

НЕЛИНЕЙНОСТЬ РОСТА БОЛЬШИХ ДАННЫХ ПО НАДЕЖНОСТИ НАКОПИТЕЛЕЙ ИНФОРМАЦИИ В DATA-ЦЕНТРАХ

Насыров И.Н., д.э.н., к.ф.-м.н. Набережные Челны

Насыров И.И., к.т.н. Набережные Челны

Насыров Р.И. Набережные Челны

Анализ находящихся в открытом доступе значений параметров накопителей информации data-центров компании Backblaze за длительный период позволил выявить нелинейность роста их количества, приводящую к проблеме «больших данных», существенно ограничивающей возможность исследования их надежности.

Введение

Актуальность исследования связана с ускоряющимся ростом генерируемых в цифровой экономике данных и возникающими при этом трудностями их сохранения и обработки. В создаваемых системах централизованного хранения данных (data-центрах) применяются все более емкие накопители как на жестких дисках (HDD – hard disk drive) для «холодного» хранения, так и твердотельные накопители (SSD – solid state drive) для оперативного использования информации. Соответственно для своевременной оценки и прогнозирования надежности накопителей информации регулярно записываются данные об их состоянии. Научной проблемой является нелинейный рост объема этих данных, затрудняющий проведение их изучения. Цель исследования состоит в выявлении указанной нелинейности увеличения их количества и выработке предложений по повышению оперативности оценки и прогнозирования надежности накопителей информации.

Методы

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 сбоев, в связи с чем исследователи всего мира активно пытаются использовать их в своей работе [1]. В качестве метода исследования выбран анализ значений параметров продолжающих функционировать, снятых досрочно и

отказавших накопителей информации. Были скопированы 29 архивных файлов с указанного сайта с данными за период с 10.04.2013 по 30.06.2022 и распакованы в 3369 файлов формата csv. Эти файлы были сгруппированы в папки для обработки по частям программой Matlab. Ввиду невозможности считывания пустых файлов этой программой три из них за 02.11.2014, 01.11.2015 и 30.01.2017 были временно исключены из рассмотрения. Каждый параметр SMART в базе записан дважды: в сыром и нормализованном виде. К ним в начале строк добавлены пять общих параметров: дата измерения, серийный номер, модель, емкость, признак работоспособности 0 или отказа 1.

Результаты

Получено, что всего за указанный период имеется 325415056 строк с данными во всех файлах. Число ежедневно записываемых параметров (включая пять добавленных) за это время увеличилось с 85 до 179 штук (рис. 1). Конкретный перечень параметров опубликован в [2]. На этом и последующих рисунках также приведены уравнения полиномиальной аппроксимации и достоверность R^2 .

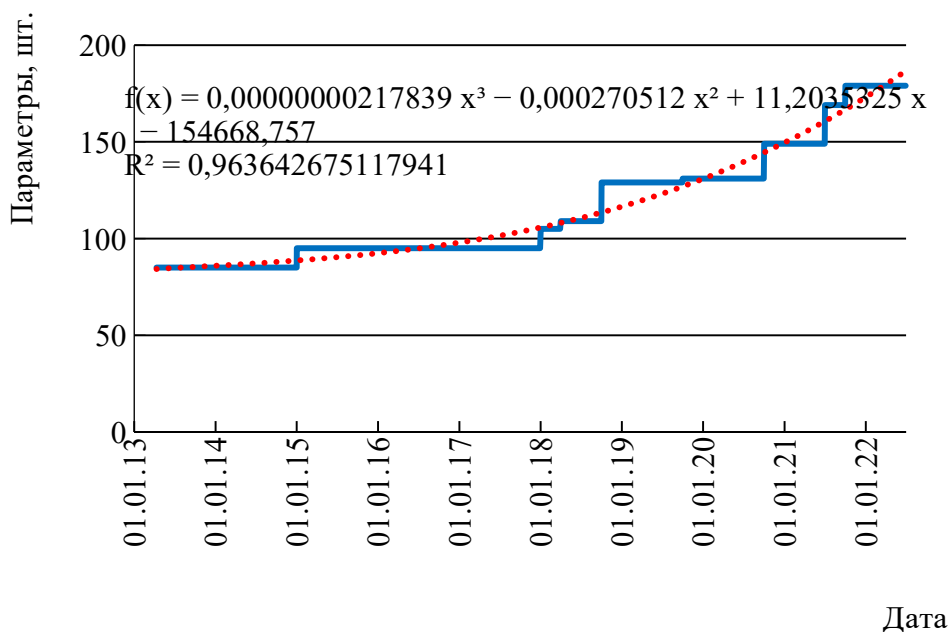


Рисунок 1. Число ежедневно записываемых параметров по годам (синяя сплошная линия) и их аппроксимация (красная точечная линия)

Число накопителей информации на каждый день за указанный период увеличилось с 21195 до 219444 штук (рис. 2). Не всегда информация записывалась со всех накопителей, а в три отмеченных выше дня записи не было вообще, что на рисунке отражается провалами.

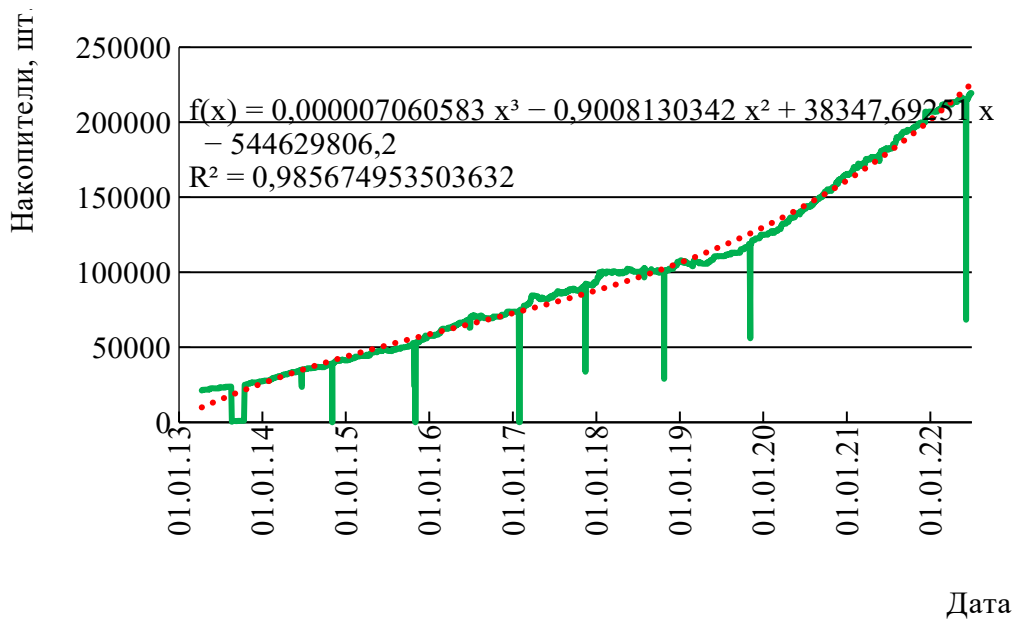


Рисунок 2. Число накопителей информации на каждый день по годам (зеленая сплошная линия) и их аппроксимация (красная точечная линия)

Число ежедневно записываемых данных, полученных перемножением числа параметров на число накопителей информации, за указанный период увеличилось с 1801575 до 39280476 штук (рис. 3).

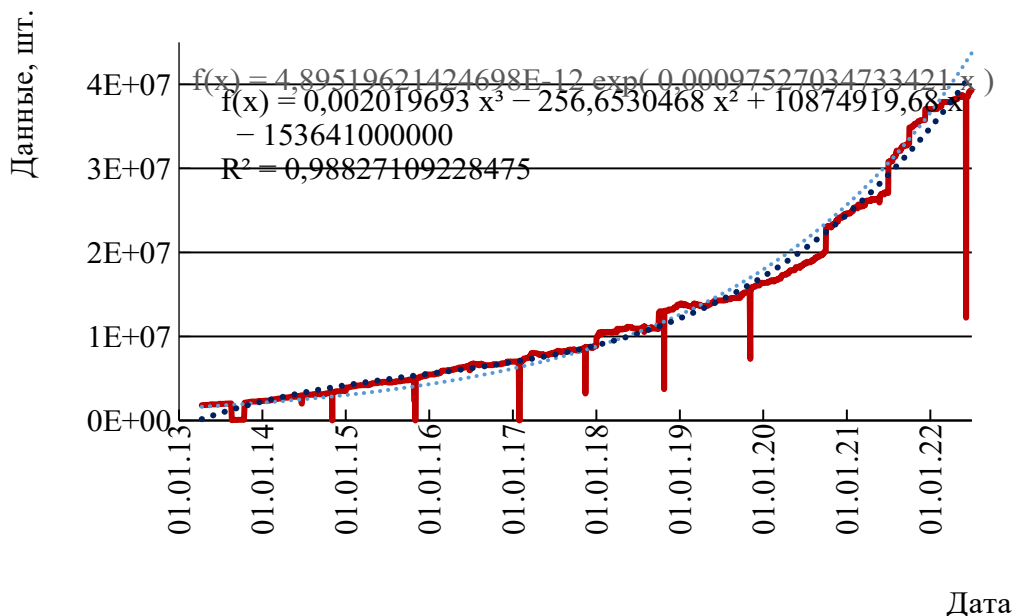


Рисунок 3. Число ежедневно записываемых данных по годам (коричневая сплошная линия) и их аппроксимация (фиолетовая точечная линия)

При увеличении степени полинома достоверность аппроксимации R^2 от времени

для числа параметров, накопителей и данных увеличивается (табл. 1). При этом начиная с третьей степени достоверность для числа данных становится выше, чем для накопителей и параметров. И как видно из рисунков 1-3 по сравнению с ними нелинейность у числа данных действительно заметно больше.

Таблица 1. Зависимость достоверности аппроксимации от степени полинома, отн. ед.

Степень	Достоверность полиномиальной аппроксимации R^2					
	1	2	3	4	5	6
Параметры	0,8620	0,9608	0,9636	0,9637	0,9645	0,9653
Накопитель и	0,9486	0,9774	0,9857	0,9863	0,9880	0,9896
Данные	0,8508	0,9687	0,9883	0,9896	0,9899	0,9911

В сложившихся условиях ускоренного роста обычный способ обработки больших данных путем разделения их на порции приемлемого размера пришлось модифицировать. Если вначале порция могла содержать данные за целый год, то в конце даже за 15 дней уже вызывала переполнение оперативной памяти в 16 Гб. Вследствие этого для уменьшения числа обрабатываемых данных нормализованные параметры пришлось удалить, после чего файлы были преобразованы к единому (максимальному) количеству столбцов и сохранены в формате mat. Пустые позиции заполнялись принятым в программе Matlab для этого случая сочетанием символов «NaN».

Далее обработка происходила следующим образом. Сохраненные файлы по очереди считывались, содержимое ранжировалось по серийному номеру и дате, из них выбирались строки на последнюю дату эксплуатации накопителя, и эти ужатые до предела сведения записывались в отдельные файлы. Эта процедура для новых файлов проводилась еще раз пока не образовался всего один набор сведений для всех 318826 накопителей (на 30.06.2022), имеющих отличающиеся друг от друга серийные номера. Затем они были сгруппированы в 151 модель, приведенные в таблице 2. Указано наименование модели, число нормально работающих, снятых досрочно и отказавших накопителей информации на последнюю дату измерения, а также процент отказов для каждой модели.

Таблица 2. Нормально работающие, снятые досрочно и отказавшие накопители информации, сгруппированные по моделям, шт.

модель	всего	норма	досроч.	отказ	% отк.
1	2	3	4	5	6
00MD00	2	0	2	0	0,00
CT250MX500SSD1	294	272	21	1	0,34
DELLBOSS VD	351	351	0	0	0,00
HGST HDS5C4040ALE630	118	26	88	4	3,39
HGST HDS724040ALE640	42	1	39	2	4,76
HGST HMS5C4040ALE640	8716	3664	4853	199	2,28
HGST HMS5C4040BLE640	16348	12728	3284	336	2,06
HGST HMS5C4040BLE641	1	1	0	0	0,00
HGST HUH721010ALE600	20	20	0	0	0,00
HGST HUH721212ALE600	2629	2605	3	21	0,80
HGST HUH721212ALE604	13226	13138	15	73	0,55
HGST HUH721212ALN604	10973	10802	9	162	1,48
HGST HUH728080ALE600	1203	1126	48	29	2,41
HGST HUH728080ALE604	79	76	0	3	3,80
HGST HUS726040ALE610	50	37	13	0	0,00
Hitachi HDS5C3030ALA630	4664	0	4518	146	3,13
Hitachi HDS5C3030BLE630	1	0	1	0	0,00
Hitachi HDS5C4040ALE630	2719	6	2625	88	3,24
Hitachi HDS722020ALA330	4774	0	4543	231	4,84
Hitachi HDS723020BLA642	11	0	8	3	27,27
Hitachi HDS723030ALA640	1048	0	976	72	6,87
Hitachi HDS723030BLE640	10	0	9	1	10,00
Hitachi HDS724040ALE640	3	0	3	0	0,00
Hitachi HDT721010SLA360	15	0	15	0	0,00
Hitachi HDT725025VLA380	1	0	1	0	0,00
HP SSD S700 250GB	110	0	2	108	98,18
MTFDDAV240TCB	97	89	1	7	7,22
MTFDDAV240TDU	2	0	0	2	100,00
SAMSUNG HD103UJ	14	0	14	0	0,00
SAMSUNG HD154UI	4	0	3	1	25,00
Samsung SSD 850 EVO 1TB	10	0	0	10	100,00
Seagate BarraCuda 120 SSD ZA250CM10003	1116	1106	2	8	0,72
Seagate BarraCuda 120 SSD ZA500CM10003	3	3	0	0	0,00
Seagate BarraCuda SSD ZA2000CM10002	4	3	0	1	25,00
Seagate BarraCuda SSD ZA250CM10002	569	559	1	9	1,58
Seagate BarraCuda SSD ZA500CM10002	18	18	0	0	0,00
Seagate IronWolf ZA250NM10002	9	9	0	0	0,00
Seagate SSD	109	106	0	3	2,75
SSDSCKKB480G8R	4	0	4	0	0,00

ST10000NM001G	22	21	0	1	4,55
ST10000NM0086	1280	1180	22	78	6,09
1	2	3	4	5	6
ST1000LM024 HN	1	1	0	0	0,00
ST12000NM0007	38838	1288	35607	1943	5,00
ST12000NM0008	20607	20033	31	543	2,64
ST12000NM001G	12531	12389	3	139	1,11
ST12000NM0117	30	9	15	6	20,00
ST14000NM0018	41	40	1	0	0,00
ST14000NM001G	10880	10728	11	141	1,30
ST14000NM0138	1689	1571	2	116	6,87
ST1500DL001	2	0	2	0	0,00
ST1500DL003	116	0	26	90	77,59
ST1500DM003	5	0	5	0	0,00
ST16000NM001G	18181	16860	1207	114	0,63
ST16000NM005G	26	26	0	0	0,00
ST18000NM000J	64	60	0	4	6,25
ST2000DL001	12	0	7	5	41,67
ST2000DL003	17	0	9	8	47,06
ST2000DM001	8	0	7	1	12,50
ST2000VN000	10	0	8	2	20,00
ST250LM004 HN	39	0	30	9	23,08
ST250LT007	9	0	1	8	88,89
ST3000DM001	4707	0	3347	1360	28,89
ST31500341AS	787	0	578	209	26,56
ST31500541AS	2188	0	1804	384	17,55
ST3160316AS	64	0	59	5	7,81
ST3160318AS	53	0	41	12	22,64
ST32000542AS	385	0	353	32	8,31
ST320005XXXX	18	0	11	7	38,89
ST320LT007	96	0	14	82	85,42
ST33000651AS	351	0	322	29	8,26
ST3500320AS	1	0	0	1	100,00
ST4000DM000	37037	18359	14012	4666	12,60
ST4000DM001	425	0	391	34	8,00
ST4000DM004	20	20	0	0	0,00
ST4000DM005	90	37	47	6	6,67
ST4000DX000	222	0	143	79	35,59
ST4000DX002	7	0	3	4	57,14
ST500LM012 HN	807	405	225	177	21,93
ST500LM021	33	33	0	0	0,00
ST500LM030	293	232	1	60	20,48
ST6000DM001	15	4	11	0	0,00
ST6000DM004	3	1	2	0	0,00
ST6000DX000	1939	886	966	87	4,49
ST8000DM002	10299	9630	17	652	6,33

ST8000DM004	7	0	0	7	100,00
ST8000DM005	27	25	0	2	7,41
1	2	3	4	5	6
ST8000NM000A	54	54	0	0	0,00
ST8000NM0055	15385	14370	34	981	6,38
ST9250315AS	64	0	53	11	17,19
ST9320325AS	25	0	23	2	8,00
TOSHIBA DT01ACA300	60	0	53	7	11,67
TOSHIBA HDWE160	10	4	6	0	0,00
TOSHIBA HDWF180	23	18	3	2	8,70
TOSHIBA MD04ABA400V	150	97	46	7	4,67
TOSHIBA MD04ABA500V	47	0	45	2	4,26
TOSHIBA MG07ACA14TA	38831	38205	84	542	1,40
TOSHIBA MG07ACA14TEY	511	501	0	10	1,96
TOSHIBA MG08ACA16TA	2494	2488	0	6	0,24
TOSHIBA MG08ACA16TE	6012	5946	2	64	1,06
TOSHIBA MG08ACA16TEY	4051	4032	3	16	0,39
TOSHIBA MQ01ABF050	589	259	19	311	52,80
TOSHIBA MQ01ABF050M	452	278	16	158	34,96
WDC WUH721414ALE6L4	140	139	0	1	0,71
WDC WUH721816ALE6L0	99	99	0	0	0,00
WDC WUH721816ALE6L4	20	20	0	0	0,00
WDC WD1000FYPS	1	0	1	0	0,00
WDC WD1001FALS	2	0	2	0	0,00
WDC WD10EACS	109	0	101	8	7,34
WDC WD10EADS	550	0	488	62	11,27
WDC WD10EADX	29	0	25	4	13,79
WDC WD10EALS	1	0	1	0	0,00
WDC WD10EARS	5	0	5	0	0,00
WDC WD10EARX	3	0	3	0	0,00
WDC WD15EADS	1	0	1	0	0,00
WDC WD15EARS	3	0	3	0	0,00
WDC WD1600AAJB	9	0	3	6	66,67
WDC WD1600AAJS	125	0	112	13	10,40
WDC WD1600BPVT	4	0	1	3	75,00
WDC WD20EFRX	167	0	152	15	8,98
WDC WD2500AAJB	2	0	1	1	50,00
WDC WD2500AAJS	3	0	3	0	0,00
WDC WD2500BEVT	1	0	1	0	0,00
WDC WD2500BPVT	9	0	8	1	11,11
WDC WD2500JB	1	0	0	1	100,00
WDC WD30EFRX	1335	0	1169	166	12,43
WDC WD30EZRS	18	0	16	2	11,11
WDC WD30EZRX	500	0	482	18	3,60
WDC WD3200AAJB	3	0	2	1	33,33
WDC WD3200AAJS	6	0	4	2	33,33

WDC WD3200AAKS	1	0	1	0	0,00
WDC WD3200BEKT	1	0	0	1	100,00
1	2	3	4	5	6
WDC WD3200BEKX	14	0	9	5	35,71
WDC WD3200LPVX	2	0	2	0	0,00
WDC WD40EFRX	50	0	46	4	8,00
WDC WD5000AAJS	1	0	1	0	0,00
WDC WD5000BPKT	24	9	13	2	8,33
WDC WD5000LPCX	57	52	1	4	7,02
WDC WD5000LPVX	350	193	106	51	14,57
WDC WD5002ABYS	6	0	4	2	33,33
WDC WD5003ABYX	2	0	2	0	0,00
WDC WD60EFRX	499	3	424	72	14,43
WDC WD800AAJB	23	0	13	10	43,48
WDC WD800AAJS	30	0	15	15	50,00
WDC WD800BB	56	0	44	12	21,43
WDC WD800JB	19	0	12	7	36,84
WDC WD800JD	1	0	0	1	100,00
WDC WD800LB	2	0	1	1	50,00
WDC WDS250G2B0A	43	42	0	1	2,33
WDC WUH721414ALE6L4	8306	8269	1	36	0,43
WDC WUH721816ALE6L0	2606	2603	0	3	0,12
WDC WUH721816ALE6L4	1180	1179	0	1	0,08

Группировка по торговым маркам нормально работающих, снятых досрочно и отказавших накопителей информации в штуках с указанием минимального, максимального и среднего времени эксплуатации в часах приведена в таблице 3.

Таблица 3. Нормально работающие, снятые досрочно и отказавшие накопители информации в штуках с указанием минимального, максимального и среднего времени эксплуатации в часах, сгруппированные по торговым маркам

торговая марка	всего	норма	доср.	отказ	% отк.	мин.	макс.	средн.
00MD00	2	0	2	0	0,00	14476	14476	14476
CT	294	272	21	1	0,34	33	5415	2752
DELLBOSS	351	351	0	0	0,00	0	0	0
HGST	53405	44224	8352	829	1,55	7	70595	31516
Hitachi	13246	6	12699	541	4,08	191	68438	44629
HP SSD	110	0	2	108	98,18	191	3001	1398
MTFDDAV	99	89	1	9	9,09	3317	14787	13159
SAMSUNG HD	18	0	17	1	5,56	5839	42380	35590
Samsung SSD	10	0	0	10	100,00	145	685	594
Seagate SSD	1828	1804	3	21	1,15	79	32134	15168
SSDSCKKB	4	0	4	0	0,00	1330	1800	1536
ST	179810	108262	59451	12097	6,73	0	72153	27724

TOSHIBA	53230	51828	277	1125	2,11	0	62446	12930
WDC	16419	12608	3279	532	3,24	14	274512	16131

SSD накопители торговой марки DELL BOSS (boot optimized storage solution – оптимизированное для загрузки решение для хранения данных) – это решение от Dell, которое было разработано для разделения операционных систем на разных физических дисках. Идея заключалась в том, чтобы создать аппаратный RAID (redundant array of independent disks – избыточный массив независимых дисков) в виде очень простой конфигурации RAID 1 с ограниченным бюджетом, которая создает полное зеркало диска. Это не повышает производительность, но если что-то случится с диском операционной системы, то существует резервная копия, которая продолжает работать без каких-либо проблем. В отличие от других SSD по ним никаких значений SMART-параметров не записывалось.

Максимальное время эксплуатации свыше 31 года у накопителей торговой марки WDC действительно имело место быть. Ведь и раньше выпускались модели, отличавшиеся особой длительностью работы, у которых головки записи/считывания ни при каких обстоятельствах не касались поверхности дисков, и даже емкость для тех лет у некоторых из них была вполне приличная. Однако с течением времени они все же морально устаревали, а часть просто вышла из строя. Поэтому в конце концов их сняли с эксплуатации (табл. 4).

Таблица 4. Дата окончания эксплуатации, серийный номер, модель, емкость, признак отказа и время эксплуатации в часах отдельных накопителей информации

дата	серийный номер	модель	емкость, байт	отказ	время, ч.
12.11.2015	WD-WCAU45409452	WDC WD10EADS	1,0002E+12	0	274512
30.04.2014	WD-WCAMD2817456	WDC WD800BB	80026361856	0	163730
04.06.2015	WD-WCAMD2252340	WDC WD800BB	80026361856	1	141415
13.08.2014	WD-WCAMD2603237	WDC WD800BB	80026361856	1	137899
29.06.2015	WD-WCAMD2819726	WDC WD800BB	80026361856	0	114239
16.03.2016	WD-WCADW2290554	WDC WD800LB	80026361856	1	90477

Обсуждение и выводы

Полный список всех моделей накопителей в виде, приведенном в таблице 2, публикуется впервые. Усеченные списки со значениями, обработанными в удобном для сравнения накопителей с разной емкостью и временем работы между собой, имеются на сайте компании Backblaze. Однако согласно высказанным там комментариям существует значительная потребность в получении именно полных данных и как раз в том виде, как в настоящей работе.

Однако среди аналитиков данных по надежности накопителей информации, использующих указанную базу, имеются проблемы в их использовании из-за того, что они действительно большие [3]. Как следствие, основная часть исследований выполнялась на данных за период с 2013 по 2017 годы [4]. При прогнозировании только по статистическим данным с окном в 600 дней без учета физической природы главные параметры, обычно применяемые в этом случае [5], оказались неизменными. Выделение по торговым маркам для HDD показало, что одна из них (ST или Seagate) сильно отличается от остальных четырех (HGST, Hitachi, Toshiba, WDC). Она же составляет основную долю всех накопителей. Для уменьшения количества рассматриваемых данных колонки с параметрами, у которых свыше одной трети строк были пустыми, удалялись. Тем не менее, итоговой рекомендацией все равно было снижение собираемых данных даже хотя бы за счет оптимизации частоты их сбора.

Для изучения физических причин отказа HDD накопителей группа китайских исследователей из Школы программного обеспечения и микроэлектроники Пекинского университета и корпорации ZTE применила способ синхронизации по времени изменений параметров SMART [6]. Ими были обнаружены скачки в значениях и показана их взаимосвязь для некоторых параметров. Нами выявлено, что часть этих резких ступенчатых изменений происходит непосредственно перед выходом накопителей из строя [7-9]. Данное обстоятельство можно применить для разработки экономически эффективной системы оперативной оценки и прогнозирования надежности накопителей.

Отсюда первый вывод состоит в том, что проблема больших данных действительно ограничивает возможности исследователей по применению всей имеющейся информации. Понятно, что вследствие нелинейного роста этих данных с течением времени проблема будет только усугубляться.

Второй вывод заключается в том, что хотя по отдельным моделям накопителей одной и той же торговой марки имеется широкий разброс в надежности, тем не менее можно все таки заметить различие и между разными торговыми марками. Это позволяет конкретизировать предпочтения по надежности при приобретении накопителей.

Третий вывод можно сделать о том, что начали широко использоваться SSD накопители. Появились в открытом доступе и продолжают накапливаться по ним SMART-данные. А перспектива дальнейших исследований проистекает из необходимости отдельного анализа HDD и SSD накопителей.

Заключение

Таким образом, главной причиной проблемы «больших данных» по надежности

накопителей информации в data-центрах 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. Указанное обстоятельство негативно отражается на возможности использования для предсказания выхода из строя оборудования хорошо зарекомендовавших себя методов машинного обучения, включая нейронные сети. Поэтому предлагается разработать новый матричный метод многопараметрической оперативной оценки и прогнозирования надежности только по двум последним значениям, основанный в том числе и на физических процессах деградации накопителей, а не только на статистике, что позволяет преодолеть указанные трудности.

Литература

1. Diallo M.S., Mokeddem S.A., Braud A., Frey G., Lachiche N. Identifying benchmarks for failure prediction in industry 4.0 // Informatics. 2021. 8 (4). P. 68. <https://doi.org/10.3390/informatics8040068>
2. Насыров И.Н., Насыров И.И., Насыров Р.И. Прикладные проблемы обеспечения эффективности хранения информации в data-центрах // Социально-экономические и технические системы: исследование, проектирование, оптимизация. 2022. № 1 (90). С. 67-76. URL: https://kpfu.ru/staff_files/F1651550418/SETS._1_90_2022_67_76.pdf
3. 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://kpfu.ru/staff_files/F171761770/Bolshie_dann_po_nadezhn_nakop_inf_v_data_centrah.pdf
4. Mashhadi A.R., Cade W., Behdad S. Moving towards Real-time Data-driven Quality Monitoring: A Case Study of Hard Disk Drives // Procedia Manufacturing. 2018. Vol. 26. P. 1107-1115. <https://doi.org/10.1016/j.promfg.2018.07.147>
5. Pinheiro E., Weber W.D., Barroso L.A. Failure trends in a large disk drive population // Proceedings of the 5th USENIX Conference on File and Storage Technologies (FAST'07). San Jose, California, USA, 13-16 February. 2007. P. 17-28. URL: https://www.usenix.org/legacy/events/fast07/tech/full_papers/pinheiro/pinheiro.pdf
6. Yang H., Li Z., Qiang H., Li Zh., Tu Y., Yang Y. ZTE-predictor: Disk failure prediction system based on LSTM // 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks: Supplemental Volume, DSN-S 2020. 9159170. P. 17-20. <https://doi.org/10.1109/DSN-S50200.2020.00017>. URL:

<https://ieeexplore.ieee.org/document/9159170>

7. Nasyrov I.N., Nasyrov I.I., Nasyrov R.I., Khairullin B.A. Positioning errors indication by Seek error rate and other HDD parameters // Journal of Advanced Research in Dynamical and Control Systems. 2019. Vol. 11. Is. 8 Special Issue. P. 1797-1805. URL: <https://www.jardcs.org/abstract.php?id=2522>

8. Nasyrov I.N., Nasyrov I.I., Nasyrov R.I., Khairullin B.A. Reallocated sectors count parameter for analysing hard disk drive reliability // Journal of Computational and Theoretical Nanoscience. 2019. Vol. 16, No. 12. P. 5298-5302. <https://doi.org/10.1166/jctn.2019.8602>. URL: <https://www.ingentaconnect.com/content/asp/jctn/2019/00000016/00000012/art00063>

9. Nasyrov I.N., Nasyrov I.I., Nasyrov R.I., Khairullin B.A. Spin retry count relation with other HDD parameters // Journal of Computational and Theoretical Nanoscience. 2019. Vol. 16, No. 12. P. 5303-5306. <https://doi.org/10.1166/jctn.2019.8603>. URL: <https://www.ingentaconnect.com/content/asp/jctn/2019/00000016/00000012/art00064>

References in Cyrillics

1. Nasyrov I.N., Nasyrov I.I., Nasyrov R.I. (2022) Prikladnye problemy obespecheniya effektivnosti hraneniya informacii v data-centrah // Social'no-ekonomicheskie i tekhnicheskie sistemy: issledovanie, proektirovanie, optimizaciya. 2022; (1); 67-76. https://kpfu.ru//staff_files/F1651550418/SETS._1_90_.2022_67_76.pdf

2. Nasyrov I.N., Nasyrov I.I., Nasyrov R.I. (2022) Bol'shie dannye po nadezhnosti nakopitelej informacii v data-centrah // Cifrovaya ekonomika. 2022; (2); 33-37. https://kpfu.ru//staff_files/F171761770/Bolshie_dann_po_nadezh_nakop_inf_v_data_centrah.pdf

*Насыров Искандар Наилович,
профессор, д.э.н., к.ф.-м.н. (ecoseti@yandex.ru)
Набережночелнинский институт (филиал)
ФГАОУ ВО «Казанский (Приволжский) федеральный университет»
ORCID 0000-0003-3293-6965*

*Насыров Ильдар Искандарович,
руководитель группы, к.т.н. (ildarec@mail.ru)
ООО «Глобал Дата Консалтинг энд Сервисез»
ORCID 0000-0002-0186-2871*

*Насыров Рустам Искандарович,
руководитель портфеля проектов (rinasyrov@gmail.com)
ООО «Газпромнефть – Цифровые решения»
ORCID 0000-0002-4923-4532*

Ключевые слова

большие данные, накопитель информации, data-центр, надежность.

Iskandar NasYROV, Ildar NasYROV, Rustam NasYROV. Big data growth nonlinearity on storage devices reliability in data centers

Keywords

big data, data storage, data center, reliability

JEL classification: C55 – Большие объемы данных: моделирование и анализ

Abstract

The analysis of the publicly available values of Backblaze data centers storage devices for a long period revealed their number growth nonlinearity, leading to the problem of «big data», significantly limiting the possibility for studying their reliability.

УДК 004.33:338