

Data mining for information storage reliability assessment by relative values

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2016 Authors. The data ambiguity problem for heterogeneous sets of equipment reliability indicators is considered. In fact, the same manufacturers do not always unambiguously fill the SMART parameters with the corresponding values for their different models of hard disk drives. In addition, some of the parameters are sometimes empty, while the other parameters have only zero values. The scientific task of the research consists in the need to define such a set of parameters that will allow us to obtain a comparative assessment of the reliability of each individual storage device of any model of any manufacturer for its timely replacement. The following conditions were used to select the parameters suitable for evaluating their relative values: 1) The parameter values for normally operating drives should always be greater or lower than for the failed ones; 2) The monotonicity of changes in the values of parameters in the series should be observed: normally working, withdrawn prematurely, failed; 3) The first two conditions must be fulfilled both in general and in particular, for example, for the drives of each brand separately. Separate averaging of the values for normally operating, early decommissioned and failed storage media was performed. The maximum of these three values was taken as 100%. The relative distribution of values for each parameter was studied. Five parameters were selected (5 - "Reallocated sectors count", 7 - "Seek error rate", 184 - "End-t-end error", 196 - "Reallocation event count", 197 - "Current pending sector count", plus another four (1 - "Raw read error rate", 10 - "Spin-up retry counts", 187 - "Reported uncorrectable errors", 198 - "Uncorrectable sector counts"), which require more careful analysis, and one (194 - "Hard disk assembly temperature") for prospective use in solid-state drives, as a result of the relative value study of their suitability for use upon evaluating the reliability of data storage devices. © 2018 Authors.

Keywords

Estimation, Information, Parameter, Reliability, Storage

References

- [1] S.M.A.R.T. From Wikipedia, the free encyclopedia. URL: <https://en.wikipedia.org/wiki/S.M.A.R.T>. Checked on 10/03/2018
- [2] Hard Drive Data and Stats / Backblaze. URL: <https://www.backblaze.com/b2/hard-drive-test-data.html>. Checked on 10/03/2018
- [3] Beach B. Reliability Data Set For 41, 000 Hard Drives Now Open Source. URL: <https://www.backblaze.com/blog/hard-drive-datafeb2015/>. Checked on 10/03/2018

- [4] Nasyrov R.I., Nasyrov I.N., Timergaliev S.N. Cluster analysis of information storage devices that failed during operation in a large data center // Information technology. Automation. Updating and solving the problems of training highly qualified personnel (ITAP-2017): materials of the international scientific-practical conference on 19 May, 2017.-Naberezhnye Chelny: KFU, 2017.-P3. 95-102. URL: <https://cloud.mail.ru/public/LBcn/phxM8D1S5>
- [5] Beach B. How long do disk drives last? URL: <https://www.backblaze.com/blog/how-long-do-disk-drives-last/>. Checked on 10/03/2018
- [6] Nasyrov R.I. Grading indicators for information storage devices by reliability degree // VIII Kama Readings: a collection of reports of the All-Russian Scientific and Practical Conference on 22 April, 2016-In 2 parts-Part 1.- Naberezhnye Chelny: CPI NCHI KFU, 2016-124.URL: <https://cloud.mail.ru/public/JPMr/8qr2jXKFC>
- [7] Klein A. Hard Drive Reliability Stats for Q1 2015. URL: <https://www.backblaze.com/blog/hard-drive-reliability-q1-2015/>. Checked on 10/03/2018
- [8] Pinheiro E., Weber W.-D., Barroso L.A. Failure Trends in a Large Disk Drive Population // The Proceedings of the 5th USENIX Conference on File and Storage Technologies (FAST'07). San Jose, California, USA, February 13-16, 2007. URL: http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/ru//archive/disk_failures.pdf
- [9] Nasyrov R.I., Nasyrov I.N. Choice of parameters for the method of forecasting the reliability of data storage devices in large data centers. // Quality. Innovation. Education.-2017.-No. 5 (144).-Pp. 40-48. URL: <https://library.ru/item.asp?id=29869743>
- [10] Rincón C.A.C., Paris J.-F., Vilalta R., Cheng A.M.K., Long D.D.E. Disk failure prediction in heterogeneous environments // Proceedings of the International Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS 2017. Seattle, WA, USA, July 9-12, 2017. URL: <http://ieeexplore.ieee.org/document/8046776/>
- [11] Qian J., Skelton S., Moore J., Jiang H. P3: Priority based proactive prediction for soon-to-fail disks // Proceedings of the 10th IEEE International Conference on Networking, Architecture and Storage, NAS 2015. Boston, MA, USA, August 6-7, 2015.-7255224.-p. 81-86. URL: <http://ieeexplore.ieee.org/document/7255224/>
- [12] Botezatu M.M., Giurgiu I., Bogojeska J., Wiesmann D. Predicting disk replacement towards reliable data centers // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. San Francisco, California, USA, August 13-17, 2016.-p. 39-48. URL: <https://dl.acm.org/citation.cfm?doid=2939672.2939699>
- [13] Chaves I.C., de Paula M.R.P., Leite L.G.M., Queiroz L., Pordeus J.P., Machado J.C. BaNHFaP: A Bayesian Network Based Failure Prediction Approach for Hard Disk Drives // Proceedings of the 5th Brazilian Conference on Intelligent Systems, BRACIS 2016. Recife, Pernambuco, BR, October 9-12, 2016.-7839624.-p. 427-432. URL: <http://ieeexplore.ieee.org/document/7839624/>
- [14] Gaber S., Ben-Harush O., Savir A. Predicting HDD failures from compound SMART attributes // Proceedings of the 10th ACM International Systems and Storage Conference, SYSTOR '17. Haifa, Israel, May 22-24, 2017.- Article No. 31. URL: <https://dl.acm.org/citation.cfm?doid=3078468.3081875>
- [15] Gopalakrishnan P.K., Behdad S. Usage of product lifecycle data to detect hard disk drives failure factors // Proceedings of the ASME International Design Engineering Technical Conference. Cleveland, Ohio, USA, August 6-9, 2017. URL: <http://proceedings.asmedigitalcollection.asme.org/proceeding.aspx?articleid=2662132>