

ユーザー数は実測値じゃない！？

～ GA4の数値の仕組み ～

TOPIC

1. GA4のユーザー数・セッション数は「推定値」
2. レポート用識別子によって、ユーザー数は変わる
3. BigQueryとの数値差異



この講座の内容

- GA4のユーザー数・セッション数は「推定値」
- レポート用識別子によって、ユーザー数は変わる
- BigQueryとの数値差異

ユーザー数は実測値じゃない！？ ～GA4の数値の仕組み～

GA4のユーザー数・セッション数は 「推定値」



GA4のユーザー数・セッション数は「推定値」

知ってた？

GA4のセッション数・ユーザー数は、実測値ではなく推定値
「HyperLogLog++」(HLL++) という統計のアルゴリズムで算出された値

なんで？

大量のデータを処理するには、膨大な時間がかかる...。
素早くレポートを表示するために、
効率と精度向上、エラー率改善を目的として
「HyperLogLog++」で値を算出している！

スピード重視
ってことだね！



レポート用識別子によって、 ユーザー数は変わる

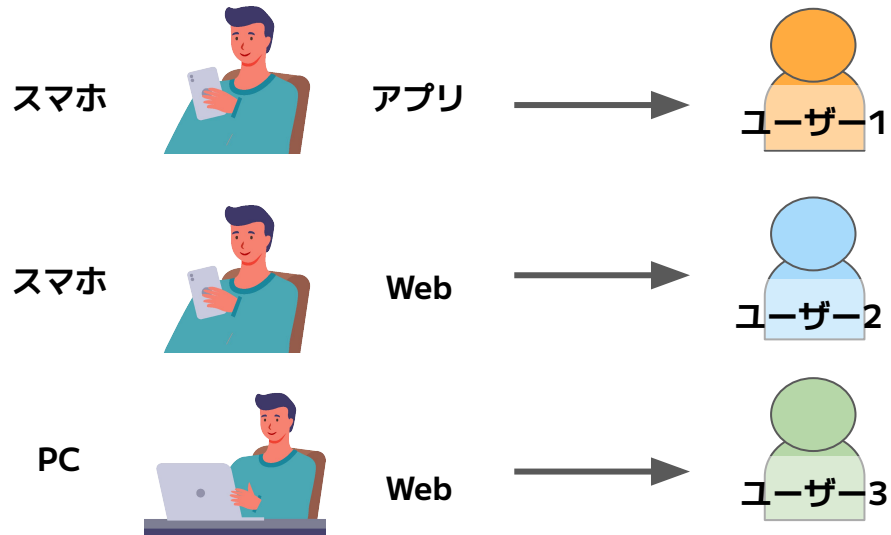


レポート用識別子ってなに？

おさらい 1 ユーザーを判別・統合する手法は...

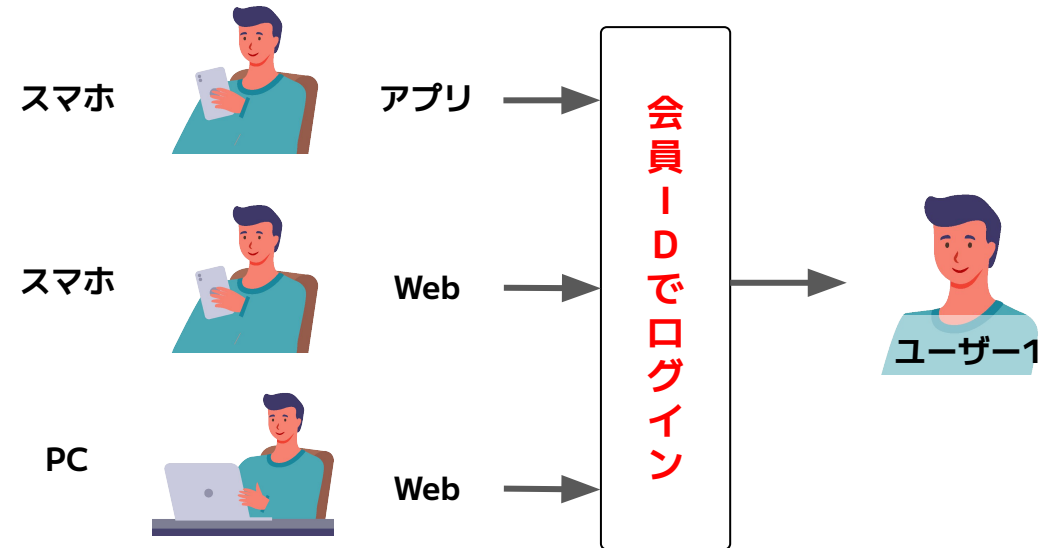
デバイスID

デバイス・ブラウザごとに別ユーザーとして集計



User-ID

会員ID (User-ID) を使って
デバイス・ブラウザが違うユーザーを統合



レポート用識別子ってなに？

おさらい 2

ユーザーを判別する方法「レポート用識別子」とは...

デフォルト

ハイブリッド

User-ID、デバイス ID、モデル化データの順で識別 **(傾向値！)**

計測データ

User-ID、デバイス IDの順で識別

デバイスベース

デバイス IDのみで識別

もっと詳しく知りたい人は、基本編 集中講座「1-1.はじめてのGA4 前編」「1-2.UA経験者向け はじめてのGA4 前編」をチェック！

レポート用識別子によって、ユーザー数は変わる

知ってた？

レポート用識別子を変えると、ユーザー数が変わる

なんで？

ユーザーを判別・統合する手法が変わるから

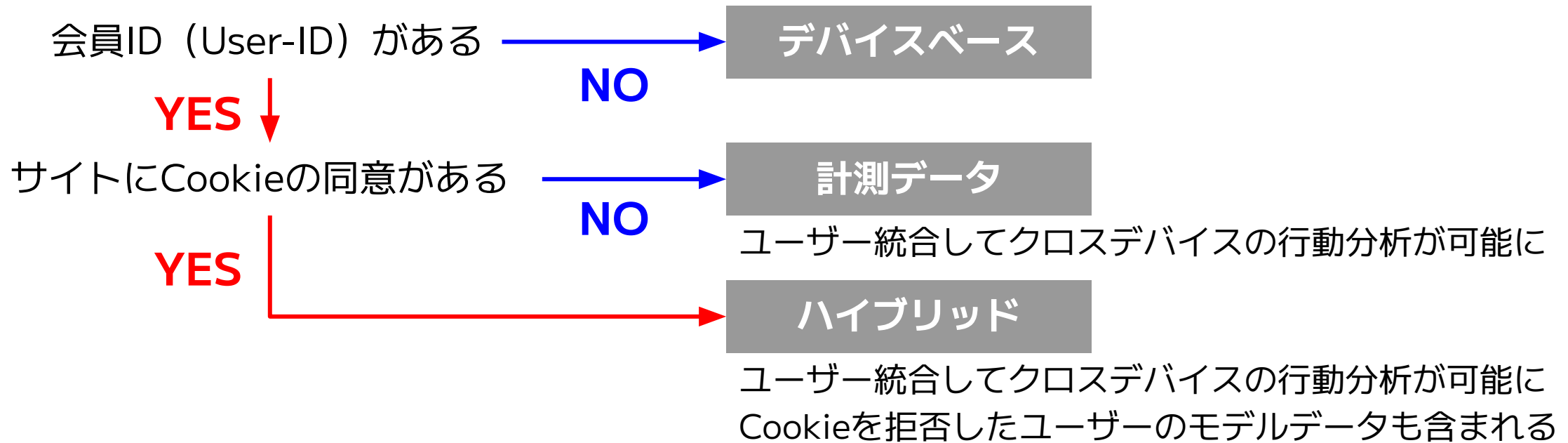
スマホとPCでアクセスした会員を、
デバイス毎に判別して【2人】とカウントするか、
会員IDで統合して【1人】とカウントするかによって、
変わるんだね！



レポート用識別子によって、ユーザー数は変わる

ちなみに

デフォルトは「ハイブリッド」。これが一番良いの??



注意

レポート用識別子の変更は、計測されたデータには影響を及ぼしませんが、対象プロパティ全体に変更が反映されます。利用中の他のユーザーのレポート数値にも影響しますので、あらかじめ周知の上ご対応ください。

ユーザー数は実測値じゃない！？ ～GA4の数値の仕組み～

BigQueryとの数値差異



BigQueryとの数値差異

GA4 HLL++による推定値

- ユーザー
 - デバイスIDから (★) ユニークユーザー数を推定
- セッション
 - ユーザーごとのユニークセッションIDの数を推定



★ポイント

BigQueryと比較する場合、レポート用識別子は「デバイスベース」にする！

2%程度の乖離が発生する

BigQuery 実測値

- ユーザー
 - 「user_pseudo_id のユニーク数」をカウントするクエリを使用
- セッション
 - 「ユーザーごとのユニークセッションID (★) の数」をカウントするクエリを使用



★ポイント

「session_start」イベントの数でカウントしない！

ご視聴 ありがとうございます

