

BigQuery 最新アップデート速報！

Google Cloud Next(米国開催)
発表内容から見る
データ分析の未来



Tokyo

Proprietary



山田 雄

Yu Yamada

Customer Engineer
Data Analytics Specialist



01 Why BigQuery

02 主要アップデート

AI関連

DWH周辺

データパイプライン

03 BigQuery と共に創るデータの未来



01. Why BigQuery

226

2024 年における BigQuery の Update 数

226

2025 年 8 月現在

179

“サイロ化されたインサイト” から “統合インテリジェンス” へ

データ ウェアハウス

80 年代後半 – 2000 年初期

- 構造化データのみ
- 高い所有コスト
- スケール制限

決まった分析, 限られたデータ

何が起こったかの報告や事前定義された問いへの回答に最適

データレイクと データ ウェアハウス

2000 年初期 – 2010 年中期

- 全てのデータタイプ
- データサイエンスの台頭
- スケーラブル & コスト効率

データ種類に対応,
サイロ化が進む

データサイエンス、探索的分析、ML
モデル

レイクハウス

2010 年中期 – 2020 年初期

- 統合データプラットフォーム
- オープンデータフォーマット
- 統合ガバナンス & セキュリティ

レイクとウェアハウスの
いいところどり、オープン

構造化データと非構造化データを組み合わせて、何が起こったのかという問いに答え、ML/AI モデルを構築

自律的

Data & AI platform

2020 年 – 現在

- AI ネイティブ
- マルチモーダル
- Agentic

統合, エージェント、柔軟性

マルチモーダル データのリアルタイムインテリジェンスと、すべてのデータおよびビジネス プロフェッショナル向けのエージェント型 AI エクスペリエンス

BigQuery

データの取り込みから AI による分析まで、データのライフサイクル全体を支援して、手作業の労力を最小限に抑える

BigQuery と Vertex AI のシームレスな統合により、ML/AI モデルのライフサイクルを迅速に加速し、多様なチームをエンパワー

自律的な data-to-AI platform.

複雑なタスクを自動化し、問題を積極的に特定して解決し、専門のアナリスト、エンジニア、保護者として機能する AI エージェント

マルチモーダル データ、AI 駆動型分析、セマンティック インテリジェンス、LLM コプロセッシング、多様なエンジン、会話型 BI、リアルタイム機能を統合した統合プラットフォーム

BigQuery イノベーションのテーマ



AI によるイノベーション

会話型のインテリジェント検索と新しいエージェントエクスペリエンスを備え、セマンティックレイヤーで精度を向上。

エージェント



統合された Data & AI プラットフォーム

シームレスな分析、AI 共同処理、マルチモーダルデータに関するリアルタイムの洞察を実現

統合



柔軟かつ将来性

ローコストの AI と 3rd party およびオープンソースとのシームレスな相互運用性

柔軟性

The BigQuery Advantage



進化する “BigQuery”

Open Lakehouse + GenAI + Governance

データ分析



BigQuery Studio
統合ユーザインターフェース

Query

Notebook

Pipeline

Data Canvas



Looker / Looker Studio
可視化・自由分析

AI/ML



BigQuery ML
機械学習

データ ガバナンス

データ処理エンジン



BigQuery SQL
アドホック分析



Spark in BigQuery
BigQuery + Spark 統合



Apache Spark
Lightning Engine



Open source engine
spark trino TensorFlow

メタデータ・カタログ



Dataplex Universal Catalog
統合カタログ(ビジネス・テクニカル・ランタイム)



BigLake Metastore
統合ランタイムメタストア



BigMetadata(CMETA)
メタデータ高速分散処理

ストレージ



BigLake

DWH + データレイク 統合ストレージエンジン (抽象化レイヤー)

ICEBERG



BigQuery Managed Storage
構造化データ
半構造化データ



Google Cloud Storage
非構造化データ
(画像・動画・音声 etc)



Management

Data & Model Lineage

Data Security & Access

Data Quality

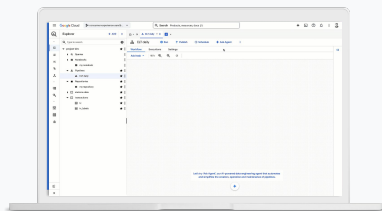
Data Discovery

02. 主要アップデート ~ AI 関連 ~

あらゆるユーザーのための 特化型エージェント

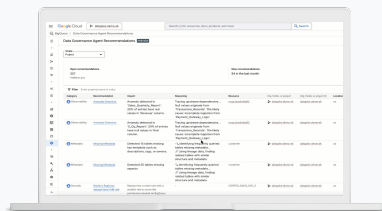
Specialized Agents for Every User

データ エージェント ファミリー



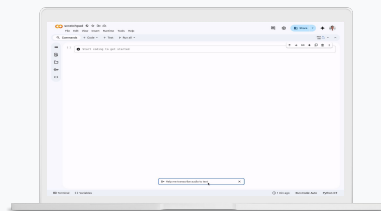
Data Engineering Agent

データ準備、データ取り込み、データクリーニング、ブロンズ→シルバー→ゴールドの自動化



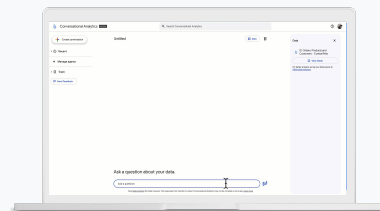
Data Governance Agent

データモデリング、メトリック定義、メタデータ生成



Data Science Agent

Colab ノートブック エージェント、探索的データ分析 (EDA)、データ可視化、予測モデリング



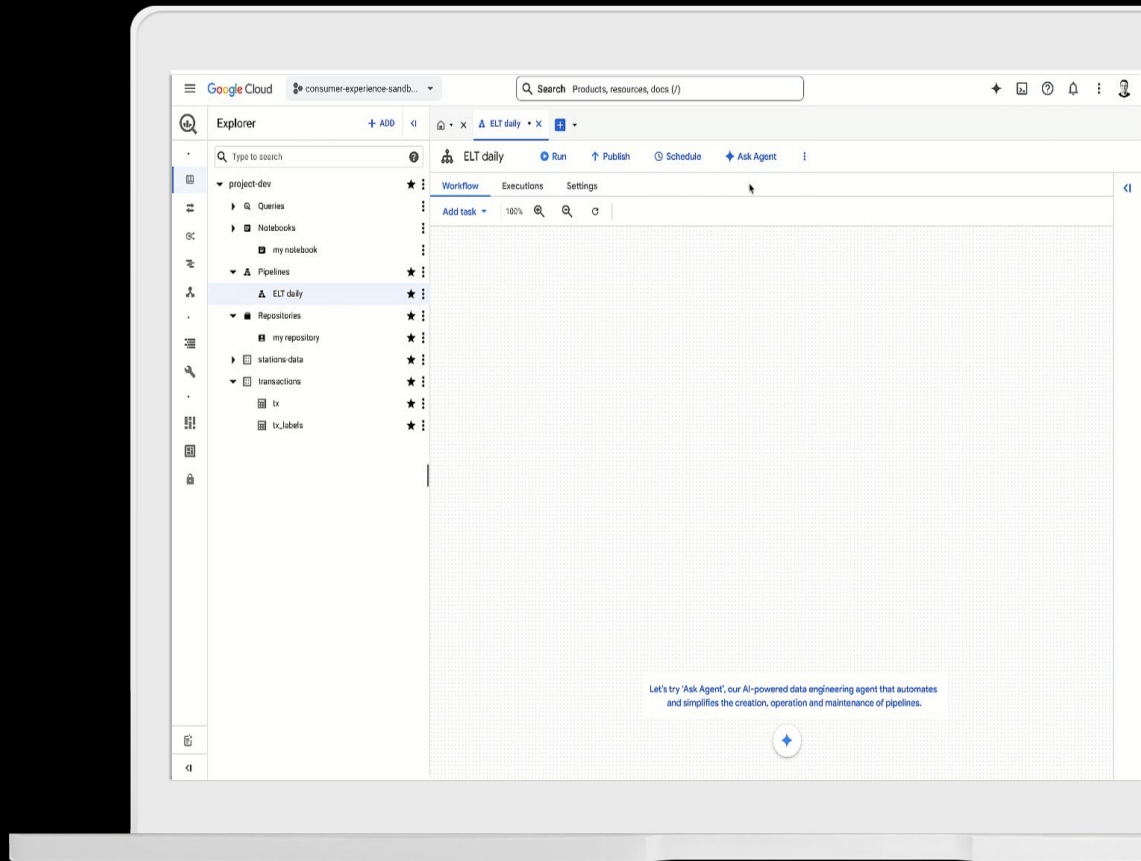
Conversational Analytics Agent

Data QnA, フォーキャスト, マルチモーダル検索

Data Engineering Agent

自動化されたデータ パイプラインの
作成、キューレーション、最適化。
BigQuery とのシームレスな連携

- データ パイプラインを大幅に加速
- 全てのデータ関連ステークホルダーを支援
- データ品質とガバナンスを保証
- データ エンジニアリング コストを削減
- ビジネス インサイトを即座に解き放つ

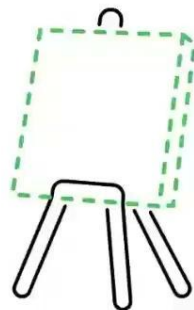




>| 無題の... ライン



無題のパイプライン



BigQuery Pipelines へようこそ

SQL と Notebooks を使用してデータ パイプラインを作成し、オーケストレーションを行います。

☒ demo@torusasaki.altostrat.com のユーザー認証情報で実行

[プレビュー](#)

☐ 選択したサービス アカウントで実行

[開始](#)
[詳細](#)

ジョブ履歴

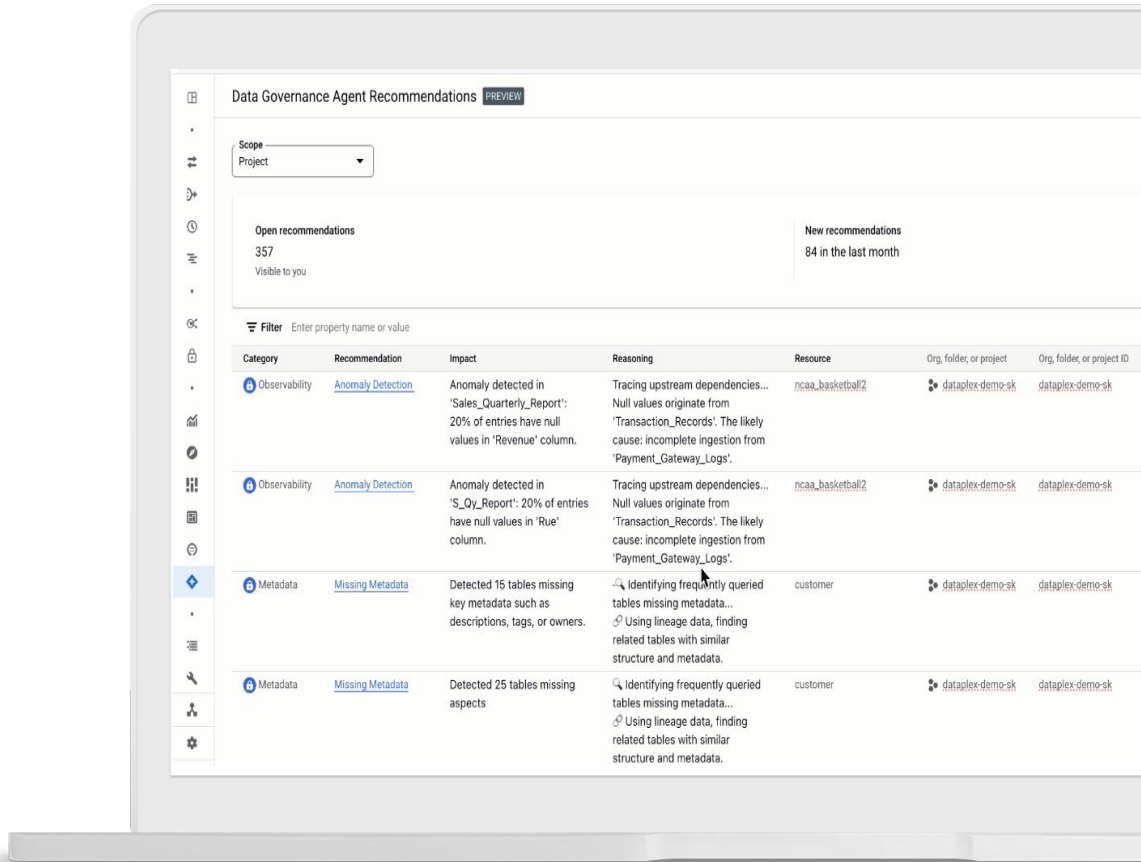
更新



Data Governance Agent

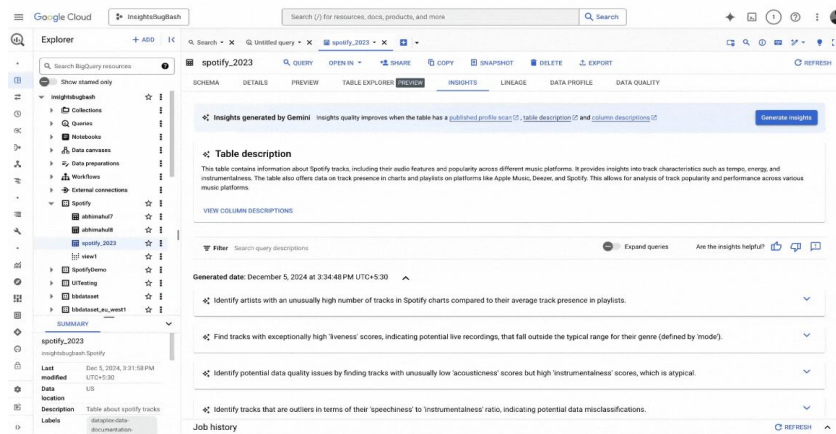
データ ガバナンス タスクを自動化

- **Proactive Governance:** コンプライアンス上の不備が発生するのを待つのではなく、システムがデータ分類、アクセス制御、データリネージ追跡、ポリシー更新を積極的に提案
- **Data enrichment:** テーブルや列のメタデータを自動的に生成し、拡充
- **Semantic grounding for agents:** エージェントにとって、統制の取れた信頼できる唯一の情報源 (SoT: Source of Truth)
- **Knowledge Graph-Powered Insights:** ナレッジグラフを利用してデータセット、メタデータ、ポリシー間の関連性を解明(マッピング)し、リアルタイムでガバナンスに関する推奨事項を提供



Automated metadata curation

- 自動メタデータ キュレーションは、明確なドキュメントの欠如や、エラーが発生しやすい手動メタデータ作成に起因する課題に対処します。特にデータセットの規模と複雑さが増すにつれて、こうした課題が顕在化します
- BigQuery データ インサイト スキャンは、**テーブルと列の説明**、そしてクエリの推奨事項を自動生成します
- ユーザーは必要に応じて説明を確認、編集し、スキーマタブと詳細タブにコピーできます



自動的なデータ異常探知

データの信頼性を向上

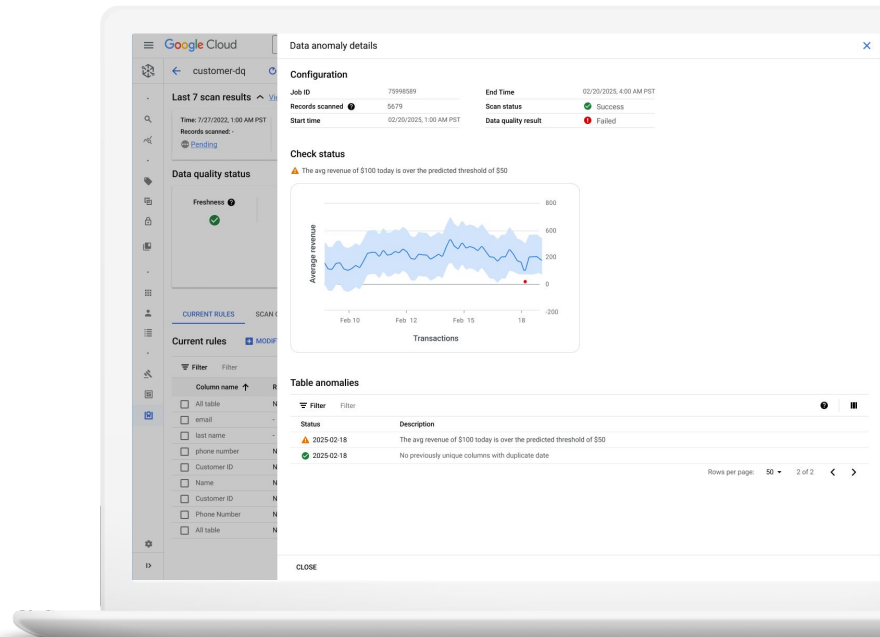
BigQuery の Data Quality 機能で、
BigQuery テーブルとビューのデータの問題を監視して
自動的に検出する自動データ異常検出のサポートを開始

能動的なリスク対処が可能に

異常を特定し、コストのかかるミス、不正行為、
コンプライアンス違反がビジネスに影響する前に防止

AI の力を解き放つ

高度な機械学習アルゴリズムを活用して、
従来のルールベースのシステムでは見逃される
可能性のあるパターンや異常を自動的に検出



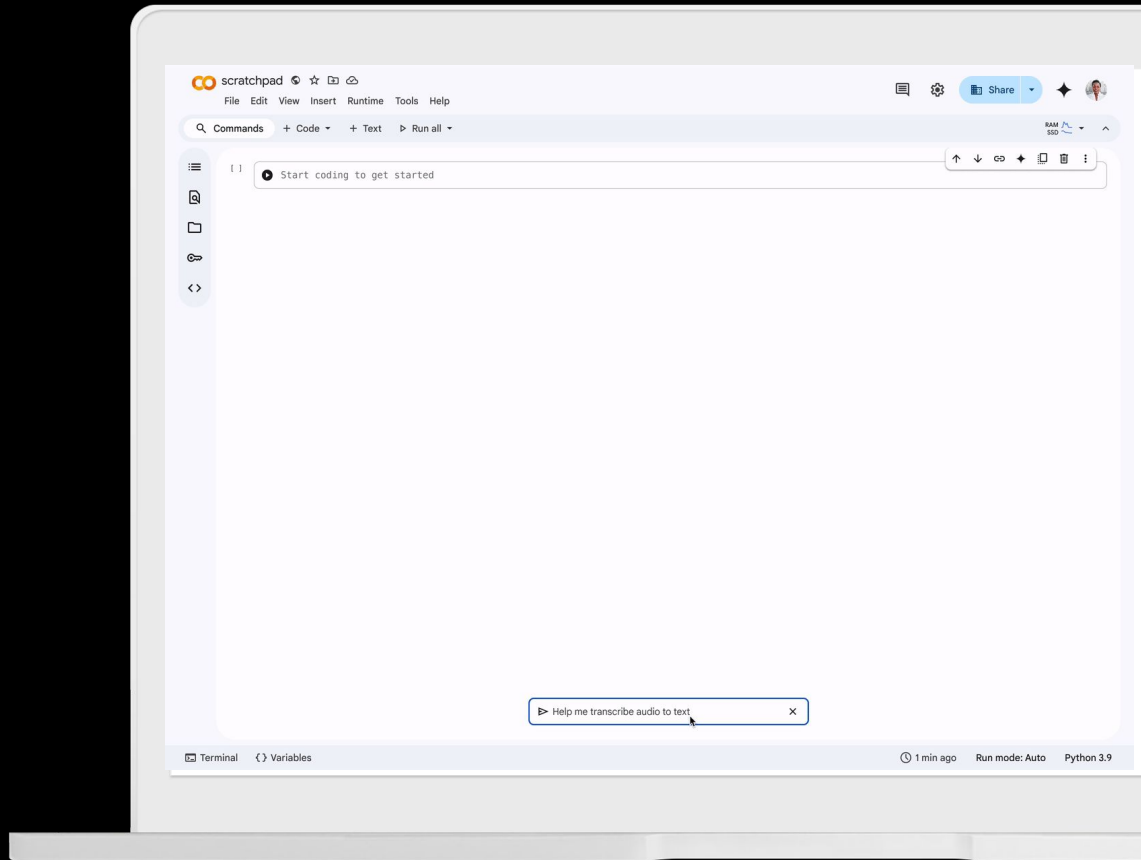
Data Science Agent

データサイエンスのタスクを自動化する 1st Party エージェント

以下の作業を自動化：

- **データ ラングリング**：データフレームの自動取り込みとクリーニング
- **データ探索**：ネイティブの視覚化と自動探索分析
- **AIによる予測**：Google Research の TimesFM などの最先端のモデルを含む AI を活用した予測
- **モデル評価**：精度と生産準備状況をより深く理解するためのモデル評価

Callout Agent session



プレビュー

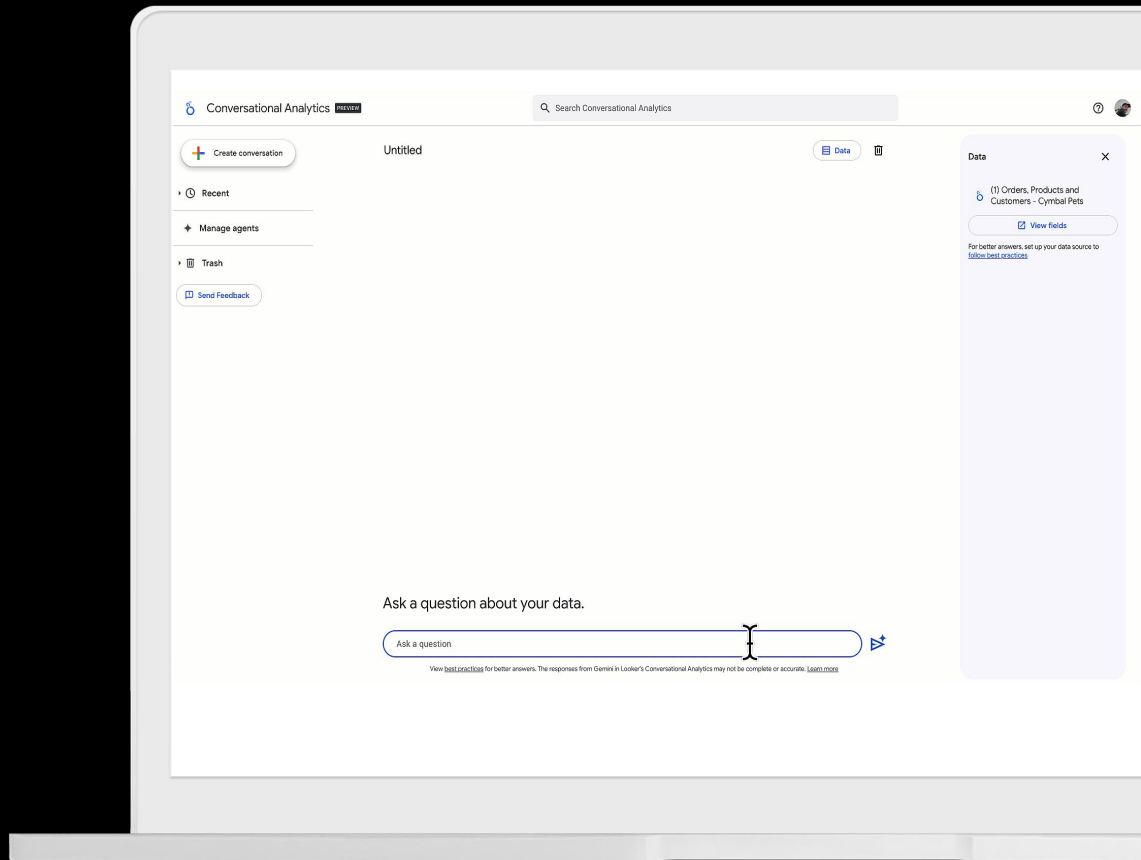
Conversational Analytics Agent

データ分析タスクを自動化する
1st Party エージェント

以下の作業を自動化：

- 分析の民主化：自然言語ベースのインサイト
- **Custom data agents:** 迅速な導入を可能にするクイック スタートを備えた柔軟な API
- **Direct connect to BigQuery:** BigQuery でデータを直接探索するか、Looker のセマンティック レイヤーと組み合わせる
- **Python の Code interpreter:** 複雑なデータ サイエンス タスク用の Python コードを生成

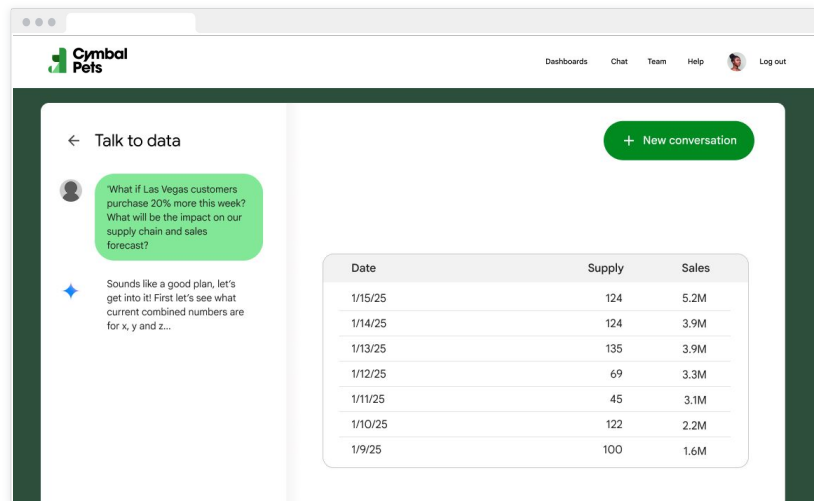
ビジネスレイヤーの人にもデータ分析を





Conversational Analytics API

エンタープライズ横断で Data Agent をカスタマイズ & デプロイ



BigQuery



Looker Model



カスタム エージェントを迅速に開発

クイックスタートを備えた
柔軟な API で迅速な導入を実現



分析目的に特化した API

ビジネス インテリジェンス、データ探索、
分析向けにチューニングされたAI サービス



BigQuery へ直接接続可能

BigQuery でデータを直接探索するか、
Looker のセマンティックレイヤと組み合わせる



**再利用可能でスケーラブルな
セマンティック レイヤー**

精度の向上と高度なマルチテーブル分析

動画あり
アーカイブ動画をご視聴ください

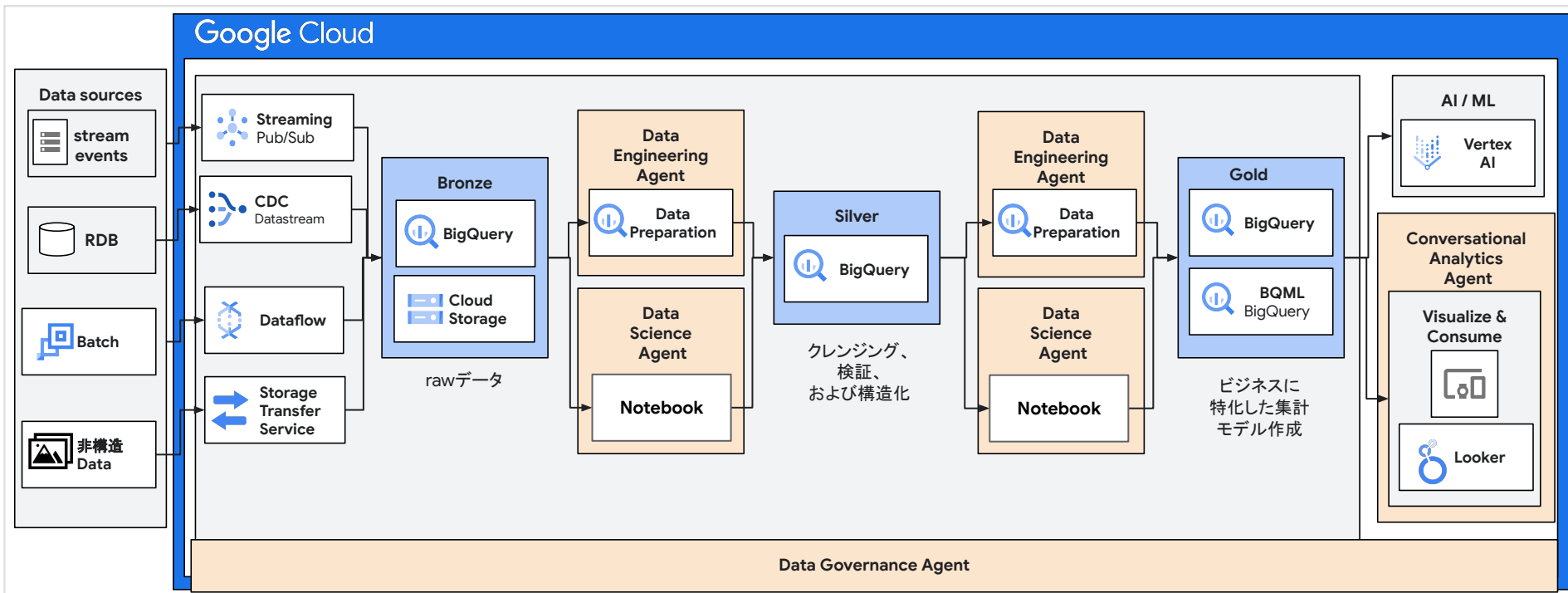
BigQuery Knowledge Engine

- BigQuery Knowledge Engine は、**Knowledge Graph** の力を活用して、高度なデータインサイトとデータ連携を実現します。
- 多様なデータセットとメタデータを統合することで、ユーザーはテーブル間の**隠れた関係性**を発見し、テーブル間クエリを生成し、より深い分析インサイトを得ることができます

The screenshot displays the Google Cloud BigQuery Knowledge Engine interface. The top navigation bar includes the Google Cloud logo and a search bar. The main content area is divided into several sections:

- Knowledge graph generated by Gemini:** This section provides a dataset description and a knowledge graph visualization. The graph shows relationships between entities such as 'Payment', 'customer', 'product', 'Inventory', 'Location', and 'legal_entity'.
- Relationships:** This section displays a table of selected relationships. The table has columns for 'Table 1', 'Table 2', and 'Relationship'. The selected relationship is 'Customers' (Table 1) and 'CustomerProfiles' (Table 2), with the relationship 'Customers.customerID = CustomerProfiles.customerID'. The table also includes a 'Write a query' button.
- Insights:** This section displays generated insights based on the knowledge graph. The insights include: 'What are the top 5 most popular program genres and how many times were they accessed?', 'Find the total number of escalated cases for each product', and 'What are the top 5 most popular program genres and how many times were they accessed?'.

メダリオン アーキテクチャ with Data Agents



Gemini in BigQuery - AI for everyone

Gemini の主要機能が無料で利用可能に

- SQL コード生成 / 説明 / 補完
- Python コード生成 / コード補完
- Data canvas
- Data preparation

channel	impressions	clicks	bookings	campaign_date	local_temp	hotel_name	phone
string	integer	integer	integer	string	string	string	string
social_media	560000	12500	485	15-07-2013	82°F	The Liberty Inn	(212) 555-1212
email	120000	8700	230	22-12-2023	28°F	Outback Oasis Resort	+61 2 9676 5432
social_media	560000	12500	485	15-07-2023	82°F	Redwood Lodge and Cabins	1-707-555-9800
email	120000	8700	230	22-12-2023	28°F	Wallaby Creek B&B	02 6123 4567
display_ads	340500	4200	375	09-08-2023	26°C		
partner_site	85000	9500	350	28-06-2023	31°C	Ryokan Sakura	03-5555-2222
search_ads	410000	18500	620	10-11-2023	-5°C	Maple Leaf Lodge	(416) 555-3210
social_media	280000	11300	315	14-02-2023	9°C	The Crown and Sceptre	020 7948-0055
display_ads	610000	8600	540	28-07-2023	33°C	Hacienda del Sol	+52 (624) 143-5555
email	220000	6300	290	25-09-2023	25°C	Pousada da Floresta	(11) 4002-8922
partner_site	170000	14100	410	27-12-2023	30°C		
email	92000	7200	340	18-03-2023	12°C	Kyoto Tea House Inn	(975) 351-0345
search_ads	350000	22700	875	20-08-2023	84.2°F	The Jazz Corner Hotel	+1-504-555-9080
social_media	420000	6400	280	05-07-2023	30°C	Villa Toscana	+39 055 555566
display_ads	375000	5800	210	01-12-2023	4°C	Gasthaus zum Adler	(089) 22 33 44
partner_site	670000	32000	940	12-02-2023	28°C	The Taj View Hotel	91-562-222-1234
search_ads	450000	16400	620	15-08-2023	32°C		
email	140000	6700	185	28-12-2023	22°C	Kiwi Country Lodge	+64 3 543 9876
social_media	380000	20100	790	25-08-2023	24°C	Hockey Haven Hotel	1-404-HOCKEY
display_ads	385000	5300	190	20-07-2023	35°C	Santorini Sunset Villas	30-22860-22548
partner_site	110000	10200	360	05-08-2023	68°F	Crossroads Motel	505-555-1234
search_ads	290000	13700	430	20-06-2023	10°C	Finca Los Gauchos	+54-11-5555-7777
social_media	420000	16100	515	15-02-2023	25°C	Playa del Carmen Paradise	(984) 876-5432

高度なコンテキストとプロアクティブな管理ツールを柔軟な価格帯から選択

SQL の生成・補完 / 色々な呼び出し方



① コーディング サポートツールで生成

```
1 # weather.data を使い、東京の一日ごとの平均気温を計算して
2 SELECT
    DATE(datetime) AS date,
    AVG(temperature) AS average_temperature
FROM
    `event24-438905.weather.data`
WHERE city = 'tokyo'
GROUP BY 1;
```

② SQL エディタのコメントで生成

```
SELECT
  city,
  DENSE_RANK() OVER (PARTITION BY city ORDER BY temperature DESC) AS temperature_rank,
  DENSE_RANK() AS weather_rank
FROM
  `event24-438905.weather.data`
FROM `event24-438905.weather.data`
```

③ SQL エディタで入力中に自動補完 (灰色部分が補完候補)

Preview

SQL の翻訳 Interactive SQL Translator

Proprietary

- 異なる DB の SQL から GoogleSQL への翻訳を支援
- イメージとしては Google 翻訳の SQL 版

The screenshot displays the Interactive SQL Translator interface. On the left, the '変換元: HiveQL' (Source: HiveQL) tab is active, showing the following SQL code:

```
1 SELECT
2   user_id,
3   COLLECT_SET(item) AS unique_items
4 FROM user_table
5 GROUP BY user_id
6 ;
```

On the right, the 'GoogleSQL' tab is active, showing the translated SQL code:

```
1 SELECT
2   user_table.user_id,
3   array_agg(DISTINCT user_table.item) AS unique_items
4 FROM
5   user_table
6 GROUP BY 1
7 ;
8
```

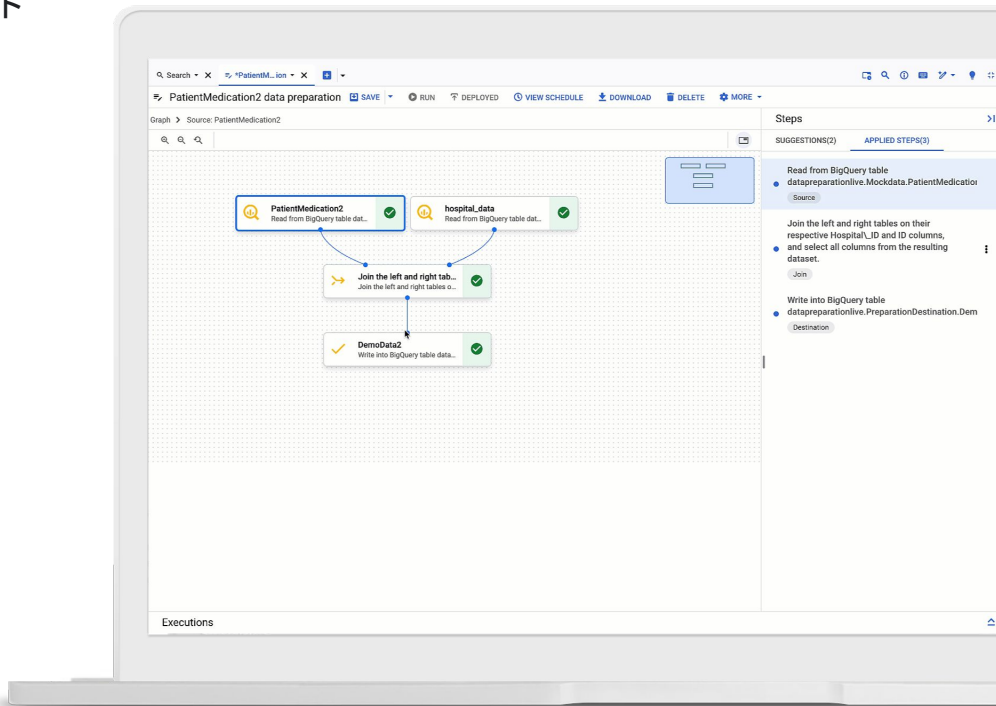
The interface includes various controls at the top, such as '変換' (Convert), '展開' (Expand), '実行' (Execute), 'カスタマイズ' (Customize), and '保存' (Save).

HiveQL から GoogleSQL への翻訳

BigQuery Data Preparation (データの準備)

BigQuery Studio 内のデータプリパレーションエージェント

- Gemini の支援により
SQL データパイプラインを数分で開発
- コンテキストを考慮した
AI 生成の変換機能により、データの検出、
クレンジング、変換を実行可能に
- スキーマ マッチング、
ビジネスとデータ品質のアサーション生成を
自動化することで、時間を節約
- GitHub との連携により、BigQuery で
デプロイとオーケストレーションを実行可能



Data Preparation / 注文データをクレンジング

orders data preparation SAVE スケジュール ダウンロード 削除 その他

Graph > ソース: orders

データ スキーマ

	purchased_date STRING	prefecture STRING
1	09/30/2024	愛知県-名古屋市
2	01/10/2024	愛知県-名古屋市
3	01/10/2024	滋賀県-大津市
4	09/30/2024	栃木県-宇都宮市
5	2024_09_30	山梨県-甲府市
6	2024_10_01	神奈川県-横浜市
7	2024_09_30	三重県-津市
8	09/30/2024	埼玉県-さいたま市
9	2024_09_30	東京都-東京
10	2024_09_30	茨城県-水戸市
11	2024_10_01	滋賀県-大津市
12	2024_10_01	愛知県-名古屋市

ページあたりの行数: 50 ▼ 1 - 28 / 28 < > >>

ステップ

提案 (10) 適用されたステップ (1)

✧ Gemini による提案

AI による分析情報を活用してデータを準備します。

変換

Converts column purchased_date from type string to date with the following format(s):
"%m/%d/%Y";"%Y_%m_%d"

```
COALESCE(SAFE.PARSE_DATE("%m/%d/%Y",
purchased_date),SAFE.PARSE_DATE("%Y_%m_%d",
purchased_date))
```

この提案を評価してください

プレビュー 適用

+ ステップを追加

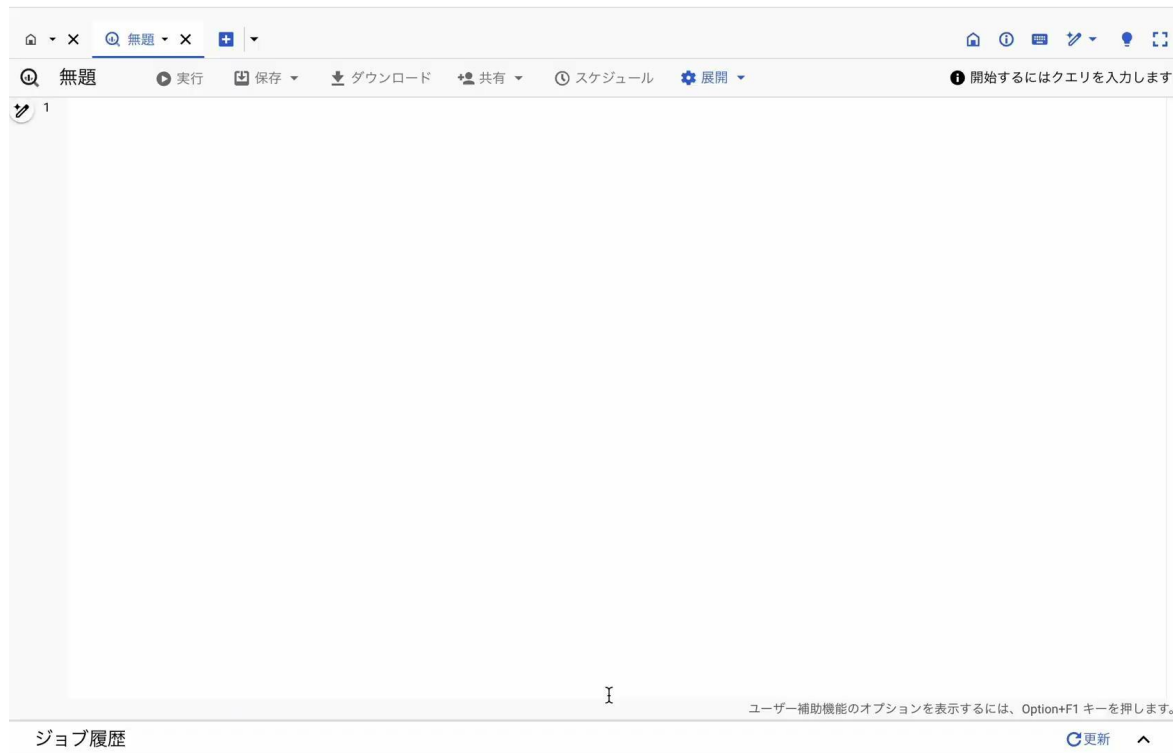
Data Canvas

自然言語も組み合わせながら
視覚的にデータを探索できる

Gemini は

- データの検索
- SQL の生成
- チャートの生成

を支援する



BigQuery AI Query Engine

LLM の力をデータ プラットフォームに組み込み

セマンティックな洞察を抽出するのは 困難

何百万もの製品レビューを考慮すると:

- 最も頻繁に表面化する、繰り返し発生する問題は?
- どの代替製品が推奨されるか?
- どのレビューが安全性を懸念しているか?



```
SELECT *  
FROM reviews JOIN products  
ON AI.IF(f"{review} recommends  
{product} as a better  
alternative")
```

AI Query Engine が作業を 簡素化

- 使い慣れた SQL と自然言語プロンプト
- アナリストが簡単に使用でき、AIの専門知識は不要
- BigQuery DataFrames の Pandas ライクな Python インタフェース
- **SQLで実現が難しい分析に**

02. 主要アップデート ~ DWH周辺 ~

BigQuery - 新しい Workload Management

Flexibly assign reservations

ユーザーが Runtime Reservation を選択可能に

/ 特徴

- デフォルト予約を上書きして選択可能に
- 選択するには IAM ロールの付与が必要
- IAM Condition で特定の予約に制限可能

```
7 | name, gender
8 | ORDER BY
9 | total DESC
10 | LIMIT
11 | 16;
```

✓ This query will process 99.95 MB when run.

Processing location: US X

Reservation override: projects/argo-tsutsum-bq-admin/locations/US/reservations/user-us-s-1 X

Location

☐ Automatic location selection ⓘ

Location type ⓘ

☐ Region

Specify a region to colocate your datasets with other Google Cloud services.

☒ Multi-region

Allow BigQuery to select a region within a group to achieve higher quota limits.

Multi-region *

US (multiple regions in United States) ▼

Reservation

Filter Type to filter

Default project assignment

On-Demand

us-1

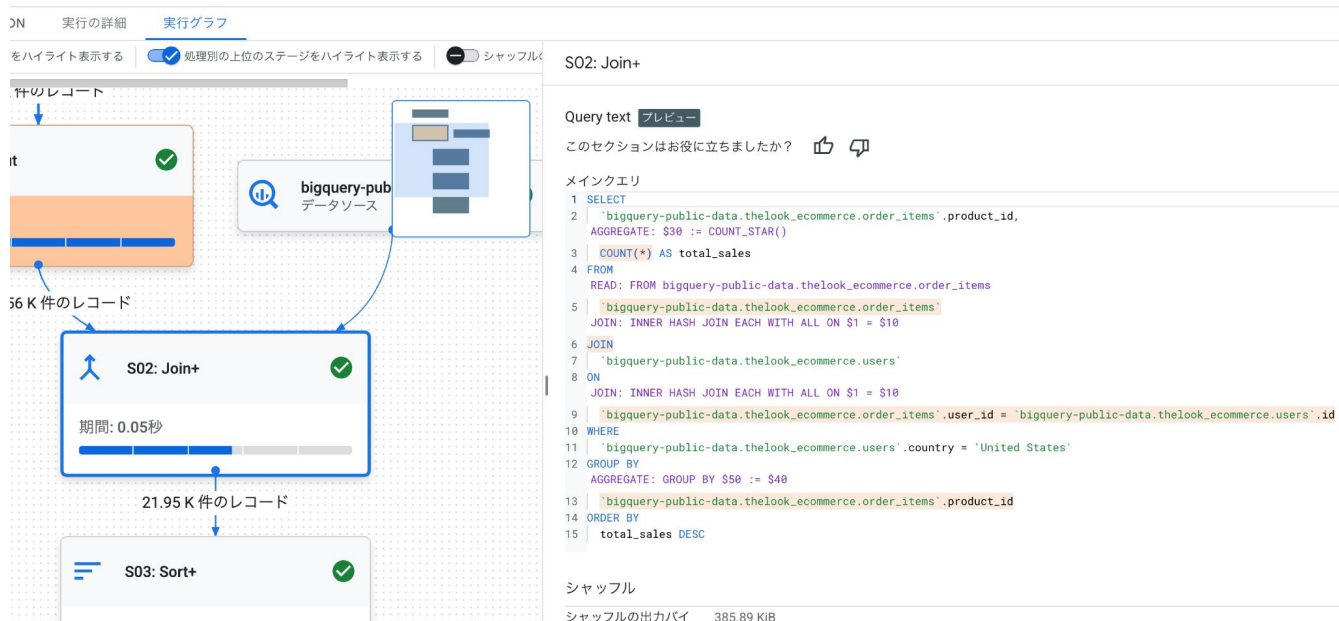
projects/argo-tsutsum-bq-admin/locations/US/reservations/us-1

user-us-s-1

projects/argo-tsutsum-bq-admin/locations/US/reservations/user-us-s-1

BigQuery クエリ実行グラフと SQL 文の対応可視化が可能に

- 実行グラフでノードをクリックすると、このノードが SQL 文のどの箇所に対応するのかハイライト表示される (クエリ実行プランも紫色で、SQL 文の対応箇所に補われる)
- 実行プランの読み解きがより簡単に : 高度なパフォーマンス チューニングが容易に



GROUP BY 句のいくつかの拡張が GA

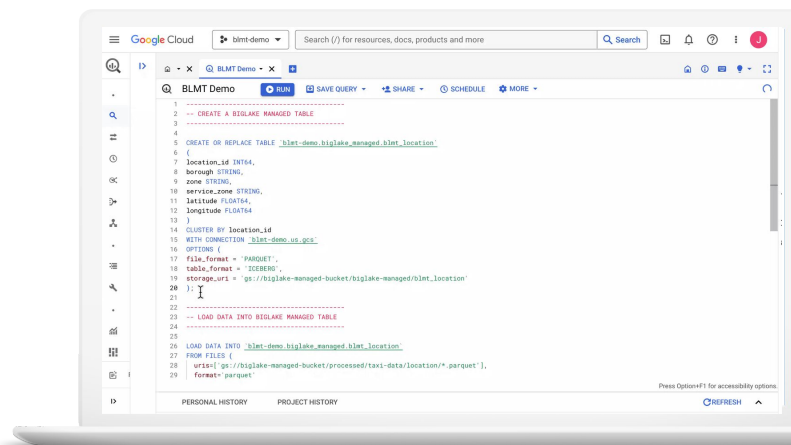
- GROUP BY STRUCT
- GROUP BY ARRAY
- GROUP BY ALL

```
SELECT
  year,
  homeTeamName,
  SUM(attendance)
FROM
  `yu-bq-demo.yu_test.baseball_schedules`
GROUP BY ALL;
```


BigLake tables for Apache Iceberg in BigQuery GA

旧 BigQuery tables for Apache Iceberg がリネームして BigLake tables for Apache Iceberg in BigQuery という名前で GA しました

- ✓ セルフ マネージドとフルマネージド の Iceberg エクスペリエンス を選択可能
- ✓ 優れた費用対効果を実現 するために Iceberg データを自動最適化
- ✓ BigQuery と OSS エンジン全体で読み取り遅延なし で 1 秒あたり 10Gbps のスループットにスケール
- ✓ BigQuery ML Vertex AI 統合は、BQ ネイティブ テーブルと同様に Iceberg テーブル上で動作
- ✓ Iceberg テーブルに対するきめ細かいセキュリティとガバナンスが、複数のエンジン間で適用可能



BigLake metastore がリリース (GA)

旧 BigQuery metastore がリネームして BigLake metastore という名前で GA しました



単一の共有メタストア

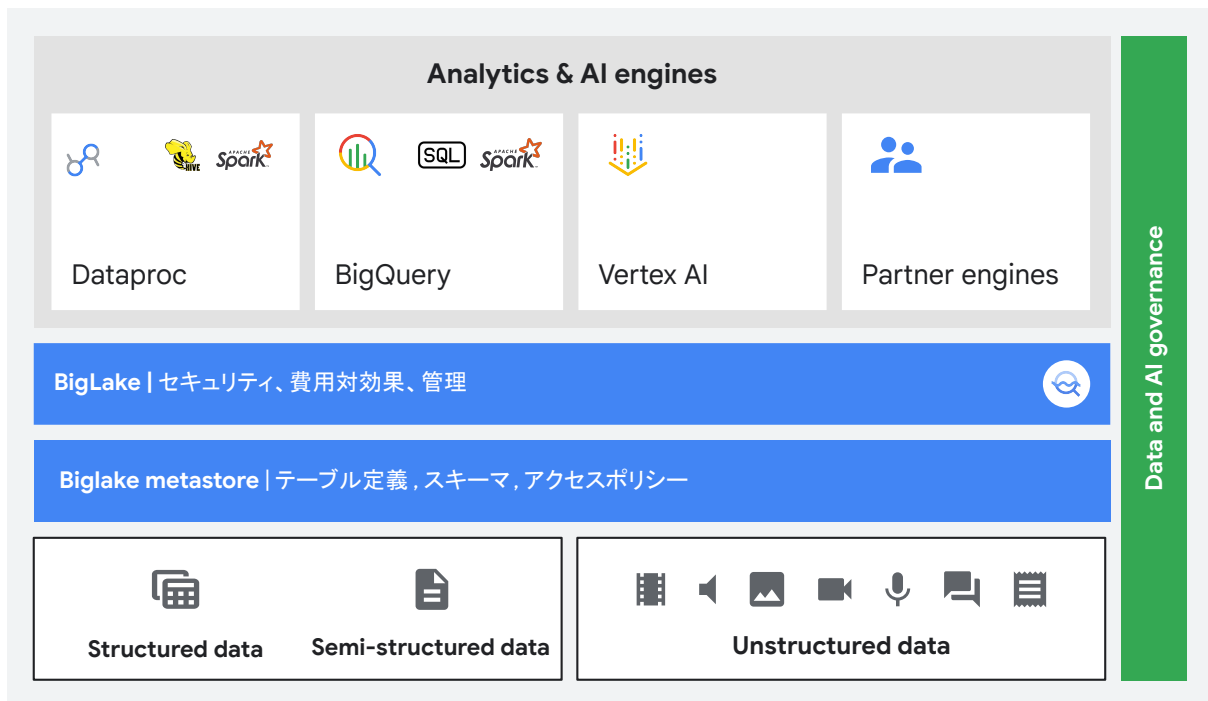
SQL、オープンソース エンジン、AI/ML のためのランタイム メタデータとコネクタ

構造化データと非構造化データのサポート

同じランタイム メタストアを使用して、構造化テーブルとオブジェクト テーブルの両方をクエリ可能

統合ガバナンス

すべてのメタデータが大規模なポリシー管理に利用可能



BigQuery アドバンスド ランタイムが使用可能に

アドバンスド ランタイムは、BigQuery クエリ プロセッサに拡張ベクトル化を提供します。拡張ベクトル化は、BigQuery ストレージでのフィルタ評価、専用のデータエンコード、最適化手法など、クエリ処理の重要な側面にベクトル化を適用することで、ベクトル化されたクエリの実行を拡張します。

/ 特徴

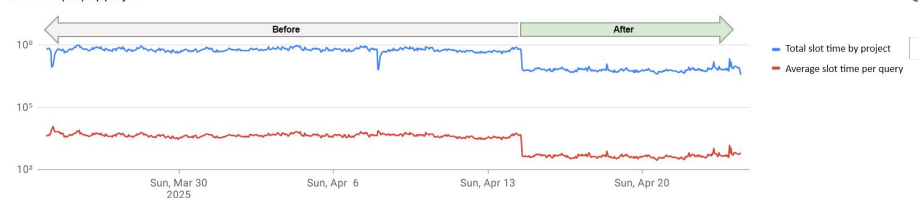
- スキーマやコードの変更なしで、クエリを最大200倍高速化
- スロットタイムを全体で最大40%削減
- ショートクエリのスロットタイムを最大10倍削減
- 完全自動化により、他のクエリに影響を与えることなく、対象となるクエリを安全に高速化
- ベクトル化実行とショートクエリ実行の大幅な機能強化
- 既存のオブザーバビリティダッシュボードまたは INFORMATION_SCHEMA.JOBSを使用して、パフォーマンスへの影響をモニタリング

高度なランタイムを有効にする

プロジェクトで高度なランタイムを有効にするには、[ALTER PROJECT](#) ステートメントを使用してデフォルト構成を変更します。ステートメントで、`query_runtime` 引数を 'advanced' に設定します。次に例を示します。

```
ALTER PROJECT PROJECT_NAME
SET OPTIONS (
  'region-LOCATION', query_runtime = 'advanced'
);
```

Slot time (ms) by project



設定画面からデフォルト設定を設定可能に

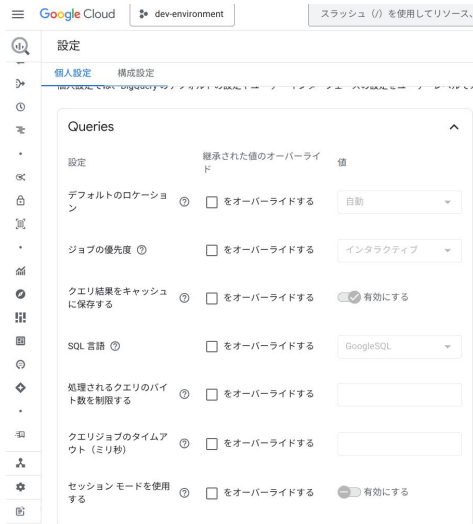
設定画面から個人/構成設定を設定することが出来るようになった

/ 特徴

以下の 2 種類の設定が可能

- 個人設定: BigQuery Studio でセッションを開始したときに適用されるデフォルト設定
一部の設定はプロジェクトまたは組織から継承されますが、「個人設定」ページで上書きできる
- 構成設定: 管理者は、選択したプロジェクトまたは組織内のユーザーに対して BigQuery Studio をカスタマイズできます。
これは、「結果の保存 > CSV (Google ドライブ)」、「開く > Looker Studio」、「エクスポート > スプレッドシートで探索」などのユーザー インターフェース要素の表示 / 非表示を切り替えられます

※これらの設定は、BigQuery Studio で非表示になっている場合でも、基盤となるデータやツールへのアクセスを制限するものではありません。組織レベルで変更を行うにはbigquery.config.get 権限が必要です。



BigQuery Tables の CMETA が GA に

BigQuery tables と external tables の両方でメタデータ インデックスを作成することが可能に

/ 特徴

- INFORMATION SCHEMA を介して CMETA の状態を監視可能
- 1 GiB を超える BigQuery テーブルのメタデータを自動的にインデックスに登録
- インデックスの鮮度要件が厳しい場合は、CMETA 用のスロット予約(検索インデックスのメンテナンスなどの他のシステムタスクと共有可能)を作成可能 (Bring-Your-Own-Slots)とも呼ばれます。

BigQuery 継続的クエリ

継続的な SQL 処理により
新しいインテリジェントな
データパイプライン能力を解き放つ

SQL ベースの 継続的な分析処理

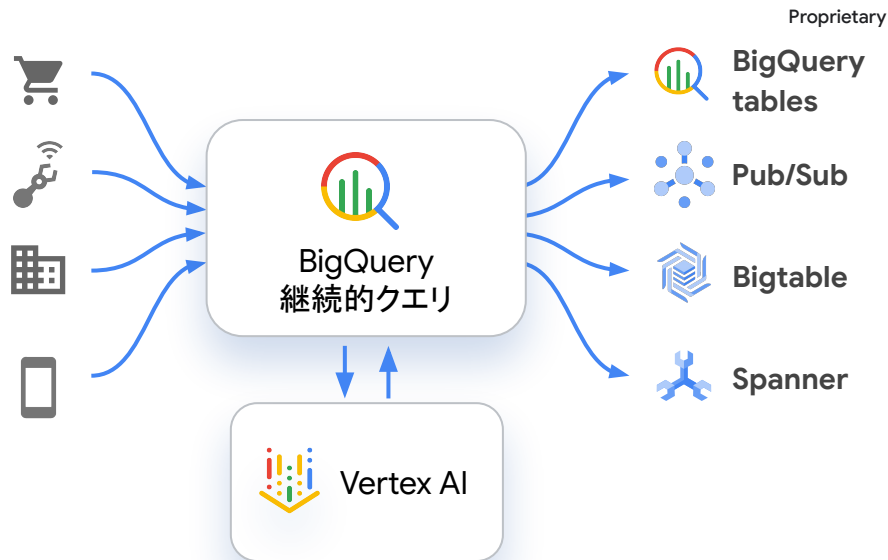
BigQuery スケールで、SQL を使用して受信データのストリームに対して無制限のサーバーレス分析を実行

リアルタイム アプリ開発への生成 AI

リアルタイムの感情分析、推奨事項、異常検出などの AI + ML 機能を備えた自動化パイプラインを開発

オペレーショナル システム への リバース ETL 連携

BigQuery のデータをリアルタイムでプログラマ的に変換し、Pub/Sub、Bigtable、Spanner、別の BigQuery テーブル、または幅広い検証済みのパートナー統合に複製



BigQuery - Pipe syntax

- パイプ構文のサポートによって、データ操作を任意の順番で記述できるように
- 複雑な SQL を直感的に分かりやすく記述できる
- SQL のデバッグも可能

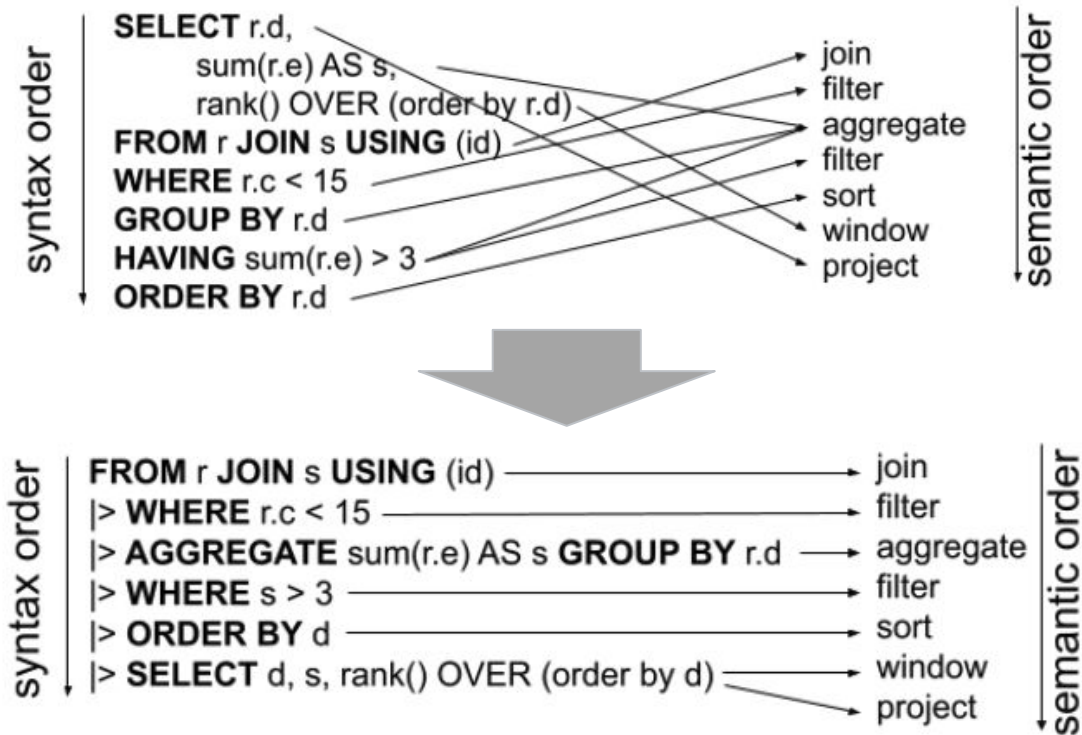
従来の構文

```
SELECT state, item_type, AVG(total_cost) AS average
FROM
(
  SELECT
    SUM(cost) AS total_cost,
    customer_id,
    state,
    item_type,
    COUNT(*) OVER (PARTITION BY customer_id) AS num_orders
  FROM mydataset.customer_orders
  GROUP BY customer_id, state, item_type
  QUALIFY num_orders > 1
)
GROUP BY state, item_type
ORDER BY state DESC, item_type ASC;
```

Pipe syntax

```
FROM mydataset.customer_orders
|> AGGREGATE SUM(cost) AS total_cost, GROUP BY customer_id, state, item_type
|> EXTEND COUNT(*) OVER (PARTITION BY customer_id) AS num_orders
|> WHERE num_orders > 1
|> AGGREGATE AVG(total_cost) AS average GROUP BY state DESC, item_type ASC;
```

```
/*-----+-----+-----+
| state | item_type | average |
+-----+-----+-----+
| WA    | clothing  | 35.0    |
| WA    | food      | 3.0      |
| NY    | clothing  | 16.0     |
| NY    | housewares| 22.0     |
+-----+-----+-----+*/
```



02. 主要アップデート ～ データ パイプライン ～

BigQuery - DTS が MySQL と PostgreSQL に対応

/ 特徴

- BigQuery Data Transfer Service の MySQL/PostgreSQL コネクタを使用すると、インスタンスから BigQuery にデータを取り込むことができます
- オンプレミス環境、Cloud SQL、および Amazon Web Services (AWS) や Microsoft Azure などの他のパブリック クラウド プロバイダーでホストされているインスタンスをサポートしています。
- また、BigQuery Data Transfer Service の コネクタを使用すると、データベースから BigQuery にデータを転送するためのオンデマンドおよび定期的なデータ転送ジョブを作成できます。

/ ユースケース

- MySQL もしくは PostgreSQL からデータを移行したい場合

BigQuery - SpannerへのExport

/ 特徴

- BigQuery のデータを SQL で Spanner へ Export 可能に
- Enterprise or Enterprise Plus のみ対応 (on-demandも非対応)

/ ユースケース

- Spanner への ReverseETL

/ 制限事項

- 1 行の上限は 1MiB
- 6 時間以内に終わるクエリ

```
EXPORT DATA OPTIONS
```

```
(
```

```
  uri
```

```
= "https://spanner.googleapis.com/projects/PROJECT_ID/instances/INST  
ANCE_ID/databases/DATABASE_ID",
```

```
  format = 'CLOUD_SPANNER',
```

```
  spanner_options = ""{ "table": "TABLE_NAME" }""
```

```
)
```

```
AS
```

```
SELECT * FROM mydataset.table1;
```

[Spanner external datasets](#)

BigQuery - External Dataset の Spanner サポート (GA)

/ 特徴

- Spanner のデータベースにリンクされた BigQuery の「外部データセット」を作成可能に
- デフォルトで Spanner の Data Boost が有効になる (変更不可)
- 外部データセットは読み取り専用で、編集不可
- 制限
 - Analytics Hub や Dataplex の Data Profile、Quality、マテビューは使用不可
 - 行・列セキュリティや列マスキングは使用不可
 - SQL auto completion 使用不可

```
CREATE EXTERNAL SCHEMA DATASET_NAME   
  OPTIONS (  
    external_source = 'SPANNER_EXTERNAL_SOURCE' ,  
    location = 'LOCATION' );
```

[Spanner external datasets](#)

BigQuery - Dataplex automatic discovery

BigQuery & Dataplex - Discover and catalog Cloud Storage data が GA になりました

特徴

- 指定した GCS のパスのデータをスキャンして、カタログの自動作成、Biglake テーブル、外部テーブル、オブジェクト テーブルを自動作成
- CSV, Parquet, Avro, ORC, JSON (newline-delimited) などの構造化・半構造化データ、および画像や動画などの非構造化データに対応
- 定期的なスキャンまたはオンデマンド スキャンが実行可能

ユースケース

- 煩雑に置かれているデータをスキャンするのではなく、ある程度データを整理した上でスキャンし、csv から biglake テーブルを作成することを自動化、あるいは画像データのメタデータを Object Table として登録することを自動化

The screenshot displays the BigQuery console interface for creating a new table from Cloud Storage data. The 'Source' section on the left lists various data sources, with 'Google Cloud Storage' highlighted by a red box. Below this, a list of other sources is shown. The 'Table configuration' section on the right allows users to specify the table name, ID, and scan frequency. The 'Scan frequency' section is set to 'On-demand'. The 'JSON or CSV' section is set to 'JSON'. The 'CSV options' section is set to 'CSV parser options are enabled'. The 'Header row' is set to '0'. The 'Delimiter' is set to 'comma'. The 'Encoding' is set to 'UTF-8'. The 'Run now' button is visible at the bottom right.

Pub/Sub メッセージを直接変換する SMT 機能が GA

メッセージの内容や属性を直接変更できる「シングルメッセージ変換 (SMT)」機能が一般提供

/ 特徴

- トピックへの公開時、またはサブスクリプションからの受信時に軽量な変換処理を挟むことができ、簡単なデータ加工のために Cloud Functions や Dataflow などの別サービスを用意する必要がなくなる
- 変換には JavaScript のユーザー定義関数 (UDF) を利用する

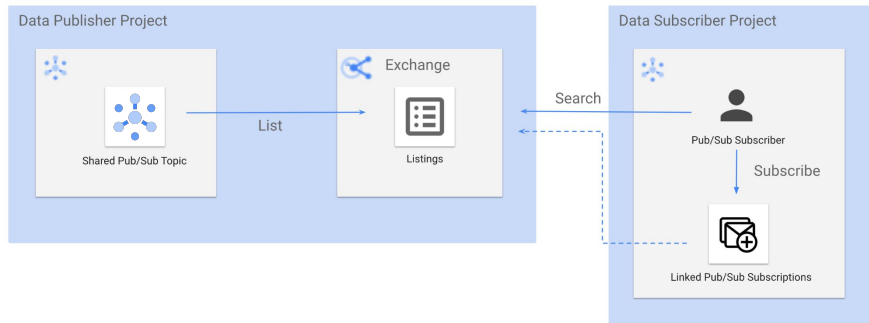
/ ユースケース

- データ マスキングと除去: クレジット カード番号や PII などのフィールドをマスキングまたは除去して機密データを保護し、データ プライバシー規制に準拠します。
- データ形式の変換: ダウンストリーム システムとの互換性を確保するために、異なる形式間でデータを変換します。
- メッセージのフィルタリング: コンテンツまたは属性に基づいて不要なメッセージを除外し、関連するメッセージのみを処理します。SMT を使用すると、Pub/Sub の組み込みフィルタよりも複雑なフィルタ条件を指定できます。
- シンプルなデータ変換: 文字列操作、日付のフォーマット設定、数学演算などの基本的なデータ操作タスクを実行します。

BigQuery - BigQuery sharing で Pub/Subトピックを共有

/ 特徴

- BigQuery sharing(Analytics Hub) を通じて Pub/Subトピックを共有し、社内外の組織にストリーミングデータのライブラリを配布。



/ ユースケース

- 金融**
 - 金融商品の価格・相場・注文情報をリアルタイムで共有
 - マネーロンダリングや決済詐欺の検知
- 小売**
 - 店舗の在庫管理
 - マーケティングとカスタマー サポートのパーソナライズ
 - ダイナミックプライシング
- ヘルスケア**
 - リアルタイムで患者のリスク分析
 - ウェアラブル デバイスによるバイタル モニタリング
- 通信**
 - ネットワークの健全性監視・障害予測

Google Maps Places データが BigQuery で利用可能に

BigQuery Data Clean Room で Google Map の豊富な POI データと自社データを統合

セキュアにデータを結合

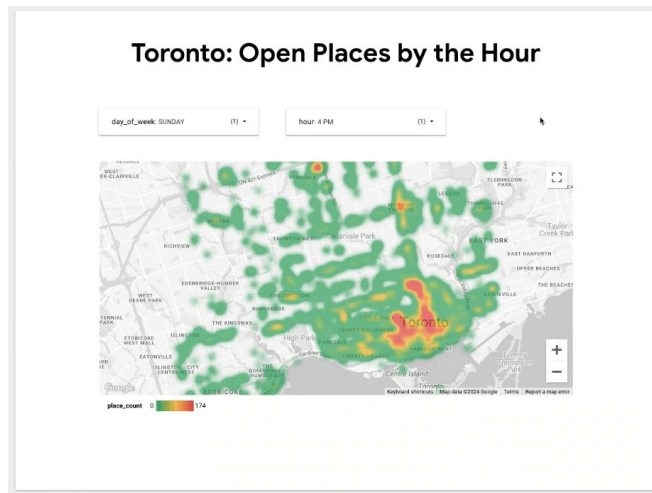
自社の独自データと Google Map の Advanced Places Insights データを統合。Data Clean Room により、データと分析のプライバシーが確保される

定期的な更新 & 包括的なインサイト

2 億 5000 万以上の企業や場所について、場所の種類、評価、営業時間などの主要な POI 属性をクエリすることで、最新かつ包括的なインサイトを獲得。ビジネスニーズに合わせた統計的なインサイト(例: 場所の密度)も発見可能

複雑な分析を可能にする使い慣れたツール

Places データを既存のワークフローに統合し、クラウドや BigQuery エコシステム内の他のツールと連携可能。BigQuery のスケールとパフォーマンスを活用して、あらゆる規模の分析を強化



An example using Advanced Places Insights to analyze what businesses are open by time of day and day of week and visualized in Looker Studio

03.BigQueryと共に創るデータの未来

Your ideal Data Warehouse.

今日もこれからも



PAST & PRESENT

サーバレス

ストレージとコンピューットの分離

データの民主化



Why BigQuery?

圧倒的なシンプルさ

拡張性

知性・自律性



FUTURE

生成AIとの完全な融合

Agentによるデータとの対話

真の民主化

組織の意思決定の中軸に

『データと対話する』
インテリジェント・プラットフォームへ