

Chief Editor

Dr. A. Singaraj, M.A., M.Phil., Ph.D.

Editor

Mrs.M.Josephin Immaculate Ruba

EDITORIAL ADVISORS

1. Prof. Dr.Said I.Shalaby, MD,Ph.D.
Professor & Vice President
Tropical Medicine,
Hepatology & Gastroenterology, NRC,
Academy of Scientific Research and Technology,
Cairo, Egypt.
2. Dr. Mussie T. Tessema,
Associate Professor,
Department of Business Administration,
Winona State University, MN,
United States of America,
3. Dr. Mengsteab Tesfayohannes,
Associate Professor,
Department of Management,
Sigmund Weis School of Business,
Susquehanna University,
Selinsgrove, PENN,
United States of America,
4. Dr. Ahmed Sebihi
Associate Professor
Islamic Culture and Social Sciences (ICSS),
Department of General Education (DGE),
Gulf Medical University (GMU),
UAE.
5. Dr. Anne Maduka,
Assistant Professor,
Department of Economics,
Anambra State University,
Igbariam Campus,
Nigeria.
6. Dr. D.K. Awasthi, M.Sc., Ph.D.
Associate Professor
Department of Chemistry,
Sri J.N.P.G. College,
Charbagh, Lucknow,
Uttar Pradesh. India
7. Dr. Tirtharaj Bhoi, M.A, Ph.D,
Assistant Professor,
School of Social Science,
University of Jammu,
Jammu, Jammu & Kashmir, India.
8. Dr. Pradeep Kumar Choudhury,
Assistant Professor,
Institute for Studies in Industrial Development,
An ICSSR Research Institute,
New Delhi- 110070, India.
9. Dr. Gyanendra Awasthi, M.Sc., Ph.D., NET
Associate Professor & HOD
Department of Biochemistry,
Dolphin (PG) Institute of Biomedical & Natural
Sciences,
Dehradun, Uttarakhand, India.
10. Dr. C. Satapathy,
Director,
Amity Humanity Foundation,
Amity Business School, Bhubaneswar,
Orissa, India.



ISSN (Online): 2455-7838

SJIF Impact Factor : 6.093

EPRA International Journal of

Research & Development (IJRD)

Monthly Peer Reviewed & Indexed
International Online Journal

Volume: 4, Issue:5, May 2019



Published By
EPRA Publishing

CC License





DATA PROCESSING USING MAP REDUCE

I.Karthika

Assistant Professor, Department of CSE , Bannari Amman Institute of Technology

G.Ranjith

Assistant Professor, Department of CSE , Bannari Amman Institute of Technology

S. Manisha Sree

PG Scholar, Department of CSE , Bannari Amman Institute of Technology

ABSTRACT

Objective: Increasing data processing speed by using Map-reduce concept. **Analysis:** Map-reduce concept is parallel processing, map-reduce takes only less time for task completion. **Findings:** In existing system , there is a complexity in coordinating threads. But in proposed system Apache hadoop's map-reduce is open source and parallel processing, it eliminate shared state completely. **Improvement:** There is no complexity in the coordination when processing large volume of data

1. INTRODUCTION

Map-reduce is a concept of parallel processing, it will take only less time for processing. Map -reduce is made along with yarn in hadoop version 2.yarn will take care sending corresponding processing API to storage unit. Map-reduce consist of partitioner and combiner for further processing. Map reduce programming model consist of driver code, mapper code and reducer code. Apache hive and Apache pig tools is writing its own map-reduce by using an y-smart tool.

2. EXISTING SYSTEM

Multi threading is the main way of doing of parallel programming .It will acts according to the threads. Major complexity in multithreading is co-

ordination between the threads to share data. There is a need of semaphores and locks with greater care otherwise it results in deadlock.

3.PROPOSED SYSTEM

Apache hadoop introduces the concept called map-reduce.it will run in parallel manner. It will distribute the tasks across multiple nodes. Map reduce process will take care by job tracker and task tracker. job tracker is master service and task tracker is a slave service. Each data node in the hadoop has task tracker in it.

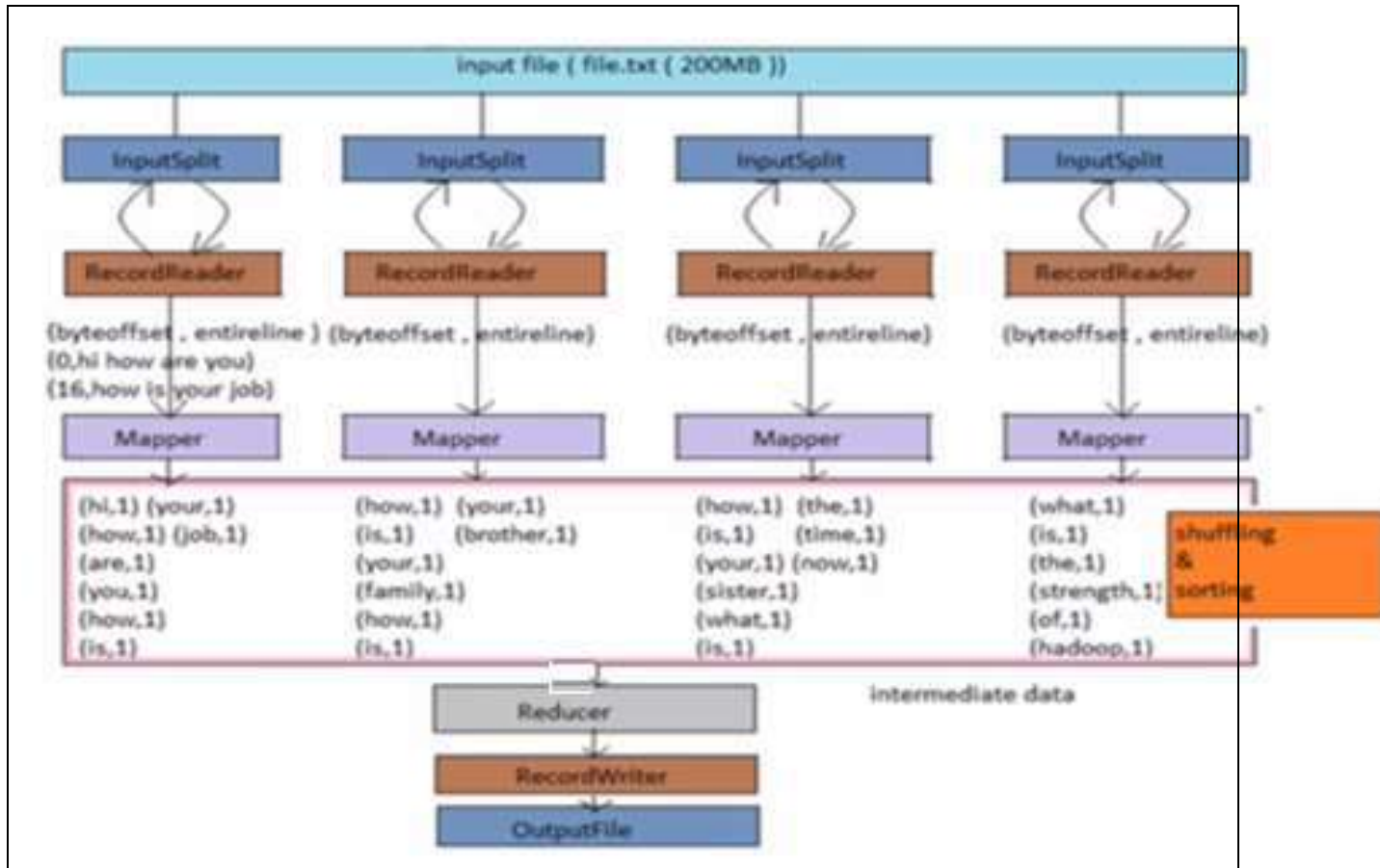


Fig 1 : Mapper Reducer

3.1.1 Input split:

Raw input file is divided into default 64 MB input splits as an meta data by the name node. This meta data is given as response to the client for storing the file in the data node. Every time heartbeat and block report send by the data node to name node.

3.1.2 Record reader:

This phase resides in between the mapper and reducer. mapper can understand only (key, value) pairs. Record reader will convert the text file to key, value pairs. At a time record reader can read only one line, (key,value) pairs segregated in the form (byte offset, Entire line).Byte offset is indexed of the text and value is an entire line. there is no need of external coding for record reader, by default it configured during installation.

3.1.3 Map phase:

Mapper gets record reader output as an input.Input splits given as an input to the Task tracker through job tracker. Input is mapped with an task tracker for processing. So number of input splits is

equal to the number of mapper. Mapper processing the (key, value) pairs and produce and an output also as an (key, value) pairs. Multiple mappers run in parallel manner. number of mapper output obtained is equal to the number of input splits.

3.1.4 Reduce phase:

Reduce phase is also only handle with an (key, value) pairs .It consists of shuffling and sorting phase. Shuffle phase is responsible for shuffling all inputs from mapper and sorting will sort the input.the output from the mapper is get and sorted in the reducer phase. Reducer output is also in the form of (key,valu) pairs.

3.1.5 Partitioner :

It is an intermediate Map (key,value) pairs output. By default hash partitioner will take of reducer phase. Partitioner will be used for dividing the data based on user conditions. Total number partitioner is equal to total number of reducer. partitioner is an optional, programmer can configure the partitioner based on the user condition.

3.1.6 Combiner:

It will act as an semi-reducer. It will accept the input from the map class and send that to reducer class. It will summarize the mapper output records

with same key. Main usage of combiner will reduce amount of data transfer between mapper and reducer. combiner is an optional, programmer can configure the combiner volume of data.

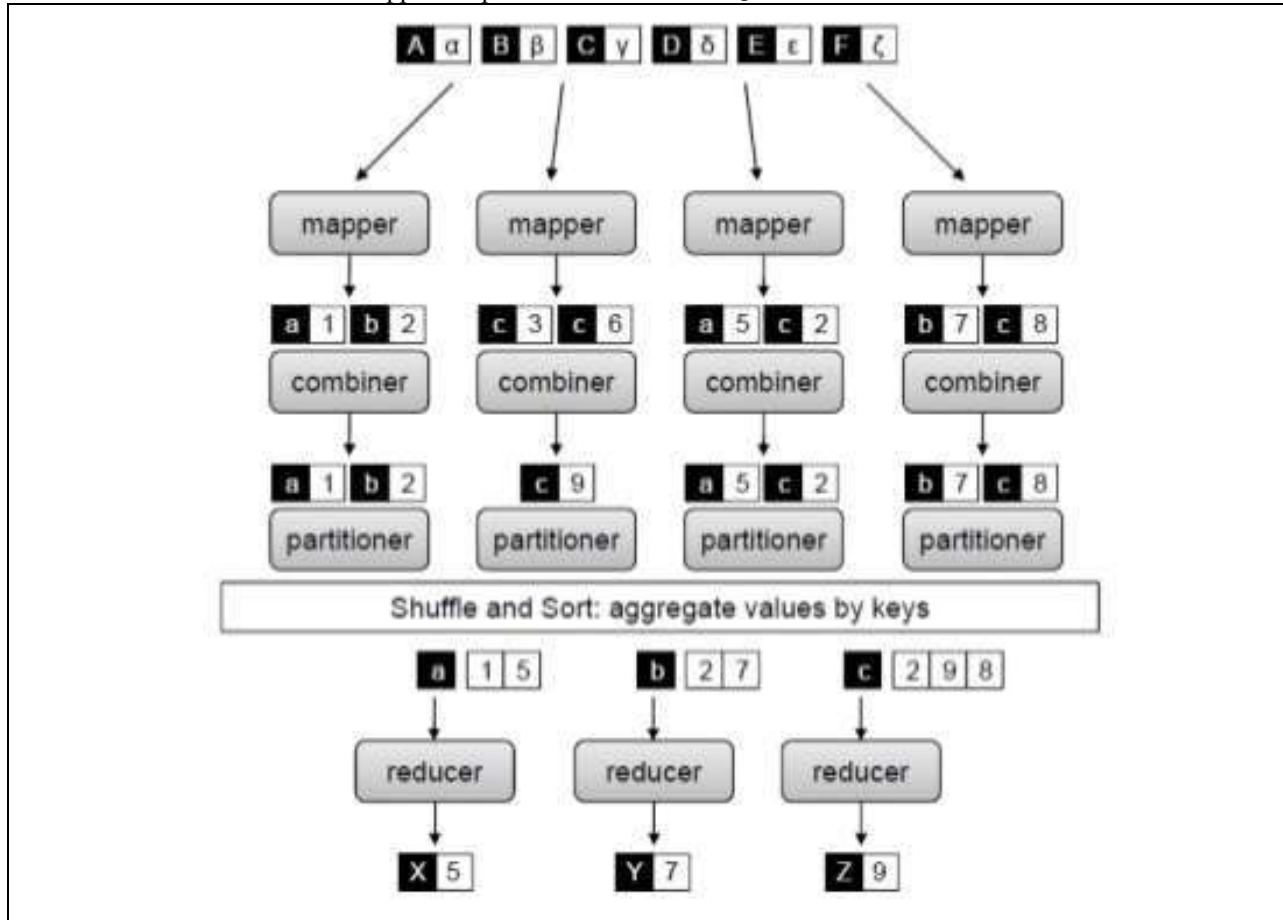


Fig 2 : Map -Reduce phase

3.1.6 Record writer:

It will get input from reducer and convert that from the (key, value) pairs to text. It is configured also by default during installation.

4.RESULT

Program has driver code. Driver code will be configured with mapper , reducer, input file format and output file directory.

\$hadoop jar testjar Drivercode file.txt TestOutput

5. CONCLUSION

Map-reduce process will give fast processing of data when compared to the multi threading concept. For Further fast process combiner and partitioner is used, it also increase accuracy. it also eliminates shared state complexity. Data get shared across the datanode and procesing with tasktracker.

6.REFERENCES

1. Anumol Johnson,Havinash P.H,Vince Paul,Mr.Sankaranarayanan P.N " Big Data Processing Using Hadoop MapReduce Programming Model" International Journal of Computer Science and Information Technologies, Vol. 6 (1) , 2015, 127-132
2. Map-Reduce: Simplified Data Processing on Large Clusters, by Jerey Dean and SanjayGhemawat,fromGoogle Research.
3. Thilagamani, S. and N. Shanthy, 2010. Literature survey on enhancing cluster quality. Int. J. Comput. Sci. Eng., 2: 1999-2002.
4. S. Chitra, B. Madhusudhanan, G. Sakthidharan, P. Saravanan, Local Minima Jump PSO for Workflow Scheduling in Cloud Computing Environments, Springer, ISBN 364241673X, 1225-1234, 2014.
5. E.T. Venkatesh , P. Thangaraj , and S. Chitra , " An Improved Neural Approach for Malignant and Normal Colon Tissue Classification from Oligonucleotide Arrays ," European J. Scientific Research , vol. 54 , pp. 159 - 164 , 2011

6. Hadoop Distributed File System <http://hadoop.apache.org/hdfs/> [3] Borthakur, D. (2007) *The Hadoop Distributed File System: Architecture*
7. S Saravanan, V Venkatachalam, "Advance Map Reduce Task Scheduling algorithm using mobile cloud multimedia services architecture" *IEEE Digital Explore*, pp21-25, 2014.
8. S Saravanan, V Venkatachalam, "Enhanced bosa for implementing map reduce task scheduling algorithm" *International Journal of Applied Engineering Research*, Vol 10(85), pp60-65, 2015.
9. I. Karthika, K. P. Porkodi, "FRAUD CLAIM DETECTION USING SPARK" *International Journal Of Innovations In Engineering Research And Technology* ISSN: 2394-3696 VOLUME 4, ISSUE 2, Feb.-2017
10. I. Karthika, K.P. Porkodi, "AUTOMATIC MONITORING AND CONTROLLING OF WEATHER CONDITION USING BIG DATA ANALYTICS", *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 6, Issue 1, January 2017
11. I. Karthika, P. Gokulraj, S. Saravanan "Prediction of sales using Big data analytics" *Journal Of Advances In Chemistry* Vol 12, No 20
12. I. Karthika, S. Priyadharshini "Survey on Location based sentiment analysis of Twitter data" © 2017 *IJEDR* | Volume 5, Issue 1 | ISSN: 2321-9939
13. I. Karthika, S. Priyadharshini "Overlapped Latent Fingerprint Matching Using Segmentation" *Journal of Chemical and Pharmaceutical Sciences* ISSN: 0974-2115
14. D. Uma Priya, I. Karthika "A GIS-BASED FRAMEWORK FOR CATASTROPHE MANAGEMENT", *International Journal of Modern Trends in Engineering and Science* Volum 3 ISSN no: 2348-3121 2016
15. I. Karthika "Safeguarding Scada Network by Captious State Based Filter", *International Journal of Innovative Research in Computer and Communication Engineering* Volume 2 Special Issue 1 -ISSN(Online): 2320-9801 2014