

E-4

データドリブンな組織を目指す AWSを活用したデータ分析基盤の取り組み

多田 貞剛 @tada_infra
SRE
株式会社スナックミー



- 多田 貞剛(@tada_infra)
- 2020年9月株式会社スナックミーに中途入社
- SRE として業務に従事
- 筋トレ -> サウナ -> サ飯 のトライセットを決めるのがマイブーム

会社紹介

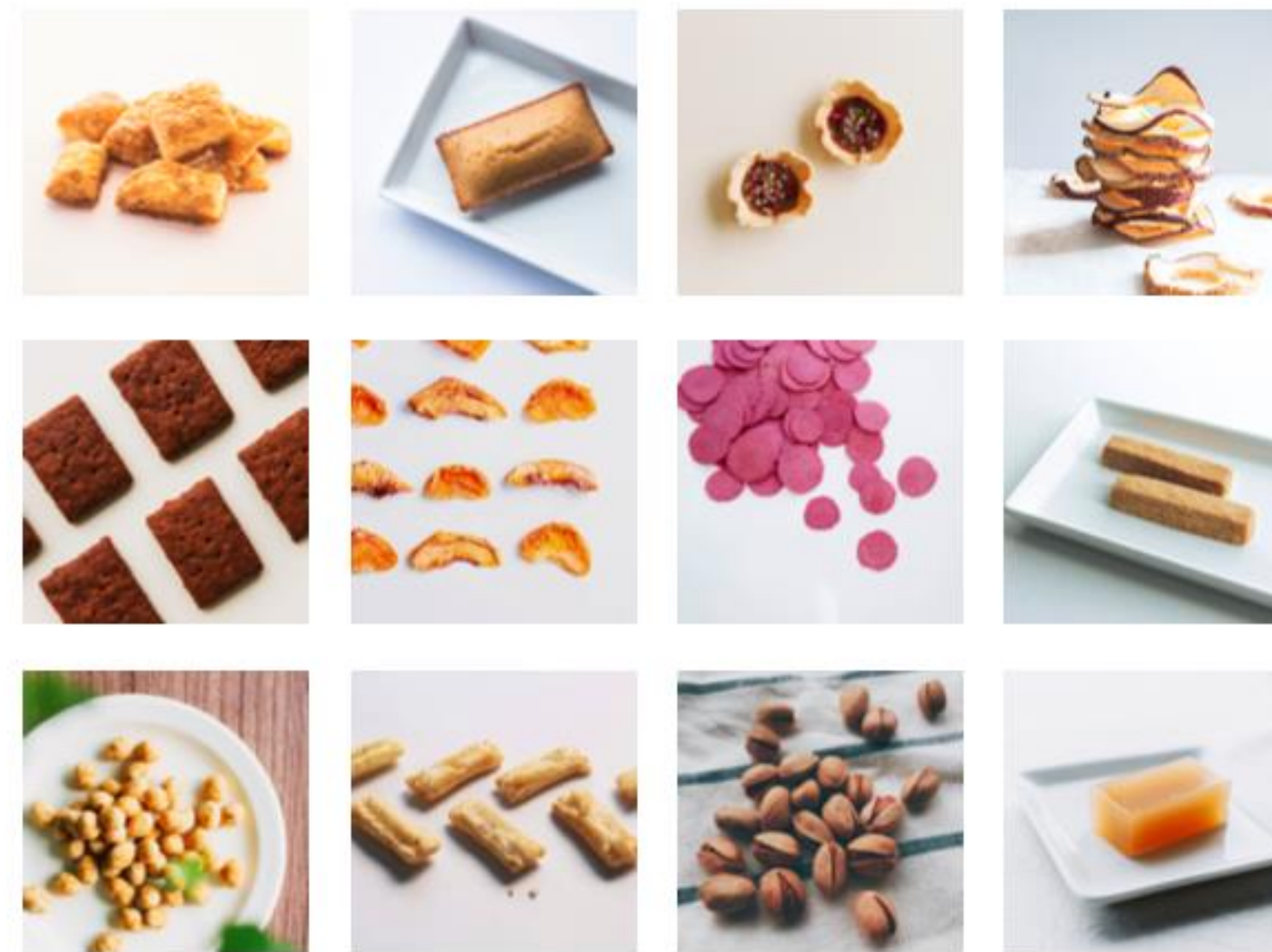


s n a q . m e

『新しいおやつ体験を創造し、おやつ時間の価値をあげる』

栄養価が高く、それでいて美味しさに妥協しないスナックによって、おやつそのものの質を上げる。
また、モノだけではなく、新しいおやつ体験をデザインし、おやつ
の時間の価値を向上させる。
それによって、おやつ時間が彩りとなり、豊かな生活を実現する。

- おやつ体験 BOX 『snaq.me』
 - 月額 1,980円(税込、送料込)
 - 4週 or 2週毎に**100種類以上**の商品からお客様にパーソナライズした8種のおやつをお届け



おやつ体験 = 7つのme



ぴったりme

パーソナライズ
してもらえる



いろいろme

いろいろな種類を
たくさん試せる



わくわくme

何が入っているか
わからないので開ける時に
ワクワク



はっけんme

知らなかった食べ方など
おやつの新しい楽しさに
出会える



あんしんme

体に不要なものは
入っていないので
安心できる



おいしいme

地方の美味しい
お菓子が食べられる



らくちんme

サブスクリプションで
受け取れる

おやつ診断、商品リクエスト、商品評価などのデータを取得してお届けする、パーソナライズされたおやつが特長

おやつ診断

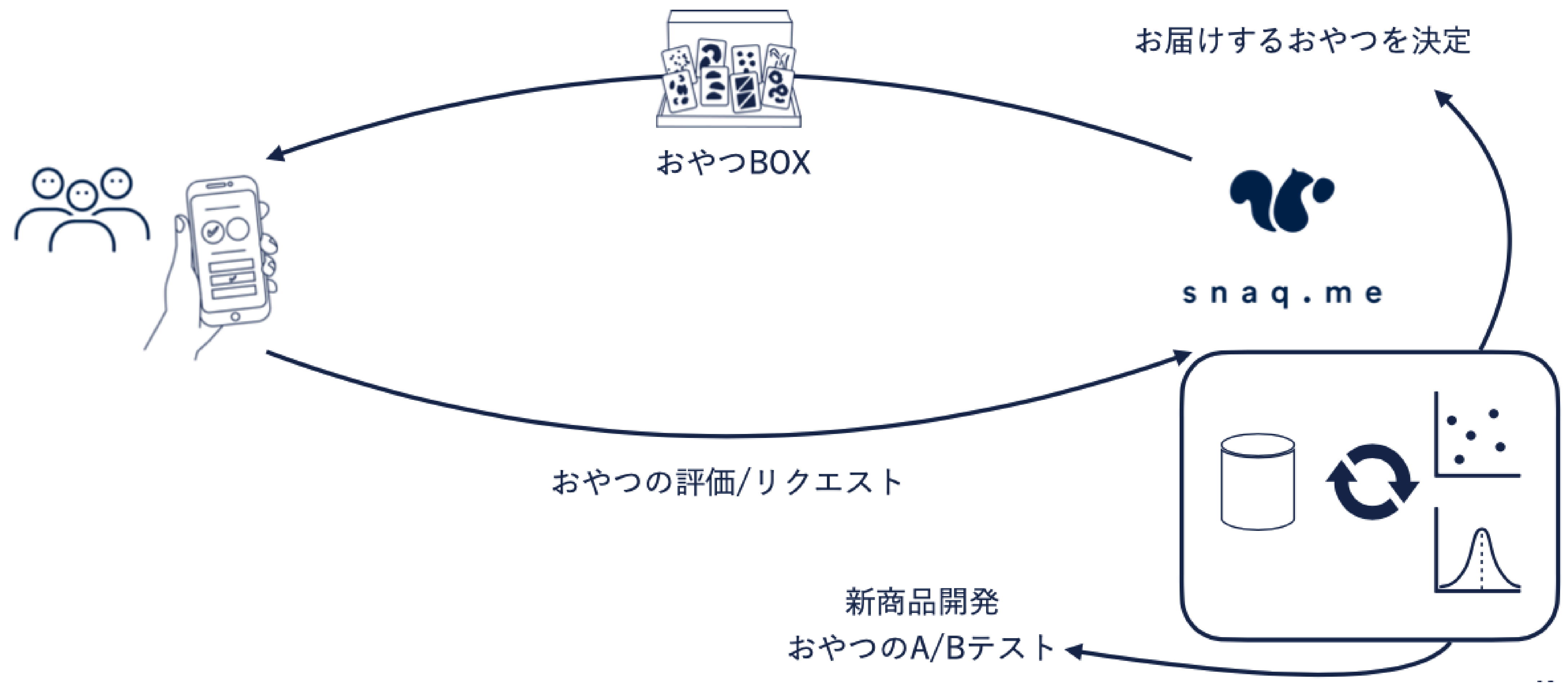


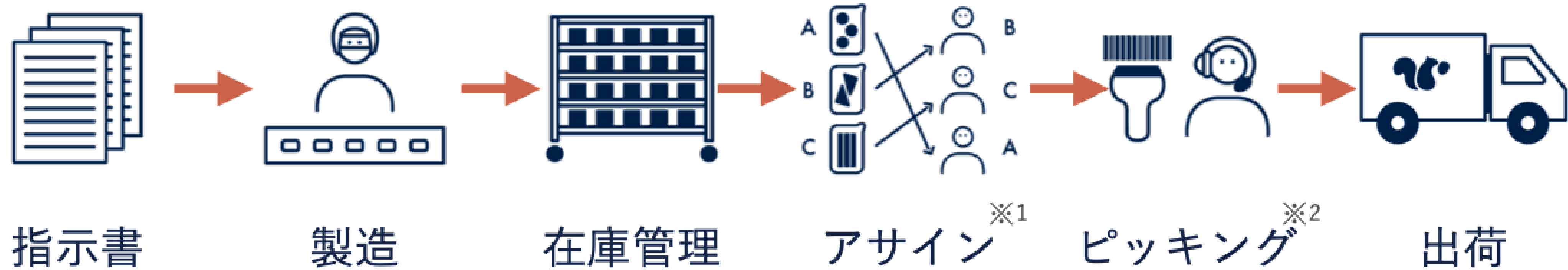
評価



リクエスト







※1 アサイン：お届けするおやつを選びこと

※2 ピッキング：配送するおやつをBOXに入れること

本日本話すること

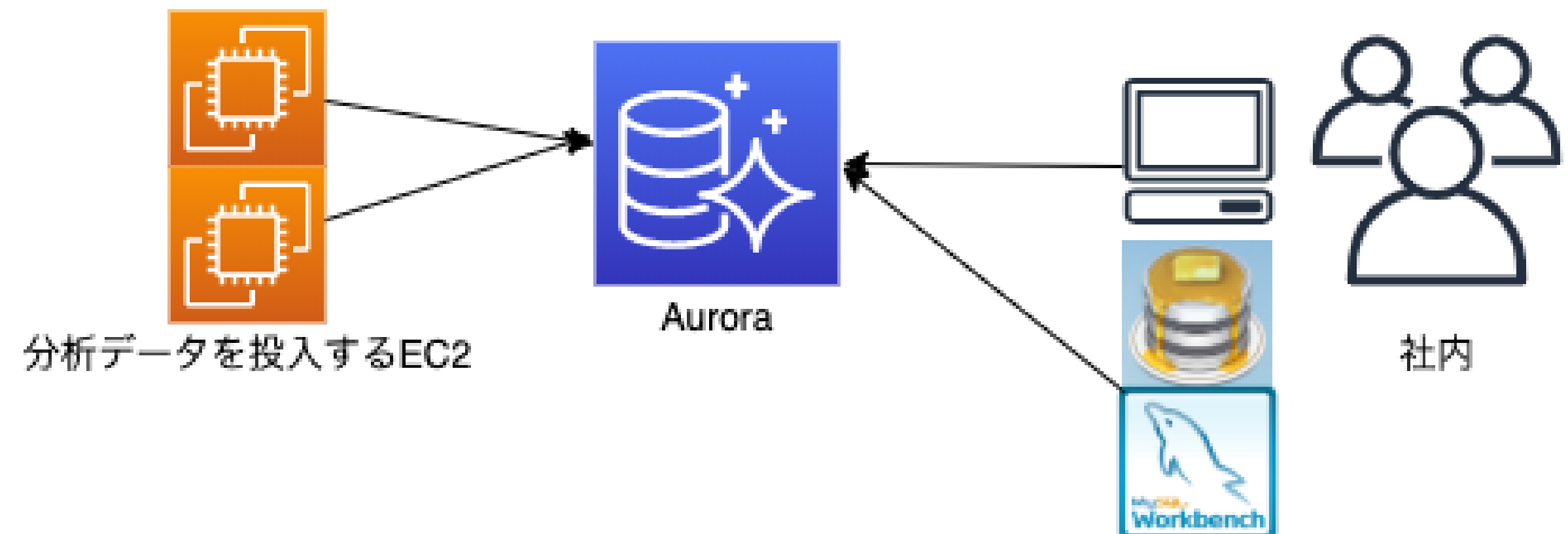
- 会社のデータにまつわる課題
- 課題に対する取り組み状況
- データドリブンな組織を目指すためのデータ基盤の形

本日はお話ししないこと

- データ基盤の導入フェーズであるため、導入以降のデータ基盤の運用
- データ基盤チームの運営

スナックミーのデータにまつわる課題

- 弊社のデータ利用の現状
 - スナックミーではサービスに関する様々なデータをデータベース(Aurora MySQL)で扱ってる
 - データを使って KPI、マーケティング、製造・配送の効率など各部門で分析や業務活用を行なっている
- データ閲覧は社内データ閲覧サイトやSQLクライアントで適宜行なっている





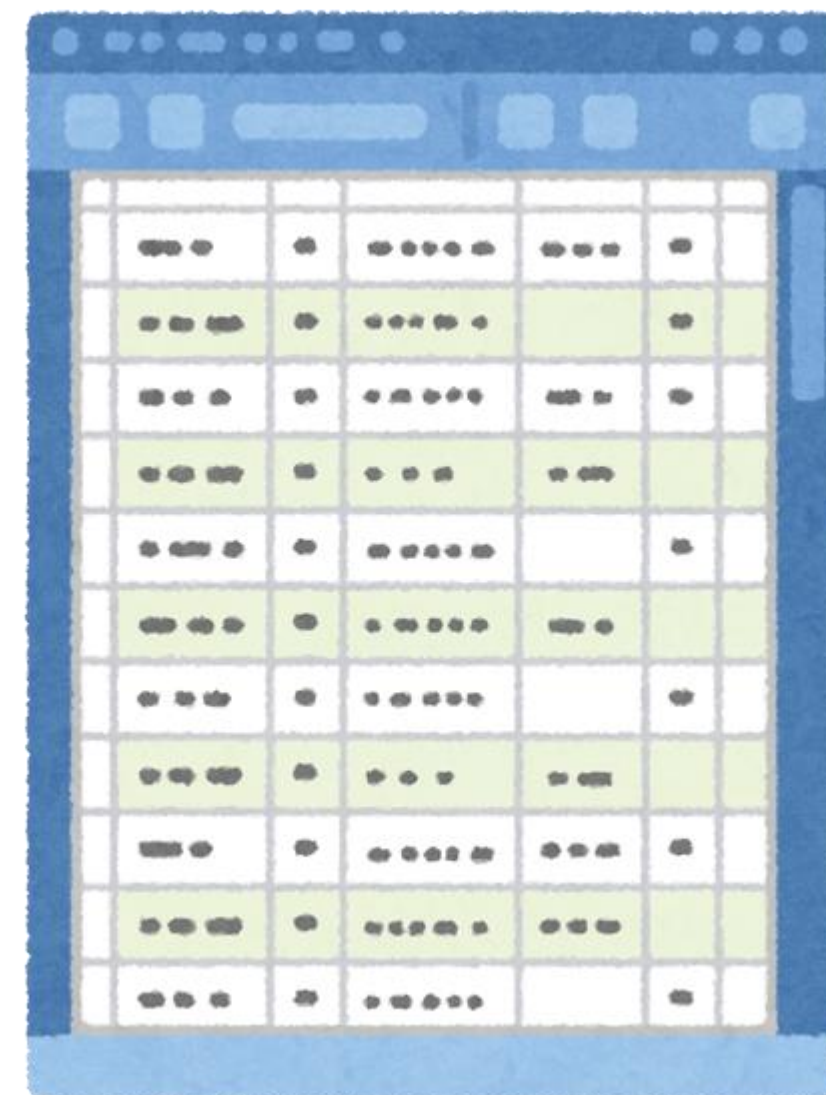
- 週次の全体 MTG で KPI に関わる数値やお客様の声を共有
- データを活用して今後の施作やマーケティングのアクションを決めている

- 大きく3つの課題があった
 - ①社内サイトのデータ閲覧における課題
 - ②業務で必要なデータが整え切れてない課題
 - ③データ分析業務における SQL の課題
-

- 社内サイトのデータ閲覧における課題
 - 欲しいデータへのアクセスに手間がかかる
 - 閲覧したいデータを表示するのに時間を要す
 - 表示データが誤っているのを修正できてないためそのページが見られない



- 業務で必要なデータが取れてなく、個々の独自スプレッドシートでデータが管理されて属人化
- 仮に退職した場合にその人しかわからないことが発生する



秘伝のタレ的
スプレッドシート



- データサイズが大きくなっていることで SQL クエリをかけたも長い場合は1時間以上クエリの結果が帰らない
- データベースはサービスでも使っており、負荷が高い状態が連日続く危険な状態

aws APP 12:53 PM

CloudWatch Alarm | [REDACTED] | ap-northeast-1 | Account: [REDACTED]

Threshold Crossed: 1 out of the last 1 datapoints [99.39619914086735 (05/10/20 03:51:00)] was greater than or equal to the threshold (70.0) (minimum 1 datapoint for OK -> ALARM transition).

Alarm State	Namespace
ALARM	AWS/RDS
Metric	EngineName
CPUUtilization	aurora

Oct 5th

aws APP 10:50 AM

CloudWatch Alarm | [REDACTED] | ap-northeast-1 | Account: [REDACTED]

Threshold Crossed: 1 out of the last 1 datapoints [88.0327868840245 (07/10/20 01:48:00)] was greater than or equal to the threshold (70.0) (minimum 1 datapoint for OK -> ALARM transition).

Alarm State	Namespace
ALARM	AWS/RDS
Metric	Role
CPUUtilization	WRITER
DBClusterIdentifier	
[REDACTED]	

Oct 7th

データの課題に対する取り組み

- 課題に対する取り組み状況
 - ①社内サイトのデータ閲覧における課題
 - ②業務で必要なデータが整え切れてない課題
 - ③データ分析業務における SQL の課題
-

- 課題に対する取り組み状況

- ①社内サイトのデータ閲覧における課題

- **BI を作ってデータを統一的かつ即座に表示する**

- ②業務で必要なデータが整え切れてない課題

- ③データ分析業務における SQL の課題

- **Athena と Aurora AutoScaling の利用**
-

社内サイトのデータ閲覧への課題対応

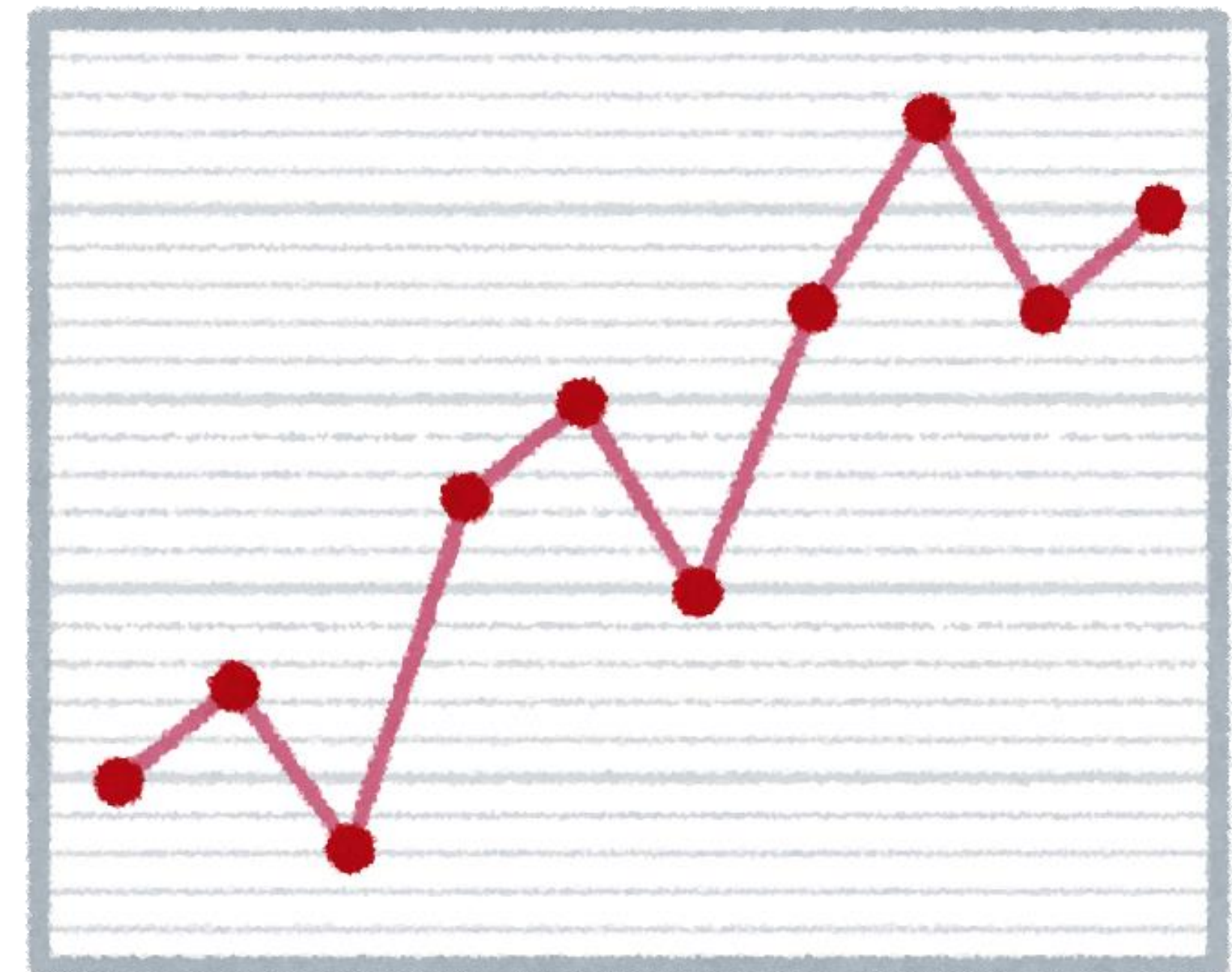
- 会社で追う指標と個別チームメンバーが業務で見たいデータを集約したページを BI として提供
 - データベースを参照しなくても欲しいデータへすぐアクセスでき、且つ業務効率化への寄与も狙える

- 各部署のデータ活用状況と内部インタビューを行った
 - 業務の中でデータ周りで困っていることやどんなデータを見て業務を行っているか等を確認
 - 特に社内サイトを業務で活用しているのがオペレーションチーム



- データを表示するための課題がいくつもあった
 - 個人情報が入ったデータの取り扱い
 - 表示したいデータの更新時間
 - インタビューした時の内容が時間がたって変化し、欲しいデータが変わっていた
-

- BI に載せた指標の一例
 - サービス全体の売上推移
 - ユーザー数の推移
 - 商品ごとの売上状況 etc



- BI を作ってみて感じたこと
 - 利用者にプロトタイプをみせて会話する
 - 利用者が業務でいつ、どんなデータが存在していればよいかを確認する
 - 経営層との認識すり合わせは必ずミーティングの場で全員で意識統一する
-

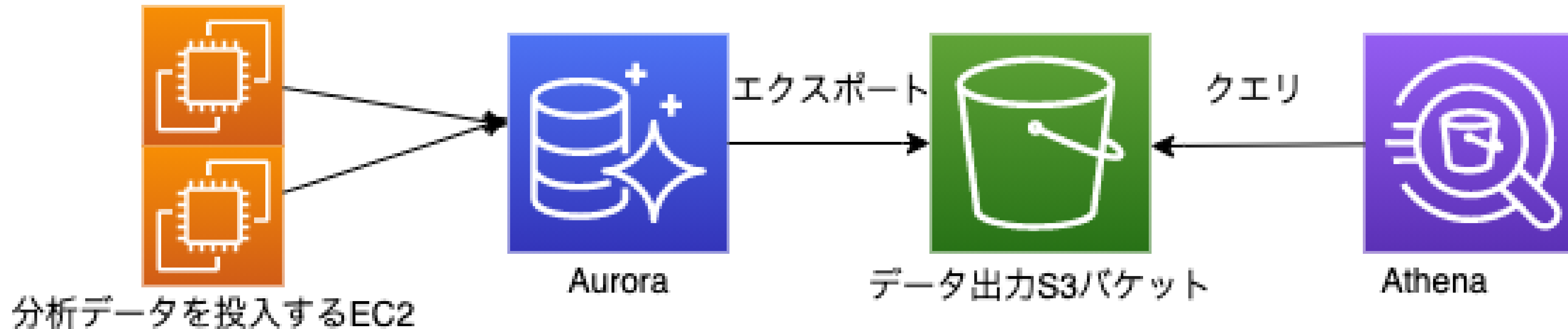
データ分析業務における SQL の課題対応



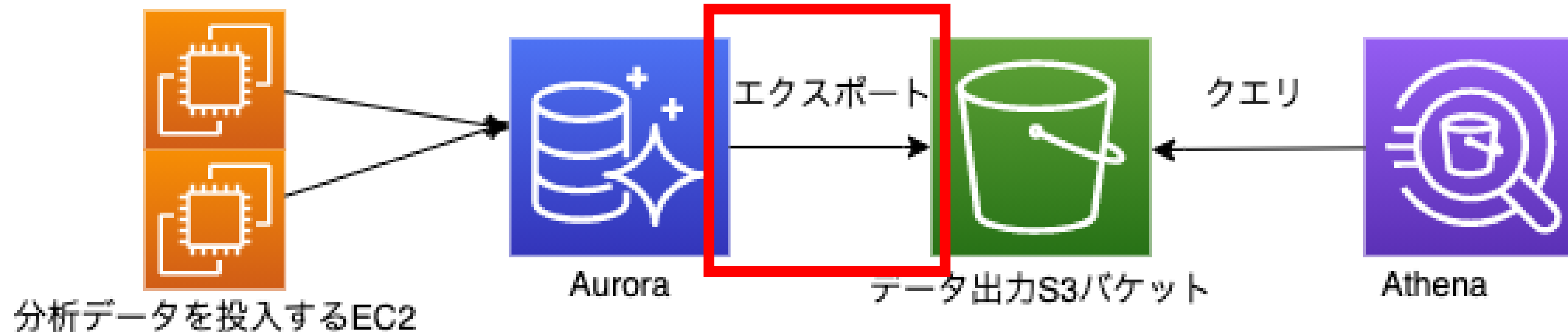
Amazon Athena

- 標準 SQL を S3 のデータに対して発行できる
- サーバーレスでインフラの構築・管理は不要
- CSV、JSON、ORC、Avro、Parquet のファイル形式に対応

- 処理が重い SQL クエリは Athena を使って対応
- Aurora のスナップショットからデータを S3 に Parquet 形式でエクスポートし、分析業務において Athena でクエリをかけるよう変更



- Aurora のスナップショット からのデータエクスポート は時間を要するのでスピードを重視する場合はこのパターンは避けた方が良い
- データサイズによるもののおおよそ2時間ほどかかる



- Athena に変更した効果としてデータベースの負荷が軽減
- データベースに直接 SQL を投げた時1時間かかっていた処理が**10秒**以内で完了するようになった



クエリの実行

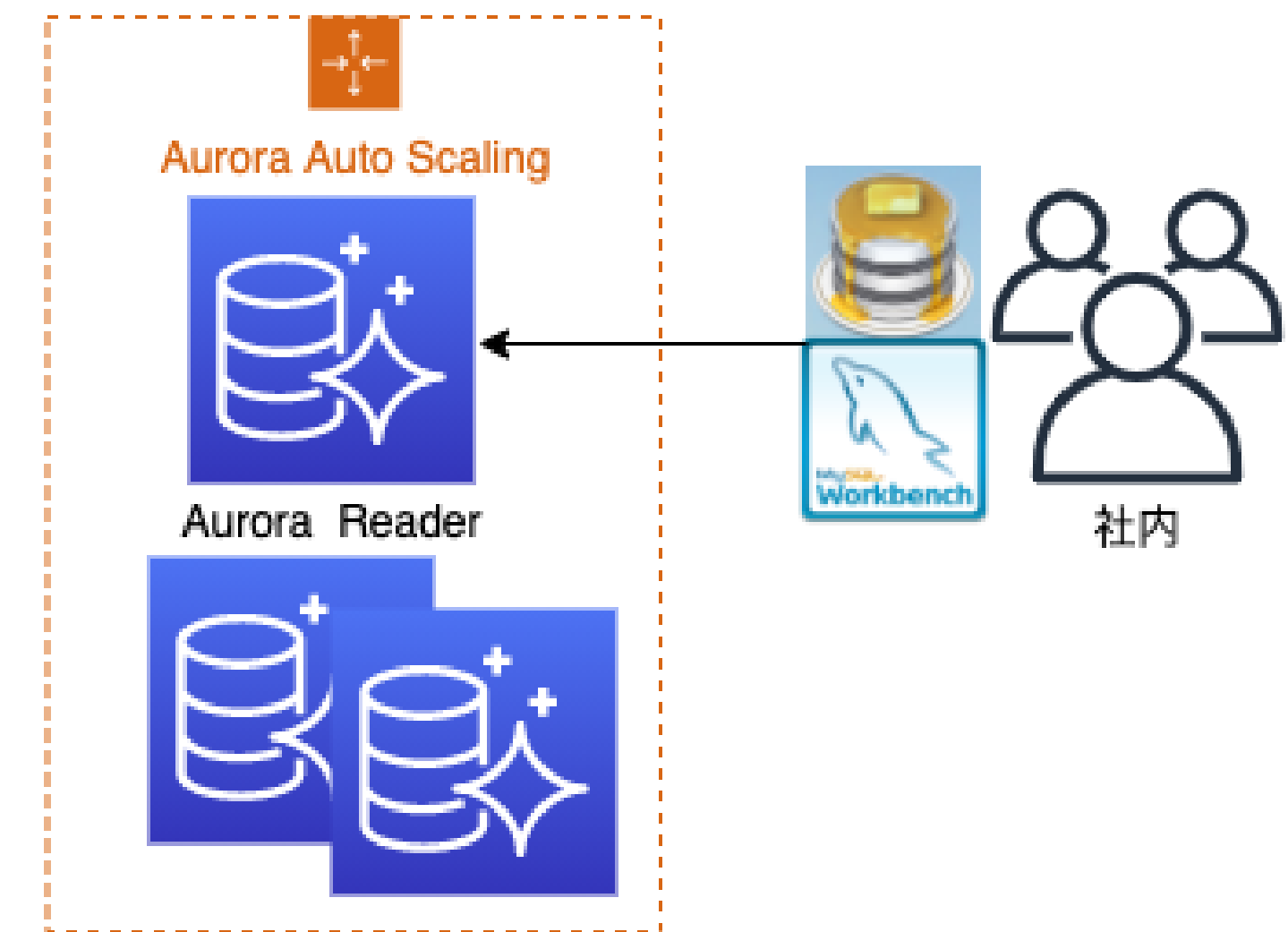
名前を付けて保存

作成

(実行時間: 3.97 秒, キャンセルしたデータ: 926.46 MB)

クエリの実行には Ctrl + Enter、オートコンプリートには Ctrl + Space を使用します

- Athena には順次クエリを置き換えているものの全ての分析用クエリを置き換えられてない
- 置き換えられてないクエリは Aurora AutoScaling を設定した Aurora の Reader に対してクエリを発行
- Aurora の負荷を気にせず分析可能



今後目指していききたいデータ基盤の形

- 属人化したデータや勘や経験に依らない意思決定をサポートするデータ基盤にしていきたい
 - 今後の活動として次のことを考えている
 - ①不足してるデータの収集と正確でないデータを整える
 - ②データ基盤に関する運用を各部門と握る
 - ③機械学習機能が搭載されたサービスの活用
-

- 不足してるデータの収集
 - データベースのテーブルを設計して収集
 - 入力手段を代替してデータをデータベースに収集
 - 正確でないデータを正規化
 - SQL ロジックを今のデータに合わせて変更
-

- データ基盤に関する運用の取り決めを利用部門と握り、運用フローを構築
 - データの表示内容、何時にそのデータが必要で、出ない場合の業務影響範囲を確認し、品質目標を合意する
 - 品質目標が達成されない場合の対応を整理する
 - 定期的に品質目標の達成度を計測してチェックする
-

- 弊社では過去データから近い未来はどのようにデータが推移していくのかも知りたい
 - ユーザーデータは日々刻々と変化しており、データを揃えつつ未来の予測を参考に戦略を立てたい
 - AWS の機械学習機能が搭載されたサービスの活用を今後のデータ基盤で検討していきたい
-



Amazon QuickSight

- 機械学習機能が搭載されたサービスの活用として QuickSight を検討している
- QuickSight の ML Insights を活用して商品の注文数や売上げの予測推移を BI として出していきたい
- 会社の週次ミーティングで BI として使ってもらえるようにしていきたい

まとめ

- 誤ったデータや収集できてないデータ等を整備しつつ BI 活用を促進して、データによる意思決定をサポートしていく
- 将来の予測を立てる参考値として QuickSight の利用を視野に入れる
- クエリが重い処理は Athena に置き換える価値あり Aurora AutoScaling も分析業務に活用中
- データを出すだけでなく、基盤の品質を向上させるために運用の決め事も行っていく

- Amazon Athena
 - https://docs.aws.amazon.com/ja_jp/athena/latest/ug/what-is.html
 - Amazon Aurora Auto Scaling
 - https://docs.aws.amazon.com/ja_jp/AmazonRDS/latest/AuroraUserGuide/Aurora.Integrating.AutoScaling.html
 - Amazon QuickSight
 - <https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>
 - データマネジメントが30分でわかる本
 - <https://www.amazon.co.jp/dp/B085W4YSZJ>
-

A group of five people (three men and two women) are standing on a waterfront promenade, smiling and looking towards each other. They are positioned behind a metal railing. In the background, there are several modern buildings and a body of water under a clear sky. A semi-transparent white box is overlaid on the image, containing text and a logo.

Thank You

We're hiring

新しいおやつ体験を創造し、おやつの時間の価値を高める

<https://engineers.snaq.me/>



snaq.me

Thank you!

多田 貞剛 @tada_infra
株式会社スナックミー