

分散処理アプリ演習 第14回

HBaseスキーマ設計 (演習回答解説)

(株)NTTデータ



回答例解説

■ TwitterImport.java

Configurationインスタンス作成

// HTable**インスタンス作成**

```
Configuration conf = HBaseConfiguration.create();  
HTable TwitterTable = new HTable(conf, "TwitterTable");
```

confとテーブル名を指定して、HTableインスタンス作成



回答例解説

■ TwitterImport.java

ツイートID、userIDをツイートログから取得

// 行キー設定

```
Number tweetId = (Number) jsonObj.get("id");
Number userId = (Number) ((Map) (jsonObj.get("user"))).get("id");
long tweetTime = (long) (df.parse(
    (String) jsonObj.get("created_at")).getTime());

byte[] rowkey = Bytes.toBytes(userId.toString() + "-" +
    tweetId.toString());
```

ツイートID、userIDを“-”でつなぎ、行キーを作る。



回答例解説

■ TwitterImport.java

// Put **インスタンス作成**

```
long tweetTime = (long) (df.parse(  
    (String) jsonObj.get("created_at")).getTime());
```

```
Put put = new Put(rowkey, tweetTime);
```

ツイート時間(ミリ秒)をツイートログから
取得

行キーとタイムスタンプを指定して、Putインスタンス作成



回答例解説

■ TwitterImport.java

```
// Put インスタンスにツイート本文とHashTagを追加
```

```
// tweet
```

```
String text = (String) jsonObj.get("text");
```

```
put.add(Bytes.toBytes("tweet"),  
        Bytes.toBytes("text"),  
        Bytes.toBytes(text));
```

tweet:text列にツイート本文を追加

```
// hashtag
```

```
Map entities = (Map) jsonObj.get("entities");
```

```
List hashtags = (List) entities.get("hashtags");
```

```
int hashCount = 1;
```

```
if (hashtags != null) {
```

```
    for (Object hashtag : hashtags) {
```

```
        put.add(Bytes.toBytes("tweet"),
```

```
                Bytes.toBytes("hashtag"+hashCount),
```

```
                Bytes.toBytes(((String) ((Map) hashtag).get("text")).toString()));
```

```
        hashCount++;
```

```
    }
```

```
}
```

tweet:hashtagN列にHashTagを追加

(HashTagが複数の場合、

修飾子をそれぞれhashtag1、hashtag2・・・とする)



回答例解説

■ TwitterImport.java

```
// TwetterTableにデータを追加  
TwitterTable.put(put);
```



回答例解説

■ 実行

```
$ HADOOP_CLASSPATH="/root/hadoop_exercise/14/lib/json-simple-1.1.jar" ¥  
hadoop jar ~/workspace/Class14/target/hbaseImport-0.1.jar ¥  
com.example.dpap.class14.TwitterImport ¥  
~/hadoop_exercise/14/data/tweets.log
```

■ 実行結果確認

```
$ hadoop jar ~/workspace/Class14/target/hbaseImport-0.1.jar ¥  
com.example.dpap.class14.TwitterTableExport 2 2> /dev/null  
100021-10009004      tweet:hashtag1      eco  
100021-10009004      tweet:text           もうすぐ10時 #eco  
100158-10013707      tweet:hashtag1      idol  
100158-10013707      tweet:text           まあまあです #idol
```