



Athena

指向アナリティクス

真面目に手を抜き

価値を得よ

JapanTaxi Co.Ltd.

技術基盤チーム

梅田昌太 @ebisusurf

**Japan Taxiという
タクシースターの会社から
来ました**





navigation

・我々のサービスの概要と歴史

・Japan TaxiがAthenaを選ぶ理由

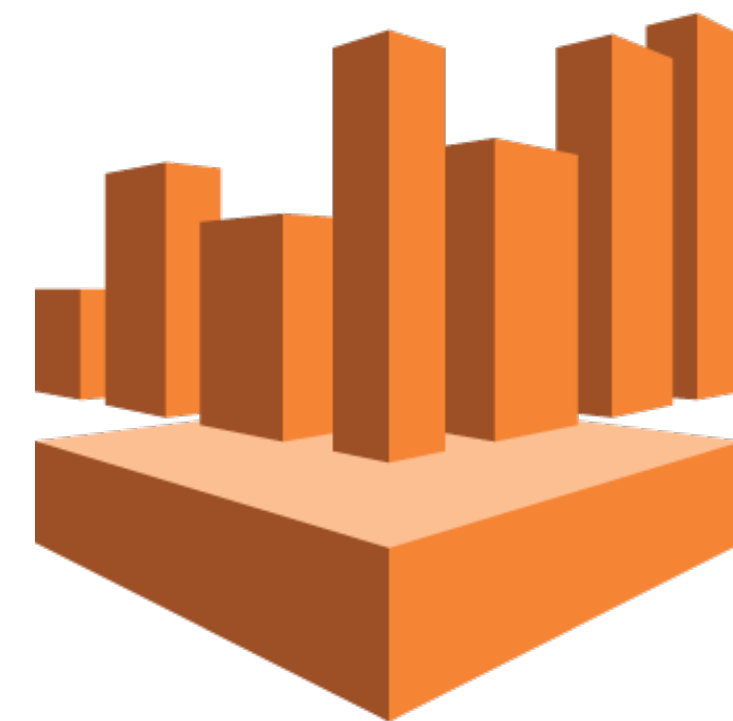
・Japan TaxiでのAthena使用例



About me

- 会社 -> インフラメイン
- 家 -> 3児の父
- 好き -> ログ, 乗り物





我々のサービス

概要と歴史



JapanTaxi

全国タクシードライバーという
アプリを
開発しています!



全国タクシー





全国タクシシー

- ・タクシシー会社が内製
- ・位置情報サービス
- ・移動で人を幸せにする



サービス規模

- ・47都道府県
- ・提携タクシース事業者322社
- ・タクシー合計台数33,552台



実は他にも色々やっています





その他のサービス

- ・パートナー様向けサービス
- ・乗務員の方向けアプリ
- ・運行管理



サービスの歴史



サービスの歴史

- ・2011年ローンチ
- ・2016年初のリニューアル
- ・色々と深い経緯がありつつ進化



サービスの歴史

- ・2011年ローンチ

今日のメイン！

- ・2016年初のリニューアル

- ・色々と深い経緯がありつつ進化



インフラの特性



インフラの特性

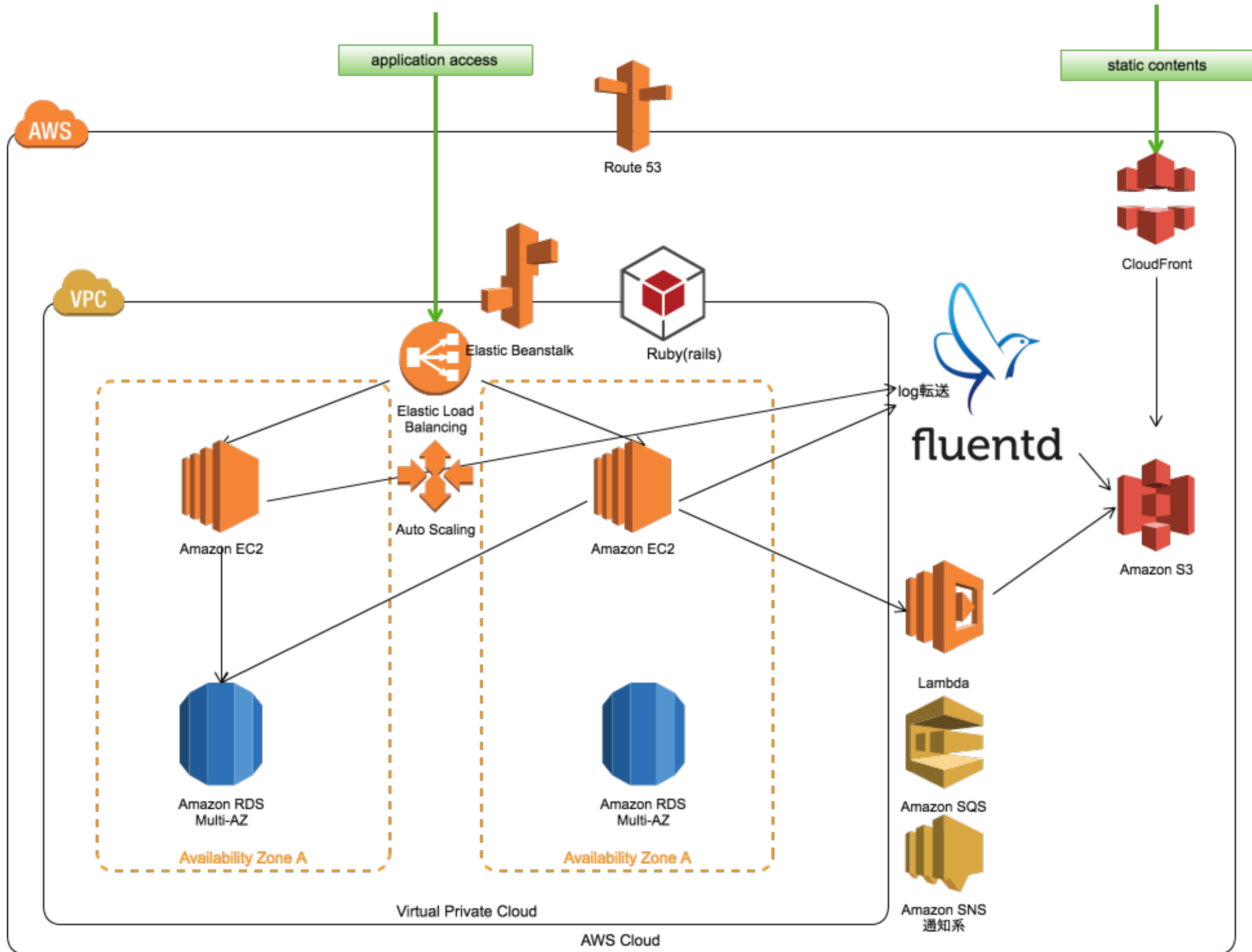
- ・ハイブリッドインフラ
- ・AWS, Azure, GCP
- ・用途に応じて使い分け



JapanTaxi

AWS

API基本構成





監視



- **Datadog(メトリック)**
- **NewRelic (パフォーマンス)**
- **CloudWatch(Datadogへ)**



JapanTaxiが Athenaを 選ぶ理由



Japan Taxi的 Athena 概要(解釈)



Athena概要

- ・雑に使えるpresto(某DWH似)
- ・課金対象がスキヤン量のみ
- ・転送量も気にせず



Athena概要

- S3がそのままデータソース
- データロード不要(某DWH違)
- json, csv, parquet...etc



つまり



S3の資産に
そのまま乗れる
のがAthenaの魅力



ちなみに



経験的なEMRとの差

- ・Athenaはあくまでマネージドなpresto
- ・前後処理がしたいとかストリーム処理がしたいならEMR一択
- ・領域が重なる事は殆ど無いのが体感値



経験的なEMRとの差

- ・共通点はファイルがS3なところ位
- ・EMRはAthenaではできないことをやる
- ・ゼロベースならまずはAthenaを検討



Japan Taxiと Athena の関係



2016年リニューアル時

最も重視したこと

(私の立場として)



ログです！



**リニューアールに並行し
てログの基盤開発を
地道にやっていた**



**この段階で
重要なこと**

(一人でやれる)



この段階で重要なこと

- ・早い
- ・安い
- ・うまい



もう少し具体的に

- 早い → 道具が揃ってる
- 安い → イニシャル不要の従量課金
- うまい → ハイパフォーマンス



ログは全て
S3に集める
(異論は無いはず)



とにかく集める

- 散らばってるログをS3に集約
- Windows, Linux, SQLServer
- FileStorage



S3なら満たせる

- ・OSSのETLツール
- ・安心のAWS内包（実は結構重要）
- ・安心のEMR



S3と
EMRの

話になってますね



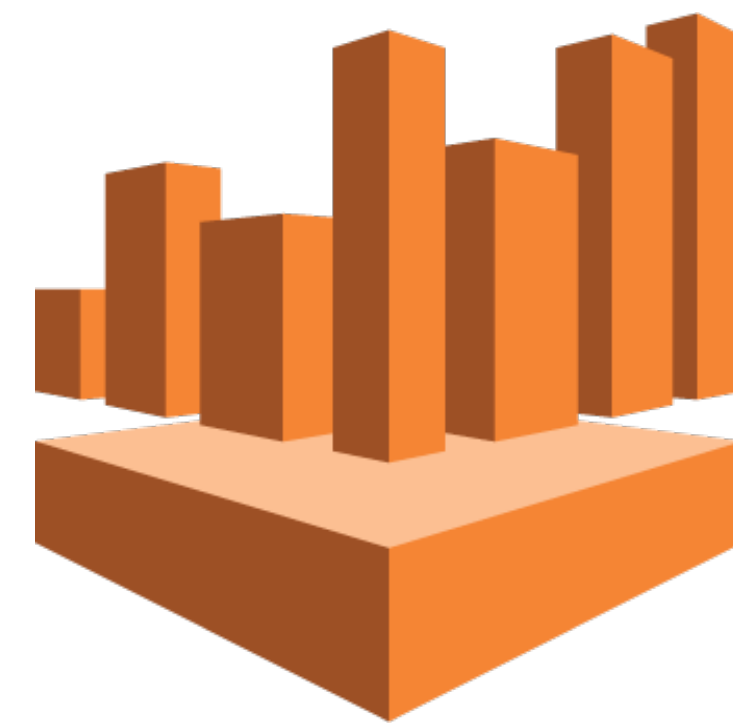
リニューアル当初
Athenaは
無かったのです



逆に言うと



**S3の設定が
Athenaを意識されて
なくても問題無い！**



Japan Taxi でのAthena 使用例

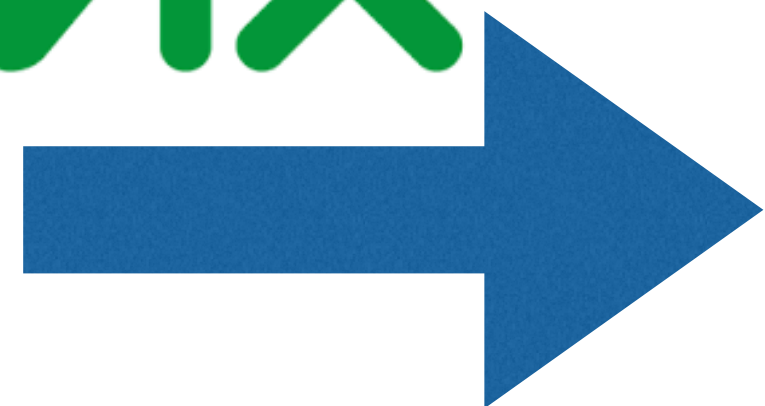


Japan Taxi的 アプローチ

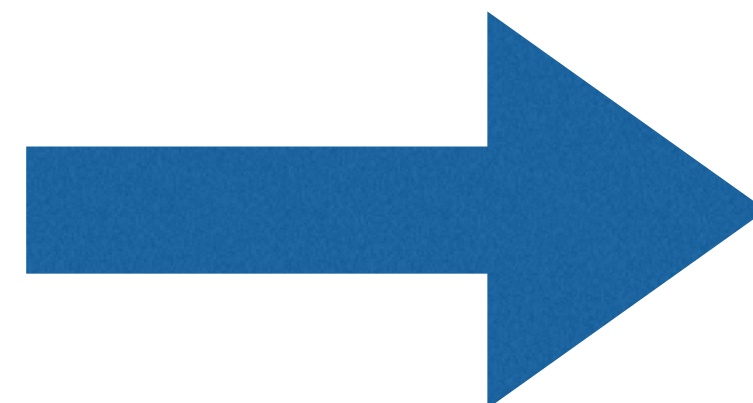


api連携&クラウド連携

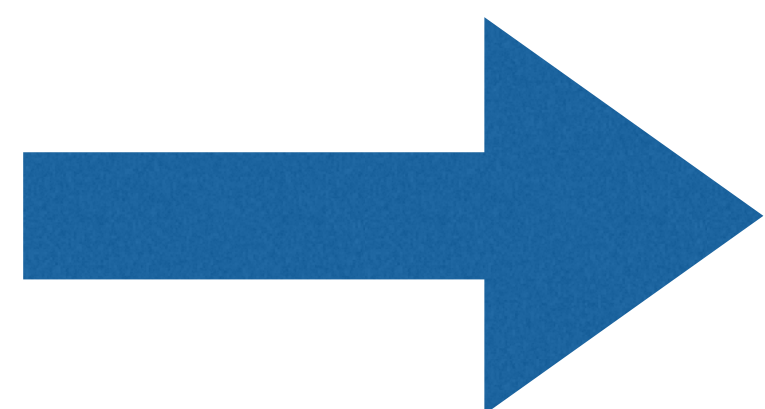
NGINX



fluentd



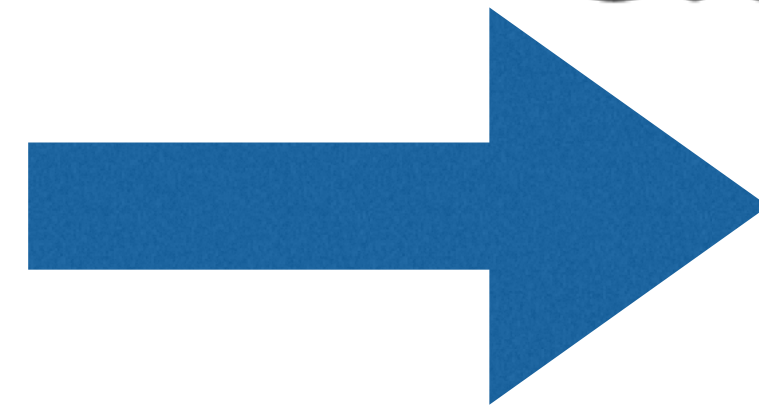
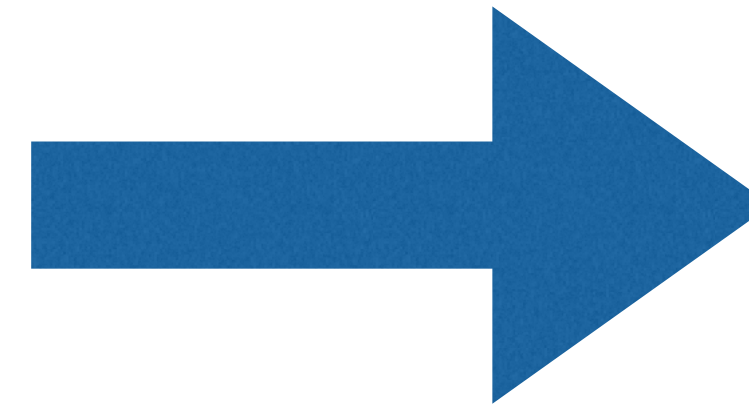
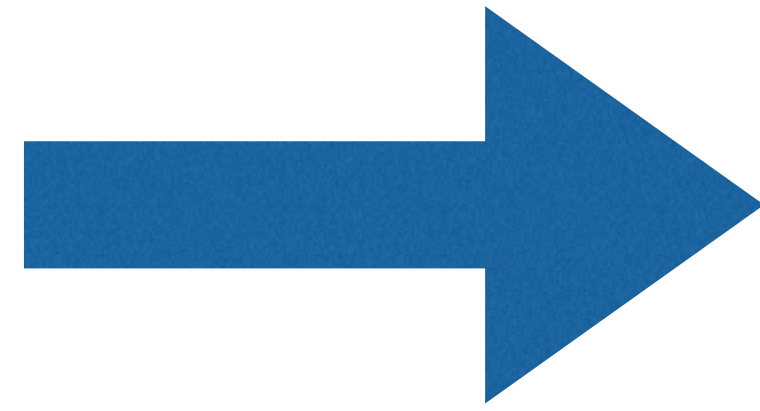
Amazon S3



Microsoft®
.NET



api間連携&クラウド連携





データ量

- probe 1.5G/daily + app log
- 古いものがおおよそ5年分
- 1年分をscanしても40秒くらいで
終わる



どうやって見るか？



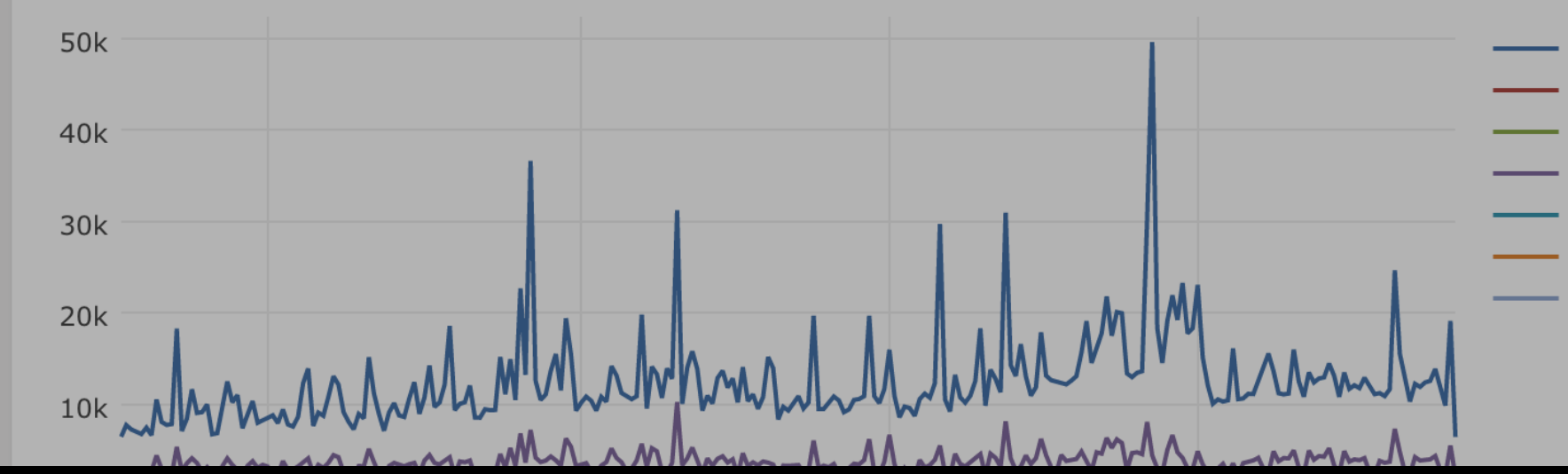
UIとして最も手軽な のはre:dash

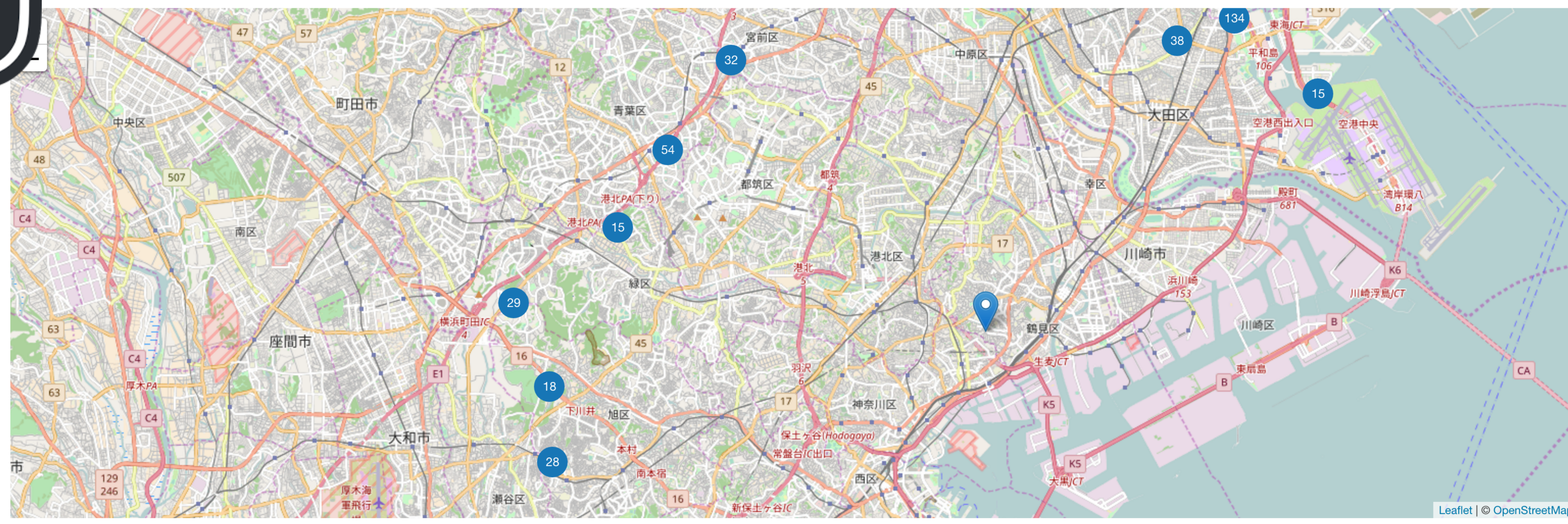
Updated: 3 minutes ago

[監視用]全国タクシーの直近10分間の注文件数
直近10分間の注文件数

76
直近10分間の注文件数

クライアント別注文数



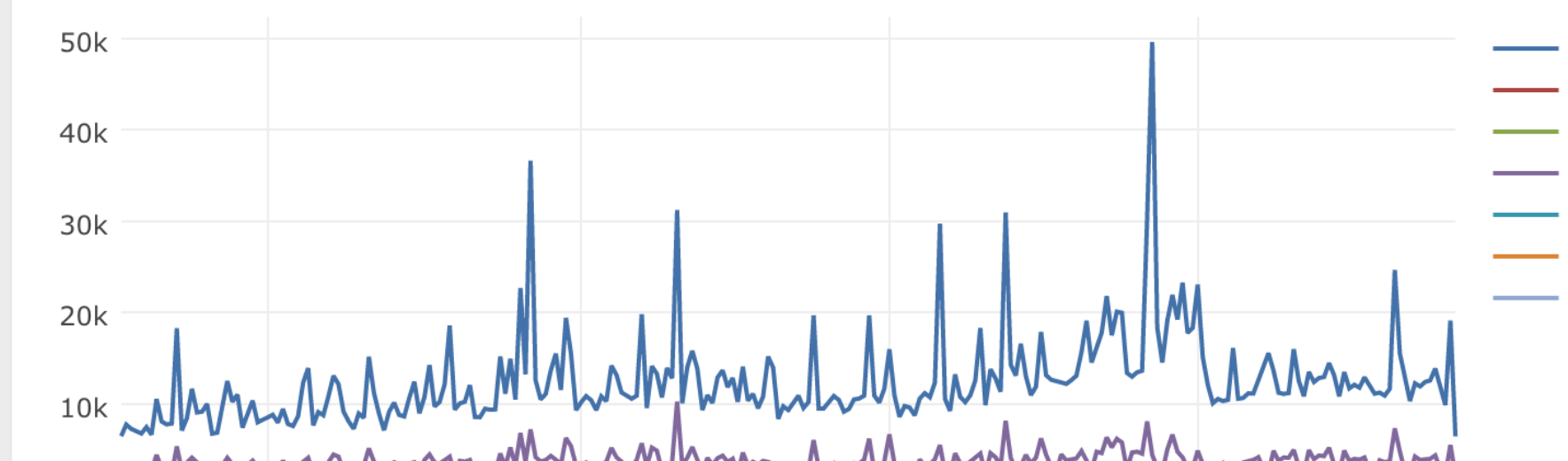


Updated: 3 minutes ago

[監視用]全国タクシーの直近10分間の注文件数
直近10分間の注文件数

76
直近10分間の注文件数

クライアント別注文数





athenaとre:dash

- iam発行するだけ
- 状態監視 (alertサポート)
- 誰もが安全にデータへアクセス

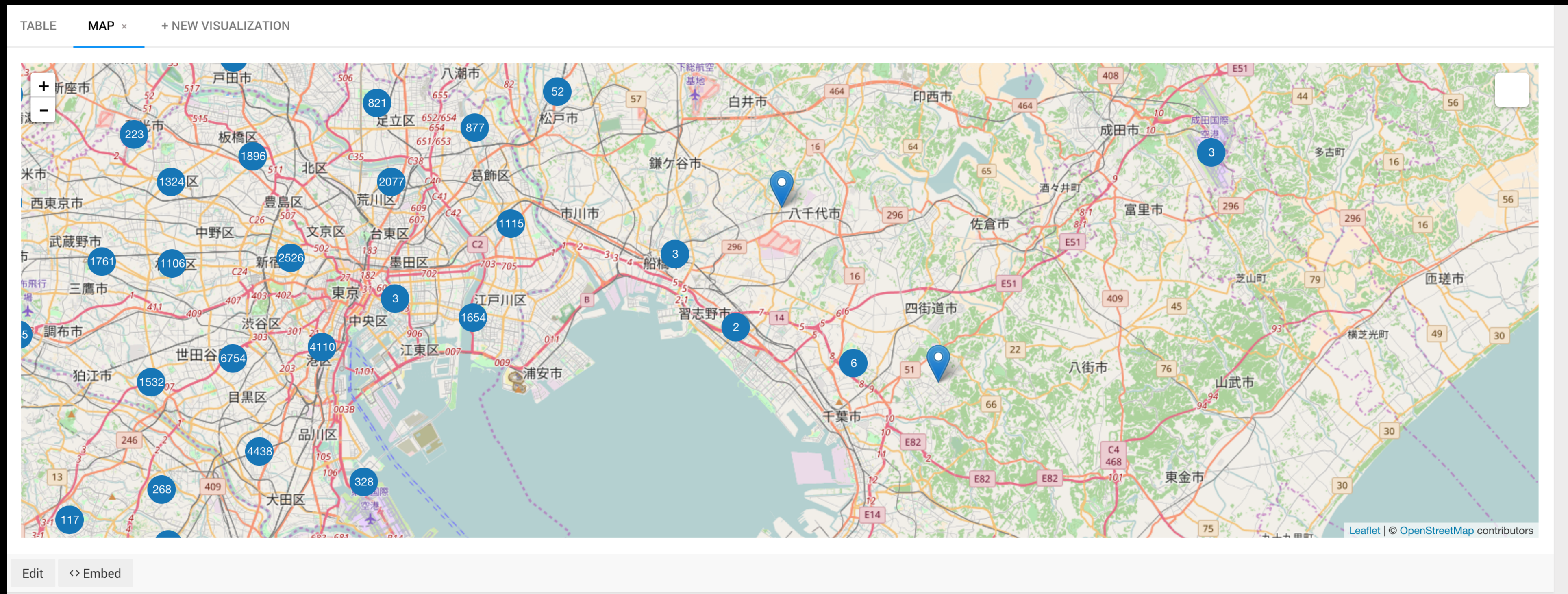




予約注文の課題を 可視化する



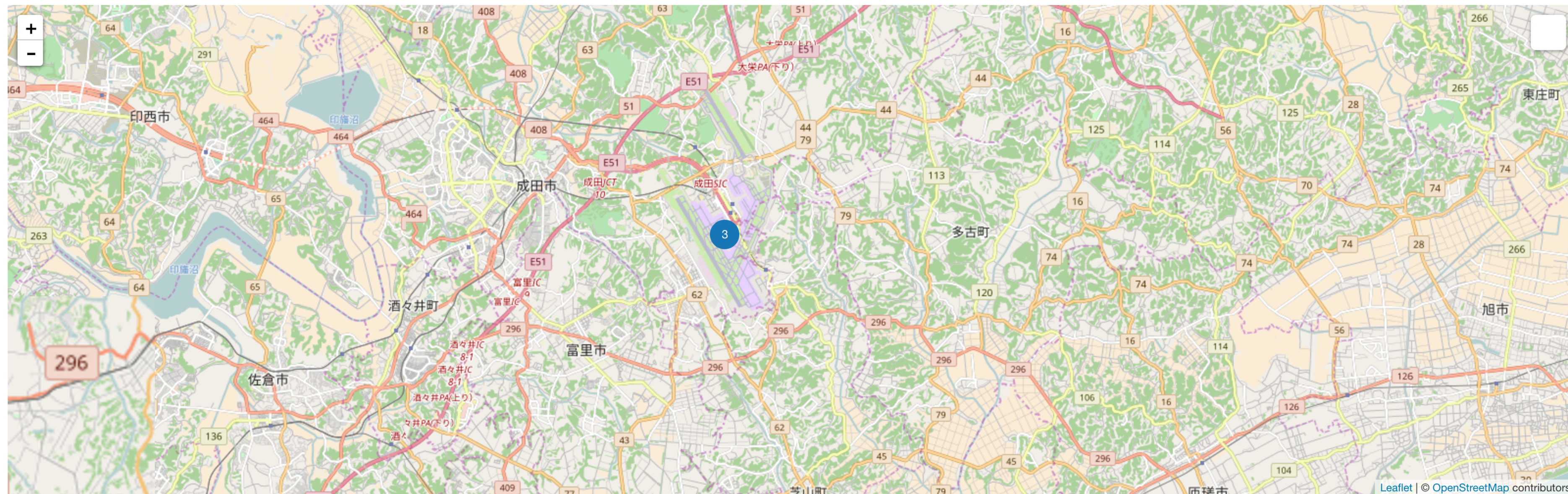
re:dashとathena でリアルタイム監視





ここに注目

TABLE MAP × + NEW VISUALIZATION



Edit <> Embed



通常予約と定額枠

- ・ 実際は同じなのに別枠だった
- ・ 枠の統一
- ・ 注文ロスの排除



case:2



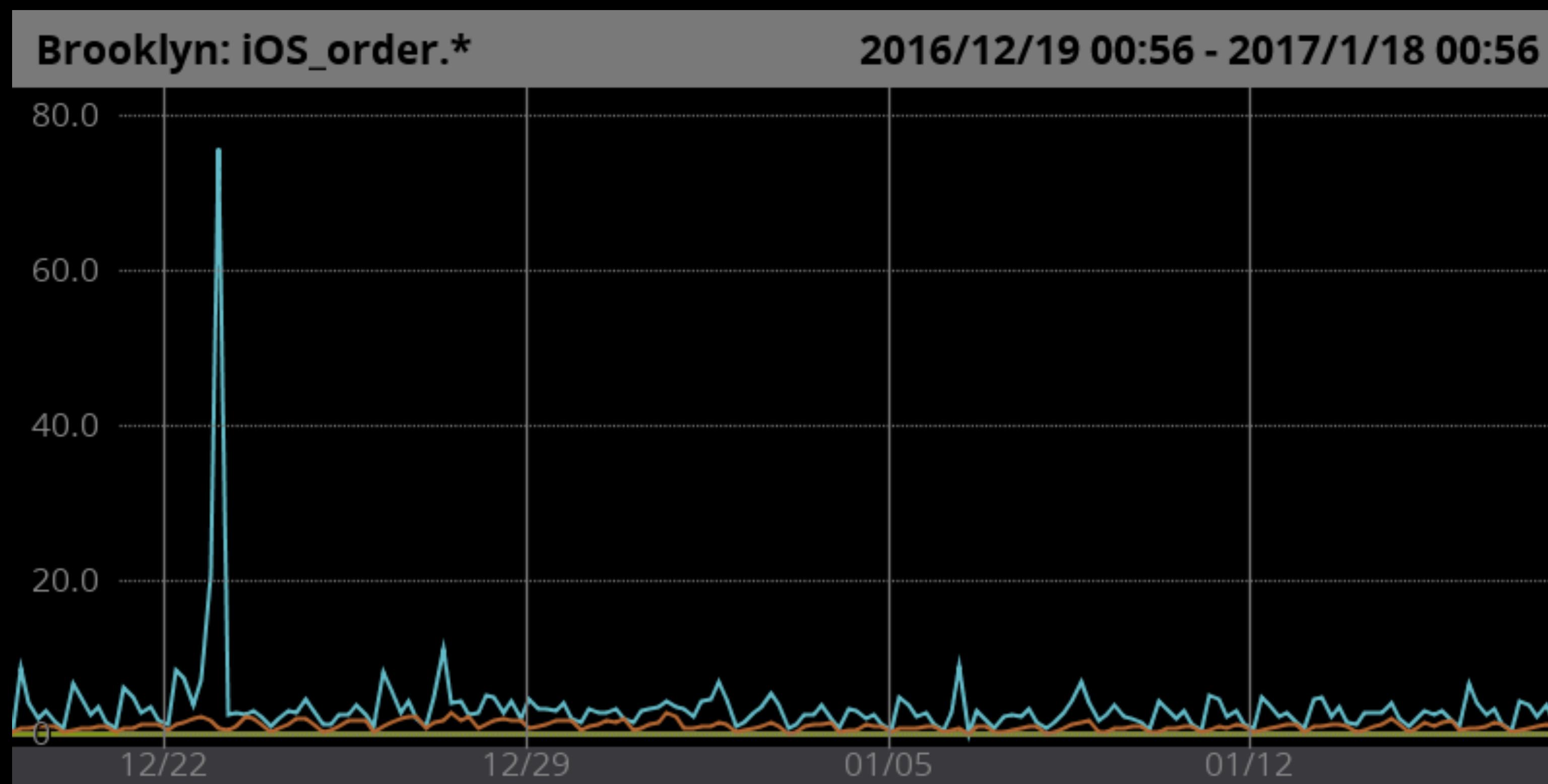
時系列データを
可視化する



TAXIのピークタイム

- ・streamで注文ピークタイムの計測
- ・NOTアクセス数(オートスケールしても車は増えない)
- ・天候・イベント事(主に雨とお酒)

一例





**ETL + aws cli +
地図サービスで
知見を可視化**



ツールチェーン

• **digdag** -> スケジューラ



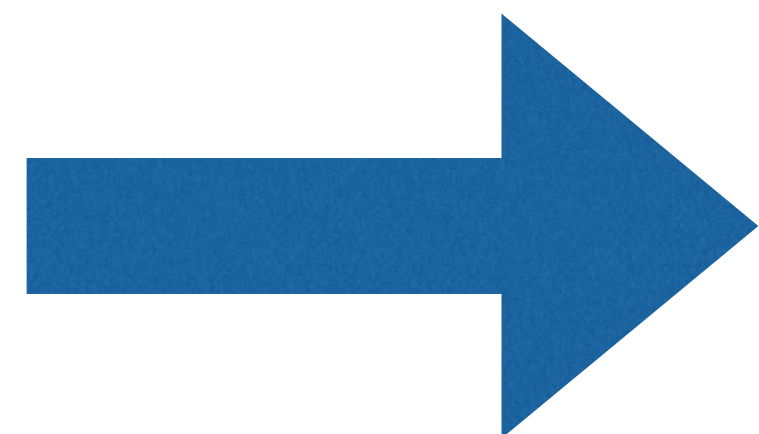
• **embulk** -> バルクインポート



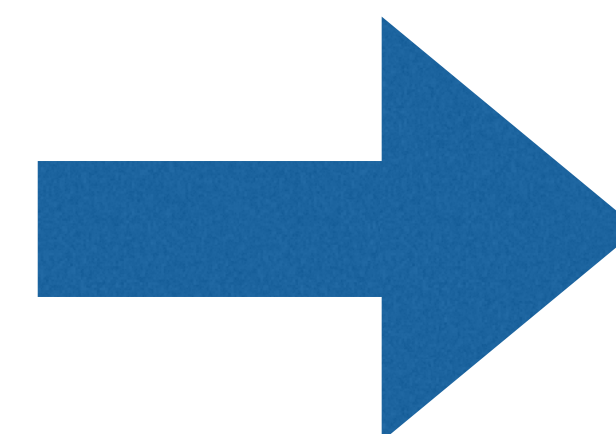
• **aws cli** -> athena操作

embulk

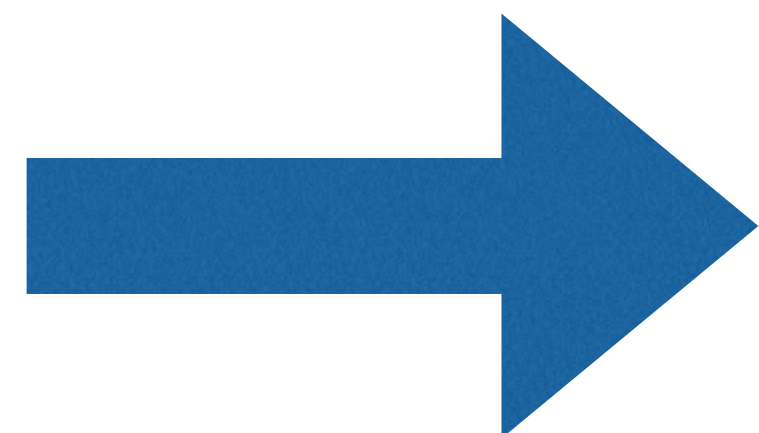
• **curl** -> apiでcsv投げる



aws cli



可視化



curl



DEMO



最後に

これから始める方へ



Location of Input

Data Set

はハマリどころなので

注意

Databases > Add table

Step 1: Name & Location

Step 2: Data Format

Step 3: Columns

Step 4: Partitions

Database

Choose an existing database or create a new one by selecting "Create new database"

Name of the new database

Table Name

Name of the new table. Table names must be globally unique. Table names must correspond to the directory where the data will be stored.

Location of Input Data Set

Encrypted data set [?](#)

Input the path to the data set you want to process on Amazon S3. For example if your data is stored at s3://input-data-set/logs/1.csv, please enter s3://input-data-set/logs/. If your data is already partitioned, e.g. s3://input-data-set/logs/year=2004/month=12/day=11/ just input the base path s3://input-data-set/logs/

External

Note: Amazon Athena only allows you to create tables with the EXTERNAL keyword. Dropping a table created with the External keyword does not delete the underlying data.

Next

リージョン入れると
エラー



今後の期待

- 東京リージョン
- UDF
- **output resultの選択**



まとめ



Athenaの嬉しいところ

- ・S3ツールチェーンという巨人の肩
- ・既存ワークフローの破壊不要
- ・SDK, APIの公開による連携



全国タクシー

JapanTaxi Wallet

ここから乗る

1,000円分のクーポンGETのチャンス!
ご自身のクーポンコードを書き込んでお使いくだ

ステツカー
持ってきてます！



ありがとう

ございました



We love

Athena !

