大数据第7周

1. eclipse开发环境配置

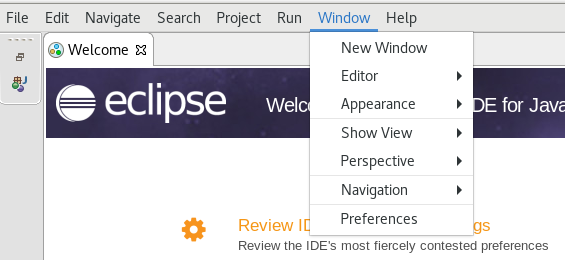
下载，解压eclipse，验证是否能够正常运行。

将hadoop-eclipse-plugin-2.6.0.jar拷贝到eclipse的dropins文件夹下。如果已经启动了eclipse，需要关闭重启eclipse。

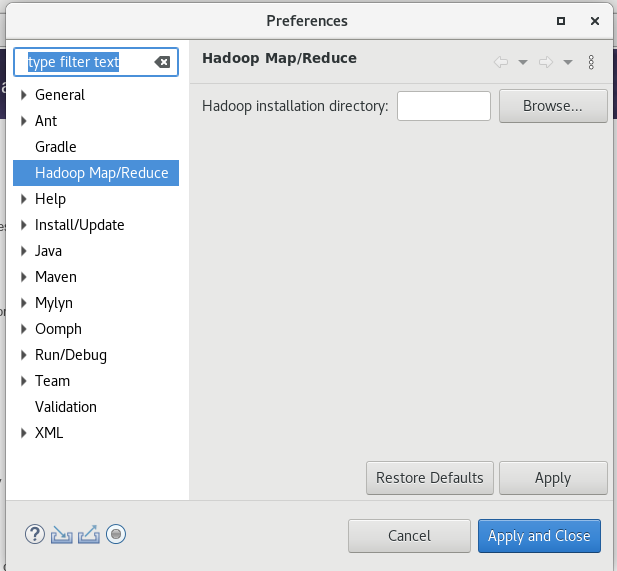
备注：应当使用hadoop2.7.7的插件，但是我没有下载到，只是上课时演示过程。

1.1配置hadoop路径

在主菜单上选择“window”，如下图，然后选择“Preferences”。

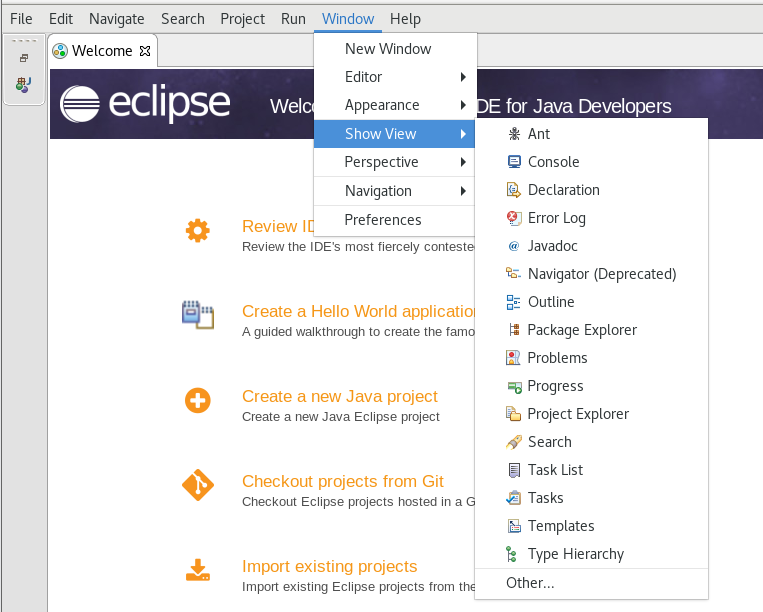


在如下界面左边选择“HadoopMap/Reduce”，在右边输入Hadoop的路径（hadoop\_home路径）。输入后点击“Apply and Close”。

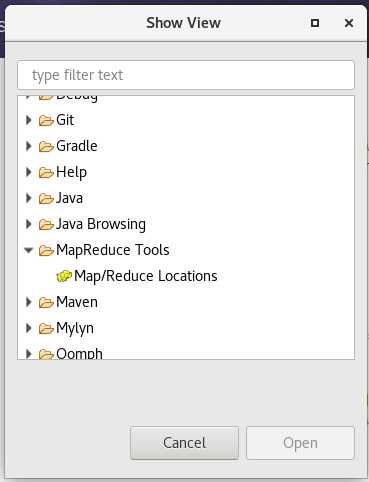


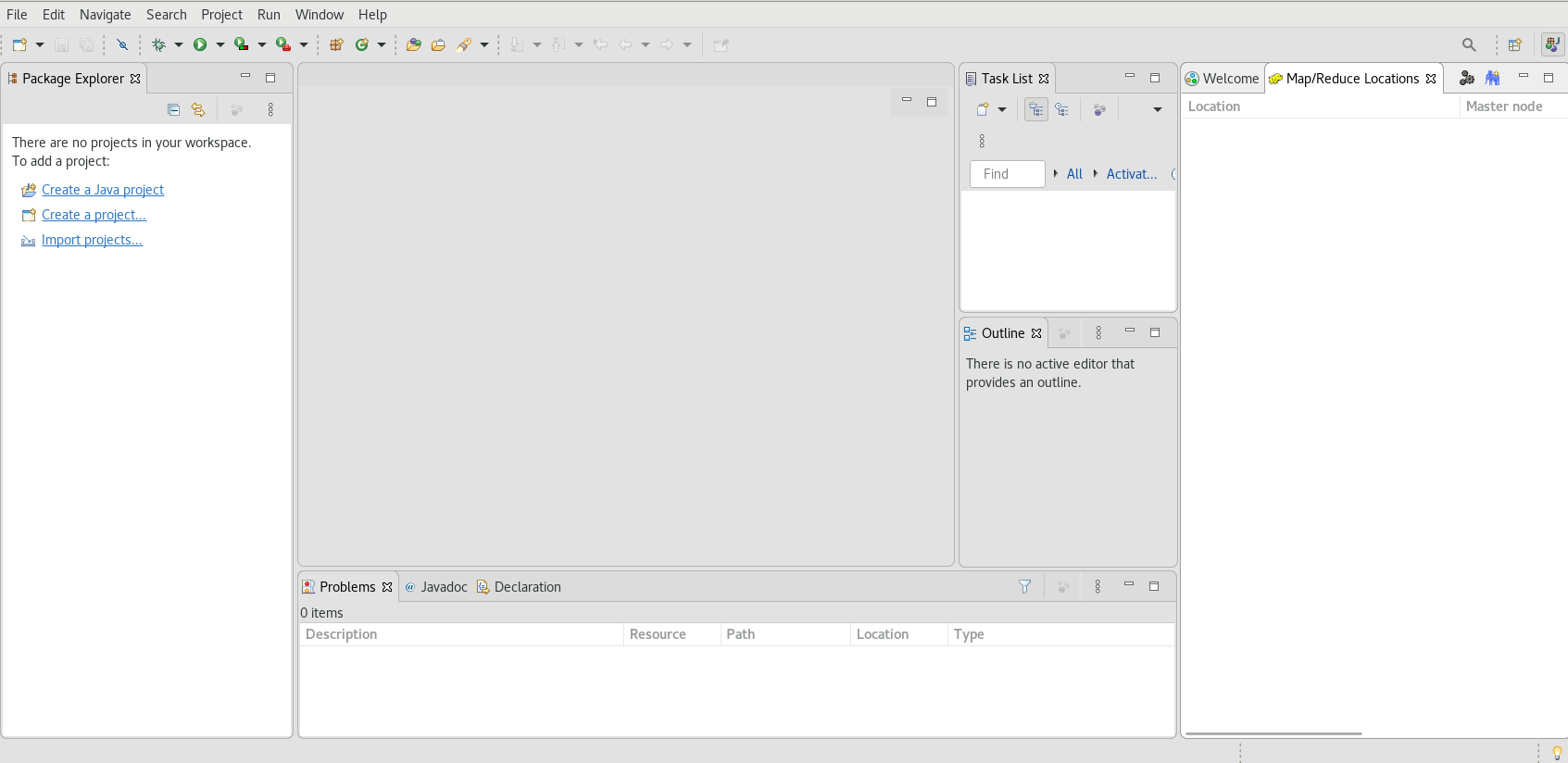
* 1. 配置Map/Reduce Locations

如下图，从主菜单开始，然后选择“Other”。

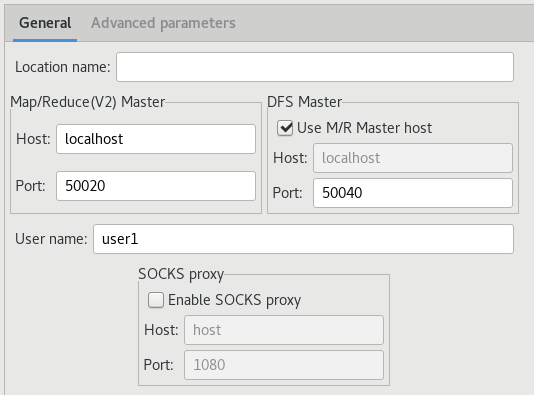


在以下界面中选择“Map/Reduce Locations”，然后点击”Open”。



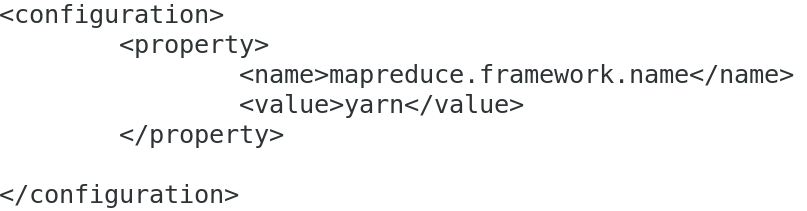


点击界面右侧上方的，新建hadoop location。

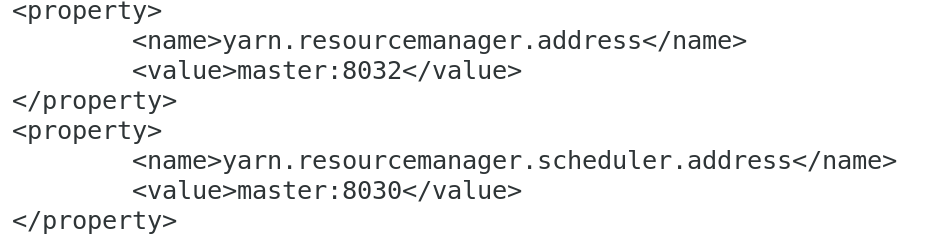


在此页面上，Location name栏输入名字，随便一个字符串都可以，除了系统保留字。

Map/Reduce(V2) Master标签页中，Host要输入master节点的ip地址，我当前master节点ip地址是10.0.0.20；Port要根据配置文件输入。打开mapred-site.xml：

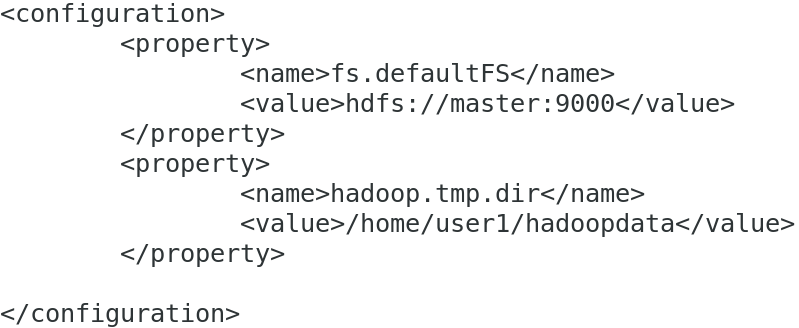


此文件中没有Map/Reduce的端口配置，只是指定了用yarn进行管理，所以需要打开yarn-site.xml:

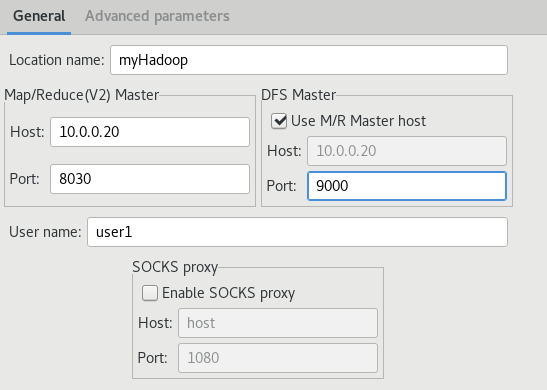


可填入如上两个端口中的一个。

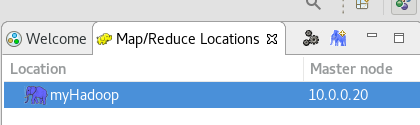
DFS Master标签页中，端口号也需要根据配置文件填入，打开dfs配置文件hdfs-site.xml，发现文件内容里没有相应的端口号配置，打开core-site.xml



可以看到端口号是9000。按照我的集群配置，填好后如下图。点击右下角“Finish”。

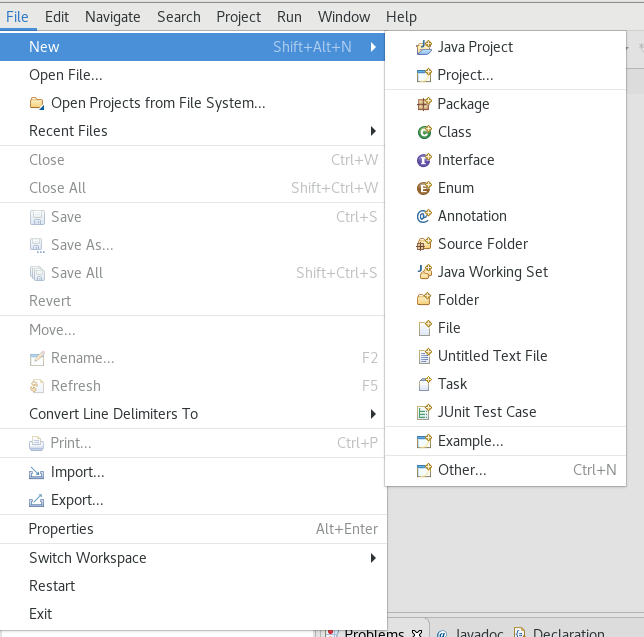


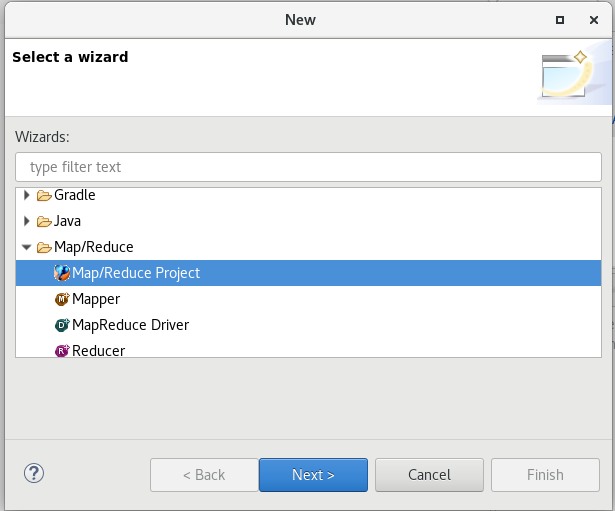
配置完成后，界面上会出现如下图内容：

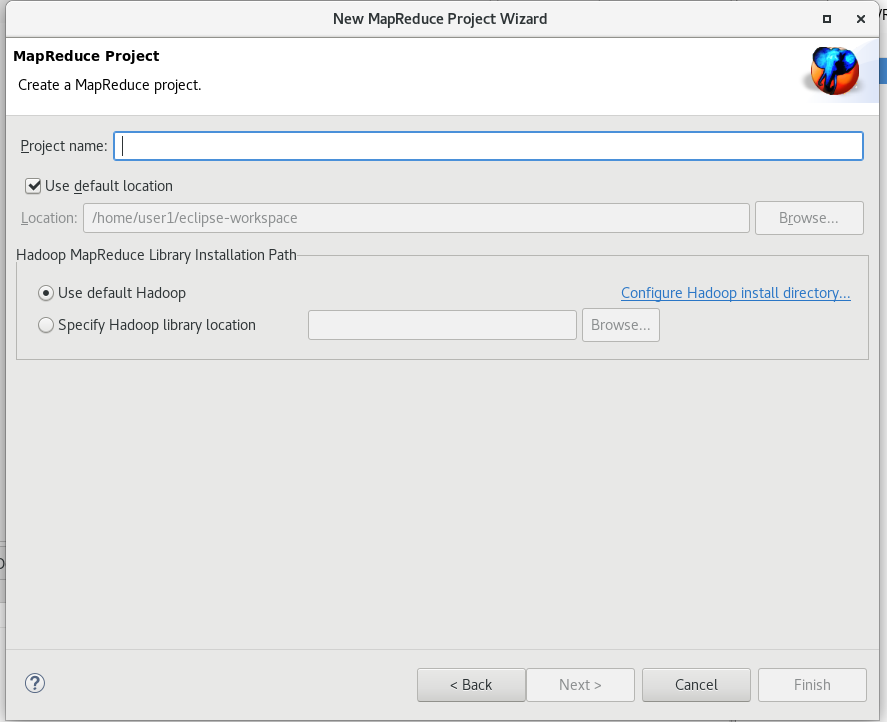


1. 建立java的project

从主菜单开始，如下图所示，然后选择“Other”。







给自己的project起个名字，然后点击“Finish”。

1. WordCount程序解释

3.1 main方法

public static void main(String[] args) throws Exception {

从main程序看，运行时需要输入参数，参数类型是String[]，由于形参数组大小不定，是根据实参确定的，所以，输入的数组元素的个数任意。

输入参数的意义？

Main的输入参数是文件路径，其中最后一个是输出文件的路径，其他的是输入文件的路径。

3.2 TokenizerMapper

3.3 IntSumReducer

下次课运行