

Mr oonga Meetup

HAYASHI Kentaro クリアコード

2014/06/29





はじめに

- 開催の目的
- 自己紹介
- Mroongaとは



はじめに(2)

- インストール
- 課題と演習
- Any Questions?
- 片付け撤退



開催の目的

「全文検索をしたい人の最初の一步を踏みだしやすくしよう」



自己紹介タイム

- 参加のきっかけ
- Groonga族とのかかわり



Mroongaとは

- Tritonnの後継
 - Sennaを組み込んだプロダクト
- Tritonnの制約を解消



Tritonn vs Mroonga(1)

- TritonnはMySQLへのパッチ前提
- Mroongaはプラグインとして導入できる



Tritonn vs Mroonga(2)

- Tritonnはテーブルロックでスケールしない
- Mroongaは参照ロックフリー
 - 検索/更新性能の向上



Mrroongaとパッケージ

- Debian
- Ubuntu
- CentOS
- Fedora
- Windows



Mroonga on Debian

- MySQL 5.5系
 - Debian 7.4 (wheezy)
 - Debian 8 (jessie)
 - Debian unstable (sid)



Mrroonga on Ubuntu

- PPA (Launchpad) でリリース
- MySQL 5.5系
 - Ubuntu 12.04 (Precise)
 - Ubuntu 13.10 (Saucy)
 - 来月でサポート終了
 - Ubuntu 14.04 (Trusty)



PPA on Launchpad

■ PPAでのインストール

```
sudo add-apt-repository -y universe
sudo add-apt-repository -y ppa:groonga/ppa
sudo apt-get update
sudo apt-get install -y -V mysql-server-mroonga
```



Mroonga on CentOS 5

- CentOS 5
 - 以前はMySQLが古すぎたので、MySQL 5.6対応パッケージを提供
 - MySQL 5.5 (SCL) 対応パッケージに移行



Mroonga on CentOS 6

- CentOS 6
 - MySQL 5.1しか使えないというのは古い情報です！
 - MySQL 5.5対応(SCL)パッケージに移行



Mroonga on CentOS

- CentOS 5やCentOS 6でMySQL 5.6
 - wingリポジトリをお勧めしています

<http://wing-repo.net/>



Mroonga on Fedora

- Fedora 20
 - MySQL 5.5 & MariaDB 5.5対応
 - MariaDB 5.5対応がつらいのでやや古い
 - Fedora公式リポジトリでは未リリース
 - Mroonga含めてインストールしたいならGroongaリポジトリを登録する



Mroonga on Windows

- 従来zipのみ提供
- ~~ビルド~~協力者増加によりmsiも提供

Mrroongaの動作モードの違い

- ラッパーモード
 - データは他のストレージエンジンに格納
 - 全文検索をMrroongaでがんばる
 - トランザクションでロールバックすると不整合発生
 - チューニングはMySQLの範囲で

Mroongaの動作モードの違い (2)

- ストレージモード
 - 全部Mroongaでがんばる
 - トランザクション何それ？



ストレージモード

- ENGINE=mroongaと覚えましょう



ラッパーモード

- コメントでエンジンを指定

```
ENGINE=mroonga COMMENT='engine "innodb"'
```



パーサーの指定

- コメントでパーサーを指定
- Groongaの用語だとトークナイザ
ー

```
FULLTEXT INDEX (content) COMMENT 'parser "TokenMecab"'
```



ノーマライザーの指定

- コメントでノーマライザーを指定

```
FULLTEXT INDEX (content) COMMENT 'normalizer "NormalizerAuto"'
```

パーサー&ノーマライザーの指定

- 両方指定したいときは、で区切る

```
FULLTEXT INDEX (content) COMMENT 'parser "TokenMecab", normalizer "NormalizerAuto"'
```

MATCH AGAINSTとBOOLEAN MODE

- QUERY EXPANSION MODE
 - MySQLにおまかせの類似文書検索
- BOOLEAN MODE
 - 明示的に検索を制御



UDF (ユーザ定義関数)

- `last_insert_grn_id`
- `mroonga_snippet`
- `mroonga_escape`
- `mroonga_command`



last_insert_grn_id

- 最後に挿入したレコードのIDを返す関数
- ストレージモード用



mroonga_snippet

- スニペットを返す関数



mroonga_escape

- クエリを適宜エスケープするための関数
- 複数のキーワードは別途がんばる必要あり



mrroonga_command

- MrroongaからGroongaの機能を直接叩くための関数



実践してみよう

- 位置情報検索
- エスケープ
- 重みづけして検索してみよう

`http://packages.groonga.org/
tmp/qiita-article.txt`



位置情報検索

☒ Qiitaの位置情報検索の記事を試してみよう

<http://qiita.com/groonga/items/7abca5004d2ae386d388>

mroonga_escapeを使ってみよう

🕒 Qiitaのmroonga_escapeの記事を試してみよう

<http://qiita.com/groonga/items/0ae7f637525c7b5f2cb1>

重みづけして検索してみよう

⌘ QiitaのWプログラマの記事を試してみよう

<http://qiita.com/groonga/items/8bcd5f9647f62b96584f>



Groonga vs Mroonga

- 既存資産があるならMroonga
- `mroonga_command`を活用

大規模なデータを使いたい

- Wikipedia
- 情報学研究データリポジトリ



Wikipedia

- ウィキペディア日本語版
 - <http://ja.wikipedia.org/wiki/Wikipedia:データベースダウンロード>

情報学研究データリポジトリ (1)

- 国立情報学研究所
- <http://www.nii.ac.jp/cscenter/idr/index.html>
- データセットをいくつか提供

情報学研究データリポジトリ (2)

- ニコニコデータセット
 - ニコニコ動画コメント等データ
 - ニコニコ大百科データ

ニコニコ動画コメント等データ(1)

- 動画メタデータ
 - 2012年11月初旬までのデータ
 - IDなど個人を特定しないデータ
 - タイトル, 説明文, タグ, 投稿日時, 再生数, コメント等
 - 約12GBほど

ニコニコ動画コメント等データ(2)

- ニコニコ大百科データ
 - 2014年2月上旬までのデータ
 - 約300GB



感謝

- WEICさん
会場提供ありがとうございます！



おわり

- Any Questions?