International Journal of Emerging Technology and Innovative Engineering
Volume I, Issue 5, May 2015 (ISSN: 2394 – 6598)

MULTI LINGUAL DICTIONARY USING MAPREDUCE

1st Author
A Betty
MTech, CSE, KMIT
Betty.kmit@gmail.com

ABSTRACT

A **dictionary** is collection of words in one or more specific languages, often listed alphabetically (or by radical and stroke for ideographic languages), with usage of information, definitions, etymologies, phonetics, pronunciations, translation, and other information; or a book of words in one language with their equivalents in another, also known as a lexicon. It is a lexicographical product designed for utility and function, curated with selected data, presented in a way that shows inter-relationships among the data.

General Terms

Hadoop

MapReduce

Hadoop Distributed File System

Keywords

MR- MapReduce

HDFS-Hadoop Distributed File System

Jar –Java Archive

1. INTRODUCTION

Apache Hadoop is a set of algorithms (an open-source software framework) for distributed storage and distributed processing of very large data sets (Big Data) on computer clustersbuilt from commodity hardware. All the modules in Hadoop are designed with a fundamental assumption that hardware failures (of individual machines, or racks of machines) are commonplace and thus should be automatically handled in software by the framework.

The core of Apache Hadoop consists of a storage part (Hadoop Distributed File System (HDFS)) and a processing part (MapReduce). Hadoop splits files into large blocks (default 64MB or 128MB) and distributes the blocks amongst the nodes in the cluster. To process the data, Hadoop Map/Reduce transfers code (specifically Jar files) to nodes that have the required data.

.

2. Existing System

A **dictionary** is collection of words in one or more specific languages, often listed alphabetically (or by radical and stroke for ideographic languages), with usage of information, definitions, etymologies, phonetics, pronunciations, translation, and other information; or a book of words in one language with their equivalents in another, also known as a lexicon. It is a lexicographical product designed for utility and function, curated with selected data, presented in a way that shows inter-relationships among the data. Hadoop is an open source project for processing large datasets in parallel with the use of low level commodity machines.

3. Proposed System

In the proposed system we come up with the solutions that we were facing in the existing systems. Here we use the apache Hadoop and MapReduce which is use to deal with the Big Data problems. In the proposed system we write a MapReduce application which takes input (Big Data) a word with meanings in different languages and finally generates an output which facilitates the users by giving a single Dictionary which is Multi lingual.

4. Screen Shots

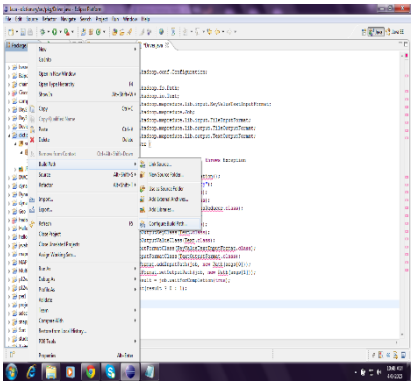

Fig 4.1Creating a Hadoop environment

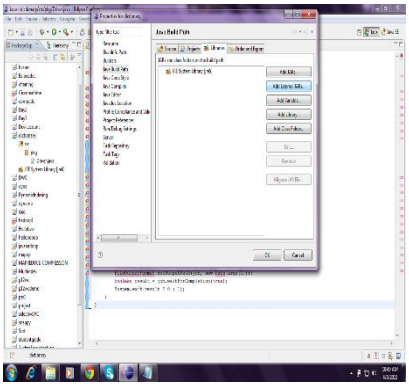

Fig 4.2 Adding External Jars

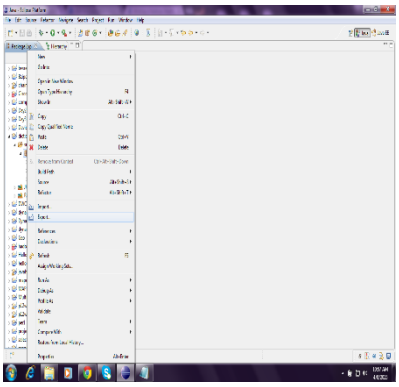

Fig 4.3 Exporting the code

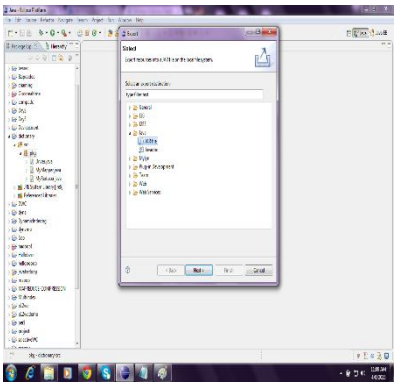

Fig 4.4Creating a .jar file of our code

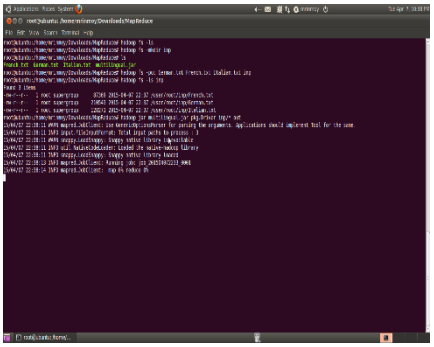

Fig 4.5 MapReduce Triggered

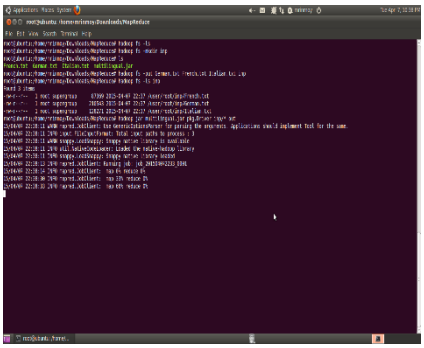

Fig 4.6MapReduce getting started

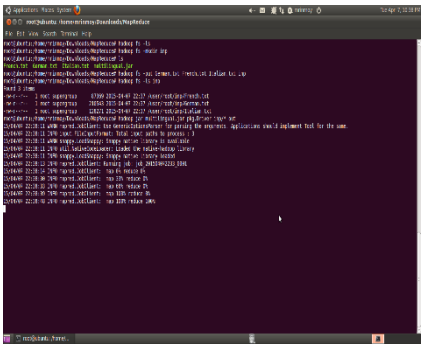

Fig 4.7 MapReduce finished

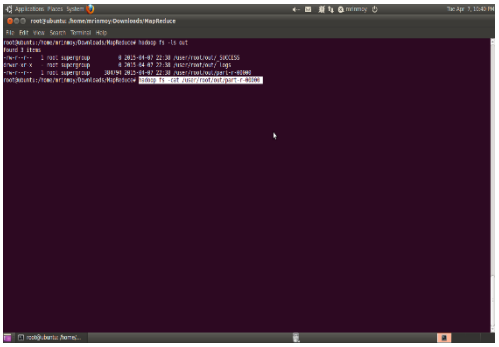

Figure 4.8 seeing the contents of output part-r-00000 file

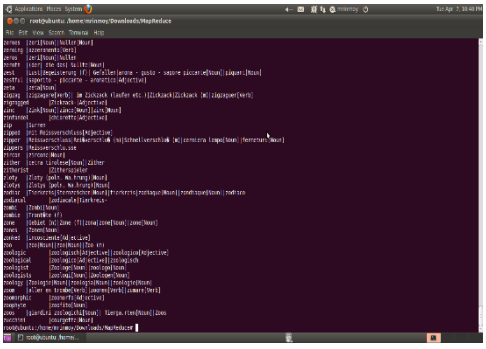

Figure 4.9 final output of the multilingual dictionary.

5. SYSTEM CONFIGURATION

Hardware System Configuration

Processor	-	Pentium-III
Speed	-	1.1 Ghz
RAM	-	1 GB (min)
Hard Disk	-	20 GB
Key Board	-	Standard Windows Keyboard
Mouse	-	Two or Three Button Mouse
Monitor	-	SVGA

Software System Configuration

Operating System	-	Linux (Preferably Fedora)
Server Cluster	-	Hadoop-1.2.1
Tools Used	-	Java 1.6

6. REFERENCES

[1] Hadoop MapReduce Cookbook, SrinathPerera, ThilinaGunarathne, 2013, ISBN 978-1-84951-728-7

[2] Understanding BigData, Paul Zikopoulos, 1976, ISBN 978-0-07-179053-6.

[3] Planning for BigData, O'Reilly Radar Team, 2012, ISBN 978-1-449-32967-9.

[4] http://hadoop.apache.org

[5] http://wiki.apache.org/hadoop/GettingStartedWithHadoop

[6] http://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html