

分散処理アプリ演習 第15回

HBase演習

(演習回答解説)

(株)NTTデータ



回答例解説

■ HashTagCountJob.java

TableMapReduceUtilを利用し、
TableInputFormatを設定

```
TableMapReduceUtil.initTableMapperJob(  
    "TwitterTable",                // InputTableName  
    s,                             // Scan  
    HashTagCountMapper.class,      // Mapper  
    Text.class,                   // MapOutput (Key)  
    IntWritable.class,            // MapOutput (Value)  
    this);  
  
TableMapReduceUtil.initTableReducerJob(  
    "HashTagCountTable",           // OutputTableName  
    HashTagCountReducer.class,     // Reducer  
    this);
```

TableMapReduceUtilを利用し、
TableOutputFormatを設定



回答例解説

■ HashTagCountMapper.java

```
private static final IntWritable one = new IntWritable(1);
private Text hashtag = new Text();

protected void map(ImmutableBytesWritable key, Result value, Context context)
                    throws IOException, InterruptedException {

    for (KeyValue kv : value.raw()) {

        byte[] qualifier = kv.getQualifier();

        if (!Bytes.toString(qualifier).equals("text")) {
            hashtag.set(kv.getValue());
            context.write(hashtag, one);
        }
    }
}
```

Resultクラスのraw()により、
KeyValueクラスを取得

取得した行の修飾子を格納

修飾子がtextでない(hashtag1等)の場合
hashtag名をキーにして、出力



回答例解説

■ HashTagCountReducer.java

hashtagの出現回数をカウント

```
protected void reduce(Text key, Iterable<IntWritable> values, Context context)
    throws IOException, InterruptedException {

    int sum = 0;
    for (IntWritable value : values) {
        sum += value.get();
    }
    Put p = new Put(Bytes.toBytes(key.toString()));
    p.add(Bytes.toBytes("count"),
        HConstants.EMPTY_BYTE_ARRAY,
        Bytes.toBytes(sum));
    context.write(NullWritable.get(), p);
}
```

行キーをhashtag名にする

"count:"列に回数を格納

HBaseのテーブルに出力



回答例解説

■ 実行

```
$ hadoop jar ~/workspace/Class15/target/hbaseCount-0.1.jar ¥  
com.example.dpap.class15.HashTagCountJob
```

■ 実行結果確認

```
$ hadoop jar ~/workspace/Class15/target/hbaseCount-0.1.jar ¥  
com.example.dpap.class15.HashTabCountTableExport 10 2> /dev/null ¥  
| sort -t "    " -k3 -rn  
eco count: 495  
lrnchat count: 381  
climate count: 369  
earthtweet count: 362  
jobs count: 361  
business count: 358  
humantrafficking count: 341  
retail count: 319  
teachers count: 308  
networking count: 307  
cause count: 306  
health count: 296
```