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

REAL-TIME WAREHOUSE ARCHITECTURE PROPOSAL FOR OIL AND GAS INDUSTRY

[1] AYAH F. LEERI, [2] MOHAMMAD A. ALKANDARI

[1] Computer Engineering Department, Kuwait University; Kuwait Oil Company. Kuwait

[2] Computer Engineering Department, Kuwait University, Kuwait

E-mail: [1]aleeri@kockw.com, [2]m.kandari@ku.edu.kw

ABSTRACT

Oil, gas and resources industry plays a dominant role as an income resource for many major companies and countries. To increase the Income through raising the production, many real-time data captured needs to be analyzed and decisions needs to be taken rapidly to optimize well performance and increase production. Currently, data captured are being stored in various systems which forward different tags to multiple applications. In this paper, we study the requirements and nature of Oil, gas and resources industry. We then propose an architecture for centralized real-time warehouse that shall consolidate all data (real-time and static) that are gathered from different resources such as wells, pipelines and other production related information. The aim of this paper is to provide a green, scalable and secure infrastructure for data storing and data analyses, taken into consideration the nature of Oil, gas and resources environment.

Keywords: *Software Architecture, Real-Time Warehouse Architecture, Data Warehouse Architecture, Software Requirements.*

1. INTRODUCTION

Due to the criticality of the Oil and gas sector, Data captured from wells and gathering centers are processed through a real-time Databases (RTDB). Depending on the policies of the Company, One or multiple vendors are responsible of managing and maintaining those RTDBs. Each RTDB is designed by a high-end international company that retails a full package system starting from RTUs which collects data wirelessly from the wells through the Fiber connectivity to the RTDB. That is located in the company's Datacenter. However, as the oil and gas market is a raising industry where Oil production is expected to increase to accommodate the needs or the market, more Oil wells are being drilled and others are decommissioned on regular bases. In an Information technology business, this will result in growth of Data stored and analyzed, change in network topology, and might be adding further systems or applications to the overall Company structure.

The nature of Oil and Gas sector defines the data transition path into three steps: first Data gathered from on-site field sensors are captured and transmitted wirelessly to the nearest RTU. Data are then pushed through Fiber links into a SCADA (Supervisory control and data acquisition) Database servers in the field for controlling and monitoring. Lastly data are copied into historians existing in the datacenter for further processing and backup. In the current situation, Data analytics are being held on the application level. The application which handles the analytics pulls the data from different RTDBs within the datacenter. Thus, each system is currently connected to multiple databases and each time, a new system is introduced, it has to be connected to all needed databases as shown in figure 1-[1a]. Correspondingly, Switches' ports are being fully occupied by the multiconnection, firewalls are needed to isolate each system from the others as each data is considered confidential, and Network bandwidth is unable to tolerate the real-time movement of the data.

The current roadmap of most companies around the globe is to go green; that is to virtualize all solutions in data center and to eliminate all duplicated systems and replace them with a single comprehensive system. In this paper, we propose a new real-time data warehouse architecture that should replace all duplicated RTDB historians with a single RTWH that is able to serve all requirements as shown in figure 1-[1b]. In the following sections, we shall discuss in detail the business requirements of such solution and the resulting architecture which will address those requirements.

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

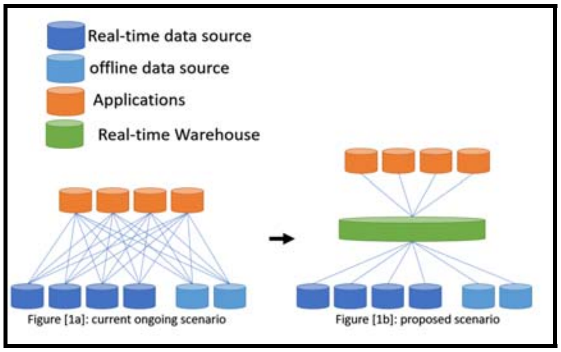

Figure 1: Comparison between overall System's communications before and after proposed system deployment.

2. LITERATURE REVIEW

2.1 In the field of Warehouse architecture

The field study of Real-Time warehouses is growing rapidly in the last decade. Many researchers surveyed the different parameters that defines the RTWH and proposed various models for implementing the same. The study of RTWH was initially derived from RTDB where researchers were aiming to combine similar RTDB info single repository.

Investigating the delay of near real-time data integration of temporal active databases was the lead motive behind real-time data warehouse for many researchers. As a result, it has been found that temporal faithfulness for transactions is to be preserved in order to maintain the serialization order of a subsystem. Despite the lagging which causes the data to be near real-time, the system is able to grantee the serialization order by assuring a (at least temporarily) faithfulness manner of the system [1].

To outcome the delay issue of data being Near real-time, Zero delay Data warehouse concept is proposed through a combination of Real-time data warehouse loading methodology (CORSO) as a middleware and General Database notification Gateway (GONG). Gong, which is a Tecco product, offers bidirectional replication of data between distributed databases. The replication can be integrated between homogenous and/or heterogenous databases in different scenarios such as mirroring, backup and distribution of data between a central warehouse and scattered databases [2, 3, 4].

Motivated by the medical industry, where numbers of patients (records) are huge and each patient file (parameters) keep growing in multiple directions (different medical fields), different architectures and models are to be used to design a data warehouse for the medical environment of hospitals and pharmacies. However, as the huge amount of information and data gathered within the warehouse are of different parameters, the data warehouse should be flexible to accommodate even future needs with minimum changes required. [5, 6].

Similar research study was conducted in [23] regarding data warehouses and its features. However, the research does not take in consideration the security measurement or the Real-time requirement of the data. Furthermore, the research focuses on the extract, transform and load ETL which is the heart of the real-time warehouses. The author surveys the existing technologies that ease the development of the warehouses in terms of software and hardware but never discuss the warehouse as a system. Many comparisons are made between various existing ETL tools based on their non-functional requirements such as availability, scalability and many other metrics.

Building a warehouse in practical requires a workflow or a framework that identifies the key steps for configuring and customizing a warehouse. Steps are initiated from finding a data source to customizing an ETL tool to extract, transform and load the data [24]. However, the research is limited to establishing a library data warehouse and does not consider existing ETL tool for easier and cost-efficient implementation

2.2 In the field of Oil and Gas Industry

Few studies were published in relation to wells or oil production data storing in relation to the digital world. However, if we examine the nature of oil, gas and resources industries to its bottom level, it is found that technology can optimize well production and help increase the productivity of a given system. The general system architecture of the digital oil fields is initiated at the rig level where wireless sensor networks are integrated with the real-time database along with other network components to gather and analyses well data. Therefore, WSN are the current key technology for monitoring well performance which helps the industry in terms of productivity, HSE and managements aspects [7].

Researchers' trend nowadays is to map Big Data studies to databases and Oil production researches. The exploration of bigdata effect on Oil and Gas production reveals that the impact of real-time data and technology on various states of petroleum production such as exploration, drilling, production and export can increase the efficiency of production data analysis and prediction [8].

Kuwait Intelligent Digital Oil Fields (KWIDF) are one of the biggest demonstration on the success of introducing technology to oil production. Several conferences and research papers were published in this regard [25, 26, 27]. KWIDF projects are established by Kuwait Oil Company to enhance and optimize wells' production and systems' monitoring. Basically, the projects allow field engineer and the production operation team to monitor and control the wells and the facilities. Furthermore, it captures all data and forwards them to the petroleum engineers responsible for field development and reservoirs managements to analyze the gathered data and plan a better strategy for optimizing production. The outcome of the field development team's study is then forwarded back to the production operation team to take action and wait for wells' response and crude production enhancement. The system was furtherly upgraded to include some dashboard which can be shared with higher management to summarize the well's production status. Additionally, the equipped wells' performance is also monitored through several systems so that any failure would trigger an alarm for Well Surveillance team to check and fix and failure related to the well and its' equipment. The collaborative work that the KWIDF project delivers was highly appreciated such that it spread along all assets and is adopted for most of the wells controlled by the company. The success is also shared as best practice in international SPE conferences.

3. ANALYZING REQUIREMENTS:

In order to introduce a successful design architecture of the needed system, requirement gathering and analysis is to be conducted. Furthermore, current network structure, software and hardware standards and security regulation must be introduced as they set the system's architecture limitation and shape the design's boundaries. In the next subsections, the system's requirements and boundaries shall be discussed in detail:

3.1. System's architecture must adapt to the current network structure and topology

The proposed system architecture shall correspond with the current existing network structure shown in figure [2] such that minimum costless change in to be implemented on the current network. Further, most of the network nodes to be affected within the project's implementation are critical real-time systems or databases that cannot endure any downtime nor exception for amendment of system and network topology.

Current network topology is segregated into 3 main layers where the field layer is presented in RED, the DMZ layer is presented in GREEN and the corporate network layer is presented in YELLOW. Field layer is the only layer with control-ability. Maximum physical and logical security measures are taken into consideration at this layer. The Field layer is isolated from any external interference. Data gathered from different and distanced wells are captured through the Remote Terminal Unit RTU and then forwarded to the Master Terminal Unit MTU via WiMAX. MTUs act like cluster heads as they aggregate the data from the wells to the nearest facility. Data are then forwarded to the respective historian resining in the DMZ layer in the Datacenter. DMZ Zone is an isolated network layer where all data is stored. All servers within the DMZ zone are hosted in the Datacenter and are only accessible through verified and approved user accounts. The purpose of the high security measures taken to secure the DMZ zone is that, though it has no control-ability, it stores critical and confidential data related to the core business. Thus, all real-time systems and Core raw confidential data are limited to DMZ users and systems. Corporate layer in the other hand, is the public network with internet access. All users are able to access corporate network from any PC within the corporate. All applications hosted in the corporate network are non-critical applications that capture non-confidential data. Thus, less security measures are addressed to provide the user with his requirements of more flexibility and liberty.

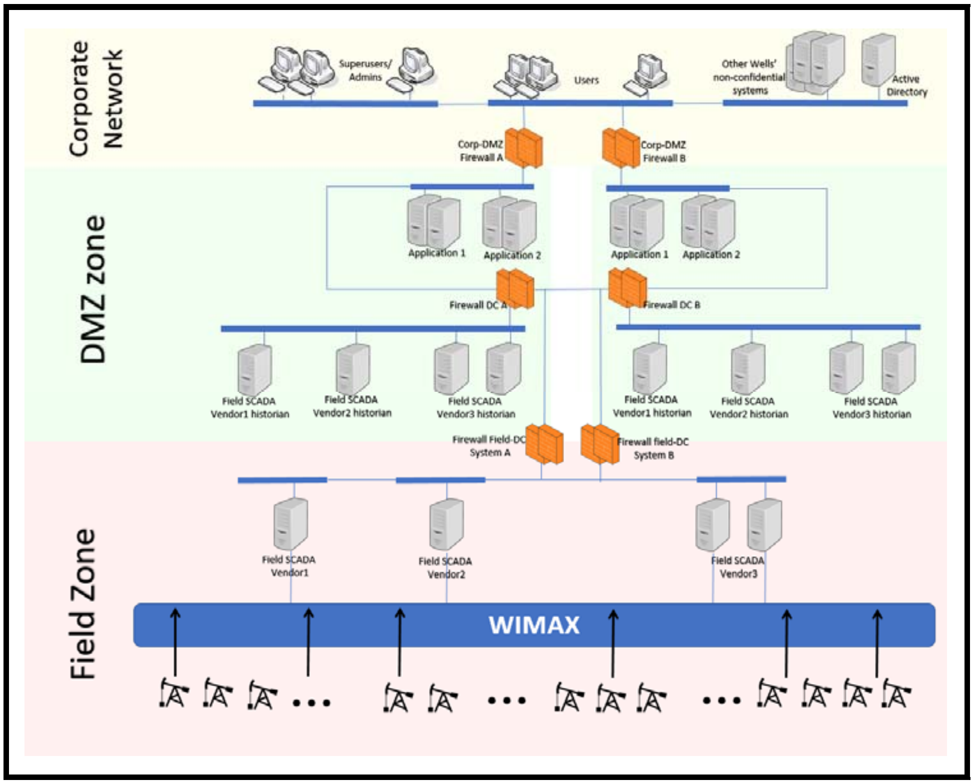

Figure 2. Existing System Architecture (assumption)

3.2. System's architecture must offer high availability and disaster recovery (site B)

Due to the criticality of the system as it shall be the single source of information for all production optimization systems, the system shall be high available such that it cannot tolerate more than 1 hours of downtime. The system should also be highly available in an active-passive manner. A secondary System (also known as disaster recovery site DR) should be a duplicate of system A (main site) in an active-passive or active-active manner such that at no single time both system shall be off.

3.3. Performance guarantee and Instantaneous Response

When designing a Real-time Warehouse, architect must consider the criticality of data Response. System must indorse the real-time data forwarded by the RTDBs and historians store them and forward them to the real-time applications within the accepted timeframe. Any delay that causes the system to be near real-time cannot be accepted.

Furthermore, the warehouse shall compose of an emergency buffer system within its application level, such that in case of network downtime or any miss functionality in the RTWH system, the RTU and MRU planted in the field along with the local Database aggregating the data in the facility are all assumed to be equipped with a cache that stores the data before insuring that it reached the historian or warehouse system. Any failure in assuring the data receipt, the last confirmed data host shall forward the missing data to the RTU through a separate emergency port. The warehouse shall mitigate this feature by accepting multiple data entry from the same host acting as a buffer which is able to restore the lost data and store them back in the RTWH. The emergency port shall read the missing data and store them in the right order such that historical data is always complete and in order.

3.4. RTWH should be secured within its network layer

Oil, gas and resource data are considered as a valuable asset to every company running this business. Therefore, the data need to be secured within its network such that no foreign entity is able to read, write or manipulate these data. As the system is somehow connected to the Supervisory Control and Data Acquisition (SCADA) servers which have the control over the wells pipelines, extra security measurements need to be taken to ensure that no hack attempts ever reach the SCADA server.

Additionally, the data is to be encrypted such that in case of any breach of the network, the data will still be unreadable and unable to be processed. Due to the criticality of the system, all read-time data is considered confidential.

Security measurement should not affect the real-time nature of the system. Time variance between the warehouse timestamp and the RTU timestamp is due to data crossing many hops such as servers and firewalls. The architect shall secure the data and grantee the performance as per section 3.3.

3.5. The system shall be Scalable in both directions such that it can connect to additional databases (data sources) and additional applications (data destinations)

Real-time warehouse is not to be built as a fit to purpose model, it shall serve the environment for the upcoming 10 to 20 years. As the data sources (such as wells and facilities) are growing in number, system architecture should be equipped with the needed to support any expansion to the system with no extra cost, delay or performance issues. Furthermore, The RTWH as a source for other systems should be able to forward its data to multiple numbers of systems regardless of bandwidth limitation or connectivity issues.

3.6. New RTWH implementation is to be migrated into the current network with zero down time seen by end-user

Most of the stakeholders (specifically end-users) are full time employees who are working 24/7 on monitoring the conditions and production of wells and their equipment. Thus, they are unable to accept any deliberative downtime for any system migration. The new warehouse implementation shall take transition in the backend so the end-user shall not be affected. All network connections shall be duplicated at time of migration. Once confirming the health of the new connection, the old link is to be revoked. This implies that network ports are to be added and configured ahead. All current systems shall be also reconfigured for such shift.

3.7. Operating System and other licenses

The operating system of the proposed system's servers shall be as aligned with the corporate standards. All other software shall be procured and maintained by the corporate. Furthermore, the admin rights of the Operating system level shall be exclusive to the system's owner. However, the Application admin privileges can be granted for the vendor as per need and proper justification.

3.8. Corporate standards

System design and specification must indorse corporate standards as provided. The standards include but not limited to: security standards, hardware standards, network configuration standards, ports configuration standards and user access protocols. Furthermore, the proposed system design shall be tailored to the corporate standards in term of hardware and software technologies.

3.9. Open Platform Communications (OPCs), protocols and interfaces

The warehouse application shall be compatible with all interfaces used by vendors within their respective historians and applications. The warehouse system is expected to read data from all historians' servers and field servers. Additionally, the warehouse system shall also export the data in a standard version and extension that is readable by all applications systems. The standard for the same in the market is to use an Open Platform Communications server-to-server as a mediator between different protocols and vendors' segregated systems. The open platform communication also known as OLE for Process Control is a set of standardizations and specifications for industrial communication. It is used to standardize the real-time communication between the control devices and the various manufactures and vendors. Furthermore, any web application must be configured to use HTTPS protocol to assure a smooth secure performance.

4. PROPOSED ARCHITECTURE

After studying and analyzing the given requirement in section [3], the proposed architecture, as shown is figure [3], summarizes the high-level design of the warehouse system. Following that, we will elaborate on the changes and additions that this paper is applying to the existing architecture:

In the proposed architecture, we introduce three main components as follow:

• Clustered Warehouse: the warehouse is suggested to be dispersed among multiple physical servers and virtual machines. This will assure the redundancy of the system such that if for any reason one server fails the rest of the servers can continue providing the service.

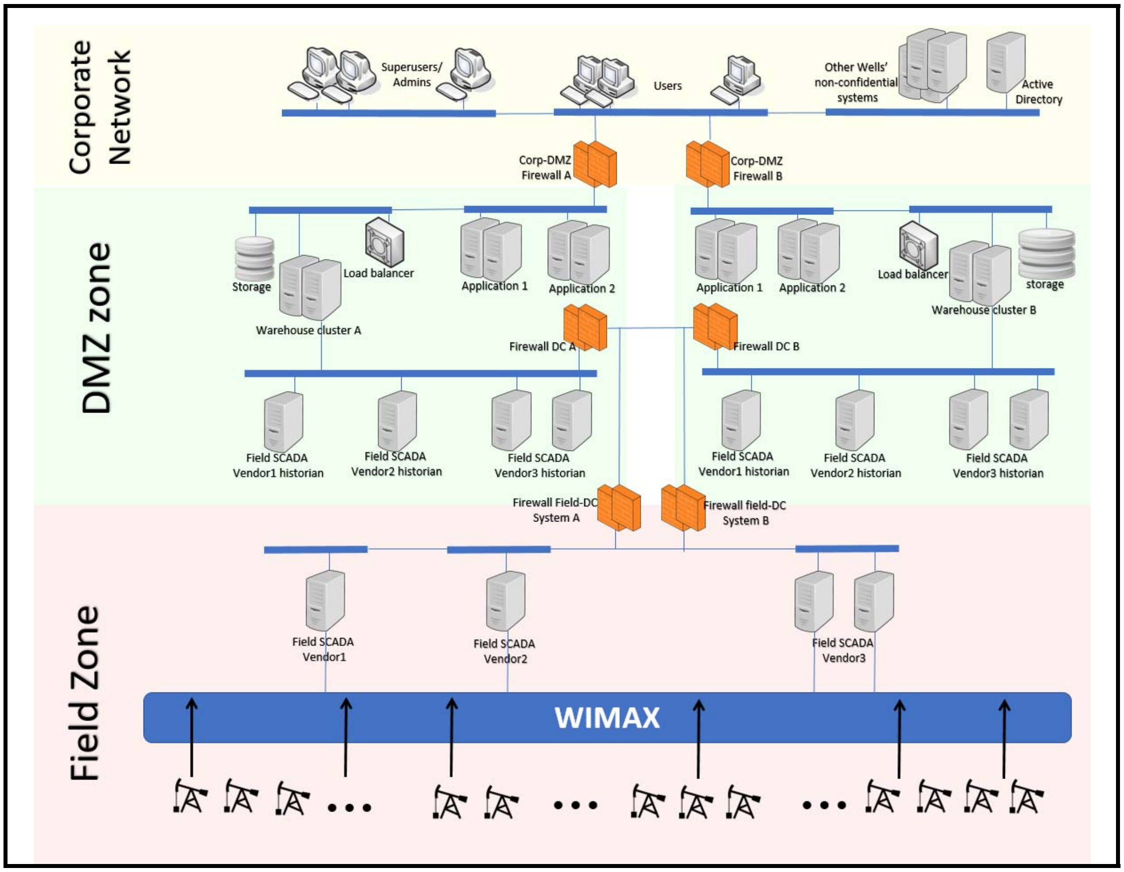

Figure 3. proposed System Architecture

Furthermore, it is suggested to host the warehouse on High-performance computers (HPC) which will assure the high performance of the system.

• Data Server Load Balancer: the purpose of the load balancer is to balance the network traffic demanding the services. Data Load balancer assures that requests doesn't focus on one server such that it crashed out of high demand and low memory space. Again, this is to assure the high performance of the system through managing servers' traffic and communication bandwidth.

• Storage system: purpose of the storage system is to backup all data within the warehouse. As the size of the real-time data is very huge, the warehouse system is to keep record of last 24 hours' data only. Any previous data is to be fetched from the storage if needed.

• Separate VLAN: system is configured on a separate VLAN to strength the security measurements taken. As the DMZ layer is fully secured, Ports needs to be configured between different applications and the warehouse. However, the warehouse is able to communicate directly with its storage as they are within the same VLAN.

The system is duplicated in two different sites such that is one site fails the second site is able to provide continuity of services. System A is to be mirrored in system B. each system shall have its own load balancer, firewalls, servers and storage such that both systems are absolutely independent of the other in terms of operating and functioning. This will assure the disaster recovery presence of the system.

As for the security, the warehouse system is surrounded with firewalls such that it is secure from all levels of the network. Hosting the warehouse in the DMZ zone prohibits any internet connections or external links. This implies that the only threat to the warehouse is through the corporate network or physically accessing one of the field nodes. However, all fields are highly secured and fenced by the company. The threats are limited to human factor within the corporate network layer. Thus, Firewall is planted between the corporate network and the DMZ zone to assure the security and integrity of the data within the warehouse system.

On the long run, once the warehouse system proves its best practice theory, the warehouse system is to replace all historians. This shall safe money and effort maintaining and supporting various servers and application from different vendors. Additionally, it shall reduce delay on data due to crossing many firewalls and applications.

5. TECHNOLOGIES IN THE MARKET

Next, we shall compare top leading Technology Brand for our proposed solution. This includes Hardware such as Racks, servers and network equipment and Software such as Operating systems and Databases technology. We shall compare them based on their differences and major features. However, the choice of technology is also related to the corporate standards. Therefore, the final choice of technology is based on the corporate discussion. This paper proposes the leading top technologies and their advantages and disadvantages for the user to decide upon. User shall consider the maintenance, support and configuration of the listed hardware and software. Furthermore, the following listed technologies are not comprehensive. Thus, additional technologies are to be used if they pass the minimum fit to purpose requirements.

5.1. Hardware & Servers

Top leading server Brands that support High-performance computing are Lenovo, Dell and HP. Many comparisons have been published to compare their performance. Following, in table [1], we will summarize top factors of that comparison

Table 1: Hardware Technologies comparison

	IBM x3650 M5 [11]	Dell R730 [10]	HP DL380 Gen 9 [9]
Height	2U	2U	2U
Diagnostic Panel	Exist	Status/ error message	No
Component Level LEDs	monitor 6 components	monitor 3-4 components	Blinking lights only.
Max RAID Support	4 RAID adapters	1-2 RAID Adapters	1-2 RAID Adapters
Processor	IntelXeon® E5-2600 v4 series	IntelXeon® E5-2600 v4 series	IntelXeon® E5-2600 v4 series
Memory (Max)	1.5TB LRDIMMs/R DIMM 2400 MHz	32GB RDIMM, 2400MT/s	3.0TB With 128 GB DDR4

Taking in consideration the amount of data to be stored, analyzed and forwarded by the warehouse, High performance computers (HPC) servers are highly recommended. HPC is provided by most of the major servers' providers. The choice of the hardware needs also to consider the corporate existing infrastructure and availability of maintenance contracts or expertise.

5.2. Virtualization environment

The most popular Virtualization environments are VMware and Hyper-V as per Gartner [14]. VMware is an independent company that can adapt to any operating system. While Hyper-V is a Microsoft solution. In table [2] we shall compare between the two available virtual machines.

Table 2: Virtual machine technology comparison

	Hyper-V	VMware
Host scalability	64 cores, 1 TB, 384 VMs	64 cores, 1 TB, 256 VMs
VM scalability	4 cores	8 cores, 255 GB
Support of Storage	Independent of the solution	Offers an integrate with Storage solutions

In order to understand the differences between Hyper-V and VMware, an intensive study in each's architecture is required. MS Hyper-V is founded on micro-Kernelized hypervisors in a parent-child matter. The parent also known as the host operating system must have at least one child which is the Virtua machine. MS Hyper-V isolated different virtual machines in partitioning matter. VMware vSphere package offers many software components such as vCenter, ESXi and vSphere client. The hypervisor (VMware ESXi) hosts all virtual machines on its server. To install, update or maintain these virtual machines an additional layer (vSphere Client) needs to be installed [30-31].

Research field is rich with comparisons between different hypervisors specially Hyper-V and VMware as they represent the most two leading vendors in the field [32-34]. Choosing the right hypervisor needs to be thoughtfully reviewed as it depends on various factors such as the Virtual machine performance, the memory management, high availability in terms of disaster prevention and disaster recovery, live migration, network interface, storage support and security measurements. Through an intensive test study, it has been found that VMware vSphere offers superior performance compared to Hyper-V in a fully virtualized environment. VMware vSphere outperformed Microsoft Hyper-V by 18.9%. it offers betters scalability and more consistent resource allocation [36].

A very intensive comparison can be found in [15]. However, as the study is published by VMware themselves, it's only fair enough to search

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

for wait for a similar study to be published by Microsoft.

5.3. Operating System

Three Major Operating Systems are leading the industry: Windows in its 7th version and Unix Redhat and Linux [12,13]. In table [3] we shall discuss the major available operating systems.

Table 3. Operating System technologies comparison

	Windows 7	Unix/ Redhat	Linux
Compatibility	High	Low	Low
Viruses	lots	Low	Few
Price	Relatively high	FREE/ inexpensive	FREE

Referring to section 5.3, Databases and warehouse compatibility issues narrows the operating system tradeoff to be between Windows and Unix. However, as the real-time data warehouse should be accessed by users whom, some of, are inexpert in computers and technologies, it is recommended to user Windows platform as it is user-friendly and well known to most of the users.

5.4. Database /warehouse

Warehouses existing solutions are surveyed in [23]. Based on the study of [23], it can be concluded that SQL and Oracle are the key solutions for building a warehouse. It shall be noted that both databases are not compatible with Unix Operating System.

Oracle published several studies and practices for real-time warehousing [16, 17, 18]. Oracle has launched a new product called Oracle Data Integrator. The Oracle data integrator consists of two environments, development and production. However, Oracle Data Integrator is a tool for a higher entity called Oracle warehouse builder. Oracle warehouse builder is an end-to-end solution for warehouse designing and Business Intelligence tool integration. The popularity of Oracle Warehouse Builder is that it depends more on graphical views and friendly user interface to build warehouses. It provides data superiority, data auditing, entirely integrated relational and dimensional modeling, and full lifecycle management of data and its related information [19, 20].

On the other hand, [21] discusses how the introduction of SQL Server Integration Services SSIS by Microsoft is considered a huge step toward the ELT market. The new Microsoft SQL Server Business Intelligence application requires a new approach for data capture which is names “Change data capture” CDC. CDC approach to data integration considers the change made to data with reference to its raw form which is captured from the historians. CDC takes care of extracting data and reducing latency when forwarding the data. The research also provides a use cases for Real-time warehouses using SQL server integration services and the deliverables that the SSIS adds to the system.

Furthermore, [22] discusses CDC in detail and how it provides an efficient extract, transform and load ETL processes. It also related CDC approach to Data warehouses design and runtime analyses.

5.5. Storage

There are currently 3 models for storages that are commonly used:

- Direct-Attached Storage (DAS): which is a block of desk directly attached to the Host/Server. Due to its setting, DAS is internally shared storage that cannot be shared. No network is involved in configuring DAS. However, a file system needs to be placed upon it prior to using the DAS.

- Network-Attached Storage (NAS): is an externally located storage in a dedicated server that is connected through Local Area Network (LAN) for shared access. A big advantage of NAS is that it allows storage sharing between multiple hosts connected within the same network. Also, it can support multiple Operating Systems and different File systems.

- Storage Area Network (SAN): is a high speed, isolated and external Storage system. San is usually linked through Fiber channel to assure the speed performance. SAN offers storage in three different modules: server-to-storage, server-to-server and storage-to-storage. SAN storage has many advantages over NAS and DAS as it offers the best speed performance compared to the rest. Furthermore, it offers LAN-Free backups and it’s overcomes the distance issues associated with communication establishment. SAN also provides server and storage consolidation which assures the high availability of data in the storage even if the server failed [28-29].

Due to the real-time feature of the warehouse, Storage technology must be fast, adaptive and it must tolerate the large amount of data fetch per second. Storage Area Network (SAN) is our best recommendation for the business need. SAN is located externally such that any failure in the server should not affect the storage

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

performance. Additionally, as the storage is separately located and configured it is scalable, upgradeable and sharable. This feature eliminates the Single point of failure factor from the server side. Furthermore, SAN provides maximum data backup performance in terms on speed and network communication which is an essential requirement of any real-time system

6. SYSTEM VALIDATION AND TEST PLAN

Testing the system is more focused on testing its performance rather than its actual successful implementation. As assumed in this research paper, an alternative solution of the warehouse is already defined and used. Thus, the research is designed to introduce a better solution of the existing in terms of the functional and non-functional requirements as shown in section [3]. We designed the Test plan to test the success implementation of the RTWH and test its basic functionalities. The test is to be configured on an existing network topology without decommissioning the existing systems represented by the various scattered Real-time databases as shown in figure [2]. The aim of keeping both systems (new and existing) up and running at the same time is to fairly compare the output and performance of both systems under the same identical environment and circumstances. Derived from the above-mentioned requirements, procedure of testing the system is to be such as follow:

6.1. Establishing connections and configuring system

First step of any system is to procure its hardware and establish connections as shown in figure [3] during the test, we shall user the following hardware and operating system:

- Lenovo X3650 M5 servers
- Windows 7 operating system
- Oracle warehouse solution
- Cisco 5585 Firewall
- Cisco c3650 switches
- Lenovo thinkStation P900
- VMware machine environment

Once all connections are established, we shall check the average number of ports utilized in the switches by each application and each database. Our estimates show that we shall expect a major decrease in number of ports as our proposed system is aimed to aggregate all connections into one system.

6.2. Warehouse configuration and user access

We need to assure that the warehouse is configured correctly before pushing data into it. Furthermore, we also need to check user accessibility to the warehouse. The test procedure must confirm user ability to log into the warehouse and view database configuration. Once data is stored in the warehouse, the user shall be able to view the respective data and confirm its availability.

6.3. Data integrity and time stamp validation

As the system proves its data availability, admins are to verify the data integrity with respect to the existing historians. Furthermore, as we are more concerned on system performance, we need to validate the timestamp on data such that the warehouse system must prove its efficiency. Time stamp from different existing databases must be compared to the warehouse data's timestamp.

6.4. Data consistency

Real-time data is a "continues flow of information" shaped as bits and bytes. System must be able to ensure the flow of data without any disturbance or loss of services. The warehouse system should be able to store and forward real-time data with no delays or denial of services. Validating this property is easily demonstrated using analysis applications to plot the flow of data. As the real-time warehouse is the single consolidated source of data, any failure in plotting the data that is presented in discrete plot is to be assumed as a fault in the RTWH functionality.

6.5. System's redundancy

As the system is configured on high availability requirements such as system is redundant in a secondary location, we must validate the high availability feature of the system by confirming that if system A if down, system B is to take over. To achieve that, we must force each of the system components to failure dependently of the others and confirm the continuity of service in despite of the failure.

Further Detailed test plan is included in Appendix [1].

7. FUTURE WORK AND PLANS

In this research, we focused on the storing and extraction feature of the warehouse. However, warehouses implementation offers much more possibilities such as data analysis and data

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

visualization. Once the warehouse proofs its purpose and performance, the research then is to consider implementing data analytics and performing some data workflows that can aid the business instead of using some third-party applications. Furthermore, in future we shall propose a detailed method of implementing the system. Time plan and schedule is to be considered. Warehouse's detail configuration manual need to be proposed and recognized. Additionally, the research in future, is to discuss in detail the way forward with the ETL unit of the warehouse and its real-time related complications as stated in many previously mentioned research papers. The research then will discuss the major vendor in the technology market. However, many newly formed companies or technologies have a solid fingerprint in the marker such as Citrix Xen in the hypervisors and virtual machines market and Virtual SAN storage which is the new era of storing data. In addition to the new releases that are to come in the future. Thus, the future release of this research needs to convoy the latest technology updates and consider additional unmentioned technologies that would benefit the research.

Finally, post implementing the system and testing it as stated in section [6], the research is to study the test outcomes and compare them with the theoretically assumed results. In case of any fluctuation, a detailed analysis and study must take place as part of this research or as a new research field.

REFERENCES

[1] Roddick, J. F., & Schrefl, M. (2000). Towards an accommodation of delay in temporal active databases. In Database Conference, 2000. ADC 2000. Proceedings. 11th Australasian (pp. 115-119). IEEE.

[2] Kühn, E. (2003, September). The zero-delay data warehouse: mobilizing heterogeneous database. In Proceedings of the 29th international conference on Very large data bases-Volume 29 (pp. 1035-1040). VLDB Endowment.

[3] Santos, R. J., & Bernardino, J. (2008, September). Real-time data warehouse loading methodology. In Proceedings of the 2008 international symposium on Database engineering & applications (pp. 49-58). ACM.

[4] T. Binder ,3002, Gong User Manual, Tecco Software Entwicklung AG.

[5] Jaber, M. M., Ghani, M. K. A., Suryana, N., Mohammed, M. A., & Abbas, T. (2015). Flexible Data Warehouse Parameters: Toward Building an Integrated Architecture. International Journal of Computer Theory and Engineering.

[6] Sahama, T. R., & Croll, P. R. (2007, January). A data warehouse architecture for clinical data warehousing. In Proceedings of the fifth Australasian symposium on ACSW frontiers-Volume 68 (pp. 227-232).

[7] Reza Akhondi, Mohammad, et al. "Applications of wireless sensor networks in the oil, gas and resources industries." Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on. IEEE, 2010.

[8] Baaziz, A., & Quoniam, L. (2015). How to use Big Data technologies to optimize operations in Upstream Petroleum Industry.

[9] Just the Facts: Lenovo HV Servers vs. HP and Dell Hewlett Packard Enterprise Development LP, "HPE ProLiant DL380 Gen9 Server", from: www.hpe.com.

[10] Dell, "PowerEdge R730", From :www.dell.com.

[11] Lenovo , "System x3650 M5 Rack Server", from; www3.lenovo.com.

[12] N Economides, "Linux vs. Windows: A Comparison of Application and Platform Innovation Incentives for Open Source and Proprietary Software Platforms".

[13] Rob Enderle,"enterprise windows vs. linux vs. unix server platforms", From: e-technologymanagement.com.

[14] Bittman, T. (2009). Virtualization With VMware or Hyper-V: What You Need To Know. Gartner Inc.[Electronic source] November.

[15]"vSphere 6.0 Advantages Over Hyper-V", From: http://www.vmware.com.

[16] Mark Rittman," An Introduction to Real-Time Data Integration", from: www.oracle.com.

[17] Real-Time Data Integration for Operational Data Warehousing,(2008) , from: http://www.oracle.com.

[18] Accessibility, Documentation. "Oracle® Warehouse Builder." (2010).

[19] Rittman, Mark. "Data Profiling and Automated Cleansing Using Oracle Warehouse Builder 10g Release 2." (2006).

[20] Borowski, Esther, and Hans-J. Lenz. "Design of a workflow system to improve data quality using Oracle Warehouse Builder." Journal of Applied Quantitative Methods 3.3 (2008): 198-206.

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

[21] Attunity, 2008, "Real Time Data Integration".

[22] Attunity, 2009, "Efficient and Real Time Data Integration with Change Data Capture".

[23] Yousuf, S., & Rizvi, S. S. A Compatrive Study of ETL Tools.

[24] Wah, T. Y., Peng, N. H., & Hok, C. S. (2007, November). Building data warehouse. In Proceedings of the 24th South East Asia Regional Computer Conference (Vol. 15, pp. 51-56).

[25] Desai, S. F., Pandey, M., Al-Shammari, B., Rane, N. M., Al-Rabah, A. A., Almater, D. S., ... & Jagannathan, R. (2016, November). Benefits Realized from Implementing Integrated Digital Field Technology in a Sector of Greater Burgan Field of Kuwait. In Abu Dhabi International Petroleum Exhibition & Conference. Society of Petroleum Engineers.

[26] Saleem, H., Al-Shammari, R., Al-Mai, N., Robert, H., Saxena, M., Joy, S., ... & Ismail, M. (2015, October). Kuwait Digital Integrated Field-Enabling Virtualization technology for a step Change. In SPE Kuwait Oil and Gas Show and Conference. Society of Petroleum Engineers.

[27] Al-Jasmi, A. K., Goel, H. K., Nasr, H., Querales, M., Rebeschini, J., Villamizar, M. A., ... & Saputelli, L. (2013, June). Short-term production prediction in real time using intelligent techniques. In EAGE Annual Conference & Exhibition incorporating SPE Europec. Society of Petroleum Engineers.

[28] Padhy, Patra, "Moving Towards San Storage: An Enterprise Perspective." http://www.rroij.com/open-access/moving-towards-san-storage-anenterprise-perspective-21-29.pdf.

[29] NetApp Storage Solutions http://www.netapp.com/us/products/storage-systems.

[30] Ashraf Al-Dabbas, 2013, "Real Hyper-V vs. VMware comparison: What you actually get for free".

[31] Contel Bradford,2013, "irtualization Wars: VMware vs. Hyper V: Which is Right For Your Virtual Environment?".

[32] Microsoft, "Hyper-V Architecture," [Online].Available: http://msdn.microsoft.com/enus /library/cc768520%28v=bts.10%29.aspx.

[33] Scott Lowe, Mastering VMware Vsphere 5, Publisher: Wiley India/Goels Computer Hut (2012).

[34] Fayyad-Kazan, H., Perneel, L., & Timmerman, M. (2013). Benchmarking the Performance of Microsoft Hyper-V server, VMware ESXi and Xen Hypervisors. Journal of Emerging Trends in Computing and Information Sciences.

[35] Hu, W., Hicks, A., Zhang, L., Dow, E. M., Soni, V., Jiang, H., ... & Matthews, J. N. (2013, August). A quantitative study of virtual machine live migration. In Proceedings of the 2013 ACM cloud and autonomic computing conference (p. 11). ACM.

[36] VMware, Inc.: Principled Technologies, Test report summary, Virtualization performance: VMware vSphere 5 vs. Microsoft Hyper-V R2 SP1. http://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/products/vsphere/vmware-vsphere-vshyper-v.pdf.