Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CONSTRUCTION OF A NOVEL ONLINE DATA WAREHOUSE ARCHITECTURE USING STORAGE AREA NETWORK WITH FIBRE CHANNEL FABRIC

[1]M. SARAVANAMUTHU, [2]Dr. G. M. KADHAR NAWAZ

[1]Associate Professor, PG Department of Computer Applications, Sree Saraswathi Thyagaraja College, Pollachi, Tamilnadu, India.

[2]Director, Department of MCA, Sona College of Technology, Salem, Tamilnadu, India.

E-mail: [1]msaravanamuthu@gmail.com , [2]nawazse@yahoo.co.in

ABSTRACT

The major issues of data warehouse architecture are higher cost, lower performance and limited access. The reasons for these problems focus the architectural design. The solutions are proposed with the new modified data warehouse design say online data warehouse architecture with the recent technological support from cloud computing say storage area networks through the use of high speed fibre channel based storage devices with fibre channel switches. Finally the benefits and advantages of proposed design were discussed with the cost, performance and access analyses.

Keywords: *Data Warehouse Architecture, Cloud Computing, Storage Architecture, Storage Area Network, Network Storage, Fibre Channel Protocol, Fibre Channel Switch/Fabric.*

1. INTRODUCTION

The problems in existing data warehouse architectures were gradually solved stage by stage by the data warehouse designers. Mostly the issues in business intelligence (BI) applications running with data mining (DM) tools are still completely unsolved. First issue is the cost of the data warehouse system is heavy in the existing models due to the huge amount of storage requirements for data and the memory for data analysis. The second issue is the performance of the BI and DM tools used by the data warehouse system are not satisfactory due to the complex data warehouse architecture design and more time consumption for data analysis. The third issue is the access to the data stored in the data warehouse is limited in terms of both users and geographical areas due to LAN (Local Area Network)/WAN (Wide Area Network) based systems. So, the data warehouse design should be modified as per the technological developments available in the industry and to satisfy the business needs.

This research paper proposes a new way of designing the data warehouse architecture using cloud computing technology with the help of fibre channel storage by reducing the overall cost with improvement in the data warehouse performance through global access for all kinds of users.

2. LITERATURE REVIEW

The literature review section is divided into four major parts namely data warehousing, cloud computing, storage area networks, and fibre channel respectively.

2.1. Data Warehousing

The first part focuses on the core components of data warehousing, various data warehouse design approaches, key factors for success, influences to maintain quality and issues in the design.

2.1.1. Components

The data warehouse consists of several components. Among these, the major components [1] are operational data store, load manager, warehouse manager, metadata, data mart, and query manager. Operational data store (ODS) is a repository of current and integrated operational data used for analysis. Load manager performs all the operations associated with the extraction and loading of data into the warehouse. Warehouse manager performs all the operations associated with

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the management of the data in the warehouse. Metadata of the warehouse stores all the metadata definitions used by all the processes in the warehouse. Data mart is a subset of a data warehouse that supports the requirements of a particular department or business function. Query manager performs all the operations associated with the management of user queries.

2.1.2. Development approaches

Inmon model [2] [13] is based on enterprise data warehouse approach, subject-oriented, integrated, top-down, relational database tools, nonvolatile, and time-variant. Kimball model [3] [13] is based on data mart approach, bottom-up, dimensional modeling, plan big, build small. Independent and dependent data mart models were also used in this design.

Real-time (active) data warehousing is the process of loading and providing data via a data warehouse as they become available which supports operational decision-making. Virtual data warehousing approach is a set of views over operational databases and only some of the possible summary views may be materialized. It is also called Management Information Systems approach.

The four views of a business analysis framework [6] are: Top-down view - allows selection of the relevant information necessary for the data warehouse. Data source view - exposes the information being captured, stored, and managed by operational systems. Data warehouse view - consists of fact tables and dimension tables. Business query view - sees the perspectives of data in the warehouse from the view of end-user.

2.1.3. Factors determining data warehousing architecture

The major factors determining data warehousing success [5] are: Cross function, campus-wide support, Consistency and integration across subject areas, Security, access control, Support for operational users and Support flexible access for decision-makers.

The influencing factors determining data warehouse quality [4][14] are: Completeness, Non-redundancy, Enforcement of (business) rules, Data reusability, Stability and flexibility, Simplicity and elegance, Communication and effectiveness and Performance.

The major problems in data warehouse design [1][14] are: Underestimation of resources for data loading, Hidden problems with source systems, Required data not captured, Increased end-user demands, Data homogenization, High demand for resources, Data ownership, High maintenance, Long duration projects and Complexity of integration.

2.2. Cloud Computing

The second part focuses on the features, properties, benefits and advantages of modern cloud computing.

2.2.1. Modern on-demand computing

The on-demand computing [11] model evolved to overcome the challenge of being able to meet fluctuating resource demands efficiently. Because demand for computing resources can vary drastically from one time to another, maintaining sufficient resources to meet peak requirements can be costly. Overengineering a solution can be just as adverse as a situation where the enterprise cuts costs by maintaining only minimal computing resources, resulting in insufficient resources to meet peak load requirements. Concepts such as clustered computing, grid computing, utility computing, etc., may all seem very similar to the concept of on-demand computing, but they can be better understood if one thinks of them as building blocks that evolved over time and with techno-evolution to achieve the modern cloud computing model we think of and use today.

2.2.2. Beyond the desktop

Cloud computing [8] signifies a major change in how we store information and run applications. Instead of running programs and data on an individual desktop computer, everything is hosted in the "cloud" a nebulous assemblage of computers and servers accessed via the internet. Cloud computing allows us to access all our applications and documents from anywhere in the world, freeing us from the confines of the desktop and making it easier for group members in different locations to collaborate. The desktop-centric notion of computing that we hold today is bound to fall by the wayside as we come to expect the universal access, 24/7 reliability, and ubiquitous collaboration promised by cloud computing. It is the way of the future.

Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.2.3. Key properties

The following are the key properties [8] of cloud computing.

User-centric: Once we as a user are connected to the cloud, whatever is stored there - documents, messages, images, applications, whatever - becomes ours? In addition, not only is the data ours, but we can also share it with others. In effect, any device that accesses our data in the cloud also becomes ours.

Task-centric: Instead of focusing on the application and what it can do, the focus is on what we need done and how the application can do it for us., Traditional applications - word processing, spreadsheets, email, and so on - are becoming less important than the documents they create.

Powerful: Connecting hundreds or thousands of computers together in a cloud creates a wealth of computing power impossible with a single desktop PC.

Accessible: Because data is stored in the cloud, users can instantly retrieve information from multiple repositories. We are not limited to a single source of data, as we are with a desktop PC.

Intelligent: With all the various data stored on the computers in a cloud, data mining and analysis are necessary to access that information in an intelligent manner.

Programmable: Many of the tasks necessary with cloud computing must be automated. For example, to protect the integrity of the data, information stored on a single computer in the cloud must be replicated on other computers in the cloud. If that one computer goes offline, the cloud's programming automatically redistributes that computer's data to a new computer in the cloud.

2.2.4. Benefits

The benefits [9] of cloud computing are:

On-demand self-service: A client can provision computer resources without the need for interaction with cloud service provider personnel.

Broad network access: Access to resources in the cloud is available over the network using standard methods in a manner that provides platform-independent access to clients of all types. This includes a mixture of heterogeneous operating systems, and thick and thin platforms such as laptops, mobile phones, and PDA.

Resource pooling: A cloud service provider creates resources that are pooled together in a system that supports multi-tenant usage. Physical and virtual systems are dynamically allocated or reallocated as needed. Intrinsic in this concept of pooling is the idea of abstraction that hides the location of resources such as virtual machines, processing, memory, storage, and network bandwidth and connectivity.

Rapid elasticity: Resources can be rapidly and elastically provisioned. The system can add resources by either scaling up systems (more powerful computers) or scaling out systems (more computers of the same kind), and scaling may be automatic or manual. From the standpoint of the client, cloud computing resources should look limitless and can be purchased at any time and in any quantity.

Measured service: The use of cloud system resources is measured, audited, and reported to the customer based on a metered system.

2.2.5. Advantages

The advantages [9] of cloud computing are:

Lower costs: Because cloud networks operate at higher efficiencies and with greater utilization, significant cost reductions are often encountered.

Ease of utilization: Depending upon the type of service being offered, we may find that we do not require hardware or software licenses to implement our service.

Quality of Service: The QoS is something that we can obtain under contract from our vendor.

Reliability: The scale of cloud computing networks and their ability to provide load balancing and failover makes them highly reliable, often much more reliable than what we can achieve in a single organization.

Outsourced IT management: A cloud computing deployment lets someone else manage our computing infrastructure while we manage our business. In most instances, we achieve considerable reductions in IT staffing costs.

Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Simplified maintenance and upgrade: Because the system is centralized, we can easily apply patches and upgrades. This means our users always have access to the latest software versions.

Low barrier to entry: In particular, upfront capital expenditures are dramatically reduced. In cloud computing, anyone can be a giant at any time.

2.3 Storage Area Network

The third part emphases on the storage area network as universal storage connectivity and its components.

2.3.1. The universal storage connectivity

The cloud computing technology deals two fundamental storage models [7] namely (1) Network Attached Storage (NAS) and (2) Storage Area Network (SAN). Among these SAN model has several variants with Fibre Channel Protocol (FCP) support which is a high speed storage device with huge data transfer rate. A typical SAN architecture is shown in figure 1.

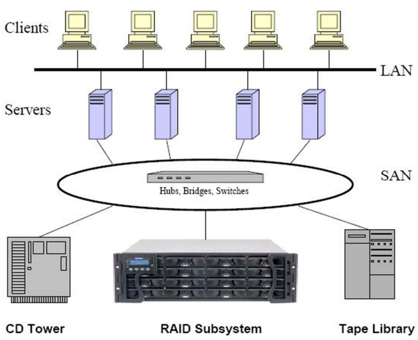

Figure 1: A typical SAN architecture

The key feature of SAN is any-to-any connectivity of computers and storage devices. According to the Storage Networking Industry Association (SNIA), "A storage area network is any high-performance network whose primary purpose is to enable storage devices to communicate with computer systems and with each other". SAN is Universal Storage Connectivity.

2.3.2. Components

The major components [10] of SAN are:

Network part: SANs provide a separated network over and above what the client/server or mainframe networks utilize in connecting clients, terminals, and other devices, including NAS. This network is based on the Fibre Channel Protocol with standards.

Hardware part: SANs depend on specific hardware devices just like other networks. Given that SANs provide a packet-switched network topology, they adhere to a standard layer of processing that the hardware devices operate within. Again, to avoid confusion between data communications packet topologies, the FC protocol relies on frames and architecture more compatible with I/O channel operations.

Software part: SANs operate within a separate set of software called a fabric that makes up the FC network. Additional software is required at the server connection where the average OS drivers must communicate within a SAN environment. The same thing is true for storage devices that must communicate with the SAN network to provide data to the servers.

Connectivity part: Much of a SAN's value derives from its connectivity, which is made up of several hardware and software functions. Given the complete shift in I/O operations from bus level communications to network communications, many components must change or be modified to operate within this environment. Connectivity options within the SAN determine performance and workload applicability.

2.4. Fibre Channel

The last part deals with fibre channel, design goals, topologies, protocol stack, switch/fabric, and advantages.

2.4.1. High-speed interconnect

Storage networking has been maturing since the introduction of Fibre Channel (FC) [7] in the early 1990s. At first, FC interfaces to storage devices were basically Server Attached Storage (SAS) - faster, fatter pipe that would allow storage to be located at a greater distance from the server, but it was still a point-to-point connection. All of the storage was dedicated to a single server, and the data was locked to the server.

High-speed interconnect technology has made SANs based on both FC and Gigabit Ethernet a reality. Most of the public attention given to SANs is focused on the interconnect that allow universal storage connectivity and the storage devices and computers that connect to them. A

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SAN device is a block-access that connects to its clients using FC and a block data access protocol such as SCSI.

2.4.2. Design goals

By coincidence, the design goals [12] of FC are covered by the requirements of a transmission technology for storage networks such as: Serial transmission for high speed and long distances, Low rate of transmission errors and Low delay (latency) of the transmitted data.

Implementation of the Fibre Channel Protocol (FCP) in hardware on Host Bus Adaptor (HBA) cards to free up the server CPUs.

2.4.3. Topologies

The topologies [7] of FC are:

Point to point, in which two devices are directly connected to each other by a FC cable as in figure 2(a).

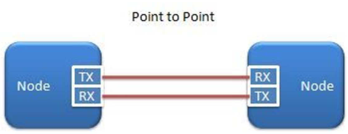

Figure 2(a): Point to point

Arbitrated loop, in which a number of devices time-share a single transmission path. A FC arbitrated loop can connect up to 126 devices. In practice, arbitrated loops are most often used to connect strings of disks to host bus adapters or RAID controller ports and are typically configured with 20 or fewer devices. It is shown in figure 2(b).

Switched or fabric, in which each device is directly connected to a central connecting point called a switch. The switch makes and breaks momentary connections between pairs of devices that need to communicate with each other. Like Ethernet switches, FC switches are high-performance devices that are capable of intermediating several simultaneous communications between pairs of devices. This is shown in figure 2(c).

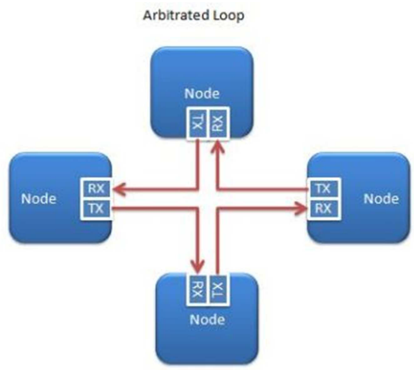

Figure 2(b): Arbitrated loop

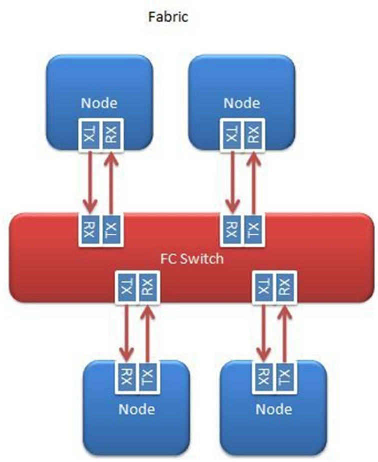

Figure 2(c): Switch or fabric

The features of individual FC topologies are shown in the following table 1.

Table 1: Features of FC topologies

Attribute	Point-to-Point	Arbitrated loop	Switched fabric
Max ports	2	127	~16777216 (2^{24})
Address size	N/A	8-bit ALPA	24-bit port ID
Side effect of port failure	Link fails	Loop fails (until port bypassed)	N/A
Mixing different link rates	No	No	Yes
Frame delivery	In order	In order	Not guaranteed
Access to medium	Dedicated	Arbitrated	Dedicated

2.4.4. FC-0: cables, plugs and signal encoding

FC-0 [12] defines the physical transmission medium (cable, plug) and specifies which physical signals are used to transmit the bits '0' and '1'. In contrast to the SCSI bus, in which each bit has its own data line plus additional control lines, FC transmits the bits sequentially via a single line. In general, buses come up against the problem that the signals have a different transit time on the different data lines, which means that the clock rate can only be increased to a limited degree in buses. The different signal transit times can be visualized

Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

as the hand rail in an escalator that runs faster or slower than the escalator stairs themselves.

FC therefore transmits the bits serially. This means that, in contrast to the parallel bus, a high transfer rate is possible even over long distances. The high transfer rate of serial transmission more than compensates for the parallel lines of a bus. The transmission rate of actual components increases every few years. Table 2 shows the market entry and the roadmap for new higher transfer rates. FC components are distinguished as Base2 (upper table) and Base10 (lower table) components.

Table 2: FC components

Product Naming	MByte/s (per direction)	T11 Spec Completed	Market Availability
1GFC	100	1996	1997
2GFC	200	2000	2001
4GFC	400	2003	2005
8GFC	800	2006	2008
16GFC	1,600	2009	2011
32GFC	3,200	2012	Market demand
64GFC	6,400	2016	Market demand
128GFC	12,800	2020	Market demand
Product Naming	MByte/s (per direction)	T11 Spec Completed	Market Availability
10GFC	1,200	2003	2004
20GFC	2,400	2008	2008
40GFC	4,800	2009	2011
80GFC	9,600	future	Market demand
100GFC	12,000	future	Market demand
160GFC	19,200	future	Market demand

When considering the transfer rate it should be noted that in the fabric and point-to-point topologies the transfer is bi-directional and full duplex, which means that, for instance for 2GFC components the transfer rate of 200 MByte/s is available in each direction. The FC standard demands that a single bit error may occur at most once in every 1012 transmitted bits. On average, this means that for a 100 Mbit/s connection under full load a bit error may occur only every 16.6 minutes. The error recognition and handling mechanisms of the higher protocol layers are optimized for the maintenance of this error rate. Therefore, when installing a FC network it is recommended that the cable is properly laid so that the bit error rate of 10−12 is, where possible, also achieved for connections from end device to end device, i.e. including all components connected in between such as repeaters and switches. Fiber-optic cables are more expensive than copper cables. They do have some advantages: Greater distances possible than with copper cable, Insensitivity to electromagnetic interference, No electromagnetic radiation, No electrical connection between the devices, No danger of cross-talking and Greater transmission rates possible than with copper cable.

2.5. FC switch/fabric

A FC switch [7] looks similar to a hub. It is an enclosure with a panel containing several sockets used to connect cables to the switch's internal port logic. A switch performs quite differently from a hub. Figure 3 depicts the key characteristics of a FC switch. For simplicity, a four-port switch is depicted. Actual switches contain more ports.

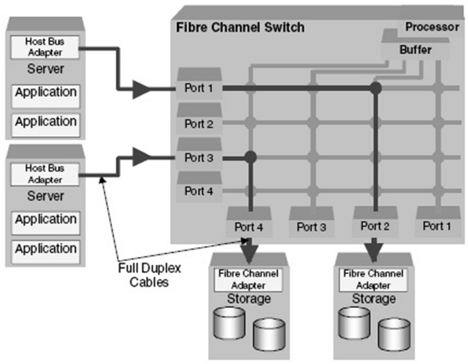

Figure 3: FC switch/fabric

Buffering within the switch provides elasticity, resolving small instantaneous differences between device transmission and reception rates, but, more important, it enables devices of different technology levels to intercommunicate. The FC industry is in the midst of a transition from first-generation 1-gigabit devices to second-generation 2-gigabit devices. Many users have large installed bases of first generation technology devices and a desire to make the SAN technology transition over a long period of time. The switch buffers transmissions from the transmitting device at its native speed and retransmits them to the receiving device at its native speed. The FC transmission protocols are designed so that switch ports and the devices attached to them negotiate to determine the highest possible speed that both devices support.
FC switches implement a standard network routing algorithm called find shortest path first (FSPF) to route data through the SAN.

The fabric topology [12] is the most flexible and scalable of the three FC topologies. A fabric consists of one or more FC switches connected together. Servers and storage devices are connected to the fabric by the FC switches. In theory a fabric can connect together up to 15.5 million end devices. Now, medium-sized installations comprise between 500 and 1,000 ports and large installations comprise several thousand

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ports. However, most installations are in a range below 200 ports. End devices connected to the various FC switches can exchange data by means of switch-to-switch connections (Inter Switch Links-ISL). Several ISLs can be installed between two switches in order to increase the bandwidth.

A transmitting end device only needs to know the Node ID of the target device; the necessary routing of the FC frame is taken care of by the FC switches. FC switches generally support so-called cut-through routing which means that a FC switch forwards an incoming frame before it has been fully received.

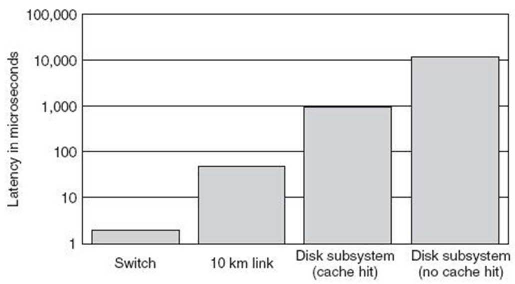

Figure 4: The latency of the FC

The latency describes the period of time that a component requires to transmit a signal or the period of time that a component requires to forward a frame. Figure 4 compares the latency of different FC SAN components. Light requires approximately 25 microseconds to cover a distance of 10-kilometres. A 10-kilometre long FC cable thus significantly increases the latency of an end-to-end connection. For hardware components the rule of thumb is that a FC switch can forward a frame in 2–4 microseconds; a FC HBA requires 2–4 milliseconds to process it. Additional FC switches between two end devices therefore only increase the latency of the network to an insignificant degree.

One special feature of the fabric is that several devices can send and receive data simultaneously at the full data rate. All devices thus have the full bandwidth available to them at the same time. A prerequisite for the availability of the full bandwidth is good design of the FC network.

2.5.1. Advantages

The advantages [7] of FC switch are:

Cost - FC is attractive for connecting disks to server and RAID controllers because it enables each server or controller port to accommodate lots of disks. But to be practical the per disk connection cost must also be low. Fibre Channel Arbitrated Loop (FCAL) topology minimizes the costs of connecting large numbers of disks to servers and RAID controllers.

Performance - FC is an ideal RAID subsystem-to-server interconnect, because it enables cross-connection of servers and RAID subsystems for increased performance as well as path failure tolerance. But for a FC fabric to be practical for this purpose; overall fabric I/O performance must be very high. Since there are few RAID subsystem and host connections in a typical data center (relative to disk connections), cost is a secondary consideration for this application.

Distance - FC is ideal for connecting tape drives mounted in robotic media handlers to computers, because the interconnect distances it supports allow backup tapes for several widely dispersed application servers to be placed in safe, environmentally controlled, professionally staffed facilities. Again, the number of tape connections in a typical data center is relatively small, so cost is a secondary consideration for this application.

Overhead - It seems that FC would also be ideal for connecting file servers to application servers because the efficient protocol designs and largely hardware implementations enable high I/O performance with low server and storage device processing overhead. For that to be practical, FC would have to support file access protocols as well as the block access FCP with which it was initially designed.

3. THE PROPOSED DESIGN OF DATA WAREHOUSE ARCHITECTURE

This research work proposes a new architecture for online data warehousing by using storage area networks in order to solve the issues namely higher cost, lower performance and limited access faced in the earlier data warehousing designs.

3.1. The Various Generations of Data Warehouse Architectures

First generation data warehouse architecture: During the invention of personal computers, all software developed was standalone and unsecured. The first generation data warehouse architecture design was also in the same model. Application developers developed software to

Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

access the data in the data warehouse directly through the simple versions of data mining tools without any kind of security. These data mining tools extracted knowledge and generated various forms of data using visualization techniques. The first generation data warehouse architecture was based on two tier model as in figure 5(a).

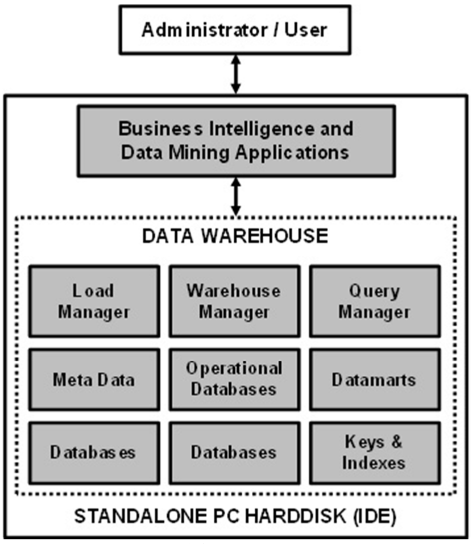

Figure 5 (a): First generation

Second generation data warehouse architecture: The rapid development of networking technology, introduced the client/server computing. The second generation data warehouse architecture was based on this model with security over data. It also allows the data access in a geographically distributed locations say servers. The applications developed for this kind of data warehouse systems were run on client systems in a secured manner. The second generation data warehouse architecture was based on three tier model as shown in figure 5(b).

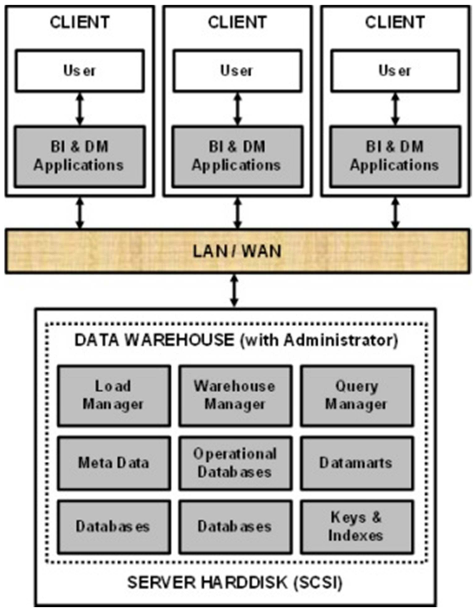

Figure 5 (b): Second generation

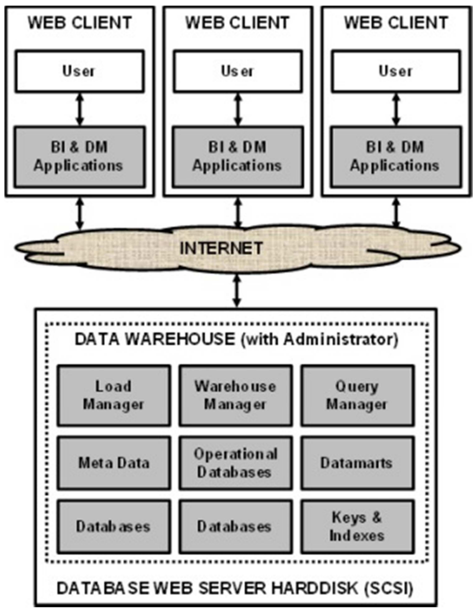

Figure 5 (c): Third generation

Third generation data warehouse architecture: The improvements over client/server computing technology reached the web-based applications using internet technology. The third generation data warehouse architecture is shown in figure 5(c). But, BI and DM applications running on internet were not satisfied with performance due to the protocols used and bandwidth reasons. Because, BI and DM applications require huge amount of RAM space and HardDisk storage space while working with data warehouse databases.

Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2. The Next (Proposed) Generation Data Warehouse Architecture

The need of modified data warehouse design and the features of network storage together form the new online data warehouse architecture. The proposed design is going to use the storage area network model for data warehouse storage. The framework of proposed design is shown in figure 5(d) as next (proposed) generation data warehouse architecture. The fibre channel fabric/switch plays a major role in between servers and storage devices.

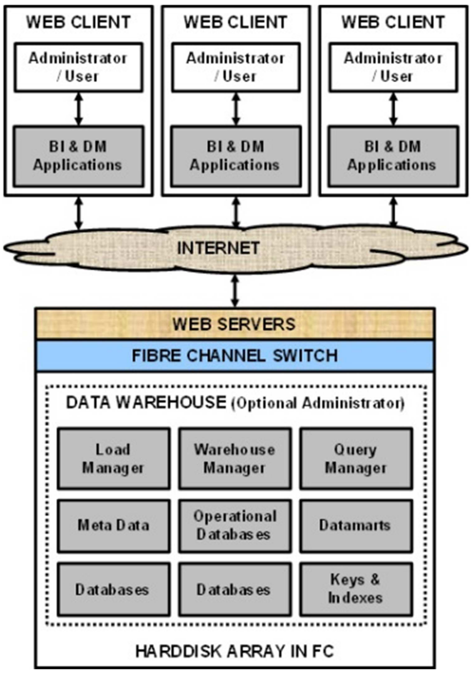

Figure 5 (d): Next (proposed) generation

3.3. Comparison of First, Second, Third and Next (Proposed) Generation Data Warehouse Architectures

The various existing generations of data warehouse architecture designs are compared with the proposed online data warehouse design on several technical features (nearly 20; namely number of tiers, number of users, data storage, storage technology, data transfer rate, processor load, network load, data accessibility, load balance, response time, application execution, warehouse performance, fault tolerance, recovery time, security, scalability, cost and services) as in table 3. From the table, it is clear that the proposed model have more solid reasons than other models.

3.4. Technology and Business Benefits of SAN

The storage networking is a mechanism for delivering services of benefit to the enterprise [7]. The deployment of storage networking should ultimately be justified in terms of the business benefits it delivers.

A SAN can reduce the capital cost of storage: Because all SAN-attached servers have direct access to all SAN-attached data, there is no business need to keep any more than one logical copy of a business data set online in a data center. The information processing benefit is lower capital expenditure for physical storage. The business benefits include lower information services capital budgets, lower management cost, and greater data reliability.

A SAN can reduce storage management cost: Eliminating redundant storage capacity and unneeded processing steps not only reduces capital and operational cost, it also reduces the number of objects the data center has to manage. Moreover, because SANs are fully interconnected, it becomes possible to consolidate management across the data center and around the world. The information processing benefits are lower operating cost and better quality of service. The business benefit is better quality more up-to-date information at lower cost to the business.

A SAN can reduce I/O bandwidth requirements: With only one logical copy of data, there is no need to copy data for other applications' use. This eliminates both I/O bandwidth and elapsed copying time from the data center work flow. The information processing benefits are lower capital cost (because less storage and I/O infrastructure is required) and improved work flow. The resulting business benefit is lower information services cost.

A SAN can improve the accuracy and timeliness of business data: With a fully interconnected SAN, all applications process the real data, not out-of-date copies. Moreover, one less step in the data processing chain is one less opportunity for procedural error. Fewer steps – particularly fewer human-assisted steps—mean fewer errors in the long run, as well as fewer opportunities for data loss or theft. Fewer procedural errors ultimately mean less cost to the business, as well as better quality and more timely information with which to run the business.

Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A SAN can extend the useful life of capital equipment: SAN interconnectivity makes it easy to repurpose equipment – to move older disks or RAID subsystems to another system –when, for example, storage devices are upgraded. Not only does the standards-based nature of SANs make storage devices independent of servers, but repurposing is generally done via administrative actions performed at a console. Because everything is interconnected, expensive and disruptive physical reconfiguration is seldom necessary.

4. RESULTS AND DISCUSSION

Figure 6 compares the CPU load of TCP/IP and FC data traffic [12]. The reason for the low CPU load of FC is that a large part of the FC Protocol stack is realized on the FC HBA. By contrast, in current network cards a large part of the TCP/IP protocol stack is processed on the server CPU. The communication between the Ethernet network card and the CPU takes place via interrupts. This costs additional computing power, because every interrupt triggers an expensive process change in the operating system. The CPU load of FC is low because its protocol is mainly processed on the HBA.

The foregoing SAN benefit analysis considers only information accessibility strategies. A more accurate picture of the business benefits of SANs may be obtained by considering the overall business value of information accessibility and reconciling this view with the data availability enabled by SANs.

In electronic business, the ability to process information can mean enterprise survival because the competition is literally "a click away". Table 4 lists the cost of inability to process information for various periods for two businesses for which online data processing essentially is the business. A typical home shopping operation processes large numbers of relatively low-value transactions. With a stock brokerage, on the other hand, individual transaction values can be much greater, so the cost per minute of downtime is correspondingly higher.

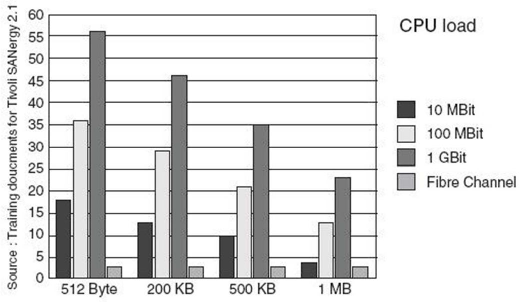

Figure 6: TCP/IP data traffic places a load on the server CPU. (The y-axis shows the CPU load of a 400MHz CPU).

The rows of table are extrapolated from the top row by multiplication. In reality, other factors, such as inability to meet financial obligations, would cause even larger consequences than the financial cost suggested by the table 4.

Table 4: Cost of downtime – Some examples

PERCENTAGE OF UPTIME	ANNUAL DOWNTIME	ANNUAL COST (HOME RETAIL)	ANNUAL COST (BROKERAGE)
99.9999%	30 seconds	$950	$53,750
99.999%	5 minutes	$9,417	$537,500
99.99%	52 minutes	$98,000	$5,590,000
99.9%	8.75 hours	$988,750	$56,000,000
99.5%	43.7 hours	$5,000,000	$280,000,000
99%	87.6 hours	$10,000,000	$560,000,000
98%	180+ hours	$20,000,000+	$1,000,000,000+
95%	450+ hours	$50,000,000+	$3,000,000,000+

The value of uptime may differ for different enterprises, but the overall message of Table 4 that downtime is cripplingly expensive – is a universal one. SAN technology enables new information processing techniques that can sharply reduce downtime, which, as can be seen from Table, can mean very significant cost savings or even survival for an enterprise.

When lost opportunity cost due to computer system downtime is considered, SAN adoption appears in a different light. It may well be possible to justify accelerated SAN adoption not just with direct benefits, but with the reduced financial impact that comes from higher uptime. Considering storage networking as an enabling technology, there is a continuum of techniques that can elevate data and application accessibility to the very nearly continuous. Figure 7 illustrates several data and application availability techniques that increase accessibility at increasing cost.

The availability baseline illustrated in figure is an adequately managed but otherwise unremarkable computer system that is backed up on

Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

a reasonably frequent and regular basis. For example, the system might serve the needs of a retail operation that does business only during the normal business day. The 96.5 percent uptime suggested in figure would amount to a total of about three days of annual downtime for a system with a duty cycle of five 8-hour days per week.

With basic data availability, backup is the only failure protection mechanism. The implicit assumption is that the duty cycle of the system allows for an adequate backup window. For any failure that affects data, including system crash, application fault, or disk storage subsystem failure, recovery consists of repairing the fault, restoring a backup, and restarting operation.

Basic data availability may be adequate for an 8×5 system. Such systems are a dying breed in the electronic business era, where 24×7 operations have become the norm.

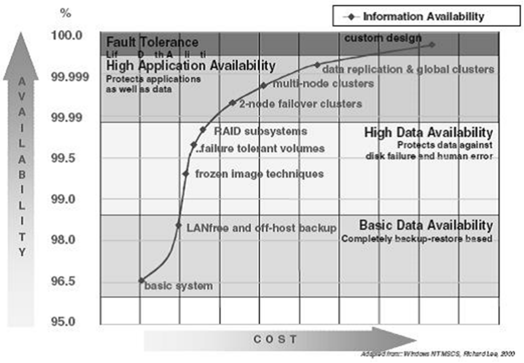

Figure 7: Degrees of system availability.

For a system that runs 24 hours per day and 7 days per week, 96.5 percent uptime would mean an average of about 12 days per year of downtime. This is clearly inconsistent with 24×7 operation. So the first lesson to draw from figure is that basic data availability, defined as periodic backup, is inconsistent with anything even approaching continuous operation.

Figure suggests that LAN-free and off-host backup, both of which are enabled by SANs, increase uptime. Both of these techniques reduce backup impact on applications, which in turn enables more frequent backups. More frequent backups means shorter restore times and less journal playback or other forms of data reconstruction, which in turn means less time until applications are back in operation after a failure.

As figure also suggests, this improvement comes at a cost. The cost comes from a combination of incremental SAN hardware and software that enables tape drive sharing and, for off-host backup, the transfer of online data ownership between application server and backup media server.

5. CONCLUSION

The proposed online data warehouse architecture design with storage area network model through fibre channel fabric shows that the cost of infrastructure for storage and computing devices are being drastically reduced by replacing them on rental basis instead of buying, which is the core of cloud computing service called Storage as a Service (SaaS). If we prefer to for buy option then the initial investment cost would be high and the maintenance cost would be less and we should update regularly both the software and hardware as per change in the industry which has several headaches as maintenance. Next, the overall performance of the proposed data warehouse architecture design is more satisfactory than the previous designs because the new design uses high speed storage devices (RAID) with very high data transfer rate (fibre channel fabric) along with scalability as discussed earlier. Finally the data access into the data warehouse is clearly examined that the cloud computing and storage area networks provides any to any connectivity as default. So, any authorized user might access the data warehouse from anywhere in the globe around the clock without any access limit on distance and time. Thus, the solutions suggested in this research work could definitely satisfy the business needs in terms of business intelligence and data mining tools which are being used for data analysis in data warehouse.

REFERENCES:

[1] "Data Warehousing Concepts Transparencies", Pearson Education Limited 1995, 2005.
[2] William H. Inmon, "Building the Data Warehouse", Fourth Edition, Wiley Publishing, 2005.
[3] Ralph Kimball and Margy Ross, "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling", John Wiley & Sons, 3rd edn. 2013.
[4] Graeme C. Simsion and Graham C. Witt, "Logical Database Design", Elsevier Inc. 2005.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[5] Peter Cava (UC Berkeley, Manager Data Warehouse Services), "Enterprise Data Warehouse" Oct'5, 2006.
[6] Jiawei Han, Micheline Kamber and Jian Pei, "Data Mining: Concepts and Techniques", Morgan Kaufmann (Elsevier) Publishers, 3rd edn. 2012.
[7] Richard Barker and Paul Massiglia, "Storage Area Network Essentials-A Complete Guide to Understanding and Implementing SANs", Wiley Computer Publishing, 2010, pp. 4, 29-47, 300-301, 303, 304.
[8] Michael Miller, "Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online", Que Publishing, 2009, pp. 12-15.
[9] Barrie Sosinsky, "Cloud Computing Bible", Wiley Publishing, Inc. 2011.
[10] Robert Spalding, "Storage Networks-The Complete Reference", McGraw-Hill/Osborne, 2003, pp. 182-183.
[11] John W. Rittinghouse and James F. Ransome, "Cloud Computing: Implementation, Management, and Security", CRC Press Taylor & Francis Group, 2010, pp. 36-37.
[12] Ulf Troppens, Wolfgang Muller-Friedt, Rainer Wolafka, Rainer Erkens, and Nils Haustein, "Storage Networks Explained-Basics and Application of Fibre Channel SAN,NAS, iSCSI, InfiniBand and FCoE", John Wiley & Sons, 2nd edn. 2009, pp. 66-69, 71-72, 89-90, 114-115.
[13] Ian Abramson, "Data Warehouse-The Choice of Inmon versus Kimball", IAS Inc.
[14] Sean Kelly, "Data Warehousing in Action", Wiley India, 2007, pp. 112-113.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3: Analysis Of Various Generations Of Data Warehouse Architectures

Data Warehouse Architecture Feature	*First Generation*	*Second Generation*	*Third Generation*	*Next Generation (proposed)*
1.Number of tiers	Two tier model	Three tier model	Three tier model	Multi-tier model
2.Number of users	Single user model (or multiple users but, one at a time)	Limited multi-user model	Unlimited multi-user model with network bandwidth limitations	Unlimited multi-user model without any bandwidth limitations
3.Data storage	Local standalone PC hard disk with space limitations	Shared client/server hard disk with space limitations	Shared database web server hard disk with space limitations	Shared array of hard disk (RAID) without any space limitations
4.Storage technology	IDE or Enhanced IDE	Parallel SCSI	Ultra SCSI	FC (or iSCSI)
5.Maximum data transfer rate	IDE - 3.3 to 25 MBps Enhanced IDE 16.7 to 133 MBps	Parallel SCSI 5 to 20 MBps	Ultra SCSI 40 to 640 MBps	1GFC to 40GFC 100 to 4800 MBps per direction
6.Processor load	Heavy due to multi-tasking in PC	Shared heavily but limited users	Shared heavily but limited bandwidth	Shared lightly with FC management (shared by HBA)
7.Network load	Not applicable	Shared with limited users and traffic	Shared with limited bandwidth	Shared without any restrictions
8.Data accessibility	Fast due to local access	Based on network traffic	Based on network and internet bandwidth	Normally fast due to FC
9.Load balancing	Possible with local buffering only	Possible with C/S and network buffering	Possible with C/S, network and internet buffering	Possible with C/S, network, internet, FC and HBA
10.Response time	Fast due to local processor	Medium because shared	Medium because shared	Fast due to FC
11.Application execution	Local	Client	Client or server	Client or server
12.Overall data warehouse performance	Initially good, but poor when data volume increases due to single user model	Initially good, but not satisfactory when data volume increases	Moderate, but not satisfactory when user volume increases	Excellent due to FC
13.Fault tolerance	Not possible	Possible with backup server	Possible with backup web server	Intelligent FC switch, RAID and HBA
14.Recovery time	Usually long time, but based on backup policy	Moderate, but based on backup policy	Moderate, but based on backup policy	Negligible due to FC, RAID and HBA
15.Security	0%	Not 100% due to C/S policies	Not 100% due to C/S, network and internet	100% assured
16.Scalability	Possible only by connecting limited number of hard disks	Possible only by connecting more hard disks in both C/S	Possible only by connecting more hard disks in Web Server	Default and automatic due to RAID and FC fabric
17.Cost	Low	Medium	High	As per usage (daily / monthly / yearly)
18.Services available	Nil (local services only)	C/S services only	Limited web services	Unlimited web services with cloud model