Emerging Trends in management information systems: The use of new Parallel Distributed Technologies

Tagreed S. Alsulimani[1], Najeh K. Almazmomi[2] and Khedija B. Arour[3]

College of Business, Department of Management Information Systems, University of Jeddah, Saudi Arabia.

Summary
New processors like Graphical Processors Units (GPU) and multicore processors underwent significant transformations in the field of information technology, represented by high-performance processors characterized by high parallelism and nonvolatile memory with hybrid storage hierarchies. Today, discovering information that identify relationships among sets of items from a large and various data is an important problem in any information system. Hence, utilizing new processors from parallel architectures has been a viable means for improving information system performance.
In this paper, we provide a survey of different techniques and solutions used for parallel and distributed management information system. We discuss the motivation for and challenges involved when we use the parallel processors like multicore and GPU also distributed technologies.

Key words:
Multicore processors, GPU processors, Information System, Large scale systems.

1. Introduction

Like all the other fields of data processing, the modern information systems have integrated the results of the advanced technologies of the last decades. These systems must collect and process data which it will be necessary exploited by using different approaches and solutions data Mining techniques. For example, discovering knowledge which trends to find interesting association or correlation relationships among large amounts of data is one of these techniques. The overall performance of processing data becomes the focus problem mostly with the volume and the variety of this data. This step is expensive with high demands for computation and data access. Parallel and distributed computing seem to have a natural role to play since parallel computers provides scalability. In this paper, we examine the issue of information systems under the new technologies. Especially that currently most of the machines and servers have multi-core processors and graphical processors. We propose a we present a state of the art of the existing systems and how new technologies are exploited.
In this paper, we provide an extensive survey of techniques and architectures proposed for management information system. We first discuss the different notions and concepts involved in distributed and parallel information systems. Also, we clarify the terminology used in this research field (Section 2). In Section 3, we discuss some existing solutions both parallel CPU and GPU have evolved over years to identify important trends. We then analyze the research works from several perspectives to highlight the principles problems.

2. Concepts and notions

2.1 Management Information System MIS

Management Information System (MIS) is the use of information technology, people, and business processes to record, store and process data to produce information that decision makers can use to make day to day decisions. There are many needs for an MIS like:

- Decision makers need information to make effective decisions.
- Record keeping
- MIS systems facilitate communication within and outside the organization

They exist many types of information systems like:

- Decision Support Systems (DSS)
- Management Information Systems (MIS)
- Transaction Processing Systems (TPS)

The following figure presents the different types of Information Systems (see Figure 1).

Manuscript received June 6, 2020
Manuscript revised June 20, 2020

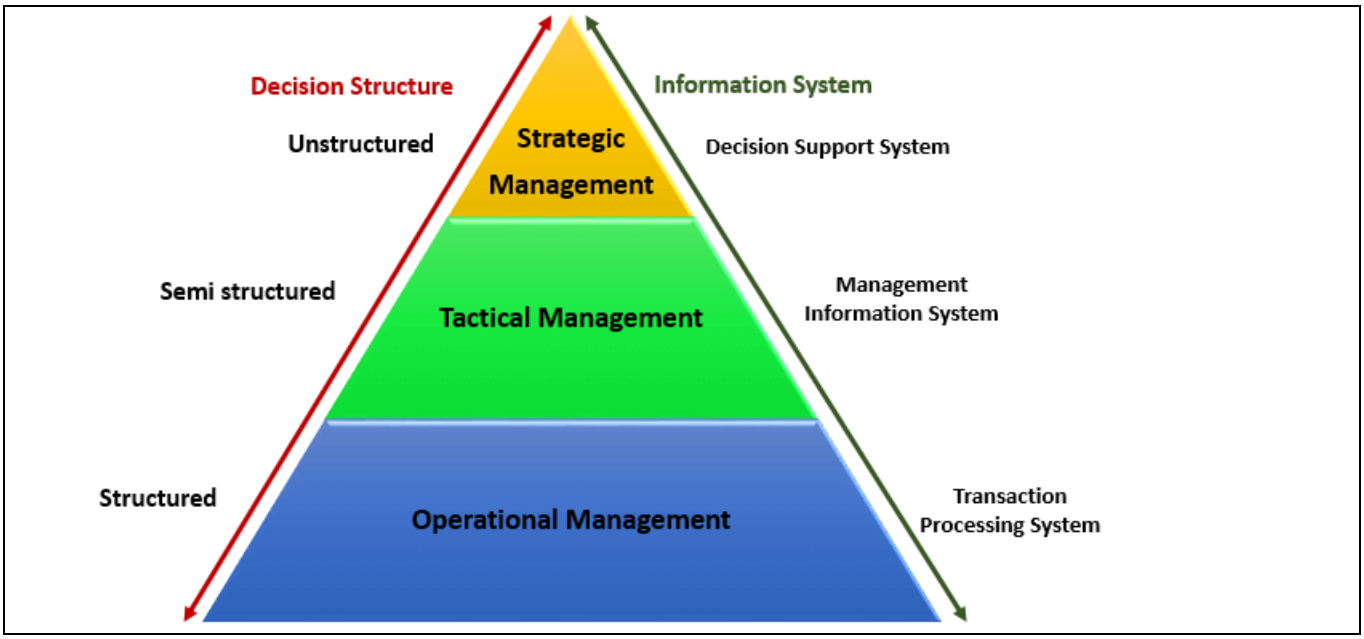


Fig. 1 Types of Information systems

2.2 Parallel and distributed architectures

Parallel and distributed computing emerged as a solution for solving complex problems by first using multiple processing elements and then multiple computing nodes in a network. The transition from sequential to parallel and distributed processing offers high performance and reliability for applications. But it also introduces new challenges in terms of hardware architectures, technologies for communication, and the system design. Parallelism is achieved by leveraging hardware capable of processing multiple instructions in parallel. Different architectures exploit parallelism to increase the performance of a computing system, depending on whether parallelism is realized on data, instructions, or both. The development of parallel applications often requires specific environments and compilers that provide transparent access to the advanced capabilities of the underlying architectures.

Parallel architectures

Parallel computing is a type of computation in which many calculations or the execution of processes are carried out simultaneously. Large problems can often be divided into smaller ones, which can then be solved at the same time. The three models that are most commonly used in building parallel computers include synchronous processors each with its own memory, asynchronous processors each with its own memory and asynchronous processors with a common memory called a shared memory [1]. Different types of parallelism can be exploited

Parallelism in Hardware (Uniprocessor)

- Parallelism in a Uniprocessor – Pipelining
- SIMD instructions, Vector processors, GPUs
- Clusters of processors
- Multiprocessor: Symmetric shared multiprocessors, multicores processors, …

Parallelism in Software

- Instruction level parallelism
- Task-level parallelism
- Data parallelism

Depending on data and/or instructions, we can exploit the four following types (see Figure 2).

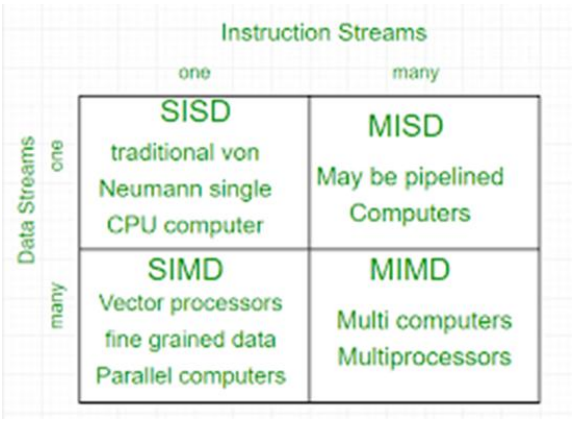


Fig. 2 Parallelism types

Multicores Processors

A multicore processor is an integrated circuit (IC) to which two or more processors have been attached for enhanced performance, reduced power consumption, and more efficient simultaneous processing of multiple tasks [1]. The following figure (see Figure 3) presents the multicore processor.

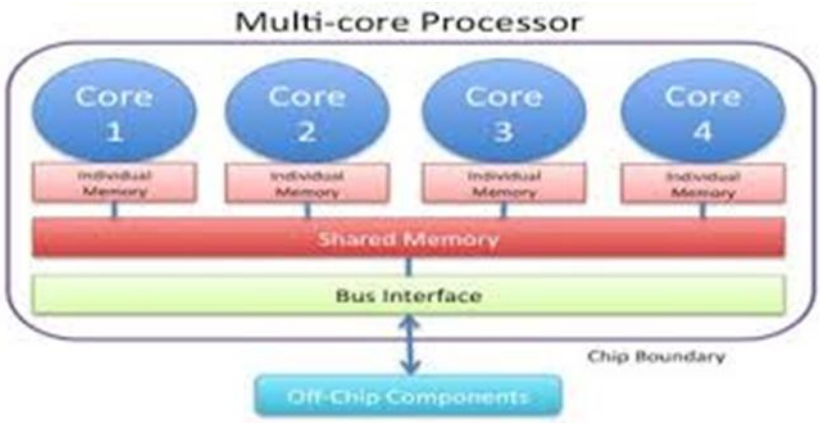


Fig. 3 Mulicore processing

GPU processors

A graphics processing unit (GPU) is a computer chip that performs rapid mathematical calculations, primarily for the purpose of rendering images but they are technically capable of doing more.

Graphical processor units (GPU) have always been the subject of a lot of investment and improvements, given the high demands for performance of these components and the competition that exists between different producers like Nvidia and ATI/AMD. The graphical processing is massively parallel. This is why GPUs are suitable for parallel computing. These GPUs are in fact composed of a large quantity of processors which allow the execution of instructions on a large number of data simultaneously, and thus considerably reduce the graphics processing times. GPU computing is the use of a GPU (graphics processing unit) as a co-processor to accelerate CPUs for general-purpose scientific and engineering computing.

The following figure (see Figure 4) gives us an idea about the processing under GPU.

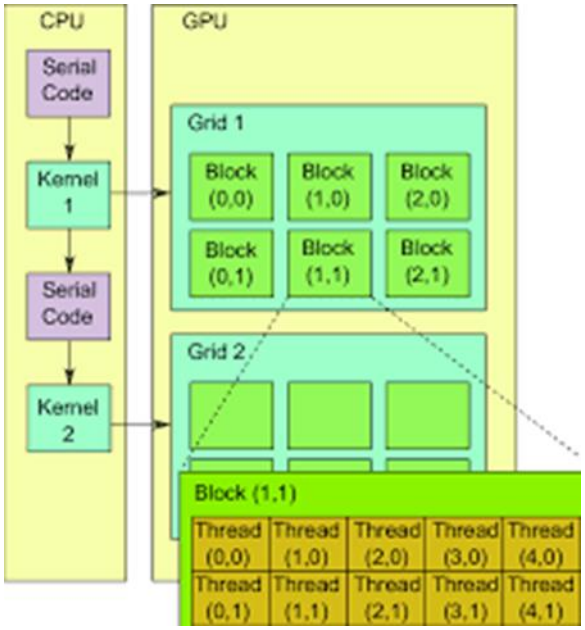
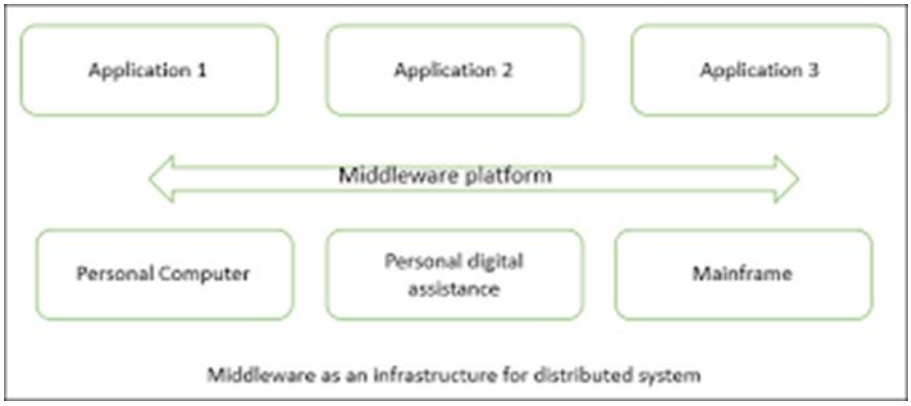


Fig. 4 GPU processing

Distributed architecture

In distributed architecture, components are presented on different platforms and several components can cooperate with one another over a communication network in order to achieve a specific objective or goal. ... It sits in the middle of system and manages or supports the different components of a distributed system. We present below a figure about distributed architecture.



Fig, 5 Distributed Architecture

With this architecture, information processing is distributed over several computers rather than confined to a single machine. Then a distributed computing is a network with multiple computers to achieve a goal of a good performance and a large used data. Each computer in the network performs a section of the overall tasks. This method helps to obtain results faster than using a single computer. It provides advantages such as scalability, redundancy and resource sharing [1]. Client-server and distributed architectures are the most known used architectures.

3. Parallel and distributed Information Systems: state of the art

We discussed the evolution of technologies supporting parallel and distributed processing information systems and we try to introduce the major reference models for designing and implementing these systems.

Unification of parallel and distributed computing allows one to harness a set of networked and heterogeneous computers and present them as a unified resource. Distributed systems constitute a large umbrella under which several different software systems are classified. Architectural styles help categorize and provide reference models for distributed systems. More precisely, software architectural styles define logical organizations of components and their roles, whereas system architectural styles are more concerned with the physical deployment of such systems.

For example, The AMARANTH system designed under Amaranth project [2] provide multi-dimensional, adaptive, assured Quality of Service (QoS) for heterogeneous distributed computing systems. In particular, Amaranth will enable systems to provide guarantees of QoS along multiple

dimensions, with each dimension having several potential levels of potential assured service like Real time deadlines, dependability and application performance.

The authors on their work presented on [8], describe the design, development and deployment challenges facing an implementation of an enterprise-wide distributed web-based Planning, Budgeting and Reporting Control Management Information System for a large public utility organization. The system serves the needs of all departments of the company's General Division of Production.

PASIS is an innovative framework for demonstrating perpetually available information systems that guarantee the survivability of information under malicious attacks or system component failures [2]. PASIS is based on a novel architecture which breaks all information into "chunks" and distributes these "information chunks" in novel ways by using information replication and dispersal methods. This enables PASIS to not have any single point of failure and thereby achieve a very high degree of security and resiliency against failures and attacks.
Performance analysis tools for enterprise distributed real-time and embedded (DRE) systems require instrumenting heterogeneous sources such as application- and system-level hardware and software resources [3].
The authors on [4] work, present a distributed web crawlers system based on Hadoop and used it to do large-scale information management.
Novel database management system architectures based on GPGPU have been proposed [5,6] to meet the challenges of querying large-scale data. Commercial systems such as Massively Parallel Database (MapD) [7] has seen success in the business world. MapD is a next-generation data analytics platform designed to process billions of records in milliseconds using GPUs.
In the work [9], authors discuss medical Big Data analysis in distributed hospital information system, related platforms and technologies of medical Big Data processing, and advanced Big Data processing technologies are detailed.

IBM DB2 BLU [10] is a GPU-aware database will offload some operations to the GPU, like a co-processor. IBM DB2® with BLU Acceleration is a revolutionary technology, delivered in DB2 for Linux, UNIX, and Windows Release 10.5. Using dynamic in-memory columnar technologies, BLU Acceleration enables organizations to perform analytic queries at breakthrough speeds. The following figure give us a detailed architecture of IBM DB2 BLU.

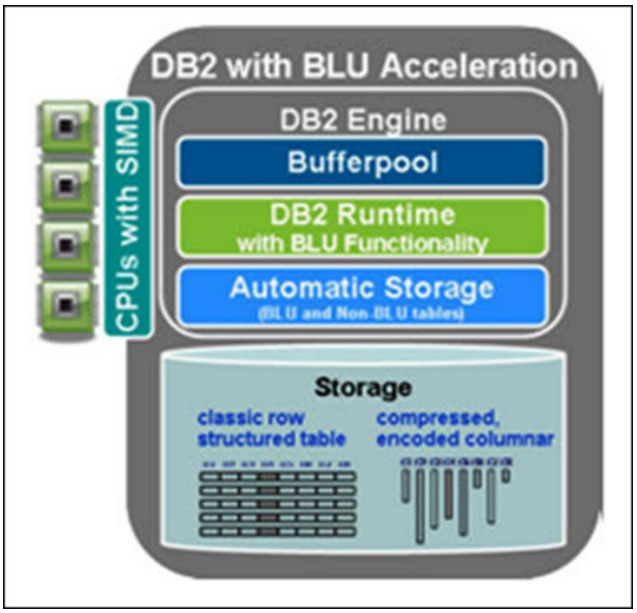


Fig. 6 DB2 engine with BLU Acceleration

Kenneth Ross et al [11] involve the development of algorithms and data structures designed for the efficient parallel execution of generic code fragments. The primary focus is on data intensive operations as would typically be found in an in-memory database engine. Critical research questions include how to design generic multi-threaded operators that can be applied to a range of computations, how to avoid cache thrashing, and how to implement the framework in a way that works on a variety of hardware platforms. Performance improvements in throughput of an order of magnitude are expected relative to naive solutions that suffer from contention. The project aims to achieve performance close to that of hand-tailored expert-written parallel code, with far less coding effort.

4. Synthesis

Parallel computing is a type of computation in which many calculations or the execution of processes are carried out simultaneously. Large problems can often be divided into smaller ones, which can then be solved at the same time. Even given today's rich hardware platforms, computation-intensive algorithms and applications, such as large-scale distributed information systems, are still challenging to run with acceptable response times. Hence, effectively mapping distributed software models to GPU/multicore is a non-trivial endeavor. Then, we must investigate ways of improving execution performance by using suitable models by means of relevant mechanisms, which are suitable for the parallel/distributed hardware.

5. Conclusion

In recent years, CPUs and GPUs are increasingly being seen as indispensable co- processors, instead of substitutes for each other. As a result, heterogeneous computing has been actively researched and utilized in variety of applications ranging from natural sciences to engineering etc. Several challenges still remain, and we believe that the research in near future will provide solutions to them. Many approaches to the acceleration of management of information systems have been presented so far, using either distributed, multicore processor unit (CPU) or Graphics Processor Unit (GPU). Moreover, there is now evidence that Fied-Programmable Gate Array (FPGA) is more likely to become a promising potential for big data processing.

References

[1] John Van Zandt, Parallel Processing in Information Systems: With Examples and Cases. Wiley Professional Computing, John Wiley & Sons Inc (June 1992).
[2] https://www.cs.cmu.edu/~aml/research/dis.html
[3] James Hill; Hunt Sutherland; Paul Stodinger; Thomas Silveria; Douglas C. Schmidt; John Slaby. OASIS: A Service-Oriented Architecture for Dynamic Instrumentation of Enterprise Distributed Real-Time and Embedded Systems. 2010 13th IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing. Carmona, Seville, Spain, 5-6 May 2010
[4] Bo Wen. Application of Distributed Web Crawlers in Information Management System. Informatica (Slovenia), 42 (1), 2018
[5] Breß S, Heimel M, Siegmund N, Bellatreche L, Saake G. Gpu-accelerated database systems: Survey and open challenges. In: Transactions on Large-Scale Data-and Knowledge-Centered Systems XV. Springer; 2014. p. 1–35.
[6] Tu YC, Kumar A, Yu D, Rui R, Wheeler R. Data Management Systems on GPUs: Promises and Challenges. In: 25th International Conference on Scientific and Statistical Database Management. SSDBM. New York, NY, USA: ACM; 2013. p. 33:1–33:4.
[7] Mostak, "An Overview of MapD (Massively Parallel Database), CSAIL, MIT,2012
[8] Christou I.T., Potamianos S. (2007) A Distributed Planning & Control Management Information System for Multi-site Organizations. In: Elleithy K. (eds) Advances and Innovations in Systems, Computing Sciences and Software Engineering. Springer, Dordrecht
[9] Jing-Song Li, Yi-Fan Zhang and Yu Tian, Medical Big Data Analysis in Hospital Information System. IntechOpe, July 20th 20
[10] L. Wu, O. Polychroniou, R. J. Barker, M. A. Kim, K. A. Ross. Energy Analysis of Hardware and Software Range Partitioning. ACM Transactions on Computer Systems, September 2014.