

Module 3: Hadoop MapReduce Framework - I

Assignment

edureka!

edureka!

© 2014 Brain4ce Education Solutions Pvt. Ltd.

Module 3: Hadoop MapReduce Framework - I

Assignment – Calculate the number of sub-patents

Table of Contents

Introduction	2
Problem statement	2

edureka!

Introduction

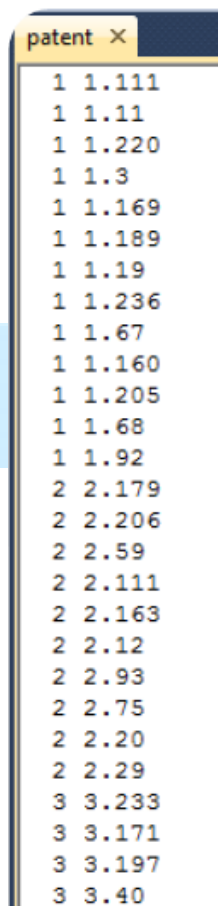
Apply your MapReduce programming knowledge and write a MapReduce program to process a dataset with patent records. You need to calculate the number of sub-patents associated with each patent.

The dataset for this problem is the '**patent**' records file available in your LMS.

Problem statement

Let's understand the problem through a subset of patent records as shown in the following figure:

FIGURE 1-1 PATENT RECORDS



1	1.111
1	1.11
1	1.220
1	1.3
1	1.169
1	1.189
1	1.19
1	1.236
1	1.67
1	1.160
1	1.205
1	1.68
1	1.92
2	2.179
2	2.206
2	2.59
2	2.111
2	2.163
2	2.12
2	2.93
2	2.75
2	2.20
2	2.29
3	3.233
3	3.171
3	3.197
3	3.40

Your task in this assignment is to process the 'alphabets' text file using MapReduce program. Your program should calculate the size of each word (Number of alphabets in the word) and count the number of words of that size in this text file.

You can review the solution in your LMS.