

大数据开发技术

东北林业大学

卢洋

第三章

MapReduce 框架原理

NLineInputFormat

NLineInputFormat

- 如果使用NLineInputFormat，每个map进程处理的InputSplit不再按Block块去划分，而是按NLineInputFormat指定的行数N来划分；
- 即：
输入文件的总行数/ N =切片数；
- 如果不能整除，余数部分生成一个切片。

示例

有4行输入

Rich learning form

Intelligent learning engine

Learning more convenient

From the real demand for more close to the enterprise

如果N是2
则每个输入分片包含两行
开启2个MapTask

1. 切片1:

(0,Rich learning form)

(19,Intelligent learning engine)

2. 切片2:

(47,Learning more convenient)

(72,From the real demand for more close to the enterprise)