# Implementation of efficient Clouds using MapReduce

**Rupali D. Korde** **Dr. V. M. Thakare**

***Abstract* — Today, Cloud Computing is having large attention because of provision of configurable computing resources. MapReduce (MR) is one of the popular and widely used frameworks for data intensive distributed computing of batch job. Now a day, Cloud Map Reduce (CMR) is widely used because it is more efficient & run faster than other implementations of the MR framework. Running MapReduce programs in the cloud has big problem of optimization of resource provisioning to reduce the economic cost or job finish time for a specific job, power management and performance. BStream is cloud bursting framework for MapReduce that combine stream processing in the external cloud (EC) with Hadoop in the internal cloud (IC). Stream processing in EC allows pipelined uploading, processing and downloading of data to minimize network latency. To make these systems manageable and scalable, an important research problem is performance. As MapReduce framework works on large datasets which contains some form of information and computation. There are many translators are available to transfer SQL query to MapReduce program but translator for Matlab language is at initial stage. Matlab language is essential for mathematical programming .This paper propose the Advance Translator that handles all types of Matlab Commands.**

***Keywords* — cloud, mapreduce, M2M convertor.**

## I. INTRODUCTION

Cloud MapReduce (CMR) is an implementation of MapReduce framework on Amazon Web Services. With the help of queues, CMR can parallelize the Map and the Shuffle stages. Amazon's visibility timeout mechanism is use to implement fault-tolerance. CMR is totally distributed architecture, and there is no single point of failure and scalability bottleneck. BStream is newly cloud bursting framework to address many difficulties. It uses stream processing engine called Storm3 in External Cloud and YARN in Internal Cloud.map and reduce operations process on the incoming stream of data as and when it arrives in EC Using Storm. BStream uses a analytical model to gauge which portions of MapReduce job to burst, when to burst job and when to start the reducers, to meet job deadline. MapReduce, and its open source implementation, Hadoop is widely used to support large computation tasks over big datasets to improve system performance and scalability. For example, Facebook build a system infrastructure on Hadoop to handle large volume of requests from Internet. To help these methods and to further improve the cloud performance, these paper introduce the translator for matlab language and reduce the execution time for mathematical task.

## II. BACKGROUND

MapReduce is a popular parallel programming model, first introduced by Google, which is designed to handle and generate large scale data sets in distributed database environment. It provides efficient way to parallelize data and analysis process. An advantage includes robustness, scalability, conveniences. The basic idea of MapReduce is to break the large input data set into many small chunks and assigned small task for each chunk to different devices. The process of MapReduce consists of two major parts, first is Map function and second is Reduce function. The input data files will be automatically split and copied to different computing nodes for processing. After that, the inputs will be transferred to Map function in (key-value) format. The Map functions processes the input pairs and generate intermediate (key-value) pairs which provides as inputs to Reduce function. The Reduce function then combines the inputs that have the same keys and produce the final output. The final output then is written into the distributed file system.

This MapReduce framework when used in Cloud Environment, it surely improves its performance. Hence it is necessary to develop such methods which works well in cloud environment using MapReduce Framework and develop an efficient cloud for use. There are various technology are exists such as CMR, CRESP, M2M, BStream etc. which work for cloud environment and this paper also propose new translator to improve result of Mapreduce program.

## III. PREVIOUS WORK DONE

Author Devendra Dahiphale, et al[1] works on Advance Mapreduce: Cloud MapReduce, enhancement and Application, which introduce new method named as CMR for efficient cloud.

Author Keke Chen, et al.[2] proposed method CRESP: Towards Optimal Resource Provisioning for MapReduce Computing in Public Clouds.

Author Junbo Zhang, et al.[3] proposed method M2M: A Simple Matlab-to MapReduce Translator for Cloud Computing, to help traditional programmers simply deploy an application to cloud systems through by translating sequential codes to MapReduce codes.

Author Nan Zhu, et al.[4] works on "Towards a Cost-Efficient MapReduce: Mitigating Power Peaks for Hadoop Clusters" for improving the performance of cluster using Mapreduce.

Author SriramKailasam, et al.[5] works on Extending MapReduce across Clouds with BStream. This BStream system works on incoming data as and when entered in system.

## IV. EXISTING METHODOLOGY

There are many technologies exists in Cloud environment for processing large data sets. With the help of MapReduce, method like Cloud MapReduce (CMR)[1] which Support stream data processing in addition to batch data by parallelizing the Map and Reduce phases through a pipelining model.

BStream is another in cloud bursting framework which addresses many difficulties. For mitigating power peak problem PowerMumak simulator is used.

Cloud RESource Provisining (CRESP) method [2] is used for optimization of resources in cloud.
matlab-to-mapreduce method[3] is used to translate matlab code to mapreduce code for processing in mapreduce framework

## V. Analysis And Discussion

### A. Analysis on Existing methodology:

The methods which are highlighted above are used in cloud environment for different purposes, which have some advantages Disadvantages. They improves the performance of system using their specialize features. Method like CRESP is used for optimization of resource but quality of model is not improved. In BStream user needs to write Hadoop and Storm code for job. CMR uses the network for Input Output and for local storage. It is also necessary that workload on each nodes in cloud is below threshold value.M2M method is still in its early stage and it supports only some basic Matlab commands. It does not support visualization and complex matrix operations.

### B. Attributes and Parameters Consider:

Input parameters:-
J - job-profile type
M- number of map tasks within a job
R- number of reduce tasks within a job
T - deadline/completion time of job
bEC- inter-cloud bandwidth

Output parameters:-
Ts-duration of shuffle phase
Tr- duration of reduce phase
Tbs- time at which shuffle starts
Tbr-slack time for burst portion to return
Tst- start time for bursting tasks
A -number of slots allocated in IC
MIC -number of maps executed in IC
MEC -number of maps burst to EC
Cm- number of maps completed before start of shuffle phase.

### C. Effects of outcomes and various attributes & parameters:

Using this attributes and parameters, model provides resource allocation plan to meet job deadline by calculating equations. It estimates parameters as no. of resources to be provisioned in IC or EC, no. of maps to burst, when to burst map, and when to start reducers in Internal Cloud. Using parameter A, model estimates allocation for submitted job to meet deadline.

### D. Trends of improvement:

Following list include trends of improvement and performance.

- Fault tolerance
- Improve quality model
- Power consumption & Workload scheduling
- Reduce latency & duplicate message

### E. Comparison and Drawback:

Batch implementation of Hadoop in a multi-cloud environment incurred huge startup latencies for map computation in External Cloud. However, if maps in EC can Stream reads from Hadoop Distributed File System in IC, then the startup latencies can be reduced. But there is one difficulty in scaling Hadoop MapReduce using cloud bursting framework is that it requires entire data before processing and so on.

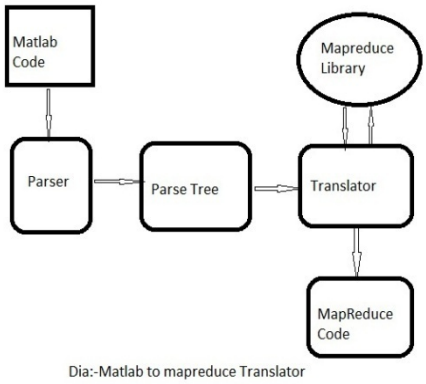
## VI. Proposed Methodology

As there are many high level languages are used in programming and it necessary to understand this languages before using this. MapReduce Framework is used in Cloud system and to make this cloud system more efficient, it is necessary that it can process all types of data entered in system. There are many translators are available to convert SQL query to Mapreduce Program. For processing numerical operation Matlab language is use. There is only one translator is used to convert matlab code to MapReduce code and this is in its initial stage. It is necessary to create translator that supports full functionality of Matlab language and improves the performance of mathematical task in mapreduce.

It is necessary to create mathematical library and other commands of Matlab language based on MapReduce framework to improve the performance of

Mathematical task. This is the proposed method of this paper.

Whenever any Matlab Program wants to execute in Mapreduce framework, it first goes through parser where tokens are generated and scan. Parse tree is given to the translator which translates Matlab code to MapReduce code with the help of Mapreduce library. MapReduce library contains all the Math operation based on MapReduce command. By using this library Appropriate Code is generated.




Dia:-Matlab to mapreduce Translator

## VII. POSSIBLE OUTCOMES AND RESULT

By creating such translator for MapReduce framework, the performance of mathematical task is improves in cloud system. This also reduces the programmer's time to learn Mapreduce

language before using it. It is also possible that making MapReduce Framework so that it can support all types of programming language. Conducting more experiments on different MapReduce programs to improve performance of cloud system.

## CONCLUSION

This paper evaluates the performance of cloud using Mapreduce. It is possible to create efficient cloud using MapReduce. There are many methods are existing to improve the performance by providing parallelism between different task and reduce time to execute it. This paper propose advance translator that is suitable for mathematical task and it improves the performance. Performance improvements are significant, reliable and promising. Cloud Computing is fast gaining attention in the industry, and the methods will add significant benefit to applications.

## REFERENCES

1] DevendraDahiphale, RutvikKarve, Huan Liu, ZhiweiYu,AmitChhajer, Jianmin Wang, and ChaokunWang,"An Advanced MapReduce: cloud MapReduce, Enhancements and Applications", IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, VOL. 11, NO. 1, MARCH 2014.

2] Keke Chen, James Powers, ShuminGuo, and FengguangTian," CRESP: Towards Optimal Resource Provisioning for MapReduceComputing in Public Clouds", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 6, JUNE 2014.

3] Junbo Zhang, Dong Xiang, Tianrui Li, and Yi Pan," M2M: A Simple Matlab-to-MapReduce Translator for Cloud
Computing", IS SN 1 0 07 - 0 2 14 0 1 / 1 2 p p 1- 9 Volume 18, Number 1, February 2013.

4] Nan Zhu, Xue Liu, Jie Liu, and Yu Hua,"Towards a Cost-Efficient MapReduce: Mitigating Power Peaks for
Hadoop Clusters**",**ISSN 1007-0214 03/10 pp24-32 Volume 19, Number 1, February 2014.

5] SriramKailasam, PrateekDhawalia, S. J. Balaji, GeetaIyer," Extending MapReduce across Clouds with BStream" IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 2, NO. 3, JULY-SEPTEMBER 2014.

6] Pedro Costa, Marcelo Pasin, AlyssonNevesBessani, and Miguel P. Correia," On the Performance of Byzantine Fault-Tolerant MapReduce", IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 10, NO. 5, SEPTEMBER/OCTOBER 2013.

## AUTHOR'S PROFILE

**Rupali D. Korde** has completed B.E. Degree in Computer Science and Engineering from Sant Gadge Baba Amravati University, Amravati, Maharashtra. She is persuing Master's Degree in Computer Science and Information Technology from P.G. Department of Computer Science and Engineering, S.G.B.A.U. Amravati. (e-mailid: rplkorde@gmail.com)

**Dr. V. M. Thakare**
Dr. Vilas M. Thakare is Professor and Head in Post Graduate department of Computer Science and engg, Faculty of Engineering & Technology, SGB Amravati university, Amravati. He is also working as a coordinator on UGC sponsored scheme of e-learning and m-learning specially designed for teaching and research. He is Ph.D. in Computer Science/Engg and completed M.E. in year 1989 and graduated in 1984-85. He has exhibited meritorious performance in his studentship. He has more than 27 years of experience in teaching and research. Throughout his teaching career he has taught more than 50 subjects at various UG and PG level courses. He has done his PhD in area of robotics, AI and computer architecture. 5 candidates have completed PhD under his supervision and more than 8 are perusing the PhD at national and international level. His area of research is Computer Architectures, AI and IT. He has completed one UGC research project on "Development of ES for control of 4 legged robot device model.". One UGC research project is ongoing under innovative scheme. At PG level also he has guided more than 300 projects/discretion. He has published more than 150 papers in International & National level Journals and also International Conferences and National level Conferences. He has also successfully completed the Software Development & Computerization of Finance, Library, Exam, Admission Process, and Revaluation Process of Amravati University. He has also worked as member of Academic Council, selection Committee member of various Other University and parent university, Member of faculty of Engineering & Science, BOS (Comp. Sci.), Member of IT Committee, Member of Networking Committee, Member of UGC, AICTE, NAAC, BUTR, ASU, DRC, RRC, SEC, CAS, NSD etc committees. . He has also worked chairman of many committees like BOS, Monitoring and Control. 3 times he has received National Level excellent paper award at National Conference, Gwalior and at other places. He has also received UGC fellowship and a major UGC project. (e-mailid: vilthakare@yahoo.co.in)