生成比较结果

#df1.withColumnRenamed("rank", "rank\_topn").crossJoin(df1).show()  
#包含自身的排名前后2位  
df1.withColumnRenamed("rank", "rank\_topn").crossJoin(df1).filter("(rank>=rank\_topn and rank<=rank\_topn+2) or (rank>=rank\_topn-2 and rank<rank\_topn)").show()  
#不包含自身的排名前后2位  
df1.withColumnRenamed("rank", "rank\_topn").crossJoin(df1).filter("(rank>rank\_topn and rank<=rank\_topn+2) or (rank>=rank\_topn-2 and rank<rank\_topn)").show()

创建hive表

PARTITIONED BY (   
  `dt` string,   
  `item\_third\_cate\_cd` string)  
ROW FORMAT DELIMITED   
  FIELDS TERMINATED BY '\t'  
STORED AS INPUTFORMAT   
  'org.apache.hadoop.hive.ql.io.orc.OrcInputFormat'   
OUTPUTFORMAT   
  'org.apache.hadoop.hive.ql.io.orc.OrcOutputFormat'

PARTITIONED BY (

`dt` string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

stored as orc;