



AWS Summit

Tokyo





Amazon CloudSearch Deep Dive

Amazon Data Services Japan 篠原英治

2015年6月2日



■ Gold Sponsors



Empowered by Innovation



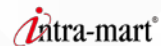
■ Global Sponsors



■ Silver Sponsors



■ Bronze Sponsors



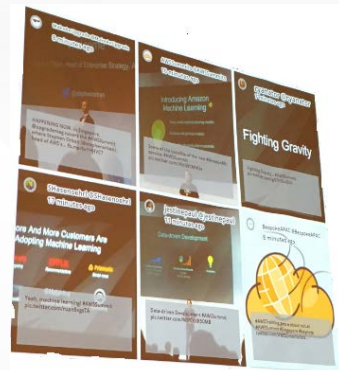
■ Global Tech Sponsors



■ Logo Sponsors



ハッシュタグ **#AWSSummit**
で、皆さんのツイートが展示エリア
の大画面に表示されます



公式アカウント **@awscloud_jp**
をフォローすると、ロゴ入り
コースターをプレゼント



【コースター配布場所】

メイン展示会場、メイン会場1F受付、デベロッパーカンファレンス会場



Agenda

- Amazon CloudSearchとは
 - CloudSearchはフルマネージドなクラウド型検索サービスです
- Inside Amazon CloudSearch
 - CloudSearchの特徴を内部の挙動から理解しましょう
- Amazon CloudSearch事例
 - 日本における様々な事例をご紹介します

自己紹介

- 篠原英治 / しのはらえいじ
 - Twitter: @shinodogg
 - Blog: <http://shinodogg.com>
- AWSソリューションアーキテクト
 - Startup企業様向けの技術支援担当
- Amazon CloudSearch Subject Matter Expert
 - Austinで開催されるLucene/Solr Revolution 2015のCall for Papersに、日本語の言語解析 & CloudSearch事例紹介で応募しました



本セッションにお越し頂きありがとうございます！

- ランチセッションということで
 - 前半はカジュアルな話題から入って
 - 後半に進むにつれてDeep Diveしていきます
 - 是非お食事もごゆっくりお楽しみください



eiji shinohara
@shinodogg

あきんどスシロー lunch session! Sushi box is #yum #awssummit

View translation



eiji shinohara
@shinodogg

Lunch session at #awssummit
Andersen(Bakery from Hiroshima) sandwich box is #yum
Andersen runs JEAN-PAUL HÉVIN JAPON!



ランチと言えば。明日のお弁当どうしよう？



アンパンマンとパイキンマンのおにぎり♡

♡ 487 💬 28 📦 55

load more




スクランブルエッグおにぎりで♪ひよこ親子の愛情弁当♡

♡ 399 💬 11 📦 52

冷蔵庫にウィンナーが残ってる

Dish / Recipe ▾ ウィンナー 🔍 Sign up Login ?

“ウィンナー”で検索




ホットプレートでパーティーオムライス

113 likes 1 comment 6 shares

オムライスバリエーション、連続紹介中♪ こともの日も近いので、最後はこんなオムライス。みんなで囲むテーブルの上で出来上がっていく楽しさ！！

Ingredients / Seasoning

ウィンナー	6本
ミックスベジタブル	120g
カゴメ「トマトケチャップ」	大さじ8
ごはん	600g
卵	4個




こいのぼり弁当

79 likes 1 comment 7 shares

Ingredients / Seasoning

アスパラ	1本
きゅうりの浅漬け	5センチ
卵焼き	1個
赤ウィンナー	1本
海苔	少々

load more



クリスマスパーティにピッタリのデコおにぎらず

303 likes 13 comments 45 shares

お弁当にも使ってくださいー！(≧▽≦)/

Ingredients / Seasoning

ご飯	茶碗3杯強
スパム	缶詰め1/5
卵	1個

レシピや作り方も



こいのぼり弁当

キャラ弁 こどもの日グランプリ2015

576 ● 16 days ago

Recipe Close

Ingredients / Seasoning (1 serving)

アスパラ	1本
きゅうりの浅漬け	5センチくらい
卵焼き	1個
赤ウインナー	1本
海苔	少々
スライスチーズ	少々

Directions

- きゅうりの浅漬けと卵焼きの片端を三角に切り落とし、スライスチーズと海苔で目を作る。
- きゅうりの浅漬けの方は、エラ部分を包丁で切込みを入れ、鱗部分はストローで模様をつけました。

<http://snapdish.co/d/KW8vma>

SnapDish ! 便利で楽しい !



Amazon CloudSearchに関するセッション



<http://snapdish.co/search/?keyword=ウインナー>

SnapDish

Dish / Recipe ▾

ウインナー



ウインナー Result 1,729 Dishes

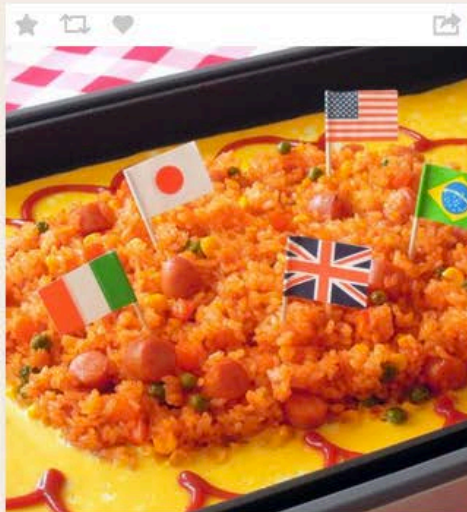
All Dish

Recipe

“ウインナー”で検索

Select ingredient

- 玉ねぎ (185)
- 塩コショウ (115)
- 卵 (114)
- ケチャップ (102)
- 水 (84)
- バター (80)
- 塩 (80)
- 牛乳 (79)

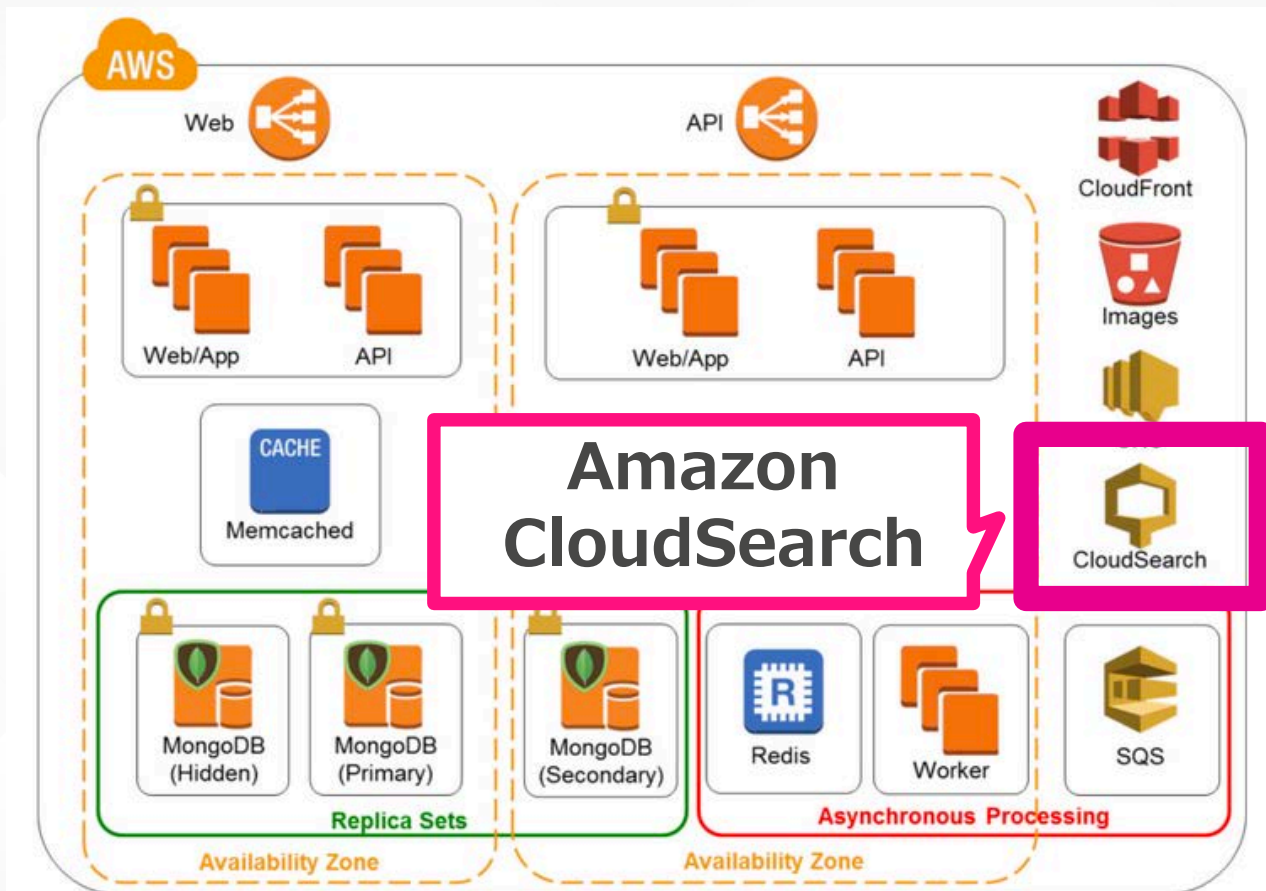


ホットプレートでパーティーオムライス



こいのぼり弁当

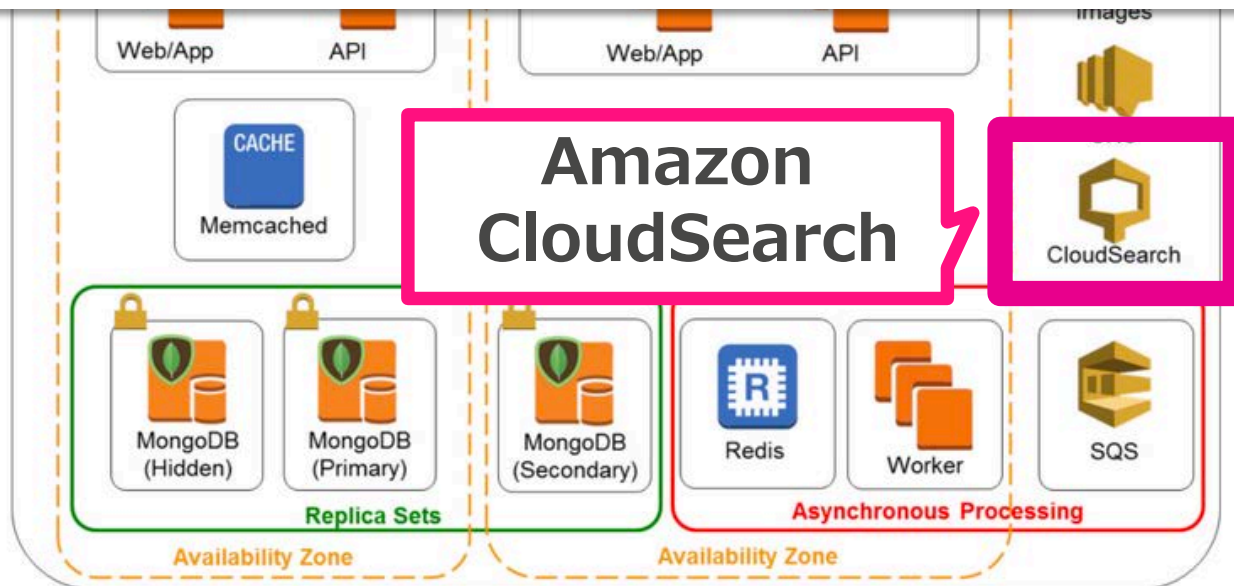
SnapDish システム構成図



<http://aws.amazon.com/jp/solutions/case-studies/snapdish/>

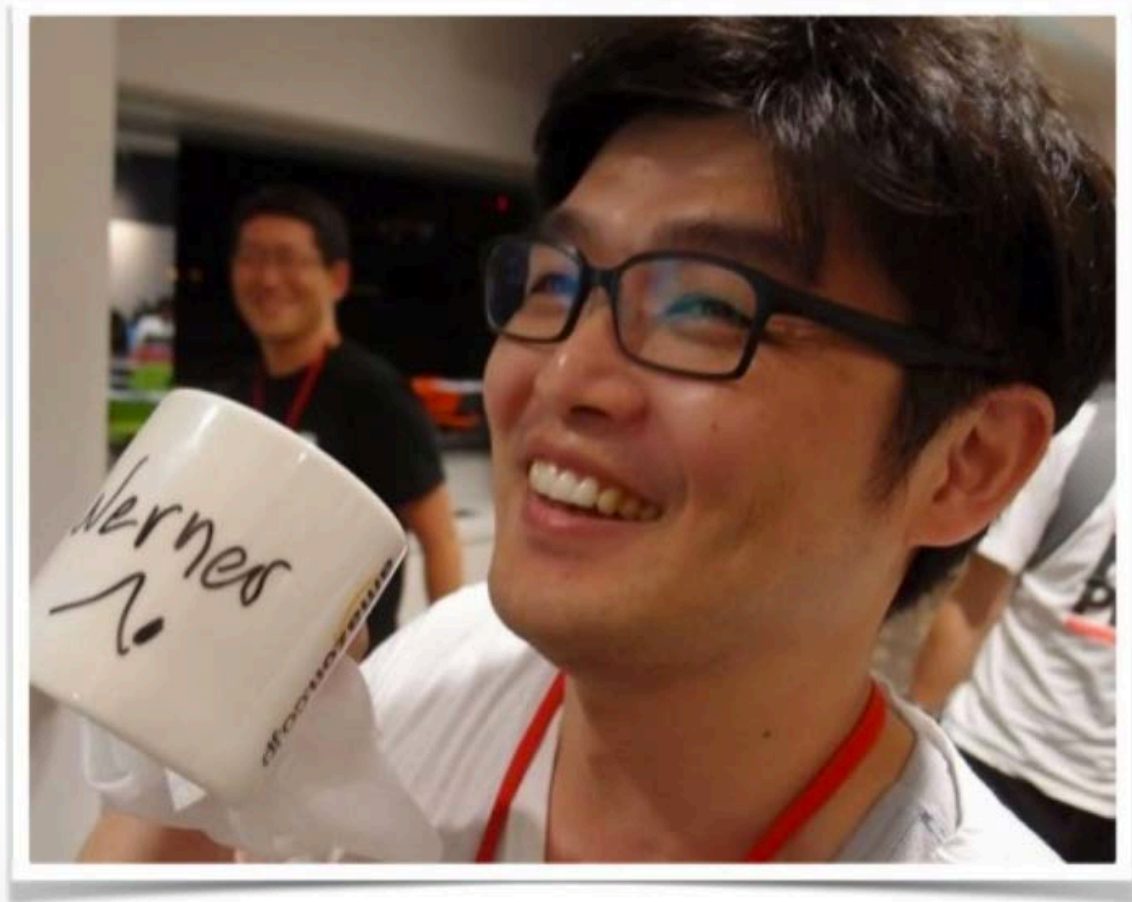
SnapDish システム構成図

また、検索機能では AWS のクラウド型検索サービスである Amazon CloudSearch を利用したことにより、サーバーの構築・運用から開放され、日本語解析処理を自分たちで実装することができました。 Amazon SNS や Amazon SQS は Amazon



<http://aws.amazon.com/jp/solutions/case-studies/snapdish/>

Vuzz株式会社 CTO 清田さん



Amazon CTOのWerner Vogelsを招いて行った Startup CTOsによるプレゼンバトルの優勝者！



Agenda

- **Amazon CloudSearchとは**
 - CloudSearchはフルマネージドなクラウド型検索サービスです
- **Inside Amazon CloudSearch**
 - CloudSearchの特徴を内部の挙動から理解しましょう
- **Amazon CloudSearch事例**
 - 日本における様々な事例をご紹介します

Amazon CloudSearch

- フルマネージドなクラウド型検索エンジン
- 検索ソリューションを容易に導入可能
- 2015年6月現在 34 の言語をサポート
- 豊富な検索機能
 - ハイライト表示
 - サジェスト(AutoComplete)
 - 地理空間検索



Amazon CloudSearch



Amazon CloudSearch – A9

amazon.co.jp
プライム

Amazonポイント: 0

マイストア | ギフト券 | タイムセール | 出品サービス | ヘルプ

カテゴリ
からさがす

すべて

AWS

検索

こんにちは、**篠原英治さん**
アカウントサービス

今すぐ登録
プライム

5
カート

ほしい物
リスト

検索結果 16,738件中 1-24件 "AWS"

カテゴリを選択して並べ替え

カテゴリ

Kindleストア

コンピュータ・IT

+ 続きを見る

本

コンピュータ・IT

ビジネス・経済

洋書

Computer Network
Programming

Computers & Technology

Web Services

+ 全26カテゴリ

絞り込み

配送オプション (詳細)



AWS クラウドを 1年間無料お試し

アマゾン ウェブ サービス (AWS) は、15製品を無料でお試しいただける1年間無料利用枠をご提供いたします。
今すぐ無料でお試しください。

関連サーチ: [amazon web service](#), [aws 資格](#), [azure](#).



Amazon Web Services クラウドデザインパターン
設計ガイド 改訂版 2015/5/28
玉川憲、片山曉雄



Getting Started with AWS (English Edition)
2012/4/23
Amazon Web Services



Amazon Web Services パターン別構築・運用ガイド
2015/3/25
佐々木 拓郎

前のページ 1 2 3 ... 20

検索結果を評価する

お探しの商品は見つかりましたか？

はい

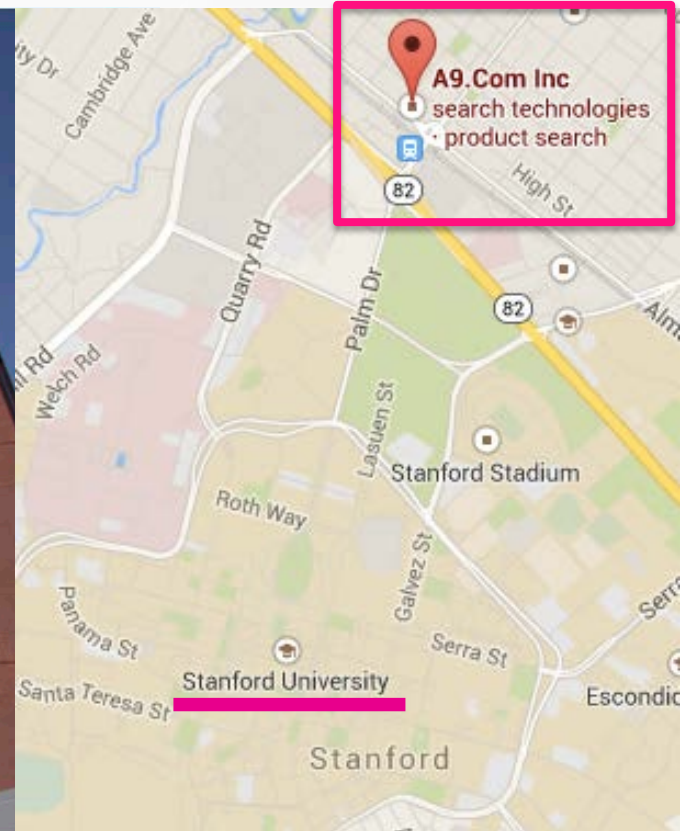
いいえ

ご質問、ご不明な点がある場合は、[ヘルプページ](#)をご確認ください。

サーチエンジン A9

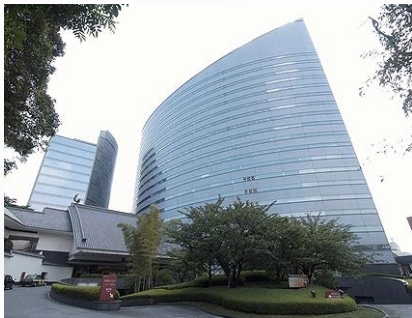
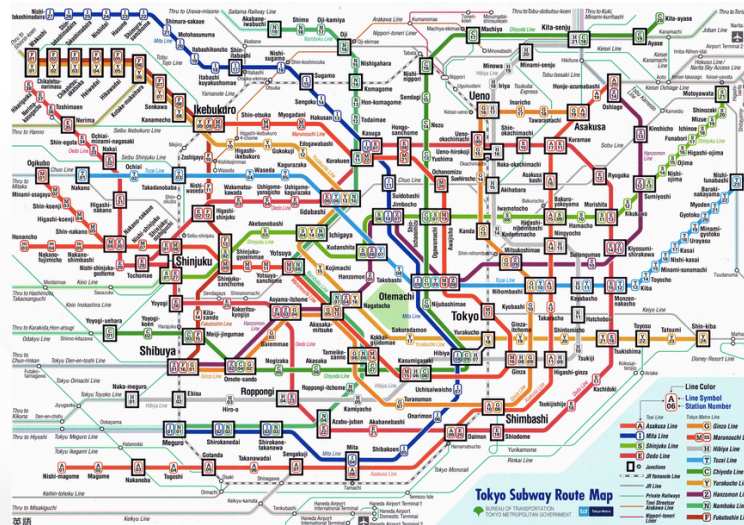


Amazon CloudSearch – A9



Amazon CloudSearch - Getting Started

- 目黒
- 白金台
- 白金高輪
- 麻布十番
- 六本木一丁目
- 溜池山王
- 永田町
- 四ツ谷
- 市ヶ谷
- 飯田橋
- 後樂園
- 東大前
- 本駒込
- 駒込
- 西ヶ原
- 王子
- 王子神谷
- 志茂
- 赤羽岩淵



Amazon CloudSearch - Getting Started

- 検索ドメインの作成

Create New Search Domain

NAME YOUR DOMAIN CONFIGURE INDEX REVIEW INDEX CONFIGURATION SET UP ACCESS POLICIES CONFIRM

Enter a name for your search domain. The name must start with a letter or number and be at least 3 and no more than

*Search Domain Name:

Prepare your domain for a large volume of data or traffic.

If you have a large amount of data to upload or anticipate a large volume of search requests, you can preconfigure your replication count based on the volume of traffic you expect. CloudSearch will still automatically scale your domain up or down.

Desired Instance Type:

Desired Replication Count:

Amazon CloudSearch - Getting Started

- 検索用データファイル

```
[ {
  "type" : "add",
  "id" : "nanboku_01",
  "fields" : {
    "code" : "N-01",
    "name" : "目黒",
    "lines" : ["東急目黒線", "都営三田線", "JR山手線", "東京メトロ南北線"]
  }
},
{
  "type" : "add",
  "id" : "nanboku_02",
  "fields" : {
    "code" : "N-02",
    "name" : "白金台",
    "lines" : ["都営三田線", "東京メトロ南北線"]
  }
}
```

駅コード

駅名

路線

Amazon CloudSearch - Getting Started

- スキーマ定義

Name	Type	Search	Facet	Return	Sort	Highlight	Analysis Scheme
code	literal	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
name	text	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Japanese
lines	text-array	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	Japanese

*_hoge のように指定する
Dynamic Field機能も使えます

日本語でのトークナイズ

Amazon CloudSearch - Getting Started

- マネージメントコンソール上で検索

Search:

Sort by: (view raw: [JSON](#) or [XML](#))

- 1. nanboku_14**

_score:	3.6249976
code:	N-14
lines:	J R山手線 東京メトロ南北線
name:	駒込
- 2. nanboku_01**

_score:	2.559809
code:	N-01
lines:	東急目黒線 都営三田線 J R山手線 東京メトロ南北線
name:	目黒

Amazon CloudSearch - Getting Started

- OR検索

Default Operator:

Search:

Sort by: (view raw: [JSON](#) or [XML](#))

- 1. nanboku_05**

_score:	3.502205
code:	N-05
lines:	東京メトロ南北線
name:	六本木一丁目
- 2. nanboku_04**

_score:	1.7511024
code:	N-04
lines:	東京メトロ南北線 都営大江戸線
name:	麻布十番

Amazon CloudSearch - Japanese Text Processing

- 形態素解析(Morphological Analysis)

- 彼(名詞-代名詞)/は(助詞-係助詞)/エンジニア(名詞-一般)/だ(助動詞)
- 日本語はスペースで区切られていないので解析する必要がある

名詞として扱われるべきであるが4つの単語に分割
⇒ 辞書のメンテナンスが必要

きゃりーぱみゅぱみゅ

きゃりーぱみゅぱみゅ (1993年1月29日^[1] -) は、日本のファッションモデル、歌手。正式な芸名は「きゃららいんちゃるんぶるっぶきゃりーぱみゅぱみゅ」^[2]。『KERA』や『Zipper』などのファッション雑誌を舞台に活動する「青文字系」モデルの代表的な人物^[3]。日本の「カワイイ」文化の世界的アイコンと目される人物でもある^[4]。所属事務所はアソビシステム^[5]。

ファッション雑誌『KERA』を舞台に2009年から読者モデルの活動を本格始動、2011年に「きゃりーぱみゅぱみゅ」名義で歌手活動を開始し、2012年にデビューシングル「ついまつける」をリリース、以後も音楽アーティストとして「ファッションモンスター」や「にんじ

きゃ
[e3 81 8d e3 82 83]

0

2

1

word

<

動詞-非自立

verb-auxiliary

り

[e3 82 8a]

2

3

1

word

⊙

助動詞

auxiliary-verb

ー

[e3 83 bc]

3

4

1

word

⊙

名詞-固有名詞-一般

noun-proper-misc

ぱみゅぱみゅ

[e3 81 b1 e3 81 bf e3 82 85 e3 81 b1 e3 81 bf e3 82 85]

4

10

1

word

⊙

名詞-一般

noun-common

<http://ja.wikipedia.org/wiki/きゃりーぱみゅぱみゅ>

Amazon CloudSearch - Japanese Text Processing

- 形態素解析(Morphological Analysis)
 - 辞書のカスタマイズできます

Stopwords **Stemming** **Synonyms** **Tokenization Dictionary**

A tokenization dictionary is a collection of entries where each entry specifies a set of characters, how they show up in a part of speech tag. You specify the dictionary as a string that specifies a JSON representation of the dictionary. An entry is an array of strings.

The example below the editor shows the correct JSON format.

Text:	<input type="text" value="きゃりーぱみゅぱみゅ"/>	?
Tokens:	<input type="text" value="きゃりー ぱみゅぱみゅ"/>	?
Readings:	<input type="text" value="キャリー パミュパミュ"/>	?
Part of Speech:	<input type="text" value="名詞,固有名詞,人名"/>	?

Amazon CloudSearch - Japanese Text Processing

- 形態素解析(Morphological Analysis)
 - 辞書のカスタマイズできます: AWS CLI/SDKからも操作可能

```
{
  "AnalysisSchemeName": "jascheme",
  "AnalysisSchemeLanguage": "ja",
  "AnalysisOptions": {
    "JapaneseTokenizationDictionary": "[
      ¥\"きゃりーぱみゅぱみゅ¥\",¥\"きゃりー ぱみゅぱみゅ¥\",¥\"キャリー パミュパミュ¥\",¥\"名
      詞,固有名詞,人名¥\"]"
  ]"
}
```

```
$ aws cloudsearch define-analysis-scheme --region us-east-1 --domain-name mydomain ¥
--analysis-scheme file://jascheme.txt
```

Amazon CloudSearch - Japanese Text Processing

- シノニム・類義語(Synonym Addition)

- 例えば Venice

- ベニス
- ベネチア
- ヴェネチア
- ヴェネツィア

- Alias

- Pupilで検索 => Studentもヒット
- Studentで検索 => Pupilはヒットしない

- Group

- 1st, first, one => どれで検索しても全てヒット

- 上記全て、マネージメントコンソールもしくははCLI/SDKで操作可能



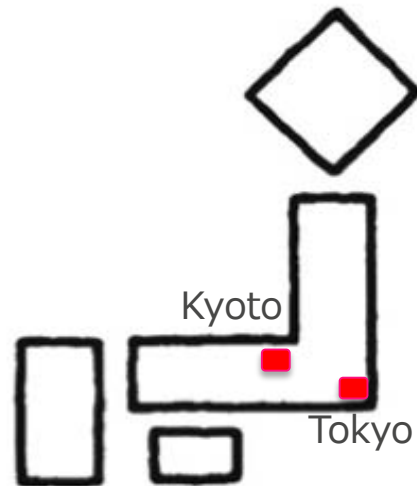
<http://ja.wikipedia.org/wiki/ヴェネツィア>

Amazon CloudSearch - Japanese Text Processing

- Bi-gramでのIndexing

- “Multiple Languages”を選択
 - CJK(中国語/日本語/韓国語)の場合Bi-gramに
- 2文字ずつで区切ってしまうので検索ノイズが発生
 - 例えば、『東京都』に対して『京都』でヒット

Type ⓘ	Search ⓘ	Facet ⓘ	Return ⓘ	Sort ⓘ	Highlight ⓘ	Analysis Scheme ⓘ
int	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
text	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Japanese
text	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Multiple Langu



Amazon CloudSearch - Japanese Text Processing

- 形態素解析とBi-gramを組み合わせて使う
 - 形態素解析での検索結果を優先したいが、
 - 検索の取りこぼしはしたくないのでBi-gramも使いたい
 - Batchアップロードの際に同じデータを2フィールド分用意したくない
 - **CloudSearchにはSourceFieldという機能**があります
 - 元フィールドからデータをコピーしてトークナイズする機能
 - 下記の例だと形態素解析のフィールドからBi-gramのフィールドでコピー

Type ⓘ	Search ⓘ	Facet ⓘ	Return ⓘ	Sort ⓘ	Highlight ⓘ	Analysis Scheme ⓘ	Default Value ⓘ	Source Field ⓘ
text ▾	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Multip ▾	<input type="text"/>	japanese [edit]

Amazon CloudSearch - Japanese Text Processing

- 形態素解析とBi-gramを組み合わせて使う
 - ^(caret)を用いてソート順を調整(形態素解析のスコアを上にする)

Search: (or japanese:'京都' multi:'京都')

Filter Query:

Query Parser: Structured Default Operator:

Search Fields: japanese^5

Sort by: Document Score Descending (view range)

1. todofuken.csv_26
_score 5.784237
japanese 京都府
multi 京都府
id 26
2. todofuken.csv_13
_score 2.690316
japanese 東京都
multi 東京都

“京都”で検索した場合、
“京都府”が“東京都”より上に

Amazon CloudSearch - Japanese Text Processing

- ソート順
 - スコア(_score)によるソート

Run a Test Search in Domain imdb

Search: Suggester: (

Sort by: (view raw: [JSON](#) or [XML](#))

1	#1228705
	_score: 14.135233
	rating: 7.0
	genres: Action; Adventure; Sci-Fi
	title: Iron Man 2
	release_date: 2010-04-26T00:00:00Z
	plot: Tony Stark has declared himself Iron Man and installed world peace... or so he... more...
	rank: 276
	running_time_secs: 7440
	directors: Jon Favreau
	image_url: http://ia.media-imdb.com/images/M/MV5BMTM0MDgwNjMyMI5BMTI5BanBnXkFtZTcwNTg3NzAzMw@@._V1
	year: 2010
	actors: Robert Downey Jr. Mickey Rourke Gwyneth Paltrow
2	#0120744
	_score: 13.176632

Amazon CloudSearch - Japanese Text Processing

- スコア(_score)は **tf-idf**

- あるドキュメントによく出てきて、他ドキュメントにあまり出てこない
- 上記のようなキーワードが高スコアになる

tf-idf

<http://ja.wikipedia.org/wiki/Tf-idf>

tf-idfは、文書中の単語に関する重みの一種であり、主に情報検索や文章要約などの分野で利用される。

tf-idfは、**tf** (英: Term Frequency、単語の出現頻度) と **idf** (英: Inverse Document Frequency、逆文書頻度) の二つの指標にもとづいて計算される。

$$\text{tfidf} = \text{tf} \cdot \text{idf}$$

$$\text{tf}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

$$\text{idf}_i = \log \frac{|D|}{|\{d : d \ni t_i\}|}$$

$n_{i,j}$ は単語 t_i の文書 d_j における出現回数、 $\sum_k n_{k,j}$ は文章 d_j におけるすべての単語の出現回数の和、 $|D|$ は総ドキュメント数、 $|\{d : d \ni t_i\}|$ は単語 t_i を含むドキュメント数である。そのため、idfは一種の一般語フィルタとして働き、多くのドキュメントに出現する語（一般的な語）は重要度が下がり、特定のドキュメントにしか出現しない単語の重要度を上げる役割を果たす。

Amazon CloudSearch - Japanese Text Processing

- Expressions

- 検索結果のチューニングに

Arithmetic	+ - * / %
Bitwise	& ^ ~ << >> >>>
Boolean	&& ! ? :
Comparison	< <= == >= >
Mathematical functions	abs ceil exp floor ln log2 log10 logn max min pow sqrt pow
Trigonometric functions	acos acosh asin asinh atan atan2 atanh cos cosh sin sinh tanh tan
haversin distance function	haversin(38.958687,-77.343149,latitude,longitude)

Amazon CloudSearch - Japanese Text Processing




- A/Bテスト
 - マネージメントコンソール上で比較

Compare Expressions in Domain imdb

Search: [More Parameters](#)

Expression 1 <small>▼ Expressions</small>	Expression 2 <small>▼ Expressions</small>
(rating ^1.5)	

Enter the terms you want to search for and click GO. << 1 to 9 of 9 Results >>

Lolita rating 6.7 genres Drama; Romance release_date 1997-09- <small>DocId: tt0119558 Score: 7.0</small>	 Iron Man rating 7.9 genres Action; Adventure; Sci-Fi release_date 2008-04- <small>DocId: tt0371746 Default Score: 9.562802</small>
The Man in the Iron Mask rating 6.3 genres Action; Adventure release_date 1998-03- <small>DocId: tt0120744 Score: 7.0</small>	 The Sessions rating 7.2 genres Comedy; Drama; Romance release_date 2012-01- <small>DocId: tt1866249 Default Score: 9.275442</small>
Outlander rating 6.2 genres Action; Adventure; Sci-Fi release_date 2008-07- <small>DocId: tt0462465 Score: 7.0</small>	 The Man in the Iron Mask rating 6.3 genres Action; Adventure release_date 1998-03- <small>DocId: tt0120744 Default Score: 8.543669</small>

Amazon CloudSearch – Field Types

Shop by Department Search Contract Cell Phones smartphone Go Hello. Sign in Your Account Try Prime Cart Wish List

Cell Phones & Accessories All Electronics Contract Cell Phones No-Contract Cell Phones Unlocked Cell Phones Accessories Best Sellers Deals

Departments

- Any Category
- Cell Phones & Accessories
 - Cell Phones
 - Contract Cell Phones

Wireless Carrier

- Verizon Wireless (44)

Brand

- Samsung

Cell Phone Display Size

- 3.9 Inches & Under (11)
- 4.0 to 4.4 Inches (27)
- 4.5 to 4.9 Inches (37)
- 5.0 to 5.4 Inches (24)
- 5.5 Inches & Over (16)

Cell Phone Internal Memory

- 4 GB (2)
- 8 GB (3)
- 16 GB (25)
- 32 GB (29)
- 64 GB

Avg. Customer Review

- ★★★★★ & Up (92)
- ★★★★☆ & Up (113)
- ★★★☆☆ & Up (115)
- ★★☆☆☆ & Up (117)

Literal

Double

Signed Integer

Date

HTC One, Silver 32GB (Verizon Wireless)

Text

Product Details

- Display: 4.7-inches
- Camera: HTC UltraPixel
- Processor Speed: 1.7 GHz
- OS: Android 4.2 (Jelly Bean)

Cell Phones > Contract Cell Phones > "smartphone"

Detail Image

Sort by Relevance New and Popular Price: Low to High Price: High to Low Avg. Customer Review Newest Arrivals

Samsung Galaxy S III, white 16GB (Verizon Wireless) \$0.01 with contract

Samsung Galaxy S III, Blue 16GB (Verizon Wireless) \$0.01 with contract

HTC One, Silver 32GB (Verizon Wireless) \$0.01 with contract

★★★★☆ (130) Product Details

- Display: 4.8-inches
- Camera: 8-MP
- Processor Speed: 1.5 GHz
- OS: Android 4.1 (Jelly Bean)

See newer version of this item

★★★★☆ (161) Product Details

- Display: 4.8-inches
- Camera: 8-MP
- Processor Speed: 1.5 GHz
- OS: Android 4.1 (Jelly Bean)

Amazon CloudSearch – Field Types

Type	説明
date	yyyy-mm-ddT00:00:00Zの形式。timestamp含む
date-array	複数のdate
double	double-precision 64-bit 浮動小数点数
double-array	複数のdouble
int	64-bit 符号付き整数
int-array	複数のint
latlon	latitude(緯度)longitude(経度)ペア。
literal	完全一致用文字列
literal-array	複数のliteral
text	言語指定可能な全文検索用フィールド
text-array	複数のtext

nanapi Tech Blog

nanapi TechBlog

エンジニアリング

CakePHPからAmazon CloudSearchを使う

Tweet 56 いいね! 47 G+1 B! 58 CLIP

Amazon CloudSearch



生活の知恵があつまる情報サイト

nanapi ナナビ

掲載中のライフレシビ数: 100,000件以上

わからないこと・しりたい

「京都 お土産」に関する検索結果

カテゴリ: int型

カテゴリ名: text型

「京都 お土産」に関する検索結果

カテゴリ

カテゴリ指定がない

- 国内旅行 (127)
- グルメ (6)
- おでかけ (5)
- 海外旅行 (3)
- 行事・祭事・イベント (2)
- デート (2)
- 歴史 (2)
- ショッピング (2)
- コミュニケーション (2)
- 家事 (1)
- 占い・スピリチュアル (1)
- 家族・親族 (1)
- 生活の雑学・トリビア (1)
- カップル (1)
- エンタメ (1)
- 移動・交通 (1)

158件の記事が見つかりました

158件中 1 - 20 件を表示

京都北山(マールブランシュ)

人気商品 茶の葉など京都北山のお菓子のオンラインショップ。喜ばれギフトなら!

www.malebranche-shop.com

アウトライン: literal-array型

京都で生まれた新しいお土産

フランスパン職人が作ったラスクから新しい焼き菓子が出来ました。

www.petit-pochi.com/

タイトル: text型

関連度順

人気

京都でのおすすめのお土産

京都のお土産 / 実際に買ったことのあるお土産 / 京都のお土産についてのアドバイス / 代表的なお土産 / 人気のお土産 / 地域限定のお土産 / 出町ふたばの豆餅 / 京野菜の漬物 /

京都に通算で200回は行っている著者が、京都の美味しいお土産について、代表的なお土産 ...

本文: text型

すべて読む

by パースたか

View数: int型

著者名: literal型

<http://blog.nanapi.co.jp/tech/2014/08/27/cakephp-cloud-search/>

Amazon CloudSearch - Suggestions

すべて ▾	クラウドデザイン
クラウド型 絵プレーヤー	クラウドデザインパターン
	クラウドデザインパターン 和書
	クラウドデザインパターン Kindleストア
	クラウドデザインパターン 実装ガイド

`/suggest?q=ir&suggester=title_sug`

```
"suggest": {"query": "iro", "found": 5,
"suggestions": [
{"suggestion": "Iron Man",... "id": "tt0371746"},
{"suggestion": "Iron Man 2",... "id": "tt1228705"},
...
]
```

Run a Test Search in Domain hoge

Search: おお

Enter the

大崎
大塚

ひらがな や カタカナ で検索しても、
漢字の候補をサジェスト

Amazon CloudSearch – Multi AZ

- 高可用性強化
 - Multi Availability-Zone

Availability Options for Domain test

Enabling the Multi-AZ option provisions and maintains extra instances in the event of a service disruption in one Availability Zone, all traffic is served by the remaining instances.

Important: Enabling the Multi-AZ option doubles the number of instances running your domain.

Your availability options are **ACTIVE**.

Multi-AZ: Off

Amazon CloudSearch – IAMによる管理

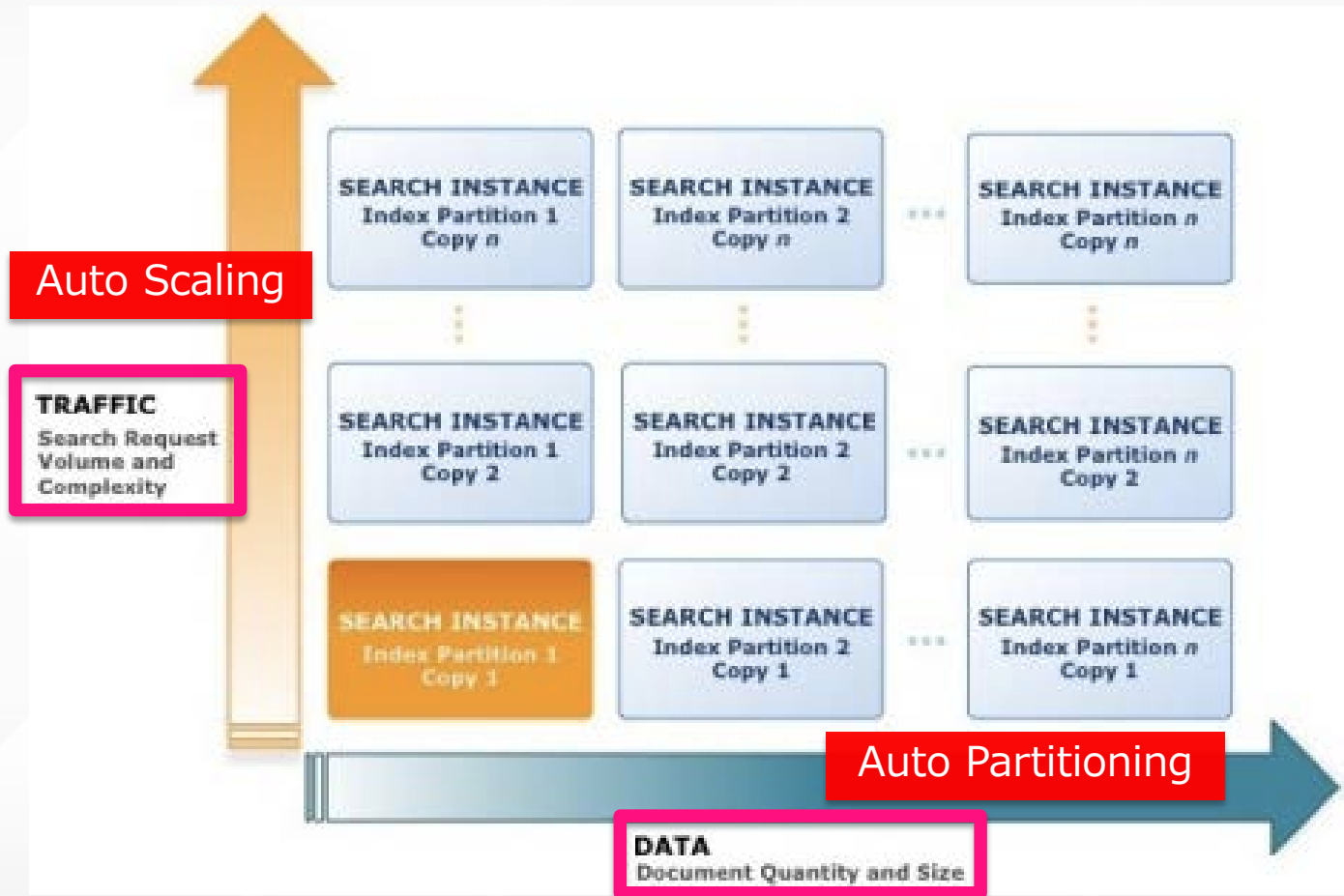
- ユーザーやリソースに対してきめ細やかな制御が可能
 - AWS Security Blog - <http://amzn.to/1uYcDKi>
 - CloudSearch: Now with More Granular Access Control for Domains
 - 設定例: 管理者,オペレーター,開発者,ビジネスオーナー,アプリケーション

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": ["cloudsearch:search",
                "cloudsearch:suggest"],
      "Resource": "arn:aws:cloudsearch:*:111122223333:domain/prod-imdb"
    }
  ]
}
```

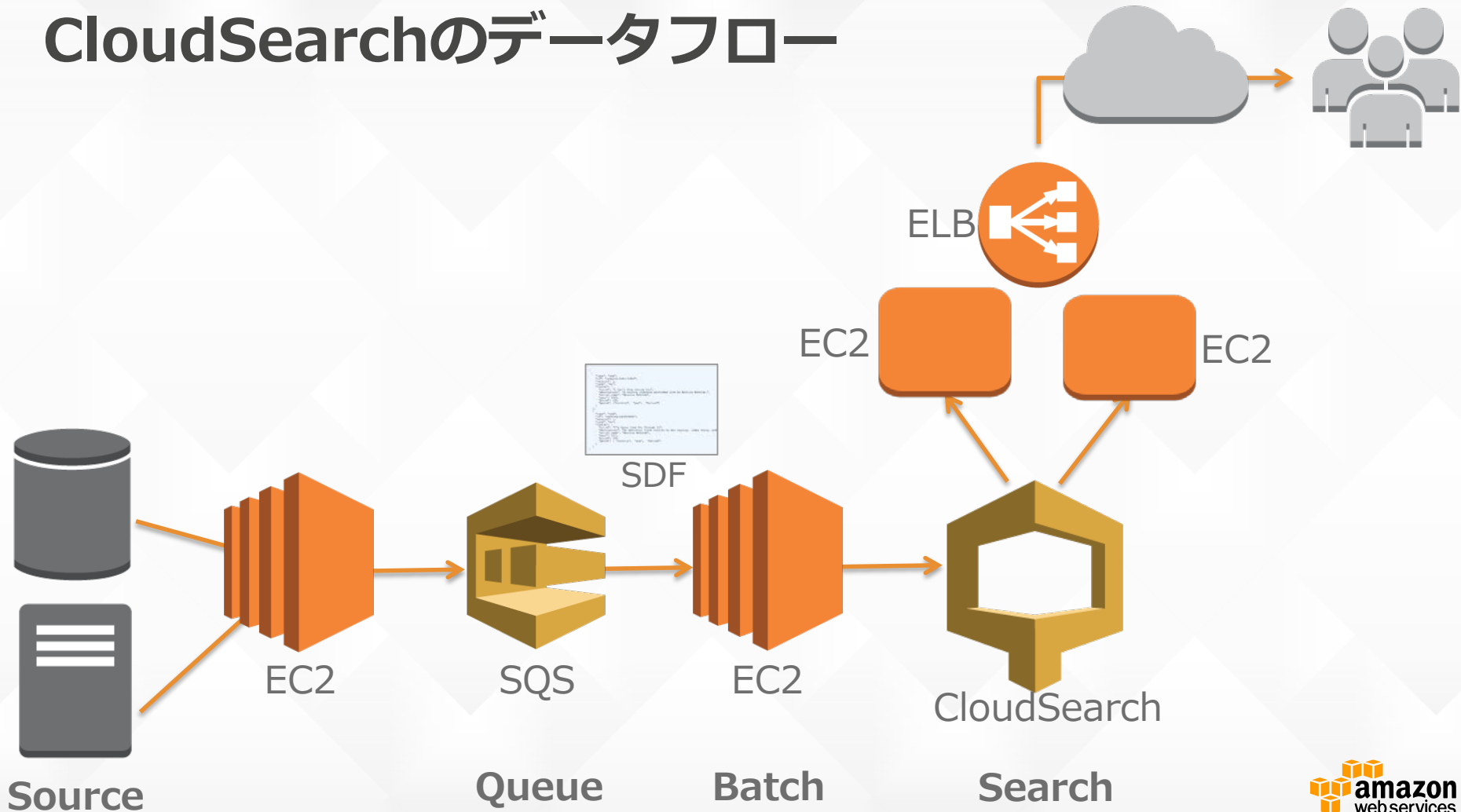
Agenda

- Amazon CloudSearchとは
 - CloudSearchはフルマネージドなクラウド型検索サービスです
- **Inside Amazon CloudSearch**
 - CloudSearchの特徴を内部の挙動から理解しましょう
- Amazon CloudSearch事例
 - 日本における様々な事例をご紹介します

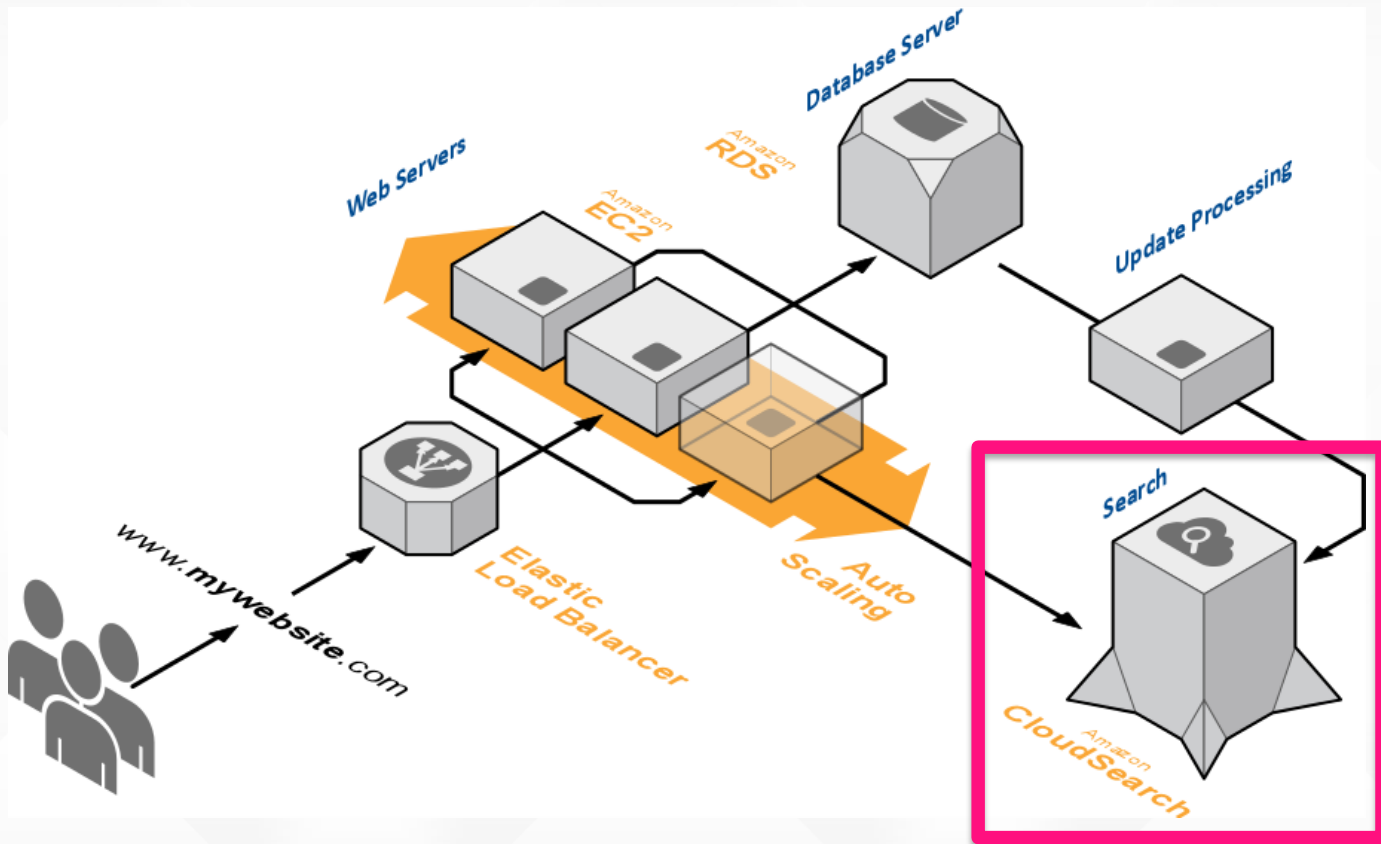
Amazon CloudSearchはフルマネージド



CloudSearchのデータフロー



CloudSearchのReference Architecture

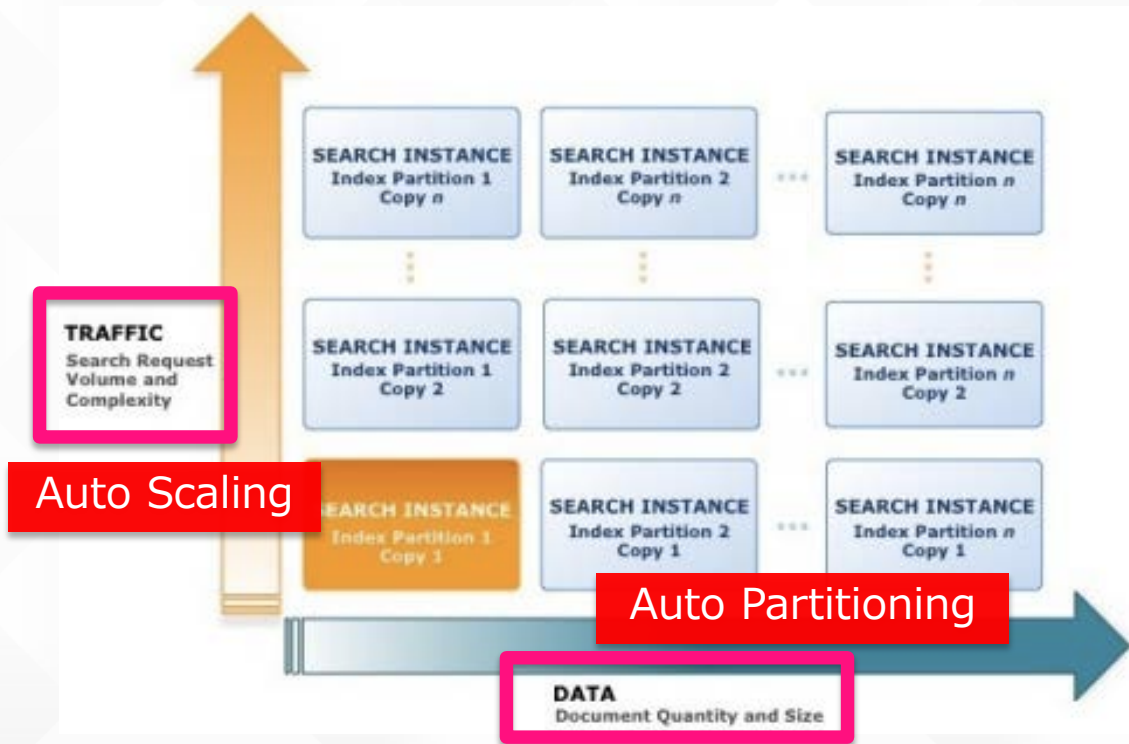


Inside Amazon CloudSearch



Inside Amazon CloudSearch

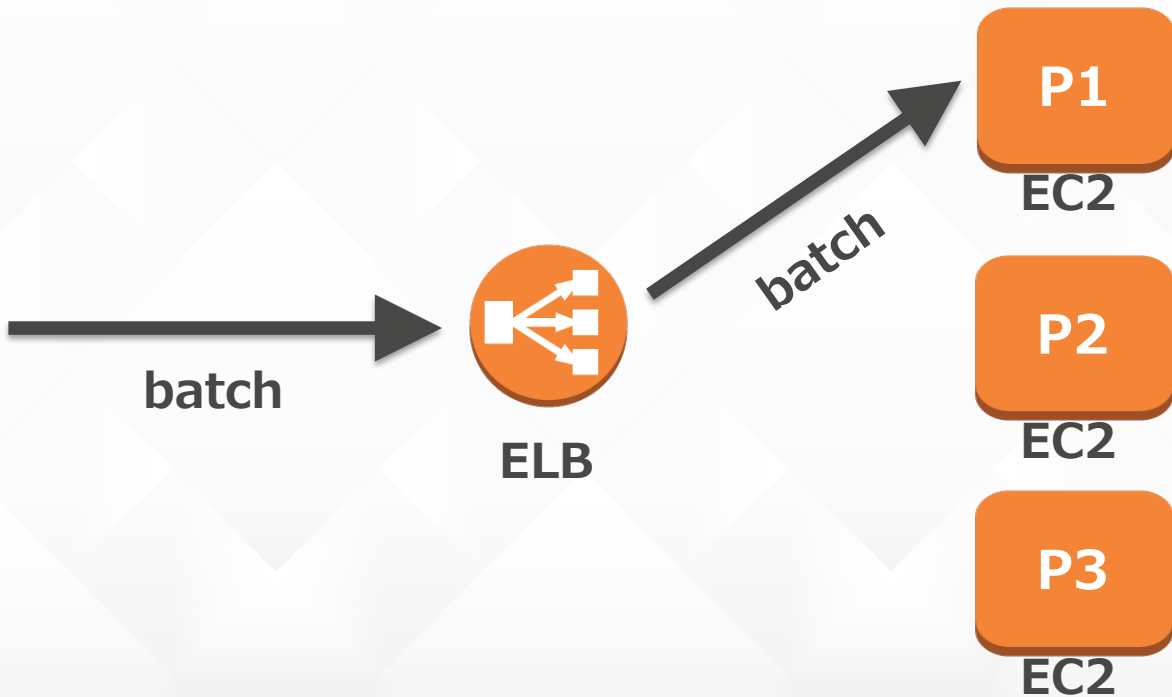
- Indexing
- Query
- Auto Scaling
- Auto Partitioning



Inside Amazon CloudSearch

- Indexing

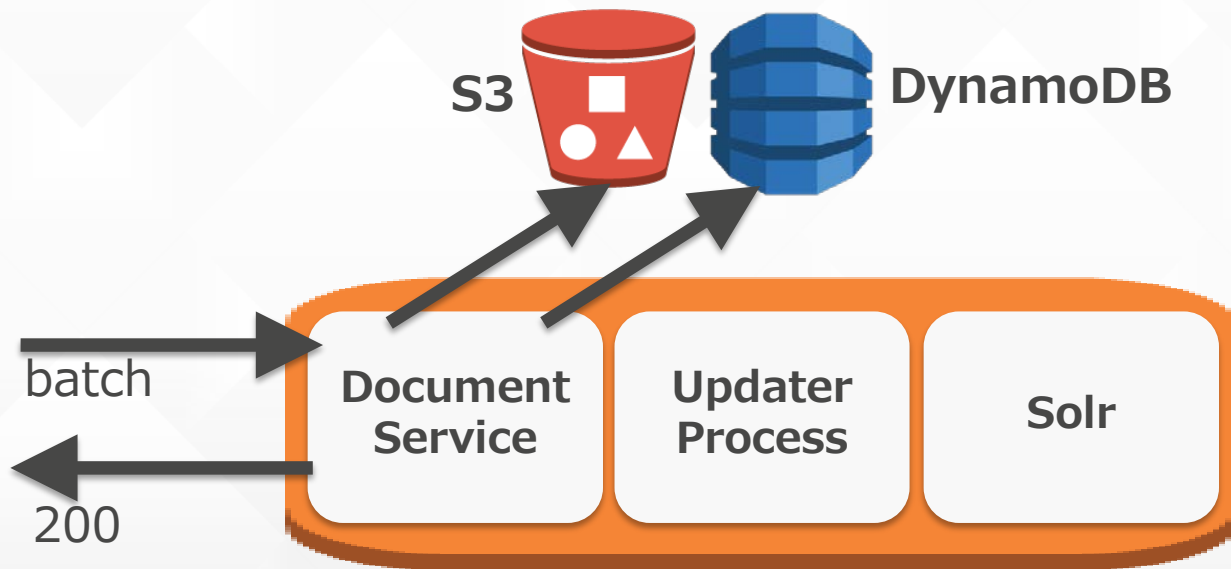
- データを保持している全てのノードにELB経由で均等にアップロード



Inside Amazon CloudSearch

- Indexing

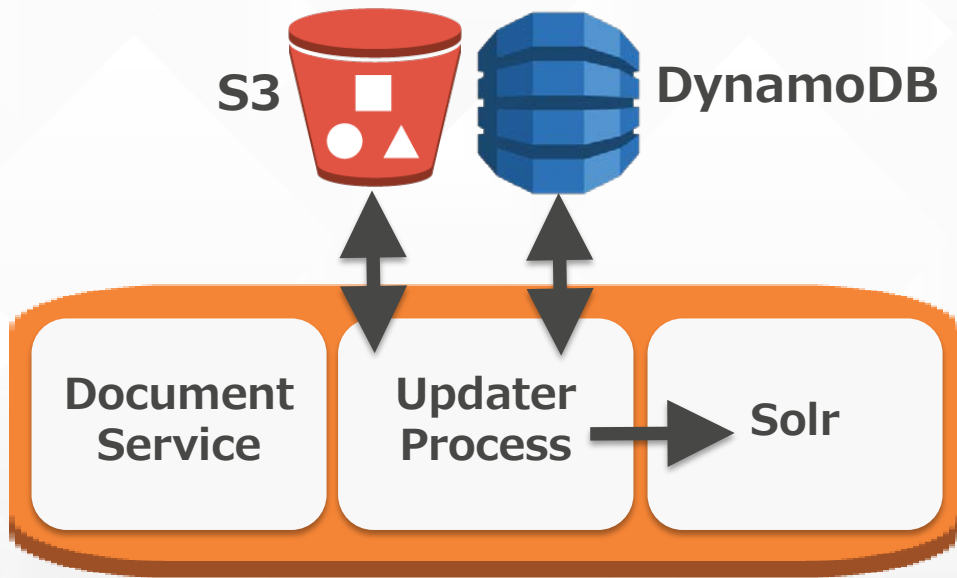
- データを受け取ったノードは、
- ファイルをS3に保存し、メタ情報をDynamoDBに保存した後、
- クライアントにHTTPステータスコード200(正常終了)を返す



Inside Amazon CloudSearch

- Indexing

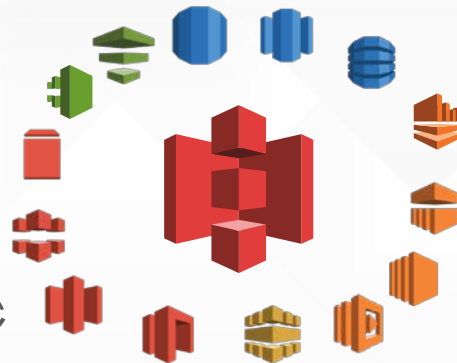
- S3とDynamoDBにポーリングを行い、
- Indexing対象があれば処理を行いS3に配置 & DynamoDBを更新
- 自Partition担当のIndexバイナリがあればローカルのSolrへ配置



Inside Amazon CloudSearch

- Amazon Simple Storage Service

- 高い堅牢性 99.999999999%
- 格納容量無制限。利用した分のみ課金
- 様々なAWSサービスと連携するセンターストレージ



- Amazon DynamoDB

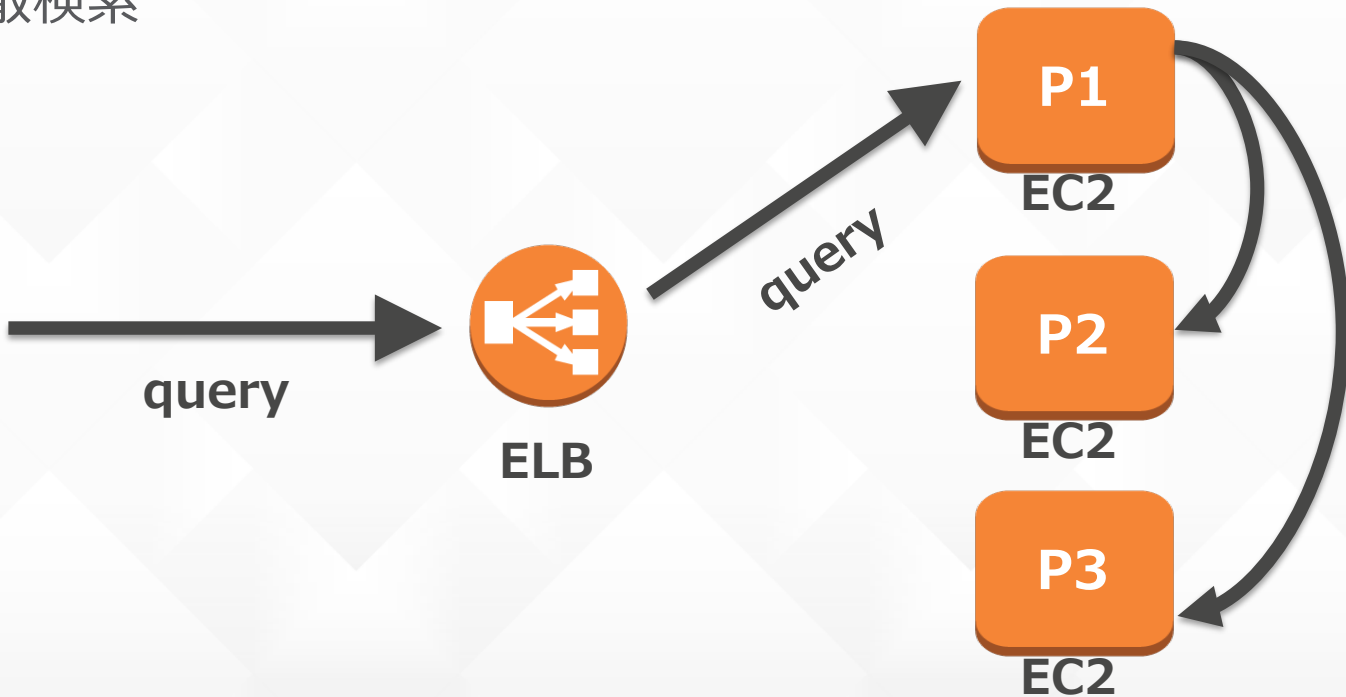
- 高い信頼性、スケーラビリティ、低レイテンシ、安定した性能を兼ね備えたNoSQLデータベースサービス
- 必要スループットを決めるだけで利用可能。ストレージ容量は事前に決める必要がなく、自動的にプロビジョンされる



Inside Amazon CloudSearch

- Query

- インデックス処理と同様にELBで均等に割り振り
- 分散検索



Inside Amazon CloudSearch

- Auto Scaling

- 大量の検索リクエストをハンドリングする必要がある場合は、
- AutoScaling(ELB + EC2)で対応

Auto Scaling Group

P1

Auto Scaling Group

P2

Auto Scaling Group

P3

Inside Amazon CloudSearch

- Auto Scaling

- 大量の検索リクエストをハンドリングする必要がある場合は、
- AutoScaling(ELB + EC2)で対応

Auto Scaling Group

P1

P1

P1

Auto Scaling Group

P2

P2

P2

Auto Scaling Group

P3

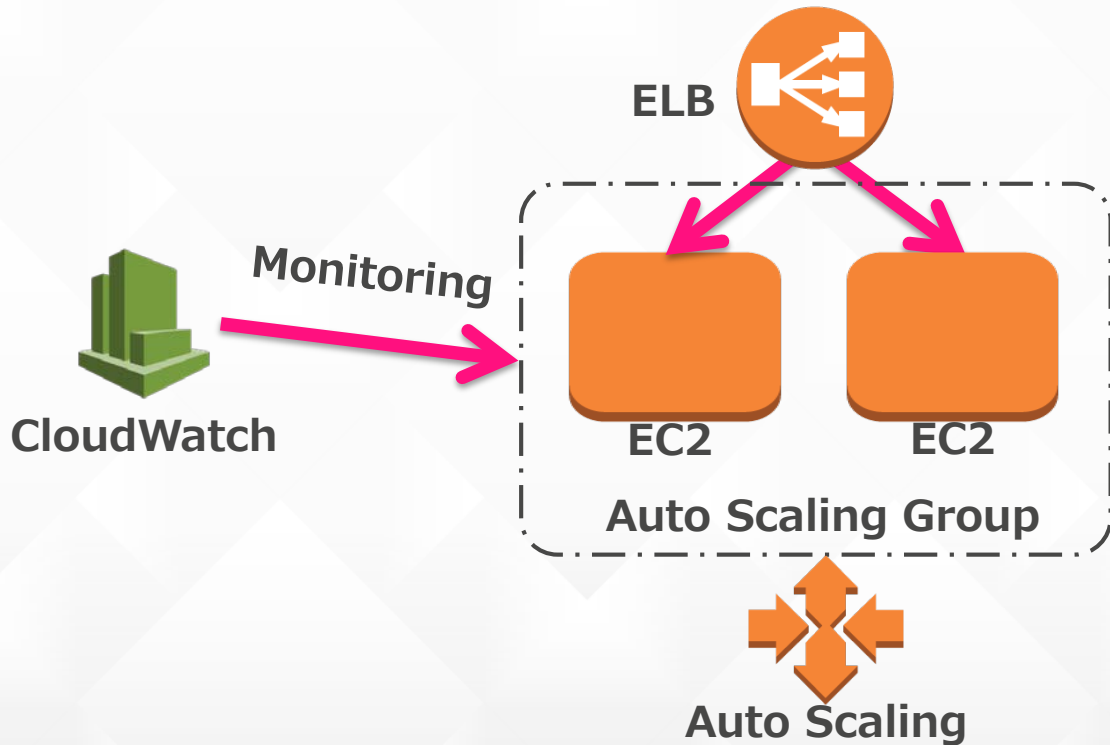
P3

P3

Inside Amazon CloudSearch

- Auto Scalingの仕組み

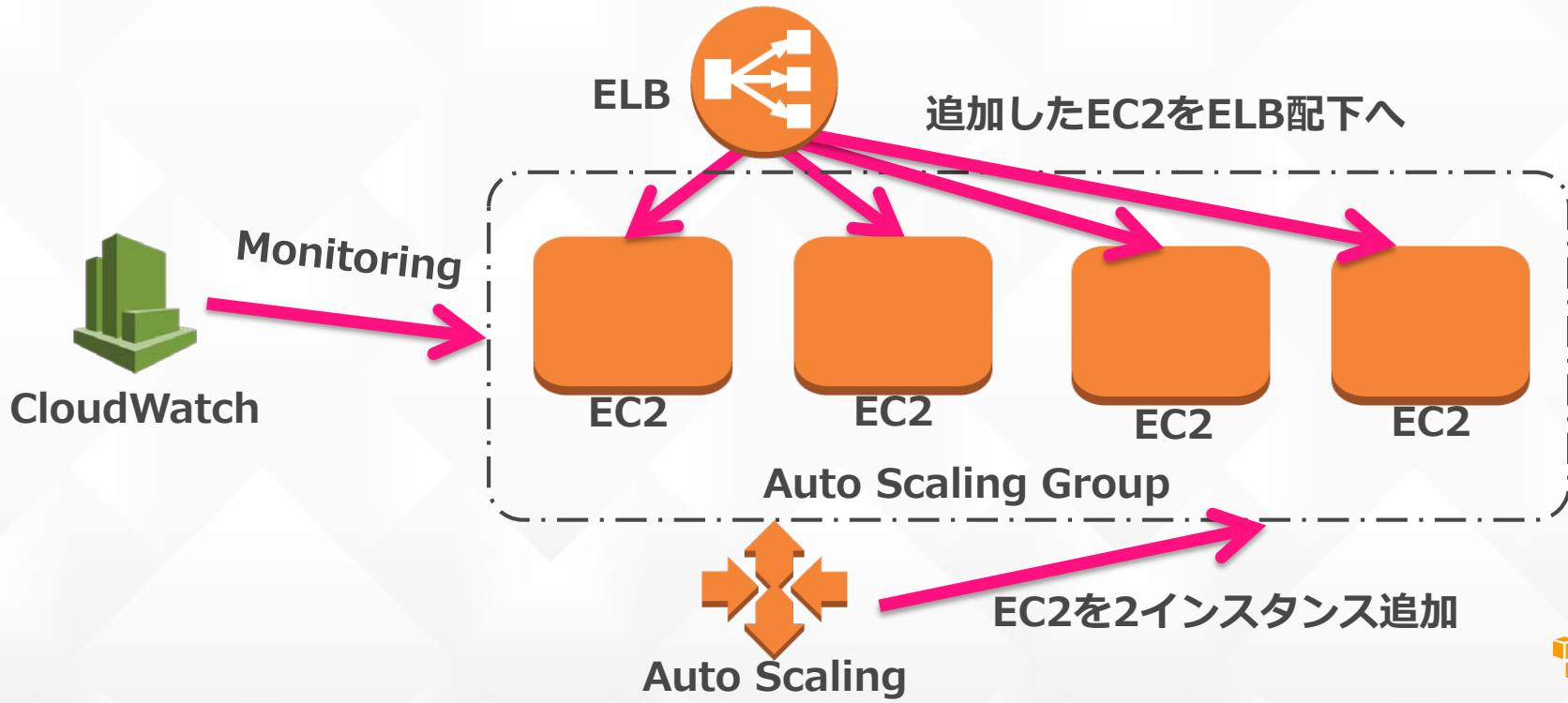
- 例)CPU利用率が5分以上70%を超える場合、EC2を2インスタンス追加



Inside Amazon CloudSearch

- Auto Scalingの仕組み

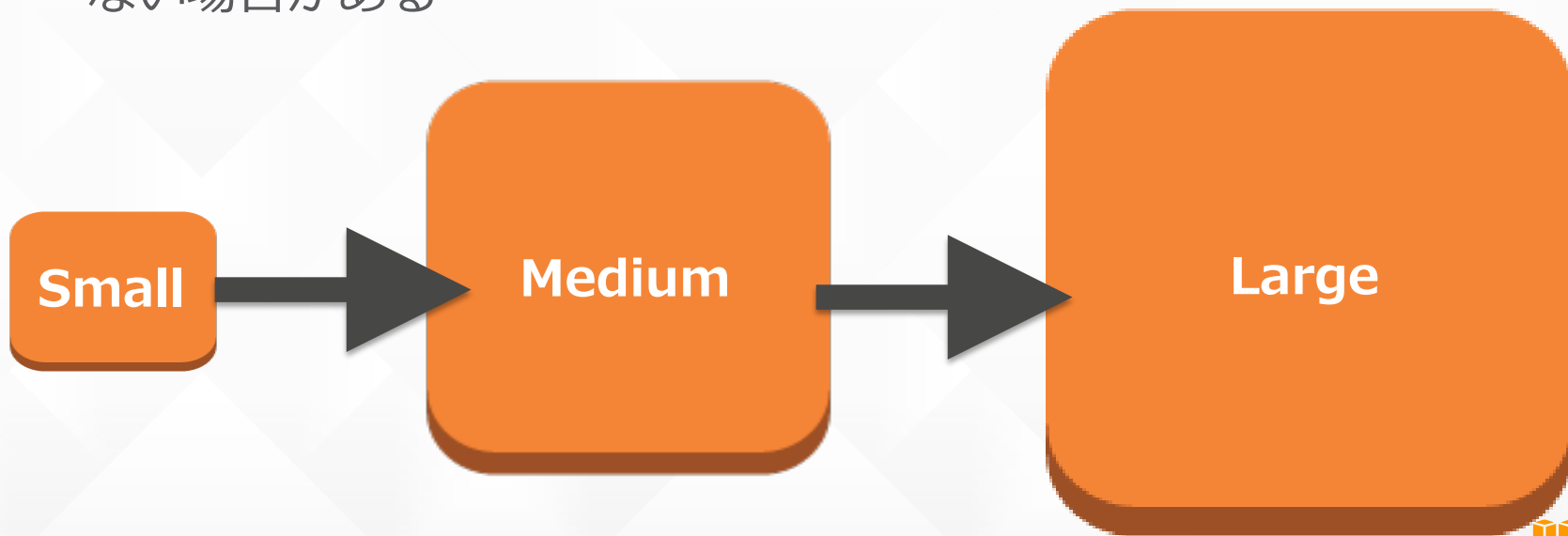
- 例)CPU利用率が5分以上70%を超える場合、EC2を2インスタンス追加



Inside Amazon CloudSearch

- Auto Partitioning

- 最初はスケールアップ
- ダウンタイムはないが切り替わり中はIndexing結果がすぐに反映されない場合がある



Inside Amazon CloudSearch

- Auto Partitioning

- 1つのノードでまかないきれなくなったらスケールアウト
- パーティションの分割処理はHadoop(EMR)で
- ダウンタイムはないが結果整合性モデル
 - S3にインデックスファイルが置かれ各ノードがポーリング



Inside Amazon CloudSearch

- 検索ドメインの設定変更を行った場合
 - 変更に伴う処理はHadoop(EMR)で
 - ダウンタイムはないが結果整合性モデル
 - S3にインデックスファイルが置かれ各ノードがポーリング



Inside Amazon CloudSearch

- 大量の初期データを投入する場合
 - 大きいインスタンスを多数並べると短時間でIndexingを実行可能

<http://www.slideshare.net/AmazonWebServices/sdd411-amazon-cloudsearch-deep-dive-and-best-practices-aws-reinvent-2014/26>

Desired Instance Type:

Desired Replication Count:

Desired Partition Count: Use default

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

... changes have not been saved

Multi-threading data load

Instance type	Instance threads	Connecting threads
search.m1.small	2	1
search.m1.large	5	3
search.m2.xlarge	9	5
search.m2.2xlarge*	17	9

Inside Amazon CloudSearch

- インスタンスタイプとデータ量の目安
 - 扱うデータによって圧縮率が異なるため保持量に差が出る

<http://www.slideshare.net/AmazonWebServices/sdd411-amazon-cloudsearch-deep-dive-and-best-practices-aws-reinvent-2014/23>



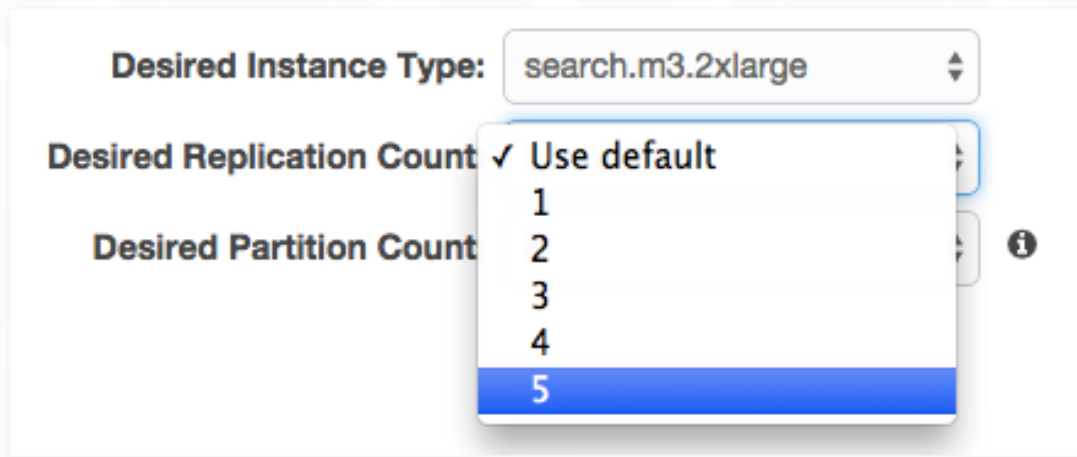
How big?

Instance type	Twitter data (Search only)	Common-crawl data (Search only)
search.m1.small	6.7 GB, 4.7 million documents	4 GB, 625 K documents
search.m1.large	26.8 GB, 18.8 million documents	16 GB, 2.5 million documents
search.m2.xlarge	53.6 GB, 37.6 million documents	34 GB, 5 million documents
search.m2.2xlarge*	107.2 GB, 75.2 million documents	64 GB, 10 million documents

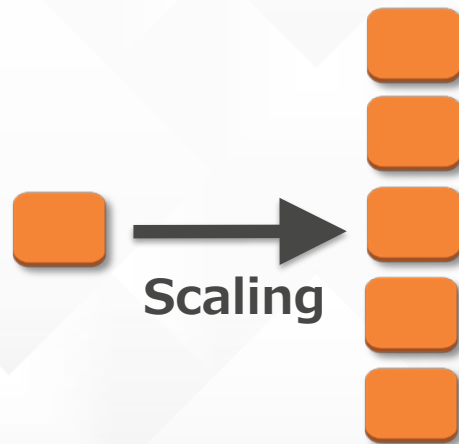
Inside Amazon CloudSearch

- 大量の検索リクエストに事前に備える
 - 事前にレプリケーションのカウンタを上げておくことで実行可能
 - マネージメントコンソールだけでなくAWS CLI/SDKからも

```
$ aws cloudsearch update-scaling-parameters --domain-name movies ¥  
--scaling-parameters DesiredInstanceType=search.m3.2xlarge,DesiredReplicationCount=5
```



The screenshot shows the Amazon CloudSearch console interface for configuring scaling parameters. The 'Desired Instance Type' is set to 'search.m3.2xlarge'. The 'Desired Replication Count' dropdown menu is open, showing options: 'Use default' (checked), '1', '2', '3', '4', and '5' (highlighted in blue). The 'Desired Partition Count' is also visible but not fully shown.



Inside Amazon CloudSearch

- インスタンスタイプとスループットの目安
 - 扱うデータによってスループットも異なる

<http://www.slideshare.net/AmazonWebServices/sdd411-amazon-cloudsearch-deep-dive-and-best-practices-aws-reinvent-2014/31>

1語、2語、3語からなる1万
個のシンプルなクエリを用
いて計測



Query scaling

Instance type	Instance threads	JMeter	Twitter throughput	Com crawl throughput
search.m1.small	20	2 hosts 10 threads	25.1 qps 397 ms	48.3 qps 206 ms
search.m1.large	20	4 hosts 20 threads	108.5 qps 183 ms	291.5 qps 68 ms
search.m2.xlarge	20	8 hosts 40 threads	419.6 qps 94 ms	665.9 qps 59 ms
search.m2.2xlarge	20	16 hosts 80 threads	566.4 qps 140 ms	985.3 qps 80 ms

Inside Amazon CloudSearch

- M3インスタンスが利用可能に - 2015年2月～
 - 最新のインスタンスを活用することで性能が向上します

最新情報: M3 インスタンスのより高いパフォーマンスで Amazon CloudSearch を実行

2015年2月10日

Amazon CloudSearch ドメインを M3 インスタンス上で実行できるようになりました。これによって旧世代のインスタンスと比べて、ほとんどのユースケースにおいて、より高く、より安定したコンピューティングパフォーマンスを提供いたします。今日から M3 インスタンス上で新しい Amazon CloudSearch ドメインを作成できます。AWS コンソール、SDK、または CLI を使って、「希望するインスタンスタイプ」設定を変更することによって、現存の Amazon CloudSearch ドメインを M3 インスタンスタイプへ変換できます。

CloudSearch ドメインでの M3 インスタンスの使用についての詳細は、Amazon CloudSearch 開発者ガイドの「[Amazon CloudSearch でスケーリングオプションを設定](#)」を参照してください。M3 インスタンスタイプの料金の詳細は、[Amazon CloudSearch 料金表ページ](#)をご覧ください。

<http://aws.amazon.com/jp/cloudsearch/whats-new/>

Agenda

- Amazon CloudSearchとは
 - CloudSearchはフルマネージドなクラウド型検索サービスです
- Inside Amazon CloudSearch
 - CloudSearchの特徴を内部の挙動から理解しましょう
- **Amazon CloudSearch事例**
 - 日本における様々な事例をご紹介します

国内のAmazon CloudSearch事例

- 爆速で導入いただいたケース: **school**



The image shows the homepage of the 'school' website. At the top, there is a navigation bar with links for 'ホーム' (Home), '授業をさがす' (Find classes), and '学部' (Department). A search bar is located on the right side of the navigation bar, containing the text '何を学びたい?' (What do you want to learn?) and a '検索' (Search) button. The search bar is highlighted with a pink border. Below the navigation bar, there is a main content area with a blue background. On the left, there is a blue button that says 'スクーってなに?' (What is school?). Below this, the main heading reads 'WEBに誕生した新しい学校のカタチ' (The shape of a new school born on the web). Underneath the heading, there is a paragraph of text: 'スクーには入学試験も、年齢制限も、卒業もありません。ここでは知的好奇心を満たすコンテンツが詰まった、オンライン上の学校。あなたの毎日に「終わらない楽しい学生生活」を提供するサービスです。' (School has no entrance exams, age restrictions, or graduation. Here, you can satisfy your intellectual curiosity with content-packed online schools. We provide a service that offers you an endless, fun student life every day.) To the right of the text, there is a logo for 'school WEB-campus' inside a white speech bubble. The logo consists of a colorful square icon followed by the word 'school' and 'WEB-campus' below it. At the bottom of the page, there is a row of four navigation links: '無料で生放送の授業を受講' (Attend live broadcasts for free), 'オンラインならではの学び' (Learning unique to online), '楽しみ方はひとそれぞれ' (Everyone has their own way of enjoying it), and '会員の種類について' (About membership types).

国内のAmazon CloudSearch事例

- 爆速で導入いただいたケース: schoo

まとめ

スクーでは**教科書通り**の標準的な使い方。

Index設計の際には、使う「**場面**」と、登録する

「**内容**」をよく考える必要があるけど、データ量と処理量に合わせてスケーラブルな拡張も**容易**。

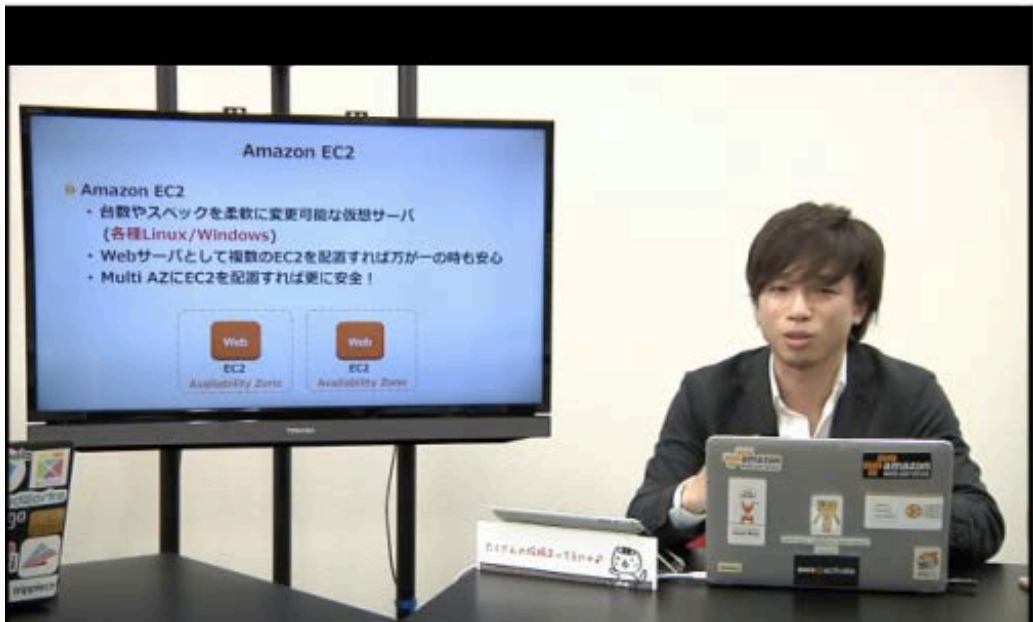
急な仕様変更での導入となったけど、リリースするまでに
1週間程度で導入を実現。

私もschoolで先生やっています！

スタートアップならおさえておきたいAWS(Amazon Web Services)入門 2 限目：…



篠原 英治



Amazon EC2

- Amazon EC2
 - 台数やスペックを柔軟に変更可能な仮想サーバ (各種Linux/Windows)
 - Webサーバとして複数のEC2を配置すれば万が一の時も安心
 - Multi AZにEC2を配置すれば更に安全!



是非schoolでAWSで検索してみてください！

The screenshot shows the school WEB-campus search interface. At the top left is the 'school WEB-campus' logo with a '本校舎β版' badge. A search bar contains 'AW|' and 'AWS', with a green '検索' button. Navigation links include '授業をさがす', '学部ホーム', and '先生/学生'. A notification icon shows '26' and a user profile picture. Below the search bar, the results are categorized into 'すべて (8)', '録画 (7)', and '先生/団体 (1)'. The main section is titled '「AWS」に関する録画 (7)'. Two video results are shown, both featuring the Amazon Web Services logo. The first result is 'スタートアップならおさえておきたいAWS(Amazon Web Services)入門 1 限目：サービス概要と基礎知識編' by 高山 博史 and 篠原 英治, from the 'プログラミング学部'. It has a green '→ 教室に入る' button with a '無料' tag and a link to '授業詳細へ'. The second result is 'スタートアップならおさえておきたいAWS(Amazon Web Services)入門 ～メディア露出時のピーク対策編～' by 高山 博史 and 今井 雄太, from the 'スタートアップ学部'. It also has a green '→ 教室に入る' button and a link to '授業詳細へ'.

school WEB-campus 本校舎β版

検索

授業をさがす 学部ホーム 先生/学生

26

検索結果

すべて (8) 録画 (7) 先生/団体 (1)

「AWS」に関する録画 (7)

amazon web services

スタートアップならおさえておきたいAWS(Amazon Web Services)入門 1 限目：サービス概要と基礎知識編

→ 教室に入る 無料

公式 高山 博史、篠原 英治

● プログラミング学部

授業詳細へ

amazon web services™

スタートアップならおさえておきたいAWS(Amazon Web Services)入門 ～メディア露出時のピーク対策編～

→ 教室に入る

公式 高山 博史、今井 雄太

● スタートアップ学部

授業詳細へ

国内のAmazon CloudSearch事例

- フルマネージドの良さを活かしたケース: nanapi

生活の知恵があつまる情報サイト



掲載中のライフレシビ数: **100,000**件以上

🔍 わからないこと・知りたいことを入力

検索

メンバー登録(無料)

🔒 ログイン

2015
5/20
(水)

これでスッキリ! 簡単で便利な収納アイデア



100均のファイルBOXを活用
棚の奥まで見渡せる収納のコツ

つっぱり棒を活用
玄関の収納アイデア

引き出しサイズに合わせて選ぶ
押入れの収納アイデア

材料費500円のできる
トイレ収納棚の作り方

天井までのつっぱり棒を活用
見せるぬいぐるみの収納術

ライフレシビのカテゴリー一覧

ハウツー情報がいっぱい! 人気順でも探すことができます

バックナンバーを見る

❤️ 恋愛

国内のAmazon CloudSearch事例

- フルマネージドの良さを活かしたケース: nanapi

CloudSearch nanapiの事例紹介



株式会社nanapi 加賀谷 祐平

CloudSearchのメリット

- デフォルトのままでいい感じに検索できる
- 簡単に日本語全文検索システムを作れる
- とにかく楽

<http://blog.nanapi.co.jp/tech/2014/08/27/cakephp-cloud-search/>



国内のAmazon CloudSearch事例

- フルマネージドの良さを活かしたケース: nanapi

 nanapi / cakephp-cloud-search

 Unwatch ▾ 11

Amazon CloudSearch DataSource Plugin for CakePHP

 11 commits

 1 branch

 5 releases

 1 contributor

 branch: master ▾ cakephp-cloud-search / + 

Optimized to remove ASCII codes from the document

 violetyk authored on Aug 25, 2014

latest commit 00d39ac6f0 

 Model

Optimized to remove ASCII codes from the document

9 months ago

<https://github.com/nanapi/cakephp-cloud-search>



国内のAmazon CloudSearch事例

- 大規模にご利用いただいているケース: ChatWork

AWS 導入事例：ChatWork 株式会社



～4 億件を越え、日々加速度的に増え続けるチャットのメッセージ検索を、わずか 3 ヶ月という短期間で構築することができました～

<http://aws.amazon.com/jp/solutions/case-studies/chatwork/>

既に5億件を越えたと伺っています！

多言語対応したバージョンのCloudSearchを初期にご採用いただき、USの開発チームも一丸となってローンチまでもっていきました！

国内のAmazon CloudSearch事例

- 大規模にご利用いただいているケース: ChatWork

Speaker Deck Published on Oct 17, 2014

開発者からみたCloudSearch

- よくわかる Amazon CloudSearch -

 chatwork 田中佑樹

Speaker Deck Published on Oct 17, 2014

運用

- ほぼメンテナンスフリーな状態で運用出来ている
- 検索速度についてはかなり改善されたのでお客様から一定の評価をいただいている。

<https://speakerdeck.com/tanakayuki/kai-fa-zhe-karamitacloudsearch>



国内のAmazon CloudSearch事例

- 様々なところでご利用いただいています！

認定ランサーをさがす | クラウドソーシング「ランサーズ」


[ログイン](#)


[無料会員登録](#)

[ヘルプ&お問い合わせ](#)


[サイトマップ](#)

Lancers

 初めての方

 仕事をさがす

 フリーランスをさがす

 仕事を相談する

仕事依頼の相談 平日10～18時

☎ 050-5893-4451

[クラウドソーシング ランサーズ](#) > [認定ランサー検索TOP](#) > 認定ランサーをさがす

認定ランサー検索

認定ランサーとはランサーズに登録している数多くのランサーの中から、ランサーズが定める条件をすべて満たしたランサーを指します。

認定ランサーに相談や依頼を行うことで、より早く・より質の高い提案を集めることが可能です。[認定基準はこちら](#)

全てのカテゴリ

例) iPhoneアプリ

 検索する

いま仕事できる

個人

法人

1週間以内にログイン

国内のAmazon CloudSearch事例

- 様々なところでご利用いただいています！

The screenshot displays the Gochikuru website interface. At the top, there is a navigation bar with the Gochikuru logo, tabs for 'お弁当' (Bento) and 'ケータリング' (Catering), and a shopping cart icon showing 0 items. Below the navigation bar, a banner features the text '宅配弁当ならごちクル' (Home-delivered bento, it's Gochikuru) and '法人・団体様向けデリバリー総合サイト' (Corporate/group-oriented delivery general site). The banner also includes a phone number '050-3785-5190' and a photo of a woman. Below the banner, there are search filters for 'お弁当・店舗を探す' (Find bento/shops) and 'ケータリング・オードブルを探す' (Find catering/aperitifs). The search filters include a dropdown for 'お届けエリアを選択' (Select delivery area) set to '都道府県を選択' (Select prefecture), a dropdown for 'エリアを選択' (Select area), a calendar for 'お届け日時はお決まりですか?' (Is your delivery date/time decided?), a price range input for 'ご予算はお決まりですか?' (Is your budget decided?), and a text input for 'お探しのキーワードはありますか?' (Do you have any search keywords?). At the bottom, there are search buttons for 'お弁当を検索' (Search bento) and 'お店を検索' (Search shops), along with a link for '詳細条件を表示' (Show detailed conditions). On the right side, there is a 'NEW OPEN' section with a photo of a bento box labeled '割烹 しんはま' (Kappo Shinhama) and a 'カタログ請求' (Request catalog) section with a '¥0 無料' (Free) tag.

国内のAmazon CloudSearch事例

- 様々なところでご利用いただいています！



Amazon Web Services

@awscloud



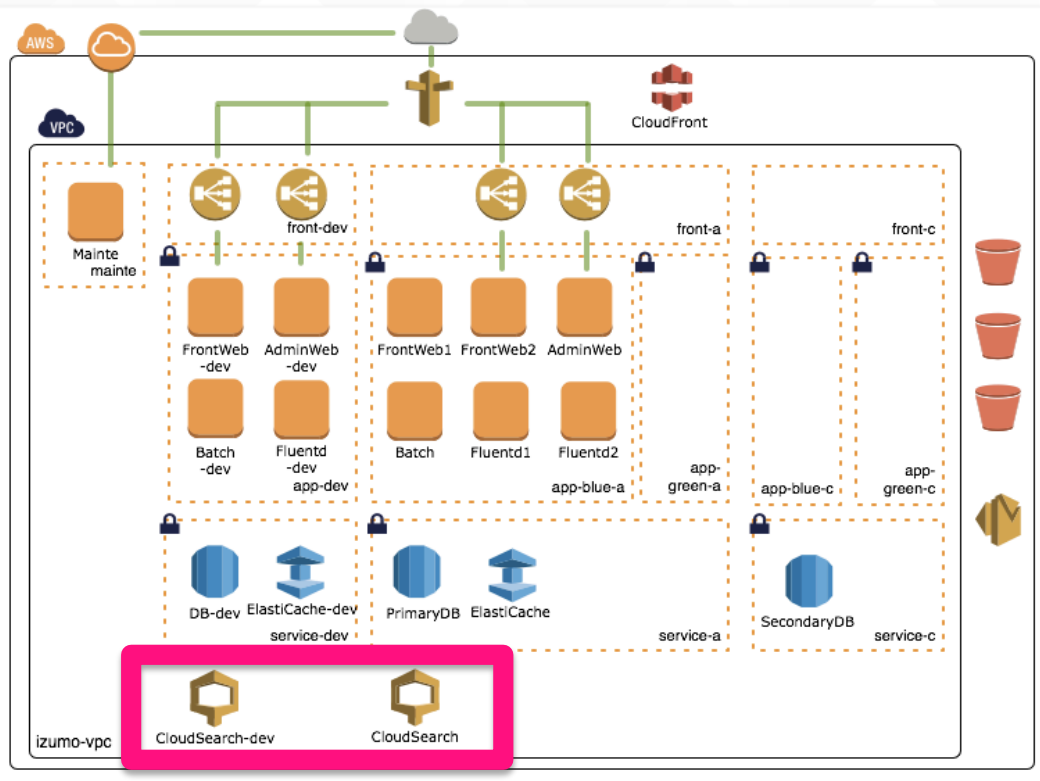
Following

SMART InSight launches Mjgen data analytics platform using Amazon CloudSearch & Redshift. oak.ctx.ly/r/2yhrc



国内のAmazon CloudSearch事例

- 様々なところでご利用いただいています！



『Connected』 - Career Search

企業情報検索

業界から探す



エリアから探す



年収から探す



活かせる仕事・スキルから探す



こだわり条件から探す



フリーワードで絞り込む

フリーワード検索

検索

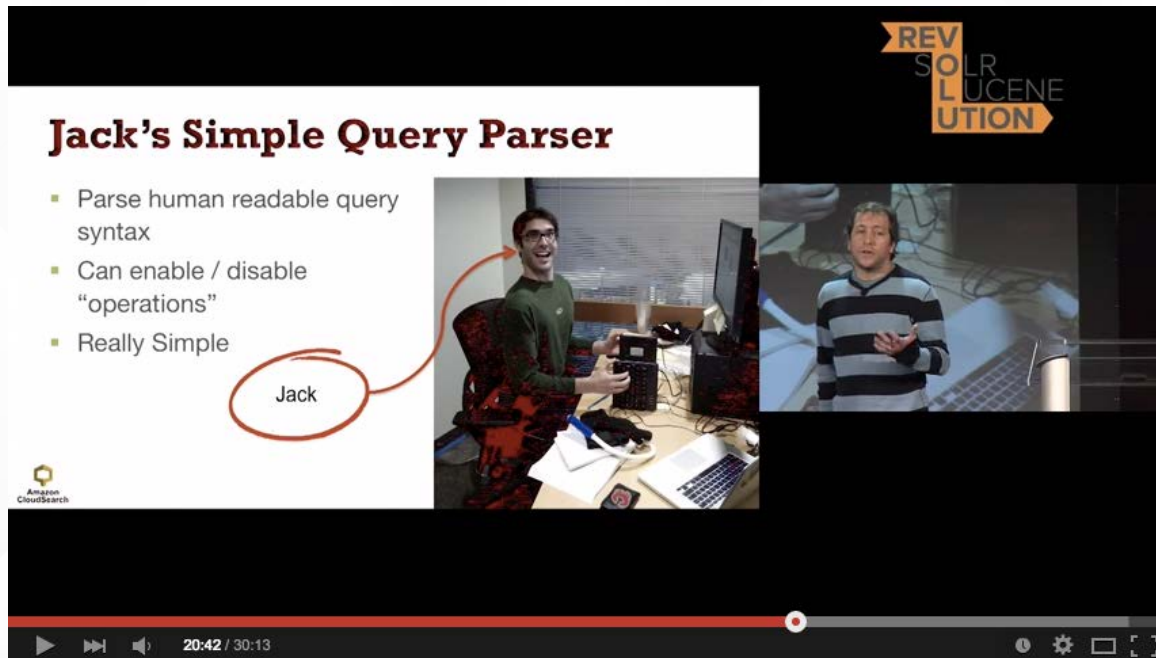
<https://pasona-connected.jp/>



<http://classmethod.jp/cases/social-recruiting-connected/>

更にDive Deepしたい方はこちら

- CloudSearch開発者による講演動画
 - Lucene/Solr Revolution 2014



REV SOLR LUCENE UTION

Jack's Simple Query Parser

- Parse human readable query syntax
- Can enable / disable "operations"
- Really Simple

Jack

Amazon CloudSearch

20:42 / 30:13

<https://www.youtube.com/watch?v=RI1x0d-yO8A>



Tomás from A9 is introducing Amazon #CloudSearch (on top of Lucene/Solr) at #LuceneSolrRev :)



更にDive Deepしたい方はこちら

- CloudSearch開発者による講演動画
 - 例えば DocValues について

DocValues

- Introduced in Lucene 4.0, but made available from Solr in Solr 4.2
- Document-to-value mapping built at index time (vs. “uninverting” at search time). Can be used for faceting, sorting, function queries, etc
- No FieldCache “insanity”
- Easy to use
 - `<field name=“foo” type=“string” indexed=“true/false” stored=“true/false” docValues=“true” />`
- Memory efficient!

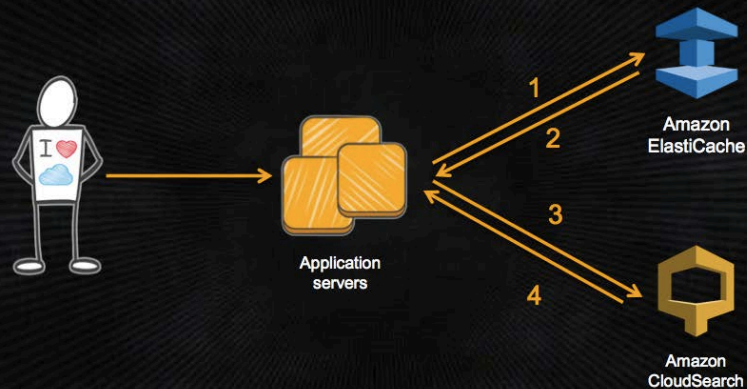
更にDive Deepしたい方はこちら

- CloudSearch SAによる講演

- AWS re:Invent 2014
- Pro Tips and Rule of thumb
- Slide: <http://goo.gl/pkIAzW>
- Youtube: <http://youtu.be/OeHaj1a66I4>



Caching



User behavior tracking



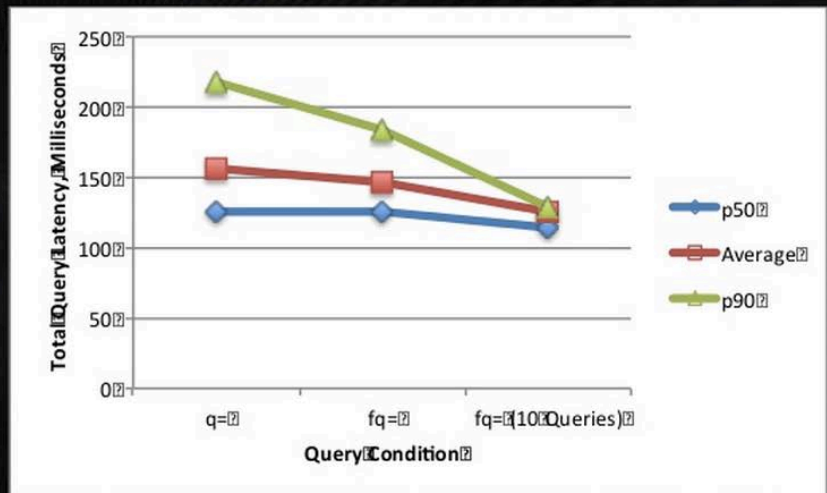
更にDive Deepしたい方はこちら

- CloudSearch SAによる講演
 - 例えば q と fq の性能差

<http://www.slideshare.net/AmazonWebServices/sdd411-amazon-cloudsearch-deep-dive-and-best-practices-aws-reinvent-2014/39>

Filtering vs. querying

- Filters do not contribute to ranking
- Filters can be faster
- Filters support the structured query language



Times include HTTP transport

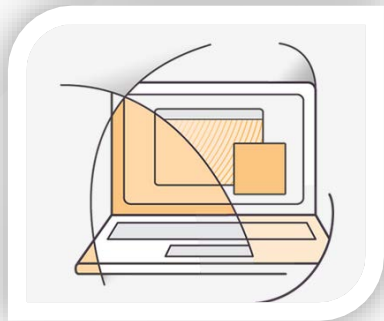
最後に

- Amazon CloudSearchは今後も多くのアップデートを予定しております
 - どんな機能が出てくるか楽しみにしててください！
 - 皆さまからのフィードバックをお待ちしておりますので、引き続きよろしくお願い致します。

Amazon CloudSearch



AWSトレーニング @ AWS Summit Tokyo

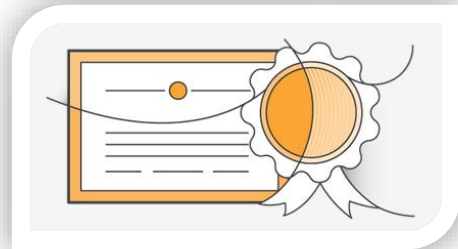


セルフペースラボ : @パミール1F 瑞光

AWS クラウドに実際に触れてみませんか？
ご自分の AWS アカウントをおつくりいただけなくても、
AWS クラウドを体験いただけます。

AWS認定試験（有償） : @ パミール1F 黄玉

特設認定試験会場を AWS Summit Tokyo 2015 会場に開設
Devopsエンジニア-プロフェッショナル認定試験を先行受験いただけます。



AWS認定資格者取得専用ラウンジ : @ パミール1F 青玉

他の AWS 認定資格をお持ちの方とのネットワーキングにぜひラウンジをご活用
ください。
お席や充電器、お飲物などを用意し、皆様をお待ちしております。



Thank You