

第5回 データ活用実践応用編 【機械学習モデルの構築とその利用】

担当: 西川

これまでやってきたこと

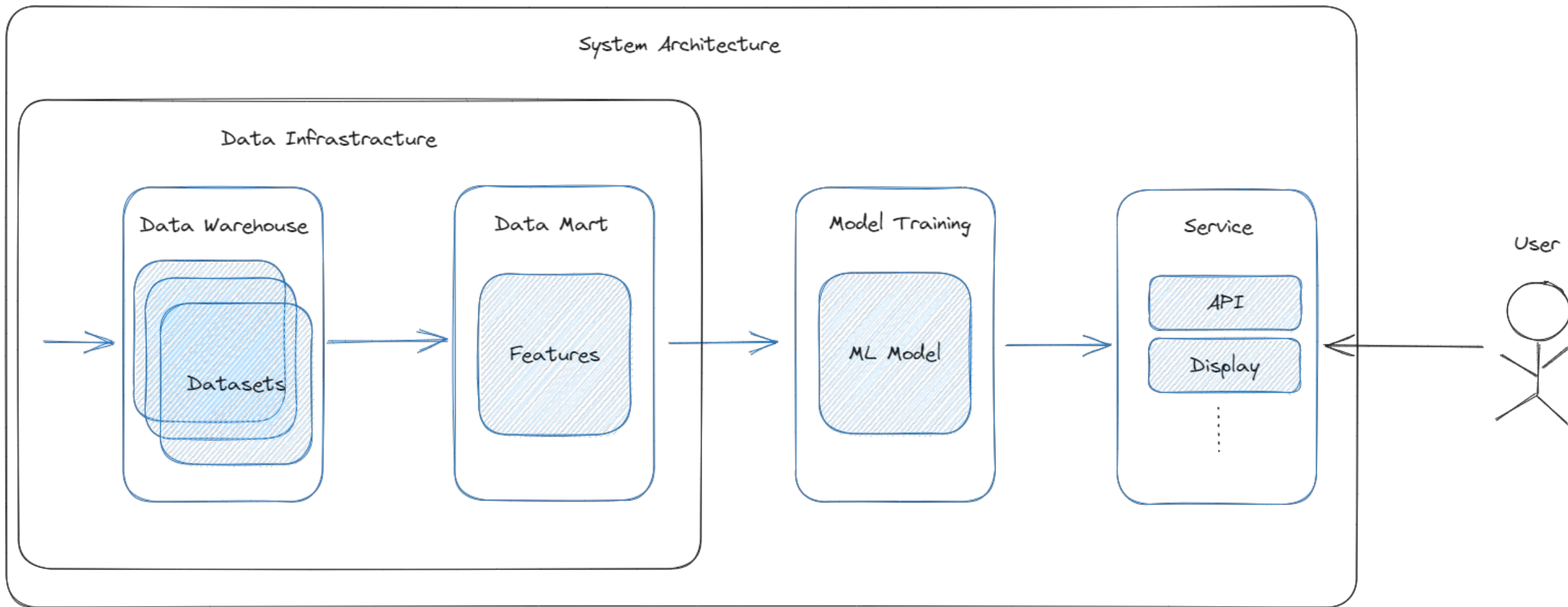
- データ基盤についての座学
- DWH の構築
 - 半構造化データ
 - ストリームデータ
- DWH からデータマートの構築
- データの活用
 - 可視化

今日やること

- データの活用
 - 推論のためのモデル構築
 - 構築したモデルの運用・MLOps(座学)

データ基盤と ML システム (最小構成)

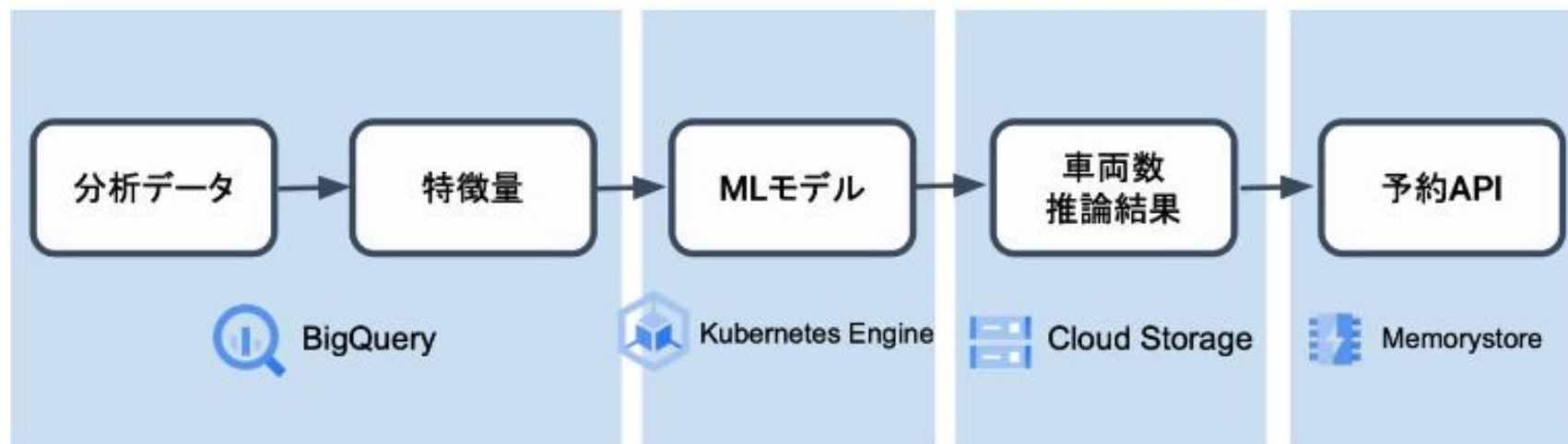
- ざっくり示すとこんな感じ
- これだけでいいわけがない



- ML モデルは適切に構築されても劣化する
 - 構築されてすぐがもっとも良い
 - 時間経過とともに劣化する
 - 劣化の原因は大抵の場合データセット
 - 学習時と運用時でデータセットに変化が生じる
 - 変化内容については省略
 - cf. Dataset Shift
- 劣化を防ぐための対応が必要
 - モニタリングする
 - モデルのメトリクス
 - データセット
 - モデルを継続的に再学習させる
- 他にもMLシステムの運用のために必要なことはたくさんある
 - モデル構築のための実験管理
 - MLに関係なく、システムとしての運用
 - コスト管理
 - etc.

タクシーアプリ「GO」における「AI予約」

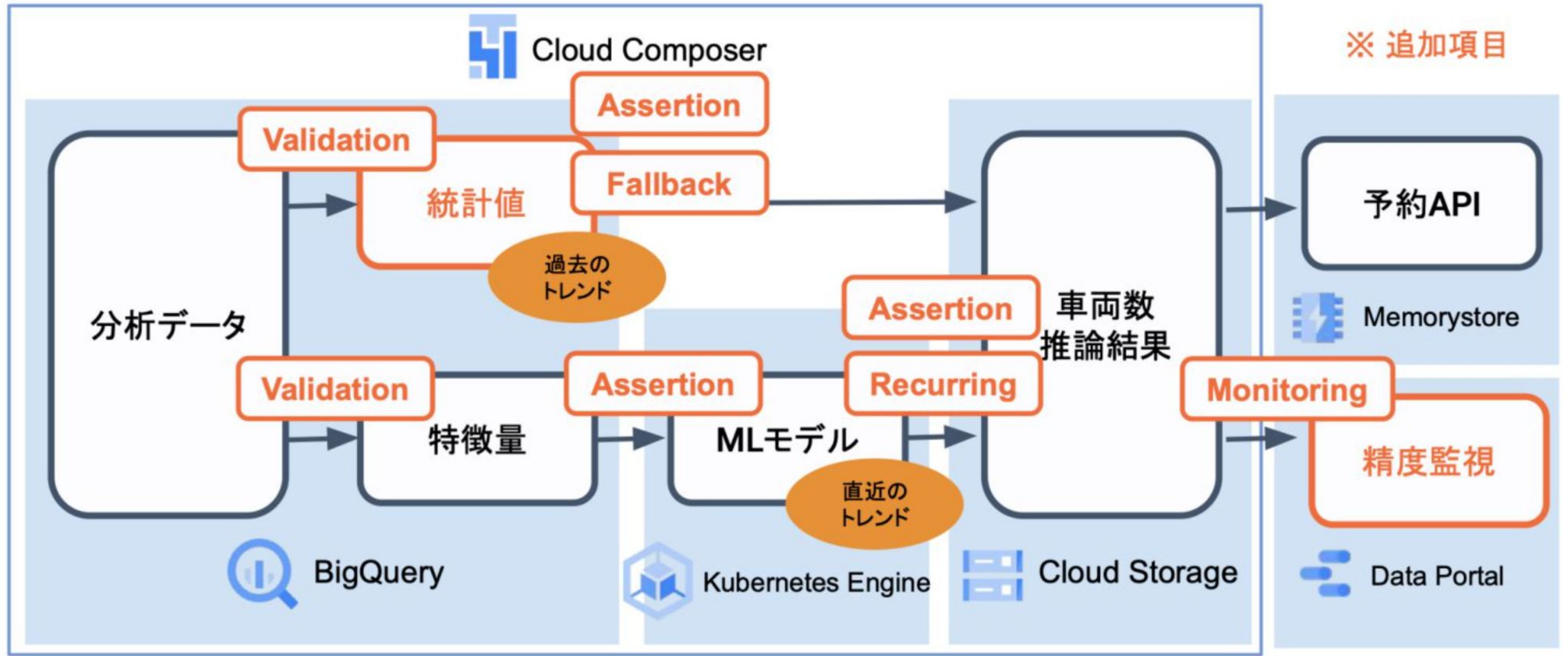
- 従来の人力での予約では事前に車両を確保する
 - 乗務員の拘束時間が長い
 - 予約枠数に限りがある
- AI 予約では AI を用いて配車可能性を予測する
 - 当日に柔軟に車両を確保する
 - 従来の10倍以上の予約依頼に対応



【出典】Mobility Technologies Engineering Blog <https://mob-mo-t.com/blog/curation-ai-reservation/>

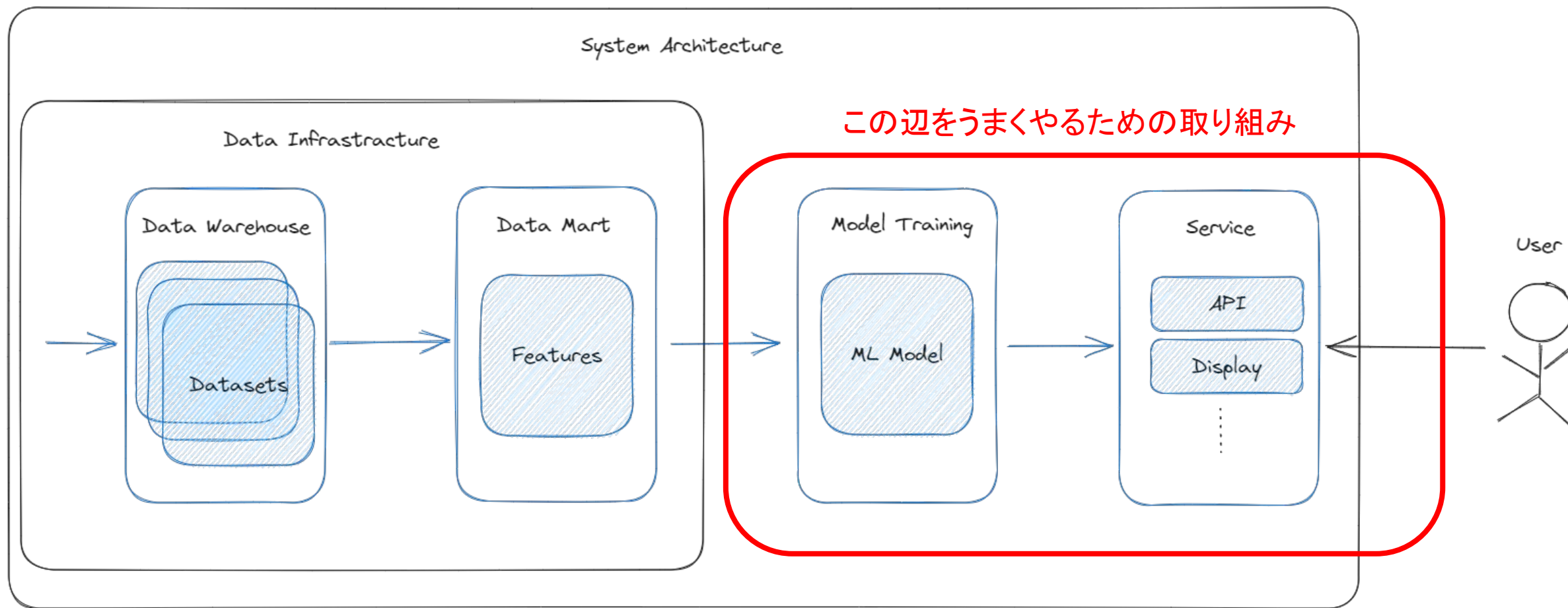


【出典】<https://go.mo-t.com/service>



【出典】Mobility Technologies Engineering Blog <https://lab.mo-t.com/blog/curation-ai-reservation>

- ML システムの開発に DevOps の考え方を導入したもの
- 定義はまだ曖昧
 - 気になる人は調べてください



1. CIFAR-10 の生データをダウンロードしてデータレイクに配置する.
2. データレイクのデータを加工して ML モデルの構築に使えるようにする.
 - パスとラベルの DataFrame を作成してテーブルとして保存する.
 - 画像を一枚ずつサーバのストレージに保存する.
3. ML モデルを構築する
 - 若干 EDA 的なことする
 - ML Flow による実験管理

