Journal of Cloud Computing and Data Base Management

Jan – Apr 2025 Vol – 11 Issue – 1

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

Journal of Cloud Computing and Data Base Management

Aims and Scope

Journal of Cloud Computing and Database Management is a peer-reviewed 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.

Journal of Cloud Computing and Data Base Management

Journal of Cloud Computing and Data Base Management

Jan - April 2025 Vol – 11 Issue - 1

Cloud Computing Rolls in Different Cloud Computing Environment Techniques

Md Shahid Anwar
M.tech (Research Scholar)
Al-Falah School of Engineering and Technology, Dhouj, Faridabad.

Ashif Ali
(Assistant Professor, CSE Deptt.)
Al-Falah School of Engineering and Technology, Dhouj, Faridabad.

Abstract

This research focused on the study of Cloud computing techniques worked on different environment and other internet-based services continue to develop rapidly, though it hasn't exactly been a direct path to get to where we are now. The current state of the industry may seem like obvious and inevitable when we look back, but just a short time ago it would have been hard to guess that this is where this is where things were actually going. The history of cloud computing has gone through a number of major changes that have made it more accessible and affordable. Like many other things, though, it's important to understand where it's been to make any sort of guess at where it's going. The current state of cloud computing rests on a strong internet backbone, but that isn't how it started or where it ends. The private cloud is now an important part of many business IT infrastructures, making elements like virtual.

Keyword: Cloud Computing, ARPANET, VM, CLBADM, FCFS.

1. Introduction

The Principles of Cloud Computing

The images of serious computing in the 50s and 60s – those pictures of row upon row of magnetic tape machines – are actually a foreshadowing of a cloud computing structure. In other words, companies were already using a lot of machines to provide more power than a single unit could and, on top of that, allowing more than one user to access the same assets. Back in the 50s, those giant mainframes were installed in schools, government organizations, and large corporations because they were the only places that could possibly house all those machines. Even then, multiple instances of the mainframe would be inconceivable, so it became normal practice to develop "dumb terminals" that allowed multiple people to access the necessary resources. This is the same principle as modern day virtualization, which puts us on the path toward cloud computing.

2. History of Virtual Machines Implementation History

The real implementation of virtual machines came in the 70s when IBM released an operating system called VM. This allowed multiple distinct computers to reside in the same processing environment, leading to the type of interactions we know call virtualization. In basic terms, it means that each individual user would have a machine with its own memory, processor, and other hardware components, but many of the resources would be shared by others. This type of "group computing" showed companies that they could start adding network solutions without actually increasing their hardware infrastructure. It was all about provisioning the resources they already had, shifting traffic as necessary, and balancing the load on the network and bandwidth to provide better services to their customers.

Telecommunications solutions were an integral part of cloud development, and this became possible with the commercialization of the internet. The network on which it is based, though, goes back to the 60s when J.C.R. Lickliker enabled the development of ARPANET (Advanced Research Projects Agency Network). This would eventually become the forerunner of the modern internet.The notion of connecting people all over the world to access programs and data from different locations became a real possibility. By the 70s, people were really delving into the potential suggested by those first experiments in the 60s. In 1971, for example, the first email was sent, and the U.S. Department of Defense continued developing ARPANET into the internet. In 1979, both CompuServe Information Services and The Source both went online, showing that it was possible for commercial service providers to host internet services. Still, it wasn't until 1993 that the Mosaic browser made the internet far more graphical – something that the average user could manage. It was soon after that when Netscape launched, and then, in 1995, both Amazon and eBay appeared.

3. Industrial Revolution – Affordable Computing

Part of the reason for the gap between 1979 and 1993 was that computers were still not affordable or compact enough for people to have in their homes or for companies to outfit their entire staff. The 80s saw the biggest boom in computers, with IBM putting out a range of affordable personal computers and Microsoft pushing its operating system out in a large scale. Then, in the 90s, there was finally sufficient bandwidth available to really make the internet available to the masses, which meant that all those companies that had outfitted their staff with computers now had a valid way to connect them all. Without this kind of high-speed bandwidth and software interoperability, this type of connected computing would never have worked.

Modern History Service-Oriented Architecture-The rise of commercial networking wasn't an easy one, and once the first bubble burst in 2000, companies had to start rethinking their business models. The lesson learned was that no matter how much money investors threw at you, you still needed a solid business plan to survive in the long run. In the search for new ways to monetize the internet, many companies started to realize that they could provide a service model to deliver usable solutions and resources. Salesforce.com was the company that really started this trend by pioneering the concept of delivering enterprise-class applications over a simple website. Next, in 2002, Amazon got on board the trend with Amazon Web

Services. This gave users the ability to access storage, computation solutions, and other apps through the internet. In 2006 they went further with the Elastic Compute cloud (EC2), which basically let developers rent space on their computers to store and run their own apps. It was an entire infrastructure that they delivered as a service. By 2009, most of the industry influencers were on board, with companies like Microsoft and Google delivering apps to the average consumer as well as businesses in the form of simple, accessible services. The ubiquity of cloud computing has led to an environment in which companies don't have to go to third parties to take advantage of this resource. The technology has developed to the point that organizations can effectively deploy their own private or hybrid clouds, rather than rely on public clouds. This can potentially increase performance and decrease certain costs in this area. More importantly, a private cloud deployment gives the IT team more visibility into the back end of their system, which is particularly useful for companies that are extremely security conscious and require direct oversight on all their assets. Private cloud deployments are becoming more prevalent because they offer a lot of the same cost and convenience benefits, and they support various platforms while allowing the organization to maintain more control. The road to get to this point has been a long one, and while it may be difficult to predict exactly what the future holds, there are currently a lot of benefits to provisioning resources over a safe, secure, private network.

4. Loads Balancing In Cloud Computing Environment

Load balancing in cloud computing provides an efficient solution to various issues residing in cloud computing environment set-up and usage. Load balancing must take into account two major tasks, one is the resource provisioning or resource allocation and other is task scheduling in distributed environment. Efficient provisioning of resources and scheduling of resources as well as tasks will ensure: a. Resources are easily available on demand. b. Resources are efficiently utilized under condition of high/low load. c. Energy is saved in case of low load (i.e. when usage of cloud resources is below certain threshold). d. Cost of using resources is reduced. For measuring the efficiency and effectiveness of Load Balancing algorithms simulation environment are required. CloudSim [12] is the most efficient tool that can be used for modeling of Cloud. During the lifecycle of a Cloud, CloudSim allows VMs to be managed by hosts which in turn are managed by datacenters. Cloudsim provides architecture with four basic entities. These entities allow user to set-up a basic cloud computing environment and measure the effectiveness of Load Balancing algorithms. A typical Cloud modeled using CloudSim consists of following four entities Datacenters, Hosts, Virtual Machines and Application as well as System Software. Datacenters entity has the responsibility of providing Infrastructure level Services to the Cloud Users. They act as a home to several Host Entities or several instances hosts' entities aggregate to form a single Datacenter entity. Hosts in Cloud are Physical Servers that have pre-configured processing capabilities. Host is responsible for providing Software level service to the Cloud Users. Hosts have their own storage and memory. Processing capabilities of hosts is expressed in MIPS (million instructions per second).

They act as a home to Virtual Machines or several instances of Virtual machine entity aggregate to form a Host entity. Virtual Machine allows development as well as deployment of custom application service models. They are mapped to a host that matches their critical characteristics like storage, processing, memory, software and availability requirements. Thus, similar instances of Virtual Machine are mapped to same instance of a Host based upon its availability. Application and System software are executed on Virtual Machine on-demand. Class diagram of Cloud architecture illustrating relationship between the four basic entities is shown in fig 1. Thus, the object oriented approach of Cloud Sim can be used to simulate Cloud Computing environment.

5. Load Balancing on the Basis of Cloud Environment

Cloud computing can have either static or dynamic environment based upon how developer configures the cloud demanded by the cloud provider.Static Environment In static environment the cloud provider installs homogeneous resources. Also the resources in the cloud are not flexible when environment is made static. In this scenario, the cloud requires prior knowledge of nodes capacity, processing power , memory, performance and statistics of user requirements. These user requirements are not subjected to any change at run-time. Algorithms proposed to achieve load balancing in static environment cannot adapt to the run time changes in load. Although static environment is easier to simulate but is not well suited for heterogeneous cloud environment. Round Robin algorithm [13] provides load balancing in static environment. In this the resources are provisioned to the task on first-cum-first-serve (FCFS- i.e. the task that entered first will be first allocated the resource) basis and scheduled in time sharing manner. The resource which is least loaded (the node with least number of connections) is allocated to the task. Eucalyptus uses greedy (first-fit) with round-robin for VM mapping. Radojevic proposed an improved algorithm over round robin called CLBDM (Central Load Balancing Decision Model) [14]. It uses the basis of round robin but it also measures the duration of connection between client and server by calculating overall execution time of task on given cloud resource.

6. Load Balancing Based on Spatial Distribution of Nodes

Nodes in the cloud are highly distributed. Hence the node that makes the provisioning decision also governs the category of algorithm to be used. There can be three types of algorithms that specify which node is responsible for balancing of load in cloud computing environment. Centralized Load Balancing In centralized load balancing technique all the allocation and scheduling decision are made by a single node. This node is responsible for storing knowledge base of entire cloud network and can apply static or dynamic approach for load balancing. This technique reduces the time required to analyze different cloud resources but creates a great overhead on the centralized node. Also the network is no longer fault tolerant in this scenario as failure intensity of the overloaded centralized node is high and recovery might not be easy in case of node failure.

7. Conclusion:

The overall conclusion of this paper based on cloud computing, Load Balancing is an essential task in Cloud Computing environment to achieve maximum utilization of resources. In this paper, we discussed various load balancing schemes, each having some pros and cons. On one hand static load balancing scheme provide easiest simulation and monitoring of environment but fail to model heterogeneous nature of cloud. On the other hand, dynamic load balancing algorithm are difficult to simulate but are best suited in heterogeneous environment of cloud computing. Also the level at node which implements this static and dynamic algorithm plays a vital role in deciding the effectiveness of algorithm. Unlike centralized algorithm, distributed nature of algorithm provides better fault tolerance but requires higher degree of replication and on the other hand, hierarchical algorithm divide the load at different levels of hierarchy with upper level nodes requesting for services of lower level nodes in balanced manner. Hence, dynamic load balancing techniques in distributed or hierarchical environment provide better performance. However, performance of the cloud computing environment can be further maximized if dependencies between tasks are modeled using workflows.

References

[1] Foster, I., Zhao, Y., Raicu, I. & Lu, S. (2008). Cloud computing and grid computing 360-degree compared. IEEE Grid Computing Environment Workshops.

[2] Youseff, L. Toward a Unified Ontology of Cloud Computing Retrieved from http://spsteve.wikispaces.asu.edu/file/view/unified+interface+

cloud.pdf.
[3] Luo, S., Lin, Z. & Chenm, X. (2011). Virtualization security for cloud computing service. 2011 International Conference on Cloud and Service Computing 978-1-4577-1637-9/11/$26.00 ©2011 IEEE. Shenzhen, China: ZTE Corporation.
[4] Zhang, Q., Cheng, L. & Boutaba, R. (2010). Cloud computing: State-of-the-art and research challenges. Journal of Internet Services and Applications, 1(1), 7-18. DOI 10.1007/s13174-010-0007-6.
[5] Open Stack: An Overview. Retrieved from www.openstack. org/downloads/openstack-overview-datasheet.pdf.
[6] White Paper (2013). Measuring the Business Value of VMware Horizon View.
[7] Huang, A. Software Architect, Citrix Systems Apache Cloud-Stack Architecture.
[8] Amazon Elastic Compute Cloud (EC2). Retrieved from http://www.amazon.com/ec2/ (accessed on April 18, 2010).
[9] Google App Engine, http://appengine.google.com (April 18, 2010).
[10] Chappell, D. (2008). Introducing the Azure services platform. White paper.
[11] CliQr Makes Cloud Benchmarking Available to All Users, Providing Critical Information Needed to Optimize Application Deployment Cloud. Retrieved fromwww.cliqr.co /assets/downloads/press/CliQrBenchmark.06.25.13.pdf.
[12] Sotomayor, B., Montero, R. S., Llorente, I. M. & Foster, I. (2009). Virtual infrastructure management in private and hybrid clouds. IEEE Internet Computing, 13(5), 14-22.
[13] Radojevic, B. & Zagar, M. (2011). Analysis of issues with load balancing algorithms in hosted (cloud) environments. In proceedings of International Convention on MIPRO, IEEE.

Developed Relational Star Schema Query and Roll of Data Warehouse in this Model

Jyoti
(M.Tech Research Scholar)
CBS Group of Institution ,Fathepuri, Jhajjar

Rajiv Munjal
Assistant Professor, CSE, Department CBS Group of Institution,
Fathepuri, Jhajjar

Abstract

This research paper based on the computing, of the Star Schema technique is the simplest style of data mart schema. The star schema consists of one or more fact tables referencing any number of dimension tables. The star schema is an important special case of the snowflake schema, and is more effective for handling simpler queries. The star schema separates business process data into facts, which hold the measurable, quantitative data about a business, and dimensions which are descriptive attributes related to fact data. Examples of fact data include sales price, sale quantity, and time, distance, speed, and weight measurements. Related dimension attribute examples include product models, product colors, product sizes, geographic locations, and salesperson names.A star schema that has many dimensions is sometimes called a centipede schema. with many table joins and makes the star schema less easy to use.

Keyword: Denormalization, Schema, Start, Schema , Data-driven DSS with online analytical processing (OLAP).

1. Introduction

The star schema gets its name from the physical model resemblance to a star with a fact table at its center and the dimension tables surrounding it representing the star's Points. In computing, the Star Schema is the simplest style of data mart schema.

The star schema consists of one or more fact tables referencing any number of dimension tables. The star schema is an important special case of the snowflake schema, and is more effective for handling simpler queries.[1]The star schema gets its name from the physical model's[2] resemblance to a star with a fact table at its center and the dimension tables surrounding it representing the star's points. Having dimensions of only a few attributes, while simpler to maintain, results in queries

2. History of Star Schema

The star schema separates business process data into facts, which hold the measurable, quantitative data about a business, and dimensions which are descriptive attributes related to fact data. Examples of fact data include sales price, sale quantity, and time, distance, speed, and weight measurements. Related dimension attribute examples include product models, product colors, product sizes, geographic locations, and salesperson names. A star schema that has many dimensions is sometimes called a centipede schema.[3] Having dimensions of only a few attributes, while simpler to maintain, results in queries with many table joins and makes the star schema less easy to use. Fact tables record measurements or metrics for a specific event. Fact tables generally consist of numeric values, and foreign keys to dimensional data where descriptive information is kept.[3] Fact tables are designed to a low level of uniform detail (referred to as "granularity" or "grain"), meaning facts can record events at a very atomic level. This can result in the accumulation of a large number of records in a fact table over time. Fact tables are defined as one of three types: Transaction fact tables record facts about a specific event (e.g., sales events) Snapshot fact tables record facts at a given point in time (e.g., account details at month end) Accumulating snapshot tables record aggregate facts at a given point in time (e.g., total month-to-date sales for a product) Fact tables are generally assigned a surrogate key to ensure each row can be uniquely identified [5].

Dimension tables usually have a relatively small number of records compared to fact tables, but each record may have a very large number of attributes to describe the fact data. Dimensions can define a wide variety of characteristics, but some of the most common attributes defined by dimension tables include: Time dimension tables describe time at the lowest level of time granularity for which events are recorded in the star schema Geography dimension tables describe location data, such as country, state, or city Product dimension tables describe products Employee dimension tables describe employees, such as sales people Range dimension tables describe ranges of time, dollar values, or other measurable quantities to simplify reporting Dimension tables are generally assigned a surrogate primary key, usually a single-column integer data type, mapped to the combination of dimension attributes that form the natural key. Star schemas are denormalized, meaning the normal rules of normalization applied to transactional relational databases are relaxed during star schema design and implementation.

The benefits of star schema denormalization are: Simpler queries - star schema join logic is generally simpler than the join logic required retrieving data from highly normalized transactional schemas. Simplified business reporting logic - when compared to highly normalized schemas, the star schema simplifies common business reporting logic, such as period-over-period and as-of reporting. Query performance gains star schemas can provide performance enhancements for read-only reporting applications when compared to highly normalized schemas. Fast aggregations - the simpler queries against a star schema can result in improved performance for aggregation operations. Feeding cubes - star schemas are used by all OLAP systems to build proprietary OLAP cubes efficiently; in fact, most major OLAP systems provide a ROLAP mode of operation which can use a star schema directly as a source without building a proprietary cube structure. The main disadvantage of the star schema is that data integrity is not enforced as well as it is in a highly normalized database. One-off inserts and updates can result in data anomalies which normalized schemas are designed to avoid. Generally speaking, star schemas are loaded in a highly controlled fashion via batch processing or near-real time "trickle feeds", to compensate for the lack of protection afforded by normalization. Star schema is also not as flexible in terms of analytical needs as a normalized data model. Normalized models allow any kind of analytical queries to be executed as long as they follow the business logic defined in the model. Star schemas tend to be more purpose-built for a particular view of the data, thus not really allowing more complex analytics. Star schemas don't support many-to-many relationships between business entities - at least not very naturally. Typically these relationships are simplified in star schema to conform the simple dimensional model. Consider a database of sales, perhaps from a store chain, classified by date, store and product. The image of the schema to the right is a star schema version of the sample schema provided in the snowflake schema article [9].

Fact_Sales is the fact table and there are three dimension tables Dim_Date, Dim_Store and Dim_Product.Each dimension table has a primary key on its Id column, relating to one of the columns (viewed as rows in the example schema) of the Fact_Sales table's three-column (compound) primary key (Date_Id, Store_Id, Product_Id). The non-primary key Units_Sold column of the fact table in this example represents a measure or metric that can be used in calculations and analysis. The non-primary key columns of the dimension tables represent additional attributes of the dimensions (such as the Year of the Dim_Date dimension).For example, the following query answers how many TV sets have been sold, for each brand and country, in 1997:

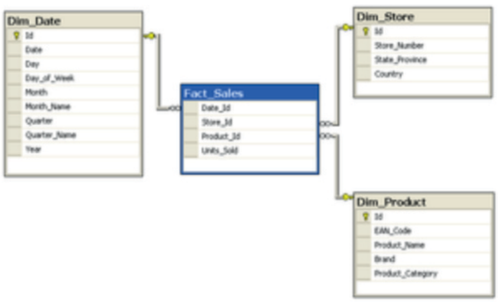


Figure 1. Start Schema Relational Diagram

3. Developed Relational Star Query :

Star schema used by example query.

```
SELECT
      P.Brand,
      S.Country AS Countries,
      SUM(F.Units_Sold)
FROM Fact_Sales F
INNER JOIN Dim_Date D ON
(F.Date_Id = D.Id)
INNER JOIN Dim_Store S ON
(F.Store_Id = S.Id)
INNER JOIN Dim_Product P ON
(F.Product_Id = P.Id) WHERE D.Year =
1997 AND P.Product_Category = 'tv'
 GROUP BY P.Brand, S.Country
```

4. Data Warehouses, Schemas and Decision Support Basics

Storing historical data is important for improving performance monitoring, conducting special decision relevant studies, improving financial reporting and providing drill down. However, a data warehouse is not a complete decision support system (DSS). Rather, a data warehouse or data mart is commonly the "driver" and dominant component for a data driven decision support system[11].The DSS adds a manager-friendly front end commonly built with a business intelligence (BI) product. A data-driven DSS emphasizes access to and manipulation of a time series of internal company data and sometimes external data. Simple file systems accessed by query and retrieval tools provide the most elementary level of functionality. Data warehouse systems that allow the manipulation of data by computerized tools tailored to a specific task and setting or by more general tools and operators provide additional functionality. Data-driven DSS with online analytical processing (OLAP) provide the highest level of functionality.

A data warehouse is an organized collection of large amounts of structured data. It is a database designed and intended to support decision making in organizations. It is usually batch updated and structured for rapid online queries and managerial summaries of its contents. According to Bill Inmon (1993), who is often called the "father" of data warehousing, "a data warehouse is a subject-oriented, integrated, time-variant, nonvolatile collection of data". Ralph Kimball (1996), another data warehousing pioneer, notes, "A data warehouse is a copy of transaction data specifically structured for query and analysis." So, the data warehouse or the single subject data mart stores data for a data-driven DSS. When a data warehouse is included as a component in a data-driven DSS, a DSS analyst or data modeler needs to develop a schema or structure for the database and identify analytic software and end-user presentation software to complete the DSS architecture and design. The DSS components need to be linked in an architecture that provides appropriate performance and scalability. In some data-driven DSS designs, a second multidimensional database management system (MDBMS) will be included and populated by a data warehouse built using a relational database management system (RDBMS). The MDBMS will provide data for online analytical processing (OLAP). It is common to build a data warehouse using an RDBMS from Oracle or IBM and then use query and reporting and analytical software from a vendor such as Micro Strategy or Business Objects as part of the overall data-driven DSS design. What some vendors call "business intelligence software" provides the analytics and user interface functionality for a data-driven DSS built with a data warehouse component.

In a data warehouse built using an

RDBMS, the most common data model is called the star schema. A related model is called a snowflake schema. A star schema is organized around a central fact table that is joined to some dimension tables using foreign key references. The fact table contains data like price, discount values, number of units sold, and dollar value of sales. The fact table usually has some summarized and aggregated data, and it is usually very large in terms of both fields and records. The basic premise of a star schema is that information can be classified into two groups: facts and dimensions. Facts are the core data elements one is analyzing. For example, units of individual items sold are facts, while dimensions are attributes about the facts. The star schema has also been called a star-join schema, data cube, data list, grid file, and multidimensional schema. The name star schema comes from the pattern formed by the entities and relationships when they are represented as an entity-relationship diagram. Metaphorically, the results of a specific business activity are at the center of the star schema database and are surrounded by dimensional tables with data on the people, places, and things that come together to perform the business activity. These dimensional tables are the points of the star. How does a snowflake schema differ from a star schema? A snowflake schema is an expansion and extension of a star schema to additional secondary dimensional tables. In a star schema, each dimension is typically stored in one table; the snowflake design principle expands a dimension and creates tables for each level of a dimensional hierarchy. For example, a Region dimension may contain the levels of Street, City, State and Country. In a star schema, all these attributes would be stored in one table; in a snowflake schema, one would expand the schema and a designer might add city and state secondary tables. Creating the data model for a data-driven DSS is a complex task. Whether DSS data is stored in a flat file, a hierarchical or multidimensional database or a relational database management system, a large, well-organized database of business facts provides the functionality for a data-driven DSS. A data warehouse is only part of such a system; but when it is used, the data component is the "driver" for decision support.

5. Roll of Data Ware House Design

Data warehouse design is a lengthy, time-consuming, and costly process. Any wrongly calculated step can lead to a failure. Therefore, researchers have placed important efforts to the study of design and development related issues and methodologies. Data modeling for a data warehouse is different from operational database data modeling. An operational system, e.g., online transaction processing (OLTP), is a system that is used to run a business in real time, based on current data. An OLTP system usually adopts Entity-relationship (ER) modeling and application-oriented database design (Han & Kamber, 2006). An information system, like a data warehouse, is designed to support decision making based on historical point-in-time and prediction data for complex queries or data mining applications (Hoffer, et al., 2007). A data warehouse schema is viewed as a dimensional model (Ahmad et al., 2004, Han & Kamber, 2006; Levene & Loizou, 2003). It typically adopts either a star or snowflake schema and a subject-oriented database design (Han & Kamber, 2006). The schema design is the most critical to the design of a data warehouse. Many approaches and methodologies have been proposed in the design and development of data warehouses. Two major data warehouse design methodologies have been paid more attention. Inmon et al. (2000) proposed the Corporate Information Factory (CIF) architecture. This architecture, in the design of the atomic-level data marts, uses denormalized entity-relationship diagram (ERD) schema. Kimball (1996, 1997) proposed multidimensional (MD) architecture.

This architecture uses star schema at atomic-level data marts. Which architecture should an enterprise follow? Is one better than the other? Currently, the most popular data model for data warehouse design is the dimensional model (Han & Kamber, 2006; Bellatreche, 2006). Some researchers call this model the data-driven design model. Artz (2006), nevertheless, advocates the metric-driven model, which, as another view of data warehouse design, begins by identifying key business processes that need to be measured and tracked over time in order for the organization to function more efficiently. There has always been the issue of top-down vs. bottom-up approaches in the design of information systems. The same is with a data warehouse design. These have been puzzling questions for business intelligent architects and data warehouse designers and developers. The next section will extend the discussion on issues related to data warehouse design and development methodologies.

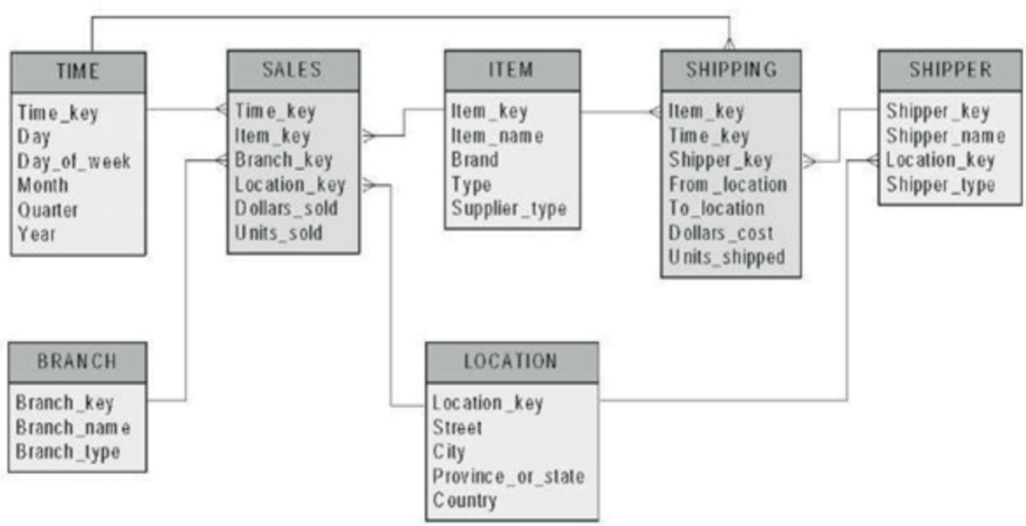
6. Design and Development Methodologies

Data Warehouse Data Modeling

Database design is typically divided into a four-stage process (Raisinghani, 2000). After requirements are collected, conceptual design, logical design, and physical design follow. Of the four stages, logical design is the key focal point of the database design process and most critical to the design of a database. In terms of an OLTP system design, it usually adopts an ER data model and an application-oriented database design (Han & Kamber, 2006). The majority of modern enterprise information systems are built using the ER model (Raisinghani, 2000). The ER data model is commonly used in relational database design, where a database schema consists of a set of entities and the relationship between them. The ER model is used to demonstrate detailed relationships between the data elements. It focuses on removing redundancy of data elements in the database.

The schema is a database design

containing the logic and showing relationships between the data organized in different relations (Ahmad et al., 2004). Conversely, a m data warehouse requires a concise, subject-oriented schema that facilitates online data analysis. A data warehouse schema is viewed as a dimensional model which is composed of a central fact table and a set of surrounding dimension tables, each corresponding to one of the components or dimensions ofthe fact table (Levene & Loizou, 2003). Dimensional models are oriented toward a specific business process or subject. This approach keeps the data elements associated with the business process only one join away. The most popular data model for a data warehouse is multidimensional model. Such a model can exist in the form of a star schema, a snowflake schema, or a starflake schema. The star schema is the simplest database structure containing a fact table in the center, no redundancy, which is surrounded by a set of smaller dimension tables (Ahmad et al., 2004; Han & Kamber, 2006).



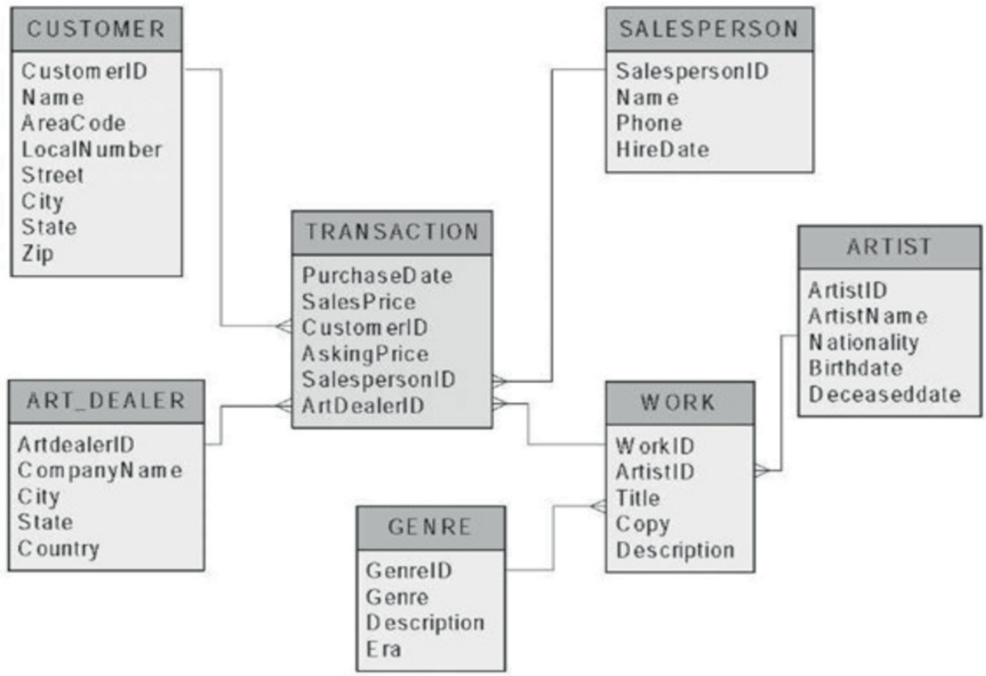


Figure 1. Example of a Star Schema

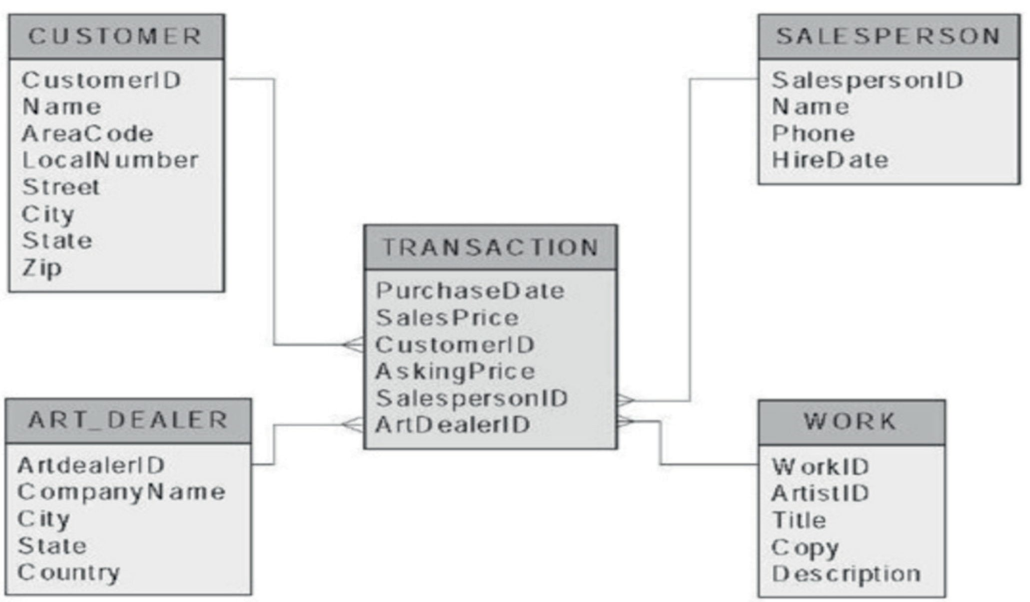


Figure 2. Example of a snowflake schema

Two major design methodologies have been paid more attention in the design and development of data warehouses. Kimball (1996, 1997) proposed multidimensional (MD) architecture. Inmon, Galemmco, and Geiger (2000) proposed the Corporate Information Factory (CIF) architecture. Imhoff et al. (2004) made a comparison between the two by using important criteria, such as scope, perspective, data flow, etc.

One of the most significant differences between the CIF and MD architectures is the definition of data mart. For MD architecture, the design of the atomic-level data marts is significantly different from the design of the CIF data warehouse, while its aggregated data mart schema is approximately the same as the data mart in the CIF architecture. MD architecture uses star schemas, whereas CIF architecture uses denormalized ERD schema. This data modeling difference constitutes the main design difference in the two architectures (Imhoff et al., 2004). A data warehouse may need both types of data marts in the data warehouse bus architecture depending on the business requirements. Unlike the CIF architecture, there is no physical repository equivalent to the data warehouse in the MD architecture.

The warehouse with business-unit needs for operational systems. It starts with experiments and prototypes (Han & Kamber, 2006). With bottom-up, departmental data marts are built first one by one. It offers faster and easier implementation, favorable return on investment, and less risk of failure, but with a drawback of data fragmentation and redundancy. The focus of bottom-up approach is to meet unit-specific needs with

7. Conclusion

The overall conclusion based on the database technique of star schema several data warehousing development and design methodologies have been reviewed and discussed. Data warehouse data model differentiates itself from ER model with an orientation toward specific business purposes. It benefits an enterprise greater if the CIF and MD architectures are both considered in the design of a data warehouse. Some of the methodologies have been practiced in the real world and accepted by today's businesses. Yet new challenging methodologies, particularly in data modeling and models for physical data warehousing design, such as the metric-driven methodology, need to be further researched and developed.

References:

[1] Inmon, W. H. "What is a Data Warehouse?" PRISM Newsletter, Center for the Application of Information Technology, Washington University in St. Louis, vol. 1, no. 1, 1993.

[2] Kimball, R. The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, 1996.

[3] Power, D. Is a Data Warehouse a DSS? What is a Star Schema? How Does a Snowflake Schema Differ from a Star Schema? DSS News, Vol. 3, No. 4, February 17, 2002."DWH Schemas". 2009.

[4] C J Date, "An Introduction to Database Systems (Eighth Edition)", p. 708

[5] Ralph Kimball and Margy Ross, The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition), p. 393.

[6] http://docs.oracle.com /html/E10312_01/dm_concepts .htm Data Mart Concepts

[7] http://datawarehouse4u .info/OLTP-vs-OLAP.html OLTP vs OLAP

[8] http://olap.com/category/ bisolutions/predictive-analytics Predictive Analysis

[9] Patil, Preeti S.; Srikantha Rao; Suryakant B. Patil (2011). "Optimization of Data Warehousing System: Simplification in Reporting and Analysis". IJCA Proceedings on International Conference and workshop on Emerging Trends in Technology (ICWET) (Foundation of Computer

Science) 9 (6): 33–37.
[10] Rainer, R. Kelly (2012-05-01). Introduction to Information Systems: Enabling and Transforming Business, 4th Edition (Kindle Edition). Wiley. pp. 127, 128, 130, 131, 133.
[11] "The Story So Far". 2002-04 Retrieved 2008-09-21.Kimball 2002, pg. 16
[12] Jones, Martyn Richard". martynjones.eu. Retrieved 2014-10-01.
[13] An architecture for a business and information system". IBM Systems Journal.
[14] Bill (1992). Building the Data Warehouse. Wiley. ISBN 0-471-56960-7.
[15] Kimball, Ralph (1996). The Data Warehouse Toolkit. Wiley. ISBN 0-471-15337-0.

A Case Study and Survey of Cloud Computing Technology

Md Shahid Anwar
M.tech (Research Scholar)
Al-Falah School of Engineering and Technology, Dhouj, Faridabad.

Ashif Ali
(Assistant Professor, CSE Deptt.)
Al-Falah School of Engineering and Technology, Dhouj, Faridabad.

Abstract

The research paper focused on case study of Cloud computing technology relies on sharing of resources to achieve coherence and economies of scale, similar to a utility over a network. At the foundation of cloud computing is the broader concept of converged infrastructure and shared services. Cloud computing, or in simpler shorthand just "the cloud", also focuses on maximizing the effectiveness of the shared resources. Cloud resources are usually not only shared by multiple users but are also dynamically reallocated per demand. This can work for allocating resources to users.

Keyword: Cloud Computing, CAPEX, Open Nebula.

1. INTRODUCTION

Cloud computing, or in simpler shorthand just "the cloud", also focuses on maximizing the effectiveness of the shared resources. Cloud resources are usually not only shared by multiple users but are also dynamically reallocated per demand. This can work for allocating resources to users. For example, a cloud computer facility that serves European users during European business hours with a specific application (e.g., email) may reallocate the same resources to serve North American users during North America's business hours with a different application (e.g., a web server). es for different applications. This approach should maximize the use of computing power thus reducing environmental damage as well since less power, air conditioning, rack space, etc. are required for a variety of functions. With cloud computing, multiple users can access a single server to retrieve and update their data without purchasing licenses for different applications. The term "moving to cloud" also refers to an organization moving away from a traditional CAPEXmodel (buy the dedicated hardware and depreciate it over a period of time) to the OPEX model (use a shared cloud infrastructure and pay as one uses it). Proponents claim that cloud computing allows companies to avoid upfront infrastructure costs, and focus on projects that differentiate their businesses instead of on infrastructure.[4] Proponents also claim that cloud computing allows enterprises to get their applications up and running faster, with improved manageability and less maintenance, and enables IT to more rapidly adjust resources to meet fluctuating and unpredictable business demand.[4][5][6] Cloud providers typically use a "pay as you go" model. This can lead to unexpectedly high charges if administrators do not adapt to the cloud pricing model.[7]The present availability of high-capacity networks, low-cost computers and storage devices as well as the widespread adoption of hardware virtualization, service-oriented architecture, and autonomic and utility computing have led to a growth in cloud computing. [8][9][10] Companies can scale up as computing needs increase and then scale down again as demands decrease. Cloud vendors are experiencing growth rates of 50% per annum.[11].

2. History of Cloud Computing

Origin of the term-The origin of the term cloud computing is unclear. The expression cloud is commonly used in science to describe a large agglomeration of objects that visually appear from a distance as a cloud and describes any set of things whose details are not inspected further in a given context.[12] Another explanation is that the old programs to draw network schematics surrounded the icons for servers with a circle, and a cluster of servers in a network diagram had several overlapping circles, which resembled a cloud.

In analogy to above usage the word cloud was used as a metaphor for the Internet and a standardized cloud-like shape was used to denote a network on telephony schematics and later to depict the Internet in computer network diagrams. With this simplification, the implication is that the specifics of how the end points of a network are connected are not relevant for the purposes of understanding the diagram. The cloud symbol was used to represent the Internet as early as 1994,[9][10] in which servers were then shown connected to, but external to, the cloud. References to cloud computing in its modern sense appeared as early as 1996, with the earliest known mention in a Compaq internal document.[11]The popularization of the term can be traced to 2006 when Amazon.com introduced the Elastic Compute Cloud. The 1950s-The underlying concept of cloud computing dates to the 1950s, when large-scale mainframe computers were seen as the future of computing, and became available in academia and corporations, accessible via thin clients/terminal computers, often referred to as "dumb terminals", because they were used for communications but had no internal processing capacities. To make more efficient use of costly mainframes, a practice evolved that allowed multiple users to share both the physical access to the computer from multiple terminals as well as the CPU time. This eliminated periods of inactivity on the mainframe and allowed for a greater return on the investment. The practice of sharing CPU time on a mainframe became known in the industry as time-sharing.During the mid 70s, time-sharing was popularly known as RJE (Remote Job Entry); this nomenclature was mostly associated with large vendors such as IBM and DEC. IBM developed the VM Operating System to provide time-sharing services .The 1990s-In the 1990s, telecommunications companies, who previously offered primarily dedicated point-to-point data circuits, began offering virtual private network (VPN) services with comparable quality of service, but at a lower cost. By switching traffic as they saw fit to balance server use, they could use overall network bandwidth more effectively. They began to use the cloud symbol to denote the demarcation point between what the providers was responsible for and what users were responsible for. Cloud computing extends this boundary to cover all servers as well as the network infrastructure. As computers became more prevalent, scientists and technologists explored ways to make large-scale computing power available to more users through time-sharing. They experimented with algorithms to optimize the infrastructure, platform, and applications to prioritize CPUs and increase efficiency for end users.[2] Since 2000 cloud computing has come into existence. In early 2008, Open Nebula, enhanced in the RESERVOIREuropean Commission-funded project, became the first open-source software for deploying private and hybrid clouds, and for the federation of clouds.[1] In the same year, efforts were focused on providing quality of service guarantees (as required by real-time interactive applications) to cloud-based infrastructures, in the framework of the IRMOS European Commission-funded project, resulting in a real-time cloud environment.[6] By mid-2008, Gartner saw an opportunity for cloud computing "to shape the relationship among consumers of IT services, those who use IT services and those who sell them"[5] and observed that "organizations are switching from company-owned hardware and software assets to per-use service-based models" so that the "projected shift to computing ... will result in dramatic growth in IT products in some areas and significant reductions in other areas."[4]. Microsoft Azure became available in late 2008.In July 2010, Rackspace Hosting and NASA jointly launched an open-source cloud-software initiative known as Open Stack. The Open Stack project intended to help organizations offer cloud-computing services running on standard hardware. The early code came from NASA's Nebula platform as well as from Rackspace's Cloud Files platform.[2] .On March 1, 2011, IBM announced the IBM Smart Cloud framework to support Smarter Planet.[3] Among the various components of the Smarter Computing foundation, cloud computing is a critical piece. On June 7, 2012, Oracle announced the Oracle Cloud.[27] While aspects of the Oracle Cloud are still in development, this cloud offering is posed to be the first to provide users with access to an integrated set of IT solutions, including the Applications (SaaS), Platform (PaaS), and Infrastructure (IaaS) layers.[8][7][5] Similar concepts-Cloud computing is the result of evolution and adoption of existing technologies and paradigms. The goal of cloud computing is to allow users to take benefit from all of these technologies, without the need for deep knowledge about or expertise with each one of them. The cloud aims to cut costs, and helps the users focus on their core business instead of being impeded by IT obstacles.[7] The main enabling technology for cloud computing is virtualization. software separates a physical computing device into one or more "virtual" devices, each of which can be easily used and managed to perform computing tasks. With operating system level virtualization essentially creating a scalable system of multiple independent computing devices, idle computing resources can be allocated and used more efficiently. Virtualization provides the agility required to speed up IT operations, and reduces cost by increasing infrastructure utilization. Autonomic computing automates the process through which the user can provision resources on-demand. By minimizing user involvement, automation speeds up the process, reduces labor costs and reduces the possibility of human errors.[3] Users routinely face difficult business problems. Cloud computing adopts concepts from Service-oriented Architecture (SOA) that can help the user break these problems into services that can be integrated to provide a solution. Cloud computing provides all of its resources as services, and makes use of the well-established standards and best practices gained in the domain of SOA to allow global and easy access to cloud services in a standardized way. Cloud computing also leverages concepts from utility computing to provide metrics for the services used. Such metrics are at the core of the public cloud pay-per-use models. In addition, measured services are an essential part of the feedback loops in autonomic computing, allowing services to scale on-demand and to perform automatic failure recovery. Cloud computing is a kind of grid computing; it has evolved by addressing the QoS (quality of service) and reliability problems. Cloud computing provides the tools and technologies to build data/compute intensive parallel applications with much more affordable prices compared to traditional parallel computing techniques.[31]

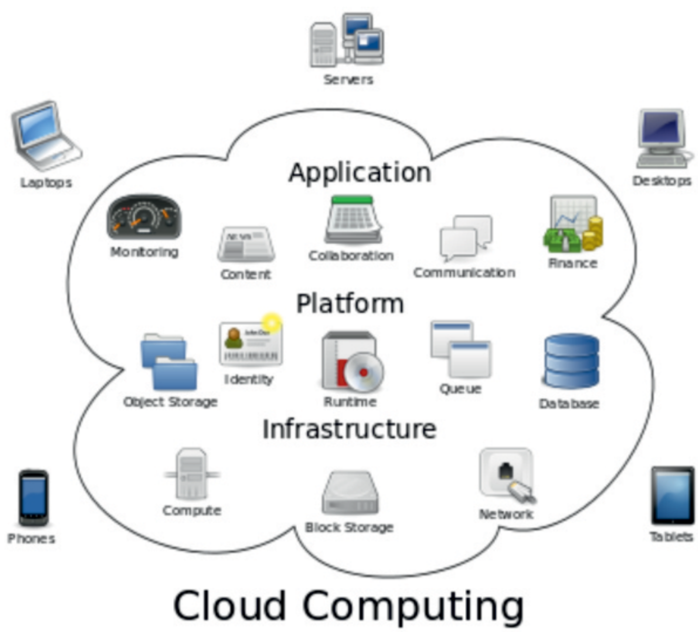


Figure 1. Cloud Computing Model

3. Characteristics of Cloud Computing

1.Client–server model Client–server computing refers broadly to any distributed application that distinguishes between service providers (servers) and service requestors (clients).[5] 2.Grid computing A form of distributed and parallel computing, whereby a 'super and virtual computer' is composed of a cluster of networked, loosely coupled computers acting in concert to perform very large tasks."3. Mainframe computer Powerful computers used mainly by large organizations for critical applications, typically bulk data processing such as: census; industry and consumer statistics; police and secret intelligence services; enterprise resource planning; and financial transaction processing.4.Utility computing The "packaging of computing resources, such as computation and storage, as a metered service similar to a traditional public utility, such as electricity."[7][8] 5.Peer-to-peer A distributed architecture without the need for central coordination. Participants are both suppliers and consumers of resources.

4. Security In Cloud Computing Security can improve due to centralization of data, increased security-focused resources, etc., but concerns can persist about loss of control over certain sensitive data, and the lack of security for stored kernels. Security is often as good as or better than other traditional systems, in part because providers are able to devote resources to solving security

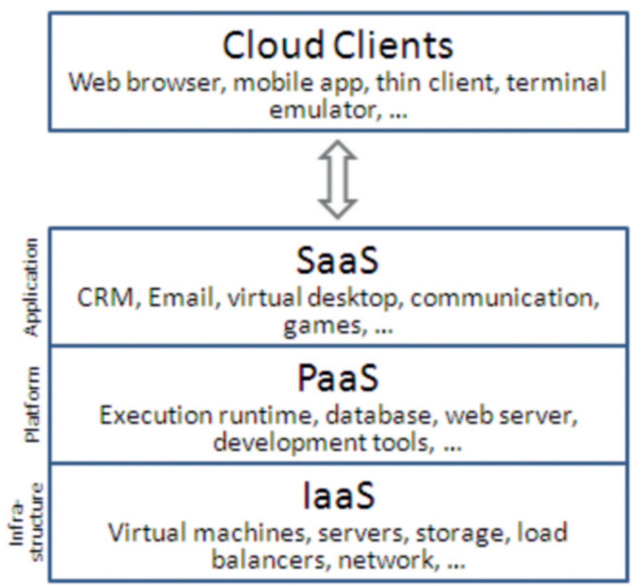


Figure 1.2 Service Models

issues that many customers cannot afford to tackle.[5] monitored, controlled, and reported, providing transparency for both the and smartphones. Some of these devices cloud clients – rely on cloud computing for all or a majority of their applications so as to be essentially useless without it. Examples are thin clients and the browser-based Chrome book.

6. Types of Cloud Computing

Private cloud-Private cloud is cloud infrastructure operated solely for a single organization, whether managed internally or by a third-party, and hosted either internally or externally.[3] Undertaking a private cloud project requires a significant level and degree of engagement to virtualize the business environment, and requires the organization to reevaluate decisions about existing resources. When done right, it can improve business, but every step in the project raises security issues that must be addressed to prevent serious vulnerabilities.[6] Self-run data centers [5] are generally capital intensive. They have a significant physical footprint, requiring allocations of space, hardware, and environmental controls. These assets have to be refreshed periodically, resulting in additional capital expenditures. They have attracted criticism because users "still have to buy, build, and manage them" and thus do not benefit from less hands-on management,[5] essentially "[lacking] the economic model that makes cloud computing such an intriguing concept".[9] Public cloud-A cloud is called a "public cloud" when the services are rendered over a network that is open for public use. Public cloud services may be free.[10] Technically there may be little or no difference between public and private cloud architecture, however, security consideration may be substantially different for services (applications, storage, and other resources) that are made available by a service provider for a public audience and when communication is effected over a non-trusted network. Saasu is a large public cloud. Generally, public cloud service providers like Amazon AWS, Microsoft and Google own and operate the infrastructure at their data center and access is generally via the Internet. AWS and Microsoft also

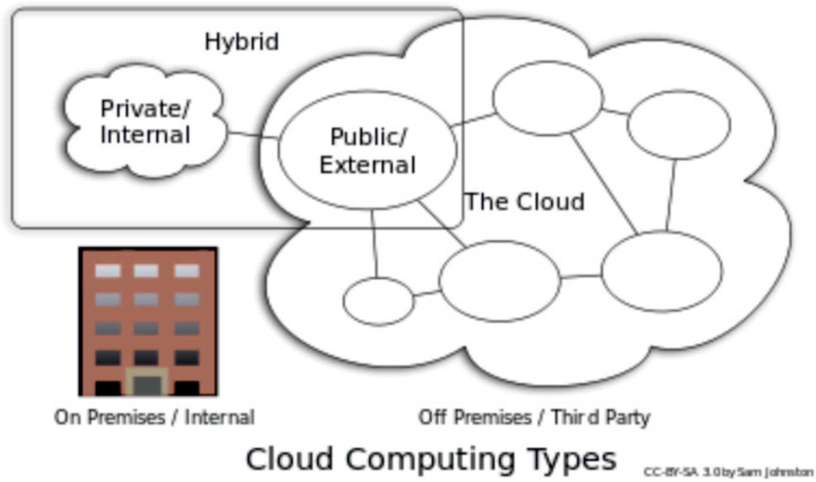


Figure 1.4 Deployment Models

offer direct connect services called "AWS Direct Connect" and "Azure Express Route" respectively, such connections require customers to purchase or lease a private connection to a peering point offered by the cloud provider.[3] bursting enables data centers to create an in-house IT infrastructure that supports average workloads, and use cloud resources from public or private clouds, during spikes in processing demands.[7]computing infrastructure is built using volunteered resources. Many challenges arise from this type of infrastructure, because of the volatility of the resources used to built it and the dynamic environment it operates in. It can also be called peer-to-peer clouds, or ad-hoc clouds. An interesting effort in such direction is Cloud@Home, it aims to implement a cloud computing infrastructure using volunteered resources providing a business-model to incentivize contributions through financial restitution [7].

7. Conclusion

The whole conclusion of this paper depend upon the survey and study of Cloud computing, or in simpler shorthand just "the cloud", also focuses on maximizing the effectiveness of the shared resources.

Cloud resources are usually not only shared by multiple users but are also dynamically reallocated per demand. This can work for allocating resources to users. For example, a cloud computer facility that serves European users during European business hours with a specific application (e.g., email) may reallocate the same resources to serve North American users during North America's business hours with a different application (e.g., a web server). es for different applications.

References

[1] Hassan, Qusay (2011). "Demystifying Cloud Computing" (PDF). The Journal of Defense Software Engineering (CrossTalk) 2011 (Jan/Feb): 16–21. Retrieved 11 December2014.

[2] Know Why Cloud Computing Technology is the New Revolution". By Fonebell. Retrieved 8 January 2015.

[3] The NIST Definition of Cloud Computing" (PDF). National Institute of Standards and Technology. Retrieved 24 July 2011.

[4] What is Cloud Computing?". Amazon Web Services. 2013-03-19. Retrieved2013-03-20.

[5] Baburajan, Rajani, "The Rising Cloud Storage Market Opportunity Strengthens Vendors," infoTECH, August 24, 2011". It.tmcnet.com. 2011-08-24. Retrieved2011-12-02.

[6] Oestreich, Ken, (2010-11-15). "ConvergedInfrastructure".CTOForum.Thectoforum.com. Retrieved 2011-12-02.

[7]. Where's The Rub: Cloud Computing's Hidden Costs". 2014-02-27. Retrieved2014-07-14.

[8] Cloud Computing: Clash of the clouds". The Economist. 2009-10-15. Retrieved2009-11-03.

[9] Gartner Says Cloud Computing Will Be As Influential As E-business". Gartner. Retrieved 2010-08-22.

[10]Gruman, Galen (2008-04-07). "What cloud computing really means". InfoWorld. Retrieved 2009-06-02.

[11]The economy is flat so why are financials Cloud vendors growing at more than 90 percent per annum?". FSN. March 5, 2013.

[12]Hongji Yang, Xiaodong (2012).

"9". Software reuse in the emerging cloud computing era. Hershey, PA: Information Science Reference. pp. 204–227.ISBN9781466608979 . Retrieved 11 December 2014.

[13]Rosenberg, Jonathan (2014). How Google Works. Grand Central Publishing. p. 11. ISBN 978-1-4555-6059-2.

[14]A network 70 is shown schematically as a cloud", US Patent 5,485,455, column 17, line 22, filed Jan 28, 1994

Various Aspect of Cloud Computing

Vineeta Jain
(Research Scholar)

Abstract

Cloud computing is the latest technology to store and retrieve information with more facilities for the information seekers. This technology is cost effective, flexible, user centric, transparent and having many feasible characters. It is a collaborate work with fine fitting for any field of the universe. The information or data incorporated with this technology could be retrieved from anywhere with its own sophisticated management. This paper will discuss this novel technology in the field of Library and Information Science.

Key words: *Cloud Computing, Information Storage, Information retrieval, novel technology, Library and Information Science.*

Introduction

A revolution is cleared as a change in the way people think and behave. By this way, cloud computing is definitely a revolution and it creating a fundamental change in computer software, architecture and tools development, in the way we store the information and retrieve. It is a different way to retrieve digital resources. The purpose of this article is to aid the library professionals in assimilating the reality of the revolution, so we can use this revolutionary invention in the computer field which can be useful in any field of Library and Information Science. Libraries can keep more and more content into the load. By using this novel technology the user would be able to browse a physical shelf of books, CDs or DVDs or choose to take out an item or scan a bar code into his/her mobile device. Historical and rare documents would be scanned into a comprehensive, easily searchable database and would be accessible to any researcher from anywhere with the networked device.

Description of Cloud Technology:

Cloud computing is associated with a paradigm for the provision of computing infrastructure. This paradigm shifts the location of this infrastructure to the network to reduce the costs associated with the management of hardware and software resources. The cloud is drawing the attention from the Information and Communication Technology community.

A definition for cloud computing can be given as an emerging computer paradigm where data and services reside in massively scalable data centres in the cloud and can be accessed from any connected devices over the internet. Cloud computing is a way of providing various services on virtual machines allocated on top of a large physical machine pool which resides in the cloud. Cloud computing comes into focus only when we think about what IT has always wanted a path to increase capacity.

The cloud computing model is comprised of a front end and a back end. These two elements are connected through a network, in most cases the Internet. The front end is the vehicle by which the user interacts with the system; the back end is the cloud itself. The front end is composed of a client computer or the computer network of an enterprise and the applications used to access the cloud. The back end provides the applications, computers, servers and data storage that creates the cloud of services.

Characteristics of Cloud Computing:

Self Healing:

Cloud computing has some possessions of self healing environment for any application or any service running on it. Without any disruption, it has a hot backup of application readily, to protect from application failure. There are multiple copies of the same application and each copy updating itself so that at the time of failure at least one copy of application might be taken over.

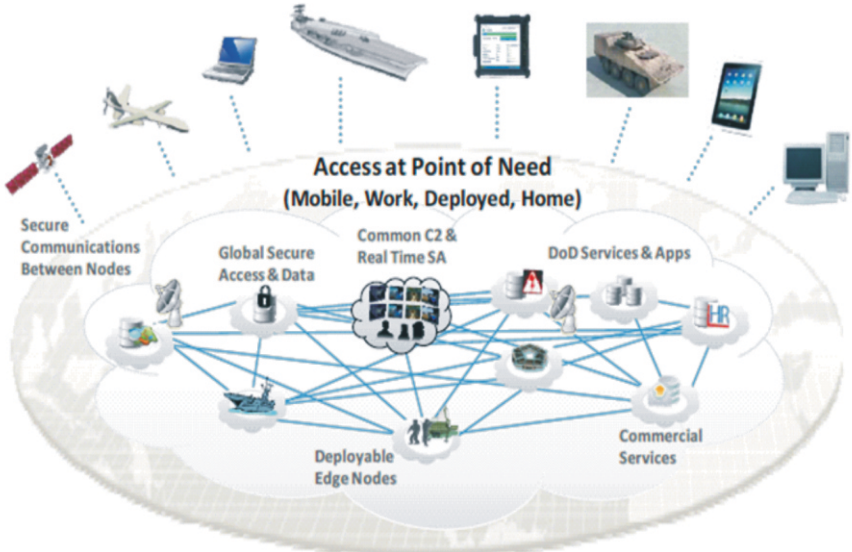


Fig -1

Multi Tenancy:

Any application supports multi tendency that is more tenants at the same time is one the character of cloud computing. Several systems are allowed the users to share the infrastructure allotted to them without any trouble to anyone who are all working under this. This facility is done by virtualizing the servers on the available machine pool and then allotting the servers to multiple users.

Linearly Scalable:

Cloud computing services are linearly scalable. The system is able to break down the workloads into pieces and service it across the infrastructure. An exact idea of linear scalability can be obtained from the fact that if can be obtained from the fact that if one server is able to process 1000 transactions per second, then two servers can process 2000 transactions per second.

Service Oriented:

Cloud computing systems are all service oriented. The systems are such that they are created out of other distinct services. Many such distinct services which are independent of each other are combined together to form this service. This allows re-use of the different services that are available and that are being created. Using this services that were just created, other such services can be created.

SLA Driven:

Usually businesses have agreements on the amount of services. Scalability and availability issues cause clients to break these agreements. But cloud computing services are SLA driven such that when the system experiences peaks of load, it will automatically adjust itself so as to comply with the service level agreements. The services will create additional instances of the applications on more servers so that load can be easily managed.

Virtualized:

The applications Cloud computing are fully decoupled from the underlying hardware. The cloud computing environment is a fully virtualized environment.

Flexible:

The other feature of the cloud computing services is that they are flexible. They can be used to serve a large variety of work load types varying from small loads of a small consumer application to very heavy loads of a commercial application.

Advantages of Cloud Computing in Libraries:

- Cost saving
- Flexibility and innovation
- User centric
- Openness
- Transparency
- Interoperability
- Representation
- Availability anytime anywhere
- Connect and Converse
- Create and collaborate

Some Examples of Cloud Computing:

1. OCLC
2. Library of Congress
3. Exlibris
4. Polaris
5. Scribd
6. Discovery Service
7. Google Docs / Scholar
8. World cat

Application of Cloud Computing in Library and Information Science

Cloud computing is offers many interesting potentials for libraries which may reduce the cost of technology as well as increase capacity reliability and performance for some type of automation activities. Cloud computing has made strong inroads into other commercial sectors and is now beginning to find more application in library science. The cloud computing pushes hardware to more abstract levels. Most of us are acquainted with fast computing power being delivered from systems that we can see and touch.

Cloud Computing in Digital Libraries

Digital Library is the most important academic and scientific research base, charges for providing information services for its users. In the past, most libraries insisted that their services are based on their own library resources. So librarians scarcely considered users' demands. But today, digital libraries have changed these views. And librarians usually need to collect as more information as they can accord users' requirements. Then they will analyze the information and sort out them.

Conclusion

In this revolutionary new era, cloud computing can provide organizations with the means and methods needed to ensure financial stability and high quality service. Of course there must be global cooperation if the cloud computing process is to attain optimal security and general operational standards. With the advent of the Could Computing the digital library, services of libraries will have a new leap in the near future. This is imperative for us all to be ready for the revolution.

References

1. Grace walker, Developer worker, Cloud Computing Fundamentals: A different way to deliver computer resource, ibm.com/developerworkers/December 2010, p.1-9
2. Nandkishor Gosavi, Seetal S.Shinde & Bhagyashree Dhakulkar, Use of cloud computing in library and

information science field, International Journal of Digital Library Services, Vol.2, No.3, p.51-59

3. Yang Mingfang, Yuan Xilin (2009), Digital Libraries under the Cloud Computing Environment Journal of Library Development, p.9

4. . en.wikipedia.org/wiki/could computing.

Cloud Computing: A Study of Flexibility in Budgeting & Emerging Trends in it Sourcing Strategy

Hitesh Keserwani

Abstract

In the evolution of computing technology, information processing has moved from mainframes to personal computers to server-centric computing to the Web. Today, many organizations are seriously considering adopting cloud computing, the next major milestone in technology and business collaboration. Cloud computing is hosted service from managed service provider's, which enables industry & enterprise businesses to focus on their key objectives, it relieves them of forming information technology strategy, and takes care to fulfill their infrastructure, platform and application requirements, delivering quality service level agreements to their employees for better efficiency & productivity. By migrating to the cloud, business organizations will be able to free up IT spends for reinvestment in mission-enabling activities or attaining organizational objectives, such as reduction in scarcity of resources. With more agile systems and faster deployment times, they will be better at supporting key business operations and providing services to customers. However apart from the benefits there are also challenges that must be tackled to achieve the real objective of implementing cloud computing. An investment today in the tools, capabilities, and processes required to overcome the obstacles to cloud migration is likely to produce a significant return in the long term. Organizations that follow a formal process to develop a business-driven sourcing strategy are more likely to achieve expected cost, performance and business outcomes, and take advantage of opportunities for innovation and low-cost IT from *outsourcing, according to Gartner, Inc. The sourcing strategy process is complex and repetitive, involving multiple steps, each with deliverables. As a result, effective execution of the sourcing strategy phase requires investment in processes, people and tools. The paper highlights the key deliverables that suggests the direction to minimize various risks and cost associated in the sourcing cycle and subsequently suggests the ways to select appropriate providers, and negotiate for a sound contract and effective deal. It also suggests the various strategies to overcome the issues like mobility, total cost of ownership (TCO), payback and ROI by discussing various case studies.*

Keywords: *Information technology, TCO, Computing, ROI, IT Strategies*

Introduction

In today's economic climate, while organizations are dealing with dynamic markets and ever-imposing regulations, CIOs are dealing with increased expectations from businesses and diminishing IT budgets. Business heads are under constant pressure to implement cost-efficient strategies that enhance business performance i.e. to do more with less. The adoption of the Cloud is expected to open up new investment and business opportunities currently obstructed by a surmounting need for huge up-front IT investments. An easy, fast and economical access to computing solutions, made possible through the use of a virtual technology environment, is attracting the business world to the Cloud. The Cloud promises to be a computing services model that is not restricted to a particular industry. This is why the Cloud is a compelling proposition even for those industry verticals that have not traditionally been at the forefront of IT adoption such as the Government, Healthcare, and Education industries. The cloud is a new business model for Facebook and Twitter too, as well as IT powerhouses such as IBM and Microsoft. Instead of buying hardware and developing applications, companies can now access software and infrastructure-based cloud services on a pay-as-you-go basis from Google, Sales force and Canadian providers such as TELUS and Bell. What makes cloud services unique is that it introduces best-in-class technologies and processes that, in the past, have been cost prohibitive to an entire class of businesses. With the cloud, these investments are primarily designed for the benefit of all companies. The cloud not just lowers but eliminates barriers to entry for enterprise-class IT services. This has tangential business benefits for the companies to become more agile, cost-effective and ultimately more competitive over time.

Literature Review

Most organizations engage in outsourcing for economic or strategic reasons. Economically, outsourcing is attractive when the tasks being outsourced can be performed by the provider at a lower total cost. Strategically, outsourcing is attractive when organizations have capacity and/or capability constraints that prevent them from servicing a market. When a firm does not have personnel of requisite quantity and skill, or sufficient physical capacity to deliver its product or services within a required time frame, it either has to postpone the work, or outsource to get the work done within the required time frame and level of quality. Apart from economic and strategic reasons, many organizations outsource because of a "herd mentality." Organizations become concerned when others in the industry outsourcing their work, potentially benefiting economically and/or strategically, while they are not. They believe they may "miss the boat" and lose out on the expected benefits that competitors may come to realize. These three rationales are supported theoretically by transaction cost theory, resource based theory and institutional theory, respectively. The following sections examine these theoretical approaches to understand the growing trend toward outsourcing.

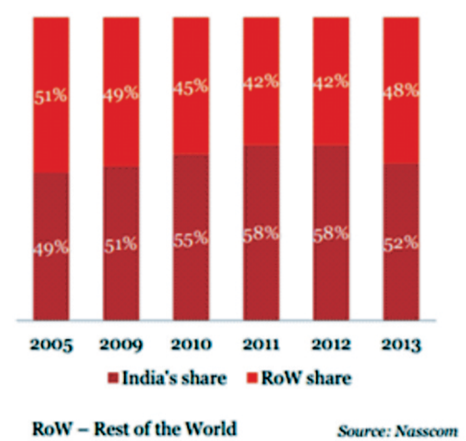
Transaction Cost Economics Theory

Cost reduction has been the primary rationale for outsourcing (DiRomualdo and Gurbaxani, 1998; Lacity and Willcocks, 1998; Takac, 1994). Transaction Cost Economics (TCE) (Coase,1937; Williamson, 1975; Williamson, 1979; Williamson, 1985) provides a theoretical foundation for addressing outsourcing from a cost perspective. Transactions are the exchanges of goods or services between firms. TCE maintains that the allocation of economic activity among firms depends on balancing each firm's internal costs against the cost of transacting for goods and services in the market (Alchian and Demsetz, 1972). This is the familiar make vs. buy argument which proposes that firms buy services from other firms (via "the market") if it is less costly than producing those services in-house (via "hierarchy"). Conversely, when the market "fails" then products and services must be produced internally; the reason why firms exist according to TCE. TCE addresses two types of costs: production and coordination (Alchian and Demsetz, 1972). Production costs represent the costs of actually producing the goods or services, and would be expected to differ among firms. Coordination costs are the costs of controlling and monitoring workers if the goods are produced internally or vendors if purchased in the market. These costs arise from the need to define, negotiate and enforce contracts, and to monitor and coordinate activities across organizational boundaries.

Resource-Based Theory

Filling the gaps in IT resources is a second major rationale for outsourcing (Lacity and Willcocks, 1998). This may be because of an increase in workload beyond an organization's current capacity (Radding, 1995), or a disparity between the existing and required IT resources (DiRomualdo and Gurbaxani, 1998, p.68). The Resource Based View of the firm (RBV) (Penrose, 1959) provides a theoretical foundation for addressing outsourcing from a resource gap perspective. Resource-based theory views a firm as a collection of productive resources (Penrose, 1959) and organizations compete based on having or controlling unique, valuable and hard-to-imitate resources (Barney, 1991). Rather than competing from a specific product/market position, a set of resources could be used to create various products for various markets. Advantage comes from being the only organization with the resources needed to create and deliver those products. Sustainability of the advantage depends on resource immobility, that is, the difficulty for others to copy, acquire, or develop those resources (Rumelt, 1984).

Table 1: India's share in global sourcing of services



Institutional Theory

The Eastman Kodak decision is regarded as a turning point in outsourcing's history. Loh and Venkatraman (1992b) examined the adoption of outsourcing before and after Eastman Kodak's decision to outsource and found that adoption of IT outsourcing was motivated more by imitative behavior, than by external influence amongst user organizations. Outsourcing is often an imitative response to the hype and publicity surrounding the subject - the so-called "bandwagon effect." (Hu, et al., 1997; Lacity and Hirschheim, 1993a; Lacity and Hirschheim, 1993b; Lacity and Willcocks, 1998). Institutional theory provides a theoretical foundation for explaining the imitative behavior regarding outsourcing (Ang and Cummings, 1997; Hu, et al., 1997). Thus, companies may make outsourcing decision based on other organizations that have already outsourced.

According to the report report "Forecast Analysis: Public Cloud Services, Worldwide, 2012-2018.The public cloud services market in India is on pace to grow 32.2 percent in 2014 to total $556.8 million, an increase from 2013 revenue of $421 million, according to Gartner, Inc. Spending on software as a service (SaaS) will total $220 million in 2014, growing 33.2 percent from last year. SaaS is the largest overall cloud market segment, followed by infrastructure as a service (IaaS), totaling $78 million and business process as a service (BPaaS), totaling $75 million.

“Growth of cloud services in India reflects the demand for new sourcing models,” said Ed Anderson, research vice president at Gartner. “We expect high growth rates across all cloud services market segments in India.”

BPaaS is expected to grow from $62.3 million in 2013 to $204 million in 2018. SaaS is expected to grow from $166 million in 2013 to $636 million in 2018, and IaaS is forecast to grow from $58 million in 2013 to $317 million in 2018. The total cloud market in India in 2013 was $421 million, and the total market for public cloud services in India is expected to reach $1.7 billion in 2018.

IT Sourcing Strategy: A case of Pharmaceutical Company

Executive Summary

Acting on an internal strategy of doubling its revenue in seven years, a multi-billion-dollar global pharmaceutical company began to examine how its IT service organization could best support anticipated product and services growth in emerging markets. With Everest Group's assistance, this organization designed an enterprise-wide IT sourcing strategy that established the economic, organizational, and governance frameworks for selecting, integrating, and managing a hybrid, multiple provider environment for both infrastructure and application services. The newly implemented structure enabled the organization to select and integrate global IT service providers in a more rapid and cost effective manner.

The Client's Challenge

There were several specific challenges involved as the pharmaceutical company embarked upon its new strategy. They included:

- Over coming limited experience in managed multiple provider environments
- Creating a flexible solution to meet demands on IT services while avoiding the up-front capital expenditures associated with creating new capabilities
- Addressing an uncertain

demand and delivery footprint (i.e., what IT services would be needed and where)

Creating a solution that would yield cost advantages over ongoing operational expenditures for current services

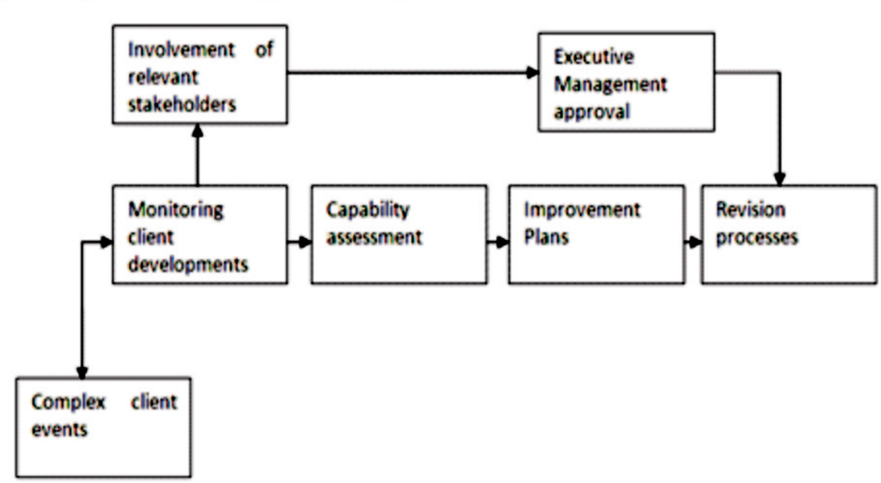
Insight to Action

Everest Group facilitated joint working sessions with the business unit and IT services executives that focused on identifying changes in IT services (e.g., performance, costs, delivery location, and regulatory requirements) required to support the global expansion strategy. With the requirements baseline in place, Everest Group worked with the IT organization to identify and assess multiple sourcing strategies and structures. Upon agreeing on a hybrid solution, Everest Group worked with the pharmaceutical company to develop operational guidelines for selecting and managing a multi-provider environment.

Impact

Everest Group enabled the pharmaceutical company to develop an effective strategy to accelerate its growth. It also helped the company avoid the pitfalls of the traditional method of splitting sourcing solutions based upon tower functions.

Figure: Proposed IT outsourcing model for organizations



Additionally, the engagement provided the pharmaceutical company with the evergreen tools and templates needed to manage the services and financials of a multi-provider solution. The final result was a sourcing solution framework that offered extensive flexibility as the company's sourcing needs evolved and changed.

References

1. Almeida, P., Song, J., and Grant, R.M. "Are firms superior to alliances and markets? An empirical test of cross-border knowledge transfer," *Organization Science* (13:2), 2002, pp. 147-163.
2. Altinkemer, K., Chaturvedi, A., and Gulati, R. "Information systems outsourcing: Issues and evidence," *International Journal of Information Management* (14:4), 1994, pp. 252-278.
3. Barney, J.B. "Firm resources and sustained competitive advantage," *Journal of Management* (17), 1991, pp. 99-120.
4. Bounfour, A. "Is outsourcing of intangibles a real source of competitive advantage?"
5. *International Journal of Applied Quality Management* (2:2), 1999, pp. 127-151.
6. Buhner, R., and Tuschke, A. "Outsourcing.," *Die Betriebswirtschaft* (57:1), 1997, pp. 20-30.
7. Carmel, E. *Global Software Teams, Collaboration Across Borders and Time Zones*, Prentice Hall, New Jersey, 1999.
8. Cavusgil, S.T., Calantone, R.J., and Zhao, Y. "Tacit knowledge transfer and firm innovation capability," *The Journal of Business and Industrial Marketing* (18:1), 2003, pp. 6-21.
9. Clark, T.D., Zmud, R.W., and McCray, G.E. "The outsourcing of information services: Transforming the nature of business in the information industry," *Journal of Information Technology* (10:4), 1995, pp. 221-237.
10. Coase, R. "Nature of the Firm,"

Economica (4), 1937, pp. 386-405.

11. Cohen, W., and Levinthal, D. "Absorptive capacity: A new perspective on learning and innovations," *Administrative Science Quarterly* (35:1), 1990, pp. 128-152.
12. Copeland, D.G., and McKenney, J.L. "Airline reservation systems: Lessons form history," *MIS Quarterly* (12:4), 1988, pp. 353-370.
13. Due, R.T. "The real costs of outsourcing," *Information Systems Management* (9:1), 1992, pp. 78-81

Instructions for Authors

Essentials for Publishing in this Journal

- Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.
- 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.
- All our articles are refereed through a double-blind process.
- 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.