1、linux常用命令

2、git命令：就说用idea集成的，add→commit→push等

3、spark分析单词的topN和词频

4、java或者scala写UDF函数对身份证前两位数据进行解析匹配地址

package com.luomk.hive;

import org.apache.hadoop.hive.ql.exec.UDF;

public class Lower extends UDF {

public String evaluate (final String s) {

if (s == null) {

return null;

}

return s.toLowerCase();

}

}

5、ods,dwd,dws,dm的概念

6、hive数据倾斜的概念，解决方案？

7、写了个游戏：缝7和3的倍数或包含3或7就喊“过”，用代码实现

8、最近在看什么书？有什么收获？

9、spark中task的划分，什么决定task的数量？Stage的划分？