



# 【技術報告】広告配信における安定して拡張性のある大量データ処理基盤の 必要性と活用

株式会社マイクロアド 大澤 昂太



大澤 昂太

株式会社マイクロアド

開発の傍ら会社のデータ基盤周りの設計などに関わってきました。

仕事で経験のある主要言語はPython、Kotlin、Java、Scalaなどです。

本日は、広告業界がいかに強力なデータベースや分散処理基盤を必要としているか説明します。

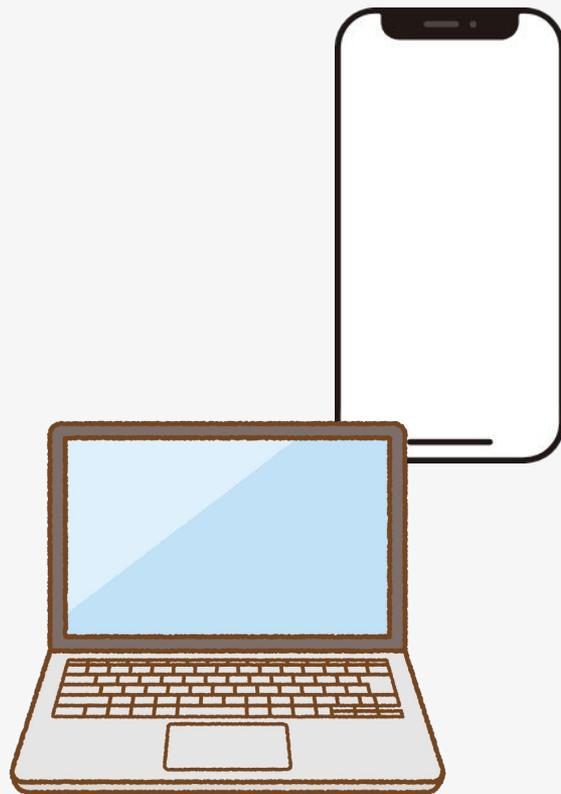


## インターネット広告について

# インターネット広告

インターネット広告とはWebサイトやスマートフォンアプリなどを対象とした広告の形態.

配信する広告を動的に変更することが可能で、これを活用して運用する広告を**プログラマティック広告**と呼ぶ.



# プログラマティック広告

広告の利用者は基本的に下記の効果を狙っている事が多い。

- 収益の増加
- 認知の拡大

プログラマティック広告では与えられた条件や予算内でこれらの効果を最大化するように自動で運用することが多い。

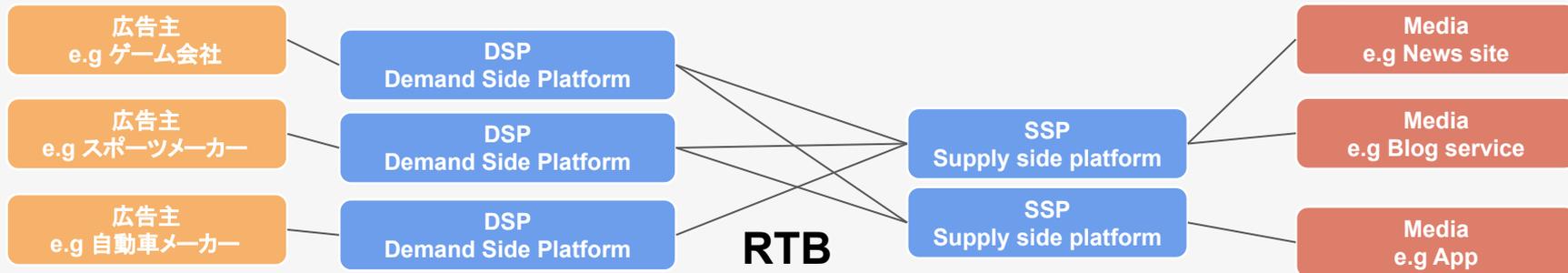
# 広告とオークション

- インターネットは広告を表示できる枠の数が非常に多い
- 手動ですべての広告枠と広告主を手動でマッチングさせるのは難しい

そのため取引を自動化する方法がある.

- 広告枠をまとめてオークションを開催するアドネットワーク.
- 企業間のネットワークに通知してオークションを開催するRTB方式.

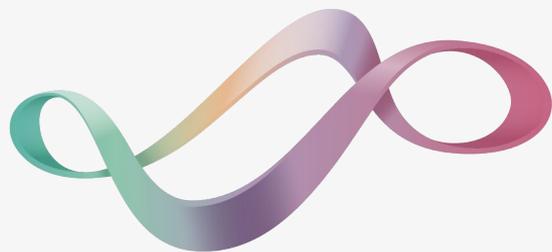
# RTB (Real Time Bidding)



DSP: 広告主の利益を最大化するためにオークションに参加するシステム

SSP: メディアの利益を最大化するためにオークションを開催するシステム

各社のシステムが接続し合いネットワーク上で広告を取引する。



**MicroAd**

Redesigning the Future Life

## 株式会社マイクロアド

インターネット広告を中心に、データを活用したマーケティングプラットフォームを提供する会社です。

インターネット広告の領域では主にRTBを中心にDSPとSSPのシステムの運用をしています。



## 主要な情報技術の課題

# 広告効果最大化

広告効果を最大化するより良いアルゴリズムを必要とする

- CTR予測
- CVR予測
- 落札額最適化
- 推薦システム

機械学習を主に統計分析やORなどの知識を活用する。

ユーザー×広告枠×広告クリエイティブ×LP×タイミング×残予算  
これだけでも10兆以上の組み合わせがある

# 高速なオークション処理

- 広告のオークション処理は50ms程度で入札を終わらせる必要がある。
- この中で複雑な配信条件のマッチングや機械学習の推論を行う必要がある。
- 弊社ではオークションを秒間50000件程度処理している。

上記の処理を行うには数十台～数百台のサーバを必要とする。性能の他に需要に応じて拡張できるスケールアウト性の良いシステムである必要がある。

# ビッグデータ処理

広告主のサイト、メディアサイトなどインターネット広告の会社では何千何万となる取引に関連するサイトのトラフィックが送られてくる。

弊社ではこれらのデータを1日数TB程度処理している。

これらのデータはリアルタイムに処理する場合もある。

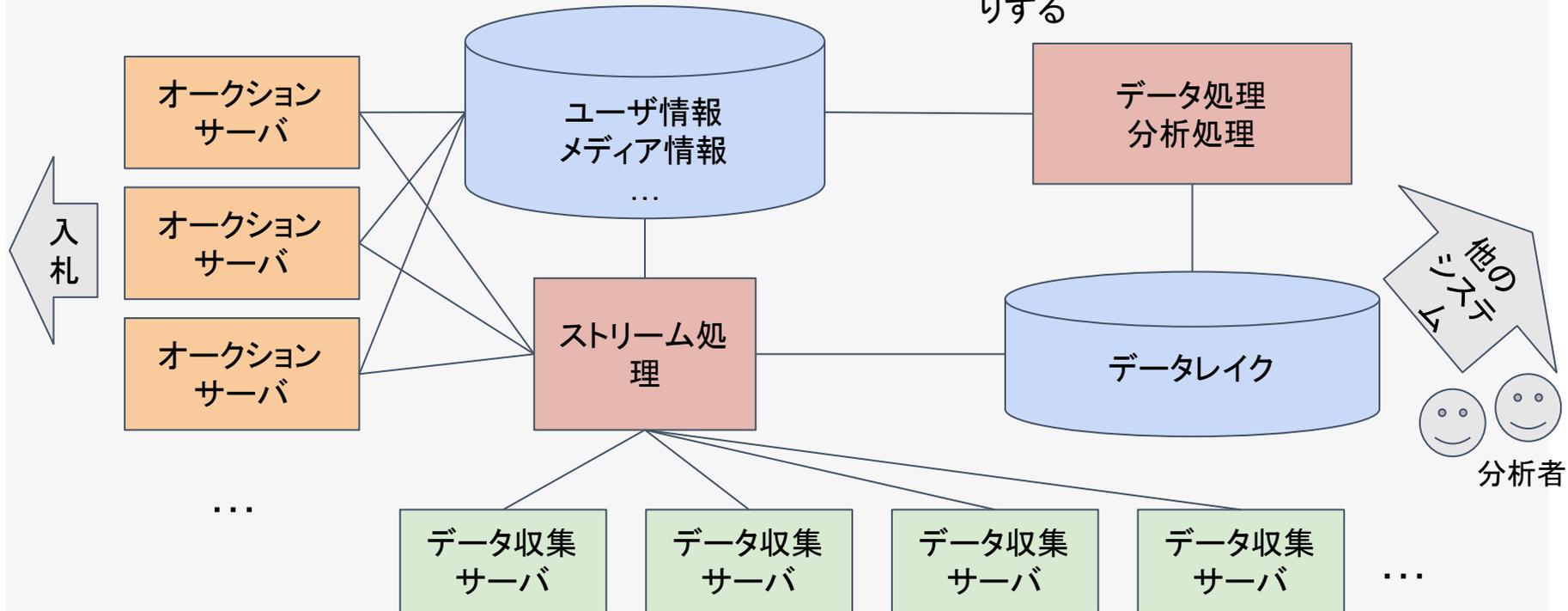
このようなデータを処理するには分散データベース、分散ストレージ、分散処理システムなどを必要とする。



## マイクロアドのデータ技術要素

# DSPシステム簡易図

各種機能は整合性が取れている必要があり、同期を行ったりする



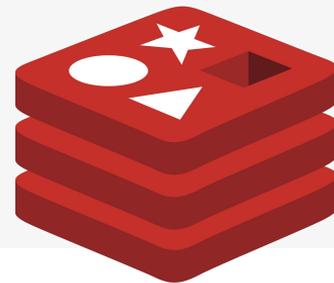
# Redis Cluster (オンメモリKVS)

- 通常のRedisはトランザクションを直列化する事でデータの一貫性を保っている
- 弱点としては処理がCPUの1スレッドに集中する.
- リーダーは1台で水平スケールを行うことができない.

Redis Clusterでは

- シャーディングで複数ノードに分散書き込み可能.
- ノード毎にCPUを使うのでマルチコアCPUを活用可能.
- オンラインで水平スケールが可能.

その代わりに強い一貫性を保証できない.



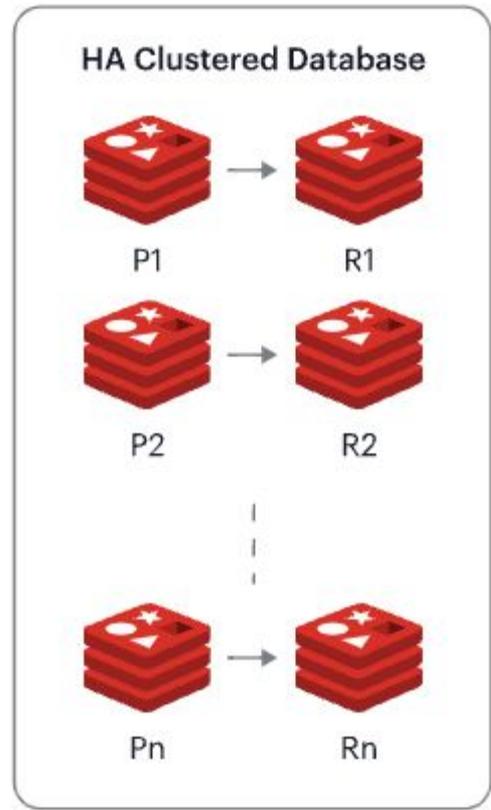
# Redis Cluster

主に配信時に参照するデータベースでは

- レスポンスが数ms
- 大量のクエリを送ってもパフォーマンスが落ちない
- 取り扱いデータは数百GB程度になる。

それらの要件にRedis Clusterはマッチしていた。

運用は非常に大変。



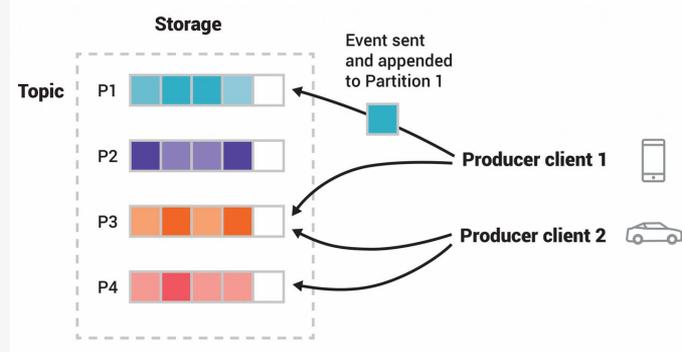
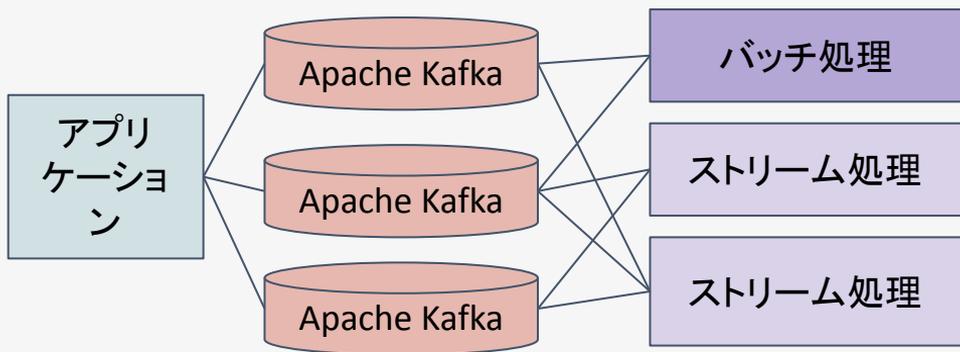
# Apache Kafka



データベースのような分散メッセージキュー。

通常メッセージキュー、PubSubはスケールアウト性が低いことが多いがKafkaは分散処理機能が強力。

最低3サーバ必要で1つのデータは最低2つのノードに書き込まれる(可用性、分断耐性は高いが一貫性は低い)。



# Apache Kafka

Apache Kafkaは1つのデータを複数のクライアントで処理させる際に威力を発揮するように作られている。

このクライアントは増やしたり、減らしたりすることも容易。

データパイプラインの柔軟性が上がるため、多くのデータ受け渡しの間で活用している。

大規模なデータを受け渡すシステムでも非常に安定して動作している。非常に頼りにしている。



# さいごに

- 今回紹介できなかったものの、分散処理システム、機械学習、大規模データ向けに設計したデータセンターなどがあります。
- 広告システムは性能要件やトラフィックの状況から、データベースや分散処理システムへの関心が高いです。
- これらの分野への関心や造詣が深い人を常に探しています。

# DEIM 企業ブースの照会

3月4日、3月5日のDEIMオンラインイベントでは  
スポンサーブースの G5 で出展しています！

気になる方は声かけてください！





株式会社 マイクロアド  
〒150-0045 東京都渋谷区桜丘町20-1  
渋谷インフォスタワー13F

Micro Ad, Inc.  
〒150-0045 東京都渋谷区桜丘町20-1  
渋谷インフォスタワー13F

[www.microad.co.jp](http://www.microad.co.jp)