ISSN Application No. 21650

Journal of Cloud Computing and Data Base Management

Volume No. 10

Issue No. 1

January - April 2024

ENRICHED PUBLICATIONS PVT. LTD

S-9, IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR-5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-47026006

ISSN Application No. 21650

Journal of Cloud Computing and Data Base Management

Aims and Scope

Journal of Cloud Computing and Database Management is a peerreviewed Print + Online journal of Enriched Publications to disseminate the ideas and research findings related to all sub-areas of Computer Science and IT. It also intends to promote interdisciplinary researches and studies in Computer Science and especially database management and cloud computing maintaining the standard of scientific excellence. This journal provides the platform to the scholars, researchers, and PHD Guides and Students from India and abroad to adduce and discuss current issues in the field of Computer Sciences.

ISSN Application No. 21650

Journal of Cloud Computing and Data Base Management

(Volume No. 10, Issue No. 1, Jan - Apr 2024)

Contents

Green Cloud Computing: An Environment Friendly Approach Of Computing

Shuchi Srivastava
Assistant Professor
Northern India Engineering College Phone: 9716974267
E-mail : myself_shuchi008@rediffmail.com

A B S T R A C T

Cloud Computing is a technology by which a shared pool of computing resources can be easily and conveniently accessed by the users from anywhere. These resources include networks, servers, storage, applications, and services connected over a network. It uses huge data centers and huge clusters for the same .With the use of this technology the organizations can lower their infrastructure cost and focus more on their core business. However with the growing use of cloud infrastructure the energy consumption of data centers has increased tremendously that causes a major threat to the environment.A high carbon emission from these centers are not environment friendly.

Energy shortages and global climate change are effecting the environment adversely so the power consumption of these data centers is a critical issue. Therefore, green cloud computing solutions are required which can save energy, thereby reducing operational costs. These energy-efficient solutions are required to minimize the impact of Cloud computing on the environment.

Nowadays, thousands of e-commerce transactions and millions of web queries take place in a day. With this increase in the number of transactions there is a huge traffic on clouds and a lot of resources are used. In this paper we will discuss the need of Green IT solutions and how green cloud computing can reduce the adverse effects of cloud computing and benefit the environment. The features of cloud enabling green computing are analyzed and the benefits of green cloud computing are discussed.

Keywords: Cloud Computing, Green Computing, Green Cloud Computing,

1. INTRODUCTION:

There is a tremendous increase in the number of transactions in recent years and With the emergence of new technologies like Cloud computing where we use a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computers, a highly scalable and cost-effective architecture for running enterprise and Web applications is required. This ever increasing demand is handled through high speed data centers. The huge data centers (DC) and huge cluster is increasing day by day so energy consumption is also increased.This High energy consumption not only affects the high operational cost but also result into high carbon emissions. Optimal energy solutions are required to curb the impact of Cloud computing on the environment. So this large amount of CO2 dissipation in environment has generated the necessity of Green computing. More processor chips generates more heat, more heat requires more cooling and cooling again

generates heats and thus we come to a stage where we want to balance the system by getting the same computing speed at decreased energy consumption. Cloud computing with green algorithm can enable more energy-efficient use of computing power. So it's required to reduce energy consumption and that we can achieve by reducing the rate of CO2. For that green computing is used nowadays which is the study and application of designing, manufacturing, utilizing, and disposing of ICT — efficiently and effectively with minimal or no impact on the environment.

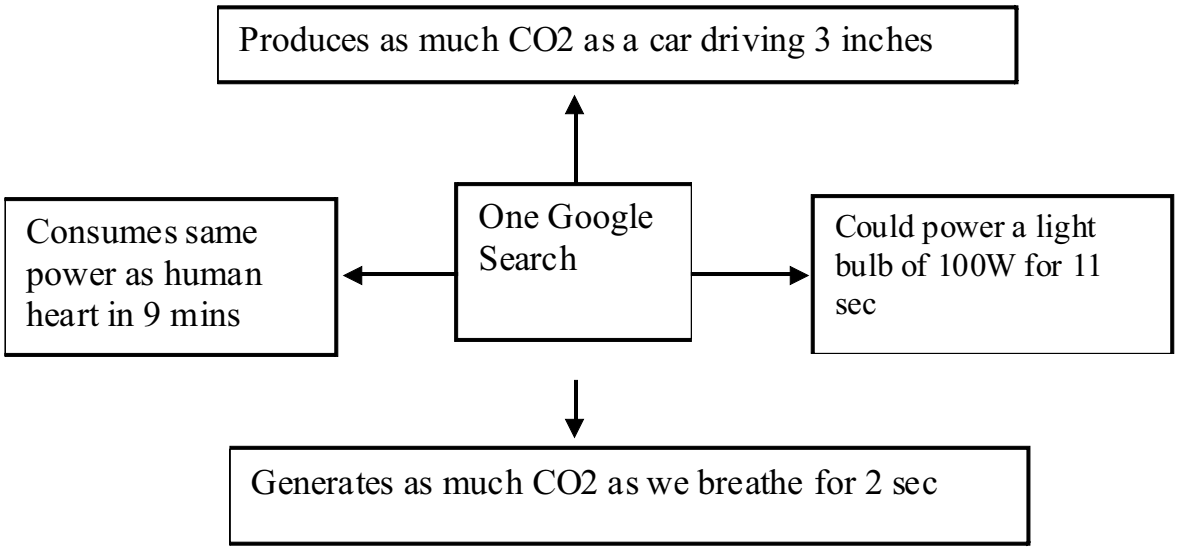


Fig. 1 Energy utilized in one Google search

The above diagram describes the energy utilized in a single google search, which clearly indicates the amount of energy consumed in the huge number of transactions.

2. GREEN COMPUTING:

Green computing is the environment friendly approach to use computers and their resources. In broader terms, it is defined as the study of designing, manufacturing/engineering, using and disposing of computing devices in a way that reduces the harm caused to the environment. It is an eco-friendly approach.

Today global warming is a major issue and Green computing represents a responsible way to address this issue. By adopting green computing practices, business leaders can protect the environment. This reduces energy and paper costs.

The following four things should be done to protect the environment

2.1 Green Use: It means reducing the energy consumption of computers and other information systems and using them in an environmentally sound manner. This may include:

Ø Hibernate or sleep mode should be used when away from a computer for long duration.

Ø Try to use flat-screen or LCD monitors, instead of conventional cathode ray tube (CRT) monitors

Ø Energy efficient notebook computers should be preferred over desktop computers

Ø Activate the power management features for controlling energy consumption

Ø Switch off computers when not in use.

2.2 Green Disposal: It includes reusing old computers and recycling unwanted computers and other electronic equipment.

Ø Proper arrangements should be made for safe electronic waste disposal

Ø Try to Refill printer cartridges, instead of buying new ones

Ø Try reuse an old computer

2.3 Green Design: Green design is thedesigning of energy efficient and environmentfriendly components, which includes computers, servers, printers, projectors and other cooling equipments.

2.4 Green Manufacturing: Manufacturing electronic components, computers and other associated sub systems with minimal impact or no impact on the environment. A green computer or green IT system involves the entire process from design, manufacture, use, and disposal which involves very little environmental impact. One of the first examples of the green computing technology was the launch of the Energy Star program way back in 1992. Energy Star is awarded to the computing products that succeed in minimizing use of energy while maximizing efficiency. Energy Star is applied to different products including computer monitors, television sets and temperature control devices like refrigerators, air conditioners, and similar items. One of the first results of green computing is the sleep or hibernate mode that places computers to power down when not in use and, therefore, save on energy impact.

Green Computing is the need of the hour coz of the Global warming which has been a major threat for the future. The following things can be done in order to Go Green.

3. WHY GREEN CLOUD COMPUTING?

Green Cloud Computing has been the answer to attenuate the danger of cloud computing. Owing to a

large number of data centers that operate under the Cloud computing model where a variety of applications ranging from those that run for a few seconds (e.g. serving requests of web applications such as e-commerce and social networks portals with transient workloads) to those that run for longer periods of time (e.g. simulations or large data set processing) on shared hardware platforms, the energy usage is very high. Multiple applications run in a data center at a time which creates the challenge of providing on-demand resource and allocation in response to time-varying workloads. Normally, data center resources are statically allocated to applications, based on peak load characteristics, in order to maintain isolation and provide performance guarantees. Until recently, data centers have been high performance and they fulfill all requests without paying much attention to energy consumption,there by not being an environment friendly option.Heavy amounts of electricity is needed to power and cool numerous servers hosted in these data centers which results in high energy costs and huge carbon footprints.Certain measures are required to be adopted by Cloud service providers to ensure that their profit margin is not drastically reduced due to high energy costs. Lowering the energy usage of data centers is a challenging and complex issue because computing applications and data are growing so quickly that increasingly larger servers and disks are needed to process them fast enough within the required time period. Green Cloud computing is visualized to achieve efficient processing and utilization of computing infrastructure with minimal energy consumption. This is essential for ensuring that the future growth of Cloud computing is justifiable. Otherwise, Cloud computing with increasing front-end client devices interacting with back-end data centers will cause an enormous rise in energy usage. To address this problem, data center resources need to be managed in an energy-efficient manner to drive Green Cloud computing.

Green cloud computing is a trend which has become popular with the emergence of internet driven services in every field of life. It refers to the prospective environmental advantages that computer based internet services can guarantee to the environment, by processing huge amount of data and information from collective resources pool. The cloud computing is emerged as an effective substitute to a traditional office based processing of services.

The following are the benefits of Green Cloud Computing:

- Reduced environmental impact (less GHG emissions, less e-waste, fewer virgin resources needed for manufacturing new devices)

- Lower energy costs.

- Green Benefits of Cloud Computing

- Longer lasting computing devices.

- Reduced health risk for computer workers and recyclers.
- Automatic Updates
- Remote Access
- Disaster Relief
- Self service Positioning
- Scalability
- Reliability
- Ease of Use
- Skills and proficiency
- Response time
- Increased storage
- Mobility

3.1 CLOUD FEATURES FOR GREEN COMPUTING:

There are certain features of cloud that can enable Green Cloud Computing.Multi-tenancy is used in Cloud computing where single instance of software runs on a client and serves multiple clients. This approach helps in reducing IT infrastructure and provides an energy efficient way that reduces carbon emission. Virtualization is another key driver technology which is the process which is the process of presenting a logical group or subset of computing resources so that they can be accessed in ways that are more energy efficient. Some underutilized servers in the form of multiple virtual machines can be grouped so that companies can save space, energy and management.

3.1.1 Multi-tenancy: With this approach, Cloud computing infrastructure can reduce overall energy usage and the associated carbon emissions with them. The SaaS providers can serve multiple companies on same infrastructure and software. Without this approach multiple copies of the software needs to be installed on different infrastructure thereby utilizing more energy. Moreover, the demand pattern of businesses is very variable, hence multi-tenancy on the same server can reduce the overall peak demand thereby minimizing the need for extra infrastructure. The smaller fluctuation in demand can result in better prediction which helps in greater energy savings.

3.1.2 Server Utilization: Virtualization can be used to host and execute multiple applications on the same server in isolation. This can lead to an increase in server utilization to a very high level. Thus, the number of active servers can be reduced drastically leading to low power usage.

3.1.3 Dynamic Provisioning: In this method, the datacenters maintain the active servers according to the current demand so that less energy is consumed. Dynamic Provisioning is used cause it is not easy to predict demand at a time specially if we take web applications where number of requests may vary depending on different conditions. So if we consider the conservative approach there will be some unutilized resources. Therefore, allocating resources as per request can save energy.

3.1.4 Datacenter Efficiency: The power efficiency of datacenters has major impact on the total energy usage of Cloud computing. Cloud providers can increase their PUE by using highly energy efficient technologies. These technologies may include advanced power management through power supply optimization, water or air based cooling, server design in the form of modular containers. Cloud service providers can achieve around 40% more power efficiency than the traditional datacenters.

4. GREEN CLOUD ARCHITECTURE:

The most important element of this architecture is the Green Broker. Its goal is to make Cloud green from both user and providers perspective. In this Green Cloud architecture, the Cloud Service request is submitted by the users through a new middleware Green Broker that manages the selection of the greenest Cloud provider to serve the users request. A user service request can be of three types i.e., software, platform or infrastructure. The services can be registered by the Cloud providers in the form of green offers to a public directory that the Green Broker access. The green offers include green services, pricing and time when it should be accessed for minimum carbon emission. The Carbon Emission Directory stores the current status of energy parameters for using the various cloud services that is obtained by the Green Broker. The Carbon Emission Directory records all the data related to energy efficiency of Cloud service. This data may include PUE(Power Usage Effectiveness) and cooling efficiency of Cloud datacenter which is providing the service, carbon emission rate of electricity and the network cost. The Green Broker is responsible for calculating the carbon emission of all the Cloud providers who are offering the requested Cloud service. Then, it selects the set of services that will result in least carbon emission and buy these services on behalf of users. The Green Cloud architecture keeps track of overall energy usage of serving a user request. For this it is dependent on two main components, Carbon Emission Directory and Green Cloud offers, which keep track of energy efficiency of each Cloud provider and also give various incentive to Cloud providers to make their service “Green”. In user's perspective the Green Broker plays a crucial role in monitoring and selecting the Cloud services based on the user quality of service requirements, and ensuring minimum carbon emission for serving a user. In general, three types of services(SaaS,PaaS,IaaS) are provided by cloud and therefore process of serving them should also be energy efficient. In other words, from the Cloud provider end, each Cloud layer needs to be “Green” conscious.

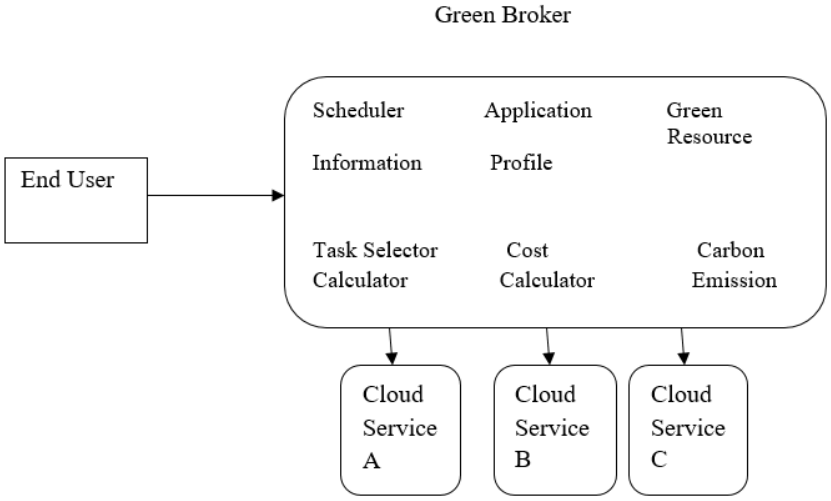


Fig. 2 Green Cloud Architecture

4.1 SaaS Level: SaaS providers mainly provide software installed on their own datacenters or resources from IaaS providers, so they need to work on their model and measure the energy efficiency of their software design, implementation, and deployments .They choose datacenters which are energy efficient and near to users.

4.2 PaaS level: PaaS providers provide platforms for application development. The platform facilitates the development of applications which ensures system wide energy efficiency. This can be done by the use of various energy profiling tools such as JouleSort .While performing External sort,this software energy efficiency benchmark measures the energy required to perform this sort. Further, the platforms can be designed to have various code level optimizations which can cooperate with underlying complier in energy efficient execution of applications. Other than application development, Cloud platforms also allow the deployment of user applications on Hybrid Cloud. In this case, to achieve maximum energy efficiency, the platforms profile the application and decide which portion of application or data should be processed in house and in Cloud.

4.3 IaaS level: The most crucial role is played by the Providers in this layer for the success of Green Architecture since IaaS level offer independent infrastructure services as well as support other services offered by Clouds. The energy consumption can be further reduced by switching-off unutilized server. Many sensors and energy meters installed which calculates the current energy efficiency of each IaaS providers and their sites. There are various green scheduling andresource provisioning policies which ensure minimum energy usage. In addition, the Cloud provider designs various green offers and pricing schemes to provide incentive to users to use their services during off-peak or maximum energy-efficiency hours.

5. SUGGESTIONS:

The following things can be done to ensure greener clouds, thereby having less harmful effect on the environment.

- Predict exactly where powers are consumed and realize the trends of where power consumption is changing. Then calculation of the financial cost for each material power consuming device should be performed. IT equipment generally should not consume most of the data center energy.

- Consolidate and virtualized servers for greater efficiency.

- Decommission, consolidate or just turn off mystery servers.

- Try to replace servers that are more than three years old with newer energy-efficient models.

- Review the CPU performance on regular basis

- Minimize storage equipment by using SANs

- Storage performance should be reviewed regularly.

- Avoid Cabinet glass doors.

- There should be proper cable management inside the cabinet and under the raised flooring to improve airflow and more efficient cooling

- Replace multiple smaller UPSs with fewer central UPSs.

6. CONCLUSIONS AND FUTURE DIRECTIONS:

Cloud computing offers customers with quick response time and high reliability support.They have the ability to handle traffic fluctuations and demand.However, they use a lot of energy. This high energy usage and carbon emissions can be reduced with the help of Green Clouds.Green Cloud Computing can offer environment friendly approach for computing.A number of ways by which we can have green clouds are discussed. However, there are still some research activities that can be done. Research can be done to further increase the performance of GreenCloud and their contribution in business so that business goals can be achieved more efficiently and effectively with minimal environmental harm. Measure should be found to ensure that Green Clouds are able to meet the requirements of real business

services including web services and Online Transaction Processing(OLTP) and OLAP(Online Analytical Processing).

REFERENCES:

1) *AlIsmail, S. M., & Kurdi, H. A. (2016). Review of energy reduction techniques for green cloud computing. Int. J. Adv. Comput. Sci. Appl, 1, 189-195.*
2) *Atrey, A., Jain, N., & Iyengar, N. (2013). A study on green cloud computing. International Journal of Grid and Distributed Computing, 6(6), 93-10.*
3) *Beloglazov, A., Buyya, R., Lee, Y. C., & Zomaya, A. (2011). A taxonomy and survey of energy-efficient data centers and cloud computing systems. In Advances in computers (Vol. 82, pp. 47-111). Elsevier.*
4) *Garg, S. K., & Buyya, R. (2012). Green cloud computing and environmental sustainability. Harnessing Green IT: Principles and Practices, 315-340.*
5) *Gondalia, A., & Vyas, H. Green Cloud Computing: An Overview.*
6) *Kalange Pooja, R. (2013). Applications of green cloud computing in energy efficiency and environmental sustainability. IOSR Journal of Computer Engineering (IOSR-JCE), 25-33.*
7) *Khajeh-Hosseini, A., Sommerville, I., & Sriram, I. (2010). Research challenges for enterprise cloud computing. arXiv preprint arXiv:1001.3257.*
8) *Kumar, A., & Bisht, L. Cloud Computing: All you need to know about.*
9) *Liang, D. H., Liang, D. S., & Chang, C. P. (2012, January). Cloud computing and green management. In Intelligent System Design and Engineering Application (ISDEA), 2012 Second International Conference on (pp. 639-642). IEEE.*
10) *Liu, L., Wang, H., Liu, X., Jin, X., He, W. B., Wang, Q. B., & Chen, Y. (2009, June). GreenCloud: a new architecture for green data center. In Proceedings of the 6th international conference industry session on Autonomic computing and communications industry session (pp. 29-38). ACM.*
11) *Pandya, S. S. Green Cloud Computing. International Journal of Information and Computation Technology, ISSN, 0974-2239.*
12) *Radu, L. D. (2017). Green Cloud Computing: A Literature Survey. Symmetry, 9(12), 295.*
13) *Srimathi, V., Hemalatha, D., & Balachander, R. (2012). Green Cloud Environmental Infrastructure. International Journal of Engineering and Computer Science, 1(3), 168-177.*
14) *Sriram, I., & Khajeh-Hosseini, A. (2010). Research agenda in cloud technologies. arXiv preprint arXiv:1001.3259.*

ABOUT THE AUTHOR

Ms Shuchi Srivastava is an Assistant Professor at Northern India Engineering College, New Delhi. She started her academia journey as a faculty member in 2007 at Babu Banarsi Das National Institute of Technology and Management, Lucknow. She is a graduate in Computer Application from Dr. Bhimrao Ambedkar University, Agra and completed her MCA degree from UP Technical University, Lucknow. She has also completed M-Tech (Computer Science) from UPTU, Lucknow. She has presented many research papers in national and international conferences.

The author can be reached on myself_shuchi008@rediffmail.com for comments or suggestions.

Analysis of Big data Technologies and WI [Web Intelligence]

[1]Saeed Anwar Jamal Ansari & [2]Tabir Ahmad

1. M.Tech Scholar,MDU saeedlko@gmail.com
2. M.Tech Scholar,MDU taabeerahmad.rizvi@gmail.co

ABSTRACT

Big Data has captured a lot of interest in industry, with anticipation of better decisions, efficient organizations and many new jobs. Much of the emphasis is on the challenges of the four V's of Big Data: Volume, Variety, Velocity, and Veracity, and technologies that handle volume, Including storage and computational techniques to support analysis (Hadoop, NoSQL, MapReduce, etc).We analyzed and incorporate a lot of data in social networks and suggest an appropriate data model. We further provided a recommendation system that can be used together with the data model for high reliability.

The Web Intelligence is building 'web-intelligence' applications exploiting big data sources arising social media, mobile devices and sensors, using new big-data platforms based on the 'map-reduce' parallel programming paradigm.

The interest in WI is growing very fast. We would like to invite everyone, who are interested in the WI related research and development activities, to join the WI community. Your input and participation will determine the future of WI.

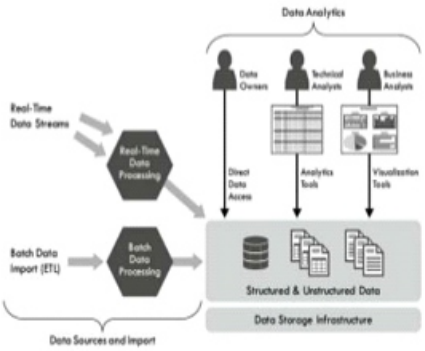
Keys: Big Data, WI [web Intelligence], Big Data Technologies, Hadoop.

Introduction:

Big Data is defined as a large amount of data which require new technologies and architecture to make possible to extract value from it by capturing and analysis process. Big data is emerged because we are living in a society which makes increasing use of data intensive technologies.

Such large size of data it becomes very difficult to performed effective analysis using the existing traditional techniques. Since Big Data is a recent upcoming technology in the market which can bring huge benefits to the business organization.

The difficulties can be related to the data capture, storage, search, analytics, sharing and visualization.



Big data due to its various properties like, velocity, volume, variety, variability, value and complexity.

Big Data:

Big data is the next generation of data warehousing and business analytics. It has many deep roots and many branches. In fact you speak with most data industry veterans; Big data has been around for decades for firms that have been handling tons of transactional data over the years –even dating back to the mainframe era.

The McKinsey study defines Big Data, refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. This definition is internationally subjective…We assume that's qualify as bid data ogy advances over time, the size of datasets that qualify as bid data will also increase. Big data in many sectors today will range from a few dozen terabytes to multiple petabytes (the thousands of terabytes).

Big data is not just a description of raw volume." The real issue is usability," From his perspective, big datasets are not even the problem. The real challenge is identifying or developing most cost-effective and reliable methods for extracting value from all the terabytes and petabytes of data now available. [7]

Big Data Characteristics:

Data Volume:

The Big Data word in Big data itself define the volume the data existing is in petabyte (10^{15}) and is supposed to increase to zetabyte (10^{21}) in nearby future.

Data volume measures the amount of data available to an organization, which does not necessarily have to own all of it as long as it can access it.

Data Velocity:

This deals with the speed of data coming from the various sources. This characteristic is not being limited to the speed of incoming data but also speed at which the data flow and aggregated.

Data Variety:

Data variety is the measure of the richness of the data representation, Text, image Audio, Vedio, etc. data being produced is not of single category as it not only include the traditional data but also the semi

structure data from various resources like web pages, Web Log Files, social media sites, e-mails, documents.

Data Value:

Data value measures to the usefulness of the data in making decisions. Data science is exploratory and useful in getting to know the data, but analytic science encompasses the predictive power of big data.

Complexity:

Complexity measures the degree of interconnectedness and independence in big data structure such that a small change in one or few elements can yields very large changes or small change.

Big Data Technologies:

A 2011 McKinsey paper suggests suitable technologies include A/B testing, association rule learning, classification, cluster analysis, crowd sourcing, data fusion and integration, ensemble learning, genetic algorithms, machine learning, natural language processing, neural networks, pattern recognition, predictive modeling, regression, sentiment analysis, signal processing,upervised and unsupervised learning, simulation, time series analysis and visualization.

Additional technologies being applied to big data include massively parallel-processing (MPP) databases, search-based applications, data-mining grids, distributed file systems, distributed databases, cloud computing platforms, the Internet, and scalable storage systems [3].

Even though there are many suitable technologies to solve Big Data challenges as indicated in the list above given by the McKinsey paper, this work will explore a handful of the more popular ones, in particular the Amazon Cloud, Hadoop's MapReduce, and three open-source parsers.[3]

1- The Elephant in the Room: Hadoop's Parallel World

Hadoop. We are in a world today where is there is exabytes of data being generated every data. Consider the following statistics.

Every minute:

- Facebook users share nearly 2.5 million pieces of content.

- Twitter users tweet nearly 400,000 times.

- Instagram users post nearly 220,000 new photos.
- YouTube users upload 72 hours of new video content.
- Apple users download nearly 50,000 apps.
- Email users send over 200 million messages.
- Amazon generates over $80,000 in online sales.

Isn't just too vast. And thus to handle this amount of data there must be some technologies in place. In order to cope, Google invented a new style of data processing known as MapReduce. A year after Google published a white paper describing the MapReduce framework, Doug Cutting and Mike Cafarella, inspired by the white paper, created Hadoop to apply these concepts to an open-source software framework to support distribution for the Nutch search engine project. Apache Hadoop is one technology that has been the darling of Big Data talk. Hadoop is an open-source platform for storage and processing of diverse data types that enables data-driven enterprises to derive the complete value from all their data.

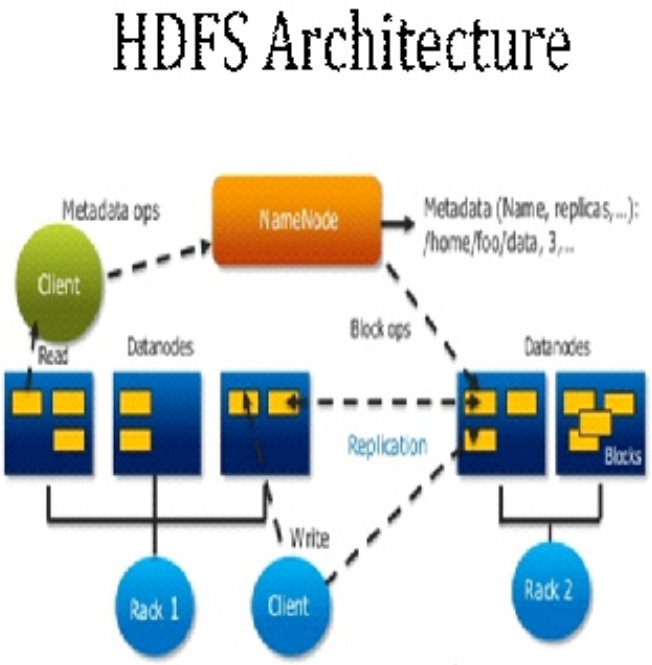
To understand Hadoop, we must understand two fundamental things about it. They are: How Hadoop stores files, and how it processes data. Imagine we have a file that was larger than our PC's capacity. We could not store that file, right? Hadoop

lets us store files bigger than what can be stored on one particular node or server. So that we can store very, very large files. It also lets us store many, many files. [1]

The two critical components of Hadoop are:

1. The Hadoop Distributed File System (HDFS).

HDFS is the storage system for a Hadoop cluster.When data lands in the cluster HDFS breaks it into pieces and distributes those pieces among the different servers participating in the cluster. HDFS breaks it into pieces and distributes those pieces among servers participating in cluster.



2. Map Reduce.

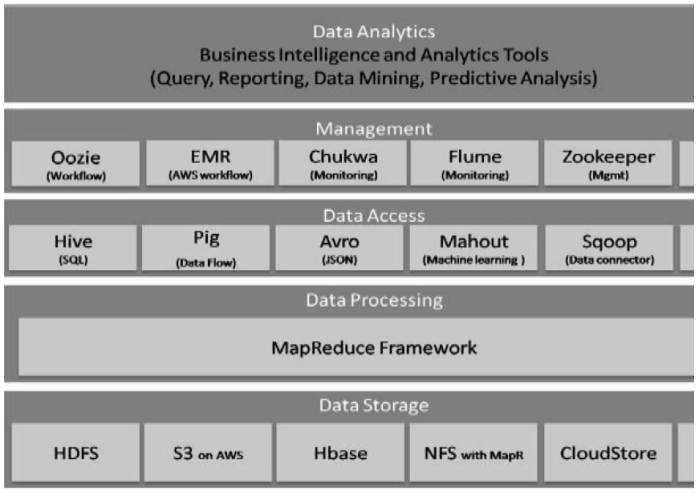
Because Hadoop stores the entire data set in small pieces across a collection of servers, analytical jobs can be distributed, in parallel, to each of the servers storing part of the data. Each server evaluates the question against its local fragment simultaneously and reports its results back for collation into comprehensive answer. Map Reduce is the agent that distributes the work and collects the results. Map and Reduce are two functions with shuffle in between which is handled by the system.

Both HDFS and Map Reduce are designed to continue to work in the face of system failure. HDFS continually monitors the data stored on the cluster. If a server becomes unavailable, a disk drive fails or data is damaged whether due to hardware or software problems, HDFS automatically restores the data from one of the known good replicas stored elsewhere on the cluster. Likewise, when an analysis job is running, Map Reduce monitors progress of each of the servers participating in the job.If one of them fails before completing its work, Map Reduce automatically starts another instance of the task on another server that has copy of the data. Thus Hadoop provides scalable, reliable and fault-tolerant services for data storage and analysis at very low cost. [3]

BIG DATAANALYSIS:

Big Data analysis tools which are used forefficient and precisedata handling the velocity and heterogeneity of data, tools like Hive, Pig and Mahout are used which are parts of Hadoop and HDFS framework. It is interesting to note that for all the tools used, Hadoop over HDFS is the underlying

architecture. Oozie and EMR with Flume and Zookeeper are used for handling the volume and veracity of data, which are standard Big Data management tools. The layer with their specified tools forms the bedrock for Big Data management and analysis framework.



Big Data Analysis Tools

WEB INTELLIGENCE (WI):

The study of Web intelligence (WI) was first introduced in several papers and books [see Refs. (1–19)]. Broadly speaking, WI is a new direction for scientific research and development that explores the fundamental roles as well as practical impacts of artificial intelligence (AI),1 such as knowledge representation, planning, knowledge discovery and data mining, intelligent agents, and social network intelligence, as well as advanced information technology (IT), such as wireless networks; ubiquitous devices; social networks; and data/knowledge grids; and the next generation of Web-empowered products, systems, services, and activities.

TheWI technologies revolutionize the way in which information is gathered, stored, processed, presented, shared, and used through electoronization, virtualization, globalization, standardization, personalization, and portals.

The new WI technologies will be determined precisely by human needs in a post-industrial era; namely (2):

- information empowerment,
- knowledge sharing,
- virtual social communities,

- service enrichment, and
- practical wisdom development. [1]

The World Wide Wisdom Web (the Wisdom Web or W4) will become a tangible goal for WI research (1, 5, 6). The new generation of the WWW will enable humans to gain wisdom of living, working, and playing in addition to information search and knowledge queries.

WEB INTELLIGENCE (WI):

WI is to enable users to gain new wisdom of living, working, playing, and learning, in addition to information search and knowledge queries. Here, the word of wisdom, according to the Webster Dictionary (Page: 1658) (17), implies the

following meanings (emphasis added):

1. The quality of being wise; knowledge, and the capacity to make due use of it; knowledge of the best ends and the best means; discernment and judgment; discretion; sagacity; skill; dexterity.
2. The results of wise judgments; scientific or practical truth; acquired knowledge; erudition. [1]

Levels of WI:

WI techniques and technologies, which cover the following four conceptual levels at least:

1. Internet-level communication, infrastructure, and security protocols. The Web is regarded as a computer-network system. WI techniques for this level include Web data perfecting systems built upon Web surfing patterns to resolve the issue of Web latency. The intelligence of the Web prefacing comes from an adaptive learning process based on the observation and characterization of user surfing behavior (1).

2. Interface-level multimedia presentation standards. The Web is regarded as an interface for human–Internet interaction. WI techniques for this level are used to develop intelligent Web interfaces in which the capabilities of adaptive cross-language processing, personalized multimedia representation and multimodal data processing are required. [1]

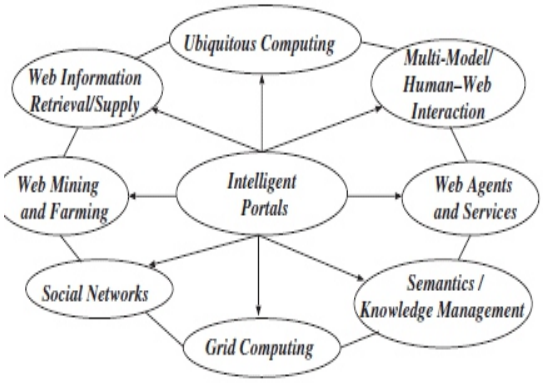
3. Knowledge-level information processing and management tools. The Webis regarded as a distributed data/knowledge base. We need to develop semantic markup languages to represent the

semantic contents of the Web available in machine-understandable formats for agent-based autonomic computing, such as searching, aggregation, classification, filtering, managing, mining, and discovery on the Web.

4. Application-level ubiquitous computing and social intelligence environments. The Web is regarded as a basis for establishing social networks that contain communities of people (or organizations or other social entities) connected by social relationships, such as friendship, co working, or information exchange with common interests. They are Web-supported social networks or virtual communities.

An Intelligent Enterprise Portal Centric Schematic Diagram of WI Technologies:

AI and IT to a totally new domain. On the other hand, the WI technologies are also expected to introduce new problems and challenges to the established disciplines on the new platform of the Web and the Internet. That is, WI is an enhancement or an extension of AI and IT.



An intelligent enterprise portals entric schematic diagram of WI technologies.

To study advanced WI technologies systematically, and to develop advanced Web-based intelligent enterprise portals and information systems, we provide schematic diagram of WI technologies from a Web-based, intelligent enterprise portals centric perspective.

ADVANCED TOPICS FOR STUDYING WI:

WI as mentioned in the section entitled "Levels of WI," the Web can be studied in several ways.

Studying the Semantics in the Web:

One of the fundamental WI issues is to study the semantics in the Web, called the semantic Web that is, modeling semantics of Web information to.

§ Allow more of the Web content (not just form) to become machine readable and processible.

§ Allow for recognition of the semantic context in which Web materials are used.

§ Allow for the reconciliation of terminological differences between diverse user communities.

Main Components of the Semantic Web. The semanticWebis a step toward intelligence of the Web. It is based on languages that make more semantic content of the page available in machine-readable formats for agent-based computing. The main components of semantic Web techniques include:

§ a unifying data model such as RDF (Resource Description Framework).

§ languages with defined semantics, built on RDF, such as OWL.

§ ontologies of standardized terminology to mark up Webresources, used by semantically rich, service-level descriptions (such as OWL-S, the OWL-based Web Service Ontology), and to support tools that assist the generation and processing of semantic markup.

Ontologies and agent technology can play a crucial role in Web intelligence by enabling Web-based knowledge processing, sharing, and reuse between applications.

CONCLUSION:

Big Data analysis tools like Map Reduce over Hadoop and HDFS, promises to help organizations better understand their customers and the marketplace, hopefully leading to better business decisions and competitive advantages. For engineers building information processing tools and applications, large and heterogeneous datasets which are generating continuous flow of data, lead to more effective algorithms for a wide range of tasks, from machine translation to spam detection. The big data model has the flexibility to be expanded to incorporate more sophisticated additional factors if needed. The experimental results using it in information recommendation and using map-reduce to process it show that it is a feasible model to be used for information recommendation.

AWI-focused scientific journal, Web Intelligence and Agent Systems: An International Journal (refer to the WIC homepage), has been providing a standard international forum for disseminating results of advanced research and development in the field of WI.

The interest in WI is growing very fast. We would like to invite everyone, who are interested in the WI related research and development activities, to join theWI community.

REFERENCES:

1- *J. Liu, N. Zhong, Y. Y. Yao, Z. W. Ras, The wisdom web: new challenges for web intelligence (WI), J. Intell. Inform. Sys.,(1): 5–9, 2003.*

2- *J. Liu, Web intelligence (WI): what makes wisdom web? Proc. Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03), 2003, pp. 1596–1601.*

3- *Ted Garcia and Taehyung ("George") Wang,"Analysis of Big Data Technologies and Methods" Department of Computer Science, 2013 IEEE Seventh International Conference on Semantic Computing California State University.*

4- *Shankar Ganes h Manikandan, Siddart h Ravi, "Big Data Analysis using Apache Hadoop".2014, IEEE.*

5- *J. Liu, New challenges in the world wide wisdom web (W4) research, in N. Zhong, et al. (eds.), Foundations of Intelligent Systems, LNAI 2871, Springer, 2003, pp. 1–6.*

6- *N. Zhong, J. Liu, and Y. Y. Yao, In search of the wisdom web,IEEE Computer, 35(11): 27–31, 2002.*

7- *Big-Data: Big analytics: Micael Minelli, Michele Chambers, Ambiga Dhiraj.*

Analysis of Characteristics of High Quality Web Site using Web Content Mining

[1]**Tabir Ahmad & [2]Saeed Anwar Jamal Ansari**

1. M.Tech Scholar,MDU taabeerahmad.rizvi@gmail.com

2. M.Tech Scholar,MDU saeedlko@gmail.com

ABSTRACT

The Quest for knowledge has led to new discoveries and inventions. With the emergence of World Wide Web, it became a hub for all these discoveries and inventions. Web browsers became a tool to make the information available at our finger tips. As years passed World Wide Web became overloaded with information and it became hard to retrieve data according to the need. If you have an existing site, or plan to develop one in the near future, it's important to understand the characteristics that can make or break the effectiveness of your online investment. An unattractive or poorly built site will do more to hurt your business than to help it. In this article, we look at the five general components involved in making a website successful. Over the years the whole SEO industry is talking about the need of producing high quality content and top experts came up with the clever quote 'Content is king' meaning that content is the success factor of any web site. While this is true, does it mean that a web site with good content is also a high quality web site? The answer is NO. Good content is not enough. It is one of the factors (the most important) that separates low from high quality sites but good content alone does not complete the puzzle of what is considered by Google as a high quality web site. Web mining came as a rescue for the above problem. Web content mining is a subdivision under web mining. This paper deals with a study of different techniques and pattern of content mining and the areas which has been influenced by content mining. The web contains structured, unstructured, semi structured and multimedia data. This survey focuses on how to apply content mining on the above data. It also points out how web content mining can be utilized in web usage mining.

Keyword: Accessibility, Optimization, Social, Speed, Design, Content, Security, Appearance.

Introduction:

The Web as been the fastest adopted technology, but often the quality of web sites is unsatisfactory, and basic web principles, like interoperability and accessibility, are ignored or scarcely considered by designers. There are several reasons for the scarce quality, in spite of the attention paid to the quality in other sectors like Software Engineering. Many of existing criteria are not easy to measure and require methods such as heuristic evaluations, and empirical usability tests. This paper aims at defining a quality model and a set of characteristics that can be measured in an automated fashion, relating internal and external quality factors and giving clues about potential problems. Search engines look for two criteria when evaluating your website links and determining your website ranking: quantity and quality. That means both the number and the relevance of your links are crucial for getting your website found faster by search engines, and by potential clients.

The advancement in technology paved the way for faster communication. The previous decade experienced a dramatic development in computer technology, such that with the press of a finger the i

nformation about a particular topic appeared in monitors within seconds. As time passed by the complexity of web increased due to enormously large amount of data. So extraction of data according to users need became a tedious task. As a result mining became an essential technique to extract valuable information from internet. And this technique was named as web mining. Web mining is further classified into three: They are Web content mining, Web Structure mining, Web Usage mining. Using the objects like text, pictures, multimedia etc. content mining is done in the web. In Web structure mining, mining is done based on the structure like hyperlinks. In the case of web usage mining, mining is done on web logs which contain the navigational pattern of users. And the study of this navigational pattern will trace out the interest of the users.

1.) Web Content Mining

Traditional technique of searching the web was via contents. Web Content mining is the extended work performed by search engines. Web Content mining refers to the discovery of useful information from web content such as text; images videos etc.Two approaches used in web content mining are Agent based approach and database approach.

The three types of agents are intelligent search agents, Information filtering / categorizing agent, and personalized web agents. Intelligent Search agents automatically searches for information according to a particular query using domain characteristics and user profiles. Information agents used number of techniques to filter data according to the predefine instructions. Personalized web agents learn user preferences and discovers documents related to those user profiles. In Database approach it consists of well formed database containing schemas and attributes with defined domains. Web content mining becomes complicated when it has to mine unstructured, structured, semi structured and multimedia data.

1.1 Unstructured Data Mining Techniques

Content mining can be done on unstructured data such as text. Mining of unstructured data give unknown information. Text mining is extraction of previously unknown information by extracting information from different text sources. Content mining requires application of data mining and text mining techniques. Basic Content Mining is a type of text mining. Some of the techniques used in text mining are Information Extraction, Topic Tracking, Summarization, Categorization, Clustering and Information Visualization.

1.1.1 Information Extraction

To extract information from unstructured data, pattern matching is used. It traces out the keyword and

phrases and then finds out the connection of the keywords within the text. This technique is very useful when there is large volume of text. IE is the basis of many other techniques used for unstructured mining? Information extraction can be provided to KDD module because information extraction has to transform unstructured text to more structured data. First the information is mined from the extracted data and then using different types of rules, the missed out information are found out. IE that makes incorrect predictions on data is discarded.

1.1.2 Topic Tracking

Topic Tracking is a technique in which it checks the documents viewed by the user and studies the user profiles. According to each user it predicts the other documents related to users interest. In Topic Tracking applied by yahoo, user can give a keyword and if anything related to the keyword pops Up then it will be informed to the user. Same can be applied in the case of mining unstructured data. An example for topic tracking is that if we select the competitors name then if at any time their name will come up in the news then this information will be passed to the company. Topic tracking can be applied in many fields. Two such areas are medical field and education field. In medical field doctors can easily come to know latest treatments. In education field topic tracking can be used to find out the latest reference for research related work. Topic tracking helps to track all subsequent stories in the news stream. Disadvantage of topic tracking is that when we search for topics we may be provided with information which is not related to our interest. For example if user sets an alert for 'web mining' it can provide us with topics related to mineral mining etc. which are not useful for user.

1.1.3 Summarization

Summarization is used to reduce the length of the document by maintaining the main points. It helps the user to decide whether they should read this topic or not. The time taken by the technique to summarize the document is less than the time taken by the user to read the first paragraph. The challenge in summarization is to teach software to analyze semantics and to interpret the meaning. This software statistically weighs the sentence and then extracts important sentences from the document. To understand the key points summarization tool search for headings and sub headings to find out the important points of that document. This tool also give the freedom to the user to select how much percentage of the total text they want extracted as summary. It can work along with other tools such as Topic tracking and categorization to summarize the document. An example for text Summarization is Microsoft word's AutoSummarize

1.1.4 Categorization

Categorization is the technique of identifying main themes by placing the documents into a predefined set of group. This technique counts the number of words in a document. It does not process

the actual information. It decides the main topic from the counts. It ranks the document according to the topics. Documents having majority content on a particular topic are ranked first.
Categorization can be used in business and industries to provide customer support.

1.1.5 Clustering

Clustering is a technique used to group similar documents. Here in clustering grouping is not done based on predefined topic. It is done based on fly. Same documents can appear in different group. As a result useful documents will not be omitted from the search results. Clustering helps the user to easily select the topic of interest. Clustering technology is useful in management information system

1.1.6 Information Visualization

Visualization utilizes feature extraction and key term indexing to build a graphical representation. Through visualization, documents having similarity are found out. Large textual materials are represented as visual hierarchy or maps where browsing facility is allowed. It helps the user to visually analyze the contents. User can interact with the graph by zooming, creating sub maps and scaling. This technique is useful to find out related topic from a very large amount of documents.

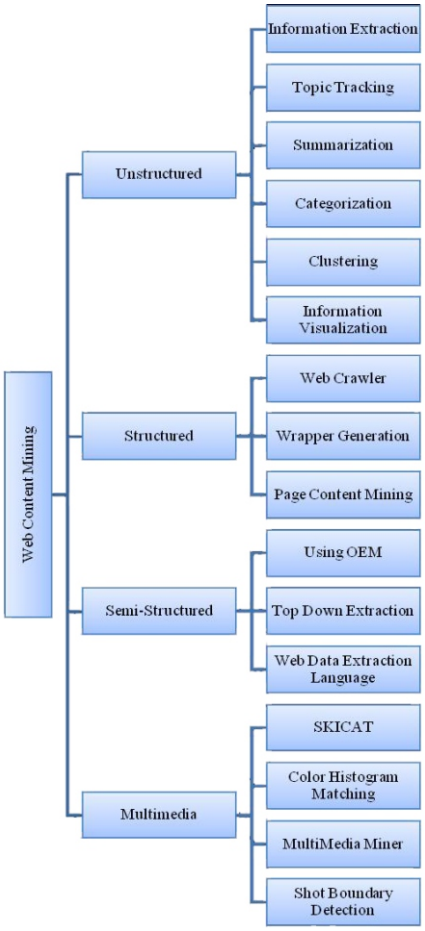


Fig 1: Web Content Mining Techniques

1.2 Structured Data Mining Techniques

The techniques used for mining structured data are Web Crawler, Wrapper Generation, Page content Mining.

1.2.1 Web Crawler

There are two types of Web Crawler which are called as External and Internal Web crawler. Crawlers are computer programs that traverse the hypertext structure in the web. External Crawler crawls through unknown website. Internal crawler crawls through internal pages of the website which are returned by external crawler.

1.2.2 Wrapper Generation

In Wrapper Generation, it provides information on the capability of sources. Web pages are already ranked by traditional search engines. According to the query web pages are retrieved by using the value of page rank. The sources are what query they will answer and the output types. The wrappers will also provide a variety of Meta information. E.g. Domains, statistics, index look up about the sources.

1.2.3 Page Content Mining

Page Content Mining is structured data extraction technique which works on the pages ranked by traditional search engines. By comparing page Content rank it classifies the pages.

1.3 Semi-Structured Data Mining Techniques

The techniques used for semi structured data mining are Object Exchange Model (OEM), Top down Extraction, and Web Data Extraction language.

1.3.1 Object Exchange Model (OEM)

Relevant information are extracted from semi-structured data and are embedded in a group of useful information and stored in Object Exchange model (OEM). It helps the user to understand the information structure on the web more accurately. It is best suited for heterogeneous and dynamic environment. A main feature of object exchange model is self describing; there is no need to describe in advance the structure of an object.

1.3.2 Top down Extraction

In top down extraction, it extracts complex objects from a set of rich web sources and converts into less complex objects until atomic objects have been extracted.

1.3.3 Web Data Extraction Language

In Web data extraction language it converts web data to structured data and delivers to end users. It stores data in the form of tables

1.4 Multimedia Data Mining Techniques

Some of the Multimedia Data Mining Techniques are SKICAT, color Histogram Matching, Multimedia Miner and Shot Boundary Detection.

1.4.1 SKICAT

SKICAT is a successful astronomical data analysis and cataloging system which produces digital catalog of sky object. It uses machine learning technique to convert these objects to human usable classes. It integrates technique for image processing and data classification which helps to classify very large classification set.

1.4.2 Color Histogram Matching

Color Histogram matching consists of Color histogram equalization and Smoothing. Equalization tries to find out correlation between color components. The problem faced by equalization is sparse data problem which is the presence of unwanted artifacts in equalized images. This problem is solved by using smoothening.

1.4.3 Multimedia Miner

Multimedia Miner Comprises of four major steps. Image excavator for extraction of image and Video's, a preprocessor for extraction of image features and they are stored in a database, A search kernel is used for matching queries with image and video available in the database. The discovery module performs image information mining routines to trace out the patterns in images.

1.4.4 Shot Boundary Detection

It is a technique in which automatically the boundaries are detected between shots in video.

1.5 Web Content Mining Tools

Web Content Mining tools are software that helps to download the essential information for users. It collects appropriate and perfectly fitting information. Some of them

are Web Info Extractor, Mozenda, Screen-Scraper, Web Content Extractor, and Automation Anywhere 5.5.

1.5.1 Web Info Extractor

This tool is helpful for mining and extracting content and monitoring content update.

1.5.2 Mozenda

Users can set up agents that regularly extract, store and circulate data to several destination.

1.5.3 Screen-Scraper

It searches a database, SQL server or SQL database, which interfaces with the software to achieve content mining requirements.

1.5.4 Web Content Extractor

It is a powerful and easy tool for web scraping, data mining and data retrieval

1.5.5 Automation Anywhere 5.5

It retrieves web data effortlessly, screen scrape from web pages or use it for web mining.

2.) Characteristics of High Quality Web Site

2.1 Unique content

Content is unique both within the site itself (i.e. each page has unique content and not similar to other pages), but also compared to other web sites. Short and organized copy: Clearly label topics and break your text up into small paragraphs. Don't bore your visitors with visually overwhelming text. You've got less than 10 seconds to hook your visitors, so grab their attention by being clear, concise and compelling. Update your content regularly: No one likes to read the same thing over and over again. Dead or static content will not bring visitors back to your site! Speak to your visitors: Use the word you as much as possible. Minimize the use of *me, us* and *us*. Consider a pro: Unless you're an especially good writer, consider using a professional to write or edit your text content.

2.2 Expertise

Content is produced by experts based on research and or experience. If for example the subject is health related, then the advice should be provided by qualified persons who can professionally give advice for the particular subject.

2.3 Accessibility

A high quality web site has versions for non PC users as well. It is important that mobile and tablet users can access the web site without any usability issues. A site must be visually appealing, polished and professional. A simple way to increase visual appeal is to use high quality photography. High quality product images are especially important for online retailers. Keep it simple and allow for adequate white space. Uncluttered layouts allow viewers to focus on your message. Don't overload your site with overly complex design, animation, or other effects just to impress your viewers.

2.4 Usability

Can the user navigate the web site easily; is the web site user friendly. A critical, but often overlooked component of a successful website is its degree of usability. Your site must be easy to read, navigate, and understand. Simplicity is the best way to keep visitors glued to your site is through valuable content, good organization and attractive design. Keep your site simple and well organized. Fast-loading pages a page should load in 20 seconds or less via dial-up; at more than that, you'll lose more than half of your potential visitors. Minimal scroll is particularly important on the first page. Create links from the main page to read more about a particular topic. Even the Search Engines will reward you for this behavior. Site layout is extremely important for usability. Use a consistent layout and repeat certain elements throughout the site.

2.5 Cross-platform / browser compatibility Different browsers often have different rules for displaying content. At a minimum, you should test your site in the latest versions of Internet Explorer (currently, versions 8 and 9), as well as Firefox and Safari.

2.6 Social

Social media changed our lives, the way we communicate but also the way we assess quality. It is expected for a good product to have good reviews, Facebook likes and Tweets. Before you make a decision to buy or not, you may examine these social factors as well. Likewise, It is also expected for a good web site to be socially accepted and recognized i.e. have Facebook followers, RSS subscribers etc.

2.7 Search Engine Optimized (SEO)

There are hundreds of rules and guidelines for effective search engine optimization, and this isn't the place to cover them all. For starters, follow these simple rules:

- Include plenty of written content in HTML format. Don't use Flash, JavaScript or image-only objects for your navigational items.

- Use your important keywords frequently and appropriately in your copy.

- Minimize the use of tables and use Cascading Style Sheets for layout and positioning; keep your HTML code clutter-free.

- Leverage your links -- make them descriptive and use your keywords in the link text

Many, many books have been written about Search Engine Optimization, and its scope is too broad to cover here.

3). Create a High Quality Web Site

3.1 Check your site for content uniqueness

Ensure that you have unique content in all the pages. (I use copy scape to check all content before publishing to ensure that it is unique). Check your existing pages and if you find duplicate content either remove it or de-index it.

3.2 Implement Google authorship

A must do for every publisher. Each and every page of your web site should 'belong' to a verified author.

3.3 Go mobile

You should provide for both a mobile version and native apps. The experience of the reader is important regardless of the platform.

3. 4 Tidy up your content

Check all articles/pages for spelling and grammar mistakes. Make correct use of formatting tags i.e.

H1, H2 and bold. Use small paragraphs to make reading easier and add images, illustrations or info graphics to make your content more interesting and appealing.

3.5 Find out what readers want.

Use the Google keyword tool to find out what people are searching and give them content they want but avoid over-optimization. You can write content around the keywords they understand but try to make your posts social as well. Avoid using optimized titles all the time but keep a balance between optimized and non-optimized content.

3.6 Minimize the amount of ads above the fold.

Google is penalizing web sites having too many ads above the fold. This is also not good for the user experience. If you have good content, ads below the fold and at the end of your articles can also perform well.

3.7 Make your web site load fast

Check your web site with the **page speed tool** and ensure that it loads as fast as possible. I know that it's not the easiest task to do but it is very important. If you are using WordPress you can **read this guide** (it's on the Thesis web site but all recommendations made are general for WordPress sites and not just web sites running Thesis).

3.8 Work on your social media promotion

Do you have a Facebook page, twitter account, Pinterest page and Google+? If not you are missing many opportunities. Read my previous post SEO boost your social media profiles on how to make the most from your social media profiles.

3.9 Check your bounce rate

One of the measures you need to monitor, as explained in the **Best Google analytics reports for beginners** is the bounce rate. Bounce rate is the percentage of people who left your web site without viewing other pages. You need to aim for a low bounce rate so check your analytics and identify which pages of your site have a high bounce rate and correct them. Although officially Matt Cutts (Head of Google Quality Team), said that the bounce rate is not a ranking factor, it is certainly a factor that can help you create better quality web sites.

3.10 Check your competition

When someone makes a complain to Google about a web site not ranking as high as it should, they sometimes reply with this: Give as examples of how your web site is better than the web sites we have in the top positions. In other words what they are telling us is go and check your competitors and try to make your content and web sites better! There is nothing wrong with that, in fact this is something you should do before creating your content. What I do before writing an article is search Google for the keyword I want to rank my article and taking a closer look on the web sites in the first page. I check things like content length, authority of the web site, authority of the writer and try to make my content better and more usable.

3.11 Keep your web site healthy and spam free

Last but not least, a high quality web site is spam free. One of the mistakes made by big web sites (think Newspapers) is not moderating their comments. As a result you see a ton of spam comments below their articles which is not a good user experience. As a rule of thumb you need to keep your web site up-to-date in terms of software updates and to monitor and delete spam or inappropriate comments.

4). Conclusions

A successful website will be search engine-friendly. Optimization includes things like using page titles and Meta tags for all of your pages. Using headers and alt tags for images can also help optimization and accessibility. There are many more characteristics that distinguish good web sites from the rest but the above guidelines are the bare minimum for producing web sites that can be trusted by Google and considered as high quality. In this paper, standardized characteristics, and about eighty directly measurable attributes for the sites on the academic domain were considered. The main goal was to establish quality requirements to arrange the list of characteristics and attributes that might be part of a quantitative evaluation, comparison, and ranking process. The proposed Web-site QEM methodology, grounded in a logic multi-attribute decision model and procedures ,is intended to be a useful tool to evaluate artifact quality in the operational phase of a WIS lifecycle. In addition, it could be also used in earlier stages as exploratory and development phases.

This paper discusses the techniques of web content mining. Web content mining has been proved very useful in the business world. The survey also discusses the techniques used for extracting information from different types of data available in the internet and how this extracted data can be used for mining purposes. Users feel difficulty in finding desired information and deciding which information is relevant to them from general purpose search engines. Web content mining solves this problem and helps the users to fulfill their needs. Topic Tracking is useful in predicting the web content related to users interest. Summarization helps the user to decide whether they should read a particular topic or

not. Categorization can be used in business and industries to provide customer support. Clustering and information visualization are the techniques frequently being used for mining. Web content mining can also be applied to business application like mining online news site and developing a suggestion system for distance learning. Content Mining helps to establish better relationship with customer by providing exactly what they need. At the end paper discusses about different tools that can be used in web content mining.

5). REFERENCES:

[1] Ahmed, S. S., Halim, Z., Blaig, R. and Bashir, S. 2008.Web Content Mining: A Solution to Consumers Product Hunt. International Journal of Social and Human
Sciences 2, 6-11.

[2] Ajoudanian, S. and Jazi, M. D. 2009. Deep Web Content Mining. World Academy of Science, Engineering and Technology 49.

[3] Bassiou, N. and Kotropoulos, C. 2006. Color Histogram Equalization using Probability Smoothening. Proceedings of XIV European Signal Processing Conference

[4] Bharanipriya, V. and Prasad, K. 2011. Web content Mining Tools: A Comparative study. International Journal of Information Technology and Knowledge Management. Vol. 4. No 1,211- 215.

[5] Cooper, M., Foote, J., Adcock, J. and Casi, S. 2003. Shot Boundary Detection via Similarity Analysis. In Proceedings of TRECVID 2003 workshop.

[6] Dunham, M. H. 2003. Data Mining Introductory and Advanced Topics. Pearson Education.

[7] Etzioni, O. 1996. The World Wide Web: quagmire or gold mine?. Communications of the ACM. Vol. 39. Issue 11. pp. 65-68.

[8]. IEEE Web Publishing Guide http://www.ieee.org/web/developers/style/

[9]. IEEE Std 1061-1992, "IEEE Standard for a Software Quality Metrics methodology"

[10]. ISO/IEC 9126-1991 International Standard,"Information technology – Software product evaluation –Quality characteristics and guidelines for their use"

[11]. Marucci, Luisa and Signore, Oreste: Evaluating Web sites quality - CMG Italia - Conferenza annuale Pisa, 19-21 maggio 2004
http://www.w3c.it/papers/cmg2004-quality/
[12]. Robertson, S.E. and K.S. Jones, Relevance weighting of search terms Journal of Documentation, 1976. 27(3): p. 129-146. 26.Jorm, A., et al., Help for depression: What works (and what doesn't) 2001, Canberra: Centre for Mental Health Research.

[17] MitChell, T. 1997. Machine Learning. McGraw Hill.

[18] Nimgaonkar, S. and Duppala, S. 2012. A Survey on Web Content Mining and extraction of Structured and Semi structured data, IJCA Journal.

[19] Oh, J. and Bandi, B. 2002.Multimedia Data Mining Framework for Raw video sequences. ACM. Third International Workshop on Multimedia Data Mining. Pp.
1-10.

[20] Pokorny, J. and Smigansky, J. 2005. Page Content Rank: An Approach to the Web Content Mining. In proceedings of IADIS International Conference Applied
Computing. Algarve, Portugal.

[21] Pol, K., Patil, N., Patankar, S. and Das, C. 2008. A Survey on Web Content Mining and extraction of Structured and Semi structured Data. IEEE First International Conference on Emerging Trends in Engineering and Technology. pp.543-546.

Analysis of Cloud Computing Risks in SAAS Applications

[1]**Subhash Chand Gupta & [2]Vikas Kumar**

1. Mewar University, Chittorgarh – 312901, Rajasthan, India
Email:- gupta_c_s@yahoo.co.uk
2. Asia-Pacific Institute of Management, 3&4 Institutional Area, Jasola, Sarita vihar ,
New Delhi – 110025, India Email:- prof.vikaskumar@gmail.com

A B S T R A C T

The objective of the present work is to identify the risks of Software-as-a-Service (SaaS) applications of cloud computing environment. In the SaaS environment, customers lose their direct control on the systems and the rigid service level agreements do not provide any detailed understanding of the involved risks. There is always an uncertainty about some elements of security, business continuity, and integration. This significantly influences the adopting organization's level of satisfaction with its overall SaaS experience. The cloud computing risks have been categorised as (a) Vendor related risk, (b) Security risk and (c) Regulatory risk. To determine the relevance of these risk dimensions, a web-based survey was conducted to take feedback from the organizational cloud decision-makers.

Keywords: Cloud Computing, SaaS, Risk assessment.

1. INTRODUCTION

Many of cloud users fear they will be suffering from confidentiality breach when using cloud services, although. Cloud computing is a new style of computing, in which dynamically scalable and often virtualized resources are provided as services over the Internet. It can be viewed as a collection of services, which are usually presented as a layered cloud computing architecture [1] - [2]. Cloud computing is a computing resource deployment and procurement model that enables an organization to obtain its computing resources and applications from any location via an Internet connection. Depending on the cloud solution model an organization adopts, all or parts of the organization's hardware, software, and data might no longer reside on its own technology infrastructure. Instead, all of these resources may reside in a technology centre shared with other organizations and managed by a third-party vendor. Cloud computing has the potential to the users [3] IT expenditure and to enable many new services to be developed. Using the cloud, even the smallest firms can reach out to ever larger markets while governments can make their services more attractive and efficient even while reining in spending. by many potential adopters of cloud computing, that the use of this technology may bring additional risks. For example, organizations may worry about business continuity in the case of service disruption whereas individuals may have concerns about what happens with their personal information. Such worries slow down the overall speed of adoption of cloud computing.

1.1 Cloud Deployment Models

The most common types of cloud computing deployment models, according to the National Institute of Standards of Technology [4] are:

- **Private cloud** – The cloud infrastructure is operated solely for an individual organization and managed by the organization or a third party; it can exist on or off the organization's premises.

- **Community cloud** – The cloud infrastructure is shared by several organizations and supports a specific community that has common interests (e.g., mission, industry collaboration, or compliance requirements). It might be managed by the community organizations or a third party and could exist on or off the premises.

- **Public cloud** – The cloud infrastructure is available to the general public or a large industry group and is owned by an organization selling cloud services.

- **Hybrid cloud** – The cloud infrastructure is composed of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability.

1.2 Cloud Service Delivery Models The cloud solutions offered by a CSP usually are referred to as cloud service delivery models, and the most common are:

- **Software as a Service (SaaS)**

Applications organizations use to perform specific functions or processes (e.g., email, customer management systems, enterprise resource planning systems, and spreadsheets). A more evolved offering of SaaS that is gaining popularity at the time of publication is known as Business Process as a Service (BPaaS). With BPaaS, entire business processes (e.g., payroll and supply-chain management) are outsourced to a third-party provider and supported by combinations of cloud service delivery solutions.SaaS is becoming an increasingly prevalent delivery model as underlying technologies that support web services and service-oriented architecture (SOA) mature and new developmental approaches become popular [5]. SaaS is also often associated with a pay-as-you-go subscription licensing model. Meanwhile, Broadband service has become increasingly available to support user access from more areas around the world. SaaS is most often implemented to provide business software functionality to Enterprise customers at a low cost while allowing those customers to obtain the same benefits of commercially licensed, internally operated software without the associated complexity of installation, management, support, licensing, and high initial cost. The

architecture of SaaS - based applications is specifically designed to support many concurrent users (multitenancy) at once. Software as a service applications are accessed using web browsers over the Internet therefore web browser security is vitally important. Information security officers will need to consider various methods of securing SaaS applications. Web Services (WS) security, Extendable Markup Language (XML) encryption, Secure Socket Layer (SSL) and available options which are used in enforcing data protection transmitted over the Internet [9]. “Within a few months of our founding, our customer base exploded,” says Joe Harrow, Director of Customer Service, Groupon. “At first, I was spending 10 percent of my time responding to customer requests. It gradually became a job for several agents. We realized we simply couldn't go on without a real ticketing solution. “Convinced that Groupon's rapid growth would continue, Harrow researched several enterprise-level support solutions. But he didn't find a good fit. “The enterprise-level solutions seemed complicated and difficult to set up,” [6] Harrow recalls. “They would have increased our efficiency, but at the cost of hampering the customer experience”. Harrow then searched the web for online support software and found Zendesk. After a quick evaluation of Zendesk, Harrow knew he had the right solution. “Right off the bat, Zendesk was intuitive to use,” Harrow says. “It seemed more powerful and robust than other online support solutions, and it had been rated very highly in reviews we'd read. Plus, we knew that because it was a web-based solution, it could easily scale to support our increasing volume. As illustrated in Figure 1.1, to qualify as a legitimate SaaS offering based on the prevailing definition, online software must be web-based, provider-owned, available only through a subscription or rental arrangement, allow the tenant to pay a periodic and predetermined usage fee, and have a cloud-based underlying infrastructure. In this SaaS pay-as-you-go arrangement, the software provider owns and maintains the hardware, software, and systems that make up a specific SaaS application.

Figure 1.1 The SaaS paradigm.

Cloud vendors and clients' need to maintain Cloud computing security at all interfaces. The sub-services of SAAS [12] are described as follows.

Communications-as-a-Service (CaaS)

CaaS is the delivery of an enterprise communications solution, such as Voice Over IP, instant messaging, and video conferencing applications as a service.

SECurity-as-a-Service (SECaaS)

SECaaS is the security of business networks and mobile networks through the Internet for events, database, application, transaction, and system .

Monitoring-as-a-Service (MaaS)

MaaS refers to the delivery of second-tier infrastructure components, such as log management and asset tracking, as a service. Organizations can access a wide range of applications, operating systems and services. These services frequently support collaborative working and the interlinking of services [12](mash ups).

Zoho Salesforce.com Basecamp Ulteo Google Apps

- Platform as a Service (PaaS) – Development environments for building and deploying applications. The CSP provides its customers with proprietary tools that facilitate the creation of application systems and programs that operate on the CSP's hosted infrastructure.

- Infrastructure as a Service (IaaS) – The CSP provides an entire virtual data center of resources (e.g., network, computing resources, and storage resources).

1.3. Cloud Computing Risks

"Risk is the possibility that an event will occur and adversely affect the achievement of objectives [7]" The types of risks (e.g., security, integrity, availability, and performance) are the same with systems in the cloud as they are with non-cloud technology solutions. An organization's level of risk and risk profile will in most cases change if cloud solutions are adopted (depending on how and for what purpose the cloud solutions are used). This is due to the increase or decrease in likelihood and impac t with respect to the risk events (inherent and residual) associated with the CSP that has been engaged

for services. The risks have been categorized as the Vendor Related Risks, Security Risks and the Regulatory Risks.

2. VENDOR RELATED RISK

Committing to a SaaS solution, only to have the vendor go out of business, could potentially cause a significant disruption to business activities. The following questions should give buyers an idea about the longevity of the vendor.

- How reliable are you - do you provide references, case studies and third party assessments?
- Do you have information available about your physical location and telephone number?
- Do you have information about your top management on your site?
- Do you have just a handful of customers, or thousands or tens of thousand?
- Are you a publicly listed company? And if not do reputable investors fund you?
- Are you well covered by traditional media and technology blogs?
- Are you active in blogs and social media sites?

a) Vendor Longevity

If the vendor goes out of business, it can be tremendously disruptive to your business. Checking out their financial standing, client references, management biographies, customer base, blog posts and media coverage can help give you a sense of their longevity. To mitigate this risk, the organization can consider what your plan would be to continue your operation if the vendor's operations are suddenly shut down.

b) Operational Reliability

With SaaS, users are at the mercy of their service provider. This means that risks with SaaS can involve inadequate uptime performance, service degradation during vendor maintenance, inadequate disaster recovery capabilities, software quality issues and security procedures, all of which should be addressed during contract negotiations.

c) Vendor Viability

SaaS customers should not overlook one of the most significant risks with SaaS—service provider viability. SaaS customers depend on the existence of their providers for virtually every routine business operation, meaning that if an SaaS company is financially volatile or encounters civil or criminal legal complications, all that company's customers can potentially go down the drain. While operational reliability issues can usually be addressed through contractual agreement, the effects of a failed SaaS partner are more difficult to mitigate. Many companies seek to create an escrow agreement that allows the customer to store backups of their software and data to guard against the danger of service provider failure. The only problem with an escrow agreement is the downtime an SaaS customer faces while scrambling to put the servers and other infrastructure in place necessary to operate their software[8]. To mitigate this risk, the organization can Perhaps the best way to deal with the SaaS provider survivability issue is to thoroughly investigate the provider's legal and financial standing prior to signing a deal.

d) Service Level Agreement

Service Level Agreements for Cloud Computing provides a unique combination of business-driven application scenarios and advanced research in the area of service-level agreements for Clouds and service-oriented infrastructures. Service Level Agreements for Cloud Computing contributes to the various levels of service-level management from the infrastructure over the software to the business layer, including horizontal aspects like service monitoring. Daily enterprise security practices count, too. “Leave all ports open, and expose all IP addresses, and enterprise data is more vulnerable in the cloud than in the data center,” said Staten. “But that's your own fault, not the fault of the cloud.” A Service Level Agreement should contain the following aspects:

- A list of the services the provider will deliver and a complete definition of each service.

- Metrics to determine if the provider is delivering the service as promised and an auditing mechanism to monitor the service.

- Responsibilities of the provider and the consumer and remedies available to both if the terms of the SLA are not met.

- A description of how the SLA will change over time.

e) Service Changes

As all IT professionals know, technology firms come and go. Cloud service providers may fail, be acquired, or change their business models and discontinue a service at any time. As a result, the organization may lose access to its data or to the service upon short notice. To mitigate this risk, the organization can

- Contractually require a specified minimum period of notice for service changes.
- Identify alternate service options that can be activated upon need.
- Maintain an updated internal copy of the data for emergency use.

f) Cost Changes

As with any outside service, costs for cloud services will change over time. Such changes can make this type of service less cost effective, jeopardizing the purpose behind using the service in the first place. To mitigate this, the organization can ensure that service costs and changes are defined in the service contract; and

evaluate the cost/benefit/risk trade-offs of the relationship during each contract renewal.

3. SECURITY RISKS

With the security risk and vulnerability in the enterprise cloud computing that are being discovered enterprises that want to proceed with cloud computing should, use the following steps to verify and understand cloud security provided by a cloud provider:-

Understand the cloud by realizing how the cloud's uniquely loose structure affects the security of data sent into it. This can be done by having an in-depth understanding of how cloud computing transmit and handles data.

Demand Transparency by making sure that the cloud provider can supply detailed information on its security architecture and is willing to accept regular security audit.

The regular security audit should be from an independent body or federal agency.

Reinforce Internal Security by making sure that the cloud provider's internal security

technologies and practices including firewalls and user access controls are very strong and can mesh very well with the cloud security measures.

Consider the Legal Implications by knowing how the laws and regulations will affect what you send into the cloud.

Pay attention by constantly monitoring any development or changes in the cloud technologies and practices that may impact your data's security.

A. Privileged User Access

Once data is stored in the cloud, the provider has access to that data and also controls access to that data by other entities (including other users of the cloud and other third party suppliers). Maintaining confidentiality of data in the cloud and limiting privileged user access can be achieved by at least one of two approaches by the data owner: first, encryption of the data prior to entry into the cloud to separate the ability to store the data from the ability to make use of it; and second, legally enforcing the requirements of the cloud provider through contractual obligations and assurance mechanisms to ensure that confidentiality of the data is maintained to required standards. The cloud provider must have demonstrable security access control policies and technical solutions in place that prevent privilege escalation by standard users, enable auditing of user actions, and support the segregation of duties principle for privileged users in order to prevent and detect malicious insider activity[13].

B. Data Security

IaaS (Infrastructure as a Service), PaaS (Platform as a Service), and SaaS (Software as a Service) are three general models of cloud computing. Each of these models possess a different impact on application security [14]. However, in a typical scenario where an application is hosted in a cloud, two broad security questions that arises are:

How secure is the Data?

How secure is the Code?

Cloud computing environment is generally assumed as a potential cost saver as well as provider of higher service quality. Security, Availability, and Reliability are the major quality concerns of cloud service users. Gens et. al. [15], suggests that security in one of the prominent challenge among all other quality challenges. These two major security issues are internally related to pretend several security threats, which are contentious challenge in cloud computing environment for confidential data. Those

threats not just affect Data security but also affect the whole cloud paradigm, where all the user as well as provider services underlying its constraints. Those security threats which bring the lack of security of data and code in cloud environments

are as follows: According to CSA (Cloud Security Alliance) the top threats in Cloud Environments are

[16]:

- Abuse and Nefarious Use of Cloud Computing
- Insecure Application Programming Interfaces
- Malicious Insiders
- Shared Technology Vulnerabilities
- Data Loss/Leakage Account, Service & Traffic Hijacking
- Unknown Risk Profile A multi-tenant cloud environment in which user organizations and applications share resources presents a risk of data leakage that does not exist when dedicated servers and resources are used exclusively by one organization. This risk of data leakage presents an additional point of consideration with respect to meeting data privacy and confidentiality requirements. To mitigate this risk, the organization can:

- Go deeper than a simple overview of their policy.
- Ask for encryption level and authentication protocol details.
- Establish how technicians are vetted and their overall data centre procedures.

C. Data Ownership

In the past it was not unusual for cloud service providers to claim ownership of all data/files submitted to their care, and to publicly state the right to redistribute all submitted documentation at will[17]. Data safety in the cloud is not a trivial concern. Some online storage vendors such as The Linkup and Carbonate have lost data, and were unable to recover it for customers. There are data access governance concerns, because there is the danger that sensitive data could fall into the wrong hands, either as a result of people having more privileges than required to do the job or by accidental or intentional misuse of the privileges they were assigned to do their job. the security of user data can be r

eflected in the following rules of implementation:

- The privacy of user storage data. User storage data cannot be viewed or changed by other people (including the operator).

- The user data privacy at runtime. User data cannot be viewed or changed by other people at runtime (loaded to system memory).

- The privacy when transferring user data through network. It includes the security of transferring data in cloud computing centre intranet and internet. It cannot be viewed or changed by other people.

- Authentication and authorization needed for users to access their data. Users can access their data through the right way and can authorize other users to access. To mitigate this risk, the organization should:

- Ensure that the service provider acknowledges the primacy of the organization's rights to the data submitted;

- Contractually obligate the service provider to limit use of the data to that approved by the organization; and

- Contractually require that the organization's data be returned and deleted upon severing the relationship.

D. Data Censorship

In some cases cloud providers retain to themselves the right to audit and censor any data submitted to the service. This can cause delays in the data posting process or changes to the data itself that may be unacceptable [18]. To mitigate this risk, the organization can

- Contractually define any such activities on the part of the service provider, and the process through which it is supported; and

- Require the service provider to report any such activities to the organization.

E. Encryption

In the current state of the cloud processing industry, implementation of encryption for data “at rest” within the cloud remains relatively rare, or is expensive when available. It has been identified within

some cloud services that are designed to meet specific regulatory requirements, but is not available within more common cloud services [18].
To mitigate this risk, the organization can:

- Use cloud services only for those data services that do not require data encryption for data privacy purposes; or

- Select a cloud service provider that can provide encryption if required;

- Confirm that the service provider has implemented appropriate encryption controls.

4. REGULATORY RISKS

i. Audit Records

In many cases the systems supporting cloud services are managed by service provider personnel through their internal processes. Service providers, as is true for most organizations, usually decline to provide internal operational details to external customers. Obtaining auditable records for the systems can therefore be difficult. In addition, the audit tools available through individual service providers vary significantly. The tools available for this purpose must be evaluated for each service provider[19]. Anytime you use a provider for cloud computing services it is important to have a 'right to audit' clause or at a minimum require the vendor to provide a report from an annual independent audit. This ensures that controls are working as they are intended.

To mitigate this risk, the organization can

- Verifying the functionality and controls of all supporting systems may be required for compliance with regulatory requirements, this becomes a significant risk.

ii. Storage Location

Typically, the service provider may store all or part of the data/files on servers where it is convenient for them. This may include transferring

the data to servers outside of the region or country of origin. Such a situation may or may not be permissible, based upon regulations under which the organization operates.To mitigate this risk, the organization can require that the service provider

- Assess the organization's legal and regulatory requirements, and determine if restrictions exist on whether it can permit its data to be stored outside of a specific legal jurisdiction;

- If the organization faces this type of restriction, include a discussion of this issue with any service providers considered as service providers;

- If possible, agree contractually to storage location restrictions that will keep the organization compliant with its regulatory needs; or

- If this is not possible, exclude the service provider from those to be considered for the service.

iii. Lack of breach notice

In addition to all the other controls that are missing, the organization using cloud services also loses insight into and control over mandated breach notifications. More often than not, the organization may not even know that a breach has occurred until it appears in the evening news.

To mitigate this risk, the organization can

- Use cloud resources for only those data applications that do not have regulatory compliance requirements.

- Contract with a cloud service provider that is willing to assert contractually that they will notify the organization at once in the event of a possible breach.

iv. Compliance

The issue of compliance validation for cloud computing applications is an open question. Little has been done in this area to date, and support for compliance requirements varies considerably between providers[20]. companies which apply cloud computing platform: An advanced platform with unified standard is provided and the quality is guaranteed. IT management becomes easier and the costs of developing products is greatly lowered. Response speed for business demand is enhanced and expandability is ensured. Existing applications and newly-emerged data-intensive applications are supported. Miscellaneous functions for expediting the speed of innovation is also provided for outsourcing service companies, colleges and universities and research institutes. To mitigate this risk, the organization can

- Use cloud resources for only those data applications that do not have regulatory compliance requirements.

- Contract with a cloud service provider that is willing to assert contractually that they will maintain compliance with the regulation in question.

- Maintain a compliance verification program that validates the service provider's compliance with the requirements.

5. CONCLUSION

The use of external resources for cloud computing in its current state involves a number of risks. The key to proper mitigation of the risks in cloud computing is to determine appropriate controls for all relevant security provider operations just as if they were internal, and then to contractually obligate the service provider to comply with those controls. In the process of mitigating its risks, any organization making use of cloud computing resources should take a number of key steps. Evaluate the risks involved in the use of cloud computing for a specific data application, and determine if the benefits to be gained offset the risks and the costs. This is especially critical if any regulatory compliance requirements are involved. Assess the available cloud computing service providers to determine if any can provide the needed service while providing appropriate support in mitigating the identified risks. Perform appropriate due diligence on the selected service provider to ensure their financial stability, and to confirm the promised support architecture is available.

REFERENCES:

[1]. Sun Microsystems, "Introduction to cloud computing architecture", White Paper, Sun Microsystems, June 2009.

[2]. Chappell D., "A short introduction to cloud platforms: An enterprise-oriented view" IEEE ITPro, pp. 23–27, August 2008.

[3]. Thomas Ristenpart, Eran Tromer, Hovav Shacham, and Stefan Savage, "Hey, You, Get Off of My Cloud: Exploring Information Leakage in Third-PartyCompute Clouds,"ACM , November2009 http://cseweb.-ucsd.edu/~hovav/dist/cloudsec.pdf..

[4]. Kretschmer, T. (2012), "Information and Communication Technologies and Productivity Growth: A Survey of the Literature", OECD Digital Economy Papers, No. 195, OECD Publishing.http://dx.doi.org/10.1787/5k9bh3jllgs7-en

[5]. Carl Brooks, "Cloud SLAs the next bugbear for enterprise IT," TechTarget.com – http://searchcloudcomputing-.techtarget.com/news/2240036361/Cloud-SLAs-the-next-bugbear-for-enterprise-IT?asrc=EM_EDA_14018952.

[6]. Cloud Security Alliance, Top Threats to Cloud Computing V1.0, March 2010, "Threat #1: Abuse and Nefarious Use of Cloud Computing," p. 8 – https:// cloudsecurityalliance .org/topthreats/csathreats.v1.0.pdf.

[

7]. Wendy Butler Curtis, Curtis Heckman, and Aaron Thorp, "Cloud Computing: eDiscovery Issues and Other Risk," Orrick eDiscovery Alert, June 28, 2010 – http:// www.orrick. com /fileupload/2740.pdf.

[8]. Hinchcliffe, D. (2009, March 3). Cloud computing: A new era of IT opportunity andchallenges. ZDNet. March 3rd, 2009. http://blogs.zdne t.com/Hinchcliffe/?p=261Hoover, J. N. (2008, August 16). Outages force cloud computing user to rethink tactics . InformationWeek. Retrieved on March 26, 2010 fromhttp://www.informationweek.com/news/services/saas/showArticle.jhtml?articleID=210004236

[9]. S. Subashini, and V. Kavitha. (2010) "A survey on security issues in service delivery models of cloud computing." J Network Comput Appl doi:10.1016/j.jnca.2010.07.006. Jul.,2010.

[10]. Peter Mell and Timothy Grance, The NIST Definition of Cloud Computing, Special Publication 800-145, http://csrc.nist.gov/publications/PubsSPs.html#800-145.

[11]. "Information Security Briefing Cloud Computing " January 2010 http://www.cpni.gov.uk/Documents/Publications/2010/2010007ISB_cloud_computing.pdf

[12]. Lucas Mearian, "How data security can vaporize in the cloud," Computerworld, October 15, 2009 – http://www.computerworld.com/s/article/9139404/How_data_security_can_vaporize_in_the_cloud_. Greene, T. (2009). New attacks on cloud services call for due diligence. Network World.South borough: Sep 14, 2009. Vol. 26, Iss. 28; pg. 8, 1 pgs. Retrieved from http://www.networkworld.com/newsletters/vpn/2009/090709cloudsec2.html International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.1, January 2011

[13]. Crowe Horwath LLP, Warren Chan ,Eugene Leung , Heidi Pili , Jume 2012 "ENTERPRISE

[14]. RISK MANAGEMENT FOR CLOUD COMPUTING " http://www.coso.org/documents/Cloud%20Computing-%20Thought%20Paper.pdf

[15]. "Cloud- 0Risks with SaaS", http://www.owasp.org/index.php/Cloud-10_Risks_with_SaaS

[16]. Joseph Granneman, "Data Protection and Access Control in the Cloud," Security and Compliance in the Cloud, ISACA Virtual Seminar, December 2010.

[17]. Carl Brooks, "Cloud SLAs the next bugbear for enterprise IT," TechTarget.com – http://searchcloudcomputing.techtarget.com/news/2240036361/Cloud-SLAs-the-next-bugbear-for-enterprise-IT?asrc=EM_EDA_14018952.

[18]. "Cloud Computing: Benefits, risks and recommendations for information security," European Network and Information Security Agency (ENISA), November 2009 – http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computing-risk-assessment.

[19]. Joseph Granneman, "Data Protection and Access Control in the Cloud," Security and Compliance in the Cloud, ISACA Virtual Seminar, December 2010.

[20]. Dr. Thomas Helbing, "How the New EU Rules on Data Export Affect Companies Running Cloud Computing and SaaS," cloudcomputing-vision.com, April 16, 2010 – http://cloudcomputing-vision.com/805/eu-rules-data-export-affect-companies-running-cloud-computing-saas/.

Rad (Rapid Application Development) Model for Mini Erp Application in Trading Company

[1]Pipit Dewi Arnesia & [2]Tristyanti Yusnitasari

1. Information System Department Gunadarma University Depok INDONESIA
pdarnesia@staff.gunadarma.ac.id
2. Information System Department Gunadarma UniversityDepok INDONESIA
tyusnita@staff.gunadarma.ac.id

A B S T R A C T

RAD (Rapid Application Development) model is an incremental model of the software development process that emphasizes system development cycle. This model breaks the project into smaller parts where each part is built with a model similar to the Waterfall model. The main objective of this model is to finish a project by section, as well as the planning process (although the initial planning is globally). As we all know that the trading company activities include export, import, and distribution. Data management is very important in the activities of the company and information technology is required to support the smooth process and decision making. In order to increase the efficiency, an application that can be integrated is required to be built. Mini ERP (Enterprise Resource Planning) application is an application that can help a trading company in conducting its business.

Mini ERP application includes purchase, sales, accounting (general ledger) and inventory stock module units. This module is a single entity that integrates its functions. Model development is done from the system engineering stage to the stage of analysis, design, programming, and testing.

Keyword: mini ERP, RAD model, UML, trading

1. INTRODUCTION

In developing an information system, there are various methodologies that can be used. Each of these methodologies will elaborate on the stages in the development of an information system. As for the goal of all of these methodologies is the success of the developed information system with timely, appropriate cost and the expected needs of the user. To examine the issues of timeliness and costs associated with the development of information systems, it is important to discuss each stage in the development of information systems as well as analyze the gaps that have the potential to be penetrated from each stage. So the failure of system development can be avoided and information . produced from the system can be guaranteed. A trading company which includes export, import and distribution activities should have a good data management so that the goal of creating a mini ERP application design is a mini ERP applications built with modules that can be integrated, easy to manage the data and support fluency in decision-making, and also improve efficiency. Application mini ERP using RAD model is a developmental process of linear sequential software that emphasizes

system development cycle in a short time (60 to 90 days) with a component-based construction approach

2. RESEARCH METHOD

Before creating a draft information system, a research method to determine what steps will be taken for the system development is needed. RAD model emphasizes the following phases:

§ **Business modeling.** At this stage, the flow of information on the business functions are modeled to know what information to control the business processes, what information is produced, who makes the information, to

§ where the information flows, and who process it.

§ **Data modeling.** The flow of information that is defined from business modeling, filtered again so that it can be used as parts of the data objects needed to support the business. Characteristics (attributes) of each object is specified along with the relationship between the object.

§ **Process modeling.** Data objects defined previously modified in order to produce a flow of information to be implemented into business functions. Description processing are made to add, modify, remove or take back the object data.

§ **Application generation.** RAD model works by using fourth generation techniques (4GT), so at this stage it is very rarely to use a conventional programming using third - generation programming languages, but with more emphasis on reuse components (if any) or create new components (if necessary). In all cases, the tools for automation is used to facilitate the creation of software.

§ **Testing and turnover.** Because of the emphasis on reuse of existing components, some of these components have been tested previously, thereby reducing the overall testing time except for the new components.

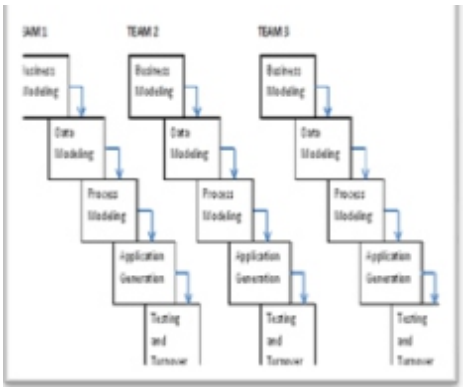


Figure 1. Rapid Application Development (RAD

An analysis of the data observation results is done using the concepts and methods of object-oriented system with the tool in the form of UML (Unified Modelling Language).

3. RESULT AND DISCUSSION

3.1 Use Case Diagram

Use case diagrams are representations of the system from the point of view of the system user, so the creation of use case more emphasized on functionality in the system, not based on the flow or sequence of events. In this case, there are some actors who perform activities in the system and some use cases that describe what the actors do in the system. Actors and use cases are connected to each other which is characterized by the presence of stereotype << uses >> or often referred to as the stereotype << include >>. The relationship is described by Use Case Diagrams can be seen in figure 2.

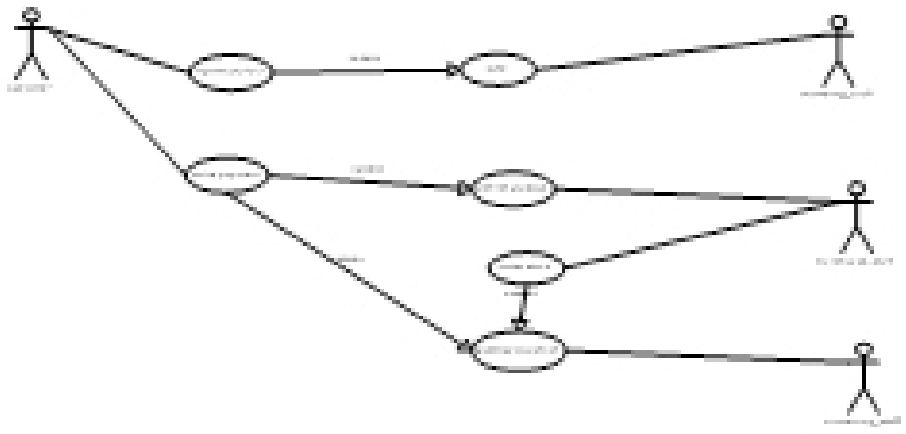


Figure 2. Use Case Diagram

3.2 Class Diagram

Class diagram illustrates how the objects / classes of different systems can connect each other. In other words, class diagrams describe the static structure of the system. In this system there are some classes thatare related. Each class has an attribute name and a different method. The main classes in the system are Customer class, Product class, Supplier class, and Staff class that are inherited into classes such as Marketing Staff, Accounting Staff and Warehouse staff in accordance with the principles of object inheritance. In this system, several classes of interfaces also need to be involved to be able to bridge the dialogue between the user with a system that is generally a form field (form). Interface class is a class

that has a method but does not have attributes (properties). Class interfaces can be distinguished from the stereotype << interface >> on behalf of the class. For more details can be seen in Figure 3.

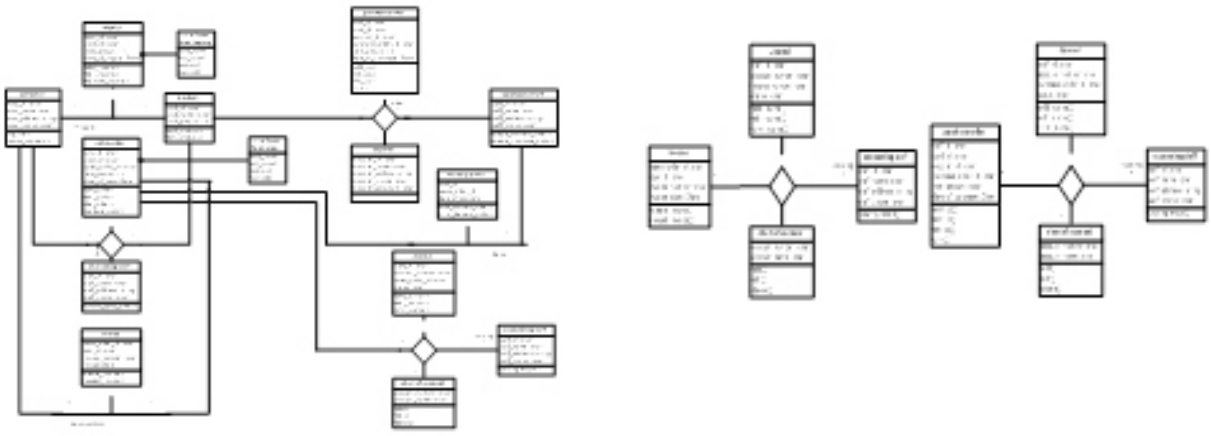


Figure 3. Class Diagram

3.3 Activity Diagram

Activity diagrams illustrate the flow of events in the system that is being designed, how each flow begins, a decision that may occur, and how they ended. Activity diagrams can also describe the parallel processes that may occur on some executions. Activity diagram is a special state diagram, where most of the state are actions and most of the transition are triggered by the completion of the previous state (internal processing). Therefore the activity diagram does not describe the internal behavior of a system (and interactions between subsystems), but rather describes the processes and the activities of top level in general, describe business processes and sequence of events in a process, and used in business modeling to show the sequence of activities of a business process.

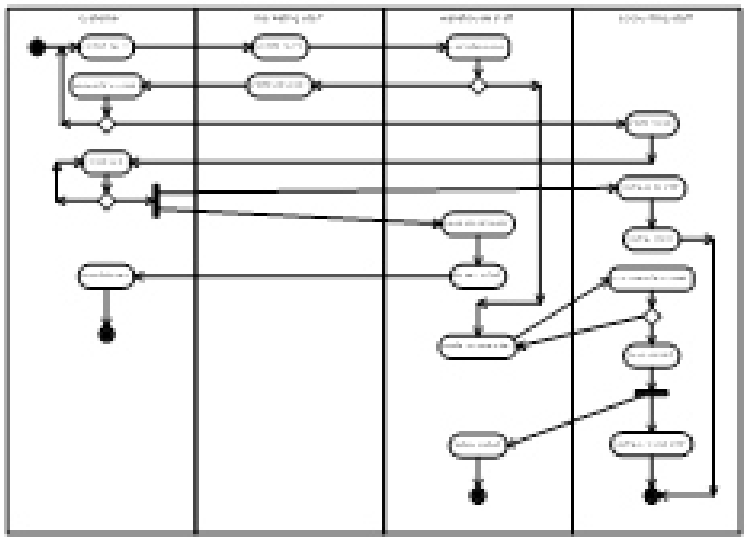


Figure 4. Activity Diagram

4. APPLICATION INTERFACE

After designing a system, carried out the construction of proper application as supporting this system. Application software and made by using PHP and mySQL. This application is divided into six main modules namely module card, inquiry, sales, purchase, project, cash management and general ledger.

Figure 5. Main Menu

Sales and Receivables module contains several menu as shown in Figure 6.

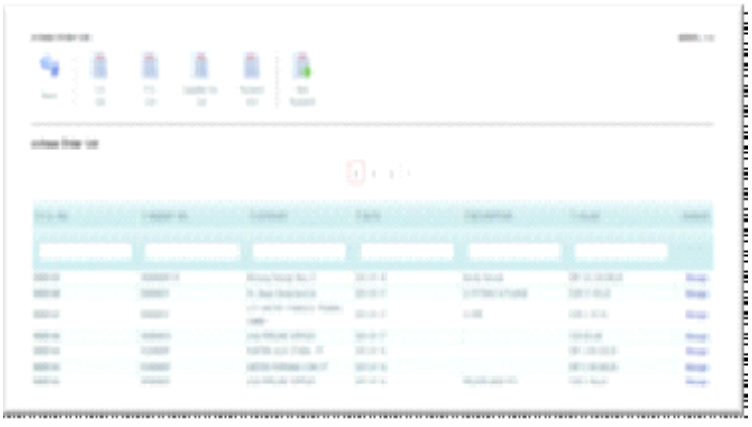


Figure 6. Sales and Receivables Menu

Menu on the module purchase and payable is shown in Figure 7

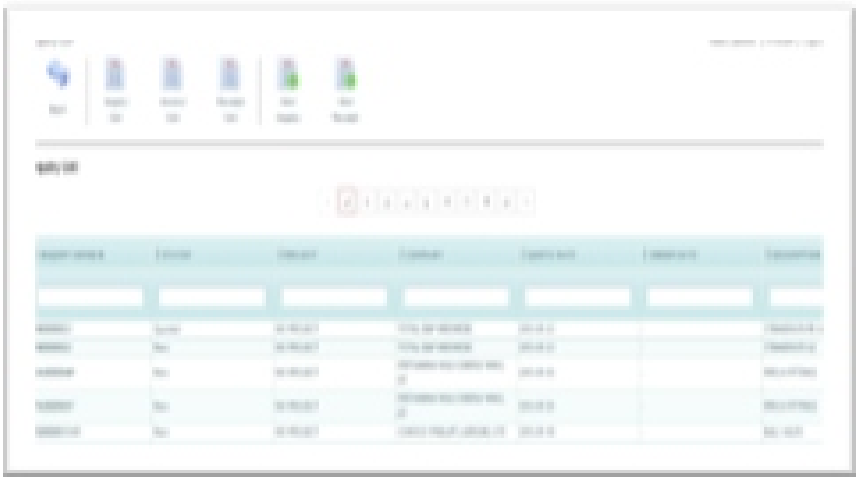


Figure 7. Purchase and Payable Menu

5. CONCLUSION

In object-oriented systems architecture, in this case is the preparation of mini ERP application, requires a methodology that aims to develop the success of information systems with timely, appropriate cost and the expected needs of the user. The model used here is RAD. By leveraging the advantages of the RAD model which is RAD model faster than Waterfall model if the needs and limitations are already known and also if the project allow for modularized, so that mini ERP can be built with RAD model. To facilitate the creation of the application, this system used graphic diagrams of the Unified Modeling Language (UML) which consists of Use Case Diagram, Activity Diagram, and Class Diagram. Interface was created to facilitate the users.

6. REFERENCES

[1]. Alexis Leon., 2006. Database Development Life Cycle, http://www.leon- leon.com/wp/2005/11/21/ddlc.html (June 01, 2008).

[2]. Boehm, B (1999), 'Making RAD work for your Project', IEEE Computer, March, pp113-117.

3]. Darryl Green and Ann DiCaterino (February 1998), "A Survey of System Development Process Models" (http://www.ctg.albany.edu/publications/reports/survey_of_sysdev)

[4]. J. R. McBride; Copyright 2002 Prentice-Hall, Inc, "Introduction to Systems Analysis, Topic 19, Rapid Application Development"(http://www.csc.uvic.ca/~jmcbride/c375t19.pdf)

[5]. Linda Night, Theresa Steinbach, and Vince Kellen (November 2001), "System Development Methodologies for Web Enabled E-Business: A Customization Paradigm" (http://www.kellen.net/SysDev.htm)

[6]. Paul Fisher, James Mc,Daniel, and Peter Hughes, "System Development Life Cycle Models and Methodologies" Canadian Society for International Health Certific ate Course in Health In formation Systems, Module 3: System Analysis & Database Development, Part 3: Life Cycle Models and Methodologies; (http://famed.ufrgs.br/pdf/csih/mod3/Mod_3_3.htm)

[7]. Paul Beynon-Davies; Kane Thompson Centre(Desember 1998), "Rapid Application Development: A Review and Case Study" (http://www.comp.glam.ac.uk/SOC_Server/research/gisc/RADbrf1.ht)

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).
2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);
3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);
4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

Address of the Editorial Office:

Enriched Publications Pvt. Ltd.
S-9,IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR -5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-45525005

Notes: