

A Migration Strategy based on Reliability Disk and Hotspot Data in Information Systems

Yin Yang^{1,a}, *Wei Liang^{2,b} and Wenyi Li^{1,c}

¹School of Accounting, Wuhan Textile University, Wuhan, China

²The 722 Institute of China Shipbuilding Industry Corporation, Wuhan, China

^acs_yangyin@hust.edu.cn, ^b38863835@qq.com, ^cnewsone260@126.com

Keywords: data migration; data prediction; disk health; Zipf

Abstract. This paper proposes a Migration strategy based on Reliability disk and Hotspot data called MRH. MRH uses reliability disk estimation to confirm disk reliability based on SMART technology. It computes the heat degree of data and the utilization of disk at the present time, so we can protect the whole data in one disk or current hotspots data. MRH uses hotspot data prediction to predict the hotspots data based on Zipf-like in the future, so we can also protect predicted new hotspots data. Our experiment shows that reliability disk estimation satisfies actual usage and hotspot data prediction achieves very high prediction rate.

Introduction

Disk vendors have been working hard to improve the reliability of products, and they have made significant step about the prediction of disk reliability: SMART. Pinheiro et al [1] present that out of all failed drives, over 36% of them have no count in any of the all remaining SMART parameters. In other words, models based only on those signals can never predict more than 70% of the failed drives. Hughes et al [2] and Hamerly and Elkan[3] have attempted to correlate SMART parameters with failure statistics. They derive failure models that achieve false positive rates of about 0.2%. From these results, prediction models achieve better accuracies for predictable failures.

There is a very strong and consistent correlation between high utilization and higher failure rates [4]. Although Pinheiro et al [1] indicate a much weaker correlation between utilization levels and failures, if the disk is low reliability disk, and in long time high utilization state, which can aggravate disk failure. So given how strongly correlated some SMART parameters were found to be with higher failure rates, and it is possible that models that use parameters beyond those provided by SMART could achieve significantly better accuracies. There is also a notable characteristic of data access pattern in storage system: 80% IO requests only access 20% data which known as "Pareto principle" [5]. This feature leads to hotspots data in information system, and these hotspots data has direct influence on the storage system's reliability.

Base on the observations above, in this paper, we propose a novel Migration strategy based on Reliability disk and Hotspot data called MRH, which includes Reliability Disk Estimation (RDE) and Hotspot Data Prediction (HDP). RDE confirms disk reliability based on SMART technology. By computing the heat degree of data and the disk utilization in terms of the data access frequency of all the data in one disk at the present time, HDP provides a hotspot data prediction model based on Zipf-like distribution which can predict the new hotspots data dynamically.

The Key Technology of MRH

Reliability disk estimation. The reliability disk estimation method is based on SMART technology. There are two health degrees: Parameter Health Degree (PHD) and Disk Health Degree (DHD). According to the data (DT), threshold (TH) and attributes value (AV), we can calculate the PHD. AV

*Corresponding Author

has been set to the maximum normal value as default. TH is the fault limit value set by the manufacturers. Users monitor the DT of the SMART parameter in regular time. We have:

$$PHD = 100 \times ((DT - TH) / (AV - TH)) \quad (1)$$

For this formula, the value of AV and TH does not change, but the value of DT change all the time, meanwhile the value of AV and DT can compare with the value of TH at the same time, so AV and TH as the denominator, DT and TH as numerator. DHD is calculated by the weight of a single parameter (the importance of parameters) and PHD.

We first select the SMART parameters which used by every manufacturer and have actual threshold. According to analyzing the significance of SMART parameters[1, 2] and observing the really actual test results of DT have changed of many disks, then we determine the weight of these parameters. The five parameters of RSC (Reallocated Sectors Count), SRC (Spin-up Retry Count), RER (Raw Read Error Rate), PCC (Power Cycle Count) and SER (Seek Error Rate) have greater influence on the disk failure and their DT can be compared with the TH. We confirm the RSC weight is 20%, SRC weight is 20%, RER weight is 40%, PCC weight is 10% and SER weight is 10%. The formula of disk health degree is:

$$DHD = 0.2 \times RSC + 0.2 \times SRC + 0.4 \times RER + 0.1 \times PCC + 0.1 \times SER \quad (2)$$

Hotspot data prediction. Zipf observed long time ago [6] that the distribution of word frequencies in English, the rank-r word has the probability Pr, the probability distribution is

$$P_r = \frac{C}{r^\alpha}, \sum_{r=1}^M P_r = 1 \quad (3)$$

From formula (3), we can get:

$$C = \left[\sum_{r=1}^M \frac{1}{r^\alpha} \right]^{-1} \quad (4)$$

If there have M data in one storage system, D_r represents the data ranked r. Meanwhile, f_r and P_r represent the access frequency and access probability of the data. The total number of access frequencies of all data is denoted by F, and the hotspot queue length is Q_h .

The value of α in one application environment is a constant, and from the formula (4), we can get:

$$C(kT) = \left[\sum_{j=1}^{M(kT)} \frac{1}{j^\alpha} \right]^{-1} = C(T) \quad (5)$$

At the time T and kT, if a data in the hotspot data, from the formula (3), the D_r and $D_{r'}$ are:

$$P_r(T) = \frac{C}{r^\alpha} = \frac{f_r(T)}{F(T)} \text{ and } P_{r'}(kT) = \frac{C}{r'^\alpha} = \frac{f_{r'}(kT)}{F(kT)} \quad (6)$$

From the formula (5) and formula (6), we can get:

$$\frac{r'^\alpha}{r^\alpha} = \frac{f_r(T)}{F(T)} \cdot \frac{F(kT)}{f_{r'}(kT)} \quad (7)$$

We assume that the value of M remain unchanged from time 0 to kT and the value of F is a linear function over a long period which only decided by time in the practical application environment. We can get:

$$\frac{F(kT)}{F(T)} = k \quad (8)$$

From the formula (7) and formula (8), we need get the ratio of $f_r(T)/f_{r'}(kT)$. The value of D_r doesn't increase from the time T to kT. We can get:

$$\frac{r'^\alpha}{r^\alpha} = \frac{f_r(T)}{f_{r'}(kT)} \cdot k \Rightarrow \frac{r'^\alpha}{r^\alpha} = k \quad (9)$$

From the formula (9), at $t=kT$, the rank of $D_{r'}$ is $r' = k^{\frac{1}{\alpha}} \cdot r$. We can get:

$$r' \leq Q_h \Rightarrow r \leq Q_h \cdot k^{-\frac{1}{\alpha}} \quad (10)$$

From the formula (10), we can predict whether D_r is a new hotspot data.

Evaluation methodology

We have use two workloads to drive our experiments: SPC I/O traces and standard I/O benchmark. We can get Financial-1, Financial-2 and Web from UMass Trace Repository and Storage Performance Council, and the write ratio of these three traces are about 77%, 18% and 0%, respectively.

From the formula (1) and (2), the reliability results of the ten disks are calculated. The DHD of the ten disks are 59, 75, 80, 79, 74, 81, 78, 77, 64 and 54, respectively. We can know D_1 - D_9 of ten disks is normal, but disk 10 is not reliable. According to the value of DHD in ten disks, the disk reliability is divided into the following three levels: if $\text{threshold} \leq \text{DHD} \leq 55$, the disk is low reliability disk (disk 10); if $55 < \text{DHD} \leq 65$, the disk is medium reliability disk (disk 1 and disk 9); if $\text{DHD} > 65$, the disk is high reliability disk.

Table 1 The rank of data and whether data is in the hotspot queue.

data number	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
D_r at the present	24	26	62	79	25	29	74	185	60	64
Hotspot?	yes	yes	no	no	yes	yes	no	no	no	no
data number	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}	d_{16}	d_{17}	d_{18}	d_{19}	d_{20}
D_r at the present	75	77	162	76	66	63	101	19	94	57
Hotspot?	no	no	no	no	no	no	no	yes	no	no

We combine with F_2 to interpret migration strategy of MRH. The data to be distributed among disks consists of 20 data. At the present time, the rank of data and whether data is in the hotspot queue are listed in Table 1, the utilization of disk are listed in Table 2. For HDP technology, in the future, the rank of data and whether data is still in the new hotspot queue are listed in Table 3; we use one prediction k_5T to analyze the hotspots. For simplicity of presentation, we assume: the 20 data have the same service time; consider one situation of data's placement in the ten disks, and one DHD to judge the reliability of disk, the threshold is 50.

Table 2 The placement of data and the utilization of disks.

disks number	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}
data placement	d_1 d_2 d_3	d_4 d_5	d_6 d_7	d_{16} d_9	d_{10}	d_{11} d_{12}	d_{13} d_{14}	d_{15}	d_8 d_{17} d_{18}	d_{19} d_{20}
Utilization	681.67	570	617.5	408	393	331.5	221	380	410	338.5

Table 3 The rank of data and whether data is still in the new hotspot queue.

data number	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
D_r in the	33	48	84	32	19	58	164	49	68	78
Hotspot?	yes	no	no	yes	yes	no	no	no	no	no
data number	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}	d_{16}	d_{17}	d_{18}	d_{19}	d_{20}
D_r in the	76	138	73	74	55	75	23	96	180	46
Hotspot?	no	no	no	no	no	no	yes	no	no	no

Our next experiment is to measure the accuracy of HDP. Fig. 1 shows that HDP achieves very high prediction rate such as F_2 and Web. The experiments show that HDP efficiently predicts the hotspots.

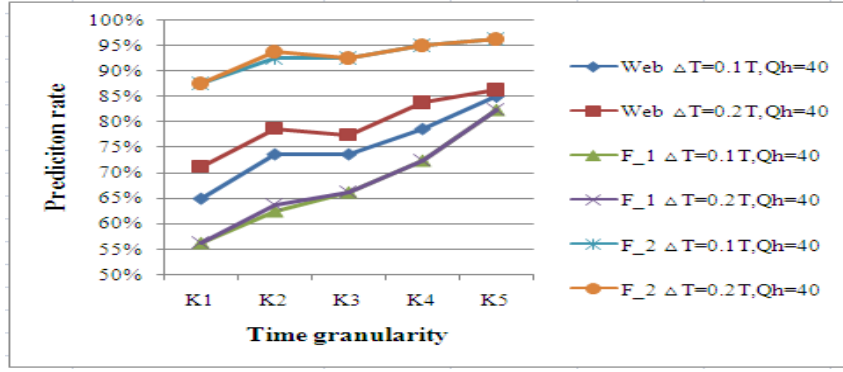


Fig. 1 Correct percentage of hotspots predicted by HDP.

The results of migration strategy from the Table 1 to Table 3 are described in detail in Fig. 2.

In Fig. 2, the low reliability disk is D_{10} ; the medium reliability disks are D_1 and D_9 . At the present time, the hotspot data includes: d_1, d_2, d_5, d_6 and d_{18} , and the high utilization disks are D_1 and D_3 . In the future, the new hotspot data includes: d_1, d_4, d_5 and d_{17} .

Step1: D_{10} is low reliability disk, so the whole data migrate to backup disk; Step2 and Step3: D_1 is medium reliability and high utilization disk, d_1 and d_2 are hotspots at the present time, so d_1 and d_2 from D_1 migrate to backup disk; Step4: D_1 is medium reliability disk and d_1 is hotspot in the future, so d_1 from D_1 migrate to backup disk; Step5: D_9 is medium reliability disk and d_{17} is hotspot in the future, so d_{17} from D_9 migrate to backup disk; Step6: Although D_9 is medium reliability disk and d_{18} is hotspot at the present time, D_9 is also low utilization, so the d_{18} don't need to migrate.

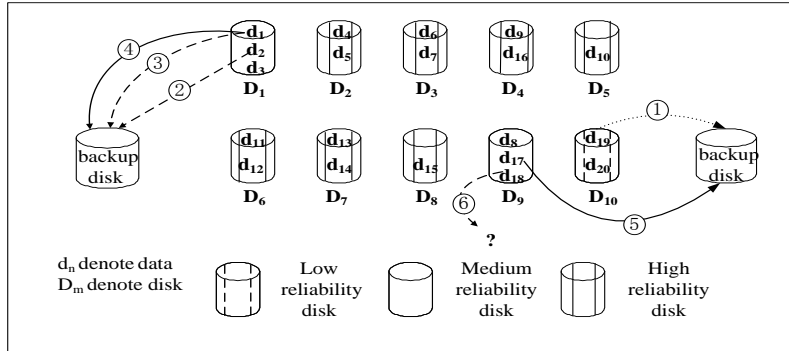


Fig. 2 Data placement, disk reliability and migration strategy of MRH.

Summary

In this paper we propose a novel Migration strategy based on Reliability disk and Hotspot data called MRH. Reliability Disk Estimation (RDE) confirms the reliability of disk. When the disk is in low reliability, we need data migration operation to protect the whole data in unhealthy disk. Because the medium reliability disk is in high utilization at present time or still in use from the present time to the future time may become the low reliability disk. Hotspot Data Prediction (HDP) predicts new hotspots data in the future, when the two situations which make medium reliability disk turns to low reliability disk happen; our method can protect the hotspots data in the medium reliability disk. We demonstrate RDE satisfies the actual usage and HDP achieves very high prediction rate, and have used a case to interpret migration strategy of MRH.

Acknowledgements

This work was financially supported by the Natural Science Foundation of Hubei Province (2015CFB235).

References

- [1] E. Pinheiro, W. D. Weber and L. A. Barroso, in: Proceeding 5th USENIX Conference on File and Storage Technologies (2007).
- [2] G. F. Hughes, J. F. Murray, K. Kreutz-Delgado and C. Elkan: IEEE Transactions on Reliability Vol. 51 (2002), p.350-357
- [3] G. Hamerly and C. Elkan, in: Proceedings of the Eighteenth International Conference on Machine Learning (2001).
- [4] C. Gerry, Seagate Technology Paper TP-338.1. (2000)
- [5] L. Cherkasova and M. Gupta: IEEE/ACM Transactions on Networking Vol. 12 (2004), p.781-794
- [6] G. K. Zipf: *Human behavior and the principle of least-effort* (Cambridge, Mass, Addison-Wesley Press, England 1949).