二、填空题（10 分）

1、YARN 的架构组成：Container(容器) 、ResourceManager(资源管理器) 、

NodeManager(节点管理器) 、ApplicationMaster (主应用) 。

2、MapReduce 的思想，任务组成

两大核心思想：Map(映射)、Reduce(化简)（分而治之，先分后合）

任务组成：Map阶段(分)：负责对数据进行预处理

Reduce阶段(合)：负责对数据进行聚合处理

3、大数据的核心问题：数据存储(HDFS) 、数据运算(MapReduce) 。

4、HDFS 的 edits (日志)文件和 fsimage (镜像)文件的区别

fsimage 是保存最新的元数据的信息，当 fsimage 数据到一定的大小时会去生成一个新的文件来保存元数据的信息，这个新的文件就是 edit，edit 会回滚最新的数据。

5、HDFS 的架构组成和作用

Client(客户端)：负责切分文件

Namenode(名称节点)：负责管理HDFS的元数据，处理Client请求

DataNode(数据节点)：处理Client发来的数据块并执行读写操作

SecondaryNameNode(辅助名称节点)：镜像备份(备份fsimage文件)

6、Hadoop 生态圈

三大核心组件：HDFS(分布式文件系统)、MapReduce(分布式计算框架)、 YARN(分布式资源管理框架)

其它组件：ZooKeeper(分布式协调服务)、MySQL(数据库管理系统)、 Hive(数据仓库)、HBase(分布式数据库)、

Flume(数据采集工具)、Sqoop(数据传输工具)

7、Hadoop 安装好后的进程

Master节点上： NameNode、SecondaryNameNode、JobTracker

Slave节点（每一个）：DataNode、TaskTracker