

2016.02.08

Spark Streamingで作る、 つぶやきビッグデータのクローン

GMOインターネット 次世代システム研究室

(所属コミュニティ: 秋葉原IT戦略研究所)

野田純一

Overview

- 1.自己紹介 私とSparkの関わり
- 2.目的
- 3.Spark Streamingについて
- 4.検証サービス説明
- 5.Spark Streamingを使用したオンラインTwitter解析によるレコメンドシステム
「Mikasa」 -三笠
- 6.まとめ

自己紹介

所属

GMOインターネット

業務分野

ビッグデータ、機械学習
アドテク



- 実業務でSpark使っています

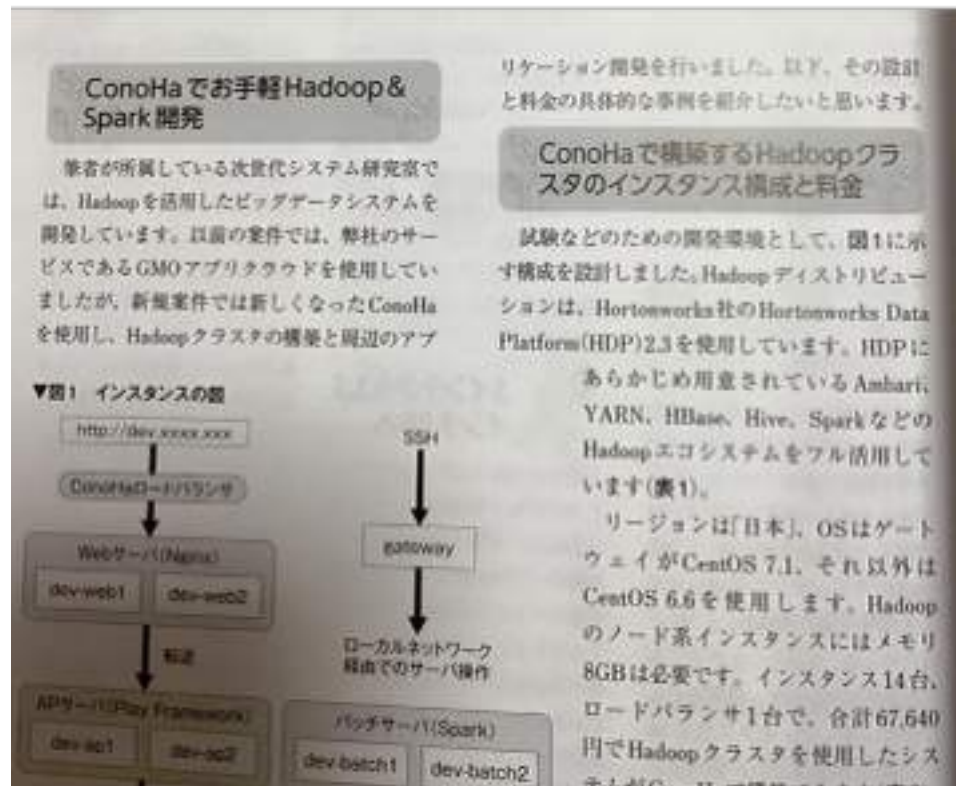
WEBメディアのユーザー回遊を最大化！
Growth & Monetizeレコメンドエンジン



金額シミュレーション



- Sparkについて少し執筆しました



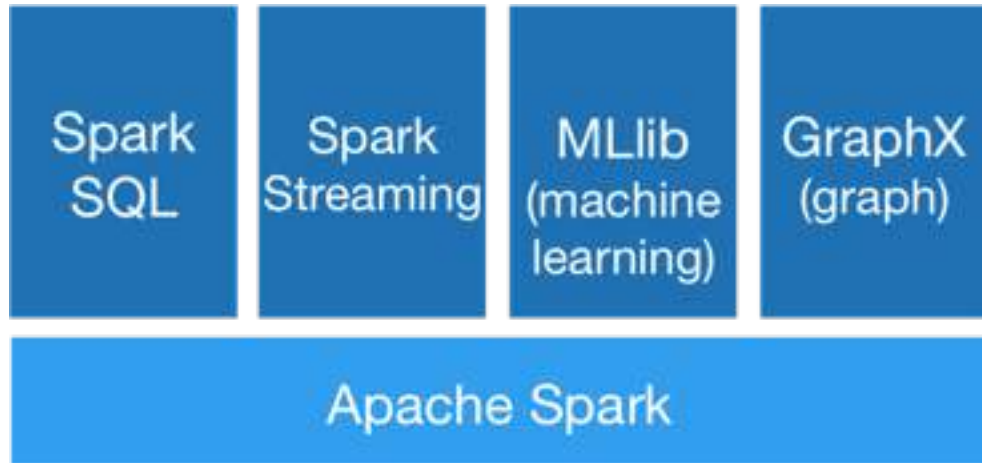
- 「Software Design 2015年11月号」
- ConoHaを使ったHadoopクラスタの構成例とSparkSQLをつかったデータ処理の実例が記載。

↓これが作りたい



2014 7月22日 NHK

3. Spark Streaming について



- Sparkサブシステムの一つでファイルからでなくリアルタイムに流れてくるストリームソースに対しての集計を対象としている。
- 直近1時間のツイート数を毎分集計する、直近3時間でアクセスが多いIPを集計する・・・などのウィンドウ集計が可能。

3. Spark Streaming ～データソースについて



その他、`ssc.socketTextStream("localhost", 7777)`

とするとローカルポートをソースにできるのでサーバーのログを流し込んでリアルタイムにエラー、クラックIPの集計などが可能。

4. Spark Streamingを使用したオンライン Twitter解析によるレコメンドシステム 「Mikasa」 - 三笠

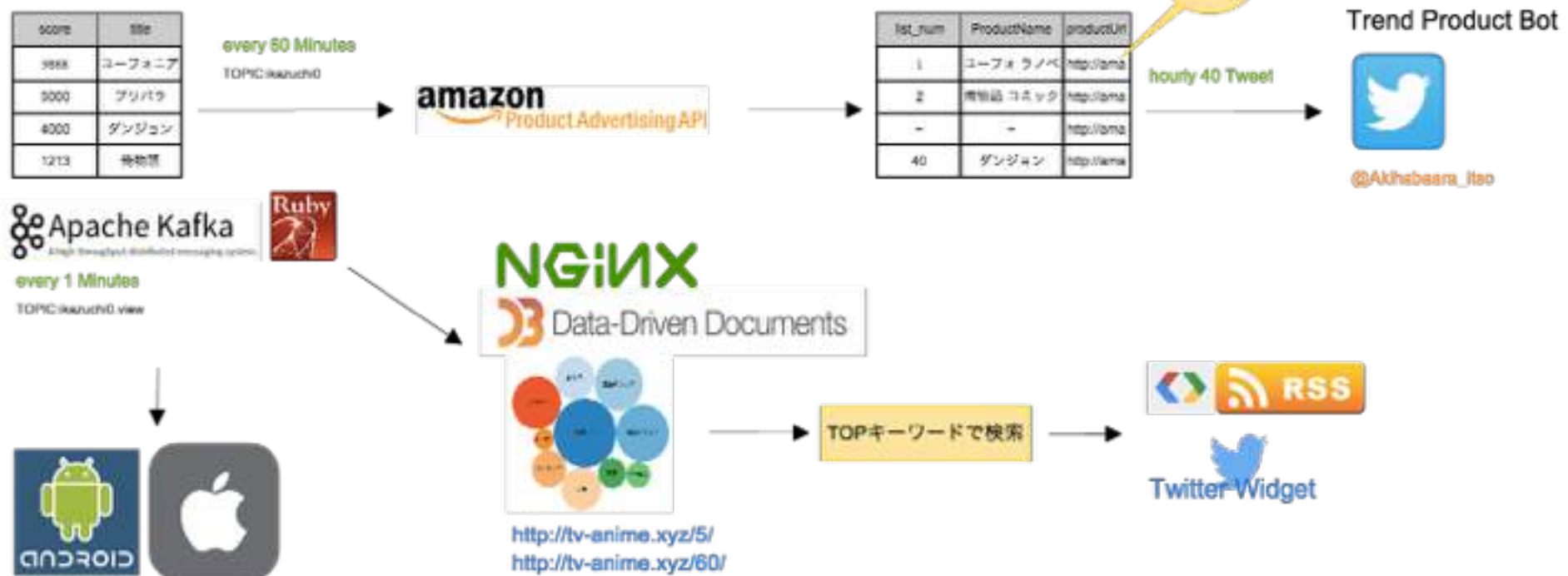
https://github.com/AKB428/mikasa_ol

https://github.com/AKB428/mikasa_rs

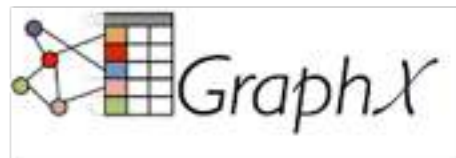
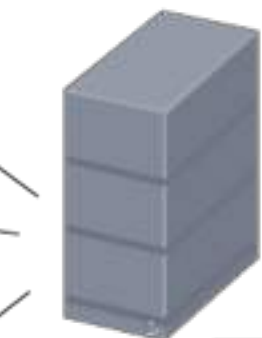
Online Layer



Recommendation System



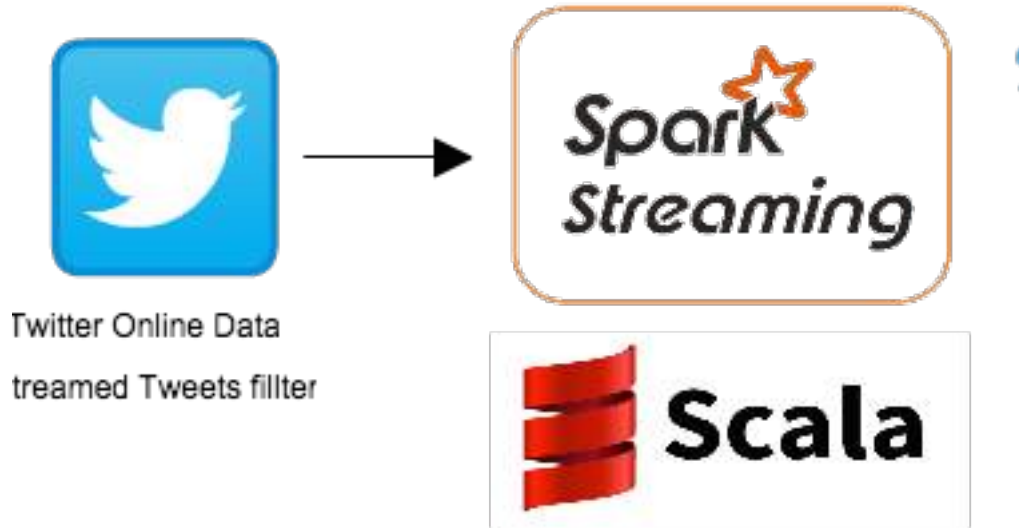
完全スタンドアロン構成



Online Layer



Recommendation System



- Twitterストリームからデータを取得する。
- Twitter Streaming APIを使う。
- 400の検索キーワードが指定可能。
- Spark Streamingでリアルタイム集計を行う。

Recommendation System

score	title
9888	ユーフォニア
5000	プリバタ
4000	ダンジョン
1213	怪物話

every 60 Minutes
TOPIC:ikazuchi0



list_num	ProductName	productUrl
1	ユーフォ ノベ	http://ama
2	怪物話 コミック	http://ama
-	-	http://ama
40	ダンジョン	http://ama

-22

hourly 40 Tweet

TOPIC:ikazuchi0.view

Trend Product B



@Akihabara_itso



every 1 Minutes
TOPIC:ikazuchi0.view

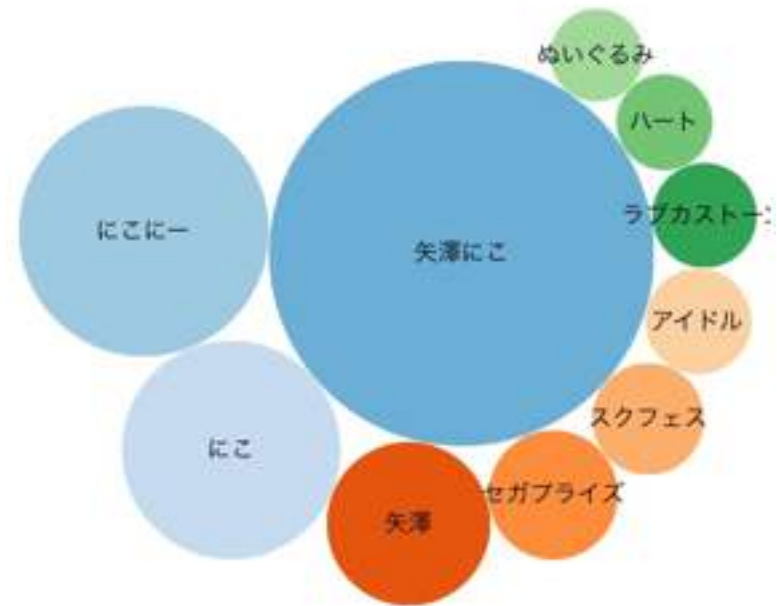


TOPキーワードで検索



<http://tv-anime.xyz/5/>
<http://tv-anime.xyz/60/>

2015年7月22日



直近5分の発言数の大きさ。更新日時 2015年07月22日 09時38分(1分に一度更新)

画面デモ（直近5分を毎分集計）

(F1)2015年冬期 TVアニメ <http://tv-anime.xyz/5/>

(F2)ラブライブ <http://lovelive-net.com/5/>

(F2)2016年春期TVアニメ <http://akiba-net.com/5/>

(F2)関東TV番組 <http://telev.net/5/>

F1=ユーザー辞書に登録してある単語のみ集計

F2=ユーザー辞書外の単語も集計

画面デモ（直近60分を毎分集計）

(F1)2015年冬期TVアニメ <http://tv-anime.xyz/60/>

(F2)ラブライブ <http://lovelive-net.com/60/>

(F2)2016年春期TVアニメ <http://akiba-net.com/60/>

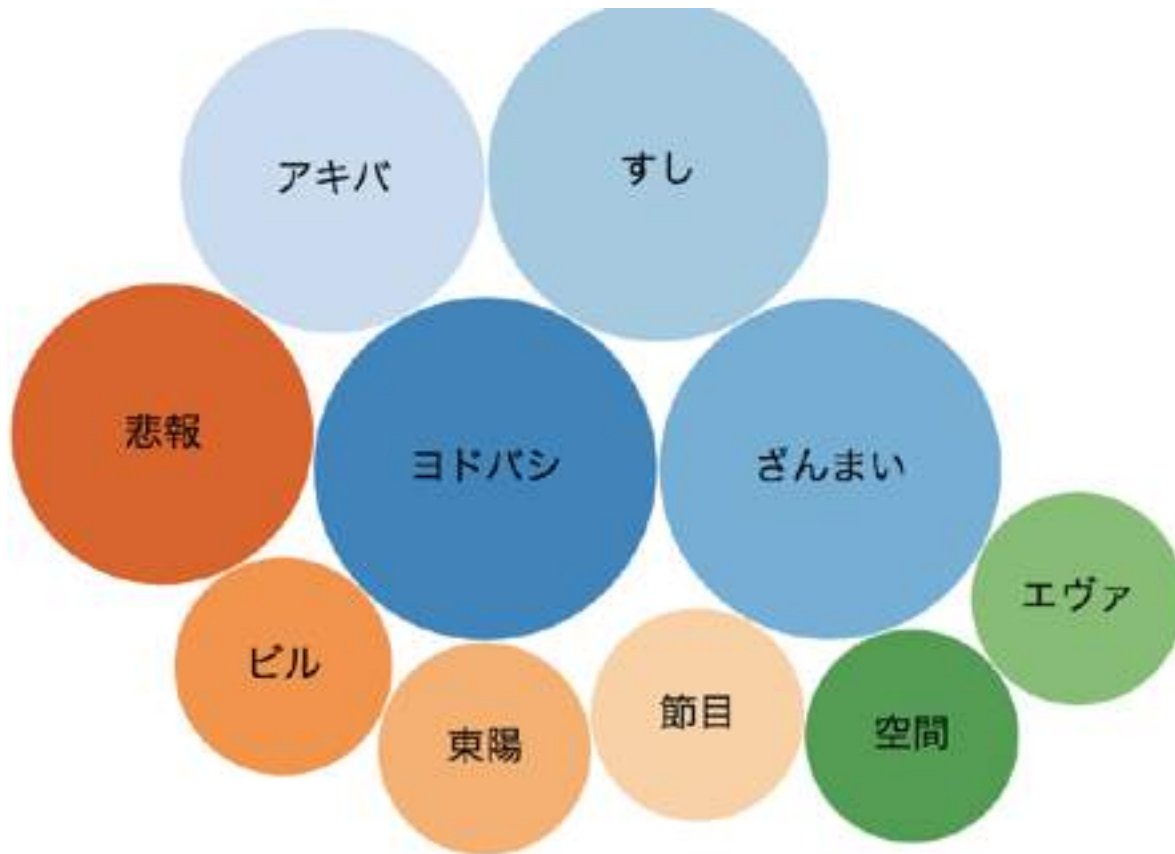
(F2)関東TV番組 <http://telev.net/60/>

F1=ユーザー辞書に登録してある単語のみ集計

F2=ユーザー辞書外の単語も集計

システム・デモ

画面デモ 過去の履歴キャプチャ



ヨドバシアキバ1F寿司屋「すしざんまい AKIBA店」、8月16日で閉店！ 8F「廻るすしざんまい AKIBA店」も10月1日まで

アキバ総研 | 2015年07月23日 21:03

タグ：[グルメ](#) [開店/閉店/移転/リニューアル](#)

ツイート 110 いいね! 49 B! 1 G+1

ヨドバシカメラ マルチメディアAkibaで発見。

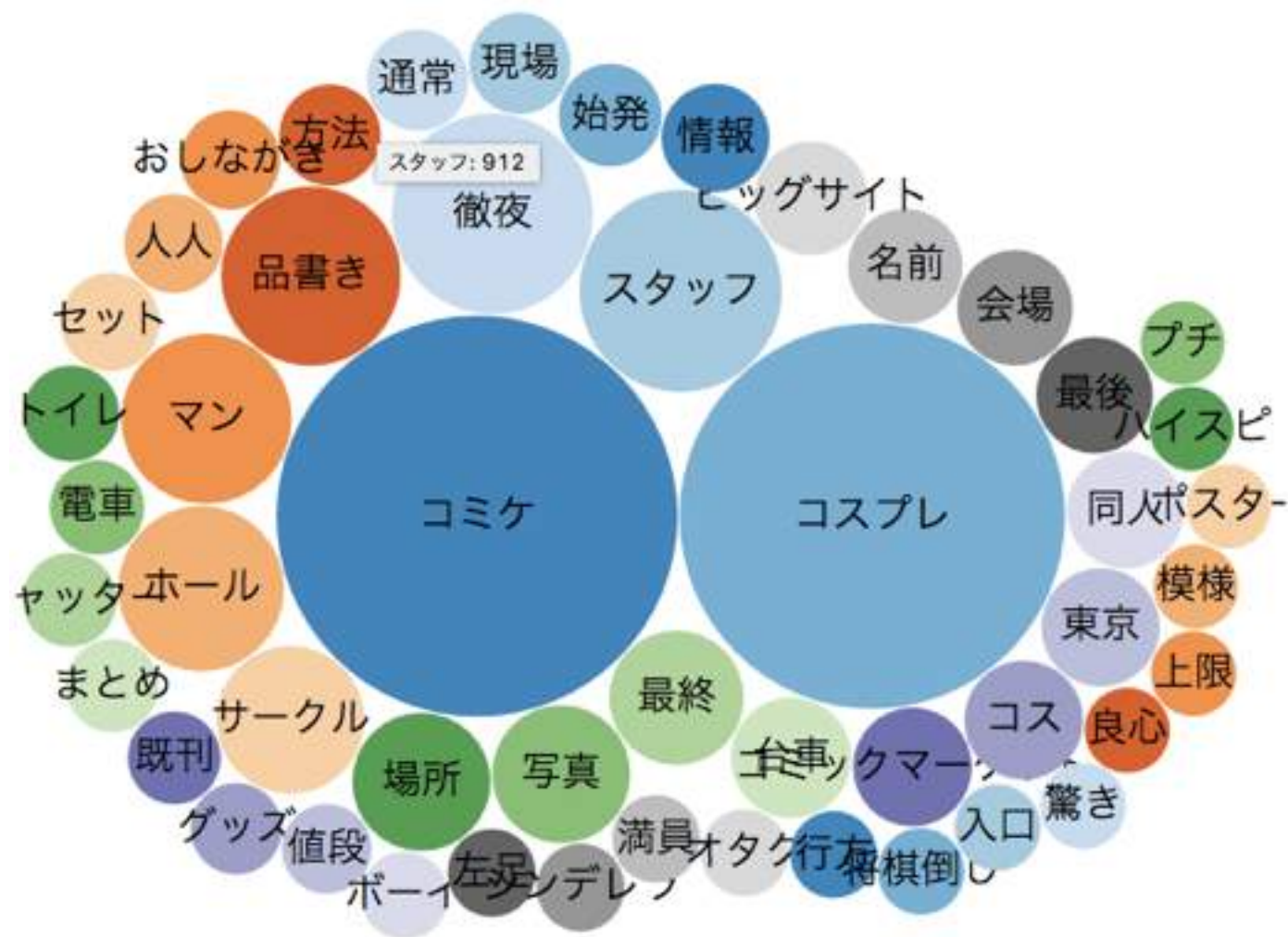
1Fに入っている寿司チェーン「すしざんまい AKIBA店」が8月16日で閉店する模様。

閉店理由は、テナント契約の期間満了。

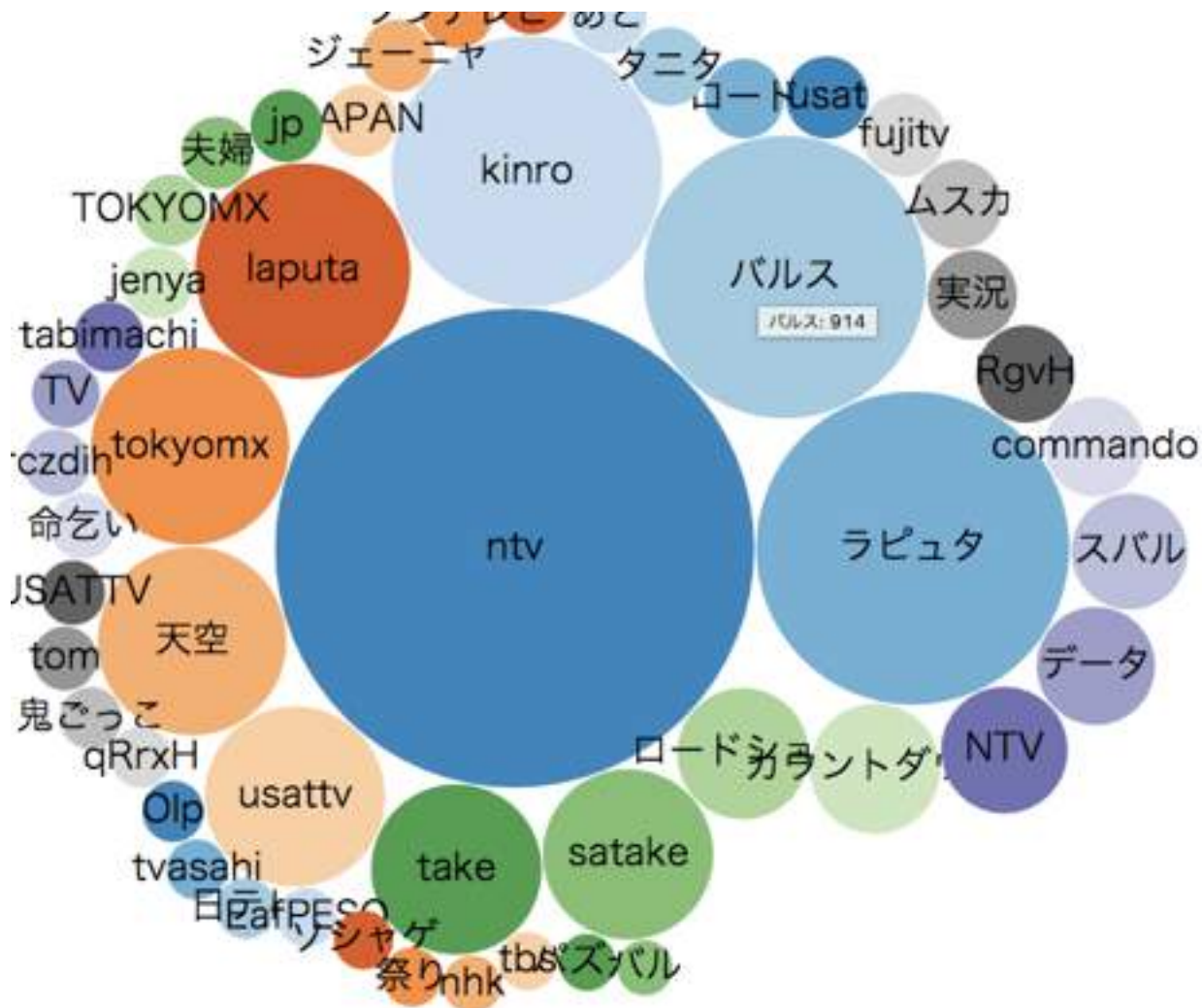
同店は、ヨドバシアキバの開業当時（2005年9月）から、24時間/年中無休で営業していた。



- コミケ 2015 12/31 朝8時頃



• 最近のバルス



直近の動き

直近5分の発言数の大きさ。更新日時 2016年01月15日 23時24分(1分に一度更新)

デモ：リコmendTwitterボット

https://twitter.com/Akihabaara_itso

<https://twitter.com/AnimeBigdata/>



ツイート 13,354 フォロワー 40 フォロワー 9

ツイート ツイートと返信 画像 / 動画

 **akiba-itso** @Akihabaara_itso · 15分
シドニアの騎士 / KNIGHTS OF SIDONIA:
SEASON 1 価格: ¥ 8,497 発売日:
[amazon.co.jp/dp/B00U0WY482/...](https://amazon.co.jp/dp/B00U0WY482/)

実際の売上

売上レポート

[目録](#)

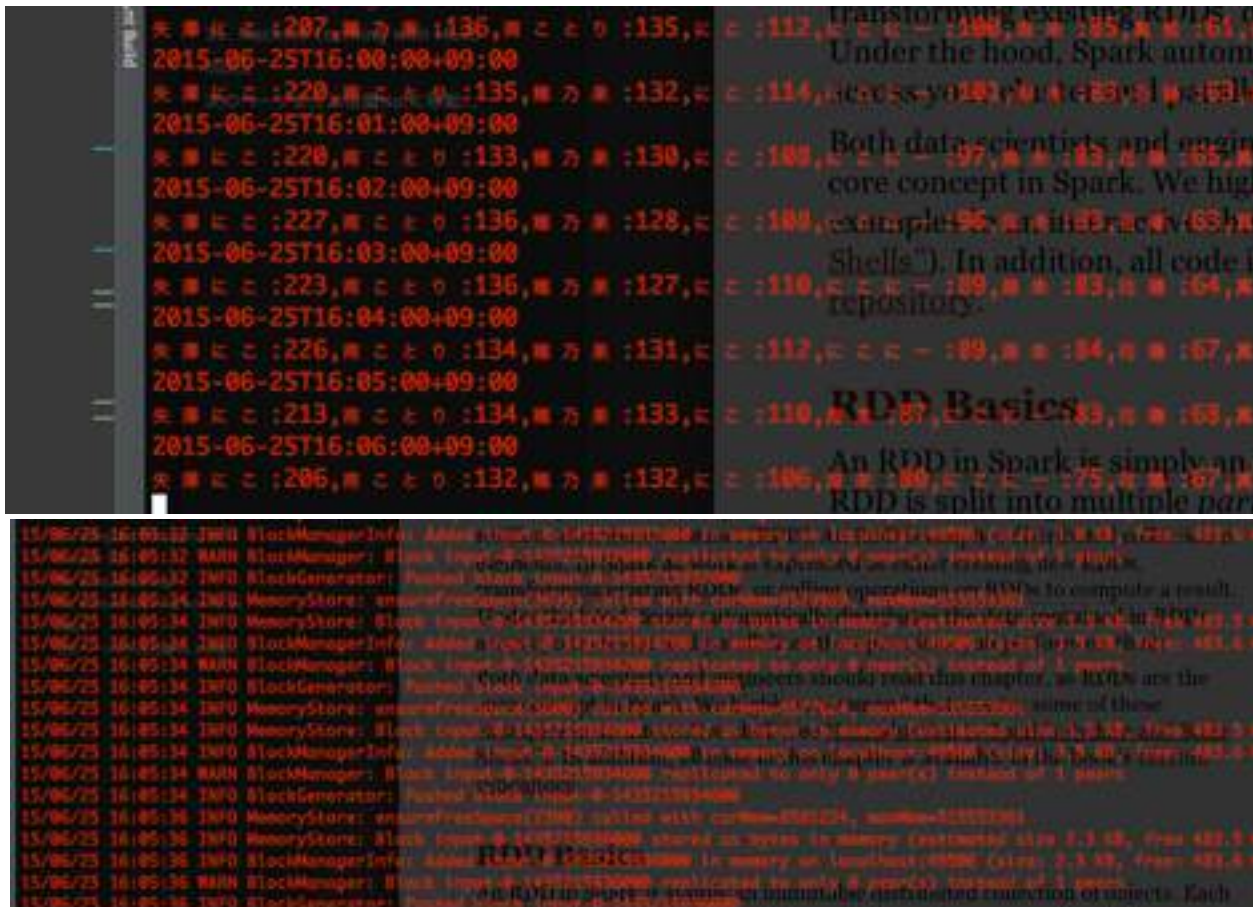
2015年7月1日 - 2015年7月30日

[サマリー](#)

商品名 ▼	出品者 ▼	価格 ▼	紹介料率 ▼	発送済み商品 ▼	売上 ▼	紹介料 ▼	デバイスの種類 ▼
	マーケットプレイス	¥1,852	3.02%	1	¥1,852	¥56	パソコン
	マーケットプレイス	¥2,204	2.99%	1	¥2,204	¥66	パソコン
	マーケットプレイス	¥3,990	7.99%	1	¥3,990	¥319	パソコン
	Amazon.co.jp	¥657	3.04%	1	¥657	¥20	モバイル
	マーケットプレイス	¥1	0.00%	1	¥1	¥0	パソコン

サーバーデモ

動作中のサーバーのコンソールを公開



Mikasaインストールについて

インストールマニュアル

<https://gist.github.com/AKB428/c30bc6a979e05fa3a022>

- TwitterAPIとAmazonAPIのアカウントがあれば**1時間**でセットアップ完了。
- AmazonAPIはリコメンドしない場合は不要。
- TwitterAPIのアカウント取得も**10分程度**で可能。

Mikasa OL

https://github.com/AKB428/mikasa_ol

Mikasa RS

https://github.com/AKB428/mikasa_rs

• 10. まとめ

- 作ったシステムはスタンドアロンでも安定して動く
(4ヶ月放置しても安定動作 Spark1.4、Spark1.5)
- 視覚的にわかりやすい、作りやすいのでSpark入門用の教材に最適、会社ではインターン生にCDH、HDPクラスタを構築してもらい、クラスタ上で動作するよう作ってもらったりしている。(分散処理を体験したい学生向け)
- イベントのハッシュタグを監視し、イベントの展示モニタなどに使うといい感じ。(OSCでは実際使ってもらっている。
- 辞書はmecab-ipadic-neologdを導入したい。

4月9日(土)にイベント主催します！

The screenshot shows the Connpass event page for 'ShangirLa Meetup 第3回'. The page is in Japanese and features a navigation bar at the top with 'connpass' logo, a search bar, and links for 'ダッシュボード', 'カテゴリ一覧', and '新着イベント'. Below the navigation bar, there are buttons for 'イベント管理' and 'イベント作成'. The main content area includes social media sharing options for Bitly, Google+, Facebook, and Twitter. The event title is 'アニメと秋葉原をITでハックする「ShangirLa Meetup 第3回」', organized by '秋葉原IT戦略研究所'. A featured image shows a speaker at a presentation with text about 'オタク業界のIT化促進' and 'オタク産業のIT化は黒船企業に勝ちたい'. The event date is '2016/04/09(土)' from 16:00 to 18:00. A table at the bottom shows the event details: '募集内容' (Recruitment Content), '参加料' (Participation Fee) of 1000 yen (venue included), and '先着順' (First-come, first-served) with 4/12 spots available. A sidebar on the right provides group information for '秋葉原IT戦略研究所', including 'イベント数 1回' and 'メンバー数 9人'. A red button 'このイベントに申し込む' (Apply for this event) is prominently displayed, with a '募集期間' (Recruitment Period) from 2016/03/06 to 2016/04/08. A 'イベントへのお問い合わせ' (Contact for event) link is also present.

connpass
イベント検索

ダッシュボード カテゴリ一覧 新着イベント

イベント管理 イベント作成

公開中 申込者进行管理する メッセージを送る イベント統計を見る イベントを告知する イベントを編集する

Bitly 0 Google+ 0 いいね! 0 ツイート

4月 9 アニメと秋葉原をITでハックする「ShangirLa Meetup 第3回」

主催: 秋葉原IT戦略研究所

オタク業界のIT化促進
ビッグデータからアニメに関するトピックを解析する際に必要となる今期アニメ作品のマスターが取得できるAPIの開発

オタク産業のIT化は黒船企業に勝ちたい
「僕(日本人)がアニメをうまく扱えるんだ!!」

同人イベントでのエンジニア率の向上

ハッシュタグ: #akibalab

募集内容	参加料 1000円(会場払い)	先着順 4/12人
------	--------------------	--------------

グループ

秋葉原IT戦略研究所

イベント数 1回
メンバー数 9人

2016/04/09(土)
16:00 ~ 18:00

Googleカレンダー iCalendar

このイベントに申し込む

募集期間
2016/03/06(日) 18:30 ~
2016/04/08(金) 18:00

イベントへのお問い合わせ

会場

会場はもちろん秋葉原です

AP CONVENTION ROOM
貸し会議室

貸し会議室のことなるAP貸し会議室へ

▶ 利用規約 ▶ 拠点情報 ▶ 会社概要

▶ お気に入りに追加

お問い合わせ



HOME

会議室を探す

会議室一覧

サービスの強み

ご利用の流れ

活用事例の紹介

お客様の声

よくあるご質問

AP貸し会議室 > 関東エリア > AP秋葉原

CONVENTION ROOM
AP秋葉原

多目的ビジネススペース

AP秋葉原

東京都 台東区 秋葉原 1番1号 秋葉原ビジネスセンター 1-4・5階(貸付1階)

AP秋葉原へのお電話はこちら
お気軽にご連絡下さい。

☎ 03-5289-9109



ハッシュタグ： #akibalah

募集内容

参加枠
1000円 (会場払い)

先着順
4/12人

📄 イベントの説明

会場費**2万**ぐらいかかります



3月26日 アニメxSparkについて話します

Peatix イベント検索

新規登録 / ログイン

当日は参加できないけど、イベントは応援したい! そんなときはカンパしよう! ?

共有用の URL: <http://niconicodatasig8.peatix.com>

ツイート いいね! シェア 50



第八回データ研究会

人工知能と根性で挑むコンテンツの世界

第八回ニコニコ学会βデータ研究会 ～人工知能と根性で挑むコンテンツの世界～

詳細

ユーザ参加型学会組織「ニコニコ学会β」の分科会として「データは面白い」をテーマに研

2016/03/26 (土)

12:30 - 19:00 JST

5/28 名古屋OSCで出展します

オープンソースの「今」を伝える

オープンソースカンファレンス 2016 Nagoya

メインメニュー

トップページ
セミナープログラム
(準備中)
展示ブース一覧
懇親会 (調整中)
お問い合わせ
OSCトップページ

ログイン

ユーザ名:

パスワード:

パスワード紛失

開催概要

日程：2016年5月28日(土) 10:00-18:00 (展示は16:00まで)

会場：名古屋中小企業振興会館 吹上
(地下鉄桜通線 吹上駅から徒歩10分)
[アクセスマップ]

費用：無料

内容：オープンソースに関する最新情報
・展示 - オープンソースコミニティ
・セミナー - オープンソースの未来

主催：オープンソースカンファレンス実行委員会

企画運営：株式会社びびねっと

団体・コミュニティ・企業 出展

OSPN Pressのご案内

メールアドレスを入力



6/18 北海道OSCで出展します

オープンソースの「今」を伝える

オープンソースカンファレンス 2016 Hokkaido

メインメニュー

[トップページ](#)
[セミナープログラム
\(準備中\)](#)
[展示ブース一覧
\(2日目\)](#)
[懇親会 \(調整中\)](#)
[お問い合わせ](#)
[OSCトップページ](#)

ログイン

ユーザ名:

パスワード:

開催概要

日程：2016年6月17日
6月18日

会場：札幌コンベンションセンター
[アクセスマップ]

費用：無料

内容：オープンソース
・展示 - オープンソース
・セミナー - オープンソース

主催：オープンソース

企画運営：株式会社び

団体・コミュニティ



ご清聴ありがとうございました