

# Improving Efficiency of Map Reduce Paradigm with ANFIS for Big Data

**Gor Vatsal H.**

*Department of Computer Science and Engineering  
NarNarayan Shashtri Institute of Technology Jetalpur ,  
Ahmedabad , India*

**Prof. Vatika Tayal**

*Department of Computer Science and Engineering  
NarNarayan Shashtri Institute of Technology Jetalpur ,  
Ahmedabad , India*

## Abstract

As all we know that map reduce paradigm is became synonyms for computing big data problems like processing, generating and/or deducing large scale of data sets. Hadoop is a well know framework for these types of problems. The problems for solving big data related problems are varies from their size , their nature either they are repetitive or not etc., so depending upon that various solutions or way have been suggested for different types of situations and problems. Here a hybrid approach is used which combines map reduce paradigm with anfis which is aimed to boost up such problems which are likely to repeat whole map reduce process multiple times.

**Keywords: Big Data, fuzzy Neural Network, ANFIS, Map Reduce, Hadoop**

## I. INTRODUCTION

Initially, to solve problem various problems related to large crawled documents, web requests logs, row data , etc a computational processing model is suggested by jeffrey Dean and Sanjay Ghemawat is Map Reduce in 2004[1]. MapReduce programming model is inspired by map and reduce primitives which are available in Lips and many other functional languages. It is like a De Facto standard and widely used for solving big data processing and related various operations. In 2005, Hadoop named framework was developed for solving large data set processing in distributed manner by using map reduce by Doug Cutting and Mike Cafarella at YAHOO for the Nutch search engine project [2]. In simple MapReduce is combination of map function , shuffle function and reduce function. Here ANFIS is combined with MapReduce paradigm which stands for Adaptive Neuro Fuzzy inference system. ANFIS combines Fuzzy logic with Neural Network. There exist such situation where hadoop processes are needed to run repeatedly, for that purpose such concept which introduced here may useful.

## II. CURRENT ISSUES IN MAP REDUCE / HADOOP AT A GLANCE

In general there exist numerous problems in current hadoop/Map Reduce square measure like Performance issues, Programming model problems, Configuration and Automation problems, Performance improvement.[3] Here additionally one defect is static procedure of map-reduce-shuffle that is materialized lack of support for unvaried information sets with minor changes as well as major changes and equalization of state between jobs too. We take associate example for social networking sites like google+ or facebook[4] that recommend suggestion of events, ads, activities of friends and relative post as per users' profiles. Now smaller amendment in this users' profiles or interest want to once more run these whole process repeatedly that is time consuming and costlier. thus for these sorts of state of affairs if we have a tendency to use already computed map perform output for regenerate suggestions for any changes in input, thus we've got to calculate only changed information rather than whole method once more.

## III. AVAILABLE TECHNOLOGIES FOR BIG DATA PROCESSING

There are variety of framework are available like MapReduce Online , EARL , Hadoop++, MRshare, ReStore, HaLoop, Incoop, Starfish, Maninal, Hive, Pig, WibiData, PLATFORA, apache spark and so on. But for iterative processing suggested technology or can say methodology/technique introduced in HaLoop , it is intended to provide solution for recursive/iterative data analysis task such as descendant queries , web page rank and Hits , k-means clustering etc.

## IV. PROPOSED METHODOLOGY

In this paper the approach is used which combines Map function only from current map reduce paradigm and fed as input to the ANFIS which stands for Adaptive Neuro Fuzzy Inference System. The <key,value> pair generated by Map function is fed as input to the ANFIS. ANFIS is combination of Fuzzy logic and neural network. Fuzzy logic stands for If... THEN rules and which are computed with the help of neural network. So Steps are as follow,

- 1) Step 1 load files to Hadoop file (HDFS)

- 2) Step 2 Evaluate data by applying Map function
- 3) Step 3 Intermediate key, value pairs are generated by Map function
- 4) Step 4 Rules are created based on key, value pairs
- 5) Step 5 Intermediate key, values pairs are fed as input in ANFIS
- 6) Step 6 finally output is generated by ANFIS structure same as in Reduce phase after shuffle in current map reduce paradigm

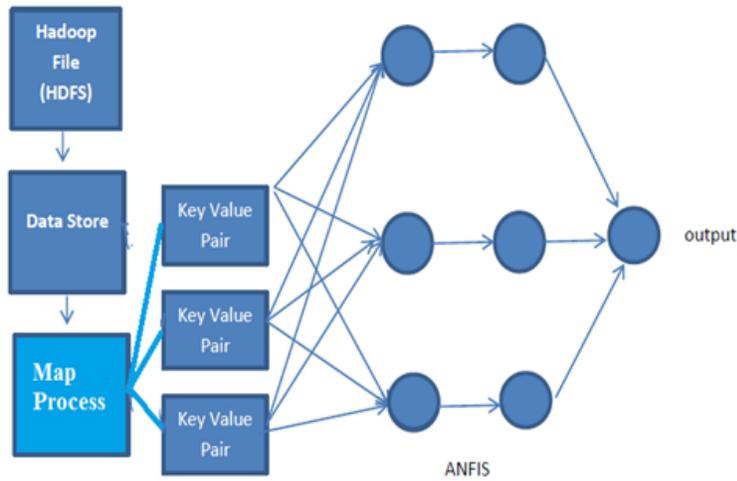
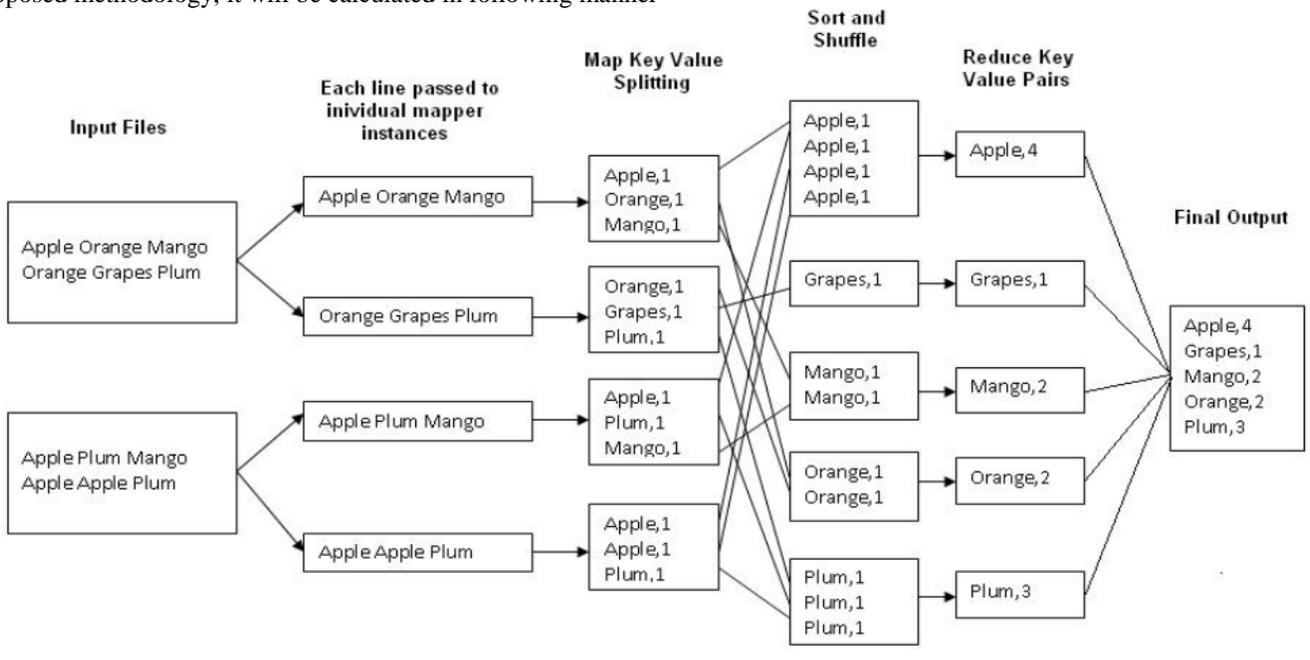


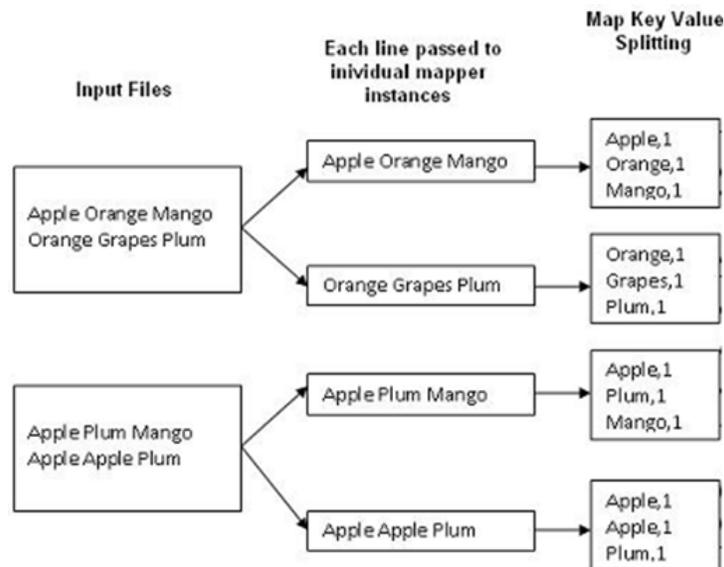
Fig. 1: proposed model

To understand above steps, here a small example is illustrated in which as input fruits are listed, then they are divided in chunks same as in map reduced paradigm before map function. Then map function is applied to that chunks of data and intermediate key value pairs are generated.

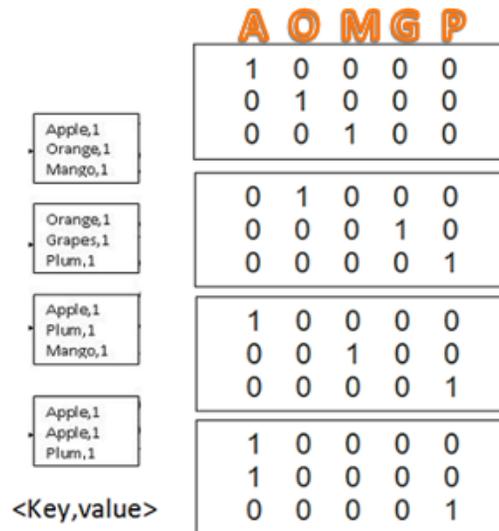
As shown in above fig demonstrates how map reduces paradigm runs in simple way. Instead of this, as mentioned above proposed methodology, it will be calculated in following manner



- Intermediately key value pair generation from map function



- Preparing output from map function for ANFIS

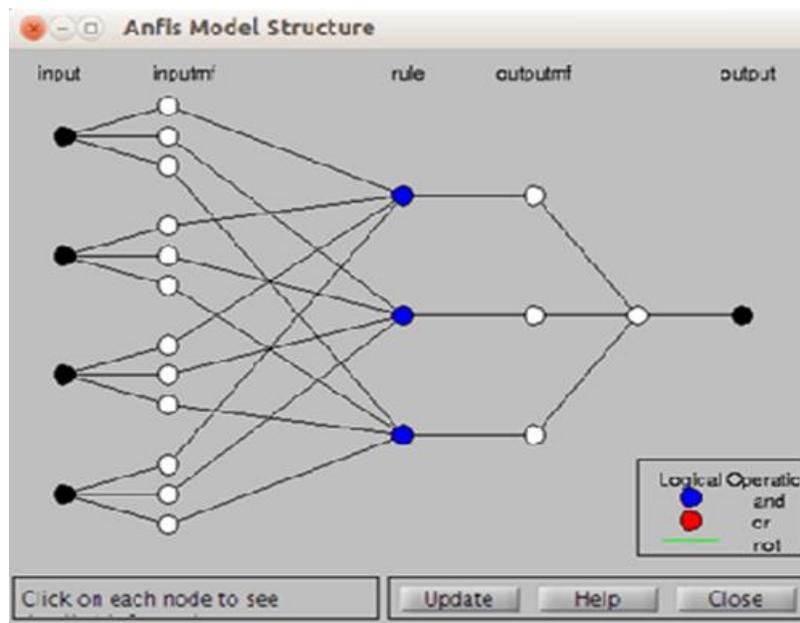


- Input file for ANFIS would be As follow in this particular case

```

1 0 0 0 0
0 1 0 0 0
0 0 1 0 0
0 1 0 0 0
0 0 0 1 0
0 0 0 0 1
1 0 0 0 0
0 0 1 0 0
0 0 0 0 1
1 0 0 0 0
1 0 0 0 0
0 0 0 0 1
    
```

- This will generate ANFIS structure as follow



## V. CONCLUSION

Here is suggest a totally different approach from existing map reduce paradigm to solve problems related to extremely large data sets and when they needed iterative methods or processes. So in such cases this totally new approach will be very useful.

## VI. FUTURE ENHANCEMENT

Here suggested approach is in initial stage, A suitable framework is needed that combines map and anfis calculation processes under same roof in distributed manner and as well as single node. So it needed a framework that enough able to handle and execute processes as per suggested methodology.

## REFERENCES

- [1] Jeffrey Dean, Sanjay Ghemawat "MapReduce: Simplified Data Processing on Large Clusters", appear in OSDI , page 1-14 , in 2014
- [2] "Michael J. Cafarella". Web.eecs.umich.edu. Retrieved 2013-04-05.
- [3] Anam Alam, Jamil Ahmed. "Hadoop Architecture and its issues", 2014 International Conference on Computational Science and Computational Intelligence @2014 IEEE ISBN – 978-1-4799- 3010-4
- [4] Vasiliki Kalavri , Vladimir Vlassov. "MapReduce: Limitation , Optimizations and Open Issues" 12th International Conference on Trust , Security and Privacy in Computing and Communications 2013