Research Plan

Performance analysis of complex parallel and distributed computer systems

Maher F.Alhesen

Abstract:

Quite recently the need for high performance computing is highly on the rise, due to large-scaled applications that call for more capacity and less execution time to perform the required tasks fairly well, the purpose of this study is to define parallel and distributed computing methodologies and identifying related performance metrics, what fields are overlapping and might have a correlation with this kind of systems, also how it might have an impact on its performance and scalability, and thus analyzing performance factors will give us a notion on how to develop the concurrent methods to overcome gaps in speed and efficiency, by minimizing amount of time required for the completion of the whole process stages operations without violating the precedence constraints.

Living In this digital era, almost everything is being connected into a very big network, and new trends started to reserve a space in this congested market, such as big data, cloud computing, grid computing and Internet of things, along with these interdisciplinary areas, more and more speed and performance are needed increasingly in order to accommodate the concurrent technological trends, and this necessity can be met by high performance computing methods which are mainly characterized by parallel and distributed computing, furthermore there are some aspects that have to be considered when talking about high performance computing, is that some emerging and existing fields are overlapping with this methodology and thus affecting speedup and time complexity of the service provided by such systems.

The increasing complexity, heterogeneity and dynamism of the current applications, in addition to the newly emerging environments had absolutely provided further importance and greater focus on parallel and distributed computing, for the purpose of solving new consequent issues and providing superlative progress in quality and performance. As no matter how powerful individual computers will become, there will be indeed the necessity of harnessing the power of multiple computational units. Even though computer scientists are constantly performing advancements in computer hardware and specifically processors, still we are living in a world full of parallel operations which undoubtedly cannot be processed by a single CPU.

As for traditional uses of distributed computing, some graphics applications like 3dsmax may require days or months to transform some 3d mesh into a rendered image or animation using a single PC, while the time for this operation can be significantly reduced using parallel and distributed computing, by assigning each decomposed task to a distinct thread and tasks as a result are wholly sent to different nodes in the system to be processed separately. Researchers and scientists of astrophysics, nuclear engineering and many other areas as well, do need large capacity and vast numbers of calculations in some occasions, when in fact HPC and parallel computing is the definitive solution. On a larger-scale basis there are several complex and interrelated events that are happening simultaneously, like real world phenomena such as traffic or weather and climate change, and its related simulations are preferably have parallel computing much more in demand compared to serial computing. On the other hand high performance computing is no more limited to scientists, academia, or engineers, yet the use has extended to include more parties and further applications, because of the emerging of some trends like big data or cloud computing.

Impacts of parallel and distributed computing have been experienced in many areas, as some of the fields involved are benefiting enormously from PDS in various aspects including improvement in productivity, minimizing the need for prototypes by implementing simulations alternatively, besides that it has an influence on economy overall, all these impacts will be detailed more clearly in the research, relatively speaking, it is worth mentioning that these impacts might have different influence on either computer science, businesses, or society overall, thus depicting these forms of interrelated effects will help us imagine what capabilities can be provided, and what innovative solutions can be introduced in the future to further enrich the human experience with a more valuable systems.

In this research, as of using an empirical research approach, performance analysis of parallel and distributed computing will be defined by studying the contemporary PDS approaches, analyzing changes which have been implemented over the past few years, implementing different operations and simulations using existing HPC tools to verify and examine the outcomes, and finally suggesting effective measures on existing systems, tools or algorithms. While some theoretical questions can be fundamentally answered, likely there is a room for improvement by constructing systems with more adjustments to accommodate larger-scale applications.

The development of processors and making faster clock speeds is difficult due to the cost and heat limitations, besides that it is expensive to attach large memory on a single processor, therefore the use of parallel computing is the ultimate solution. For larger-scale applications, using distributed computing is exceedingly valuable, regardless of releasing superior processors by hardware vendors, various hardware dedicated accelerators or even integrated flash memories for data caching, as eventually combining both systems of parallel and distributed computing will divide tasks' processing into distinct nodes, and so gathering results once computations are carried out independently from widely dispersed locations in some cases. However, in this study the distinguishing between parallel and distributed systems, in addition to the challenges that continuously face these methodologies will be addressed, like concurrency control, fault tolerance, algorithmic efficiency, and communication. Trying to separate these issues into different pools and investigating of what kind of solutions or updates can be installed to smarten the adequacy of the whole system, moreover defining of what criteria upon which to choose among diverse parallel computing models.

Any kind of performance measurement is fundamentally based on both hardware architectural design and software algorithmic complexity, thus the main pillars for performance improvement basically relies on various elements which are designated in accordance to hardware and its evolution, mainly processors and memory models as aforementioned, which basically represents how much latency can be avoided, how much bandwidth it can handle, and how much capacity the system units already accommodate. Moreover the homogeneity or heterogeneity of systems can play an essential role in identifying the level of performance and communication interoperability between devices. Hovering over the ways to determine the proper architectural design and the capacity needed will be included in the research, as generally it depends on the application itself.

Full Analysis of performance metrics for both processors and memory, in addition to the performance metrics of parallel computing, as well as the operational performance metrics of distributed computing will be explained properly, as this will lead to a full understanding of what approaches to be used for further performance upgrade.

Networks and interconnections efficiency for distributed systems are considered to be very critical regarding high performance evaluation, as apparently, variables with respect to availability of sufficient bandwidth, network routing optimality and traffic forwarding speed, take part in determining the performance of the system as a whole, along the similar vein, if we have a glance at cisco devices, we can notice the importance of the network devices physical characteristics like number of ports, interfaces speed, compatibility, capacity and some other related features. On the other hand, the role of the IOS, algorithms, protocols and data-carrying techniques used in network devices cannot be disregarded, for its true effects on networks operations efficiency. Software modules are to be developed and integrated collaboratively with hardware resources, to effectively exploit performance and scalability which can be acquired through the different components of hardware platform, and as a result getting greater delay reduction.

Finally if we dig deeper in the soft components of a system, there are various approaches with which we can enhance speedy of communication and therefore time required for the whole data processing, at the same time, we can realize the role of some functions and algorithms for tasks re-adjustments and re-scheduling for better performance and further scalability enhancement, as by using the application-driven approach we can analyze the inner application structure, determining data dependencies and visualizing norms, through which, we can make a vital difference in data processing by parallel and distributed computing systems. Related areas involve resources allocation, prioritizing and task scheduling, and load balancing will be intensively examined throughout the study.

Computation power is increasing constantly, as there is an extreme shift towards multicore technologies for the high capabilities they can afford, however as of using such technologies, new challenges in software development and algorithms optimization techniques emerged for the purpose of gaining more performance, by exploiting parallelism in all system levels including processors and the addition of dedicated accelerators, and utilizing the maximum cores as possible, as systems heterogeneity of hardware platforms have an explicit implication on the efficiency we can achieve through parallel and distributed computing, through the research the impact of different parameters on performance metrics will be examined, taken into consideration that knowing adequate information about processors, memory models, and resources of the system will help applications examiners and programmers to obtain the maximum performance, adding that by using some algorithms with a high degree of parallelism, there will be an efficient use of large number of processors and cores, and so adding more scalability to the system.

A critical factor that represents transition coordination time of data and tasks to be processed, is the parallel overhead, which inclusively caused by synchronization and data communication, and at this point we can realize the importance of the parallel programs design to exploit system resources efficiently, and reduce the time required for some parallelism procedures regarding data communication cost, data dependencies, costs of creating and scheduling processes, and make an optimal use of processors in the system by load balancing and hyper-threading. Many issues will arise when designing a parallel system or deciding of what to choose among the different parallel systems programming models, through the research there will be an analysis of the issues to which parallel programs designing are exposed, and after selecting and understanding the application and the related problem, there will be an investigation of what programming models are the most common? What limitations are there for each model, including advantages and disadvantages?

Over and above that, when getting to solve a problem by creating parallel solutions, there are some questions we have to investigate like identifying whether the problem can be parallelized? Checking what partitioning or decomposing approach can be implemented? If there is a need for synchronization or data communication? And checking for any existence of data dependency within the problem analysis, as finally through analyzing the problem and defining its related variables, it will be more convenient and reasonable before starting to parallelize a system, some cases will be introduced in the research to further explore different problems aspects, identifying what methodologies and models to be used and seeking for any optimization mechanisms that can be implemented in the solution.

Despite the fact that there are some factors that affect performance and scalability, either by hardware related factors including, Memory-CPU bus bandwidth, communications network bandwidth, processor clock speed, or by adding more resources like increasing number of processors. However there are still various approaches by which we can decrease memory usage thus allowing more complex environments to be addressed, or speeding up the system by developing parallel programs in order to make an efficient use of the parallel system and make further optimization for the whole system, as usually, communication time is very high compared to processing time.

Tightly coupled and loosely coupled systems are actually old and new terminologies, and nowadays more industries are turning to high performance computing technologies in order to tackle big data analytics workloads more efficiently, and we are apparently heading towards a new era in supercomputing.