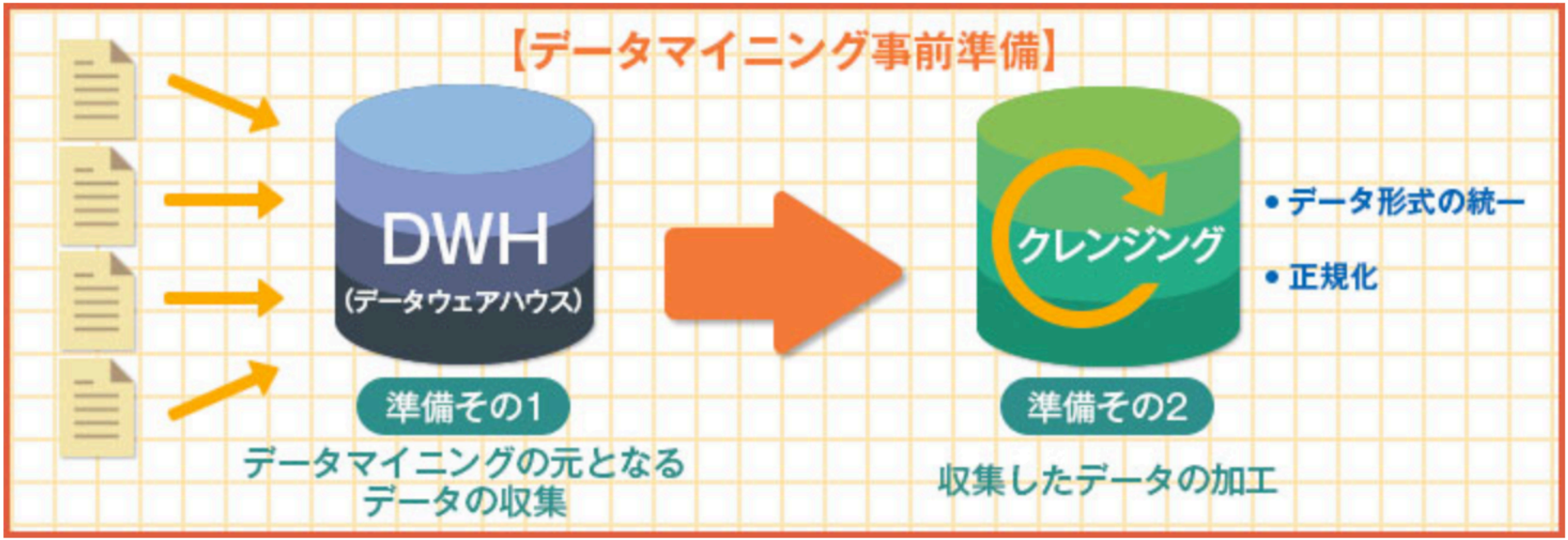
- データマイニングとは？基本から分析手法までを解説！｜ITトレンド
  - **データマイニングとは？**
    - **大量のデータを統計学や人工知能などの分析手法を駆使して、「知識」を見出すための技術**
    - DIKW
      - ▪データ（Data）：整理されていない数値
      - ▪情報（Information）：「データ」を整理・カテゴライズしたもの
      - ▪知識（Knowledge）：「情報」から得られる傾向・知見
      - ▪知恵（Wisdom）：「知識」を利用して人が判断する力
  - **データウェアハウス（DWH）とは**
    - データウェアハウスはデータベースとは異なり、登録されたデータの削除や更新を原則として行いません。データウェアハウスは、データを蓄積することが目的だからです。
  - **データマイニングとデータウェアハウス（DWH）の関係**
    - データマイニングを行う上では、データウェアハウスの存在は欠かせない
    - データマイニングを利用する上でデータウェアハウスが必須となるわけではない
  - **準備その１： データマイニングの元となるデータの収集**
    -

  - **準備その２： 収集したデータの加工**
    - =クレンジング
    - データ形式」は、原則としていずれかに統一する必要があります
    - 正確な結果を出すには、「正規化」などを施す必要もあります
  - **データマイニング２つの具体的手法**
    -

    - **事前に仮説を用意しない「機械学習」**
      - データの中からコンピュータが自分で学習しながら相関関係などを導き、人ではできない新しい分類（顧客セグメント）などを発見
      - 得意分野
        - 複雑な条件が絡む課題の分析や最適化
        - 「個客ごとの最適なマーケティングシナリオの選定」
        - 「ある事象の原因の特定」
    - **事前に仮説をたてる「統計分析」**
      - 代表的な統計手法
        - 回帰分析
        - 主成分分析
        - 因子分析
  - データマイニングに必要な３要素は？
    - 大量の元データ
    - データマイニング用のシステム
    - スキルを持った人
  - **データマイニングの分析手法**
    - クラスタリング
      - グルーピング
      - 活用例：顧客のセグメント化
    - ロジスティック回帰分析
      - YES/NO課題に有効
      - 活用例：顧客はある製品を買ってくれるか否か
    - マーケット・バスケット分析
      - =アソシエーション分析
      - 相関性を見出す
      - 例：ビールとおむつ
  - データマイニングの可能性
    - 「顧客の購買情報にもとづいた営業戦略」などをさらに高度に実行
    - Web検索技術の発達をうまく活用すれば「Web上の検索ワードに対してデータマイニングを行い、人々が何に関心を持っており、どのような視点で関心の対象を理解しているのかを知ること」も可能
    - これまであまり重要視されていなかったデータにも光を当て「ショッピングモール内の温度、湿度、照明光度と売上との関係」といった、新たな視点からの相関関係も見出されるかもしれません


【データマイニング事前準備】
DWH
(データウェアハウス)
クレンジング
• データ形式の統一
• 正規化
準備その1
データマイニングの元となる
データの収集
準備その2
収集したデータの加工

【データマイニング具体的手法】
…
時間
店舗
天気
製品
納期
売上
顧客
…
在庫
Beer
ビール
と
オムツ
新しい
発見!!
仮説A
仮説B
どっちが
正解?
仮説A
仮説B
手法その1
事前に仮説を用意しない
「機械学習」
手法その2
事前に仮説をたてる
「統計分析」