

ソーシャルゲーム×

データ解析

KLab勉強会 2012-06 資料

自己紹介

高田敦史

開発部アーキテクトG+Kラボラトリー所属です
ソーシャルゲーム開発

PHP、Ruby、Python、JavaScript

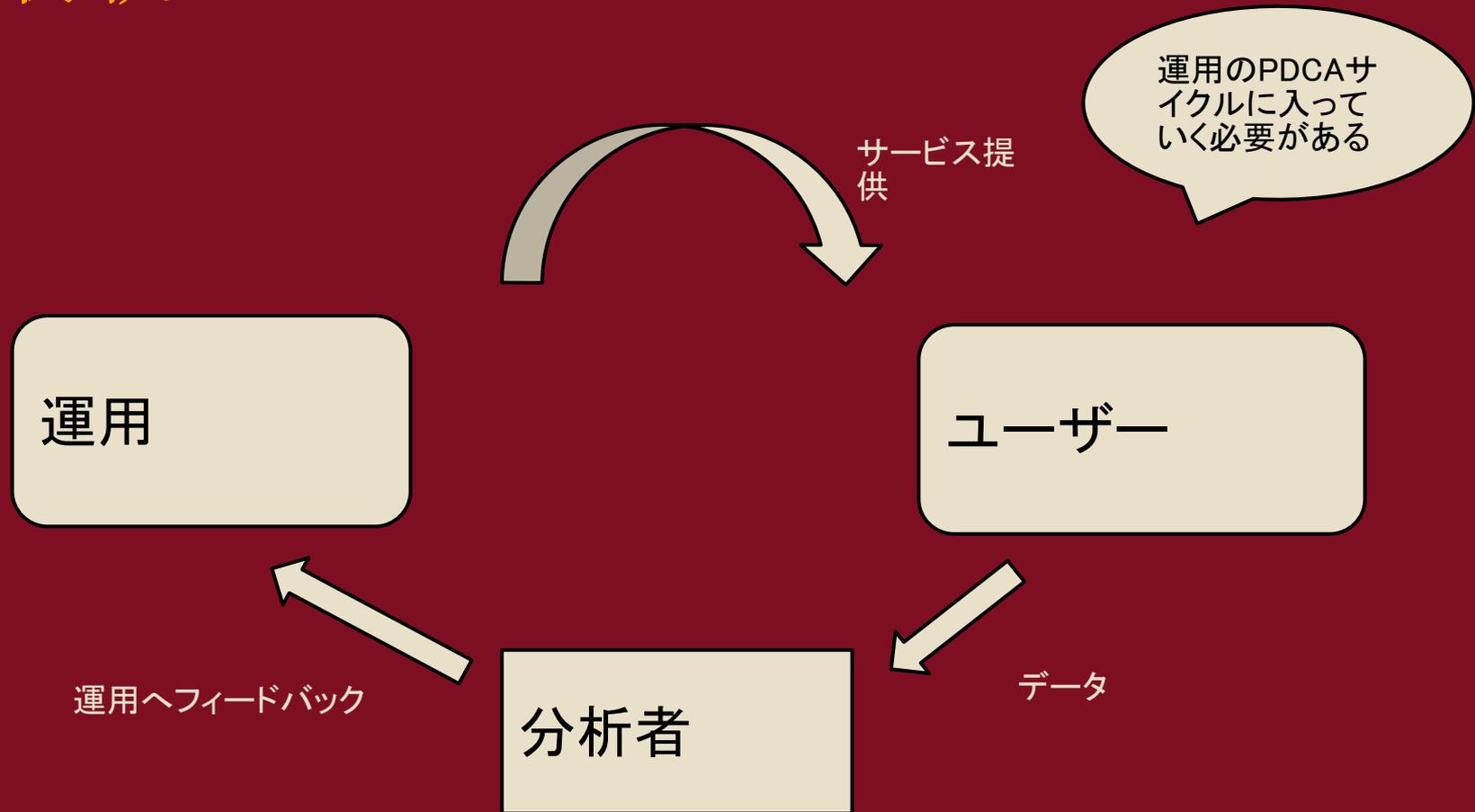
最近はデータマイニングに興味があります

最初に

何のためにデータの収集・解析が必要なのか

- ゲームの運用改善、サービス向上に役立てるため
- 運用・企画のフローの一部に組み込まなければならない
- 技術だけで完結する話でもない

最初に



KLabのソーシャルゲーム

- 12タイトル
- 33サイト
 - モバゲー、GREE、mixi、iPhoneアプリなど
- 1サイトあたり、最大数千万pv/日

↓
膨大なログを要約/可視化する必要がある

体制

BI用ツール(kg_report):

WEBサービス

定常的な売上管理

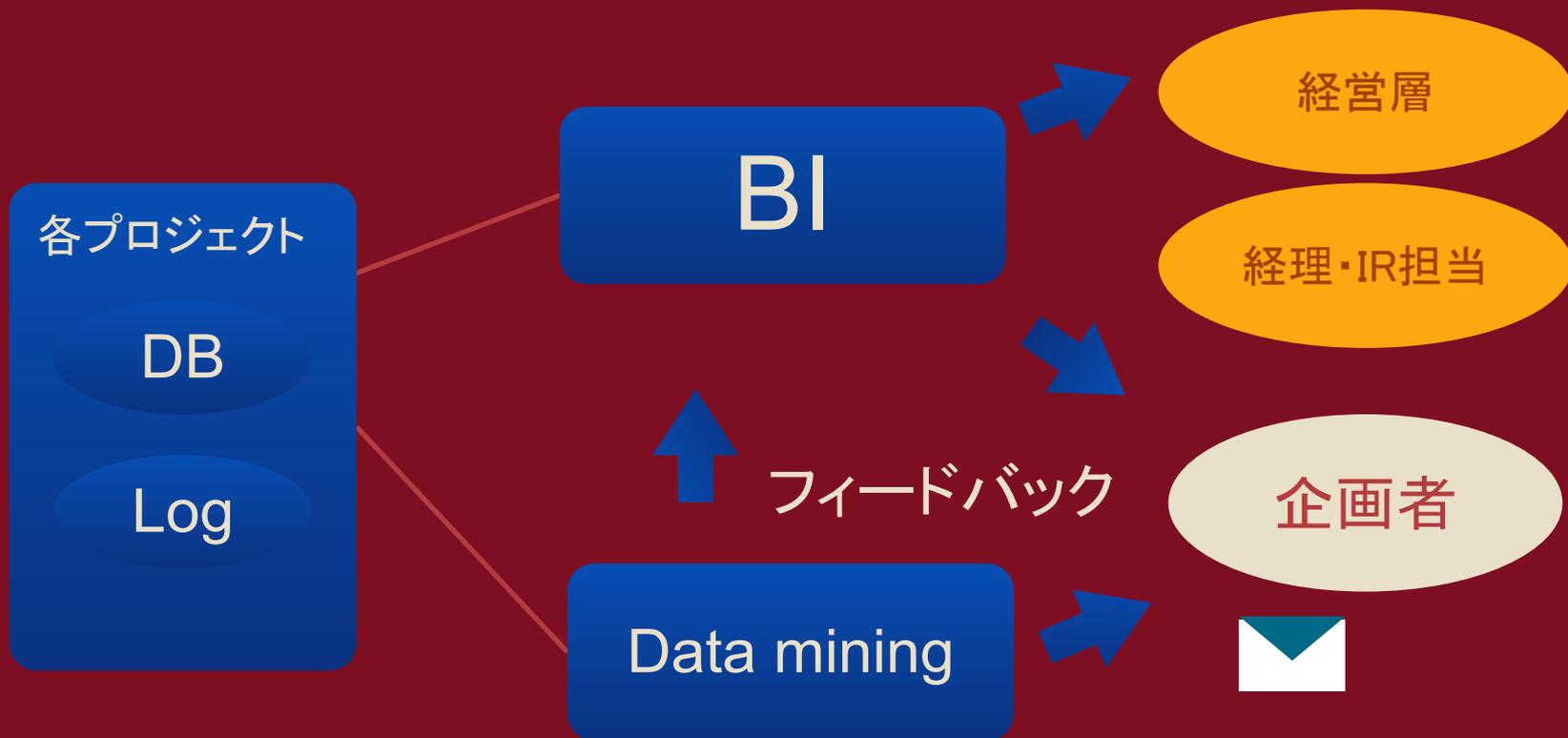
データマイニングチーム(kg_kpi):

マーケティング担当者 + 開発

各ゲームの詳細なデータを収集 + 蓄積 + 分析

新しい分析軸の模索

体制



BIチームのミッション

- kg_reportの開発・運用
 - 社内限定のWebサービス
 - 経理方面の人々が売上の確認に利用する
 - ゲームの企画・運用にも利用されているが、メインではない。

BI: ビジネスインテリジェンス

売上管理の側面

- 全プロジェクト網羅的であることが重要。
- 正確であること重要。
- 必要な数字は合計だけで、厚い情報は必要ない。
- 担当者が確認できるツールが必要。

kg_report

TOP カレンダー ユーザーレポート 売上レポート アンケート

mxixi 各アプリ合算

売上レポート(継続ユーザ別)

集計単位 日毎 週毎 毎月

期間 2012/05/31 ~ 2012/06/14

アイテムカテゴリ 全体

検索

2012/05/31 (木) ~ 2012/06/14 (木)

日付	SNSポイント	インストール数	UU	課金UU	課金率(%)	ARPPU	ARPU	PV	1日継続ユーザ			3日継続ユーザ			7日継続ユーザ			14日継続ユーザ			30日継続ユーザ			メモ
									ARPPU	ARPU	売上	ARPPU	ARPU	売上	ARPPU	ARPU	売上	ARPPU	ARPU	売上	ARPPU	ARPU	売上	
2012/05/31 (木)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/01 (金)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/02 (土)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/03 (日)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/04 (月)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/05 (火)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/06 (水)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/07 (木)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/08 (金)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/09 (土)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/10 (日)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/11 (月)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/12 (火)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/13 (水)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	
2012/06/14 (木)	1,000,000	10,000	100,000	100,000	10%	100,000	10,000	1,000,000	100,000	100,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	1,000,000	1,000,000	10,000,000	

※売上（SNSポイント）に関して

- ・ 1円 = 1 SNSポイントです。
- ・ 消費税込みの数値となります。
- ・ あくまで概算のため、PFへの各種手数料等は考慮していません。
- ・ モンスターバスターは、モンパスゴールドは含まれていません。

※各数値に関して

課金UU = 課金したユニークユーザ数, 課金率 = 課金UU ÷ UU, ARPPU = SNSポイント ÷ 課金UU, ARPU = SNSポイント ÷ UU

※インストール数に関して

・ Appleのアプリにつきましては、厳密には「インストール数」ではなく、「初回起動数」となります。

データマイニングチームのミッション

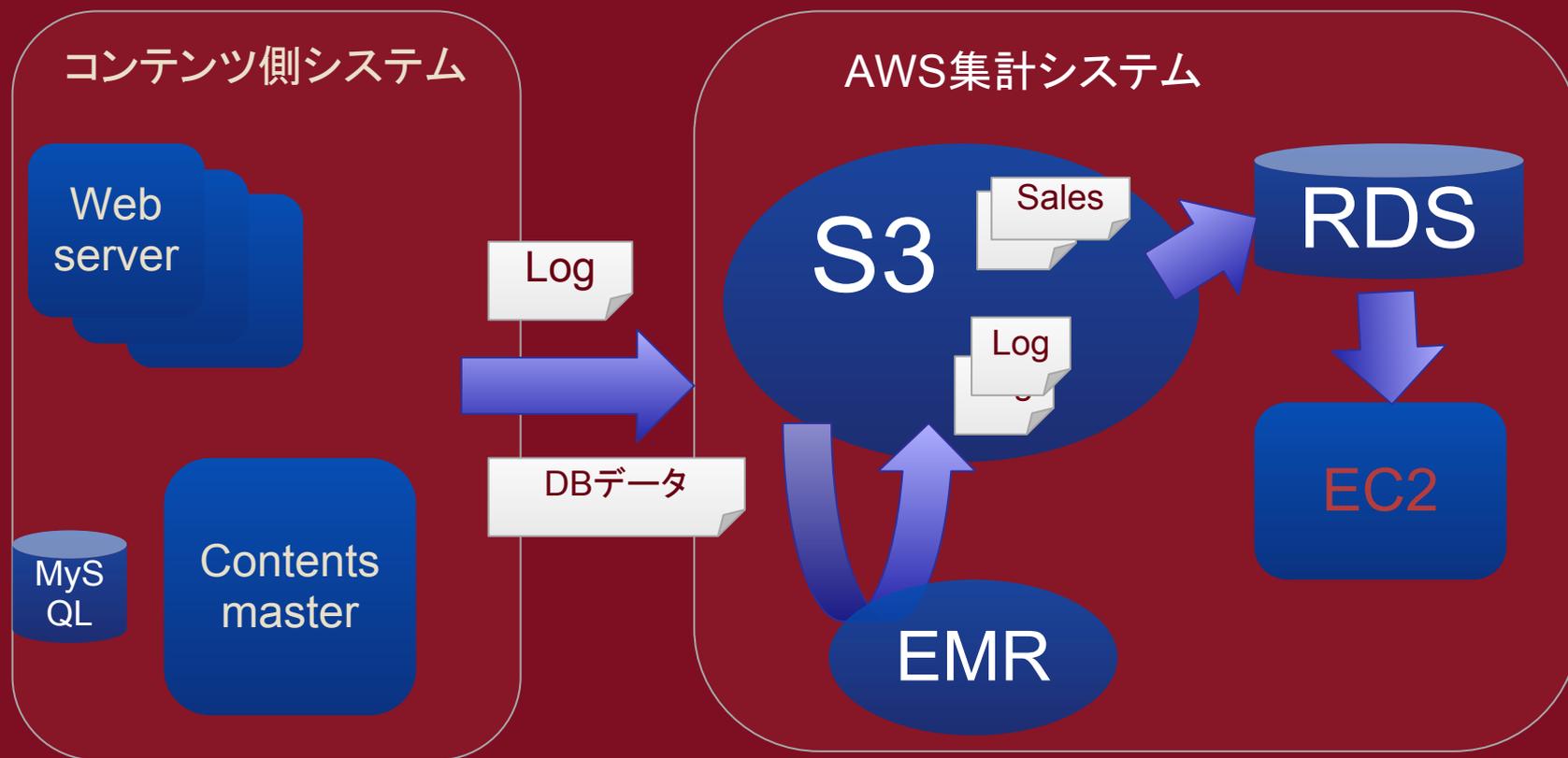
- 開発
 - 多くのゲームのデータを統一フォーマットで収集
 - ・蓄積 → 分析に利用できるように
 - 定常的なレポートの整備
- 企画
 - データを分析し、企画・運用チームへフィードバック → 改善へ
 - 定常的なレポート項目の追加

※ KPI: key performance indicator

ゲームの運用に役立つ情報

- 分析に必要なになったときに、使える状態に
 - 最低限蓄積されていればよい
- 深い・細かい情報
- 標準化されたフォーマットで蓄積

KPIシステム構成



システム構成

- 集計スクリプト(python)
 - 各プロジェクトのサーバーで稼動し、アクセスログ+DBデータをシリアルライズ→送信
 - データフォーマット
 - メタデータ + タブ区切り
 - 売上ログ
 - アクセスしたユーザー
 - 招待
 - 新規インストール
 - レベルのスナップショット
 - 登録日、スマートフォン/フィーチャーフォン、UAなど
 - アクセスログ

システム構成

- AWSデータストレージ
 - EC2: スクリプト動作用
 - 集計・データのチェック
 - メールレポート送信
 - 分析用マシンとしても稼動
 - S3: アクセスログおよびKPIデータをファイルで格納
 - RDS: 分析用にデータを格納
 - EMR(Elastic MapReduce):
 - ログをバイナリフォーマットに変換

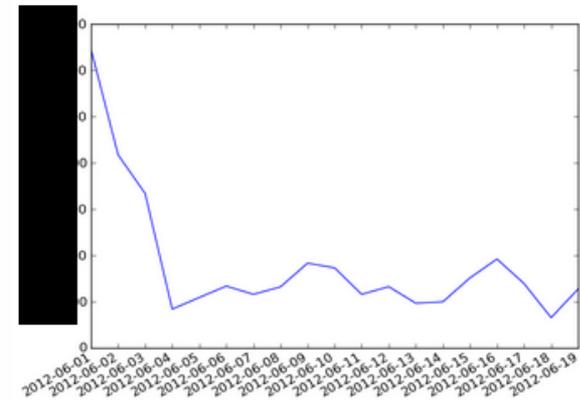
HTMLレポート

- HTML静的ファイル
 - グラフと基本データの表
- 基本KPI
 - 売上、DAU、新規登録など
- アクセスデータ
 - 滞在時間
 - パス別滞在時間
- 登録日別データ
 - 登録日別売上、UU

HTMLのメリット

- 迅速に項目追加
- HTMLテーブルはエクセルにコピペできる
 - 分析者が利用しやすい
- 図を載せられる

売上(日別)

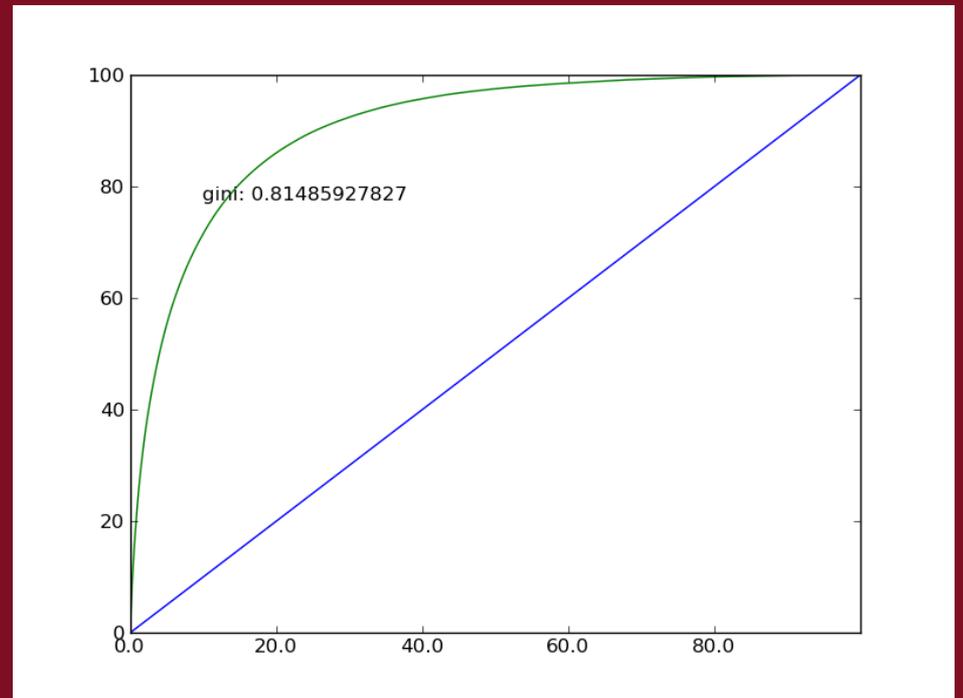


日付	売上	DAU	課金者数	課金率	ARPPU	ARPU
2012-06-01						
2012-06-02						
2012-06-03						
2012-06-04						
2012-06-05						
2012-06-06						
2012-06-07						
2012-06-08						
2012-06-09						
2012-06-10						
2012-06-11						
2012-06-12						
2012-06-13						
2012-06-14						
2012-06-15						

ジニ係数

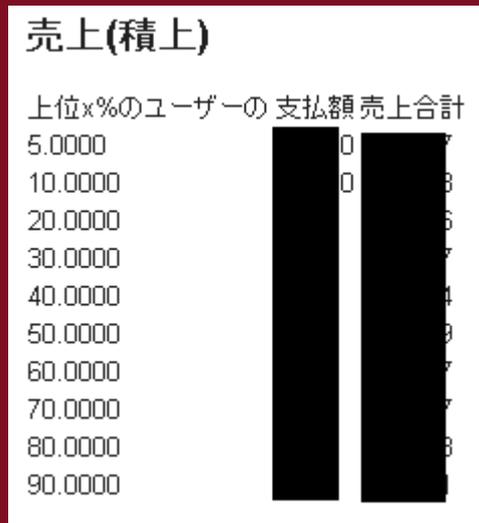
ジニ係数:
主に所得格差の指標として
使用される

ここでは購入額の差を数
量化



デシル分析

上位x%のユーザーで売上の何%をしめるか



アクセスログのバイナリ化

msgpack:

JSONのような構造化バイナリフォーマット

<http://msgpack.org/>

- opensocial_viewer_idごとにファイルを100分割
- 4カラムのバイナリデータに変換

ログ検索
の高速化

時刻 (unixtime)	opensocial_vi ewer_id	URL (path)	GET param
------------------	--------------------------	---------------	--------------

Elastic MapReduce

Amazonのインフラ上で動くHadoop

13台の仮想マシンによる分散処理で、ログを変換

\$0.06 per hour × 13

統計パッケージの利用

- **R**
 - オープンソースの統計パッケージ
 - 独自の言語を使用
 - プログラマーには扱いやすい
- **SPSS**
 - IBM製、有償
 - GUIで統計解析を実行
 - 企画・マーケティングチームの利用を想定

統計パッケージの役割

- 統計パッケージ上で、データを組み合わせ、解析にかけることで、様々な仮説をスピーディに検証できる
 - 前提として、データが取得できる状態にならない
- 有益なレポートは定常レポートへ
 - Python: numpy + matplotlib
- 運用チームへのフィードバックも重要

質問など