

Hadoop Java API

该部分说明适用于PHBS 2020-2021秋季学期M1的大数据分析课程使用

更详细的使用说明请参考Apache hadoop Java API官方文档以及hadoop3.1.4的[官方文档](#)

client的HDFS配置

client上传的时候文件的replication和block size请配置对应的hdfs-site.xml文件，细节可参考上面hadoop-3.1.4的官方文档

client的host配置

1. 完整的windows host文件配置（请加入自己的host文件）。以下配置来源于金融实验室，仅供课程参考

localhost	127.0.0.1
10.0.50.139	hadoop-namenode
10.0.50.137	hadoop-datanode2
10.0.50.138	hadoop-datanode3
10.0.50.136	hadoop-datanode4
10.0.50.135	hadoop-datanode5

配置java依赖文件

1. 新建一个java项目
2. 添加java依赖文件（以intellij IDEA为例）

找到File - Project Structure. **首先**在Project Settings - Project中首先确认自己的Project SDK以及Project language level, 以jdk14为例, SDK是 14 (java version '14.0.2') , Project language level是14. **接下来**, 在Project Settings - Modules里面添加Dependencies, 所有需要的jar都在hadoop的目标路径下, 添加（当然也可以一次性添加）

```
YOUR_HADOOP_PATH\hadoop-3.1.4\share\hadoop\common,  
YOUR_HADOOP_PATH\hadoop-3.1.4\share\hadoop\common\lib,  
YOUR_HADOOP_PATH\hadoop-3.1.4\share\hadoop\hdfs,  
YOUR_HADOOP_PATH\hadoop-3.1.4\share\hadoop\mapreduce,  
YOUR_HADOOP_PATH\hadoop-3.1.4\share\hadoop\yarn.
```

添加完之后别忘了点击Apply, **[Optional]**如果之后想要输出jar包, 请同时设置Project Settings-Artifacts的Output layer(要先点击"+"新建)

3. 在上一步结束之后, 在src目录下新建java class开始coding, 注意, 此时使用 `import org.apache.hadoop.` 之后会出现自动填充(in IntelliJ IDEA)

尝试链接namenode

```
import java.io.BufferedReader;  
import java.io.IOException;
```

```

import java.io.InputStreamReader;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.fs.FSDataInputStream;

public class TestRead{
    public static void main(String[] args) throws IOException {
        Configuration conf = new Configuration();
        conf.addResource(new Path("E:\\bigDataAnalysis2\\hadoop-
3.1.4\\etc\\hadoop\\core-site.xml")); //写自己的path
        BufferedReader br = new BufferedReader(new
InputStreamReader(System.in));
        System.out.println("Enter the file path...");
        String filePath = br.readLine();

        Path path = new Path(filePath);
        FileSystem fs = path.getFileSystem(conf);
        FSDataInputStream inputStream = fs.open(path);
        System.out.println(inputStream.available());
        fs.close();
    }
}

```

选择Run这个java class, 输入 `hdfs://hadoop-namenode:9082/fileName_8MB.txt` 或其他存在的文件名会返回file size