

大数据开发技术

东北林业大学

卢洋

第三章

MapReduce 框架原理

TextInputFormat

TextInputFormat

- **TextInputFormat**是默认的**FileInputFormat**实现类，按行读取数据。
- 键：存储该行在整个文件中的起始字节偏移量，**LongWritable**类型。
- 值：这行数据的内容，不包括任何终止符(换行符和回车符)，**Text**类型。

示例

Rich learning form

Intelligent learning engine

Learning more convenient

From the real demand for more close to the enterprise

❶ 示例：

Rich learning form

Intelligent learning engine

Learning more convenient

From the real demand for more close to the enterprise

❷ 每条记录表示为键/值对：

(0, Rich learning form)

偏移量，行

(19, Intelligent learning engine)

(47, Learning more convenient)

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