

東芝テレビ REGZA

全国150万台の視聴データ分析基盤

東芝映像ソリューション株式会社
クラウド事業センター アナリティクス研究開発グループ
菊池 匡晃

2021年2月18日

レグザなら選べる

2020年 ALL NEW 有機ELレグザ、フルラインアップで堂々完成

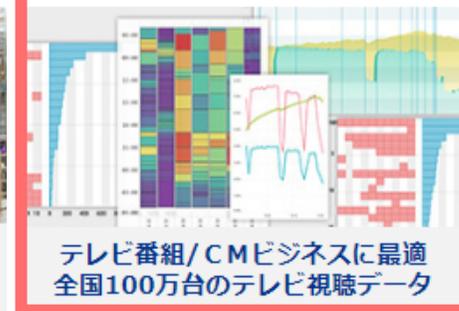


ネットワーク対応テレビ「レグザ」

画質と、ネット連携する高度な番組録画機能が自慢です

<https://www.toshiba.co.jp/regza/>

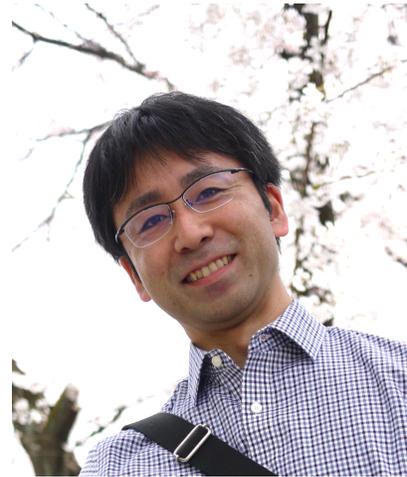
- 東芝ブランドの映像関連事業を承継。
- テレビ「レグザ」、レコーダー「レグザブルーレイ」「レグザサーバ」、「デジタルサイネージ」、「テレビ視聴データ」の各種映像ソリューションを提供。



クラウド事業センター

テレビやレコーダのB2Cクラウドサービス、B2B向け視聴データ分析サービス等の企画・開発・運営

- **アナリティクス研究開発グループ グループ長**
 - テレビ/レコーダからのデータを受信するところから先を担当
 - プレイングマネージャーとして、手も動かしています
- **技術領域**
 - データエンジニア
 - AWS(Redshift, EMR, 他)/GCP(BigQuery)/SQL/Ruby
- **二児の父**
 - 在宅勤務で家族との時間が増えてうれしい
 - 子供と一緒に鬼滅の刃、ポケモンにはまり中



Eight QR

1. サービス紹介とデータ活用事例

B2C 東芝テレビ向けクラウドサービス、B2B 視聴データ分析サービス

2. 視聴データ分析基盤の変遷

EMR、Amazon Redshift、BigQuery

3. Redshift + BigQuery の苦労話

移行、あるいは組み合わせ時に意識すること

1. 東芝テレビ/レコーダユーザー向けサービス

(機器内から呼び出して使用)

[MAU 約160万台に対して提供しています]

あなた**好みの番組を**
レグザが**まとめてアーカイブ**



リアルタイムで見たいスポーツ中継、ずっと応援しているタレント。レグザなら、あなたが見逃したくない番組を、まとめて保存。その中からあなたにぴったりの番組を、AIが教えてくれます。



音声認識 > 意図理解 > 操作/検索



2. 放送局/広告会社向け

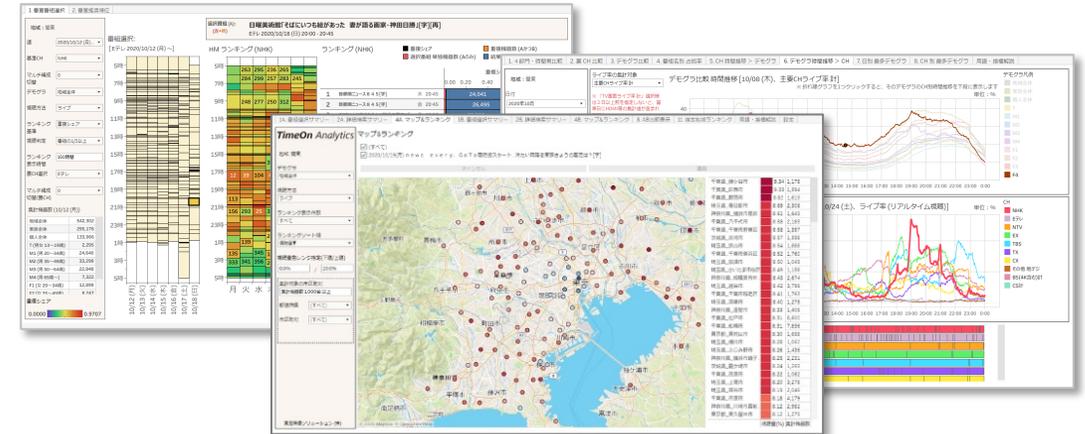
視聴データ分析サービス

TimeOn Analytics

[50社以上でご利用頂いています]



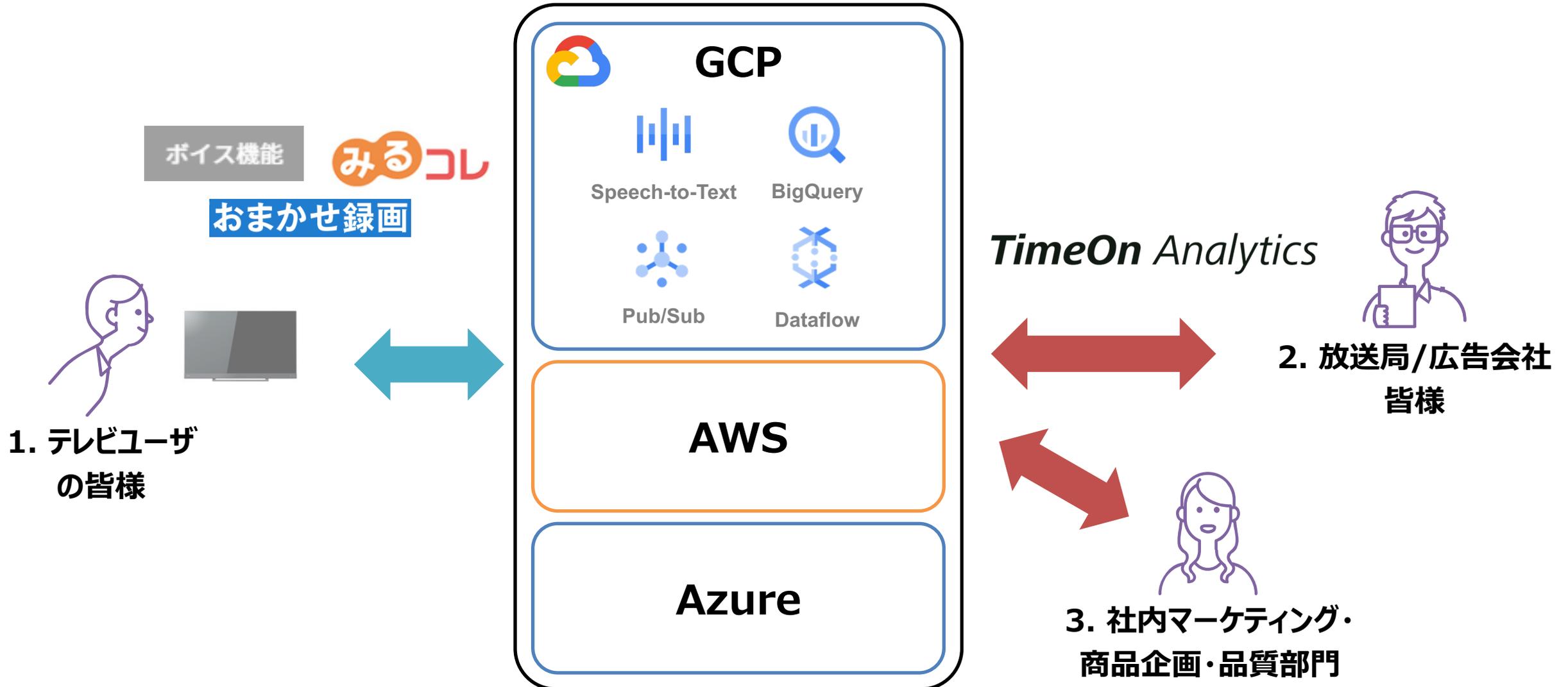
- 47都道府県150万台の**前日の番組視聴を翌朝までに集計・可視化**
- リアルタイム視聴に加え、録画再生視聴も合わせて集計
- 1秒単位**での集計のため、**CM視聴状況も把握可能**

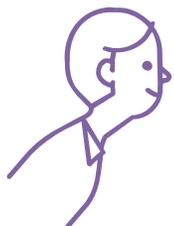
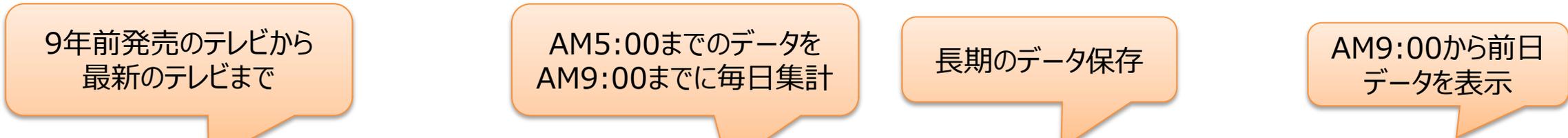


<https://www.toshiba.co.jp/tvs/tvdata/>

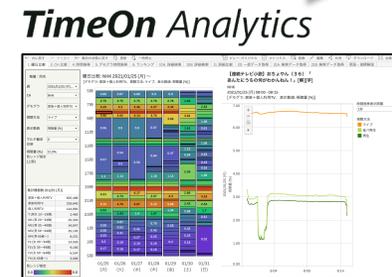
3. 社内マーケティング・商品企画・品質部門向け 機器利用状況分析サービス

2012年から運用中。現在、AWSを中心に、GCPやAzureも併用したシステムになっています。





視聴ログ

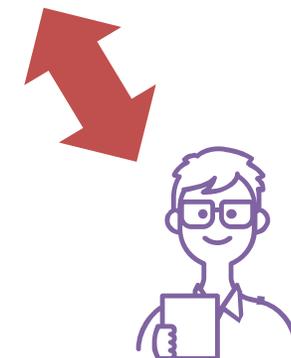


1. データ利用を
許諾いただいた
テレビユーザの皆様

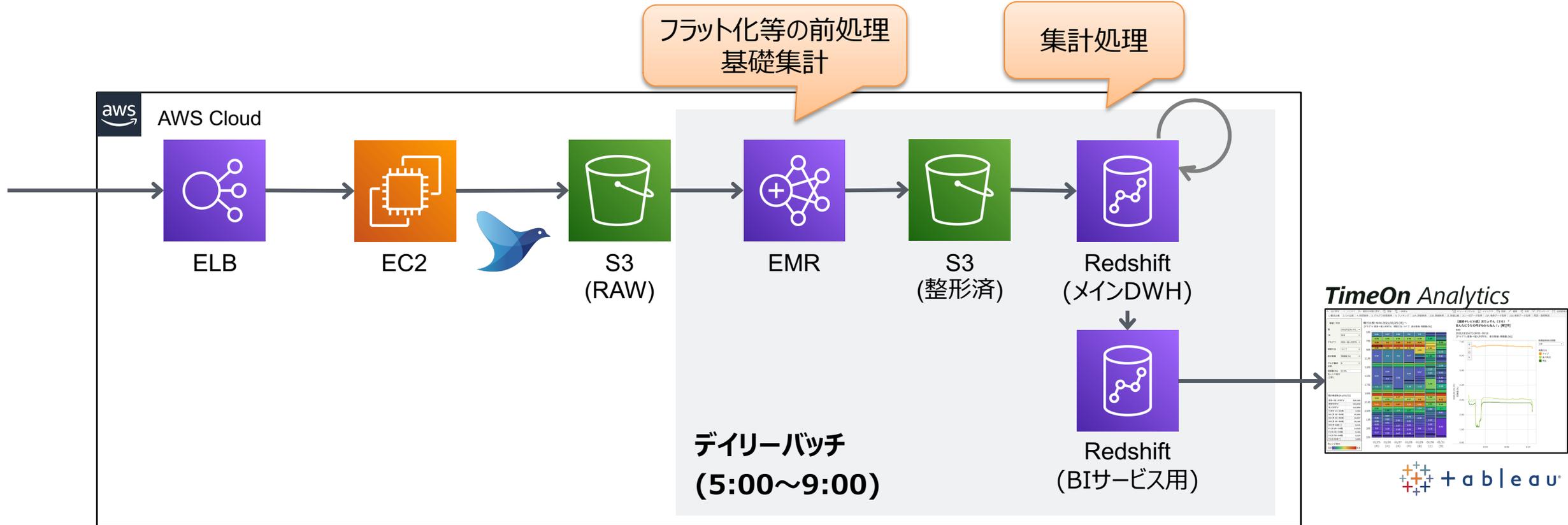
Ch	Begin	End
A局	2021-02-18 18:03:04	2021-02-18 18:10:58
...		

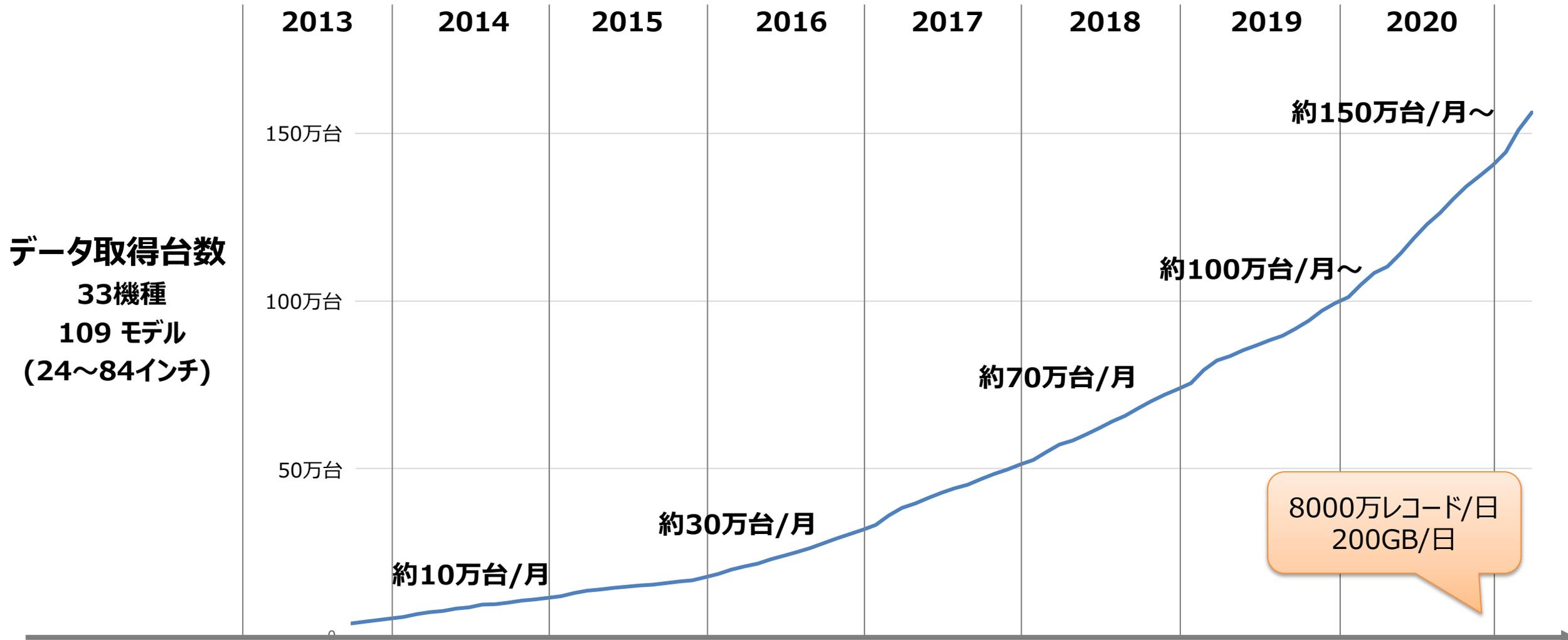


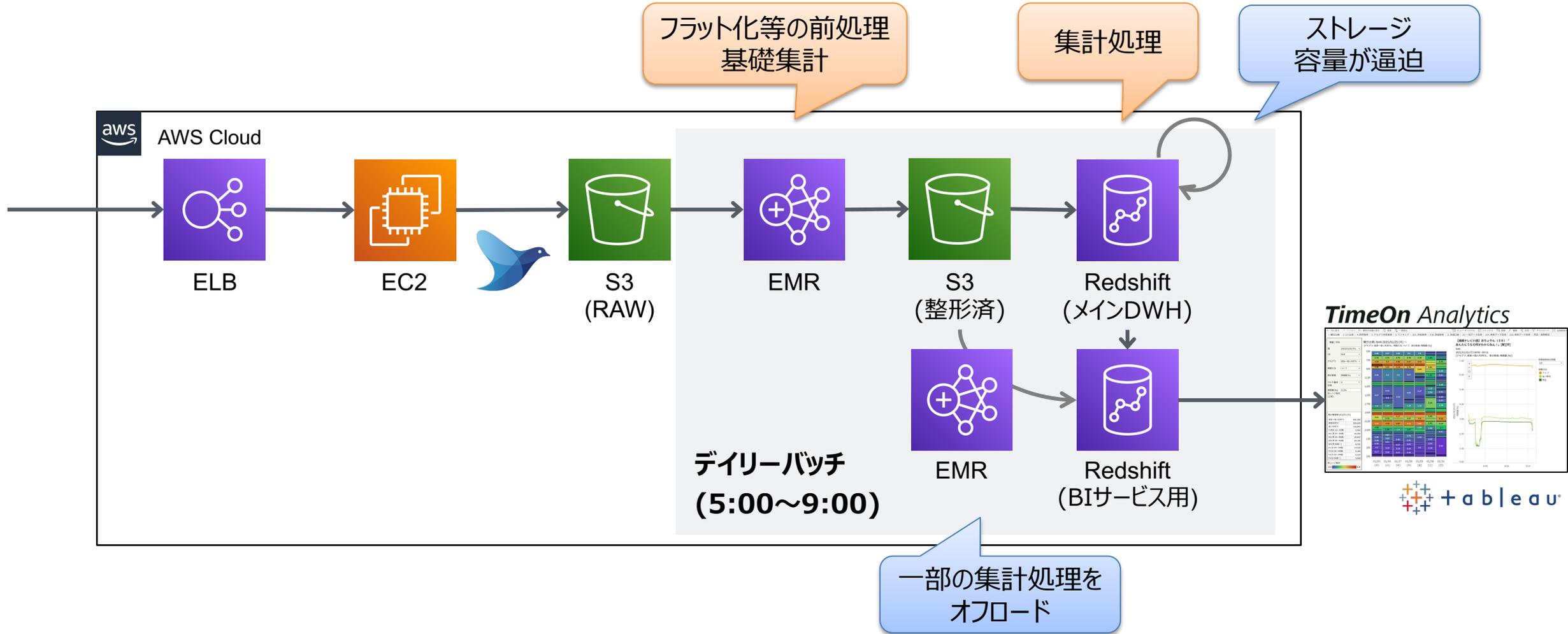
3. 社内マーケティング・
商品企画・品質部門



2. 放送局/広告会社
皆様



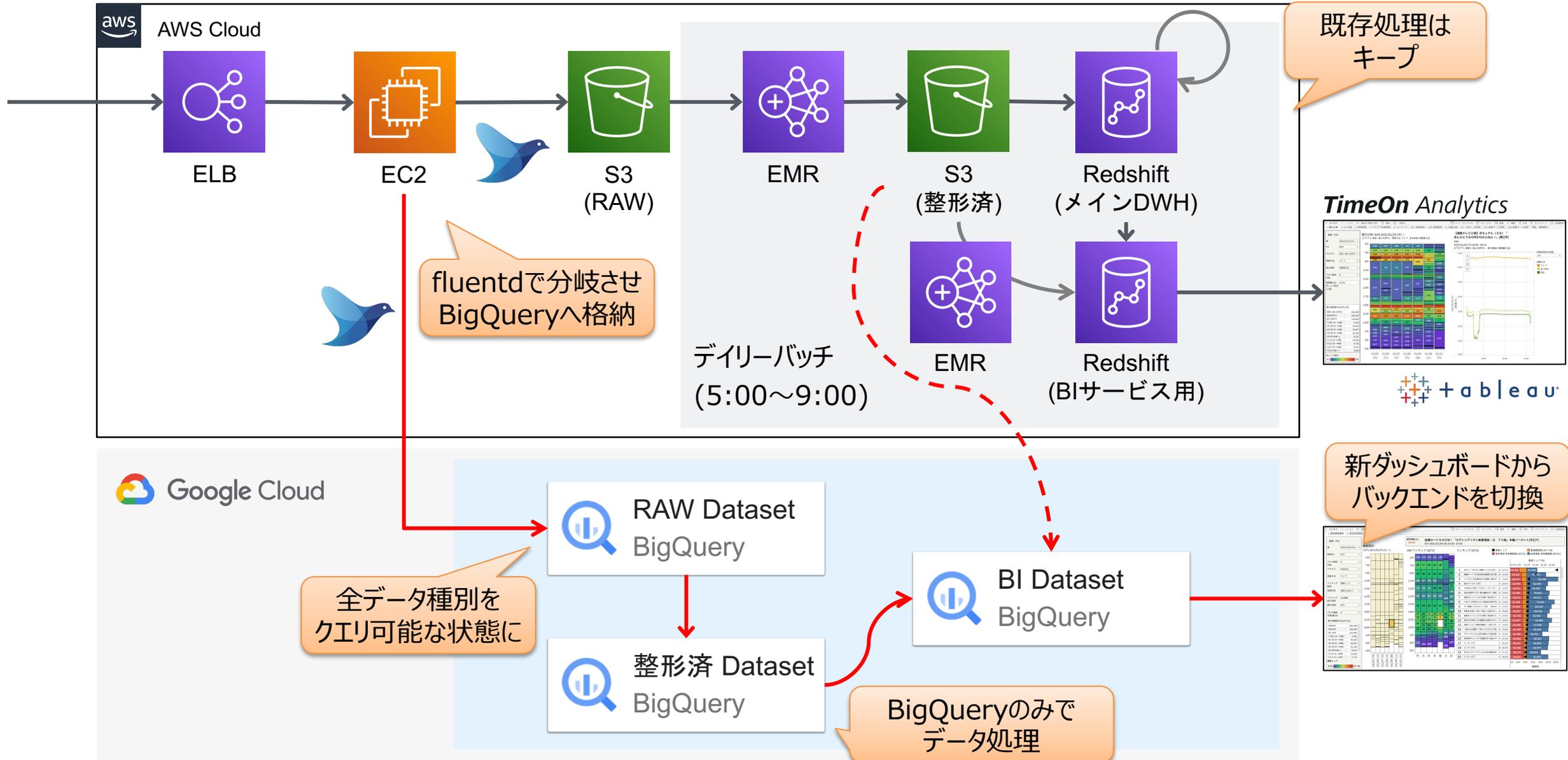




- 課題1: 複数の構成要素が複雑に入り組んでいる
 - Redshift SQL, HiveQL / Spark on EMR
- 課題2: Redshift の料金体系
 - クラスターサイズによる買い切り料金
- 課題3: Redshift の運用
 - リサイズに長時間かかる

→ **BigQuery に置き換える**

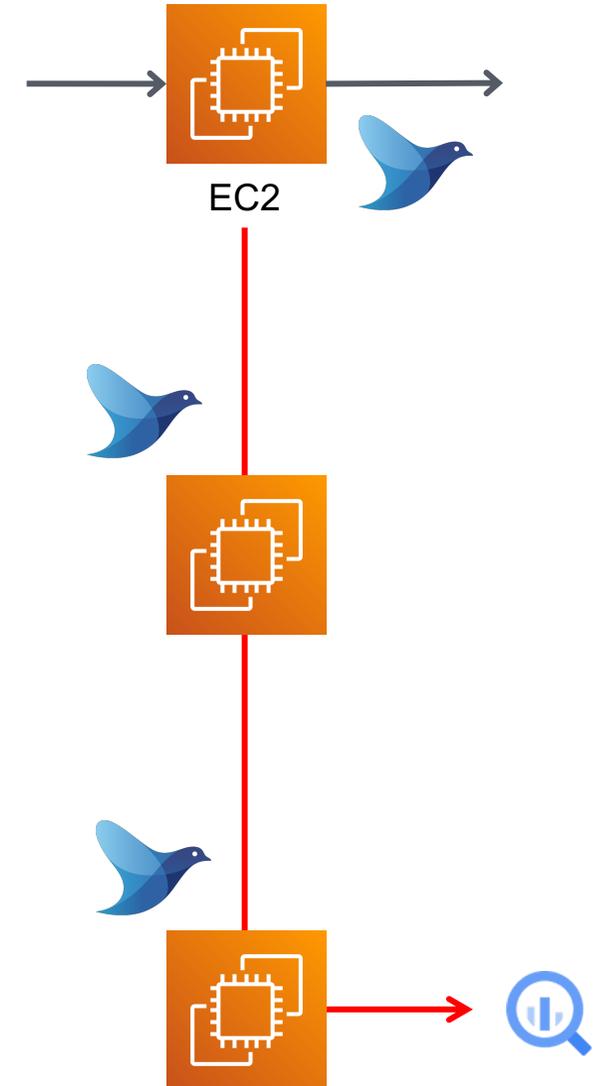
データ分析基盤の構成 (BigQuery移行中)



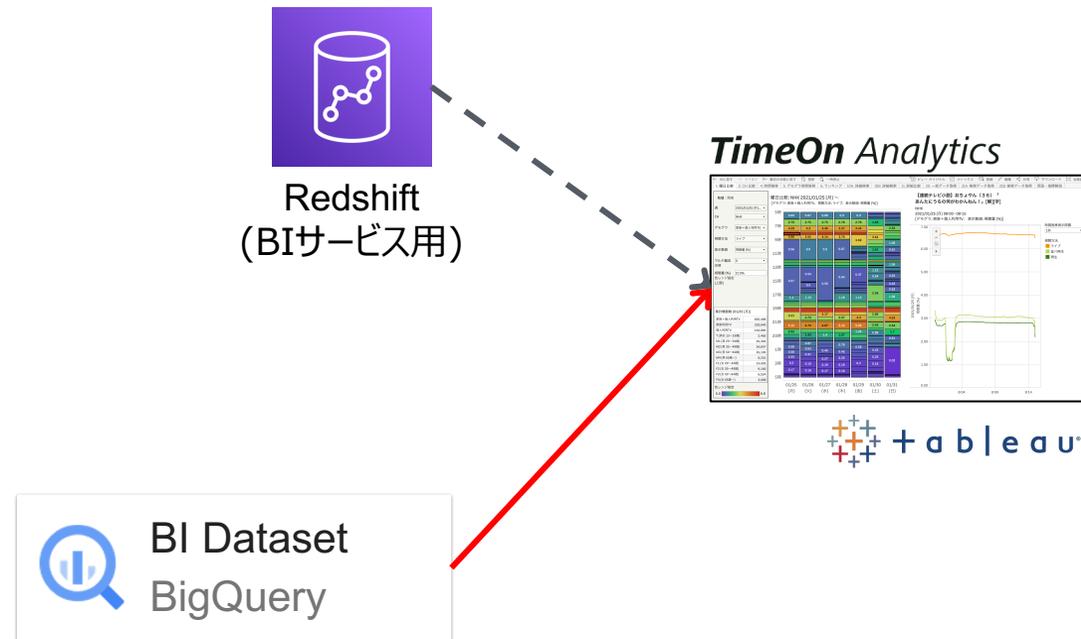
- **BigQuery のみの構成要素で、シンプルに**
 - 前処理なしで取り込める
 - BigQuery に処理を集約できる
- **参照頻度の低い過去データを保持するコストが低い**
 - 最新データまでシームレスに扱うことができる
- **運用作業不要**
 - ストレージ上限なし

- **データの社内活用が大幅に広がった**
 - 死蔵していたデータを格納
 - RAWデータをクエリ可能にするまでが簡単
 - Repeat型、Struct型、String型
- **速報型のデータ分析サービスの提供が可能に**
 - Stream Insert でほぼリアルタイムにデータを触れるように
- **クエリ料金の心配にもガードレールあり**
 - パーティションフィルタ必須オプション、課金バイト数制限オプションで安心
- **Redshiftと並行運用してもお財布に優しい**

- 多段 fluentd の制御が難しい
 - AWS 内の従来処理に影響しないように分岐させた
 - しかし、
 - fluentd やプラグインの更新でボトルネックが変わる
 - サーキットブレーカーが機能しなかった
 - 解析を諦めて毎回本番相当の流量で負荷試験を実施
 - Pub/Sub 等を挟む構成を推奨



- BI ツールのバックエンド切り替え
 - 同じテーブル構成でもチューニングの仕方が異なる
 - パーティションキー、クラスターキーが効くデータの持ち方



- **BigQuery : スクリプティングをもっと便利に**
 - SQL で操作を記述できるように
 - Google Cloud Storage からデータロード
 - Parquet 形式で Export
 - テーブルの Rename
 - データセット名のパラメタ化
 - 同じクエリをデータセットを切り替えて実行したい
 - Staging と Production など
 - 料金体系の維持、値下げ
 - 特に課金方針を変更されると厳しい

- **東芝テレビのデータ分析基盤の事例と変遷を紹介**
 - AWS + Google Cloud でサービスを組み合わせた基盤を9年間運用
 - BigQuery 導入のメリット
 - BigQuery により構成要素がシンプル化
 - 従量課金であることがデータ利用スタイルにマッチ
 - RAWデータがクエリ可能になり活性化
 - 速報型サービスへの展開
 - 既存のデータ基盤と併用しながら移行

コンテンツ系データ分析を自らいじり倒したい方募集！ (分析基盤は既にあります) データサービス開発、機械学習によるインサイト発見、など、応用分野取り組み中です



フレックス制 (コアタイム無)、在宅勤務、土日祝日休、有休20日以上、気軽な Chat/mtg. 環境
(コロナ禍以前からリモートワーク可能、現在、在宅勤務率9割以上、柔軟な勤務が可能)

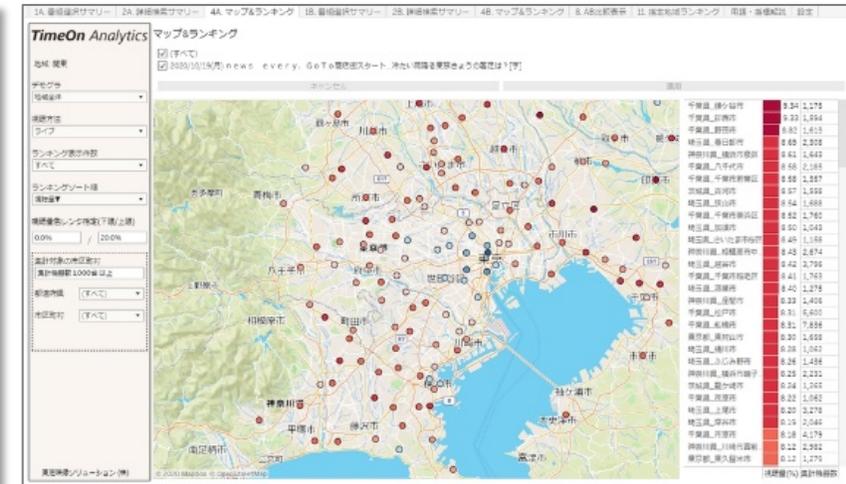
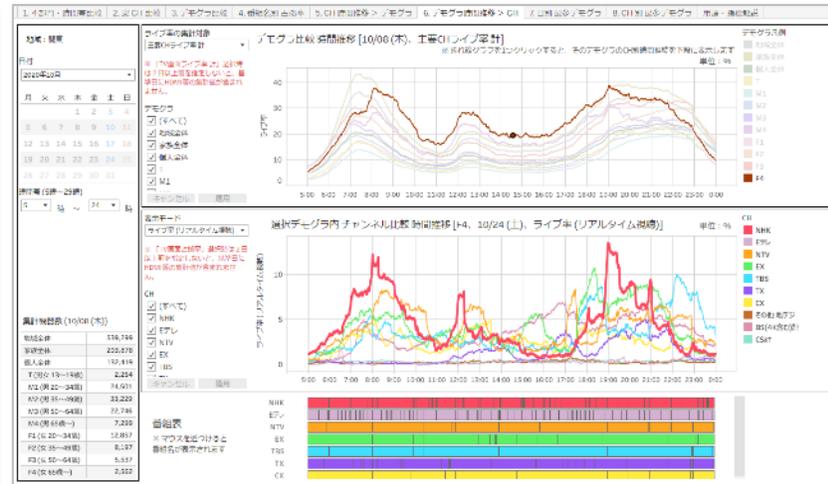
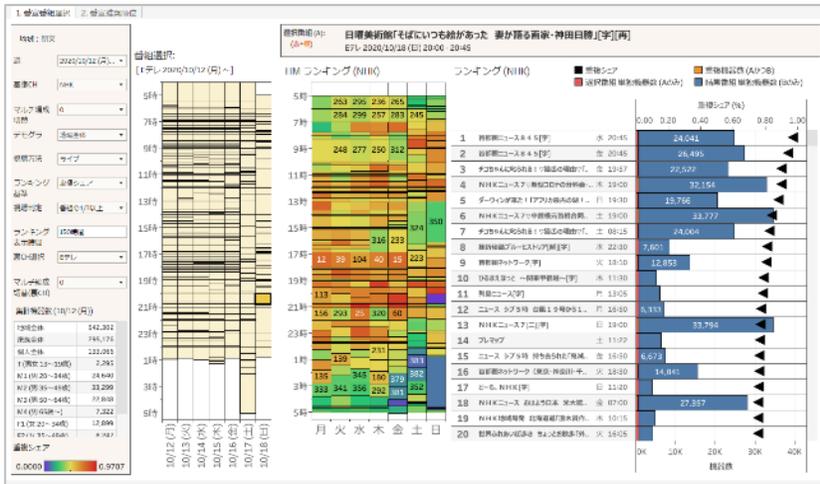
出勤する際の場所：JR新川崎駅 (横須賀線 / 湘南新宿ライン) 徒歩3分、JR鹿島田駅 (南武線) 徒歩2分

Eight QR



気軽にご連絡ください

masaaki.kikuchi@toshiba-visual.com



REGZA

TimeOn Analytics