

Snowflakeで始めるData Clean Room！

DCRってなに？から構築テクニックまで伝授しちゃおうよ～

DCRユースケースの輪を広げよう！

エクスチュア株式会社 喜田紘介



自己紹介

喜田 紘介 @kkkida_twtr

エクスチュア株式会社
日本PostgreSQLユーザ会 理事長

最近の仕事

約半年間、Snowflakeのデータクリーンルーム構築案件、社内データ基盤の設計、社内の情報整備やメンバー育成に従事。

データベースが好き、ITが好き、コミュニティが好きを活かして、コミュニティ盛り上げの一助になれば幸いです。



本セッションでは

19:05

バーチャル

対面

エクスチュア 松村拓さん「データクリーンルームって何?データプライバシーの潮流と、企業同士でのデータ相互活用で活躍するDCRについて」

19:35

バーチャル

対面

KDDIアジャイル開発センター株式会社 福田陽一さん「Snowflake DCRの紹介」

19:50

バーチャル

対面

KDDIアジャイル開発センター株式会社 篠原 誠さん「Snowflake DCRをTerraform化した話・Snowflake DCRの未来」

20:05

バーチャル

対面

エクスチュア 喜田 紘介さん「DCRユースケースの輪を広げよう！」

データクリーンルームの必要性を理解



Snowflake DCRの仕組み
実装手順を理解



どんなデータがあるとどう使えるの?
利用イメージをもっと膨らませて
「相手探し」の第一歩を踏み出そう!

本セッションでは

うちのデータこういう風に整えれば売れるんだ！
うちならではの強みってなんだろう？



買ったらすぐに活かせるデータの形って？
データを活かすビジネスモデルを描く



SoR から SoE、SoIの時代へ



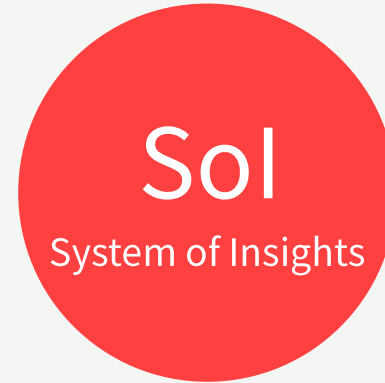
事実を正確に記録し、間違わない
失わない、流出しないためのシステム

- ✓ 会計
- ✓ 契約管理
- ✓ 在庫管理



顧客とのエンゲージ（約束・結びつき）を
主眼に置いたCXのためのシステム

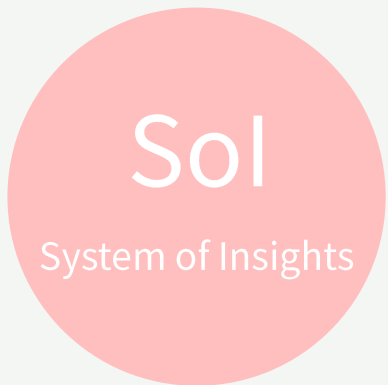
- ✓ コールセンター
- ✓ リコメンド
- ✓ CDP



蓄積されたデータを分析し新たな洞察を得る
ためのシステム

- ✓ ML/AI
- ✓ パブリックデータ

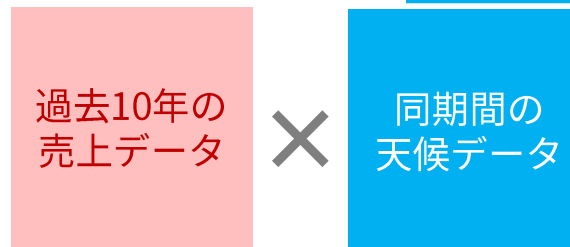
SoE、SoIの時代 - どのようなデータに価値があるか -



事業全体の改革につながる新たな知見を得たい

- ✓ 年単位の鮮度
- ✓ 超大規模データから特徴抽出 (ML/AI)
- ✓ パブリックデータとの突合

その道の専門機関が
収集したデータ



顧客の指向を理解し、要望にフィットする対応したい

- ✓ 個人に紐づく履歴
- ✓ 界隈のトレンドも加味
- ✓ 日単位の鮮度

多くの属性をとらえる

mail_address	実績	A	B	C	D	E
aaa@mail	成約	✓		✓	✓	✓
bbb@mail	トライアル	✓	✓			
ccc@mail	失注	✓				✓
ddd@mail	成約	✓		✓	✓	✓



個人に紐づく属性を充実させてデータドリブンな施策へ

購入歴のある人に
DMを送ろう

経済系youtuberと
コラボしてみようよ

旅行絡めた新たなブラン
ディングで新CMやろう

会員ID	購入歴あり			経済	コ	エンタメ	スポーツ	旅行
aaa@mail	✓			✓	✓		✓	✓
bbb@mail	✓			✓				
ccc@mail	✓			✓				✓
eee@mail	✓			✓	✓		✓	✓
fff@mail	✓							
hhh@mail	✓			✓	✓	✓	✓	✓

会員IDに紐づく属性が充実している状態

何らかの集計・分析を経て施策へ

分析：会員全体のうち該当者が多い属性は？

キーワード	会員数	該当者	割合
経済	20000	16000	80%
IT	20000	4000	20%
エンタメ	20000	100	2%
スポーツ	20000	2000	10%
旅行	20000	12000	60%

サードパーティデータを購入して属性を充実

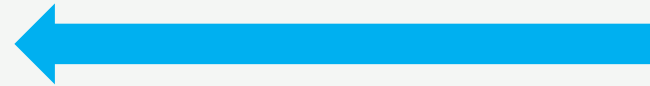
分析：A社会員のうち、X社に該当した人で多い属性は？

キーワード	マッチした数	該当者	割合
経済	5000	4000	80%
IT	5000	1000	20%
旅行	5000	3000	60%

ABC社

会員ID	購入歴あり		
aaa@mail	✓		
bbb@mail	✓		
ccc@mail	✓		
eee@mail	✓		
fff@mail	✓		
xxx@mail	✓		
yyy@mail	✓		
hhh@mail	✓		

XYZ社作成のレポート納品



XYZ社

mail_address	経済	コ	エンタメ	スポーツ	旅行
aaa@mail	✓	✓		✓	✓
bbb@mail					
ccc@mail		✓			✓
ddd@mail	✓	✓			✓
eee@mail	✓	✓			
fff@mail					
ggg@mail		✓			
hhh@mail	✓	✓	✓	✓	✓
iii@mail		✓	✓	✓	✓
jjj@mail				✓	
kkk@mail			✓	✓	

XYZ社にリストを提出



プライバシーポリシーの同意が必要

ヒットしないデータもある

仮にデータを直接クエリできたとして（NGな例）

会員に紐づく属性が直接見えてしまう！

会員ID				経済	IT	エン	スポ	旅行
aaa@mail	✓			✓	✓		✓	✓
bbb@mail	✓			✓				
ccc@mail	✓			✓				✓

これが許されてきたのが
3rd Party Cookie

XYZ社

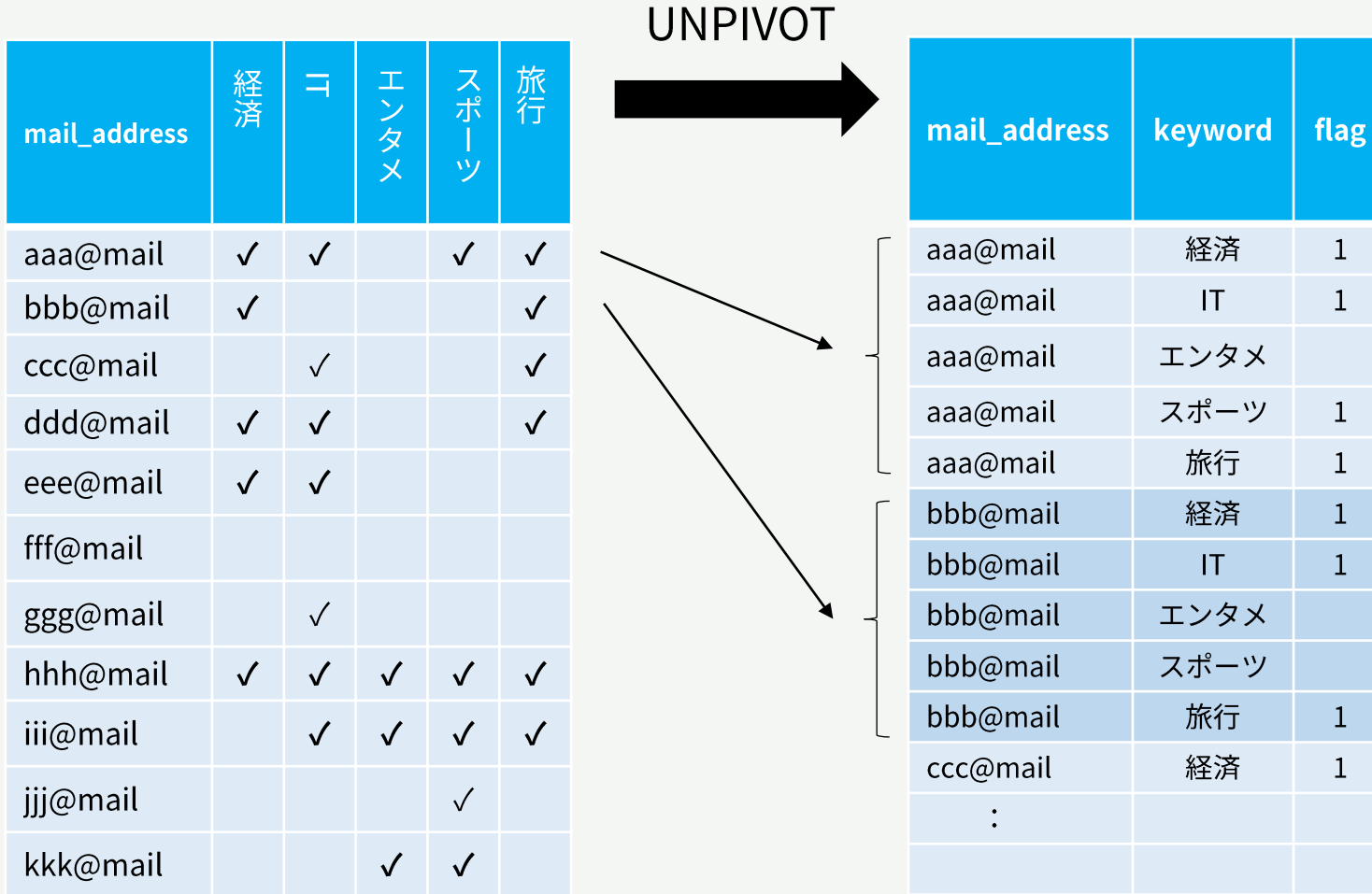
mail_address	経済	コ	エンタメ	スポーツ	旅行
aaa@mail	✓	✓		✓	✓
bbb@mail					
ccc@mail		✓			✓
ddd@mail	✓	✓			✓
eee@mail	✓	✓			
fff@mail					
ggg@mail		✓			
hhh@mail	✓	✓	✓	✓	✓
iii@mail		✓	✓	✓	✓
jjj@mail				✓	
kkk@mail			✓	✓	

ABC社

会員ID	購入歴あり		
aaa@mail	✓		
bbb@mail	✓		
ccc@mail	✓		
eee@mail	✓		
fff@mail	✓		
xxx@mail	✓		
yyy@mail	✓		
hhh@mail	✓		

```
SELECT * FROM  
list_x  
JOIN  
list_a  
ON 会員ID=mail_address
```

GROUP BYしやすい形に整える



```
SELECT  
  keyword, count_if(flag)  
FROM  
  list_x  
GROUP BY  
  keyword
```

keyword	Countif
経済	2000000
IT	1000000
エンタメ	3000000

再掲：仮にデータを直接クエリできたとして（NGな例）

会員に紐づく属性が直接見えてしまう！

会員ID				経済	IT	エン	スポ	
aaa@mail	✓			✓	✓		✓	✓
bbb@mail	✓			✓				
ccc@mail	✓			✓				✓

ABC社

会員ID	購入歴あり		
aaa@mail	✓		
bbb@mail	✓		
ccc@mail	✓		
eee@mail	✓		
fff@mail	✓		
xxx@mail	✓		
yyy@mail	✓		
hhh@mail	✓		

```
SELECT * FROM  
  list_x  
JOIN  
  list_a  
ON 会員ID=mail_address
```

XYZ社

mail_address	経済	コ	エンタメ	スポーツ	旅行
aaa@mail	✓	✓		✓	✓
bbb@mail					
ccc@mail		✓			✓
ddd@mail	✓	✓			✓
eee@mail	✓	✓			
fff@mail					
ggg@mail		✓			
hhh@mail	✓	✓	✓	✓	✓
iii@mail		✓	✓	✓	✓
jjj@mail				✓	
kkk@mail			✓	✓	

データクレンジングで実行しやすいSQL

A社のリストを考慮したキーワード別レポートが取得できた

全体のマッチ数と割合は別途考慮

キーワード	マッチした数	該当者	割合
経済	5000	4000	80%
IT	5000	1000	20%
旅行	5000	3000	60%

ABC社

会員ID	購入歴あり		
aaa@mail	✓		
bbb@mail	✓		
ccc@mail	✓		
eee@mail	✓		
fff@mail	✓		
xxx@mail	✓		
yyy@mail	✓		
hhh@mail	✓		

```
SELECT
  keyword, count_if(flag)
FROM
  list_x
JOIN
  list_a
ON
  会員ID=mail_address
GROUP BY keyword
```

XYZ社

mail_address	keyword	flag
aaa@mail	経済	1
aaa@mail	IT	1
aaa@mail	エンタメ	
aaa@mail	スポーツ	1
aaa@mail	旅行	1
bbb@mail	経済	1
bbb@mail	IT	1
bbb@mail	エンタメ	
bbb@mail	スポーツ	
bbb@mail	旅行	1
ccc@mail	経済	1
:		

keyword	Countif
経済	2000000
IT	1000000
旅行	1500000

```
SELECT
  keyword
  count_if(flag)
FROM
  list_x
GROUP BY
  keyword
```

売り手側の視点：どのようなデータが強みになるか

うちのデータこういう風に整えれば売れるんだ！

うちならではの強みってなんだろう？



豊富な属性

トレンドを反映した属性

ニッチな属性（地域性、趣味/指向）

多くのユーザー

mail_address	経済	コ	エンタメ	スポーツ	旅行
aaa@mail	✓	✓		✓	✓
bbb@mail					
ccc@mail		✓			✓
ddd@mail	✓	✓			✓
eee@mail	✓	✓			
fff@mail					
ggg@mail		✓			

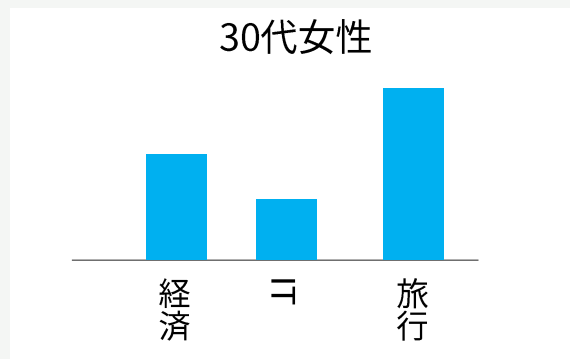
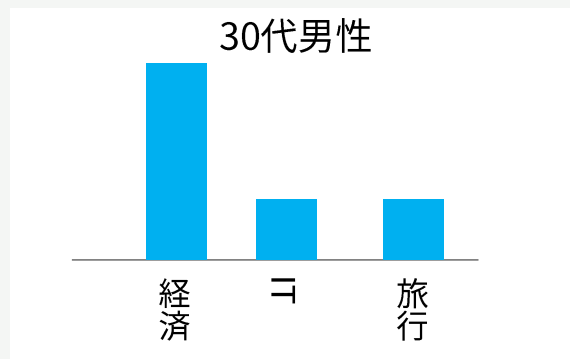
買い手側の視点：すぐに活かせるデータとは

結合に使えるユーザー一覧があれば
すぐに始められる



会員ID	購入済	年齢層	性別
aaa@mail	✓	20代	男
bbb@mail	✓	30代	女
ccc@mail	✓	30代	男
eee@mail	✓		
fff@mail	✓		
xxx@mail	✓		
yyy@mail	✓		
hhh@mail	✓		

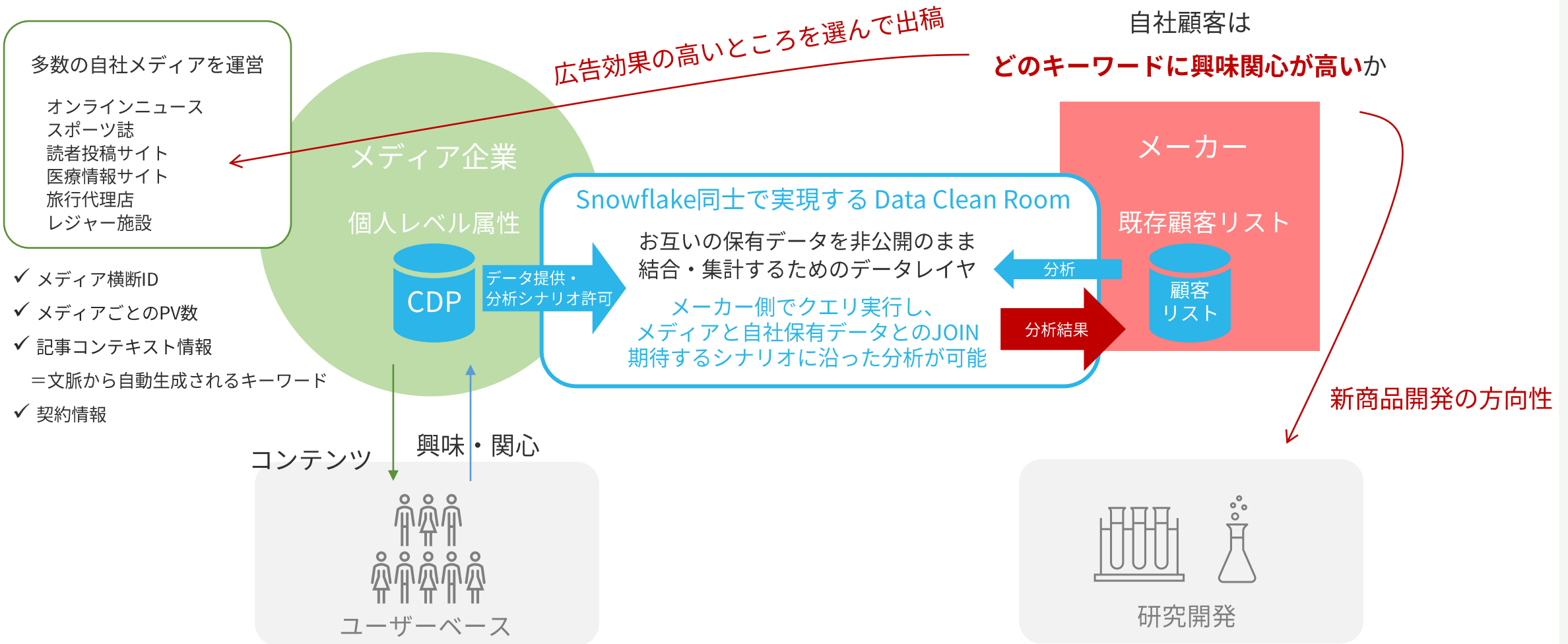
追加で分析できる属性を持つ
→複数DCRの掛け合わせ？



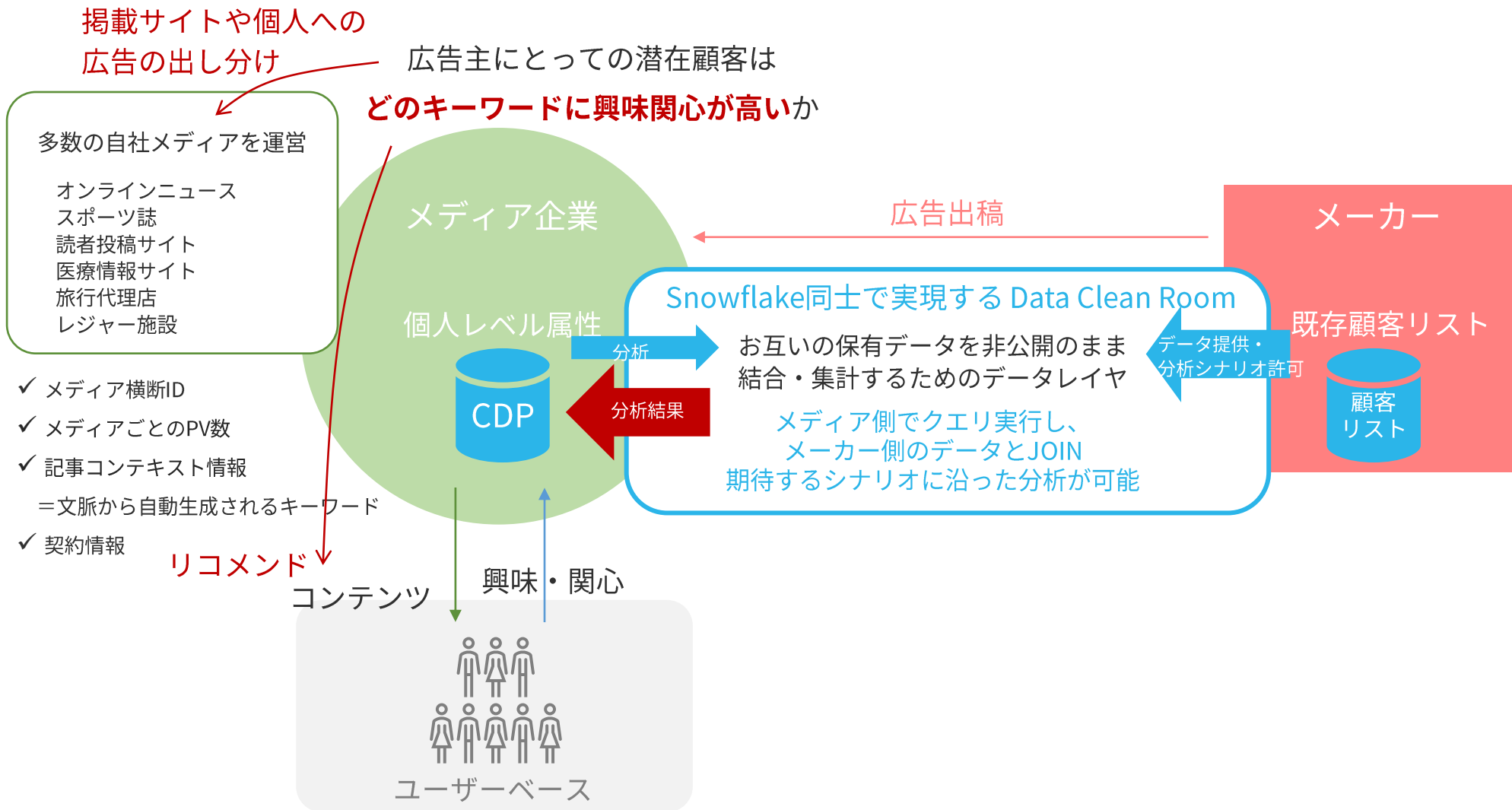
買ったらずきに活かせるデータの形って？
データを活かすビジネスモデルを描く



Data Clean Roomによるコンシューマデータの強化



Data Clean Roomによる広告価値向上



まとめ

うちのデータこういう風に整えれば売れるんだ！
うちならではの強みってなんだろう？

誰もがプロバイダーになれる時代
できることを知り、データの価値創出を模索しましょう

買ったらずくに活かせるデータの形って？
データを活かすビジネスモデルを描く

分析を始めるまでのデータ整備のハードルは低い
はじめの一步を踏み出しましょう！

豊富な属性

トレンドを反映した属性

ニッチな属性（地域性、趣味/指向）

mail_address	経済	コ	エンタメ	スポーツ	旅行
aaa@mail	✓	✓		✓	✓
bbb@mail					
ccc@mail					
ddd@mail					
eee@mail					
fff@mail					
ggg@mail		✓			

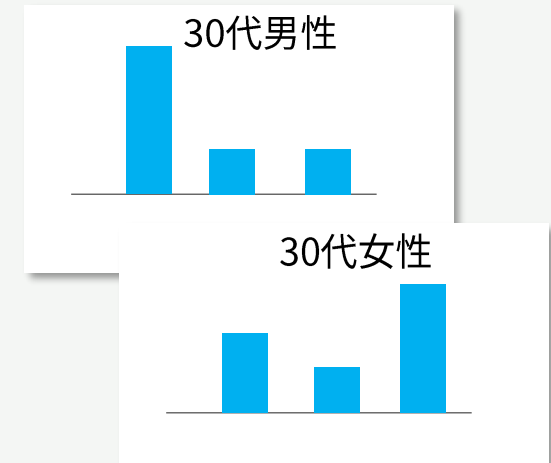
多くのユーザー

売り手も買い手も、良い相手を求めている
Japan - Data Clean Roomコミュニティを活用しましょう

結合に使えるユーザー一覧があればすぐに始められる

会員ID	購入済	年齢層	性別
aaa@mail	✓	20代	男
bbb@mail	✓	30代	女
ccc@mail		30代	男
ddd@mail			
eee@mail			
fff@mail			
ggg@mail			
yyy@mail	✓		
hhh@mail	✓		

追加で分析できる属性を持つ
→複数DCRの掛け合わせ？



Data to the People

すべての人にデータを