

Вестник НГУЭУ. 2023. № 3. С. 132–143
Vestnik NSUEM. 2023. No. 3. P. 132–143

Научная статья
УДК 004.052:004.33:658.155
DOI: 10.34020/2073-6495-2023-3-132-143

РАСПРЕДЕЛЕНИЕ ОТКАЗАВШИХ НАКОПИТЕЛЕЙ ИНФОРМАЦИИ ОТ ВРЕМЕНИ ЭКСПЛУАТАЦИИ В DATA-ЦЕНТРАХ

Насыров Искандар Наилович¹, Насыров Ильдар Искандарович²,
Насыров Рустам Искандарович³

¹ Казанский (Приволжский) федеральный университет

² ООО «Телеком Интеграция»

³ ООО «Газпромнефть – Цифровые решения»

¹ ecoseti@yandex.ru

² ildarec@mail.ru

³ rinasyrov@gmail.com

Аннотация. Изучено распределение вышедших из строя HDD и SSD накопителей информации различных производителей от времени эксплуатации в data-центрах одной из крупнейших в мире компаний облачного хранения Backblaze за длительный период. В результате исследования определены временные диапазоны с максимальной долей отказавших накопителей. Обнаружено, что наряду с максимумами, соответствующими заявленным техническим характеристикам накопителей, практически для всех торговых марок имеется большое количество отказов также и в начальный период времени эксплуатации. Предлагается способ учета данного обстоятельства при оценке эффективности деятельности data-центров.

Ключевые слова: большие данные, накопитель информации, data-центр, время эксплуатации, надежность, эффективность

Для цитирования: Насыров И.Н., Насыров И.И., Насыров Р.И. Распределение отказавших накопителей информации от времени эксплуатации в data-центрах // Вестник НГУЭУ. 2023. № 3. С. 132–143. DOI: 10.34020/2073-6495-2023-3-132-143.

Original article

DISTRIBUTION OF FAILED DATA STORAGE DEVICES FROM OPERATION TIME IN DATA CENTERS

Nasyrov Iskandar N.¹, Nasyrov Ildar I.², Nasyrov Rustam I.³

¹ Kazan (Volga region) Federal University

² Telecom Integration LLC

³ Gazpromneft – Digital Solutions LLC

¹ ecoseti@yandex.ru

² ildarec@mail.ru

³ rinasyrov@gmail.com

Abstract. The distribution of failed HDD and SSD drives from various manufacturers by operation time in one of the world's largest cloud storage companies Backblaze data centers for a long period has been studied. As a result of the study, time ranges with the maximum share of failed drives were determined. It was found that along with the maxima corresponding to the declared drives technical characteristics, for almost all brands there is a large number of failures also in the initial operation period. A method of taking this circumstance into account when evaluating the data centers' effectiveness is proposed.

Keywords: big data, data storage device, data center, operation time, reliability, effectiveness

For citation: Nasyrov I.N., Nasyrov I.I., Nasyrov R.I. Distribution of failed data storage devices from operation time in data centers. *Vestnik NSUEM*. 2023; (3): 132–143. (In Russ.). DOI: 10.34020/2073-6495-2023-3-132-143.

Введение

Актуальность исследования обусловлена необходимостью оценки эффективности функционирования data-центров, представляющих из себя системы централизованного хранения создаваемых в цифровой экономике больших данных [5]. Для этого надо изучить состав и динамику затрат, требующихся на замену вышедших из строя накопителей информации. Исходя из этого предлагается рассмотреть распределение отказавших накопителей на жестких дисках HDD (hard disk drive) и твердотельных на микросхемах SSD (solid state drive) от времени эксплуатации, сгруппировав их по разным торговым маркам различных производителей.

Научной проблемой является отсутствие детализированного учета затрат на замену отказавших накопителей при оценке эффективности деятельности data-центров. Цель исследования состоит в разработке способа детализации учета данного обстоятельства. Соответственно задачами исследования являются разработка порядка раздельного учета затрат на материалы, оплату труда, амортизацию и обесценивание.

Методы

Информационной базой исследования послужили ежедневно записываемые SMART-данные (self-monitoring, analysis and reporting technology – технология самоконтроля, анализа и отчетности) накопителей, находящи-

еся в свободном доступе на сайте одной из крупнейших в мире компаний облачного хранения Backblaze (<https://www.backblaze.com/b2/hard-drive-test-data.html>) [8]. Они удовлетворяют всем требованиям для оценки состояния накопителей [10], в связи с чем исследователи со всего мира активно используют их в своей работе в самых разных областях [6]. В качестве методов исследования выбраны группировка и визуализация распределения по времени эксплуатации продолжающих функционировать, снятых досрочно и отказавших накопителей информации за период с 10.04.2013 по 30.06.2022 г. Исходя из имеющего место различия по торговым маркам распределение по значениям параметров состояния изучалось дифференцированно. Число вышедших из строя единиц оборудования рассматривалось в процентах к суммарному числу продолжающих функционировать, снятых досрочно и отказавших накопителей.

Результаты

Число продолжающих работать, снятых досрочно и отказавших накопителей каждой торговой марки, которые есть в наличии, а также их процентная доля от общего числа накопителей приведены в табл. 1.

Таблица 1

Число накопителей разных торговых марок в штуках и их доля, %
Drives number of different brands in pieces and their share, %

Торговая марка	Всего, шт.	Работа, шт.	Досрочно, шт.	Отказ, шт.	Работа, %	Досрочно, %	Отказ, %
00MD00	2	0	2	0	0,00	100,00	0,00
HGST	53 405	44 224	8352	829	82,81	15,64	1,55
Hitachi	13 246	6	12 699	541	0,05	95,87	4,08
Samsung	18	0	17	1	0,00	94,44	5,56
ST	179 810	108 261	59 451	12 098	60,21	33,06	6,73
Toshiba	53 230	51 828	277	1125	97,37	0,52	2,11
WDC	16 419	12 608	3279	532	76,79	19,97	3,24
Всего HDD, шт.	316 130	216 927	84 077	15 126			
CT	294	272	21	1	92,52	7,14	0,34
DELLBOSS	351	351	0	0	100,00	0,00	0,00
HP	110	0	2	108	0,00	1,82	98,18
MTFDDAV	99	89	1	9	89,90	1,01	9,09
Samsung	10	0	0	10	0,00	0,00	100,00
Seagate	1828	1804	3	21	98,69	0,16	1,15
SSDSCKKB	4	0	4	0	0,00	100,00	0,00
Всего SSD, шт.	2696	2516	31	149			

Изменение суммарных долей работающих, снятых досрочно и отказавших HDD и SSD накопителей от времени эксплуатации, представлено на рис. 1 и 2.

Шаг по времени эксплуатации выбран равным 1000 ч. Сначала вычислялась доля накопителей, попадающих во временные диапазоны каждого шага раздельно по значениям параметров состояния и по торговым маркам.

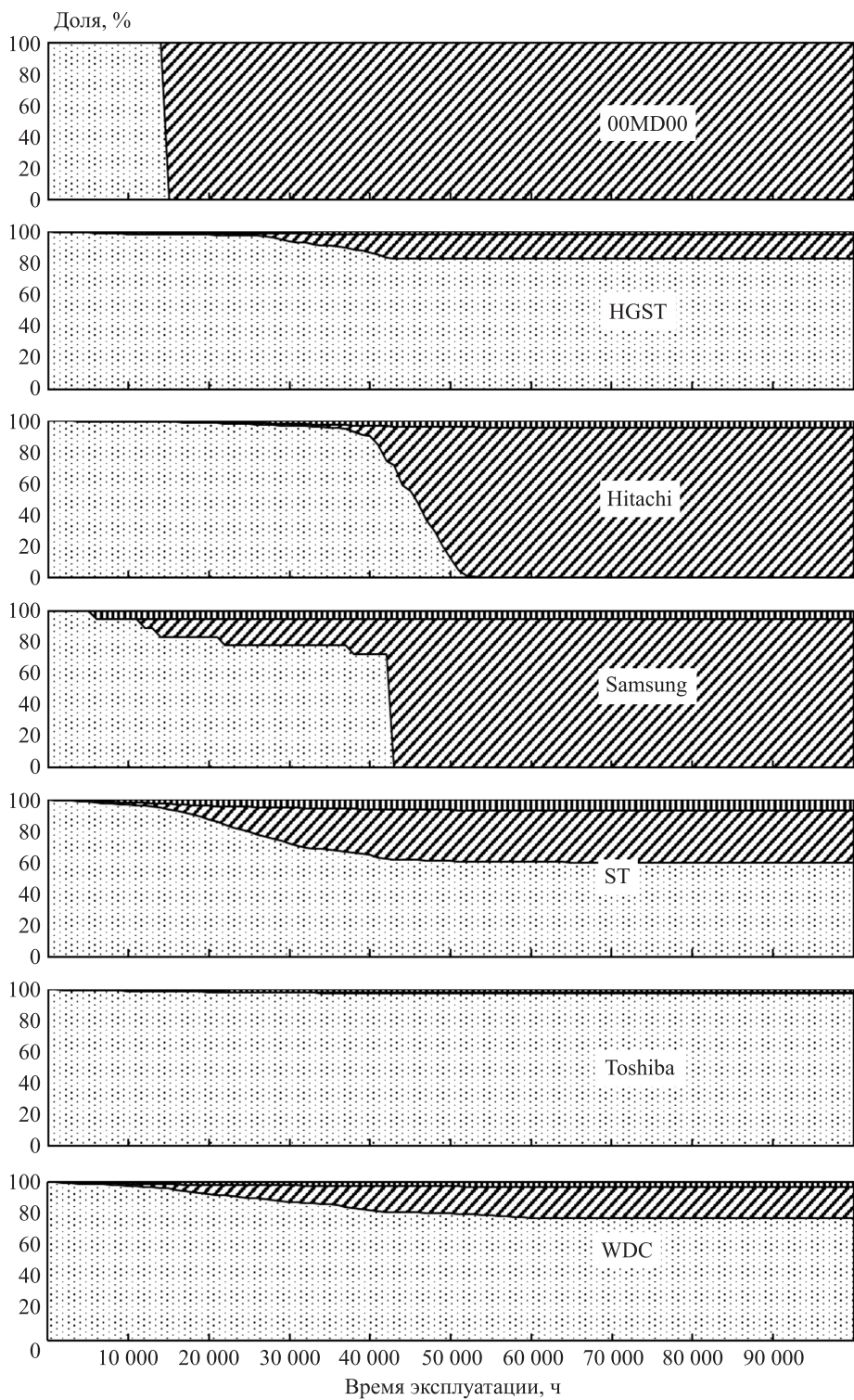


Рис. 1. Изменение доли работающих (точки), снятых досрочно (наклонные линии) и отказавших (вертикальные линии) HDD накопителей от времени эксплуатации, %
Change in the proportion of working (dots), prematurely removed (slanted lines) and failed (vertical lines) HDD drives from operation time, %

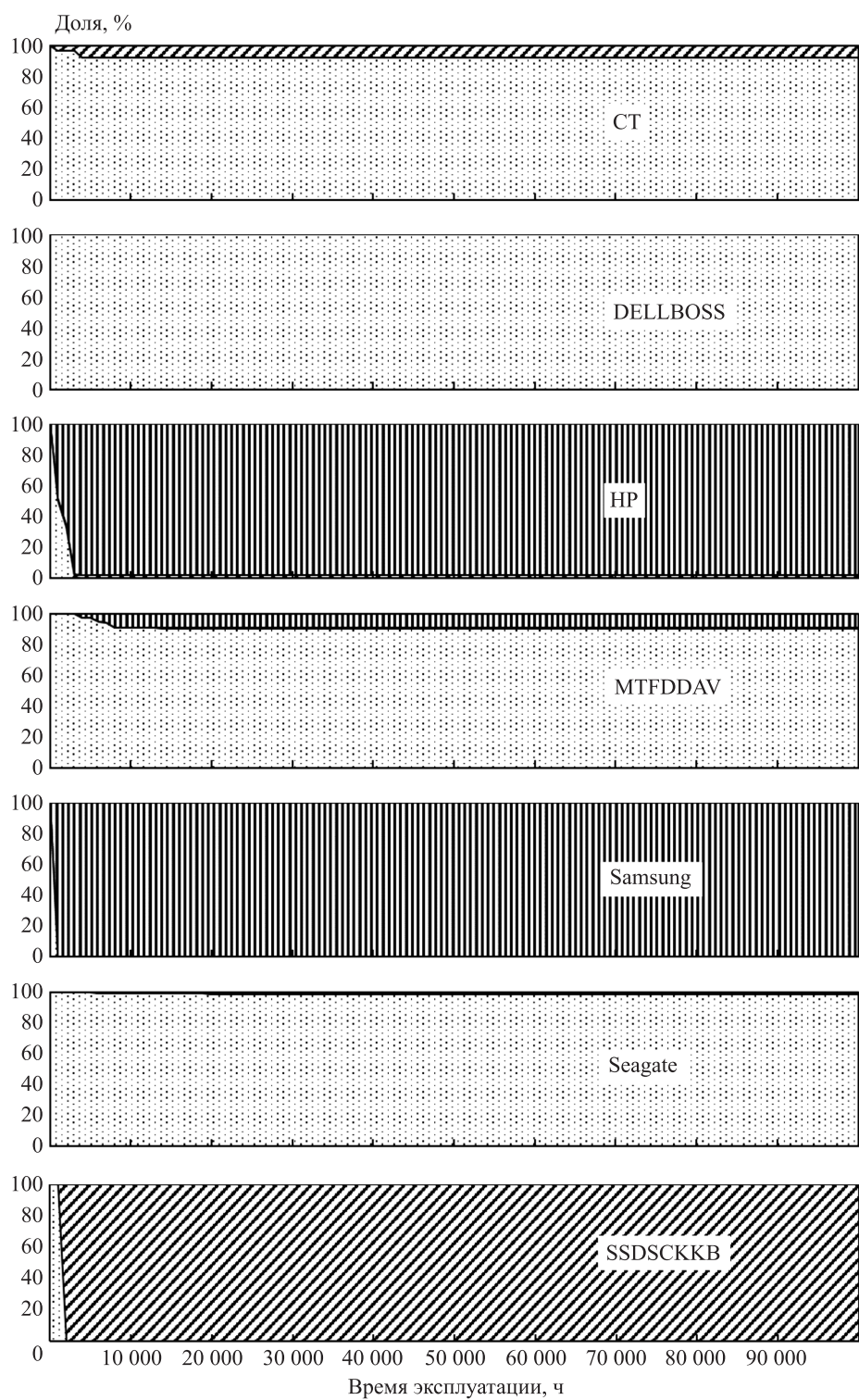


Рис. 2. Изменение доли работающих (точки), снятых досрочно (наклонные линии) и отказавших (вертикальные линии) SSD накопителей от времени эксплуатации, %
Change in the proportion of working (dots), prematurely removed (slanted lines) and failed (vertical lines) SSD drives from operation time, %

Затем рассчитывалась суммарная доля снятых досрочно и отказавших накопителей по возрастающей от нулевого значения времени эксплуатации. Потом эти накапливаемые доли вычитались от исходных 100 % и таким образом определялась доля продолжающих работать накопителей. Как видно из рис. 1, после окончания гарантийного срока в 5 лет (43 800 ч) происходит интенсивное снятие HDD накопителей с эксплуатации. Для торговых марок 00MD00, Hitachi, Samsung их доля уменьшается вплоть до нуля. Если число накопителей 00MD00 и Samsung было просто мало и их применяли в основном для тестирования, то накопители Hitachi скорее всего морально устарели и были заменены на более современные и емкие торговой марки HGST (Hitachi Global Storage Technologies). Остальные накопители торговых марок ST (Seagate Technology), Toshiba, WDC (Western Digital Corporation) продолжают присутствовать на рынке и коммерчески доступны в требуемых для обеспечения функционирования data-центров количествах.

Исходя из рис. 2, для SSD накопителей подобного интенсивного послегарантийного прекращения эксплуатации не наблюдается. Вместо этого имеются заметные доли досрочного снятия у торговых марок CT (Crucial) и SSDSCKKB (Intel), а также отказов у HP (Hewlett Packard), MTFDDAV (Micron), Samsung в начальный период времени. Незначительные доли досрочного снятия и отказов у Seagate раскиданы по времени. У DELLBOSS (boot optimized storage solution – оптимизированное для загрузки решение для хранения данных) все накопители функционировали нормально.

Так как доля отказов мала [9], то представляет интерес детализированное изучение вышедших из строя накопителей. На рис. 3 и 4 представлено пошаговое распределение отказавших HDD и SSD накопителей от времени эксплуатации. Для наглядности оси ординат были растянуты для каждой торговой марки индивидуально. Пустые диаграммы для 00MD00 на рис. 3 и для DELLBOSS и SSDSCKKB на рис. 4 означают отсутствие отказавших накопителей (см. табл. 1). Однопиковые диаграммы для Samsung на рис. 3 и для CT на рис. 4 означают один единственный отказ, а для Samsung на рис. 4 – что все десять SSD накопителей перестали работать на первом же шаге начального периода.

Некоторые отказавшие HDD накопители торговой марки WDC имеют большие значения по времени эксплуатации и поэтому вышли за пределы отображаемого диапазона оси абсцисс. Для SSD накопителей только у торговой марки Seagate есть что-то наподобие времени наработки на отказ в районе 20 000 ч согласно определению надежности. У остальных отказ происходит в начальный период эксплуатации. У значительной доли HDD накопителей также наблюдается аналогичное явление [12].

Обсуждение и выводы

Так называемая «младенческая смертность» накопителей информации, заключающаяся в их выходе из строя в начальный период эксплуатации, с точки зрения материальных затрат ущерба не приносит, так как покрывается гарантией. По статье «оплата труда» затраты уже имеют место быть,

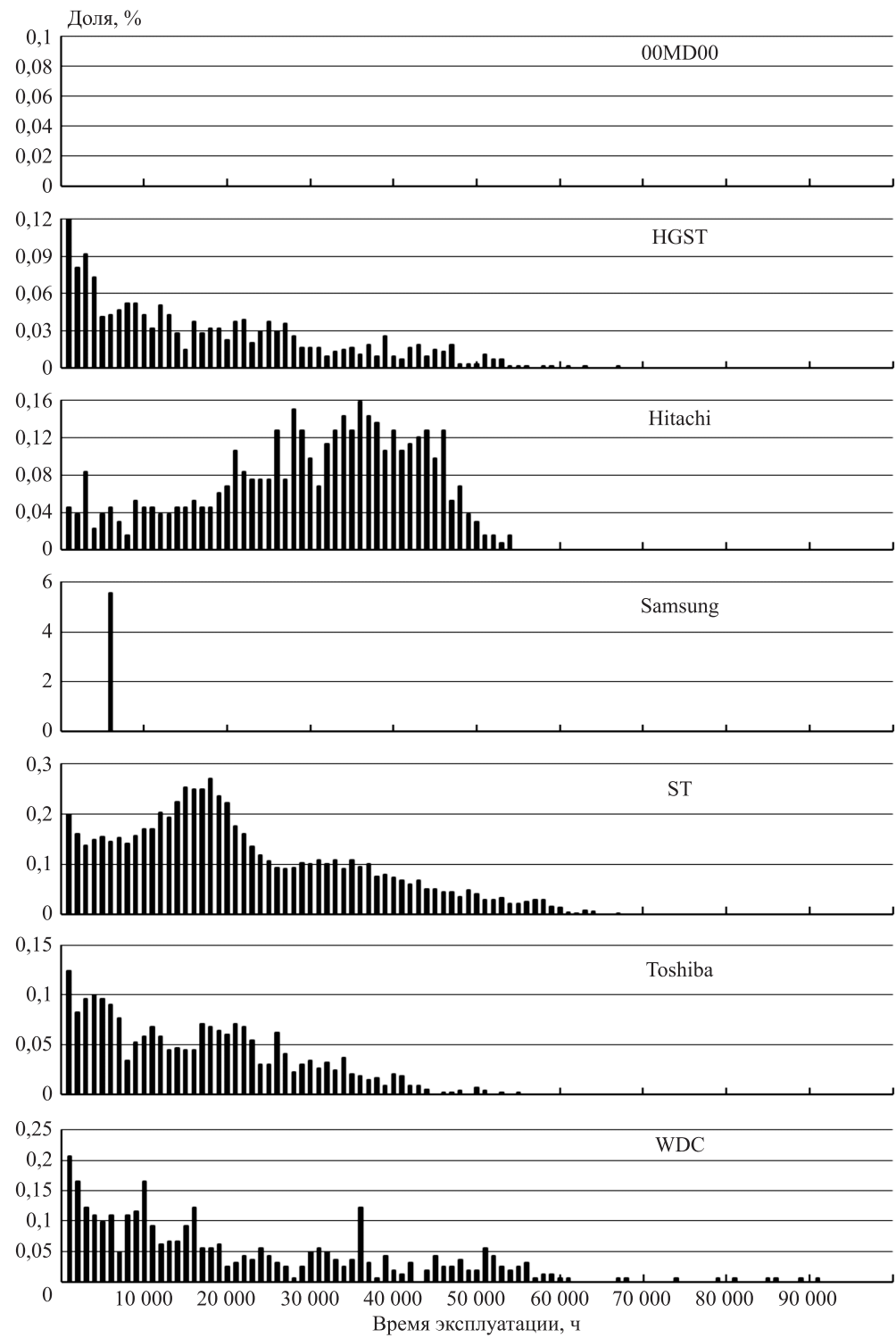


Рис. 3. Распределение отказавших HDD накопителей различных торговых марок от времени эксплуатации, %
Distribution of failed HDD of various brands from operation time, %

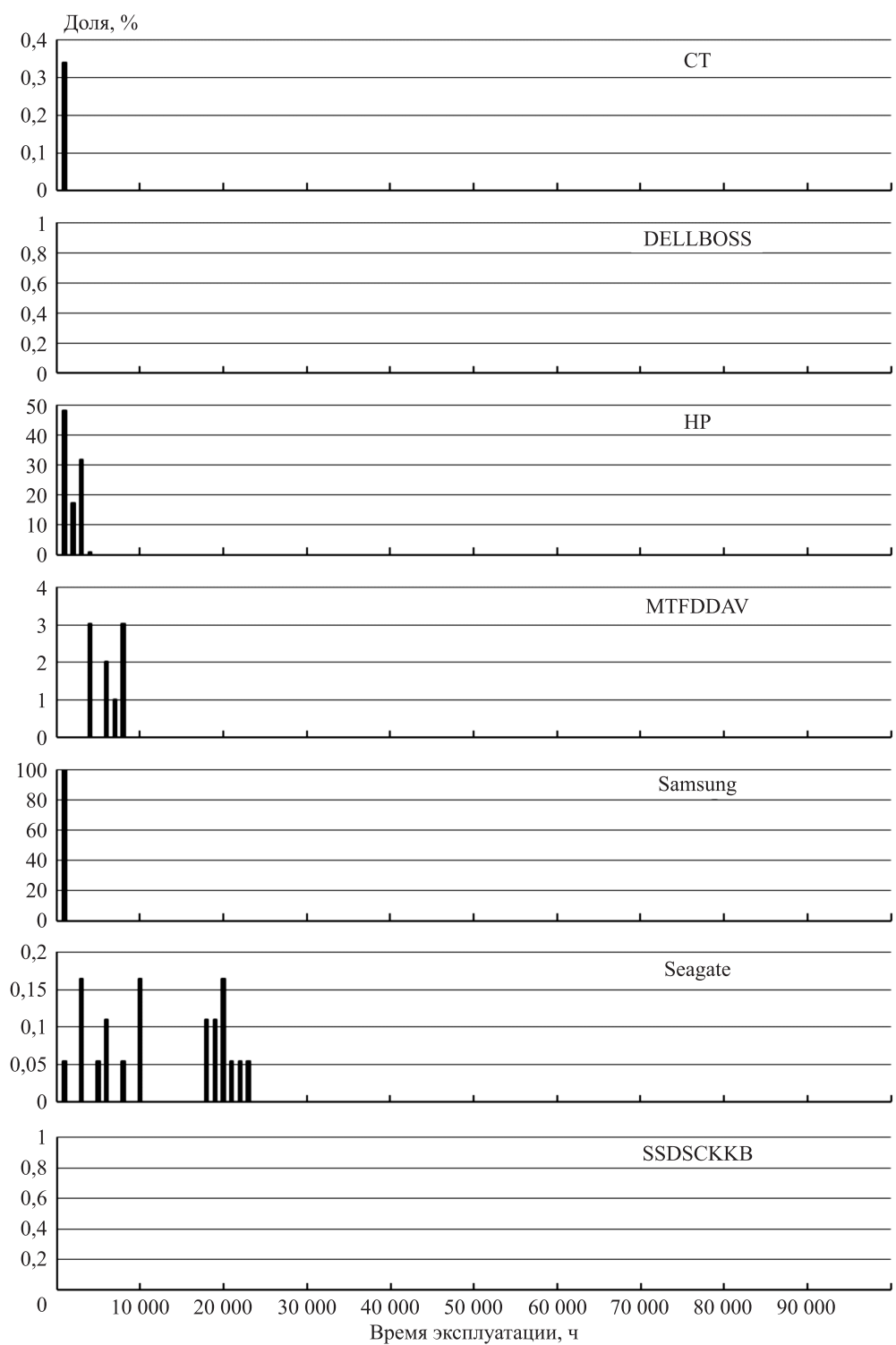


Рис. 4. Распределение отказавших SSD накопителей различных торговых марок от времени эксплуатации, %
Distribution of failed SSD of various brands from operation time, %

в связи с появлением необходимости проведения ручных операций по замене оборудования. Предположим, что на каждый накопитель тратится полчаса. Амортизации, т.е. списания по частям затрат на приобретение оборудования, а не сразу целиком в момент их совершения, здесь нет, потому что накопители проходят по статье «материалы» ввиду их относительно низкой стоимости. Величина обесценивания за время обмена по гарантии (пусть это будет 1 месяц) по сравнению с суммарным временем эксплуатации (например, 5 лет) составляет ориентировочно 1,7 % от стоимости накопителя.

Тогда формула для оценки дополнительных затрат на замену одного накопителя вследствие отказа в начальный период эксплуатации будет иметь следующий вид:

$$\text{Доп. затраты} = k \times (0,5 \times \text{почасовая оплата труда} + 0,017 \times \text{стоимость накопителя}),$$

где k – доля подобных отказов. Ее можно вычислить, например, как сумму долей за первую половину гарантийного срока эксплуатации, т.е. за 2,5 года. Результаты такого расчета приведены в табл. 2.

Таблица 2

Доля отказов накопителей разных торговых марок в начальный, конечный и послегарантийный период эксплуатации, отн. ед.

The share of drives failures of different brands in the initial, final and post-warranty operation period, rel. units

Торговая марка	Число, шт.	Всего	В начале	В конце	После
00MD00	2	0,0000	0,0000	0,0000	0,0000
HGST	53 405	0,0155	0,0103	0,0042	0,0010
Hitachi	13 246	0,0408	0,0108	0,0251	0,0050
Samsung	18	0,0556	0,0556	0,0000	0,0000
ST	179 810	0,0673	0,0416	0,0202	0,0054
Toshiba	53 230	0,0211	0,0151	0,0057	0,0002
WDC	16 419	0,0324	0,0201	0,0076	0,0047
Всего HDD	316 130	×	×	×	×
CT	294	0,0034	0,0034	0,0000	0,0000
DELLBOSS	351	0,0000	0,0000	0,0000	0,0000
HP	110	0,9818	0,9818	0,0000	0,0000
MTFDDAV	99	0,0909	0,0909	0,0000	0,0000
Samsung	10	1,0000	1,0000	0,0000	0,0000
Seagate	1828	0,0115	0,0109	0,0005	0,0000
SSDSCKKB	4	0,0000	0,0000	0,0000	0,0000
Всего SSD	2696	×	×	×	×

Как видно из табл. 2 у длительно работающих HDD накопителей торговых марок HGST, ST, Toshiba, WDC, а также SSD накопителей торговых марок CT, DELLBOSS, MTFDDAV, Seagate доли отказов в первую половину гарантийного срока эксплуатации сильно различаются между собой. Но оптимизировать закупки оборудования по данному параметру, на наш взгляд, не следует, все-таки сама общая доля отказов мала и выбор будет

осуществляться в первую очередь на основе стоимости накопителей, и только во вторую – надежности [7]. Тем не менее учет дополнительных затрат на замену вышедших из строя в начальный период единиц желательно выполнить хотя бы для оценки необходимых для этого резервов.

При стремительном росте информационной индустрии, обусловленном снижением стоимости хранения и передачи информации [2], неопределенность в результате от цифровизации бизнеса является главным препятствием на пути к переходу к новому экономическому укладу [1].

Отсюда вытекает следующий вывод: на первом этапе резерв для замены подобных накопителей можно создать универсальный, а в дальнейшем уточнить его размер по фактическим затратам согласно номенклатуре устанавливаемого в конкретном data-центре оборудования.

Заключение

Таким образом, в результате исследования выявлено, что в распределении выходов из строя накопителей значительную часть составляют отказы в начальный период времени эксплуатации. Особенно заметен этот эффект для SSD накопителей. Предложен способ учета дополнительных затрат на замену оборудования вследствие указанного обстоятельства. В результате учета и оценки этих затрат будет получена предварительная величина создаваемых резервов для замен.

Также ввиду того, что статистика для SSD накопителей в Backblaze велась только начиная с 2018 г., то для продолжения исследований требуется пополнение информационной базы свежими данными. Инструменты для этого имеются [4].

Заметная доля отказов в начальный период создает препятствия для применения хорошо зарекомендовавших себя в других случаях методов машинного обучения при прогнозировании выхода из строя накопителей информации, включая нейронные сети [3, 11].

Список источников

1. *Алешин А.В., Алешин В.А., Никитаева А.Ю.* Финансовые аспекты цифровой трансформации промышленности // Вестник Академии знаний. 2019. № 35 (6). С. 37–43. URL: <https://doi.org/10.24411/2304-6139-2019-00007>.
2. *Асташова Н.Д., Бобкова Т.В.* «Неэкономические» модели цифровой экономики // Вестник Томского гос. ун-та. 2021. № 470. С. 72–79. URL: <https://doi.org/10.17223/15617793/470/8>.
3. *Демидова Л.А., Филатов А.В.* Разработка модели классификации состояния жестких дисков на основе LSTM-нейронных сетей // Высокопроизводительные вычислительные системы и технологии. 2021. Т. 5, № 1. С. 37–42.
4. *Елизова Н.А., Ефимов А.В.* Анализ эффективности и функциональных возможностей программ по выявлению и обработке отказов SSD дисков // Инновации и научно-техническое творчество молодежи: сб. мат. Российской науч.-техн. конф. Новосибирск, 2022. С. 127–131.
5. *Лисичкин И.А.* Анализ сервисов резервного копирования и перспективы их использования // Инновационное развитие: потенциал науки и современного образования: сб. ст. V междунар. науч.-практ. конф. В 2 ч. Ч. 1. Пенза: Наука и просвещение, 2019. С. 73–75.

6. Насыров И.Н., Насыров И.И., Насыров Р.И. Большие данные по надежности накопителей информации в data-центрах // Цифровая экономика. 2022. № 2 (18). С. 33–37. URL: <https://doi.org/10.34706/DE-2022-02-04>
7. Савин И.В. Пять основных составляющих надежности файловых хранилищ данных // Наука, техника и образование. 2018. № 11 (52). С. 64–67.
8. Ходжаева Д.Ф. Лучшие облачные сервисы, используемые в сфере образования // Проблемы современной науки и образования. 2022. № 1 (170). С. 24–27.
9. Юрков А.А., Петров А.В. Мониторинг здоровья жестких дисков // Современные технологии в теории и практике программирования: сб. мат. науч.-практ. конф. Санкт-Петербург, 22 апреля 2021. СПб.: Санкт-Петербургский политехнический университет Петра Великого, 2021. С. 228–229.
10. Diallo M.S., Mokeddem S.A., Braud A., Frey G., Lachiche N. Identifying benchmarks for failure prediction in industry 4.0 // Informatics. 2021. Vol. 8, no. 4. P. 68. URL: <https://doi.org/10.3390/informatics8040068>
11. Filatov A., Demidova L. Application of Recurrent Networks to Develop Models for Hard Disk State Classification // Communications in Computer and Information Science. 2022. Vol. 1526. P. 380–390. URL: https://doi.org/10.1007/978-3-030-94141-3_30.
12. Nasyrov I.N., Nasyrov I.I., Nasyrov R.I., Khairullin B.A. Dependence of reallocated sectors count on HDD power-on time // International Journal of Engineering and Technology (UAE). 2018. Vol. 7, Is. 4.7, Special Issue 7. P. 200–203. URL: <https://doi.org/10.14419/ijet.v7i4.7.20544>

References

1. Aleshin A.V., Aleshin V.A., Nikitaeva A.Ju. Finansovye aspekty cifrovoj transformacii promyshlennosti [Financial aspects of digital transformation of industry], *Vestnik Akademii znaniy* [Vestnik of the Academy of Knowledge], 2019, no. 35 (6), pp. 37–43. URL: <https://doi.org/10.24411/2304-6139-2019-00007>
2. Astashova N.D., Bobkova T.V. «Nejekonomicheskie» modeli cifrovoj jekonomiki [“Non-economic” models of digital economy], *Vestnik Tomskogo gos. un-ta* [Vestnik of Tomsk State University], 2021, no. 470, pp. 72–79. URL: <https://doi.org/10.17223/15617793/470/8>
3. Demidova L.A., Filatov A.V. Razrabotka modeli klassifikacii sostojaniya zhestkih diskov na osnove LSTM-nejronnyh setej [Development of the hard discs state classification model based on LSTM-neural networks], *Vysokoproizvoditel'nye vychislitel'nye sistemy i tehnologii* [High-performance computing systems and technologies], 2021, vol. 5, no. 1, pp. 37–42.
4. Elizova N.A., Efimov A.V. Analiz jeffektivnosti i funkcional'nyh vozmozhnostej programm po vyjavleniju i obrabotke otkazov SSD diskov [Analysis of efficiency and functional capabilities of the programs for detection and processing of SSD discs failures]. Innovacii i nauchno-tehnicheskoe tvorchestvo molodezhi: sb. mat. Rossijskoj nauch.-tehn. konf. Novosibirsk, 2022. Pp. 127–131.
5. Lisichkin I.A. Analiz servisov rezervnogo kopirovaniya i perspektivy ih ispol'zovaniya [Analysis of backup services and prospects of their use], *Innovacionnoe razvitiye: potencial nauki i sovremennogo obrazovaniya*: sb. st. V mezhdunar. nauch.-prakt. konf. V 2 ch. Ch. 1. Penza: Nauka i prosveshhenie, 2019. Pp. 73–75.
6. Nasyrov I.N., Nasyrov I.I., Nasyrov R.I. Bol'shie dannye po nadezhnosti nakopitelej informacii v data-centrah [Big data on the reliability of information storage devices in data centres], *Cifrovaja jekonomika* [Digital Economy], 2022, no. 2 (18), pp. 33–37. URL: <https://doi.org/10.34706/DE-2022-02-04>
7. Savin I.V. Pjat' osnovnyh sostavljajushhih nadjozhnosti fajlovyyh hranilishh dannyh [Five main components of reliability of file data storages], *Nauka, tehnika i obrazovanie* [Science, Technology and Education], 2018, no. 11 (52), pp. 64–67.

8. Hodzhaeva D.F. Luchshie oblačnyje servisy, ispol'zuemye v sfere obrazovanija [Best cloud services used in education], *Problemy sovremennoj nauki i obrazovanija* [Problems of modern science and education], 2022, no. 1 (170), pp. 24–27.
9. Jurkov A.A., Petrov A.V. Monitoring zdorov'ja zhestkih diskov [Monitoring of hard discs health]. *Sovremennye tehnologii v teorii i praktike programirovanija: sb. mat. nauch.-prakt. konf. Sankt-Peterburg, 22 aprelja 2021. Saint Petersburg: Sankt-Peterburgskij politehnicheskij universitet Petra Velikogo, 2021. Pp. 228–229.*
10. Diallo M.S., Mokeddem S.A., Braud A., Frey G., Lachiche N. Identifying benchmarks for failure prediction in industry 4.0, *Informatics*, 2021, 8, no. 4, p. 68. URL: <https://doi.org/10.3390/informatics8040068>
11. Filatov A., Demidova L. Application of Recurrent Networks to Develop Models for Hard Disk State Classification. *Communications in Computer and Information Science*, 2022, vol. 1526, pp. 380–390. URL: https://doi.org/10.1007/978-3-030-94141-3_30
12. Nasyrov I.N., Nasyrov I.I., Nasyrov R.I., Khairullin B.A. Dependence of reallocated sectors count on HDD power-on time. *International Journal of Engineering and Technology (UAE)*, 2018, vol. 7, is. 4.7, special issue 7, pp. 200–203. URL: <https://doi.org/10.14419/ijet.v7i4.7.20544>

Сведения об авторах:

И.Н. Насыров – доктор экономических наук, профессор, Казанский (Приволжский) федеральный университет, Казань, Российская Федерация.

И.И. Насыров – кандидат технических наук, ведущий сервис-менеджер, ООО «Телеком Интеграция», Казань, Российская Федерация.

Р.И. Насыров – руководитель портфеля проектов, ООО «Газпромнефть – Цифровые решения», Санкт-Петербург, Российская Федерация.

Information about the authors:

I.N. Nasyrov – Doctor of Economics, Professor, Kazan (Volga Region) Federal University, Kazan, Russian Federation.

I.I. Nasyrov – Candidate of Technical Sciences, Service Manager, Telecom Integration LLC, Kazan, Russian Federation.

R.I. Nasyrov – Head of the Project Portfolio, Gazpromneft – Digital Solutions LLC, St. Petersburg, Russian Federation.

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

<i>Статья поступила в редакцию</i>	<i>10.04.2023</i>	<i>The article was submitted</i>	<i>10.04.2023</i>
<i>Одобрена после рецензирования</i>	<i>13.05.2023</i>	<i>Approved after reviewing</i>	<i>13.05.2023</i>
<i>Принята к публикации</i>	<i>18.05.2023</i>	<i>Accepted for publication</i>	<i>18.05.2023</i>