IBM Storage

IBM Content Aware Storage

Z tym to się je



Maryia Shulhach
Infrastructure Technical Leader Poland & Baltics

Wyzwanie ~~Problem~~: AI nie zna danych Twojej firmy

Dane nieustrukturyzowane, RAG i wektory w prostym języku



Modele AI znają internet, ale nie Twoją firmę

<1%

danych firmowych trafiło do treningu dużych modeli językowych

PDF · e-mail · prezentacje · audio · wideo



Niedokładne odpowiedzi

Niepełne i nieaktualne dane



Wysokie koszty

Wielokrotne kopiowanie danych



Ryzyko bezpieczeństwa

Więcej kopii, większa ekspozycja

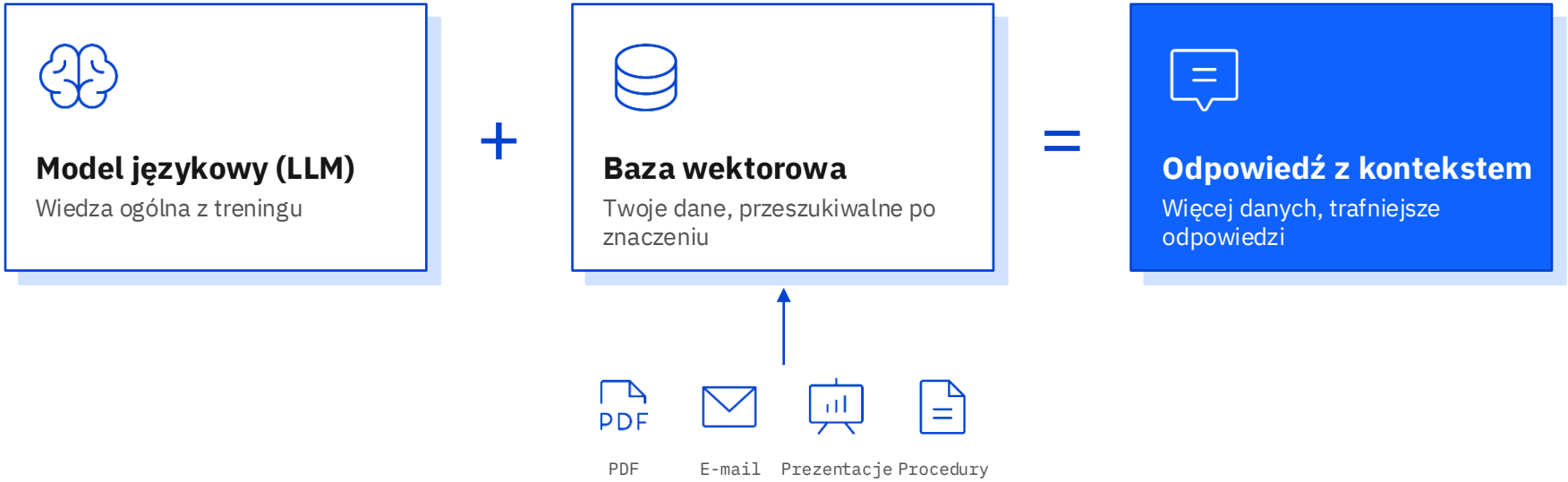


Złożoność operacyjna

Infrastruktura GPU wymaga ekspertyzy

RAG: model odpowiada na podstawie Twoich dokumentów

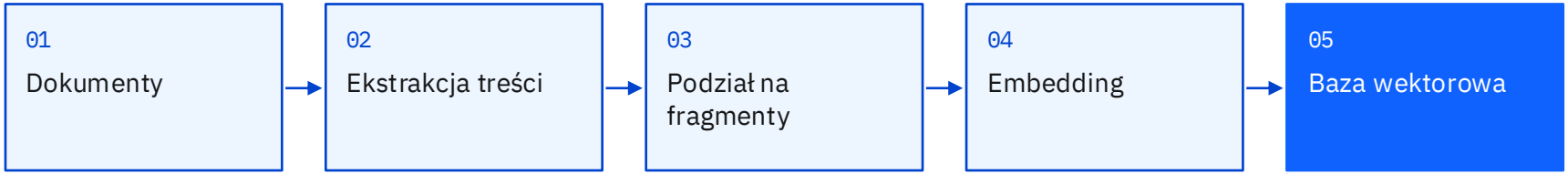
Retrieval Augmented Generation = wyszukaj → dołącz do pytania → wygeneruj odpowiedź



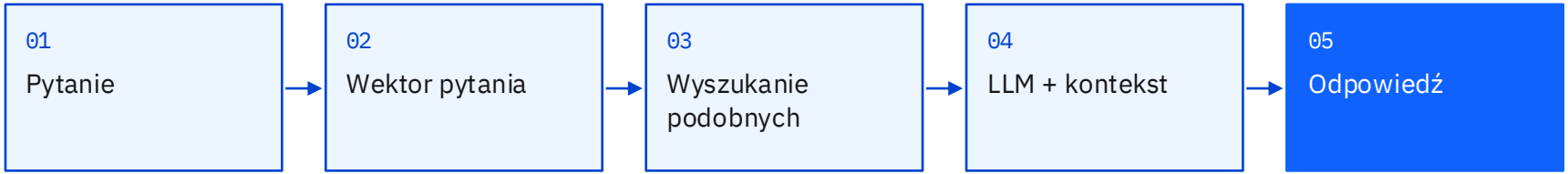
Model się nie zmienia – zmienia się to, co dostaje razem z pytaniem.

RAG: model dostaje właściwy kontekst w chwili pytania

A · Przygotowanie danych

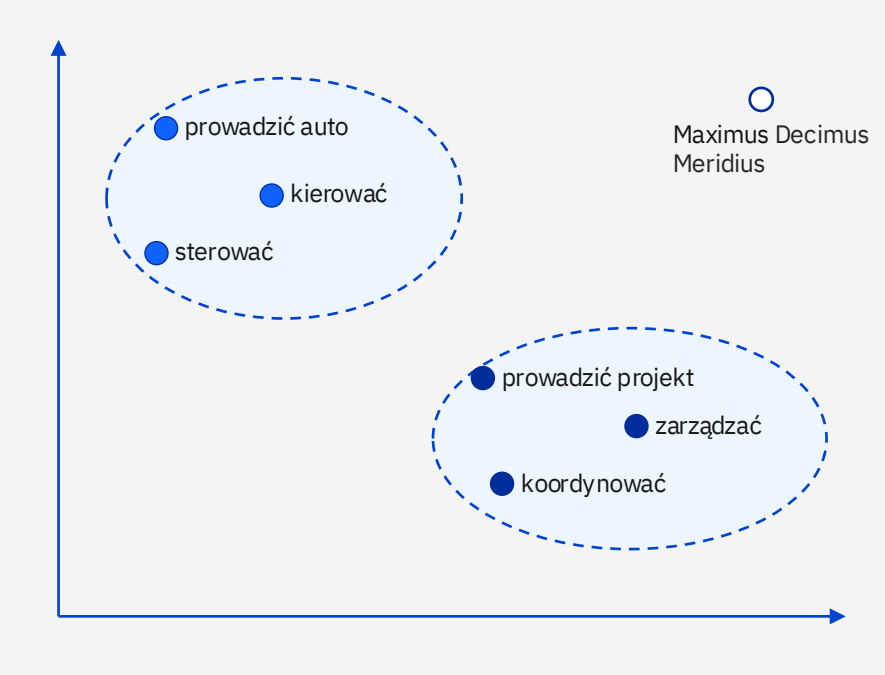


B · Zapytanie



IBM Storage | © 2026 IBM Corporation

Wektor to współrzędne znaczenia, nie słowo kluczowe



Ilustracja poglądowa (2D zamiast setek wymiarów)

Podobne znaczenia leżą blisko siebie. Na przykład: Prowadzić

Wyszukiwanie porównuje odległości wektorów, więc znajduje treści trafne znaczeniowo, nawet bez tego samego słowa.

```
embedding: tekst → [ liczba, liczba, … ]
```

Bez RAG model traci kontekst – i trafność

Bez RAG model może pomyśleć, że to…

Red Hat →



Czerwony kapelusz
dodatek do stroju ?

IBM Fusion →



Fuzja jądrowa ?

Maximo →

IBM maximo

Może literówka w Maximus Decimus Meridius ?



02

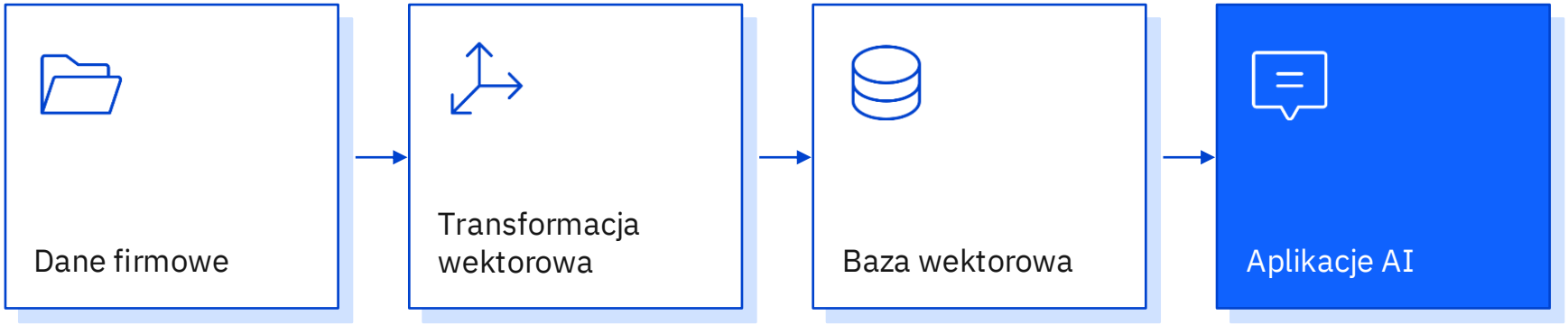
Czym jest IBM Content Aware Storage

Definicja, działanie krok po kroku, zastosowania i wartość



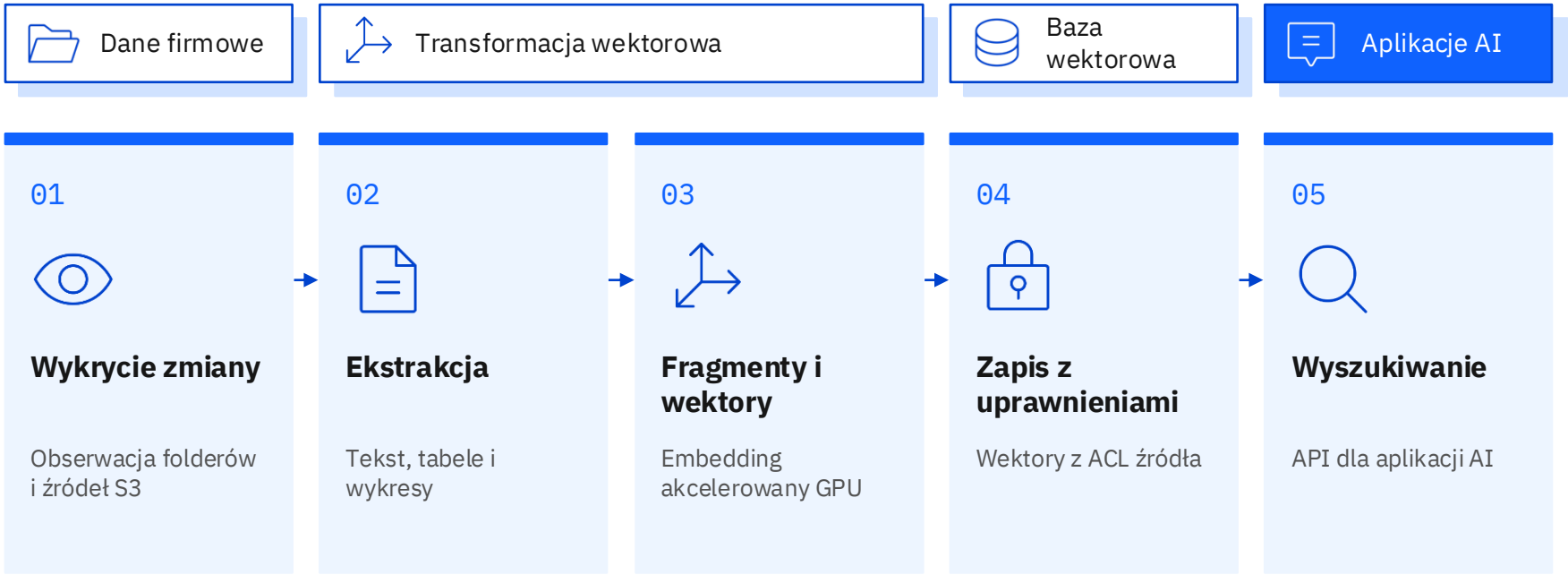
CAS czyni pamięć masową aktywnym uczestnikiem potoku AI

Baza wektorowa, repozytorium treści i narzędzia dostępu, dzięki którym aplikacje AI korzystają z danych firmowych.



IBM Fusion · IBM Storage Scale · Red Hat OpenShift · GPU NVIDIA

Od zmiany w pliku do odpowiedzi AI w pięciu krokach



Docling · NVIDIA NeMo Retriever · IBM Storage Scale

Cztery typowe zastosowania CAS



Asystenci i agenci AI

Odpowiedzi oparte na aktualnej wiedzy firmy



Bieżąca synchronizacja

Zmiana w danych aktualizuje bazę wektorową



Uproszczone potoki danych AI

Mniej przenoszenia danych i opóźnień



Wyszukiwanie semantyczne

Po znaczeniu, nie tylko po słowach kluczowych

Szybsza wiedza, niższy koszt, prostsza operacja

Wartość biznesowa



Szybszy czas do wiedzy

Przyrostowe aktualizacje bazy wektorowej



Niższe koszty

Mniej kopii danych i mniejsze zapotrzebowanie na GPU

Wartość techniczna



Wyższa wydajność

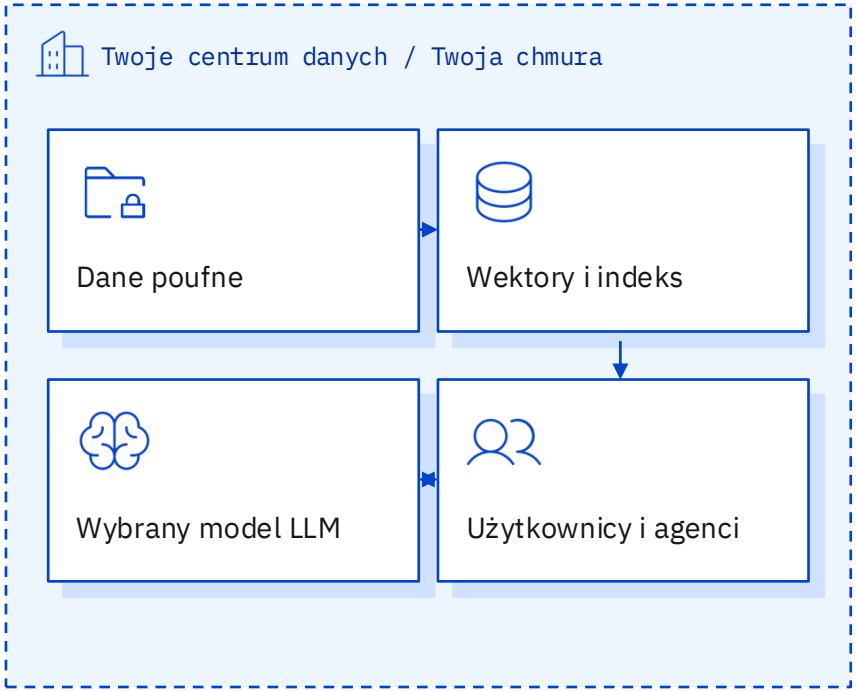
NVIDIA GPUDirect: do 8x większa przepustowość*



Prostsza architektura

Potok i baza wektorowa w systemie storage

RAG on-premises: dane poufne zostają pod Twoją kontrolą



Wiesz, gdzie są dane

Dokumenty, wektory i odpowiedzi

Niezależność

Ty wybierasz model i możesz go wymienić

Kontrola dostępu

Uprawnienia i audyt w Twojej organizacji

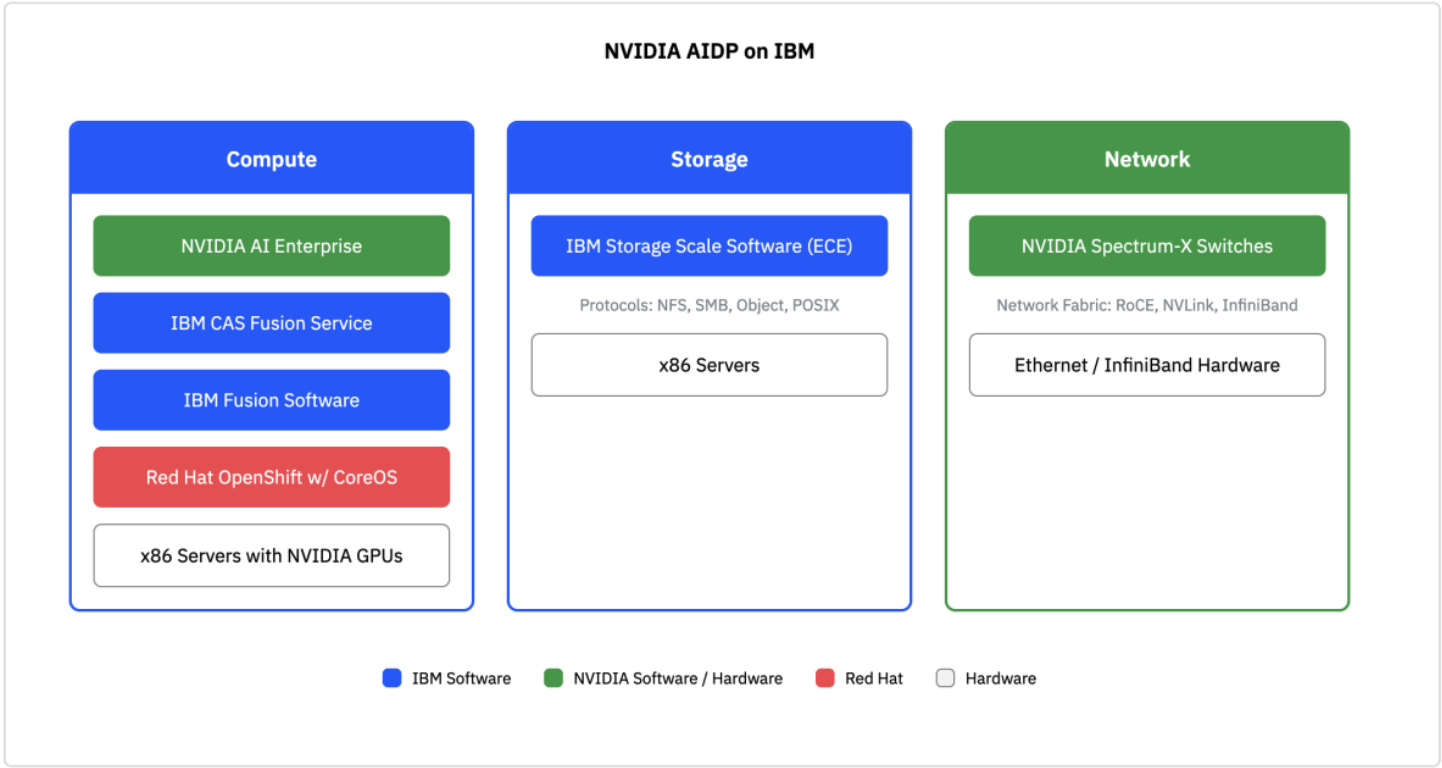
Zgodność

Łatwiej spełnić wymogi regulacyjne

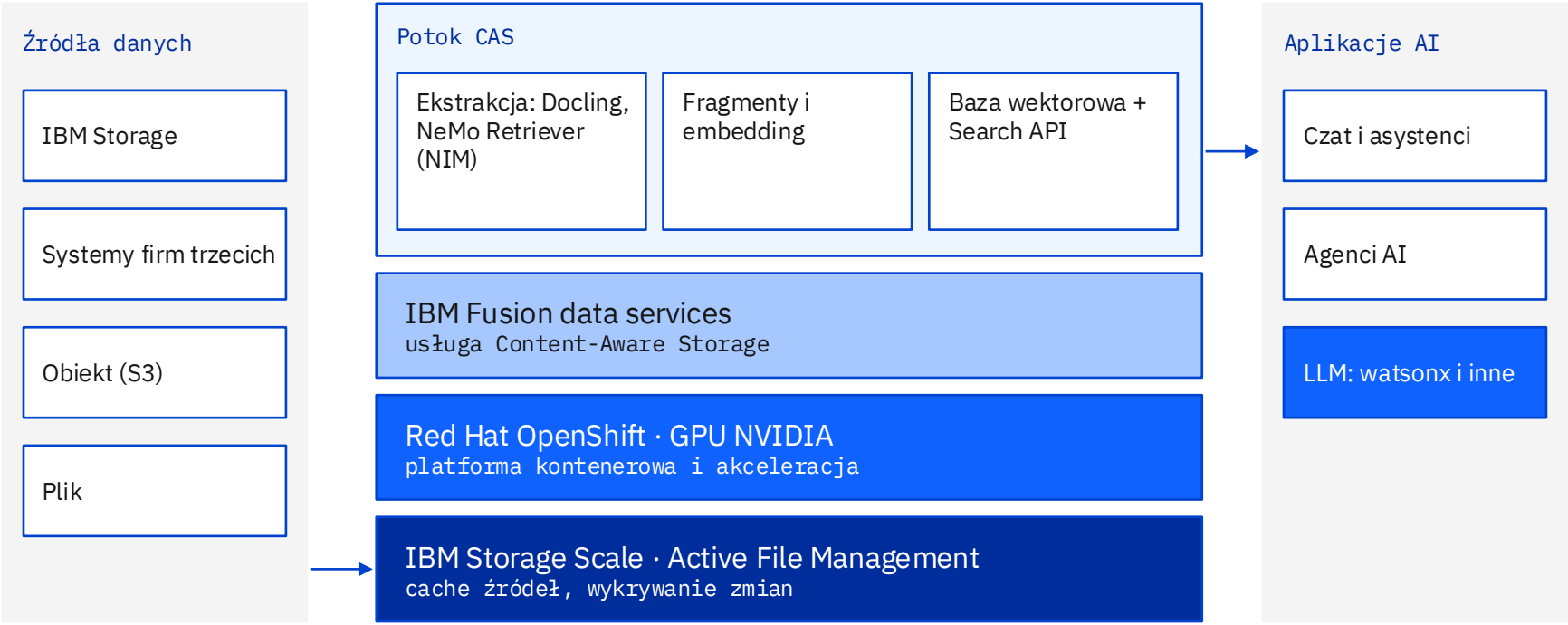
Klasyczny RAG przenosi dane do AI, CAS – AI do danych

	Klasyczny RAG	**Content Aware Storage**
Gdzie działa potok	Poza storage, osobne systemy	W warstwie storage
Aktualność	Wsadowo: dzień lub tydzień	Przyrostowo, po każdej zmianie
Kopie danych	Wielokrotne kopiowanie	Bez kopiowania i replikacji
Uprawnienia	Odtwarzane osobno	Dziedziczone ze źródła

Architektura referencyjna: stack



Architektura referencyjna: cztery warstwy wokół danych



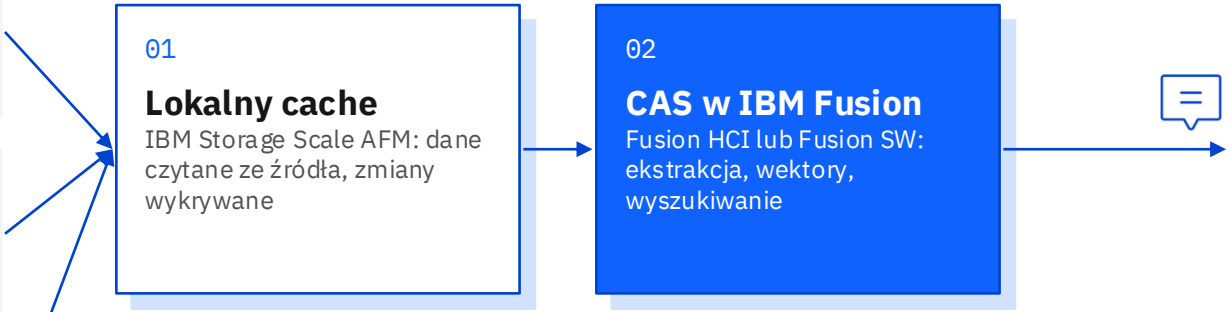
CAS pracuje z danymi tam, gdzie już leżą

Twoje obecne źródła

NAS (NFS)
także innych producentów

Obiekt S3
on-prem i chmura: AWS, IBM Cloud, S3-compatible

IBM Storage Scale
ESS / Storage Scale System



Bez migracji danych na nowy storage
Do 25 źródeł danych na instancję CAS
Zmiana w źródle → ponowne przetworzenie

Ten sam stos CAS, cztery sposoby wdrożenia

	IBM Fusion HCI	IBM Storage Scale System 6000 (AI Data Platform)	Fusion HCI + ESS 6000	IBM Fusion SW
Charakter	Appliance do trenowania, inferencji, RAG i innych zastosowań.	Rozwiązanie RAG, AI Data Platform	Appliance CAS z dodatkowym storage	Oprogramowanie: chmura lub własna infrastruktura
Kiedy	Prostota, jeden dostawca, wspólny blok dla aplikacji, wirtualizacji i AI	Istniejące GPU, skala, tiering, potrzebujemy wydajnego storage.	Duże wolumeny danych i intensywny ingest	CAS na infrastrukturze DIY lub w chmurze

Wszystkie opcje: ten sam stos oprogramowania CAS; licencje OpenShift i NVIDIA AI Enterprise dostępne w IBM



IBM Fusion HCI

- Appliance with Fusion Software
- IBM Fusion HCI
- IBM Fusion SW
- Parade Vector Database based on PostgreSQL
- OpenShift and NVAIE licensing available
- Control Nodes (3 recommended)
- Workers Nodes (1 GPU server minimum)
- 4 Network Switches
  - 2 data network switches
  - 2 admin network switches



IBM Storage Scale System 6000 Solution

- Engineered Solution
- Start with IBM Storage Scale System 6000
- IBM Fusion Software
- Parade Vector Database based on PostgreSQL
- OpenShift and NVAIE licensing available
- GPU Nodes with Accelerators (minimum of 1)
- Enterprise Management Server (EMS) Utility node
- 1 Active File Management node for all CAS functions
- 1 Ethernet management switch
- Utility nodes as needed for protocol (optional)
- Utility nodes for multi site sharing (optional)

IBM Storage Scale System 6000 jako AI Data Platform

Enterprise data
Structured i unstructured data

AI/ML workloads
Training, fine-tuning, inferencing

01

Content Aware Storage

- AI-ready data
- Zero-copy integration



IBM Storage Scale System 6000

AI Data Platform

02

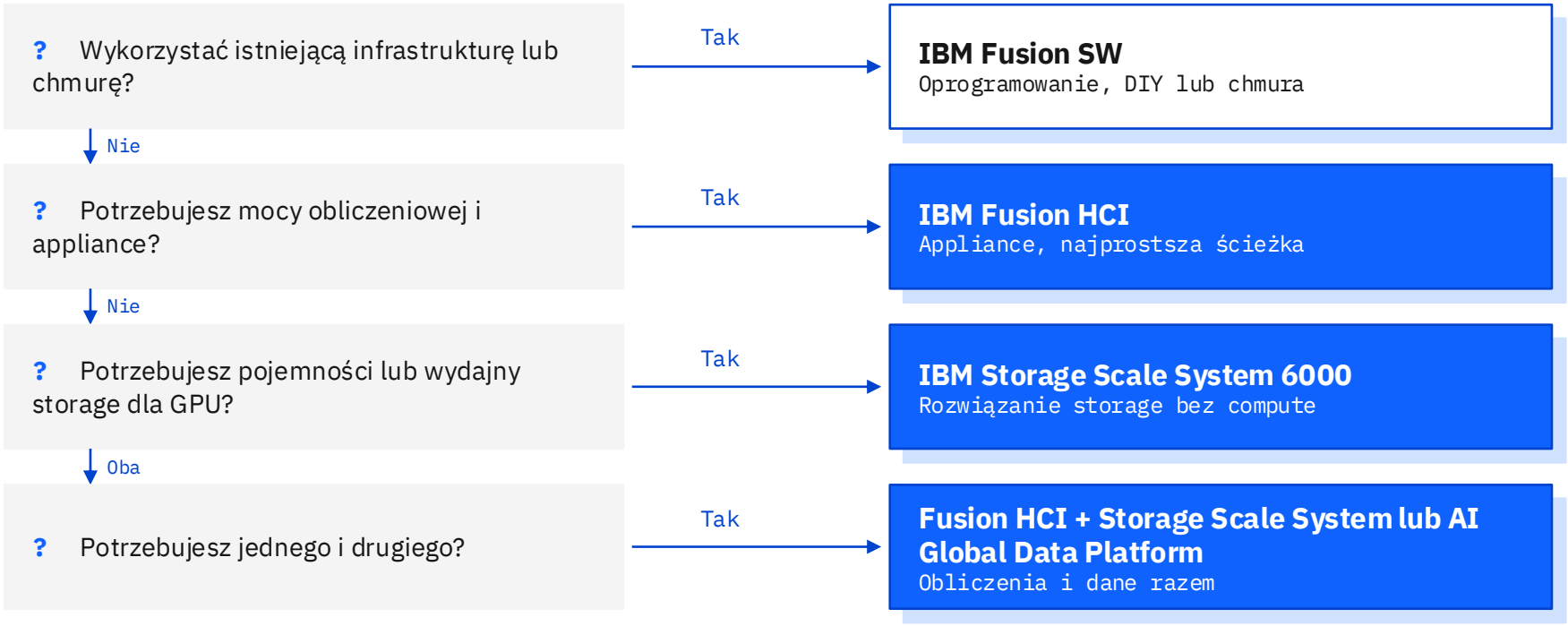
NVIDIA RTX PRO 6000

- High-performance GPU
- GPU-accelerated RAG pipeline

01 AI model training → 02 Data ingestion → 03 Embedding & preprocessing → 04 Real-time inferencing

Secure, scalable, high-performance AI data pipeline – w jednej platformie

Trzy pytania prowadzą do właściwej opcji



Masz już Fusion HCI lub SSS 6000? Najkrótsza droga to rozbudowa tego, co działa.

Podsumowanie

Trzy rzeczy do zapamiętania

1 CAS przenosi AI do danych, a nie dane do AI.

2 Odpowiedzi są aktualne i bezpieczne: przyrostowe zmiany i uprawnienia źródła.

3 Wdrożenie dopasowane do potrzeb: Fusion HCI, ESS 6000, kombinacja lub software.

Q&A

Pytania i dyskusja

Dziękuję za uwagę



Maryia Shulhach
Infrastructure Technical Leader Poland & Baltics