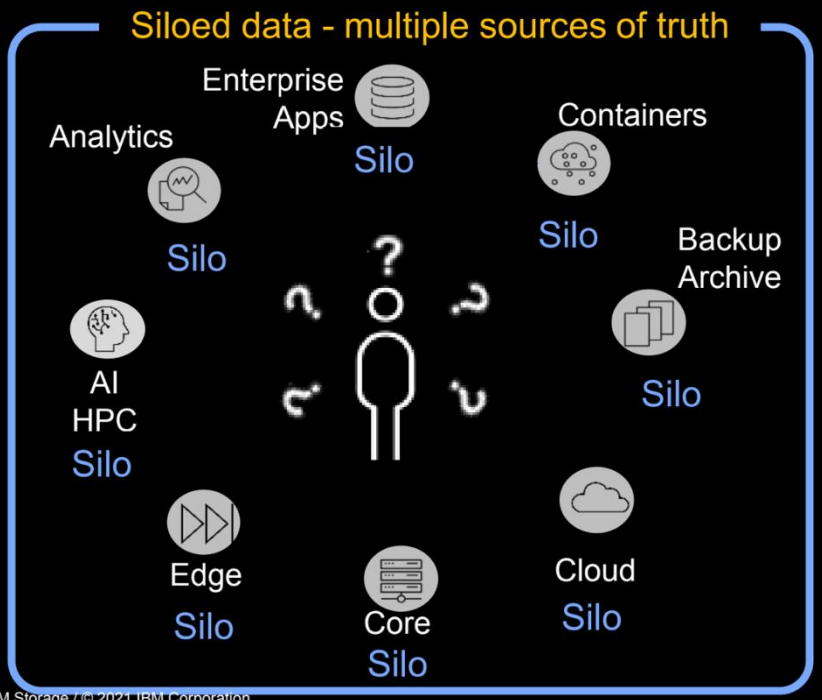
IBM Storage Scale

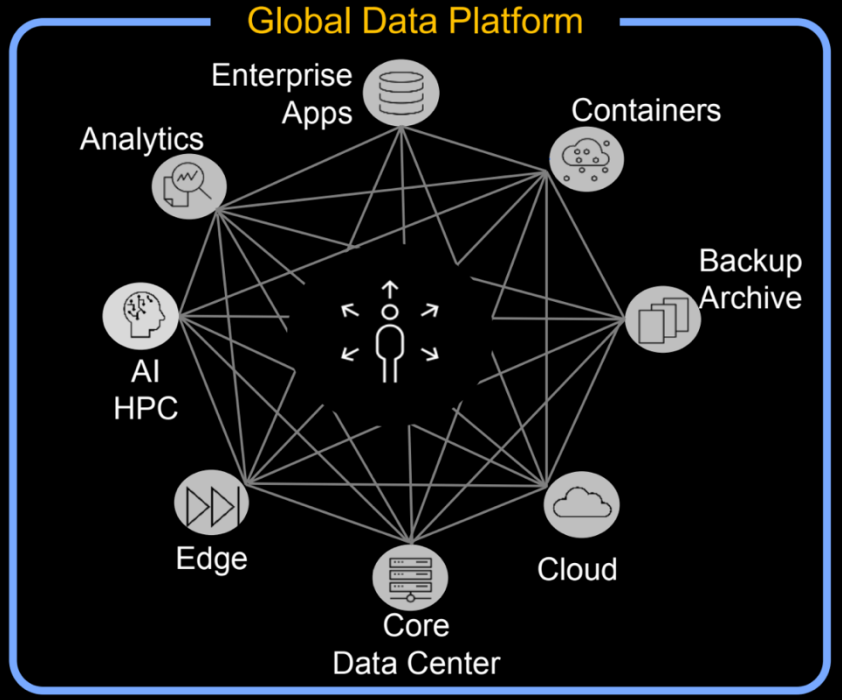
Akcelaracja dostępu do danych AI & BigData
Globalna Platforma Danych



Globalna Platforma Danych IBM

Jedno źródło dla wszystkich danych – bez ograniczeń, bez silosów





IBM Storage / © 2021 IBM Corporation

IBM Storage Scale

Wysoce skalowalna, wydajna, ujednolicona pamięć masowa dla pliku i obiektu z wbudowaną akceleracją danych

Wydajność poprzez eliminację wąskich gardeł

- Skalowalny, wieloprotokołowy równoległy system pamięci masowej
- 2.5TB/s rzeczywistej przepustowości systemu

Prostota zarządzania poprzez globalną współpracę

- System rozproszonej pamięci masowej ze wspólną przestrzenią nazw (namespace)
- Globalna Platforma Danych dla Pliku, Obiektu, AI i Bigdata

Ekonomia poprzez optymalizację kosztów

- Automatyczny rozkład danych
- Mechanizmy: Thin-provisioning, TRIM i QOS dla wszystkich danych

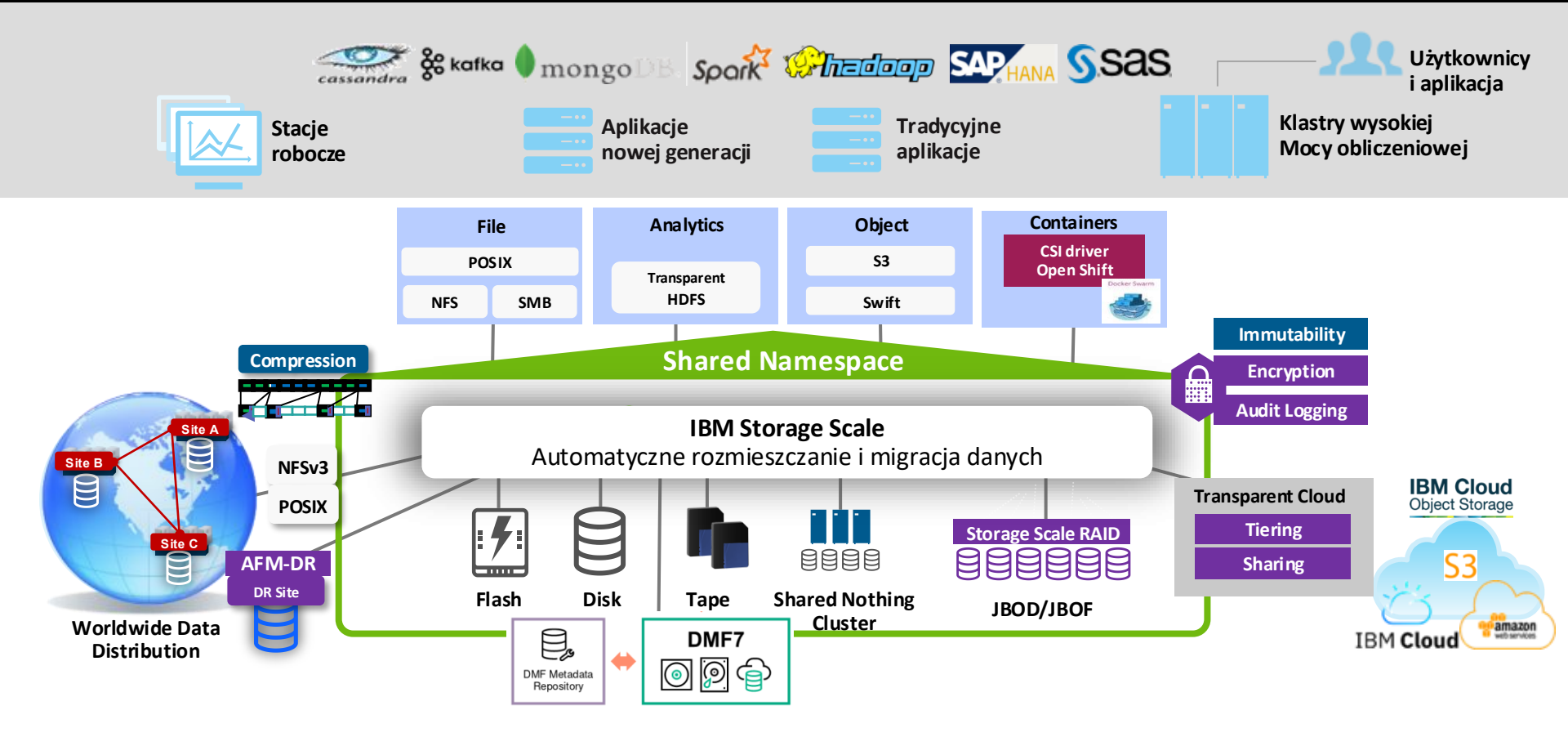
Wiarygodność danych poprzez integralność i bezpieczeństwo

- Wbudowane EC (erasure coding), replikacja, kopie migawkowe i enkrypcja
- Weryfikacja sum kontrolnych end-to-end, certyfikowane bezpieczeństwo NIST/FIPS



IBM Storage Scale

Nowa odsłona pamięci masowej na skalę globalną



IBM Storage Scale

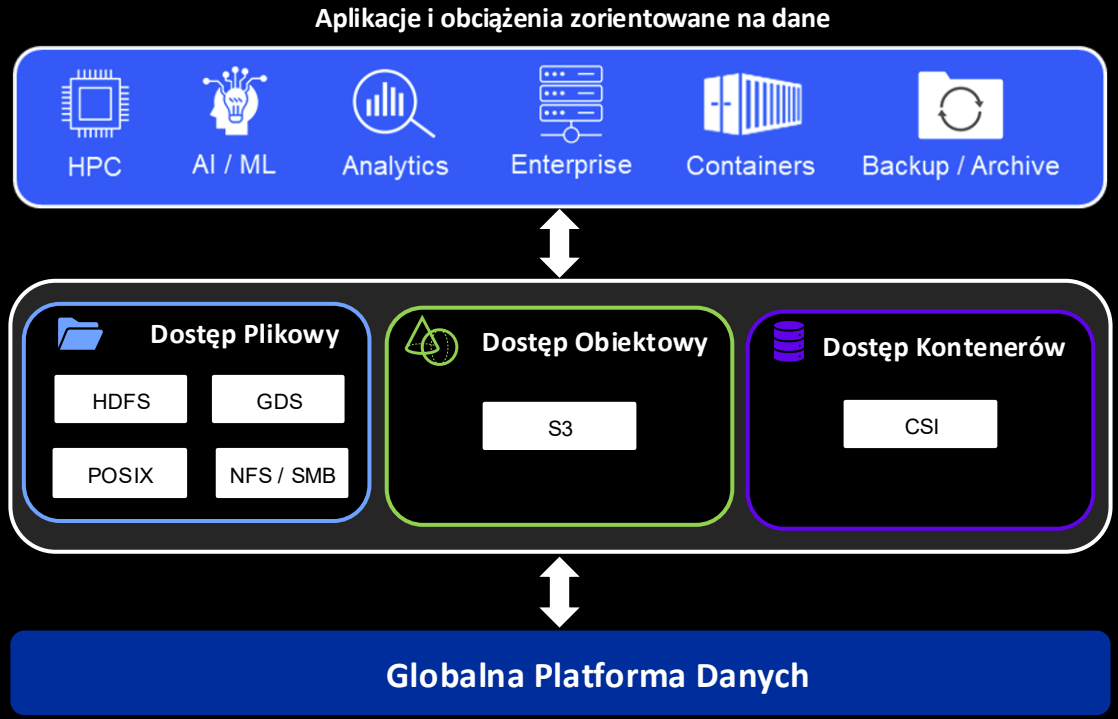
Usługi Dostępu Danych aby aplikacja definiowała interfejs użycia

Złożone środowiska wymagają prostego dostępu do danych

- Jedno źródło wspólnych danych – bez niepotrzebnych kopii, odrębnych silosów
- To aplikacja decyduje o protokole dostępu
- Wykorzystanie jednego źródła danych dla AI / Analityki, nawet z zewnętrzynych macierzy – szybko, wydajnie, bezpiecznie

Korzyści

- Znacznie większa wydajność przez zrównoleglony / wielościeżkowy dostęp do danych
- Elastyczność aplikacji poprzez wieloprotokołowy dostęp do tych samych danych
- Elastyczność zasobów dzięki wielu lokalizacjom, które mogą uzyskiwać dostęp do tych samych danych



Globalna platforma danych dla danych plik/obiekt

HPC | AI / ML | Analytics | Enterprise | Containers | Backup / Archive

1 Usługi Dostępu Danych

Big Data	Extreme Performance File	High performance object	Network attached	High Performance Containers
HDFS	GDS POSIX	S3	NFS / SMB	CSI CSNA

2 Usługi Buforowania Danych

Global Data Platform
(Storage Scale)

Local Cache | Local Cache | Local Cache | Local Cache | Local Cache

Ochrona inwestycji
File & Object Storage (3rd Party)

Macierz Obiektowa
IBM Cloud Object Storage

Macierz Plikowa
IBM Storage Scale

Kontenerycja
IBM Storage Fusion

Archiwum
LTFS / Taśmy

3 Usługi Zarządzania Danymi

4 Usługi Bezpieczeństwa Danych

- Identyfikacja
- Zabezpieczanie
- Wykrywanie
- Reakcja
- Przywracanie

IBM Storage Scale

Usługi Zarządzania Danymi dla prostoty i ekonomi przechowywania danych

Katalogowanie i orkiestracja Danych

- Katalogowanie metadanych dla Pliku i Obiektu
- Automatyczna akceleracja wyszukiwania i orkiestracji danych na wszędzie: Brzeg sieci, Rdzeń, Chmura Publiczna i Hybrydowa

Zarządzanie cyklem życia danych / informacji

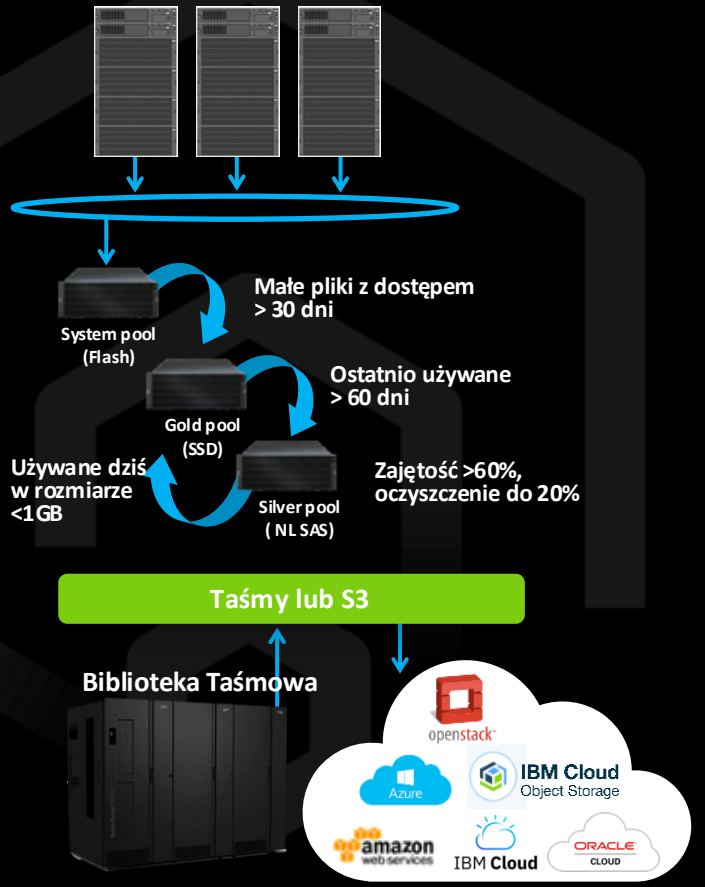
- definiowanie reguł optymalizacji czasu przechowywania plików i danych obiektowych w różnych klasach infrastruktury pamięci masowej od pamięci trwałej: NVMe, HDD, bibliotek taśmowych do chmury

Bezpieczeństwo danych

- Kopie zapasowe i odtwarzanie
- Wysoka dostępność i odtwarzanie po awarii (DR)
- Ochrona przed cyber atakami

Retencja Danych i Archiwizacja

- Mechanizmy ułatwiające retencję i archiwizację danych plikowych i obiektowych



IBM Storage Scale

Usługi Bezpieczeństwa Danych dla aktywnej ochrony wszystkich danych

IDENTYFIKACJA



- Narzędzie do oceny odporności na cyberataki,
- Ocena odpowiedzi pamięci masowej na incydenty cyberbezpieczeństwa (CIRSA)

ZABEZPIECZANIE



- Autentykacja wieloetapowa, RBAC,
- Multi-tenatowość– pełna izolacja
- Kopie bezpieczeństwa Safeguarded Copy
- Logiczna separacja danych (air gap)
- Storage Scale RAID (Erasure Coding)
- Logowanie aktywności Administratorów i Użytkowników



PRZYWRACANIE



- Odzyskiwanie wielo-petabajtowych systemów plików w ciągu kilku godzin
- CyberVault
- QRadar Incident Forensics
- Kopie zapasowe i DR

WYKRYWANIE



- File Audit Logging, Watch Folders
- Analiza kopii danych pod kątem oznak oprogramowania ransomware
- Raportowanie luk bezpieczeństwa: QRadar

REAKCJA



- Zautomatyzowana akcja po wykryciu zagrożenia
- Alerty automatycznie ustalają priorytety na podstawie wagi zagrożenia i krytyczności zaangażowanych zasobów

Czym jest multi-tenancy w IBM Storage Scale?

Wspólna infrastruktura storage, ale oddzielne klastry, przestrzenie nazw, polityki i odpowiedzialności tenantów.

TENANT A

AI / GPU cluster

TENANT B

Analytics / HPC

CENTRALNY STORAGE CLUSTER

IBM Storage Scale /
Storage Scale System 6000

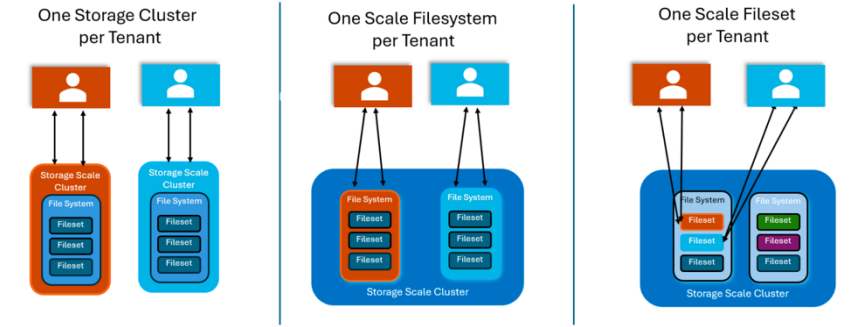
TENANT C

Department / project

TENANT D

Customer / service

Remote cluster mounts



Co zapewnia rozwiązanie?

Izolacja nie jest pojedynczą funkcją. Powstaje z połączenia namespace, szyfrowania, limitów, QoS i kontroli dostępu.

1 Izolacja danych

Oddzielny filesystem lub fileset oraz własny namespace dla każdego tenanta.

2 Szyfrowanie per tenant

Dedykowane klucze szyfrowania.

3 Kontrola pojemności

Quota ogranicza capacity i liczbę plików w modelu fileset; filesystem size w modelu filesystem.

4 Izolacja wydajności

QoS ogranicza bandwidth i IOPS, ograniczając efekt noisy neighbor.

5 Kontrola dostępu

ACL, RBAC oraz RFAC dla ograniczenia widoczności filesetów z tenant cluster.

6 Monitoring i audyt

Metryki tenantów, logi audytowe oraz integracja z Grafana, Prometheus i SIEM.

Efekt: bezpieczne współdzielenie infrastruktury z przewidywalnym zużyciem zasobów i rozdzieleniem odpowiedzialności provider–tenant.

Dwie opcje: Filesystem vs Fileset multi-tenancy

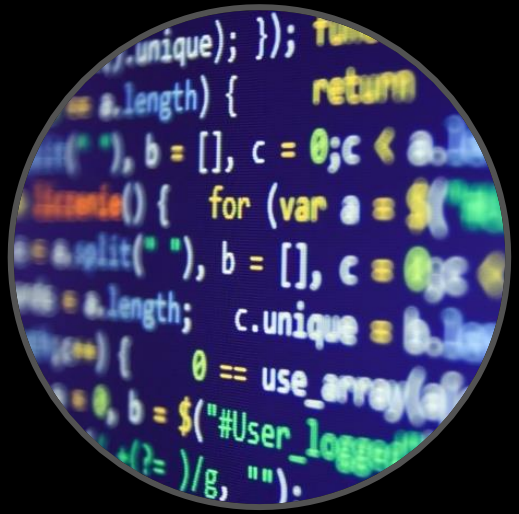
Wybór modelu jest kompromisem pomiędzy siłą izolacji a skalą, elastycznością capacity i efektywnością współdzielenia.

FILESYSTEM PER TENANT		FILESET PER TENANT	
Izolacja	Ścisła izolacja bezpieczeństwa i silniejsza izolacja awarii.	**Izolacja**	Logiczna izolacja namespace; słabsza izolacja bezpieczeństwa i awarii.
Skala	Najlepszy dla dziesiątek tenantów.	**Skala**	Setki lub tysiące tenantów.
Capacity	Dedykowany zestaw dysków / rozmiar filesystemu; zmiana pojemności wymaga zmian w filesystemie.	**Capacity**	Quota-based, szybkie i granularne przydzielanie pojemności oraz inode.
QoS	Limit na poziomie filesystemu.	**QoS**	Limit na poziomie filesetu.
Kiedy wybrać	Regulowane środowiska, klienci z wymaganiem mocnej separacji.	**Kiedy wybrać**	Departamenty, projekty i środowiska wymagające wysokiej efektywności współdzielenia.

Ważne: RFAC ogranicza widoczność filesetów, ale dokument jasno wskazuje, że nie stanowi twardej granicy bezpieczeństwa przeciw administratorowi root tenant cluster.

IBM Storage Scale

Formy wdrożenia



Oprogramowanie



Appliance
Scale Storage
System



Chmura
publiczna

IBM Storage Scale

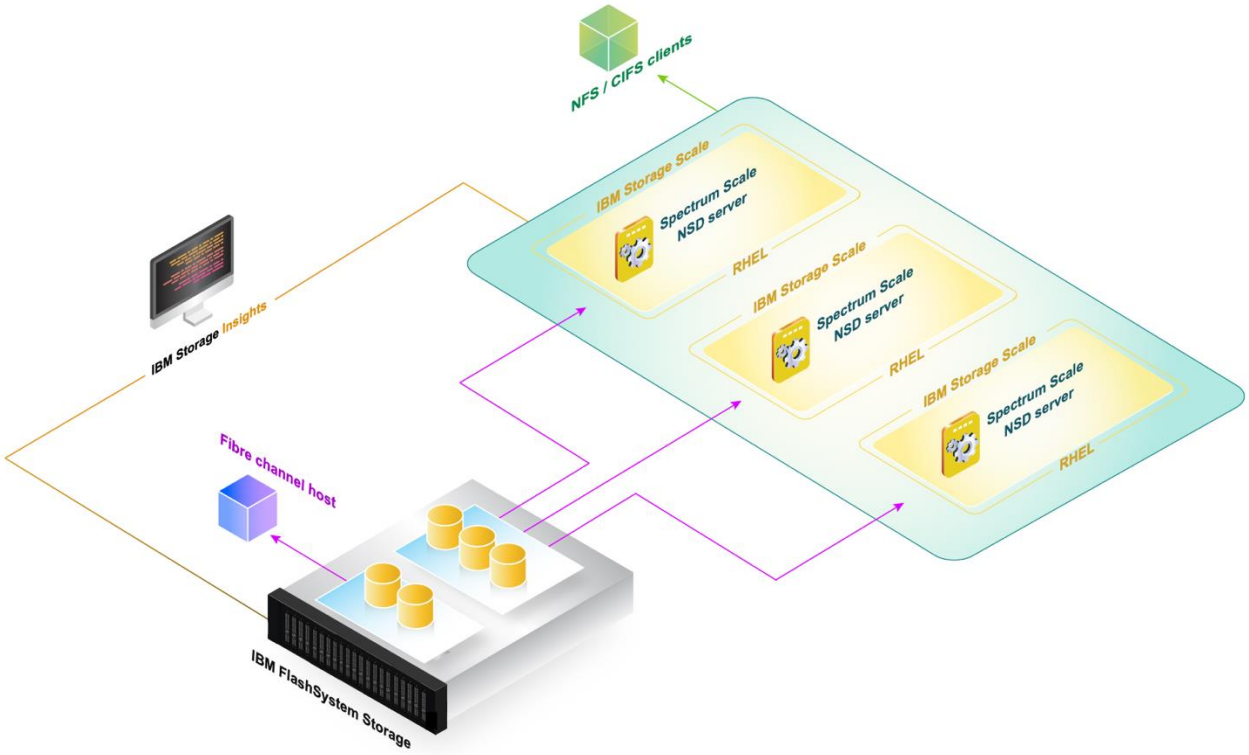


Figure 1. High level high performance file and block access solution architecture

IBM Storage Scale System

Gotowy Appliance dla wydajności i/lub pojemności

Optymalny element składowy wydajnej, skalowalnej i niezawodnej pamięci masowej Storage Scale dla przedsiębiorstw

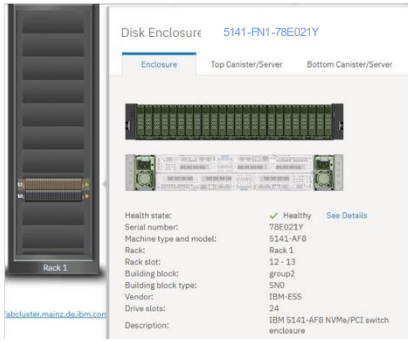
- Szybszy dostęp do danych dzięki możliwości skalowania (scale-up, scale-out)
- Łatwe do wdrożenia klastry z ujednoliconym GUI
- Uproszczone zarządzanie globalną platformą danych

Jedno rozwiązanie dla wszystkich danych

- Pojedyncze repozytorium danych z ujednoliconą obsługą plików i obiektów
- Dostęp do danych z dowolnego miejsca, dzięki obsłudze wielu protokołów przy użyciu: NFS 3/4.0/4.1, SMB, Object i HDFS
- Idealne dla SI / Analityki, Big Data, baz Hadoop i aktywnej archiwizacji

Gotowy na krytyczne dane biznesowe

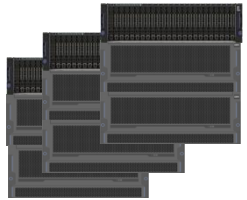
- HA / DR z wbudowaną synchroniczną i asynchroniczną replikacją
- Większe bezpieczeństwo i szybszy czas odbudowy danych po awarii dzięki Storage Scale RAID (erasure coding)
- 5x 9 (99.999%) dostępności systemu, bezprzerwowa rozbudowa i aktualizacje



Uproszczone i ujednolicone GUI



ISS 3500

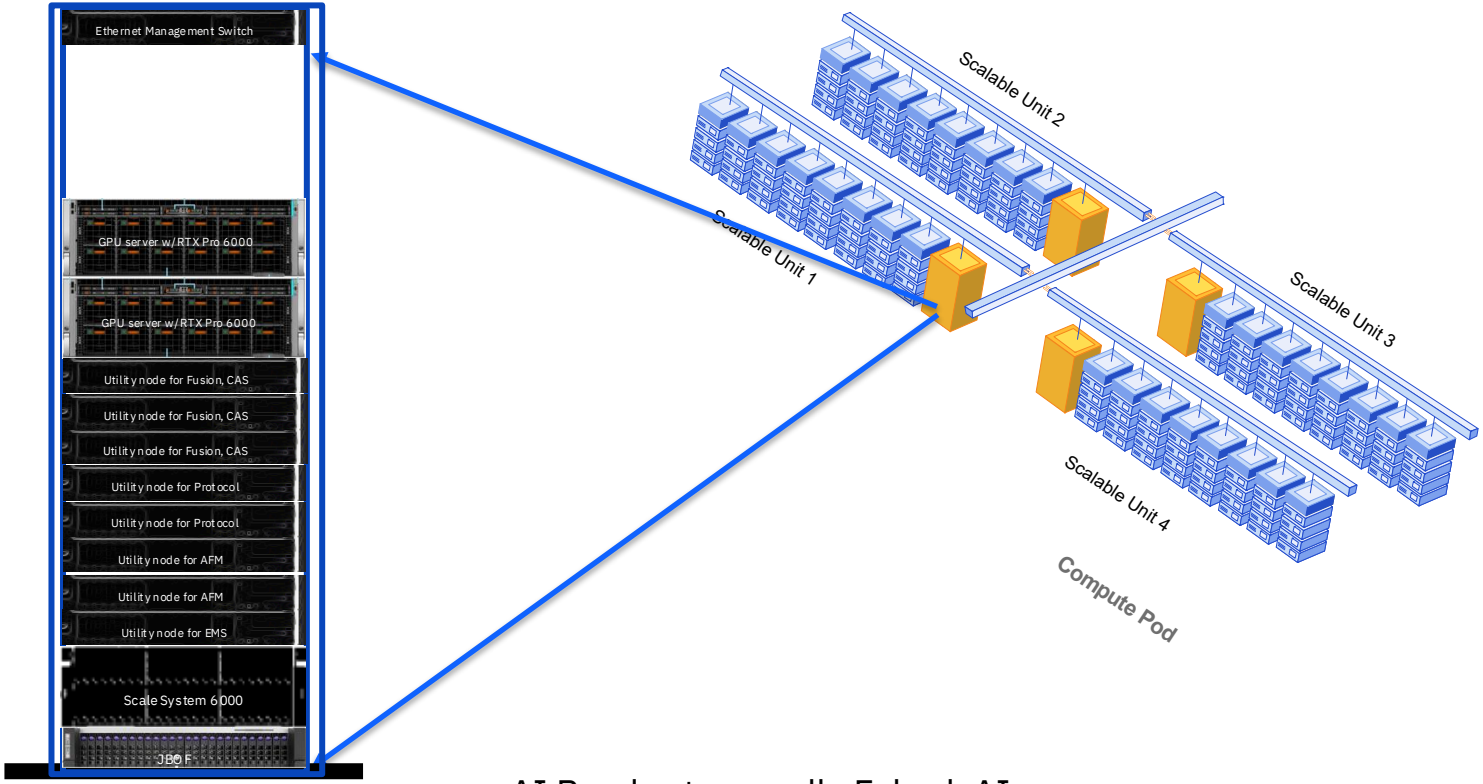


ISS 3500 cluster



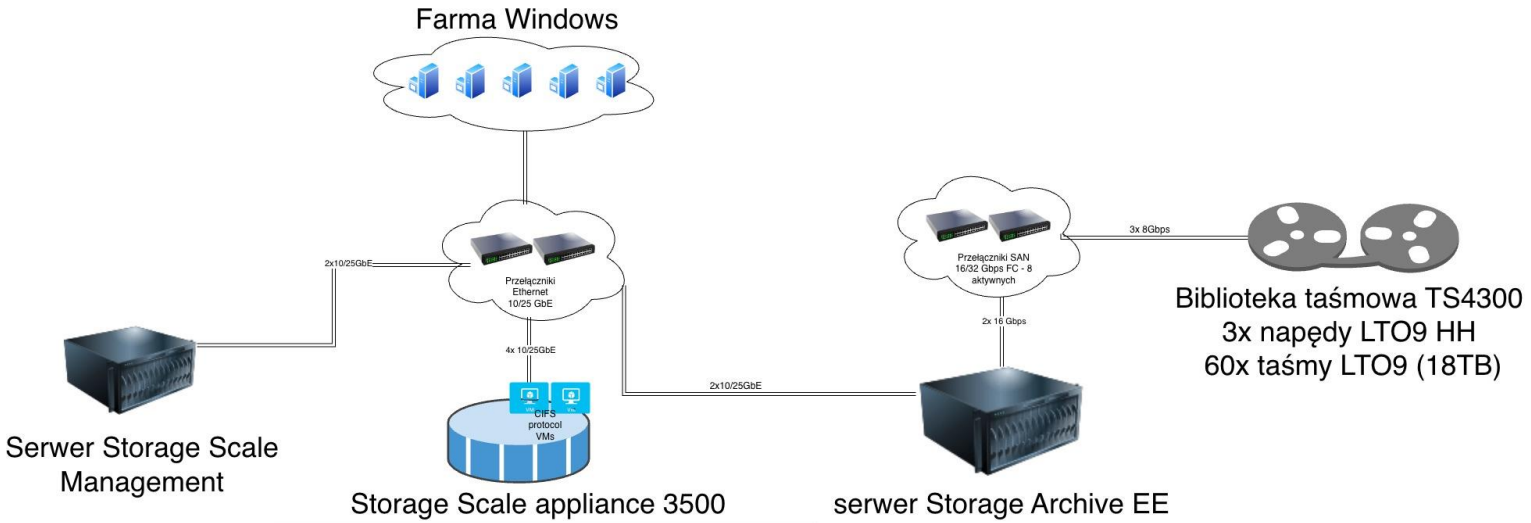
ISS 6000

IBM Storage Scale System 6000 AIDP



AI Ready storage dla Fabryk AI

Use case – platforma plikowa – aktywne archiwum



Enclosures		Drives			Usable Capacity	
Type	Quantity	Type	Quantity	RAID Type	System	Total
HDD	1	10 TB	102	8 + 2P	760.64 TB	760.64 TB
					760.64 TB	760.64 TB

Summary

ESS 3500 C1 (Capacity Model - 1 Enclosure)

Drives
102x HDD 10 TB

Total Usable Capacity 760.64 TB

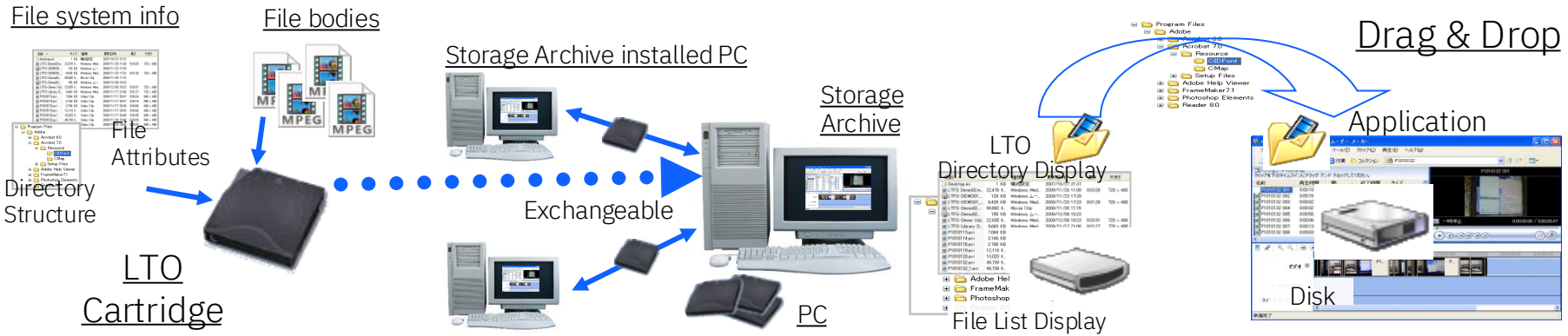
Total Throughput
100% Read Output HDD 9.42 GB/s
100% Write Output HDD 6.14 GB/s

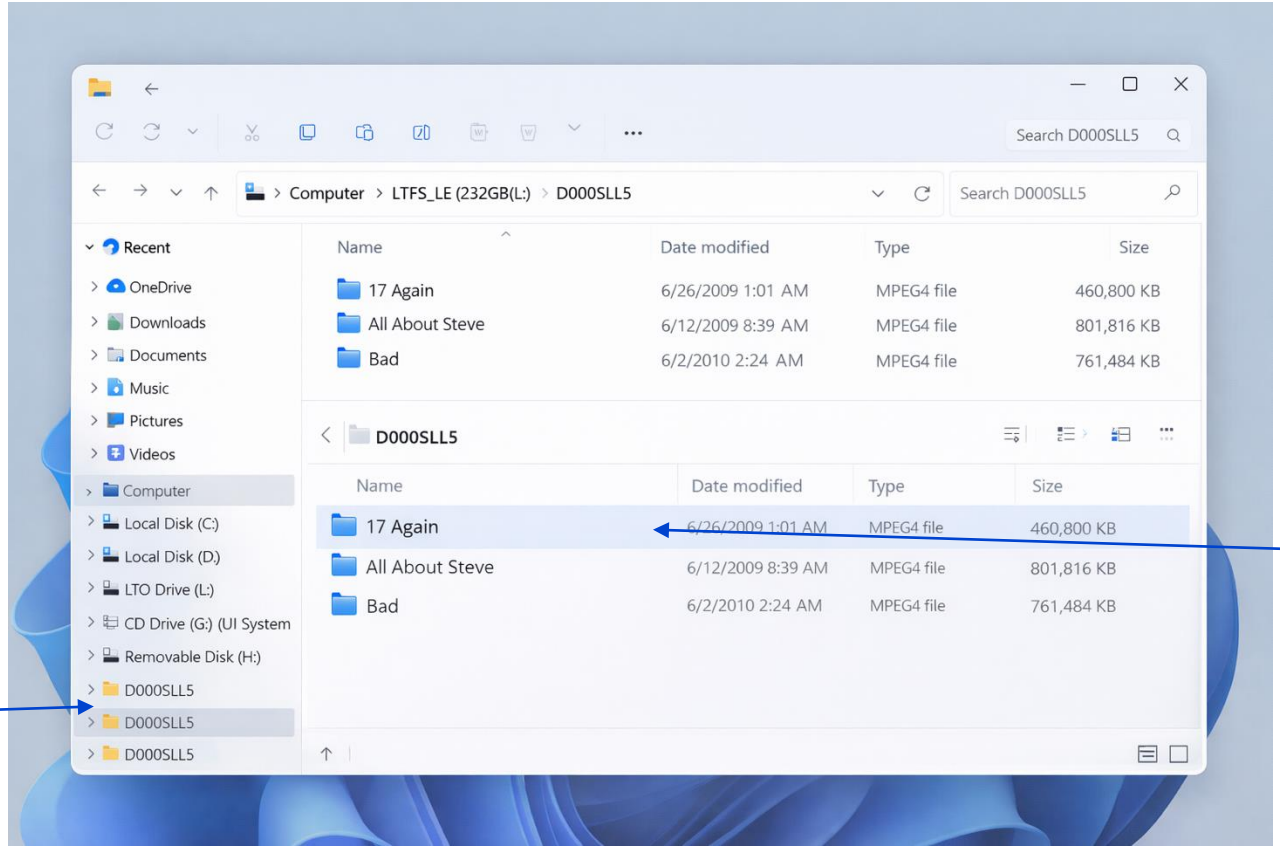
IBM Storage Archive

- File system dostępny z napedami LTO oraz IBM Enterprise
- IBM Enterprise Tape Drive TS1140 and <
- LTO 5 and <

- Dostęp plikowy na taśmie tak jakby to był dysk lub usb
- Dostęp w przeglądarce lub drag & drop na pulpicie.
  - File Open, Write, Read, Append, Delete i Close z aplikacji

- Szybszy oraz bardziej intuicyjny dostęp do danych



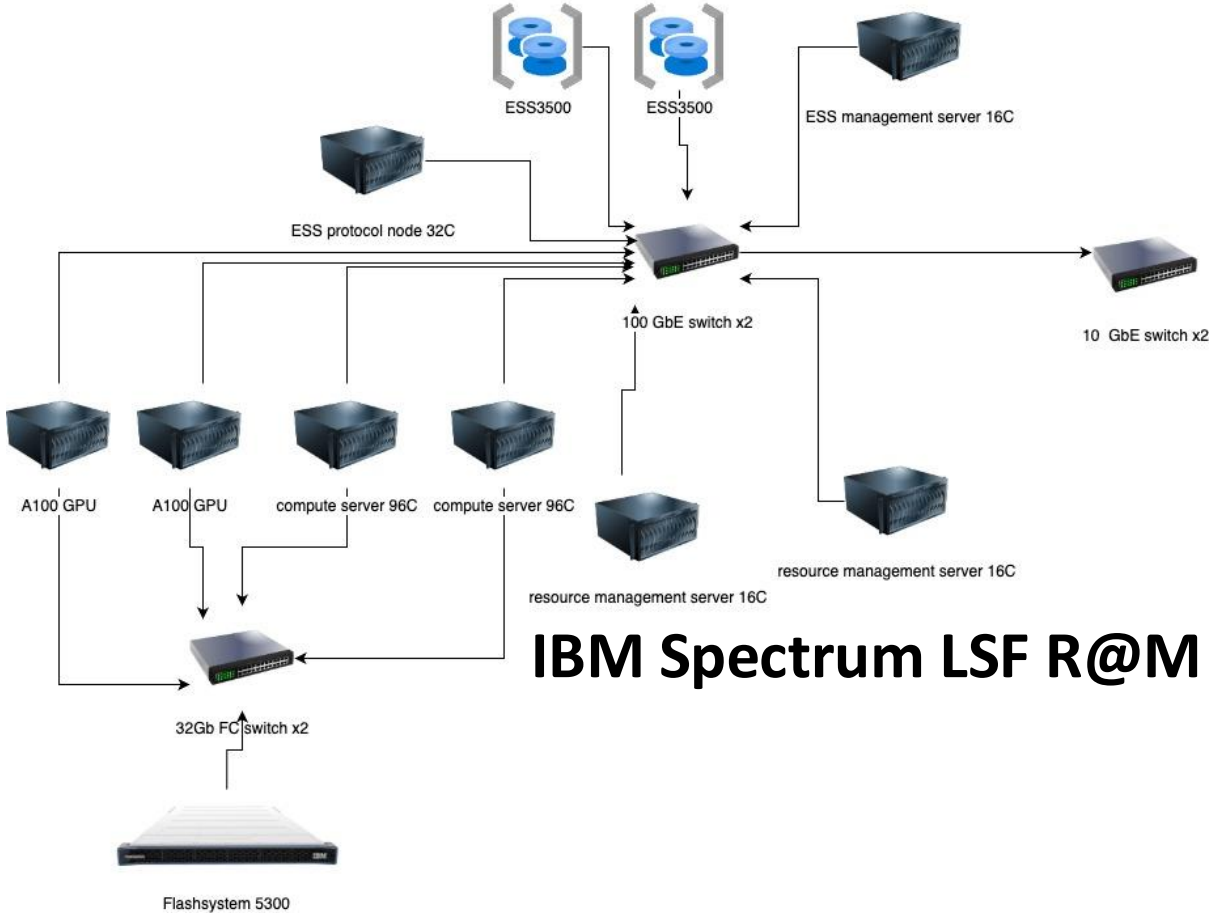


Tape in the library appear as directories in explorer

Files on tape are directly accessible from explorer

LTFS LE manages the library interface

Use case – analiza danych medycznych



IBM Spectrum LSF R@M by Matic S.A.

Jak to działa — architektura

Trzy źródła danych, jeden serwer aplikacyjny (członek klastra GPFS), przeglądarka. Bez agentów instalowanych na stałe i bez kluczy root do ESS.

Host LSF · lsfm

- lsb.acct — zadania zakończone
- bjobs / bhosts / lsload / lshosts
- bmod / bswitch / bkill (zadania PEND)
- hook POST_EXEC: dane zadania + pula
- agent wysyłany przez SSH przy zapytaniu

IBM Spectrum Scale / ESS

- REST API GUI — konto tylko do odczytu
- mm* tylko do odczytu lokalnie na lsfa
- mmapplypolicy: migracje i skan pul
- migawki i wyniki na /gpfs/fs01/lsfshared
- opcjonalnie: klucz z wymuszoną komendą

Serwer aplikacyjny · lsfa

- serwer WWW w Pythonie, paczka offline
- SSH do lsfm jako zalogowany użytkownik
- raporty HTML / Excel, cennik, dzienniki
- harmonogram: migawka 04:00, skan pul
- hasło API zaszyfrowane (Fernet)

Przeglądarka

- logowanie kontem z serwera LSF
- 6 zakładek, bez wtyczek
- Excel jednym przyciskiem
- działa w sieci wewnętrznej
- uprawnienia administratora z config.json

Hasła użytkowników nie są zapisywane — służą tylko do sesji SSH. Każdy widzi i zmienia w LSF to, na co pozwala mu jego własne konto; cennik, ustawienia i migracje — tylko administratorzy.

Co rozliczamy — każde zadanie w sześciu wymiarach

Wszystko z danych, które LSF już zapisuje, plus pomiar danych zadania na dysku (z rozpoznaniem puli) i opcjonalnie mocy.

CPU	czas CPU procesów zadania; rdzeniogodziny = sloty × czas trwania; efektywność = czas CPU / rdzeniogodziny	**rdzeniogodziny × stawka**
GPU	liczba przydzielonych kart GPU × czas trwania = GPU-godziny	**GPU-godziny × stawka**
RAM	maksymalne RSS zadania × czas trwania = GB-godziny	**GB-godziny × stawka**
Dane NVMe	dane pozostawione przez zadanie w puli gold (hook: mmlsattr największego pliku → pula); GB-dni = GB × dni retencji	**GB-dni / 30 × stawka NVMe**
Dane SAS	dane w puli silver (HDD) — ta sama miara, inna cena; bez hooka: klasa wg ścieżki (pool_by_path) lub domyślna	**GB-dni / 30 × stawka SAS + GB × stawka jednorazowa**
Energia	całka mocy węzła z BMC co 60 s, dzielona między zadania wg slotów; bez pomiaru — model zastępczy	**kWh × PUE × cena kWh**

Rozliczenia cykliczne i pomiary na hoście LSF

Skrypty uruchamiane z crona oraz hook mierzący dane każdego zadania — działają także bez portalu.

`rozliczenie_tydzien.py / _miesiac.py`

Excel z żywymi formułami: Parametry, Zadania, Projekty, Użytkownicy, Kolejki, Dziennie, Hosty; kolumny „Pula danych”, „Dane NVMe / SAS [GB]” — zmiana ceny przelicza cały skoroszyt.

`raport_tydzien.py`

Raport HTML z wykresami i kafelkami per pula do wysłania mailem; bez openpyxl — zestaw CSV, rozliczenie nigdy nie zostaje bez wyniku.

`zadanie_dane_hook.sh`

POST_EXEC w kolejkach: pliki zmienione w trakcie zadania, realne ścieżki, pula największego pliku (mmlsattr), dziennik, duplikaty liczone raz, zawsze kod 0.

`rozliczenie_magazyn.py`

Rozliczenie przestrzeni użytkowników za okres: GB-dni osobno dla NVMe i SAS ze skanów pul i próbek zajętości; dwie ceny.

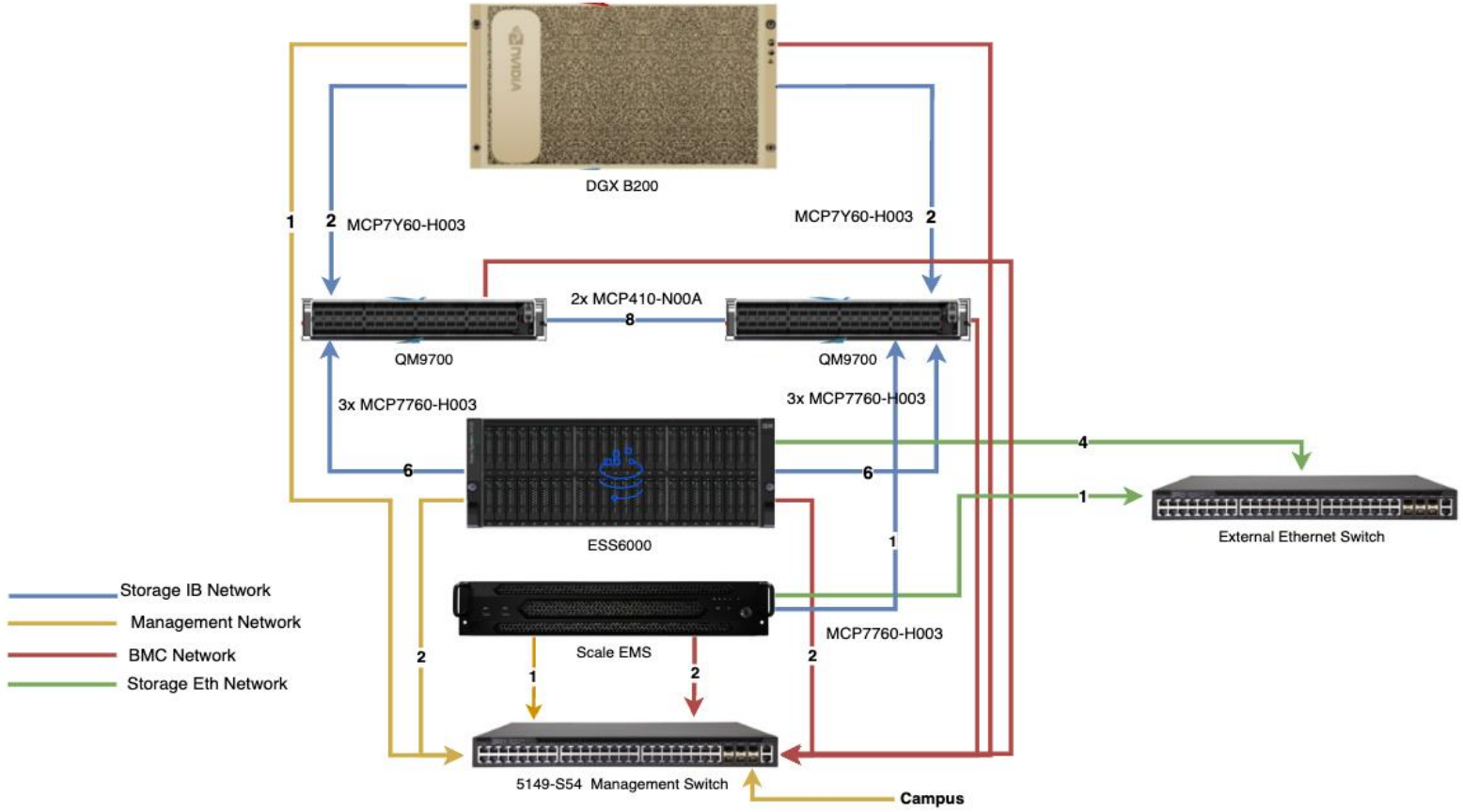
`przypisz_dane_zadaniom.py`

Przypisanie wsteczne danych do zadań po czasie modyfikacji, z pulą pliku — dla okresu sprzed uruchomienia hooka.

`power_collector.py / storage_collector.py / inwentarz_wezla.py`

Moc węzłów z BMC i próbki lsload co 60 s; próbki zajętości katalogów; konfiguracja sprzętowa węzłów (model, SN, CPU, RAM, GPU).

Use case – Nvidia DX B200 - GPU direct



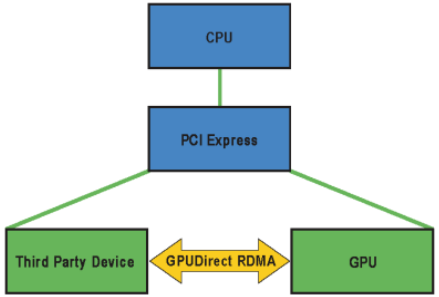
GPU direct RDMA do komunikacji z NVIDIA DGX B200

IBM SSS6000 GPUDirect Storage



GPUDirect RDMA to technologia w NVIDIA GPU, która umożliwia bezpośrednią wymianę danych między procesorami graficznymi a urządzeniem zewnętrznym za pomocą interfejsu PCI Express.

https://docs.nvidia.com/cuda/gpudirect-rdma/



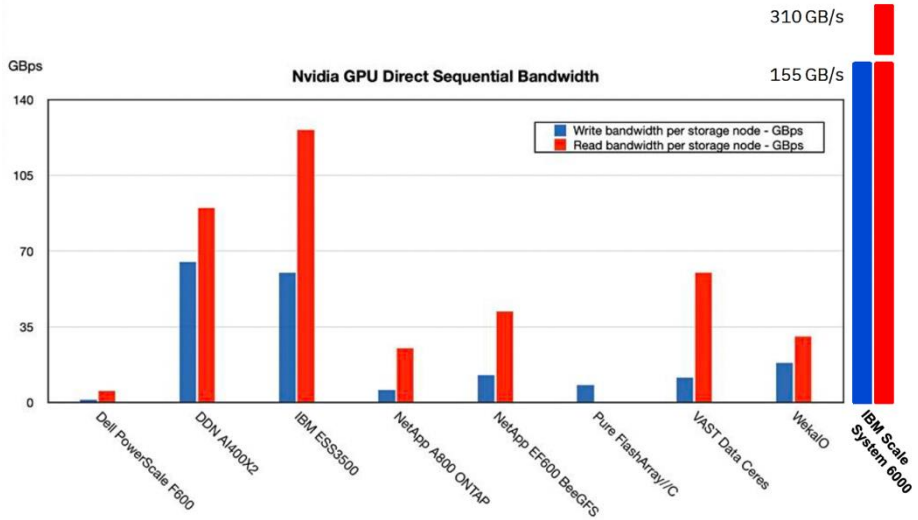
A100 / H100 / GH200



Do 3.5x większa przepustowość; 50% mniejsze opóźnienia;

IBM Storage Scale: rozwiązania GPU AI & NVIDIA

Pełne spektrum skalowalnych rozwiązań AI



IBM Scale System 6000

to do 5 razy lepsza wydajność niż jakiekolwiek inne rozwiązanie pamięci masowej z certyfikatem GDS i SuperPOD

Źródło: https://blocksandfiles.com/2023/08/15/ibm-nvidia-gpu-data-delivery/



NVIDIA and IBM Storage

IBM Spectrum Scale Guide

DGX SuperPOD & IBM Guide

NVIDIA evaluations show that the IBM ESS 3200 storage systems with Spectrum Scale file system meets the DGX SuperPOD performance and functionality requirements. It is a great choice to pair with a DGX SuperPOD to meet current and future storage needs.

As storage requirements grow, ESS 3200 building blocks can be added to seamlessly scale capacity, performance, and capability. The combination of NVME hardware and IBM Spectrum Scale parallel file system architecture provides excellent random read performance, often just as fast as sequential read patterns. Testing has validated that each ESS3200 can deliver the highest levels of per node performance and meet all of our application performance requirements. The IBM Spectrum Scale parallel file system provides a platform that is fully supported with the NVIDIA DGX SuperPOD.

https://www.nvidia.com/en-us/data-center/resources/ibm-spectrumai-reference-architecture/
https://www.ibm.com/downloads/cas/DOPRJPVX

IBM Storage Scale
edycje i licencjonowanie

Edycje oprogramowania IBM Storage Scale:

- Data Access Edition (DAE)
- Data Management Edition (DME)
- Free Developer Edition (DE) – do 12TB danych za darmo
- Erasure Code edition (ECE) - z myślą o środowiskach hyper-scale

Licencja per pojemność, z myślą o prostocie:

- Licencja per 1TB przestrzeni, dla łatwości planowania i rozbudowy
- Nielimitowana ilość klientów IBM Storage Scale jak i serwerów dostępowych

Feature	Data Access or Standard Edition	Data Management, or Developer Edition	Erasure Code Edition
Multi-protocol scalable file service with simultaneous access to a common set of data	Yes	Yes	Yes
Facilitate data access with a global namespace, massively scalable file system, quotas and snapshots, data integrity and availability and filesets	Yes	Yes	Yes
Simplify management with GUI	Yes	Yes	Yes
Improved efficiency with QoS and compression	Yes	Yes	Yes
Create optimized tiered storage pools based on performance, locality, or cost	Yes	Yes	Yes
Simplify data management with Information Lifecycle Management (ILM) tools that include policy-based data placement and migration	Yes	Yes	Yes
Enable worldwide data access using AFM asynchronous replication	Yes	Yes	Yes
Immutability (WORM / Write Once Read Many)	Yes	Yes	Yes
Asynchronous multi-site Disaster Recovery		Yes	Yes
Hybrid cloud (Transparent Cloud Tiering)		Yes	Yes
Protect data with native software encryption and secure erase, NIST compliant and FIPS certified		Yes	Yes
File audit logging		Yes	Yes
Watch folder		Yes	Yes
Erasure coding	ESS only	ESS only	Yes

A Global Data Platform: Anyone. Anywhere. Everywhere.

"We did testing with AI and machine learning workloads with Storage Scale and [it] outperformed all other file systems that we tested against."[1]

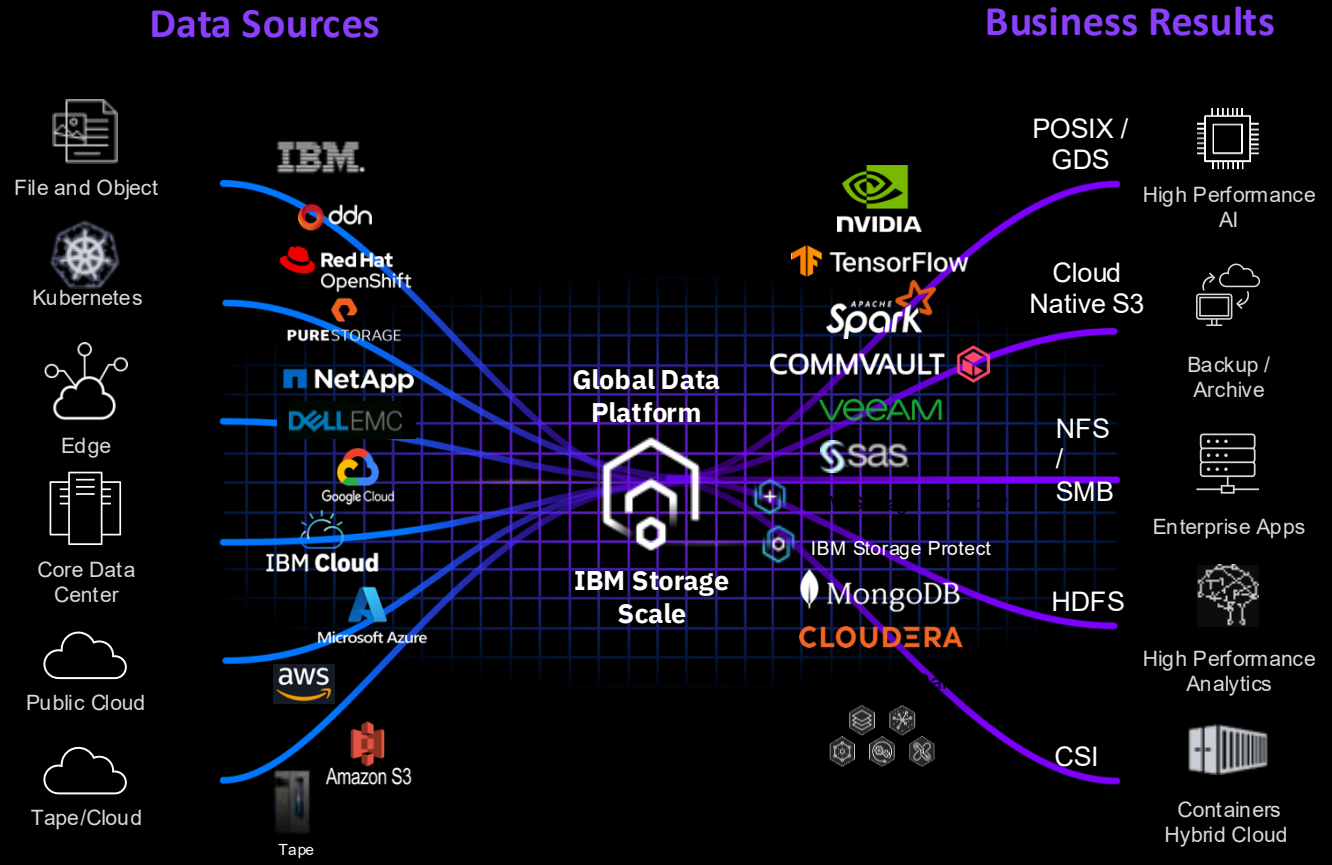
Senior system administrator

Return on investment

380%[1]

"We've managed to unify the entirety of the data storage where we're putting everything into one holistic platform."[1]

CTO of research and advanced computing




IBM

Rodzina rozwiązań IBM Storage

Budowa współczesnych środowisk Hybrid-Cloud | Multi-Cloud

