

# 跨版本distcp报Check-sum错误

报错信息:

```
Caused by: java.io.IOException: Check-sum mismatch between
hdfs://192.168.81.30:8020/user/hive/warehouse/xy_ods.db/ods_verification_cardno_d_incr/pk_year=2017/pk_mont
07/pk_day=2017-07-23/000011_0 and
hdfs://172.20.85.39:8020/user/hive/warehouse/xy_ods.db/ods_verification_cardno_d_incr/pk_year=2017/.distcp.
Source and target differ in block-size. Use -pb to preserve block-sizes during copy. Alternatively, skip
checksum-checks altogether, using -skipCrc. (NOTE: By skipping checksums, one runs the risk of masking
data-corruption during file-transfer.)
    at
org.apache.hadoop.tools.mapred.RetriableFileCopyCommand.compareChecksums(RetriableFileCopyCommand.java:221)
    at
org.apache.hadoop.tools.mapred.RetriableFileCopyCommand.doCopy(RetriableFileCopyCommand.java:134)
```

解决方法:

迁移到的hadoop集群版本比较高, 最好设置-skipcrccheck选项也-update选项, skipcrccheck忽略FileChecksum校验, 因为版本的升级可能带来Checksum值不一样, cdh6与cdh5就是这样.

命令为:

```
hadoop distcp -D mapreduce.job.queueName=xy_yarn_pool.development -D ipc.client.fallback-to-simple-
auth-allowed=true -update -skipcrccheck
hdfs://192.168.81.30:8020/user/hive/warehouse/xy_ods.db/ods_verification_cardno_d_incr/pk_year=2017/pk_r
07/pk_day=2017-07-23/000011_0
hdfs://172.20.85.39:8020/user/hive/warehouse/xy_ods.db/ods_verification_cardno_d_incr/pk_year=2017/pk_m
07/pk_day=2017-07-23/000011_0
```