B.C.A. SCIENCE : SEMESTER-IV

GRID AND CLOUD COMPUTING

Dr. Ms. MANISHA BHARAMBE
Mrs. VEENA GANDHI

NIRALI PRAKASHAN

Text Book Of

GRID AND CLOUD COMPUTING

For
B.C.A. Science : Semester - IV

As Per New Syllabus

Dr. Ms. Manisha Bharambe
M.Sc (Comp. Sci.), M.Phil, Ph.D (Comp. Sci.)
Associate Professor, Computer Science Dept.,
MES Abasaheb Garware College,
Pune - 4

Mrs. Veena Gandhi
M.C.S., M.Phil (Comp. Sci.), UGC-NET
Assistant Professor, Computer Science Dept.,
Abeda Inamdar Senior College,
Pune - 1

Price ₹ 70.00

NIRALI PRAKASHAN
ADVANCEMENT OF KNOWLEDGE

N3880

Grid and Cloud Computing

First Edition : January 2018

Published By :

NIRALI PRAKASHAN

Abhyudaya Pragati, 1312, Shivaji Nagar

Off J.M. Road, PUNE – 411005

Tel - (020) 25512336/37/39, Fax - (020) 25511379

Email : niralipune@pragationline.com

Printed By :

STAR COPIERS PVT. LTD.

Kumthekar Road, Sadashiv Peth

Pune - 411 030

Tel - (020) 24479201

➢ DISTRIBUTION CENTRES

PUNE

Nirali Prakashan : 119, Budhwar Peth, Jogeshwari Mandir Lane, Pune 411002, Maharashtra
Tel : (020) 2445 2044, 66022708, Fax : (020) 2445 1538
Email : bookorder@pragationline.com, niralilocal@pragationline.com

Nirali Prakashan : S. No. 28/27, Dhyari, Near Pari Company, Pune 411041
Tel : (020) 24690204 Fax : (020) 24690316
Email : dhyari@pragationline.com, bookorder@pragationline.com

MUMBAI

Nirali Prakashan : 385, S.V.P. Road, Rasdhara Co-op. Hsg. Society Ltd.,
Girgaum, Mumbai 400004, Maharashtra
Tel : (022) 2385 6339 / 2386 9976, Fax : (022) 2386 9976
Email : niralimumbai@pragationline.com

➢ DISTRIBUTION BRANCHES

JALGAON

Nirali Prakashan : 34, V. V. Golani Market, Navi Peth, Jalgaon 425001,
Maharashtra, Tel : (0257) 222 0395, Mob : 94234 91860

KOLHAPUR

Nirali Prakashan : New Mahadvar Road, Kedar Plaza, 1st Floor Opp. IDBI Bank
Kolhapur 416 012, Maharashtra. Mob : 9850046155

NAGPUR

Pratibha Book Distributors : Above Maratha Mandir, Shop No. 3, First Floor,
Rani Jhanshi Square, Sitabuldi, Nagpur 440012, Maharashtra
Tel : (0712) 254 7129

DELHI

Nirali Prakashan : 4593/21, Basement, Aggarwal Lane 15, Ansari Road, Daryaganj
Near Times of India Building, New Delhi 110002 Mob : 08505972553

BENGALURU

Pragati Book House : House No. 1, Sanjeevappa Lane, Avenue Road Cross,
Opp. Rice Church, Bengaluru – 560002.
Tel : (080) 64513344, 64513355,Mob : 9880582331, 9845021552
Email:bharatsavla@yahoo.com

CHENNAI

Pragati Books : 9/1, Montieth Road, Behind Taas Mahal, Egmore,
Chennai 600008 Tamil Nadu, Tel : (044) 6518 3535,
Mob : 94440 01782 / 98450 21552 / 98805 82331, Email : bharatsavla@yahoo.com

Preface...

We take an opportunity to present this book entitled as **"Grid and Cloud Computing"** to the students of **B.C.A. Science, Semester-IV** as per the New Syllabus, June 2017-2018.

The book covers theory of Grid Computing-An overview, Benefits of Grid Computing, Cloud Computing-An overview and Abstraction and Virtualization.

A special words of thank to Shri. Dineshbhai Furia, Mr. Jignesh Furia for showing full faith in us to write this text book. We also thank to Mr. Amar Salunkhe and Mrs. Prachi Sawant of M/s Nirali Prakashan for their excellent co-operation.

We also thank to Mr. Ravindra Walodare, Mr. Sachin Shinde, Mr. Nilesh Deshmukh, Mr. Ashok Bodke, Mr. Moshin Sayyed and Mr. Nitin Thorat.

Although every care has been taken to check mistakes and misprints, any errors, omission and suggestions from teachers and students for the improvement of this text book shall be most welcome.

Authors

Syllabus...

1. Grid Computing - An Overview (12 L)

- High-Performance Computing
- Cluster Computing
- Peer-to-Peer Computing
- Internet Computing
- Grid Computing - What Grid Computing is, Peer-to-Peer Networks and Grid Computing, Cluster Computing and Grid Computing, Internet Computing and Grid Computing
- Grid Computing Models
- Open Grid Services Architecture
- Types of Grids - Departmental Grids, Enterprise Grids, Extraprise Grids, Global Grids, Compute Grids, Data Grids, Utility Grids
- Grid Networks - Grid Network Peering Points
- Grid Applications Characteristics

2. Benefits of Grid Computing (03 L)

- Exploiting Underutilized Resources
- Parallel CPU Capacity
- Virtual Resources and Virtual Organizations for Collaboration
- Access to Additional Resources
- Resource Balancing
- Reliability
- Management

3. Cloud Computing - An Overview (11 L)

- Defining Cloud Computing
- The NIST Model
- Deployment Models - Public, Private, Hybrid
- Service Models - Infrastructure as a Service (IaaS), Platform as a service (PaaS), Software as a Service (SaaS), Cloud Reference Model.
- Examining the Characteristics of Cloud Computing
- Benefits of Cloud Computing
- Disadvantages of Cloud Computing

4. Abstraction and Virtualization (04 L)

- Using Virtualization Technology
- Load Balancing and Virtualization - The Google Cloud
- Understating Hypervisors - Virtual Machine Types
- Exploring SaaS - Salesforce.com, PaaS - Force.com, IaaS - Amazon EC2

❖❖❖

Contents...

❖❖❖

CHAPTER 1

Grid Computing - An Overview

Contents...

Objectives...

After reading this chapter you will be able,

- ❑ To understand High-performance Computing
- ❑ To learn Cluster Computing and Internet Computing
- ❑ To study Peer-to-Peer Computing
- ❑ To understand Basic Concepts of Grid Computing (Need, Definition etc.)

1.1 INTRODUCTION

- In today's incredibly complex world of computational power, very high speed machine processing capabilities, complex data storage methods, next-generation telecommunications, new-generation operating systems and services and extremely advanced networking services capabilities - we are

entering a new era of computing. Today industry, businesses, and home users are placing more complex and challenging demands on the networks.

- Grid computing is one of the most innovative aspects of computing techniques in recent years. It mainly focuses on resource sharing among geographically distributed sites and the development of innovative, high-performance oriented applications.
- Grid computing is a rapidly maturing technology that will address many of the financial and operational inefficiencies of today's Information Technology (IT) infrastructure.
- After developing strong roots in the global (world-wide) academic and research communities over the last decade, grid computing has successfully entered the commercial world.
- Grid computing is a set of standards and technologies that academics, researchers, and scientists around the world are developing to help organizations take collective advantage of improvements in microprocessor speeds, optical communications, raw storage capacity, and the Internet. By using the technique to disaggregate their computer platforms and distribute them as network resources, companies can vastly increase their computing capacity.
- As grid computing concept empowered the development of large computing system spread across wide geographic location, collaboration capability was the next need of such system where members of a team dispersed geographically could work together on a common task.
- In this chapter we study, High-performance Computing, Cluster Computing, Internet Computing, Grid Computing etc., in detail.

1.2 HIGH-PERFORMANCE COMPUTING

- Billions of people use the Internet every day. As a result, supercomputer sites and large data centers must provide high-performance computing services to huge numbers of Internet users concurrently.
- High-performance computing is the use of distributed computing facilities for solving problems that need tremendous computing power. Historically, supercomputers and clusters are specifically designed to support high-performance computing applications.
- Supercomputer is a general term for computing systems capable of sustaining high-performance computing applications that require a large number of processors, shared or distributed memory, and multiple disks.
- High-performance computing aims to deliver the computational power and resources necessary to address issues with performance requirements that would require expensive resources or would otherwise be unavailable.
- High-Performance Computing (HPC) generally refers to what has traditionally been called supercomputing. There are hundreds of supercomputers deployed throughout the world. Key parallel processing algorithms have already been developed to support execution of programs on different, but co-located processors.
- High-performance computing is defined as, “the practice of aggregating tremendous computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in order to solve large/complex problems in various fields like science, engineering, or business”.
- High-performance computing was the desire to share high-performance computing resources amongst researchers that led to the development of grid computing technology and some of its fundamental infrastructure.

- Modern supercomputers are built on the principles of Grid Computing (GC) incorporating many smaller computers into a larger whole.
- The idea of GC is to handle the computation management, data movement, storage management, and other infrastructure that could handle large grids without restricting themselves to specific hardware and requirements.
- In grid HPC related to the parameters speed, performance and throughput of computing. Grid computers (supercomputers) are belongs to high performance.
- High-performance mainly relies on parallel processing and speedup. We improve the performance in domains like processor, memory, I/O devices. In the form of improving these one we can achieve the high performance.
- Grids are super Internets for high-performance computing, world-wide collections of high-end resources such as supercomputers, storage, advanced instruments, and immersive environments. These resources and their users are often separated by great distances and connected by high-speed networks.
- High-performance computing system deployment, contrary to popular belief is not limited to research or academic institutions. In fact, more than half of supercomputers deployed in the world today are in use at various corporations like healthcare, banking, finance and so on.
- In high-performance computing systems, a pool of processors (processor machines or Central Processing Units [CPUs]) connected (networked) with other resources like memory, storage, and input and output devices, and the deployed software is enabled to run in the entire system of connected components.
- The industries in which high-performance systems are deployed are numerous in nature. Following table shows the distribution of the top 500 supercomputers by their industries.

Sr. No.	Industry Area	Sample Companies/Organizations
1.	Information Services	EDS
2.	Database	Starbucks, State Farm
3.	Finance	Charles Schwab
4.	Geophysics	Shell, Aramco
5.	Telecommunication	Duetsche Telekom, Sprint
6.	Pharmaceutics	Aventis Pharma
7.	Energy	Centrica
8.	Manufacturing	Alcoa
9.	Transportation	Oy Saimaa Lines
10.	Mechanics	Hitachi
11.	Aerospace	Dassault Avaition
12.	Worldwide Web	Amazon, Newsky
13.	Chemistry	Bayer
14.	Electronics	Motorola, Cisco
15.	Automotive	GM, BMW

Advantages of High-Performane Computing:

1. High-performance computing resources scale with demand and are available with no capital outlay-only the resources used are actually paid for.
2. Experts in high-performance computing help, setup, and optimize the software environment and can help troubleshoot issues that might occur.
3. Computing costs are reduced, particularly where the workflow has spikes in demand.

1.3 CLUSTER COMPUTING

- Due to the high prices of supercomputers, which made those systems out of reach for many research projects the ideal cost-effective solution the term cluster computing emerged.
- A cluster is a collection of parallel or distributed computers which are interconnected among themselves using high-speed networks, such as Gigabit Ethernet.
- Clusters are high-performance, massively parallel computers built primarily out of commodity hardware components, running a free-software operating system such as FreeBSD or Linux.
- Clusters are interconnected by a private high-speed network. It consists of a cluster of PCs, or work-stations, dedicated to running high-performance computing tasks.
- The nodes in the cluster do not sit on user's desks, but are dedicated to running cluster jobs. A cluster is usually connected to the outside world through only a single node.
- In cluster computing, a group of computers are connected through a network switch as shown in Fig. 1.1. Specialized high-speed interconnections were developed for cluster computing.
- A cluster is a collection of stand-alone computers connected using some interconnection network. In cluster computing, systems are loosely coupled in a way that support them to carry out different tasks that are collectively considered to be completed by a single system.
- Nodes are the subsystems of the cluster and are usually connected with each other by an LAN network. Each node runs on its operating system and can be used as a server.

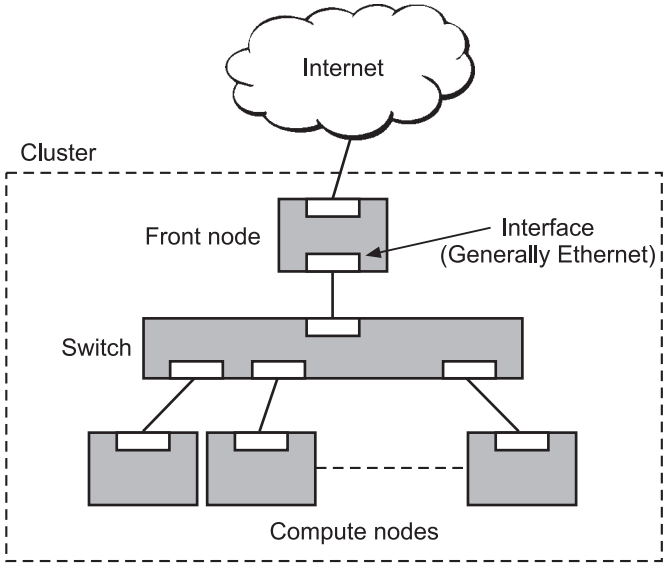

Fig. 1.1: Typical Cluster Computing Configuration

- Cluster computing allows for cluster-based as well as node-based system management for each node of the cluster.

- The computing cluster can be as simple as two machines forming two connected nodes. The concept of cluster computing allows for very advanced processing and provides a high-performance environment.
- A cluster can be formed for different purposes, from a simple enterprise-based cluster to a university-wide network to advanced processing research laboratories, and they can work as efficiently as high-performance computers.
- Cluster computing has been around since 1994 when the first 'Beowulf' clusters were developed and deployed run and manage clusters. Platform Computing, an organization that is now a leader-computing, developed many of the early load-balancing tools for clusters.
- Additionally, tools have also been developed to adapt applications run in the cluster environment. For example, ForgeExplorer tool, can check if particular applications are suitable for parallelization and determine if they would be suitable run on a cluster.
- Over the last decade the exponential growth in microprocessor speeds, has made it possible to create truly impressive clusters. Clusters are widely deployed in industries such as finance, digital entertainment, life sciences and so on.
- An example of a cluster computing system, is shown in Fig. 1.2.

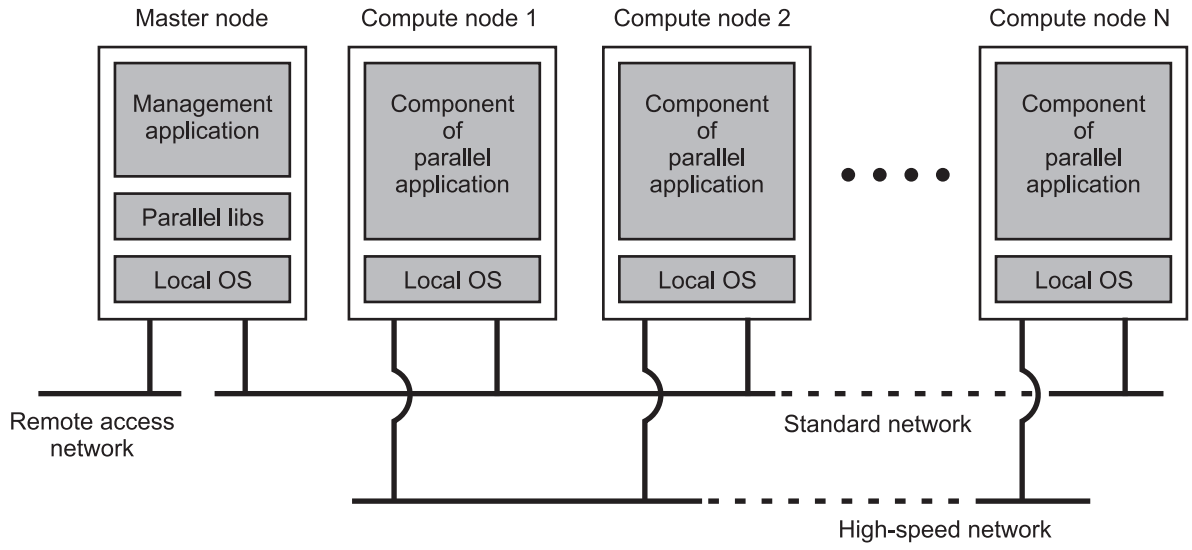

Fig. 1.2

Advantages and Disadvantages of Cluster Computing:

1. Clusters are local to the domain and constructed to solve inadequate computing power.
2. A cluster is related to the pooling of computational resources to provide more computing power by parallel execution of the workload.
3. Clusters are limited in scope with dedicated functionality and local to the domain, and are not suitable for resource sharing among participants from different domains.
4. One of the advantages of cluster computing is to make available, at low cost, the computational power to address in manageable time some categories of problems.
5. The nodes in a cluster are centrally controlled and the cluster manager is aware of the state of node. This forms only a subset of the grid principle of more widely available, intra/inter-domain, communication and resource sharing.

1.4 PEER-TO-PEER COMPUTING

- The earliest method of co-operatively using computers connected to the Internet is peer-to-peer in computing. The recent growth of user-friendly file-sharing networks like Kazaa or Gnutella or Napster, has only now brought P2P (Peer-to-Peer) networks and file sharing into the public eye, methods for transferring files and information between computer have been, in fact, around almost as long as computing itself.
- P2P file sharing has become one of the most popular Internet activities. Peer-to-Peer communication/ interaction involves a direct connection between the sender and the receiver.
- In peer-to-peer computing, there is no client-server relationship among computers connected to a network. A computer can be a client or a server, depending on the circumstances.
- Peers can share files or computing time among themselves. A group of peers can also be used to cooperatively solve compute intensive problems by donating idle CPU cycles available in their computers.
- A P2P distributed system is a system whereby individual users or nodes can communicate with each other autonomously. The computing load is equally shared among the peers.
- A P2P can be defined as, "a set of software applications that consume distributed resources (computing power, content, network bandwidth, and presence - computers, human, and other resources) to perform functions in a centralized/decentralized manner. These functions may include distributed computing, data and content sharing, communication and collaboration."
- P2P is a network-based computing model for applications where computers share resources via direct exchanges between the participating computers. P2P computing has two fundamental properties, the direct communication and the sharing resources between peer users.
- Fig. 1.3 shows peer-to-peer computing. Peer X requests for some data that Peer Y has. However, Peer X has to first locate Peer Y through other peers in the P2P network. Once, Peer Y is located, Peer X deals directly with Peer Y.

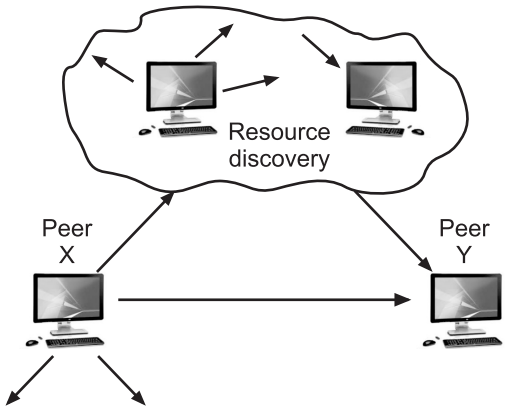

Fig. 1.3

- Today, there are a number of advanced P2P file sharing applications, and the reach and scope of peer networks have increased dramatically.
- The two main models of P2P that have evolved are the centralized model (used by Napster) and the decentralized model (used by Gnutella).

1. Centralized Model:

- In this model, file sharing is based around the use of a central server system that directs traffic between individual registered users.

Grid And Cloud Computing

Publisher : Nirali Prakashan

Author : Dr. Ms. Manisha Bharambe,Mrs. Veena Gandhi

Type the URL : http://www.kopykitab.com/product/19708