

大数据开发技术

东北林业大学

卢洋

第一章

MapReduce概述

1.4

官方WordCount源码

源代码在哪?

- 反编译;
- src文件;
- GitHub.


```
https://github.com/apache/hadoop/tree/trunk/hadoop-  
mapreduce-project/hadoop-mapreduce-examples/src/  
main/java/org/apache/hadoop/examples
```

```
https://github.com/apache/hadoop/blob/  
branch-2.7.2/hadoop-mapreduce-project/hadoop-  
mapreduce-examples/src/main/java/org/apache/  
hadoop/examples/WordCount.java
```


1.5

常用数据的序列化类型

Hadoop自己定义重写的类型 ，可序列化

Java类型	Hadoop Writable 类型
boolean	BooleanWritable
byte	ByteWritable
int	IntWritable
float	FloatWritable
long	LongWritable
double	DoubleWritable
String	Text
map	MapWritable
array	ArrayWritable