Standard Input Formats provided by Hadoop :

|  |  |
| --- | --- |
| **Input Format** | **Description** |
| TextInputFormat | Each line in the text files is a record. Key is the byte offset of the line, and value is the content of the line.key: LongWritable value: Text |
| KeyValueTextInputFormat | Each line in the text files is a record. The first separator character divides each line. Everything before the separator is the key, and everything after is the value. The separator is set by the key.value.separator.in.input. line property, and the default is the tab (\t) character.key: Text value: Text |
| SequenceFileInputFormat<K,V> | An InputFormat for reading in sequence files. Key and value are user defined. Sequence file is a Hadoop specific compressed binary file format. It’s optimized for passing data between the output of one MapReduce job to the input of some other MapReduce job.key: K (user defined) value: V (user defined) |
| NLineInputFormat | Same as TextInputFormat, but each split is guaranteed to have exactly N lines. The mapred.line.input.format. linespermap property, which defaults to one, sets N.key: LongWritable value: Text |