

Ordeal 0 Part2

🕒 Created	@October 10, 2022 8:34 PM
🏷️ Tags	

概要

Part 1ではデータが主に人の行動や事象から発生すると学んだが、実際にデータはどこにあり、どのように保存されているのかを学習する。学習後、データを取得するうえでの注意点を把握し、分析前の準備に活かしていくことができるようになる。

参照

・[出典：KT Channel] 「データはどこにある？」

<https://www.youtube.com/watch?v=q11RLh3Z4yY>

#DataSaberBridge

#Tableau

#データ

#データベース

データはどこにある？

データはデータベースにある。

ところで私たちが見るデータはデータベース上のデータか？

→ いいえ。

私たちが基本的に見るデータ（報告書など）： 取捨選択、加工、集計済みの状態。

→レポートなどの報告書ではデータの傾向を確認できる一方で、

- データを深掘りする
- 柔軟な見方でデータを見る

ことができない。

→データを深掘し柔軟な見方でデータを見るには、データベース上のデータにアクセス必要がある。

データベースにあるデータの特徴

①コンピュータが記録しやすい型、名前、値で記録されており、人が目視で見るのが難しい

（例：商品ID）

②正規化されている

[正規化：データベース上で管理しやすいように、情報（項目）の重複がない状態で保存されること。]

③件数が多くかつ処理が重い

④データの入力に干渉すると入力処理に影響がある。

※生データ：分析に適さない、構造化されていないデータ。

もし私たちが「商品の売れ行きを詳細に分析できるようなデータ」が欲しいのであれば、

購買履歴といった明細データとデータ提供相手に伝えたほうがベスト。

データを保存する情報システム（データベース）の変化

時代を経るごとに変化するニーズに応じて、

データを保存する情報システム（データベース）も変化。

■データウェアハウス（DWH, Data Warehouse。Warehouse：倉庫）

■データマート（DM, Data Mart。Mart：小売）

- 売上データやマーケティングデータなどの生データを保存
- 生データの整理、項目の追加・整理

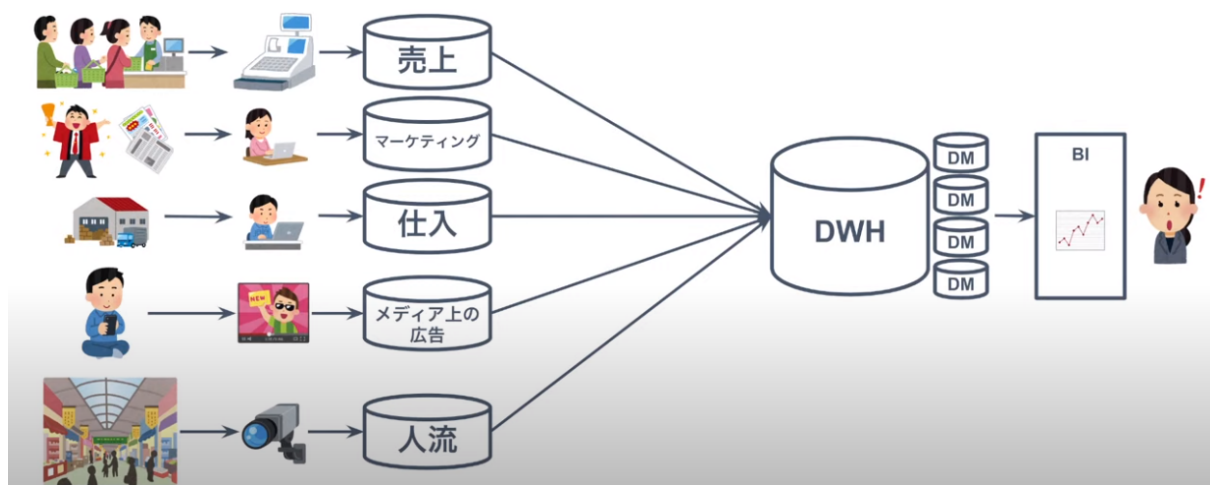
を行う**データウェアハウス**に保存。

BI（**B**usiness **I**ntelligence。意思決定のためのデータ分析）を担当する人や部署ごとに

アクセスできるデータを揃えた**データマート**で用意。

- メリット：大規模データを保存したデータウェアハウスから必要なデータをデータマートに持ち込むことができ、比較的早く分析がしやすくなる。
- デメリット：BI側のニーズに合わせたデータマートの設計負担増。

そこで登場する情報系システム



■データレイク

売上やマーケティングデータなどのデータを生データの状態で**データレイク**に保存。

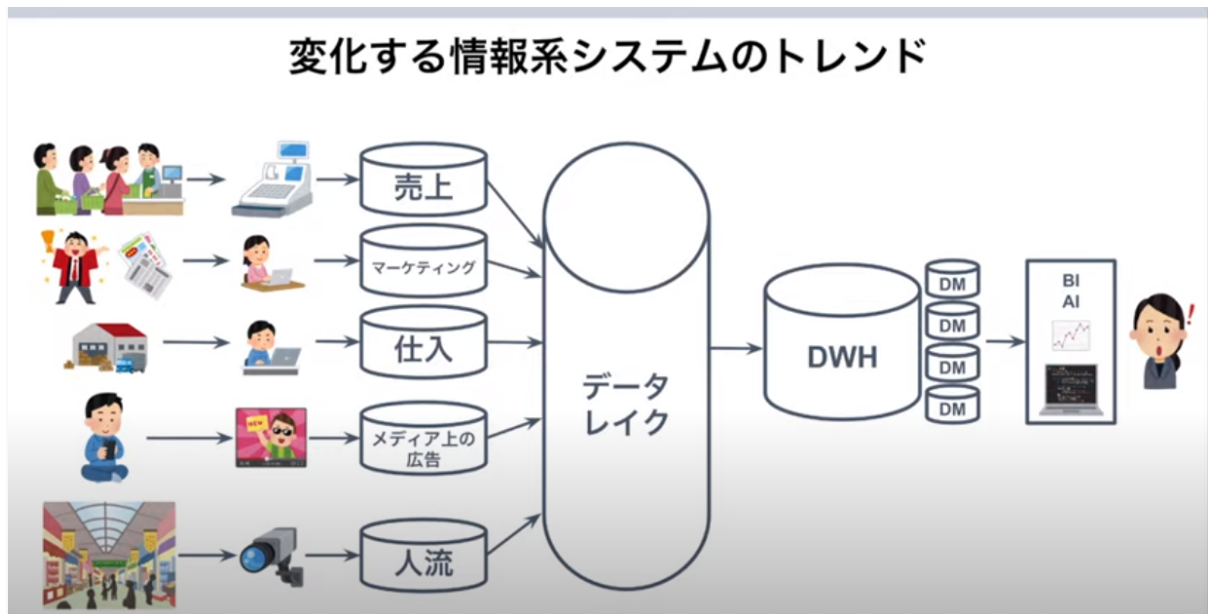
データウェアハウスに整理したデータを持ち込み、

データマートで必要なデータをBI側に提供する。

- メリット：データウェアハウスに直接生データを入れ込んだ後加工される状態が解消され、データの不足がなくなった。

●デメリット：データレイクからデータウェアハウスにデータを持ち込むまでに時間がかかる。

→結局は素早く正規化されたデータにアクセスできればよい。



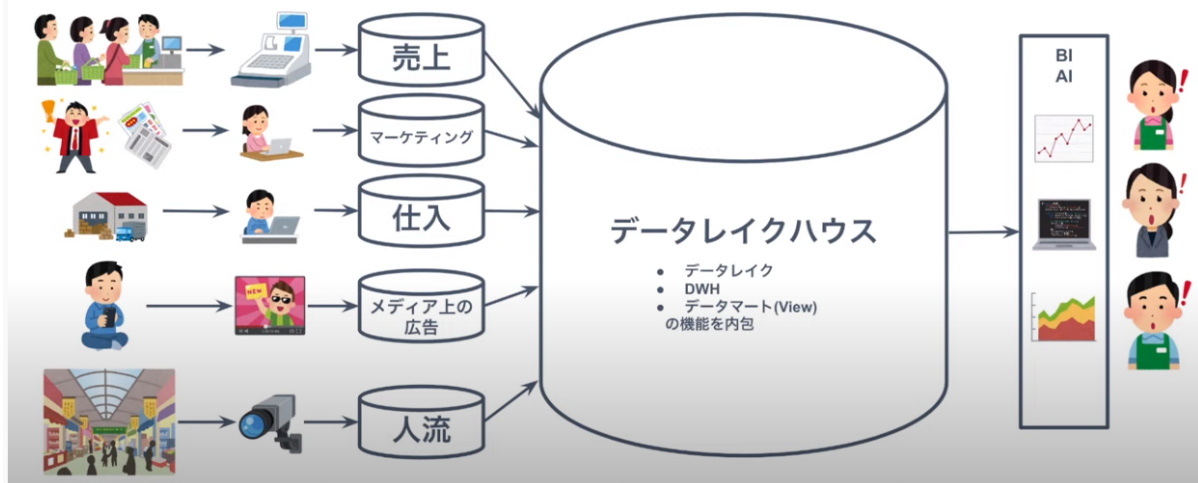
■データレイクハウス（データレイク＋データハウス＋データマート）

分断されていた各種データベースを統合し、

- 生データを保存
- 素早く正規化されたデータにアクセス

できるデータレイクハウスの展開。

さらに進化する情報系システムのトレンド

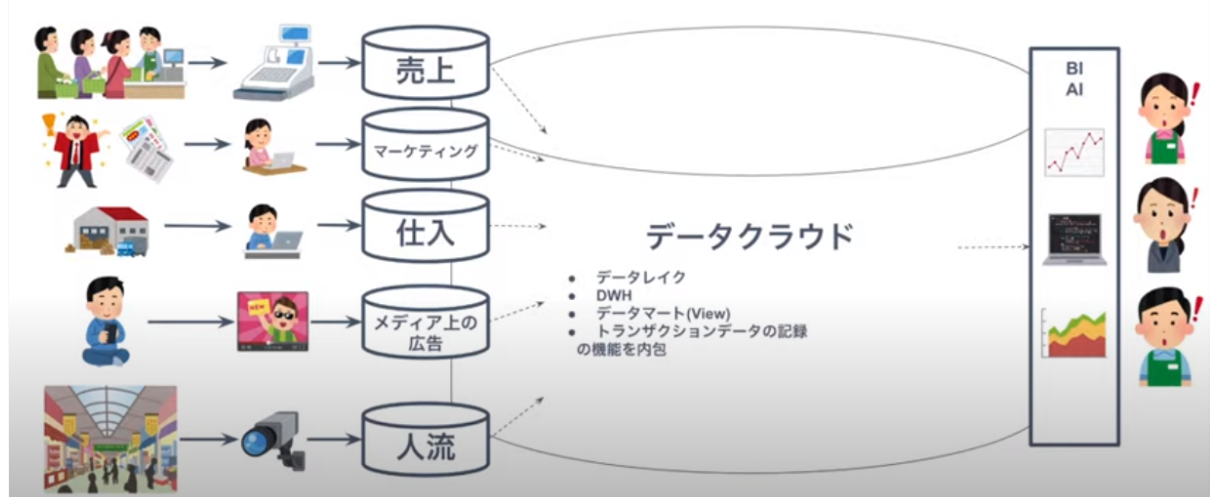


■データクラウド

データレイクハウスにトランザクションデータ（リアルタイムに記録されるもの）を保存し、

アクセスできるような機能を追加したデータクラウドの登場。

さらに??? リアルタイムを叶える情報系システム



役割ごとに決まっているデータへのアクセス

組織ごとに異なるが、大まかには以下の役割ごとに各データへアクセスするよう設定される。

※役割名（アクセスできる範囲例）と記載。

- 経営層（各部門の売り上げ状況など）
- データエンジニア（データの管理者として必要なデータすべて）
- データサイエンティスト（データ加工のために必要なデータレイク上の生データなど）
- データアナリスト・営業事務・マーケティング（分析のためにBI上でアクセスできるデータ）
- 現場メンバー（在庫管理のために必要な分析結果を記載したデータ）

どこにアクセスできるかはみんなの役割と組織による



経営層



データ
エンジニア



データ
サイエンティスト



データアナリスト
営業事務
マーケティング



現場のメンバー

※組織によって誰にどこまでアクセスさせるかは異なる

データを要求・分析するうえで心掛けておくこと

例えば、A社に所属するTableauユーザの社員が「A社の商品の日別販売状況」を知りたいとする。

①知りたいことは何か

→ 「商品の売れ行きを日ごとに知りたい」

②そのデータはどうやって記録されているか？

→「例：A社のデータベース上の明細テーブルに購買履歴が記録されている」など

③誰の持ち物か

→「例：A社内にある」

④（自分が）アクセス可能かどうか、どうやってアクセスするか

→「例：データベースから購買履歴ファイルをダウンロードしてTableauでアクセスする」など

⑤どの範囲まで見れたら、事足りるか

→「例：商品名と商品が売れた日付、売れた場所。特に日付は前四半期分のみ」など

データの在処に思いを馳せられるかどうか

- 知りたいことは何か
- そのデータはどうやって記録されているか
- 誰の持ち物か
- アクセス可能かどうか、どうやってアクセスするか
- どの範囲まで見れたら、事足りるか

