Chapter 5
Challenges and Opportunities in High Performance Cloud Computing

Manoj Himmatrao Devare

ⓘ https://orcid.org/0000-0002-9530-3914

Amity University Mumbai, India

ABSTRACT

The scientist, engineers, and researchers highly need the high-performance computing (HPC) services for executing the energy, engineering, environmental sciences, weather, and life science simulations. The virtual machine (VM) or docker-enabled HPC Cloud service provides the advantages of consolidation and support for multiple users in public cloud environment. Adding the hypervisor on the top of bare metal hardware brings few challenges like the overhead of computation due to virtualization, especially in HPC environment. This chapter discusses the challenges, solutions, and opportunities due to input-output, VMM overheads, interconnection overheads, VM migration problems, and scalability problems in HPC Cloud. This chapter portrays HPC Cloud as highly complex distributed environment consisting of the heterogeneous types of architectures consisting of the different processor architectures, inter-connectivity techniques, the problems of the shared memory, distributed memory, and hybrid architectures in distributed computing like resilience, scalability, check-pointing, and fault tolerance.

INTRODUCTION TO HPC

High-Performance Computing (HPC) is the utilization of tools, techniques, and available architectures for execution of the parallel processing scientific and advanced application with a target to achieve high speed, efficiency, reliability, fault tolerance, and security during the execution. The HPC term applies especially to the systems that function above a teraflop or 10^{12} floating-point operations per second (FLOPS). The HPC is a synonym for supercomputing, and some supercomputers work at more than a Petaflop or 10^{15} floating-point operations per second. The HPC is the practice of aggregating computing power in a way that the resulting computation evolves with more power than a single desktop com-

puter. There are massive problems in science, engineering, or business, which needs crunching of the numbers regarding the linear equations. The users of HPC systems are scientific researchers, engineers, academic institutions, some government agencies, the military, medical, health science, environmental sustainability also rely on HPC for multiple potpourri applications execution and their outputs. High-performance systems often use custom-made components in addition to commodity components. The processing power is highly needed in the various emerging technologies like Neural Networks with a massive amount of neurons in many hidden layers, running Deep-Learning applications, Block-Chain applications, Big-data Applications and many interdisciplinary science applications.

Historically, in 1970s CRAY introduced Vector Supercomputer which is oldest, having pipeline arithmetic and stream data from memory and back. There are Parallel Vector Processing (PVP), and then Symmetric Multi-Processing (SMP) with large shared memory. The SMPs are having one thread for each processor and needs synchronization among the multiple threads. Massively Parallel Processing (MPP) connected by customized network and protocol, PC cluster, Non-Uniform Memory Access (NUMA), Grid Computing, and the combination of CPU and GPU. Few of them became obsolete. An HPC machine is having a more complex architecture than a simple desktop computer. It has all of the elements like on processors, memory, disk, network connections and operating system (OS). HPC of interest to small and medium-sized businesses with clusters of computers. Each computer in a commonly configured small cluster has between one and four processors, and today's processors typically have two to four cores. The single computer in a cluster called a node. A cluster of interest to small business could have as few as eight nodes or 32 cores. A typical cluster size in many enterprises is between 16 and 64 nodes, or from 64 to 256 cores. The HPC systems have a prominent choice of operating systems such as Linux and Windows. Linux currently dominates HPC installations, but this in part due to HPC's legacy in supercomputing, large-scale machines, and UNIX. The HPC is closely related with the distributed computing paradigm where the non-serial tasks are divided into many independent tasks, and final output delivered. The HPC infrastructure consisting of 1000s of nodes. Each node is having its processor, RAM, disk, boot disk, and network connectivity. The network fabric is mostly popular IB. The nodes communicate with each other by the message passing mechanism.

There are several parallelization approaches like programs, language extensions, APIs which cannot be ideal because of the several reasons like the overhead of the processes, thread synchronization, and communication. Bottlenecks in the parallel computer design, are memory or network bandwidth limitations. The Load imbalance which occurs due to one processor has more work than the others causing them to wait. As per the Amdhal's Law, Serial parts in the program may not be parallelized at all perhaps cause of the limitation during the code conversion.

A typical job in the parallel system can be characterized by the attributes like job run time length, number of slots, and amount of memory. Additionally, wall time, node hour, core hour, node day, storage requirements, is also necessary to understand the total turn-around time of the application. The HPC infrastructure connected with fast Infiniband (IB) data network, which solves the multi-threaded and distributed parallel computations. The HPC runs long-running jobs that require massive compute resources, like many CPUs or memory. HPC systems make use of batch non-interactive jobs to allow many users to run the program at the same time. In a batch system, users describe the workflow of their application, and once a job submitted to the system, it runs independently of any user input until it finishes. In this case, the tasks can be monitored, but there is no interaction.

Compute jobs running on the HPC can operate in parallel using popular frameworks like OpenMP and Message Passing Interfaces (MPI). Many users write and compile their software to run on the HPC,

28 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/challenges-and-opportunities-in-high-performance-cloud-computing/225714

Related Content

Evolution of Fog Computing Applications, Opportunities, and Challenges: A Systematic Review
Hewan Shrestha, Puviyarai T., Sana Sodanapalliand Chandramohan Dhasarathan (2021). *International Journal of Fog Computing (pp. 1-17).*
www.irma-international.org/article/evolution-of-fog-computing-applications-opportunities-and-challenges/284861

Securing Business IT on the Cloud
Bina Ramamurthy (2014). *Security, Trust, and Regulatory Aspects of Cloud Computing in Business Environments (pp. 115-125).*
www.irma-international.org/chapter/securing-business-it-on-the-cloud/100840

A P2P Architecture for Social Networking
Michele Tomaiuolo, Monica Mordoniniand Agostino Poggi (2019). *Applying Integration Techniques and Methods in Distributed Systems and Technologies (pp. 220-245).*
www.irma-international.org/chapter/a-p2p-architecture-for-social-networking/229171

Fog Computing to Serve the Internet of Things Applications: A Patient Monitoring System
Amjad Hudaiband Layla Albdour (2019). *International Journal of Fog Computing (pp. 44-56).*
www.irma-international.org/article/fog-computing-to-serve-the-internet-of-things-applications/228129

Multi-Layer Token Based Authentication Through Honey Password in Fog Computing
Praveen Kumar Rayani, Bharath Bhushanand Vaishali Ravindra Thakare (2018). *International Journal of Fog Computing (pp. 50-62).*
www.irma-international.org/article/multi-layer-token-based-authentication-through-honey-password-in-fog-computing/198412