**学习日志**

姓名： 陈昭宇 学号：1812977 日期：6.24

**学习内容安排：**

上午：

Hadoop的I/O

下午：

MapReduce的应用开发

**学习反馈：**

已掌握知识：

常用错误检测代码CRC-32

用open传递false给setVerifyChecksum() -get -copyToLocal -ignpreCrc

xgexksumFileSystem

压缩

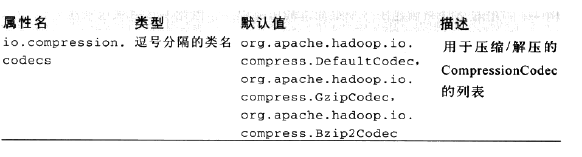


Gzip -1 file(-1,表示速度最优，-9表示空间最优)

编码。解码器

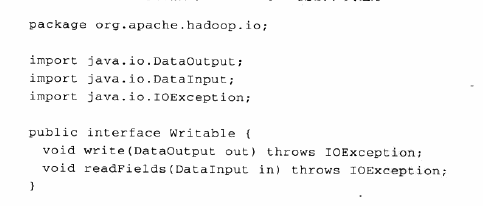






序列化

Writable



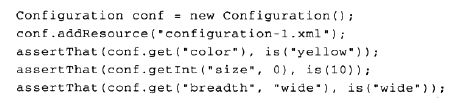
Writeable

Text类

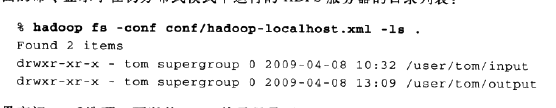
自定义Writeable类

MapReduce应用开发

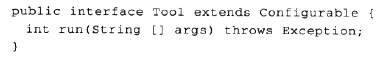
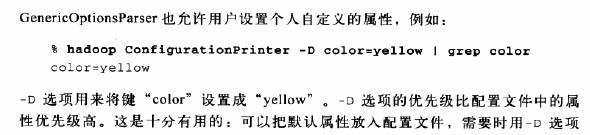
API的配置

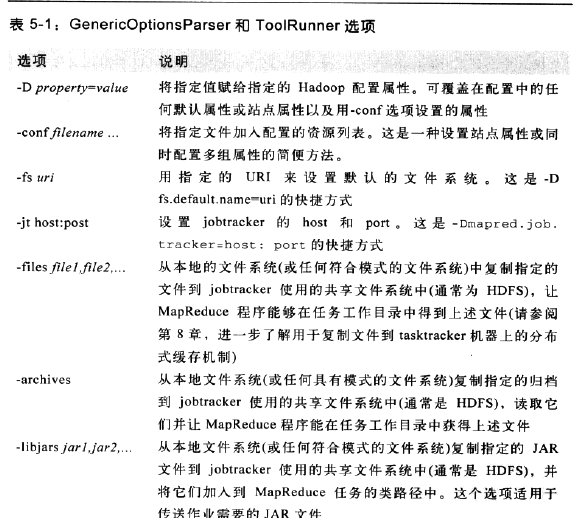


后定义的属性会覆盖之前的属性标记为final不能被覆盖

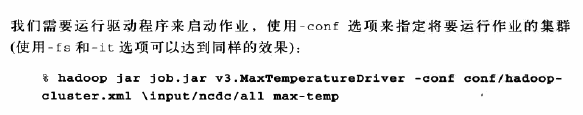


Hadoop的辅助类

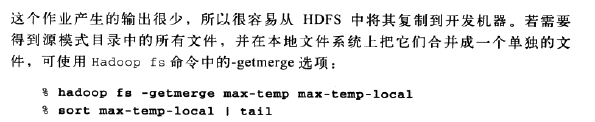


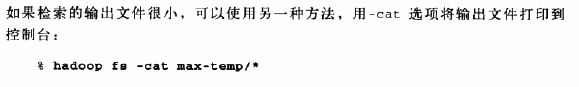


在集群上运行



将utility类应用于parse记录的mapper当作业数过多10000时后他们的ID为把作业前缀替换成任务前缀，再加一个后缀



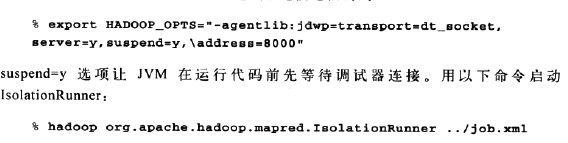


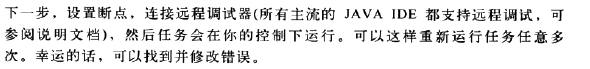
通过map输出流来调试作业

使用远程调试器









MapReduce的工作流

将复杂的问题分解成MapReduce

运行独立的作业：

对于线性的，一个接一个运行作业。

更复杂的借助类库

未掌握知识： SequuenFile, 测试驱动器

学习心得：感觉看的不太明白。就今天主要的学习是编写和I/O，就看的明白他的思路，但到自己写就无从下手。还要多体会一下

。