

НАУЧНЫЙ ЖУРНАЛ

ЭКОНОМИКА. ИНФОРМАТИКА

SCIENTIFIC JOURNAL

ECONOMICS. INFORMATION TECHNOLOGIES

2024. Том 51, № 2



ЭКОНОМИКА. ИНФОРМАТИКА

2024. Том 51, № 2

До 2020 г. журнал издавался под названием «Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика».

Основан в 1995 г.

Журнал включен в Перечень ВАК рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней кандидата и доктора наук (1.2.2. Математическое моделирование, численные методы и комплексы программ; 2.3.1. Системный анализ, управление и обработка информации; 2.3.3. Автоматизация и управление технологическими процессами и производствами; 2.3.4. Управление в организационных системах; 2.3.8. Информатика и информационные процессы; 5.2.3. Региональная и отраслевая экономика; 5.2.4. Финансы; 5.2.6. Менеджмент). Журнал зарегистрирован в Российском индексе научного цитирования (РИНЦ).

Учредитель: Федеральное государственное автономное образовательное учреждение высшего образования «Белгородский государственный национальный исследовательский университет».

Издатель: НИУ «БелГУ» Издательский дом «БелГУ».

Адрес редакции, издателя: 308015, г. Белгород, ул. Победы, 85.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ ЖУРНАЛА

Главный редактор

Е.Г. Жиликов, доктор технических наук, профессор, профессор кафедры информационно-телекоммуникационных систем и технологий института инженерных и цифровых технологий (НИУ «БелГУ», Белгород, Россия)

Заместитель главного редактора

Е.А. Стрябова, доктор экономических наук, доцент, заведующая кафедрой прикладной экономики и экономической безопасности института экономики и управления (НИУ «БелГУ», Белгород, Россия)

Ответственные секретари

Ю.В. Лыщикова, кандидат экономических наук, доцент, доцент кафедры прикладной экономики и экономической безопасности института экономики и управления (НИУ «БелГУ», Белгород, Россия)

Е.В. Болгова, кандидат технических наук, доцент кафедры прикладной информатики и информационных технологий института инженерных и цифровых технологий (НИУ «БелГУ», Белгород, Россия)

Члены редколлегии:

А.В. Богомолов, доктор технических наук, профессор (Центральный научно-исследовательский институт Военно-воздушных сил Минобороны России, Москва, Россия)

О.В. Ваганова, доктор экономических наук, профессор, заведующая кафедрой инновационной экономики и финансов института экономики и управления (НИУ «БелГУ», Белгород, Россия)

М.В. Владыка, доктор экономических наук, доцент, профессор кафедры прикладной экономики и экономической безопасности, заместитель директора по научной работе института экономики и управления (НИУ «БелГУ», Белгород, Россия)

В.П. Волчков, доктор технических наук, профессор (Московский технический университет связи и информатики, Москва, Россия)

В.П. Воронин, доктор экономических наук, профессор, профессор кафедры торгового дела и товароведения (Воронежский государственный университет инженерных технологий, Воронеж, Россия)

В.С. Голиков, доктор технических наук, профессор (Universidad Autónoma del Carmen (UNACAR), Мексика)

О.А. Ивацук, доктор технических наук, профессор, заведующая кафедрой информационных и робототехнических систем (НИУ «БелГУ», Белгород, Россия)

А.В. Коськин, доктор технических наук, профессор, профессор кафедры информационных систем и цифровых технологий (Орловский государственный университет им. И.С. Тургенева, Орел, Россия)

Н.А. Кулагина, доктор экономических наук, профессор, профессор кафедры государственного управления, экономической и информационной безопасности, директор инженерно-экономического института (Брянский государственный инженерно-технологический университет, Брянск, Россия)

А.С. Молчан, доктор экономических наук, профессор, профессор кафедры бизнес-аналитики (Кубанский государственный технологический университет, Краснодар, Россия)

Т.В. Никитина, доктор экономических наук, профессор, профессор кафедры банков, финансовых рынков и страхования, директор Международного Центра исследований финансовых рынков (Санкт-Петербургский государственный экономический университет, Санкт-Петербург, Россия)

А.А. Сирота, доктор технических наук, профессор, заведующий кафедрой технологий обработки и защиты информации (Воронежский государственный университет, Воронеж, Россия)

В.Б. Сулимов, доктор физико-математических наук, старший научный сотрудник (Московский государственный университет имени М.В. Ломоносова, Научно-исследовательский вычислительный центр, Москва, Россия)

В.М. Тумин, доктор экономических наук, профессор, профессор кафедры менеджмента (Московский политехнический университет, Москва, Россия)

Т.Л. Тен, доктор технических наук, профессор, проректор по цифровым технологиям и инновациям (Карагандинский экономический университет Казпотребсоюза, Караганда, Казахстан)

А.А. Черноморец, доктор технических наук, доцент, профессор кафедры прикладной информатики и информационных технологий института инженерных и цифровых технологий (НИУ «БелГУ», Белгород, Россия)

ISSN 2687-0932

Журнал зарегистрирован в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации средства массовой информации ЭЛ № ФС 77-77834 от 31.01.2020.

Выходит 4 раза в год.

Выпускающий редактор Ю.В. Мишенина. Корректур, компьютерная верстка и оригинал-макет Ю.В. Мишенина. Гарнитура Times New Roman, Arial Narrow, Arial. Уч.-изд. л. 25,0. Дата выхода 30.06.2024. Оригинал-макет подготовлен отделом объединенной редакции научных журналов НИУ «БелГУ». Адрес: 308015, г. Белгород, ул. Победы, 85.

СОДЕРЖАНИЕ

РЕГИОНАЛЬНАЯ И МУНИЦИПАЛЬНАЯ ЭКОНОМИКА

- 265 **Горочная В.В., Михайлова А.А.**
Континентально-океаническая дихотомия в развитии туристического бренда западных регионов России в период геоэкономической турбулентности
- 283 **Куприянов С.В., Климов А.В., Герасимова Н.А.**
Отдельные аспекты реализации поведенческого механизма управления развитием муниципальных образований
- 296 **Михайличенко М.Ю.**
Методика оценки сбалансированного пространственного развития региона
- 309 **Мухаметова А.Д.**
Сценарии регионального экономического развития на основе использования человеческого капитала региона
- 320 **Стрябкова Е.А., Скребова А.В.**
Оценка пространственной интеграции макрорегионов Российской Федерации

ИНВЕСТИЦИИ И ИННОВАЦИИ

- 330 **Митенков А.В.**
Концепция и инструменты фасилитации для управления конфликтными противоречиями заинтересованных сторон при трансформации систем управления
- 344 **Петрова Е.А., Калинина В.В.**
Информационно-аналитическая система мониторинга развития цифровых технологий и оценки их влияния на динамику структурных сдвигов экономики России
- 355 **Малыхина И.О., Широкий И.В.**
Современные тенденции и перспективные направления осуществления инновационной деятельности в Белгородской области

ОТРАСЛЕВЫЕ РЫНКИ И РЫНОЧНАЯ ИНФРАСТРУКТУРА

- 368 **Волоцков А.А.**
Национальный проект «Туризм и индустрия гостеприимства» как инструмент развития территорий: первые итоги

ФИНАНСЫ ГОСУДАРСТВА И ПРЕДПРИЯТИЙ

- 379 **Карловская Е.А., Ваганова О.В., Мельникова Н.С., Быканова Н.И.**
Прогнозное моделирование развития рынка кредитования и краудфандинга в условиях увеличения ключевой ставки Банка России
- 393 **Мусацкая Я.С.**
Внутренний аудит в системе обеспечения кибербезопасности коммерческого банка

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

- 402 **Жиляков Е.Г., Гайворонская Д.И., Прохоренко Е.И., Балабанова Т.Н.**
О краткосрочном прогнозировании временных рядов на основе субполосных представлений

СИСТЕМНЫЙ АНАЛИЗ И УПРАВЛЕНИЕ

- 413 **Салтанаева Е.А., Васильева С.В., Эшлиоглу Р.И.**
Оптимизация информационных процессов в области маркетинговых исследований с применением технологий искусственного интеллекта
- 425 **Гуляева К.А., Артемьева И.Л.**
Инструментальные средства поддержки процессов создания систем понятий и баз знаний для предметных областей со сложно-структурированными объектами (на примере задачи определения реакционных способностей химических соединений)
- 444 **Гуль С.В.**
Алгоритм построения системно-объектной трёхмерной классификации
- 453 **Полещенко Д.А., Коврижных О.А.**
Разработка интеллектуальной системы поддержки принятия решений на основе прецедентов для оперативного планирования сталеплавильного цеха

ИНФОКОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

- 466 **Кудинов В.А., Бородин М.Е.**
Об обнаружении ошибочного употребления термина в русскоязычном тексте на основе онтологии предметной области
- 479 **Насыров И.Н., Насыров И.И., Насыров Р.И.**
Типичные ошибки в больших данных по надежности накопителей информации в data-центрах
- 489 **Сумин В.И., Лукин М.А.**
Оптимизация передачи сообщений в сетевых информационных системах

УДК 004.052:004.33:658.155
DOI 10.52575/2687-0932-2024-51-2-479-488

Типичные ошибки в больших данных по надежности накопителей информации в data-центрах

¹ Насыров И.Н., ² Насыров И.И., ³ Насыров Р.И.

¹ Казанский (Приволжский) федеральный университет
Россия, 420008, Республика Татарстан, г. Казань, ул. Кремлевская, д. 18

² ООО «Телеком Интеграция»
Россия, 420015, Республика Татарстан, г. Казань, ул. Подлужная, д. 60

³ ООО «Газпромнефть – Цифровые решения»
Россия, 190013, г. Санкт-Петербург, ул. Киевская, д. 5
E-mail: ecoseti@yandex.ru, ildarec@mail.ru, rinasyrov@gmail.com

Аннотация. Цифровая экономика основана на сборе и использовании больших данных. Однако с течением времени их объем в централизованных хранилищах (data-центрах) становится настолько велик, что даже только сведения о состоянии накопителей информации сами начинают из себя представлять большие данные. Содержащиеся в них ошибки создают определенные проблемы при оценке надежности оборудования. Целью данного исследования является выявление и анализ указанных ошибок. Рассматривались находящиеся в открытом доступе значения параметров накопителей информации data-центров компании Backblaze за длительный период. В результате был выявлен ряд ошибок, типизация которых укладывается в пять крупных групп: по оформлению, по содержанию, по времени, корректируемые и некорректируемые. Показано, что число ошибок и скорость их прироста можно значительно снизить путем корректировки.

Ключевые слова: data-центр, накопитель информации, показатель надежности, большие данные, ошибка, скорость накопления

Для цитирования: Насыров И.Н., Насыров И.И., Насыров Р.И. 2024. Типичные ошибки в больших данных по надежности накопителей информации в data-центрах. Экономика. Информатика. 51(2): 479–488. DOI 10.52575/2687-0932-2024-51-2-479-488

Typical Errors in Big Data on Storage Devices Reliability in Data Centers

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

¹ Kazan (Volga region) Federal University
18 Kremlyovskaya St, Kazan, Tatarstan, 420008, Russia

² Telecom Integration LLC
60 Podluzhnaya St, Kazan, Tatarstan, 420015, Russia

³ Gazpromneft – Digital Solutions LLC
5 Kievskaya St, St. Petersburg, 190013, Russia
E-mail: ecoseti@yandex.ru, ildarec@mail.ru, rinasyrov@gmail.com

Abstract. The development of the digital economy is based on the collection and use of big data. However, over time, their volume in centralized storages (data centers) becomes so large that even only information about the state of storage devices themselves begin to represent big data. The errors contained in them create certain problems when evaluating the reliability of equipment. The purpose of this study is to identify and analyze such errors. The publicly available parameter values of the data storage devices states in Backblaze data centers for a long period were considered. As a result, a number of errors were identified, the typing of which fits into five large groups: by design, by content, by time, correctable and uncorrectable. It is shown that the number of errors and the rate of their increase can be significantly reduced by adjusting. Also due to the increase over time in the data centers of both the number of drives themselves and the number of recorded



parameters of their state, collectively leading to an accelerated nonlinear increase in the total number of analyzed data, investigation must be continued.

Keywords: data center, storage device, reliability parameter, big data, error, increase rate

For citation: Nasyrov I.N., Nasyrov I.I., Nasyrov R.I. 2024. Typical Errors in Big Data on Storage Devices Reliability in Data Centers. Economics. Information technologies. 51(2): 479–488 (in Russian). DOI 10.52575/2687-0932-2024-51-2-479-488

Введение

Актуальность исследования обусловлена необходимостью своевременной оценки надежности накопителей информации в системах централизованного хранения генерируемых в цифровой экономике больших данных (data-центрах). Для этого обычно регулярно снимаются и записываются сведения об их состоянии [Юрков, 2021]. На их основе разрабатываются различные методы прогнозирования выхода из строя накопителей: на основе нейронных сетей [Demidova, 2020; Демидова, Филатов, 2021а, б; Демидова, Фурсов, 2021; Shi et al., 2021; Demidova, Fursov, 2022b], моделей экстремального обучения [Demidova, Fursov, 2022a], бинарной классификации [Демидова, Филатов, 2022; Филатов, 2022; Filatov, Demidova, 2022], стохастической [Peng et al., 2021], адаптационной [Lan et al., 2021], других моделей машинного обучения [Kuznietsova, Kuznietsova, 2020; Demidova, Fursov, 2021].

Научной проблемой является наличие ошибок в исходных сведениях, затрудняющих проведение их анализа. Цель исследования состоит в выявлении подобных ошибок и их типизации для последующего гарантированного устранения. При этом вследствие увеличения с течением времени как числа самих накопителей, так и числа записываемых параметров их состояния, в совокупности приводящих к нелинейному росту общего числа анализируемых данных, исследования необходимо продолжать.

Объекты и методы исследования

Информационной базой исследования послужили ежедневно записываемые SMART-данные (self-monitoring, analysis and reporting technology – технология самоконтроля, анализа и отчетности) накопителей, находящиеся в свободном доступе на сайте одной из крупнейших в мире групп коммерческих data-центров компании Backblaze (<https://www.backblaze.com/b2/hard-drive-test-data.html>). Они удовлетворяют всем требованиям для прогнозирования сбоя, в связи с чем исследователи всего мира активно используют их в своей работе [Diallo et al., 2021]. В качестве методов исследования выбраны сортировка, подсчет и анализ значений параметров, продолжающих функционировать, снятых досрочно и отказавших накопителей информации за период с 10.04.2013 по 30.06.2022. Накопители на жестких дисках (HDD – hard disk drive) для «холодного» хранения и твердотельные (SSD – solid state drive) для оперативного использования информации рассматривались отдельно.

Было проведено сжатие 325415056 строк с данными за указанный период в 318826 строку путем отбора накопителей, имеющих отличающиеся друг от друга серийные номера, на последнюю дату их упоминания. Из них у 219444 индикатор состояния показывал, что они продолжали функционировать нормально, 84108 были сняты досрочно, 15274 вышли из строя.

Результаты

В результате проведенных исследований надежности накопителей информации были выявлены следующие ошибки в данных об их состоянии:

Во-первых, получено, что один из накопителей модели 00MD00 не имеет серийного номера. Всего накопителей такой модели 2 штуки. Сравнение значений параметров показало, что это все-таки различающиеся друг от друга накопители. Оба были сняты с эксплуатации досрочно.

Во-вторых, у накопителей моделей ST4000DM000 (2 шт.) и HGST HMS5C4040ALE640 (1 шт.) не были указаны емкости на момент отказа. В последнем случае и значения параметров были сбойными. Для корректировки все эти величины были взяты из значений параметров за предыдущие до отказа дни.

В-третьих, у накопителей моделей HGST HMS5C4040ALE640 (1 шт.), HGST HUH721212ALN604 (1 шт.), ST12000NM0007 (59 шт.), ST16000NM001G (6 шт.), ST4000DM000 (28 шт.), ST8000NM0055 (1 шт.), TOSHIBA MD04ABA400V (1 шт.), TOSHIBA MG07ACA14TA (1 шт.), WDC WUH721414ALE6L4 (1 шт.) емкости были отрицательными, равными –1. Значения параметров отсутствовали. Из этих моделей только у HGST HUH721212ALN604 (1 шт.), ST12000NM0007 (57 шт.), ST8000NM0055 (1 шт.), TOSHIBA MD04ABA400V (1 шт.) для корректировки нужные величины можно было взять из значений параметров за предыдущие дни. У остальных моделей HGST HMS5C4040ALE640 (1 шт.), ST12000NM0007 (2 шт.), ST16000NM001G (6 шт.), ST4000DM000 (28 шт.), TOSHIBA MG07ACA14TA (1 шт.), WDC WUH721414ALE6L4 (1 шт.) значения параметров за предыдущие дни отсутствовали. Поэтому корректировка была проведена лишь величины емкости, да и то по аналогичным накопителям из той же серии.

Из этих моделей ST12000NM0007 (1 шт.), ST16000NM001G (6 шт.), ST8000NM0055 (1 шт.), TOSHIBA MD04ABA400V (1 шт.), TOSHIBA MG07ACA14TA (1 шт.), WDC WUH721414ALE6L4 (1 шт.) были помечены как вышедшие из строя, остальные HGST HMS5C4040ALE640 (1 шт.), HGST HUH721212ALN604 (1 шт.), ST12000NM0007 (58 шт.), ST4000DM000 (28 шт.) – как снятые с эксплуатации досрочно.

В-четвертых, в дополнение к предыдущим у накопителей моделей DELLBOSS VD (всего 351 шт.), HGST HMS5C4040ALE640 (3 шт.), HGST HMS5C4040BLE640 (1 шт.), ST12000NM0007 (9 шт.), ST12000NM0008 (8 шт.), ST12000NM001G (8 шт.), ST14000NM001G (3 шт.), ST14000NM0138 (1 шт.), ST16000NM001G (3 шт.), ST4000DM000 (11 шт.), ST8000DM002 (7 шт.), ST8000NM0055 (8 шт.), TOSHIBA MG07ACA14TEY (2 шт.), TOSHIBA MG08ACA16TE (1 шт.) значения всех SMART-параметров отсутствовали. Из них у HGST HMS5C4040ALE640 (1 шт.), ST4000DM000 (8 шт.) еще и величины емкостей были неверными. У модели DELLBOSS (boot optimized storage solution – оптимизированное для загрузки решение для хранения данных) всегда все значения отсутствуют, поэтому в дальнейшем ее исключили из списка рассматриваемых.

Из оставшихся после исключения DELLBOSS только у моделей HGST HMS5C4040ALE640 (3 шт.), HGST HMS5C4040BLE640 (1 шт.), ST12000NM0007 (6 шт.), ST12000NM0008 (8 шт.), ST12000NM001G (4 шт.), ST14000NM001G (1 шт.), ST14000NM0138 (1 шт.), ST16000NM001G (2 шт.), ST4000DM000 (9 шт.), ST8000DM002 (3 шт.), ST8000NM0055 (7 шт.), TOSHIBA MG07ACA14TEY (2 шт.), TOSHIBA MG08ACA16TE (1 шт.) для корректировки нужные величины можно было взять из значений параметров за предыдущие дни. Остальные модели ST12000NM0007 (3 шт.), ST12000NM001G (4 шт.), ST14000NM001G (2 шт.), ST16000NM001G (1 шт.), ST4000DM000 (2 шт.), ST8000DM002 (4 шт.), ST8000NM0055 (1 шт.) даже если в предыдущие дни и упоминались, однако значения параметров у них отсутствовали, т. е. они сразу после установки вышли из строя или вовсе не заработали. Поэтому корректировка была проведена в сторону даты первого упоминания таких накопителей.

Также из этих моделей HGST HMS5C4040ALE640 (3 шт.), HGST HMS5C4040BLE640 (1 шт.), ST12000NM0007 (7 шт.), ST12000NM0008 (6 шт.), ST12000NM001G (8 шт.), ST14000NM001G (2 шт.), ST14000NM0138 (1 шт.), ST16000NM001G (3 шт.), ST4000DM000 (9 шт.), ST8000DM002 (7 шт.), ST8000NM0055 (8 шт.), TOSHIBA MG07ACA14TEY (2 шт.), TOSHIBA MG08ACA16TE (1 шт.) были помечены как вышедшие из строя, остальные ST12000NM0007 (2 шт.), ST12000NM0008 (2 шт.), ST14000NM001G (1 шт.), ST4000DM000 (2 шт.) – как снятые с эксплуатации досрочно. Причем один накопитель модели ST12000NM0008 вначале был помечен как отказавший, а перед заменой – как нормально работающий.



Естественно, что в итоге он был учтен нами как отказавший. Этот особенный вид ошибок наблюдался впервые.

В-пятых, во всех случаях корректировки даты смещались на день раньше для моделей HGST HMS5C4040ALE640 (3 шт.), HGST HMS5C4040BLE640 (1 шт.), HGST HUH721212ALN604 (1 шт.), ST12000NM0007 (62 шт.), ST12000NM0008 (3 шт.), ST12000NM001G (4 шт.), ST14000NM001G (1 шт.), ST16000NM001G (1 шт.), ST4000DM000 (6 шт.), ST8000DM002 (2 шт.), ST8000NM0055 (4 шт.), TOSHIBA MD04ABA400V (1 шт.), на два дня раньше для моделей ST12000NM0007 (1 шт.), ST12000NM0008 (3 шт.), ST4000DM000 (1 шт.), ST8000DM002 (2 шт.), ST8000NM0055 (1 шт.), TOSHIBA MG07ACA14TEY (1 шт.), на три дня раньше для моделей ST12000NM0008 (1 шт.), ST12000NM001G (1 шт.), ST14000NM001G (2 шт.), ST4000DM000 (1 шт.), на четыре дня раньше для моделей ST12000NM001G (1 шт.), ST4000DM000 (2 шт.), ST8000NM0055 (2 шт.), на пять дней раньше для моделей ST12000NM0007 (1 шт.), ST14000NM0138 (1 шт.), ST8000NM0055 (1 шт.), на семь дней раньше для моделей HGST HMS5C4040ALE640 (1 шт.), ST16000NM001G (1 шт.), TOSHIBA MG08ACA16TE (1 шт.), на восемь дней раньше для модели ST12000NM0008 (1 шт.), на десять дней раньше для моделей ST8000DM002 (1 шт.), ST8000NM0055 (1 шт.), TOSHIBA MG07ACA14TEY (1 шт.). У остальных накопителей даты при корректировке не смещались – это у моделей 00MD00 (1 шт.), HGST HMS5C4040ALE640 (1 шт.), ST12000NM0007 (4 шт.), ST12000NM001G (2 шт.), ST16000NM001G (7 шт.), ST4000DM000 (31 шт.), ST8000DM002 (2 шт.), TOSHIBA MG07ACA14TA (1 шт.), WDC WUH721414ALE6L4 (1 шт.).

Распределение ошибок в данных о надежности накопителей информации по типам приведено в таблице 1. Выделены пять групп: по оформлению, по содержанию, по времени, корректируемые и некорректируемые. Приоритет отдан содержанию. Если содержание некорректируемо, то рассматривается оформление. Временной сдвиг при корректировке повторно касается всех тех же случаев, что и по оформлению и содержанию, поэтому вынесен отдельно.

Таблица 1
Table 1

Распределение ошибок в данных о надежности накопителей информации по типам, шт.
Distribution of errors in data on information storage devices reliability by type, pcs.

Ошибки	тип	корректируемые	некорректируемые	всего
Отсутствие серийного номера	оформление	0	1	1
Отсутствие значения емкости	содержание	1	0	1
	оформление	2	0	2
Значение емкости отрицательно	содержание	60	39	99
Отсутствие значения параметров	содержание	48	17	65
Смещение даты	время	118	50	168

Распределение ошибок в данных о надежности по торговым маркам накопителей информации приведено в таблице 2. Для HDD за исключением накопителей марок 00MD00 и SAMSUNG HD, доля которых в общем количестве относительно мала, наибольший процент ошибок имеется у накопителей марки ST (Seagate). Последние составляют и максимальную долю среди всех накопителей. Для SSD наибольший процент ошибок имеется у накопителей марки DELLBOSS, если ошибками считать то, что у них с самого начала никакие значения SMART-параметров не записывались. Для остальных накопителей процент ошибок нулевой.

Таблица 2
 Table 2

Распределение ошибок в данных о надежности по торговым маркам накопителей информации, шт.

Distribution of errors in reliability data by information storage brands, pcs.

Накопители	торговые марки	всего	ошибки	% ошибок
HDD	00MD00	2	1	50,00000
	HGST	53405	7	0,01311
	Hitachi	13246	0	0,00000
	SAMSUNG HD	18	0	0,00000
	ST	179810	154	0,08565
	TOSHIBA	53230	5	0,00939
	WDC	16419	1	0,00609
	Итого	316130	168	0,05314
SSD	CT	294	0	0,00000
	DELLBOSS	351	351	100,00000
	HP SSD	110	0	0,00000
	MTFDDAV	99	0	0,00000
	Samsung SSD	10	0	0,00000
	Seagate SSD	1828	0	0,00000
	SSDSCKKB	4	0	0,00000
	Итого	2696	351	13,01929

Распределение числа ошибок в записях значений параметров накопителей информации по годам приведено на рисунке 1. В последнем временном отрезке данные имеются только за половину 2022 года.

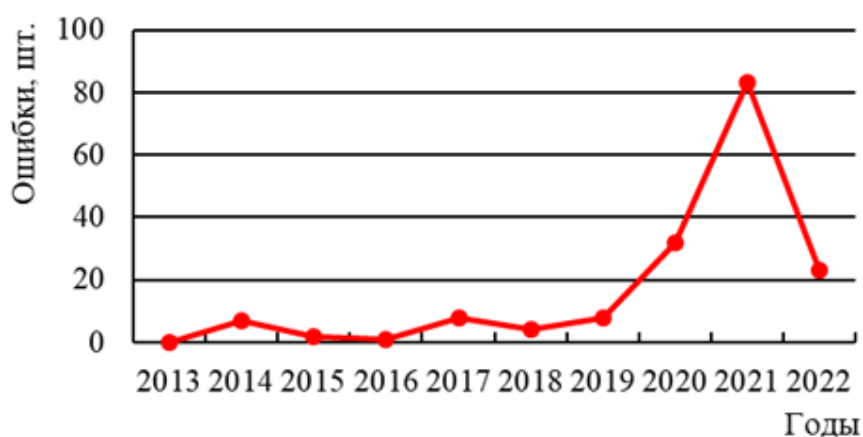


Рис. 1. Число ошибок в записях значений параметров накопителей информации по годам
 Fig. 1. Errors number in the records of data storage parameters values by year

Распределение числа накопителей информации в зависимости от временного сдвига при корректировке приведено на рисунке 2. Задержка записи значений параметров равная нулю означает, что накопители вышли из строя сразу после установки или вовсе не заработали.

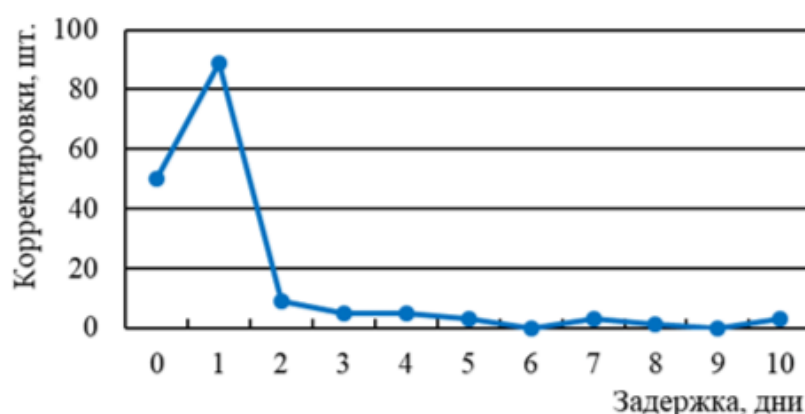


Рис. 2. Распределение числа накопителей информации по дням задержки записи значений параметров

Fig. 2. Distribution of data storage devices number by days of delay in recording parameter values

Обсуждение и выводы

В работе [Cahyadi, 2021] накопители модели DELLBOSS VD, у которых для всех параметров SMART значения отсутствовали, тоже не были включены в дальнейшие анализы. Также была обнаружена аномалия, заключающаяся в том, что для некоторых накопителей в определенный момент времени емкость указана как отрицательная. В тех случаях, когда это происходило, все значения SMART-параметров для этого конкретного накопителя отсутствовали. Но корреляции между отрицательной емкостью и отказом накопителя не обнаружено. На этом основании авторами было принято решение удалить все наблюдения, содержащие отрицательную емкость накопителя, чтобы уменьшить потенциальный шум.

Как вытекает из таблицы 1, доля 168 ошибок в сжатых в 318826 строк данных составляет всего 0,0527 %. И даже относительно только 15274 отказавших накопителей эта доля равна 1,10 %. Поэтому обычный прием при стандартной обработке больших данных, заключающийся в отбрасывании такого незначительного числа ошибок, выглядит вполне оправданным.

Максимальная доля накопителей торговой марки ST (Seagate) и одновременно наибольшая доля ошибок в таблице 2 может быть связана с происходившей в этот период их массовой заменой. Неожиданным на первый взгляд результатом является полное отсутствие ошибок у остальных SSD накопителей, за исключением DELLBOSS. Но если исходить из статистики, то из них вероятность ошибки должна была быть всего для одного накопителя. Поэтому здесь вроде бы всё в порядке.

Единственным тревожным моментом была зависимость скорости их появления от времени. Действительно, согласно рисунку 1, вплоть до 2021 года происходил ускоренный рост числа ошибок. Однако в следующем году он не только приостановился, но даже уменьшился по сравнению с предыдущим. Это вытекает из сопоставления увеличенного в два раза числа ошибок за 2022 год, т. к. данные приведены всего за полгода, с 2021.

Зависимость числа накопителей информации от временного сдвига, связанного с задержкой записи значений параметров при корректировке, приведенное на рисунке 2, имеет вид нормального статистического распределения, обычно получающегося при социологических опросах. Скорее всего это обусловлено тем, что все рассмотренные ошибки вызваны так называемым человеческим фактором, когда даже помеченные как отказавшие накопители не заменялись по несколько дней [Насыров и др., 2022б].

Возможность корректировки большей части ошибок позволяет сделать вывод о пригодности рассматриваемого образца больших данных для построения и тестирования системы оценки и прогнозирования надежности накопителей информации в data-центрах.

Заключение

Таким образом, в результате исследования выявлено, что число ошибок и скорость их ежегодного прироста не является главной причиной проблемы для анализа больших данных по надежности накопителей информации в data-центрах ввиду незначительности их итоговой величины после корректировки. А таковой причиной считаем увеличение с течением времени как числа самих накопителей, так и числа записываемых параметров их состояния, в совокупности приводящих к ускоренному нелинейному росту общего числа данных. Эта проблема отрицательно влияет на возможности применения для предсказания отказа оборудования таких хорошо зарекомендовавших себя в иных случаях инструментов, как нейронные сети и другие методы машинного обучения [Насыров и др., 2022a]. Вследствие указанного обстоятельства предлагается использовать новый матричный метод оперативной оценки и прогнозирования надежности [Насыров и др., 2021a, б; Nasurov et al., 2021], позволяющий преодолеть эту проблему.

Перспектива дальнейших исследований связана с переходом к широкому использованию для хранения информации в data-центрах SSD накопителей [Елизова, Ефимов, 2022]. Хотя в настоящее время они все еще относительно дороже, чем HDD накопители, однако их преимущество в быстродействии при операциях записи/считывания выходит с точки зрения эксплуатации на первый план.

Список литературы

- Демидова Л.А., Филатов А.В. 2021. Контроль и классификация состояния жестких дисков с применением рекуррентных нейронных сетей. Контроль. Диагностика. Т. 24, № 10(280): 36-43. DOI: 10.14489/td.2021.10.pp.036-043. EDN VOVAXJ.
- Демидова Л.А., Филатов А.В. 2022. Разработка модели бинарной классификации состояний дисковых накопителей на основе алгоритма случайного леса с учетом критерия важности признаков. Актуальные проблемы прикладной математики, информатики и механики: сб. тр. межд. науч. конф. (Воронеж, 13–15 декабря 2021). Воронеж: ООО "Вэлборн", 1557-1563. EDN FLBIUQ.
- Демидова Л.А., Филатов А.В. 2021. Разработка модели классификации состояния жестких дисков на основе LSTM-нейронных сетей. Высокопроизводительные вычислительные системы и технологии. Т. 5, № 1: 37-42. EDN WGUDRH.
- Демидова Л.А., Фурсов И.А. 2021. Разработка модели прогнозирования остаточного срока службы накопителей данных с использованием технологии рекуррентных нейронных сетей. Высокопроизводительные вычислительные системы и технологии. Т. 5, № 1: 43-48. EDN CXSKCF.
- Елизова Н.А., Ефимов А.В. 2022. Анализ эффективности и функциональных возможностей программ по выявлению и обработке отказов SSD дисков. Инновации и научно-техническое творчество молодежи: мат. Российской науч.-техн. конф. (Новосибирск, 20–21 апреля 2022). Новосибирск: Сибирский государственный университет телекоммуникаций и информатики, С. 127-131. EDN CXFBAA.
- Насыров И.Н., Насыров И.И., Насыров Р.И. 2022. Большие данные по надежности накопителей информации в data-центрах. Цифровая экономика. 2 (18): 33-37. DOI: 10.34706/DE-2022-02-04. EDN PVGOEX.
- Насыров И.Н., Насыров И.И., Насыров Р.И. 2021. Метод многопараметрической оценки надежности жестких дисков. Приборы. № 2: 13-19. EDN LUFJN.
- Насыров И.Н., Насыров И.И., Насыров Р.И. 2021. Модель для многопараметрической оценки жестких дисков по риску отказа. Автоматизация в промышленности. № 1: 38-42. DOI: 10.25728/avtprom.2021.01.06. EDN EKUFND.
- Насыров И.Н., Насыров И.И., Насыров Р.И. 2022. Прикладные проблемы обеспечения эффективности хранения информации в data-центрах. Социально-экономические и технические системы: исследование, проектирование, оптимизация. № 1 (90): 67-76. EDN LEXFPM.
- Филатов А.В. 2022. Подготовка данных и подбор гиперпараметров модели машинного обучения в задачах классификации. Новые информационные технологии в научных исследованиях: мат. XXVII всеросс. науч.-техн. конф. студентов, молодых ученых и специалистов. (Рязань, 07–09



- декабря 2022). В 2-х томах. Том 1. Рязань: Рязанский государственный радиотехнический университет им. В.Ф. Уткина. 272-273. EDN PBQUSA.
- Юрков А.А., Петров А.В. 2021. Мониторинг здоровья жестких дисков. Современные технологии в теории и практике программирования: сб. мат. науч.-прак. конф. (Санкт-Петербург, 22 апреля 2021). Санкт-Петербург: Федеральное государственное автономное образовательное учреждение высшего образования "Санкт-Петербургский политехнический университет Петра Великого". 228-229. EDN ORSHGC.
- Cahyadi, Forshaw M. 2021. Hard disk failure prediction on highly imbalanced data using LSTM network. Proc. IEEE International Conference on Big Data (Big Data, 15-18 December 2021, Orlando, FL, USA). P. 3985-3991. DOI: 10.1109/BigData52589.2021.9671555.
- Demidova L.A., Fursov I.A. 2021. Aspects of Feature Engineering in the Problem of Predicting the Service Life of Hard Drives. Proc. 3rd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA 2021: 3. Lipetsk, 10–12 November 2021). Lipetsk. P. 1188-1191. DOI: 10.1109/SUMMA53307.2021.9632089. EDN QAXGJS.
- Demidova L.A., Fursov I.A. 2022. Disk Drives Remaining Useful Life Prediction Using the Extreme Learning Machine. Hybrid methods of modeling and optimization in complex systems: Proc. International Workshop "Hybrid methods of modeling and optimization in complex systems" (HMMOCS 2022. Krasnoyarsk, 22–24 November 2022) / Krasnoyarsk Regional Science and Technology City Hall, Russia Siberian Federal University, Reshetnev Siberian State University of Science and Technology. London, United Kingdom: European Proceedings. P. 304-312. DOI: 10.15405/epct.23021.37. EDN FLGYEF.
- Demidova L., Fursov I. 2022. Software Implementation of Neural Recurrent Model to Predict Remaining Useful Life of Data Storage Devices. Communications in Computer and Information Science. Vol. 1526. P. 391-400. DOI: 10.1007/978-3-030-94141-3_31. EDN YYPNEN.
- Demidova L.A. 2020. Recurrent Neural Networks' Configurations in the Predictive Maintenance Problems. IOP Conference Series: Materials Science and Engineering: 2019 Workshop on Materials and Engineering in Aeronautics (Moscow, 16–17 October 2019). Vol. 714. Moscow: Institute of Physics Publishing. P. 012005. DOI: 10.1088/1757-899X/714/1/012005. EDN GSGYWN.
- Diallo M.S., Mokeddem S.A., Braud A., Frey G., Lachiche N. 2021. Identifying benchmarks for failure prediction in industry 4.0. Informatics. 8(4). P. 68. DOI: 10.3390/informatics8040068.
- Filatov A., Demidova L. 2022. Application of Recurrent Networks to Develop Models for Hard Disk State Classification. Communications in Computer and Information Science. Vol. 1526. P. 380-390. DOI: 10.1007/978-3-030-94141-3_30. EDN MPRDYP.
- Kuznietsova N., Kuznietsova M. 2020. Data Mining Methods Application for Increasing the Data Storage Systems Fault-Tolerance. IEEE 2nd International Conference on System Analysis and Intelligent Computing (SAIC 2020: 2. Kyiv, 05–09 October 2020). Kyiv. P. 9239222. DOI: 10.1109/SAIC51296.2020.9239222. EDN ADWUKB.
- Lan X., Feng M., Ng D., Liu Y., Liu J., Xu F., He C. 2021. Adversarial Domain Adaptation with Correlation-Based Association Networks for Longitudinal Disk Fault Prediction. Proc. International Joint Conference on Neural Networks (Virtual, Shenzhen, 18–22 July 2021). Virtual, Shenzhen, DOI: 10.1109/IJCNN52387.2021.9533383. EDN JKAIY.
- Nasyrov I.N., Nasyrov I.I., Nasyrov R.I., Khairullin B.A. 2021. Method for HDD Reliability Multiparametric Assessment. Revista San Gregorio. Is. 44, Special edition. P. 167-178. EDN JBJRZ.
- Peng Y., Wang Y., Wang G., Tsui K.-L. 2021. Doubly Stochastic Cumulative Damage Model for RUL Prediction of HDDs in Uncertain Operating Environments. IEEE Transactions on Industrial Electronics. Vol. 68, No. 9. P. 8743-8752. DOI: 10.1109/TIE.2020.3013777. EDN XQDZHR.
- Shi C., Wu Z., Lv X., Ji Y. 2021. DGTL-Net: A Deep Generative Transfer Learning Network for Fault Diagnostics on New Hard Disks. Expert Systems with Applications. Vol. 169. P. 114379. DOI: 10.1016/j.eswa.2020.114379. EDN HGACRB.

References

- Demidova L.A., Filatov A.V. 2021. Monitoring and classifying the state of hard disks using recurrent neural networks. Testing. Diagnostics. Vol. 24, No. 10 (280): 36-43. DOI 10.14489/td.2021.10.pp.036-043. (in Russian).
- Demidova L.A., Filatov A.V. 2022. Razrabotka modeli binarnoy klassifikatsii sostoyaniy diskovykh nakopiteley na osnove algoritma sluchaynogo lesa s uchetom kriteriya vazhnosti priznakov [Development of a binary classification model of disk drive states based on a random forest algorithm,

- taking into account the criterion of features importance]. Aktual'nyye problemy prikladnoy matematiki, informatiki i mekhaniki: sbornik trudov mezhdunarodnoy nauchnoy konferentsii. (Voronezh, 13–15 December 2021). Voronezh: Velborn LLC. P. 1557-1563.
- Demidova L.A., Filatov A.V. 2021. Development of a hard disk state classification model based on the LSTM architecture of recurrent neural networks. High-performance computing systems and technologies. Vol. 5, No. 1. P. 37-42. (in Russian).
- Demidova L.A., Fursov I.A. 2021. Development of a model for predicting the residual useful life of data drives using recurrent neural network technology. High-performance computing systems and technologies. Vol. 5, No. 1. P. 43-48. (in Russian).
- Elizova N., Efimov A. 2022. Analysis of the effectiveness and functionality of programs for detecting and processing SSD disk failures. Innovatsii i nauchno-tekhnicheskoye tvorchestvo molodezhi: materialy Rossiyskoy nauchno-tekhnicheskoy konferentsii. (Novosibirsk, 20–21 April 2022). Novosibirsk: Siberian State University of Telecommunications and Information Science, P. 127-131. (in Russian).
- Nasyrov I.N., Nasyrov I.I., Nasyrov R.I. 2022. Big data on storage devices reliability in data centers. Cifrovaya ekonomika. No. 2 (18). P. 33-37. DOI 10.34706/DE-2022-02-04. (in Russian).
- Nasyrov I.N., Nasyrov I.I., Nasyrov R.I. 2021. Method for HDD reliability multiparametric assessment. Instruments. No. 2. P. 13-19. (in Russian).
- Nasyrov I.N., Nasyrov I.I., Nasyrov R.I. 2021. A model for HDD multiparametric assessment by failure risk. Automation in Industry. No. 1. P. 38-42. DOI 10.25728/avtprom.2021.01.06. (in Russian).
- Nasyrov I.N., Nasyrov I.I., Nasyrov R.I. 2022. Applied problems of ensuring information storage devices efficiency in data centers. Social'no-ekonomicheskie i tekhnicheskie sistemy: issledovanie, proektirovanie, optimizatsiya. No. 1 (90). P. 67-76. (in Russian).
- Filatov A.V. 2022. Podgotovka dannykh i podbor giperparametrov modeli mashinnogo obucheniya v zadachakh klassifikatsii [Data preparation and selection of hyperparameters of the machine learning model in classification tasks]. Novyye informatsionnyye tekhnologii v nauchnykh issledovaniyakh: materialy XXVII Vserossiyskoy nauchno-tekhnicheskoy konferentsii studentov, molodykh uchenykh i spetsialistov. (Ryazan, 07–09 December 2022). In 2 vol. Vol. 1. Ryazan: Ryazan State Radio Engineering University named after V.F. Utkin, P. 272-273.
- Yurkov A.A., Petrov A.V. 2021. Monitoring zdorov'ya zhestkikh diskov [Monitoring the health of hard drives]. Sovremennyye tekhnologii v teorii i praktike programmirovaniya: sbornik materialov nauchno-prakticheskoy konferentsii (Saint Petersburg, 22 April 2021). Saint Petersburg: Peter the Great St. Petersburg Polytechnic University, P. 228-229.
- Cahyadi, Forshaw M. 2021. Hard disk failure prediction on highly imbalanced data using LSTM network. Proc. IEEE International Conference on Big Data (Big Data, 15-18 December 2021, Orlando, FL, USA). P. 3985-3991. DOI: 10.1109/BigData52589.2021.9671555.
- Demidova L.A., Fursov I.A. 2021. Aspects of Feature Engineering in the Problem of Predicting the Service Life of Hard Drives. Proc. 3rd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA 2021: 3. Lipetsk, 10–12 November 2021). Lipetsk. P. 1188-1191. DOI: 10.1109/SUMMA53307.2021.9632089. EDN QAXGJS.
- Demidova L.A., Fursov I.A. 2022. Disk Drives Remaining Useful Life Prediction Using the Extreme Learning Machine. Hybrid methods of modeling and optimization in complex systems: Proc. International Workshop “Hybrid methods of modeling and optimization in complex systems” (HMMOCS 2022. Krasnoyarsk, 22–24 November 2022) / Krasnoyarsk Regional Science and Technology City Hall, Russia Siberian Federal University, Reshetnev Siberian State University of Science and Technology. London, United Kingdom: European Proceedings. P. 304-312. DOI: 10.15405/epct.23021.37. EDN FLGYEF.
- Demidova L., Fursov I. 2022. Software Implementation of Neural Recurrent Model to Predict Remaining Useful Life of Data Storage Devices. Communications in Computer and Information Science. Vol. 1526. P. 391-400. DOI: 10.1007/978-3-030-94141-3_31. EDN YYPNEN.
- Demidova L.A. 2020. Recurrent Neural Networks' Configurations in the Predictive Maintenance Problems. IOP Conference Series: Materials Science and Engineering: 2019 Workshop on Materials and Engineering in Aeronautics (Moscow, 16–17 October 2019). Vol. 714. Moscow: Institute of Physics Publishing. P. 012005. DOI: 10.1088/1757-899X/714/1/012005. EDN GSGYWN.
- Diallo M.S., Mokeddem S.A., Braud A., Frey G., Lachiche N. 2021. Identifying benchmarks for failure prediction in industry 4.0. Informatics. 8(4). P. 68. DOI: 10.3390/informatics8040068.
- Filatov A., Demidova L. 2022. Application of Recurrent Networks to Develop Models for Hard Disk State



- Classification. Communications in Computer and Information Science. Vol. 1526. P. 380-390. DOI: 10.1007/978-3-030-94141-3_30. EDN MPRDYP.
- Kuznietsova N., Kuznietsova M. 2020. Data Mining Methods Application for Increasing the Data Storage Systems Fault-Tolerance. IEEE 2nd International Conference on System Analysis and Intelligent Computing (SAIC 2020: 2. Kyiv, 05–09 October 2020). Kyiv. P. 9239222. DOI: 10.1109/SAIC51296.2020.9239222. EDN ADWUKB.
- Lan X., Feng M., Ng D., Liu Y., Liu J., Xu F., He C. 2021. Adversarial Domain Adaptation with Correlation-Based Association Networks for Longitudinal Disk Fault Prediction. Proc. International Joint Conference on Neural Networks (Virtual, Shenzhen, 18–22 July 2021). Virtual, Shenzhen, DOI: 10.1109/IJCNN52387.2021.9533383. EDN JKAIY.
- Nasyrov I.N., Nasyrov I.I., Nasyrov R.I., Khairullin B.A. 2021. Method for HDD Reliability Multiparametric Assessment. Revista San Gregorio. Is. 44, Special edition. P. 167-178. EDN JBJRZ.
- Peng Y., Wang Y., Wang G., Tsui K.-L. 2021. Doubly Stochastic Cumulative Damage Model for RUL Prediction of HDDs in Uncertain Operating Environments. IEEE Transactions on Industrial Electronics. Vol. 68, No. 9. P. 8743-8752. DOI: 10.1109/TIE.2020.3013777. EDN XQDZHR.
- Shi C., Wu Z., Lv X., Ji Y. 2021. DGTL-Net: A Deep Generative Transfer Learning Network for Fault Diagnostics on New Hard Disks. Expert Systems with Applications. Vol. 169. P. 114379. DOI: 10.1016/j.eswa.2020.114379. EDN HGACRB.

Конфликт интересов: о потенциальном конфликте интересов не сообщалось.

Conflict of interest: no potential conflict of interest related to this article was reported.

Поступила в редакцию 29.11.2023

Received November 29, 2023

Поступила после рецензирования 04.03.2024

Revised March 04, 2024

Принята к публикации 05.05.2024

Accepted May 05, 2024

ИНФОРМАЦИЯ ОБ АВТОРАХ

INFORMATION ABOUT THE AUTHORS

Насыров Искандар Наирович, доктор экономических наук, доцент, профессор кафедры экономики предприятий и организаций Набережночелнинского института (филиала), Казанский (Приволжский) федеральный университет, г. Казань, Россия

Iskandar N. Nasyrov, Doctor of Economics, Associate Professor, Professor of the Department of Economics of Enterprises and Organizations of the Naberezhnye Chelny Institute (branch) Kazan (Volga Region) Federal University, Kazan, Russia

Насыров Ильдар Искандарович, кандидат технических наук, сервис-менеджер, ООО «Телеком Интеграция», г. Казань, Россия

Ildar I. Nasyrov, Candidate of Technical Sciences, Service Manager of Telecom Integration LLC, Kazan, Russia

Насыров Рустам Искандарович, руководитель портфеля проектов, ООО «Газпромнефть – Цифровые решения», Санкт-Петербург, Россия

Rustam I. Nasyrov, Head of the Project Portfolio of Gazpromneft – Digital Solutions LLC, St. Petersburg, Russia