

# 目次

<b>第 1 章</b>	<b>データベースを使わない世界</b>	<b>3</b>
1.1	ケース 1: 販売履歴の記録をはじめる . . . . .	3
1.2	ケース 2: サイトの認知度向上につき, 得られるデータも膨大に!?	5
1.3	おわりに . . . . .	6
<b>第 2 章</b>	<b>データベースの概念</b>	<b>9</b>
2.1	データベースとは . . . . .	9
2.2	データ処理の際に求められる機能 . . . . .	10
2.3	データベースを用いるメリット . . . . .	14
2.4	データのモデリング . . . . .	15
2.5	関係データモデル (導入) . . . . .	16
2.6	クイズ . . . . .	18
<b>参考文献</b>		<b>19</b>



## 第 1 章

# データベースを使わない世界

データベースと聞いて何を思い浮かべるだろうか。大抵の人は、データベースといえば「大きなデータの集まり」といったイメージを持つのではないだろうか。このイメージはあながち間違いではない。

さて、本講義のようにわざわざ科目を立ててまで、データベースについて学ぶことはあるのだろうか？結論としては、大規模データに携わる IT エンジニアやデータ分析者を目指す人であれば、「大いにあり」である。1960 年代から今日に至るまで、データベース技術は盛んに研究開発が行われてきた。大きなデータの集まりを扱うには、対処しなければならない問題が思った以上に数多く存在するのである。

本講義では 10 数回にわたってデータベース技術について解説するが、この第 1 講ではデータベースのことはいったん横に置いておいておく。今回は（割と）大きなデータを扱うときに遭遇する問題について考えてみよう。

### 1.1 ケース 1: 販売履歴の記録をはじめ

以下は、山畑さんという架空の人物のお話である。

山畑さんは家族で小さな小売店を営んでいる。個人経営ながら山畑さんのお店は繁盛している。とはいえ、街には大手チェーン小売店が進出してきており、このまま順調に経営を続けられるか、不安が募っている。何か手を打たなければならない。

2020 年の 4 月、山畑さんは念願のショッピングサイトを立ち上げた。言うまでもない。ショッピングサイトを立ち上げたのは、オンラインの場にも顧客獲得の機会を求めるためだ。サイトは順調に立ち上がり、注文もポツポツ入ってきている。

ところで、最近「データサイエンス」なるものが世間の注目を集めているらしい。データを活かせばビジネスチャンスが広がるとのことだ。山畑さんは、Excel シートに記録を取り始めた販売履歴を分析してみようと思いついた。

山畑さんが使っている Excel シートには、「いつ、誰が、何を、いくらで購入したか」の情報が記録されている。ショッピングサイトは立ち上がったばかりであり、Excel シートには 200 行しかデータが入っていない。しかし、今後データが貯まっていけば、売り上げを増やすための課題が見えるかもしれない。いずれがっつりとデータ分析をやるためにも、山畑さんは手持ちのデータを用いて分析の練習に取り組むことにした。

#### Q1. データの確認

こちらの URL ([https://dbnote.hontolab.org/data/purchase\\_small.xlsx](https://dbnote.hontolab.org/data/purchase_small.xlsx)) から、上のケース 1 で山畑さんがデータ分析の練習に使おうとしている Excel ファイル (`purchase_small.xlsx`) をダウンロードし、中身を確認しなさい。

なお、Excel シートの各列の意味は以下の通り：

- `purchased_at`: 購買（販売）日時
- `customer`: 商品を購入した人物の氏名
- `gender`: 商品を購入した人物の性別
- `product`: 購入された商品名
- `sale`: 販売価格

#### Q2. あの人は何回買い物をしている？

「岡田 真綾」という人物が何回買い物をしていたかを数えなさい。

**Q3. 商品 X を購入しているのは誰？**

Excel のオートフィルタ機能を使って、「ビタミン補助剤」を購入している人をリストアップしなさい。

**Q4. 総売上金額**

Excel 関数の SUM を用いて、現時点での総売上金額を計算しなさい。

**Q5. 最も売れた商品は？**

Excel のピボットテーブル機能を使って、集計期間中に

- 最も購買回数が多かった商品
- 最も売上金額の合計が大きかった商品

をそれぞれ求めなさい。

## 1.2 ケース 2: サイトの認知度向上につき、得られるデータも膨大に!?

現時点では手持ちのデータは少ないものの、販売履歴データの分析に将来性を感じた山畑さん。販売履歴データを有効活用できるよう、ショッピングサイト運営により力を入れる決意を固めたのであった。

以下は、山畑さんのその後の話（架空の話）である。

ショッピングサイト立ち上げ以降、順調に利用者数も増えていった。やはりメディアに取り上げられたのが大きかったのだろう。あのタイミングでサイトの認知度が一気に高まり、サイトの利用者数や利用頻度も加速度的に増えていった。それに伴い、サイト運営に関わるスタッフも増員した。

販売履歴の管理は、当初は山畑さんが一人で担当していたが、さすがに一人では対応しきれなくなった。そこで、ある時点から数名体制で販売履歴の記録を行うことになった。これまで販売履歴の管理に使ってきた Excel シートをクラウドストレージに置き、記録担当スタッフの PC 間で同期を取る仕組みを

導入。同じ Excel ファイルの上で、スタッフ全員で販売履歴を記録できるようにしたのである。

2年後、サイト事業は軌道に乗った。十分な量の販売履歴データが蓄積されたと判断した山畑さんは、いよいよ大規模な販売履歴データの分析に取りかかることを決意した。立ち上げ当初は200~300行しかなかった Excel シートであったが、シートを開きその行数を数えてみると…なんとその数90万行以上! データの量に小躍りした山畑さんは、Excel シートの扱いに詳しいスタッフと共に、意気揚々とデータ分析に取りかかったのであった。

#### Q6. データの再確認

こちらの URL ([https://dbnote.hontolab.org/data/purchase\\_large.xlsx](https://dbnote.hontolab.org/data/purchase_large.xlsx)) から、上のケース2で山畑さんが分析しようとしている Excel ファイル `purchase_large.xlsx` をダウンロードしなさい。またダウンロードしたファイルを用いて下記課題（演習1と同じ）に取り組み、データ分析上の課題（困ったこと）を議論しなさい。以下、課題1の内容を再掲する。

- ・「岡田 真綾」という人物が何回買い物をしていたかを数えよ
- ・「ビタミン補助剤」を購入している人をリストアップせよ
- ・総売上金額を計算せよ
- ・集計期間中に「最も購買回数が多かった商品」「最も売上金額の合計が大きかった商品」を求めよ

もし、`purchase_large.xlsx` ファイルがうまく開けない場合は、こちらの URL ([https://dbnote.hontolab.org/data/purchase\\_medium.xlsx](https://dbnote.hontolab.org/data/purchase_medium.xlsx)) からダウンロードできる `purchase_medium.xlsx` を用いなさい。なお、ダウンロードできる Excel シートの構造はケース1で用いた `purchase_small.xlsx` と同じである。

### 1.3 おわりに

ケース1および2で用いた Excel ファイルは、販売履歴データの集まりであった。一般的な認識からすると、このようなデータの集まりは「データベース」とい

The screenshot shows an Excel spreadsheet with the following data (rows 2-21):

	A	B	C	D	E
			gender	product	sale
2	2020-04-01	山本 裕美子	F	お茶飲料	160
3	2020-04-01	渡辺 真綾	F	果汁飲料	120
4	2020-04-01	高橋 くみ子	F	油菓子	200
5	2020-04-01	橋本 零	M	タバコ	1780
6	2020-04-01	佐藤 和志	M	果汁飲料	120
7	2020-04-01	木村 七夏	F	機能性飲料ドリンク	240
8	2020-04-01	前田 舞	F	タバコ	1780
9	2020-04-01	鈴木 稔	M	チーズ	250
10	2020-04-01	山本 裕美子	F	お茶飲料	160
11	2020-04-01	山本 裕美子	F	お茶飲料	160
12	2020-04-01	山本 裕美子	F	お茶飲料	160
13	2020-04-02	藤井 香織	F	蒲鉾	310
14	2020-04-02	吉田 千代	F	コンニャク	150
15	2020-04-02	渡辺 直子	F	調理補助器具キッチンシール	580
16	2020-04-02	渡辺 桃子	F	弁当	100
17	2020-04-02	中村 加奈	F	弁当	100
18	2020-04-02	中村 あすか	F	プレミアムアイス	270
19	2020-04-03	加藤 淳	M	歯磨き	350
20	2020-04-03	渡辺 翼	M	畜産珍珠	100
21	2020-04-03	渡辺 太一	M	牛乳	150

Annotations on the image:

- Excelが固まる…
- 日付の書き方が統一されていない…
- 全角数字が混じっている…
- 1マスに複数の商品が…
- 途中から数字がおかしい…? (疑心暗鬼に)

図 1.1: 大きな表データを複数人で Excel で扱うときの悲劇.

うことになるだろう.

ところで, 上記演習, とりわけケース 2 に取り組んでみてイライラしなかっただろうか? 数万件, 数十万件ある表データを Excel で扱おうとすると, さまざまな不都合が生じる (図 1.1). これは, 本来 Excel は個人用の表計算アプリケーションであって, 大規模データの管理や処理を前提として設計されていないためである.

では, 大規模なデータを管理・処理するためにはどうすればよいだろうか? そのための技術こそが「データベース」である. 以降, 本書では大規模データを効率よく管理・処理するための「データベース」技術について学習する.





## 第2章

# データベースの概念

今日、データベースは様々な業務やアプリケーションで利用されている。例えば、Amazon<sup>\*1</sup>や楽天市場<sup>\*2</sup>といったオンラインショッピングサイトにおいては、購買履歴管理や在庫管理を行うためにデータベースが用いられている。Instagram<sup>\*3</sup>やTiktok<sup>\*4</sup>といった、多数のユーザがコンテンツを投稿するソーシャルメディアにおいてもデータベースは欠かせない。個人用ブログのような比較的小規模なウェブサイトでも、記事データを管理するためにデータベースが用いられている（例: WordPress<sup>\*5</sup>）。データを語る上で、データベースはなくてはならないものである。

本章では、データベースの概念を説明した後、最も代表的なデータベースであり、かつ本書の主要テーマでもある**関係データベース**の概要について述べる。

### 2.1 データベースとは

データを収集・利用する文化が根付いてきたこともあり、データベースという言葉が市民権を得つつある。しかし、「データベース」という語が指す意味は、一般人とIT屋とは大きく異なる。一般人にとってのデータベースとは、「データの集まり」くらいの意味である。一方、IT屋にとってのデータベースは、

---

<sup>\*1</sup> Amazon. <https://www.amazon.co.jp/>

<sup>\*2</sup> 楽天市場. <https://www.rakuten.co.jp/>

<sup>\*3</sup> Instagram. <https://www.instagram.com/>

<sup>\*4</sup> Tiktok. <https://www.tiktok.com/>

<sup>\*5</sup> WordPress. <https://ja.wordpress.org/>

“複数の応用目的での共有を意図して、組織的にかつ永続的に格納されたデータ群（北川博之著「データベースシステム」[1]より）”

あるいは

“データの正しさを管理する主体によって体系的に整理され、計算機によって永続的に格納されたデータの集まり（吉川正俊著「データベースの基礎」[2]より）”

を意味する。特に「データの正しさを組織的に管理する」という概念が重要である。

データベースを扱うためのシステムは、**データベース管理システム（database management system, DBMS）**と呼ばれる。本来データベースとはDBMSによって管理されるデータ群を指すが、DBMS そのもの（あるいはDBMS とそれによって管理されるデータ群）をデータベース（DB）と呼ぶこともある。

増え続ける多様かつ大量のデータと付き合っていくためには、データのうまい管理・処理方法が必要になる。場当たりのデータを作ってはExcelシートやフォルダ（ディレクトリ）に突っ込んでおくというやり方は、扱うデータが増え、データを使う人間やアプリケーションが増えるにつれて、やがて破綻する。IT屋が考えるデータベースは、こういった事態を未然に防ぐための強力なツールである。

## 2.2 データ処理の際に求められる機能

大量のデータを管理・処理するには、以下のような要件が求められる。

### 多様かつ大規模なデータの管理

ビジネスをはじめとする現場では、多種多様かつ大量のデータが時々刻々と発生している。例えば、2020年1月1日から2020年12月31日までの12か月間に、Amazon.co.jpで購買された商品の数は5億点以上にのぼるとされている[3]。

小規模な表データであれば、Excelのような表計算ソフトでも対応できる。しかし、データの規模が大きくなり、さらに表データの登録・更新を担当する人員が増えていくと…データの管理が破綻するのは想像に難くない（そもそもExcelは1つの表につき最大100万行程度しか扱えない）。Excelのようなスプレッドシートでは、購買データのように多様かつ大規模なデータを効率よく集積、管理することは

取引ID	顧客ID	品目	単価	個数	売上
A001	c001	鶏玉丼	440	1	440
A002	c001	はーいお茶	120	2	260
A003	c003	ポテト	1	1	150

↓ データが数百万行レベルに増える  
& 表の数も増えると…

取引ID	顧客ID	品目	単価	個数	売上
A001	c001	鶏玉丼	440	1	440
A002	c001	はーいお茶	120	2	260
A003	c003	ポテト	1	1	150
		⋮			
X57925	x67b2	きのこの里	200	3	600

Excelの動作が重くなる  
& そもそも数が多すぎて  
管理ルール等が破綻する  
(間違いが発生する)

図 2.1: 管理する表データの規模や数が増えると破綻する。

困難極まりない (図 2.1)。

### データの正しさの保証

管理するデータが大量かつ多様になってくると、データを正しく保つことが難しくなってくる。管理対象となるデータに誤りが混入すると、大変なことになる。例えば、図 2.2 のように、購買データの中に

- ・ 現実にはあり得ない売り上げ金額が記録されている
- ・ 入力形式が異なる日付が混在している
- ・ 取扱商品リストにないはずの商品が購買履歴に含まれている

といったことが起きると、オンラインショッピング事業においては一大事である。それゆえ、データ管理においては、格納されるデータの正しさ、データ間の矛盾のなさを保証する機能が求められる。

購買ID	購買日	品名	売り上げ(円)
1123	2019/10/05	コーヒー	350
1124	2019/10/06	ホットミルク	300円
1125	2019年10月7日	野菜ジュース	100
1126	2019/10/08	コーヒー	3百万円

存在しない商品の取引がある

ありえない金額が入力されている

日付の入力形式が異なる

図 2.2: バッドデータ (not ビッグデータ) の例.

### 高速で効率的なデータ処理

データが大量に格納できたとしても、対象となるデータを高速に処理できなければ使い物にならない。たとえ数百万件の書籍情報を格納しているシステムがあっても、ニーズを満たす書籍リストの検索結果を出力するのに数分待たされるようでは、そのようなシステムは使われない。

ユーザが増えるに従って、システムにかかる負荷も増える。例えば、Amazon.co.jp では毎分平均 900 件以上の商品取引が行われている [4]。このように大量のデータのやりとりが発生するケースにおいても、高速にデータ処理されることが求められる (図 2.3)。

### 同時実行 (並列処理)

関連して、大勢の人が同時にデータを検索、追加、更新、削除しても、システムが問題なく動作することも重要である。

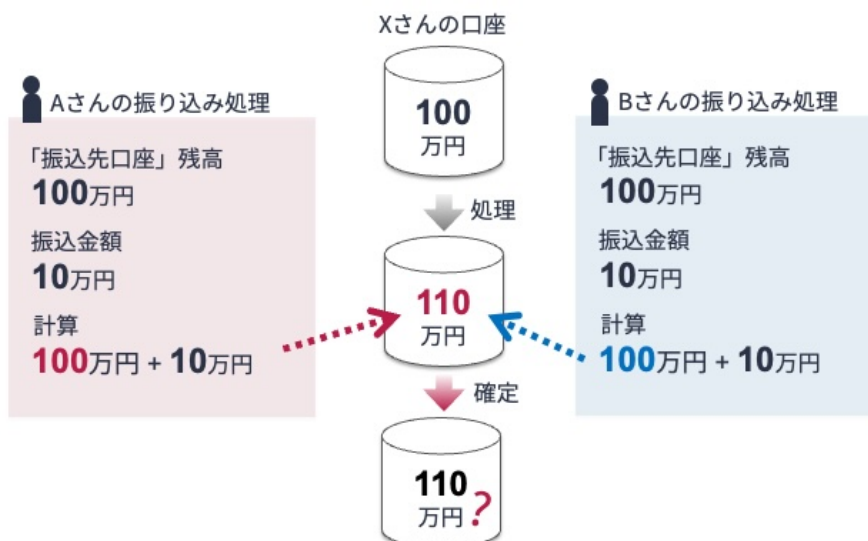
例えば、X さん、A さん、B さんがとあるネット銀行の口座を利用しているとして、X さんの口座残高は 100 万円であるとする。振り込みがあった際、振込先口座の残高の計算は

$$\text{残高} = \text{振り込み時の「振込先口座」の残高} + \text{振込額}$$

例えば100万個ある商品から該当する商品IDのものを先頭から愚直に探すと、最悪100万回チェックする必要あり

商品ID	名称	単価	登録日
AB0023	鶏玉丼	440	2010/07/15
AC0027	はーいお茶	130	2010/09/25
AD0002	ポテト	150	2011/10/02
	⋮		
QR1127	きのこの里	200	2023/04/01

図 2.3: 線形探索では大規模データに対する高速なアクセスは難しい。



振込元口座の残高の計算は

$$\text{残高} = \text{振り込み時の「振込元口座」の残高} - \text{振込額}$$

となる。これはは至極当然の処理であるが、次のような状況を考えてみよう。

ある日、AさんとBさんがXさんの口座に一秒の狂いもなく、まったく同時に10万円送金したとする。この際、前述の計算式をそのまま適用すると、図 2.4 のように誤った処理が行われてしまう。AさんとBさんが振込みを行った後、Xさん

の口座の残高は 110 万円になりました – こんなことが起きたら大変である。この例における問題点は、たとえ振込み処理が同時に発生したとしても、A さんの振り込み処理を待ってから B さんの処理を行うべきであったという点である。

このように、データ処理の内容によっては、別の処理が完了したことを保証してから次の処理を行う必要がある。さらに、処理の途中で何らかのエラーが起きた場合は、すべての処理をキャンセルして最初の状態に戻すことが求められる。同時にアクセスがあった場合でも、データを矛盾なく処理できることが重要である。

### アクセス権限のコントロール

データによっては、誰でも自由に閲覧してよいものもあれば、特定の立場のユーザしかアクセスできないようにすべきものも存在する。多種多様なデータのやりとりが発生する環境においては、ユーザの属性ごとにデータの閲覧、作成、更新、削除といったアクセス権限を制御する機能が必要となる。

## 2.3 データベースを用いるメリット

一人で扱えるほどデータの規模が小さければ、Excel などの表計算用のソフトウェアでデータ管理しても問題はない。しかし、扱うデータが多様かつ大規模になると、要求されるデータ処理の質が変わる。そのため、データの管理方法や処理方法を見直す必要がある。

本稿で学ぶデータベースを用いれば、**データの一元管理**が可能となり、データの管理や利活用において様々な恩恵を受けられる。具体的には、データベースが備える以下のような特性あるいは関連技術を用いることで、前節で述べた「データ処理に求められる要件」をクリアすることができる。

- 大規模なデータの管理：「**物理的データ格納方式**」の工夫によって対応
- データの正しさの保証：「**一貫性制約**」によって対応
- 高速で効率的なデータ処理：「**索引づけ**」「**問い合わせ最適化**」などで対応
- 同時実行，データの共同利用：「**トランザクション**」などで対応
- アクセス権のコントロール：「**ロール管理**」によって対応

## 2.4 データのモデリング

### モデリングとは？

科学やビジネスの世界では、モデルあるいはモデリングという用語がしばしば登場する。ビジネスプロセスモデル、物理モデル、統計モデルなど、世の中には様々なモデルが存在する。モデルとは、複雑な仕組みや現象、状態などを表現・分析・操作しやすくするために、本質的でない要素を取り除き、関心のある側面のみを抽出し抽象化したものである。**モデリング**とはモデル化、つまりモデルを作る行為である。

### データモデリング

データベースで何らかのデータを扱う場合、まずデータをモデリングする。世の中に存在するデータは多種多様であり、データを統一的に整理し、計算機で処理しやすい形に抽象化、すなわちモデリングする必要がある。

**データモデリング**とは、データベース化すべき情報を取捨選択し、対象とするデータとその操作に関する枠組み（**データモデル**; **data model**）を設計する行為である。対象とする事象やアプリケーションに応じて適切なデータモデルを設計することで、データモデルに従って実データを格納し、操作することが可能となる。

一般に、データモデルは以下の3つの要素を含む：

- データの構造
- データの制約条件
- データの操作

本講義の主要テーマである**関係データベース (relational database)** は、**関係データモデル (relational data model)** <sup>\*6</sup>に基づき設計されたデータベースである。関係データモデルの詳細については、次章で説明する。

データベース管理システムで扱われるデータモデルとしては、関係データモデルのほかにも以下のようなものがある：

- ネットワークモデル [5]

---

<sup>\*6</sup> 関係モデル (relational model) と呼ぶこともある。

- 階層型データモデル [6]
- オブジェクト指向モデル [7]
- キー・バリュー (key-value) モデル [8]
- グラフデータモデル [9]

## 2.5 関係データモデル (導入)

関係データモデルは、最も代表的なデータモデルである。1970 年に Edgar F. Codd 氏により提案されたもので、単純ながらその背後には強力な数学的基盤をもつ [10]。関係データモデルでは、あらゆるデータを**表 (table)** としてモデル化する。

図 2.5 は、関係データモデルを用いて構築された、授業の履修状況・成績を管理する関係データベースの例である。この関係データベースには、以下の 3 つの表<sup>\*7</sup>が存在する。

- 学生テーブル：学籍番号、氏名、入学年度、所属といった学生に関する情報を格納
- 科目テーブル：科目 ID、科目名、開講年度といった科目に関する情報を格納
- 履修テーブル：どの学生が何の科目を履修し、どのような成績であったかに関する情報を格納

この例だけ見ると「なんだ、関係データモデルとは単なる表なのか」と思われたかもしれないが、**ただの表ではない**。関係データモデルに基づいて表現された(表)データは、あらかじめ定義された「データの構造」「データの制約」「データの操作」に関する規則に従ってデータが作られ、関係データベース内に格納される。例えば、以下のような規則が考えられる：

- 学生テーブルは「学生 ID」「姓」「名」「入学年度」「所属」という見出しをもつ
- 履修テーブルには、科目名や学生の氏名を格納しない
- 履修テーブルの成績には「優」「良」「可」「不可」のいずれかしか登録でき

---

<sup>\*7</sup> テーブルと呼ぶこともある。



学生					科目		
学生ID	姓	名	入学年度	所属	科目ID	科目名	開講年度
s00001	川澄	桜	2023	A学部	c0001	線形代数	2023
s00002	山畑	滝子	2024	B学部	c0002	線形代数	2024
s00003	田辺	通	2024	C学部	c0003	統計学入門	2024
		⋮				⋮	

履修		
科目ID	学生ID	成績
c0001	s00001	不可
c0002	s00001	良
c0002	s00002	優
c0003	s00003	可
	⋮	

図 2.5: 関係データベースの例.

ない

- 履修テーブルに現れる科目 ID および学生 ID は、必ず学生テーブルと科目テーブルに存在する
- 学生テーブル、科目テーブルの各テーブルにおいて、同じ科目 ID は存在しない

このような規則に従い、授業の履修状況や成績を管理するための情報が、複数のテーブルに分割され格納される。

次章以降では、

- このような規則、すなわち関係データモデルをどう設計するのか
- 一見ただの表にしか見えない関係データモデルが、どのように数学的に定式化されているか
- 関係データモデルを用いると、なぜ大規模データの管理が効率的になるのか

などについて詳しく述べていく。

## 2.6 クイズ

### Q1. メジャーなデータベース管理システム

ウェブ検索エンジンを用いて、世の中にあるメジャーなデータベース管理システムを調べなさい。また、調べたデータベース管理システムを「関係データベースを扱うもの」と「そうでないもの」に分類せよ。

### Q2. 線形探索

100 万件ある商品リストの中に特定の商品が含まれているかを確認したい。商品リストの先頭から末尾まで順に商品名を確認していくと、平均で何回（何件）の確認で商品の有無を確認できるか？

### Q3. データベース管理システムの利用例

普段利用しているサービスのうち、データベース管理システムを用いていると思われるものを3つピックアップしなさい。

### Q4. CSV/TSV ファイル

CSV ファイルおよび TSV ファイルとは何かを調べなさい。また、（データベース管理システムでデータを管理する場合と比較して）CSV/TSV ファイルに存在する欠点を挙げなさい。

### Q5. チューリング賞

関係データモデルを提唱した Edgar F. Codd 氏は、計算機科学分野のノーベル賞といわれるチューリング賞の受賞者である。Codd 氏以外で、データベースに関する功績でチューリング賞を受賞した人をピックアップしなさい。

## 参考文献

- [1] 北川博之：データベースシステム, オーム社, 改訂 2 版 (2020).
- [2] 吉川正俊：データベースの基礎, オーム社 (2019).
- [3] Amazon Inc., : What is ecommerce? Definition, advantages and disadvantages of ecommerce websites and ecommerce business: <https://sell.amazon.co.jp/learn/what-is-ecommerce>.
- [4] Amazon Inc., (2021), <https://press.aboutamazon.com/jp/news/中小企業支援/2021/10/amazon-中小規模の販売事業者様の販売状況を発表>.
- [5] Bachman, C. W.: Data structure diagrams, *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, Vol. 1, No. 2, pp. 4–10 (1969).
- [6] Tsichritzis, D. C. and Lochovsky, F. H.: Hierarchical Data-Base Management: A Survey, *ACM Computing Surveys*, Vol. 8, No. 1, pp. 105–123 (1976).
- [7] Atkinson, M., Dewitt, D., Maier, D., Bancilhon, F., Dittrich, K. and Zdonik, S.: *The object-oriented database system manifesto*, pp. 946–954, Morgan Kaufmann Publishers Inc. (1994).
- [8] DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., Sivasubramanian, S., Voshall, P. and Vogels, W.: Dynamo: amazon’s highly available key-value store, in *Proceedings of Twenty-First ACM SIGOPS Symposium on Operating Systems Principles (SOSP 2007)*, pp. 205–220, Association for Computing Machinery (2007).
- [9] neo4j, : What is a graph database: <https://neo4j.com/docs/getting-started/graph-database/>.
- [10] Codd, E. F.: A relational model of data for large shared data banks,

---

*Communications of the ACM*, Vol. 13, No. 6, pp. 377–387 (1970).