Dyski: HDD, SSD, SAS i NVMe, wytrzymałość i stan zdrowia

Poziom: średnio

Dotyczy: dyski w serwerach, Windows Server 2022 (S2D), Linux, smartmontools 7.x, nvme-cli

Czas: ~55 min

Dyski zużywają się i psują częściej niż jakakolwiek inna część serwera, więc umiemy rozróżnić ich rodzaje i odczytać stan zdrowia, zanim dysk padnie. Opisujemy dyski HDD, SSD, SAS i NVMe, wytrzymałość (TBW i DWPD), SMART oraz rolę dysków w puli S2D. Komendy z Fiszki wykonujemy jako administrator (PowerShell) albo root (bash), a wyniki w przykładach są poglądowe.

Skróty użyte na tej stronie (17)

Skrót	Co znaczy
CRC	Cyclic Redundancy Check: suma kontrolna dopisana do odpowiedzi czujnika, dzięki której wykrywamy przekłamany odczyt
CSV	w Windowsie: Cluster Shared Volume, dysk widoczny naraz dla wszystkich węzłów klastra; w danych: Comma-Separated Values, plik z wartościami oddzielonymi przecinkami
DWPD	Drive Writes Per Day: ile razy dziennie można zapisać całą pojemność dysku SSD przez okres gwarancji
EDSFF	Enterprise and Datacenter Standard Form Factor: rodzina nowych kształtów dysków serwerowych (E1.S, E3.S)
HDD	Hard Disk Drive: tradycyjny dysk twardy z wirującymi talerzami i głowicą, wolniejszy od SSD, ale tańszy za gigabajt, typowo używany jako warstwa pojemnościowa
IBM	International Business Machines: firma, w której laboratorium powstał model relacyjny i pierwszy działający silnik SQL
NAND	rodzaj pamięci flash, z której zbudowane są dyski SSD; komórki zużywają się przy kasowaniu
NVMe	Non-Volatile Memory Express: protokół komunikacji z dyskami SSD podłączonymi przez złącze PCIe, znacznie szybszy niż starsze SATA i SAS, typowy dla szybkich dysków systemowych
PCIe	PCI Express: szybka magistrala w komputerze, do której podłącza się karty i dyski NVMe

Skrót	Co znaczy
RAID	Redundant Array of Independent Disks: sposób łączenia kilku fizycznych dysków w jeden, widziany przez system jako pojedynczy dysk, zwykle z dodatkową odpornością na awarię jednego z nich
S2D	Storage Spaces Direct: programowa macierz Windows Server z dysków wielu serwerów
SAS	Serial Attached SCSI: złącze i magistrala dyskowa w serwerach, następca równoległego SCSI, pozwala podłączyć wiele dysków i macierzy przez jeden kontroler z większą przepustowością niż SATA; w Azure także Shared Access Signature: podpisany adres z tokenem, który daje czasowy dostęp do danych w koncie magazynu
SATA	Serial ATA: popularne złącze dyskowe znane z komputerów biurowych i tańszych serwerów, prostsze i wolniejsze niż SAS, typowe dla dysków klasy konsumenckiej
SCSI	Small Computer System Interface: standard podłączania dysków
SMART	Self-Monitoring, Analysis and Reporting Technology: wbudowany w dysk mechanizm zbierający liczniki o jego kondycji, na przykład temperaturę i liczbę realokowanych sektorów, odczytywany programami takimi jak smartctl
SSD	Solid State Drive: dysk półprzewodnikowy
TBW	Terabytes Written: łączna ilość danych, którą producent gwarantuje, że dysk SSD wytrzyma zapisać

Fiszka

□□ TYLKO ODCZYT

Wylistowanie dysków fizycznych w Windows: Get-PhysicalDisk

```
Get-PhysicalDisk | Sort-Object Usage | Select-Object FriendlyName, MediaType, BusType,
Usage, OperationalStatus, HealthStatus, Size
```

Kolumna `BusType` mówi, czy dysk jest podłączony jako SATA, SAS czy NVMe, a `Usage` pokazuje rolę w puli. Dyski pamięci podręcznej S2D mają `Usage` równe `Journal`.

Przykład (poglądowy, hv02):

```
FriendlyName    MediaType BusType Usage       OperationalStatus HealthStatus Size
------------    --------- ------- -----       ----------------- ------------ ----
NVMe Cache 1    SSD       NVMe    Journal     OK                Healthy      960 GB
PhysicalDisk1   HDD       SAS     Auto-Select OK                Healthy      1.82 TB
PhysicalDisk3   HDD       SAS     Auto-Select Predictive Failure Warning     1.82 TB
```

SATA = Serial ATA: popularne złącze dyskowe znane z komputerów biurowych i tańszych serwerów, prostsze i wolniejsze niż SAS, typowe dla dysków klasy konsumenckiej · **SAS** = Serial Attached SCSI: złącze i magistrala dyskowa w serwerach, następca równoległego SCSI, pozwala podłączyć wiele dysków i macierzy przez jeden kontroler z większą przepustowością niż SATA; w Azure także Shared Access Signature: podpisany adres z tokenem, który daje czasowy dostęp do danych w koncie magazynu · **NVMe** = Non-Volatile Memory Express: protokół komunikacji z dyskami SSD podłączonymi przez złącze PCIe, znacznie szybszy niż starsze SATA i SAS, typowy dla szybkich dysków systemowych · **S2D** = Storage Spaces Direct: programowa macierz Windows Server z dysków wielu serwerów

TYLKO ODCZYT

Odczyt liczników zużycia i błędów w Windows: Get-StorageReliabilityCounter

```
Get-PhysicalDisk | Get-StorageReliabilityCounter | Select-Object DeviceId, Temperature, Wear, PowerOnHours, ReadErrorsUncorrected, WriteErrorsUncorrected
```

`Wear` to zużycie w procentach (dla SSD i NVMe), `Temperature` to temperatura w stopniach Celsjusza, a `PowerOnHours` to czas pracy. Nie każdy dysk i sterownik wypełnia wszystkie pola, więc pusta kolumna oznacza brak danych, a nie zero [4].

Przykład (poglądowy):

```
DeviceId Temperature Wear PowerOnHours ReadErrorsUncorrected WriteErrorsUncorrected
-------- ----------- ---- ------------ --------------------- ----------------------
1        38                31544        0                     0
3        44                49310        12                    0
10       36          14    12877        0                     0
```

SSD = Solid State Drive: dysk półprzewodnikowy · **NVMe** = Non-Volatile Memory Express: protokół komunikacji z dyskami SSD podłączonymi przez złącze PCIe, znacznie szybszy niż starsze SATA i SAS, typowy dla szybkich dysków systemowych

ZMIENIA SYSTEM

Wskazanie dysku w obudowie: Enable-PhysicalDiskIdentification

```
Get-PhysicalDisk -FriendlyName "NazwaDysku" | Enable-PhysicalDiskIdentification
```

> (gdzie `NazwaDysku` to `FriendlyName` z `Get-PhysicalDisk`, na przykład `PhysicalDisk3`; komenda zapala diodę lokalizacji na kieszeni, jeśli obudowa to obsługuje)

Wyłączamy ją po pracy poleceniem `Disable-PhysicalDiskIdentification` tak samo wywołanym. Pełny przebieg wymiany dysku w puli opisuje strona S2D: wymiana dysku bez przestoju, gdy pula jest prawie pełna.

S2D = Storage Spaces Direct: programowa macierz Windows Server z dysków wielu serwerów

TYLKO ODCZYT

Wylistowanie dysków w Linuksie: lsblk

```
lsblk -d -o NAME,ROTA,SIZE,MODEL,TRAN
```

Kolumna `ROTA` równa 1 oznacza dysk wirujący (HDD), a 0 SSD lub NVMe. `TRAN` pokazuje transport: `sata`, `sas` albo `nvme`.

HDD = Hard Disk Drive: tradycyjny dysk twardy z wirującymi talerzami i głowicą, wolniejszy od SSD, ale tańszy za gigabajt, typowo używany jako warstwa pojemnościowa · **SSD** = Solid State Drive: dysk półprzewodnikowy · **NVMe** = Non-Volatile Memory Express: protokół komunikacji z dyskami SSD podłączonymi przez złącze PCIe, znacznie szybszy niż starsze SATA i SAS, typowy dla szybkich dysków systemowych · **lsblk** = list block devices: pokaż dyski i partycje

TYLKO ODCZYT

Odczyt stanu zdrowia dysku SATA lub SAS: smartctl -a

```
sudo smartctl -a /dev/sdX
```

> (gdzie `/dev/sdX` to dysk do sprawdzenia, na przykład `/dev/sda`; dyski za kontrolerem RAID wymagają dodatkowej opcji `-d` zgodnej z kontrolerem)

Przykład (poglądowy, wycinek dla dysku w bkp01):

```
  5 Reallocated_Sector_Ct   0x0033   100   100   010    Pre-fail  Always       -       0
  9 Power_On_Hours          0x0032   062   062   000    Old_age   Always       -
```

```
34712
194 Temperature_Celsius     0x0022   064   045   000    Old_age   Always       -
36
197 Current_Pending_Sector  0x0012   100   100   000    Old_age   Always       -       0
198 Offline_Uncorrectable   0x0010   100   100   000    Old_age   Offline      -       0
199 UDMA_CRC_Error_Count    0x003e   200   200   000    Old_age   Always       -       0
```

Najwięcej mówią atrybuty 5, 197 i 198 (sektory uszkodzone i czekające na realokację). Atrybut 199 rośnie przy złym kablu lub złączu, a nie przy złym dysku.

SATA = Serial ATA: popularne złącze dyskowe znane z komputerów biurowych i tańszych serwerów, prostsze i wolniejsze niż SAS, typowe dla dysków klasy konsumenckiej · **SAS** = Serial Attached SCSI: złącze i magistrala dyskowa w serwerach, następca równoległego SCSI, pozwala podłączyć wiele dysków i macierzy przez jeden kontroler z większą przepustowością niż SATA; w Azure także Shared Access Signature: podpisany adres z tokenem, który daje czasowy dostęp do danych w koncie magazynu · **RAID** = Redundant Array of Independent Disks: sposób łączenia kilku fizycznych dysków w jeden, widziany przez system jako pojedynczy dysk, zwykle z dodatkową odpornością na awarię jednego z nich · **smartctl** = narzędzie z pakietu smartmontools do odczytu i testowania danych SMART z dysków ATA/SATA i NVMe z linii poleceń, wywoływane pod spodem przez wtyczkę SMART agenta2

TYLKO ODCZYT

Szybka ocena dysku: smartctl -H

```
sudo smartctl -H /dev/sdX
```

Wynik `PASSED` oznacza tylko, że żaden atrybut nie przekroczył progu producenta. Dysk potrafi go mieć na kilka dni przed awarią, więc czytamy też liczniki.

smartctl = narzędzie z pakietu smartmontools do odczytu i testowania danych SMART z dysków ATA/SATA i NVMe z linii poleceń, wywoływane pod spodem przez wtyczkę SMART agenta2

TYLKO ODCZYT

Wylistowanie dysków NVMe: nvme list

```
sudo nvme list
```

NVMe = Non-Volatile Memory Express: protokół komunikacji z dyskami SSD podłączonymi przez złącze PCIe, znacznie szybszy niż starsze SATA i SAS, typowy dla szybkich dysków systemowych · **nvme** = NVM Express: narzędzie nvme-cli do odczytu stanu i zdrowia dysków NVMe, na przykład nvme smart-log

◫◫ TYLKO ODCZYT

Odczyt zdrowia NVMe: nvme smart-log

```
sudo nvme smart-log /dev/nvme0
```

Przykład (poglądowy):

```
critical_warning                    : 0
temperature                         : 41 C
available_spare                     : 100%
available_spare_threshold           : 10%
percentage_used                     : 7%
data_units_written                  : 18,402,113
media_errors                        : 0
num_err_log_entries                 : 0
```

`percentage_used` to zużycie według producenta (może przekroczyć 100), `available_spare` to zapasowe komórki, a `media_errors` liczy błędy, których dysk nie poprawił [5]. Te same dane daje `smartctl -a /dev/nvme0`.

◫◫ **NVMe** = Non-Volatile Memory Express: protokół komunikacji z dyskami SSD podłączonymi przez złącze PCIe, znacznie szybszy niż starsze SATA i SAS, typowy dla szybkich dysków systemowych · **nvme** = NVM Express: narzędzie nvme-cli do odczytu stanu i zdrowia dysków NVMe, na przykład nvme smart-log

Z życia wzięte

◫◫ Z ŻYCIA WZIĘTE

Wtorek, 2:14 w nocy. Zabbix budzi nas alarmem o dysku w `hv02`: jeden z dysków HDD ma status `Predictive Failure`. Rano czytamy `Get-StorageReliabilityCounter` i widzimy, że dysk ma 49 tysięcy godzin pracy i dwanaście niepoprawionych błędów odczytu, a dwa dyski obok mają po zero. Wiemy więc, że to nie wspólna awaria kontrolera, tylko jeden dysk, i zamawiamy wymianę, zanim padnie drugi.

Skąd się wzięło

Pierwszy dysk twardy, IBM 350 RAMAC, wszedł na rynek w 1956 roku: pięć milionów znaków na pięćdziesięciu talerzach o średnicy 24 cali, a całość ważyła ponad tonę [1]. Dysk NVMe to ostatni etap długiej drogi: dyski SSD z pamięcią NAND zaczęły zastępować wirujące w serwerach na przełomie pierwszej i drugiej dekady XXI wieku, a protokół NVM Express w wersji 1.0 opublikowano w 2011 roku [2].

Różnica dla administratora jest taka: dyski wirujące zawodzą mechanicznie, zwykle po kilku latach i z zapowiedzią (rosnące błędy), a SSD zużywają się przy każdym zapisie i potrafią paść bez słowa. Dlatego liczniki zdrowia czytamy inaczej dla każdego z nich.

Wnętrze dysku twardego Seagate ST33232A: talerz, głowica na ramieniu i silnik wrzeciona. Cały dys

Wnętrze dysku twardego Seagate ST33232A: talerz, głowica na ramieniu i silnik wrzeciona. Cały dysk ma kilka części ruchomych, które się zużywają. Źródło: Wikimedia Commons (autor i wolna licencja podane na stronie pliku).

Rodzaje i zasada działania

HDD: dysk wirujący

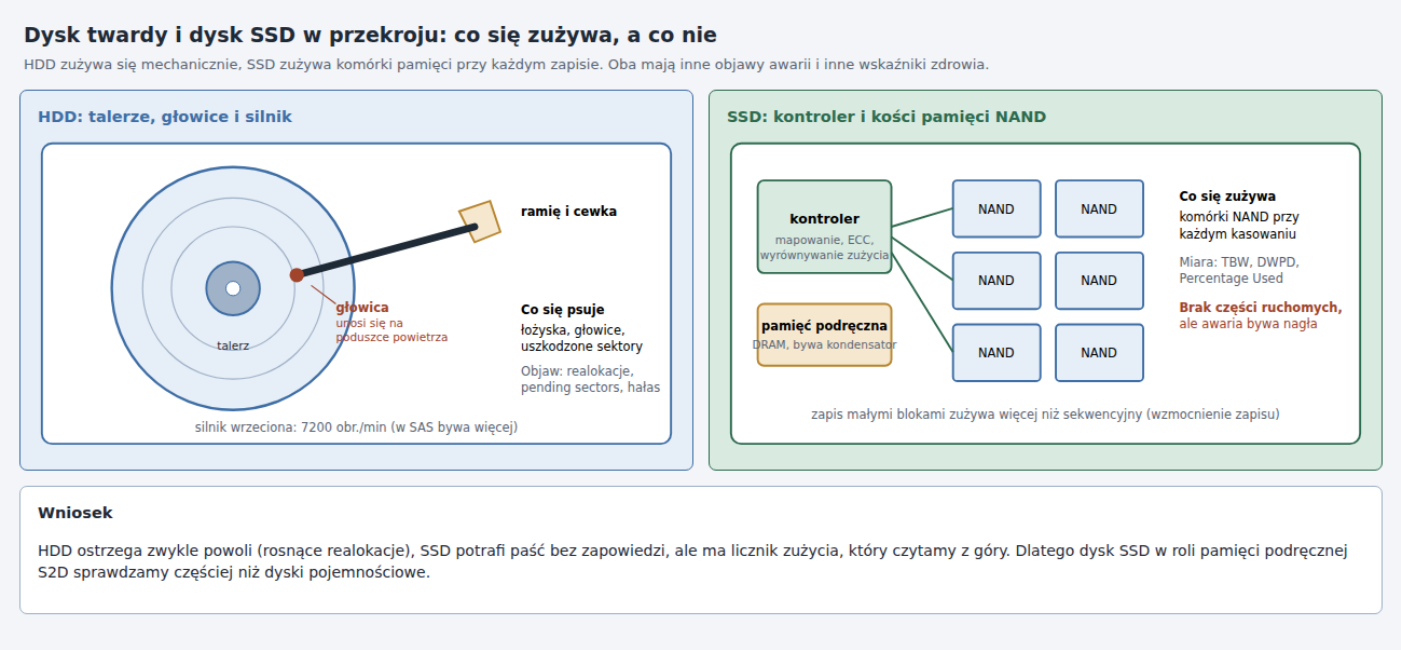
> **Pojęcie: HDD (Hard Disk Drive).** Dysk, który zapisuje dane magnetycznie na obracających się talerzach, a głowica na ramieniu czyta i zapisuje na nich w odpowiedniej ścieżce. Czas dostępu zależy od ruchu głowicy i obrotu talerza.

Dyski serwerowe obracają się z prędkością 7200 obr./min (pojemnościowe) albo 10 000 i 15 000 (dyski SAS pod szybsze zadania). Zaletą jest cena za terabajt i duża pojemność, wadą mała liczba operacji na sekundę (kilkaset), wrażliwość na wstrząsy i hałas. W S2D dyski HDD są warstwą pojemności, a przed nimi stoi warstwa szybkich dysków.

SSD: dysk półprzewodnikowy

> **Pojęcie: SSD (Solid-State Drive).** Dysk bez części ruchomych, w którym dane leżą w kościach pamięci NAND, a kontroler zarządza mapowaniem, korekcją błędów i równomiernym zużyciem komórek.

Kość NAND zużywa się przy każdym kasowaniu bloku, więc kontroler rozkłada zapisy po wszystkich komórkach (wyrównywanie zużycia) i trzyma zapasową pulę. Dyski serwerowe mają zwykle kondensatory, które pozwalają dokończyć zapis przy zaniku prądu (ochrona przed utratą zasilania). Dysk biurkowy takich kondensatorów zwykle nie ma, co jest jedną z przyczyn, dla których dysków z półki sklepowej do serwerów nie bierzemy.



Rysunek 1. Dysk HDD i SSD w przekroju. Po lewej mechanika, która się zużywa, po prawej komórki NAND, które zużywają się przy zapisie. (kliknięcie otwiera pełny rozmiar)

SATA, SAS i NVMe: interfejsy

> **Pojęcie: interfejs dysku.** Standard złącza i protokołu, którym dysk rozmawia z kontrolerem. Wpływa na szybkość, liczbę dysków na jednym kontrolerze i odporność na awarie ścieżki.

SATA (do 6 Gb/s, około 550 MB/s w praktyce) to najtańszy i najpopularniejszy interfejs, jedna ścieżka do dysku. SAS (Serial Attached SCSI, 12 Gb/s w wersji SAS-3) jest przeznaczony do serwerów: kontrolery SAS obsługują także dyski SATA, a dyski SAS mają zwykle dwa porty, więc awaria jednej ścieżki nie odcina dysku. NVMe (NVM Express) rozmawia z dyskiem przez magistralę PCIe, więc jeden dysk daje kilka gigabajtów na sekundę i dziesiątki razy więcej operacji na sekundę niż SATA.

Dysk SAS 2,5 cala. Złącze na brzegu mieści dwa porty danych, więc dysk można podłączyć dwiema

Dysk SAS 2,5 cala. Złącze na brzegu mieści dwa porty danych, więc dysk można podłączyć dwiema niezależnymi ścieżkami. Źródło: Wikimedia Commons (autor i wolna licencja podane na stronie pliku).

Kształty NVMe: M.2, U.2, U.3, E1.S i E3.S

M.2 to mała płytka wkładana do gniazda na płycie, popularna w komputerach i w serwerach jako dysk systemowy (rozmiary typu 2280 i 22110). U.2 to dysk w obudowie 2,5 cala z wymianą na gorąco z przodu serwera, a U.3 jest jego rozszerzeniem: ta sama kieszeń i złącze przyjmują dysk SATA, SAS albo NVMe, zależnie od kontrolera. E1.S i E3.S to nowsze formaty z rodziny EDSFF, w których dysk jest cienką „linijką", mieszczącą się gęściej w serwerach 1U i 2U.

M.2 w wersjach SATA i NVMe. Oba mają ten sam kształt, ale różnią się wycięciem w złączu i protoko

M.2 w wersjach SATA i NVMe. Oba mają ten sam kształt, ale różnią się wycięciem w złączu i protokołem, więc nie są wymienne. Źródło: Wikimedia Commons (autor i wolna licencja podane na stronie pliku).

1 TB NVMe w formacie M.2 2280. Taki dysk bywa pojedynczym dyskiem systemowym w serwerze, n

1 TB NVMe w formacie M.2 2280. Taki dysk bywa pojedynczym dyskiem systemowym w serwerze, najczęściej dwa w mirrorze. Źródło: Wikimedia Commons (autor i wolna licencja podane na stronie pliku).

ZRZUT EKRANU DO UZUPEŁNIENIA

Zdjęcie do uzupełnienia: dysk U.2 albo U.3 w szufladzie z przodu serwera i porównanie z dyskiem E3.S, z opisem kieszeni i kolorem uchwytu. Plik na Wikimedia Commons nie został potwierdzony, więc wstawiamy zdjęcie własne.

Wytrzymałość: TBW i DWPD

Pojęcie: TBW (Terabytes Written) i DWPD (Drive Writes Per Day). TBW to całkowita ilość danych, jaką producent gwarantuje, że dysk wytrzyma. DWPD to ile razy dziennie można zapisać całą pojemność dysku przez okres gwarancji (zwykle pięć lat).

Zależność jest prosta: DWPD = TBW / (pojemność razy 365 razy lata gwarancji). Dysk 1,92 TB z TBW równym 3504 TB wytrzyma 3504 / (1,92 razy 365 razy 5), czyli 1 zapis całego dysku dziennie przez pięć lat. Typowe klasy dysków serwerowych to „read intensive" (około 1 DWPD), „mixed use" (około 3) i „write intensive" (kilka do kilkunastu).

Microsoft zaleca dla dysków pamięci podręcznej w S2D wytrzymałość co najmniej 3 DWPD albo co najmniej 4 TBW na dobę [3]. To dlatego pamięć podręczna S2D nie powinna siedzieć na tanim dysku konsumenckim, który ma ułamek DWPD.

SMART i stan zdrowia NVMe

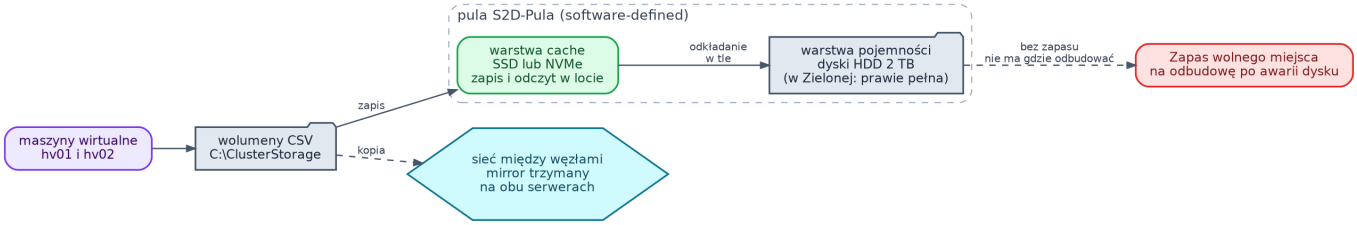
> **Pojęcie: SMART (Self-Monitoring, Analysis and Reporting Technology).** Mechanizm w dysku, który zbiera liczniki własnego stanu: sektory uszkodzone, czas pracy, temperaturę, błędy. Odczytujemy je programem, a dysk sam niczego nie zgłosi.

Dla dysków SATA patrzymy na atrybuty 5 (Reallocated), 197 (Pending), 198 (Offline Uncorrectable) i 199 (błędy CRC kabla). Dla SAS najbardziej mówi licznik „grown defect list" (lista uszkodzonych sektorów dodanych od produkcji). Dla NVMe jest osobny dziennik: `critical_warning`, `available_spare`, `percentage_used`, `media_errors`. Wartości progowe są w karcie katalogowej producenta i w dokumentacji `smartctl`.

Dyski w S2D: pamięć podręczna i pojemność

> **Pojęcie: S2D (Storage Spaces Direct).** Technologia Windows Server, która łączy lokalne dyski kilku serwerów w jedną wspólną pulę, odporną na awarię dysku i całego serwera.
>
> Opisujemy ją szerzej na stronie Storage Spaces i warstwy magazynu.

W puli S2D szybkie dyski (NVMe albo SSD) pełnią rolę pamięci podręcznej, a wolniejsze (HDD albo SSD) są warstwą pojemności. Zapisy najpierw trafiają na szybkie dyski, a w tle są odkładane na pojemnościowe. Microsoft wymaga co najmniej dwóch urządzeń pamięci podręcznej na serwer i podaje, że przy dwóch serwerach z mieszanką HDD potrzeba co najmniej dwóch szybkich i czterech pojemnościowych dysków (łącznie w klastrze) [3].



Rysunek 2. Warstwy puli S2D: maszyny wirtualne piszą na wolumen CSV, zapis trafia na dyski cache, a w tle jest odkładany na dyski pojemnościowe. Przy prawie pełnej puli nie ma miejsca na odbudowę po awarii dysku. (kliknięcie otwiera pełny rozmiar)

W Zielonej `r1-05` pokazuje w puli `S2D-Pula` dyski HDD po 2 TB i prawie pełną pulę. Czy `hv01` i `hv02` mają też dyski cache, z samego tego przykładu nie wynika, więc sprawdzamy to komendą `Get-PhysicalDisk` (rola `Journal`). Jeżeli ich nie mają, jest to luka do omówienia, bo wymagania Microsoftu zakładają szybką warstwę przy dyskach HDD.

Wszystkie rodzaje w jednej tabeli

Cecha	HDD (SATA lub SAS)	SSD SATA	SSD SAS	NVMe
Medium	talerze magnetyczne	NAND	NAND	NAND
Interfejs	SATA 6 Gb/s, SAS 12 Gb/s	SATA 6 Gb/s	SAS 12 Gb/s	PCIe
Typowa szybkość	do około 250 MB/s	do około 550 MB/s	do około 1 GB/s	od kilku GB/s
Operacje na sekundę	rzędu setek	dziesiątki tysięcy	dziesiątki tysięcy	setki tysięcy i więcej
Dwa porty	SAS tak, SATA nie	nie	tak	zależy od modelu
Wymiana na gorąco	tak, w kieszeni	tak	tak	tak (U.2, U.3, E3.S), nie dla M.2
Rola w S2D	pojemność	pojemność albo cache	pojemność albo cache	cache albo pojemność
Cena za TB	najniższa	średnia	wyższa	najwyższa

W serwerowni

Dyski w `hv01` i `hv02` w szafie R2 wymieniamy tylko dyskami tego samego rodzaju, o tej samej lub większej pojemności i z klasą wytrzymałości nie niższą niż stary dysk. Nie mieszamy rozmiarów sektorów (512e i 4Kn) bez sprawdzenia dokumentacji, a dysk zapasowy trzymamy w szafie lub w magazynie, zamiast zamawiać po awarii.

Monitoring: dyski w puli S2D czytamy przez `Get-PhysicalDisk` i `Get-StorageReliabilityCounter`, a nie przez SMART agenta Zabbixa, bo tak opisuje to strona Zabbix: SMART dysków na Windows przez agenta 2. SMART w Zabbixie zostaje dla dysków systemowych i dla Linuksa, na przykład `bkp01`. Progi robocze do dostrojenia: zużycie SSD powyżej 80 procent to ostrzeżenie, powyżej 90 to alarm, a temperatura HDD powyżej 50 stopni to ostrzeżenie.

Poza serwerownią te same zasady stosujemy w NAS-ach domowych (dyski do pracy ciągłej, a nie biurkowe), w kamerach monitoringu (dyski zapisujące bez przerwy 24 godziny na dobę) i w laptopach (dysk NVMe M.2 z licznikiem zużycia).

Co sprawdzamy u dostawcy

- Rodzaj i interfejs każdego dysku w ofercie: HDD, SSD albo NVMe, SATA czy SAS, a nie samo „dysk 2 TB".
- Klasę wytrzymałości SSD (DWPD albo TBW) zapisaną w ofercie, a dla cache S2D co najmniej 3 DWPD.
- Ochronę przed utratą zasilania w dyskach SSD (kondensatory) i tryb zapisu po awarii.
- Listę zgodności dysków z kontrolerem i serwerem oraz fakt, że dysk jest dostarczony w kieszeni (carrier) do wymiany na gorąco.
- Rozmiar sektora (512e albo 4Kn) i zgodność z S2D.
- Dyski zapasowe w zestawie: ile i jakiego rodzaju.
- Odczyt liczników zdrowia po montażu jako punkt odniesienia (`Get-StorageReliabilityCounter`, `smartctl`).

Mity i sprostowania

Obiegowa wersja	Jak jest naprawdę
SSD psuje się szybko, więc HDD jest bezpieczniejszy	SSD ma licznik zużycia, który czytamy z góry, a HDD zawodzi mechanicznie, więc oba wymagają monitoringu
SMART ostrzega przed każdą awarią	Część awarii przychodzi bez zapowiedzi, a status `PASSED` mówi tylko, że progi producenta nie zostały przekroczone
RAID zastępuje kopię zapasową	RAID chroni przed awarią dysku, a nie przed skasowaniem plików, błędem człowieka i oprogramowaniem szyfrującym
NVMe to po prostu szybszy SSD	NVMe to inny protokół na PCIe, z innym złączem i innym dziennikiem zdrowia
Dysk z półki sklepowej wystarczy do serwera	Dyski biurkowe nie mają ochrony zasilania, mają niższą wytrzymałość i nie są przewidziane do pracy ciągłej

Oś czasu

Data	Wydarzenie
1956	IBM 350 RAMAC, pierwszy dysk twardy
2011	Opublikowanie specyfikacji NVM Express 1.0
2016	Windows Server 2016 z Storage Spaces Direct

Częste błędy

Objaw	Przyczyna	Co robimy
`smartctl` nie widzi dysków za kontrolerem	dyski są ukryte za RAID-em	używamy `-d` zgodnie z kontrolerem albo narzędzia producenta
Pole `Wear` jest puste	sterownik albo dysk nie zgłasza	sprawdzamy `nvme smart-log` albo narzędzie producenta
Pula S2D nie odbudowuje się po awarii dysku	za mało wolnego miejsca w puli	zwalniamy miejsce albo dokładamy dysk, patrz r1-05
Dysk w nowej kieszeni nie jest widoczny dla puli	stary podział albo metadane z innego systemu	czyścimy dysk według procedury z r1-05 po sprawdzeniu, że to właściwy
Alarm temperatury dysków tylko w jednej szafie	zatkany wlot lub zły przepływ powietrza	sprawdzamy ΔT szafy według strony o przepływie powietrza

□□ Ćwiczenia

1. Dysk SSD 3,84 TB ma w karcie katalogowej 7000 TBW i gwarancję pięć lat. Ile wynosi DWPD?

Odpowiedź

7000 / (3,84 razy 365 razy 5) daje około 1 DWPD. To klasa „read intensive", za słaba na cache S2D, który według Microsoftu ma mieć co najmniej 3 DWPD.

1. `smartctl -H` pokazuje `PASSED`, a atrybut 197 ma wartość 24. Czy dysk jest zdrowy?

Odpowiedź

Nie wiemy tego na pewno. Atrybut 197 to sektory czekające na realokację, więc wartość różna od zera jest sygnałem ostrzegawczym, nawet gdy ogólny wynik to `PASSED`. Obserwujemy trend i planujemy wymianę.

1. Dlaczego dla dysków w puli `S2D-Pula` czytamy `Get-StorageReliabilityCounter`, a nie SMART przez agenta Zabbixa?

Odpowiedź

Dyski za sterownikiem puli nie oddają się `smartctl` jako zwykłe, pojedyncze nośniki, więc odczyty SMART bywają niespójne, jak opisuje strona o SMART w Zabbixie. Moduł Storage zna pulę i zwraca liczniki niezawodności oraz stan zdrowia w sposób, który jest wiarygodny.

Powiązane: Serwer od środka: od szafy mainframe'u do bursztynowej diody na hv02 · RAID i kontrolery · Firmware i wymiana na gorąco · S2D: wymiana dysku bez przestoju · Zabbix: SMART dysków na Windows

Źródła:

1. en.wikipedia.org: IBM 305 RAMAC i dysk IBM 350
2. en.wikipedia.org: NVM Express
3. Microsoft Learn: Storage Spaces Direct hardware requirements
4. Microsoft Learn: Get-StorageReliabilityCounter
5. Strony podręcznika: nvme-smart-log i smartctl

Revision #14
Created 2026-10-03 15:34:19 UTC by Tomasz Konieczny
Updated 2026-10-26 14:22:35 UTC by Tomasz Konieczny