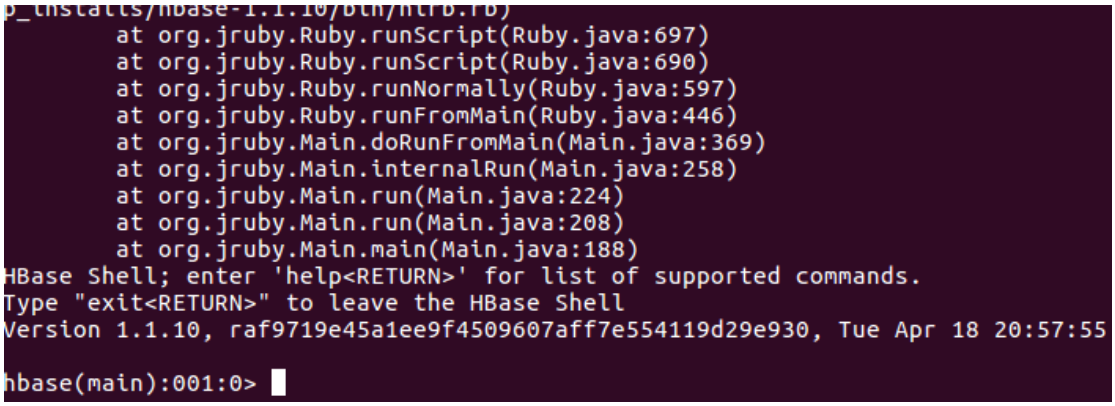
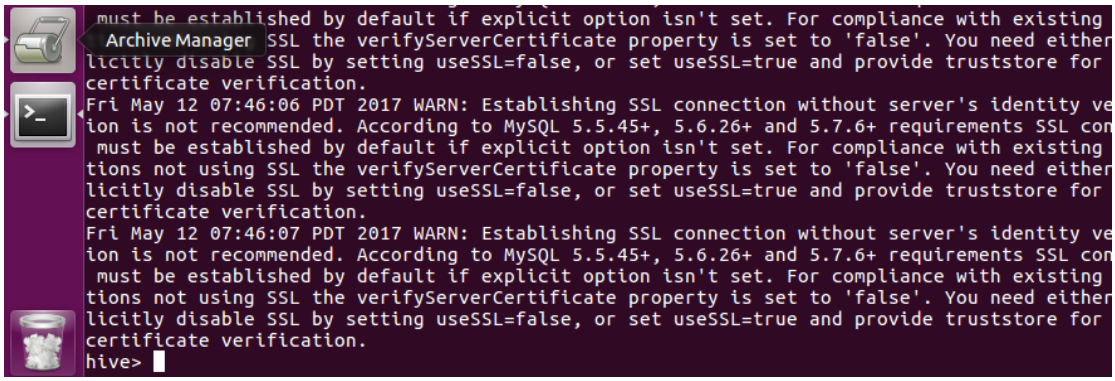
实验四 Hbase，Hive 安装与使用

141070010，141180016，141290008

1、 在自己电脑正确安装和运行 HBase 和 Hive

下载安装包后，依照网上教程修改配置文件即可





2、 在 Hbase 中创建表 Wuxia：

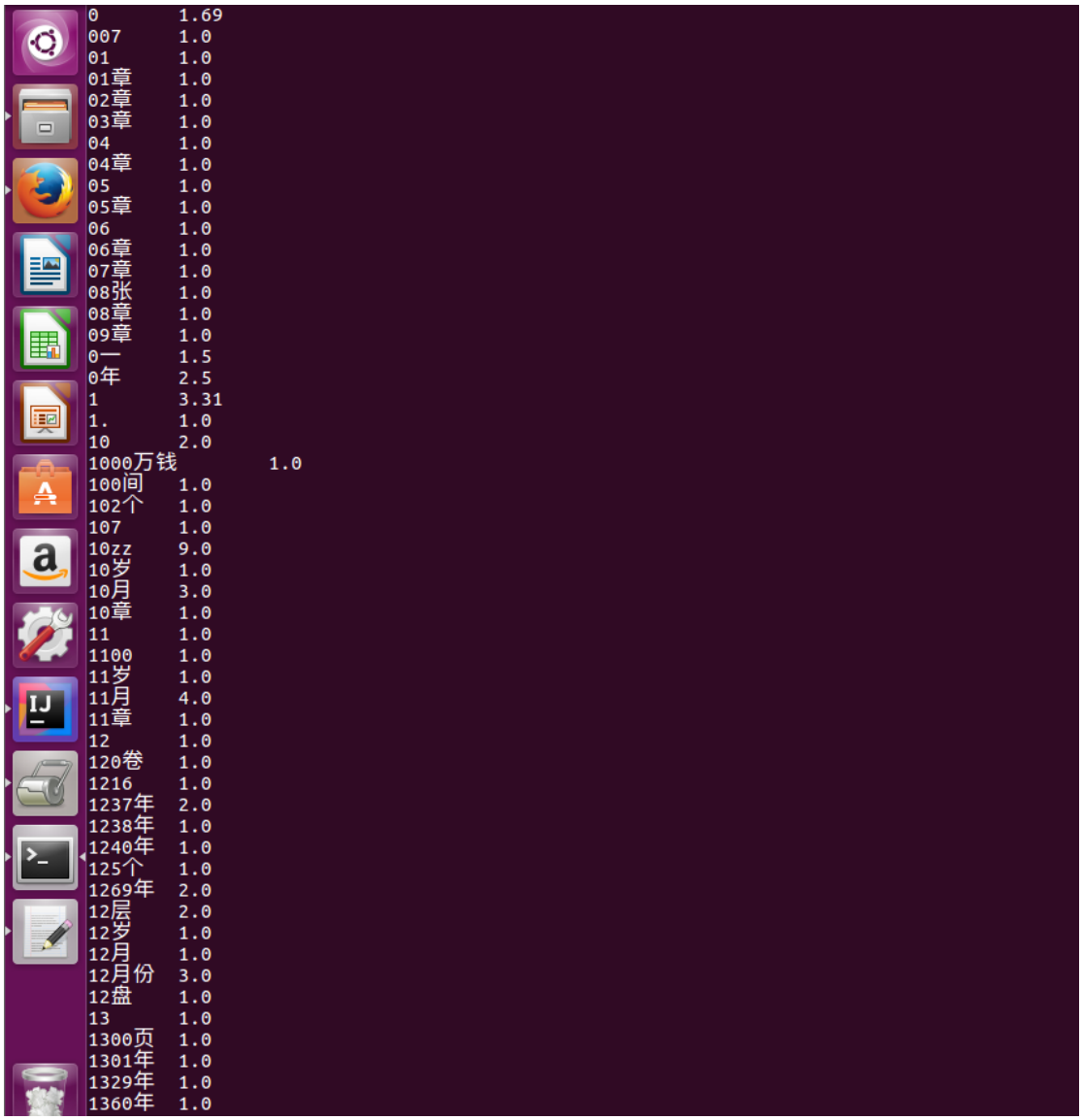
Create ‘Wuxia’,’datas’

3、 将 mapreduce 结果写入 Hbase 的表 Wuxia 中

```
if(!preword.equals(word) && !preword.equals(new Text(" "))){

                StringBuilder out = new StringBuilder();
                long size = postingList.size();
                double average = wordsum / (double) size;

                HTable table =new HTable(HBASE_CONFIG, tableName);
                 //以当前单词作为 Row_keys
                Put put = new Put(preword.getBytes());
                 //参数分别对应列族，修饰符，列值
                put.add(Bytes.toBytes("datas"), Bytes.toBytes("average"),
                        Bytes.toBytes(String.format("%.2f", average)));
                table.put(put);
            table.close();
```

4、 遍历 Hbase 中的表

```
        String tableName = "Wuxia";
        Configuration HBASE_CONFIG = HBaseConfiguration.create();

        Scan scan = new Scan();
        HTable table =new HTable(HBASE_CONFIG, tableName);
        ResultScanner resultScanner = table.getScanner(scan);

        for(Result r:resultScanner) {
            //由上知每行的行键就是单词
            String word = Bytes.toString(r.getRow());
            //获取出现频率
            byte[] valueBytes = r.getValue(Bytes.toBytes("datas"),
                    Bytes.toBytes("average"));
            String value = Bytes.toString(valueBytes);
            String fileName = "Wuxia.txt";
            FileWriter writer = new FileWriter(fileName, true);
            writer.write(word + "\t" + value+"\n");
            writer.close();
        }
    }
```



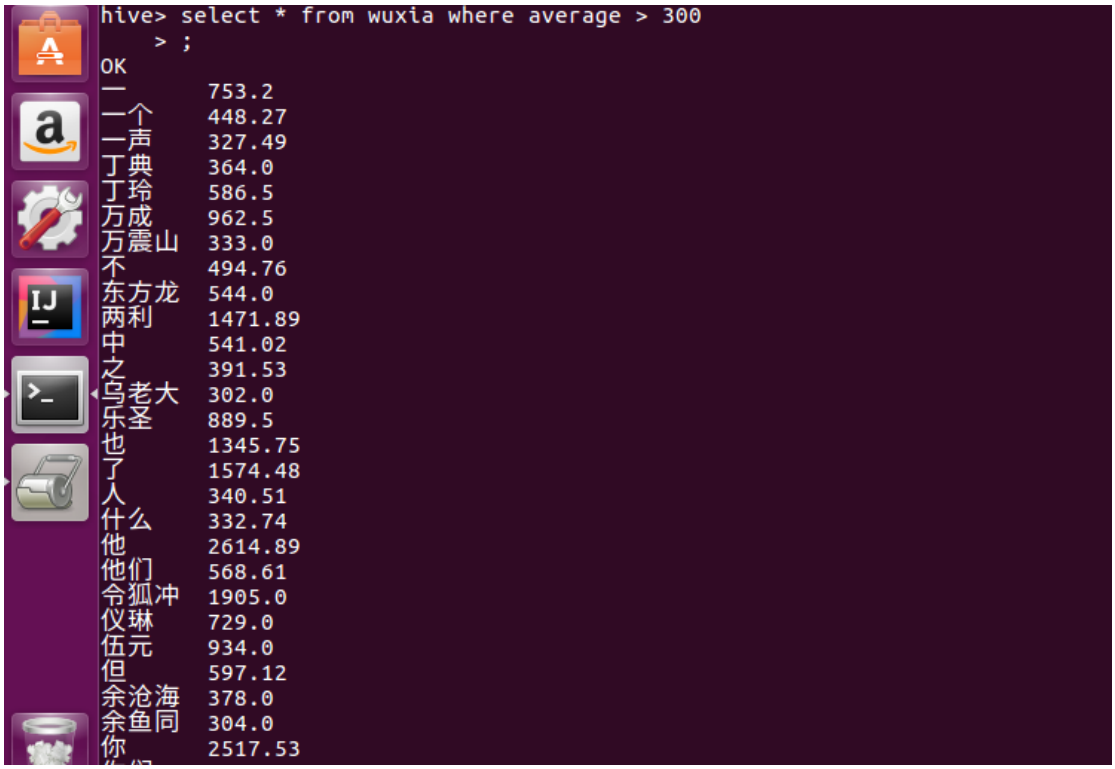
5、

（1）创建表并导入数据：

```
hive> create table Wuxia(word string,average double)
    > row format delimited
    > fields terminated by '\t'
    > lines terminated by '\n';
OK
Time taken: 2.272 seconds
hive> load data local inpath '/home/hadoop/Wuxia.txt' into table Wuxia;
Loading data to table default.wuxia
Table default.wuxia stats: [numFiles=1, totalSize=1897164]
OK
Time taken: 1.335 seconds
```

（2）查询出现次数大于 300 的词语：

Select * from Wuxia where average > 300



（3）前 100 个出现次数最多的词：

Select * from Wuxia order by average desc limit 100