EFFICIENT DATA PROTECTION FOR ENTERPRISE DATA CENTERS

Khaled Aldossari

EXPEC Computer Center, Saudi Aramco, Saudi Arabia

ABSTRACT

This paper addresses the major challenges that large organizations face in protecting their valuable data. Some of these challenges include recovery objectives, data explosion, cost and the nature of data. The paper explores multiple methods of data protection at different storage levels. RAID disk arrays, snapshot technology, storage mirroring, and backup and archive strategies all are methods used by many large organizations to protect their data. Furthermore, efficiency technologies — including: data compression and deduplication, and tiered storage —used with modern data protection systems are explained. The paper proposes certain criteria that can be used to evaluate enterprise-level backup and archive solutions in the market today. The evaluation criteria cover all business needs and help to tackle the key issues related to data protection. Finally, this paper provides insight on data protection mechanisms and proposes guidelines that help organizations to choose the best backup and archive solutions.

KEYWORDS

Data Protection, Data Loss, Data Recovery, Backup, Archive

1. INTRODUCTION

In any organization, the requirement to store digital data has been growing exponentially year after year. To cope with this increasing data requirement, larger amounts of bigger and faster storage devices need to be installed in data centers around the world. The downside with having more hardware installed is that it also increases the chance of losing data due to user and hardware error or malfunction. Losing data can be costly for organizations both legally and financially. Below are some statistics that show the potential results from losing data:

- The cost associated with lost data for the energy business is $2.8 million of lost revenue per hour. [1]
- The cost of recreating just 20 MB of engineering data is 42 days and $98,000. [2]
- In less than a year after they faced a major data loss, 70 percent of small companies stop business permanently. [2]
- Among companies that lost data in 2012, only 33 percent were able to recover 100 percent of their data. [3]

To avoid such impacts, a successful data protection strategy has to keep data available and accessible against possible losses caused for any reason. In fact, data loss can happen for different reasons, including:

- Hardware or system malfunctions, such as power failure, media crash and controller failure
- Human errors, such as accidental deletion of files or physical damage caused by dropping storage devices

119

- ▪ Software corruption, including software bugs and software crashes while editing
- ▪ Computer viruses and malware
- ▪ Natural disasters, such as earthquakes, floods and fires

The chart below shows the percentage of data loss incidents due to each leading cause according to Kroll Ontrack Inc. [1]
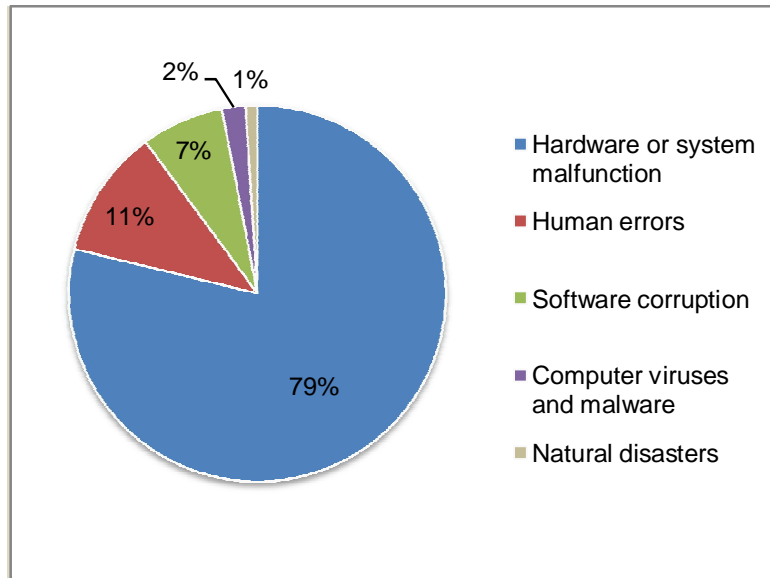

Figure 1. Causes of data loss

The next sections cover, in detail, the availability and recoverability aspects of data protection solutions at the enterprise level.

2. DATA RECOVERY CHALLENGES

Today, large organizations face challenges when they plan and implement data recovery solutions. These challenges could make the recovery of data in the event of data loss more difficult. As a result, it is important to understand and address these challenges before implementing a data protection strategy. Major challenges are listed and explained below:

2.1. Recovery Objectives

The recovery time objective (RTO) and recovery point objective (RPO) are two critical business concepts related to data recovery. RTO is the maximum time period by which the data must be restored after a data loss event. On the other hand, RPO is defined as a point in time prior to a data loss event where data can be restored. Each organization has its own RTO and RPO that should be defined clearly and carefully based on the business needs and regulatory compliance requirements. Meeting recovery objectives is one of the challenges that faces data protection solutions.

2.2. Data Explosion

As the volumes of data continue to grow exponentially, the scalability and performance of data protection solutions become significant challenges. The backup and restore system has to have

enough performance to meet the recovery objectives. It also has to be able to scale out to encounter future data expansion.

2.3. Cost

The cost associated with data protection is one of the greatest challenges. The total cost of ownership of the backup and recovery infrastructure and its operational cost are proportional to the amount of data to be protected. As a result, the data protection solution should be cost-effective while maintaining the recovery objectives.

2.4. Nature of Data

Enterprise organizations usually have different types of data that reside on heterogeneous systems. Structured data, such as databases, requires backup and recovery technologies that are different from those used with unstructured data, such as images and videos. The change rate of data affects the frequency of its backup. With equal total size, a large number of small files has more of an impact on the performance of backup and recovery systems than a small number of large files. Therefore, different types of data require different backup and recovery technologies, which make data protection a challenge

3. LEVELS OF DATA PROTECTION

To address the aforementioned challenges, enterprises implement data protection at different levels. Data protection technologies use the principle of redundancy to prevent a total loss of data by creating another copy of it [4].

The most common technologies of data protection are explained below.

3.1. RAID Disk Arrays

Redundant array of inexpensive/independent, disks (RAID) is a storage technology that combines multiple disks, known as a RAID set, and presents the disks as a single logical disk. [5] With this technique, data spread across the RAID set enhances reliability and increases I/O performance. [5] Different architectures of RAID exist to provide different levels of fault tolerance and performance. The most common levels of RAID include RAID 0, RAID 1, RAID 5, and RAID 10. [6]

RAID is perfect for protecting against hardware failures but cannot protect against other types of risks such as human errors, software corruption, malware, virus attacks and natural disasters.

3.2. Snapshot Technology

Snapshot technology is an instance copy of a defined collection of data—a set of files, directories, or volumes—at a particular point in time. [7] Apparently, a snapshot provides another level of redundancy. Different storage vendors offer various implementations of snapshot. The most popular snapshot implementations are copy-on-write, redirect-on-write, split-mirror, log structure file architecture, copy-on-write with background copy and continuous data protection. [8]

Snapshots provide protestation against human errors, software corruption, and malware and virus attacks but still cannot protect against natural disasters.

3.3. Storage Mirroring (Replication)

As mentioned above, RAID and snapshot cannot protect data against natural disasters. Here is where storage mirroring, also known as replication, comes in handy. Today, many storage manufacturers offer replication solutions, which can be used for disaster recovery. Typically, replication solutions copy data from the primary storage system over long distances to another storage system. [4] With this technique, an up-to-date copy of the data is maintained at a remote site in case the primary copy is lost. Replication solutions operate in either synchronous or asynchronous mode. With synchronous replication, data is written to the primary and secondary storage systems at the same time. [9] Asynchronous replication, on the other hand, writes data to the secondary storage system with a delay. [9]

3.4 Backup and Archive Strategies

Although storage replication enables business continuity in case one site is lost, it introduces extra overheads on writing. Also, it requires more disk space, which needs more power and space. Most enterprises implement backing up and archiving to cheaper media, mainly tapes, to keep data protected and, at the same time, lower associated cost.

The backup process creates a redundant copy of the original data on a different location, or storage media, for the purpose of recovery in case of data loss. The most common types of backup are full and incremental backup. With full backup, the entire file system is copied to the backup destination. It allows fast recovery of the file system in case it becomes inaccessible. Backing up the entire file system is slow and requires more backup media. Incremental backup, on the other hand, provides faster backup with less capacity by copying only files that are created or modified since the last backup. [10]

Archiving is the process of moving a selected collection of data, usually inactive, to another storage system for long-term retention. Archives are kept for long periods of time to meet regulatory compliance and/or for future reference.

Backup and archive play a major role in almost every data protection plan. Today, most backup systems also provide archive capabilities. Large organizations typically use low-price disks or tapes as the destination storage system for their backups and archives to reduce the cost. It has always been a good practice to send the backup/archive media to a remote place as a part of the disaster recovery plan.

The table below summarizes the possible data recovery methods to protect against each type of data loss threat:

Table 1. Recommended Data Recovery Methods

Data Loss Cause	Data Recovery Method
Hardware or system malfunctions	RAID, storage mirroring (replication), backup and archive
Human errors	Snapshots, backup and archive
Software corruption	Snapshots, backup and archive
Computer viruses and malware	Snapshots, backup and archive
Natural disasters	Storage mirroring (replication), backup and archive

4. EFFICIENCY TECHNOLOGIES FOR DATA PROTECTION

Now, as we have all these methods that ensure data protection, how to improve their efficiency and reduce associated costs? Modern data protection strategies use different technologies to achieve such objectives. These technologies include data compression and deduplication, and tiered storage. These technologies are explained in more details in the following sections.

4.1. Data Compression and Deduplication:

Data compression is a way of encoding data using less data than the original, to reduce its size. [15] Another key point to mention here is that data compression is implemented at the binary level within a block. [16] In general, there are two types of data compression: lossy compression and lossless compression. In lossy compression, losing a portion of the data is acceptable and it is used for certain types of applications such as video streaming and digital images. Lossless compression ensures that the entire content of the original data is preserved. [15] For data protection systems, the only accepted type of compression is lossless compression.

Many backup and archive solutions provide options to enable data compression at the hardware and the software levels. In hardware compression, the I/O devices, including tape drives, use special hardware to implement the compression algorithm. In software compression, the backup server or client perform the compression on data before writing it to the backup storage. In comparison, hardware compression is much faster than the software counterpart but it costs more. Enabling both hardware and software compression is not recommended since it might cause the size of the compressed data to be larger than the original one. Instead, enable only the hardware compression if it is supported by the backup storage device. [17] Depending on the type of data and the used compression algorithm, the compression ratio can approximately range from 1.2:1 to as high as 10:1. [18]

Data deduplication is another technology used in data reduction. In deduplication, multiple copies of data are replaced with pointers to a unique instance of the data to save storage space, bandwidth, or both. [19] Different implementations of deduplication are used to remove duplicate data across files or blocks within files. Data deduplication is available as hardware appliance such as NAS appliances and virtual tape libraries, VTL. It is also available as a feature in many backup software packages. [18] Furthermore, deduplication process can be performed before writing the data into the storage, called inline deduplication, or after the data has been initially stored, called post deduplication. Inline deduplication runs on the fly, which introduces some processing overhead and potential latency. Modern storage controllers are shipped with faster processors to help mitigate the negative performance impact of inline deduplication. Post-deduplication can be scheduled during off hours to avoid the processing overhead impact involved in the inline deduplication. The drawback of post-deduplication is that it requires extra space capacity, more than the inline deduplication to store the undeduplicated data. [20]

Deduplication algorithms require high speed random access to reconstitute the deduplicated data within an accepted time period. Because of the sequential access nature of tape, applying deduplication to tape is not recommended. [21] Generally, the aggregate ratio of deduplication falls in the range between 2:1 to 80:1 depending on the type of data and which implementation is being used. [18]

Overall, choosing the best approach to data compression and/or deduplication depends mainly on the business needs, the budget, backup policies, and provided features by that approach. It is also important to consider the compatibility with the current infrastructure including backup software and hardware such as disk array, VTL, and tape drives. Therefore, before implementing any

approach of data reduction technology, it is always recommended to test it first and compare it to what is currently in place. This approach will ensure that the chosen solution is efficient and has no negative impact on the existing environment.

In the end, modern data protection solutions benefits from the data reduction involved within compression and deduplication in different areas. These technologies offer organizations choices to reduce the required storage capacity, which consequently saves on cost as well as space inside data centers. Moreover, these technologies significantly improve the backup and recovery performance and hence better RTO and RPO can be achieved. Compression and deduplication are also used to reduce the amount of data to be transmitted, which help organizations to implement storage replication over an existing network more effectively.

4.2. Tiered Storage:

Different applications require different levels of performance, availability, and data retention. Treating all data the same can increase the risk of losing critical data, raise the total cost of their protection, or degrade the overall performance. Hence, many organizations are implementing the concept of tiered storage. [22] In tiered storage architecture, storage systems are classified into distinct categories, based on price, performance, or other attributes. [23] This concept is adopted by modern data protection solutions to increase efficiency, improve scalability, and meet regulatory compliance requirements with affordable prices. [24] High priority or mission-critical data can be backed up more frequently to faster, but more expansive, data storage systems. Less priority or rarely used data can be backed up less frequently to slower, but much cheaper, data storage systems.

To secure successful implementation of tiered storage architecture, a strategic, step-by-step plan should developed before apply it in any organization. First, all data in the organization should be classified based on business needs, operating environments, and regulatory compliance requirements. Business needs include its criticality and required level of availability. Operating environments involve users, groups and privileges. Regulatory compliance requirements include data retention and the location of backed up or archived data. Depending on its classification of data, the organization can maintain different of storage tiers. Although creating more tiers can provide more control, it introduces more complexity. So, deciding on the number of tiers should be studied wisely. The second step is to determine the storage needs that meet each tier's requirements with low costs. Specifications of storage systems, onsite or remote locations, and backup frequencies are factors in choosing the type of storage system for each tier. The ultimate goal is to backup or archive each type of data to the lowest-cost storage system. The last step in the plan is to test the plan to see if it is efficient enough and can meet the business requirements while saving the organization money. [25]

Generally, backup and archive systems intelligently combine disk arrays and tape media creating tiered storage. Disk-based backup and archive leverage the high I/O speed of disks to shorten the RTO and RPO. Usually, disk storage systems are used as staging to ensure backup can finish within its window. On the contrary, tape-based backup and archive exploit the low prices of tape media to store larger amount of backup and archive data for longer time with less cost. The intelligent combination of disk and tape helps enterprises to get the best of both worlds. Moreover, most backup and archive systems can automatically move data across different storage tiers, which reduces administrative complexity.

A very important point to remember while planning for any tiered storage implementation is to consider the data protection requirements, not just for today, but also for the next few years. As

technologies evolve over time, the storage architecture for each tier should be reevaluated to meet the minimum data protection requirements more efficiently.

5. EVALUATING BACKUP AND ARCHIVE SOLUTIONS

Backup and archive strategies are the most common data recovery methods used by most enterprises. Therefore, in this section, we propose different criteria to help enterprises choose backup and archive solutions that meet their requirements. These criteria were used by the EXPEC Computer Center at Saudi Aramco Oil Company to evaluate some of the most popular backup and archive systems available in the market.

We identified 19 criteria that cover every aspect of data backup and recovery. The importance of each criterion varies from one organization to another depending on its requirements. These criteria are:

5.1. Architecture

Different backup and archive solutions have different architecture. Certainly, the system architecture has an impact on its performance, scalability, reliability and other features.

In one-tier architecture, all components of the system exist on a single server. The advantage of using a one-tier architecture backup and archive system is its simplicity and ease of management. The scalability and overall performance of this type of architecture does not help large organizations meet their backup and archive requirements. Two-tier architecture consists mainly of clients and a server residing on different hosts. In this type of backup and archive system, the client moves backup or archive data to the server and the server only keeps track of metadata. Client-server systems have the advantage of flexibility and can provide better performance, yet they have the single server bottleneck. Three-tier architecture, on the other hand, involves clients, a server and data movers. Clients send their backup or archive data to the data movers, which move them to the backup storage. The role of the server is just to monitor the whole backup and archive environment and to execute some administrative tasks. This architecture delivers better performance and more scalability but it might increase the complexity of the solution. Sharing resources like tape drives or the host memory in three-tier architecture requires more effort.

5.2. Scalability

The backup and archive solution has to be scalable enough to meet the continuing growth of data. As mentioned before, the architecture of the solution has a direct impact on its scalability. In addition, the internal structure of the software and its associated database affects solution scalability.

5.3. Reliability

Is the solution highly available with no single point of failure? By looking into each component of the solution, you can pinpoint the possible cause of failure and, therefore, service disruption.

5.4. Performance

The overall performance of the solution should be sufficient to cover the RTO and RPO requirements. Two measurements can be used to evaluate the performance of the system: backup or archive speed (TB/hr) and restore speed (TB/hr).

5.5. Storage Efficiency

Technologies such as data compression, deduplication, and tiered storage can improve the storage efficiency and help organizations to reduce the total cost of their data protection strategy. Supporting these technologies should be considered when large organizations want to evaluate any backup or archive solution.

5.6. Supported operating systems

Each organization has its own preferable operating system. Therefore, it is important to know operating system platforms for which backup and archive solutions support every component.

5.7. Simplicity

The backup and archive solution should be easy enough to implement and to manage. Losing data by itself is troublesome — the data recovery should not be.

5.8. Security

Data security is a critical feature in any backup and archive system. The system shall provide different levels of user access control such as administrator, operator and users. In some cases, integrating the operating system permissions and ownership with the backup system is necessary.

5.9. Intelligence

Is the backup and archive system smart enough to fix damaged data in its backup storage? Is it capable of performing some data analysis and share the result with the administrator?

5.10. Data Policy Management

Depending on the business requirements, particular data policies are needed. Examples of these policies include flexible data retention policy per dataset, ability to extend data retention on the fly, automatic data expiration process, automatic data or tape media replication feature for disaster recovery, and automatic media transcription.

5.11. Open Standards

Does the backup and archive solution support open standards? More specifically, does the system write data to tape media in open format readable by other applications? Can the system export and import data to/from an open format such as LTFS? Is the system capable of rebuilding the system from tape without any additional outside information?

5.12. Metadata Search Engine Capability

In many cases, especially for archives, a fast and reliable search engine capability is very helpful. Users can navigate through the command line interface (CLI) or Web interface to identify and retrieve any achieved data by searching one or more key fields in the database such as dataset name, size, tags and age.

5.13. Tape Vault Management

Most large organizations use tape as their backup storage media. Usually, the organization maintains a large number of tape media that exceed the automated tape library capacity. As a result, the backup and archive solution has to be able to vault tapes for disaster recovery, and to track and report data on these tapes. Also, it has to track tape media outside the library and notify the operator console when there is a need to insert a tape into the tape library.

5.14. Tape Operator Console

Besides the previous requirement, the solution shall provide a centralized tape operator console. The console can be a Web or GUI-based interface that provides real-time monitoring of solutions. The console should present helpful information to the operator such as the health status of the tape library, tape drives, online tape media, capacity (online/vaulted), tape drive activities (busy/idle), errors and alerts, and a list of tapes on the shelf. It also has to show actions waiting for operator input such as inserting tapes.

5.15. Reporting

One of the important features of any enterprise solution is its reporting capability. In the backup and archive environment, the system shall be able to report the status of backup and archive jobs. It also has to generate reports about the health status of its storage systems, servers and clients. Customized reporting on data utilization per user group or class will also be useful.

5.16. Support Services

Regardless of the level of expertise an organization has, the vendor should support the solution. In addition, having official documentation is necessary for any proposed solution. Large community support can also help system administrators to resolve related problems and come up with new ideas. Official training is required to build the administrator's skills and expertise for the solution.

5.17. System Popularity

Some enterprises look for a solution that is more popular and used by a wide range of businesses. It provides an enterprise with more confidence and allows it to find needed resources easily and cost-effectively.

5.18. Other Features

Depending on business needs, different organizations might require other features. In some cases, supporting multiple automated tape libraries within the system domain is needed. Limitations on the number of files to manage or the maximum size of a single file might impact the selection of the system. Spanning a single large file over two or more tape media is important to consider if the environment has very large files.

5.19. Total Cost of Ownership

The capital and operational expanses of the backup and archive system are important factors when evaluating different solutions. It is also important to consider the license model of the solution (per TB, number of servers, number of hosts to backup or others), which indeed affects

its cost in the long term. The key point here is to choose the most cost-effective and affordable solution that meets the minimum backup and archive requirements.

6. SUMMARY

The value of digital data in any organization has increased. Without successful data protection strategies, data loss can be costly to an organization. We discussed challenges that face data recovery. These challenges include recovery objectives, data explosion, associated cos, and the nature of data. Then, we reviewed the different technologies used by most enterprises to overcome these challenges at different levels. Such technologies are RAID disk arrays, snapshot technology, storage replication, and backup and archive strategies. In addition, we explained different efficiency technologies used to improve the overall performance and reduce the total cost of the aforementioned data protection solutions. Because backup and archive systems are used by most enterprises, we focused on this method. We proposed 19 criteria that cover every aspect of backup and archive systems. These criteria can be used by any organization as a template when weighing different backup and archive solutions on the market.

ACKNOWLEDGEMENTS

REFERENCES

[1] Kroll Ontrack, 'Understanding Data Loss'. [Online]. Available: http://www.ontrackdatarecovery.com.au/understanding-data-loss/. [Accessed: 23- Sep- 2014].

[2] M. Foster, 'Save your business with data backup', *NetSource Technologies*. [Online]. Available: http://www.netsourceinc.com/blog/save-your-business-with-data-backup. [Accessed: 23- Sep- 2014].

[3] Kroll Ontrack, 'Kroll Ontrack study reveals 40 percent of companies lose data annually from their virtual environments', 2013. [Online]. Available: http://www.krollontrack.com/company/news-releases/?getPressRelease=62077. [Accessed: 23- Sep- 2014].

[4] C. Chang, 'A Survey of Data Protection Technologies', *2005 IEEE International Conference on Electro Information Technology*, p. 6, 2005.

[5] M. Dutch, *A Data Protection Taxonomy*. The Storage Networking Industry Association, 2010, p. 20.

[6] R. Natarajan, 'RAID 0, RAID 1, RAID 5, RAID 10 Explained with Diagrams', *The Geek Stuff*, 2010. [Online]. Available: http://www.thegeekstuff.com/2010/08/raid-levels-tutorial/. [Accessed: 23- Sep- 2014].

[7] M. Staimer, 'Backup in a snap: A guide to snapshot technologies', *Storage Technology Magazine*, 2009. [Online]. Available: http://searchstorage.techtarget.com/magazineContent/Backup-in-a-snap-A-guide-to-snapshot-technologies. [Accessed: 23- Sep- 2014].

[8] StoneFly, 'Exploring Storage Snapshot technology'. [Online]. Available: http://www.iscsi.com/resources/Storage-Snapshot-Technology.asp. [Accessed: 23- Sep- 2014].

[9] D. Bradbury, 'Remote replication: Comparing data replication methods', *ComputerWeekly*, 2011. [Online]. Available: http://www.computerweekly.com/feature/Remote-replication-Comparing-data-replication-methods. [Accessed: 23- Sep- 2014].

[10] A. Chervenak, V. Vellanki and Z. Kurmas, 'Protecting file systems: A survey of backup techniques', in *Joint NASA and IEEE Mass Storage Conference*, 1998.

[11] P. Dorion, 'Backup vs. archive', *Search Data Backup*, 2008.[Online]. Available: http://searchdatabackup.techtarget.com/tip/Backup-vs-archive. [Accessed: 23- Sep- 2014].

[12] H. Garcia-Molina, C. Polyzois and R. Hagmann, in *Compcon Spring '90. Intellectual Leverage.Digest of Papers. Thirty-Fifth IEEE Computer Society International Conference*, 1990, pp. 573-577.

[13] L. Black, 'The Importance of Data Backup', *The Livingston Business Journal*, 2014. [Online]. Available: http://www.livingstonbusiness.com/2014/07/20/the-importance-of-data-backup/. [Accessed: 23- Sep- 2014].

[14] Software Testing Class, 'What is Difference Between Two-Tier and Three-Tier Architecture?', 2013. [Online]. Available: http://www.softwaretestingclass.com/what-is-difference-between-two-tier-and-three-tier-architecture/. [Accessed: 23- Sep- 2014].

[15] The 2014 SNIA Dictionary, 'Data Compression', 2014. [Online]. Available: http://www.snia.org/education/dictionary/d. [Accessed: 03- Feb- 2015].

[16] V. Stewart, 'Data Deduplication and Compression Means Double Savings', Pure Storage Blog, 2013. [Online]. Available: http://www.purestorage.com/blog/data-deduplication-and-compression-means-double-savings/. [Accessed: 03- Feb- 2015].

[17] C. Konas, 'Compression confusion', Tech Target, 2002. [Online]. Available: http://searchstorage.techtarget.com/tip/Compression-confusion. [Accessed: 03- Feb- 2015].

[18] M. Staimer, 'A comparison of data compression and data deduplication technologies for SMBs', Tech Target, 2009.[Online]. Available: http://searchsmbstorage.techtarget.com/tip/A-comparison-of-data-compression-and-data-deduplication-technologies-for-SMBs. [Accessed: 03- Feb- 2015].

[19] The 2014 SNIA Dictionary, 'Data Deduplication', 2014. [Online]. Available: http://www.snia.org/education/dictionary/d. [Accessed: 03- Feb- 2015].

[20] S. Lowe, 'How Does Data Deduplication Work?', Enterprise Storage Guide, 2014. [Online]. Available: http://www.enterprisestorageguide.com/how-data-deduplication-works. [Accessed: 03- Feb- 2015].

[21] S. Waterhouse, 'Tape and Deduplication', The Backup Blog, 2011. [Online]. Available: http://thebackupblog.typepad.com/thebackupblog/2011/07/tape-and-deduplication.html. [Accessed: 03- Feb- 2015].

[22] M. Staimer, 'Tiered data backup storage strategies', Tech Target, 2010. [Online]. Available: http://searchdatabackup.techtarget.com/tip/Tiered-data-backup-storage-strategies. [Accessed: 04- Feb- 2015].

[23] The 2014 SNIA Dictionary, Tiered Storage, 2014. [Online]. Available: http://www.snia.org/education/dictionary/d. [Accessed: 04- Feb- 2015].

[24] C. Loringer, 'Build A Tiered Backup Architecture', Field Technologies Online, 2006. [Online]. Available: http://www.fieldtechnologiesonline.com/doc/build-a-tiered-backup-architecture-0001. [Accessed: 04- Feb- 2015].

[25] Iron Mountain, 'Reduce Your Data Backup and Recovery Costs With Tiered Storage', 2015. [Online]. Available: http://www.ironmountain.com/Knowledge-Center/Reference-Library/View-by-Document-Type/General-Articles/R/Reduce-Your-Data-Backup-and-Recovery-Costs-With-Tiered-Storage.aspx. [Accessed: 04- Feb- 2015].

[26] J. Singh, 'Understanding Data Deduplication', Druva Blog, 2009. [Online]. Available: http://www.druva.com/blog/understanding-data-deduplication/. [Accessed: 03- Feb- 2015].

[27] P. Gupta and M. Pegah, 'A new thought paradigm: delivering cost effective and ubiquitously accessible storage with enterprise backup system via a multi-tiered storage framework', Proceedings of the 35th annual ACM SIGUCCS conference on User services - SIGUCCS '07, 2007.

[28] D. Meister, J. Kaiser, A. Brinkmann, T. Cortes, M. Kuhn and J. Kunkel, 'A study on data deduplication in HPC storage systems', Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, p. 7, 2012.

Author

Khaled M. Aldossari works with the data storage support group at the EXPEC ComputerCenter, Saudi Aramco. For more than eight years of experience, Khaled led major projects to evaluate, design, and implement different data protection solutions. He worked also on supporting large-scale high performance storage. Khaled attained a distinguished Bachelor Degree in Computer Engineering from KFUPM University. He also received his Master's degree in Computer Science from California State University.In addition, Khaled is a SNIA Certified Storage Professional.