15th International Research/Expert Conference
"Trends in the Development of Machinery and Associated Technology"
TMT 2011, Prague, Czech Republic, 12-18 September 2011

INFORMATION LIFECYCLE MANAGEMENT AS SOLUTION FOR CONTINUOUS DATA AVAILABILITY

Dražena Gašpar
University of Mostar, Faculty of Economics
Matice Hrvatske bb, Mostar
Bosnia and Herzegovina

ABSTRACT
Paper presents and explains ways of using Information Lifecycle Management (ILM) in order to enable continuous data availability and ensure accomplishing of business processes. It is stressed that ILM is not just technology, because ILM integrates business processes and IT in order to determine how data flows through an organization, enabling users and managers to manage data from the moment it is created to the time it is no longer needed. ILM involves all aspects of dealing with data, starting with user practices, rather than just automating storage procedures and enables more complex criteria for storage management than data age and frequency of access. In the paper is showed that ILM is solution that could offer resolving of main continuous data availability problems like the cost of managing storage, ineffective use of storage, storage growth related to data backup, replication, disaster recovery continuance and so on.
Keywords: data availability, ILM, continuous data availability

1. INTRODUCTION
Today's huge amount of different variety of data stored in voluminous databases and data warehouses has to be readily accessible by Internet and sophisticated communications network. Gathering customer data, vendor information, minute financial measurements, product data, retail sell-through data, and manufacturing metrics become one of the most important corporate goals. The final result is that corporation had to accumulate terabytes of data on increasingly large storage systems, with the main purpose to enable end users that by different software applications use that data in order to improve their business activities and resolve potential business problems. Data relevant to business are stored on different hardware, inside different databases, different data warehouses, inside as well as outside of organization (public databases, web pages, Internet and so on). Accuracy, quality, timeliness and especially availability of such distributed data make crucial role in the process of business managing. Now, one of the biggest challenges is to ensure continuous (24/7/365) access to that data, even in critical and disaster situations.
There are different approaches and solutions that IT community developed and used in order to protect data and enable its continuous availability. Among standard approaches there are that using high availability (HA) devices with redundant systems, backing up data regularly to tape, and data duplication techniques. Also, there are some more sophisticated methods, including remote mirroring and remote copy (data vaulting), hot (near-line) backup, Data Lifecycle Management (DLM), Information Lifecycle Management (ILM) and so on. This paper is focused on Information Lifecycle Management (ILM) as a solution for continuous data availability.

2. DATA AVAILABILITY
Availability can be defined as the time that a system or resource, in this case data, is available for use. The definition of high availability is typically measured according to its percentage of absolute

availability where 100 % means that the resource is available all of the time and there is no downtime [7]. But, it is important to stress that 100% availability is used as ideal situation that is almost impossible to achieve in practice. Instead, the closest practical measure of very high availability is five 9s or 99.999 %. Expressed mathematically, availability can be defined as:

Percentage of availability (%) = ((total elapsed time – sum of downtime)/total elapsed time) (1)

where the percentage of system availability equals the total elapsed time, minus the total time the system is not available, divided by the total elapsed time [1].

If each year has a total of 8,760 hours (24 hours per day times 365 days per year) of available uptime, it means that a total of 8,760 hours of uptime could be translate to 100% of the available uptime for that year.

But, continuous availability of information is not just about technology. It is also about company staff, both in-house and staff located at remote locations. Without well educated, trained and responsible staff it is impossible to maintain maximum availability.

3. CHALLENGES TO CONTINUOUS DATA AVAILABILITY

The role of technology in creating a highly available environment has multiple levels. From a system standpoint, technological solutions address routine maintenance as well as the various types of failures that can occur such as site failures, server failures, and database corruption. From a business perspective, technology also influences the people, policies, and processes in business environment. For example, the hardware and software solutions a company chooses will determine both the skills needed by the staff and the particular processes the company will need to set up to manage that technology. The maintenance of those skills is a very important factor that can influence system availability. Continued training is vital to enable operations personnel to keep their skill level current, thereby ensuring that they can perform both routine and emergency procedures correctly and efficiently.

Different factors, either planned or unplanned usually influenced on decreasing of the percentage of availability. For example, it's common for a system to have one day (eight hours) of scheduled downtime for monthly maintenance so that IT staff can perform hardware upgrades, apply system patches, or perform other routine maintenance activities. In this scenario, the percentage of availability of the system is shown in the following expression:

$$98.9 = ((8760 – (8 \times 12)/8760)$$

In other words, one day of downtime per month results in an overall availability number of 98.9 percent. It is obvious that one day of downtime per month results in an overall availability of 98.9 percent which rates at one 9 out of a possible five 9s [7].

One of the challenges in designing a high available database is examining and addressing all the possible causes of downtime. There are different classifications of downtime causes and Figure 1. [7] shows one of them. The main principle of this classification is that causes of downtime could be planned and unplanned. Planned downtime is related to regular data maintenance activities such as backups, hardware and software upgrades, applying system patches and other. Unplanned downtime is primarily the result of some kind of failure, either computer (hardware failure), communication or data failure.

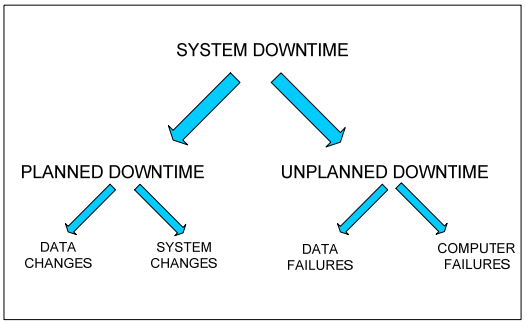

Figure 1. Causes of database downtime

In today's 24/7/365 business planned downtime is just as disruptive to operations as unplanned ones. Although Internet has made it easy to share data globally, it brings new challenges and requirements for data availability making regular (planned) maintenance operations have all but evaporated. Namely, there are no longer any windows of time during which users are not affected. Unplanned database downtime is something that is hard to predict and avoid, but it is for sure that it'll happened. As figure 1. shows there are two main kinds of unplanned database downtime: computer failures and data failures. Computer failure can be defined as an unplanned event that causes the server system to be inaccessible by users. Computer failures can result from a variety of hardware and software causes including hardware failure (CPU, RAM, storage, I/O, or power supply), operating system or device driver failure, database server failure. One of the first steps in protection against hardware failure is to invest in a hardware platform that provides redundancy of key components (hot-swappable RAM and RAID drives, redundant power supplies, built-in Uninterruptible Power Supply – UPS and so on). It is clear that users couldn't influence or completely avoid computer failures, they only could have implemented mechanisms for data protecting, e.g. they have to ensure that data couldn't be lost because of computer failures.

4. INFORMATION LIFECYCLE MANAGEMENT

Information life cycle management (ILM) is a comprehensive approach to managing the flow of an information system's data and associated metadata from creation and initial storage to the time when it becomes obsolete and is deleted. Unlike earlier approaches to data storage management, ILM involves all aspects of dealing with data, starting with user practices, rather than just automating storage procedures and in contrast to older systems (for example hierarchical storage management – HSM), ILM enables more complex criteria for storage management than data age and frequency of access [4].
It is important to stress that ILM is not just technology. ILM integrates business processes and IT in order to determine how data flows through an organization, enabling users and managers to manage data from the moment it is created to the time it is no longer needed.
ILM is, first and foremost, a strategic process for dealing with information assets. Usually, ILM is expressed as a strategy, which is then used to generate policies upon a set of rules is created and used by the organization or software to comply with the policies. ILM processes take into account what the information is, where it is located, what relationships it has to other information, and the lifecycle of the information [2].
Although terms data life cycle management (DLM) and information life cycle management (ILM) are sometimes used interchangeably, ILM is often considered as a more complex.
Data classification by business value is an integral part of the ILM process. Namely, the ILM approach recognizes that the importance of any data does not rely solely on its age or how often it is accessed. ILM expects that users and managers specify different policies for data that declines in value at different rates or that retains its value throughout its life cycle. A path management application, either as a component of ILM software or working in conjunction with it, makes it possible to retrieve any data stored by keeping track of where everything is in the storage cycle [6].

5. USING ILM IN ENSURING DATA AVAILABILITY

Successful and efficient implementation of IML needs that organization identifies critical data security requirements and includes them in their data classification processes. Data users, both individuals and applications, should be identified and categorized on the basis of the needs associated with their function.
Some of best practices related to IML implementation include: clearly definition of the ILM project mission and objectives; starting with business understanding; analyzing different models; using models for data classification and business alignment; using virtualisation techniques to group, move and map data efficiently; considering archiving solutions that can be integrated with existing backup solutions; considering a cross-platform data management; considering backup solutions that offer advanced features and can easily integrate with compliance solutions; maximizing current employee resource training in lifecycle management solutions; modelling business requirements; and starting with a subset of the infrastructure and build upon success [3].

Of course, each organization will need to develop and implement its own unique storage security solution, which must continue to evolve, adapting to new opportunities, threats, and capabilities. On Figure 2. is presented an example of conducting the ILM process [3].

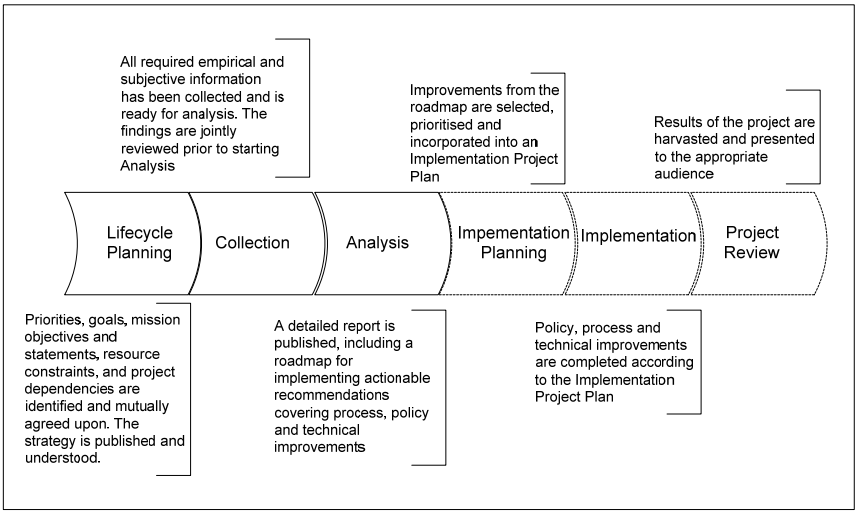

Figure 2. ILM process model

6. CONCLUSION

Data availability is a key component of business continuity with a main goal to safeguard important data from destruction, alteration, or loss, and to ensure 24/7/365 availability. It is achieved through a combination of technology, business processes, and best practices. Core components of a data protection strategy are backup and recovery, remote data movement, storage system security, clustering, mirroring and etc.

It is important to stress that high availability doesn't just happen. It is only achieved through strengthening the combination of people, processes, and technology. A focus purely on technology will never achieve very high levels of availability because many of the significant factors that affect availability stem from the interaction of people and processes. Enabling the proper hardware, software and database platform is only a starting point. ILM stress that high availability is the result of good planning and practices in combination with the appropriate technologies.

7. REFERENCES

[1] Patterson, D.A.: A Simple Way to Estimate the Cost of Downtime. Retrieved June, 22, 2009, from http://roc.cs.berkeley.edu/papers/Cost_Downtime_LISA.pdf

[2] Petrocelli,T. Data Protection and Information Lifecycle Management, Prentice Hall, USA, 2005.

[3] Reid R.S, Fraser-King G., Schwaserer W.D.: Data Lifecycles – Managing Data for Strategic Advantage, John Wiley&Sons, Ltd, USA, 2007.

[4] SearchStorage: Information Lifecycle Management. Retrieved, May, 10, 2011, from http://searchstorage.techtarget.com/sDefinition/0,,sid5_gci963635,00.html#

[5] Snedaker, S.: Business Continuity and Disaster Recovery Planning for IT Professionals, Burlington, USA, Syngress Publishing, Inc., 2007.

[6] Tomić D., Markić B.: Continuous Database Availability, Chapter 8 in „Always-On Enterprise Information Systems for Business Continuance- Technologies for Reliable and Scalable Operations", IGI Global, USA, 2009.

[7] Tomić D., Markić B, Mabić M.: Continuous Database Availability as a Precondition For Business Continuity, International Conference on Business and Information (BAI2007), Tokyo, Japan, July, 10-13, 2007.