紹介

- 👉 ビズアイユ こと おーがみ。
- 👉 表の顔はBIエンジニア、裏の顔は秘密。
- 👉 結構シブなDB2エキスパート資格持ち。
(↑でも、業務では全然触ってない…)
- 👉 残りの情報は二次会でどうぞ。





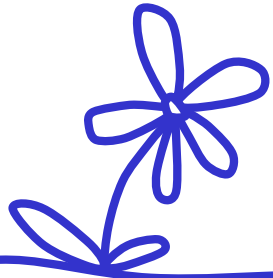
高次正規形とは？

一般的には、**第四正規形以降**のCodd氏以外が定義した正規形を指す模様。

Normal form	Defined by
First normal form (1NF)	Two versions: E.F. Codd (1970), C.J. Date (2003) ^[12]
Second normal form (2NF)	E.F. Codd (1971) ^[13]
Third normal form (3NF)	E.F. Codd (1971) ^[14] ; see +also Carlo Zaniolo's equivalent but differently-expressed definition (1982) ^[15]
Boyce-Codd normal form (BCNF)	Raymond F. Boyce and E.F. Codd (1974) ^[16]
Fourth normal form (4NF)	Ronald Fagin (1977) ^[17]
Fifth normal form (5NF)	Ronald Fagin (1979) ^[18]
Domain/key normal form (DKNF)	Ronald Fagin (1981) ^[19]
Sixth normal form (6NF)	C.J. Date, Hugh Darwen, and Nikos Lorentzos (2002) ^[5]

今回の主トピックは
DKNFと6NFです



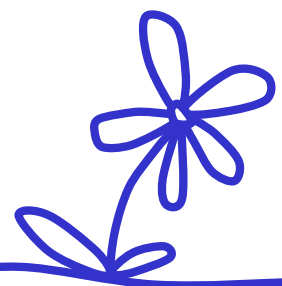


関係モデルのおさらい

- 関係：属性の定義域の直積の部分集合
- 組：関係の属性値の一つの組み合わせ
- 集合：一意に識別できる要素の集まり
- 候補キー：関係の中で組を一意的に識別できる最小構成の属性

候補キーのないテーブルは
厳密には関係と言えない





第一正規形のおさらい

定義：関係Rの全ての属性が
 スカラ値（単一）であり、
 繰り返し項目を持たない。

Not NF

ID	First Name	Family Name	Telephone Number
1	Robert	Ingram	555-861-2025
2	Jane	Wright	555-403-1659, 555-403-1660, 555-776-4100
3	Maria	Fernandez	555-808-9633

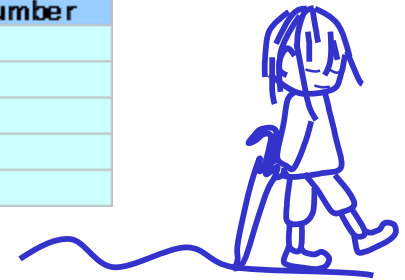
Date氏的には
 これも非正規形

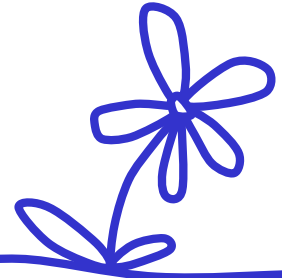
ID	First Name	Family Name	Tel #1	Tel #2	Tel #3
1	Robert	Ingram	555-861-2025		
2	Jane	Wright	555-403-1659	555-403-1660	555-776-4100
3	Maria	Fernandez	555-808-9633		

Not NF

1NF

ID	Line	First Name	Family Name	Telephone Number
1	1	Robert	Ingram	555-861-2025
2	1	Jane	Wright	555-403-1659
2	2	Jane	Wright	555-403-1660
2	3	Jane	Wright	555-776-4100
3	1	Maria	Fernandez	555-808-9633





第二正規形のおさらい

定義：第一正規形の条件を満たし、
 全ての非キー属性が候補キーに
 部分関数従属していない。

1NF

Employee	Skill	Department	Current Work Location
Jones	Typing	General Affairs	114 Main Street
Jones	Shorthand	General Affairs	114 Main Street
Jones	Whittling	General Affairs	114 Main Street
Bravo	Light Cleaning	Production	73 Industrial Way
Ellis	Alchemy	Production	73 Industrial Way
Ellis	Flying	Production	73 Industrial Way
Harrison	Light Cleaning	Production	73 Industrial Way

候補キーが複数の属性で
 構成される場合のみ
 部分関数従属が起こる

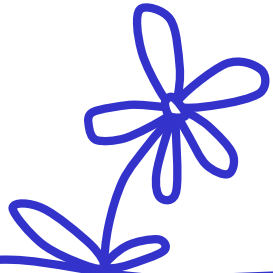
2NF

Employee	Department	Current Work Location
Jones	General Affairs	114 Main Street
Bravo	Production	73 Industrial Way
Ellis	Production	73 Industrial Way

Employee	Skill
Jones	Typing
Jones	Shorthand
Jones	Whittling
Bravo	Light Cleaning
Ellis	Alchemy
Ellis	Flying
Harrison	Light Cleaning

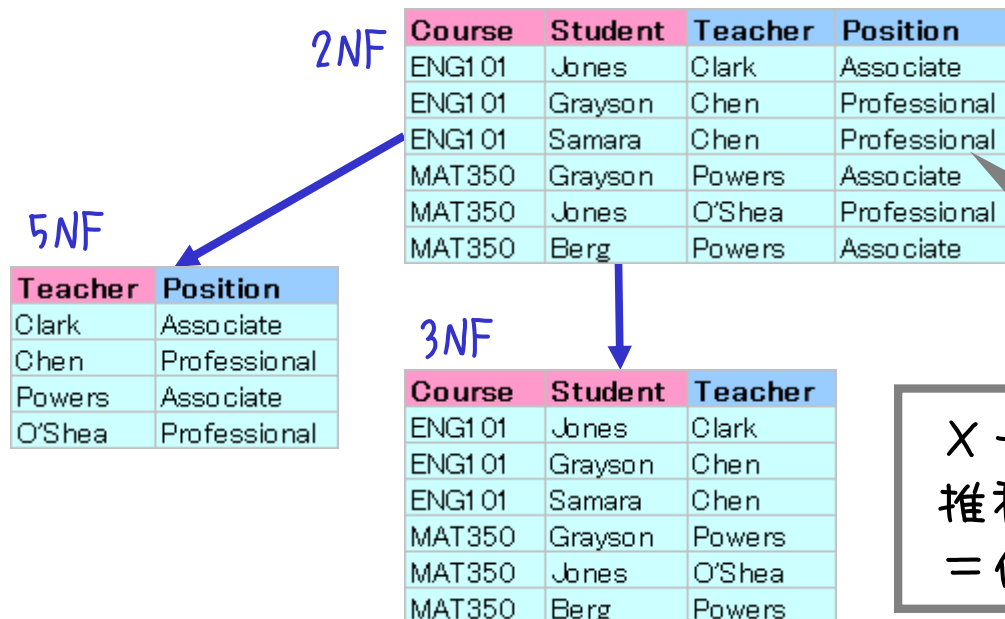
5NF





第三正規形のおさらい

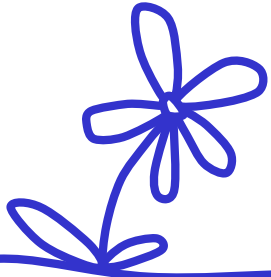
定義：第二正規形の条件を満たし、
全ての非キー属性が候補キーに
推移的関数従属していない。



非キー属性間に関数従属
= 推移的関数従属アリ

$X \rightarrow Y \rightarrow Z$ でも $Y \rightarrow X$ なら
推移的関数従属にならない
= 代理キーは使ってもOK





属性の従属性のおさらい

⑤ 関数従属性：

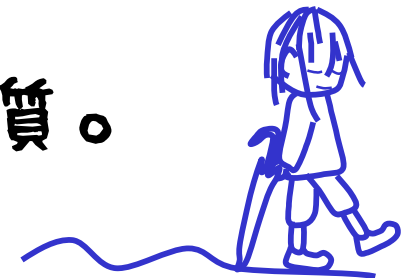
属性Xの値が決まると
属性Yの値も一意に定まる性質。

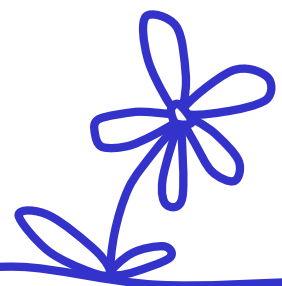
⑤ 多値従属性：

属性Xの値が決まると
属性Yの値も複数に定まる性質。

⑤ 結合従属性：

属性Xによる結合を行っても
属性Yの情報量が増加しない性質。





Boyce-Codd正規形のおさらい

定義：第三正規形の条件を満たし、
 非キー属性から候補キーの
 一部分に対する関数従属がない。

3NF

Course	Student	Teacher
ENG101	Jones	Clark
ENG101	Grayson	Chen
ENG101	Samara	Chen
MAT350	Grayson	Powers
MAT350	Jones	O'Shea
MAT350	Berg	Powers

{講座、学生} → 講師
 講師 → 講座

結合従属性を維持して
 正規化を行うには
 関係を三つに分ける

Course	Student
ENG101	Jones
ENG101	Grayson
ENG101	Samara
MAT350	Grayson
MAT350	Jones
MAT350	Berg

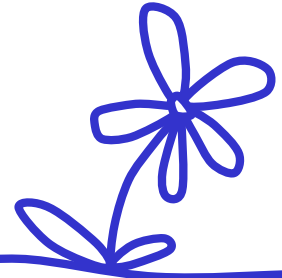
Student	Teacher
Jones	Clark
Grayson	Chen
Samara	Chen
Grayson	Powers
Jones	O'Shea
Berg	Powers

Teacher	Course
Clark	ENG101
Chen	ENG101
Powers	MAT350
O'Shea	MAT350

BCNF (5NF)

元の関係にあった
 関数従属性
 (ビジネスルール)
 は失われる





第四正規形のおさらい

定義：BCNFの条件を満たし、
 対象性のある（複数の）
 多値従属性を含まない。

BCNF

Restaurant	Pizza Variety	Delivery Area
A1 Pizza	Thick Crust	Springfield
A1 Pizza	Thick Crust	Shelbyville
A1 Pizza	Thick Crust	Capital City
A1 Pizza	Stuffed Crust	Springfield
A1 Pizza	Stuffed Crust	Shelbyville
A1 Pizza	Stuffed Crust	Capital City
Elite Pizza	Thin Crust	Capital City
Elite Pizza	Stuffed Crust	Capital City
Vincenzo's Pizza	Thick Crust	Springfield
Vincenzo's Pizza	Thick Crust	Shelbyville
Vincenzo's Pizza	Thin Crust	Springfield
Vincenzo's Pizza	Thin Crust	Shelbyville

Restaurant	Delivery Area
A1 Pizza	Springfield
A1 Pizza	Shelbyville
A1 Pizza	Capital City
Elite Pizza	Capital City
Vincenzo's Pizza	Springfield
Vincenzo's Pizza	Shelbyville

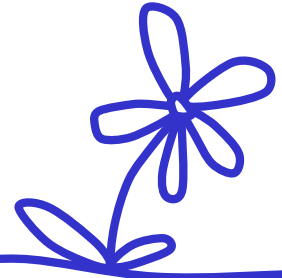
レストランによって
 配達できる地域が
 複数に定まる

Restaurant	Pizza Variety
A1 Pizza	Thick Crust
A1 Pizza	Stuffed Crust
Elite Pizza	Thin Crust
Elite Pizza	Stuffed Crust
Vincenzo's Pizza	Thick Crust
Vincenzo's Pizza	Thin Crust

レストランによって
 作るピザの種類が
 複数に定まる

4NF





第五正規形のおさらい

定義：BCNFの条件を満たし、
 自明でない結合従属性を含まない。

BCNF

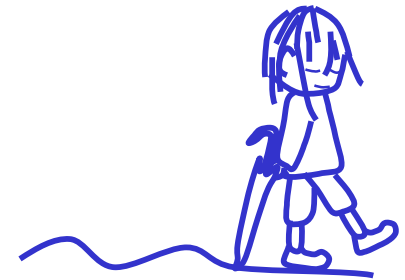
Restaurant	Pizza Variety	Delivery Area
A1 Pizza	Thick Crust	Springfield
A1 Pizza	Thick Crust	Shelbyville
A1 Pizza	Thick Crust	Capital City
A1 Pizza	Stuffed Crust	Springfield
A1 Pizza	Stuffed Crust	Shelbyville
A1 Pizza	Stuffed Crust	Capital City
Elite Pizza	Thin Crust	Capital City
Elite Pizza	Stuffed Crust	Capital City
Vincenzo's Pizza	Thick Crust	Springfield
Vincenzo's Pizza	Thick Crust	Shelbyville
Vincenzo's Pizza	Thin Crust	Springfield
Vincenzo's Pizza	Thin Crust	Shelbyville

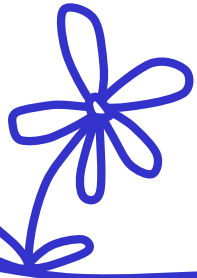
Restaurant	Delivery Area
A1 Pizza	Springfield
A1 Pizza	Shelbyville
A1 Pizza	Capital City
Elite Pizza	Capital City
Vincenzo's Pizza	Springfield
Vincenzo's Pizza	Shelbyville

Pizza Variety	Delivery Area
Thick Crust	Springfield
Thick Crust	Shelbyville
Thick Crust	Capital City
Stuffed Crust	Springfield
Stuffed Crust	Shelbyville
Stuffed Crust	Capital City
Thin Crust	Capital City
Thin Crust	Springfield
Thin Crust	Shelbyville

Restaurant	Pizza Variety
A1 Pizza	Thick Crust
A1 Pizza	Stuffed Crust
Elite Pizza	Thin Crust
Elite Pizza	Stuffed Crust
Vincenzo's Pizza	Thick Crust
Vincenzo's Pizza	Thin Crust

5NF





ここまでのまとめ

👉 候補キーを単一の属性で構成すると、関係をシンプルに正規化できる。

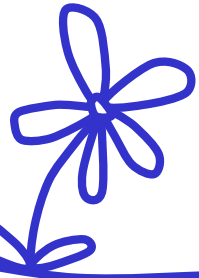
☆ 候補キーが単一の属性の場合、

Not NFか 2NFか 5NFにしかない。

👉 エンティティの関連を一つの関係に複数盛り込むと、話が厄介になる。

☆ 二つの属性で構成された候補キーが組である関係は必ず 5NFとなる。





なぜ正規化を行うのか？

- ≡ 構文論的 (Syntactic) な視点：
素直な関係演算を可能にするため。
(↑ NULL を取り除くことも含む。)

Not NF

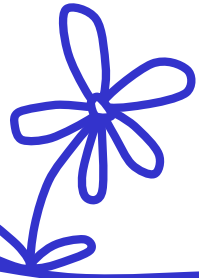
Account	Amount #1	Amount #2	Amount #3	Amount #4
A	200	300	1,000	400
B	600			
C	1,000	200		

5NF

Account	Line	Amount
A	1	200
A	2	300
A	3	1,000
A	4	400
B	1	600
C	1	1,000
C	2	200

500円以上の商品が
購入された取引明細を
求めるSQLを考えると...





なぜ正規化を行うのか？

意味論的 (Semantic) な視点：

一つの関係に対応するエンティティを一つに限定し、更新時異状を防ぐため。

1NF

Emp ID	Emp Name	Post ID	Post Name
001	Korematsu	001	President
101	Tanaka	100	Manager
102	Sato	110	Senior
103	Inoue	111	Associate
104	Suzuki	111	Associate

従業員エンティティと
役職エンティティが
関係内に混在したのが
更新時異状の原因

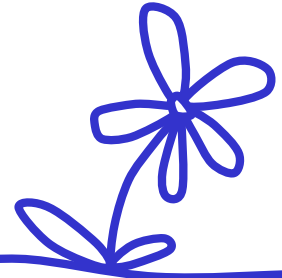
5NF

Emp ID	Emp Name	Post ID
001	Korematsu	001
101	Tanaka	100
102	Sato	110
103	Inoue	111
104	Suzuki	111

5NF

Post ID	Post Name
001	President
100	Manager
110	Senior
111	Associate





ドメイン/キー正規形とは？

定義：関係Rの全ての静的な属性値が
キーとドメインの制約に紐付く。
(↑更新時異状が起こらない。)

5NF

Wealthy Person	Wealthy Person Type	Net Worth in Dollars
Steve	IT Millionaire	124,543,621
Roderick	Car Billionaire	6,553,228,893
Katrina	IT Billionaire	8,829,462,998
Gary	Car Millionaire	495,565,211

DKNF

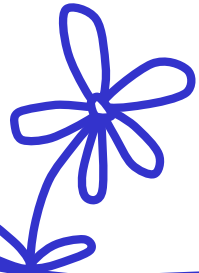
Wealthy Person	Wealthy Person Type	Net Worth in Dollars
Steve	IT	124,543,621
Roderick	Car	6,553,228,893
Katrina	IT	8,829,462,998
Gary	Car	495,565,211

DKNF

Status	Minimum	Maximum
Millionaire	1,000,000	999,999,999
Billionaire	1,000,000,000	999,999,999,999

5NFでも更新時異状が起こりうる
(Garyの財産が10億を超えると...)



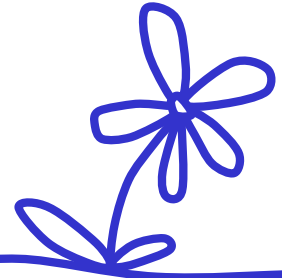


DKNFの理想と現実

- ⦿ 属性値の制約条件が全てDB側にあれば、アプリ側に制約を実装しなくてよい。
(↑コードをI/O化して除去できる。)
- ⦿ キーによるものの以外の制約を現実的に全て実装するのは困難と思われる。
- ⦿ DKNFは概念的な正規形と言える。

<取り除ける更新時異状の原因>
5NFまで：C-異常、DKNF：F-異常





第六正規形とは？

定義：候補キーに対し非キー属性が一つしか存在しない。
 (↑エンティティを最小化する。)

5NF

ID	Name	Job	Salary
1234	Anne	Lawyer	100,000
1235	Boris	Banker	
1236	Cindy		70,000
1237	Davinder		

6NF

ID	Name
1234	Anne
1235	Boris
1236	Cindy
1237	Davinder

ID	Job
1234	Lawyer
1235	Banker

ID	Salary
1234	100,000
1236	70,000

NULLが消えた

名前と職業と収入はそれぞれ独立の事象(エンティティ)である
 ↓
 独立の事象なので発生タイミングも別々
 ↓
 一つの関係にまとめるとNULLが現れる原因に

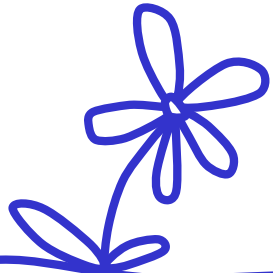




第六正規形の理想と現実

- ⌚ 時間的 (Temporal) な値の変化を想定。
- ⌚ NULLが完全に除去されるので、
二値論理で関係演算を完結できる。
- ⌚ エンティティが明確になるので、
概念スキーマを深く検討できる。
- ⌚ そのまま論理スキーマにしてしまうと、
JOINばかりで辛くなると思われる。





第六正規形と時間変化

5NF

ID	Date	Name	Job	Salary
1234	28-Nov-1979	Anne		
1234	01-Apr-2001	Anne	Lawyer	
1234	31-Mar-2002	Anne	Lawyer	80,000
1234	31-Mar-2004	Anne	Lawyer	100,000

新しい情報を追加する際、
それとは関係ない過去の情報も
含めないと意味が変わってしまう。

6NF

ID	Date	Name
1234	28-Nov-1979	Anne

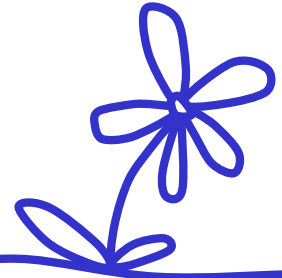
ID	Date	Job
1234	01-Apr-2001	Lawyer

ID	Date	Salary
1234	31-Mar-2002	80,000
1234	31-Mar-2004	100,000

新しい情報の追加だけで問題ない。

そもそも第六正規形は
Temporal Databaseという
論考の中で定義されました





DKNFと6NFの厄介な状況

👉 6NFがDKNFの意味で使用されている。

☆5NFの次にDKNFが定義されたのが原因。

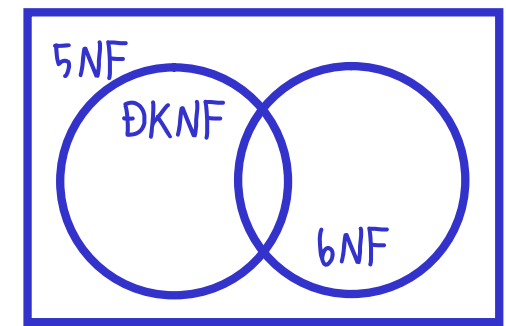
☆文献を調べるときにも非常に紛らわしい。

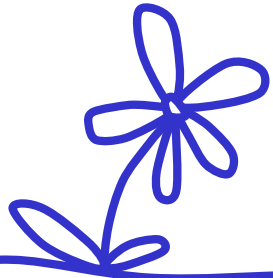
👉 より高次な正規形がどちらか分からない。

☆DKNFは静的な属性が念頭の正規形。

☆6NFは時間変化する属性が念頭の正規形。

☆どちらが高次という話ではなく、異なる概念である。

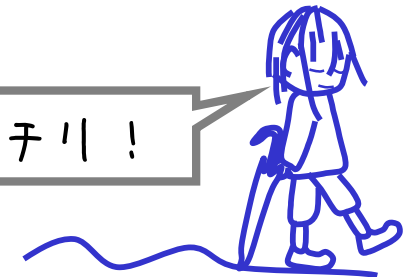


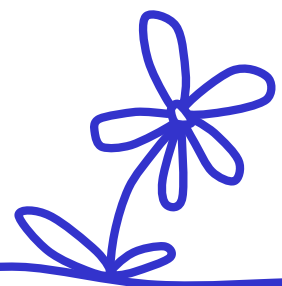


正規形を一発で見抜く表

	繰り返し属性	部分関数従属	推移的関数従属	非キーへの関数従属	対称な多値従属	自明でない結合従属	F-異常の更新時異状	値の発生の時間的ズレ
非正規形								
第一正規形	除去							
第二正規形	除去	除去						
第三正規形	除去	除去	除去					
BC正規形	除去	除去	除去	除去				
第四正規形	除去	除去	除去	除去	除去			
第五正規形	除去	除去	除去	除去	除去	除去		
DK正規形	除去	除去	除去	除去	除去	除去	除去	
第六正規形	除去	除去	除去	除去	除去	除去		除去

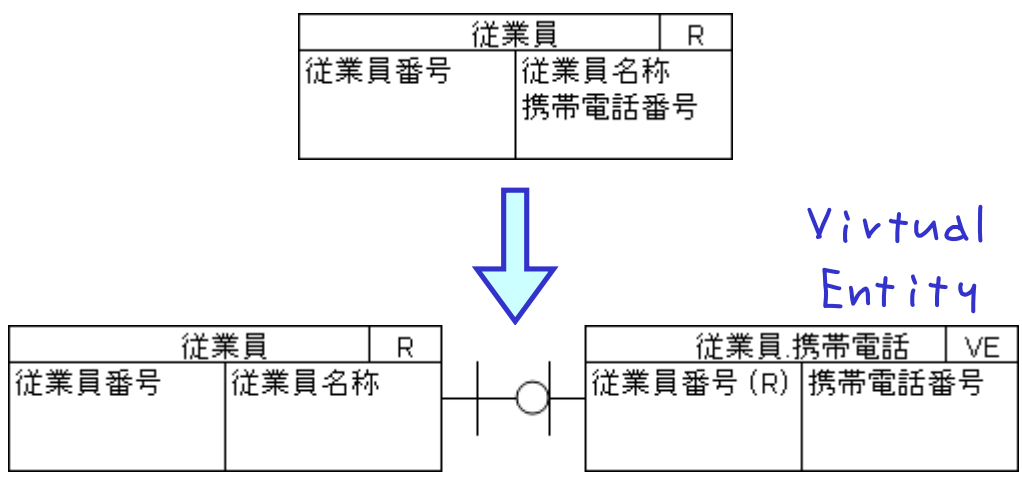
これでデスぺ対策もバッチリ！





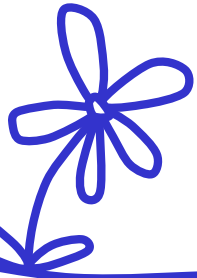
6NFとT字型ER手法

⑤ Date氏が2002年に6NFを定義する以前、
 1998年に発表された日本発のデータ
 モデリング手法『**T字型ER手法**』には、
 6NFを応用した技術が存在する。



携帯電話番号は
 NULLを含む
 ↓
 従業員と携帯電話は
 別の性質の資源
 ↓
 みなしエンティティ

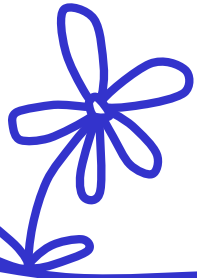




更なる深淵へ

- 👉 第七正規形を考えた人が日本にいる。
(↑ 誰か知り合いの方はいませんか?)
- 👉 DKNFをググると若干危険な雰囲気。
- 👉 もっとT字型ER手法に突撃してみる。
- 👉 SuperSQLに最適な正規化のレベルは?





グループはこちら

👉 ブログ:

流離と見聞録

👉 mixi と gree と twitter:

全部 **ビズアイユ** で検索

👉 接近戦:

結構な割合で **ClubDB2** にいる

