The Future of Data Storage: A Case Study with the Saudi Company

Azzah Al Ghamdi[1], Thomas Thomson[2]

[1]College of Computer Science and Information Technology, Computer Information Systems Department, Imam Abdalrhman Bin Faisal University, Dammam, Kingdom of Saudi Arabia

[2]School of Computer Science, Information Technology Department, University of Manchester, Manchester, United Kingdom

Email address:
azghamdi@iau.edu.sa (A. Al Ghamdi), Thomas.thomson@manchester.ac.uk (T. Thomson)

Abstract: The age of big data has emerged. These data are generated from online transactions, emails, posts, videos, search queries, etc. People also produce data by using the Internet of Things (IoT) applications and devices. Storing these massive quantities of data has become one of the most important and critical issues for big companies like Google, LinkedIn, Yahoo, and for the digital society in general. Traditional data storage methods such as Relational Database Management Systems (RDBMSs) are coming under increase pressure due to their capability limitations. However, many of new technical solutions have proved their efficiency in storing big data for large companies; some examples of these solutions include NetApp, Hadoop, SAN, the cloud, data centres, etc. The storage, accessibility, and security of big data issues are not only a computer science concern; it has become a topic of interest in many fields such as healthcare, E-commerce, and business in general. This project is investigating the data storage methods and future requirements for one of the largest oil companies in the world, Saudi SA Oil Company. A survey was carried out to understand current storage problems, and collect requirements to implement effective storage methodologies that overcome most of SA's storage difficulties.

Keywords: Big Data, SA Company, NetApp Storage, Two Surveys, Analysing Tools
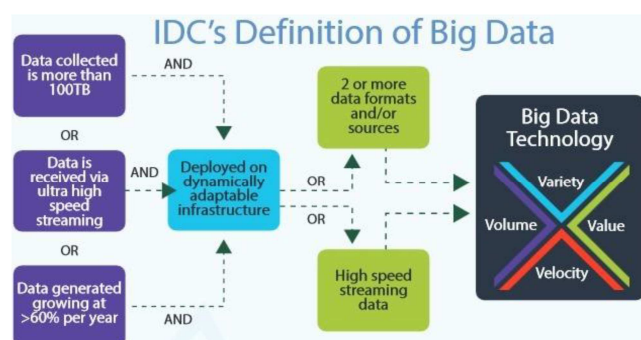
1. Introduction

Figure 1. Big data definition by IDC.

The term *big data* is used to describe the rapid and increasing growth of data. It can be expressed by the following 4Vs: Velocity, Volume, Variety and Value. A simple example which could be examined for a better understanding about the growth and generation of big data is Twitter, which is one of the most popular social networking websites. Of interest is how many people around the world might tweet at the same time, as well as how many hash tags people interact with at the same time as well, which likely represent enormous amounts of data. Figure 1 shows the definition of big data by IDC [4].

1.1. Contribution

a. To identify the best data storage method for the next years for one of the Saudi Arabian Oil companies, called SA Company. Cost, security, and speed are the most important criteria that should be considered to ensure the safest option for data storage [21].

b. To investigate the most critical problems the company faces related to their current storage methods. Analysing two surveys responses before (launching the main survey) and after (launching the follow up survey)

using the potential solution; which is the NetApp storage system in SA

c. To determine the company's future requirements by administering a survey to experts in SA asking about their knowledge, predictions, and expectations for the future of big data storage

d. To present a potential solution to optimize data storage;

e. To apply NetApp model to SA's IT infrastructure, observing the new results and finding outs to guarantee that it is a suitable data storage method for the IT future of SA

1.2. Paper Organization

The paper consists of eight sections as follows: the first section introduced the overview of the big data and the project contributions. The second section outlines some of related work in this area. Thus followed by the third section that illustrates different data collecting methods and analysing tools used for the project's two surveys (the main and follow up survey). Section four discusses an actual application of this project in a real business environment for a one of the largest oil company in Saudi called SA. The progresses are including a site visit to their IT data centre, and launching the main survey. Consequently; the fifth chapter shows the main survey results, and discovers more storage problems and requirements of SA. The sixth section presents the potential solution to be used in SA IT storage, which is NetApp storage system with some useful storage properties. The evaluation of the proposed solution will be explained in the seventh section in terms of including different evaluations strategies; such as a comparison between Sun and NetApp, launching a follow up survey. The paper will end with the conclusion section.

2. Related Work

There is an ongoing discussion about big data in large companies, international banks, and healthcare from various aspects; cost, time reducing, organizational structures, team members' skills, and big data analytics [6, 26]. More details about big data analytics in the real time were given by Barlow and Smith [1, 24]. A mention of big data challenges in IT departments is important in this project, in addition to guidance to IT members with essential roles [13, 25]. In terms of optimizing data storage method in large companies, Schlumberger, the oil and gas Exploration Company, and CATIA V5 Deployments have chosen their own data storage method to be performed by NetApp storage solutions to handle large quantities of data [7, 9]. Finally, different architectures have discovered to deal with big data from various aspects, such as storage, analyzing, mining, and backup and safety. These architectures use different storage techniques like: Hadoop, the cloud, data centers, SAN, NetApp [18, 19, 7, 10].

A further valuable source of data storage requirements emerged from a survey conducted by Intel in 2012. This survey aimed at two-hundred IT managers from some of largest organizations in the world, in order to ask them different questions related to handling and analyzing big data within their organizations. Overall, this survey presented important findings and studies for many other companies [11]. Those two-hundred IT managers were asked about the top three data sources within their companies via that survey. Their answers were business transactions, business documents, and emails, including semi-structured or unstructured data [11]. Figure 2 shows these sources percentages of big data for 200 IT managers from different large companies. Figure 3 also displays the percentages of big data standards according to Intel's survey [11].
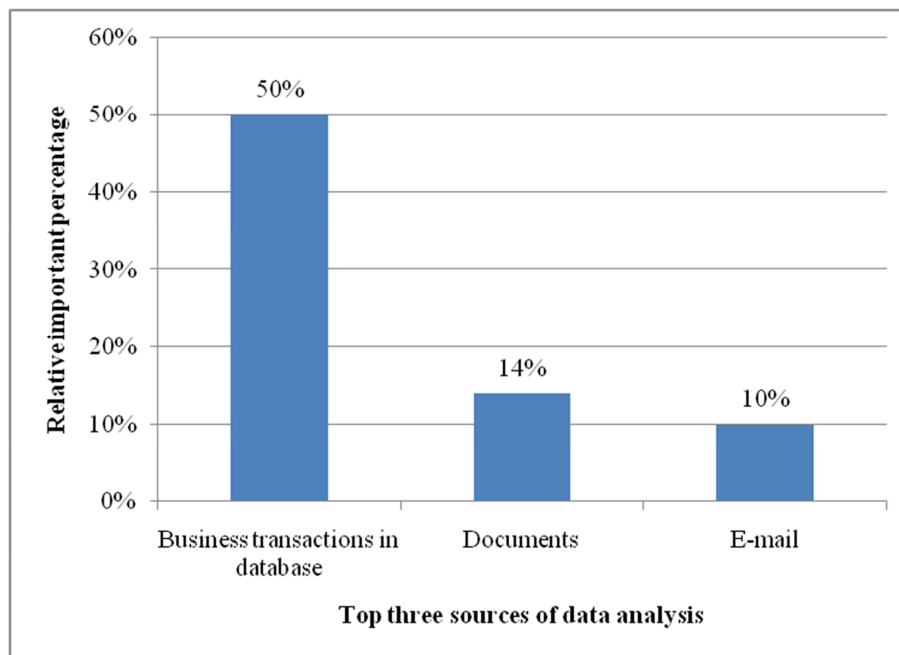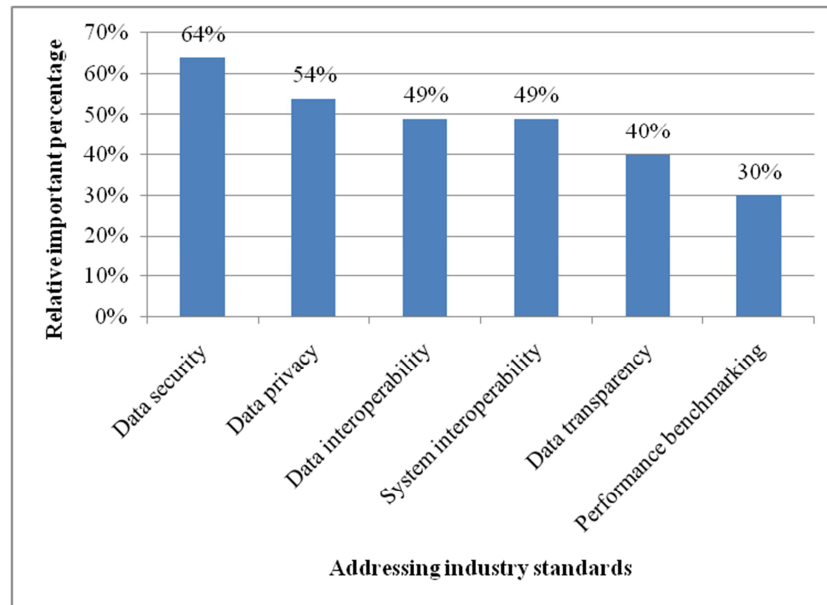

Figure 2. *Big data sources based on Intel's survey, for the question: "Which are the top three sources?" after [11].*

Figure 3. Big data analytics standards based on Intel's survey, for the question: "What standards for big data analytics are you most interested in having the industry address?" after [11].

3. Data Analysis Methodologies

This section focuses on different data collecting and analyzing methods and tools. These tools such as Google form, QuestionPro are used to analyse results of published surveys. The results of the project's two surveys are multiple choices and open text answers. The followings are a description for each of these different tools and data collecting methods.

3.1. Data Collecting Methods

Survey responses are stored in the database of the analyzing tool used for this project (QuestionPro). In addition to the ability to extract these responses into different file formats, such as Microsoft Word, Microsoft Excel, PDF and Text files, used for different purposes. Voice calls were not permitted to record (via Skype) with the SA storage team, due to confidentiality policies. However, useful information was extracted from these calls by taking detailed notes. This information was not included in the surveys, but it was used at different stages of the project, such as identifying storage problems and determining SA's future storage requirements.

3.2. Data Analyzing Tools

a. QuestionPro tool

The QuestionPro tool is an effective way to create and analyse surveys and polls, with many advanced features. These features are: the simplicity, usability, graphical user interface (GUI) support suitable for geographers, biologists and non-technical staff. In addition to using: (i) a secure HTTP throw secure sockets layer (SSL) channel (HTTPS), (ii) having a security authentication through HMAC-SHA1 (Single Sign-On); (iii) DES Encrypted Custom Variables.

This tool also provides the ability to save and continue later, putting a logo for a survey, setting the countdown timer, and customizing the URL of the survey with a meaningful name. The survey can be displayed either in interactive mode or none interactive mode. Interactive mode means all questions are appeared in multiple web pages. However, none interactive mode means all questions are appeared in a single web page. Regarding to importing and exporting data, this tool has the property of extracting responsive data from different file formats such as Word, PDF, Excel and CSV. Beside to supports a variety of question types (e.g. multiple choices, open text, rating, upload files option, QR code reader option, etc.) and languages as well.

b. NVivo

NVivo is a qualitative data analysis (QDA) computer software used for analysing qualitative data resulting from a conducted interview (multimedia information, e.g. audio files, videos, digital photos) or a published survey. A user can install a trial version of the NVivo for 60 days on his/her device. This tool imports data as a dataset from different file formats such as Word, PDF, spread sheets, EndNote and so on to deal with structured data such as tables. After importing the data, this software allows a user to perform very advanced analysis, association, queries, classify, sort, and clustering processes. However, the user should make a link between questions and answers after examining relationships in the data. In contrast to the first tool (QuestionPro), the linking between questions and answers had been performed automatically. QuestionPro and NVivo have the World Cloud property. This property is a visual representation for the text data. However, NVivo tool has also another option called Tree Map. This is a method for displaying hierarchical data by using nested rectangles.

c. Seobook free tool used with Microsoft Excel

This tool was developed to address the issue of the World Cloud property, which captures the whole input of words without any concern for the most significant words that are associated with the probabilities of answers for each question. For example, in main survey, the responses to an open text question, the question no 2 in particular; which is 'Can you estimate how often you reconsider your data storage systems in your organisation?', the word 'month' was in the largest proportion, even though it is not the targeted word to be considered. This is because the month is the considered period of time to be measured. Seobook offers a Keyword Density Analyser tool that is used to analyse an inserted text based on word frequencies. The output of this tool can be exported as.CSV file extensions. After that, the file can be opened by using Microsoft Excel, which contains three different sheets and organises the exported values from them into tables and charts. Finally, the generated charts are shown the frequencies for the common words after getting rid of unnecessary phrase frequencies (meaningless). This means a filtering phase happens to eliminate from duplicated words or correct mistaken ones. These responses' results will be presented in subsection 5.1 for the main survey and in subsection 7.3 for the follow up survey. Figure 4 views the flow of using these two tools together; Seobook and Excel [22].
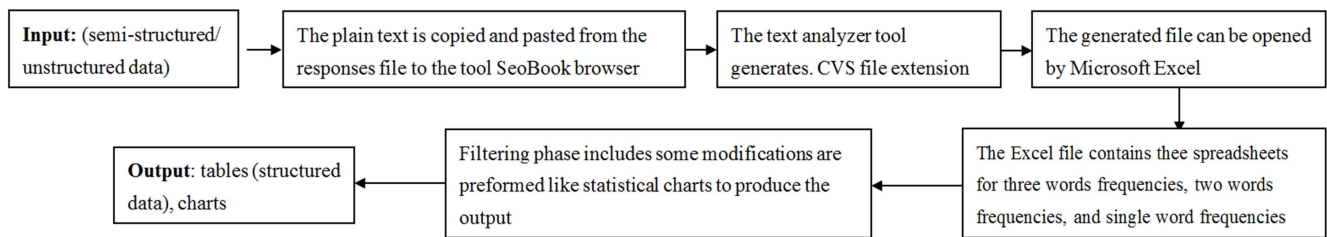

Figure 4. The combination processes between Seobook tool and Microsoft Excel.

4. Application of Big Data: Saudi Arabian Organization Case Study

This section is divided into two stages. The first stage explores the early communications with SA, starting from February and going to May 2015. These include an official visit to the SA IT data centre, and meeting the data storage team members in the SA IT department. The second stage covers the main survey creation and its progress. With a link to the previous work of this project that performed several months ago. Figure 5 provides an overview of the four important phases that have been covered in this project, starting from the analysis phase and ending with the evaluation phase.
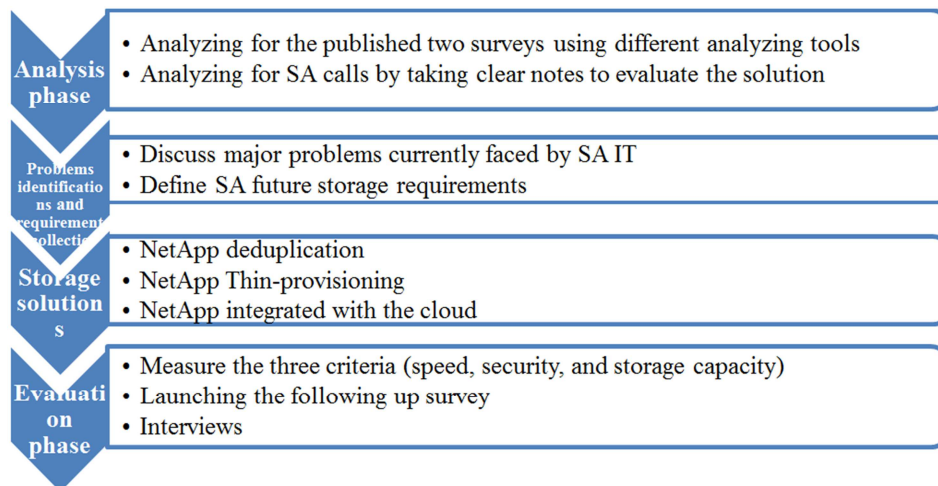

Figure 5. The prominent phases of the project implementation.

4.1. Initial Study of the SA Data Storage Environment

This project involves a real business environment that deals with huge amounts of data. The company in question is Saudi SA Oil Company, which is located in the Kingdom of Saudi Arabia. It is established in 1933. SA now is considered one of the largest oil-producing companies in the world with many branches and production lines [7, 20, 21]. Saudi SA Company has many different divisions dealing with multiple types and volumes of data. These divisions have different data storage methods based on diverse requirements and conditions. The focus of this project is on the data storage methods currently used in this company, and to decide which of them are most effective for SA's IT department in the future.

4.2. Main Survey Creation Using QuestionPro

The QuestionPro tool is used to create the two surveys (the main survey and the follow up survey) of this project. It is an excellent way to create surveys; due to the large number of the effective properties (had covered in section 3.2). The main survey was created by using the English language, because it is the common language among Saudi and non-Saudi staff in SA. This survey is consisted of eight questions, mostly multiple choices. Each multiple-choice question had four possible answers, rather than just three, to avoid any ambiguity between the answers, and not five choices, to avoid any similarity between the answers, which can lead to difficulty in the selection. The whole questions were not mandatory, because some of the staff have no experience regarding to some questions' answers. Additionally, the property of saving and continue later was not enabled. This is because the survey is short and it needs approximately five minutes to complete the survey filling. All the survey's questions are appeared in interactive mode. Figure 6 shows the main interface of the main survey including: the survey's logo, the header, and the first question. The survey link is: https://datastorageexperts.questionpro.com. The questions and answers of the main survey are shown in the Appendix 1. The main purposes of these questions were to study the current storage system in SA; including the major problems, suggest different ways to develop storage methods in this organization, and predict the future perspective of data storing in SA.

Figure 6. *The main interface of the main survey shows the first question, by clicking on the next button the second question will appear.*

5. The First Survey Results and In-Depth Study of Saudi Organization Data Storage

This section is divided into two subsections. The first subsection releases the results of the main survey, which its progress has been covered in the section four (the questions are in appendix 1). These results are analysed using the analysing tools, which covered in the section three. The second subsection discusses the in-depth study of the SA data storage environment which the early communications were covered in the previous section.

5.1. The Project's Main Survey Results

5.1.1. Results of the Multiple-Choice Questions

Q1's answer: A high percentage of the responses were for SAN, which indicates that SAN is the current data storage method used in SA.

Q3's answer: The proportion indicates that the current data storage system in SA is easy to work with, and performs most of the back-up and retrieves operations properly and within a minimal time.

Q4's answer: It appears that SA hires only qualified employees, especially when they are employed in the IT department. This is because they are going to work with sensitive data (such as oil information) and use very important systems.

Q5's answer: Most SA IT members believe that their storage systems are protected against cyber-attacks, with the exception of the virus attack against the SA system in August 2012 that affected about 30,000 workstations [3].

5.1.2. Results of the Open Text Questions

Q2's answer: For this question, 29 of the respondents saying that they reconsidered the storage systems on an average between every 3 and 4 months.

Q6's answer: The current responses indicate that the majority of the problems are easily solved within a short length of time, and most of them are related to a power failure or cashing.

Q7's answer: A large percentage of the responses suggested using: (i) a hybrid cloud computing model which combines between the advantages of the private cloud in terms of providing the required privacy and data confidentiality; and the public cloud in terms of the cost-efficiency, (ii) or moving to a NetApp storage system, as it is a recommended solution by Gartner. Some of the responses suggested making an improvement in the backup mechanism or increasing the storage capacities by using virtual machines.

Q8's answer: The last question was about the prediction of the future of the data storage. Most of the responses focused on the future of data storage in SA in particular and stated that SA is going to move to NetApp storage systems. This solution is recommended from the point of view of this project, as well as by Gartner, alongside which SA works for most of the company's IT issues.

5.2. In-Depth Study of the SA Organization Data Storage Environment

5.2.1. Feeding Back the Main Survey Results to SA

After the process of analysing and obtaining the first survey's results, SA was informed of the current findings, as is mentioned in the project's survey results in the previous subsection. However, further communications with SA were conducted to obtain more information about storage requirements and problems for this project, in order to provide a suitable solution to deal with these requirements and solve these problems.

5.2.2. Further Investigating the Storage Problems

The aim of this subsubsection is to extract further critical problems that the survey has not covered. A discussion of the current IT problems of SA was conducted via Skype calls and interviews, in which the major issues were considered on the basis of clear notes taken from the interviewees' comments. (For reasons of confidentiality, their voice calls were not able to be recorded). Additionally, further valuable resources have been considered in order ensuring the accuracy of the problems presenting, and supporting their arguments. Most of these problems are related to the increasing power consumption of the data centres, from which SA and most large companies have suffered. These problems can be summarised as follows [2]:

1- Consolidation of servers and storage: With a usage percentage of 50%, servers and cooling systems are the largest consumers of the energy coming into the data centre. Storage systems are the second-largest consumer with 27% of the power being consumed by direct-attached storage. For that reason, companies with large IT infrastructures are highly recommended to use network storage systems, such as NAS or SAN instead of DAS, in terms of power saving and increasing the number of available watts in data centers. In addition to increasing the number of disks instead of the number of servers to reduce the cost, simplify data management and enhance the storage performance.

2- The problem of drives with low capacities that are still used in SA IT.

3- The problem of unused storage spaces, which is a waste of power. When the storage administrator allocates storage space to a particular volume, two significant problems arise. First, the size of this space cannot be changed. Second, the storage space is used for only one application. Typically, the requester will ask for more storage space since he/she could not predict the actual space size needs of his/her application. This will lead to asking for more storage space than is actually needed, which may lead to some storage space going unused.

4- Backup issues that 'do more with less': Backup issues are one of the most significant problems faced by SA. Most of these problems are associated with the efficiency of backup systems, such as backup speed, duplication data during the backup, etc.

5- The problem of increasing the risk percentage of data loss and downtime by the current type of RAID used in SA, which is RAID 5 (single-party RAID protection with SATA or FC disk drives). RAID 5 is most recommended for use within small organizations and non-mission-critical applications because RAID 5 cannot cope with double disk or media failures. However, if a large organization that deals with big data such as SA continues using RAID 5, it will face serious problems regarding data loss, downtime and financial issues in the long term.

5.2.3. Determining the Storage Requirements

This subsubsection shows the most significant data storage requirements that SA IT infrastructure should meet to gain the required performance level from the new storage system, which is NetApp. These requirements are captured via Skype calls, because they were not covered from the published survey. Further valuable resources have been considered in order ensuring the accuracy of the SA requirements presenting. SA requirements are summarised for performance, cost and capacity aspects [14, 23]. These include:

a. storage capacity to support big data applications and a massive number of large files; performance efficiency;

b. supporting heterogeneous environments by connecting different types of operating systems such as Linux, UNIX, Windows, and Mac through a SAN or NAS;

c. support of file sharing across various environments; supporting multiple data types;

d. data protection and data integrity, which are among the most important requirements for SA to ensure the data protection and integrity by using a checksum even for the backup copies. From an industry perspective, StoreNext is the standard version of SAN system for big data, with the capacity to deploy about 60,000 file system clients and 500 PB of data under license. StoreNext is selected because SAN is currently used in SA.

6. Potential Solution to the Saudi Organization Future Data Storage: NetApp

This section is divided into two subsections. This first subsection shows the installation process of the NetApp in the SA infrastructure. The second subsection illustrates different solutions provided by the NetApp components for the discussed storage problems in the previous subsubsection 5.2.2.

6.1. Installing the NetApp in SA IT Infrastructure

Saudi SA Company is going to install the NetApp storage system upon the recommendation of this project progress and from the Gartner Company as well [8]. This company will benefit from the resulting huge system metrics and gain a great deal of support and help from the NetApp. The following steps that illustrate the installation process from the beginning are shown below:

a. the storage administrator of SA was contacted by the NetApp team through their website;
b. the NetApp team responded promptly to the SA request. The responses to the forth question of the follow up survey also said that (covered in the subsection 7.3.1).
c. after that, onsite surveys were conducted by NetApp engineers at SA's IT location and the data centre;
d. several meetings were conducted between both parties, in order to satisfy SA's requirements (mentioned in 5.2.3) and collect the systems information relevant to SA, such as SAP, Oracle, SQL, other applications, disk types either SAS or SATA, the protocols used such as Fiber Channel and iSCSI, etc.;
e. in case of hardware damage, a software failure or maintenance, and any upgrading issues, NetApp is providing a maintenance contract for several years to SA IT.
f. further, the responses to the fifth question of the follow up survey believe this (covered in the subsection 7.3.1);
g. finally, this project took approximately from six to seven weeks for visits to SA's data centre, implementation, ordering of hardware, installing, deployment, testing, etc.

6.2. Solving the Discussed Problems by NetApp Technologies

1- For the first problem, that was the consolidation of servers and storage: one NetApp storage system can perform and support all capabilities and advantages rather than using hundreds of servers supported by direct-attached storage systems.

2- The second problem identified that was using low capacities drives: in terms of reducing the power consumption by half (50% less power per TB), using SATA disk drives with a higher capacity is preferable to using an equivalent capacity of fibre channel drives. Other benefits can be gained from SATA disk drives, such as providing the highest available storage density per drive, thus making SATA the first choice for NetApp and for other large companies rather than other drives.

3- Increased utilisation is the solution for the third problem: NetApp has provided a very elegant solution for the problem of unused storage spaces, which are wasteful of power. This solution is called NetApp FlexVol®. By using NetApp FlexVol®, a thin provision technique enables the storage administrator to resize the storage spaces based on the actual size of each application and return the unused storage space to a storage pool to be used for another application. This property will increase storage utilisation by 60% and reduce a corresponding number of required disks, in addition to enhancing the performance level of the storage system.

4- The fourth problem identified the backup issue; NetApp also has a very effective solution, called NetApp Snapshot™ [2, 7]. With this technology, copies are created within seconds with two significant advantages. Firstly, Snapshot copies consume less storage space, since they save only changes of data. Secondly, these copies will enable a user to identify which single copy of his/her data is used for multiple uses. The NetApp backup system is not only used for backup purposes. It used also for compliance and disaster recovery purposes, instead of each having an independent system, with different storage system requirements and specifications, leading to further huge consumptions of power.

5- To solve the last problem, that was the increasing the percentage of data loss: with NetApp RAID-DP (double-party, RAID 6 implementation), the percentage of data loss will be decreased as well as any related costs reduced, providing better performance and without ever compromising data integrity.

7. Evaluation Methodologies for the Potential Solution

This section concentrates on the evaluation methodologies used after installing the NetApp in the SA IT infrastructure. The evaluation methodology is measured by comparing the previous storage methods that SA was using, such as the Sun storage with the new solution, which is the NetApp storage system. In addition to ask the targeted individuals, the storage team members, about the storage system performance and functionalities after the installation of NetApp by quick interviews via Skype. Finally, publishing the follow up survey to receive the feedback about the NetApp performance.

7.1. The First Evaluation Method: Comparison Table

Table 1 shows this comparison between Sun and NetApp storage systems.

Table 1. *A Comparison Between Sun and the NetApp Storage Mechanisms in Terms of the Most Significant Storage Factors.*

	Security concern	Backup/retrieval speed (SFP)	Storage capacity saving	Cost per year (Saudi Riyal)
Sun Storage method	No Encryption (Clarke, 2009)	Slow as they were using 4G small form-factor pluggable (SFPs) (Lillibridge et al., 2013).	No deduplication/ Compression	The contract cost was 1.5 million (SR) every year without the maintenance included (additional fees for the maintenance)
NetApp storage system solution	Encryption Available using NetApp SnapVault	Fast as they are using 8G SFPs	Good saving as they have the deduplication/ Compression feature	The total amount of support is two million (SR) for five years (0.4 million in each year) plus the hardware and software support and maintenance

7.2. The Second Evaluation Method: Interviews

This method is based on receiving feedback from the storage administrator and technical support team members after deploying the NetApp storage system as follows:

'The previous storage that we were using was very slow and had much less storage space. Also the space was not consumed elegantly as it had no feature of compression/deduplication. We also were in danger of security as it had no encryption feature as well.' Sameer Khan, Data storage Administrator. One of the storage team members said: 'NetApp is a great solution to have; it is very easy to manage and very reliable.'

7.3. The Third Evaluation Method: Launching the Follow up Survey

Beside the main survey of this project, a follow up survey was launched to determine how well the NetApp solution met the SA storage requirements. In addition to measuring the degree of satisfaction used the same group of the storage team-members in SA. This group was asked about the NetApp storage performance and the quality of storage operations. These operations included the speed of data backup and retrieval, data security and an overall impression of the NetApp system features and functionalities. Finally, the results of this survey will be considered to decide whether the NetApp solution will be used for the future of SA or not. Figure 7 shows the main interface of the follow up survey including; the follow up survey's logo, the header, and the first question. The survey link is: https://followup.questionpro.com. This survey's questions and answers are shown in the Appendix 2.

Figure 7. *The Main Interface of the Follow up Survey Including the First Question, with a Property of Exiting the Survey.*

7.3.1. Follow up Survey Results

Q1's answer: The large proportion of the response confirms that the storage system's performance after using the NetApp is better than the previous Sun storage system.

Q2's answer: More than half of the respondents said that

problems occurred once a month, which is considered a long time between occurrences of a problem, indicating to the efficiency of the NetApp.

Q3's answer: No significant problems were encountered after deploying the NetApp storage system. The problems of power failure and caching have been hidden after deploying the NetApp.

Q4's answer: The majority of the storage team members who work in the SA data storage department praised the NetApp's quick response to the company's problems. None of them said that the NetApp is not supportive.

Q5's answer: Most of the respondents said that the SA contract with NetApp Company included the hardware maintenance and software support for any problems or technical issues.

Q6's answer: Regarding the use of NetApp, most of the respondents believe that they are going to use this solution for the future, which confirms the validity of this solution to be used for the future storage of this organisation. These question responses match with the eighth question responses in the main survey, when this group asked about the future of the data storage. Most of the responses said the NetApp is one of the expectations.

Q7's answer: After analysis of the storage team members' responses to this question, they have no additional comment about NetApp, although two of the responses said 'It is a well-established program.' which indicate to the staff satisfaction by using this solution.

7.3.2. Result after Installing NetApp

This subsubsection shows the efficiency of the NetApp components. Table 2 summarises the most significant NetApp components that cope with most of storage challenges and make NetApp an excellent choice for many large companies.

Table 2. A Summary of NetApp Effective Solutions and Features for Most IT Challenges in SA after [15, 16, 17].

Challenges	NetApp-provided solutions
1. Data warehouse availability	NetApp storage system is able to maintain the availability of hardware and software without the additional costs of traditional storage solutions (Sun storage system).
2. Long time for backup and restoration of data	NetApp SnapShot and SnapVault can easily and quickly store and restore huge amounts of data as compared to tape-based solutions.
3. Disaster and Recovery is always complicated and expensive	NetApp SnapMirror is a technique to perform disaster and recovery operations more efficiently with less cost and time.
4. Minimize production impact	NetApp SnapMirror and FlexClone® techniques are easy solutions to minimise the impact of production systems.
5. Keeping an organisation's data protected when upgrading storage drives	NetApp Storage Encryption (NSE) provides a very effective solution to protect data with the property of Self-encryption drives, which prevents unauthorised users from accessing data.

8. Conclusion

Based on the evaluations of this project's proposed solution (were covered in the previous chapter), NetApp is the recommended solution to be used in the SA future storage infrastructure. Moreover, most of NetApp's features are optimised by the NetApp Company, in terms of high performance, reducing the time and cost, etc. [7]. To conclude, new applications and forms of modelling that are used in oil and gas companies are creating the amounts of data that need better data storage and management solutions. For example, NetApp FAS systems provide a comprehensive solution that ensures the highest levels of performance and data protection. Accordingly, most organisations choose NetApp for the following reasons [7]:

a. industry-leading functionality;
b. seamless application integration;
c. replication from all-flash to hybrid systems;
d. warranty and customer experience; and
e. a comprehensive portfolio provides multiple options.

Acknowledgements

Appendix

Appendix 1

The main survey questions and answers. This survey launched on 29th March 2015 for the storage team-members at SA. The purpose of this survey was in order to obtain a general overview about the current storage system used in SA infrastructure.

Q1. In your organisation, what methods of storing data are currently used?

A1. [Hadoop, SAN, NAS, DAS, other (please specify)]

Q2. Can you estimate how often you reconsider your data storage systems in your organisation?

A2. [Open text answer]

Q3. How easy is it to retrieve and back up the data within your storage systems?

A3. [Very easy, Easy, Not very easy, Complicated]

Q4. Are data storage team members well-trained and qualified before they start dealing with these storage systems?

A4. [Yes, they are; Sometimes; No, they are not; No idea]

Q5. In your opinion, is the storage system in your

organisation protected against cyber-attacks?

A5. [Yes, well protected; Yes, protected to some extent; Basic protection; I do not know]

Q6. Have you ever faced a serious problem within the data storage system in your organisation? If yes, please specify.

A6. [Open text answer]

Q7. Briefly, can you suggest two possible ways to improve the data storage provision in your organisation?

A7. [Open text answer]

Q8. From your point of view, what is the long-term future of data storage?

A8. [Open text answer]

Appendix 2

The follow up survey questions and answers. This survey launched on 25th June 2015 for the storage team-members at SA. The purpose of this survey was in order to obtain the feedback and a general impression after using the NetApp in SA IT infrastructure.

Q1. How did the system perform after the installing NetApp framework?

A1. [Better than before, same as before, worse than before, no idea]

Q2. How often have you faced problems with NetApp storage system?

A2. [Once a day, Once a week, Once a month, no idea]

Q3. Describe the type of problems that you encounter?

A3. [Open text answer]

Q4. How responsive has the NetApp Company been?

A4. [They were very supportive and response to us quickly, They were supportive but they took long time to response to us, They were not supportive and did not response to our queries, no idea]

Q5. Did the contract with NetApp including hardware maintenance and software support?

A5. [Yes, it was including the hardware maintenance and software support, It was including only the hardware maintenance, It was including only the software support, None of them]

Q6. Do you think that you are going to use NetApp after five years?

A6. [Yes, maybe yes, maybe no, no]

Q7. Do you have any addition comment concerning your experiences with NetApp?

A7. [Open text answer]

References

[1]   Barlow, M., (2013) *Real-Time Big Data Analytics: Emerging Architecture*. 1st edn. Sebastopol: O'Reilly Media. [Online] Available at: http://www.pentaho.com/assets/pdf/CqPxTROXtCpfoLrUi4Bj .pdf. (Accessed: 30 January 2015).

[2]   Battles, B., Belleville, C., Grabau, S., and Maurier, J. (2007) *REDUCING DATA CENTER POWER CONSUMPTION THROUGH EFFICIENT STORAGE*. [Online] Available at: http://www.it-executive.nl/images/downloads/reducing-datacenter-power.pdf. (Accessed: 22 April 2015).

[3]   Bronk, C., and Tikk-Ringas, E. (2013) *The Cyber Attack on Saudi Aramco*. 55(2) pp. 81-96. [Online] Available at: http://www.tandfonline.com/doi/pdf/10.1080/00396338.2013. 784468. (Accessed: 11 June 2015).

[4]   Chen, M., Mao, S., Zhang, Y., and Leung, V. C. (2014*) 'Big Data Related Technologies, Challenges and Future Prospects'*. 1st edn. United Kingdom: Cham: Springer.

[5]   Clarke, R. (2009) *Overview of Storage and Data Management Industry Trends in Long Term Information Retention and Preservation.* [Online] Available at: http://web.stanford.edu/group/dlss/pasig/PASIG_September20 09_DC/GMW_OverviewofStorageandDataManagementTrend s.odp_0.pdf. (Accessed: 11 June 2015).

[6]   Davenport, T. and Dyche', J. (2013) *Big Data in Big Companies.* [Online] Available at: http://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf. (Accessed: 26 January 2015).

[7]   David, L., and Ledoux, T. (2009) Optimizing Data Storage and Management for Petrel Seismic Interpretation and Reservoir Modeling A White Paper for Upstream Oil and Gas Data Managers. *Schlumberger.* [Online] Available at: http://www.netapp.com/us/media/wp-optimizing-data-storage-for-petrel.pdf. (Accessed: 2 April 2015).

[8]   Gartner. (2015) [Online] Available at: http://www.gartner.com/technology/home.jsp. (Accessed: 31 March 2015).

[9]   Godbole, R., and Davied, M. (2006) *CATIA V5 Deployments with NetApp Solutions,* pp. 1-26. [Online] Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.127 .1458&rep=rep1&type=pdf. (Accessed: 19 June 2015).

[10] Greenberg, A., Hamilton, J., Jain, N., Kandula, S., Kim, C., Lahiri, P., Maltz, D., Patel, P., Sengupta, S. (2009) *VL2: A Scalable and Flexible Data Center Network,* Microsoft Research, pp. 51-62. [Online] Available at: http://delivery.acm.org/10.1145/1600000/1592576/p51-greenberg.pdf?ip=130.88.0.55&id=1592576&acc=PUBLIC& key=BF07A2EE685417C5%2EA30F01C2334F920C%2E4D4 702B0C3E38B35%2E4D4702B0C3E38B35&CFID=6496849 49&CFTOKEN=38709119&__acm__=1428331216_1a9b893 cbce57acf93d8be89a59f05fe. (Accessed: 6 April 2015).

[11] Intel IT Centre. (2012) *Peer Research Big Data Analytics Intel's IT Manager Survey on How Organizations Are Using Big Data.* [Online] Available at: http://www.intel.com/content/dam/www/public/us/en/docume nts/reports/data-insights-peer-research-report.pdf. (Accessed: 26 March 2015).

[12] Lillibridge, M., Eshghi, K., and Bhagwat, D. (2013) *Improving Restore Speed for Backup Systems that Use Inline Chunk-Based Deduplication,* USENIX Association 11th USENIX Conference on File and Storage Technologies (FAST '13) HP Labs, HP Storage. pp. 183-197. [Online] Available at: https://www.usenix.org/system/files/conference/fast13/fast13-final124.pdf. (Accessed: 5 June 2015).

[13] Mohammad, A., Mcheick, H., and Grant, E. (2014) *Big Data Architecture Evolution: 2014 and Beyond* [Online]. Available at: http://dl.acm.org/citation.cfm?doid=2656346.2656358. (Accessed: 27 February 2015).

[14] Moniruzzaman, A., and Hossain, S. (2013) *NoSQL Database: New Era of Databases for Big data Analytics –Classification, Characteristics and Comparison*, International Journal of Database Theory and Application. pp. 1-14, 6(4). [Online] Available at: http://arxiv.org/ftp/arxiv/papers/1307/1307.0191.pdf. (Accessed: 30 April 2015).

[15] NetApp. (2009) *Application Solution NetApp Data Management for Decision Support Systems* [Online] Available at: http://www.netapp.com/us/media/ds-2923.pdf. (Accessed: 22 May 2015).

[16] NetApp. (2011) *NetApp Storage Encryption (NSE) Full disk encryption that protects data at rest with no operational impact.* [Online] Available at: https://www.hastorage.com/files/NetApp_Sheet.pdf. (Accessed: 29 May 2015).

[17] NetApp. (2014) *Datasheet Clustered Data ONTAP Operating System Respond more quickly to business changes and new opportunities—on premises or in the cloud.* [Online] Available at: http://www.netapp.com/us/media/ds-3231.pdf. (Accessed: 5 June 2015).

[18] Poulton, N. (2014) *Data Storage Networking Real-World skills for the CompTIA Storage+™ Certification and Beyond.* 1st edn. Canada: John Wiley & Sons, Inc., Indianapolis, Indiana.

[19] SAGIROGLU, S., and SINANC, D. (2013) *Big Data: A Review*. pp. 42-47 [Online] Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6567202. (Accessed: 19 March 2015).

[20] Saudi Aramco. (2015a) *1930s.* [Online] Available at: http://www.Aramco.com/en/home/about/history/milestones/1930s.html (Accessed: 26 January 2015).

[21] Saudi Aramco. (2015b) *Our business.* [Online] Available at: http://www.saudiAramco.com/en/home/our-business.html (Accessed: 3 April 2015).

[22] SEOBOOK. (2003) KEYWORD-DENSITY ANALYZER. [Online] Available at: http://tools.seobook.com/general/keyword-density/. (Accessed: 27 May 2015).

[23] Slack, E. (2012) *Storage Infrastructures for Big Data Workflows*, Storage Switzerland White Paper. [Online] Available at: https://iq.quantum.com/exLink.asp?8615424OJ73H28I34127712. (Accessed: 29 April 2015).

[24] Smith, D. (2012) *Real-Time Predictive Analytics with Big Data from deployment to production,* Revolution Analytics. [Online] Available at: http://www.slideshare.net/RevolutionAnalytics/realtime-big-data-analytics-from-deployment-to-production. (Accessed: 15 April 2015).

[25] Villars, R., and Olofson, C. (2011) *Big Data: What It Is and Why You Should Care*, IDC Analyzed the future. [Online] Available at: http://www.emitac-ees.ae/wp-content/uploads/2014/04/IDC_AMD_Big_Data_Whitepaper.pdf. (Accessed: 3 March 2015).

[26] Yuan, D., Yang, Y., Liu, X., and Chen, J. (2010) *A Cost-Effective Strategy for Intermediate Data Storage in Scientific Cloud Workflow Systems*, pp. 1-12. [Online] Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5470453. (Accessed: 8 April 2015).