Journal of Information Technology and Society
Volume: 01 Issue: 01 June 2021 Page No.11-14
ISSN: XXX-XXXX

# An Empirical Assessment for Word Counter using Hadoop – MapReduce by Implementing of Android Application

G. Vasantha Rani
Assistant Professor, Don Bosco College, Yelagiri Hills

**ABSTRACT -** In Hadoop, MapReduce is an interaction that decreases enormous control exercises into fragmented errands that can be executed in equal across an area important to workers. The yield of assignments can be combined to finish eventual outcomes. The contribution to each stage is key-esteem sets. There are two sorts of assignments: In Map errands (Splits and Mapping), Set of information will be considered as an admission and the outcome will be an adjusted arrangement of information (key and Value - sets). In upgraded assignments, aftereffect of the Map work (set of Tuples) will be considered as an admission and Changes into a divided arrangement of tuples (Shuffling, Reducing). There are 5 stages in which work process of MapReduce falls and there are Splitting the boundary, planning the comparative words, Intermediate parting of the planned words, decreasing the comparable words and Combining equivalent to a single word. In this paper, I intend to utilize an android application for checking words, sentences, sections and characters in the content which client offers and to discover the recurrence of each word. Here, the part of Mapper is to plan the keys to the current qualities and the job of Reducer is to total the keys of regular qualities. Thus, everything is addressed as a Key-esteem pair. This examination needs to know the recurrence of each word and furthermore what that word is. Along these lines, we will execute it in an android application and investigate the recurrence of each given word. This can be tried or carried out with the gigantic information.

**KEYWORDS:** MapReduce, Intermediate splitting, splitting parameter, MapReduce function logic, set of tuples.

## I. INTRODUCTION

Hadoop MapReduce is a programming worldview at the core of Apache Hadoop for giving monstrous versatility across hundreds or thousands of Hadoop bunches on ware equipment. The MapReduce model cycles enormous unstructured informational indexes with a conveyed calculation on a Hadoop cluster.In the present information driven market, calculations and applications are gathering information day in and day out about individuals, cycles, frameworks, and associations, bringing about gigantic volumes of information. The test, however, is the way to handle this enormous measure of information with speed and proficiency, and without forfeiting significant insights.This is the place where the MapReduce programming model acts the hero. At first utilized by Google for dissecting its query items, MapReduce acquired huge prominence because of its capacity to part and handle terabytes of information in equal, accomplishing faster outcomes.

The term MapReduce addresses two isolated and unmistakable assignments Hadoop programs perform-Map Job and Reduce Job. Guide work scales take informational collections as info and interaction them to deliver key worth sets. Decrease work takes the yield of the Map work for example the key worth matches and totals them to deliver wanted outcomes. The information and yield of the guide and decrease occupations are put away in HDFS. I will talk about and execute "How MapReduce Algorithm takes care of WordCount Problem in Android Studio Project" hypothetically and Practically and furthermore going to discover the check of the quantity of events of each word accessible in a DataSet and the equivalent can likewise be tried with test sources of info and it will create precise yields. To accomplish this I have executed the MapReduce Algorithm into an android application project to test/discover the exactness of the equivalent. At long last, I have given the outcomes with examination of word include in different dialects/innovations or devices.

Journal of Information Technology and Society
Volume: 01 Issue: 01 June 2021 Page No.11-14
ISSN: XXX-XXXX

## II. LITERATURE REVIEW

It assists with parting the information informational collection into various parts. Runs a program on all information parts equal immediately. The term Map Reduce alludes to two independent and particular undertakings. The first is the guide activity, takes a bunch of information and converts it into another arrangement of information. The decrease activity consolidates those information tuples. It runs in the Hadoop climate to offer versatility, conventionality, hustle, restoration and simple outcomes for information handling.

Map Reduce is a programming system that permits us to perform dispersed and equal preparing on enormous informational indexes in an appropriated climate. These are the issues which we need to take care independently while performing equal handling of immense informational collections when utilizing conventional methodologies. To defeat those issues, we have the Map Reduce system which permits us to perform such equal calculations without worrying about the issues like dependability, adaptation to internal failure and so forth Subsequently, Map Reduce gives you the adaptability to compose code rationale without thinking often about the plan issues of the framework.

## III. METHODOLOGY & IMPLEMENTATION

**Guide Function**

Guide Function is the initial phase in Map Reduce Algorithm. It takes input undertakings and partitions them into more modest sub-errands. At that point perform required calculation on each sub-task in equal. This progression plays out the accompanying two sub-steps:1) Splitting 2) Mapping
MapReduce First Step Output = List of <key, Value> Pairs

**Mix Function**

It's otherwise called "Join Function", It plays out the accompanying two sub-steps: Merging and Sorting. It takes a rundown of yields coming from "Guide Function" and plays out these two sub-steps on every single key-esteem pair.
Mix Function Output= List of <Key, List<value>> Pairs

**Lessen Function**

It is the last advance. It performs just one stage: Reduce step. It takes rundown of <Key, List<Value>> arranged sets from Shuffle Function and perform decrease activity
Diminish Function Output= Sorted rundown of Key Value sets

**Tasks in Map Reduce Algorithm**

Input Phase, Map Phase, Combiner, Shuffle and Sort, Reducer phase and Output phase.

**Guide Reduce Component:** Mapper Class, Reducer Class, Driver Class Benefits of MapReduce: Parallel Processing and Data Locality
**MapReduce Algorithm**: Map Function, Shuffle Function . Reduce Function

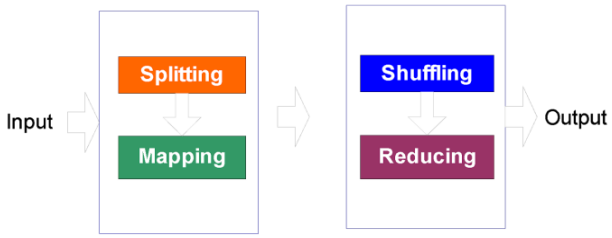


Figure 1 : Mapreduce word count process

MapReduce Word count Process (MapReduce Method) This is an Mapreduce word count process in which terms are represented in diagrammatical way. Were input is a Dataset for Map Function. This includes the

Journal of Information Technology and Society
Volume: 01 Issue: 01 June 2021 Page No.11-14
ISSN: XXX-XXXX

Splitting and Mapping process. Then, the Map Function output is given as an input for Reduce Function. Which includes shuffling and Reducing process.

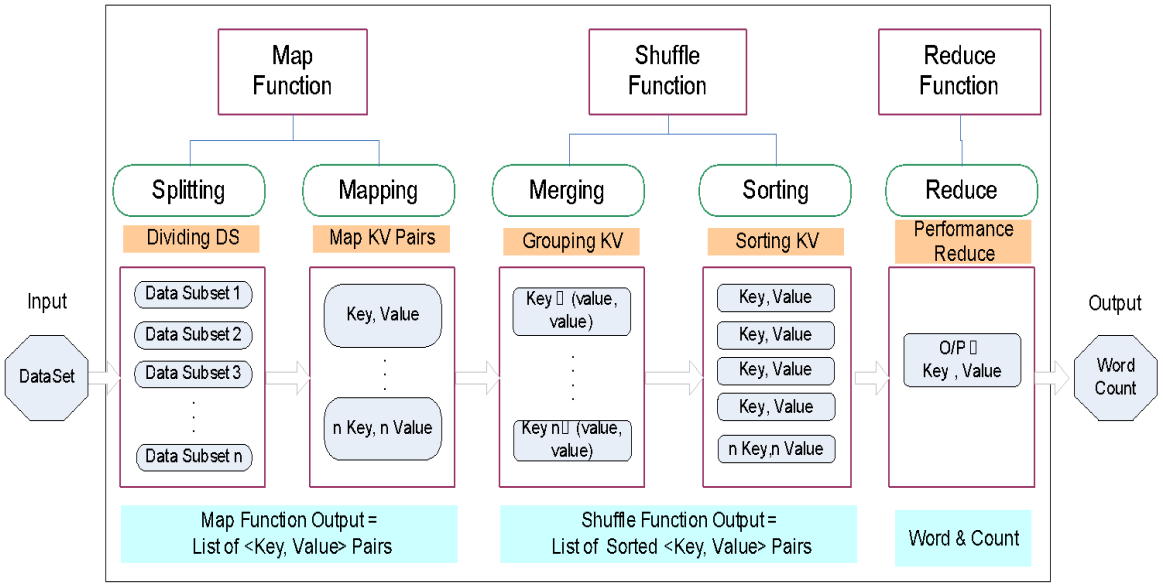


Figure 2 : MapReduce Algorithm Implementation

MapReduce Algorithm Implementation Coding

```
public static class IntSumReducer extends
Reducer<Text,IntWritable,Text,IntWritable>
{
  private IntWritable result = new IntWritable();
  public void reduce(Text key, Iterable values,Context context) throws IOException,      InterruptedException
  {
     int sum = 0;
     for (IntWritable val : values)
    {
        sum += val.get();
    }
   result.set(sum);
   context.write(key, result);
 }
}
```

Sample Dataset is given and the output of Word counter is shown based on the implementation of the algorithm.

**Input - Sample Dataset**
Welcome to Don Bosco College
I am studying in Don Bosco College
I am working in Don Bosco College
I like Don Bosco College

**Output:**
am - 2
Bosco - 4
College - 4

Journal of Information Technology and Society
Volume: 01 Issue: 01 June 2021 Page No.11-14
ISSN: XXX-XXXX

Don - 4
i - 3
in - 2
like - 1
to - 1
Welcome – 1

## IV. EXPERIMENTAL RESULTS

Map Reduce implementation for Word Count Algorithm computation. The outcome is the list of words with the count of appearance of each word. Results are possible using a Map Reduce Word Count algorithm. Research dataset is stored on a local file system. Text cleaning has been done by removing corrupted, erroneous, misleading and empty fields. It has been copied from the local file system to HDFS. The Word Count module is developed in the Java programming language and then the .JAR file is uploaded to single node storage. The data are tokenized using the Map Reduce algorithm in order to establish the interest area of research.

## V. CONCLUSION

All in all, the Map Reduce calculation is carried out and tried for a word include measure in which it checks the quantity of the incessant words in a given dataset. That is to carry out it into an android application, where java is utilized as a programming language and the upheld class documents or container records are imported. Contrasting the customary route or by doing likewise with different prospects, this method of drawing nearer gives more exactness. In this examination paper I have done a word tally from the Dataset, in future I intend to test and carry out for site page based Map Reduce.

## REFERENCES

[1] Yahoo! Inc, Hadoop Tutorial from Yahoo! Available:http://developer.yahoo.com/hadoop/tutorial/index.html

[2] Jens Dittrich and JorgeArnulfo Quian´eRuiz, " Efficient Big Data processing in Hadoop Mapreduce," Proceedings of the VLDB Endowment, Volume 5 Issue 12, August 2012, Pages 2014-2015

[3] Arnab Nandi, Cong Yu, Philip Bohannon, and Raghu Ramakrishnan, Fellow, IEEE, “Data Cube Materialization and Mining over MapReduce” TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 6, NO. 1, JANUARY 2012

[4] Abouzeid, A., Bajda-Pawlikowski, K., Abadi, D., Rasin, A., and Silberschatz, A. 2010. HadoopDB in action: Building real world applications. In Proceedings of the 36th ACM SIGMOD International Conference on Management of Data (SIGMOD‟10).

[5] MapReduce: Simplified Data Processing on Large Clusters. Available at http://labs.google.com/papers/mapreduceosdi04.pdf