

INIAD CS Essentials

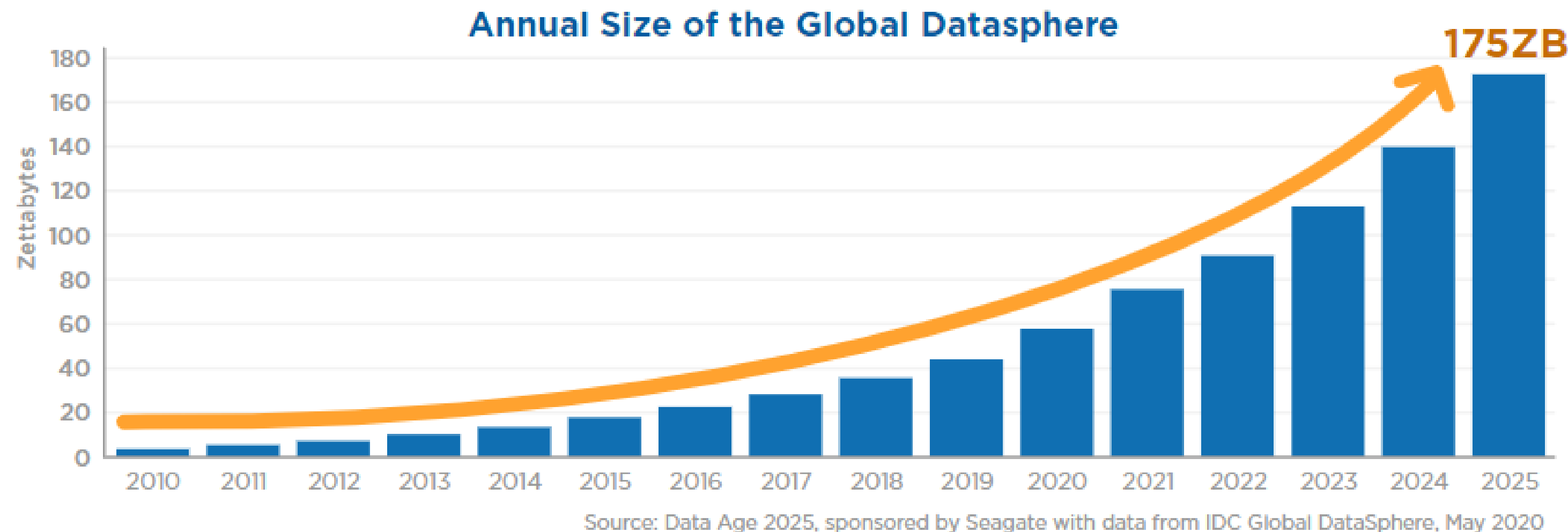
12-2: データを管理しよう

1. データとデータベース

たくさんのデータを管理するための仕組みにデータベースがあります。
データベースの利点とデータ管理に必要な要件についてみていきましょう。

情報化社会とデータ

- 現代の情報化社会では、膨大な量の「データ」が日々生み出されています
 - 2025年までに人類が生み出すデータ総量は、175ゼタバイト(1ゼタバイト = 10^{21} バイト)になると予測されています (*1)



*1: IDC "The Digitation of the World", Nov 2018

情報化社会とデータ

- たとえば、学部の学籍データ、科学実験データ、お店の毎日の売上記録、人の毎日の健康データ(身長・体重、血圧など)、携帯の位置データ、電子マネー利用記録、銀行口座情報、...
- 大量のデータを保存、管理し、必要なデータを検索したり集計したりするためには、どのような仕組みが必要でしょうか？

データをまとめて保存・管理する

- 多くの場合、形式が決まった似たようなデータがたくさん並び、またひとつひとつのデータは、いくつかの項目に分かれています。
- たとえば学籍データを保存・管理する場合は、一人一人の学生の名前や住所などの情報を集め、通し番号(学籍番号)をつけて整理します。

学生1

学籍番号:10001
姓:情報
名:太郎
かな姓:じょうほう
かな名:たろう
コース:ビジネス
住所:115-0053...

.....

学生2

学籍番号:10002
姓:連携
名:花子
かな姓:れんけい
かな名:はなこ
コース:デザイン
住所:113-0032...

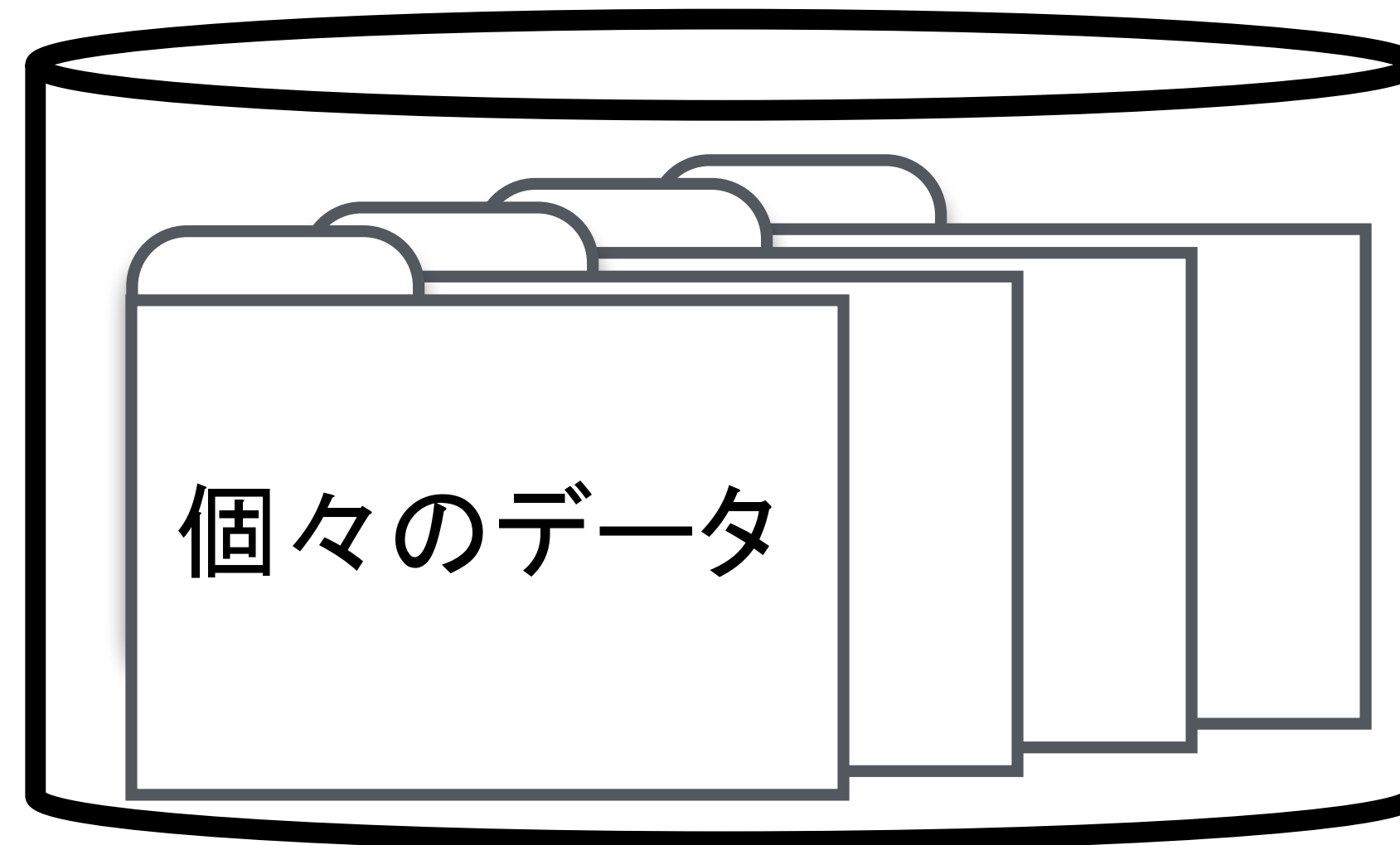
.....

...

データとデータベース

- ここから、東京都在住の学生を抽出したり、特定コースの学生全体で点数を集計したり、新入生のデータを追加したり、...とさまざまな操作を行うことになります。このように大量のデータを整理して保存・管理する仕組みを「データベース」と呼びます。
- これから数回にわたって、データベースを取り扱うための基本概念、用語、データベース専用言語SQLや、pythonを使ってデータベースを操作する方法などを学びます。

データベース



データの管理に必要なこと

- データの管理には、以下のような機能が必要になります
 - 大量のデータを保存する
 - 新しいデータを追加する
 - 既にあるデータを変更・削除する
 - 条件を満たすデータを検索・抽出する
 - ...

データの保存・管理の方法によって、各機能の容易さや作業量などが変わってきます

さらにデータベースに求められる機能

- データに対して行った操作の記録保存(ロギング, logging)
 - 障害時の復旧などで必要になります
- 一連の操作をまとめて実行 / まとめて取り消し
 - 例:「銀行口座間振込」の手順は、1. 送り手口座の金額を減らす, 2. 受け手口座の金額を増やす、の両方が成功した場合のみ、実際にデータに変更を加える(コミット, commit)。1, 2の途中でエラーが発生した場合は、これらの手順全体を取り消して元の状態に戻す(ロールバック, rollback)。
 - このようなひとまとまりの操作をトランザクション, transaction)と呼ぶ

さらにデータベースに求められる機能

- 複数ユーザによる同時編集への対応
 - あるデータの現在の値が100だったとする
 - Aさんはこのデータを「現在の値から10減らしたい」と考え、このデータを90にする編集操作を行った。
 - その操作の直前にBさんが同じデータに対して「値を80に設定」する編集操作を行った。
 - Bさんの操作はAさんの操作によって上書きされてしまい、このデータは90になった。
 - このような事態を避けるために、たとえば、「Aさんの作業中は他のユーザ(Bさん)によるデータ変更を禁止する」などの工夫が求められる(これを「**ロック**、**lock**」と呼びます)

● ...

テキストファイルや表計算ソフトでは...

- テキストファイルでは、データが明示的に項目別に分かれていないので、たとえば住所の特定文字列だけを確実に変更するのは困難です
- 表計算ソフトでは、データは項目別に分かれているので、特定項目の修正などは可能ですが、トランザクションなどは提供されていません

テキストファイル

学籍番号:10001
 姓:情報
 名:太郎
 かな姓:じょうほう
 かな名:たろう
 コース:ビジネス
 住所:115-0053...

学籍番号:10002
 姓:連携
 名:花子
 かな姓:れんけい
 かな名:はなこ
 コース:デザイン
 住所:113-0032 ...

...

表計算ソフト

学籍番号	姓	名	...
10001	情報	太郎	...
10002	連携	花子	...
...			

データベース管理システム(DBMS)を使って解決

- このように、ファイルや表計算ソフトでは、先に挙げたような、データベース管理に必要な機能をすべて実現するのは困難です。
- このため、データベース構築・管理専用のソフトウェアが数多く開発され、広く用いられています。
- これらのソフトウェアを総称してデータベース管理システム(Database Management System, **DBMS**)と呼びます。

