

1.为什么用cdh?

因为集成了大数据各种服务, 减少运维成本!

2.如何搭建cloudera manager

请参考cloudera官网, 下载指定版本bin文件, 于服务器上执行!

3.cloudera manager ip: 192.167.78.133:7180 admin:admin

4.服务器信息192.167.78.133/192.167.78.134/192.167.78.137 root:hadoop

5.hadoop 命令

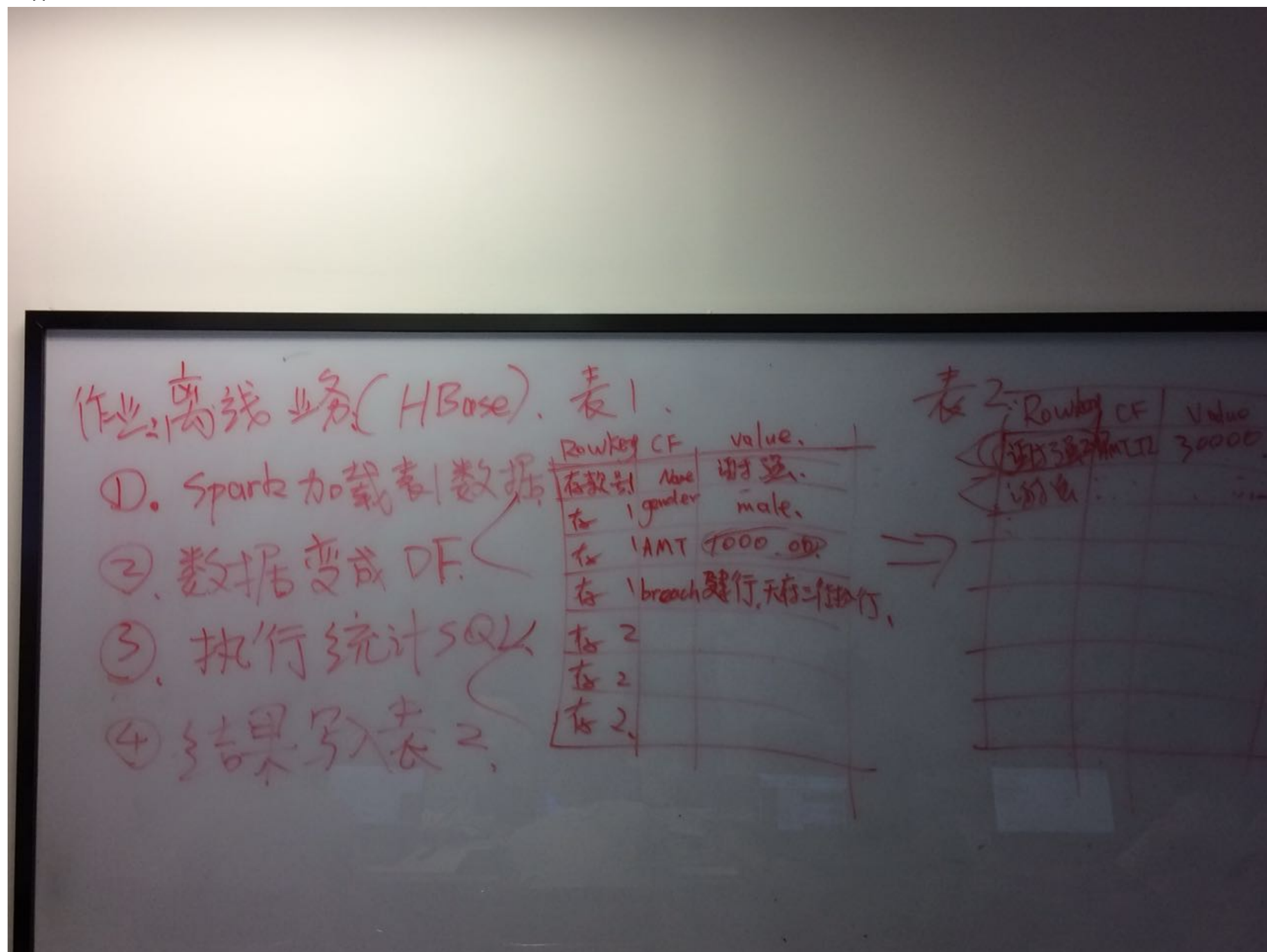
<http://blog.csdn.net/wuwenxiang91322/article/details/22166423>

6.hbase命令

hbase shell

7.数据采集: flume

8.作业:



<http://www.cnblogs.com/cssdongl/p/6238007.html>

采用第二种方式 用spark加载和写入hbase

进阶要求: 配置放在外面加载(实现更改业务逻辑只需要修改配置文件)