

「単車の虎」に於ける アクセスログ解析手法

On Hadoop !

株式会社Donuts

システム開発部

インフラエンジニア 今井陽太

Donuts Co. Ltd.

1. 自己紹介・タントラ紹介
2. サーバ構成・データ構造
3. 圧縮方式と容量・処理速度
4. 解析していること
5. Hadoopを導入してみた

Donuts Co. Ltd.

2

今井 陽太

株式会社 Donuts所属
インフラ 兼 情シス (兼雑用) 担当

出生地: 北海道帯広市
学歴: 釧路高専@北海道 ネットワーク技術専攻

その後もネットワーク機器のResellerやMaker、
某SNS m社さんのネットワークチームで働いていたNWエンジニア

が、何故かData Miningっぽいことを語ります。

Donuts Co. Ltd.

3

- 弊社メインプロダクトの一つ
 - 不良となってバイクを改造し、プレイヤー同士での「タイマン」に挑む
 - チーム「族」間での「抗争」
- RAIDボスなどを除けば基本Player vs Player のゲーム



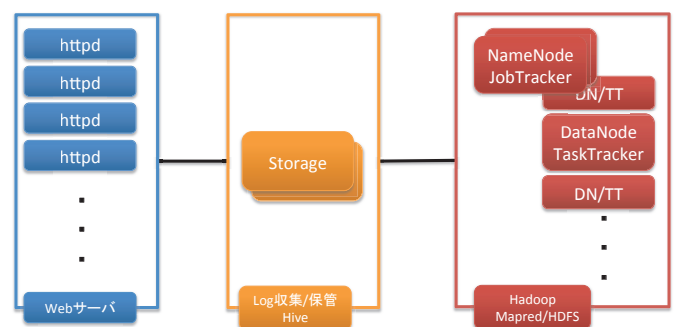
Donuts Co. Ltd.

4

サーバ構成・データ構造

Donuts Co. Ltd.

5



Donuts Co. Ltd.

6

データの形式と流れ

Donuts

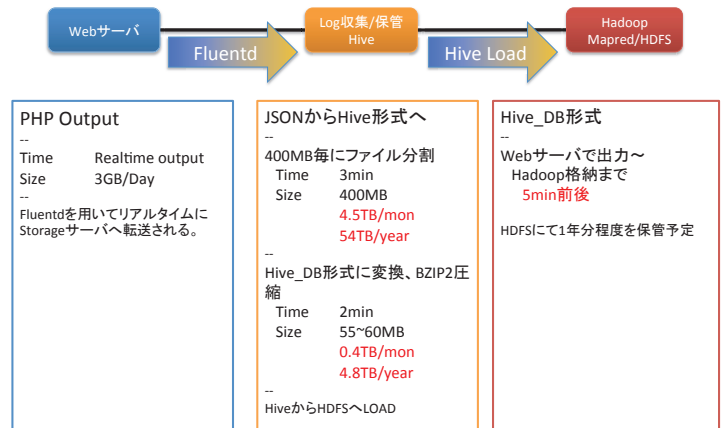


Donuts Co. Ltd.

7

データの処理時間とサイズ

Donuts



Donuts Co. Ltd.

8

圧縮形式の比較

Donuts

- 非圧縮データは処理が早いですがStorageコストが非常に高くなる。
- Bzip2はStorageコストは抑えられるが解析処理コストが高い。
- 長所短所を把握してしっかり設計してあげるとGzipがバランス良さそう。

圧縮形式	データサイズ(圧縮率)	処理時間
Raw data	1650MB (54TB/year) : 100%	31s : 100%
Gzip	200MB (6.55TB/year) : 12%	47s : 151%
bzip2	160MB (5.24TB/year) : 9%	61s : 196%

Donuts Co. Ltd.

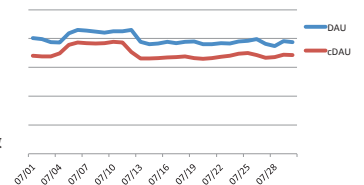
9

アクセスログ解析 by Hive

Donuts

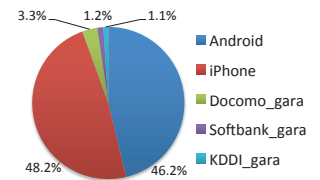
解析しているコト(定期処理)

- Daily/Monthly Active User
継続ActiveUser数
(昨日も今日もログインした人の数)



- 端末シェア
 - まだまだガラケーも健在

- 機能別UU/回数
 - タイムンUU/回数
 - ダチ申請UU/回数
 - 機能毎に数値の上下を確認して、イベント/キャンペーンの効果分析とか



Donuts Co. Ltd.

10

使っているクエリ by Hive

Donuts

解析しているコト(定期処理)

- Daily/Monthly Active User
 - `select count(distinct player_id) from access_log_201308 where dt='20130811'`
テーブルは月別 日別にPartition化
- 端末シェア
 - `select distinct a.player_id, user_agent from access_log_201307`
この結果をshellスクリプトに渡して機種切り出しやcountを別に行う
- 機能別UU/回数
 - タイムンUU/回数
 - `select count(distinct player_id) from access_log_201308 where dt='20130811' and request_uri like '%/battle/btl%'`

Donuts Co. Ltd.

11

Hadoop導入してみて - 反省とか感想とか

Donuts

- bz2圧縮でもかなり時間短縮できた
 - Dailyの解析が10分程度、Monthlyは現状6時間程度で終わる。
- NameNode壊れたら全データが飛ぶ恐怖
 - 念のため半年分はStorageサーバに別に保管してある
- 適切な設計、適切なクエリ
 - URIやQuery-StringをKeyにしたいときはOutputが大きくなりがち & 事後修正が必要になって手間
 - Monthlyで油断して適当にOutputさせると数百GBのデータがローカルに。。。メモリに収まらなくて4時間たった当たりでエラー終了したりもする。
- フクザツな処理はHiveだけでは完結しないことが多い。
 - どうしてもSQLに縛られるので、出来ないことも出てくる。

Donuts Co. Ltd.

12



ご清聴ありがとうございました！