# Implementation of Map-Reduce based Image Processing Module

Under the Guidance : Prof. Danny J. Pereira(HOD)
Divya Gandhi, Dnyaneshwari Naik, Komal Waghe, Renu Ramekar

*B.E, Department of Computer, GCOEARA, Maharashtra, India*
*B.E, Department of Computer, GCOEARA, Maharashtra, India*
*B.E, Department of Computer, GCOEARA, Maharashtra, India*
*B.E, Department of Computer, GCOEARA, Maharashtra, India*

**ABSTRACT**

*As of late, there is a quick headway in web and the developing number of individuals utilizing long range interpersonal communication administrations (SNSs) have encouraged the sharing of sight and sound information like sharing of pictures, messages. Notwithstanding, interactive media information preparing including the strategies like as transcoding and transmoding force an extensive weight on the registering base as the measure of information increments. In this way, we propose another preparing apparatus called Map-Reduce-based picture change module in distributed using so as to compute environment Hadoop to minimize the weight of processing force. The proposed framework including two sections: a capacity framework, i.e., Hadoop Distributed File System (HDFS) for picture information and a Map-Reduce program.*

**Keyword :** - *Hadoop, Map-Reduce, HDFS, Cloud Environment, Client-server architecture, Parallel processing*.

## 1.INTRODUCTION

SNS and media content suppliers are always moving in the direction of giving mixed media rich encounters to end clients. In spite of the fact that the capacity to share sight and sound articles makes the Internet more appealing to customers, customers and hidden systems are not generally ready to stay aware of this developing interest. We perform two examinations to show the proposed module's greatness in transcoding capacity. In the primary trial, we contrast the proposed module and a non-Hadoop-construct single system running in light of two unique machines. What's more, we direct the execution assessment of the proposed module as indicated by the Java Virtual Machine (JVM) reuse alternative for the issue of numerous little documents.

### 1.1. HDFS

HDFS is the primary storage system used by Hadoop applications [5]. HDFS creates multiple replicas of data blocks and distributes them on computed nodes throughout a cluster to enable reliable and extremely rapid computations. HDFS has a master-slave structure uses the TCP/IP protocol to communicate with each node. Figure 1 shows the structure of HDFS.

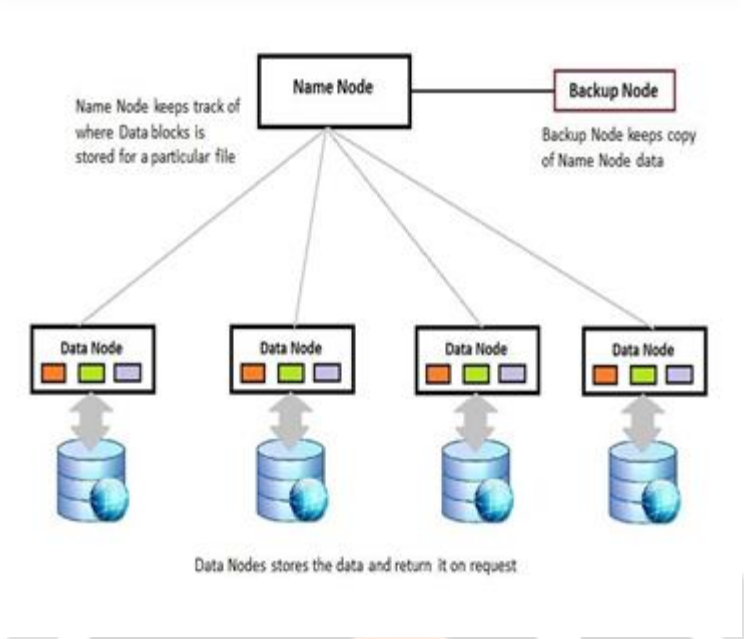


**Fig 1:** HDFS Flow.

As shown in Figure 2, NameNode manages the namespace and controls the file access by the client, and DataNode manages the storage of each node in the cluster. In addition, DataNode executes block commands issued by NameNode.

### 1.2. Map-Reduce

Map-Reduce is a programming model used for parallel processing of distributed large-scale data [6]. Map-Reduce processes an entire large -scale data set by dividing it among multiple servers. Figure 2 shows the structure of Map-Reduce.

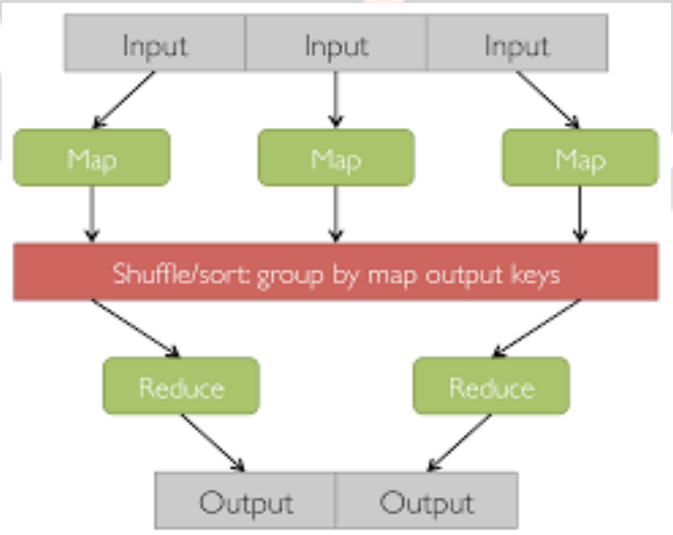


**Fig 2:** Map-Reduce Structure

Map-Reduce frameworks provide a specific programming model and a run-time system for processing and creating large amounts of datasets which is amenable to various real-world tasks [8]. Map-Reduce framework also handles automatic scheduling, communication, synchronization for processing huge datasets and it has the ability related with fault tolerance. Map-Reduce programming model is executed in two main steps, called mapping and reducing. Mapping and reducing are defined by mapper and reducer functions that are s data processing functions. Each phase has a list of key and values pairs as input and output. In the mapping , Map-Reduce input datasets and then feeds each

data element to the mapper in the form of key and value pairs. The output from mapper act as input to the reducer and thus reducer form final output

## 2.Image Conversion Architecture

Social Media Cloud Computing Service Model In this study, we designed and implemented a Map Reduce- based image conversion module in a cloud-computing environment to solve the problem of computing infrastructure overhead. Such overhead increases the burden on the Internet infrastructure owing to the increase in multimedia data shared through the Internet. The traditional approach of transcoding multimedia data usually involves general-purpose devices and offline-based processes. However, multimedia data processing is time consuming and requires large computing resources. To solve this problem, we designed an image conversing module that exploits the advantages of cloud computing. The proposed module can resize and convert images in a distributed and parallel manner. The proposed module use HDFS as storage for distributed parallel processing. The image data is distributed in HDFS. For distributed parallel processing, the proposed module uses the Hadoop Map-Reduce framework. In addition propose module uses the JAI library in Mapper for image resizing and conversion. Figure 3 shows the proposed module architecture.

## 3.Mathematical Modeling

S={s, e, X, Y, Fme, DD,NDD,θ}
Where,
s = Start of the program.
1. Log in with Hadoop.
2. Load Datasets .

e = End of the program.
To retrieve the useful pattern form information repositories.

X = Input of the program.
Images (Media/ Unstructured data)

Y = Output of the program.
Discovering meaningful new correlations, patterns and Image Conversion by sifting through large amounts of data stored in repositories.

X, Y ∈ U
Let U be the Set of System.
U= {Usr, DS, JAI, Clust}
Where
Usr, DS, JAI, Clust are the elements of the set.
Usr=User
DS= Image Dataset
JAI=Converting Image in Target Format
Clust=k-Mean Cluster

Fme = {F1, F2, F3, F4, F5}
Where,
F1= Input Datasets.
F2= Define the Cluster size.
F3= Find HACE classification.
F4= Apply K-Mean Cluster.
F5= Image Processing.
F6= Displaying the Conversion results.

DD= since Fme gives expected result, therefore problem is deterministic.

θ= Failures and Success conditions.
Failures:
1. Huge image database can lead to more time consumption to get the information.

2. Hardware failure.
3. Software failure.

Success:
1. Image Conversion and image extraction will perform from Datasets.
2. User gets result very fast according to their needs

## 4.System Architecture

In the architecture we have created a map-reduce based model for image processing. This model consist mapper and reducer which process image to and from HDFS. It is useful for big data analytic. Image related all problem are deals here. Its objective is to deal with tendentious amount of data reliably,error free and and with compatibly. Our Hadoop database including key ideas as mapper and reducer. Mapper work as change, filtering and conversion it mean it perform the accompanying capacities Record peruse, Partioner, mapper and Combiner. As same way the reducer including conglomeration and gathering. Reducer work on mean it incorporate the capacities as shuffling, sorting, reducer and yield design.

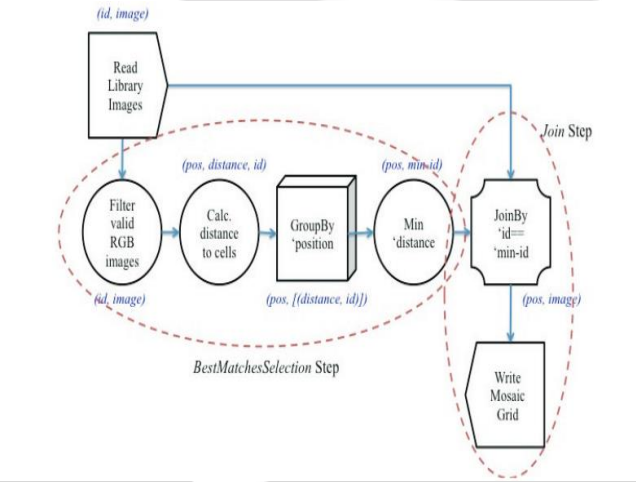


**Fig 3.System Flow**

## 5. Proposed System Definition

Execute a guide decrease based Image Processing and Image Conversion in Cloud Computing Environment Current ways to deal with preparing pictures rely on upon handling a little number of pictures having a successive handling nature. The interpersonal interaction application is intended to make social life more dynamic and fortifying. The informal organization can offer you some assistance with maintaining existing associations with individuals and offer pictures and messages, and set up new ones by contacting individuals you've never met. This application likewise gives the elements of Image transferring. The principle thought behind blogging is to impart your considerations to every one of your companions which can be perused by every one of the clients utilizing the application. This picture can be taken care of by the client as he needs. It additionally gives the components of including recordings. The administration of unstructured information is perceived as one of the major unsolved issues in the data innovation (IT) industry, the principle reason being that the apparatuses and procedures that have demonstrated so fruitful changing organized information into business insight and noteworthy data just don't work with regards to unstructured information. Unstructured information documents regularly incorporate content and mixed media content. Illustrations incorporate email messages, word preparing reports, recordings, photographs, sound documents, presentations, website pages and numerous different sorts of business archives. Note that while these sorts of documents might have an interior structure, they are still viewed as "unstructured" on the grounds that the information they contain doesn't fit perfectly in a database. Specialists appraise that 80 to 90 percent of the information in any association is unstructured. The measure of unstructured information in undertakings is becoming altogether regularly commonly speedier than organized databases are developing. New methodologies are fundamental. In this way, we need another parallel way to deal with work adequately on massed picture information. Keeping in mind the end goal to prepare countless adequately, we utilize the Hadoop HDFS to store a lot of remote detecting picture information, and we utilize Map-Reduce to handle these in parallel.

## 6. Goals and Objectives with Scope

The social networking application is designed to make social life more active and stimulating. The social network can help you maintain existing relationships with people and share pictures and messages, and establish new ones by reaching out to people you've never met before.
This application also provides the features of Image uploading. The main idea behind blogging is to share your thoughts with all your friends which can be read by all the users using the application. This image can be handled by the user as he wants. It also provides the features of adding videos.

### 6.1. Scope

The management of unstructured data is recognized as one of the major unsolved problems in the information technology (IT) industry, the main reason being that the tools and techniques that have proved so successful transforming structured data into business intelligence and actionable information simply don't work when it comes to unstructured data.
Unstructured data files often include text and multimedia content. Examples include e-mail messages, word processing documents, videos, photos, audio files, presentations, webpages and many other kinds of business documents. Note that while these sorts of files may have an internal structure, they are still considered "unstructured" because the data they contain doesn't fit neatly in a database. Experts estimate that 80 to 90 percent of the data in any organization is unstructured.
The amount of unstructured data in enterprises is growing significantly often many times faster than structured databases are growing.
New approaches are necessary. Therefore, we are in need of a new parallel approach to work effectively on massed image data. In order to process a large number of images effectively, we use the Hadoop HDFS to store a large amount of remote sensing image data, and we use MapReduce to process these in parallel.
The MapReduce programming model will be actively working with this distributed file system. It is these reasons that motivate the need for research with vision applications that take advantage of large sets of images.
Hadoop can process stores of both unstructured and structured data that are extremely large, very complex and changing rapidly.

## 7. Methodologies/Algorithm Details

### 7.1. Algorithm 1/Pseudo Code

#### 7.1.1. Gray scale image jpeg format

```
For I = 0 to height do
  For I = 0 to weight do
  {
  Conversion from gray-scale to png
  Print "conversion done"
  }
 End;
```

#### 7.1.2. Gray scale image png format

```
For I = 0 to height do
  For I = 0 to weight do
  {
  Conversion from gray-scale to png
  Print "conversion done"
  }
 End;
```

## 8. Conclusion

As of late, the wide accessibility of economical equipment, for example, PCs, advanced cameras, PDAs, and other simple to-utilize advances has empowered the normal client to fiddle with interactive media. The remarkable development of Internet innovations, for example, SNSs permits clients to scatter interactive media objects. Notwithstanding, the expanding measure of mixed media information forces an impressive weight on the web base required to process picture transformation capacity to give dependable administrations to various heterogeneous gadgets. Hence, we outlined and actualized a Map-Reduce-base picture transformation module in a distributed computing Environment. The proposed module depends on Hadoop HDFS and the Map-Reduce system for appropriated parallel preparing of expansive scale picture information. We overhauled and actualized Input Format and Output Format in the Map-Reduce system for picture information. We utilized the JAI library for changing over the picture design and resizing the pictures. We misused the benefits of distributed computing to handle mixed media information preparing. We performed two investigations to assess the proposed module. In the principal examination, we contrasted the proposed module and a non-Hadoop-based single project utilizing the JAI library. The proposed module indicates preferred execution over the single project in the second trial.

## 9. Acknowledgements

It gives us great pleasure in presenting the preliminary project report on 'Implement map-reduce based Image Processing module in Cloud Computing Environment by using hadoop'. We would like to take this opportunity to thank our HOD and internal guide Prof. Danny J. Pereira, Head of Computer Engineering Department, Government College of Engineering and Research, Avasari(Kd) for giving us all the help and guidance we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful. We would also like to thank our faculty teachers for extending their warm support towards our project. In the end, our special thanks to our Lab Assistants for providing various resources such as laboratory with all needed software platforms, continuous Internet connection, for our project.

## 10. References

[1] Sun-Moo Kang, Bu-Ihl Kim, Hyun-Sok Lee, Young-so Cho, Jae-Sup Lee, Byeong-Nam Yoon, "A study on a public multimedia sevice provisioning architecture for enterprise networks", Network Operations and Management Symposium, 1998, NOMS 98., IEEE, 15-20 Feb 1998, 44-48 vol.1, ISBN : 0 -7803-4351-4
[2] Hari Kalva, Aleksandar Colic, Garcia, Borko Furht, "Parallel programming for multimedia applications", MULTIMEDIA TOOLS AND APPLICATIOS, volume 51, number 2, 901-818, DOI: 10.1007/s11042-010-o656-2
[3] Gracia, A., Kalva, H., "Cloud transcoding for mobile video content delivery", Consmer Electronics(ICCE), 2011 IEEE International Conference on, 9-12 Jan. 2011, 379-380, ISSN : 2158-3994
[4] http://www.cloudera.com/blog/2009/02/the-small-files-problem/
[5] Hadoop Distributed File System : hadoop.apache.org/hdfs/
[6] Jeffrey Dean, Sanjay Ghemawat, "MapReduce : Simplified Data Processing on large Cluster", OSDI`04 : Sixth Symposium on Operating System Design and Implementation, San Francisco, CA, December, 2004.
[7] Java Advanced Imaging Library : java.sun.com/javase/technologies/desktop/media/jai/
[8] Shivnath Babu, "Towards Automatic Optimization of MapReduce Programs", The 1st ACM symposium on Cloud Computing, 2010.