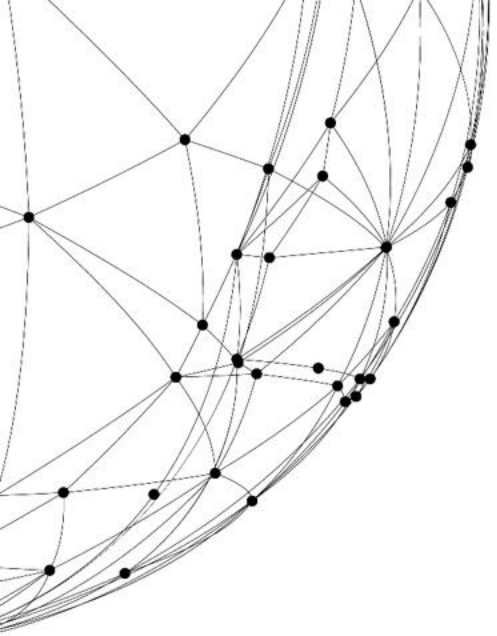
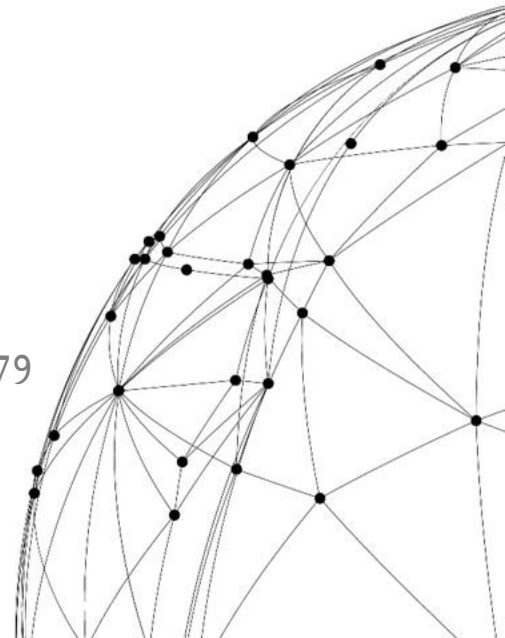


Hadoop大数据处理行为痕迹记录的创新应用

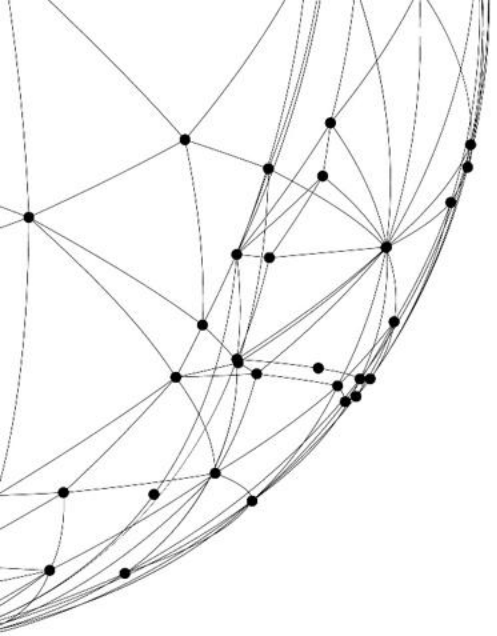


指导老师：张至柔、吴娟

负责人： 杨秉学 13756697779
组员： 刘俊龙 18072264663

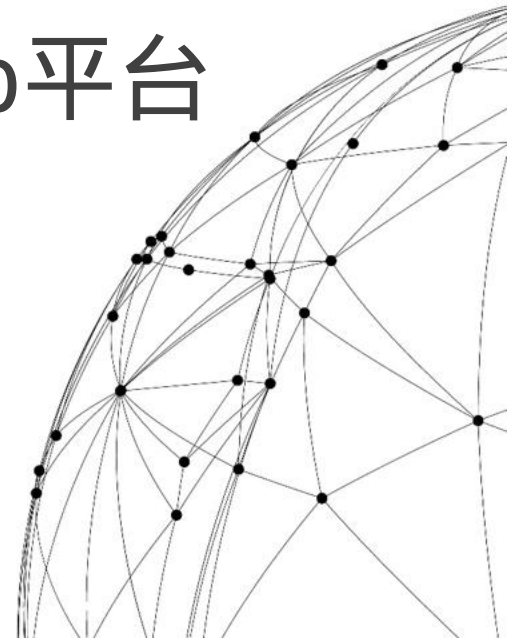
项目成果

1. 在服务器上成功部署了完全分布式Hadoop平台
2. 根据实际需求优化部署
3. 记录分析Hadoop的计算和IO开销状况与文件输出时间。
4. 升级Hadoop计算框架算法



1

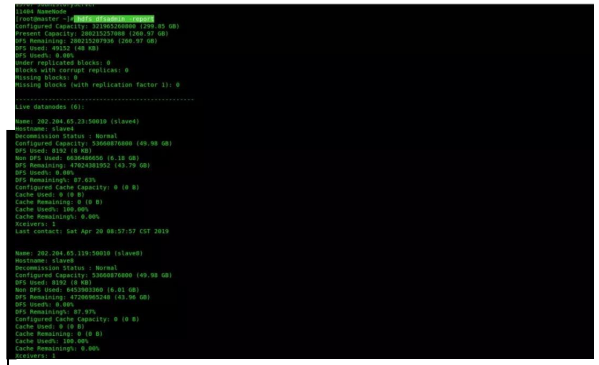
在服务器上成功部署了完全分布式Hadoop平台



```
[root@master ~]# jps
16080 Jps
12099 SecondaryNameNode
14724 ResourceManager
15707 JobHistoryServer
11404 NameNode
```

1

Master节点中NameNode进程检测



3

命令行检测DataNode启动状况

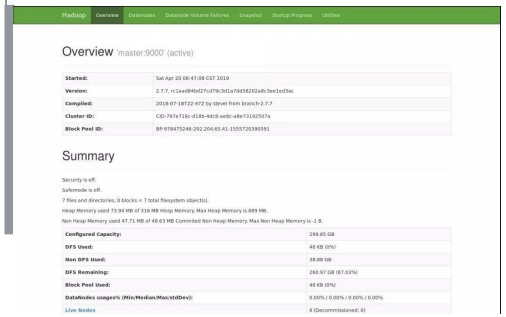
2

Slaver中DataNode进程检测

```
[root@slave8 ~]# jps
13618 Jps
129352 NodeManager
127530 DataNode
[root@slave8 ~]#
```

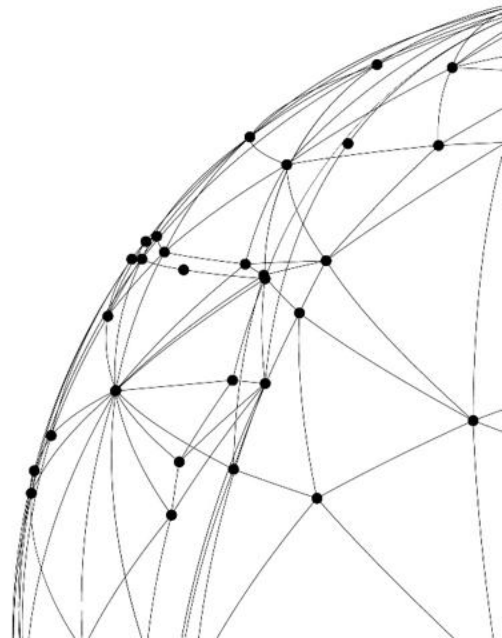
4

Web方式检测DataNode运行状况

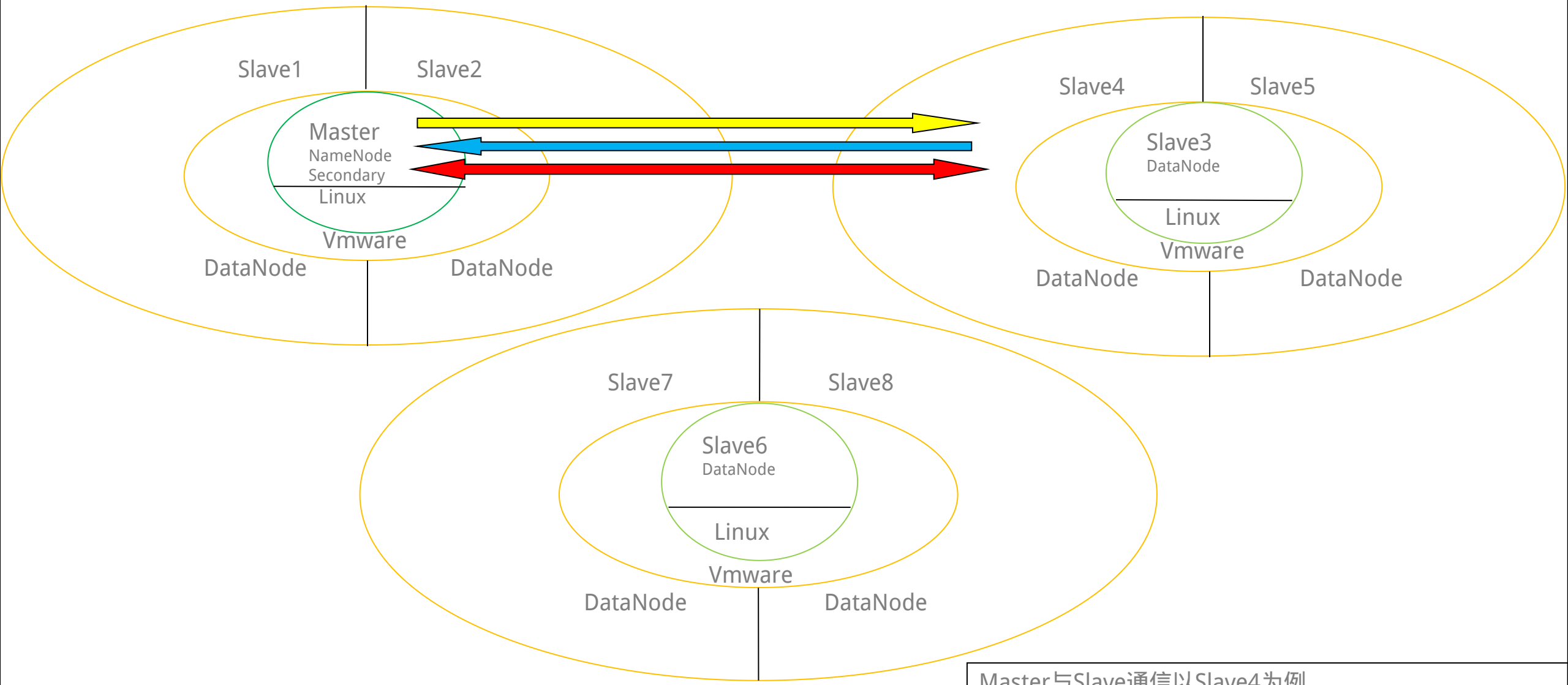


2

根据实际需求优化部署



项目初期Hadoop架构图



Master与Slave通信以Slave4为例

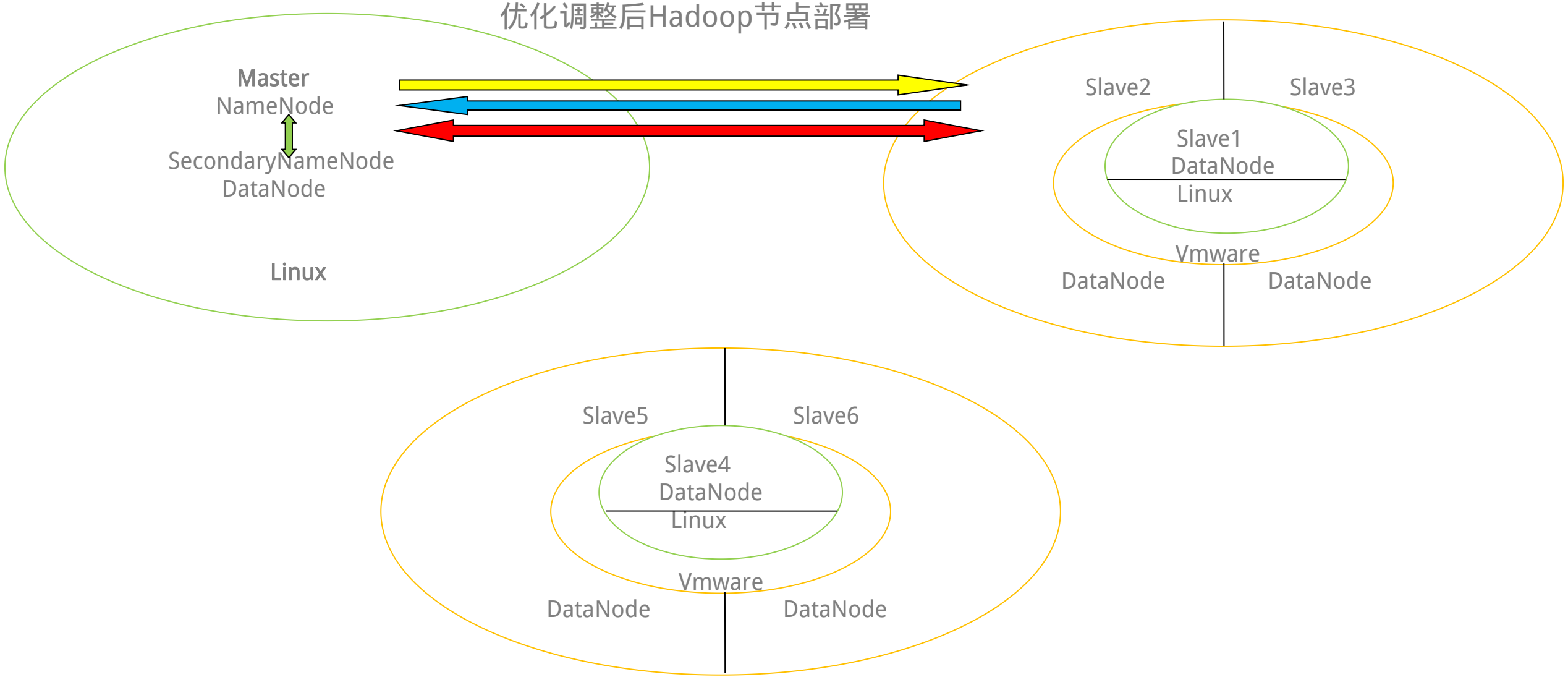
：任务调度

：结果反馈

：心跳机制

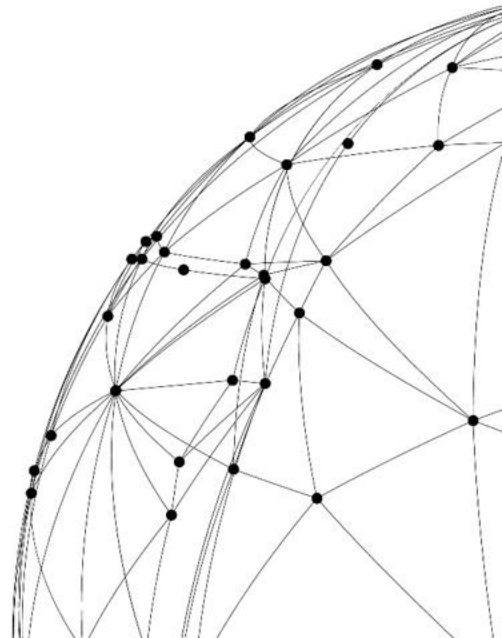
：NameNodeSecondaryNameNode通讯

优化调整后Hadoop节点部署

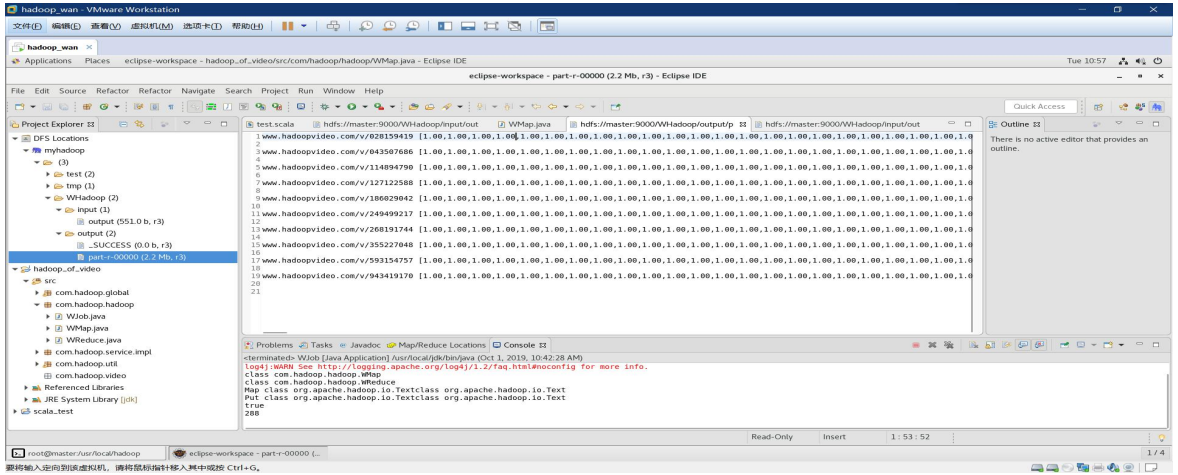


3

记录分析Hadoop的计算和IO开销状况与文件输出时间。

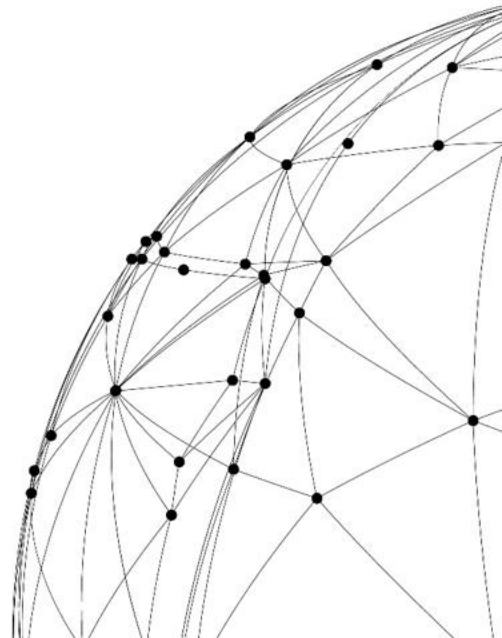


test.scala
hdfs://master:9000/WHadoop/input/out
WMap.java
hdfs://master:9000/WHadoop/output/p
hdfs://master:9000/WHadoop/input/out
1 www.hadoopvideo.com/v/943419170-0-73830-51514-61383-2.5+www.hadoopvideo.com/v/268191744-1-83418-79515-82295-2.0+www.hadoopvideo.com/v/043507686-2-
Problems
Tasks
Javadoc
Map/Reduce Locations
Console
<terminated> WJob [Java Application] /usr/local/jdk/bin/java (Oct 1, 2019, 10:42:28 AM)


hadoop_wan - VMware Workstation
eclipse-workspace - part-r-00000 (2.2 Mb, r3) - Eclipse IDE
Project Explorer
DFS Locations
myhadoop
WHadoop (2)
input (1)
output (2)
part-r-00000 (2.2 Mb, r3)
hadoop_of_video
com.hadoop.hadoop
WJob.java
WMap.java
WReduce.java
Console
<terminated> WJob [Java Application] /usr/local/jdk/bin/java (Oct 1, 2019, 10:42:28 AM)
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
class com.hadoop.hadoop.WMap
class com.hadoop.hadoop.WReduce
Map class org.apache.hadoop.io.Textclass org.apache.hadoop.io.Text
Put class org.apache.hadoop.io.Textclass org.apache.hadoop.io.Text
true
288

4

优化Hadoop计算框架算法



项目目前存在需要解决的主要问题

现在才有的算法是简单的for循环，没有经过优化
时间与空间复杂度比较高

项目主要问题解决方案

算法优化，以满足我们在实际应用场景的需要
对用户行为的分片大小进行优化
未来可以放到学校网络教学平台进行测试，推广使用

下阶段主要计划及时间安排

2020年12月至1月

争取初步完成立项时的要求

2020年2月

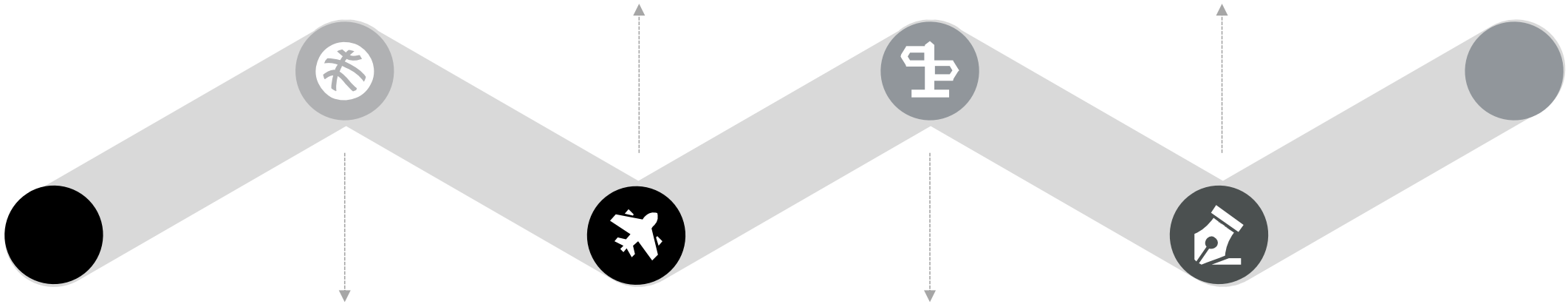
完善以及算法优化

2020年3月

完成技术说明书、撰写论文、申请软件著作权

2020年4月至5月

申请结题



谢谢聆听！