Pig

1. Install Pig

- $ sudo yum install -y pig

2. Run grunt shell

- $ pig
  grunt>

3. Load data from musicbrainz paths

```
grunt> tracks = LOAD 'musicbrainz_tracks' AS (trackid:int, releaseid:int,
trackname:chararray, position:int);

grunt> releases = LOAD '/user/cloudera/musicbrainz_releases' AS (id:int,
album:chararray, artist:chararray);
```

4. Join bags releases and tracks

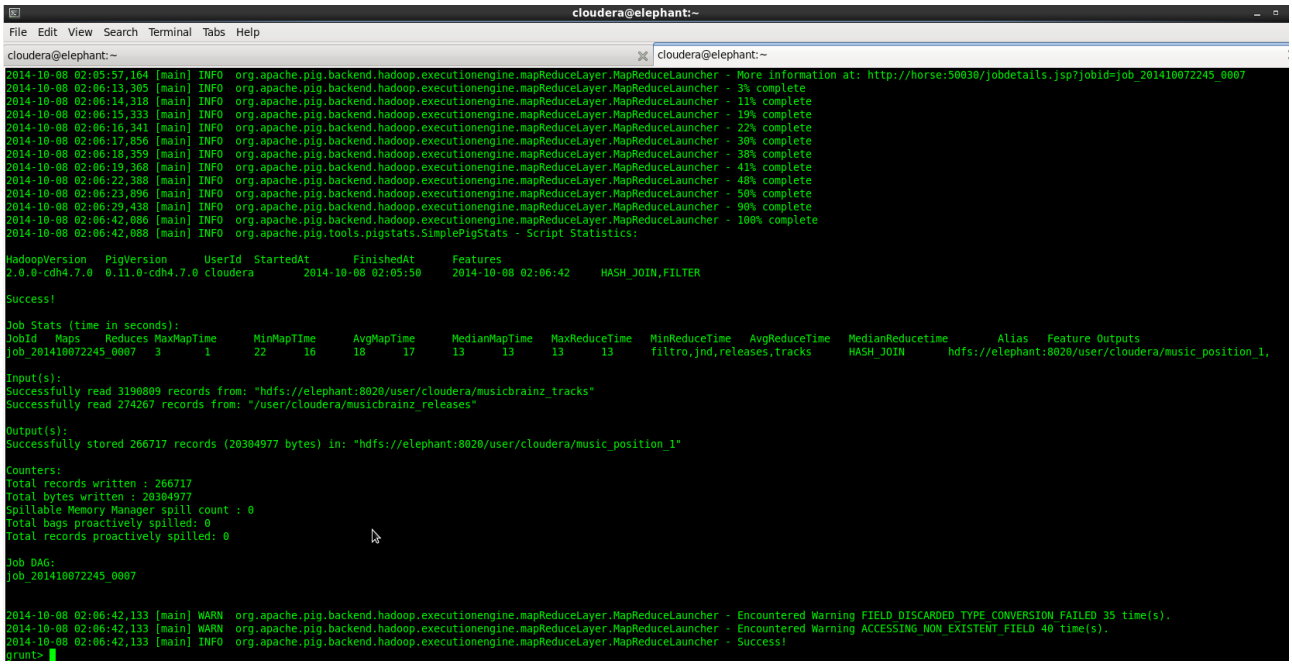
- grunt> jnd = JOIN tracks BY releaseid, releases BY id;

5. Filter by position 1 on list

- grunt> filtro = FILTER jnd BY position == 1;

6. Store result *filtro* on HDFS

- grunt> STORE filtro INTO 'music_position_1';



7. Check results in file

- grunt> quit
- $ hadoop fs -tail music_position_1/part-r-00000