

Оценка производительности современных дисковых накопителей с целью их использования для буферизации и передачи РСДБ-данных

© И. А. Безруков¹, А. И. Сальников¹, В. В. Васильев¹,
А. В. Вылегжанин^{1,2}

¹ИПА РАН, г. Санкт-Петербург, Россия

²ФТИ им. А. Ф. Иоффе РАН, г. Санкт-Петербург, Россия

Реферат

В настоящее время в ИПА РАН в трех обсерваториях («Светлое» в Ленобласти, «Зеленчукская» в КЧР, «Бадарь» в Бурятии) эксплуатируется 12 систем буферизации и передачи данных, которые включают порядка 250 накопителей на твердых магнитных дисках форм-фактора 3.5". Основная задача систем буферизации и передачи данных заключается в высокоскоростной регистрации данных наблюдений, временном хранении большого объема данных и оперативной передаче этих данных в центр корреляционной обработки ИПА РАН, а также в международных центрах обработки. Существенное увеличение объема данных наблюдений предъявляет высокие требования к надежности и производительности дисковых накопителей систем буферизации и передачи данных. Оценка последней была представлена в статье «Исследование производительности дисковой подсистемы системы буферизации и передачи данных», опубликованной в 2018 г. В соответствии с рекомендациями по результатам проведенных исследований, во всех системах буферизации и передачи данных была проведена установка накопителей SAS и NearLine SAS (NL-SAS) HDD дисков объемом 6 ТБ. Для обеспечения наблюдений с возросшим объемом данных планируется использовать современные дисковые накопители объемом более 6 ТБ на жестких магнитных дисках HDD и твердотельные дисковые накопители SSD. Следует отметить, что объем и надежность дисковых накопителей обоих типов различных производителей в настоящее время существенно выросли. Как отмечается в спецификации производителей, по параметру UER (Unrecoverable Error Rate) SSD-накопители даже надежнее HDD-дисков и проигрывают им только по объему и стоимости. Однако в IT-индустрии последняя грань постепенно стирается. В связи с новыми предложениями производителей дисковых накопителей объемом более 6 ТБ и появлением более совершенного серверного оборудования, а также новых версий операционных и файловых систем, потребовалось провести оценку возможности использования современных дисковых накопителей в системах буферизации и передачи данных.

В статье представлены результаты тестирования новых версий дисковых накопителей для системы буферизации и передачи данных, которые обеспечивают как высокую скорость чтения/записи, так и достаточной большой дисковый объем.

Ключевые слова: HDD и SAS дисковые накопители, операции ввода/вывода, IOPS, операционная система FreeBSD, буферизация и передача РСДБ данных.

Контакты для связи: Безруков Илья Алексеевич (bezrukov@iaaras.ru).

Для цитирования: Безруков И. А., Сальников А. И., Васильев В. В., Вылегжанин А. В. Оценка производительности современных дисковых накопителей с целью их использования для буферизации и передачи РСДБ-данных // Труды ИПА РАН. 2024. Вып. 70. С. 34–38.

<https://doi.org/10.32876/ApplAstron.70.34-38>

Evaluation of the Performance of Modern Disk Drives for Buffering and Transferring VLBI Data

I. A. Bezrukov¹, A. I. Salnikov¹, V. V. Vasilyev¹, A. V. Vylegzhanin^{1,2}

¹Institute of Applied Astronomy of the Russian Academy of Sciences, Saint Petersburg, Russia

²Ioffe Physical-Technical Institute of the RAS, Saint Petersburg, Russia

Abstract

Currently, the Institute of Applied Astronomy Russian Academy of Science operates 12 buffering and data transmission systems in three observatories (“Svetloye” in Leningrad Region, “Zelenchukskaya” in the Republic of KCR, “Badary” in the Republic of Buryatia), which include about 250 drives on solid magnetic disks of the 3.5" form factor. The main task of buffering and transmission systems is to provide the high-speed registration of observational data, temporary storage of a large amount of data and prompt transmission of this data to the center for correlation processing of the IPA RAS, as well as to international processing centers. A significant increase in the volume of observational data imposes high requirements on the reliability and performance of disk drives of buffering and data transmission systems. This as-

assessment was presented in the paper entitled "Performance study of the disk subsystem of the buffering and data transfer system" published in 2018. In accordance with the recommendations of the studies conducted, the installation of SAS and NearLine SAS (NL-SAS) HDD drives with a capacity of 6TB was carried out in all the buffering and data transfer systems. To provide observations with an increased amount of data, it is planned to use modern disk drives with a volume of more than 6 TB on hard magnetic disks HDD and solid-state disk drives SSD. It should be noted that the volume and reliability of disk drives of both types from different manufacturers has now grown significantly.

The paper presents the results of testing new versions of disk drives for buffering and data transfer systems that provide both high read/write speeds and have a sufficiently large disk volume.

Keywords: HDD and SAS disk drives, I/O operations, IOPS, the FreeBSD operating system, buffering and transfer of VLBI data.

Contacts: Ilya A. Bezrukov (bezrukov@iaaras.ru).

For citation: Bezrukov I. A., Salnikov A. I., Vasilyev V. V., Vylegzhanin A. V. Evaluation of the performance of modern disk drives for buffering and transferring VLBI data // Transactions of IAA RAS. 2024. Vol. 70. P. 34–38.

<https://doi.org/10.32876/AplAstron.70.34-38>

Введение

С 2016 г. ИПА РАН эксплуатирует системы буферизации и передачи данных (СБПД), на которых ежегодно регистрируются и передаются до 10 ПБ данных. В настоящее время в связи с длительной эксплуатацией СБПД и существенным увеличением объема данных наблюдений возникла необходимость замены дисковых накопителей. На основе методики оценки производительности дисковой подсистемы, описанной ранее ([Безруков и др., 2018](#)), в данной работе проведена оценка производительности операций ввода/вывода ряда современных дисковых накопителей ([Краткое описание SSD Intel](#); [Твердотельные накопители ThinkSystem](#); [Технические характеристики: Ultrastar SCSI 530](#)), включая накопители на жестких магнитных дисках (HDD) и твердотельные накопите-

ли (SSD). На основе проведенных исследований даны рекомендации по выбору дисковых накопителей для задач высокоскоростной регистрации РСДБ-данных.

Состав макета

Состав измерительного стенда включал сервер Dell с параметрами:

— CPU: Intel(R) Xeon(R) Gold 5320 CPU @ 2.20 GHz (26 core), RAM: 128 GB;

— Motherboard: Supermicro X12SPM-TF с поддержкой NVMe SSD;

— HBA Controller: Dell 12Gbps HBA (LSI SAS3008) и дисковую полку Supermicro SAS 2.0.

Характеристики тестируемых дисковых накопителей от производителей приведены в табл. 1.

Перечень тестируемых дисков

Таблица 1

| Код диска | Производитель | Модель диска | Тип диска | Емкость диска, ТБ |
|-----------|---------------|---------------------|------------|-------------------|
| 1 | TOSHIBA | MG06SCA600E | NL-SAS HDD | 6 |
| 2 | TOSHIBA | MG06SCA600E | NL-SAS HDD | 8 |
| 3 | TOSHIBA | MG06SCA600E | NL-SAS HDD | 6 |
| 4 | TOSHIBA | MG06SCA600E | NL-SAS HDD | 8 |
| 5 | HGST | HUS728T8TAL5204 | NL-SAS HDD | 8 |
| 6 | SEAGATE | ST10000NM002G | NL-SAS HDD | 10 |
| 7 | SEAGATE | ST10000NM018B | NL-SAS HDD | 10 |
| 8 | TOSHIBA | MG07ACA1 | NL-SAS HDD | 14 |
| 9 | TOSHIBA | MG08SCA16TE | NL-SAS HDD | 16 |
| 10 | WDC | WUSTR6432ASS204 | SAS SSD | 3 |
| 11 | INTEL | SSDPF2KX076TZ | NVMe SSD | 7 |
| 12 | SAMSUNG | MZWJLJ7T6HALA-00007 | NVMe SSD | 7 |

Методика тестирования

При проведении тестирования накопителей была использована разработанная авторами методика, которая подробно описана в статье «Исследование производительности дисковой подсистемы системы буферизации и передачи данных» ([Безруков и др., 2018](#)). Тестирование производилось для двух версий операционной системы (FreeBSD-11. И FreeBSD-13.1) штатной утилитой `fiio`. Длительность и размер данных подбирались так, чтобы размер доступной оперативной памяти сервера был превышен. Диски объединялись в группу с помощью ZFS в конфигурации `stripe`.

Исследования включали 4 вида экспериментов, в которых оценивались:

1) максимальная производительность скорости последовательной записи на один диск и сравнение полученной производительности скорости последовательной записи на один HDD- и SSD-диск с данными производителя дисков — Тест 1 (табл. 2);

2) максимальная производительность скорости последовательной записи группы из двух HDD-дисков — Тест 2 (табл. 3);

3) производительность операций ввода/вывода группы из двух HDD-дисков в режиме работы СБПД (одновременное чтение и запись при выставлении скорости порогов 256 Мбайт/с в режиме записи и 30 Мбайт/с в режиме чтения) и двух версий операционной системы FreeBSD — Тест 3 (табл. 4);

4) производительность операций ввода/вывода при одновременном чтении и записи для одного SSD-диска и двух версий операционной системы FreeBSD — Тест 4 (табл. 5).

Файлы заданий

При тестировании для штатной утилиты `fiio` использовались следующие параметры файлов заданий:

1. `direct = 1` (директива, предписывающая не использовать буферизацию операций ввода/вывода);

2. `ioengine = sync` (данный параметр определяет механизм операций ввода/вывода; механизм `sync` и используемая функция `write()` наиболее схожи с принципом работы программы регистрации данных СБПД);

3. `blocksize = 128 k` (размер блока ввода-вывода, обусловлен размером блока файловой системы);

4. `runtime = 600 s` (длительность задания);

5. `unlink = 1` (удаление текстового файла после каждого цикла эксперимента).

Все файлы заданий включали общую секцию `global`

```
[global]
ioengine=sync
bs=128K
direct=1
unlink=1
```

Командная строка в режимах Тест 1 и Тест 2:

```
# fio job file
[write]
directory=${MOUNTPOINT}
filename=FIO-test-write
runtime=600s
time_based
rw=write
size=400G
startdelay=20
```

Командная строка в режиме Тест 3 и Тест 4:

```
# fio job file
[write256m]
startdelay=20
directory=${MOUNTPOINT}
filename=FIO-test-write
rate=256m
size=600G
runtime=600s
time_based
rw=write

[read30m]
startdelay=20
directory=${MOUNTPOINT}
filename=FIO-test-read
rw=read
rate=30m
size=200G
runtime=600s
time_based
```

Результаты эксперимента

В табл. 2 приведены результаты сравнения измеренной скорости записи на один HDD- и SSD-диск для двух версий операционной системы FreeBSD с данными производителя.

В табл. 3 приведены результаты измеренной скорости записи на два HDD-диска для двух версий операционной системы FreeBSD.

В табл. 4 приведены результаты измеренной скорости записи с одновременным чтением на два HDD-диска для двух версий операционной системы FreeBSD.

Таблица 2

Сравнение измеренной скорости записи на один HDD- и SSD-диск с данными производителя

| Тип диска | Код диска | IO Производителя, МБ/с | IO Измеренная, МБ/с | |
|-----------|-----------|------------------------|---------------------|--------------|
| | | | FreeBSD-11.2 | FreeBSD-13.1 |
| HDD | 1 | 241 | 247 | 217 |
| | 2 | 241 | 249 | 209 |
| | 3 | 250 | 256 | 265 |
| | 4 | 260 | 264 | 248 |
| | 5 | 255 | 258 | 216 |
| | 6 | 245 | 263 | 259 |
| | 7 | 263 | 284 | 267 |
| | 8 | 260 | 275 | 272 |
| | 9 | 274 | 273 | 236 |
| SSD | 10 | 2120 | 857 | 1065 |
| SSD NVMe | 11 | 4000 | 1680 | 3995 |
| | 12 | 3500 | 2575 | 3816 |

Таблица 3

Скорость записи на группу из 2 HDD дисков

| Код диска | FreeBSD-11.2 | | | FreeBSD-13.1 | | |
|-----------|--------------|--------------|---------------------------|--------------|--------------|---------------------------|
| | IOPS | Задержка, Мс | Ср. скорость записи, МБ/с | IOPS | Задержка, мс | Ср. скорость записи, МБ/с |
| 1 | 3702 | 0.27 | 485 | 3613 | 0.27 | 474 |
| 2 | 3666 | 0.27 | 481 | 2964 | 0.33 | 389 |
| 3 | 3813 | 0.26 | 500 | 3658 | 0.27 | 479 |
| 4 | 3979 | 0.25 | 522 | 3755 | 0.26 | 492 |
| 5 | 3572 | 0.28 | 468 | 2866 | 0.34 | 376 |
| 6 | 3912 | 0.25 | 513 | 2578 | 0.39 | 338 |
| 7 | 4186 | 0.24 | 549 | 2688 | 0.37 | 352 |
| 8 | 4070 | 0.24 | 533 | 4039 | 0.25 | 529 |
| 9 | 4093 | 0.24 | 537 | 3603 | 0.28 | 472 |

Таблица 4

Производительность операций ввода/вывода группы дисков в режиме работы СВПД для двух версий ОС FreeBSD

| Код диска | FreeBSD-11.2 | | | | FreeBSD-13.1 | | | |
|-----------|--------------|--------|---------------------------|--------|--------------|--------|---------------------------|--------|
| | Задержка, мс | | Ср. скорость записи, МБ/с | | Задержка, мс | | Ср. скорость записи, МБ/с | |
| | чтения | записи | чтения | записи | чтения | записи | чтения | записи |
| 1 | 5.50 | 0.02 | 24 | 268 | 0.10 | 0.02 | 31 | 268 |
| 2 | 4.10 | 0.70 | 31 | 178 | 0.06 | 0.60 | 31 | 268 |
| 3 | 2.99 | 0.02 | 31 | 268 | 0.06 | 0.02 | 31 | 268 |
| 4 | 2.00 | 0.02 | 31 | 268 | 0.06 | 0.02 | 31 | 268 |
| 5 | 10.0 | 0.02 | 12 | 268 | 2.00 | 0.02 | 31 | 268 |
| 6 | 5.80 | 0.70 | 23 | 268 | 2.10 | 0.02 | 31 | 268 |
| 7 | 4.40 | 0.02 | 30 | 268 | 1.90 | 0.02 | 31 | 268 |
| 8 | 15.5 | 0.02 | 8.0 | 268 | 1.40 | 0.02 | 31 | 268 |
| 9 | 11.0 | 0.03 | 12 | 268 | 0.05 | 0.03 | 31 | 268 |

Производительность операций ввода-вывода SSD дисков

| Код диска | FreeBSD-11.2 | | | | FreeBSD-13.1 | | | |
|-----------|--------------|--------|---------------------------|--------|--------------|--------|---------------------------|--------|
| | Задержка, мс | | Ср. скорость записи, МБ/с | | Задержка, мс | | Ср. скорость записи, МБ/с | |
| | чтения | записи | чтения | записи | чтения | записи | чтения | записи |
| 10 | 0.34 | 0.16 | 379 | 789.0 | 0.34 | 0.16 | 381 | 764,0 |
| 11 | 0.55 | 0.06 | 238 | 2097 | 0.47 | 0.03 | 277 | 3854 |
| 12 | 0.37 | 0.04 | 349 | 3033 | 0.49 | 0.04 | 265 | 3085 |

В табл. 5 приведены результаты измеренной скорости записи с одновременным чтением на один SSD-диск для двух версий операционной системы FreeBSD.

Результаты измерений, приведенные в табл. 5, показывают, что производительности SSD-накопителей достаточно для регистрации нескольких каналов одним диском. Например, производительность NVMe- SSD-дисков позволит регистрировать поток данных 8 каналов широкополосной системы преобразования сигналов с одновременной их передачей по научному каналу связи в глобальной информационной сети Интернет в ЦКО РАН.

Заключение

В проведенных исследованиях получены следующие результаты.

1. Проведена оценка производительности операций ввода/вывода ряда современных накопителей. Результаты показали, что все диски за исключением модели MG06SCA800E обеспечивают достаточную производительность и могут быть использованы в существующих и новых версиях СБПД.

2. Один NVMe SSD-накопитель позволяет обеспечить регистрацию потока данных 8 каналов широкополосной системы преобразования сигналов с одновременной передачей их по сети в ЦКО РАН. Однако переход на SSD-накопители в будущих системах потребует изменения логики хране-

ния данных в обсерваториях и передачи их по сети в центры обработки данных.

3. Проведено сравнение двух версий ОС FreeBSD 11.2 и 13.1, в которых с точки зрения работы дисковой подсистемы произошли достаточно серьезные изменения в файловой системе ZFS. Результаты исследования показывают, что версия ZFS ОС FreeBSD 13.1 более стабильная и производительная, особенно с дисками NVMe SSD.

Литература

Безруков И. А., Сальников А. И., Яковлев В. А., Вылегжанин А. В. Исследование производительности дисковой подсистемы системы буферизации и передачи данных // Приборы и техника эксперимента. 2018. № 4. С. 1–6.

Краткое описание SSD Intel [Электронный ресурс]. URL: <https://www.simms.co.uk/PDFS/Products/SSD/ssd-dc-p5510-brief.pdf> (дата обращения: 11.06.2024).

Твердотельные накопители ThinkSystem PM1733 Entry NVMe PCIe 4.0 x4. Руководство по эксплуатации [Электронный ресурс]. URL: <https://download.semiconductor.samsung.com/resources/brochure/PM1733%20NVMe%20SSD.pdf> (дата обращения: 11.06.2024).

Технические характеристики: Ultrastar SCSI 530 [Электронный ресурс]. URL: https://documents.westerndigital.com/content/dam/doc-library/en_us/assets/public/western-digital/product/data-center-drives/ultrastar-ssd-sas-series/data-sheet-ultrastar-dc-ss530.pdf (дата обращения: 11.06.2024).