

*Review Article*

# A Review of X-to-MapReduce Translator for Evaluation of various Commands on a Cloud

Neha N. More<sup>†\*</sup> and V.B. Gadicha<sup>†</sup>

<sup>†</sup>Computer Science and Engineering, P.R. Patil COET, Amravati, Maharashtra, India

Accepted 21 March 2015, Available online 25 March 2015, Vol.5, No.2 (April 2015)

## Abstract

*In recent times, Cloud Computing is having large attention due to its necessity for configurable computing resources. MapReduce (MR) is the one of the a popular structure for data-intensive spread computing of batch jobs also for equivalent programming model for cloud computing platform and has develop into the proficient method for dispensation massive data via cluster of computers. Running MapReduce programs in cloud has a big problem of optimization of resource provisioning to decrease the economic cost or job finish time for a specific jobs, power management and performance. As MapReduce framework works on huge datasets which contains various form of information and computation, X-to-MapReduce translator provides a feasible solution to help traditional programmers simply deploy an application on cloud system through translating sequential codes to MapReduce codes. In Recent years, various SQL to-MapReduce translators appear to translate similar like-SQL queries to MapReduce codes also have a superior performance on cloud systems computation. Matlab is an extremely developed language and provides interactive environment for arithmetical computation, visualization, and programming. We propose and develop a simple X-to-MapReduce translator for cloud computing, for basic numerical computations. It translates a Matlab code with up to 100 commands to MapReduce code in only some seconds, which may, In addition, X-to-MapReduce can also identify the dependency among complex commands, which is all the time confusing during hand over coding. We implement and evaluate Matlab commands on a cloud .This paper propose the Advance Translator that handles all types of Matlab Commands.*

**Keywords:** MapReduce, Cloud Computing, Translator, Matlab, Cluster

## 1. Introduction

Nowadays, the number of data is increasing at an exceptional rate. For the function of processing in huge amounts of data, Google introduced a software framework called Map Reduce, to sustain big distributed applications using a huge amount of computers, collectively known as a cluster. Since cloud computing is one of the most important technique It introduces Map Reduce has becoming a popular computing model for cloud computing platform and is helpful to examine large amounts of data. By Google's work, many Map Reduce runtime systems have rapidly emerged. Nowadays, the volume of data is rising at an exceptional rate.

Cloud Computing is having large attention because of provision of configurable computing resources. Map Reduce (MR) is one of the popular and widely used frameworks for data intensive distributed computing of batch job.

Cloud Map Reduce (CMR) is widely used because it is more capable & run rapidly than other implementations of the MR framework.

BStream is cloud bursting framework for Map Reduce that combine stream processing in

The external cloud (EC) with Hadoop in the internal cloud (IC), it uses stream processing engine called Storm3 in External Cloud and YARN in Internal Cloud.map and reduce operations process on the arriving flow of data and it arrives in EC Using Storm.

As Map Reduce framework works on large datasets which contains some form of information and computation.

In order to improve system performance scalability MapReduce, as well as its open source implementation, In order to deliver high performance, large-scale data centers have to deal with power-related problems due to large power consumption and electricity bills. In order to deliver high performance, Map and Reduce (some jobs only contain Map phase) is used. The developers have to study about how to program by using the MapReduce model and spend some time on perceptive characteristics of different

\*Corresponding author Neha N. More is a Student and V.B. Gadicha is working as HOD

cloud platforms. Habitual translation intended for certain programming languages is a possibility particularly for various data similar to application programming languages such as SQL and Matlab. The MapReduce skeleton requires that users execute their applications by coding their individual map and reduce functions.

There are many translators available to transfer SQL query to Map Reduce 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. To help this method and to further improve the cloud performance, this paper introduces the translator for matlab language and reduces the execution time for mathematical task.

APIs with high-level programming languages, like Java, Python (S.Babu,2010) or more expressive language like HiveQL (SQL-Like)( JunboZhang, Dong Xiang, TianruiLi, and Yi Pan, *et al*, 2013). The user specific computation logic octaves in the Map() and Reduce() functions of Map and Reduce phase respectively.

## 2. Literature Review

MapReduce is a famous parallel programming model (JunboZhang, Dong Xiang, TianruiLi, and Yi Pan, *et al*, 2013), original introduced by Google, which is developed to handle and generate big scale data sets in distributed database environment. It provides efficient way to parallelize data and analysis process. (JunboZhang, Dong Xiang, TianruiLi, and Yi Pan, *et al*, 2013) to help traditional programmers simply deploy an application to cloud systems through by translating sequential codes to MapReduce codes. Advantages include robustness, scalability, conveniences.

The basic idea of MapReduce is to break the big input data set into various small chunks and assigned small task for each chunk to different devices. The process of MapReduce (Zhu, XueLiu, JieLiu, and YuHua, 2014) consists of two major parts, first is Map function and second is Reduce function. The input data records will be automatically divide and copied to various 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. (Nan Zhu, XueLiu, JieLiu, and YuHua, 2014) for improving the performance of cluster using Mapreduce.

This MapReduce framework when used in Cloud Environment, it surely improves its performance. Hence it is necessary to developed 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 (KekeChen, James Powers, ShuminGuo, and

FengguangTian, *et al*, 2014),( JunboZhang, Dong Xiang, TianruiLi, and Yi Pan ,*et al*, 2013),( SriramKailasam, PrateekDhawalialia, S. J. Balaji, GeetaIyer, *et al*,2014). This BStream system works on incoming data as and when entered in system etc. which work for cloud environment and this paper also propose new translator to improve result of Mapreduce program.

### 2.1 Existing Technologies

There are many technologies exists in Cloud environment for processing large data sets. With the help of Map Reduce, method like Cloud MapReduce (CMR) (Devendra Dahiphale, Rutvik Karve, Huan Liu, Zhiwei Yu, Amit Chhajer, Jianmin Wang, and Chaokun Wang, *et al*, 2014), which introduce new method named as CMR for efficient cloud which supports flow of data processing in adding to batch data by parallelizing the Map and Reduce phase through a pipelining model.

BStream is another in cloud bursting framework which addresses many difficulties for mitigating power peak problem Power Mumak simulator is used. (SriramKailasam, PrateekDhawalialia, S. J. Balaji, GeetaIyer, 2014). This BStream system works on incoming data as and when entered in system

Cloud Resource Provisioning (CRESP) method (Zhu, XueLiu, JieLiu, and YuHua, 2014) is used for optimization of resources in cloud. Matlab-to-Map Reduce method is used to translate Matlab code to MapReduce code for processing in MapReduce Framework. (JunboZhang, Dong Xiang, TianruiLi, and Yi Pan, *et al*, 2013), to help traditional programmers simply deploy an application to cloud systems through by translating sequential codes to MapReduce codes.

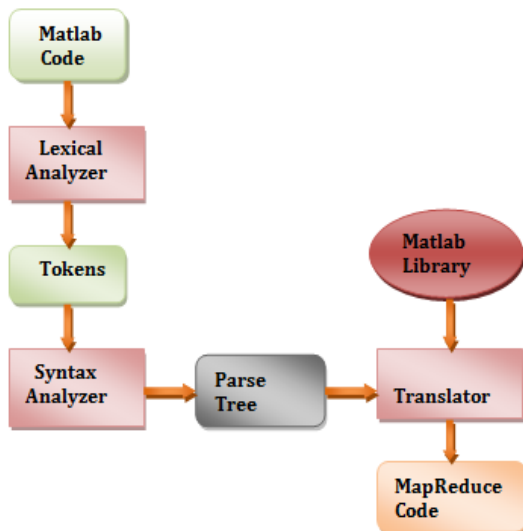
### 2.2 Shortcomings on Existing Technologies

The methods which are highlighted above are used in cloud environment for different purposes, which have some advantages, Disadvantages. Method like CRESP is used for optimization of resource but quality of model is not improved. In BStream customer needs to note down 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.

## 3. 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.



Dig- Matlab to MapReduce Translator

**Here, this flowchart giving the detailed explanation of our proposed system which is explained below by using following steps.**

(1) The initial stage is a lexical analyzer, used for token creation. The scanner begin the analysis of Matlab source code by understanding the input character by character, and assembling characters into meaningful words and tokens define by a grammar of standard expressions.

(2) The next phase is a syntax analyzer, designed for parsing. The syntax analyzers read tokens and group them into units as specific for the production of the CFG being used. Here, downward parsing is adopted in this method.

(3) The third phase is a semantic analyzer, designed for semantic parsing or analysis. It's functioning out the implication of the expression immediately validated and taking the appropriate action. At this point, focusing at the Matlab operations, an easy Math Operation Library based on MapReduce (Matlab library) is build. When the translator processed a math command, it connects to the Library, gets the corresponding MapReduce code, and combines with a main function. Its meaning notices that the similar commands only need one MapReduce code. 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.

### 3.1 Multiple independent commands to MapReduce

A real application always contains multiple commands. Matlab code by using the command min.

```
x = load ('in. data') //load data from the file 'in. data'
a = min(x).
```

### Conclusions

1] Current research work, evaluate the performance of cloud using MapReduce.

2] 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.

3] Current research work propose advance translator that is suitable for mathematical task and it improves the performance.

4] 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

- Devendra Dahiphale, Rutvik Karve, Huan Liu, Zhiwei Yu, Amit Chhajjer, Jianmin Wang, and Chaokun Wang (March 2014), An Advance MapReduce: cloud MapReduce Enhancements and Applications, *IEEE transactions on network and service management*, vol. 11, No. 1, 2014.
- KekeChen, James Powers, ShuminGuo, and FengguangTian (June 2014) ,CRESP: Towards Optimal Resource Provisioning for MapReduce Computing in Public Clouds, *IEEE transactions on parallel and distributed systems*, vol. 25, No. 6 2014.
- JunboZhang, Dong Xiang, TianruiLi, and Yi Pan,(2013), M2M: A Simple Matlab-to-MapReduce Translator for Cloud Computing, *Tsinghua Science and Technology* Volume 18, Number 1, February 2013, pp 1- 9.
- Nan Zhu, XueLiu, JieLiu, and YuHua.(2014) ,Towards a Cost-Efficient MapReduce: Mitigating Power Peaks for Hadoop Clusters, Volume 19, Number 1, February 2014, pp24-32.
- SriramKailasam, PrateekDhawalia, S. J. Balaji, GeetalYer(July-September 2014), Extending MapReduce across Clouds with BStream, *IEEE transactions on cloud computing*, vol. 2, No. 3.
- Pedrocosta,MarceloPasin,AlyssonNevesBessani,and Miguel P. Correia(2013), On the Performance of Byzantine Fault Tolerance MapReduce.
- J. Ekanayake, H. Li, B. Zhang, T. Gunarathe, S.-H. Bae, J. Qiu, and G. Fox (2010), Twister: A runtime for iterative mapreduce, *in Proc. of the 19th ACM Int. Symposium on High Performance Distributed Computing*, New York, NY, USA: ACM, pp. 810-8
- R. Lee, T. Luo, Y. Huai, F. Wang, Y. He, and X. Zhang (2011), Ysmart: Yet another sql-to-MapReduce translator, *in Distributed Computing Systems (ICDCS)*, 31st Int. Conf. on, June 2011, pp. 25-
- A. Gilat(2011), *MATLAB: An Introduction with Applications*, 4th ed. John Wiley & Son
- K. Kc and K. Anyanwu (2010), Scheduling hadoop jobs to meet deadlines, *in Cloud Computing Technology and Science (CloudCom)*, 2010 IEEE Second Int. Conf. on, Dec. 2010, pp. 388 3
- S. Agarwal, S. Kandula, N. Bruno, M.-C. Wu, I. Stoica, and J. Zhou(2012), Re-Optimizing Data-Parallel Computing, *in Proc. 9th USENIX Conf. NSDI*, p. 2
- S.Babu(2010), Towards Automatic Optimization of MapReduce Programs, *in Proc. 1st ACM Symp. Cloud Computing*, pp. 137-
- J. Dean and S. Ghemawat (2014), MapReduce: Simplified Data Processing on Large Clusters, *in Proc. OSDI*, pp.137-17
- Apache software foundation(2013), Mumak: MapReduce simulator. [https:// issues. apache.org/jira browse/ MAPREDUCE-728,2013](https://issues.apache.org/jira/browse/MAPREDUCE-728,2013).