

Google Cloud

INSIDE

Retail

#gc\_inside

Google Cloud

# データ分析を内製化したい方へ なぜ小売業界で BigQuery が選ばれるか

.....  
グーグル・クラウド・ジャパン合同会社  
カスタマーエンジニア  
唐澤 匠

Google Cloud

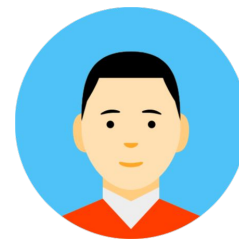
# 想定ケース

在庫を減らしたいから  
データを利用して分析して！  
ジャストインタイムってやつね  
あと、お金ないから内製でよろしく



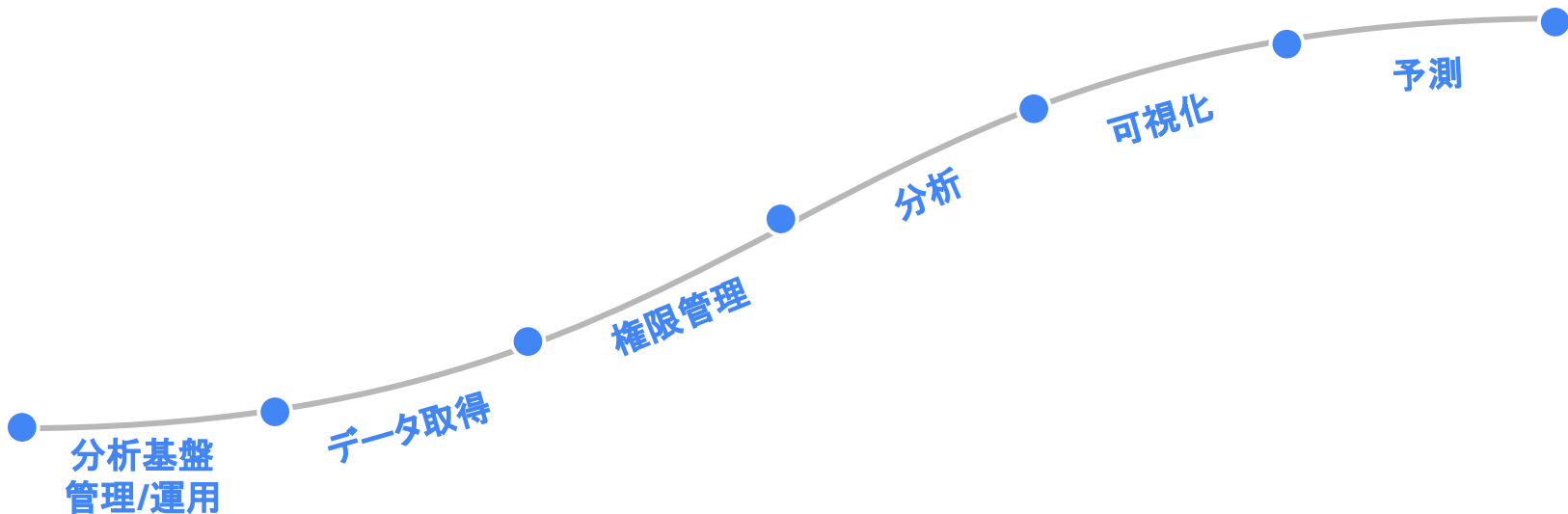
上司

???



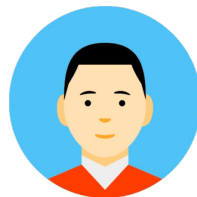
新設 DX 室  
A 君

# ゴールまでの道筋



# 小売業界の分析基盤管理/運用の課題

人がいないけれど管理/運用は  
どうしよう



顧客情報扱うからセキュリティしっかり  
したいんだけど

Google Cloud



分析基盤  
管理/運用



データ取得



権限管理



分析



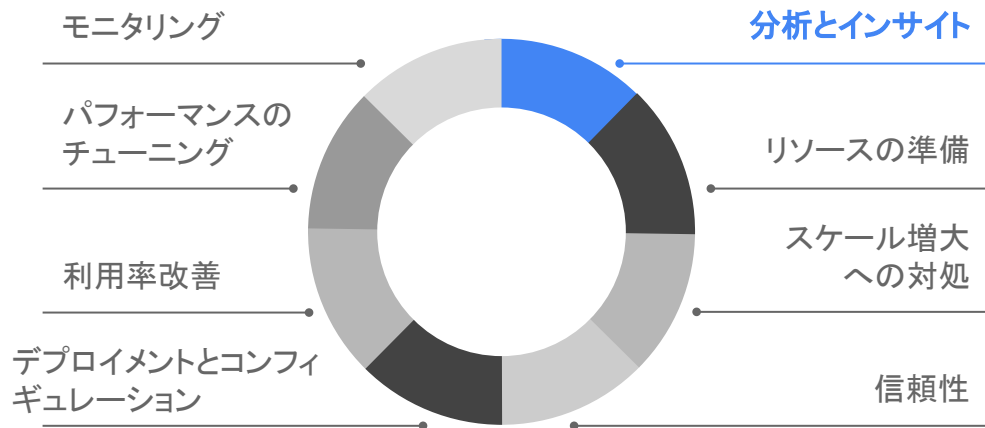
可視化



予測

# サーバーレス データ分析基盤

## 従来のデータ分析基盤



## サーバーレスの データアナリティクスモデル

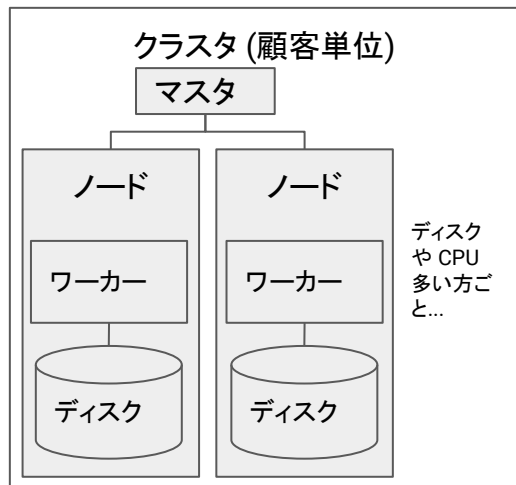


Google Cloud



# DWH アーキテクチャ比較

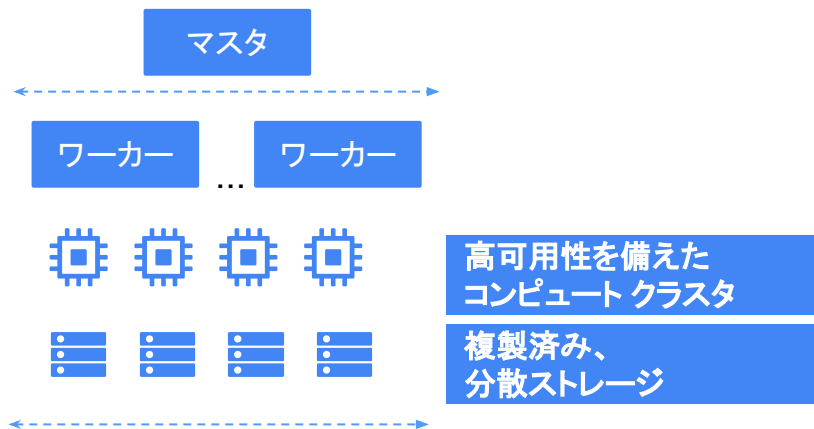
## 一般的なクラウド DWH アーキテクチャ



顧客ごとに  
クラスタ...

- 事前プロビジョニング、顧客環境ごとクラスタ  
= 規模の経済がDWHレイヤでは成立しない
- ディスクはコンピュートノードとセットでつながっているため、コンピュートとストレージが一体したスケーリング  
= どちらか利用の多い方を追加プロビジョンする必要がある

## BigQuery の アーキテクチャ



- プロビジョニング不要、リージョンレベルのクラスタ  
= 規模の経済がDWHレイヤでも成立
- ストレージ、コンピュートについて動的な割当
- コンピュートとストレージが独立したスケーリング  
= 分離して課金が可能で、ユーザーの使い方(ストレージヘビーor コンピュートヘビー)に応じ従量課金

# セキュリティ

第三者による**監査**と**認証**により  
BigQuery の各種コンプライアンス要件への準拠を保証



Google Cloud



分析基盤  
管理/運用



データ取得



権限管理



分析

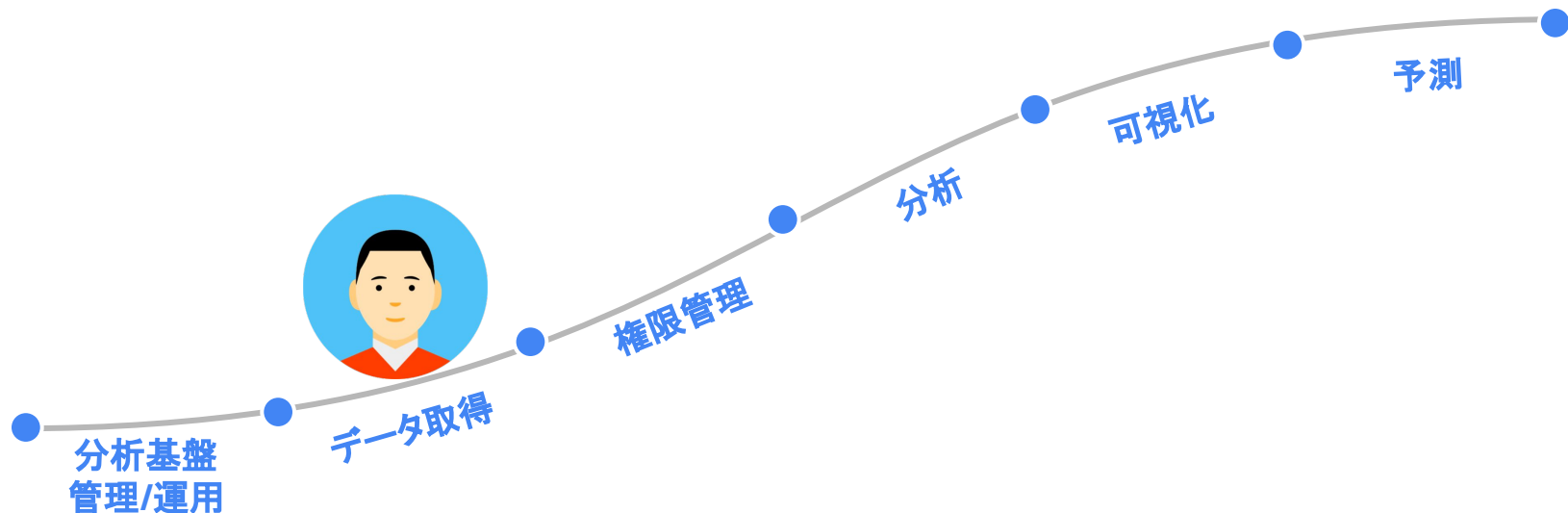


可視化



予測

# ゴールまでの道筋: データ取得

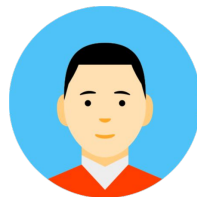


Google Cloud



# 小売業界のデータ取得の課題

在庫データはリアルタイムに取り込みたいけれど出来るかな？



Eコマースのデータって簡単に取れるのかな？

Salesforceのような営業ツールと連携して出来るのかな？

Google Cloud



分析基盤  
管理/運用



データ取得



権限管理



分析



可視化



予測

# BigQuery ストリーミング挿入

- データが読み込まれてから数秒以内でリアルタイムに分析が可能
- 標準で数百万 QPS(クエリ / 秒) Exactly once セマンティクス
- クエリ パフォーマンスの低下なし



Google Cloud



分析基盤  
管理/運用



データ取得



権限管理



分析



可視化



予測

# データ転送

Data Transfer Service: 100 以上の連携元から BigQuery へのデータロードが可能。



Adroll  
Adobe Analytics  
Apache Kafka  
Apple Search Ads  
Amazon Aurora  
Amazon Cloudfront  
Amazon DynamoDB  
Amazon Kinesis Firehose  
Amazon S3  
Asana  
Azure Blob Storage  
Azure Functions  
Azure SQL  
Bing Ads

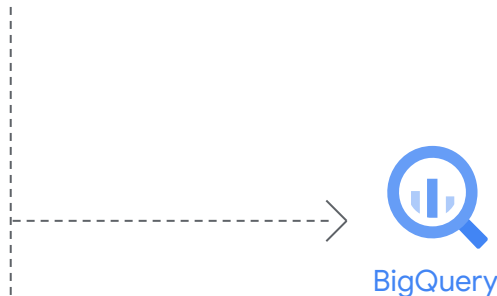
Braintree  
Criteo  
Desk.com  
Dropbox  
Eloqua  
Facebook Ads  
Freshdesk  
FTP/SFTP  
GitHub  
Greenhouse  
Help Scout  
Heroku Postgres  
Hubspot

Intercom  
iTunes Connect  
Jira  
Klaviyo  
LinkedIn  
Magento  
Mailchimp  
Mandrill  
Marketo  
MariaDB  
Microsoft TFS  
Mixpanel

MongoDB  
Moz  
MySQL  
Netsuite  
Optimizely  
Oracle RAC  
Oracle RDS  
Outbrain  
Pardot  
Pinterest  
PostgreSQL  
Quickbooks

QuoraAds  
Recharge  
Recurly  
Sailthru  
Salesforce  
Salesforce Marketing Cloud  
Salesforce IQ  
Segment  
SendGrid  
SFTP  
Shopify

Snowplow  
SQL Server  
Stripe  
Taboola  
Twitter Ads  
UserVoice  
Xero  
Yahoo Gemini  
Zendesk Chat  
Zuora



Google Cloud



# BigQuery Export

- 有償版の Google アナリティクス 360 で使用可能
- Google Analytics 4 では無償で使用可能
- GA のサンプリングされていないレポートを BigQuery に抽出
- ユーザのヒットレベル (pageview, event) のデータを統計的分析に活用可能
- 5 つ以上のタッチポイント参照
- CRM など他の社内の他のデータと統合した分析も可能



Row	fullvisitorId	transactionRevenue
1	9417857471295131045	10589140000
2	6236695646664370912	4009560000
3	9308310352918219134	2933610000
4	9417857471295131045	2548500000
5	7445235885559107095	2490000000
6	875410720669459903	2407200000
7	6052206455243513477	1794400000

Google Cloud



分析基盤  
管理/運用



データ取得



権限管理



分析

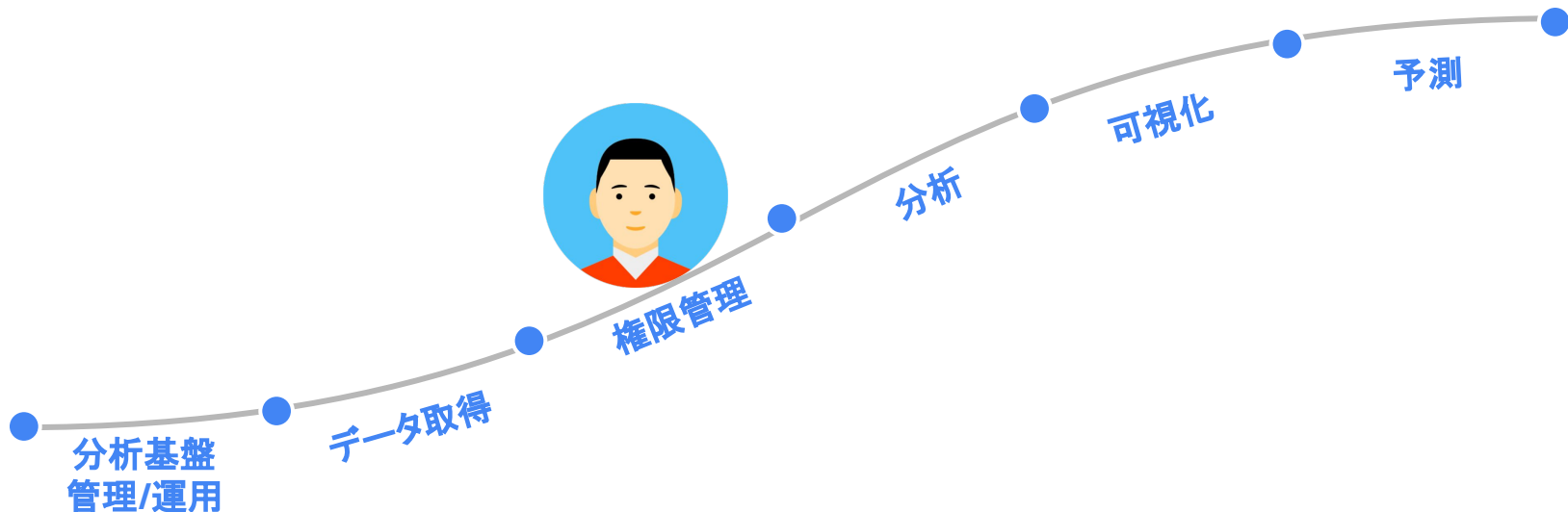


可視化



予測

# ゴールまでの道筋: 権限管理

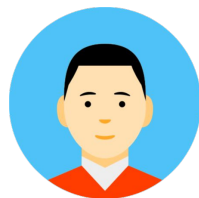


Google Cloud



# 小売業界の権限管理の課題

店舗データは特定の部署にしか  
見せたくないけれど出来るのかな



個人情報が入っていたら自動的にタ  
グが付かないかな

Google Cloud



分析基盤  
管理/運用



データ取得



権限管理



分析



可視化



予測

# 列レベルのアクセス制御

- テーブルごとに加えて、テーブル内のカラムごとにデータアクセス制御を実現
- 取り扱いのクラス分けが必須となる規制に準拠可能
- 個人情報や売上情報などのアクセス制御を実施可能

Name ↑	ID	Description
▼ Restricted	3247623653529953690	Highly Restricted Data
▼ PHI	4081878655865131464	Patient Health Information
🔒 Drug_Details	348889402753783706	Details about a drug prescribed
🔒 NHS_Number	4099447459463431825	Patient ID
🔒 Treatment_Details	6587645476172403944	Details about a treatment or condi...
▼ PII	1690556303680165819	Personally identifiable data
🔒 Email	5606010836299662298	Email address
🔒 IMEI	7077445421065241870	Cellphone hardware ID
🔒 IP_Addr	2449414728069309088	IP Address of a session/connection
🔒 Personal_Car_VIN	7187828684927708308	Vehicle Identifier
🔒 Phone_Num	8401384437536803987	Phone number
🔒 SSN	9118232350617909155	US Social Security Number
▼ Sensitive	5013925770628759512	Sensitive Data
▼ Financials	358397642325435489	Financial Data
🔒 Bank_Account	8370833355300570	International Bank account ID
🔒 Credit_Card_Num	6313828804358283165	Credit Card number
▶ Unrestricted Data	8097084282273622955	Unrestricted Data, broad access

Drug\_And\_Treatments

⚠️ 2/9 columns in table shown. 7 columns restricted because you don't have access.

Schema Details Preview

Field name	Type	Mode	Policy tags ⓘ
GUID	STRING	NULLABLE	
Drug_Company	STRING	NULLABLE	
Drug_Name	STRING	NULLABLE	Data Sensitivity : Drug_Details
Generic	STRING	NULLABLE	Data Sensitivity : Drug_Details
NDC_Code	STRING	NULLABLE	Data Sensitivity : Treatment_Details
ICD10	STRING	NULLABLE	Data Sensitivity : Treatment_Details
ICD10_Desc	STRING	NULLABLE	Data Sensitivity : Treatment_Details
NHS_Number	INTEGER	NULLABLE	Data Sensitivity : NHS_Number
SSN_Number	STRING	NULLABLE	Data Sensitivity : SSN

Google Cloud



分析基盤  
管理/運用



データ取得



権限管理



分析

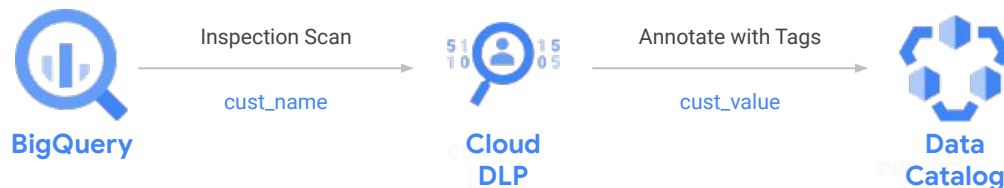


可視化



予測

# 個人情報検知

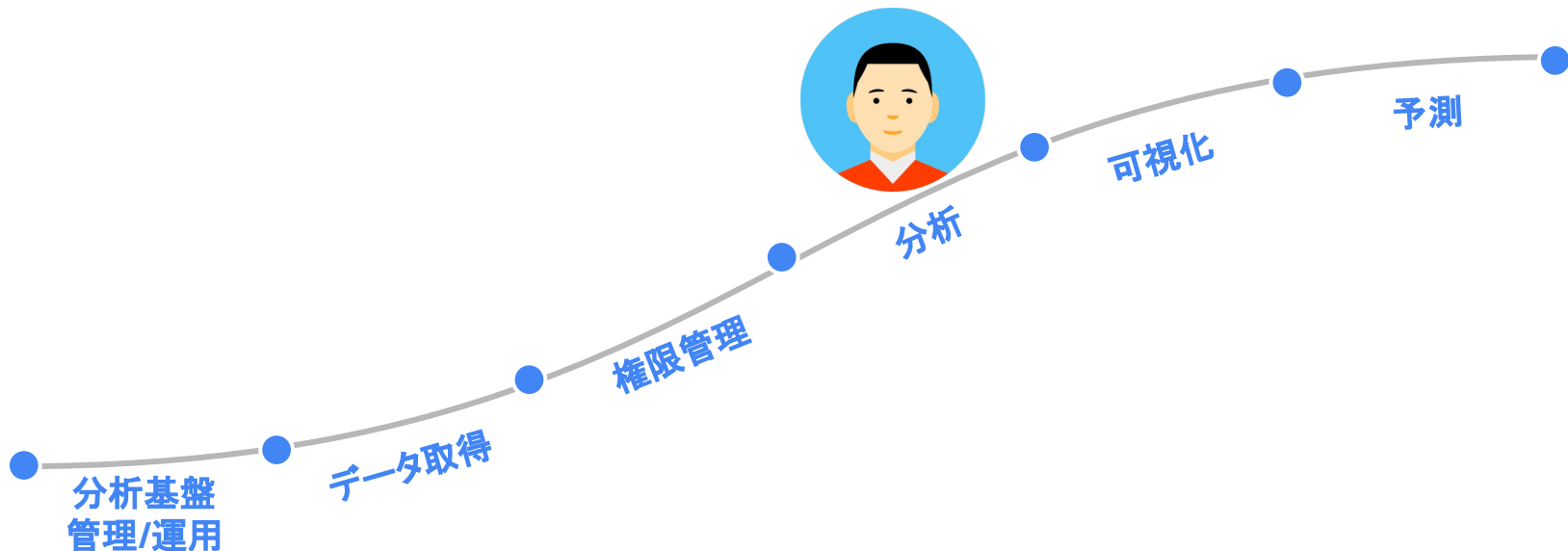


column:credit\_card\_number: DLP Result Scan Template (travel-bookings-236421.dlp\_result\_scan\_template)

Attribute	Display name	Value
scan_date_time	Date Time of Scan	04/04/2019 04:44
result	Result	CREDIT_CARD_NUMBER
has_pii	Has PII Information	true
num_rows_scan	Number of Rows to Scan in DLP	1000
result_percentual	Result percentual	1
all_result	all result	[CREDIT_CARD_NUMBER(LIKELY : 100.0%);]
num_rows_table	Number of Rows in the table scanned	13000010
result_likelihood	Results likelihood	LIKELY
num_match	Number of Results	1000

Cloud DLP により自動タグ付けされたメタデータタグ

# ゴールまでの道筋: 分析

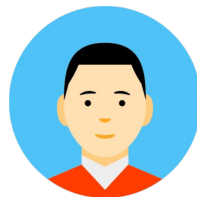


Google Cloud



# 小売業界の分析の課題

基本表計算ソフトを使っていて  
SQL 書ける人がいないんだけど  
どうしよう



店舗にお客さんが来店したときにそ  
の人の過去データとか簡単に調べら  
れないかな

Google Cloud



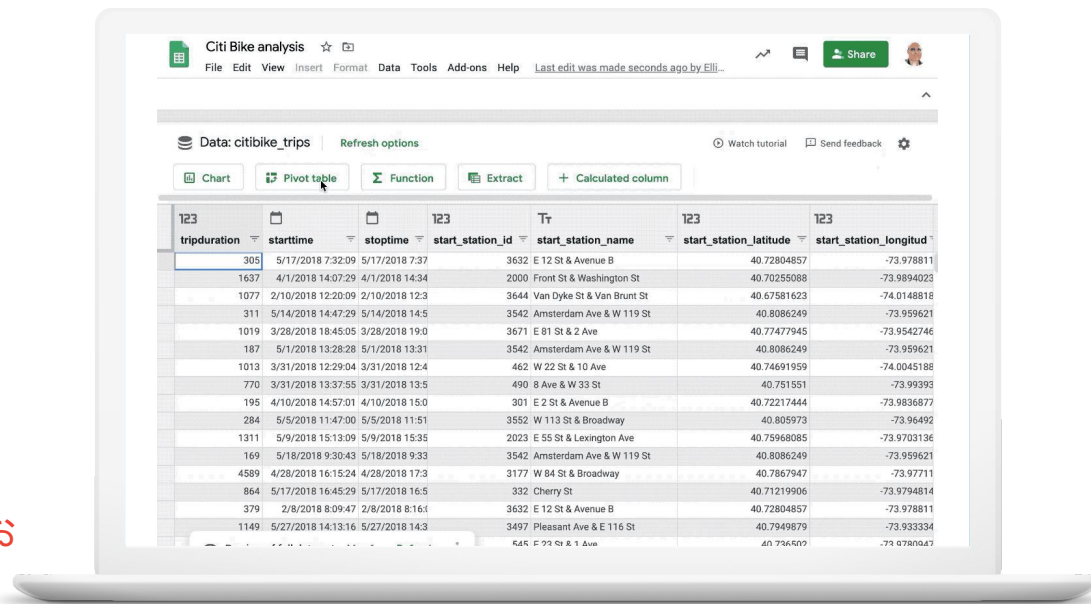
# Connected Sheet

BigQuery のパワーとスケールを使い慣れたインターフェースから利用可能

SQL の知識がなくても Google  
スプレッドシートによって数十億行の  
BigQuery データを分析できます

ピボット テーブル、グラフ、  
式のような使い慣れたツールを  
適用し、ビッグデータから簡単に  
インサイトを導き出せます

Google Workspace Enterprise SKU をお  
持ちの方は利用可能



Google Cloud



# Data QnA: 自然言語インターフェース

1

自然言語を使ったセルフサービス分析を通じてインサイトを民主化

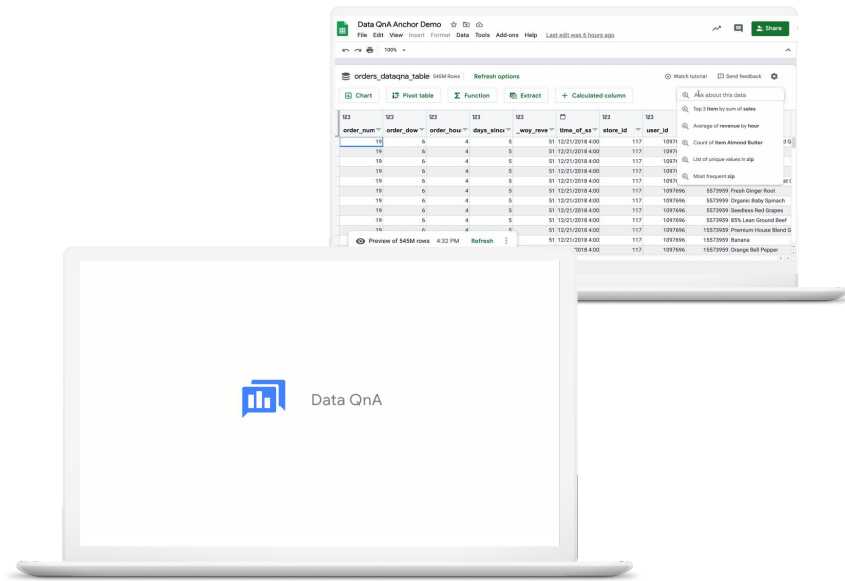
2

アドホック レポートを不要にすることで BI チームの生産性が向上

3

Google スプレッドシート、BigQuery、Chatbots、カスタム UI (API 経由)、Looker、Google Voice など、さまざまなインターフェースを介してアクセス

※現状は英語のみ対応



Google Cloud



分析基盤  
管理/運用



データ取得



権限管理



分析

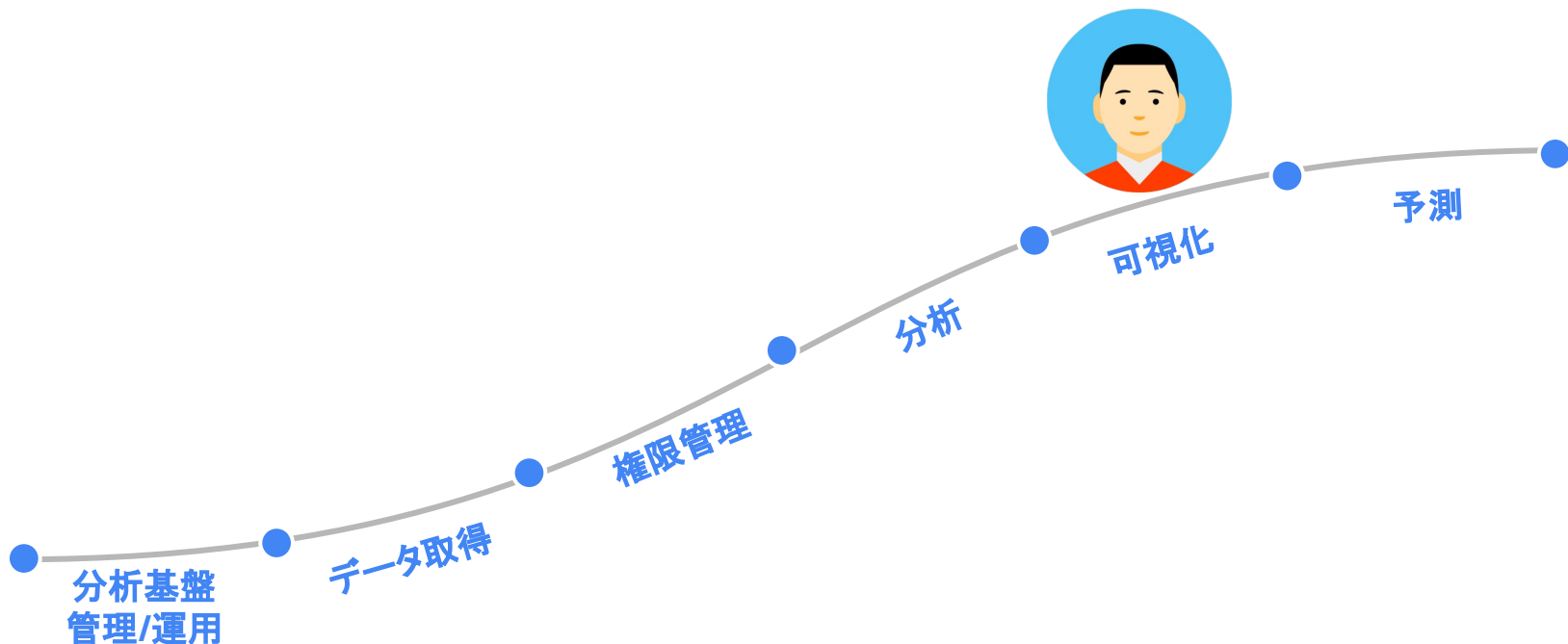


可視化



予測

# ゴールまでの道筋: 可視化

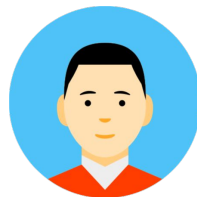


Google Cloud



# 小売業界の可視化の課題

スタッフがスマホで見れるように Web  
ベースのものがいいな  
なるべく安いのがベスト



リアルタイムに在庫を確認したいんだ  
けれど、表示に時間がかからないか  
な

Google Cloud



分析基盤  
管理/運用



データ取得



権限管理



分析



可視化



予測

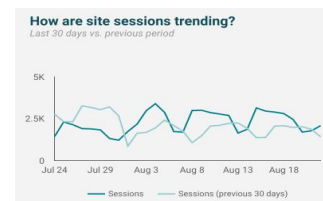
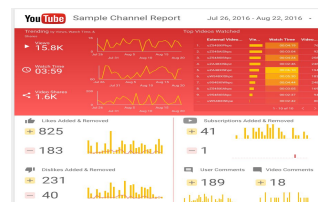
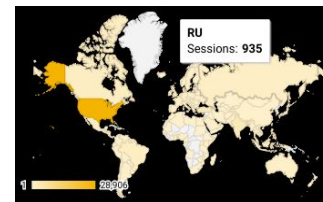
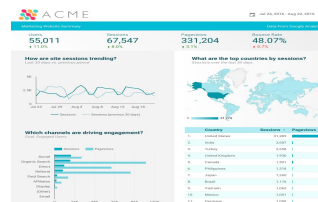
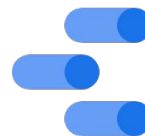
# データポータル

Web ベースのデータ可視化を無料で提供

BigQuery 含む様々なデータソースへのアクセス

コラボレーションによる作業

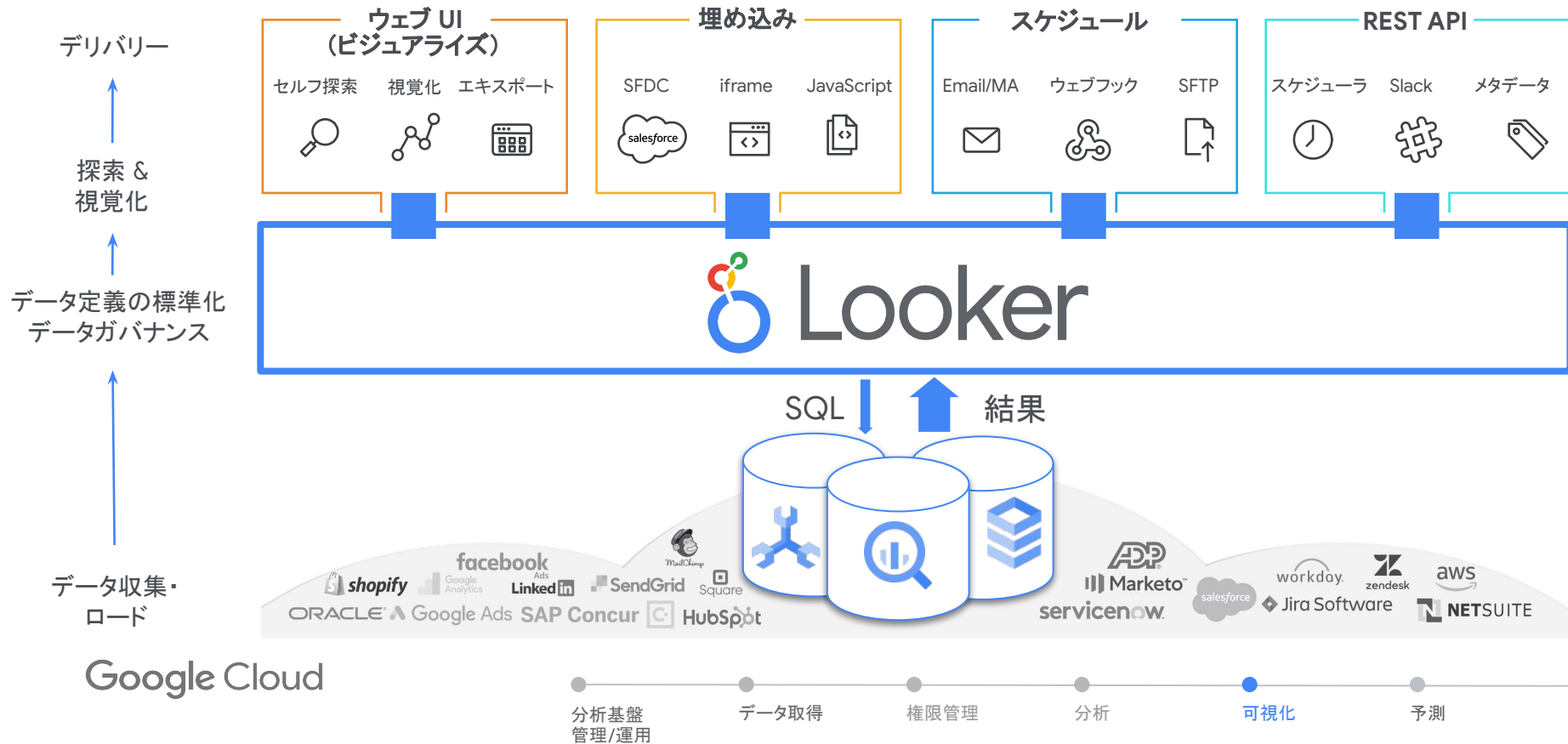
簡単なパブリッシュと共有



Google Cloud



# Looker

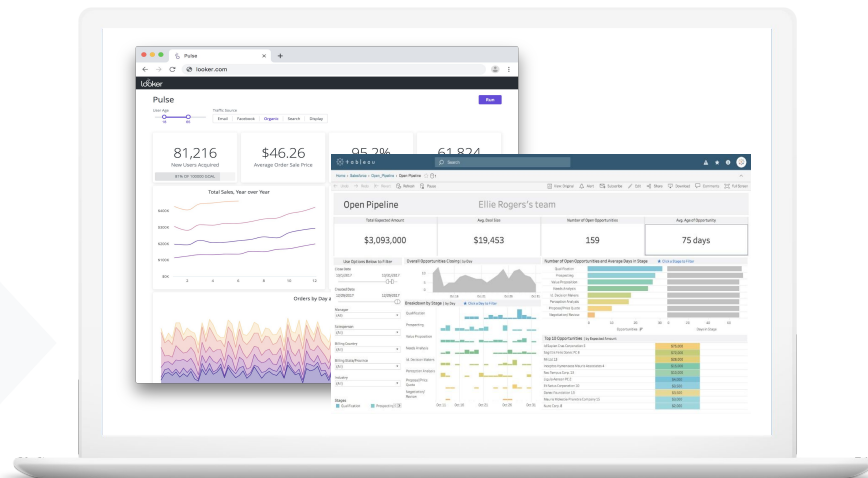


# BigQuery BI Engine



## BigQuery BI Engine

- 適応型キャッシュが、最もアクセス頻度の高いデータを自動的に BI Engine RAM に保持
- BigQuery でクエリを実行すると、結果が 1 秒未満で返される
- BigQuery の SQL を直接使用
- Data Portal、BigQuery UI に加えて **New** Looker、Tableau のプレビュー版をリリース



Google Cloud



分析基盤  
管理/運用



データ取得



権限管理



分析

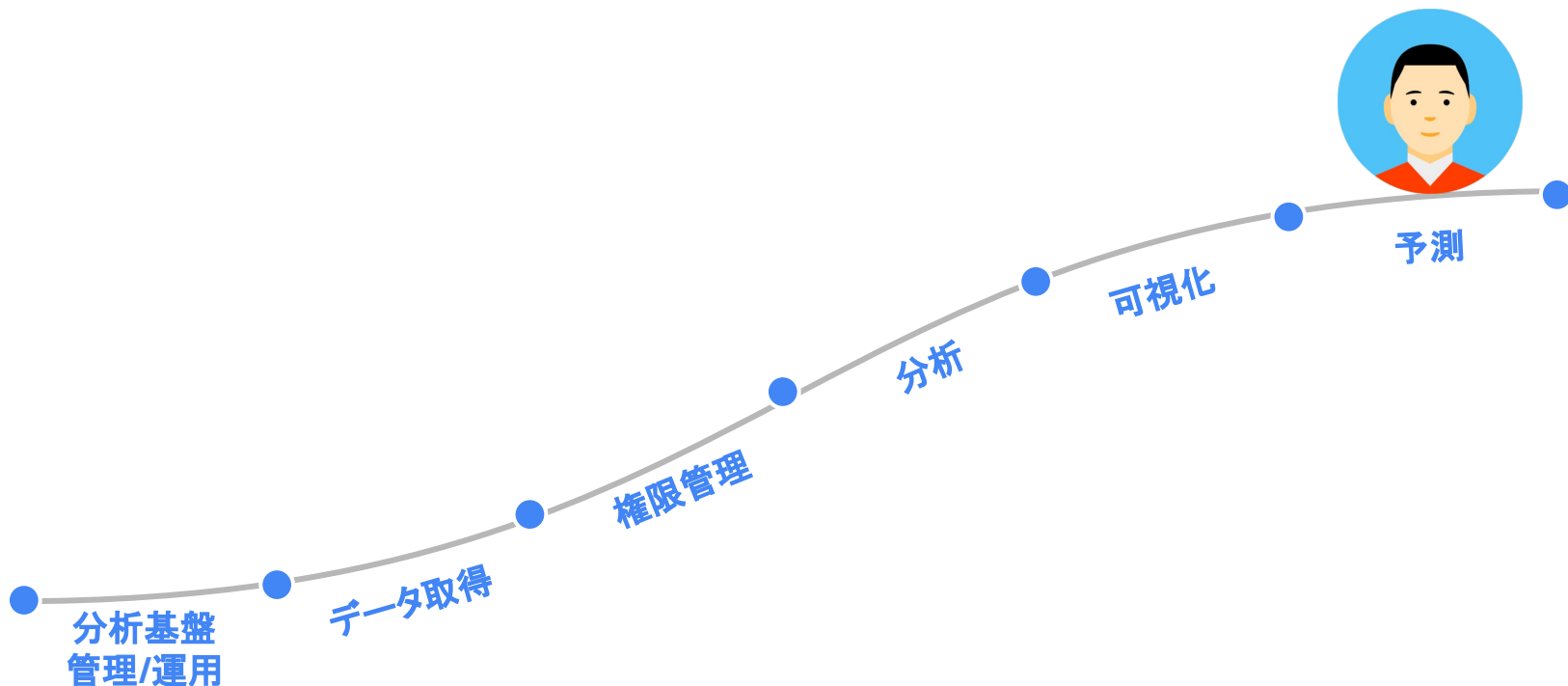


可視化



予測

# ゴールまでの道筋: 予測



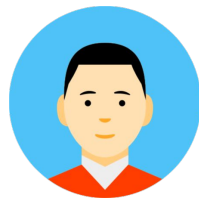
Google Cloud



# 小売業界の予測の課題

商品ごとの売上が時系列で  
どう推移するのか予測したいな

お客さんごとのおすすめの商品なん  
かを予測できないかな



そもそもエンジニアがいないから、コード  
やクエリを書かないで機械学習したいな

Google Cloud



分析基盤  
管理/運用



データ取得



権限管理



分析

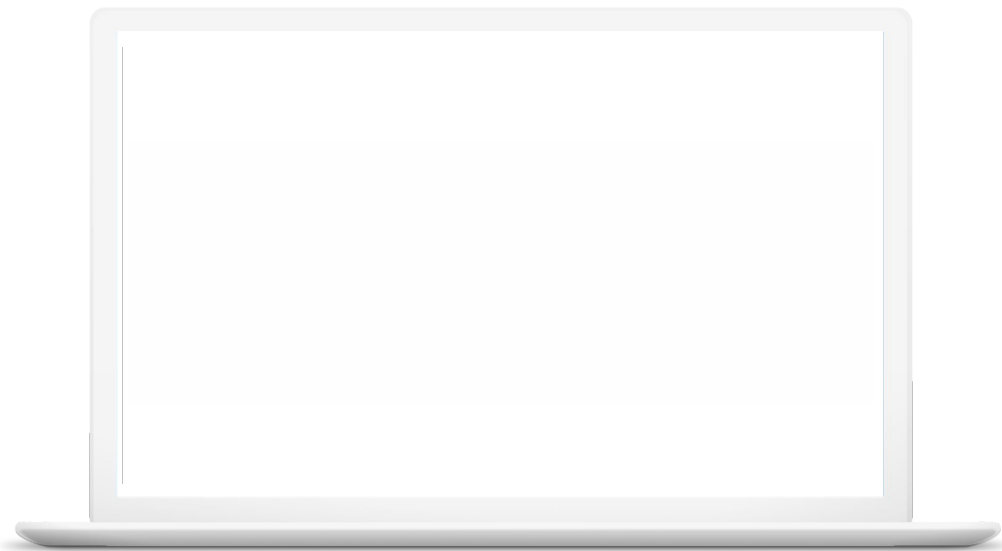


可視化



予測

# BigQuery ML



BigQuery からデータを移動せずに  
**機械学習モデル作成**

—  
使い慣れた SQL で  
**開発速度を向上**

—  
ハイパーパラメータチューニング  
などの ML タスクを**自動化**

—  
インフラの管理やセキュリティを **一元化**

Google Cloud



分析基盤  
管理/運用



データ取得



権限管理



分析



可視化



予測

# BigQuery ML のサポートモデル

## 分類

- ロジスティック回帰
- DNN 分類器 (TensorFlow)
- XGBoost
- AutoML Tables

## 他のモデル

- k 平均法クラスタリング
- 時系列予測
- レコメンデーション : マトリックス ファクタライゼーション

## 回帰

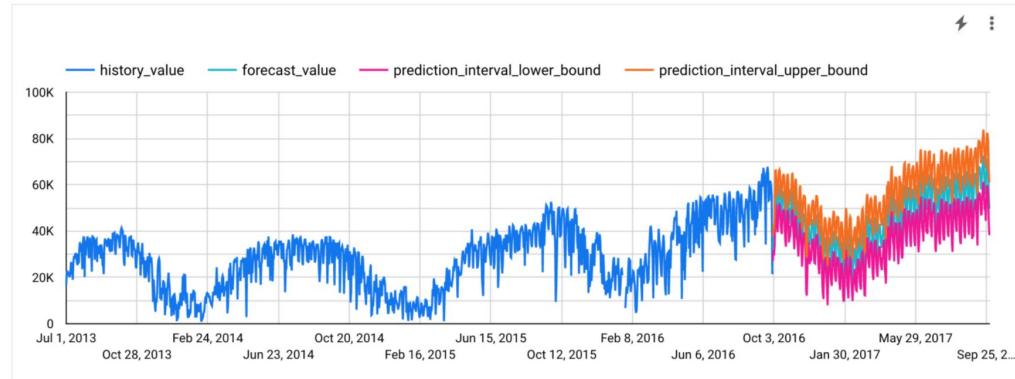
- 線形回帰
- DNN 回帰器 (TensorFlow)
- XGBoost
- AutoML Tables

## モデルのインポート / エクスポート

- バッチおよびオンライン予測のための TensorFlow モデル

# ARIMA モデル

- 時系列解析のためのアルゴリズム
- 時系列解析とは、過去と未来の関係性から将来を予測する手法
- 季節性や国ごとの休日効果を盛り込むことが可能



Google Cloud



分析基盤  
管理/運用



データ取得



権限管理



分析



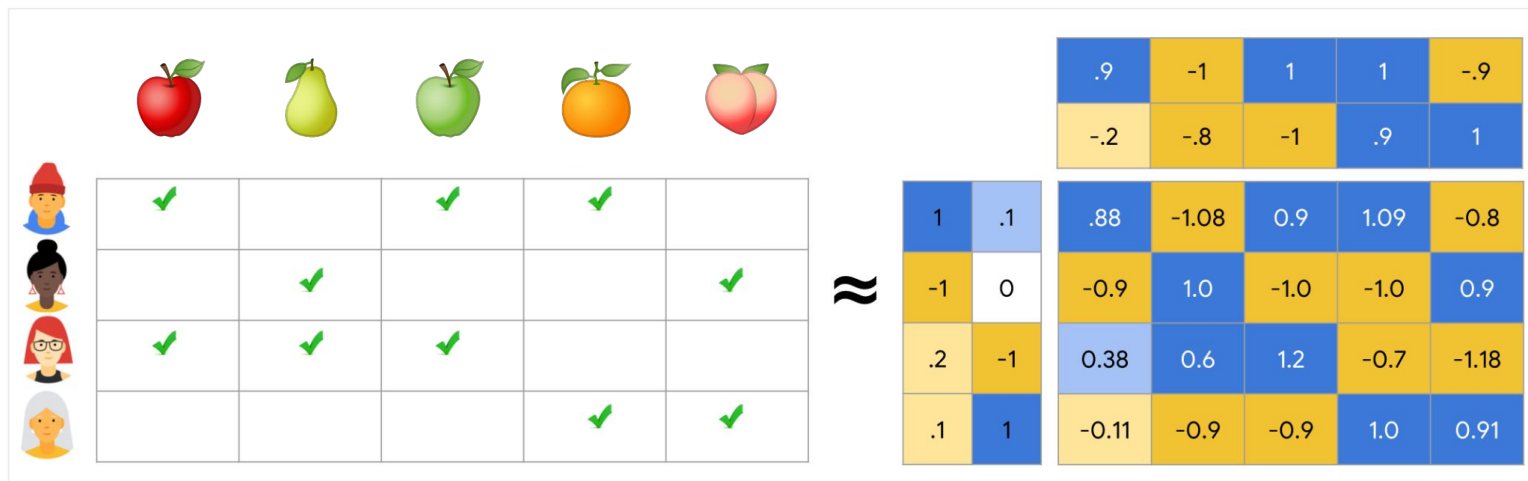
可視化



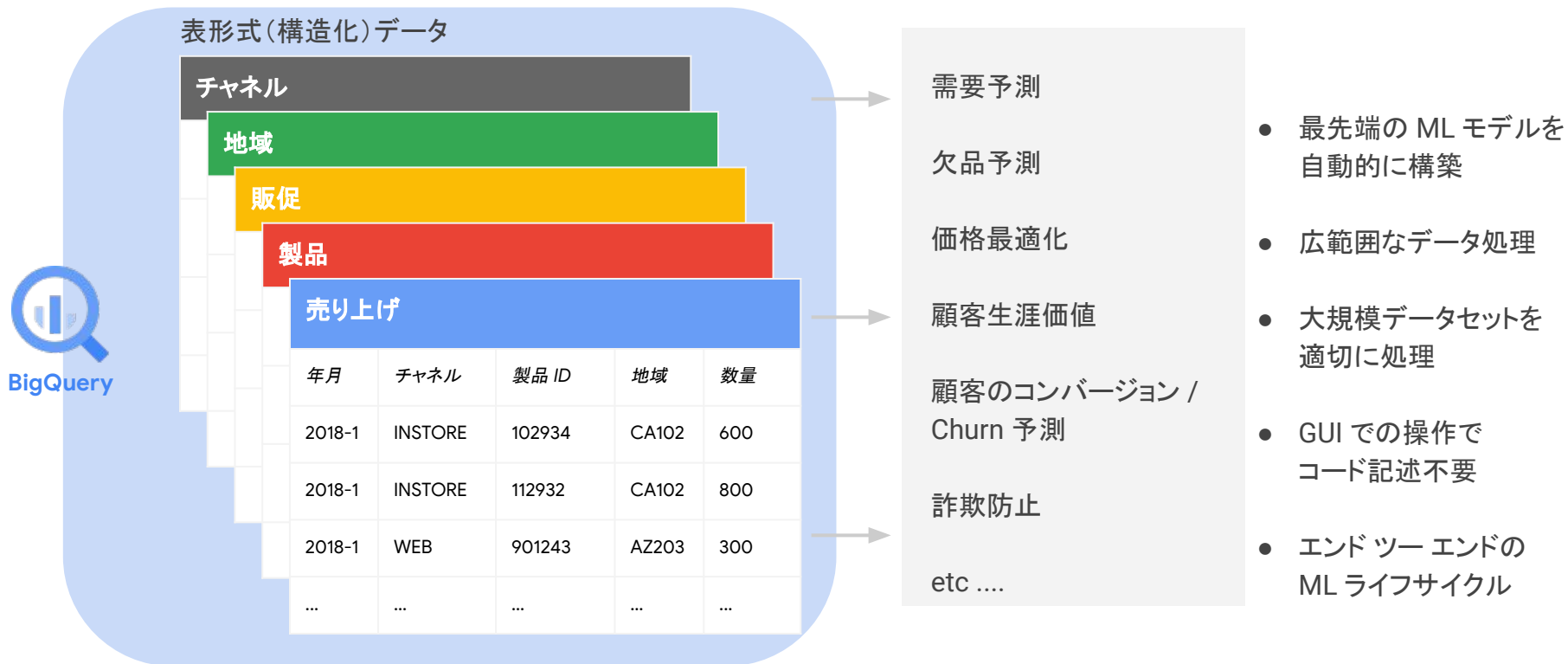
予測

# レコメンデーション(Matrix Factorization)

- 趣味や嗜好からおすすめの商品を予測するアルゴリズム
- Flat-rate(定額)をご選択の方のみ利用可能



# AutoML Tables (ベータ版)



Google Cloud



# まとめ

- 分析の内製化、スモールスタートをするので  
あれば、運用/管理が楽なサービスがおすすめ
- 非エンジニアの方でも BigQuery を通じて分析をすることが可能
- 機械学習を始めるのは難しくない

**Thank you**