

# Propreのサービスを支える システム基盤と 最新のRDB事情

株式会社Propre Japan  
CTO 根岸良多

## SECTION 2

Propreのサービスを支えるシステム基盤と最新のRDB事情

# 不動産ビッグデータとは？

- ビッグデータとは？  
→特に定義はない

1000～ 目視チェックはストレス  
10万～ Excelだと限界  
100万～ MySQLやPostgreSQLで本格的なチューニング要

- Propreの扱う不動産データ

縦：数億～数十億レコード

横：3万項目＋住所約50万(※「港区」なども重要なキーワード)  
世界中の不動産情報のため為替も入ってくる

- データの処理数

収集データで1500～2000万／日

※ <https://propre.com/> ホームページに前日の取扱数を掲載

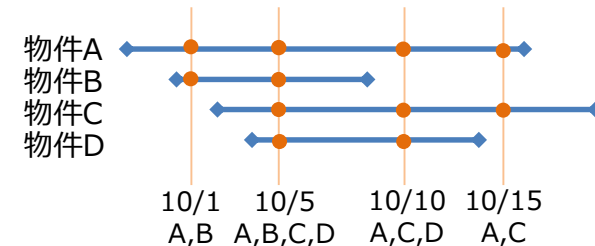
		
ターゲット	18 NOV 2020 記録	総データ数
17 カ国	15,483,944 物件	417,618,402 物件
UTC 18 Nov 2020 07:00 更新		


比較的構造化されたデータだがRDBでどう処理するか？

# Propreが行うデータ処理の特徴と仕組み

- ✓ 最終アウトプット出力のリアルタイム性を重視（即時/時間/日/週/月）
- ✓ いろいろな視点や角度で集計・分析を追加できる
- ✓ 処理のやり直しができる  
（クレンジング精度・名寄せやテキストマイニングの向上、新ロジック、バグ判明 etc.）
- ✓ ある時点のスナップショットをいつでも再現できる
- ✓ 各プロセスの全自動化（24h稼働）

## Queue Controller(QC)を開発、 レコードを管理・制御



- 
- 差分レコードのみ収集
  - データの収集とクレンジングの工程を分離
  - RAW・加工段階のデータを残していつでも取り出せる
  - 基本情報＋「タグ」という概念（タグには距離などの数値も持てる）

物件データとしてのトランザクションテーブル数は10程度

1sec約1000程度のSQL発行、半分くらいが更新系 ※集計・分析は別途

# RDS(MySQL)→Autonomous移行まで

<初期フェーズ>

約10台のRDS (MySQL)

<パフォーマンス追求フェーズ> UPDATE激速！

MySQL Cluster (インメモリDB) + RDS

管理ノード ×1

データノード ×4 (メモリ合計約1.5TB)

SQLノード ×3

分析用に自作サーバー (M.2 SSD×8並列) も構築

<管理・TCO検討フェーズ>

Autonomous Transaction Processing

← 今ここ

<発展フェーズ>

Data Science Platform

Autonomous Data Warehouse

などとの連携で、よりシームレスにデータを扱っていく

# OCI/Autonomous Databaseのメリット

OCI: Oracle Cloud Infrastructure

## クラウドのコストメリット

特にメモリとネットワークが安価

	OCI	AWS
Linuxインスタンス	2コア32GBメモリ(E3.2) \$0.0980	2コア16GBメモリ(m5.xlarge) \$0.1920 2コア32GBメモリ(r5.xlarge) \$0.2520
ネットワーク	\$0.0250	\$0.08
サポート費用	不要	必要

料金は<https://www.oracle.com/jp/cloud/economics/> より抜粋

## Autonomous Transaction Processing(ATP)のメリット

- パラレル処理 (PostgreSQLパラレルクエリ: 制約あり)
- CPU数を自由に変えられる
- メンテナンス工数の大幅な削減

当初インスタンスの単純移行を検討したもののTCOを考えてATPを採用

※INDEXはきちんとつけてヒント句で明示指定したほうがいい

# 進化を続けるRDBと今後に期待すること

階層型 → RDB → NoSQL ???

RDBは古い技術／終わった／新しい仕組みに・・・



**そんなことはない  
ここ数年のRDBの進化はすごい！**

今のRDBは表形式のデータをSQLで抽出するだけではない

- JSON対応
- Spatial（空間データ・位置情報）
- Data Science／分析ツール
- プログラムコードとの融合（データ転送不要）

etc.

クラウドならではのメリットに期待

オンライン拡張、外部連携、セキュリティー、API、、、

# 最後に・・・

## やりたいことを実現するシステムを目指しましょう

(システム開発あるある)  
それはDBが厳しい・難しいですね～  
過去データは何年分取っておきたいですか？



以前のシステム的な制約は今の技術でかなり解消されているので、  
できるだけ調べてみる、実現できないか考えてみる、**相談してみる**

## ぜひOracle Cloud Infrastructureを使ってみましょう

(利用前の不安)  
日本語ドキュメントが少ないのでは？  
管理ページの操作にも慣れないと・・・



- 基本的に同じことができる  
( & 後発なので割と嬉しいスペックや機能)
- **サポートに質問できる**
- 公式ドキュメントはそれなりに充実

Computeインスタンスを使うだけでも  
簡単でコストメリットはあるので  
ぜひ試しに使ってみてください！