Efficient Big Data Analytics and Management through the Usage of Cloud Architecture

Suryakanthi Tangirala
Department of Accounting and Finance, University of Botswana, Faculty of Business, Gaborone, Botswana
Email: suryakanthi.tangirala@mopipi.ub.bw; ch.kanthi@gmail.com

***Abstract*— Due to continuous growth of information systems to store business data and advent of new areas like mobile computing, users now need to use the enterprise applications through tablets, Smartphone, iPhone, laptops, and desktop computers. Effective decision making can be achieved with efficient information system. With the introduction of Business Intelligence tools organizations can now analyze the raw data and perform various activities like data mining, online analytical processing (OLAP), querying and reporting. Business Intelligence technology will help the managers to make better informed decisions. In order to perform data driven decision making business analytics practices are adopted. Cloud computing in recent days has gained a lot of prominence to store and process the business data. Organizations use cloud storage to manage data as they face challenges in local storage. But cloud also possesses certain challenges due to which organizations at large still find difficult to move to the cloud. Cloud computing can also be used to store and process big data. As Big data need to be analyzed for attaining maximum business value a new data model need to be proposed which have the properties those differs from traditional data model. In this paper we discuss that various challenges faced by organizations to move to cloud. We also propose that how the challenges can be overcome by the organizations so that cloud will be a promising architecture for better information management. Data model for Big data is also proposed in this paper.**

***Index Terms*—business intelligence, big data, cloud computing, data mining, OLAP, Information management**

I. Introduction

Evolution in technology leads to generation of voluminous amount of data. The data generated by the organizations need to be managed effectively in order to provide better informed decisions. Organizations have to transform to a model where the technology and business factors will influence with the transformation. Cloud based model [1] will provide such transformation which will address the information management solutions. There are various processes for information management which needs solutions. Few of them include data backup and recovery, archiving the data, data retrieval and security. Information management solutions should be capable for the cloud based architecture. With the advent of areas like mobile computing, organizations need to provide the information for various devices such as smart phones, tablets and various devices. Cloud computing architecture [2] should be able to deliver all the resources as services.

II. Data, Information and Knowledge Management

In a business organization the day to day data is associated with the functional process. The data is entered into an information system and stored in the database. A Sales information system will generate the sales data. This data stored in the sales database and can be queried later for the related report generation. These data has to be in the well structured format and also should be able to recover on failures. As we know that when data is applied for a purpose which gives specific meaning can be called as information. Data and information can be used interchangeably.

In a business context, knowledge is often linked to strategic levels of management [3] and long-term business planning, where it is associated with having a head for business or business flair. However, knowledge vital to an organization's success can come from any level within it, and needs to be recognized as an important part of organizational assets. It combines information, experience and insight into a mix that is unique to every employee. It is this mix of understandings, based on personal knowledge at a tacit level that creates the strengths and at times the vulnerability of organizations. It is important for organizations to recognize that holding knowledge at the tacit or hidden level can only have value where people are isolated from everyone else in their decision making [4]. This is neither realistic nor good business practice.

The decision making is the first key step for any organization. The second important step is to analyze and evaluate the various options. One of the challenge that organizations today face about the speed with which the mangers have to take decisions. Faster decision helps organizations to improve at various phases. Detailed investigation on information cannot be made as the information is not appropriate or available. Management decisions are made at three broad levels within the organization, and each type of decision has its own uniqueness. Operational decisions which affect the day to day business of the organization are to be of short term and are made frequently. Tactical decisions which are generally made by the middle managers are involved with

Manuscript received March 17, 2016; revised November 18, 2016.

doi: 10.12720/jait.7.4.302-307

implementation of the policies within the organization. Strategic decisions are made by top management, which affect eh whole organization business and usually they are of long term which are infrequent. All these decisions will require information, but the type of information that is needed will be different for each level of decision making.

There are various benefits of the knowledge management [5] where in new markets can be identified with high level of intelligence. Responsiveness to market needs can be enhanced by external knowledge. Creating innovative products and improving existing products based on the customer knowledge.

III. Cloud Computing

Cloud computing is a model to provide shared pool of computing resources like networks, storage, applications and various services that can be scalable as per the requirements of the organization with minimal management effort or interaction with the service provider [6]. The cloud characters tics are so flexible that organizations can adopt it with its broad range of services. Cloud provides on-demand self service where in the consumers have their computing capabilities on demand and resources are provided without any human interaction.

Cloud also provides broad network access wherein the resources so that they can be accessible on various platforms such as mobile phones, tablets, laptops, and workstations. Resource pooling allows the cloud computing resources are shared by multiple consumers. Rapid elasticity feature allows the capabilities which can be easily scaled up or down based on consumers demand. Pay as you go or by subscription are ways where in the consumers pay for the services availed by the cloud.

Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [7, 8, 9, 10, 11] are the three broad categories of service models available for cloud architecture. NIST defines them as

A. *Infrastructure as a Service (IaaS):*

It is capability provided to the consumer to provision processing, storage, networks, and other fundamental computing resources so that the consumer can deploy and run arbitrary software, which can include operating systems and applications [7]. The consumers do not require managing and controlling the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls). Instead of consumers purchasing servers, software, data centre space and network equipment, they buy resources from providers as fully outsourced service [8]. Example of IaaS is Amazon Web Services. IaaS is termed as most basic level of Cloud Computing service model [9].

B. *Platform as a Service (PaaS):*

It It is capability provided to the consumer to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. Consumers do not require managing or controlling underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment. PaaS is a way to rent hardware, operating systems, storage and network capacity over the Internet. It allows the providers to rent virtualized servers and associated services for running existing applications or developing and testing new ones. Example of PaaS includes Microsoft Azure Platform [10], Google App Engine [9].

C. *Software as a Service (SaaS):*

It is consumers' capability to use providers' applications running on cloud infrastructure. Provider provides an application to the consumer as a service on demand which includes the hardware infrastructure and the software product [8]. SaaS also termed as Application as a Service [12] is usually accessed by rich web-based interface. Applications are also accessed through program interface. An exception is where user-specific application configuration is possible. Examples of SaaS include Google Docs, Gmail and Salesforge.com. The service model for cloud is shown in Fig. 1.

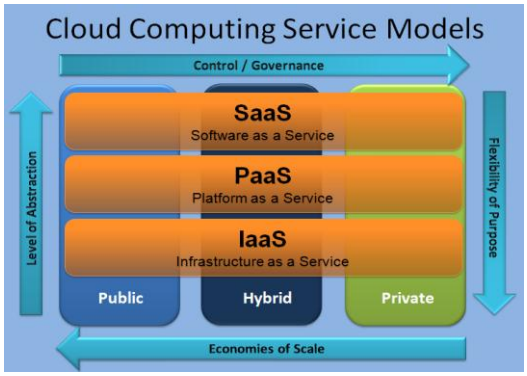

Figure 1. Service models for cloud computing

IV. Deployment Model for Cloud Computing

There are a number of different deployment models for cloud computing. A deployment model is a particular method of delivering cloud computing service. Most commonly used deployment models for cloud computing found in scholarships are Private cloud, Community cloud, Public cloud and Hybrid cloud [7, 8, 9, 10, 11] as described in sections below:

A. *Cloud Deployment Models*

Private cloud: Cloud Infrastructure is provisioned for the exclusive use of an organization. That organization comprises multiple users such as business units. The infrastructure may be owned, managed and operated by the organization or by a third party or some combination of both. It may exist on or off premises. In case the infrastructure is provided by the organization then it is usually provided by internal Information Technology (IT) or Information Systems (IS) department. As both user and

provider are part of a same organization, private clouds allow the user greater control over quality of service provided by the cloud. This control comes at price because organization has to bear full cost of cloud infrastructure.

Public cloud: The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider. Like many other scholars, Microsoft's view on difference between public and private clouds are same; based on whether cloud infrastructure is dedicated to a single organization (private cloud) or shared between many distinct organizations (public cloud) . Examples of Public cloud are Amazon Web Services and Microsoft Azure. In [12] it is referred data centre hardware and software as Cloud and they use the term Private cloud to refer to internal data centers of a business or other organization, not made available to the general public. In contrast, when the cloud is available to public in a pay-as-you-go manner, then they term it as Public cloud.

Community cloud: NIST defines community cloud as the cloud infrastructure provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns like mission, security requirements, policy, and compliance considerations. Like public cloud, it may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises. Organizations in the community collaborate for the purpose of a particular mission or concern. An example of community cloud is Google Gov.

Hybrid cloud: It is actually a composition of two or more distinct deployment models (private, community, public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability like cloud bursting for load balancing between clouds. Usually Hybrid cloud deployment model is chosen to use public cloud's capability to capture task that cannot be run easily on private cloud. Example of hybrid cloud is a private SaaS application that is based on a public IaaS. Hybrid cloud is usually focuses on driving workload to the public clouds to meet the stringent quality of service requirements. That is why in many cases Hybrid clouds are found to be private clouds, which turn to the capacity of public clouds for peak demand. The cloud deployment model is shown in Fig. 2.

Figure 2. Cloud Deployment models

V. Cloud Architecture for Information Management

In this section we describe the cloud architecture which is needed for the information management. Organizations do not need to manage the data and infrastructure locally. They can transfer the data through networks to the cloud provider as discussed in the previous sections. In the current business era information management can be carried out by various devices which include smart phones, laptops, tablets and Pcs. So cloud providers have established data centers. Fig. 3 shows that data storage and databases are stored in data centre A. In data centre B the email server and office productivity data are stored. Various applications like financial, human resource management, enterprise applications will be stored in Data centre C. all these data centers are protected with firewall as security and is connected to internet. The enterprise users through various devices can access all these applications as all these are available as services on the cloud. Enterprise applications like human resource management, financial applications, sales and marketing applications, and few specialized applications can be accessed by the clients through cloud computing. This makes ease of organizational data management as they need not worry about the technical aspects like infrastructure management, increase in storage space and various aspects.

Figure 3. Cloud data centers for information management

VI. Challenges of Cloud Computing

Though there are many benefits of adoption of cloud computing to the organizations, there are certain challenges that still cloud implementation is not adopted by most of the organizations. In this section, we discuss about few of major challenges. In the next section, we propose the solutions to overcome those challenges. In order to efficiently manage the information organizations should adopt cloud architecture by implementing the suggested solutions. Security and privacy [13] is one of the major challenge that cloud computing should able to address for the adoption by business organizations. As the data is residing the data centers remotely, how the cloud providers will provide the security to the data is a major question that organizations need to focus on. Connectivity [14] and access is another challenge as cloud access requires high speed networks. Reliability [15] is yet another challenge as the organization data moved to the cloud.

VII. Big Data Analytics

Big data [16] refers to those large data sets where the traditional database management tools no longer can efficiently handle the data for storage or for processing. The volume of data we deal has grown from Terabytes to Petabytes. We need to capture, store, share, search, analyze and visualize the data. The big data challenges can be managed by using a NoSQL databases like Hbase, MongoDB, Cassandra. Distributed computing systems like Hadoop [17] need to be employed. The traditional ways of data model is no longer applicable and we need to design a new data model which are non-relational, distributed, horizontally scalable and schema less. Traditional database management systems focus on resolving the complexities arising in the schema based data. But today data need a schema less model where it can handle unstructured or semi structured data. The next section focus on the Data model proposed for big data processing.

VIII. Data Model for Big Data on Cloud

In this section we propose a data model for big data information management. As the business data grows computing requirements become more and more complex in nature. In order to process these data a new computing paradigm called MapReduce which can reduce the data using key value pairs. The data is stored in cloud architecture and using the data model the data can be reduced using this data mode. Let us consider an enterprise data and their operational data is stored in the server. These data can be structured, unstructured or semi structured data. The data is divided into block of chunks using the Hadoop and then sent to MapReduce function to further reduce the data into blocks. The data is then sorted and shuffled to produce the intermediate data. Again an MapReduce function is passed to get final Reduced output data. This data can be used for efficient decision making by the organizations.

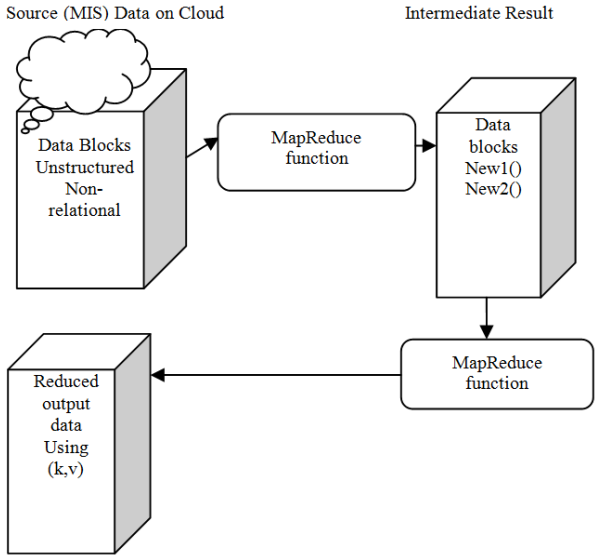

Figure 4. Data model for big data processing

IX. Benefits of Using This Data Model

While storing this data we use NoSQL databases for querying the data. The reason behind the use of NoSQL system is that data is growing rapidly. Data grows in columnar way. Document and tuple data, hierarchical data are the characteristics of the data suitable for NoSQL systems. But this system is helpful for data scalability and efficient query capabilities.

X. Proposed Architecture and Solutions to Overcome the Challenges

We propose cloud architecture for efficient information management. Organizations use Management Information Systems (MIS) for their day to day business activities. All the information is stored and processed locally within the organization. All the security measures are taken care to protect the data. The infrastructure for the organization is either maintained by internal people or can be outsourced to an organization. But when the size of data increases, the organization need to invest on the infrastructure and scale to the required level. In order to solve these issue organizations need to choose the cloud provider for the MIS as it auto scales up to whatever level the organization needs to store and process the data. As mentioned in section 6 security and privacy is a major challenge. To overcome this challenge organizations need to encrypt these data and then stored in the cloud servers. When the user needs the data decryption can be applied. The Service level agreements should be strictly imposed so that the cloud providers protect the data at various levels. The connectivity issue can be resolved by providing high speed bandwidth networks. The following architectures provide a secured architecture for the cloud for better information management. The best way to provide security on the cloud is to provide Secure Socket Layer (SSL), and Transport Layer Security. We have to make sure that no sessions are being altered by any form of attacks. The network can made secure at lower level by providing the Internet Protocol Security like IP Sec.

Figure 4 provides the secured way of users transmitting the information over a secured cloud which provides IP security. The data server resides in one of the cloud data centers. For example, Fig. 4 shows that the MIS server of an enterprise is with the cloud provider. All the enterprise data is sent to the server via secured networks. Users at the enterprise level generate data and then using an efficient encryption tools transmit the data over the secured networks which provides internet protocol security. At the other end the cloud provider process the data, store the data in secured way to ensure that the user is not worried about the confidentiality of the data. In this way the information can be efficiently managed by the usage of cloud architecture.

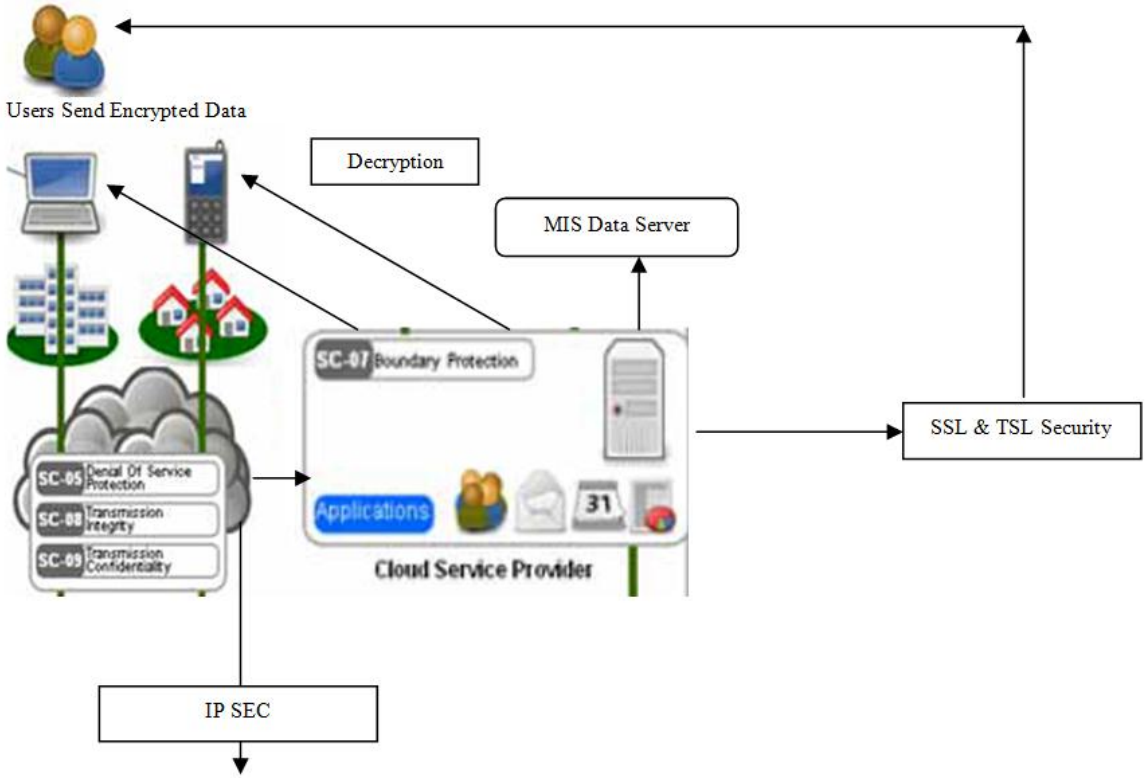

Figure 5. Secured information management by using cloud architecture

XI. Conclusion And Future Work

Cloud computing is a promising architecture for various applications which is used by the enterprises. Efficient information processing and managing can be achieved by the usage of cloud architecture. Organizations therefore should adopt cloud computing for the storage, processing and managing the enterprise data. Before adopting the cloud computing certain issues like privileged user accessibility, regulatory compliance, data location, data segregation methods, recovery of data, investigative support and long term viability of the cloud provider should be carefully studied. The service level agreements must be made after addressing all the issues. There are various solutions that cloud providers can adopt to address the issues which are of major concern by the enterprises for cloud adoption. By providing transparency in the services offered by the cloud providers organizations trust that cloud is an efficient architecture for better information management.

Cloud computing to attract all the organizational levels of trust and security a standardized model and security aspects at various service levels need to be improved. Future work may be extended to propose a standard cloud model which overcomes the various challenges need to be addressed to the enterprises for better information management.

References

[1] R. Buyya, C. S. Yeo, and S. Venugopal, "Market-oriented cloud computing: Vision, hype, and reality for delivering IT services as computing utilities," in *Proc. 10th IEEE International conference on High Performance Computing and Communications*, Dalian, Sept. 25-27, 2008, pp. 5-13.

[2] R. Sreekanth, R. R. Gondkar, and R. S. R. Babu, "Cloud computing architecture to process big data healthcare information for healthcare industries," *International Journal of Applied Engineering Research*, vol. 9, no. 21, pp. 4905-4911, 2014.

[3] R. Amit and P. J. H. Schoemaker, "Strategic assets and organizational rent," *Strategic Management Journal*, vol. 14, no. 1, pp. 33–46,1993.

[4] S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang, and A. Ghalsasi, "Cloud computing—The business perspective," *Decision Support Systems*, vol. 51, no. 1, pp. 176–189.,2011.

[5] C. W. Holsapple and Wu, *A Resource-Based Perspective on Information Technology, Knowledge Management, and Firm Performance.* In Y. K. Dwivedi, B. Lal, M. D. Williams, S. L. Schneberger, and M. Wade (Eds.), Handbook of Research on Contemporary Theoretical Models in Information Systems. IGI Global. [Online]: http://www.igi-global.com/chapter/resource-based-perspective-informationtechnology/35836, 2009

[6] S. Ahson and M. Ilyas, *Cloud Computing and Software Services: Theory and Techniques*, Boca Raton, FL: CRC Press, 2011.

[7] P. Mell and T. Grance, " The NIST definition of cloud computing", *Special Publication (NIST SP) - 800-145*, pp. 7, Sept. 28, 2011.

[8] G. Conway, "Introduction to cloud computing". *White Paper, Innovation Value Institute*, Jan 2011.
[9] A. Marinos and G. Briscoe, "Community cloud computing," in M. G. Jaatun, G. Zhao, & C. Rong, Eds., *Cloud Computing* vol. 5931, pp. 472-484). Berlin, Heidelberg: Springer Berlin Heidelberg, 2009
[10] B. Halpert, A*uditing Cloud Computing: A Security and Privacy Guide*, 1st ed., Wiley , 2011
[11] *The Cloud: Changing the Business Eco System*. KPMG, India, 2011.
[12] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, et al., "A view of cloud computing. *Communications of the ACM*," vol. 53, no. 4, pp. 50–58, 2010.
[13] T. Mather, S. Kumaraswamy, S. and Latif, *Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance*, O'Reilly Media,2009
[14] G. Avram, "Advantages and challenges of adopting cloud computing from an enterprise perspective," in *Proc. 7th International Conference Inter disciplinarily in Engineering*, Romania, Oct. 10-12, 2013, pp. 529 – 534.
[15] R. Buyya, A. Goscinski, and F. Broberg, "Introduction to cloud computing," *In Cloud computing: principles and paradigms*. Hoboken, N.J.: Wiley, 2011.
[16] Big data. [Online]. http://en.wikipedia.org/wiki/Big_data
[17] Hadoop apache. [Online]. https://hadoop.*apache.org/*
[18] S. H. H. Madni, M. S. A. Latiff, V. Coulibaly, and S. M. Abdulhamid. (2016b). "Resource Scheduling for Infrastructure as a Service (Iaas) in Cloud Computing: Challenges and Opportunities." *Journal of Network and Computer Applications*, vol. 68, pp. 173-200.
[19] S. M. Abdulhamid, S. M. Abd Latiff, and M. B. Bashir, "Scheduling techniques in on-demand grid as a service cloud: A review," *Journal of Theoretical & Applied Information Technology*, vol. 63, no. 1, pp. 10 -19, 2014.
[20] S. H. H. Madni, M. S. A. Latiff, and Y. Coulibaly, "An appraisal of meta-heuristic resource allocation techniques for Iaas cloud," *Indian Journal of Science and Technology*. vol. 9, no. 4, 2016.
[21] S. M. Abdulhamid and M. S. Abd Latiff, "League championship algorithm based job scheduling scheme for infrastructure as a service cloud," in *Proc. 5th International Graduate Conference on Engineering, Science and Humanities (IGCESH2014)*, Malaysia, Aug. 19-21. pp. 381-382, 2014.
[22] S. M. Abdulhamid, M. S. Abd Latiff, and I. Ismaila, "Tasks scheduling technique using league championship algorithm for makespan minimization in Iaas cloud," *ARPN Journal of Engineering and Applied Sciences*. vol. 9, no. 12, pp. 2528 – 2533, 2014.
[23] S. M. Abdulhamid, M. S. A. Latiff, S. H. H. Madni, and O. Oluwafemi, "A survey of league championship algorithm: Prospects and challenges," *Indian Journal of Science and Technology*. vol. 8, no. 3, pp. 101-110, 2015.
[24] M. Abdullahi, M. A. Ngadi, and S. M. Abdulhamid, "Symbiotic organism search optimization based task scheduling in cloud computing environment," *Future Generation Computer Systems*, vol. 56, pp. 640-650, 2016

Dr. S. Tangirala earned her master's degree in computer applications in 2006 from Andhra University, Visakhapatnam, India and doctoral degree in 2014 from Lingaya's University, Faridabad, India. She has worked for around 2 years in software industry and has been teaching for 3 years at University Level. She was Assistant Professor of Computer Applications at Lingaya's University and worked as Fellow at Botho University, Gaborone, Botswana. Currently she is working with University of Botswana. She has 12 research papers to her credit in various international conferences and journals. Her current research interests include Artificial Intelligence, Natural Language Processing, Machine Translation, Big data analytics and Theory of automata.