ISSN:0975-9646
Gurjit Singh Sra et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012,3360 - 3364

Performance Evaluation of Globus Grid Environment Using Avi to Flash Video Encoding

Gurjit Singh Sra[#1], Dinesh Kumar[*2]
[#]*Dept. of Computer Sc. & Engg., Igce Abhipur, Mohali (Pb.), India*
[1]gurjit013@gmail.com
[*]*Dept. of Computer Sc. & Engg., Daviet Jalandhar (Pb.), India*
[2]erdineshk@gmail.com

Abstract- Grid is a collection of heterogeneous systems that are geographically distributed over a network. Grid systems can be used for solving large-sale computational problems such as video encoding, protein folding, etc. Video encoding is conversion of video from one format into another. The encoding process requires large computational power and takes a lot of time on conventional systems. Video file can be easily partitioned into smaller parts and these parts can be rejoined together to form full video file. This factor makes video encoding feasible on Grid. In this paper, video encoding from AVI to FLV video format is attempted on Grid environment and the performance of Grid is evaluated on basis of job execution time. Grid environment is implemented on virtual desktops connected through LAN. It is observed that AVI to FLV conversion on grid takes less time as compared to on a single system. The performance of grid system is 40% better than the conventional system. Further it is observed that the performance depends on problem at hand and other factors such as local processing load and network bandwidth.

Key words - Grid, Globus, Video Encoding, FLV, AVI, Virtual Machine.

I. Introduction

The advancements in communication and computing technologies have lead to the aggregation of diverse, distributed, storage and computing resources to form systems that are commonly known as Grid[1]. Grid Computing is an infrastructure of heterogeneous computing resources that work in collaboration to complete a task. The computational resources may or may not be connected through high speed networks. With the increasing computation power and interconnection between these resources, the user can view these distributed resources as a single virtual computing resource for solving large–scale problems[13]. A large number of Grid systems have been built to create and monitor grid environment for execution of large-scale problems. The most common Grid systems built are based on Globus toolkit. Globus provides software infrastructure that enables applications to handle distributed heterogeneous computing resources as a single virtual machine[16]. This toolkit is based on Linux based systems. It consist of resource discovery, security, data management, execution management, fault detection, portability[2][10] etc.

Grids computing offer a way to solve Grand Challenge problems such as protein folding, financial modelling, earthquake simulation, and climate/weather modelling [10]. Various other project and application areas of Grid computing include CERN Large Hadrons Collider[9], Distributed aircraft engine diagnostics, Worldwide telescope, Biomedical informatics research network[3][15]. Other project areas based on grid computing are POEM @Home, Climateprediction.net, SZTAKI Desktop Grid, MilkyWay@Home, ZetaGrid[9][10].

A growing percentage of the world population uses presently image, video and audio compression technologies on a rather regular basis. These technologies are behind the success and quick deployment of services and products such as digital pictures, digital television, DVDs, and MP3 players. The main objective of digital video encoding technologies is to represent the original data into a video format that is compatible with the user applications used for viewing the video, while preserving an acceptable video quality[6].Video encoding plays a vital role in the spread and distribution of multimedia over internet. Most of the videos already available are in AVI format. Flash is most common and widely used method for playing videos on the web. With more and more websites embedding flash player in their web pages, it is desired that the existing video content be converted into flv format. Flash Video has been accepted as the default online video format by many sites. Notable users of it include YouTube, Hulu, VEVO, Yahoo! Video, metacafe, Reuters.com, and many other news providers[7].

II. Related Work

In past few years Grid computing has been successfully used to solve large number of computational problems[9]. Currently, it is a great potential technology that leads to the effective utilization of the computing resources. Meanwhile, the complex problems such as scientific, engineering and business need to utilize the huge resources that are available in vast heterogeneous computational environment. Therefore, grid computing is considered as an ultimate solution to solve these problems[4][9]. Grids are extensively being used to carry out compute and storage intensive tasks, so there is a need to evaluate the performance of grids with respect to different parameters like processing time, resource utilization, memory statistics, network statistics etc., so that user of the grid can make easy decisions regarding the configuration of the environment they have to build or use for the execution of their grid jobs[5]. Various applications areas such as medical applications are built upon grid infrastructure. The privacy and security of sensitive data in such cases can be implemented using trusted data storage[3]. Further the type of scheduling strategies adopted plays a vital role in the performance of grid system. Various scheduling and dispatching rules such as First Come First Served (FCFS), Earliest Due Date (EDD), Earliest Release Date (ERD), and an Ant Colony Optimization (ACO) can be applied. The performance of FCFS is better than others as far as total scheduling time is concerned[4].

Digital video can exist as any number of formats, using any number of settings. When a digital video file does not meet the specifications, or the file type is bad for the intended use, it must be converted to the proper format, using video encoder hardware or software. Video compression and conversion techniques have been evolving from last 20 years and there are 50% compression gains for every 5 years[6]. Parallel implementation of video encoding systems using either the software or the hardware approach has been attempted in the area of real time video encoding. Hardware based parallel video encoder[8] is implemented on bus network which makes use of Divisible load theory (DLT) paradigm for strip-wise load partitioning, balancing, load distribution. Strategies are developed to exploit the data parallelism inherent in the video encoding process. In this, both the granularity of the load partitions and all the associated overheads caused during parallel video encoding process can be explicitly considered. This significantly contributes to the minimization of the overall processing time of the video encoder[8].

III. Proposed Work

The Performance of grid is evaluated by execution of compute intensive, storage intensive or network intensive tasks. There are varieties of compute intensive problems that can be executed on grid systems. E.g. drug discovery, seismic data analysis, weather forecasting video encoding, image processing etc. The exact performance of grid is dependent upon the problem undertaken, the configuration of the environment, resources available in environment and other factors. The evaluation of grid using various parameters is highly desirable as it serves as basis for future research in related areas.

The problem undertaken in this paper is encoding of video from AVI to FLV format. The problem is computed intensive as it requires large number of compute cycles to complete. It is also network intensive as it requires large amount of video data to be transferred on the distributed system for video processing. AVI file format is the most widely used format for playing videos on desktops and televisions. With the wide spread use of Flash plug in, most of the video data is viewed on the web in the form of FLV format. AVI file format also requires large size which is not efficient enough to be transferred on web especially with low bandwidth connections. There is a requirement to convert the existing AVI format video into FLV format but this conversion process being compute intensive takes lot of time to complete. Grid systems can be exploited for execution of such jobs as the video file can be easily be split into number of fragments and each fragment is independently and simultaneously converted into FLV format on the distributed system. Each encoded fragment is then recombined to create full video in FLV format. The metrics that are used for evaluating the performance of grid system is job execution time.

IV. Methodology and Architecture

To meet the above objectives Grid environment is created using Globus toolkit, Torque cluster [11][12], Network file system (NFS) on the grid nodes. The topology implemented as star topology in which all nodes are connected through a central hub as shown in Figure1.

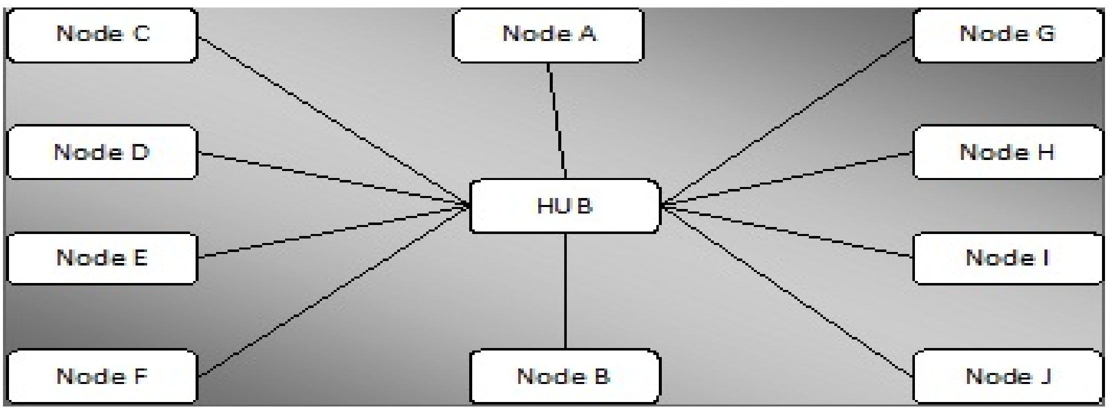

Fig. 1 Architecture of Grid system

Table I. Software Components on Grid Nodes

Software components, Nodes	IP Addresses	Globus Container Services	Globus run-ws	GridFtp Sever	NFS Server	NFS Client	Torque server, Scheduler	Torque Client
Node A	192.168.31.39		Yes	Yes				
Node B	192.168.31.40	Yes	Yes	Yes	Yes		Yes	
Node C	192.168.31.41					Yes		Yes
Node D	192.168.31.42					Yes		Yes
Node F	192.168.31.43					Yes		Yes
Node G	192.168.31.44					Yes		Yes
Node H	192.168.31.45					Yes		Yes
Node I	192.168.31.46					Yes		Yes
Node J	192.168.31.47					Yes		Yes

Various software components installed on grid nodes are as shown in Table I. Apart from that all nodes have video encoder that encodes the video from avi to flv format. The encoder used in this paper is mencoder. Node B is the central server which runs globus container. Node B implements CA (certification authority) which delegates proxy certificates over the grid nodes to authenticate grid users and is also the head node of torque cluster on which PBS server runs. The PBS server dispatches and monitors the execution of job on local cluster. The local cluster is composed of Node C up to Node J. Nodes are added into the cluster from time to time. The file sharing is done by implementing NFS on the cluster nodes i.e. Node B up to Node J. NFS server runs inside the kernel of Node B and NFS client service runs on all other nodes of the cluster. NFS is configured for optimum performance so that it does not affect the permornace of the grid system. The Node A is the grid client node running the gridftp and globusrun services for submitting the job on the grid.

The grid jobs are submitted from grid client Node A on grid server Node B. On Node B the grid job submission request received from Node A is translated automatically into job submission request on torque cluster by wsgram-pbs adapter[14]. The adapter translates the request into job submission request as executed by pbs server. The pbs scheduler schedules the job request on to the available execution nodes in the cluster. The cluster nodes from Node C to Node J are execution nodes which complete the task and return the results to cluster head node i.e. Node B. The results are then passed back from Node B to Node A by gridftp services. The file staging between node A and node B is done by gridftp server running on both Node A and Node B. The file staging between Node B and the cluster nodes is prformed by shared network file system.

The video job is submitted in the form of job submission file in xml format. Firstly the video to be encoded is partitioned into 2 minute fragments and each fragment is submitted onto the Node B as job submission file. The globus services running on Node B creates dynamic job submission command over the torque cluster. The video fragments are encoded into FLV format on the torque cluster nodes and sent back from the grid to the client node i.e. Node A. The encoded video fragments are combined into single FLV file resulting into completion of job. The splitting, joining of video fragments and creation of dynamic job submission files for each fragment is implemented in the form of python script. Different fragment/slice sizes are taken into consideration for observing the effect of job fragment size on the total job execution time. The same procedure is adopted for converting 84 minute size video from AVI to FLV.

V. Results Analysis

The experiments are done on a video of length 60 minutes in avi format. The video is fragmented into 30 slices by the job script, each of length 2 minutes and submitted over the grid. Each slice is encoded into flv format and all the encoded slices are reassembled to create full video in flv format. The parameter used is the job completion time i.e. the time when the job is submitted to the grid system and the time when the resultant complete video in flv format is available. The results are as shown in Table II:

Table II. Execution time with varying number of nodes for 60 minute video

Number of Execution nodes	Execution Time (in sec)
	Fragment size= 2 min
1	558
2	311
3	309
4	288
5	256
6	277
7	252
8	242

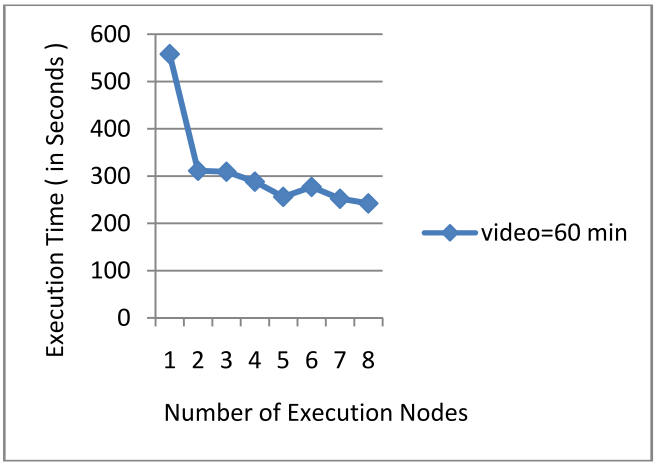

Fig. 2 Performance Comparison with varying no. of Grid Nodes for 60 minute video.

For a video of size 84 minutes, various execution times for different fragment sizes are as shown in Table III.

Table III. Execution time with varying number of nodes for 84 minute video.

Number of Execution nodes	Execution Time (in sec)
	Fragment size=2 min
1	718
2	596
3	410
4	414
5	399
6	365
7	357
8	342

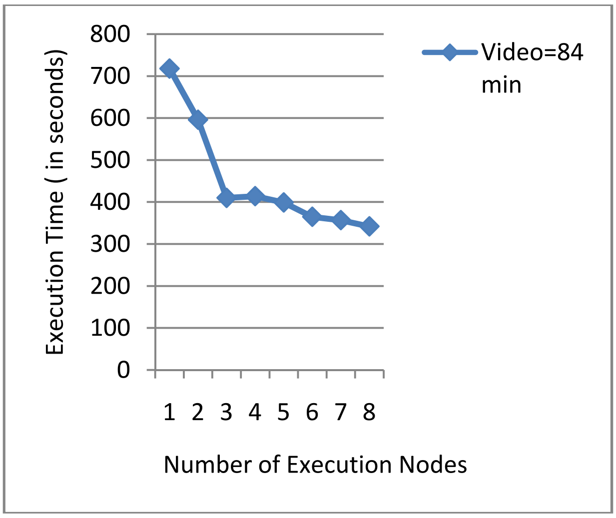

Fig. 3 Performance Comparison with varying no. of Grid Nodes for 84 minute video.

It is observed from the experimental results that avi to flv video conversion takes less time to execute on globus grid system as compared to single system. It is also observed that the performance graph is not linear and that the performance does not improve linearly with increasing number of nodes. After addition of sixth node the performance gain is saturated and in some cases there is adverse effect i.e. the job takes more time with increasing number of nodes. This is due to limited network bandwidth available for implementing network file system. Overall in both cases the average performance gain is 40% as compared to single execution system.

VI. Conclusion

It is observed that globus grid systems are far more suitable for the execution of jobs that can be partitioned and each fragment can be executed on separate execution node than running the same job on single or centralized system. Though the results show better performance than single system yet the performance is not completely linearly dependent on addition of nodes into grid system. In future, the performance can be further improved by implementing the grid system on faster network with higher bandwidth. The effect of variation in job fragment size on performance of grid system is yet to be determined. The current grid system is implemented using torque local scheduler, the performance of which can also be compared with other cluster scheduling systems such as Maui, Load Leveller, SGE etc.

References

[1] Ian F., Carl K., and Steve T., 2001, "*The anatomy of the grid: Enabling scalable virtual organizations*", International J. Supercomputer Applications, Vol. 15, Issue 3, pp 13.

[2] Rajeev W., 2008, *'Grid Computing With Globus: An Overview And Research Challenges"*, Department of Computer and Information Sciences, University of Hyderabad International Journal of Computer Science and Applications, ÓTechnomathematics Research Foundation Vol. 5, No. 3, pp 56 - 69.

[3] Guido J., Noordende V., 2008, *'A Trusted Data Storage Infrastructure for Grid-based Medical Applications"*, Eighth IEEE International Symposium on Cluster Computing and the Grid, pp 627-632.

[4] Lorpunmanee S., et.al. , 2007, *'An Ant Colony Optimization for Dynamic Job Scheduling in Grid Environment"*, Pwaset Vol. 23 Issn 1307-6884.

[5] Sidhu J. and Singh S., 2010, *'Design, Implementation and Evaluation of Grid Environment for DV to MPEG4 Video Conversion"*, International Journal of Computer Science & Communication Vol. 1, No. 2, pp. 151-155

[6] Fernando P., *2009, 'Video Compression: Discussing the Next Steps"*, 978-1-4244-4291-1/09, IEEE, pp 1582-1583.

[7]http://en.wikipedia.org/wiki/Flash_ Video

[8] Ping Li, Veeravalli B., and Ashraf A.K., *2005, 'Design and Implementation of Parallel Video Encoding Strategies Using Divisible Load Analysis"*, IEEE Transactions On Circuits And Systems For Video Technology, Vol. 15, No. 9, pp 1098-1112.

[9]http://en.wikipedia.org/wiki/ Grid_computing

[10] http://www.grid.org/

[11]http://www.adaptivecomputing.com/ products/torque.php

[12]http://www.adaptivecomputing.com/resources/docs/torque/pdf/TORQUE_Administrator_ Guide.pdf

[13]https://www.redbooks.ibm.com/red books /pdfs/sg246778.pdf

[14]http://globus.org/toolkit/docs/4.0/ execution/wsgram/WS_GRAM_Approach.html

[15]www.cloudbus.org/reports/EOSJ Article TR05.pdf

[16] Globus Project Home, Available at http://www.globus.org/