Implementing a custom InputFormat

Hadoop enables us to implement and specify custom InputFormat implementations for our MapReduce computations. We can implement custom InputFormat implementations to gain more control over the input data as well as to support proprietary or application-specific input data file formats as inputs to Hadoop MapReduce computations.

Assume we are getting Sensor Data as follows

SensonID timestamp status value1 value2

a 1386023259550 on 22 23

b 1389523259550 off 33 34

A InputFormat implementation should extend the org.apache.hadoop.mapreduce.InputFormat<K,V> abstract class overriding the createRecordReader() and getSplits() methods.

In this example, we implement a InputFormat and a RecordReader for above Sensor data.

This InputFormat will generate customised MyKey instances as keys and MyValue instances as the values.