

Why Google Cloud Analytics?

~分析システムのステークホルダーごとの課題とGoogle Cloudでの解決策~

スピーカー自己紹介: 鈴木邦明

@Google Cloud Solution & Technology Group

Data Analytics GTM Practice Lead



2002	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	2018-
------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-------

2002/04 – 2010/04

エンジニア@国内Sler

2010/05 – 2014/04

アカウントマネージャ
@ web解析ベンダ

2014/05 – 2017/12

コンサルタント
@ BIツールベンダ

2017/12 - present
Global Technical Partner
Data Analytics Specialist
@Google

本日のアジェンダ

データ分析基盤とステークホルダーの悩み	01
Google Cloud の Smart Analytics で解決できること	02

01

データ分析基盤と ステークホルダー

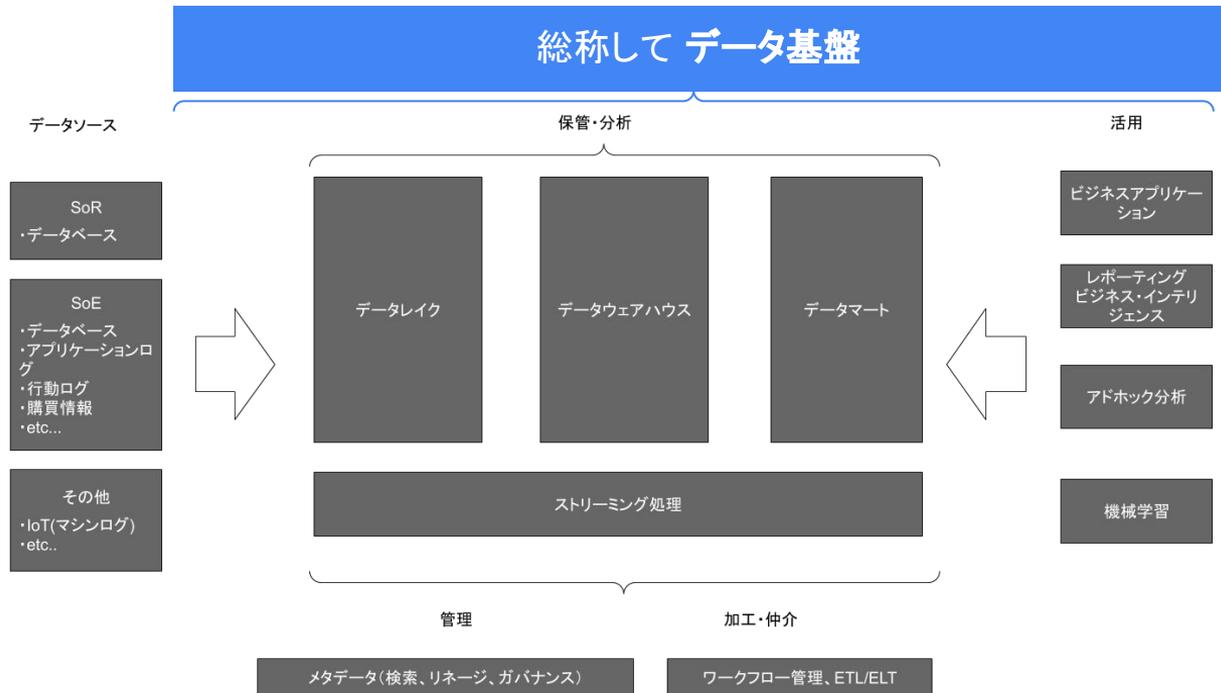
データ分析、活用を支えるデータ基盤

様々なコンポーネントが存在

- 収集、加工、仲介、管理、保管、分析、活用

データ エンジニアの貢献

- データの利用者が**信頼性のあるデータ**に対し**効率的に分析**ができる



出典: 技術評論社、Google Cloud で始める実践データ エンジニアリング (2021)

データ分析のステークホルダー



ビジネスユーザー



グロースハック/
デジタルマーケ



データアナリスト
データサイエンティスト



IT 部門



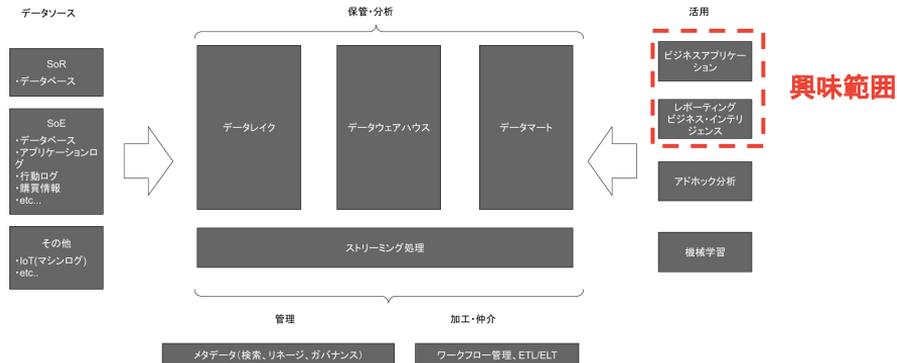
データ分析のステークホルダー



ビジネスユーザー

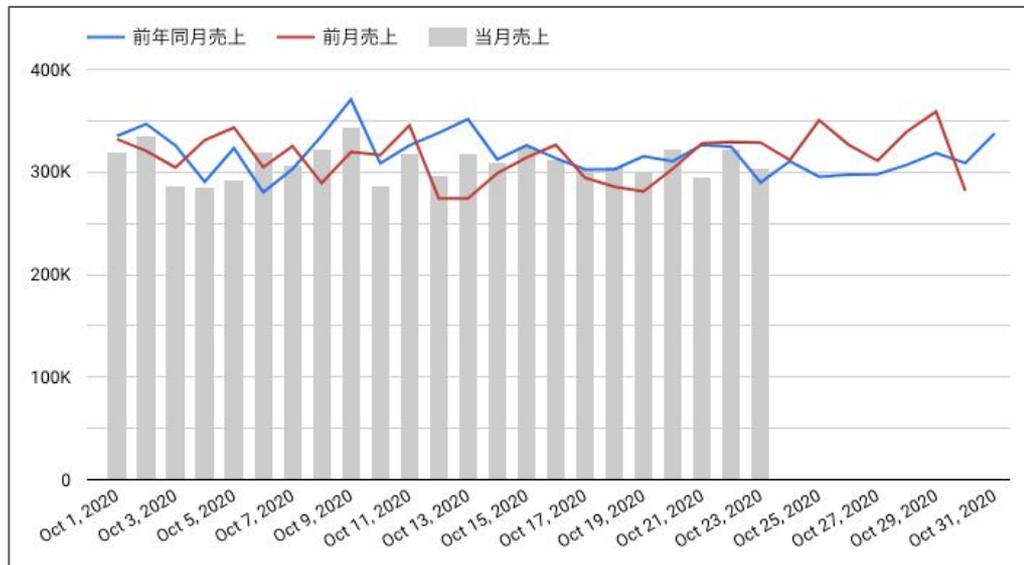
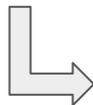
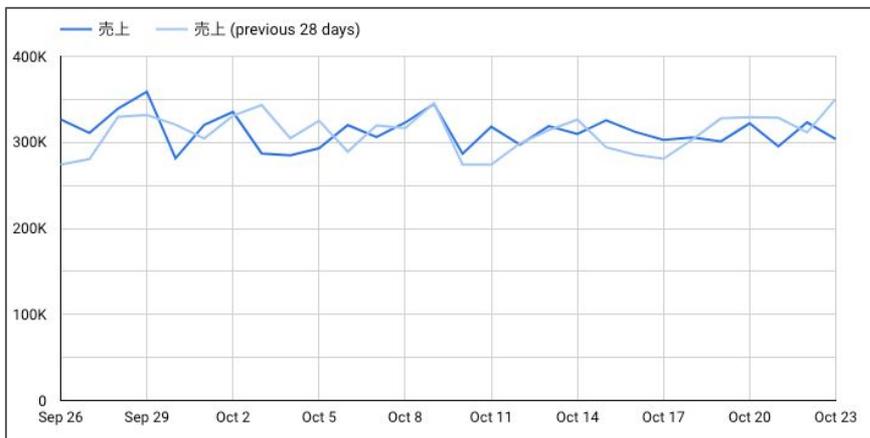
やりたいことのイメージはあるけれど、実現方法がわからない

- データの在り処やそれに対するアクセス方法が不明
- できる人でも Excel の関数やピボットテーブルが限界、SQL は無理
- IT 部門の用意してくれた BI ツールのダッシュボードでは要件が満たされない、変更を依頼しても待たされる



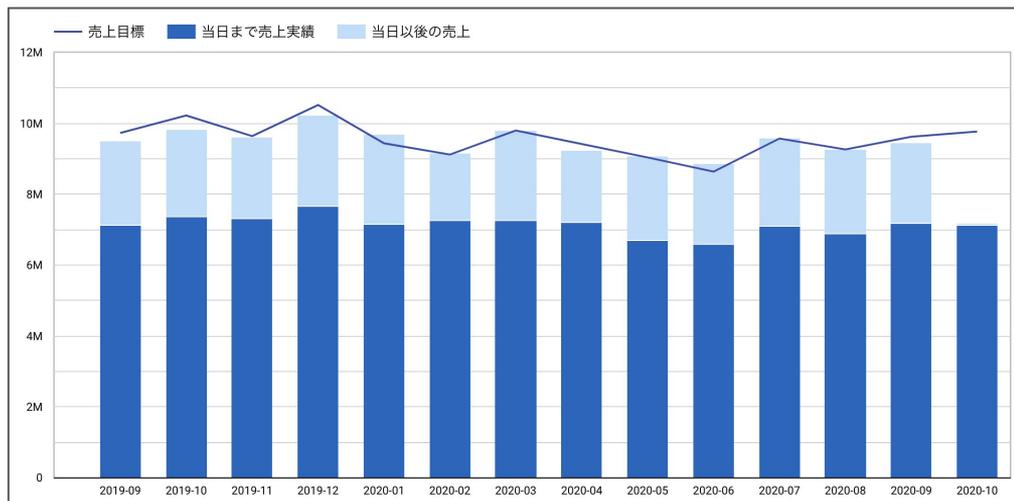
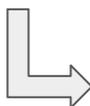
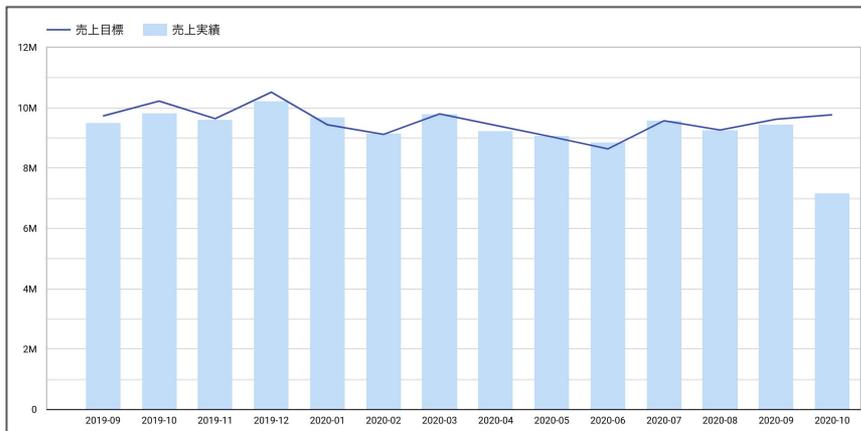
BI ツールでは表現力が足りない例 1

- 期間比較が固定・グラフ種類を変えられない・系列を増やせない
- 複雑な期間比に対応できない(旅行業などに顕著)



BI ツールでは表現力が足りない例 2

Month to date での売上を表現したい(月中でも他月と進捗を比較できる)



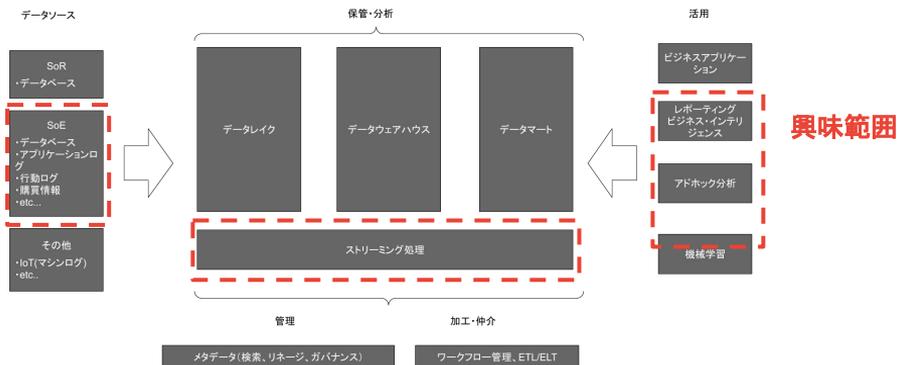
データ分析のステークホルダー



グロースハック
デジタル マーケ

主戦場は web・スマホアプリ、純粋なビジネスユーザよりもデータ慣れしている

- 専門の分析ツールには慣れている(web なら GA・アプリなら Firebase など)
- チャンネル横断での分析まで手がけたい(≒CDP - Customer Data Platform、顧客データ分析基盤がほしい)
- 業態によってリアルタイム データに対する要求が強い



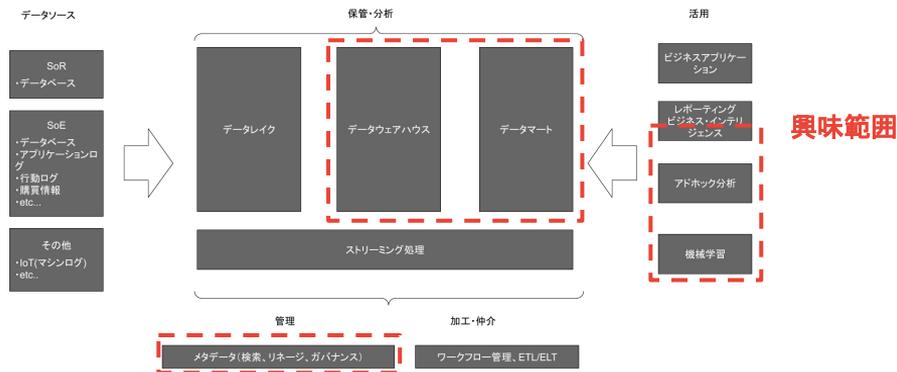
データ分析のステークホルダー



データアナリスト
データサイエンティスト

データの扱いが得意

- 自分の得意とする言語でデータを触りたい(SQL / Python + Jupyter Notebook)
- データの素性を重視する(定義・取得元・更新頻度)
- クエリ速度に生産性が大きく依存するが、IT 部門の管理するインフラに手が出ない
- 分析結果を共有する方法がない



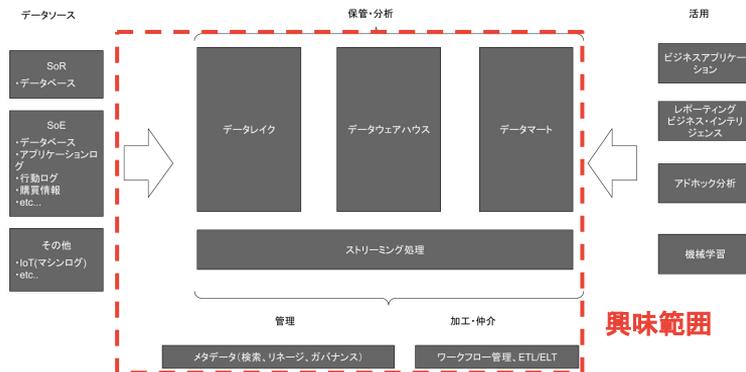
データ分析のステークホルダー



IT 部門

分析環境のインフラ・ETL ジョブの管理が主な関心事

- 将来のデータ増加予測に基づくインフラのサイジングが困難 → サイロ化
- ビジネス部門の要件が曖昧で頻繁にかわる、それに合わせた DB のチューニングが辛い
- 要件を全て実現するとストレージ枯渇やバッチ突き抜けリスクがある
- セキュリティ・ガバナンスの観点から LoB 側に勝手に BI ツールやDWHを持ってほしくない



要するに

- **[アクセス性]** スキルレベル・用途に合わせて様々なツールでデータを扱いたい
 - SQL・Python・Excel・BI ツール
 - 各部門がセルフサービスでデータを活用したい
- **[接続性]** 基幹データだけでなく種々のデータソースをサポートする必要がある
- **[パフォーマンス]** データ量の増加は予測不可だがそれに対する高いパフォーマンスが必要
 - 一方で個別のユースケースに合わせたチューニングはできない
- **[セキュリティ・ガバナンス]** 統一的な管理が行いたい
 - データ持ち出し・漏洩リスク
 - データ マネジメント

02

Google Cloud の Smart Analytics で 解決できること

Smart Analytics によるサイロ化したデータの統合

リアルタイムに、シームレスに、未来を予測。

あらゆる場所に
分散したデータ

データを一箇所に
集約

誰でも分析可能
コーディングレスでの
ML と高度な分析

SoR
データ



部門別の基幹 DB,
部門別の DWH



Analytics 360



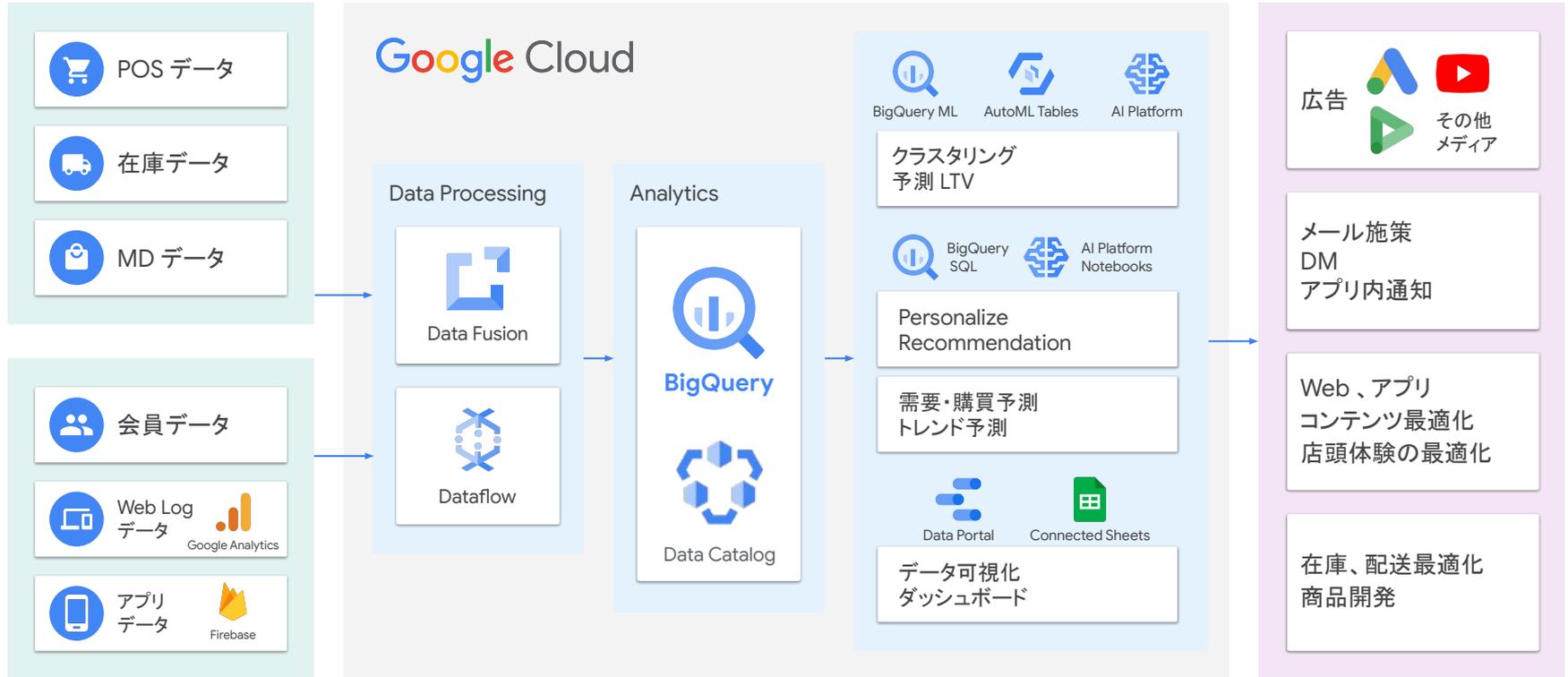
Google Cloud



SoE
データ



データの一連の流れを各コンポーネントでカバー

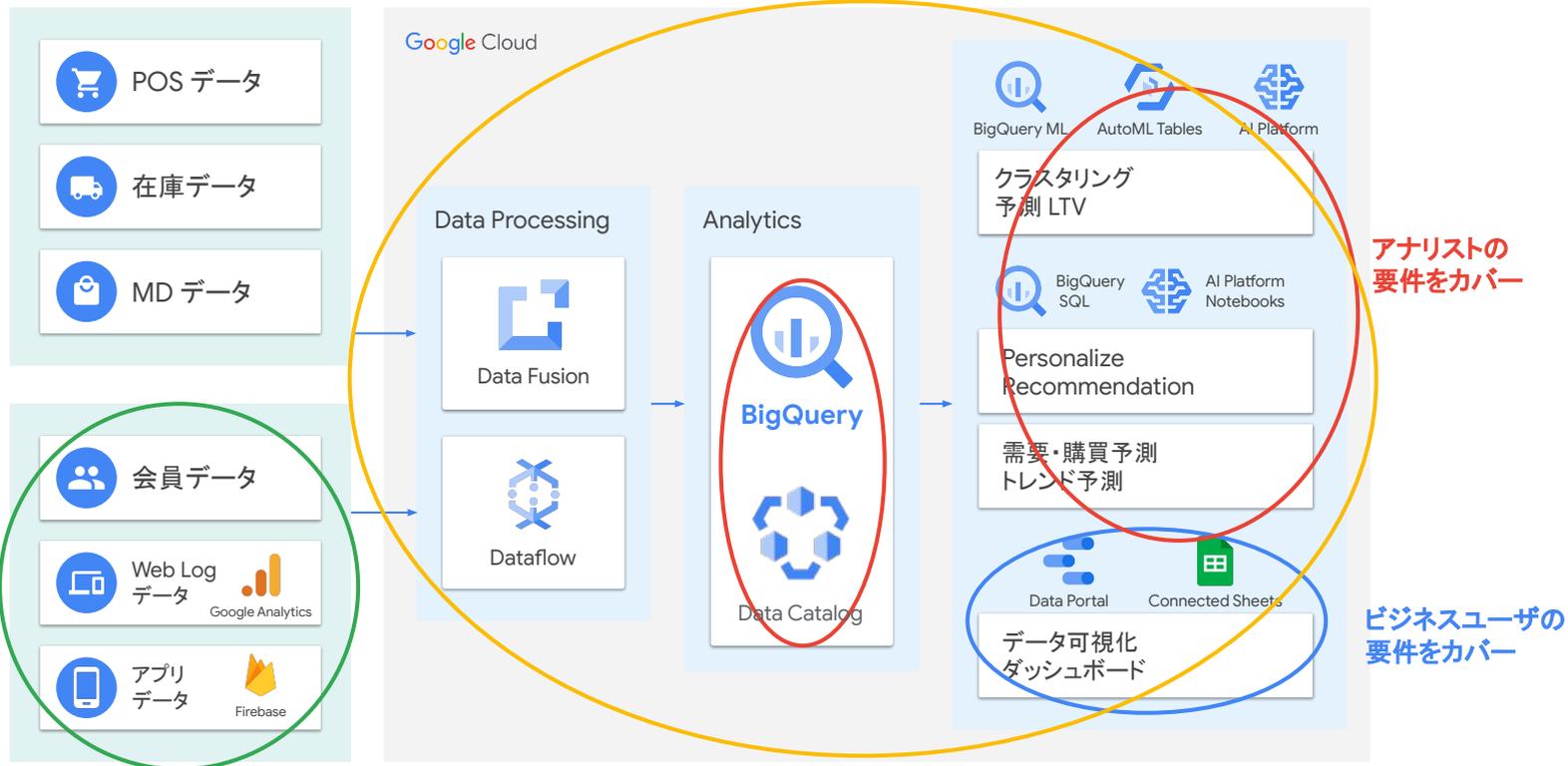


データ収集・加工

分析・予測

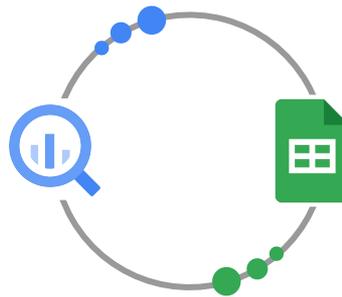
アクション
顧客体験

ステークホルダーと Google Cloud 機能の関連



IT 部門の要件をカバー

ビジネス ユーザー向け: スプレッドシートによる最新データ共有



/ ビジネス ユーザーでも最新のデータに基づき、
分析が可能に

/ SQL 不要で複雑な分析を実行

/ BigQuery のパフォーマンスをフル活用

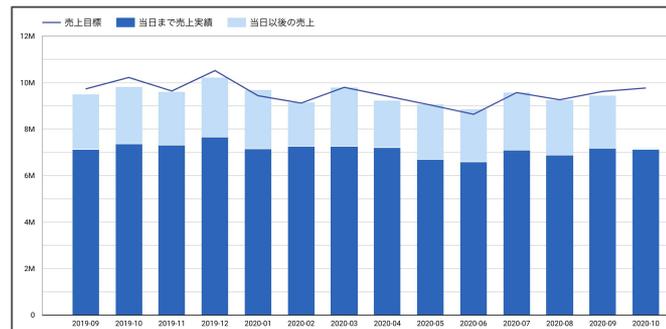
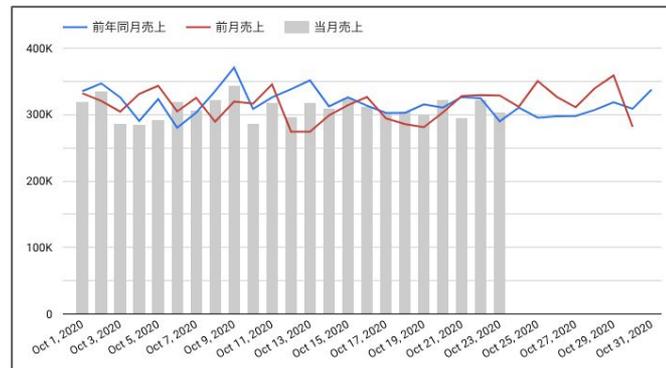
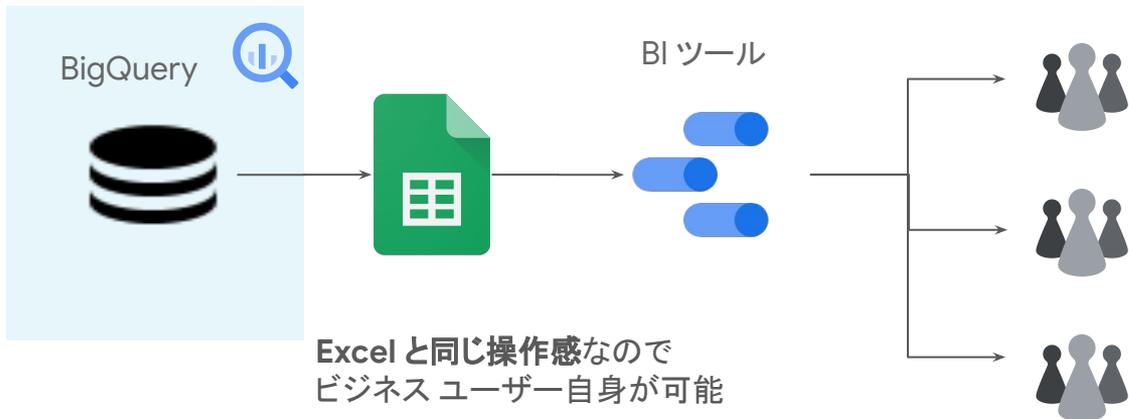
他にも BI ツールのドリルダウンを1ms 以下で返すための
インメモリ分析エンジン、BI Engine を搭載

The screenshot shows the Google Data Studio interface. At the top, it says "Quiz Analytics" with a star icon. Below that, there are menu options: File, Edit, View, Insert, Format, Data, Tools, Add-ons, Help. A search bar and a "SHARE" button are also visible. The main content area shows a data table titled "quizOrders_2017-2018" with 7.75M rows and 5.1 GB of data. The table has 16 columns and options for "Chart", "Pivot table", and "Function". A search bar asks "Ask a question about this data...". Below the table, there is a "Preview" section showing a table with columns: Tr, quizid, startDate, # questionCount, completed, Revenue, quizType, region, Tr, Tr, Tr, # timeToFinish, promoted. The table contains 20 rows of data.

Tr	quizid	startDate	# questionCount	completed	Revenue	quizType	region	Tr	Tr	Tr	# timeToFinish	promoted
	j-9545325774-927	19 Dec, 2016, 4:33:20	1	false	\$0.16	Trivia	Atlantic		a-4823444		156.24	true
	a-9647157787-462	5 Sep, 2014, 20:03:21	1	false	\$0.10	Checklist	Pacific		c-64865308		165.30	false
	k-1160504238-664	5 May, 2014, 17:18:18	2	true	\$0.04	Poll	EU-east		r-53327141		73.42	false
	s-6710999482-129	4 Oct, 2015, 20:46:13	1	true	\$0.03	Checklist	Midwest		a-56347810		196.38	true
	a-5761938468-797	8 Jun, 2016, 14:43:40	2	true	\$0.17	Checklist	Midwest		k-84174436		98.31	true
	k-7219019817-865	5 Jan, 2015, 3:10:48	1	false	\$0.13	Personality	Can		k-72990642		196.51	false
	j-9514985733-877	19 Jun, 2015, 2:25:06	1	true	\$0.11	Personality	EU-west		r-23199634		156.15	true
	j-4181956985-160	15 Sep, 2016, 22:17:28	2	false	\$0.16	Poll	Can		k-36043305		3.47	false
	s-7564267324-89	27 Jan, 2016, 1:34:31	5	true	\$0.00	Trivia	Pacific		a-16666210		13.19	true
	j-9093134192-779	18 Dec, 2016, 14:31:36	1	false	\$0.04	Poll	EU-west		a-4746032		78.17	true
	s-4498805537-498	27 Jul, 2016, 13:06:43	1	true	\$0.02	Personality	Pacific		k-18240315		156.38	true
	k-6793606482-884	5 Nov, 2015, 8:25:41	1	false	\$0.14	Personality	Atlantic		r-24104512		21.41	true
	j-6330551868-22	11 Feb, 2015, 21:21:20	2	false	\$0.15	Checklist	Atlantic		r-63349878		90.40	true
	r-4854825664-788	15 Dec, 2015, 4:12:30	1	true	\$0.15	Poll	Atlantic		r-19579309		166.27	false
	a-3465812456-418	2 Sep, 2016, 9:47:17	1	false	\$0.12	Poll	Atlantic		a-87803472		25.34	true
	r-9991310916-451	7 Dec, 2015, 12:25:27	1	true	\$0.07	Personality	Can		a-86581096		64.16	true
	a-5597962255-267	17 Nov, 2016, 18:37:36	2	false	\$0.08	Checklist	Pacific		r-42096291		126.11	false
	s-2479790589-888	20 Oct, 2014, 20:03:46	1	true	\$0.04	Poll	Midwest		a-53383421		20.44	true
			2	true	\$0.14	Poll	Can		k-50484171		84.32	false

Connected Sheets + データポータル の SQL レス BI ([詳細はこちら](#))

- コネクテッド シートで BigQuery からデータを取得し、可視化用の加工を行う
- 加工されたデータをデータポータルから参照
- **ビジネスの意思決定に必要な表現を SQL なしで実現**



ネイティブアプリ、Web 解析連携 - グロースハック、デジタルマーケティング目線

/ ワンクリックで BigQuery にデータをストリーミング連携、他のデータとの統合分析が可能

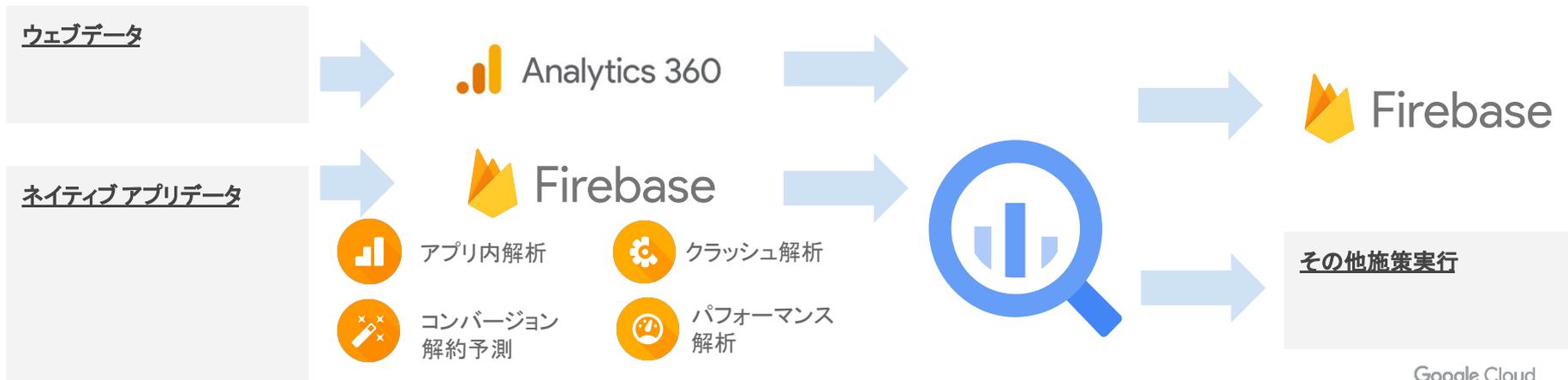
/ **Web** : Google Analytics 360, Google Analytics 4 プロパティ

/ **ネイティブ** : Google Analytics for Firebase, その他 Firebase サービス

GA / GA4F の raw data を export できるのは BigQuery のみ

/ Firebase A/B Testing, Cloud Messaging (プッシュ) などが BigQuery セグメントの定期インポートをサポート

/ 「打ち手」の実行までを自動化、マーケターでもできるように



データサイエンスプラットフォーム連携

/ Jupyter Notebooks から BigQuery を操作できる
Extension による柔軟な連携

/ データレイクハウスの構成:
レイクとウェアハウスのデータ二重もちの解消
相互データ探索

/ フェデレーション

(Cloud Storage、Cloud SQL、Google ドライブの CSV 等)

/ BigQuery Storage API とコネクタ

(Hadoop / Spark などから高速にデータを読み出し
DataFrame に格納できる)

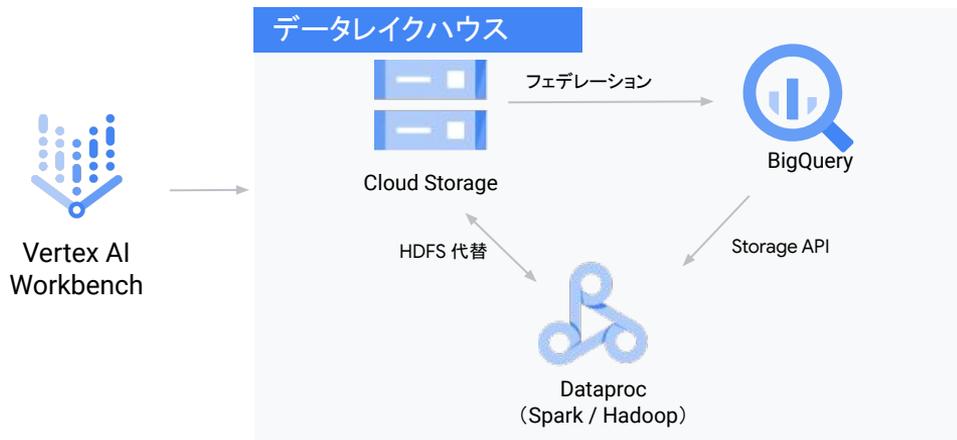
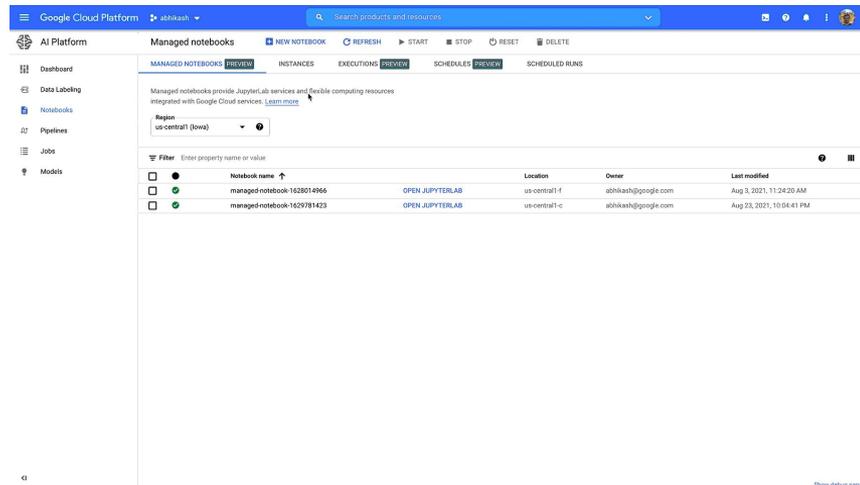
/ BigQuery ML / AutoML

SQL による機械学習モデル構築と予測取得

/ BI ツールから簡単に予測結果が取れる

/ BigQuery のパフォーマンスを生かした分散トレーニング

/ モデル指定の BigQuery ML、モデル選択も



IT 部門の目線で見ると Google BigQuery

エンタープライズ向け
分析用データ ウェアハウス

エクサバイト規模のストレージと
ペタバイト規模の SQL クエリ

セキュア、耐久性
メンテナンスフリー



フルマネージドでサーバレス

Unique

ストリーミング データの
リアルタイム分析

Unique

ML と GIS をビルトイン

Unique

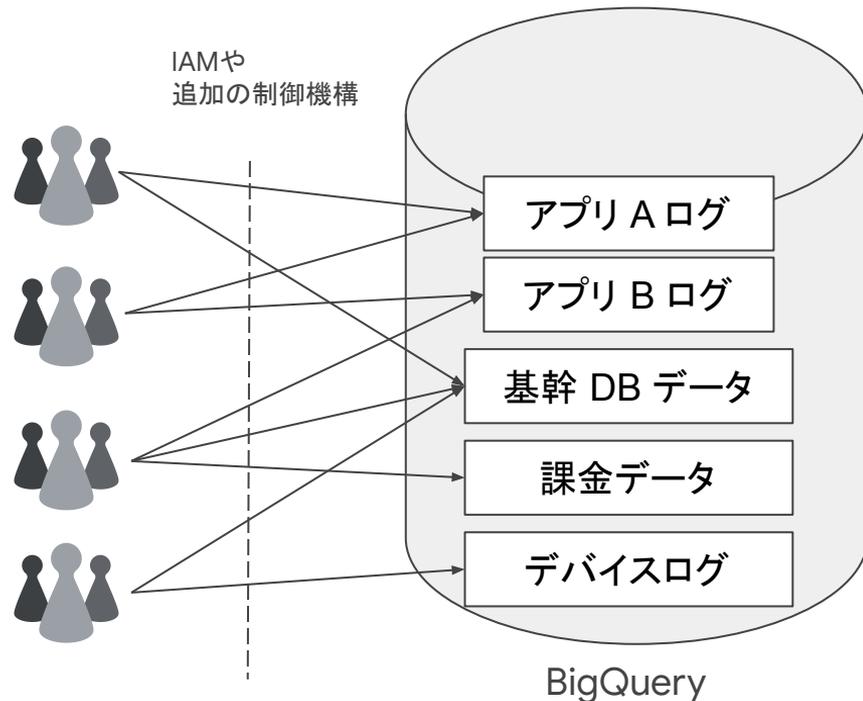
ハイスピードでインメモリの
BI エンジン

Unique

BigQuery マルチテナントの DWH

サイロを作らずデータ活用を促進、トランザクション データでも高速に

- リージョンで1つの仮想的な「箱」
- 権限で制御、社内でデータシェアしたい場合には権限を付与するだけ
- **データコピーなしにデータ共有、JOIN 可能**
- コストもクエリ・ストレージともに従量課金(≠ クラスタ課金)



IT 部門がデータ基盤に対して抱える課題(再掲)

分析環境のインフラ・ETL ジョブの管理が主な関心事

- 将来のデータ増加予測に基づくインフラのサイジングが困難 →サイロ化
→容量無制限でサイジング不要、完全従量課金で事前コミット不要
- ビジネス部門の要件が曖昧で頻繁にかわる、それに合わせた DB のチューニングが辛い
→チューニング・インフラ管理不要、100% 分析に集中できる
- 要件を全て実現するとストレージ枯渇やバッチ突き抜けリスクがある
→ストレージ管理不要・コンピュートリソースは必要に応じて数クリックで増強可能
- セキュリティ・ガバナンスの観点から LoB 側に勝手に BI ツールや DWH を持ってほしくない
→UI・ツールに関わらず一度設定したアクセス権限が適用され、監査ログも自動収集

Google Cloud Smart Analytics の capability



[アクセス性] スキルレベル・用途に合わせて様々なツールでデータを扱いたい

- SQL・Python・Excel・BI ツール
- 各部門がセルフサービスでデータを活用したい



[接続性] 基幹データだけでなく種々のデータソースをサポートする必要がある



[パフォーマンス] データ量の増加は予測不可だがそれに対する高いパフォーマンスが必要

- しかし個別のユースケースに合わせたチューニングはできない



[セキュリティ・ガバナンス] 統一的な管理が行いたい

- データ持ち出し・漏洩リスク
- データマネジメント

Thank you