


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

0

1

在服务器上成功部署了完全分布式Hadoop平台

0

2

根据实际需求优化部署

0

3

记录分析Hadoop的计算和IO开销状况与文件输出时间。

0

4

将Hadoop计算框架升级为Spark计算框架



01

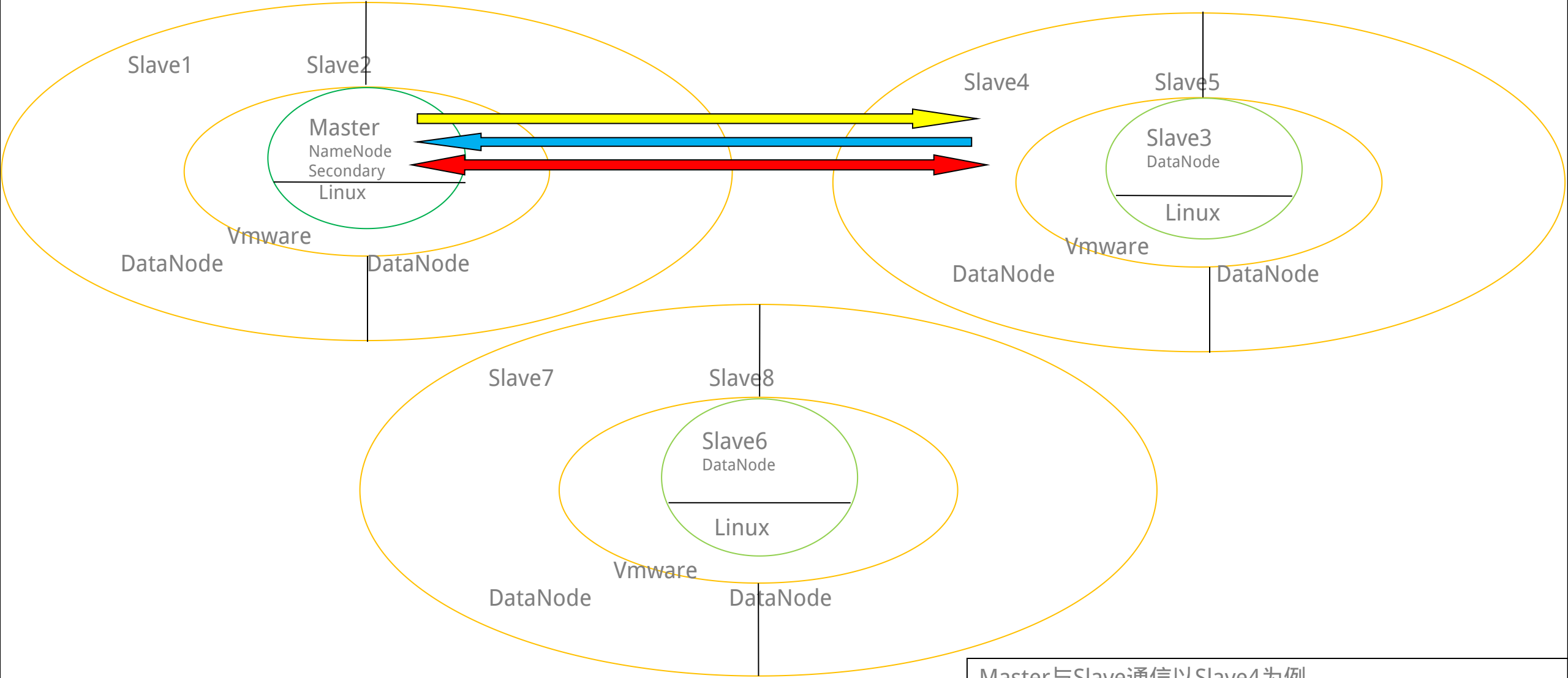
在服务器上成功部署了完全分布式Hadoop平台



02

根据实际需求优化部署

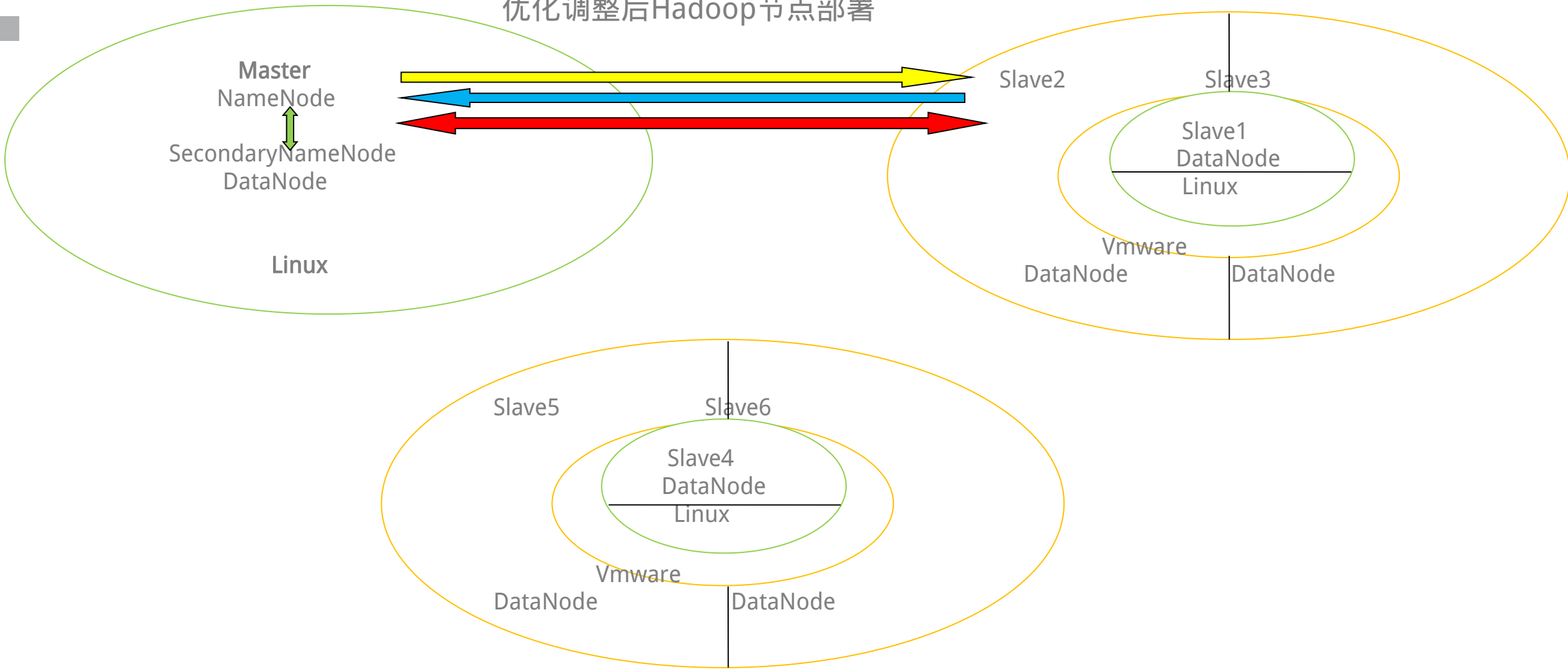
项目初期Hadoop架构图



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



Master与Slave通信以Slave2为例

- 任务调度
- 结果反馈
- 心跳机制
- NameNode与SecondaryNameNode通讯



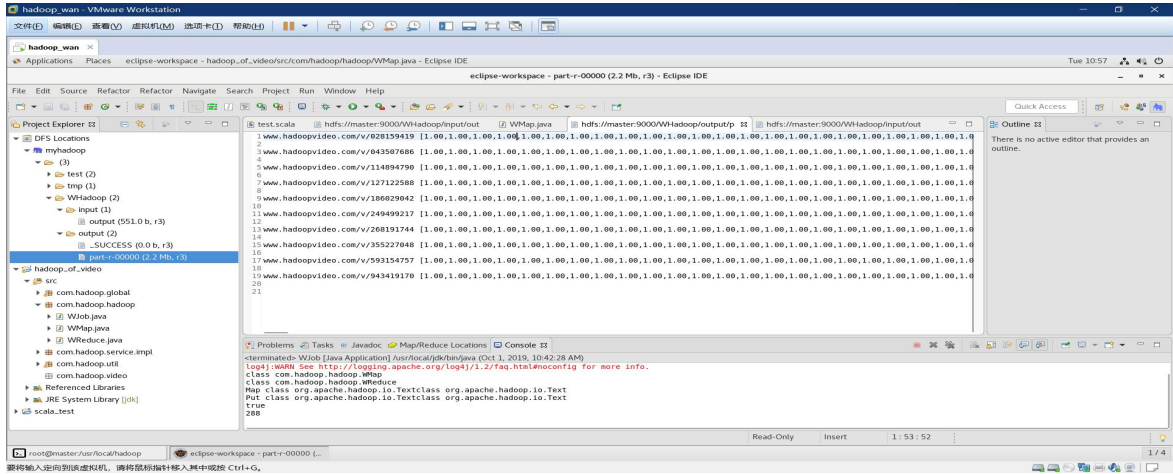
03

记录分析Hadoop的计算和IO开销状况与文件输出时间。





前端所传用户行为痕迹记录数据以及并行式处理分析用户行为痕迹记录获取视频播放总体情况数据结果



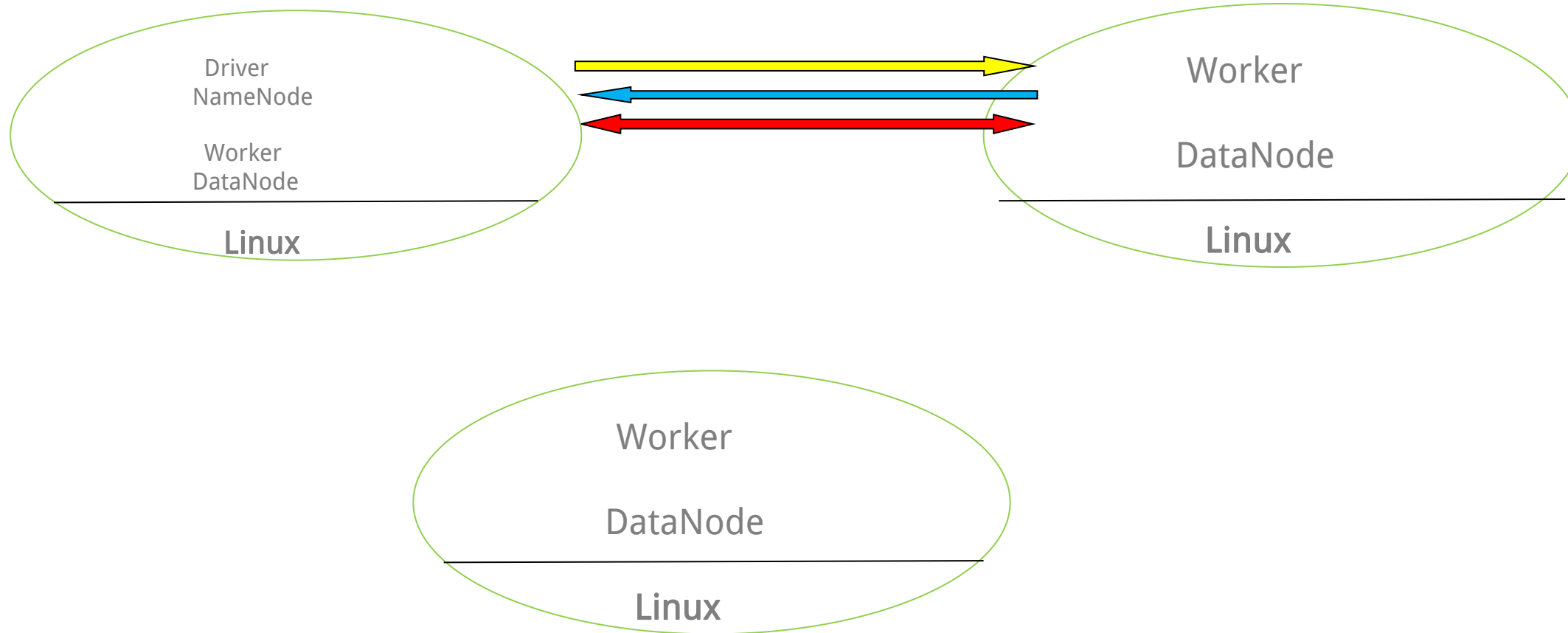


04

将Hadoop计算框架升级为Spark计算框架



项目Spark架构



Master与Slave通信以Slave2为例

- 任务调度
- 结果反馈
- 心跳机制

项目目前存在需要解决的主要问题

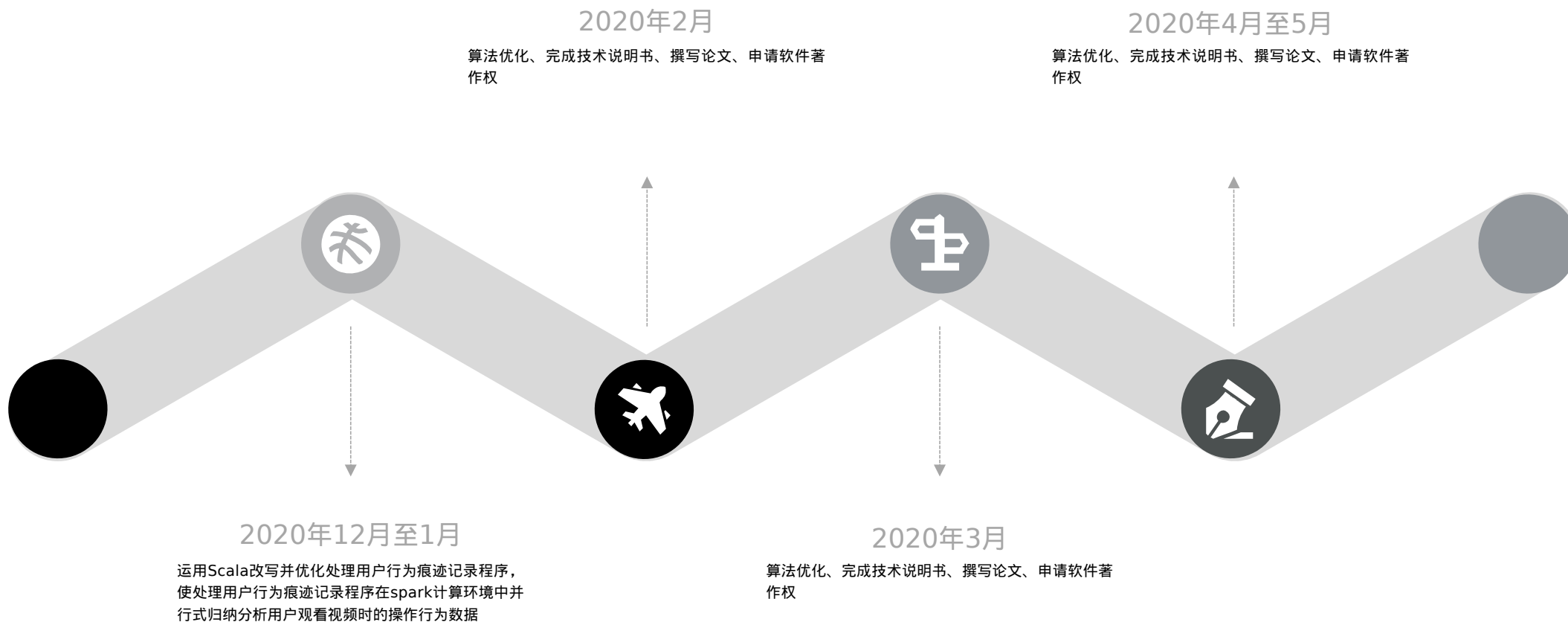
运用Java开发的hadoop程序
迁移到spark上比较繁琐

项目主要问题解决方案

改用Scala语言开发
并行spark程序进行分析



下阶段主要计划及时间安排





谢谢聆听！