Hard disk drive failure rates in datacenters

Slađana Đurašević
Faculty of technical sciences Čačak
University of Kragujevac
Čačak, Serbia
ORCID: 0000-0002-3598-5781

Uroš Pešović
Faculty of technical sciences Čačak
University of Kragujevac
Čačak, Serbia
ORCID: 0000-0001-8722-6544

Borislav Đorđević
Institute Mihajlo Pupin
Belgrade, Serbia
ORCID: 0000-0002-6145-4490

Vanja Luković
Faculty of technical sciences Čačak
University of Kragujevac
Čačak, Serbia
ORCID: 0000-0002-1887-6102

Abstract— **Data centers store large quantities of data on thousands of hard disk drives (HDD) connected into redundant arrays to prevent data loss in case of drive failures. Hard disk drives are prone to failures due to wear of their electromechanical components, defects, and external influences. This paper analyzes failure rates for several HDD models based on the SMART data set from the large data center. Obtained results show annualized failure rate and useful HDD life which represents the point from which failure rates increase due to mechanical wear.**

Keywords—Hard disk drive, failure rate, MTBF, AFR, SMART

I. INTRODUCTION

Data centers facilities are used to store an ever-growing amount of data. Instead of using traditional offline data storage, these data centers provide instant access to the store data thanks to the advancements in ICT technologies. Datacenters rely on the hard disk drive for storage, proven storage technology present for six decades. Even this mature, this technology is constantly evolving, improving the storage density. When compared to solid-state drives hard drives provide greater capacity, power price per stored data, and superb write endurance when compared to SSD, making them the prime choice for use in data centers [1,2]. Thanks to constant technological improvements, hard disk drive technology provides answers to the ever-growing demand for drive capacity even in the future. Currently, HDD manufacturers use SMR (Shingled Magnetic Recording) technology which overlays multiple tracks one on top of each other, significantly increasing storage density when compared with traditional perpendicular magnetic recording which creates guard bands between neighboring tracks. Future technologies rely on energy assistant magnetic recording such are HAMR (Heat Assisted Magnetic Recording) and MAMR (Microwave-Assisted Magnetic Recording) which will further increase storage density on hard drives.

Hard disk drives (HDDs) are the only electromechanical component in the computer system and due to their mechanical design are more susceptible to failure than other computer components. Since HDD is used as the primary storage of user data its failure leads to permanent data loss. Regular users need to perform regular data backups to prevent loss of data due to HDD failure. Datacenter used the RAID 6 scheme which uses data redundancy and could operate even if two HDD in the array fail, achieving very high reliability.

SMART (Self-Monitoring, Analysis and Reporting Technology) is used to monitor various indicators during the HDD operation, such as drive temperature, power-on hours, the number of load/unload cycles, the number of damaged sectors [3]. These SMART indicators can be used to indicate a possible imminent drive failure, when some of these exceed manufacturer set thresholds. Thus user can perform necessary actions to create backup in order to avoid loss of data. SMART parameters can be used only to detect predictable failures which are caused by processes which slowly degrade drive performance due mechanical wear and gradual degradation of storage surfaces. Indicators of these processes can be monitored to determine when such failures are becoming more likely. Other classes of failures are unpredictable failures which represent sudden drive failures, which occur due defective electronic components or sudden mechanical failures caused by improper handling or external force.

In this paper we analyzed failure rates for several HDD models based on the SMART data set from Backblaze data center. Obtained results show annular failure rate and useful HDD life which represents the point from which failure rates increase due to mechanical wear.

II. FAILURE RATE

The failure rate represents the statistically determined probability of failure of a component in a unit of time. Since it is a unit that has a very small value, it is usually presented in the form of a reciprocal value of MTBF (Mean Time Between Failures), which represents the time between two failures of a particular component in the system [4]. Typical MTBF parameter values for today's hard disk drives range from 100,000 for disks intended for personal computers to 2,500,000 for high-performance disks intended for servers. These data are obtained based on statistical measurements whereby a certain population of disks is tested under the conditions declared by the manufacturer (number of working hours per year, number of load/unload cycles, maximum read/write load). MTBF is determined as the quotient of the total number of hours of operation of the disk population, divided by the number of failures that occurred during the test. For example, if a population of 10,000 disks would run 1000 h, and in that case, four disks would fail, the MTBF would be 10,000x1,000 / 4 = 2,500,000 hours.

Such high MTBF values, as 2.5 million hours, could lead to the conclusion that the hard disk can, on average, run continuously for 285 years without failure, but this parameter does not predict the lifespan of an individual hard disk. The true meaning of this parameter is valid if the warranty period of the disk is 5 years for enterprise drives, which would mean that if the hard disk is replaced with a new one every 5 years, in 285 years there will be one failure of such a disk. Given the continuous development of computer components, it is questionable whether after 5 years the same disk model will be available with which it will be possible to replace the previous one to reach the declared MTBF. Recently, the

AFR (Annualized Failure Rate) parameter has been used, which is an indicator of the probability that the hard disk will fail during one year of use. AFR parameter can be calculated when MTBF is divided by 8766 hours present in one year, as shown in (1). This parameter also does not predict the lifetime of an individual disk but is useful for predicting disk failure rates in large data centers. Based on this parameter, it is possible to predict the number of failures of the disk population and plan the purchase of replacement disks accordingly.

$$AFR = \frac{MTBF}{8766} \quad (1)$$

The rate of disc failure changes over the lifetime of the disc population, and it follows the trend of the bathtub curve. Fig. 1. shows, bathtub curve which can be divided into periods: infant mortality, useful life, and wear-out period. The first period known as infant mortality has an increased rate of disc failure due to defects occurring during production that manifest at the very beginning of life. Most of these defects are detected by extensive tests in the manufacturer's plant, where defective discs do not pass quality control and are not delivered to users. The flat part of the curve represents the period of operation of the disk, which usually coincides with the period of warranty. During this period, the disk failure rate is constant and significantly lower than in other periods and is most often the result of external influences. The last phase of the disk life is the wear-out phase, in which there is a sudden increase in the failure rate due to the wear of the magnetic disk components. The beginning of this phase is an indication that it is necessary to start the process of replacing disks preventively to avoid data loss due to successive disk failures.

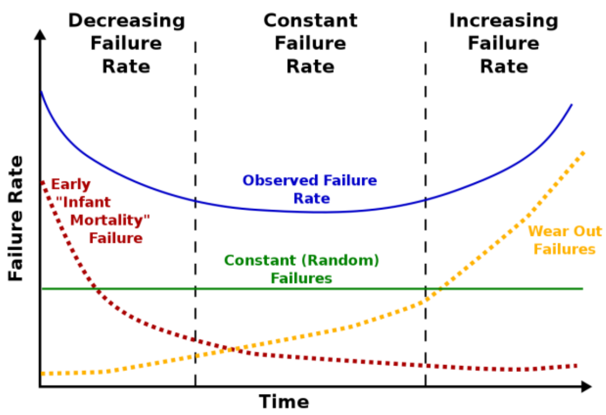

Fig. 1. Bathtub curve [5]

III. RESULTS

Backblaze is a cloud storage and data backup company, founded in 2007 targeted at both business and personal markets. It started sharing SMART statistics of the HDDs operating in their data center in 2013, publishing the SMART statistics data quarterly every year [6]. Data is published in form of .csv (comma separated values) files, where each file represents a collection of all drive parameters for a particular day of the year. These parameters include the date, serial number, model, capacity, failure indicator, and the number of SMART parameters presented in raw and normalized form. These files were imported into the SQLITE3 database, for easier access, where data used in this research is extracted using SQL queries. In our research, we analyzed failure rates for several Western Digital (HGST) and Seagate HDD models with capacities of 4 GB, 8 GB, and 12 GB, which operated in period from 10th April 2013 to 31th December 2020. This data includes date, serial number, model, failure indicator, and raw value of SMART 9 indicator which contains information about drives power-on hours. All models are enterprise-class drives intended for 24/7 operation (8760 operating hours per year) with a five-year warranty, except models ST4000DM000 and ST8000DM002 which are desktop-class drives intended for 8/5 operation (2400 operating hours per year) with a two-year warranty. The number of drives used in the research and their average age is shown in Table 1.

TABLE I. HDD MODELS USED IN SREARCH

Model	Capacity	No. drives	Av. age
HMS5C4040ALE640	4 GB	8.718	4,09
HMS5C4040BLE640	4 GB	16.345	4,10
ST4000DM000	4 GB	37.005	4,45
ST8000DM002	8 GB	10.220	4,16
ST8000NM0055	8 GB	15.038	3,37
ST12000NM0007	12 GB	38.738	2,34

For each drive model used in this research, we performed an SQL query on the database which found the instance for every serial number of the certain model which had a maximum value for drives power-on hours. When power-on hours of all instances of certain models are added, we obtained total operating hours of drive population for that model. Also using another SQL query we counted the number of all instances of certain drive model which had failed, by observing failure indicator attribute set to one. MTBF parameter for every drive model is obtained when the total operating hours of the hard disk drive population are divided by the number of failed drives, and results are shown in Table 2.

TABLE II. MTBF OF DRIVES USED IN RESEARCH

Model	Oper. hours	Failures	MTBF
HMS5C4040ALE640	312.251.116	170	1.836.771
HMS5C4040BLE640	587.612.662	274	2.144.571
ST4000DM000	1.443.678.938	4.106	351.602
ST8000DM002	372.458.895	440	846.497
ST8000NM0055	444.588.697	607	732.436
ST12000NM0007	795.282.167	1.788	444.788

According to obtained results, we can observe that the MTBF parameter for Western Digital (HGST) drive models is close to the manufacturer stated MTBF of 2500000 hours for these enterprise drives. Seagate desktop drives ST4000DM000 and ST8000DM002 surpassed the typical MTBF set for desktop drives to the 300000 hours, even they are used in 24/7 operation which they are not intended to operate. On the other hand, Seagate enterprise drive models ST8000NM0055 and ST12000NM0007 achieved fairy poor MTBF for the enterprise-class drives, which is three to five times shorter when compared to Western Digital drives.

Furthermore, based on the queried data we analyzed failure rates of these models during the population lifetime. For each failed drive in this population, we determined the number of operating drives that exceeded its power-on hours when the drive had failed. Based on the ratio of one failed drive and the number of drives that continue to operate we calculated the failure rate for the hard disk drive population. Fig. 2 shows failure rates for two models of Western Digital (HGST) enterprise drives. Both drives had a constant failure rate which started to increase due to wear of drive components in the fifth year of operation which determines the usual lifetime of these two models. Fig. 3 shows failure rates for Seagate desktop models. Both drives had constant failure rates which started to increase due to wear of drive components in the fifth and sixth year of operation which determines the usual lifetime of these two models.

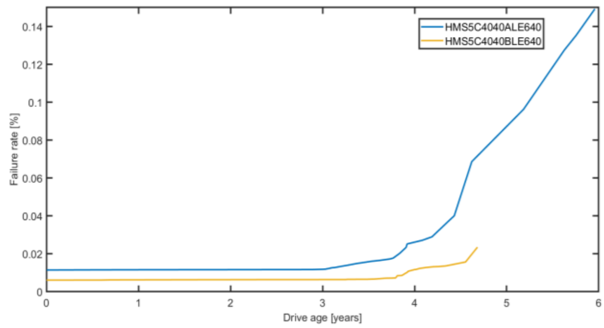

Fig. 2. Failure rate of Western Digital (HGST) enterprise drives

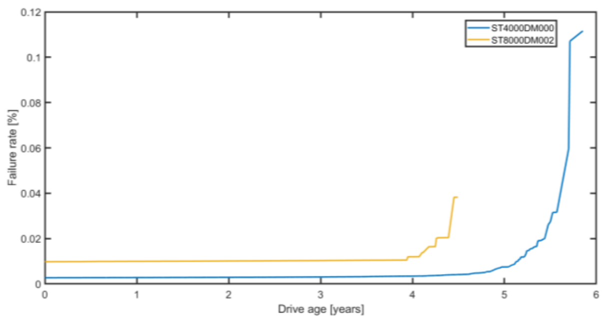

Fig. 3. Failure rate of Seagate desktop drives

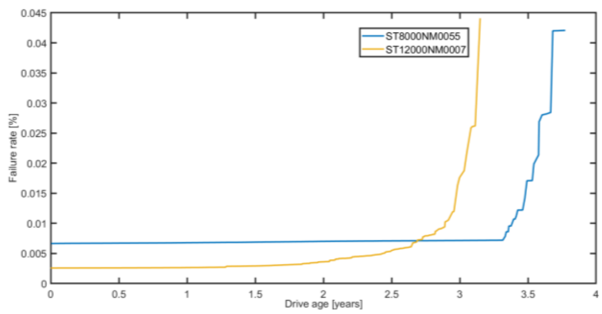

Fig. 4. Failure rate of Seagate enterprise drives

Finally, Fig. 4 shows failure rates for two models of Seagate enterprise drives. Both drives had a constant failure rate which started to increase due to wear of drive components in the third and fourth year, but it was much less when compared with previous drive models. The reason for this is the lack of more data, since these two drive models operated on average three years in the data center, compared to four years for other drive models.

In the end, the AFR parameter is calculated as a sum of the failure rate of a certain drive model during one year. Since AFR can be calculated quite coarse due to the small number of operating years, we first calculated the monthly failure rate and then calculated AFR as moving average on twelve consecutive monthly failure rates. Results in Fig. 5. shows that two models of Western Digital (HGST) enterprise drives have the lowest AFR, around 0.5% failed drives per year. We can observe the early infant mortality period during the first year of operation which gradually decreased from 1% at the beginning to 0,5% in the first year and remained constant for four years of operation. AFR stated to increase near the end of HDD lifetime which is around the fifth year.

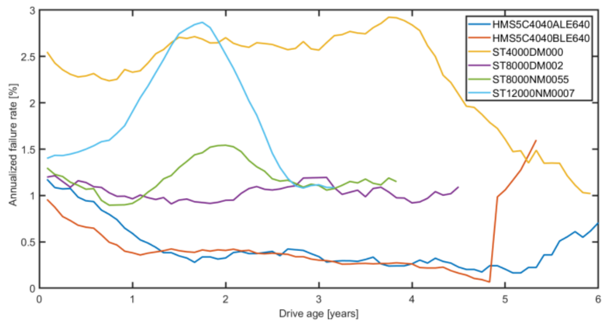

Fig. 5. Annualized failure rate drives

As expected for two of the Seagate desktop models AFR were significantly higher than for WD enterprise drives, especially for the ST4000DM000 drive which has an average AFR of 2.5% which in the last two years started to decrease to 1%. The reason for this high AFR is because these desktop drives are not intended for 24/7 operation as they are used in the data center, but they are intended to be used in desktop computers, and drives are rated for 2400 operating hours per year. The other two Seagate enterprise drives had AFR which was a bit higher than 1%, except abnormally high AFT for drive model ST12000NM0007 which peaked at 2.5% AFR in the second year, and then returned to a value of 1%.

In the end, we compared the calculated value of AFR base on equation 1 and the value which is determined as a mean value of AFR during drive lifetime. Results presented in Table 3. show that calculated AFT is consistent with AFR obtained by measurements for all drive models and results differ up to 19%.

TABLE III. CALCULATED AND MEASURED AFR OF DRIVES

Model	MTBF	Calculated AFR	Measured AFR
HMS5C4040ALE640	1.836.771	0,48	0,44
HMS5C4040BLE640	2.144.571	0,41	0,51
ST4000DM000	351.602	2,49	2,30
ST8000DM002	846.497	1,04	1,04
ST8000NM0055	732.436	1,20	1,18
ST12000NM0007	444.788	1,97	1,85

IV. CONCLUSION

In this research, we performed the analysis of the failure rate for several hard disk drive models operating in a data center. The result showed that typical AFR ranges from 0,5% to 1% for enterprise drives which is equivalent to 1 failure per year in population from 100 to 200 drives. Results have also shown that the useful life of a hard disk drive is around five years of continuous operation. Further research will be focused on SSD failure rate when these data will be publicly available from large datacenters.

ACKNOWLEDGMENT

The research in this paper is part of the project 451-03-9/2021-14/200132 funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] J. Byron, D. Long, E. Miller, “Measuring the Cost of Reliability in Archival Systems”,36th International Conference on Massive Storage Systems and Technology (MSST 2020), 29-30 October 2020, Santa Clara, CA, USA

[2] R. Pinciroli, L. Yang, J. Alter, E. Smirni, “The Life and Death of SSDs and HDDs: Similarities, Differences, and Prediction Models“, DOI: arxiv-2012.12373

[3] S. Djurasevic, B. Djordjevic, “Anomaly detection using SMART indicators for hard disk drive failure prediction”,

[4] B. Schroeder, G. Gibson, “Disk failures in the real world: What does an MTTF of 1,000,000 hours mean to you? “, 5th USENIX Conference on File and USENIX Association Storage TechnologiesFebruary 2007, San Jose, CA, USA

[5] Seagate, Understanding Reliability Metrics, Technology Paper, August 2015

[6] BackBlaze SMART data set, https://www.backblaze.com/b2/hard-drive-test-data.html, accessed 1th April 2021.

Implementation of Monetization Practices in Mobile Applications

Vlatko Efremovski
School of Computer Science and Information Technology
University American College Skopje
Skopje, Macedonia
vlatko.efremovski@gmail.com

Adrijan Božinovski
School of Computer Science and Information Technology
University American College Skopje
Skopje, Macedonia
bozinovski@uacs.edu.mk

***Abstract*— Our everyday life is increasingly influenced by technological improvements. This generation is mostly affected by the daily use of smartphones, tablets and other types of mobile devices. These devices have a huge number of different applications, which are available through the digital distribution platforms, i.e., App Store for users of Apple products and Google Play for users of Android products. Users often want to have utilitarian and practical benefits from applications, presented in a visually pleasing and high-quality user-interface which is easy to use. Although most applications are available for free and are often used as non-profitable promotional marketing tools, a large number of applications use a business model to generate profit. This paper will be based on the study of the advantages and disadvantages of different monetization strategies and business models and their use and impact on the applications that serve them. It will also includes original research, which indicates the willingness of users in Macedonia to pay for mobile applications and the reasons which fuel their decisions to do so. The gathered data suggests that there is room for improvement and potential growth when it comes to the usage and implementation of monetization strategies, in particular within the Macedonian society. Relevant targeted ads, affordable apps with significant value for money features and flexible subscription models are just some of the areas which can be improved on.**

Keywords—Monetization, Mobile Development, Applications

I. INTRODUCTION

Over the course of just a few years (2016 to the present), there has been a 113% growth in revenue from mobile app downloads, in-app marketing and in-app purchases with the revenue potentially reaching almost $190 billion by the end of 2020 [1]. While this revenue growth of 113% seems impressive, it just serves to prove how much more users are connected to their devices, and feel generally more freedom and security when it comes to the consumer power that can be invested in them.

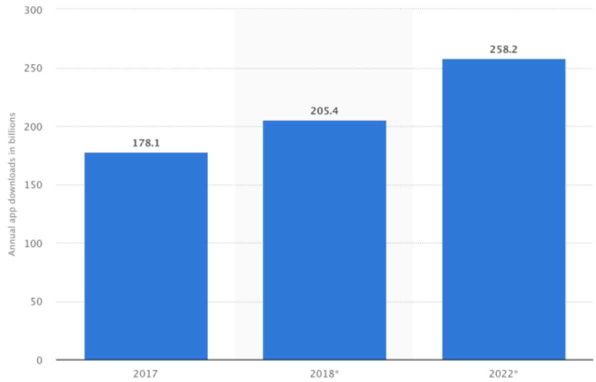

Fig. 1. Downloads of applications presented in billions [1]

The market leader is the first social network that breached the barrier of one billion registered user accounts, i.e., Facebook, which by the end of 2019 had 2.41 billion monthly active users [2]. It is estimated that more than 2 billion people use some kind of social network application, and this number is constantly growing because the companies that develop applications are adapting to the requirements, i.e., offering greater stability, security, multilingualism and an increasingly simple way to use. In addition, popular social networks have a great ratio between the number of users and the duration of using the applications during the day. Mobile games offer a wide range of mobile application downloads and stand out as the most popular category for application developers. For example, almost 25% of all mobile applications are gaming applications – 2.5 times more than business applications [1]. The most common applications developed fall into the gaming category, however only 65% of smartphone users have at least one such application on their device. On the other hand, more than 90% of all smartphone users have some type of a social networking or chat application, web browser or entertainment app installed on their devices. It is clear that the demand for gaming applications is not in line with its supply. There are 2.71 billion smartphone users in the world today. This means that in the world of wireless connections, 35.13% of the world's population today has a smartphone. According to GSMA real-