IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

# Cluster and Grid Computing: A Detailed Comparison

**A. Ahuja**

Department of Computer Science, Guru Nanak Dev University, Amritsar, India

*Corresponding Author: annieahuja89@gmail.com, Tel.: +91-8054915241

**Available online at: www.ijcseonline.org**

Accepted: 24/May/2018, Published: 31/May/2018

***Abstract***— Computing power has gained momentum in the last few years with the incorporation of cluster and grid computing models. With supercomputers, some issues need to be tackled in terms of solving the intensive computational problem tasks in an economical way. Even high performance computing but with low cost was not possible earlier with supercomputers but with the advent of cluster and grid computing technologies it is no more restricted to the multi-institutional virtual organizations but becomes a reality for the small scale organizations also. This paper discusses the cluster and grid computing technologies, and gives an insight details in terms of vital features of clusters and grids, challenges, applications, tools and simulation environment for their implementation and comparative analysis of the same.

***Keywords—*** Cluster computing, Grid computing, Computational grid, Data grid, Scavenging grid, Nimrod, Globus, Condor

## I. INTRODUCTION

The existing distributed and parallel computing paved the way for the cluster and grid computing technologies. The appearance of the terms cluster and grid computing in the era of 90's has significantly gained importance in the field of computer science. For the past many years, supercomputers were considered to be the leader in the field of computing and information technology. The computational tasks were implemented with supercomputers but that involves high scale investment, thus it is an expensive approach [1]. It was not easy for any computing technology to withstand the computing power of supercomputers. So the key requirement in the evolving technology is to have a computing power which gives high performance with low cost. This leads to the emergence of clustering approach of independent computers and hence, clusters computing. The cluster platforms brought up by a number of academic projects, such as Beowulf, Berkeley NOW, and HPVM [1].

The goal of Grid computing was to change the way of accessing resources. It involves distributed computing at its core as coordinating and sharing the resources, computing power, storage, application, data, network in a dynamic manner for the organization dispersed in different geographical locations [2]. With the passage of time, grid has significantly gained pace and provides an effective way for sharing the resources and solves problem encountered in clustering model in a dynamic manner in multi-institutional virtual organizations. Both cluster and grid computing evolution leads to the evolving cloud computing technology.

This paper reviews the basic terminology of computing models by representing theoretical findings by an author. The significant contribution made by an author lies in comparative analysis of both paradigms. The rest of the paper is structured as follows: Section II contains introduction to computing models- Cluster Computing and Grid Computing. Section III contains benefits of computing models. Section IV discusses the shortcomings of computing models. Section V contains challenges being faced by these computing models. Section VI contains area of application of computing models. Section VII contains tools that are used for implementing these computing models. Section VIII represents parametric analysis of Cluster and Grid Computing.

## II. COMPUTING MODELS

### *A. Cluster Computing*

Some areas in the field of science and engineering which could not be effectively managed using supercomputers which leads to the emergence of cluster paradigm. In computing terminology, a cluster refers to the connection between two or more computers or systems that are interconnected with high-speed networks like Gigabit Ethernet to form a single unified system that is capable enough to perform computationally intensive as well as data intensive tasks in an efficient manner. It is computationally more powerful than the single computer or system. The coordination reflects the ease in problem solving which is otherwise not possible with a single computer system. The cluster computing concept is shown in Figure 1.

© 2018, IJCSE All Rights Reserved 

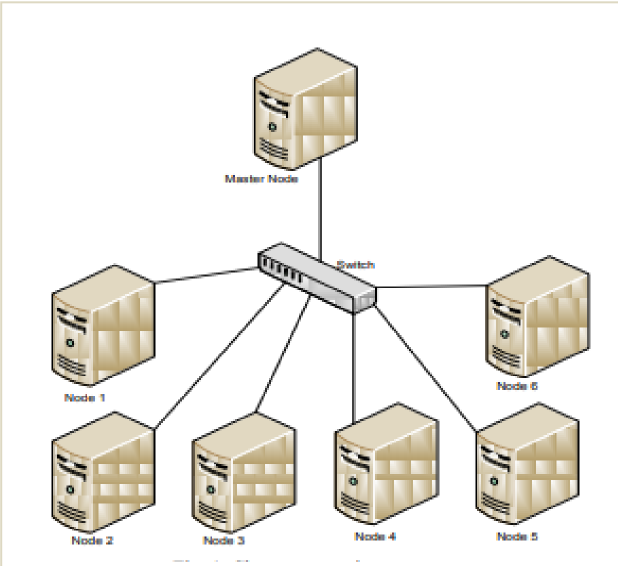


Figure 1. Cluster Computing [adapted from [3]]

The main purpose of using clusters as shown in Figure 2 is that they are capable of providing services even when one or more of the system component fails in an unpredictable manner as this paradigm maintains redundant nodes which get utilized in such a scenario. In this way, there is improved system performance and hence, it is no more an obstacle for researchers as well as for community. Moreover, workload also gets shared among multiple nodes/ computers as it gives an illusion to the user of working as a single virtual computer though multiple systems are involved [1].

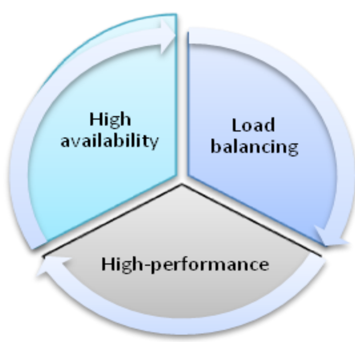


Figure 2. Types and Capabilities of Clusters

1) *Characteristics of clusters*:

The significant characteristics that make up a cluster are as follows:

a) *Availability*: As a cluster, it is always available to the end-user as it is capable enough to handle the failure of any node which forms a part of cluster.

b) *Scalability*: Clusters scalability is defined in terms of number of nodes that can be added any time whenever the demand increases. Moreover, resources can be added to cluster as the application demands.

c) *High performance*: Cluster technology has gained momentum as now it is possible to have high performance computing with cluster computers. Now, it is possible to have solution for those problems (computationally data intensive) which was not possible earlier.

d) *Processing Power*: As a cluster it is computationally more powerful than a single computer. It is able to perform a lot more processing than a single system.

e) *Unified system*: Although, it is a collection of multiple computers but for the end user it seems to be a single virtual computer.

*B. Grid Computing*

Its vision was to utilize the computer based resources that ranges from CPU cycles to the servers used for keeping data in the same way as we utilize real world utilities [2]. The origin of grid can be traced back late in the era of 1980s and early in 1990s [4]. The term 'computational grid' implies hardware and software structure that is responsible for providing reliable, persistent and cost-effective access to high-end computational tasks [5]. The size of a grid varies with the intensity of a problem to be solved. It can be as small like a network of nodes within an organization or it can be widely distributed across many organizations and networks. Grid computing is an evolving vision computing technology that basically exploits the underutilized resources. The goal of grid paradigm is to constructively utilize the computational power. Grid computing as shown in Figure 3 basically balances out the utilization by exploiting parallelism and thus providing a means for solving computationally intensive problems in real time.

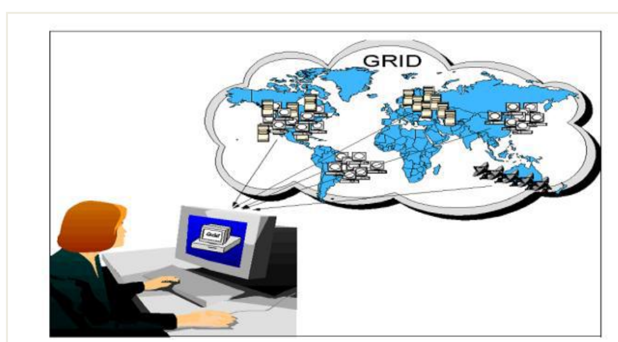


Figure 3. Grid Computing [adapted from [6]]

© 2018, IJCSE All Rights Reserved 

Grid computing is parallel distributed computing environment which is about several processors distributed globally and sharing the computational resources to solve various issues. It comes out to be an advanced computing technology for the existing distributed computing paradigm. In order to determine whether a system is a Grid, if it is satisfied with all the following features:

- To make the system implementation in a most-economical way for a given amount of computing resources
- Efficient in solving large-scale problem that otherwise cannot be solved using existing paradigms and without involving more computational power
- It also proposed that resource management and controlling of nodes/computers is towards a common goal. [5]

1) *Types of Grids*:

There are no such standard types of grids but generally grids are classified into following three types as specified by IBM as represented in Figure 4 [7].

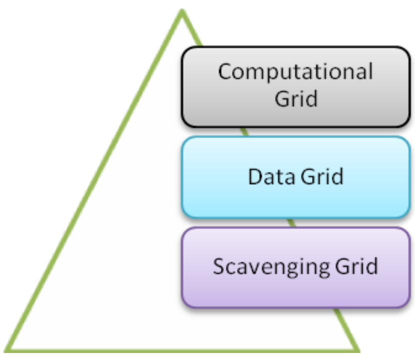


Figure 4. Types of Grids

a) *Computational grids:* Computational grids are capable enough to provide huge amount of computational capabilities that are required for solving data intensive and computational intensive problems with cost-effectiveness, which otherwise was not possible with supercomputers. Though, supercomputers can provide computational power but at the cost of high-investment which is not feasible.

b) *Scavenging Grid:* In this type of grid CPU cycles are fully utilized, even other software and special peripherals like sensors are shared when and where required. With the optimal utilization of resources, these grids move towards their common objective.

c) *Data grid:* In this grid, the main focus is to provide management and sharing for scattered data. The data whether present al local level or remote level; it provides the necessary support by replicating the data grids, thus providing most-economical solution.

2) *Characteristics of Grids:*

The significant characteristics that make up a grid are described as follows [8].

a) *Resource Sharing*: The resources can be shared in a grid paradigm; the resources which are not at your site can be accessed efficiently by the other site which leads to enhanced efficiency and thus become cost-effective approach. The resources of one organization can be used by other one when and where required.

b) *Resource Coordination*: Coordinating the resources in such a manner so as to fully utilize the potential in terms of computing capabilities.

c) Large scale: A large scale problem requires grid to have large number of resources that range from few to millions for effective solution.

d) *Heterogeneity*: The capability to handle large number of resources with wide variety is heterogeneity.

e) *Geographical distribution*: The resources in grid computing model are geographically dispersed.

f) *Multiple administrative domains*: Multiple administrative domains are required for facilitating the resource sharing among multiple organizations and each organization/company can have its own mechanism for accessing the resources for security purposes. That's why it has become even more complicated with the involvement of different security policies of each organization that comprises a grid.

g) *Secure access*: Secure access to be provided to the users who are within grid environment. They must be protected from any non-repudiation or malicious attacks.

h) *Performance*: Assurance of the best performance in terms of quality service to the users must be provided by the grid.

i) *Scalability*: According to the application demand, the number of nodes to be deployed can vary and

© 2018, IJCSE All Rights Reserved 

there must be consistency in the performance level.

j) *Adaptability*: Grid has to carefully tackle the scenario of unexpected hardware, software faults and failures as their ratio is high.

k) *Consistency*: The interfaces used in grids in order to handle heterogeneous resources and thus maintaining scalability should be based on standards so that consistent access should not suffer.

## III. BENEFITS OF CLUSTER AND GRID COMPUTING

The advantages of both Cluster and Grid paradigms are represented in Table 1 [6, 9].

Table 1. Benefits of Cluster and Grid Computing

| Sr. No. | Advantages of Cluster Computing | Advantages of Grid Computing |
|---|---|---|
| l | High Availability | Availability of additional resources in addition to CPU, storage, etc |
| 2. | Single System Image | Optimal use of computing capacity |
| 3. | Management w.r.t. number of components | High Reliability especially in case of real time systems |
| 4. | Cost-effective | Easy collaboration with other organizations |
| 5. | Load balancing | Time consumed to solve large scale problems is less |

## IV. DRAWBACKS OF CLUSTER AND GRID COMPUTING

The disadvantages of both Cluster and Grid paradigms are represented in Table 2 [6, 9].

Table 2. Drawbacks of Cluster and Grid Computing

| Sr. No. | Disadvantages of Cluster Computing | Disadvantages of Grid Computing |
|---|---|---|
| 1. | Programming Issues | Grid software and standardization mechanisms are still in evolving state |
| 2. | Difficult to handle by non-professional | Multiple Administrative Domains |
| 3. | Difficult to find fault | Connection with high-speed internet is required |

## V. CHALLENGES OF CLUSTER AND GRID COMPUTING

Every computing model as they are evolving from each other has similar or unique challenges that need to be addressed and convert these obstacles like challenges into opportunities so that potential utilization of each and every paradigm can be possible. In this section, we highlight some of the challenges of the three computing models considered.

### A. *Challenges of Cluster Computing:*

The research challenges of cluster computing are as follows: [1]

1) *Middleware*: It is quite critical to make such an interface or middleware or software environment that provides users an image of single system, although it is a collection of multiple systems.

2) *Programming Issues*: To choose a suitable programming model as per the need of application is quite a critical decision. It must be capable enough to handle division of task among the nodes/computers whenever request is made and then providing coordination and communication between them.

3) *Flexibility*: The handling of varied number of incoming service requests. The variance is in response time as it depends on the application.

4) *Scalability*: The system should be scalable enough to meet the variance in requirements of the system which can affect the performance of the system.

5) *Network Selection [proposed]*: Network should be selected such that it provides high bandwidth and low latency, so that it does not affect the performance of the application which is to be run.

### B. *Challenges of Grid Computing:*

The research challenges faced by grids include the following: [1]

*1) Different Administrative Domains*: Multiple administrative domains are involved because of collaboration between different companies which is difficult to maintain and coordinate.

*2) Support*: Technology support is required to utilize the grid.

© 2018, IJCSE All Rights Reserved 

*3) Standards*: It is still an evolving vision in terms of software and standards. Yet no proper standardization mechanisms have been there which are required for its effective realization.

*4) Lack of grid enabled software*: There is need to develop open source environment as most of the grid enabled software's lack copyright issues and source code of license.

*5) Dynamic*: As resource management is achieved by different organizations and there are constraints for their arrival and leaving the grid that poses a burden on the grid paradigm.

*6) Application Development*: Application development in such a way that it must be capable enough to disassemble and then assemble the computational elements.

*7) Programming Issues*: The geographical dispersion of nodes makes it even more complex to do the application programming.

## VI. APPLICATIONS OF CLUSTER AND GRID COMPUTING

### *A. Applications of Cluster Computing:*

Some of the applications of cluster computing in the various fields are as follows: [1]

1) The Weather Research and Forecast (WRF) is a research project in collaboration for developing a next generation regional forecast model and for the weather forecast data assimilation system is used for collecting the data and then make prediction.

2) Hadoop is an open-source framework which provides platform to data-intensive applications in a processing cluster built from commodity server hardware. Other applications such as filtering and indexing of web listings, facial recognition, etc.

3) Clusters made their existence by solving some of the challenging applications which includes weather modeling, automobile crash simulations, life sciences, image processing, electromagnetic, data mining, aerodynamics and astrophysics. It also provides a base to those applications that requires data intensive computations. For backup purposes like in banking sector it finds its application. Hosting of many service sites can be done.

4) Other application areas such as Education, commercial sectors for industrial promotion, medical field, etc [6].

### *B. Applications of Grid Computing:*

Grid computing model has wide range of applications. Some of the applications in which grid computing is utilized are the following [6].

1) Its applications in the field of Engineering Design and Automation includes Computational aerodynamics, artificial intelligence and automation, remote sensing applications, pattern recognition, computer vision, etc.

2) It is applicable in the field of Medical, Military like Polymer chemistry, medical imaging, etc.

3) In Predictive Modeling and Simulation area, its applications include flood warning, socio-economic and government use, numerical weather forecasting, astrophysics, etc.

4) Another field of application is in Energy Resource Exploration which includes Plasma Fusion power, seismic exploration, nuclear reactor safety, etc.

5) Visualization is also part of its application area like computer-generated graphics, films and animations, data visualization, etc.

6) MammoGrid is medical grid application in which prototypes are used to present the mammogram analysts, who are working in breast cancer screening can use this significant information provided by grid infrastructure to resolve image analysis problems[1].

## VII. RELATED WORK

### *A. Tools and Simulation Environment for Cluster and Grid Computing*

The various tools that are applied for implementing cluster and grid computing paradigms are represented in Table 3.

Table 3. Tools and Simulation Environment for Cluster and Grid Computing [1]

| Technology | Tools | Application Area (Key Findings) |
|---|---|---|
| Cluster | Nimrod | Parametric computing on clusters |
| | PARMON | Monitoring of system resource at 3 levels: system, node and component, Example-C-DAC PARAM 10000 supercomputer |
| | Condor | Job and resource management |
| | MPI | Parallel programming on |

© 2018, IJCSE All Rights Reserved 

| | | |
|---|---|---|
| | | clusters |
| Grid | Paradyn | Performance analysis |
| | Nimrod-G | Dynamic resource discovery and dispatching of jobs over computational grids |
| | Condor-G | Security in communication and standardized accessing, job management and provides friendly execution environment |
| | Globus | Construction of computational grids and grid based applications |
| | Gridbus | Cluster and grid middleware development for Service oriented computing |
| | Legion | Transparent core scheduling, security, fault-tolerance capability, data management |

The various tools that are used for cluster computing implementation includes Nimrod, PARMON, Condor, MPI (Message Passing Interface) and each and every tool is vital in its own way. The other cluster simulators such as Flexi-Cluster which is for a single compute cluster and VERITAS which is a server simulator, etc [1]. Similarly, Grid computing tools include Paradyn, Nimrod-G (Nimrod-Globus), Condor-G (Condor –Globus), Globus, Gridbus, and Legion. The other simulators such as Gridsim, ZENTURIO, etc are used in grid computing model [1].

### *B. Parametric Analysis of Cluster and Grid Computing*

The parametric analysis has been carried out for both the computing technologies showing in Table 4 [1, 6, 9].

Table 4. Parametric Analysis of Cluster and Grid Computing using existing parameters

| Existing Parameters | Cluster Computing | Grid Computing |
|---|---|---|
| Interoperability | Yes | Yes |
| Standardization | Yes | Yes with conflicts |
| Business Model | No | No |
| Service Oriented Architecture (SOA) | No | Yes |
| Multi-tenancy | No | Yes |
| Heterogeneous | No | Yes |
| Expensive | Yes | Yes |
| Data Locality exploited | No | Yes with data grids [Key Finding] |
| Rapid Flexibility | No | No |
| Network accessing | Yes | Yes |
| Inherent Distribution | No | Yes |

Table 5. Parametric Analysis of Cluster Computing and Grid Computing using Key Findings (Parameters) by studying literature

| Key Findings (Parameters) | Cluster Computing | Grid Computing |
|---|---|---|
| Resource Manager | Yes | Yes |
| Different hardware and Software | No | Yes |
| Multiple Ownership | No | Yes |
| Centralized User Management | Yes | No |
| Based on Virtual Organization | No | Yes |
| Single System Image | Yes | No |
| Failure Management | Yes, up to some extent | Yes, up to some extent |
| Potential for Value added solutions | Yes, up to some extent | Yes, up to some extent |

In the parametric analysis, it is observed that grid computing is more powerful than cluster computing in some aspects. It depends upon your application need that which paradigm is the most suited one. Some of the parameters have not focused by the earlier researchers while doing comparative analysis in this form as shown in Table 5 like centralized user management, computing technology that is using the concept of virtual organization, failure management, potential for value added services, etc. The fourth parameter is Service Oriented Architecture (SOA) in Table 4 which is shown by Figure 5 that the evolving computing technologies support SOA like Grid computing but Cluster computing does not support SOA.

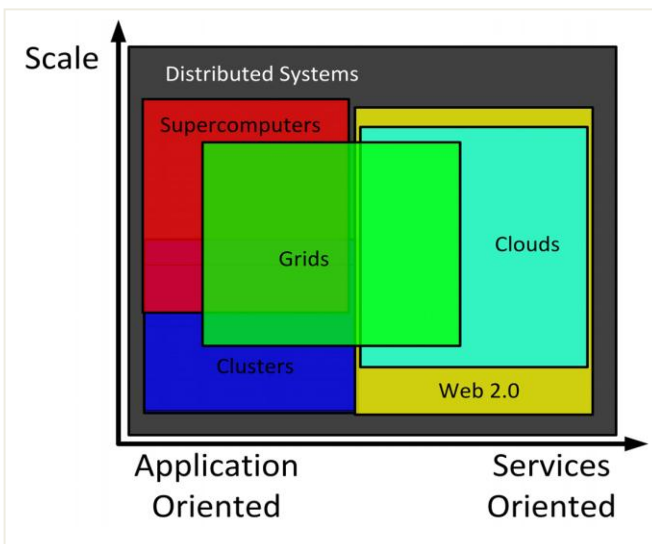


Figure 5. Evolution of Service Oriented Architecture (SOA) started from supercomputers to clouds [adapted from [2]]

© 2018, IJCSE All Rights Reserved 

## VIII. CONCLUSION and Future Scope

In this paper, author's theoretical findings in the form of detailed comparison of cluster computing and grid computing technologies are presented. The insight exploration of the clusters and grids is discussed, which form the core component of both these computing paradigms. The application areas of the various tools that are used to implement cluster and grid paradigms have also been explored for the key findings. Author believes that a detailed comparison can help the researchers and communities understand, share and evolve in this domain and also results in acceleration of cloud computing paradigm. It can be concluded that as the future of cluster computing is grid computing, similarly the future of grid computing is cloud computing. It has vast area of scope if exploited potentially. It is observed that the application oriented framework is now moving towards service oriented framework which proves to be a big success for synchronous and asynchronous applications. With the evolving technologies, business logic is getting modularized and presented as services for user applications. In the last section of this paper, parametric analysis is carried out in order to understand the technical side behind these technologies which otherwise conceptually look similar. Although Grid computing is still an evolving technology yet it is promising one to handle the large and complex problems very efficiently. Both the technologies have gained pace in recent years. Therefore, there is a scope for further research in these areas.

## ACKNOWLEDGMENT

I would like to thank my colleagues for their invaluable guidance.

## REFERENCES

[1] N. Sadashiv, S. M. D. Kumar, "Cluster, Grid and Cloud Computing: A Detailed Comparison", In the Proceedings of 2011 The 6th International Conference on Computer Science & Education (ICSSE 2011), Singapore, pp. 477-482, 2011.

[2] S. M. Hashemi, A. K. Bardsiri, "Cloud Computing Vs. Grid Computing", ARPN Journal of Systems and Software, Vol. 2 No. 5, pp. 188-194, 2012.

[3] I. Nwobodo, "Cloud Computing: A Detailed Relationship to Grid and Cluster Computing", International Journal of Future Computer and Communication, Vol. 4, No. 2, pp. 82-87, 2015.

[4] S. K. Singh, K. Kaur, A. Aggarwal, D. Verma, " Achieving High Performance Distributed System: Using Grid, Cluster and Cloud Computing", International Journal of Engineering Research and Applications, Vol. 5, No. 2, pp. 59-67, 2015.

[5] R. Geetha, D. Ramyachitra, "Security Issues in Grid Computing", In the Proceedings of 2013 International Conference on Research Trends in Computer Technologies ( ICRTCT-2013) , pp. 33-37, 2013

[6] K. Kaur, A. K. Rai, " A Comparative Analysis: Grid, Cluster and Cloud Computing", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, No. 3, pp.5730-5734, 2014.

[7] P. Samimi, A. Patel, "A Review of Pricing Models for Grid & Cloud Computing", In the Symposium of 2011 IEEE Symposium on Computers & Informatics, pp. 634-639, 2011.

[8] P. Sharma, "Grid Computing Vs. Cloud Computing" International Journal of Information and Computation Technology, Vol. 3, No. 6, pp. 577-582, 2013.

[9] R. Kumar, S. Charu, "Comparison between Cloud Computing, Grid Computing, Cluster Computing and Virtualization", International Journal of Modern Computer Science and Applications (IJMCSA), Vol. 3, No. 1,pp. 42-47, 2015.

## Authors Profile

*Ms. A Ahuja* pursued Bachelor of Computer Appli-Cations from Guru Nanak Dev University, Amritsar In 2011 and Master of Computer Applications from Punjab Technical University in 2014. She is currently Working as Assistant Professor in the Department of Computer Science, Guru Nanak Dev University, Amritsar. She is IBM DB2 certified. She has 4 years of teaching experience. Her research interest is in latest developments in Information Technology.

© 2018, IJCSE All Rights Reserved 