

# グルンガの紹介

groonga開発者から

未来検索ブラジル

森大二郎

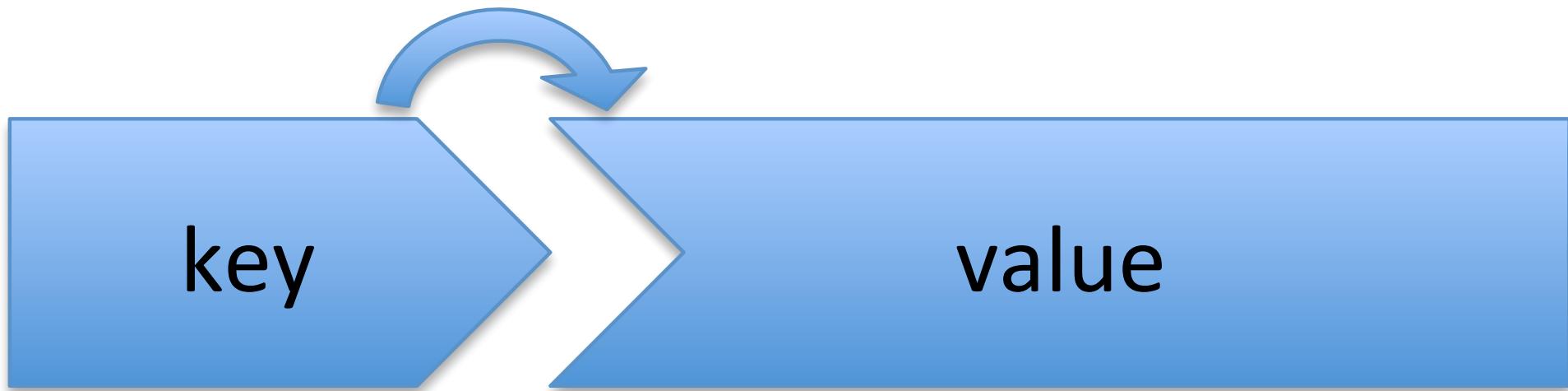
# 今日の内容

- 3分でわかるgroongaの核心
- \*roonga族の使い分け
- groonga開発予報

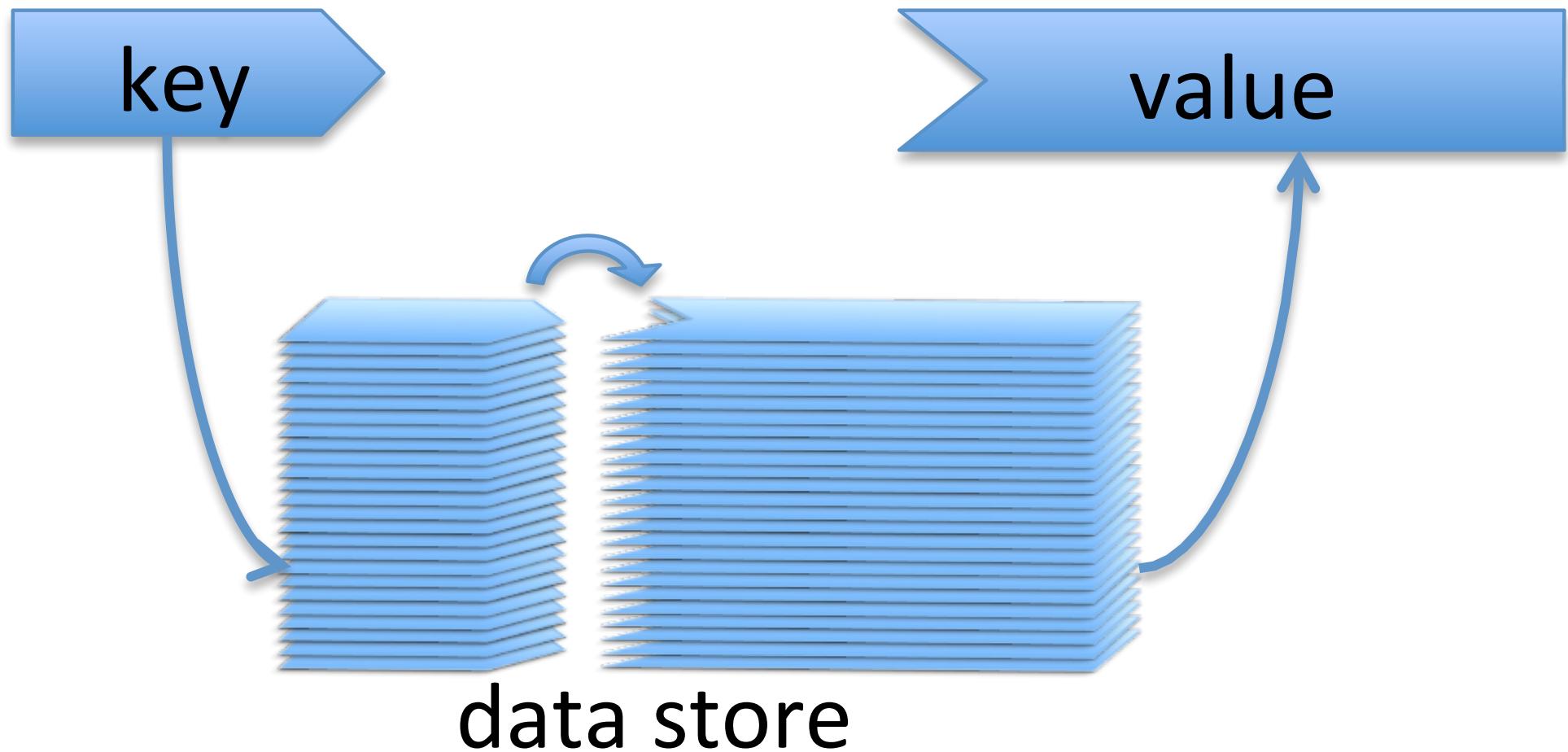
# **groonga って何ですか?**

- 全文検索が得意な
- カラム指向のデータストア
- 巍のKVSやNoSQLとどう違う?

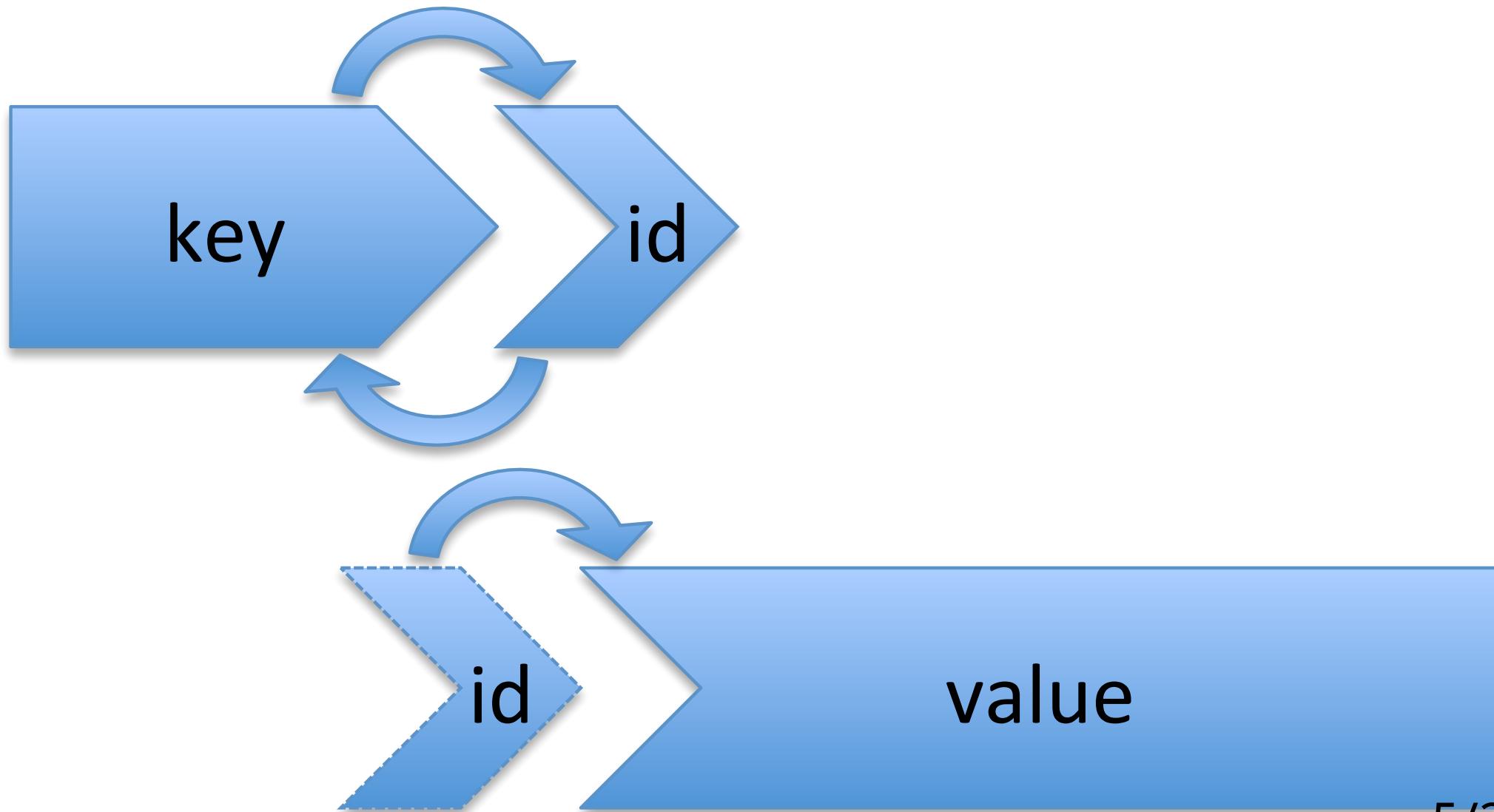
# KVSのkey-valueペア



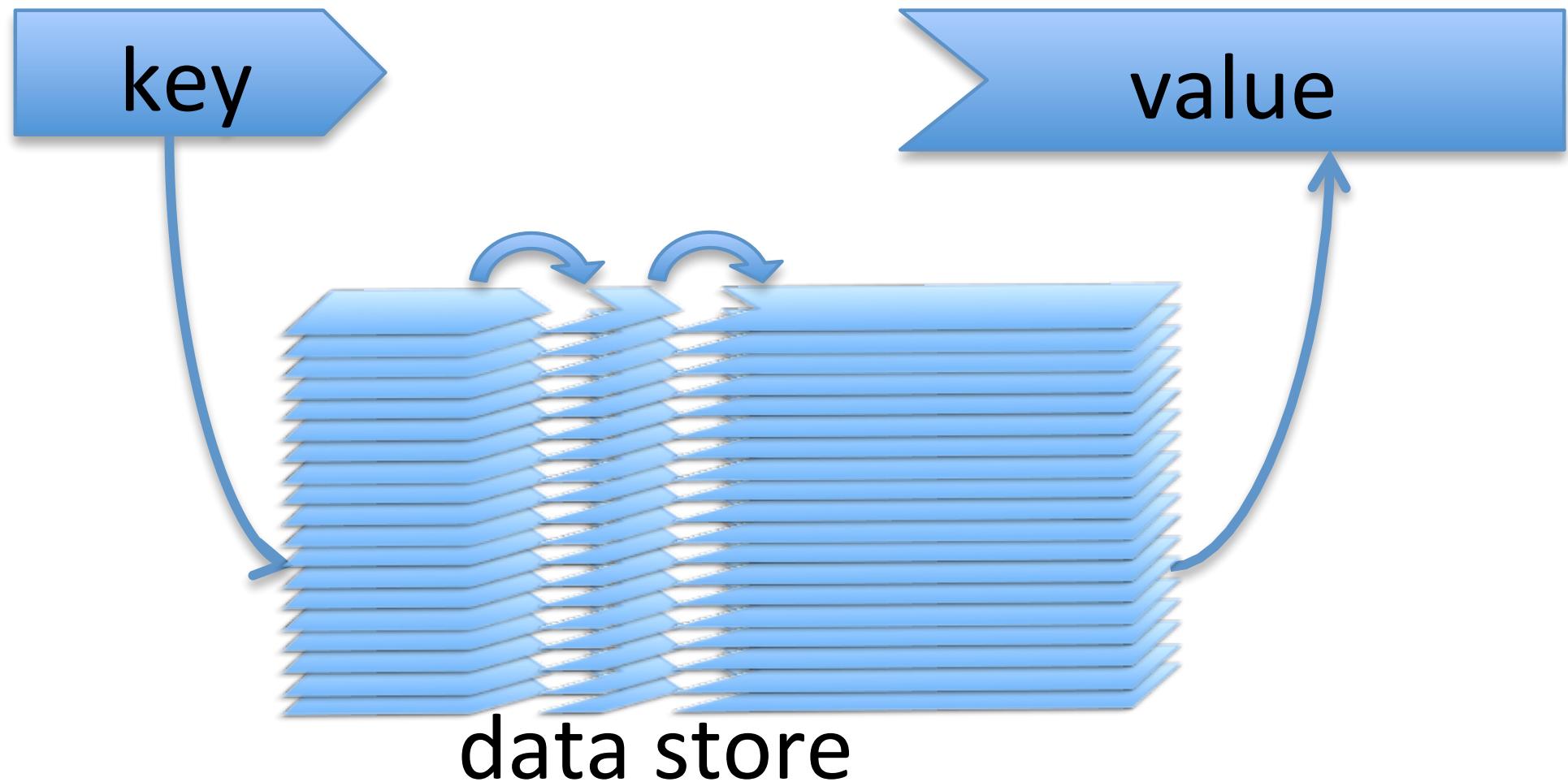
# KVSへのアクセス



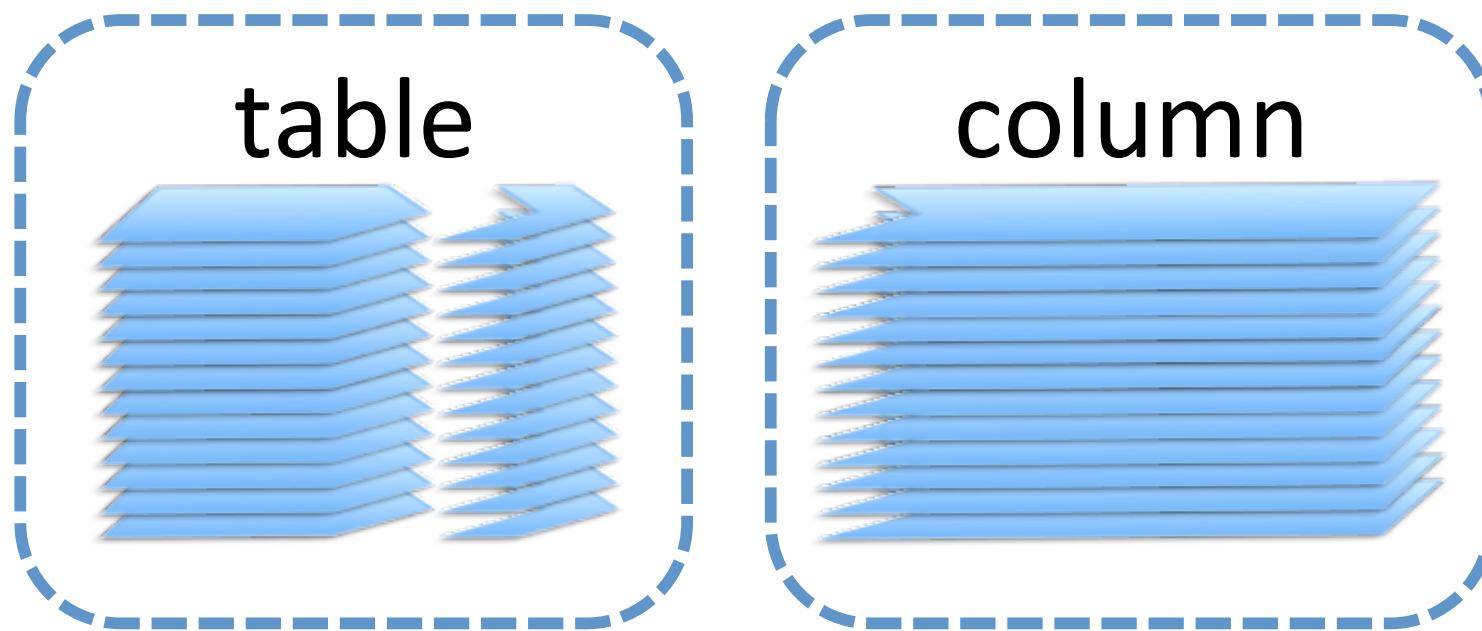
# groongaのデータモデル



# groongaへのアクセス



# groongaにおけるtableとcolumn

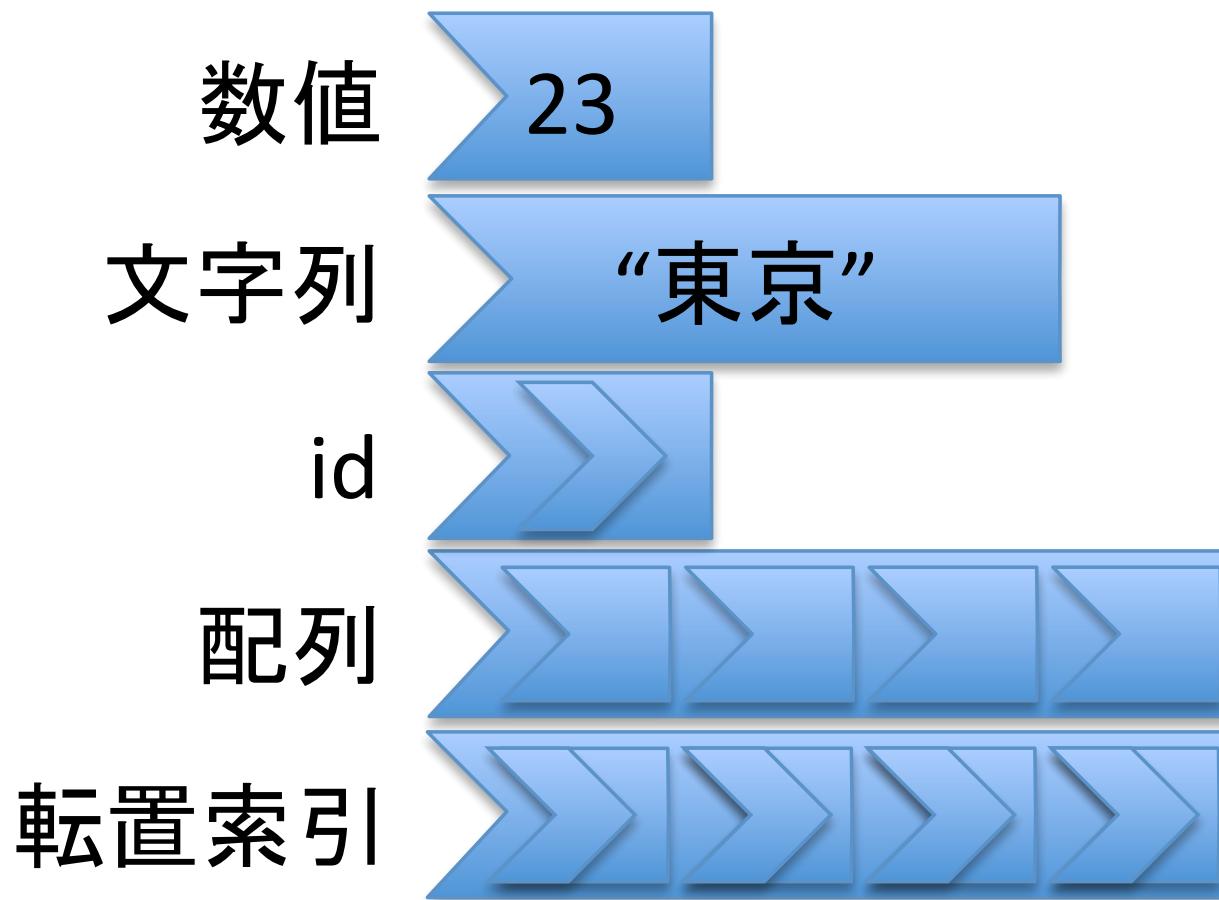


# groongaにおけるtable

- key→id, id→keyを高速に検索
- ハッシュ表, トライなど
- 機能面・性能面でそれぞれ個性がある

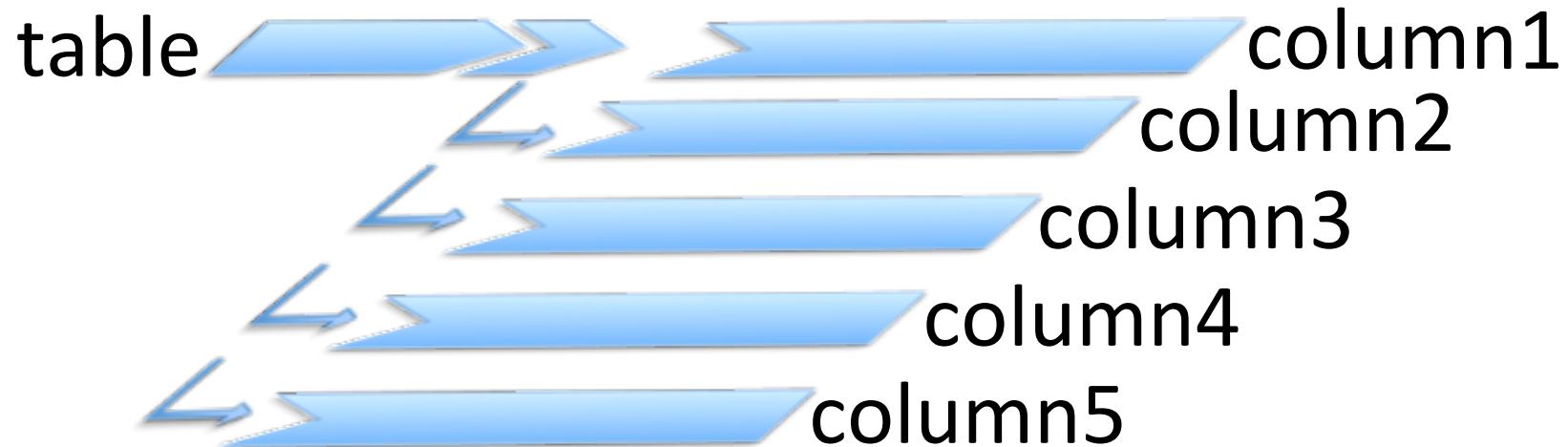
# groongaにおけるcolumn

個々に固有のデータ構造を持つ

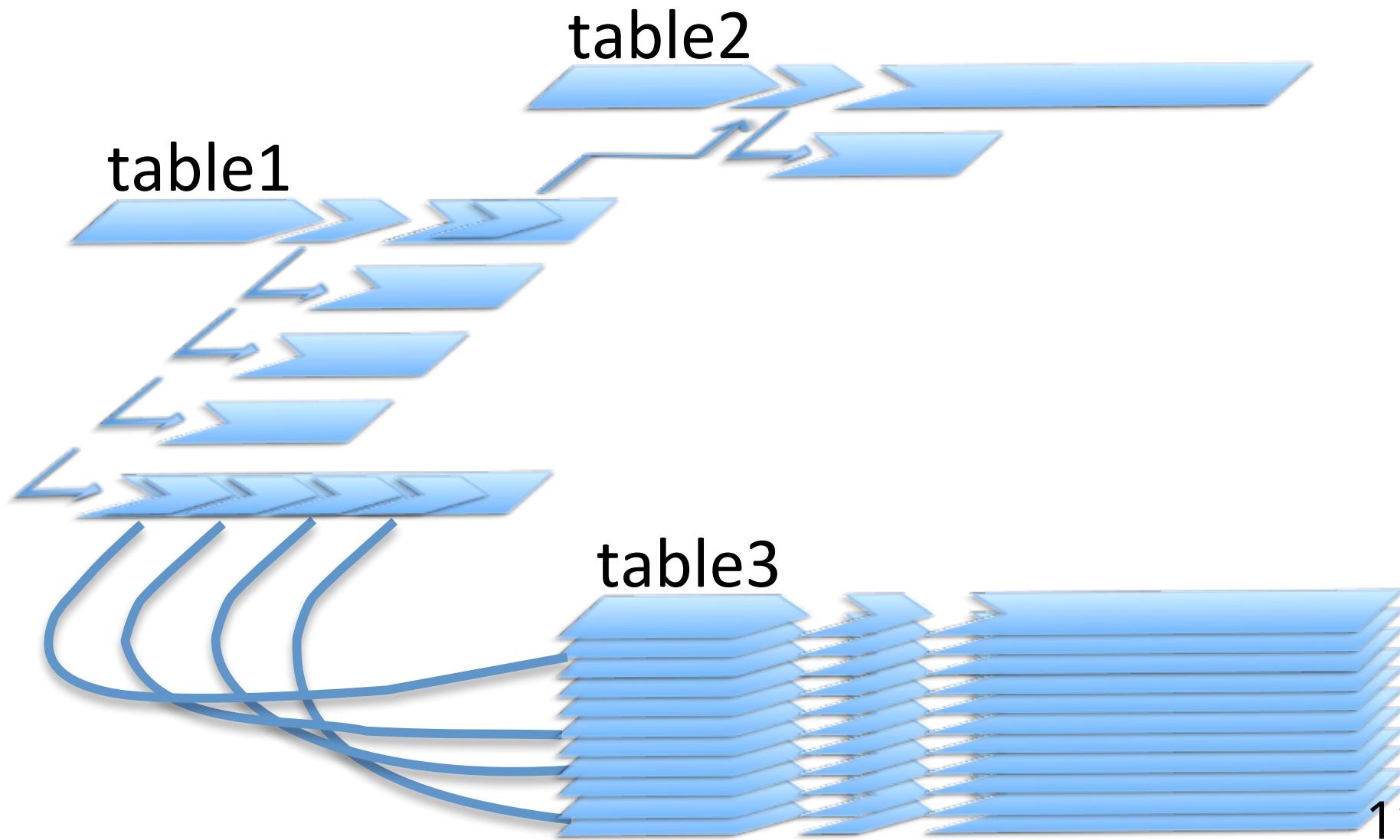


# **id**と**value**を分けることで

複数のカラムを持つてゐる



# 多様なデータ構造も表現



# groongaの特徴

- DBを複数プロセスで共有可能
- 更新処理は、
- table,column単位で
- atomicに実行

# リアルタイムウェブなどに!

- 更新も検索も頻繁な環境で
- 更新性能と検索性能を両立
- コア数の多い環境で
- CPU性能を最大限発揮

# \*roonga族の使い分け

クエリによって処理時間が異なる

- key参照 : usec単位
- index参照 : msec単位
- 全文検索 : 10m~100msec
- tablescan: sec単位

# \*roonga族の使い分け

kvs的に使うなら

---

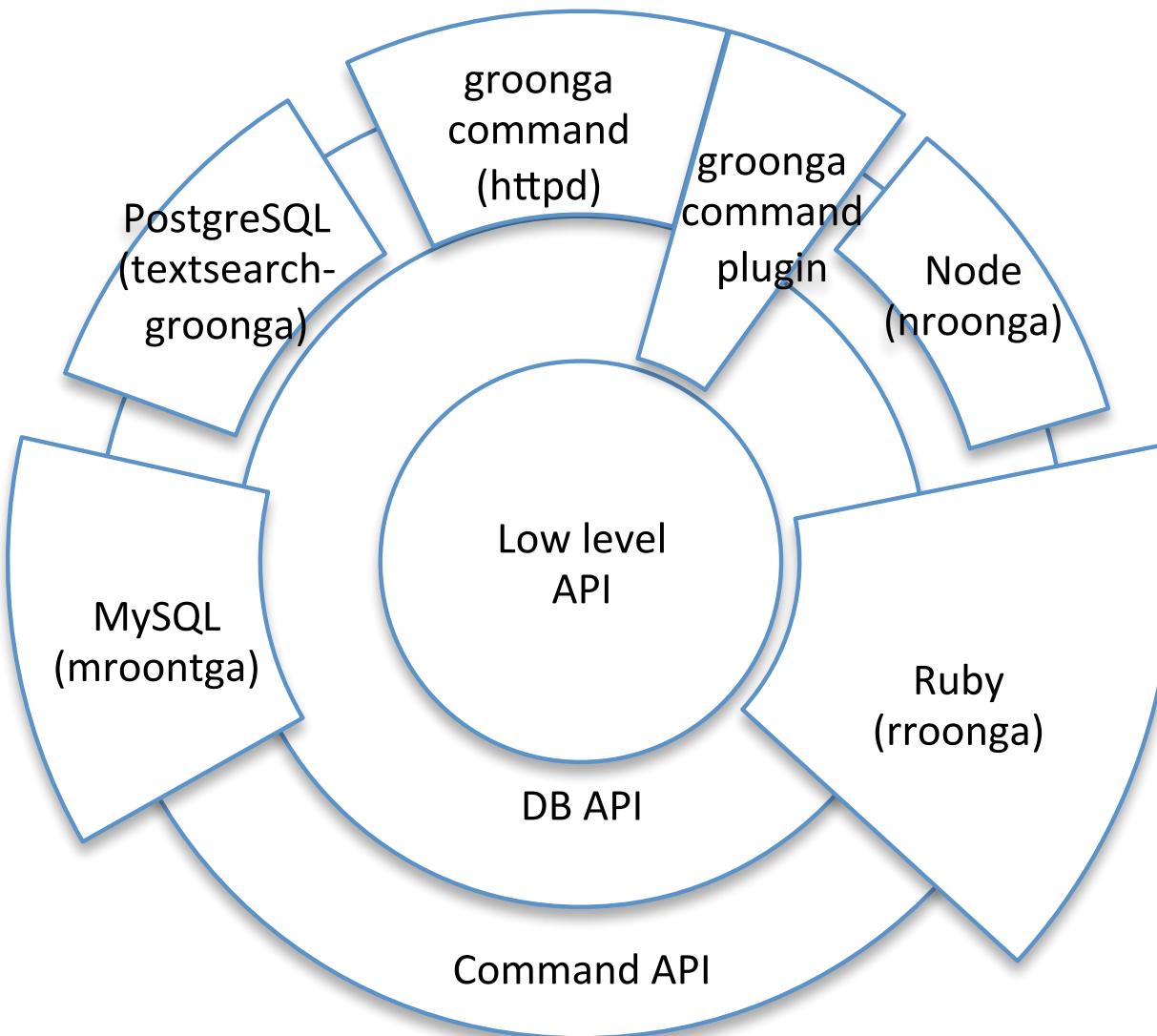
低レイヤAPIが高速

全文検索なら

---

ネットワーク経由で十分

# 使用可能なAPIの違い



# 今日のgroonga

**groonga2.0.9 リリース!**

---

高度な検索機能をサポート!

**fluent-plugin-groonga**

---

fluentを使ってお手軽レプリケーション

# groonga開発予報

- スケーラビリティ向上
- scale up, scale out どちらも

# scale up

## groonga++

最大レコード数を1000倍に拡大

静的なデータは圧縮

MARISA-trieが使用可能に

(gzipより小さく圧縮 & 高速検索!!)

<https://github.com/groonga/grnxx>

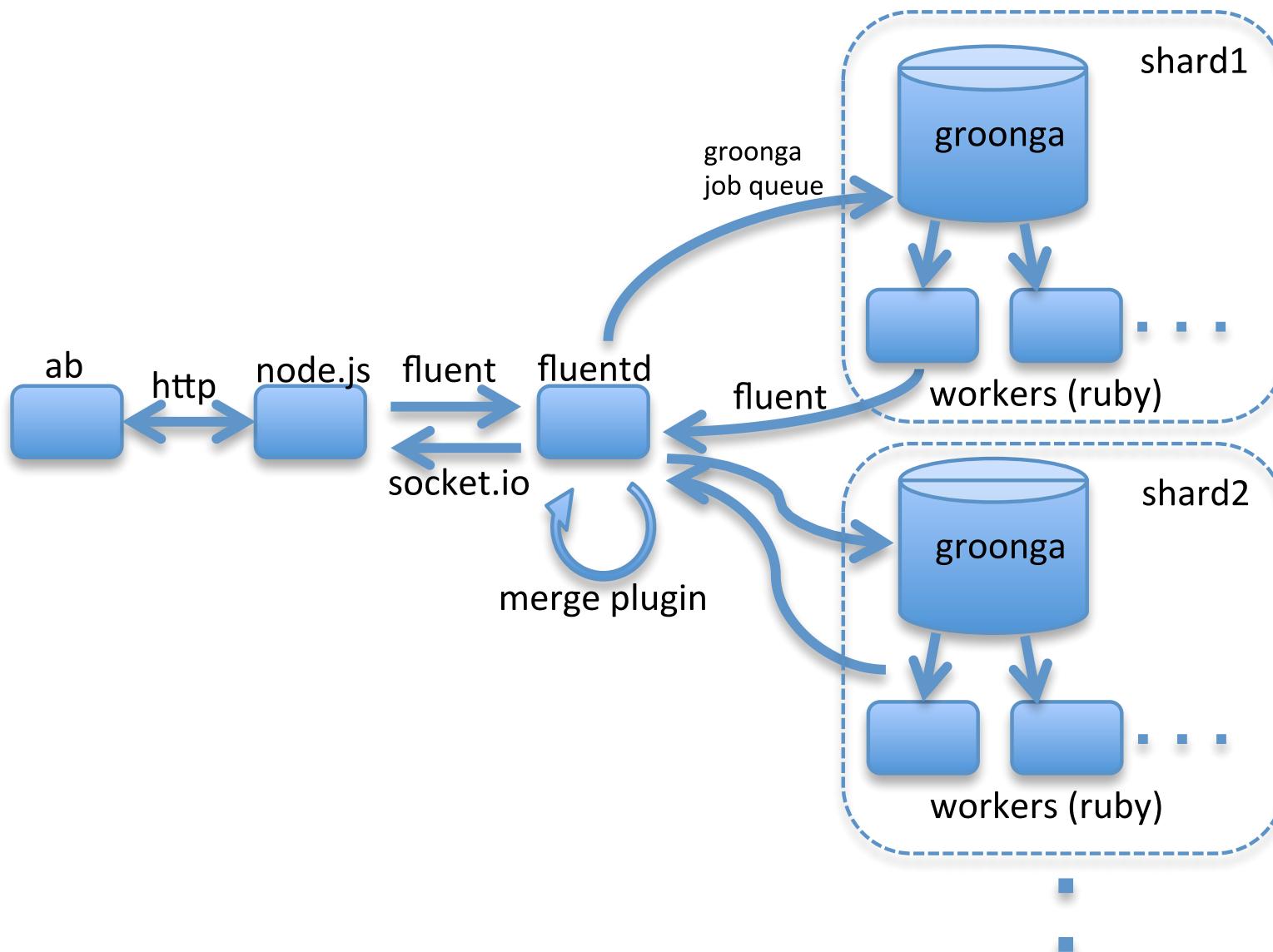
# scale out

- MySQL使いなら迷わずSpider
- MySQLでない人は..

# fluentdでsharding

- fluentdでreplicationだけでなく shardingもできないか?
- 複数へのクエリの配信と結果のマージを fluentdでできないか?
- 絶賛実験中

# 実験システムの構成



# plugin out-merge

```
<source>
  type forward
  port 24224
</source>
<match groonga.**>
  type merge
  <store>
    type groonga_stream
    database MULTIDB/wiki1
  </store>
  <store>
    type groonga_stream
    database MULTIDB/wiki2
  </store>
  <store>
    type groonga_stream
    database MULTIDB/wiki3
  </store>
  <store>
    type groonga_stream
    database MULTIDB/wiki4
  </store>
  <emit>
    type socket_io
    dest http://localhost:3000
  </emit>
</match>
```

# 実験環境

検索対象: Wikipedia(en) 4343516件(15GB)

クエリ: "test" (クエリキャッシュ無効) (95972件hit)

出力: 上位100件のスコアと主キー

サーバ(1台): CPU: Xeon X5550 2.67GHz \* 2

クライアント(1台): ab -c 100 -n 10000

# 実験結果

groonga server(单一DB): 184qps  
fluent + node (4shard): 308qps

- ・まだまだ検証が必要ですが、
- ・素のgroongaより1.6倍以上高速!
- ・近々リリースを目指します
- ・とりあえずfluentd優秀

ご→**静**聴ありが

とうございま

した