

# Droongaの はじめかた

須藤功平

クリアコード

全文検索エンジンGroonga勉強会@札幌 2014/03/29  
2014/03/29



# この勉強会とこの話の関連

- 勉強会
  - Groongaの利用事例を募集
- この話
  - Groongaを使ったプロダクト  
Droongaの話



# お知らせ1

grnga

4.0.1リリース！



# 4.0.1の新機能1

- adjuster
  - スコアブースト機能
  - 使用例: 運営のオススを上に表示
- 重み付きベクター
  - レコード毎にn個の重みを設定
  - adjusterと連携するとよい



## 4.0.1の新機能2

- データベース肥大化抑制
  - Milkodeユーザーにも効果大なはず
  - 注意：データベースの再作成が必要
  - データベースの後方互換性あり



# お知らせ2

drnga

1.0.1リリース！



# 1.0.1の新機能

- adjuster対応
- 重み付きベクトル対応
- ユーザー視点の使い勝手向上
  - grn2drn-schema, droonga-send, ...

今月は開発者としてよりもユーザーとして触っていたから



# 今日の目標

1人以上  
Droongaデビュー





# 話すこと

## Droonga 未経験者向け ヒント



# 流れ

- **Droongaとはなにか**
- Groongaの使い方
- ↑とDroongaの使い方の違い
- Droongaの考え方のイメージ



# Droongaとはなにか

droonga

Distributed Groonga



# Distributed?

スケールアウト  
可能



# Droonga

スケールアウト  
可能な  
Groonga



# Groongaが透けて見える

スケールアウト  
可能な  
Groonga



# GroongaとDroonga

- 一部はGroonga  $\neq$  Droonga
- 一部はGroonga  $=$  Droonga



# Groonga != Droonga

- APIは違う
  - DroongaにGroonga互換レイヤーあり
- 構成は違う
  - 1台でDroongaを使っても嬉しくない
- 性能特性は違う
  - 1台で捌ける量→Groongaの方が速い





# Groonga == Droonga

- クエリーの書き方は同じ
- スキーマの考え方は同じ
- 検索機能はだいたい同じ

# Droongaをはじめするために

教養として  
Groongaを  
押さえて  
おくべき！



# Groongaの使い方

- Droongaとはなにか
- **Groongaの使い方**
- ↑とDroongaの使い方の違い
- Droongaの考え方のイメージ



# なぜまずGroongaか

## Groongaはお手軽

- 1台のマシンで動く
- 設定ファイルなしで動く
- コマンドラインで動く



# Groongaを使う流れ

初心者向けの簡易版

1. **スキーマを作る**
2. インデックスを作る
3. データを入れる
4. 検索する



# スキーマを作る

- 検索したいものを**1つ**決める
  - 例：たいやき屋検索ならたいやき屋
- 検索したいものの集まりをテーブルにする
  - 例：TaiyakiShopsテーブル
- 付加情報をカラムにする
  - 例：店舗名、場所、メニューなど



# スキーマ例

```
table_create TaiyakiShops \  
  TABLE_HASH_KEY ShortText  
# 店舗名  
column_create TaiyakiShops name \  
  COLUMN_SCALAR ShortText  
# 場所  
column_create TaiyakiShops location \  
  COLUMN_SCALAR WGS84GeoPoint
```



# Groongaを使う流れ： 2

1. スキーマを作る
2. **インデックスを作る**
3. データを入れる
4. 検索する





# インデックスを作る

- 全文検索対象を決める
  - 例：店舗名
- トークンの集合を格納する  
テーブルを作る
  - トークン：最小検索単位
- ↑にインデックスカラムを作る



# インデックス例

```
# トークン用テーブル
# 初心者は常にこれでOK
table_create Tokens \
  TABLE_PAT_KEY ShortText \
  --default_tokenizer TokenBigram \
  --normalizer NormalizerAuto
# インデックスカラム
column_create Tokens taiyaki_index \
  COLUMN_INDEX|WITH_POSITION \
  TaiyakiShops name
```



# Groongaを使う流れ： 3

1. スキーマを作る
2. インデックスを作る
3. **データを入れる**
4. 検索する



# データを入れる

- フォーマットはJSON
  - GroongaのJSONパーサーは緩い
  - 多少壊れたJSONも受け付ける



# データ例

```
load --table TaiyakiShops
[
  {
    "_key": "sapporo-yanagiya",
    "name": "札幌の柳屋",
    "location": "43.11629x141.34855"
  }
]
```



# Groongaを使う流れ： 4

1. スキーマを作る
2. インデックスを作る
3. データを入れる
4. **検索する**



# 検索する

- 全文検索対象を指定する
- 検索キーワードを指定する



# 検索例

```
select TaiyakiShops \  
  --match_columns name \  
  --query "札幌"  
# [[...], ← レスポンスヘッダー  
#  [[1], ← ヒット件数  
#   ↓ 出力内容のメタデータ  
#  [["_id","UInt32"],...],  
#   ↓ マッチしたレコード  
#  [1,"sapporo-yanagiya","...","札幌の柳屋"]]]]
```



# Groongaを使う流れの再確認

1. スキーマを作る
2. インデックスを作る
3. データを入れる
4. 検索する



# 流れ3

- Droongaとはなにか
- Groongaの使い方
- ↑とDroongaの使い方の違い
- Droongaの考え方のイメージ



# Droongaを使う流れ

1. クラスタを作る ← **New**
2. スキーマを作る
3. インデックスを作る
4. データを入れる
5. 検索する



# クラスターを作る

大事だけど時間がないので

# 省略！

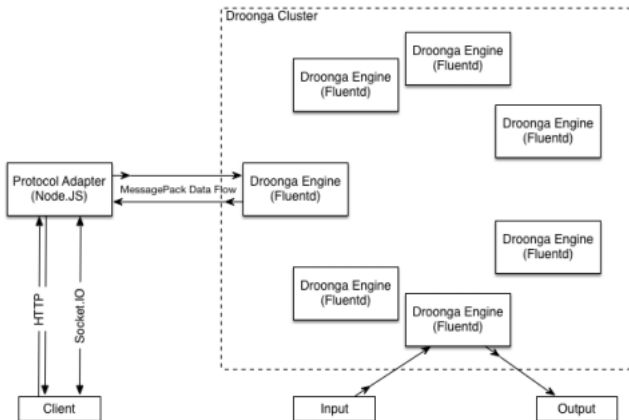


# 流れ4

- Droongaとはなにか
- Groongaの使い方
- ↑とDroongaの使い方の違い
- **Droongaの考え方のイメージ**



# Droongaの構成



<http://droonga.org/overview/>



# まとめ1

- Droonga
  - スケールアウト可能なGroonga
  - Groongaと同じことがいろいろある
  - Groongaの理解が役立つ



# まとめ2

- Groongaの使い方
  - a. スキーマを作る
  - b. インデックスを作る
  - c. データを入れる
  - d. 検索する





# まとめ3

- Droongaの使い方
  - a. クラスターを作る
  - b. 以下、Groongaと同じ



# まとめ4

## Droongaの構成

- n個のGroongaで  
大きなGroongaを構成
- Droonga=大きなGroonga



# 今日の目標の確認

1人以上  
Droongaデビュー  
を見据えてGroongaデビュー