

freee が目指す 生成 AI 時代に向けた次世代 データ プラットフォームと ガバナンスとは



Tokyo

Proprietary



城谷 信一郎

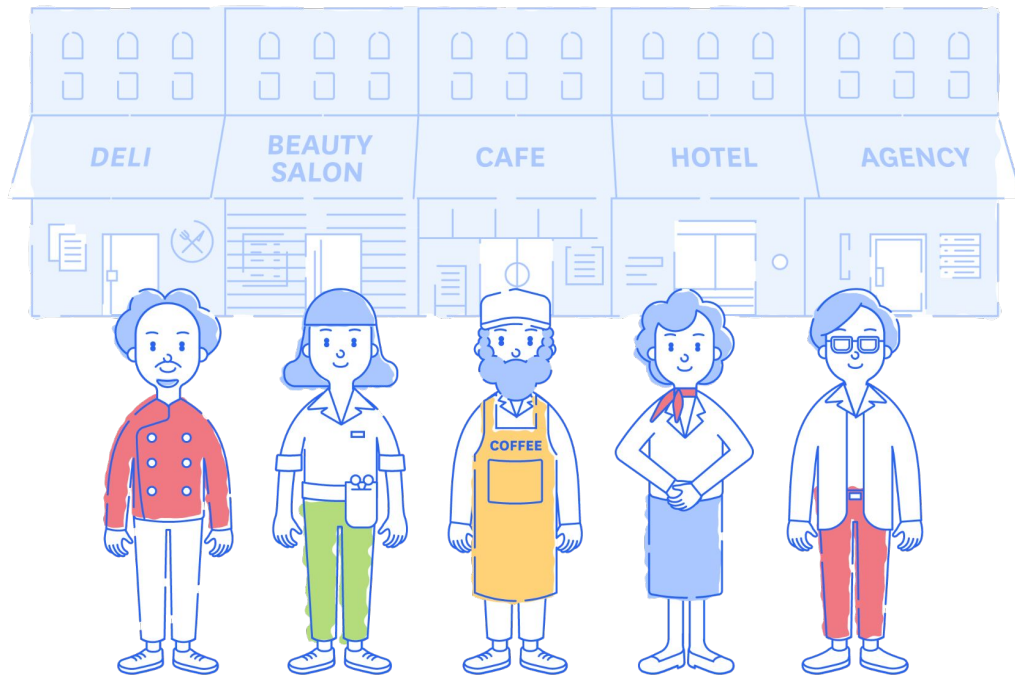
エンジニアリング基盤本部
データ部 部長



01. はじめに

私たちの事業について

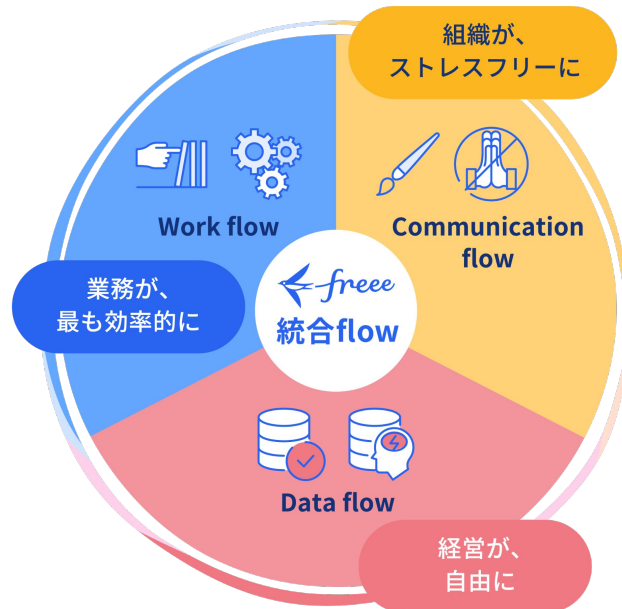
スモール ビジネスを、世界の主役に。



私たちの事業について



会計・人事労務・販売管理を核とした
統合型経営プラットフォーム



バックオフィス業務の分断をなくし、
統合体験の構築

私たちの事業について

統合型クラウド会計ソフト



2013 年 3 月～

日本のクラウド
会計ソフト市場
シェア No.1⁽²⁾

- 請求書
- 経費精算
- 決算書
- 予実管理
- ワークフロー
- 内部統制

統合型クラウド人事労務ソフト



2014 年 10 月～

スモールビジネスの
人事管理市場において
売上金額シェア No.1⁽³⁾

- 勤怠管理
- 入退社管理
- 給与計算
- 年末調整
- マイナンバー
管理

統合型クラウド販売管理ソフト



2022 年 11 月～

【国内初】
クラウド会計ソフトと
一体型で使える
販売管理サービス

その他サービス

● 支出管理



● 人事労務



● 工数管理・労務費管理



● 法人手続き



● 開業手続き



● 社宅制度の導入



● 申告書作成



● 見積・発注・請求



● クレジットカード



● 電子契約



フリーのデータ・AI 活用ニーズ



プロダクト Data



お客様との接点、属性、商談 Data



フリーの経営資源
(人、物、金)



データ
プラットフォーム



分析・レポート



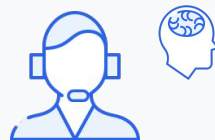
セールス・マーケティング



データ / AI プロダクト



DX / オペレーション



サクセス / サポート

フリーのデータ・AI活用ニーズ



本日本お伝えしたいこと

1. 生成 AI がもたらしたデータ活用の変化と企業戦略
2. なぜフリーは次世代のデータアーキテクチャに取り組むのか？
3. データの利便性とガバナンスをバランスするための具体的な施策

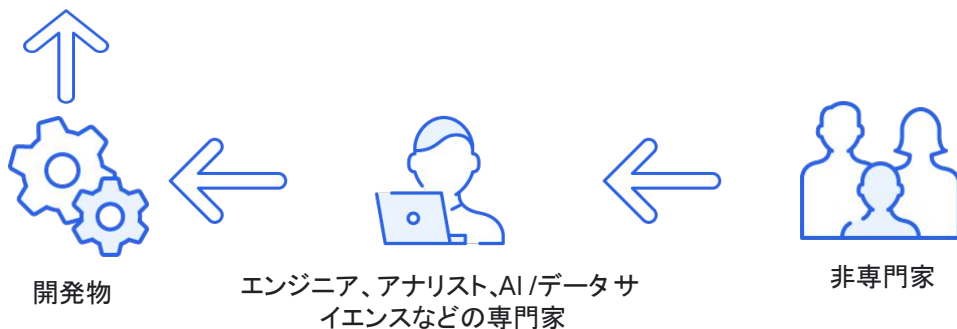
一般的なお話

フリーの話

02. 生成 AI がもたらしたデータ活用 の変化と企業戦略

生成AI がもたらしたデータ活用の変化

これまでの多くのデータ活用は専門家に依存



生成AIがもたらしたデータ活用の変化

専門性のハードルを下げ非専門家でも活用できる時代へ



生成AIを最大限活用するための条件

生成AIに自社のコンテキストを漏れなく/分断なく入れられるか



生成 AI を最大限活用するための条件

生成 AI に自社のコンテキストを漏れなく / 分断なく扱うプラットフォーム



専門家



非専門家



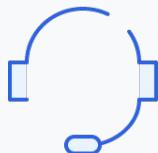
生成AI

個社ごとのビジネス ロジックを反映させたデータモデル / セマンティクス

プロダクト / サービス



事業・サービス
データ



サポート
データ



ご意見



売上データ

ビジネスデータ



顧客
データ



商談
データ



マーケティング
データ

社内データ



社内オペレー
ションデータ



ナレッジ・非構造



財務
データ



人事
データ

マルチモーダル



画像



動画、音声

省庁からも AI における示唆が出始めている

AI・ソフトウェアによる破壊。データ戦略がその競争源になる

引用

経産省: デジタル経済レポート データに飲み込まれる世界、
聖域なきデジタル市場の生存戦略

”” 聖域なきデジタル市場では、すべての企業がソフトウェアカンパニーに、そしてデータカンパニーにそのモデルを変革しなければ生存は極めて難しい。””

https://www.meti.go.jp/policy/it_policy/statistics/digital_economy_report/digital_economy_report.pdf

引用

デジタル庁: R6 年度 生成 AI の業務利用に関する技術検証、
利用環境整備 報告書

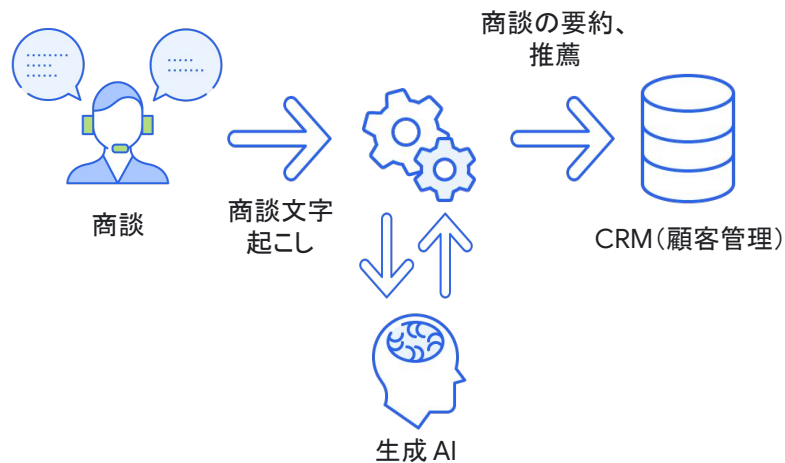
”” 目下、我が国としては、今後の AI 活用の基盤となるデータの整備等を含むインフラの整備・強化に向けた検討・取組や、AI の実態と動向を把握し、リスクと必要な対応策を特定した上で官民における適切な AI 活用に向けた検討・取組を進めることが重要である。””

https://www.digital.go.jp/assets/contents/node/information/field_ref_resources/527968c1-5f55-42d4-868c-54112776c19f/9df94519/20250602_news_generative-ai_report_01.pdf

フリーでも AI Transformation が始まっている

商談後のオペレーションを自動化するつばめAuto

今後、文字起こしなどのデータ活用は自明。横断的な活用に備える必要性



Forbes JAPAN NEW SALES OF THE YEAR 2025で
「AIトランスフォーメーション賞」を受賞



ここまでのまとめ

生成 AI がもたらした効果と企業が適応し競争力を得るには

1. 生成 AI がもたらした非専門家によるデータ・AI 活用の可能性
2. 生成 AI の価値を最大化し競争力を得るための条件
 - a. 事業の営みで発生したデータ、ナレッジを分断なく統合
 - b. 統合したデータに対するデータモデルとビジネス ロジックの整理
 - c. 合理的なユースケースを限定しないアーキテクチャと開発体制

03. なぜフリーは次世代のデータ アーキテクチャに取り組むのか？

フリーは生成 AI を最大限活用できる状態か？

データの統合・分断防止



 データの統合済み

BigQuery へ統合は実施

データモデル
セマンティクスの整備



Raw データの活用が中心

データモデル、セマンティクス
整備過渡期

合理的なユースケースの
実現



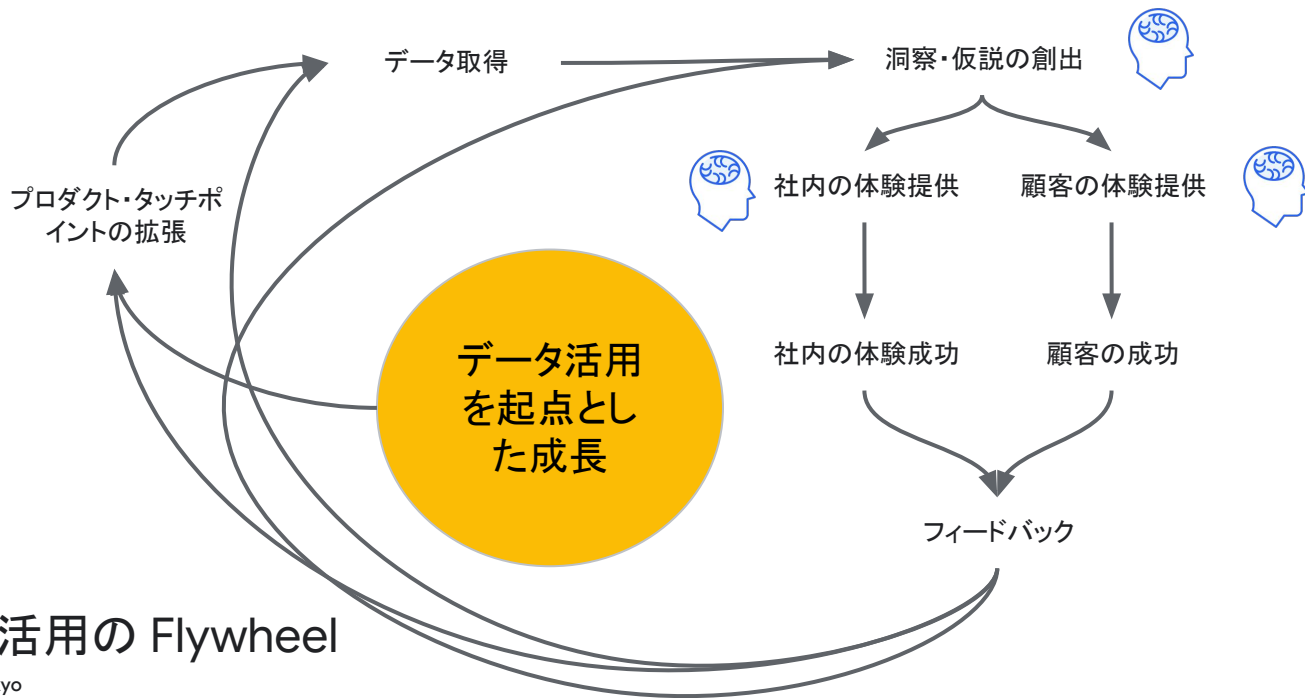
統計データの活用が中心

お客様パーソナライズなど
が出来ず、活用の幅に限界



フリーのデータ活用理想

生成 AI とデータを活用のエンジンとして事業の成長とお客様の成功を高速に実現



データ活用の Flywheel

理想状態へ向かうための壁は？



統計データの利用が中心

お客様パーソナライズなど
が出来ず、活用の幅に限界

- 統計情報ではなく、具体的な機微情報の活用を可能にする必要性
 - プライバシー ポリシー・法律と合理性判断が前提
- 利用する人・役割・リテラシーに応じた柔軟な開発環境とアクセス ポリシー
 - 一つの環境を様々な目的・組織が開発するのではなく論理的に自律
 - 自律することで、最小限のデータをサンドボックスで扱えることが理想

過去の登壇で詳細に説明

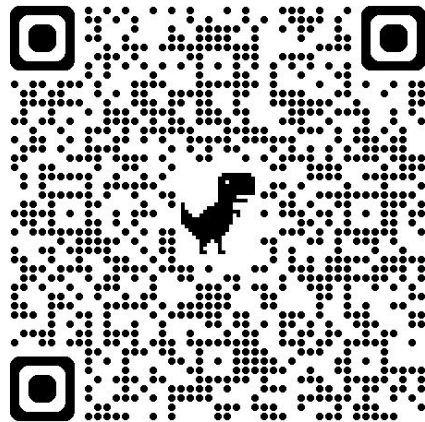
datatech-jp Casual Talks #6 にて背景を詳細を説明



datatech-jp Casual Talks #6

なぜfreeeはハブ・アンド・スポーク型の
データメッシュアーキテクチャにチャレンジするのか？

フリー株式会社 エンジニアリング基盤本部 データ部 城谷



理想状態に向かう概念アーキテクチャ

フェーズとコンテキストを言語化したメッシュ型アーキテクチャ

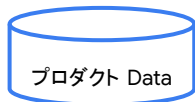
データ統合フェーズ

データ共有フェーズ

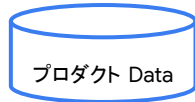
データ活用フェーズ



プロダクト Data



プロダクト コンテキスト



プロダクトデータ資産



データ モデリング コンテキスト



分析



セールス
マーケ



お客様との接点、属性、
商談 Data



ビジネス コンテキスト



ビジネスデータ資産



ビジネス コンテキスト



プロダクト



サポート



フリーの経営資源
(人、物、金)



社内・バックオフィス コンテキスト



社内データ資産



社内・バックオフィス コンテキスト

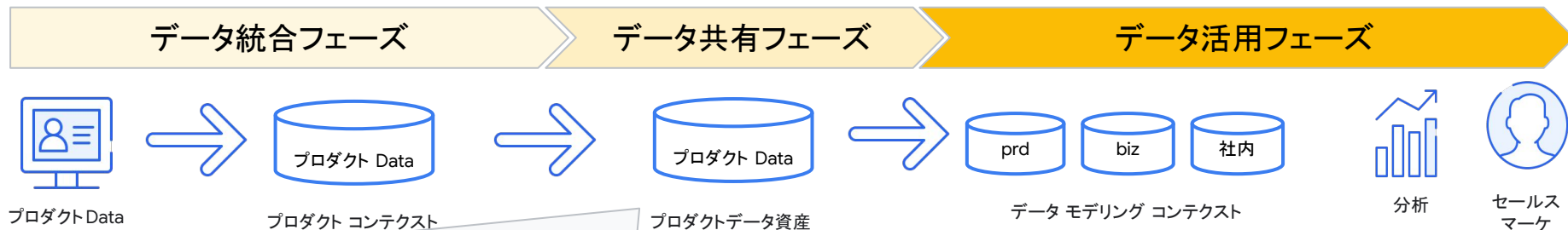
など複数のコンテキスト



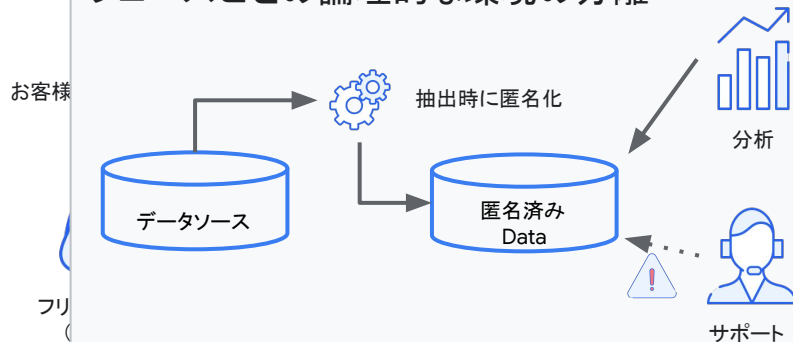
DX / Ope

理想状態に向かう概念アーキテクチャ

フェーズとコンテキストを言語化したメッシュ型アーキテクチャ



フェーズごとの論理的な環境の分離

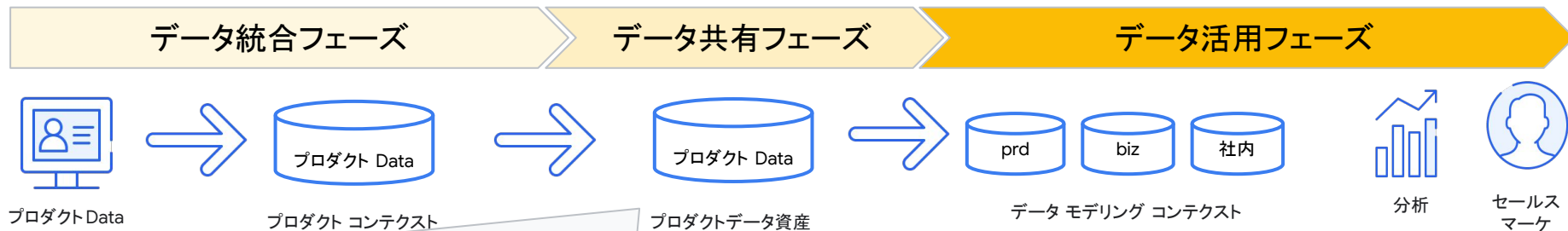


- データ基盤のよくある課題として、セキュリティを優先するためにデータ抽出時点で、匿名化・統計化を実施する
- そのため、分析など統計利用は可能だが、お客様サポートなど機微情報の合理的な利用が不可になり、別のシステム構築を行うなど全体的な生産性を低下させる原因に

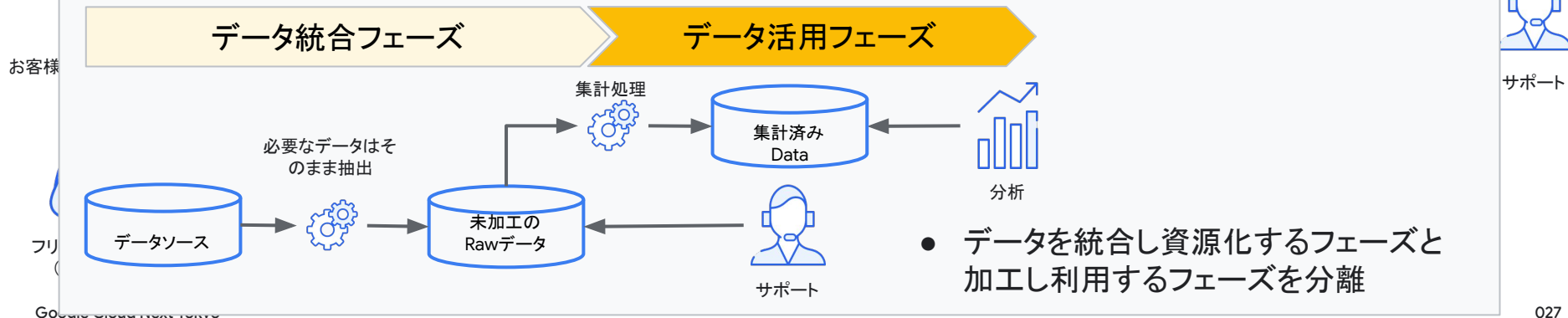


理想状態に向かう概念アーキテクチャ

フェーズとコンテキストを言語化したメッシュ型アーキテクチャ

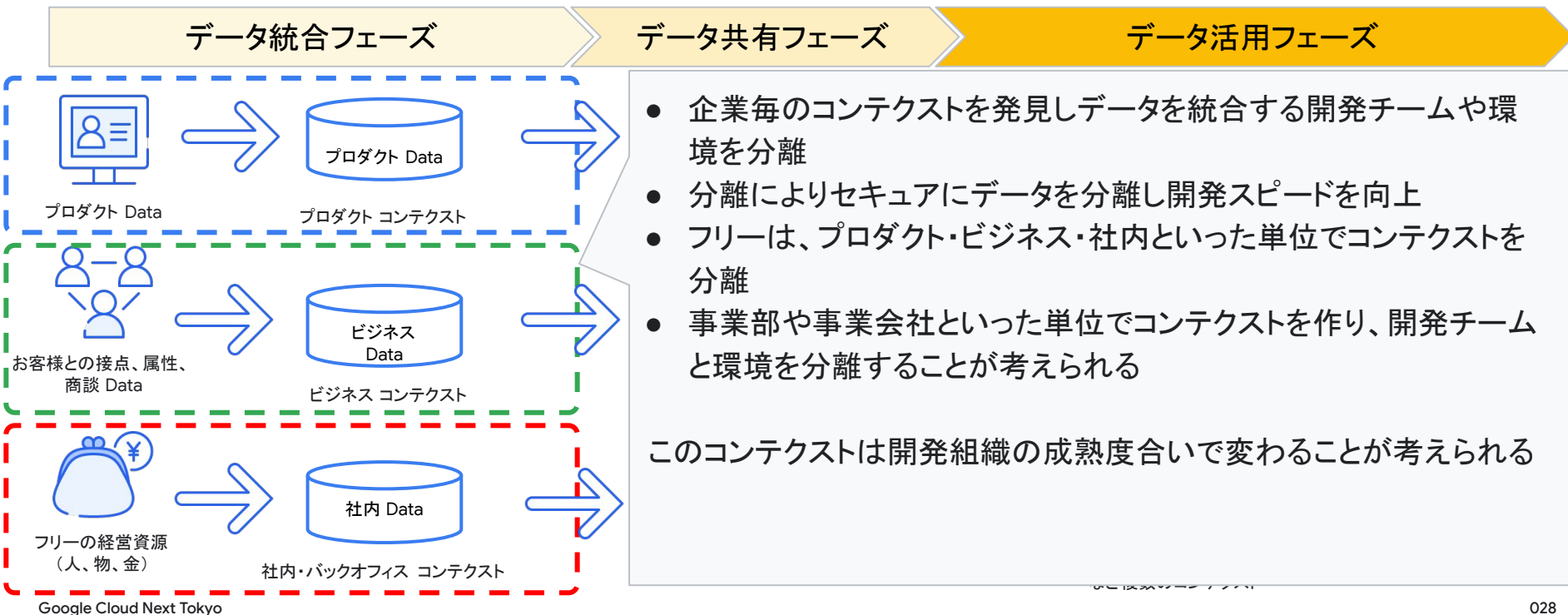


フェーズごとの論理的な環境の分離



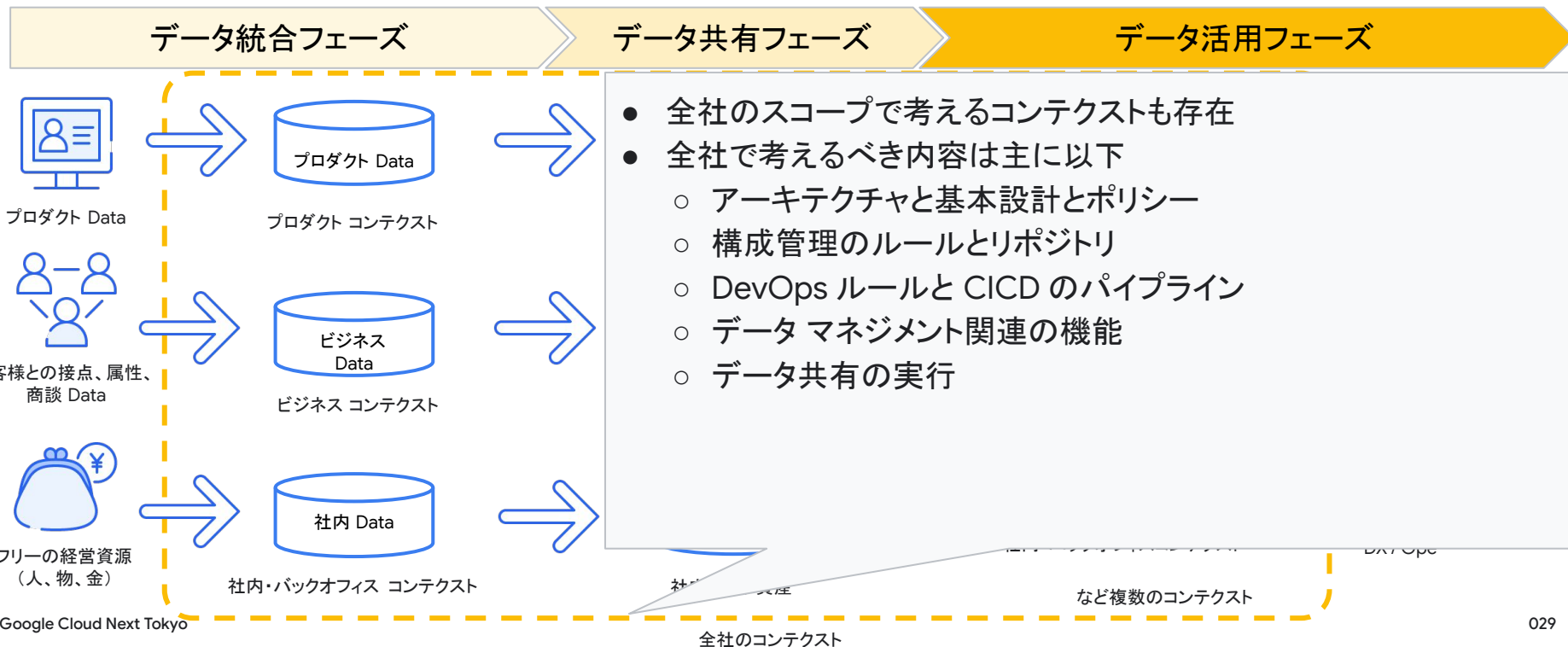
理想状態に向かう概念アーキテクチャ

フェーズとコンテキストを言語化したメッシュ型アーキテクチャ



理想状態に向かう概念アーキテクチャ

フェーズとコンテキストを言語化したメッシュ型アーキテクチャ



理想状態に向かう概念アーキテクチャ

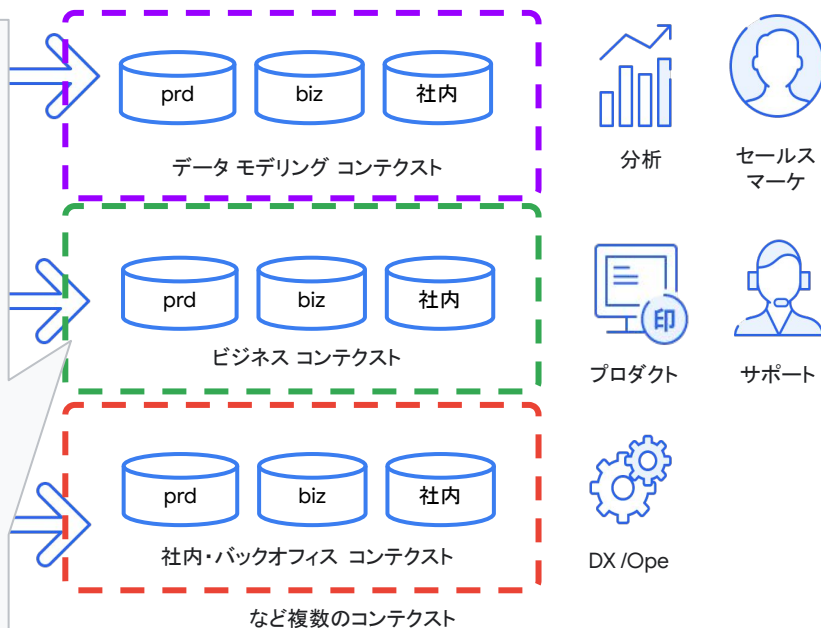
フェーズとコンテキストを言語化したメッシュ型アーキテクチャ

データ統合フェーズ

データ共有フェーズ

データ活用フェーズ

- 活用のフェーズになると、利用目的ごとにコンテキストを細分化
- プロダクト/ビジネス/社内のデータを横断して活用すること想定し活用環境を整備
- 例えば、データモデリングコンテキストには、企業の機密データから様々な統計情報含むデータモデリングを行うので、機密情報を提供しセキュアに管理
- それ以外は、必要なデータを目的に合わせて提供



理想状態に向かう概念アーキテクチャ

フェーズとコンテキストを言語化したメッシュ型アーキテクチャ



ここまでのまとめ

フリーが目指す理想の生成 AI / データ活用に向けた概念を提唱

1. データを統合・資産化するフェーズと活用フェーズを明確に分離
2. 企業の特性や短期的な未来を鑑みて、コンテキストを分離し組織作り
3. 環境が分離することで非効率化する箇所は、基本ポリシー・構成管理・DevOps
など標準ルールとツールセットを全社に提供

次ページからは具体的な Google Cloud でのアーキテクチャ実現を DeepDive

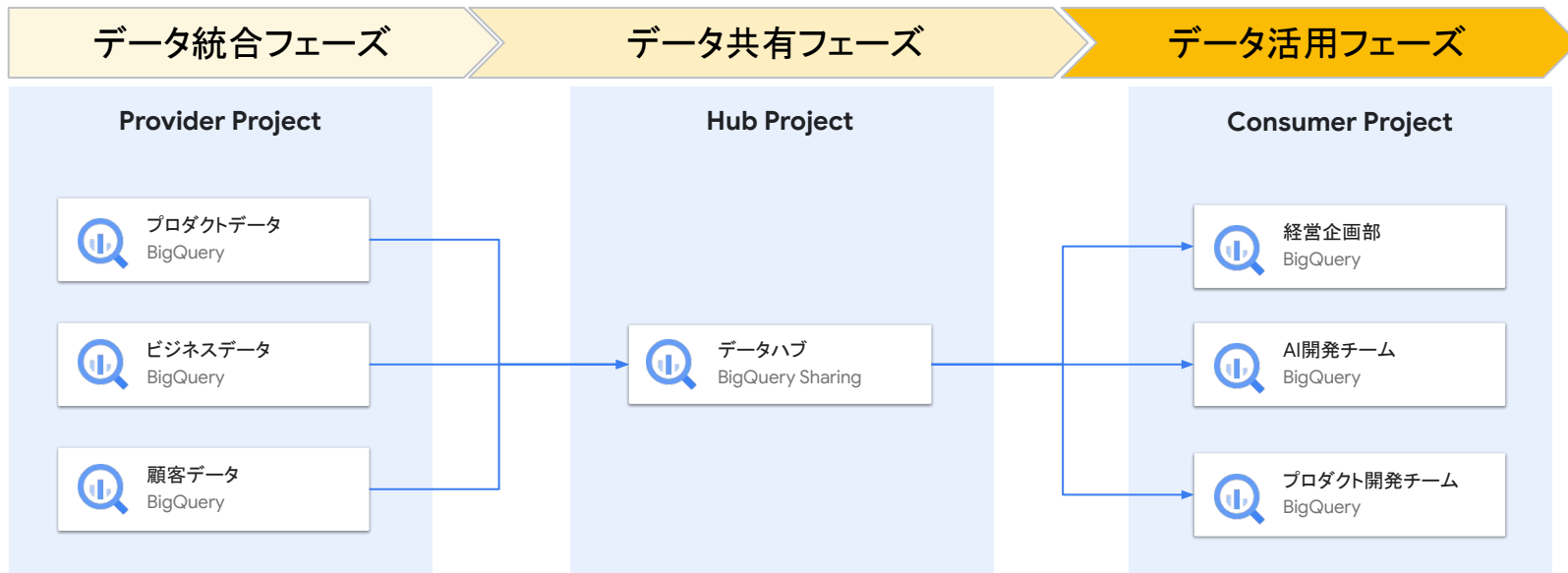
04. データの利便性とガバナンスの両立

西 朋里

エンジニアリング基盤本部
データ部 エンジニア

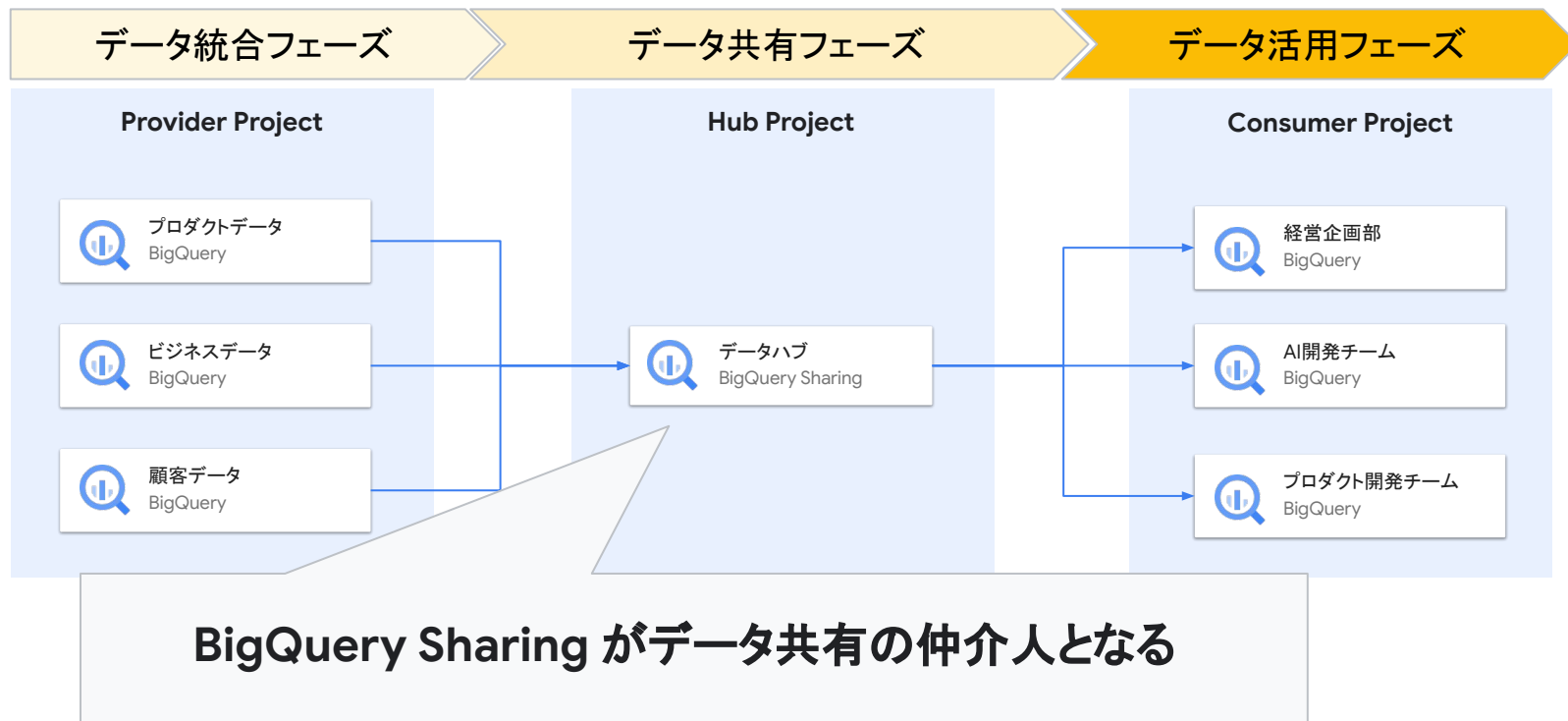


Google Cloud で構築するデータメッシュ構成



- Provider Project: 元データを保存するGoogle Cloud Project
- Hub Project: BigQuery Sharing の設定を持つGoogle Cloud Project
- Consumer Project: データ分析やマートの構築をするGoogle Cloud Project

データ共有の中核を担う BigQuery Sharing



データ共有の中核を担う BigQuery Sharing (旧 Analytics Hub)

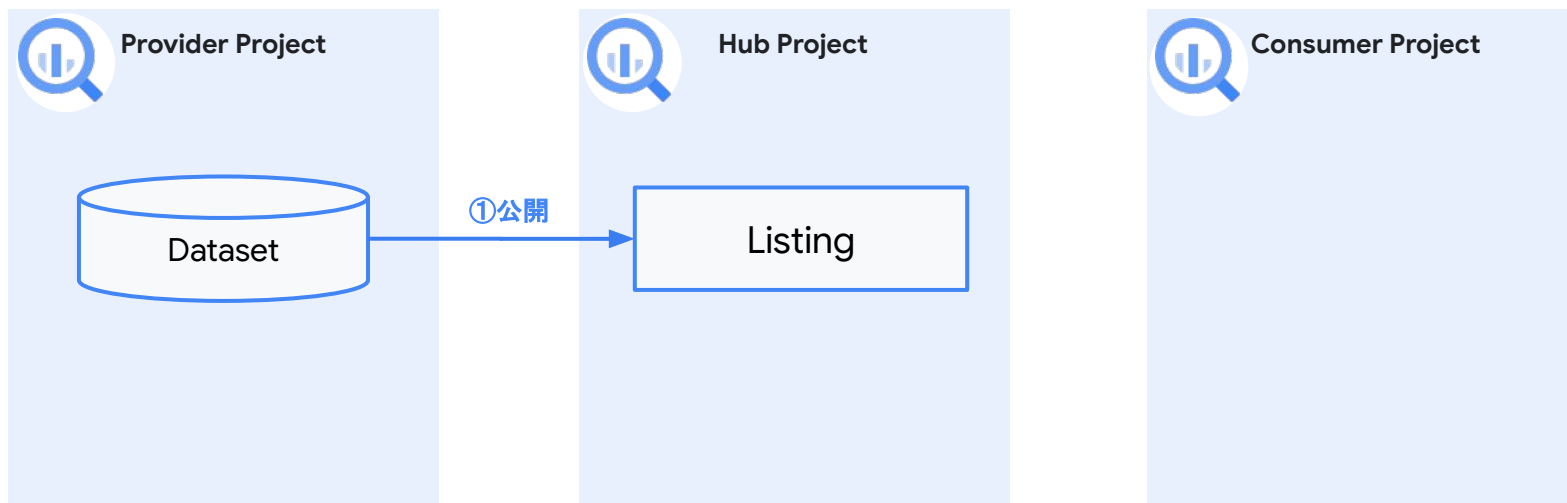


BigQuery Sharing

- グループ会社間でのデータ共有
- オープンデータの公開
- 同一組織内でのデータアクセス権限管理
- 特定のデータセットを特定の目的で他のプロジェクトに共有

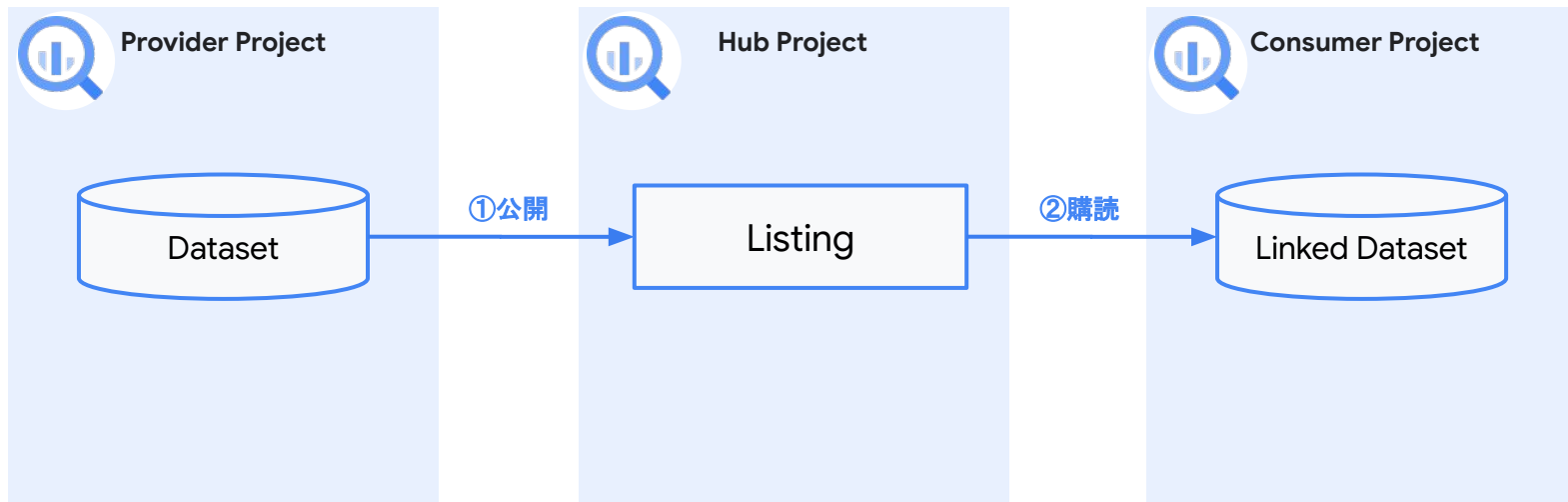
データ共有の仕組み ① 公開

データセットと紐づいた共有単位リスティングを作成



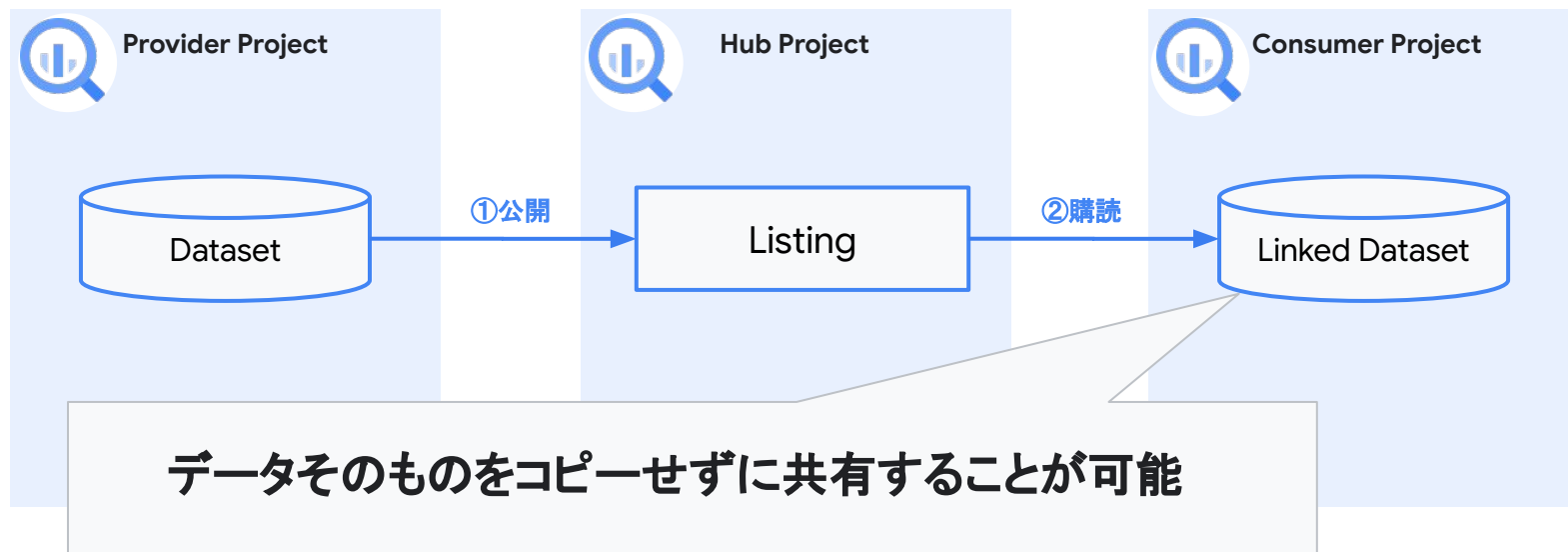
データ共有の仕組み ② 購読

リスティングを購読(サブスクライブ)し、Linked Dataset を作成



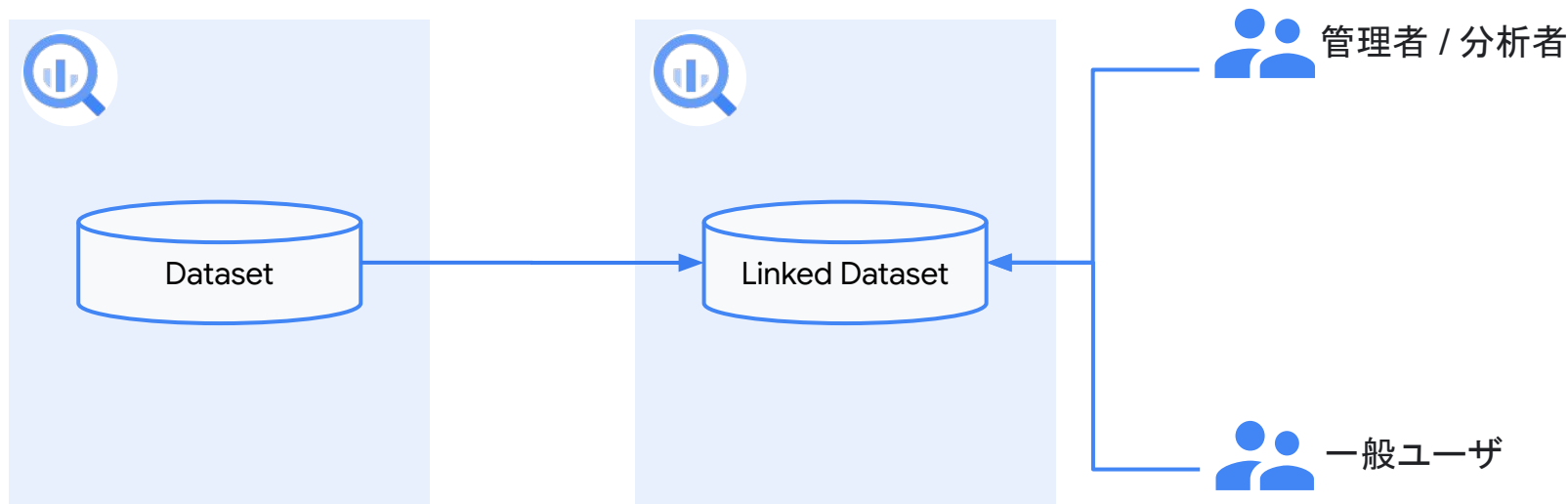
データ共有の仕組み

Linked Dataset は物理データを持たないため
シングルリソースによるデータ共有が可能



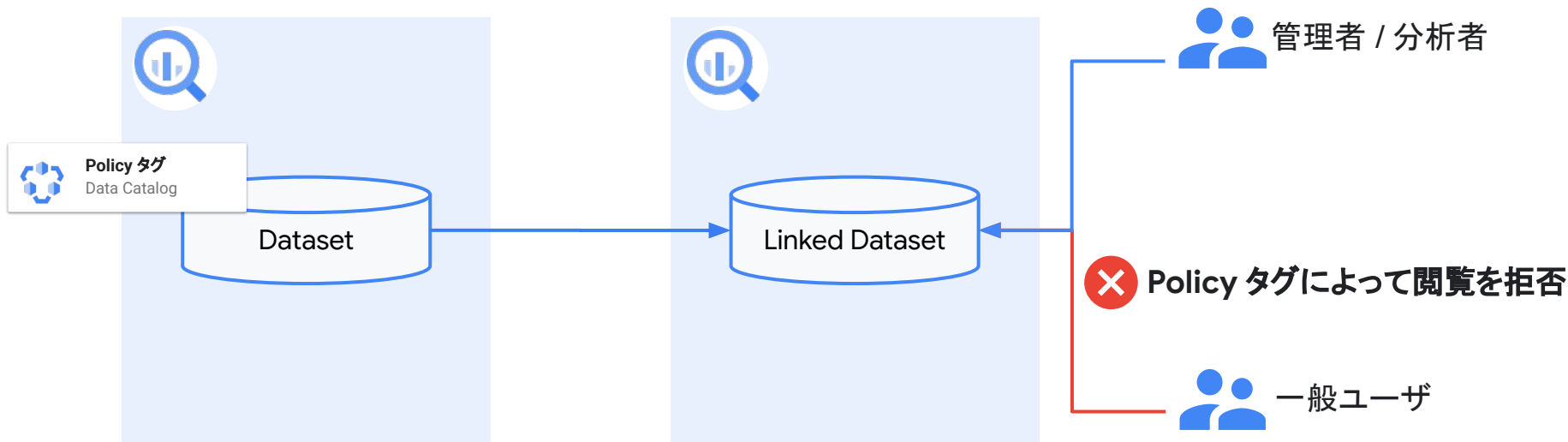
より安全なデータ共有を目指して

データセットの全てのテーブル・カラムが共有されるため
Google Cloud Project に権限を持つ全てのユーザがアクセス可能

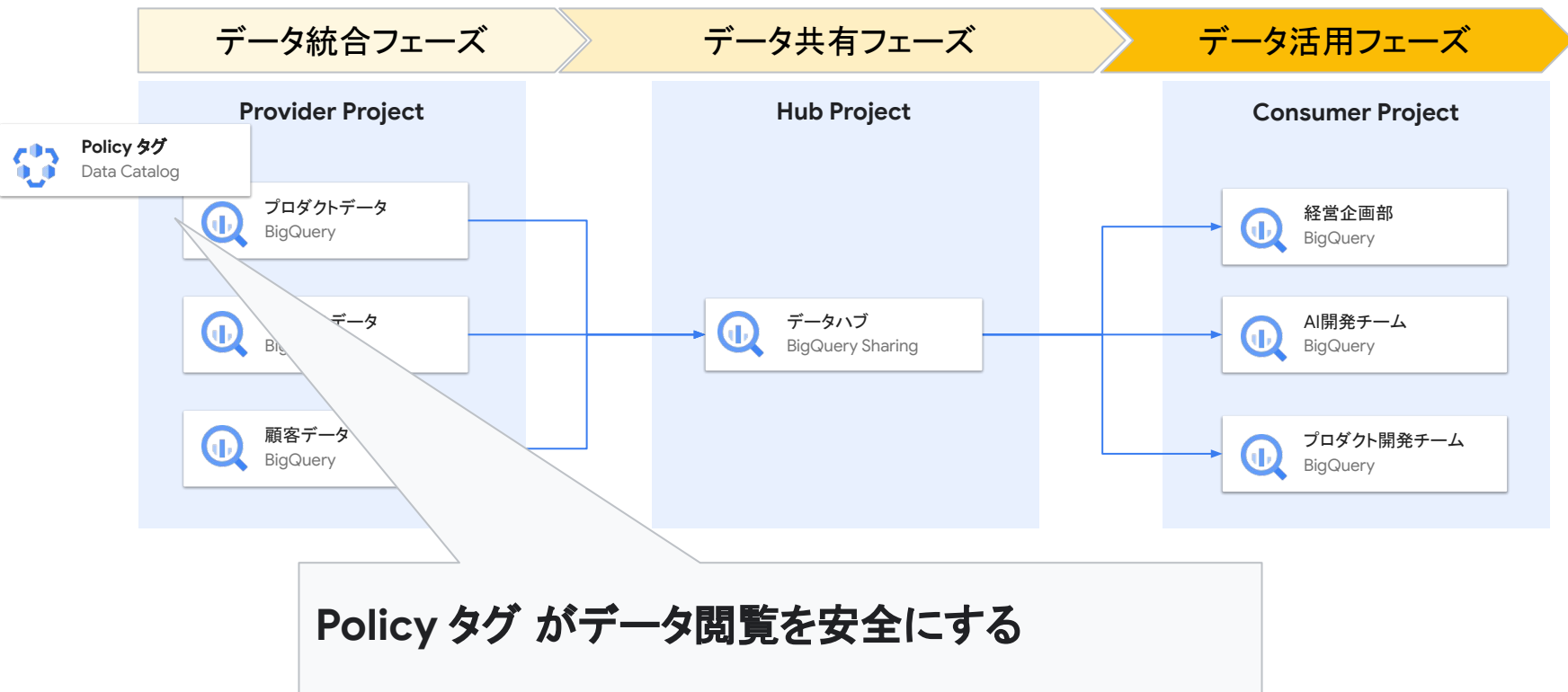


より安全なデータ共有を目指して

Policy タグを使用した閲覧制限を導入する



ガバナンスの要となる Policy タグ



ガバナンスの要となる Policy タグ

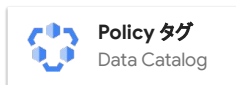


Data Catalog

- Data Catalog の一機能
- BigQuery のカラムレベルのアクセス制御を可能にする

Policy タグによる動的なアクセス コントロール

アクセス権を持つユーザーが参照する場合



Tag: id

Tag: Name

Tag: Email

Tag: accounting



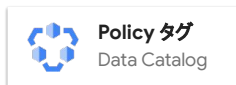
管理者 / 分析者



user_id	name	email	amount
1001	クラウド 太郎	g-taro@example.com	15,000
1002	見本 花子	m-hanako@example.com	8,800
1003	デー タ 一郎	d-ichiro@example.com	23,100
1004	テスト 次郎	t-jiro@example.com	5,400

Policy タグによる動的なアクセス コントロール

アクセス権を持たないユーザーが参照する場合



Tag: id

Tag: Name

Tag: Email

Tag: accounting



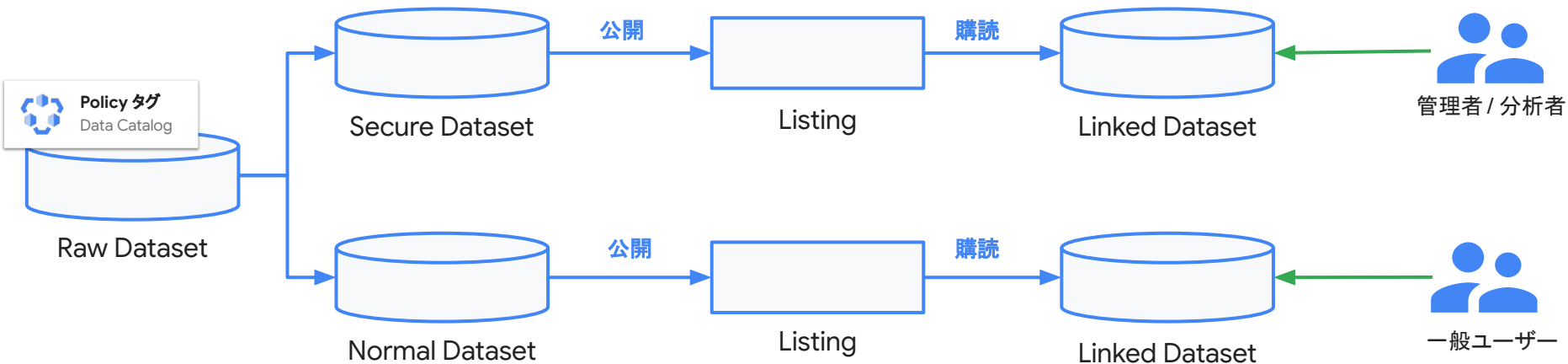
user_id	name	email	amount
1001	NULL	NULL	NULL
1002	NULL	NULL	NULL
1003	NULL	NULL	NULL
1004	NULL	NULL	NULL

Policy タグを元に共有用データセットを作成

データ統合フェーズ

データ共有フェーズ

データ活用フェーズ

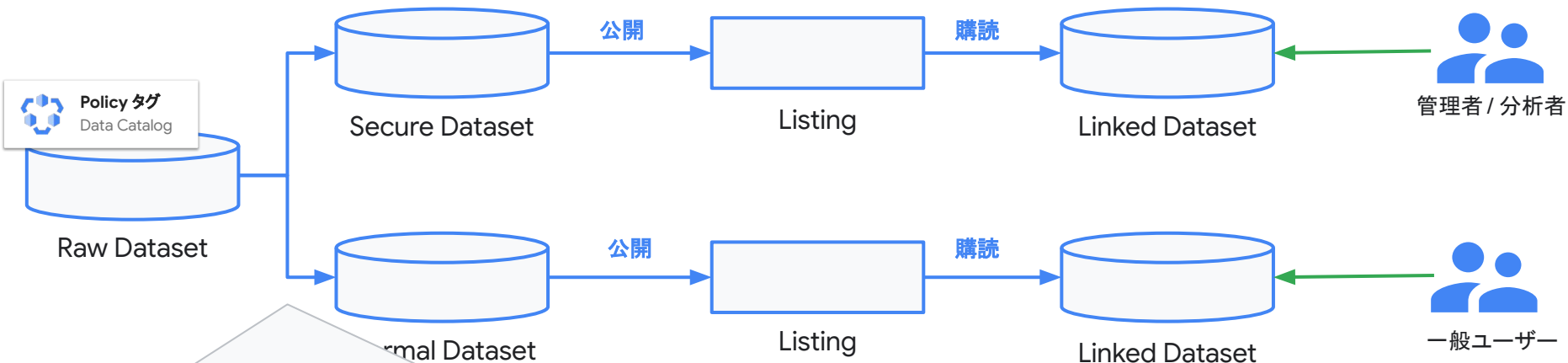


Policy タグを元に共有用データセットを作成

データ統合フェーズ

データ共有フェーズ

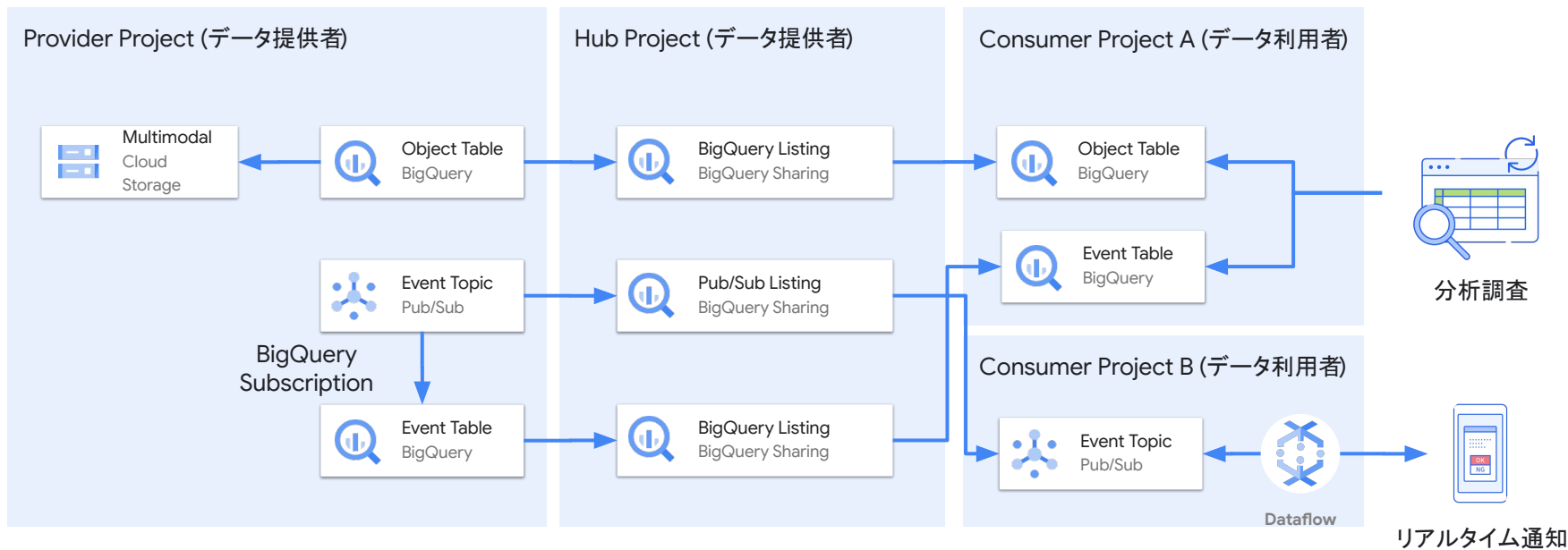
データ活用フェーズ



Policy タグを元にセキュリティレベルの異なるデータセットを自動作成

多様なデータの SSoTを維持した組織内シェア

- 画像、動画、音声などのマルチモーダルデータの組織内シェア
- Pub/Sub キューの組織内データシェア



生成 AI によるデータ分類の自動化への挑戦

お客様データについては全てのカラムに Tag を付与するのが理想



Tag: id

Tag: Name

Tag: Email

Tag: accounting

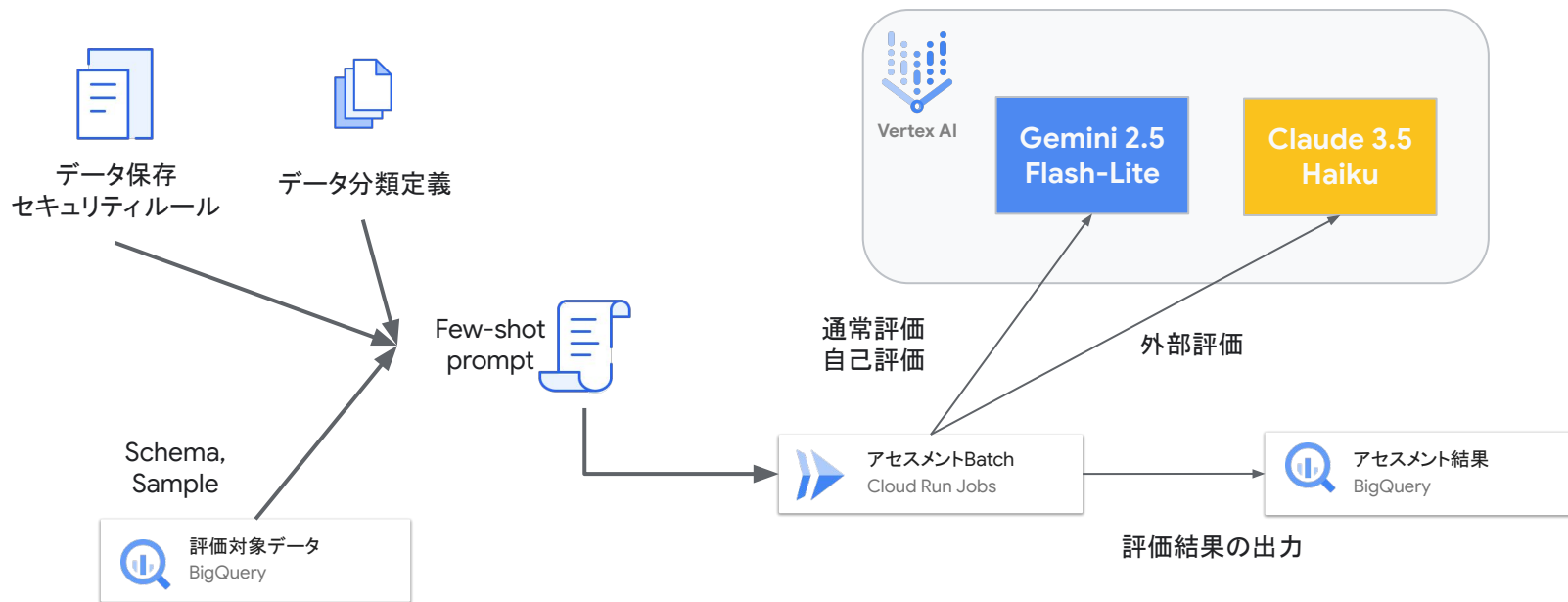
user_id	name	email	amount
1001	クラウド 太郎	g-taro@example.com	15,000
1002	見本 花子	m-hanako@example.com	8,800
1003	データ 一郎	d-ichiro@example.com	23,100
1004	テスト 次郎	t-jiro@example.com	5,400

データの分類 (Tag) が活用の大前提
しかし、15k のテーブル、800k を超える
カラムへ手動での適用は困難



生成 AI によるデータ分類の自動化への挑戦

生成 AI を活用した自動分類の概要アーキテクチャ (PoC)



生成 AI によるデータ分類の自動化への挑戦

Few-shot prompt イメージ

1. セキュリティ ルールとデータの分類定義
 - a. 基本的な全社的のデータに対するセキュリティ定義
 - b. 加えて、より具体的なデータに対する分類定義
2. 評価対象のデータとサンプル
 - a. Information_schema, 業務メタデータ
 - b. 該当カラムのサンプルデータ
 - i. TABLESAMPLE SYSTEM 関数でコストを下げながらサンプルを取得

分類定義については
Policy タグの階層を意識し設計

L1: 個人情報
└ L2: 機微な個人情報
└ L3: 識別情報
└ L3: 連絡情報
など

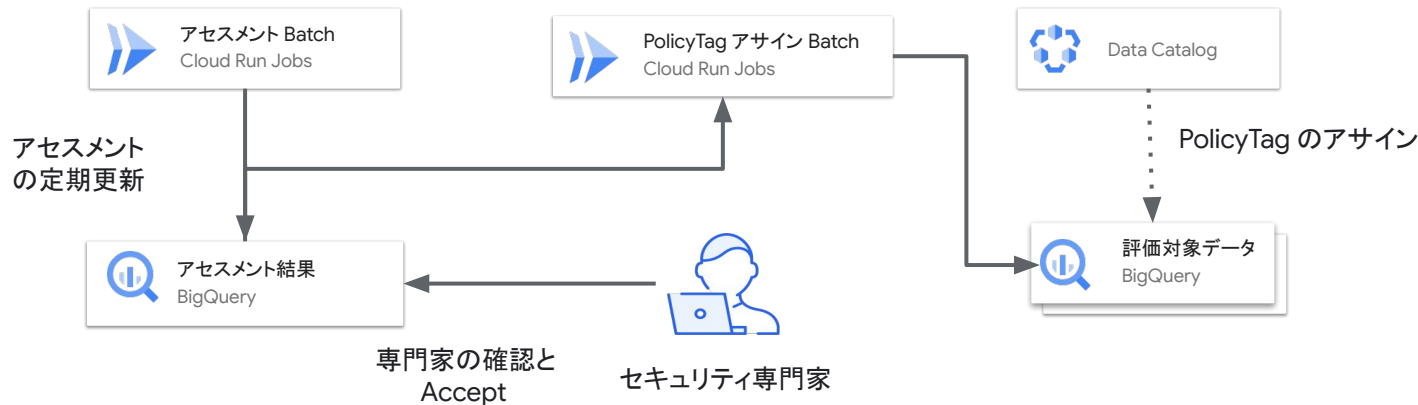
生成 AI によるデータ分類の自動化への挑戦

アセスメント結果のイメージ

1. カラム単位の生成 AI による判定スコア: self_consistency_score
2. データ分類の判定タグと理由: detail_classification, detail_classification_reason
 - a. 具体的なカラム毎に email = L1:個人情報_L2:機微情報_L3:連絡情報 といったタグを提案
3. 生成 AI が自動的に作った安全なサンプル: sample_data
4. 自己整合と外部整合の評価:
 - a. 自己整合: 同じモデルで、Temperature を 0.3-0.4 程度と回答がブレる形にして複数回評価
 - b. 外部整合: 自己整合にブレがある＝分類が誤っている可能性があるため別モデルで評価

生成 AI によるデータ分類の自動化への挑戦

ヒューマン インザループの運用構想



負荷が低い運用と確認のためのI/F 整備が今後の課題

05. まとめ

生成 AI による果実を得る企業の条件は 企業内の情報を分断なく統合できること

生成 AI は専門家でなくても活用の裾野を広げるチャンス
一方で、データ・ナレッジ整備が大前提となり成否の分かれ目
フリーでは、メッシュ型アーキテクチャと対応した組織で適応



Google Cloud を使った 企業内の情報を統合するためのプラクティス

フリーでは、BigQuery Sharing と PolicyTag をコアに実現

組織内で様々な利用部署間でシェアリングエコノミーを構築

疎結合で自律可能な全体最適の仕組みと個別最適を実現

事業のデータを評価し素早く活用するための生成 AI 活用もトライ中

シェアリング エコノミーは 組織外にもポテンシャル

