

Methodology for Anticipating Operational Budget to Ameliorate Hybrid Storage Clouds

Sasikanth Eda^{*1}, Riyazahamad M Shiraguppi², Bhushan P Jain³, Sandeep R Patil⁴

IBM India Systems and Technology Labs, Pune, India

^{*1}sasikanth.eda@in.ibm.com; ²riyaz.shiraguppi@in.ibm.com; ³bpjain@cs.stonybrook.edu;

⁴sandeep.patil@in.ibm.com

Abstract-The dramatic evolution of various disk and memory technologies has helped with the rapid establishment of hybrid storage environment comprising of heterogeneous storage units. In order to provide an improved and effective storage management, there exists a necessity of adapting S.M.A.R.T changes in the storage management stack that enables compact sensing, processing, decision making capability based on the importance of data, disk life span predicted on operational workloads and financial analysis estimating the disk replacement cost etc. This paper reviews system architecture for two faces of RAS (Reliability Availability Serviceability) storage cloud features namely disk breakage prediction independent of disk technology, host aware data tier based on disk life span and subsequently describes operational budget anticipation based on predicted disk failures.

Keywords- Hybrid Disk; Life Span; Curve Fitting; S.M.A.R.T Parameters; SSD; Operational Budget

I. INTRODUCTION

Storage in general and disks in particular are the driving force for all Information Technology (IT) enabled business. Since the storage system hosts the entire business data, it is vital to have these systems with state of the art storage characteristics. Further, characteristics like performance, reliability, backup, availability and data protection of storage subsystems like disks directly impact on the effective execution of an IT enabled business. Hence generally, significant IT budget is allocated for ever growing storage needs of a business. Along with others, this includes cost involved in:

- Upgrade or replacement of storage subsystems like disk replacement to newer disk technology.
- Adherence to changing business and legal requirements.

Newer Disk Technologies viz., Solid State Disks (SSD) and memory technologies like Phase Change and Race Track [1] offer great promise and unique characteristics, which may be typically used in Storage Cloud. Moreover, since clouds span across different geographies, it adds new elements to be considered in cost efficiency. This paper discusses a novel method for disk breakage prediction independent of disk technology on production workloads, and anticipates operational budget for predicted disk failures or upgrades. This will help to optimize the existing methodologies by taking into consideration the advantages of various disk technologies as well as different cost factors.

II. HYBRID STORAGE AND STORAGE CLOUDS

Cloud computing refers to a computing platform that is able to dynamically provide, configure, and reconfigure technology infrastructure to address a wide range of dynamic needs. There are various cloud offerings addressing specific requirements like high performance computing and image or video processing which are unstructured by nature. The amount of unstructured data is and will continue to increase exponentially due to astronomical data generated from video, audio, graphics, and web applications. Storage Clouds are one of the most popular storage systems in the enterprise world which are being used to host this unstructured data.

The new evolving disk technologies like Phase Change Memory & Racetrack memory while currently the popular Flash Class storage memory is potentially being used in providing more innovative solutions. In the light of so much research and advances in the disk technology, a complete restructuring of the backend storage to replace the older technology disks with new technology ones is neither economical nor practically feasible. It is imperative that multiple generations of disk technologies co-existing in the same storage system for effective operation and low cost of ownership give us the ground for Hybrid Storage. Thus we have a hybrid storage structure formed in which all the different disk & memory technologies are available and we must leverage this hybrid architecture.

Hybrid Storage has started becoming apparent in the industry. Building systems with the use of hybrid storage systems collectively for cloud based use cases is known as storage cloud with hybrid storage devices.

III. RELATED WORK: NESCIENT OVER HYBRID DISK TECHNOLOGY

Traditional disk replacement methodologies [2] do not consider the advantage of the underlying disk technology and cost budgeting based on data stored during the operational process. There is a strong need for such systems to do so to derive the

much needed growth in performance and to meet the challenges of being greener.

Traditionally, industry follows the disk replacement strategy of replacing problematic disks. The schemes phase out old disks and replace them with new technology [3] disk. The tools that are used typically rely on failure factors [4] reported by disks S.M.A.R.T (Self-Monitoring Analysis and Reporting Technology) reports. While this is an acceptable approach, it is not a S.M.A.R.T and business effective approach. The existing methods and tools used to strategize disk replacement policies for data centres have started to appear primitive with the advent of newer disk technologies like Flash [5, 6], Phase Change technology and Racetrack memory using spintronic science which will continue its advent. The ideal solution is to replace all the disks with new technology disks, but the cost expense prohibits that. Since regular operation of disk directly impacts on the business costs [7] (through power consumption/ heat dissipation/carbon credit utilization), it is vital to have a method that can be deployed in tools which will help to strategize and identify right disks for proactive replacement in a geographically spread cloud environment where data is replicated across disks which are located in different parts of the world.

IV. PROPOSED TECHNIQUE FOR AMELIORATED BEHAVIOUR

Hybrid Storage is very apparent due to various disk technologies. Based on the previous facts, the paper discusses methodologies explaining how we can utilize the various evolving disk technologies to ensure a S.M.A.R.T and cost effective solution. This paper proposes a methodology for disk replacement with the use of calculating the devices which needs to be replaced in priority based on the mathematical model which derives data from device characteristics. The proposed technique analyses and mathematically forms weight to each operational disk with the factors and present a sorted view of most probable disk to replace with a new technology disk to ensure a positive influence on the business ROI, efficiency and aid in a S.M.A.R.T, realistic disk replacement strategy.

Assuming that measurement of the stated factors like power consumption, heat dissipation, carbon credit utilization, etc. per disk can be determined; detailed mathematical calculation procedure of the disk replacement result is explained as follows:

A. Forming a Mathematical Variable for Individual Disk Behaviour:

Consider vendors specifications that are provided in the data sheet as the initial observations of a disk. S.M.A.R.T parameters of each disk [8] are equated to a set variable. Supposing the provided disk is a Hard Disk Drive, vendor specifications are collected and equated to the variable "SP" (Note all the S.M.A.R.T specifications that are provided by the vendor for a particular HDD is considered as the initial reading).

$SP = \{\text{Head flying flight, Data throughput performance, Spin up time, Reallocated sector count, Seek error rate, Seek time performance, Spin try recount, Drive calibrations retry count}\}.$

Similarly if the provided disk is a Solid Disk Drive, vendor specifications [9] are collected and equated to a variable "SP" (Note all the S.M.A.R.T specifications that are provided by the vendor for a particular SDD is considered as the initial reading).

$SP = \{\text{Power management, Latency specification, Random read/write input/output operations per second, Electrical characteristics, Altitude, Electromagnetic immunity, Shock and vibration}\}.$

In a heterogeneous cloud data center, the S.M.A.R.T parameters of individual disks can be equated to variable X_i (where $i = 0, 1, 2, 3 \dots n$).

After installing the disks in a production environment, the disk behaviour tends to differs from the ideal specifications provided by disk manufacturer. The differences of S.M.A.R.T parameter are exhibited by examining each individual disk that is noted as differential variable (Δx). The value of the differences may contain either positive or negative tolerances.

B. Considering the Environmental Disturbances at Production Environment:

Environmental deviations are noted between the default vendor specified temperatures to the field temperature. Here the environmental noise [10] can be assumed to be largely an accumulation of thermal noise.

C. Forming a Mathematical Equation for Individual Disk Behaviour:

A minimum of 100 hours of observation time is considered for all the disk parameters. Assumption at this state is that there are no impulse I/O load changes, environmental issues, and electrical characteristics. The noted S.M.A.R.T values per disk (initial, installed, 100 hour readings) are plotted on to a graphical sheet and traditional normalization techniques are applied on each S.M.A.R.T parameter plot. The plotted curve can follow as any one or a combination of the noted following curves exponential decay curve ($y=Ae^{-Kx}$), exponential raise curve ($y=Ae^{Kx}$), parabolic nature ($y^2=4ax$), and sum of exponential curves ($y=Ae^{ax} + Be^{bx}$). Curve fitting mechanism named "least squares" [11] is applied on the plotted readings, which are normalised to form a curve.

Considering the sample analysis, the values of time in hours vs. SP are noted as (2, 1.8), (4, 1.5), (6, 1.4), (8, 1.1), (10, 1.1), and (12, 0.9). The tabulated values are as shown in Table 1.

TABLE 1 NOTED VALUES FOR TIME AND S.M.A.R.T PARAMETERS (FOR A GIVEN DISK)

Time (hours)	SP (S.M.A.R.T parameter)
2	1.8
4	1.5
6	1.4
8	1.1
10	1.1
12	0.9

The plotted curve for the above tabulated values following a decay and numerical method technique is applied on them to achieve the curve equation.

The solved curve equation is given as:

$$0.936^{SP} = (T/2.013) \quad (1)$$

Where T denotes time in hours and SP denotes the S.M.A.R.T parameter value for observed disk. Disk breakage algorithm uses the above calculated equation to measure the S.M.A.R.T parameter values based on the time provided.

D. Considering the Weighted Approximation:

The equations that support the weighted parameters like environmental changes and importance of data stored are given below:

$$a_0 \sum_i w_i + a_1 \sum_i x_i w_i + a_2 \sum_i x_i^2 w_i = \sum_i y_i w_i \quad (2)$$

$$a_0 \sum_i x_i w_i + a_1 \sum_i x_i^2 w_i + a_2 \sum_i x_i^3 w_i = \sum_i x_i y_i w_i \quad (3)$$

$$a_0 \sum_i x_i^2 w_i + a_1 \sum_i x_i^3 w_i + a_2 \sum_i x_i^4 w_i = \sum_i x_i^2 y_i w_i \quad (4)$$

Eqs. (2), (3) and (4) are noted from 'weighted least squares curve fitting normalisation mechanism' [12], and are used in the proposal for considering the weighted factor (w_i) for environmental variations.

E. Identifying the Sudden Changes Noted via Plot:

The impulse response time (Δt) between the two consecutive time points is calculated in order to estimate the sudden changes occurred within the combined S.M.A.R.T parameter plot. The algorithm considers the slope variations derived from impulse response and modifies the weighted approximation accordingly.

F. Identifying the Contour Region Surrounding the Breakage Point:

The achieved equation is extrapolated based on the resource procurement period (differs from vendor to vendor) using Eq. (1). Apply Voronoi principles [13] to identify the closed contour region around the given point plotted on the site as shown in Fig. 1.

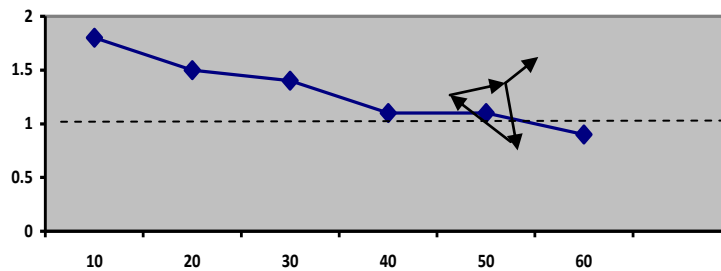


Fig. 1 Plot for Time (hours) on x-axis vs. S.M.A.R.T parameters on y-axis and disk breakage threshold

Assumption here is that the disk breakage at S.M.A.R.T parameter value equals to '1'. The plotted curve forms the extrapolation of the disk values based on the time variations. The contour region circled against the marked break point denotes alarming time gap that triggers an auto notification to the cloud administrator.

The S.M.A.R.T cloud management stack prototype achieved by using the proposal is tabulated as shown in Table 2.

TABLE 2 NOTED VARIOUS DISK PARAMETERS, CALCULATED DEGRADATION VALUES, TIME LEFT FOR REPLACEMENT AND ITS MEASURED PRIORITY

Disk no.	Degrade	Data importance	Time left	Backup	Replace Priority
SATA1	50%	4%	72hr	Yes	3
SSD1	90%	1%	30hr	Yes	6
SATA4	40%	28%	>50hr	No	4
SATA6	10%	50%	>100hr	No	5
SATA8	99%	78%	24hr	No	1
SSD7	2%	10%	>100hr	Yes	7
SSD0	75%	35%	55hr	No	2

The above table demonstrates the heterogeneous cloud scenario containing variations in disk technologies, degradation ratios, data importance factor, extrapolated time scale, and back up provided for individual disk array (in the above table the 'Time left' value is derived from plots from Fig. 1, Degrade % is derived from 'time left' value while the rest of the values are manually entered by the input system).

The proposed algorithm description is framed in a flowchart as shown in Fig. 2.

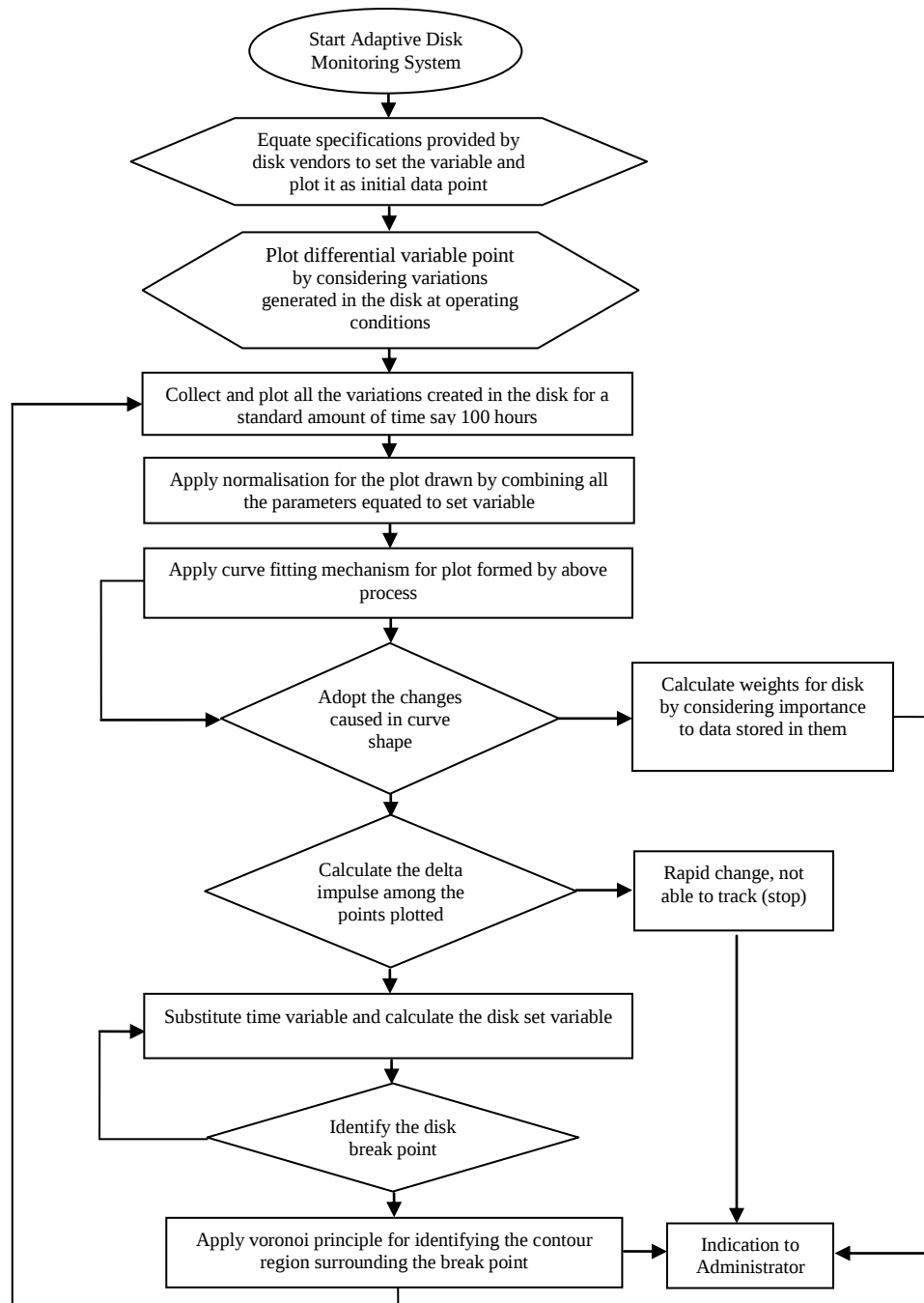


Fig. 2 Flow chart demonstrating the proposed algorithm for disk breakage prediction

The proposed disk replacement methodology calculates the degradation factor using the S.M.A.R.T parameter variations, extrapolates time left for replacement using curve fitting mechanism and provides the values to the centralised management stack. The centralised management application identifies the importance of data stored within the disks, backup option available at each disk site and forms priorities for disk replacement.

V. OPERATIONAL BUDGET ANTICIPATION FOR DISK REPLACEMENT

Even though the unit price of storage volume (per GB) is observed to be declining over the years due to effective use of techniques such as virtualization, deduplication etc., still the operational cost spent on disk manufacturing remains constant. Apart from other management overheads such as electricity, cooling, floor space, etc., data centers are still facing difficulties in terms of predicting the disk failures in advance and planning the priorities of disk replacement following the financial norms practised by the respective organisation. As demonstrated in the previous section of the paper, the calculations help administrators of data centers to keep track of disk failures predicted for future, where as the current section deals with the analysis that considers various factors contributing to financial planning (described below) and remaps the priorities of disk replacements accordingly. The proposed methodology involves consideration of multiple fluctuating factors listed below for making out effective operational budget for disk replacement.

A. Logistical Time Delay Expected for Disk Replacement

A typical problem faced by growth market data centers is ready availability of disks in markets. In the market units, disks (SATA, SAS, Near line SAS, SSD, etc.) need to be segregated from manufacturing units located at multiple locations or sometimes even from other countries which involves management approvals, export regulations, shipping delays, etc, and all these factors contribute to a considerable amount of logistical delay. The proposed methodology remaps the disk replacement priority considering the logistical delay which can be predetermined according to the vendors expected time of delivery for a specific class of disk.

B. Estimated Selective Replication Overheads

Tier-1 and Tier-2 data centers built by small and medium size businesses usually practise no redundancy or partial redundancy [14] in terms of power, cooling, etc. and selective or grained replication. The proposed methodology considers the replication overheads that take place in this topology and remaps the disk replacement priority accordingly.

C. Safe Disk Failure Count in RAID Combinations

The advantageous features of implementing RAID in data centers are failsafe access and data protection. The proposed budget anticipation involves understanding the maximum number of disks failures supported in the current RAID configuration [15] which might be any one or a combination of various RAID technologies such as no redundancy in RAID 0, single drive failure allowed in RAID 1, single drive failure allowed in RAID 1E, single drive failure allowed in RAID 5, single drive failure allowed in RAID 5EE, two drive failure allowed in RAID 6, up to one disk failure allowed in each sub-array allowed in case of RAID 10, up to one disk failure allowed in each sub-array in case of RAID 50, up to two disk failure allowed in each sub-array in case of RAID 60 and remapping of individual disk replacement priority.

D. S.M.A.R.T Parameter Variability Analysis

The section-IV of this paper discusses a technique to analyse the S.M.A.R.T parameter differences between the observed and vendor specified statistics, the proposed operational budget anticipation methodology using the same differential information to keep track of the physical location of observed high disk failures, note the S.M.A.R.T parameter that formed the root cause for this failures and analyse the cost for tuning this parameter.

The proposed methodology uses the above explained factors (influencing financial planning) that contribute to a major share in operation cost and recalculating the disk replacement priorities based on the empirical relation noted below:

$$M_p = K * \frac{N_p * R_L}{L_D * R_p} \quad (5)$$

Where in Eq. (5), M_p denotes the Modified priority of disk replacement, K represents proportionality constant, N_p denoted the priority calculated without considering the operation cost factor, R_L denotes RAID safe disk failure factor, L_D denotes respective logistical delay and R_p denotes fine grained replication overhead.

TABLE 3 NOTED VARIOUS FINANCIAL FACTORS, CALCULATED LOGISTICAL DELAY, REPLICATION OVERHEAD, FAILSAFE DISKS IN RAID COMBINATION AND MEASURED MODIFIED PRIORITY

Disk no.	Replace Priority	Logistical delay	Replication overhead	RAID	Modified Priority
SATA1	3	< 1 hour	YES	NO	6
SSD1	6	> 48 hour	NO	NO	3

SATA4	4	< 1 hour	YES	YES	7
SATA6	5	< 1 hour	YES	NO	5
SATA8	1	< 1 hour	NO	YES	1
SSD7	7	< 24 hour	YES	NO	4
SSD0	2	< 5 hour	YES	NO	2

The above table demonstrates a sample manifestation of modified priority for disk replacement priorities in a heterogeneous cloud scenario containing variations financial and logistical factors in to consideration.

VI. CONCLUSIONS

Systems and procedures are integrated to have enhanced methodologies for disk replacement strategies and operational budget anticipation. In the above sections, the device characteristics and important factors discussed are requiring for enhancement of the current disk replacement strategies. Integrating these features in storage subsystems will enhance the systems and will be a step towards making smarter data centers. In general trends disk replacements are based on disk life span determined by vendors which may not be accurate - as the workload is dynamic and varies from time to time. The proposed technique proactively helps indicate the priorities of potential disks that may need replacements, adjust the priorities of the disk replacement based on the data centre financial planning and further aids to get a realistic required budget instead of guess estimation.

REFERENCES

- [1] S. N. Piramanayagam and Tow C. Chong, Developments in Data Storage: Materials perspective, John Wiley & Sons, New Jersey, 2012.
- [2] Sandeep Patil, Abhinay Nagpal, "Agile Disk Replacement Strategy for Data Cloud Environment Using Evolutionary Algorithms," Data Storage and Data Engineering (DSDE), pp. 287-291, 2010.
- [3] E. Grochowski, R. D. Halem, "Technological impact of magnetic hard disk drives on storage systems," IBM SYSTEMS JOURNAL, vol. 42, no. 2, pp. 338-346, 2003.
- [4] Kruti Sharma, Kavita R Singh, "Online Data Back up and Disaster Recovery Techniques in Cloud Computing: A Review," International Journal of Engineering and Innovative Technology (IJEIT), vol. 2, pp. 249-254.
- [5] H. Howie Huang, Shan Li, Alex Szalay and Andreas Terzis, "Performance Modeling and Analysis of Flash-based Storage Devices," IEEE 2011.
- [6] Un-Keun Yoon, "Intelligent Data Prefetching for Hybrid Flash-Disk Storage Using Sequential Pattern Mining Technique," Computer and Information Science (ICIS), pp. 280-285, 2010.
- [7] Kliazovich, D., Audzevich, Y., Khan, S.U., "GreenCloud: A Packet-Level Simulator of Energy-Aware Cloud Computing Data Centers," Global Telecommunications Conference (GLOBECOM 2010), pp. 1-5, 2010 IEEE.
- [8] Zbigniew Chlondowski, "S.M.A.R.T. Site: attributes reference table," S.M.A.R.T. Linux, Retrieved January 17, 2007.
- [9] Intel X18-M/X2 SATA Solid State Drive product manual, May 2009.
- [10] M. Shell. (2002) IEEEtran. homepage on CTAN.
- [11] Lawrence C. Evans, An Introduction to Stochastic Differential Equations version, Department of Mathematics, US Berkely.
- [12] R. K. Jain, S.R.K. Iyengar and M. K. Jain, Numerical Methods, New Age International, 2009.
- [13] Atsuyuki Okabe, Barry Boots, Kokichi Sugihara, Dr Sung Nok Chiu, Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, John Wiley & Sons, 2009.
- [14] <http://www.colocationamerica.com/data-center/tier-standards-overview.htm> [accessed 11/04/2013].
- [15] http://www.adaptec.com/en-us/_common/compatibility/_education/raid_level_compar_wp.htm [accessed 11/04/2013].