Binning

Description du pattern

Le pattern binning, comme le pattern précédent, déplace les données dans différentes catégories, indépendamment de l'ordre de collecte.

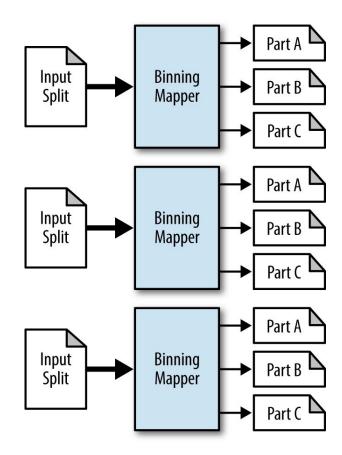
Intention

Trier chaque enregistrement de l'ensemble de données dans une ou plusieurs catégories.

Motivation et structure

- Le pattern est très similaire au pattern Partitioner
- Les données sont triées dans des fichiers de sortie (appelés "bins") pendant la phase de Mapping
- Un fichier de sortie par catégorie et par Mapper

- Utilisation de la classe MultipleOutputs
- Pas de Combiner, de Partitioner ou de Reducer



Exemple

```
public static class BinningExampleMapper
extends Mapper<LongWritable, Text, Text, NullWritable>{
   private MultipleOutputs<Text, NullWritable> mos = null;
   protected void setup(Context context) {
       mos = new MultipleOutputs(context);
   public void map(LongWritable key, Text value, Context context
           ) throws IOException, InterruptedException {
       String line = value.toString();
       double cancer latitude = 23.43671;
       double capricorne latitude = -23.43671;
       String[] attributes = line.split(",");
       if(null != attributes && attributes.length > 6) {
           double latitude = Double.parseDouble(attributes[5]);
           if(latitude > cancer latitude)
               mos.write("bins", value, NullWritable.get(), "nord");
           if(latitude <= cancer latitude && latitude >= capricorne latitude)
               mos.write("bins", value, NullWritable.get(), "tropiques");
           if(latitude < capricorne latitude)</pre>
               mos.write("bins", value, NullWritable.get(), "sud");
   protected void cleanup(Context context) throws IOException,
   InterruptedException {
       mos.close();
```

```
Found 9 items
                                              0 2019-11-12 10:55 outputsdemo/ SUCCESS
             3 ccaulonque supergroup
-rw-r--r-
-rw-r--r--
             3 ccaulonque supergroup
                                       84410587 2019-11-12 10:55 outputsdemo/nord-m-00000
             3 ccaulonque supergroup
                                       10637944 2019-11-12 10:55 outputsdemo/nord-m-00001
-rw-r--r--
             3 ccaulonque supergroup
                                              0 2019-11-12 10:55 outputsdemo/part-m-00000
-rw-r--r--
             3 ccaulonque supergroup
                                              0 2019-11-12 10:55 outputsdemo/part-m-00001
-rw-r--r--
             3 ccaulonque supergroup
                                        2741125 2019-11-12 10:55 outputsdemo/sud-m-00000
-rw-r--r--
             3 ccaulonque supergroup
                                         343689 2019-11-12 10:55 outputsdemo/sud-m-00001
-rw-r--r--
                                       47066058 2019-11-12 10:55 outputsdemo/tropiques-m-00000
             3 ccaulonque supergroup
-rw-r--r--
             3 ccaulonque supergroup
                                        5945858 2019-11-12 10:55 outputsdemo/tropiques-m-00001
-rw-r--r--
```

Fichiers de sorties ("bins")

cm, ngaundere, Ngaundere, 10,,7.3166667,13.5833333 cd,mabilatuku,Mabilatuku,10,,-4.283333,25.6 id,selaawi satu,Selaawi Satu,30,,-6.555833,107.474167 ph,marawi city,Marawi City,E1,,7.998611,124.292778 mx,el arenal,El Arenal,05,,15.083333,-92.483333 th,ban thung charong,Ban Thung Charong,49,,12.45825,102.506972 br,alto alegre,Alto Alegre,22,,-5.8225,-37.410556 sn,ngoultoune,Ngoultoune,07,,14.4833333,-16.75 pe,yuripampa,Yuripampa,06,,-7.5508333,-77.9130556 co,atunsuela,Atunsuela,29,,3.737778,-76.676389 hn,balibrea,Balibrea,04,,14.4833333,-87.8833333 cu, minas, Minas, 09,,20.1744444,-76.7572222 th,ban na waeng,Ban Na Waeng,81,,18.264258,103.550851 gh, kasuri, Kasuri, 06,,9.5166667,-1.1 br,ricardo,Ricardo,15,,-15.766667,-43.866667 sl,serekolia,Serekolia,02,,9.5333333,-11.0833333 ht,pilaurant,Pilaurant,12,,18.3166667,-73.1166667 co,la pola,La Pola,38,,10.016667,-74.516667 id,qubuk daya,Gubuk Daya,17,,-8.3808,116.6898 ng, jerdeia, Jerdeia, 24,, 13.06923, 8.390829 ph,san vicente,San Vicente,N4,,13.3564,123.6994

Morceau du fichier tropiques-m-00001

Conséquences

- Pattern performant car seul le Mapper est utilisé
- Besoin de post-processing pour utiliser les données
- Sélection possible des données classifiées pour le post-processing