



# データドリブンへようこそ。 BigQuery ML で始める購買予測

小山 航

日本情報通信株式会社 バリューインテグレーション本部  
クラウドテクノロジー部、テクニカルセールス

# スピーカー自己紹介



小山 航

日本情報通信 (NI+C)  
バリューインテグレーション本部  
クラウドテクノロジー部  
テクニカルセールス

NI+C 入社後、テキストマイニング製品の導入、チャットボット開発に携わり、AI ソリューションのハッカソン出場を機に先進技術領域の案件を担当。

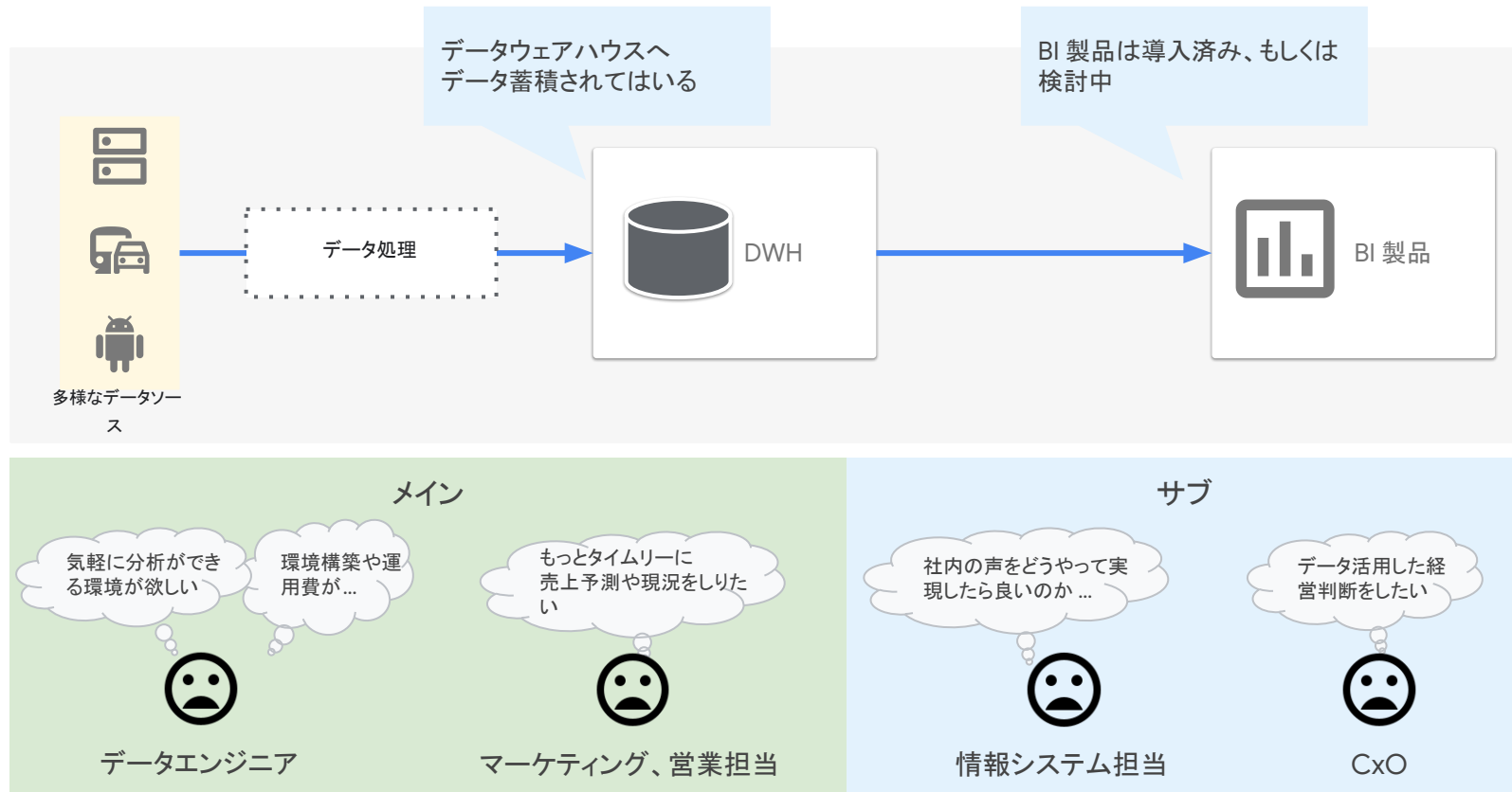
以降、Google Cloud を中心とし自然言語処理、画像認識、予知保全、IoT、Bigdata 基盤、ETL、データ可視化の分野でテクニカルセールスとして業界を問わず幅広く活動している。

製品導入、システム開発に伴うデータ利用観点の整理や評価軸を整理することが得意。

# アジェンダ

- 本セッションの対象となる方
- データドリブンへようこそ
- 購買予測をするには
- Wrapup

# 本セッションの対象となる方

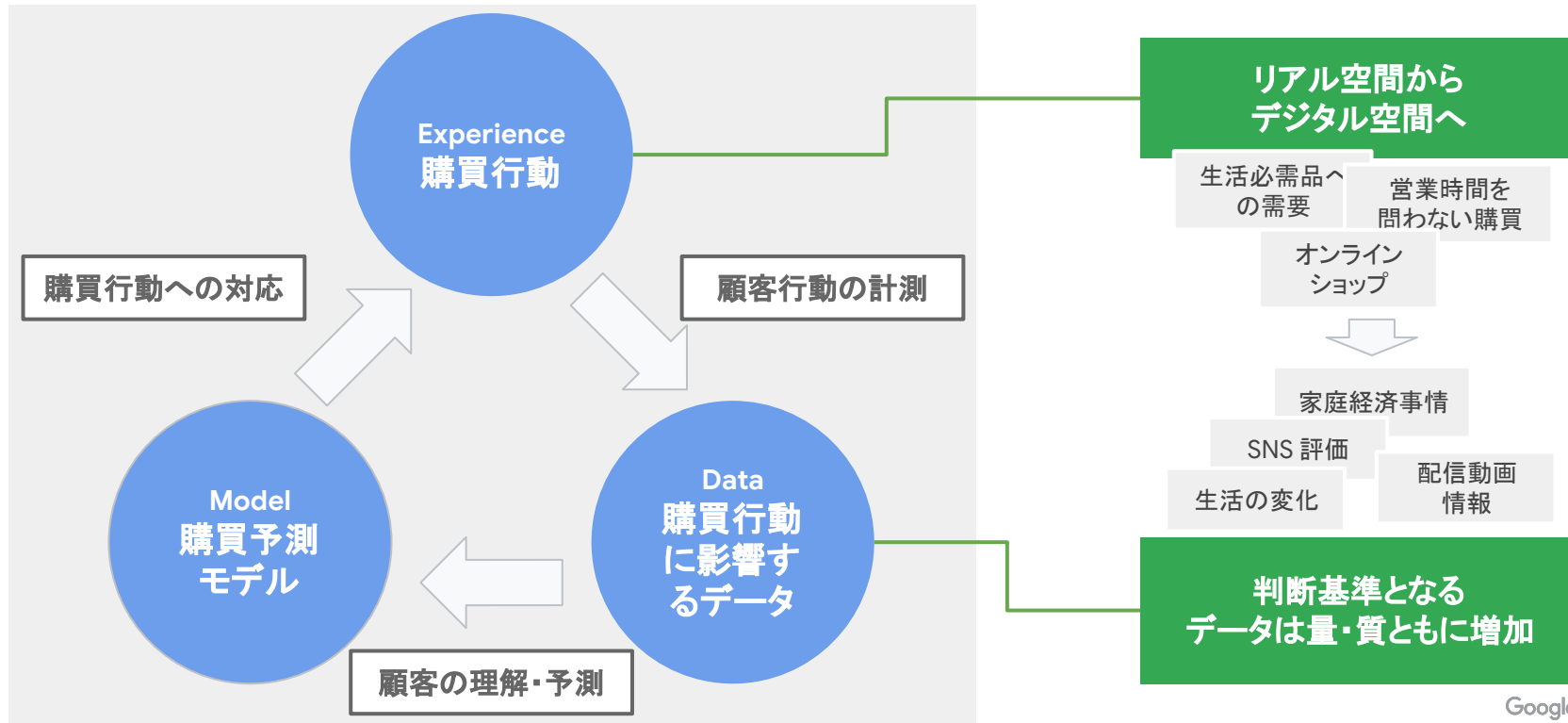




# データドリブンへようこそ

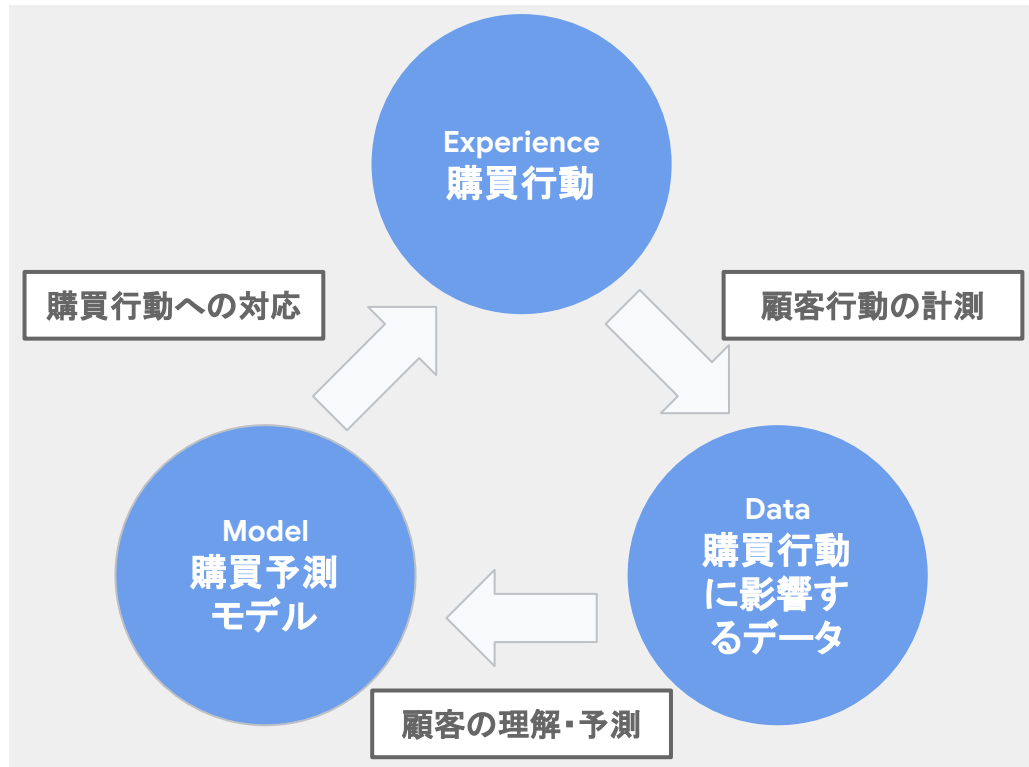
# 昨今の状況とデータドリブンな判断が重要な訳

## データドリブンをするためのサイクル



# サイクルを回すためのポイント

## データドリブンをするためのサイクル



- 膨大なデータを蓄積できる基盤
- 市場価値がある内に手が打てる素早い分析ができること

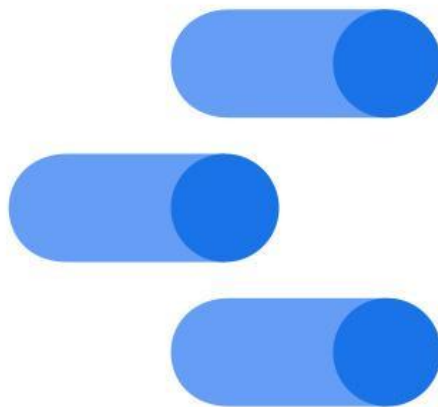


DWH 内で一気に分析までできるような仕組みがあればいいのでは...？

あとシームレスに可視化まで出来たら...

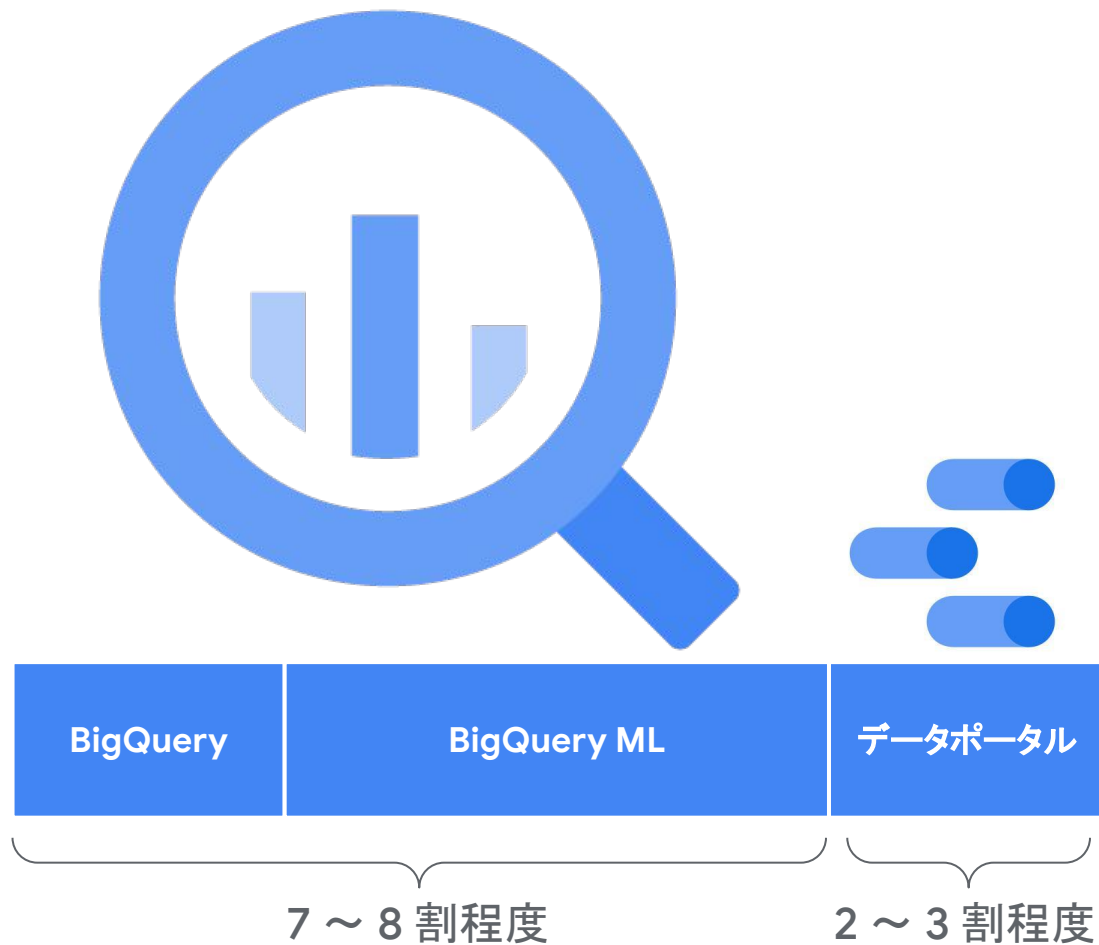


BigQuery  
(BigQuery ML)



データポータル







購買予測をするには

# 事例紹介

## お客様

業界 : 小売業

既存システム : Google Cloud ユーザー様で社内の売上などのデータを BigQuery に蓄積

保持データ : 各店舗の売上データ

## 課題・ご要望

### 課題

ニーズが掴めず  
機会損失

過去の経験や  
特定個人の判断により  
意思決定されていた



### 目的

在庫/品出し  
適正化

過去のデータから予測し  
傾向などの分析すること



### 評価設定

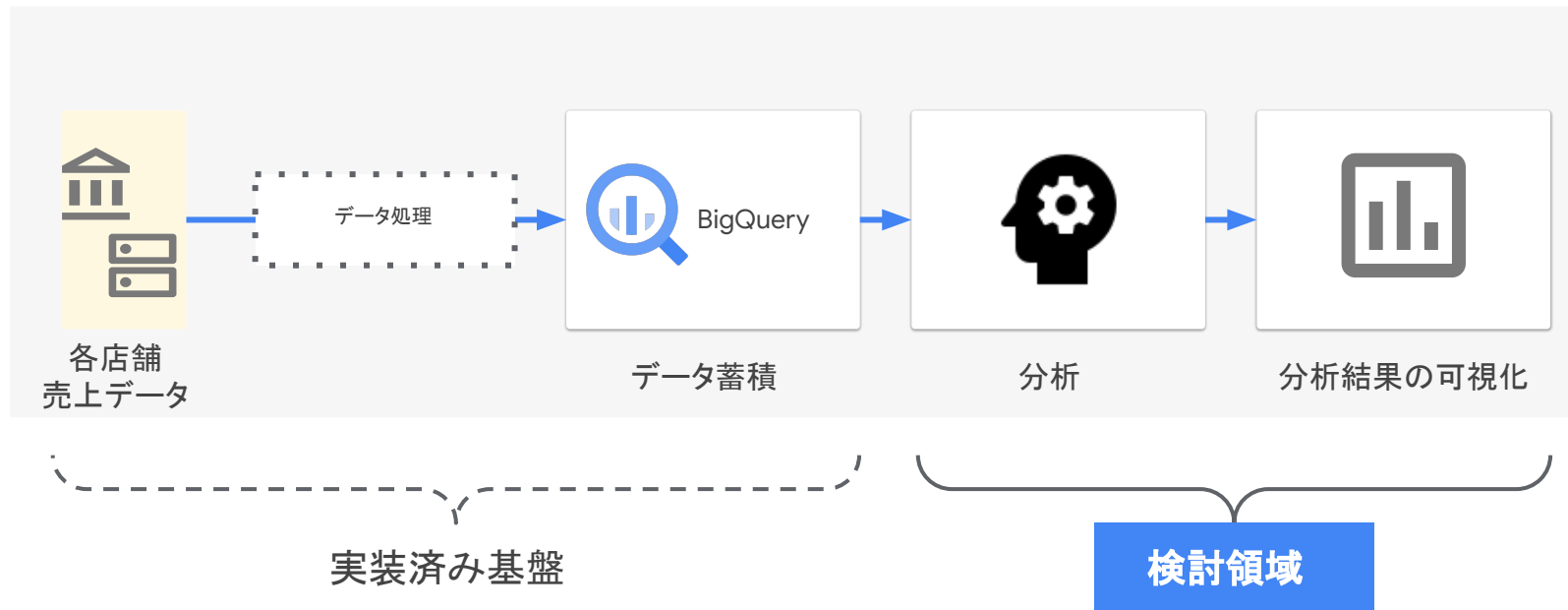
購買予測に対する  
アクションの計測

予測に対する社内アクションと効果を計測し、  
設定目標との評価する



データドリブンな  
意思決定へ

# 要望を満たす為に検討したこと



# 円滑なプロジェクト進行のための 事前準備・推奨される作業

## 目標の決定

やりたいこと、解決したい課題を明確にしておくが良いです。漠然と何か分析できるのでは？だとGoalがぶれてしまいます

## ビジネスロジック共有

社内で蓄積されているドメイン知識の共有ができる環境の整備をしておく  
と評価軸の検討、エンドユーザーの要望との齟齬が無い分析・予測ができます。部署間で仲良くしておきましょう

## 評価軸の決定

何を計測するのかをビジネスロジックな観点からも考えて置くことがよいです

## データの加工

分析したい目的によって必要となるデータが異なるので、事前に処理出来ているとよいです

ケースと必要データ例

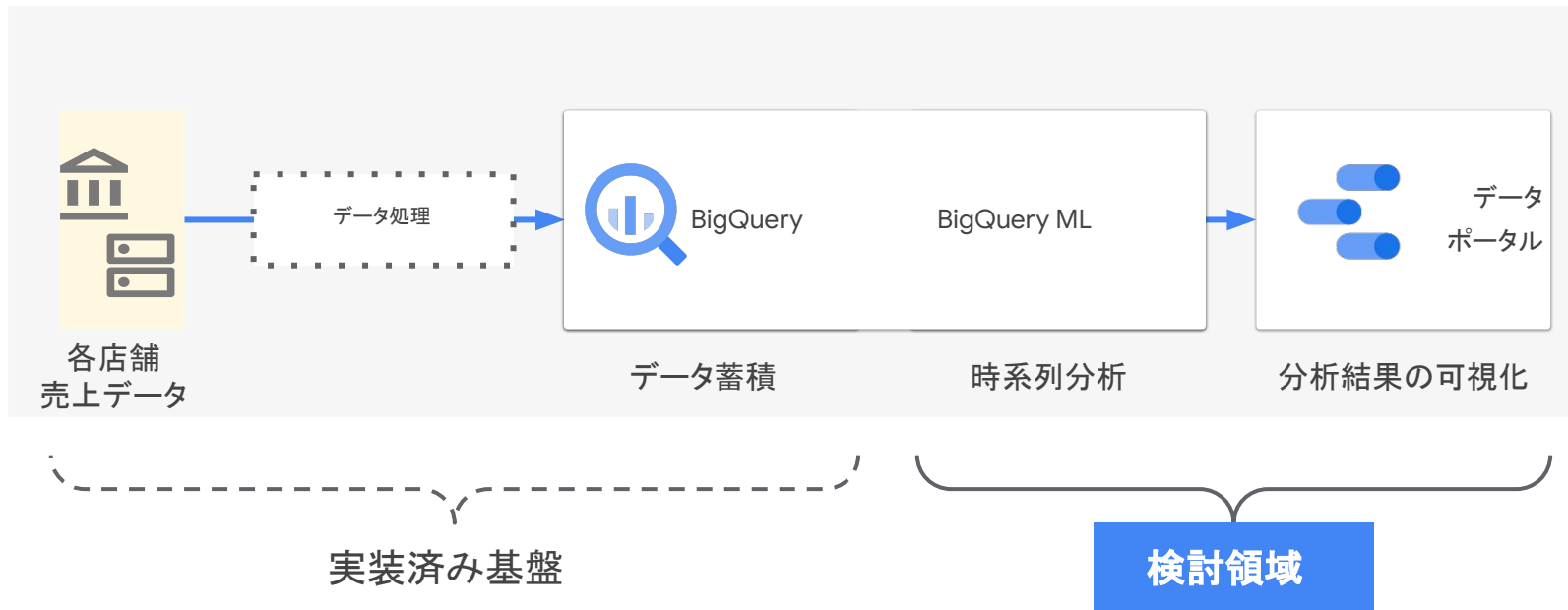
【商品レコメンド】

- ・ユーザーデータ
- ・商品データ

【違反ラベルの分類・検知】

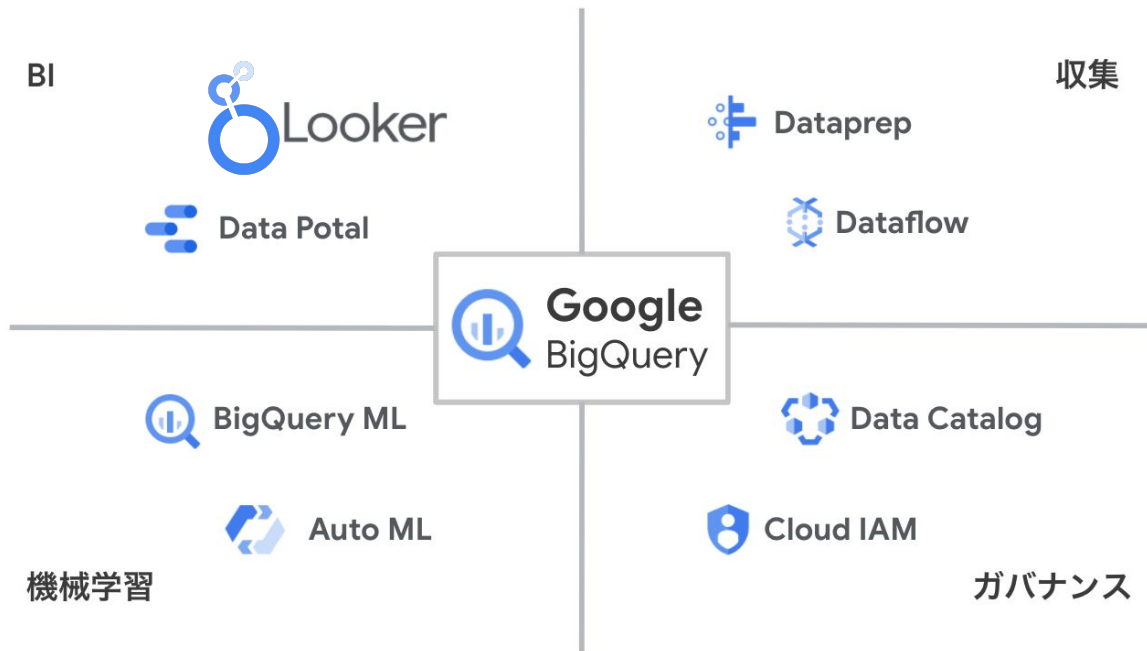
- ・各ラベルごとのデータ

# 実現方式検討結果




# BigQuery の紹介

データ活用を実施するための周辺ツールもたくさん存在しているため  
分析やデータ活用のハードルが低くなる



# BigQuery ML の紹介

## BigQuery ML



BigQuery ML ではモデルが事前に組み込まれており  
モデルのパラメータチューニングや評価の算出を  
自動で実施、モデルなどは SQL クエリによって構築可能

## Custom ML TensorFlow

BigQuery ML に組み込まれているモデル以外が必要  
になった場合、TensorFlow で書いたモデルのインポート  
が可能、モデルの管理を **BigQuery ML で一括管理が可能**

**BigQuery でデータやモデルの一括管理を実施**



# BigQuery ML の紹介

## 標準クエリでMLモデル作成と実行ができる機能

### 使った分だけの課金

学習、推論で利用したリソースのみの課金  
GPU などのリソースの管理が必要ない

### データの移動不要

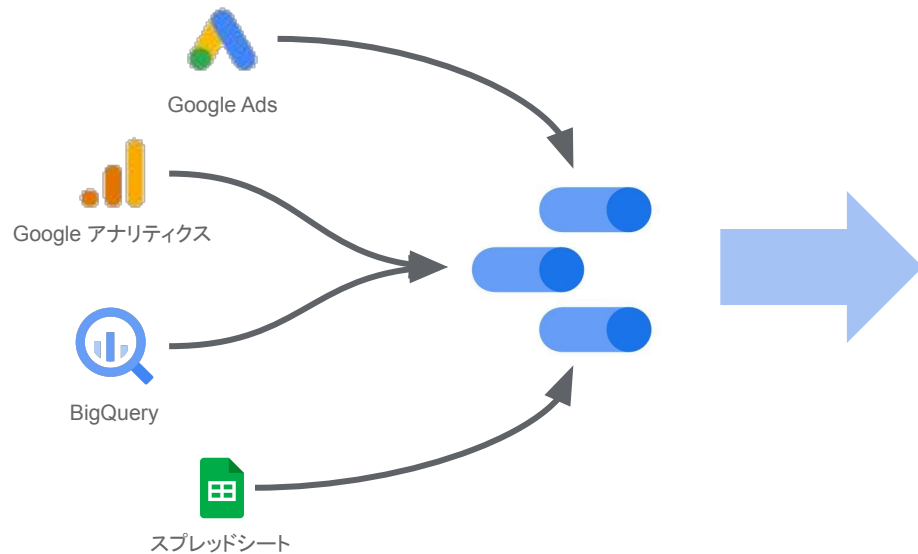
データのエクスポートは不要でモデルの開発  
とデータ探索に集中できる

### 機械学習の民主化
















スプレッドシートを使用してモデルの構築と実行  
ができる

```
CREATE MODEL
  project_name.model_name
OPTIONS
  (MODEL_TYPE = "arima_plus",
   TIME_SERIES_TIMESTAMP_COL = "TIMESTAMP_COL",
   TIME_SERIES_DATA_COL = "DATA_COL",
   TIME_SERIES_ID_COL = "CATEGORY_COL",
   HORIZON = 14,
   AUTO_ARIMA = TRUE,
   DATA_FREQUENCY = "DAILY",
   HOLIDAY_REGION = "JP"
  ) AS
SELECT
  TIMESTAMP_COL,
  DATA_COL,
  CATEGORY_COL,
FROM
  `project_name.train_dataset_name`
```

## データポータルの特徴

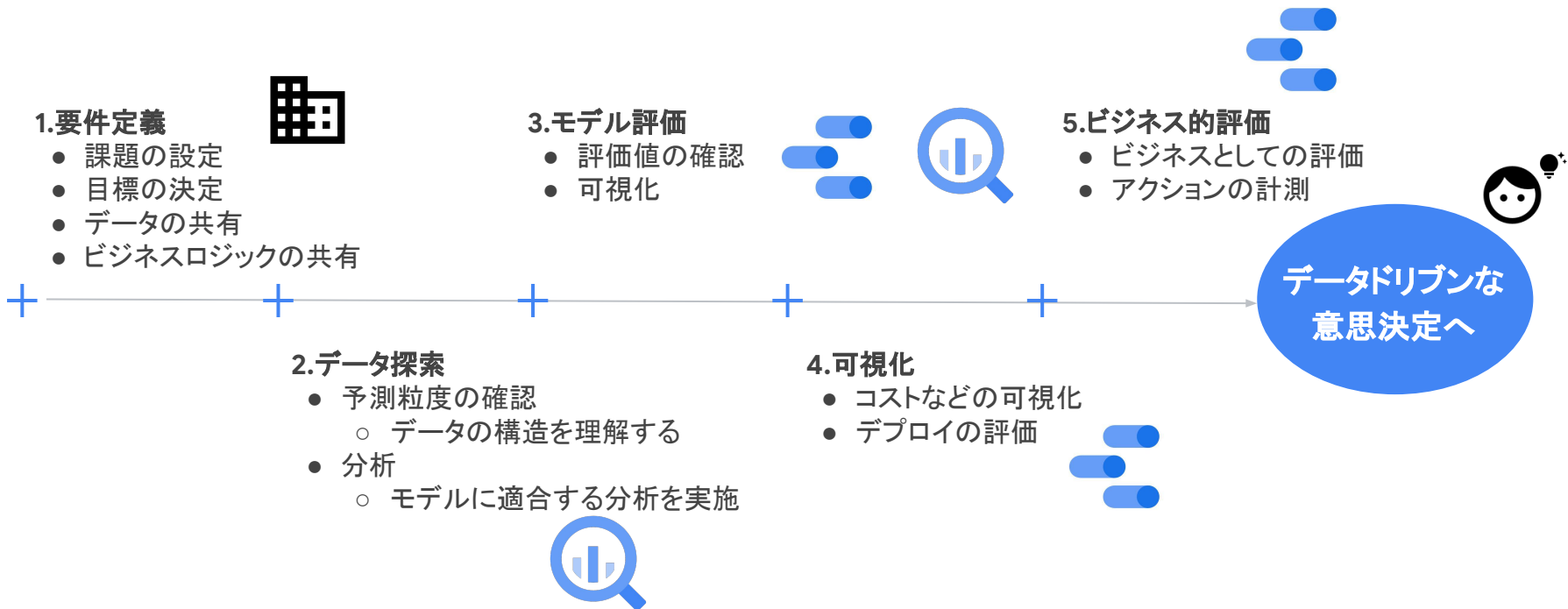


## 他の候補は...？

サービス名	機能量	開発コスト	運用コスト	分析速度	お手軽さ
BigQuery ML	 テーブルデータの モデルが主体	 標準クエリで実装が 可能で追加構成不要	 環境面でも学習・推論 環境を別にできる	 BigQuery 内 処理のため	 BigQuery に閉じてい る為迅速に分析可能
VertexAI	 画像、テキストも できる	 他コンポーネント との接続等が発生	 管理コンポーネントが 増える	 複数モデルを用いる ため時間がかかる	 機能が多い半面 設定項目が多い
Python (TensorFlow, Pytorch)	 実装次第で変動	 手組のためコスト高	 スキル保持者が必要	 実装/構成次第	 できることは非常に 多いがスキルが必要

※今回の事例でのお客様のご要望である手軽さという観点でまとめたものです

# データドリブンを実現する 購買予測/意思決定のプロセス



# データドリブンを加速する BigQuery ML のすごさ



ML 分析に必要な  
一般的な環境



データ  
蓄積環境



モデル管理  
ライブラリ



学習環境  
(モデル作成)



推論環境  
(モデル実行)

BigQuery ML の場合



BigQuery 環境とクエリだけ！

# データドリブンを加速する BigQuery ML で手軽にできる購買予測



BigQuery

## ML モデルの作成

```
1 CREATE MODEL
2   project_name.model_name
3 OPTIONS
4   (MODEL_TYPE = "arima_plus",
5    TIME_SERIES_TIMESTAMP_COL = "TIMESTAMP_COL",
6    TIME_SERIES_DATA_COL = "DATA_COL",
7    TIME_SERIES_ID_COL = "CATEGORY_COL",
8    HORIZON = 14,
9    AUTO_ARIMA = TRUE,
10   DATA_FREQUENCY = "DAILY",
11   HOLIDAY_REGION = "JP"
12 ) AS
13 SELECT
14   TIMESTAMP_COL,
15   DATA_COL,
16   CATEGORY_COL,
17 FROM
18   `project_name.train_dataset_name`
```

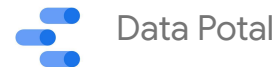
## ML モデルでの推論

```
1 SELECT
2   *
3 FROM
4   ML.FORECAST(MODEL `dataset_name.model_name`,
5               STRUCT(30 AS horizon,
6                     0.8 AS confidence_level))
```

## ML モデルの評価

```
1 SELECT
2   *
3 FROM
4   ML.EVALUATE(MODEL `dataset_name.model_name`)
5
```

# BigQuery ML を用いた購買予測結果

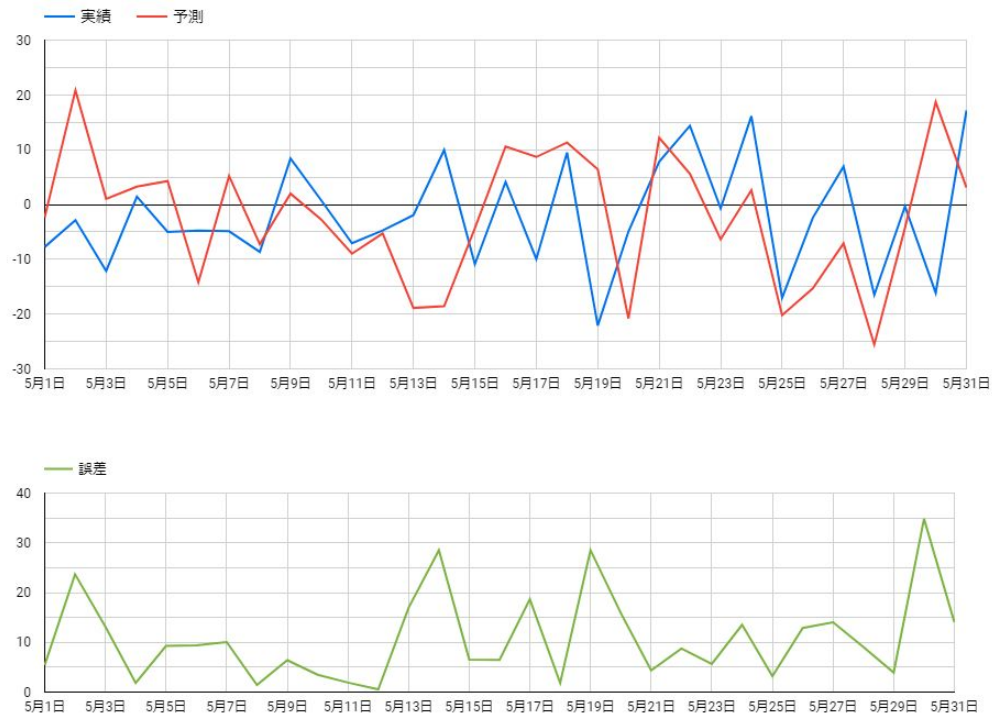


入力データ: 売上実績  
出力データ: 売上予測値

売上予測値と実績、予測誤差

日付	売上予測値	売上実績	予測誤差
2022/05/01	-2.299	-7.750	5.451
2022/05/02	20.841	-2.844	23.686
2022/05/03	1.029	-12.139	13.168

⋮

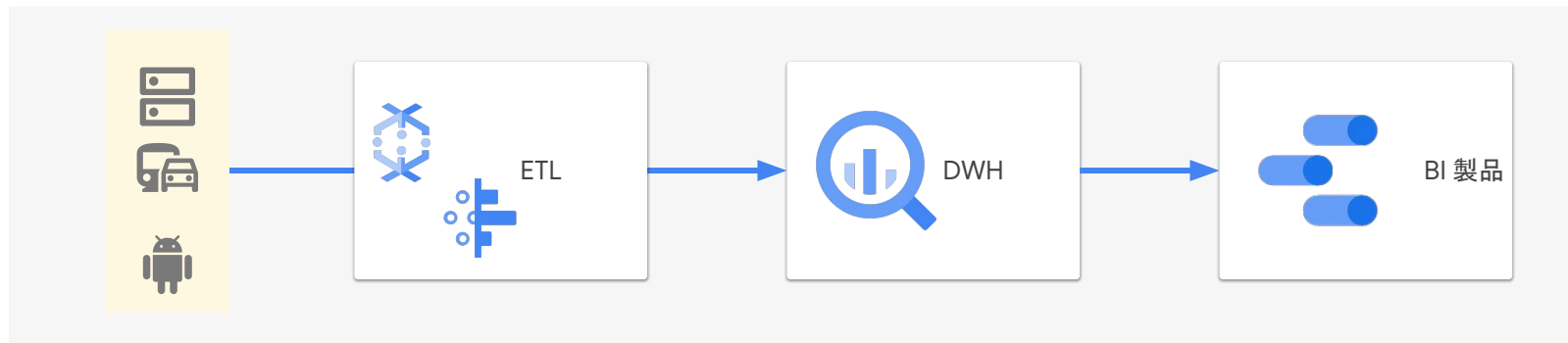




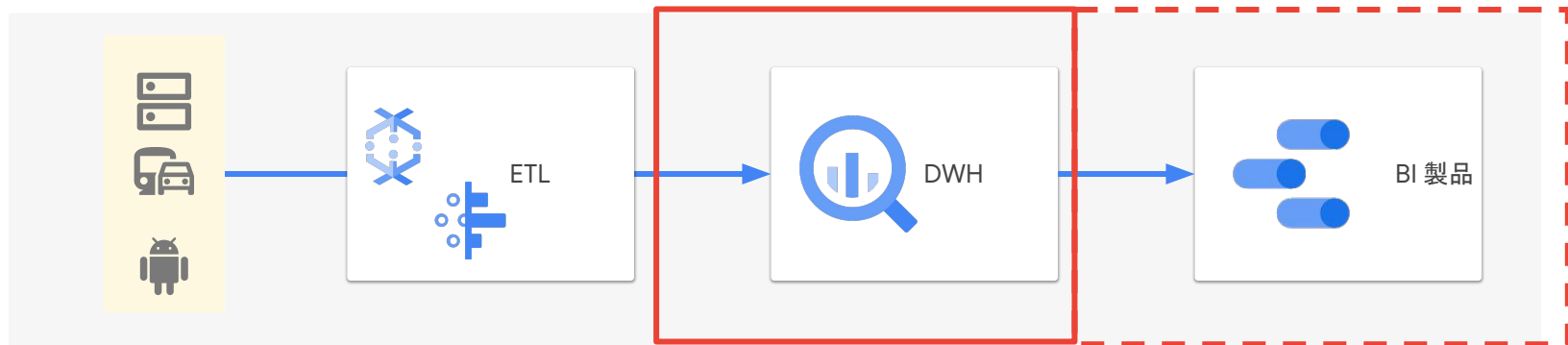
# Wrapup



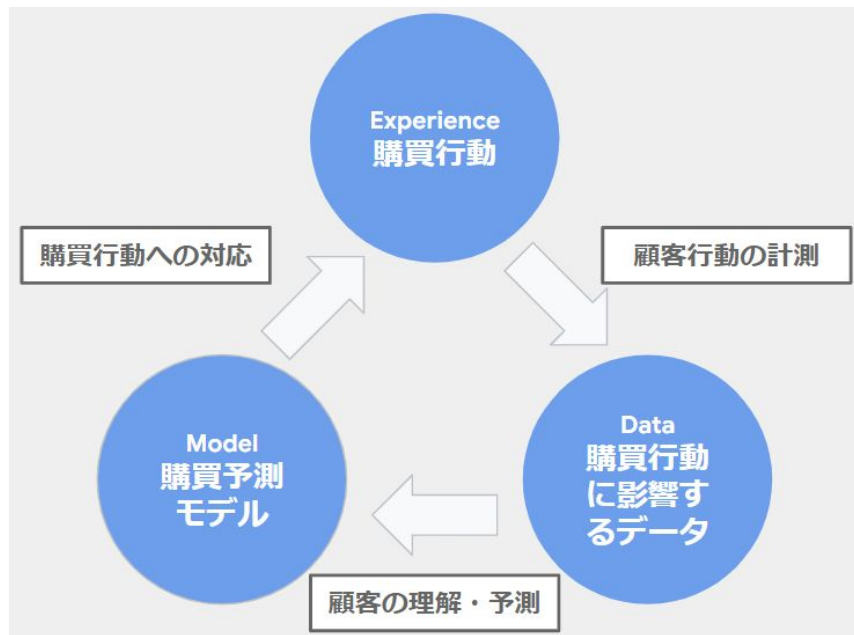
# データドリブン(購買予測)のための データ活用プロセス



# データドリブン(購買予測)のための データ活用プロセス



# 本セッションのまとめ



01

昨今のビジネスでは、  
**データドリブンな経営・判断**が重要になっている

02

データドリブンをするためのサイクル実行  
には、**BigQuery (BigQuery ML)**と**データポータル**が適切

03

データドリブンサイクル実行をご検討の企業様はぜひ **NI+C** へご相談を！

# Thank you.

