

Mr oonga 2016

高速日本語全文検索 for MariaDB
Super fast full text search for MariaDB

Kouhei Sutou

ClearCode Inc.

*MariaDB Community Event in Tokyo
2016-07-21*





Mroonga

- 読み方：むるんが

Pronunciation: múlúngá

- ストレージエンジン

Storage engine

- MariaDBバンドル

Bundled in MariaDB

- 別途インストールしなくてもよい

No need to install Mroonga separately



特徴

Characteristics

- **高速日本語全文検索** (全言語OK)
Super fast full text search for all languages
- **カラムストアによる高速処理**
Super fast processing by column store architecture
- **全文検索初心者でも使える**
Easy to use by full text search beginners
- **全文検索上級者は活用できる**
Features for full text search specialists



高速日本語全文検索

Super fast full text search

1. ベンチマーク

Benchmark

2. 速さの秘密

The reason why Mroonga is fast



ベンチマーク環境

Benchmark environment

- **対象：Wikipedia日本語版**
Target: Japanese version Wikipedia
- **レコード数：約185万件**
The number of records: About 1.85 millions
- **データサイズ：約7GB**
Data size: About 7GB
- **メモリー4GB・SSD250GB (ConoHa)**
Memory: 4GB, SSD: 250GB



補足

Supplement

■ MySQL 5.7を使用

MySQL 5.7 is used

- MariaDBのInnoDBは日本語未対応

InnoDB in MariaDB doesn't support Japanese yet

■ 他人のベンチマークは参考程度

Just refer benchmark result by others

- 検討時は実環境でベンチマークを！

Run benchmark with the real data on real env

詳細(Detail) :

<https://github.com/groonga/wikipedia-search/issues/4>



検索1

Search1

キーワード：テレビアニメ

(ヒット数：約2万3千件)

Keyword: TV animation

(N hits: About 23K)

InnoDB ngram	3m2s
InnoDB MeCab	6m20s
Mroonga: 1	0.11s



検索2

Search2

キーワード：データベース

(ヒット数：約1万7千件)

Keyword: Database

(N hits: About 17K)

InnoDB ngram	36s
InnoDB MeCab: 1	0.03s
Mroonga: 2	0.09s



検索3

Search3

キーワード : PostgreSQL OR MySQL

(ヒット数 : 約400件)

Keyword: PostgreSQL OR MySQL

(N hits: About 400)

InnoDB ngram	N/A(Error)
InnoDB MeCab: 1	0.005s
Mroonga: 2	0.028s



検索4

Search4

キーワード：日本

(ヒット数：約63万件)

Keyword: Japan

(N hits: About 630K)

InnoDB ngram	1.3s
InnoDB MeCab	1.3s
Mroonga: 1	0.21s



検索まとめ

Wrap up search

- Mroonga : 安定して速い
Always fast
- InnoDB FTS MeCab
 - ハマれば速い
Fast only for one token query
- InnoDB FTS ngram
 - 安定して遅い
Always slow



速さの秘密

The reason why Mroonga is fast

■ 最適化された転置索引実装

Optimized inverted index implementation

■ 2段階のデータ圧縮

2 level data compression

■ 高速なポスティングリスト探索

Fast posting list search

■ 検索だけでなく更新も速い

Not only search but also update is fast

11年以上開発が続いている全文検索エンジンGroongaを使用
Groonga full text search engine (11 years old) is used



もっと速さの秘密

More reasons why Mroonga is fast

■ カラムストアを活かした最適化

Optimizations based on column store architecture

- ポイント1：余計なI/Oを減らす

Point1: Reduce needless I/O

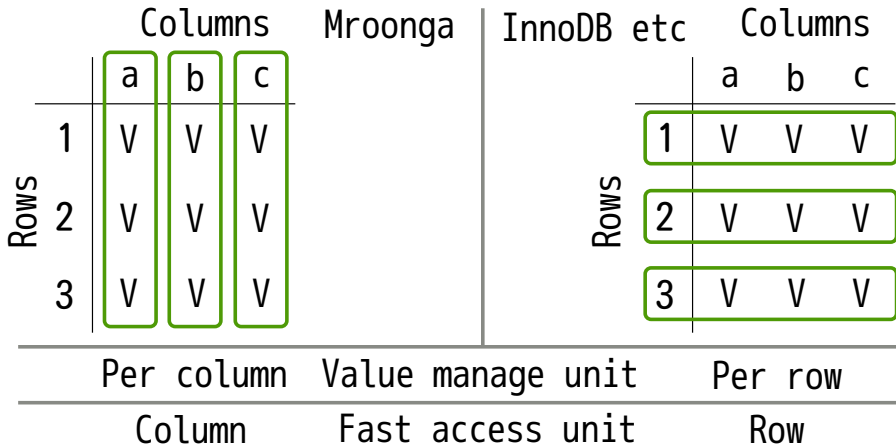
- ポイント2：I/Oを局所化

Point2: Localize I/O



カラムストア

Column store



必要なカラムのみアクセス

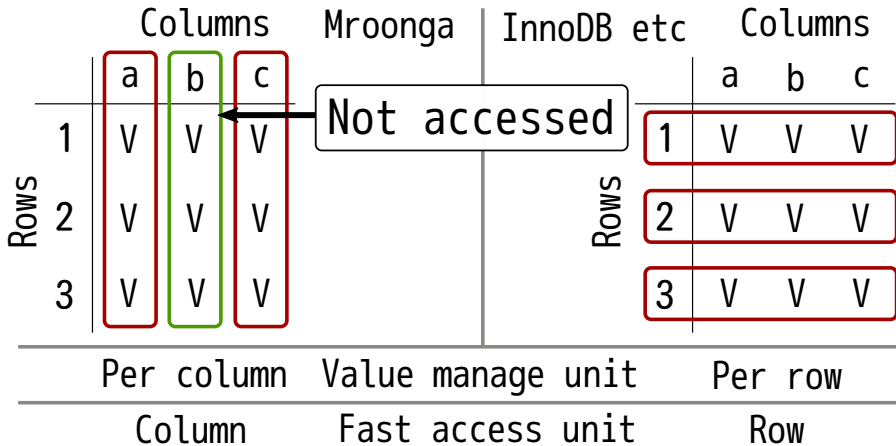
Access to only needed columns

```
-- Access to only a  
SELECT a  
FROM table  
  
-- Access to only c  
WHERE c = XXX;  
  
-- b isn't accessed
```



減ったI/O

Reduced I/O





行カウント

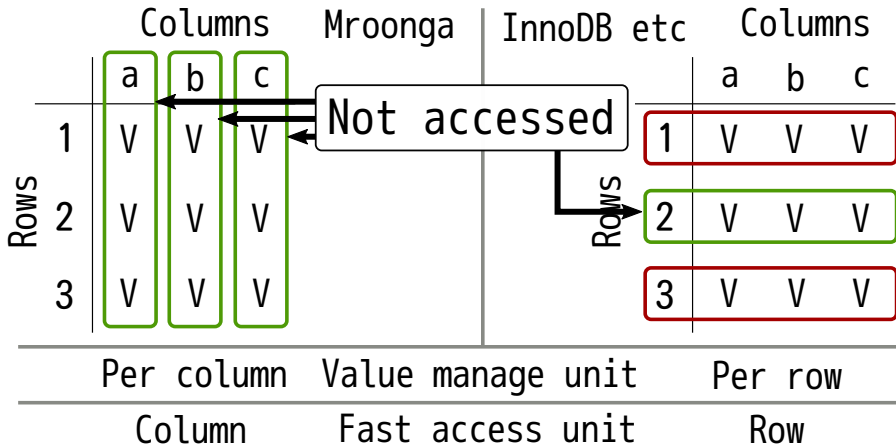
Row count

```
-- No column values are needed
SELECT COUNT(*)
  FROM table
-- Access to only full text search index of c
  WHERE MATCH(c)
        AGAINST('+keyword' IN BOOLEAN MODE);
-- a, b and c aren't accessed
```



減ったI/O

Reduced I/O





ORDER BY LIMIT

```
SELECT *  
  FROM table  
 WHERE MATCH(c)  
        AGAINST('+keyword' IN BOOLEAN MODE)  
-- Mroonga processes ORDER BY LIMIT  
-- instead of MariaDB  
-- → Mroonga returns only 10 records  
--   to MariaDB instead of all matched records  
ORDER BY a LIMIT 10;
```

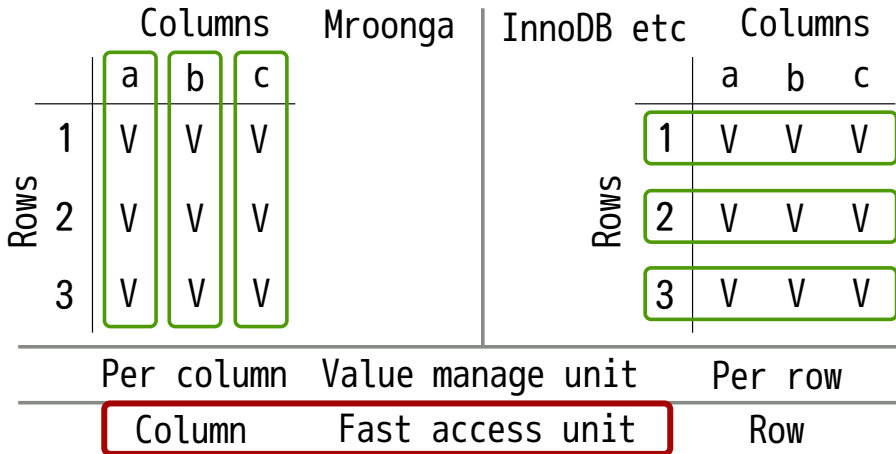
Optimized ORDER BY LIMIT

- 検索 (Search) by Mroonga
 - カラム毎の処理でI/Oを局所化
(索引非使用時)
Localize I/O by per column processing
(on no index case)
- ソート (Sort) by Mroonga
 - カラム毎の処理でI/Oを局所化
Localize I/O by per column processing
- OFFSET/LIMIT by Mroonga



カラム毎の処理は速い

Per column processing is fast





最適化のまとめ

Wrap up optimization

- 転置索引実装が速い
Inverted index implementation is fast
 - 検索も更新も速い
Both search and update are fast
- カラムストアで速い
Fast by column store architecture
 - ポイント：I/O削減・I/O局所化
Points: Reduce and localize I/O

全文検索初心者でも使える

Easy to use by beginners

- インストールが簡単
Easy to install
- MySQLの標準機能のみで使える
Usable only with MySQL standard features



インストールが簡単

Easy to install

- MariaDBバンドル
MariaDB bundles Mroonga
- Apt/Yumリポジトリ
Apt/Yum repositories
- MariaDB込みのWindowsバイナリ
Windows binary with MariaDB



標準機能のみで使える

Require only MySQL standard features

```
-- Create  
CREATE TABLE table (  
    -- . . . ,  
    FULLTEXT INDEX (column)  
) ENGINE=Mroonga;
```



標準機能のみで使える

Require only MySQL standard features

-- *Convert*

```
ALTER TABLE table
```

```
  ADD FULLTEXT INDEX (column)
```

```
  ENGINE=Mroonga;
```



標準機能のみで使える

Require only MySQL standard features

```
SELECT * FROM table
WHERE
    MATCH(column)
    AGAINST('+keyword'
            IN BOOLEAN MODE);
```



全文検索上級者向け機能

Features for specialists

■ カスタマイズ

Customizable

- デフォルト値はいい感じ

→初心者はカスタマイズなしでよい

Suitable default values

→Beginners don't need to customize

■ Groongaの機能をもっと使える

(高速・高機能)

Specialists can use more Groonga features

(Fast and high functionality)



文字正規化ルール変更

Change normalizer

```
CREATE TABLE table (  
    -- ...,  
    FULLTEXT INDEX (column)  
    --  
    -- Specify a parameter as comment  
    COMMENT='normalizer "NormalizerAuto"  
) ENGINE=Mroonga;
```



文字正規化ルール変更

Change normalizer

```
CREATE TABLE table (  
    -- ...,  
    FULLTEXT INDEX (column)  
    -- MariaDB:  
    -- Custom parameter can be used  
    NORMALIZER='NormalizerAuto'  
    ) ENGINE=Mroonga;
```

Groongaの検索機能を使う

Use full Groonga search features

```
SELECT * FROM table
WHERE
  -- "c1" is meaningless with "*SS" pragma
  MATCH(c1)
  -- "*SS" is a pragma to use
  -- full Groonga search features
  -- Multiple indexes can be used in A query
  AGAINST('*SS c1 @ "keyword" && c2 < 100'
          IN BOOLEAN MODE);
```



今後 Futures

■ 最新機能サポート

Support the latest features

■ JSONを全文検索

(JSON型のデータの読み書きは対応済み)

Full text search against JSON

(Storing/fetching JSON are already supported)

■ virtual column/generated column

■ 最新版をMariaDBにバンドル

Bundle the latest Mroonga to MariaDB



最新版をバンドル

Bundle the latest Mroonga

- Mroongaは毎月リリース
Mroonga is released monthly
- MariaDB 10.2.1 bundles
Mroonga **5.04**
 - The latest Mroonga is 6.06
 - Mroonga supports MariaDB 10.2
since **6.03**
 - How can we improve this?



まとめ1

Wrap up1

- **高速日本語全文検索** (全言語OK)
Super fast full text search for all languages
- **カラムストアによる高速処理**
Super fast processing by column store architecture
- **全文検索初心者でも使える**
Easy to use by full text search beginners
- **全文検索上級者は活用できる**
Features for full text search specialists



まとめ2

Wrap up2

- 今後もMroongaは便利になる
We continue to improve Mroonga
- MariaDBで最新Mroongaを使える
MariaDB will bundle the latest Mroonga

MariaDBで全文検索ならMroonga!
Mroonga is the best for full text search on MariaDB!