[AFB+08] *Lars Arne Turczyk, Christian Frei, Martin Behrens, Nicolas Liebau, Ralf Steinmetz;* **Simulation of Automated File Migration in Information Lifecycle Management.** In: Proceedings of the Fourteenth Americas Conference on Information Systems, Toronto, Canada August 14th-17th 2008, August 2008. Seite

Simulation of Automated File Migration in Information Lifecycle Management

Lars Arne Turczyk, Christian Frei, Martin Behrens, Nicolas Liebau, Ralf Steinmetz
Technische Universität Darmstadt
KOM – Multimedia Communications Lab
Merckstr. 25
D-64283 Darmstadt
Germany
lars.turczyk@siemens.com

ABSTRACT

Information Lifecycle Management (ILM) is a strategic concept for storage of information and documents. ILM is based on the idea that in an enterprise information have different values. Information with different values are stored on different storage hierarchies. ILM offers significant potential cost savings by tiering storage and 90% of decision makers consider implementing ILM (Linden 2006). Nonetheless, there are too few experience reports and experimenting and researching in real systems are too expensive. This paper addresses this issue and contributes to supporting and assisting IT managers in their decision-making process.

ILM automation needs migration rules. There are well-known static, heuristic migration rules and we present a new dynamic migration rule for ILM. These migration rules are implemented in an ILM simulator. We compare the performance of the new dynamic rule with the heuristics. The simulative approach has two advantages. On the one hand it offers predictions about the dynamic behaviour of an ILM migration rule and, on the other hand, it dispenses with real storage hardware.

Simulation leads to decisions under certainty. When making a decision under certainty, the major problem is to determine which is the trade-off among different objectives. Cost-benefit analysis can be used to this purpose. A decision matrix is laid where rows represent choices and columns represent states of nature. The simulated results support the choice of migration rules and help to avoid mismanagement and poor investments in advance. The results raise the awareness of choosing the best alternative.

Keywords

Information Lifecycle Management, Simulation, File Migration, Migration Rules

INTRODUCTION

Information Lifecycle Management (ILM) is a strategic concept for storage of information and documents (Peterson 2004). ILM is based on the idea that in an enterprise different information have different values. Valuable information is stored on systems with a high quality of service (QoS). The value changes over time and therefore migration of information is required to cheaper storage systems with a lower QoS. Automated migration makes ILM dynamic. Such automation requires storage systems to understand which files are important at what time so that right migration rules can be applied. In this respect ILM nowadays lacks methods and tools.

For tapping cost potentials it is necessary to obtain a broader knowledge about ILM procedures and scenarios. There are too few experience reports on this subject and experimenting and researching in real systems is not in the focus of enterprises. Techtarget found that 66% of IT managers do not have time to put together a basic cost model or a data value model for ILM projects (Foskett 2006). Therefore the aim of this paper is to compare migration rules to generate results and experiences by simulating ILM scenarios.

The research questions are "What different classes of migration rules in ILM exist?" and "How can the best fitting migration rule be selected?". To answer the first question we analyse migration rules. There are already well-known static, heuristic migration rules and we present a new dynamic migration rule. In general sophisticated migration rules have to perform better than simple heuristics. That is what we investigate in this paper.

We look at characteristic properties of migration rules. Static rules are defined by time periods. If the period is over, the file is migrated. Unlike static rules, the dynamic rule looks at the access history and determines the "probability of further accesses" for each file. This probability changes each day. Therefore we speak of a dynamic migration rule.

To answer the second research question, simulations will be used to predict performance of each migration rule. Simulation leads to decisions under certainty. When making a decision under certainty, the major problem is to determine which is the trade-off among different objectives. Cost-benefit analysis can be used to this purpose. In decision-making under certainty a matrix is laid where rows represent choices and columns represent states of nature. The simulated results are put into a decision table to select the best alternative. The simulative approach for answering the second research question has two advantages. First, it offers predictions about the long-term dynamic behaviour of an automated ILM scenario. Second, it dispenses with real storage hardware which allows comparison of migration alternatives.

The comparison intends to support the design and decision process on ILM implementations. To be truly successful with the design and implementation of an ILM solution, it is important to remember that only about 25% of the challenge is about the selection of products (Patel 2005) (Patel and Shah 2005). We focus on the other 75% and derive criteria for effective migration rules. The simulator offers the possibility of checking the performance of the migration rules under dynamic circumstances without implementing real systems. This can help to avoid mismanagement and poor investments in advance.

To create a reality-based migration rule, we conducted two case studies. First we analyzed a knowledge database from a consulting department within a large German enterprise. The results showed that, on the one hand, there is demand in industry for ILM and that, on the other hand, Microsoft (MS) Office application formats .doc, .xls and .ppt dominate the file population (Gostner et al. 2005). Afterwards we conducted a second case study focussing on MS files. There we derived distribution functions of the access behaviour of different MS Office application formats (Groepl et al. 2006).

We now use these results to derive a file valuation metric which is used here to formulate dynamic migration rules for ILM environments. The migration rule will be implemented in a simulator to check its feasibility. The paper concludes with the comparison of the derived dynamic migration rule with well-known static migration rules.

The essence of this paper is as follows:

1. We present heuristic and dynamic migration rules for ILM
2. We check the performance of each migration rule by simulations
3. We compare the simulation results to select the best alternative of the migration rules

RELATED WORK

ILM is relatively new and has its roots in Hierarchical Storage Management (HSM). In opposite to HSM which focuses on cost only ILM takes into account the whole business process and observes multiple factors. Nevertheless some results from former HSM investigations can be used for ILM. Especially work done on file observation and algorithms is still relevant. Strange examined the long-term access behaviour on files in an UNIX system (Strange 1992). His aim was to identify regularities and patterns which can be applied to automated migration strategies for HSM. Further work deals mainly with algorithms which can be used for ILM or other storage strategies. Some examinations focus on the analysis of the access behavior and the development of migration strategies. The file migration protocol listing of a supercomputer was analyzed in a study by Miller and Katz. Migration methods were developed for a corresponding system (Miller and Katz 1993). Miller and Gibson examined the access behaviour in further studies in UNIX environments and designed a "file aging algorithm" as a migration rule (Gibson and Miller 1999). Today ILM is a strict focus of research. Beigi et. al. (2005) and Tanaka et al. (2005) offered proposals for migration rules in ILM. Last but not least, (Chen 2005) focused on the valuation of files, which leads to migration rules, too. We take these migration rules and compare them with our migration rule, which is based on statistical access probabilities.

Decision support for ILM migration rules was not considered in the technical research. Nonetheless ILM is a very real topic on which IT managers have to decide. In most cases decision support is offered by cost models. Cost models for IT are found for nearly every situation. For ILM cost models are available, too. For example, Turczyk et al. (2007) presented an ILM cost model. This model supports the decision about ILM, Yes/No, and how many hierarchies have to be employed. There are no models for the decision about ILM migration rules. Here our paper supports modelling with simulations and decision-making.

In the next section we present a set of four heuristic migration rules and one dynamic migration rule.

MIGRATION RULES

There are already several heuristic migration rules for ILM. Our new rule has to compete with these heuristics. Before presenting our new migration rule we look at the heuristic rules. We distinguish four types of migration rules (trivial, optimal, static and dynamic) The first type are the trivial migration rules. They are very simple and do not really fit to ILM. Nonetheless they have to be considered because they are real options for IT managers. There are two trivial migration rules. The one is called "Do nothing!" (M1).

Definition „Do nothing!" (M1): All files are stored on the highest hierarchy.

Applying M1 means that there is no ILM.

The other trivial migration rule is called "Naive method" (M2).

Definition "Naive Method" (M2): All files are stored on the lowest hierarchy. If there is an access, the file is stored on the highest hierarchy for one day. After one day the file is stored on the lowest hierarchy again.

There is another rule, the optimal rule. It says "one day before you need a file, it is migrated to the highest hierarchy". It is no real option for IT managers because this method is based on anticipation. Nonetheless it works in laboratories and in simulations and shows the potential of ILM in a given scenario which is an important information for IT managers. Here is the definition of the optimal rule.

Definition „Method of perfect anticipation (optimal method)" (M3): All files are stored on the lowest hierarchy. If a file will be accessed the next day, it will be migrated one day before (anticipation).

Again, this method does not work in reality. This method looks similar to M2.

Beyond the mentioned migration rules there are two types of real applicable migration rules - the static rules and the dynamic rules.

Static migration rules have fixed time-thresholds which determine when the file is migrated. The most famous static migration rule is offered by Fred Moore who said "90 days on enterprise storage – 90 years on tape" (Moore 2004).

Here is the definition of the time based migration rule.

Definition "Time based migration" (M4): A file which is not accessed for 90 days is migrated to the next lower hierarchy.

Unlike static rules the dynamic rule looks at the access history and determines the "probability of further accesses" for each file. This probability changes each day. Therefore we speak of a dynamic migration rule.

Definition „Method of probability of further access" (M5): A file is migrated from hierarchy 1 to hierarchy 2 if the probability of further access is lower than 10%. A file is migrated from hierarchy 2 to hierarchy 3 if the probability of further access is lower than 2.5%.

Calculation of the probabilities is made according to our statistical results from observing file access in our second case study (Groepl et al. 2006). It was established that there is no suitable distribution model for the complete sample of MS Office files. For this reason, mixed distribution functions were introduced and differently composed subgroups of the sample were examined for their distribution. Suitable distribution models were then successfully generated for some subgroups. Table 1 gives an overview of the test results.[1] The Gamma distribution is a general distribution covering many special cases, including the Chi-squared distribution and Exponential distribution. The Weibull distribution gives the distribution of lifetimes of objects. It was originally proposed to quantify fatigue data.

[1] "Distribution model" in the table stands for the suitable continuous and truncated distribution function in a mixed distribution function.

Criterion	Class	Distribution model
Age of the file	[0 days;365 days)	W(0.35,3.5)
	[365 days;730 days)	-
	[730 days;1772 days)	
Number of accesses	[1 access;7 accesses)	-
	[7 accesses;15 accesses)	G(0.32,183)
	[15 accesses;292 accesses)	W(0.36,4.0)
File type	Doc	
	Xls	W(0.25,1.1)
	Ppt	W(0.38,14.3), G(0.19,221)
	Pdf	
	Miscellaneous	W(0.46,27.7), G(0.29,181)

Table 1. Summary of the test results. W($\hat{\alpha}$, $\hat{\beta}$) = Weibull distribution, G($\hat{\alpha}$, $\hat{\beta}$) = Gamma distribution, $\hat{\alpha}$ and $\hat{\beta}$ are the estimated values of the parameters.

Now we become familiar with 5 migration rules, M1 - M5. To decide about the rules we have to check their performance. Therefore we implemented a simulator to observe the behaviour of each migration rule. As an example we show the graphic of the simulated capacity requirements of M5 (see figure 1). Here we see the characteristic shift of capacity demand from hierarchy 1 to hierarchy 3 towards the end of the simulation period.

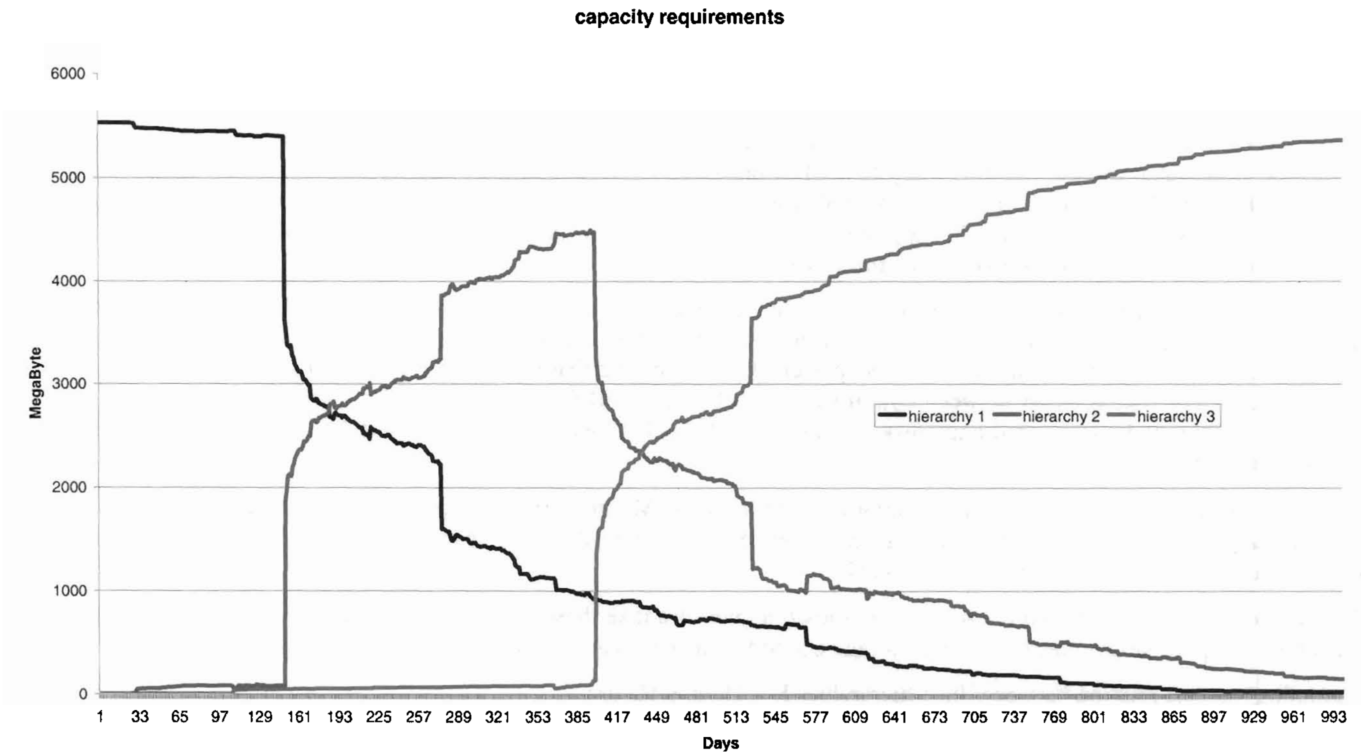

Figure 1. Capacity demand per hierarchy when applying M5

The question is "How can the best fitting migration rule be selected?". Therefore now we collect the simulation results and compare the migration rules for choosing the best alternative.

COMPARISON OF MIGRATION RULES

Industry best practice is to take a strategic view of storage architectures and determine the best solutions based on total cost of ownership (TCO) and operating expense (OPEX) minimization. This approach was followed e.g. by (Turczyk et al. 2007). They created a cost model for ILM. This cost model allows decision to be made about the architectures themselves, i.e. "ILM yes/no?", "how many hierarchies?" or "which storage technologies?". Assuming the decision in favour of ILM was made, we now have to decide about the migration rules.

Therefore we look at three criteria which are characteristic for migration rules.

The first criterion is the direct cost of storage volume which is needed when applying the migration rule.

The second criterion is the number of remigrations, i.e. how many times are files migrated upwards although the migration rule migrated them downwards already.

The third criterion is the average quality of service (QoS) per file offered by applying the migration rule.

We implemented a simulator to derive the results for each migration rule and each of the three criteria.

The basis for the comparison is an ILM scenario with three storage hierarchies and 10,000 files. The total volume of the 10,000 files is 5,529 megabytes. The simulator feigns the migration behaviour of the 5 migration rules, M1-M5, over a period of 1,000 days.

We now look at the criteria in detail and afterwards we will summarise the results for decision-making.

Direct Cost

Storage volumes are growing at phenomenal rates, yet IT organizations cannot always justify increasing storage budgets. IT organizations expect budget allocations for overall storage services, storage hardware, and software infrastructure to be constant year after year - a percentage of total IT budgets.

Businesses are looking to better manage their storage capacity use and are starting to automate some of their management functions. Automated file migration is a key of ILM. It makes ILM dynamic. 90% of decision makers consider implementing ILM (Linden 2006). They are forced with cost reductions and IT continues to be a target for reducing costs while still being required to deliver critical services. The argument is IT price erosion. In fact in storage area enterprise, modular and locally attached disk will all continue to experience price erosion.

In 2006 Sun Microsystems said prices have been declining annually at approximately 35% (SUN 2006). Prices for enterprise disk were in the $25 to $40 per gigabyte range. Midrange disk is at $12 to $20 per gigabyte, low-end disk is at $3 to $10 per gigabyte. High-end tape automation is $1 to $2 per gigabyte. Knowing that these prices will be obsolete by now, the relation is almost constant. The relation between enterprise disk (FC, SCSI, FICON, ESCON) and midrange disk (SCSI, FC) is 2:1. The price difference between midrange disk and low-cost disk (S-ATA) is 3:1. The price difference between low-cost disk and automated tape is 5:1

Horizon Information Strategies made a similar relation in 2003 (Moore 2003). The price difference between enterprise disk and midrange disk is 2.5:1. The price difference between midrange disk and low cost disk is 3:1. The price difference between low cost disk and automated tape is 6:1.

The relations tend to be relatively constant, as shown above. We take these relations for comparison. Nevertheless price per megabyte is a poor single metric to use in storage economic decisions as we will see by the simulations.

Assuming the cost relation between the 3 hierarchies H1, H2 and H3 are:

H1 = 1 cost unit/MB, H2 = 1/6*H1, H3 = 1/5*H2 (SUN 2006). Hence the direct cost can be derived from the capacity requirements offered by the simulator (see figure 1 and table 1).

Migration rule	M1	M2	M3	M4	M5
Direct cost	5529	237	237	1122	1746

Table 1. Direct Cost per migration rule

Knowing that costs are the most important issue (Short 2006), we see that the trivial rule M2 is a real option. The problem with M2 is that the files are remigrated at the very moment they are needed. This means that the user has to wait for the files. We will observe the remigration and summarize them to a single parameter which we call "jitter". The jitter is the average number of remigrations in an ILM scenario. Hence it is a criterion for quality of migration rules.

Furthermore in scenario M2 almost all files are stored on the lowest hierarchy. In enterprise the lowest hierarchies have the lowest quality of service (QoS). This leads, for example, to lower availability and longer recovery periods. This is an issue for IT managers, too. Hence we take QoS into consideration. We now look at the jitter and the QoS.

Jitter

Figure 2 shows how a single file trembles between the three hierarchies over the lifecycle of 1,000 days.

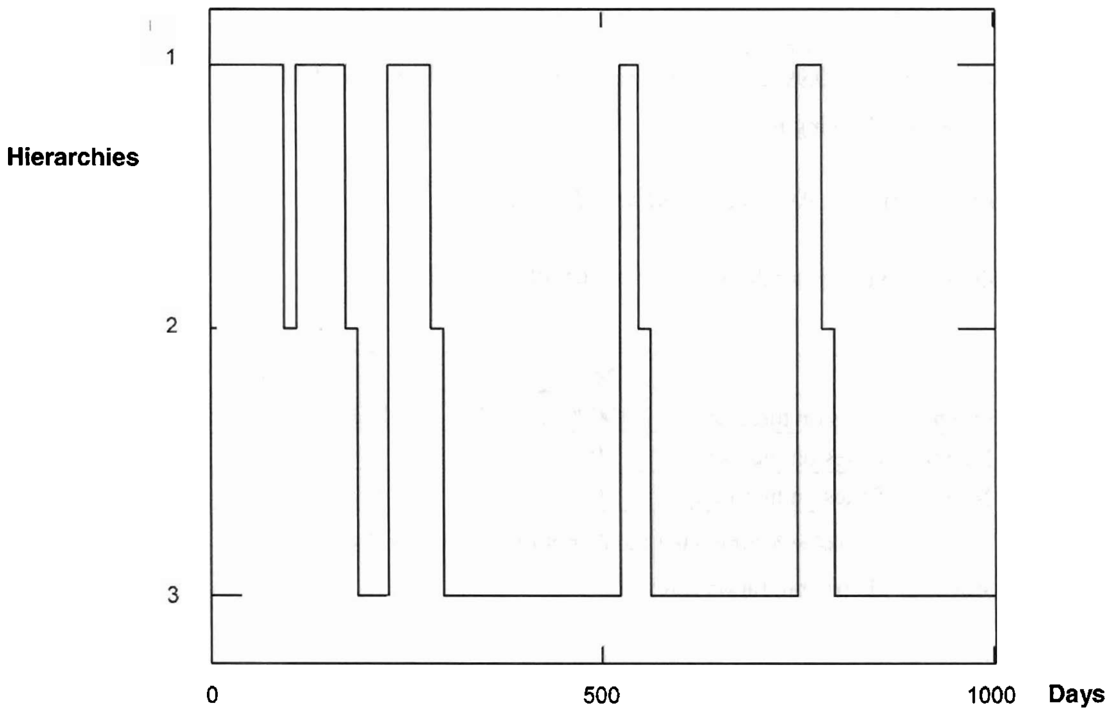

Figure 2: Integral observation of a single file

Since the file trembles between the storage hierarchies, this is a good example for the demonstration of jitter. Until the file was definitely to be stored on level 3, it was migrated back to a higher hierarchy four times in 1,000 days. Therefore it has a jitter of J(1000)=4. For the simultaneous analysis of all 10,000 files, the graphic evaluation does not make sense, therefore the jitter is an important measure to generate results over all files.

Definition (Jitter): Jitter Jk(t) is the mean number of accesses on a file k when k is stored on a lower hierarchy than hierarchy 1 within the observation period t.

The jitter says how many times in average a file was accessed when it was not on the highest hierarchy. This means for the user that he has to wait while loading the file. This is not the intention of IT-managers, because they are responsible for the IT-service. Therefore jitter is an indicator for the quality of migration rules in an ILM scenario. Table 2 shows the simulation results concerning jitter.

Migration rule	M1	M2	M3	M4	M5
Jitter	0	6.97	0	0.5	0.3

Table 2. Simulated jitter per migration rule

We see M2 ("Naive method") has the largest jitter. M3 has no jitter per definition, because here the file is loaded in advance. Hence there is no waiting time for the user. M4 and M5 have quite small jitter values.

QoS considerations

In enterprise the lowest hierarchies have the lowest quality of service (QoS). This leads for example to lower availability and longer recovery periods. This is an issue for IT managers. Hence we take availability as representative of QoS into consideration, too.

We assume there are different SLA (service level agreements) for the 3 hierarchies offering the following values for availability: Availability of hierarchy 1 = 0.99, availability of hierarchy 2 = 0.97 and availability of hierarchy 3 = 0.93.

We calculate QoS according to the following formula:

$$QoS = \frac{1}{10000} \sum_{i=1}^{3} (availability\, H_i \cdot Average\ number\ of\ files\ on\ H_i)$$

The simulator offers the following average numbers of files on the different hierarchies:

	M1	M2	M3	M4	M5
Average number of files on hierarchy $_1$	10000	1401	1401	1921	2696
Average Number of files on hierarchy $_2$	0	0	0	1195	2607
Average Number of files on hierarchy $_3$	0	9859	9859	6884	4697

Table 3. Simulated number of files per hierarchy

Here is the average availability per file per migration rule:

Migration rule	M1	M2	M3	M4	M5
QoS	0.99	0.931	0.931	0.942	0.957

Table 4. Simulated QoS per migration rule

The summarization of the three criteria per migration rule is found in table 5:

	Cost	Jitter	QoS
M1	5529	0	0.99
M2	237	6.97	0.931
M3	237	0	0.931
M4	1212	0.5	0.946
M5	1747	0.3	0.957

Table 5. Summary of simulation results per migration rule

Now we have the simulation results and we have to decide on the best alternative.

Decision Making

No matter what kind of decision one must make, the six steps for decision-making are:

1. Clearly define the problem at hand.

Here: Which is the best of the presented migration rules?

2. List the possible alternatives.

Here: M1 to M5

3. Identify the possible outcomes.

Here: The simulation results per decision criterion[2].

4. List the payoff or profit of each combination of alternatives and outcomes.

Here: Payoff table according to utility function

5. Select one of the mathematical decision theory models.

Here: Decision-Making Under Certainty (DMUC) with multiple targets. In this environment, decision-makers know with certainty the consequence of every alternative. They will choose the alternative that will maximize their well being or will result in the best outcome.

6. Apply the model and make your decision.

Here: Choosing the best alternative

As we see, steps 1 to 3 are completed. We now create the payoff table (step 4).

The utility function is $\Phi(M_i) = \sum_{p=1}^{r} g_p \cdot u_{ip}$

With $\Phi(M_i) =$ utility of choosing alternative M_i

r = number of criteria

g_p = weight of criterion p

u_{ip} = utility of M_i according to criterion p

The weighting has an influence on the decision result. Here we choose almost equal weights with stress on direct cost:

They are: $g_1 = 0.4$, $g_2 = 0.3$, $g_3 = 0.3$ [3]

When summarizing the three criteria, we have to norm them otherwise the result has no significance.

The best value of all migration rules per criterion shall become 1 and the worst shall become 0.

This leads to the normed decision table:

Target	cost		Jitter		QoS		Utility	Rank
Target weight	0.4	$g_1 \cdot u_{i1}$	0.3	$g_2 \cdot u_{i2}$	0.3	$g_3 \cdot u_{i3}$	$\Phi(a_i)$	
ILM-Method								
M1	0	0	1	0.3	1	0.3	**0.6**	**4**
M2	1	0.4	0	0	0	0	**0.4**	**5**
M3	1	0.4	1	0.3	0	0	**0.7**	**2**
M4	0.816	0.326	0.93	0.279	0.254	0.076	**0.682**	**3**
M5	0.715	0.286	0.96	0.288	0.44	0.132	**0.706**	**1**

Table 5. Normed decision table

[2] A decision criterion represents the decision maker's point of view for selecting the best alternative.

[3] IT managers have to vary the weights according to their specific storage situation. The logical decision steps remain unchanged.

We know with certainty the consequence of every alternative. The decision model is Decision Making Under Certainty (DMUC) with multiple targets (step5). The last step is to choose the best alternative. This is M5 "Method of probability of further access".

SUMMARY AND OUTLOOK

Information Lifecycle Management (ILM) is a strategic concept for storage of information and documents. ILM is based on the idea that in an enterprise information have different values. Information with different values are stored on different storage hierarchies. 90% of decision-makers consider implementing ILM (Linden 2006). Nonetheless, there are too few experience reports and experimenting and researching in real systems are too expensive. This paper focusses on the issue of automated file migration in ILM. It contributes to supporting and assisting IT managers in their decision-making process.

Automation needs migration rules. There are well-known static, heuristic migration rules and we present a new dynamic migration rule for ILM. These migration rules are implemented in an ILM-simulator. The research questions are "What different classes of migration rules in ILM exist?" and "How can the best fitting migration rule be selected?".

We identified four classes of migration rules (trivial, optimal, static and dynamic) and presented 5 migration rules. To decide about the rules we checked their performance. Therefore we implemented a simulator. Then we observed their behaviour in a simulated 1,000 days' scenario. The scenario consisted of 3 hierarchies and 10,000 files. The results according to the criteria direct cost, jitter and QoS were collected and put into a decision table. Applying a utility function, the decision for the best alternative was made.

The results were explicitly:

Result 1: The trivial migration rule M2 ("Naive method") is the best concerning cost. The jitter is more than twenty times higher than the dynamic migration rule (M5). Overall M2 is worse than the trivial migration rule M1 ("Do nothing"). In general trivial migration rules are no real alternative because they do not reflect multi-hierarchical ILM environments, because they only support two hierarchies.

Result 2: The static migration rules like M4 ("Time based migration") have a long tradition from HSM and are real options. They fit to multi-hierarchical environments, but they are too simple. Therefore, they do not really apply to the ILM storage concept. The performance according to the three criteria gained second place among the real options.

Results 3: The dynamic migration rule M5 ("Method of probability of further access") fits to the ILM concept because of its multifactor character. The performance considering the three criteria was the best among all options. This method even was better than the "optimal" method (M3), which has a strict focus on direct cost.

For tapping cost potentials it is necessary to obtain a broader knowledge about ILM procedures and scenarios. In this respect ILM nowadays lacks methods and tools. The comparison presented helps to gain experience on automation performance. This can help to avoid mismanagement and poor investments in advance.

Further work will be done in analyzing the performance of dynamic migration rules using data from real storage environments.

REFERENCES

1. Beigi, M., Devarakonda, M., Jain, R., Kaplan, M., Pease, D., Rubas, J., Sharma, U., Verma, (2005) A., Policy-Based Information Lifecycle Management in a Large-Scale File System. In Proceedings of the Sixth IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY'05) pp. 139-148
2. Chen, Y. (2005) Information valuation for Information Lifecycle Management. In: Proceedings of the Second International Conference on Autonomic Computing (ICAC'05), pages 135-146.
3. Foskett, S. (2006) Best Practices – Belt tightening continues to shift budgets away from security and information management, TECHTARGET http://searchwebservices.techtarget.com/magPrintFriendly/0,293813,sid35_gci1179438,00.html
4. Gibson, T. and Miller E. (1999) An Improved Long-Term File-Usage Prediction Algorithm. 25th Annual International Conference on Computer Measurement and Performance (CMG 99), Reno, NV, USA, pages 639-648.
5. Gostner, R., Turczyk, L., Berbner, R., Heckmann, O. and Steinmetz, R. (2005) „Analyse von Datei-Zugriffen zur Potentialermittlung fuer Information Lifecycle Management", TU Darmstadt KOM Technical Report 01/2005

6. Groepl, M., Turczyk, L., Heckmann, O. and Steinmetz, R. (2006) „Statistische Datenanalyse von langfristigem Dateizugriffsverhalten in einer Unternehmensdatenbank fuer ILM", TU Darmstadt KOM Technical Report 01/2006
7. Linden, L. (2006) Storage Budget Survey 2006, Glasshouse Technologies Inc. http://www.glasshouse.com/storage-budget-survey/registration.shtml
8. Miller, E. L., Katz, R. H. (1993) An Analysis of File Migration in a UNIX Supercomputing Environment. In: USENIX Winter, pages 421–434, 1993.
9. Moore, F. (2003) Storage - New Game New Rules, Horison Information Strategies,
10. Moore, F. (2004) Information Lifecycle Management, Horison Information Strategies,
11. Patel, C. D., Shah, A. J. (2005) Cost Model for Planning, Development and Operation of a Data Center. Internet Systems and Storage Laboratory HP Laboratories Palo Alto, HPL-2005-107(R.1), June 9, 2005
12. Peterson, M (2004)., ILM Definition and Scope - An ILM Framework, SNIA Data Management Forum, Version 2.3, July 2004
13. Short, J. E. (2006) ILM Survey: What Storage, IT and Records Managers Say. ISIC Research Report Vol. 06, No. RB06-02, Information Storage Research Center, University of California.
14. Strange, S. (1992) Analysis of Long-Term UNIX File Access Patterns for Application to Automatic File Migration Strategies. Technical Report UCB/CSD-92-700, EECS Department, University of California, Berkeley.
15. Sun Microsystems (2005) Storage Optimization – ILM. March 2006"
16. Tanaka, T., Ueda, R., Aizono, T., Ushijima, K., Naitih, I., Komoda, N. (2005) Proposal and Evaluation of Policy Description for Information Lifecycle Management Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05) 0-7695-2504-0/05
17. Turczyk, L., Behrens, M., Liebau, N., Steinmetz, R. (2007). Cost Impacts on Information Lifecycle Management Design. 13th Americas Conference on Information Systems (AMCIS 2007); Keystone, Colorado, USA.