《基于mapreduce技术的新冠肺炎数据分析报告》

1. 环境准备

1.dokcer run centos

2.在etc/profile下添加

export PATH=$PATH:/hadoop/hadoop-3.3.0/sbin:/hadoop/jdk1.8.0\_281/bin:hadoop/jdk1.8.0\_281/bin:$PATH

3.在hadoop-3.3.0/etc/hadoop-env.sh修改

export JAVA\_HOME=/hadoop/jdk1.8.0\_281

4.准备hadoop-core-1.1.2.jar到hadoop-3.3.0主目录

4.ssh-genkey -t -p

1. 实现：
2. 创建四个文件，分别是BeanTest.java ,covid.java ,covidredurce.java, covidmap.java
3. BeanTest缓冲数据内容
4. covid主类设定map类和redurce类，并设定文件输入输出地址。
5. covidredurce类实现
6. covidmap类对数据进行分类，把北京的疫情数据提取出来

三．主要代码：

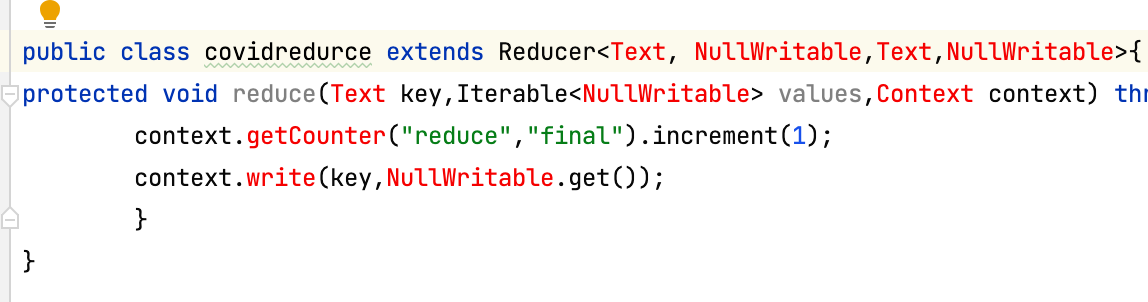
1.map阶段：

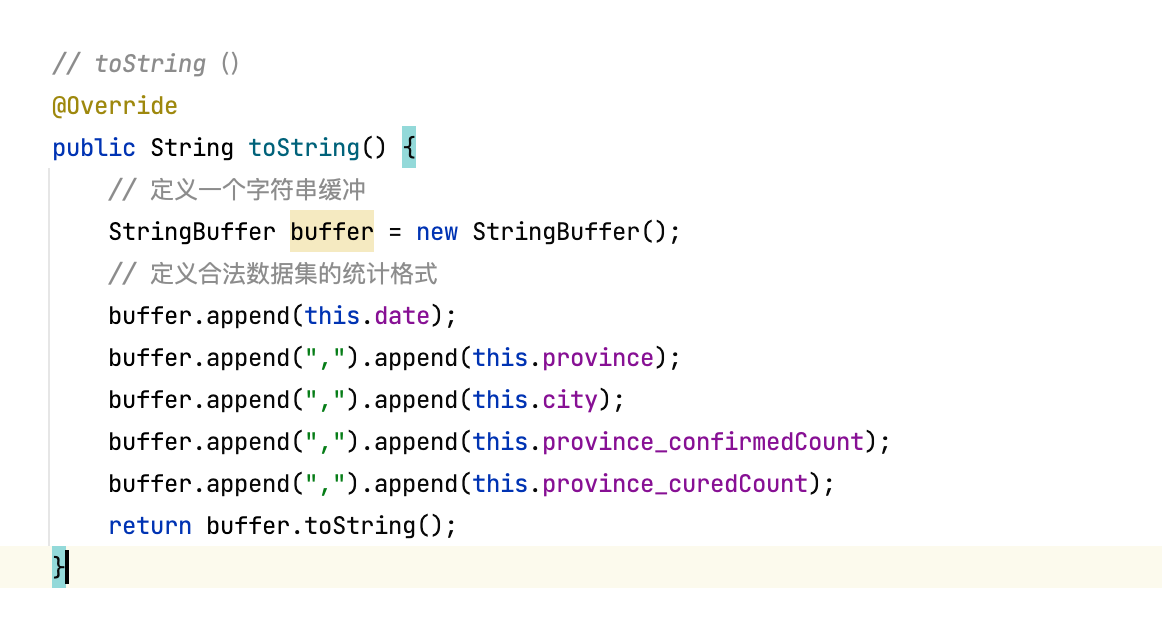
对数据进行提取处理，并把数据redurce进行分类



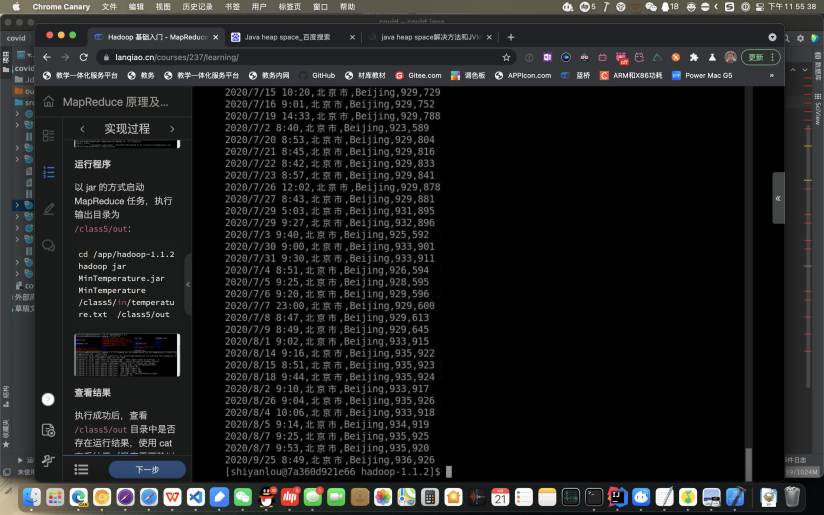
2.redurce阶段：

保存数据

3.Bean（缓冲区存储数据）：



1. 结果：

2.运行截图：