Bezpieczne rozwiązania AI z technologią AMD

Piotr Nogaś

2025.09.25

REGULACJE: KOSTKA I HIERARCHIA

EU Cybersecurity Regulatory Framework

REGULATION (EU) 2019/881 on Cybersecurity Act (CSA)

Cybersecurity Act strengthens the EU Agency for cybersecurity (ENISA) and establishes a cybersecurity certification framework for products and services.

Directive NIS2 (EU) 2022/2555

Sets cybersecurity measures on entities falling under critical infrastructure sectors.

DORA (Digital Operational Resilience Act) – Financial sector Regulation (EU) 2022/2554

Ensures that financial entities in EU remain resilient through a severe operational disruption.

Cyber Resilience Act (proposal)

Use of EU cybersecurity certifications and rules to ensure more secure hardware and software products.

RED - radio equipment directive 2014/53/EUEN

Establishes a framework for placing radio equipment on market and subjects certain categories of radio equipment to increased level of cybersecurity, personal data protection and privacy.

Regulation (EU) 2021/887 ECCC (Network of National Coordination Centres)

Boosts research excellence and the competitiveness of the Union in the filed of cybersecurity.

Directive (EU) 2019/1937 on Whistle-blower

Reports of violations of NIS requirements

Directive CER (EU) 2022/2557 (Resilience of critical entities)

strengthens the resilience of critical infrastructure to a range of threats, including natural hazards, terrorist attacks, insider threats, or sabotage.

Network Code on sector-specific rules for cybersecurity aspects of cross border electricity flows (NCCS)

Sets a standard for cross border flows in EU.

(non exhaustive list)

HPE | AMD

Source: https://assets.kpmg.com/content/dam/kpmg/be/pdf/2023/KPMG-Webinar-on-NIS2-2023-05-23.pdf

STATUS REGULACJI

- Dyrektywa dotycząca wspólnego poziomu bezpieczeństwa sieci i informacji na terytorium Unii (**dyrektywa NIS 2**)
  - Termin implementacji upłynął **17 października 2024 r.** **(nowelizacja UKSC)**
- Dyrektywa w sprawie odporności podmiotów krytycznych (**dyrektywa CER**)
  - Termin implementacji upłynął **17 października 2024 r.** **(nowelizacja Ustawy o Zarządzaniu Kryzysowym oraz Narodowy Program Ochrony Infrastruktury Krytyczne)**
- Rozporządzenie w sprawie operacyjnej odporności cyfrowej sektora finansowego (**DORA**)
  - Weszło w życie **17 stycznia 2025** r.
- Rozporządzenie (akt) dot. wymagań cyberbezpieczeństwa dla produktów z elementami cyfrowymi (**Cyber Resilience Act**)
  - **Tekst opublikowany 20 listopada 2024 r.** w dzienniku urzędowym UE. Wszedł w życie 20 dni po publikacji z okresem przejściowym **36 miesięcy** po wejściu w życie, przy czym **niektóre przepisy są stosowane wcześniej**.
- Rozporządzenie (akt) w sprawie sztucznej inteligencji (**AI Act**)
  - Obowiązuje od **lutego 2025 r.** , z okresem przejściowym upływającym w **sierpniu 2026 r.**
- Dyrektywa w sprawie odpowiedzialności za sztuczną inteligencję (**ALD**)
  - **Projekt opublikowany w 2022 r**.; prowadzone są prace legislacyjne.
- Ustawa Prawo Komunikacji Elektronicznej (**PKE**)
  - **Projekt opublikowany w 2024**. rozszerza zakres podmiotowy na podmioty nie świadczące usług

DYNAMICZNE WYMAGANIA W ZAKRESIE SYSTEMU ZARZĄDZANIA BEZPIECZEŃSTWEM INFORMACJI — ROZPORZĄDZENIA RM I NIE TYLKO

Szczegółowe wymagania dla system zarządzania bezpieczeństwem mogą określone w rozporządzeniu.

Art. 8a nowelizacji uKSC:

„Rada Ministrów może określić, ***w drodze rozporządzenia*** *odrębnie dla danego rodzaju działalności wykonywanej prz podmioty kluczowe lub podmioty ważne* ***szczegółowe wymagania dla systemu zarządzania bezpieczeństwem informacji, o którym mowa w art. 8*** *ust. 1, biorąc pod uwagę* ***rekomendacje międzynarodowe o charakterze specjalistycznym****, w tym rekomendacje Agencji Unii Europejskiej do spraw Cyberbezpieczeństwa, zwanej dalej „****ENISA****”, wielkość podmiotu,* ***skalę działalności wykonywanej przez te podmioty oraz potrzebę podejmowania przez te podmioty działań zapewniających cyberbezpieczeństwo****.”*

Dodatkowe środki bezpieczeństwa:

Zwracam szczególną uwagę na konieczność podjęcia działań prewencyjnych, celem zmniejszenia ryzyka przyszłych incydentów. Wśród nich zaleca się:

- Stosować dwuskładnikowe uwierzytelnianie z wykorzystaniem kluczy sprzętowych, szczególnie dla kont serwisowych i z uprawnieniami administratora.
- Stosować menedżery haseł do bezpiecznego przechowywania danych uwierzytelniających.
- Regularnie aktualizować oprogramowanie, zwłaszcza w systemach związanych z dostępem do sieci Internet.
- Stosować segmentację sieci celem podniesienia bezpieczeństwa wrażliwych elementów infrastruktury.

Implementacja ww. zaleceń jest również zasadna w przypadku podobnych incydentów w innych podmiotach, z którymi została nawiązana współpraca.

Krzysztof Gawkowski
Wiceprezes Rady Ministrów
Minister Cyfryzacji
/dokument podpisany elektronicznie/

Dochowanie należytej staranności wymaga także uwzględnienia rekomendacji.

HPE | AMD

REKOMENDACJE SEKTOROWEGO PODMIOTU NADZORUJĄCEGO

- szkolenia bezpieczeństwa i zgodności dla kadry i szeregowych użytkowników
- (dynamiczna) segmentacja, mikrosegmentacja sieci,
- Zero-Trust: Różne domeny autentykacji, granulacja i ograniczenie dostępu (w tym domeny i konta serwisowe),
- Ochrona i mitygacja skutków ransomware,
- Kontrola dostępu NAC, Secure Service Edge (SASE),
- Endpoint Detection and Response/eXtended Detection and Response (EDR/XDR),
- Szyfrowanie i zarządzanie kluczami,
- bezpieczne usługi (w tym DNS, wieloskładnikowe uwierzytelnienie, autoryzacja, TLS v1.3 itp.),
- Bezpieczeństwo łańcucha dostaw,
- Finansowanie pełnego audytu bezpieczeństwa i zgodności (nie tylko w zakresie niezbędnym do rozliczenia grantów/dofinansowania),
- SOC/SOAR – RBA/współpraca z CEZ-CSIRT.

SOC – Security Operation Center
SOAR – Security Operation Automatic Response
RBA – Risk Based Approach

DOSTOSOWANIE BEZPIECZEŃSTWA DO STOPNIA ZAGROŻEŃ

W odpowiedzi na stan alarmowy CHARLIE-CRP Centrum e-Zdrowia rekomenduje:

- Każda jednostka ochrony zdrowia powinna posiadać linię telefoniczną typu PSTN. Linia tego typu powinna być doprowadzona do centrum zarządzania, dyrektora jednostki lub innych osób koordynujących działania jednostki w sytuacjach kryzysowych. Linia musi być odporna na brak zasilania w jednostce.
- Przygotowanie procedur i zespołów odpowiedzialnych za ciągłość pracy systemów informatycznych oraz procedur dotyczących awaryjnego gaszenia
  i uruchamiania systemów informatycznych.
- Wykonanie przeglądu stosowanych polityk wykonywania kopii zapasowych obejmujących:
  - system obsługi "części białej", w którym są gromadzone dane medyczne,
  - system kadrowo-płacowy i finansowo-księgowy,
  - inne, krytyczne dla jednostki ochrony zdrowia systemy z punktu widzenia ciągłości działania.

  Ze względu na możliwość cyberataku kopie zapasowe systemów medycznych i danych medycznych muszą być wykonywane na bieżąco. Po wykonaniu kopii nośniki należy separować od sieci teleinformatycznej. Kopie muszą być wykonywane w trybie dobowym, jako kopie pełne lub przyrostowe z zachowaniem ostatniej pełnej kopii danych.
- Rekomendacje dla zabezpieczenia transmisji danych:
  - ograniczenie dostępów zdalnych za pomocą protokołu RDP, SSH - w szczególności z zewnątrz organizacji,
  - szyfrowanie ruchu protokołem SSL,
  - wykorzystywanie sieci VPN do połączeń zdalnych.

Źródła: https://ezdrowie.gov.pl/portal/artykul/rekomendacje-cyberbezpieczenstwa-w-zwiazku-z-ogloszeniem-stanu-alarmowego-charlie-crp
https://www.knf.gov.pl/knf/pl/komponenty/img/knf_125701_PTE_Wytyczne_IT_16_12_2014_40005.pdf
https://www.nik.gov.pl/aktualnosci/podrecznik-kontroli-systemow-informatycznych.html

HPE | AMD

OBECNA LINIA ORZECZNICTWA

Rozpiętość nałożonych kar oraz podsumowanie ich uzasadnienia:

Brak anonimizacji udostępnionych danych	1M €	https://www.uodo.gov.pl/decyzje/DKN.5130.2215.2020
Wyciek danych	644k €	https://uodo.gov.pl/decyzje/ZSPR.421.2.2019
Brak udokumentowanych testów	443k €	https://www.uodo.gov.pl/decyzje/DKN.5112.1.2020
Ransomware	235k €	https://uodo.gov.pl/decyzje/DKN.5130.1354.2020
Niezgodności z ustawą o ochronie zdrowia	22k €	https://www.uodo.gov.pl/decyzje/DKN.5131.32.2023

Samowolna decyzja zapłaty okupu (ang. ransom) jest podstawą do oskarżenia o finansowanie działalności przestępczej i/lub terroryzmu.

DOCHOWANIE NALEŻYTEJ STARANNOŚCI - ORZECZNICTWO

Uzasadnienia orzeczeń

... administrator nie przedstawił dowodu, że po wystąpieniu naruszenia ochrony danych osobowych dokonuje regularnego testowania.

... testowanie, mierzenie i ocenianie zastosowanych zabezpieczeń, musi być dokonywane w sposób regularny,

... nierzetelnie przeprowadzona analiza ryzyka.

... zarządzanie ryzykiem stanowi fundament działań związanych z ochroną danych . Zarządzanie ryzykiem ma charakter **ciągłego procesu** ... , ale także zapewnienie **ciągłości monitorowania** poziomu zagrożeń oraz zapewnienie **rozliczalności** w zakresie poziomu oraz **adekwatności** wprowadzonych zabezpieczeń.

Wnioski

W następstwie incydentu, przywrócenie przetwarzania/danych musi być poprzedzone dostosowaniem środków i procesów do najwyższego poziomu zagrożenia – **testy** powinny być wykonywane **przy uwzględnieniu najbardziej restrykcyjnych ustawień bezpieczeństwa**.

Testy muszą potwierdzać swoją skuteczność (KPI/SLI/SLA/SLO), monitorowanie usług przetwarzania oraz wszystkich wymaganych pomocniczych (bezpieczeństwo, dostępność, DR, ochrona danych itp) muszą być wykonywane regularnie

Adekwatności zabezpieczeń musi być **adekwatna do bieżącego poziomu zagrożeń**.

BEZPIECZEŃSTWO, PRYWATNOŚĆ I ZGODNOŚĆ ROZWIĄZAŃ AI

- **Atak: manipulacja danych wejściowych** (Input Manipulation)
  - polega na wprowadzeniu zmodyfikowanych danych wejściowych w celu oszukania modelu uczenia maszynowego - prowadząc do manipulacji wyników.
  - Obrona obejmuje trenowanie modeli odpornych przez walidację danych wejściowych oraz monitoring anomalii. Atak może obejmować m.in. manipulację obrazami lub ruchem sieciowym w celu uniknięcia wykrywania włamań.
- **Atak: zatrucie danych** (Data Poisoning)
  - polega na manipulacji danymi treningowymi, w celu błędnego działania modelu uczenia maszynowego.
  - Ochrona wymaga walidacji danych, bezpiecznego przechowywania oraz kontroli dostępu do danych, monitoringu anomalii, audyty, oraz ciągła walidacja modeli.
  - Przykładem ataku: wprowadzenie błędnych danych do klasyfikatora spamu.
- **Atak: inwersja modelu** (Model Inversion)
  - Polega na odwróceniu działania modelu przez atakującego, aby wydobyć z niego poufne informacje.
  - Przykład: kradzież tożsamości z modelu rozpoznawania twarzy poprzez analizę jego odpowiedzi.
- **Atak: inferencja członkostwa** (Membership Inference)
  - manipulowanie danymi treningowymi, aby ujawnić konkretne dane użyte w procesie uczenia modelu.
  - Ochrona obejmuje losowe mieszanie danych, zaciemnianiu: ukryciu krytycznych algorytmów decyzyjnych, modeli predykcyjnych, unikalnych metod przetwarzania oraz stały monitoring.
  - Przykładem ataku jest próba uzyskania poufnych danych, poprzez potwierdzenie, że określony niezanonimizowany rekord był częścią zbioru danych treningowych.
- **Atak: kradzież modelu** (Model Theft)
  - Zdobycie przez atakującego dostępu do parametrów modelu, co pozwala na jego skopiowanie i wykorzystanie niezgodnie z przeznaczeniem.
  - Ochrona obejmuje stosowanie szyfrowania, kontroli dostępu, regularne kopie zapasowe, zaciemnianie/obfuskację kodu oraz znaki wodne.
  - Przykład ataku to kradzież modelu przez konkurenta w celu uzyskania przewagi rynkowej.
- **Atak: łańcuch dostaw uczenia maszynowego** (ML Supply Chain)
  - Atak na platformy MLOps, oprogramowania do zarządzania danymi, modele i inne wykorzystywane w procesie Uczenia Maszynowego (ML).
  - Ochrona wymaga weryfikacji integralność pakietów, stosowanie bezpiecznych źródeł, aktualizacja oprogramowanie oraz kontrolę dostępu w modelu Zero Trust.
  - Ataki prowadzą do kompromitacji infrastruktury i wycieku danych.
- **Atak: transfer nauczania** (Transfer Learning)
  - Dostosowanie modelu zbudowanego do realizacji zadania, do innych zadań w celu spowodowania jego niepożądane działanie.
  - Ochrona wymaga monitoringu, aktualizacji i stosowania wyłącznie zaufanych danych zbiorów danych, izolację modeli, implementacji prywatności różnicowej oraz regularnych audytów bezpieczeństwa.
  - Atak prowokuje błędne wyniki, naruszenia poufności, szkody na reputacji oraz problemy prawne.
- **Atak: przekształcenie modelu** (Model Skewing)
  - Manipulacja rozkładem danych treningowych, aby spowodować niepożądane działanie modelu.
  - Ochrona wymaga silnej kontroli dostępu (Zero Trust), weryfikacji autentyczności danych, technik walidacji i czyszczenia danych, detekcji anomalii oraz regularne monitorowanie wyników modelu.
  - Ataki prowokują błędne decyzje, prowadzą do straty finansowych, uszczerbku na reputacji oraz szkód w krytycznych aplikacjach np. do diagnostyka medyczna.
- **Atak: integralność wyników** (Output Integrity)
  - Manipulacja wynikami modelu uczenia maszynowego, aby wpłynąć na jego zachowanie lub zaszkodzić systemowi.
  - Ochrona wymaga stosowania metod kryptograficznych, bezpiecznych kanałów komunikacyjnych, walidację wyników, regularne aktualizacje oprogramowania oraz audyt wyników.
  - Ataki prowadzą do utraty zaufania do wyników, strat finansowych lub zagrożeń w krytycznych aplikacjach np. systemach zarządzania kryzysowego.
- **Atak: zatrucie modelu** (Model Poisoning)
  - Manipulacja parametrami modelu, w celu zmanipulowania jego wyników.
  - Ochrona wymaga stosowanie technik rekuraryzacji, projektowanie odpornych modeli oraz kryptograficznej ochrony parametrów modelu.
  - Ataki prowadzą do błędnych wyników, np. w systemach bankowych, gdzie model identyfikuje numery kont, w celu defraudacji.

Więcej: https://nflo.pl/baza-wiedzy/obfuskacja-zaciemnianie-kodu-co-to-jest-jak-dziala-i-jak-ja-wykryc/

Źródła: owasp.org, atlas.mitre.org, nask.pl

ZGODNOŚĆ NIS2/CER/RODO ETC

W metodyce IT, rola architekta systemów polega na umiejętnym przełożeniu potrzeb biznesowych na model przetwarzania, procesy i zasoby informatyczne, zapewniając wymagane poziomy bezpieczeństwa, monitorowania i jakości w całym okresie życia produktu. Architekt łączący wiedzę biznesową z informatyczną jest od lat naturalną częścią IT.

Brak, w przeszłości, szczegółowych regulacji prawnych w zakresie informatyki w większości sektorów sprawił, że analogiczna umiejętność łączenia wiedzy prawnej i informatycznej jest obecnie domeną niewielkiego grona audytorów oraz biegłych sądowych.

ODKRYJ I KONSOLIDUJ ŚRODOWISKA VMWARE Z HPE CLOUDPHYSICS

Wgląd z HPE CloudPhysics

Audyt zgodności, oraz możliwości optymalizacji dostępne dla wszystkich klientów.

Wynik: Klienci zyskują wgląd w

Licencje VMware | sprzęt | poprawne skalowanie VM | Możliwości konsolidacji

HPE | AMD

NOWOCZESNA STRATEGIA DLA DANYCH

- **Zarządzanie danymi:**
  Solidne praktyki zarządzania danymi są kluczowe dla nowoczesnej strategii danych. Zarządzają jakością danych, integralnością i bezpieczeństwem - podstawa dla wiarygodnych analiz i wglądu.
- **Zbieranie danych:**
  z różnych, wewnętrznych i zewnętrznych źródeł. Krytycznym jest, aby gromadzenie danych było zwinne i skalowalne w czasie.
- **Integracja danych:**
  Priorytetyzacja integracji danych. Elastyczna konsolidacja i analiza danych w różnych systemach lub platformach.
- **Transformacja danych:**
  Przystosowanie surowych danych do narzędzi i technik analitycznych w celu zapewniania odpowiedniej wiedzy wspierającej proces decyzyjny na każdym poziomie organizacji.
- **Kultura danych:**
  Efektywna strategia danych filarem kultury podejmowania decyzji w oparciu o kompletny zestaw informacji – kluczowe dla innowacji i optymalizacji procesów na każdym poziomie.

Źródło: https://services.google.com/fh/files/misc/modern-data-strategy-2023.pdf

ZGODNOŚĆ — MAPOWANIE NIEZBĘDNYCH FUNKCJONALNOŚCI

Więcej: https://www.gov.pl/web/baza-wiedzy/bezpieczenstwo-filarem-zgodnosci---poradnik-hpe

Lp	Technologia	PCBE	EDF	StoreOnce	Veeam	Network	Wyjaśnienia
1.	Immutability	X	X	X	X		Zintegrowana retencja repozytoriów (Veeam ISV-DI) – eliminacj ludzkiego; akcelerowane sprzętowo. Repozytoria zabezpieczone przed skasowaniem (immutable)
2.	RTO (100TB)	<15 min	<4h*	<4h*	<4h*		RTO<15min niezbędne dla usług wymagających wysokiej dostępności. oraz dla mitygacji ransomware.
3.	2 składnikowa autentykacja	X	X	X	X	X	Specyfikowane w DORA, CER, NIS 2, NSC, zalecenia CeZ
4.	2 składnikowa autoryzacja	x		X			Dodatkowe zabezpieczenie przed błędem ludzkim i ransomware
5.	(mikro/dynamiczna) segmentacja	X	X	X		X	EDF z SPIFFE/SPIRE – mikrosegmentacja CX-10000 (mikro)segmentacja + dynamiczna segmentacja; EDR/XDR itp.
6.	Detekcja i mitygacja ransomware	X		X	X		Dzięki wbudowanym algorytmom detekcji, zdublowanym narzędziom przywrócenia w tym natychmiastowego równoległego przywracania danych.
7.	szyfrowanie	X	X	X	X	X	Transmisji i składowanie danych; CX 10000 – 800Gbps sprzętowo szyfrowanych tuneli.
8.	Śluza dla backupu	X		X	X		Wszystkie repozytoria i systemy kopii zapasowych w każdym ośrodku są w stale odseparowane (okna synchronizacji są zbędne) i zapewniają Immutability.
9.	Ochrona danych (backup)	X			X		Zwielokrotnione narzędzia odzyskiwania danych (koegzystencja). Paradygmat 3-2-1 akcelerowany sprzętowo.
10.	Dystrybuowany FW L4-L5	X				X	(Dynamiczna) segmentacja oraz mikrosegmentacja akcelerowana sprzętowo w przełączniku ToR zintegrowanym z platformą wirtualizacyjną.
11.	Izolowane środowiska przywracania (IRE)	X	X	X	X	X	Pełna automatyka, wsparcie dla Infrastructure as Code (IaC), ponad 10-krotne skrócenie testów przywracania/mitygacji skutków ransomware bez wpływu na produkcję.
12	Security Service Edge/SASE					X	Secure Web Gateway (SWG), Cloud Access Security Broker (CASB), Zero Trust Network Access (ZTNA) and Firewall as a Service (FWaaS) – SASE
13	Akceleracja sprzętowa/Sztuczna inteligencja	X	X	X		X	Sprzętowa akceleracja przetwarzania (CPU, GPU, FPGA, APU) i bezpieczeństwa DPU, Immutability, remediacja ransomware, mikrosegmentacja, DDOS, VPN.

TAKTYKA: ZINTEGROWANA PLATFORMA GOTOWA NA KAŻDE WYZWANIE

HPE w 3-**5 zintegrowanych produktach dostarcza wszystkie niezbędne technologie**

Integrated Data Inteligence (HPE DF)
Cold Storage (HPE DF, Artesca)
DNA Storage
Autonomus Storage (PCBE)
Storage Platform for Gen AI (PC AI)
Edge Storage (HPE DF)
Cyber Storage (PCBE)
Distributed Hybrid Infrastructure (PCBE, HPE DF)
Infrastructure Platform Consumption Services (PCBE)
Enterprise Storage Platform (PCBE)
Hybrid Cloud Storage (HPE DF, Artesca)
Data Storage Management Services (PCBE)
Function Accelerator Cards (Aruba CX-10000)
Hybrid Cloud File Data Services (HPE DF, Artesca)
Nearline Solid State Drive (PCBE)
Container Native Storage (PCBE)
Storage as a Service (PCBE)
NVMe-oF (PCBE)
Open Source Storage Software
Software Defined Storage (HPE DF)
Distributed File Systems (HPE DF)

HPE | AMD

infrastructure leadership

Data Center CPUs

5th Generation
AMD EPYC™ “Turin”

Data Center GPUs

AMD Instinct™
MI325X and MI350
Series

Networking

AMD Pensando™
Pollara 400/Salina 400

AI PCs

3rd Generation
AMD Ryzen™ AI PRO

HPE | AMD

Architektura usług scentralizowanych

Scentralizowane usługi starfull firewall:

- zarządzane z poziomu pojedynczego appliance
- zapewniają przełączanie awaryjne (dla klastra HA)
- zbierają logi zapory sieciowej
- Zapora sieciowa zapewnia funkcję bramy domyślnej w sieci w celu kierowania ruchu
- Schemat adresowania IP musi odzwierciedlać separację dla egzekwowania polityki stanowej

Przekierowanie do zapory sieciowej:

- zwiększa opóźnienia
- zużywa przepustowość sieci
- zwiększa złożoność
- Zwiększenie kosztów zakupu I utrzymania architektury

Akceleratory AMD w rozwiązaniach chmurowych

AMD DPUs and Software

Delivering Meaningful Outcomes for Our Customers

ORACLE CLOUD	Microsoft Azure	IBM Cloud	DXC TECHNOLOGY	vmware
5X SDN Performance[2]	**100X** CPS Gains[3]	Bare-Metal and VSI	TCO Savings of $65M[5]	**3X** App Performance Improvement[6]
Multi-Service	Cost Savings	TCO Savings[4]		**30%** App Reduced Latency[6]
Future Ready		Regulated Industries		

Aruba CX10000: Stateful Fabrics

Przełącznik usług rozproszonych: L2/L3/L4

1. Płynna integracja/migracja w ramach L2/L3
2. Bezpieczeństwo wbudowane w Fabric - prosta, tańsza segmentacja i szyfrowanie
3. Pełna widoczność sieci
4. Wyższy stopień automatyzacji sieci

Architektura usług rozproszonych

Decentralizowane usługi typu starfull:

- Dostarczane już na trasie pakietów (w obrębie przełącznika typu access/leaf), bez przekazywania do scentralizowanej zapory sieciowej
- Programowalne, dedykowane to tego celu DPU zapewniają filtrowanie ruchu zgodnie z zadanymi politykami

Policy & Services Manager (PSM)

- Zapewnia scentraliozwane zarządzanie politykami i usługami
- Dystrybuuje konfigurację do wszystkich przełączników CX 10K
- Zbiera wszelkie logi FW, alarmy, telemetrię I statystyki
- **Integruję się z Aruba Fabric Composer**

Centralne zarządzanie rozproszonym środowiskiem

Dashboard view
Includes information about fabrics, switches, hosts, VMs, and security

Workflow automations and guided setup
Point and click GUI network and security services automation

Distributed services

Network and switch visualization
Hosts, MAC, neighbors, switch inventory, health status

API, MAC integrations with various environments
Including HPE, Aruba, VMware vSphere,ESX, NSX and Nutanix

Źródło: https://www.arubanetworks.com/assets/so/SO_Aruba-CX-10000-Integration-with-Akamai-Guardicore.pdf

Centralne zarządzanie rozproszonym środowiskiem

Dashboard view
Includes information about fabrics, switches, hosts, VMs, and security

Workflow automations and guided setup
Point and click GUI network and security services automation

Distributed services

Network and switch visualization
Hosts, MAC, neighbors, switch inventory, health status

API, MAC integrations with various environments
Including HPE, Aruba, VMware vSphere,ESX, NSX and Nutanix

HPE | AMD

Aruba CX10X00 Stateful Fabrics

Płynna integracja/migracja w ramach L2/L3

Bezpieczeństwo wbudowane w Fabric - prosta, tania segmentacja i szyfrowanie (do 2000 tuneli)

Pełna widoczność sieci

Pełna automatyzacja sieci

Pomijalne opóźnienia <5μs (sieci OT)

Dystrybuowane bezpieczeństwo centralne zarządzanie.

Stateful LB/NAT/etc.	Gen 1 with CPU (52 Intel cores)	Gen 2 with DPU (2 AMD Pensando DPUs)
Bandwidth	200 G/s	800 G/s
Sessions	100 M	100 M
Latency	~100 μs	~2 μs
PPS	~70 Mpps	~140 Mpps
Elephant Flow	10 G/~1 Mpps per core	400 G/70 Mpps per DPU
Power Draw	5X	80% reduction

HPE | AMD

Źródło: https://www.amd.com/content/dam/amd/en/documents/pensando-business-docs/solution-brief/csp-case-study.pdf

infrastructure leadership

Data Center CPUs

5th Generation
AMD EPYC™ “Turin”

Data Center GPUs

AMD Instinct™
MI325X and MI350
Series

Networking

AMD Pensando™
Pollara 400/Salina 400

AI PCs

3rd Generation
AMD Ryzen™ AI PRO

HPE | AMD

AMD EPYC™ 9005 Series Processors

- Increased core density
- Energy efficient
- Broad OPN stack

Cores	AMD EPYC	CCD (Zen5/ Zen5c)	Base/ Boost* (up to GHz)	Default TDP (W)	L3 Cache (MB)	Price (1 KU, USD)
192 cores	9965	"Zen5c"	2.25 / 3.7	500W	384	$14,813
160 cores	9845	"Zen5c"	2.1 / 3.7	390W	320	$13,564
144 cores	9825	"Zen5c"	2.2 / 3.7	390W	384	$13,006
128 cores	9755 9745	"Zen5" "Zen5c"	2.7 / 4.1 2.4 / 3.7	500W 400W	512 256	$12,984 $12,141
96 cores	9655 9655P 9645	"Zen5" "Zen5" "Zen5c"	2.6 / 4.5 2.6 / 4.5 2.3 / 3.7	400W 400W 320W	384 384 256	$11,852 $10,811 $11,048
72 cores	9565	"Zen5"	3.15 / 4.3	400W	384	$10,486
64 cores	9575F 9555 9555P 9535	"Zen5" "Zen5" "Zen5" "Zen5"	3.3 / 5.0 3.2 / 4.4 3.2 / 4.4 2.4 / 4.3	400W 360W 360W 300W	256 256 256 256	$11,791 $9,826 $7,983 $8,992
48 cores	9475F 9455 9455P	"Zen5" "Zen5" "Zen5"	3.65 / 4.8 3.15 / 4.4 3.15 / 4.4	400W 300W 300W	256 256 256	$7,592 $5,412 $4,819
36 cores	9365	"Zen5"	3.4 / 4.3	300W	192	$4,341
32 cores	9375F 9355 9355P 9335	"Zen5" "Zen5" "Zen5" "Zen5"	3.8 / 4.8 3.55 / 4.4 3.55 / 4.4 3.0 / 4.4	320W 280W 280W 210W	256 256 256 128	$5,306 $3,694 $2,998 $3,178
24 cores	9275F 9255	"Zen5" "Zen5"	4.1 / 4.8 3.25 / 4.3	320W 200W	256 128	$3,439 $2,495
16 cores	9175F 9135 9115	"Zen5" "Zen5" "Zen5"	4.2 / 5.0 3.65 / 4.3 2.6 / 4.1	320W 200W 125W	512 64 64	$4,256 $1,214 $726
8 cores	9015	"Zen5"	3.6 / 4.1	125W	64	$527

HPE | AMD

Leverage HPE To Solve Your Customer's Data Center Problems

Compute

Industry-Standard servers

HPE ProLiant DL325 Gen11
4th and 5th Gen EPYC

HPE ProLiant DL345 Gen11
4th and 5th Gen EPYC

HPE ProLiant DL365 Gen11
4th and 5th Gen EPYC

HPE ProLiant DL385 Gen11
4th and 5th Gen EPYC

Retail and Edge

DX325 Gen10 Plus v2
3rd Gen EPYC

New!

HPE ProLiant DL145 Gen11 EPYC 8000

Gen 12 Coming Soon

HPE ProLiant DL325 Gen12
Turin (500W), 2dpc

HPE ProLiant DL345 Gen12
Turin (500W), 2dpc

HPE ProLiant ML/DL
AMD EPYC 4005

Networking / Security

HPE Aruba CX10000
Pensando

HPE Aruba CX10400
Pensando

HPE Aruba CX9300S
Embedded CPUs

HPE Aruba CX5400
HPE Aruba CX8325
Embedded CPUs

HPE Aruba EdgeConnect
5th Gen EPYC

Storage

HPE Alletra MP
2nd & 3rd Gen Embedded EPYC

HPE Alletra 6000
2nd Gen Embedded EPYC

HPE Alletra dHCI
4th and 5th Gen EPYC

SimpliVity 325 Gen11
4th Gen EPYC

NUTANIX
DX Platforms
3rd & 4th Gen EPYC

HPE | AMD

infrastructure leadership

Data Center CPUs

5th Generation
AMD EPYC™ “Turin”

Data Center GPUs

AMD Instinct™
MI325X and MI350
Series

Networking

AMD Pensando™
Pollara 400/Salina 400

AI PCs

3rd Generation
AMD Ryzen™ AI PRO

HPE | AMD

Delivering on Annual Roadmap Commitment

AMD INSTINCT™

2023

AMD INSTINCT™
MI325X

2024

AMD INSTINCT™
MI35X SERIES

2025

AMD INSTINCT™
MI400 SERIES

2026

HPE | AMD

Roadmap subject to change

MI350 Series GPUs

AMD INSTINCT	Instinct™ MI350X	Instinct™ MI355X
MEMORY	288 GB HBM3E	288 GB HBM3E
MEMORY BANDWIDTH	8 TB/s	8 TB/s
PEAK FP64*	72 TF	79 TF
PEAK FP16	4.6 PF	5 PF
PEAK FP8	9.2 PF	10 PF
PEAK FP6	18.4 PF	20 PF
PEAK FP4	18.4 PF	20 PF
TBP	Up to 1000W	Up to 1400W

HPE | AMD CDNA 4

Industry's Highest Performing GPUs for AI and HPC

AMD INSTINCT MI355X GPU VS NVIDIA B200/GB200

Instinct™ MI355X	vs. GB200	vs. B200
MEMORY CAPACITY	~1.6x	1.6x
MEMORY BANDWIDTH	1x	1x
PEAK FP64*/PEAK 32	~2x	~2x
PEAK FP16	1x	1.1x
PEAK FP8	1x	1.1x
PEAK FP6	2x	2.2x
PEAK FP4	1x	1.1x

HPE | AMD CDNA 4

Equal GPU Count for Architecture-Level Comparison with GB200

*All Matrix Core numbers except FP64 with sparsity

See endnote MI350-006, MI350-009

Deep collaboration with developer community

PyTorch

Day 0 support for latest features

Triton

Vendor agnostic compiler support

Hugging Face

Nightly CI/CD ensuring all models work out-of-box

vLLM

SGLang

JAX

Tensor Flow

ONNX Runtime

OpenXLA

DeepSpeed

MLIR | IREE

Increasing open-source contributions and expanding footprint

Use or mention of third-party marks, logos, products, services, or solutions herein is for informational purposes only and no endorsement by AMD is intended or implied. GD-83

Broad and Growing Instinct Portfolio

AI SYSTEMS

HPE Cray XD665	HPE Cray XD675	HPE ProLiant Compute XD685	HPE DL385 Gen11
2x 4th Gen EPYC	2x 4th Gen EPYC	**2x 5th Gen EPYC**	**2x 4th and 5th Gen EPYC**
4x Nvidia GPUs	8x AMD MI300X GPUs	**8x AMD MI325X / 300X / 355X** 8x H200/B200	4x/8x NVIDIA PCIe GPUs
4U Chassis	8U Chassis AC	6U AC / 5U DLC Chassis	2U
HPC Simulations, AI Training and Deep Learning	Larger AI Models, AI Training	Larger AI Models, AI Training	Smaller AI Models, ML, Big Data Analytics
	Shipping	**May/June**	**Shipping**

HPE | AMD

Analiza praktyki wdrożeń AI.

Badania i komentarz Gartnera na temat trendów w dziedzinie sztucznej inteligencji w 2025 roku:

„Według raportu Gartnera ponad 40% projektów z zakresu sztucznej inteligencji opartej na agentach zostanie anulowanych do końca 2027 roku z powodu rosnących kosztów i niejasnej wartości biznesowej”

…MIT odkrył, że największy zwrot z inwestycji (ROI) można uzyskać dzięki automatyzacji zaplecza — eliminacji outsourcingu procesów biznesowych, obniżeniu kosztów agencji zewnętrznych i usprawnieniu operacji.

…„Prawie wszędzie, gdzie działaliśmy, przedsiębiorstwa próbowały zbudować własne narzędzie” – powiedział, ale dane pokazały, że **komercyjne rozwiązania przyniosły bardziej wiarygodne rezultaty**.

źródła: https://www.reuters.com/business/over-40-agentic-ai-projects-will-be-scrapped-by-2027-gartner-says-2025-06-25/

https://crn.pl/aktualnosci/gartner-tonuje-nastroje-wobec-agentow-ai/

https://crn.pl/aktualnosci/mit-ostrzega-banka-ai-moze-peknac-95-projektow-bez-zyskow/

Akceleratory GPU AMD vs NVIDIA

Kiedy AMD Instinct (GPU):

- ekosystemy open source (bez vendor lock-in)
- Wnioskowanie (w tym wsadowe) na dużą skalę
- brak zależności od narzędzi: ROCm
- cena, efektywność energetyczna, czas dostarczenia (brak ograniczeń eksportowych).
- Infinity Fabric Link, wydajne połączenia z niskim opóźnieniem.

Kiedy NVIDIA (GPU):

- maksymalna kompatybilność i wydajności
- Trening modeli głębokiego uczenia (DL) na dużą skalę
- Polegasz na natywnych narzędziach CUDA (np. cuDNN, TensorRT)
- Chcesz obsługi plug-and-play we wszystkich frameworkach

W debacie dotyczącej GPU AMD i NVIDIA najlepszy wybór zależy od harmonogramu, celów, budżetu oraz tolerancji na złożoność konfiguracji. NVIDIA to złoty standard w dziedzinie głębokiego uczenia dzięki zbudowanemu zestawowi oprogramowania oraz szerokiej kompatybilności.
AMD jest realną alternatywą, zwłaszcza dla organizacji i aplikacji wrażliwych na koszty, zgodnych z open source (suwerenność).

Niezależnie od tego, czy rozbudowujesz klaster AI, czy przeprowadzasz wnioskowanie na sprzęcie konsumenckim, zrozumienie kompromisów między AMD a NVIDIA jest kluczowe dla świadomych decyzji i redukcji ryzyka.

Dojrzałość rozwiązań

IDC MarketScape Worldwide Private AI Infrastructure Systems Vendor Assessment

Source: IDC, 2025

Źródło: https://www.gartner.com/doc/reprints?id=1-2LITSZ3Z&ct=250724&st=sb

Gartner

HPE | AMD

Platforma HPE Private Cloud AI

Sztuczna Inteligencja pod pełną kontrolą

Nie tylko Chat GPT

CHATGPT

BARD AI

BIELIK.AI

Turnkey AI optimized Hardware | Private Cloud AI

Best for

Unified experience through HPE GreenLake cloud

Inferencing	Inferencing + RAG	Inferencing + RAG + Fine-tuning
Small	**Medium**	**XLarge**
4 or 8 NVIDIA L40S GPUs	8 or 16 NVIDIA L40S GPUs	16 or 32 NVIDIA H200 NVL GPUs
109TB File/Object	217TB File/Object	670TB File/Object
100GbE NVIDIA Networking	200GbE NVIDIA Networking	400GbE NVIDIA Networking
up to 8 kW rack	up to 17.7 kW	up to 35 kW x 2

HPE | AMD

Kluczowe funkcje programowalnych barier (Guardrails)

— Budowa zaufanych, bezpiecznych i chronionych aplikacji wykorzystujących LLM z wykorzystaniem ram zabezpieczających i prowadzących konwersację:

- **dostosowanie zachowania aplikacji w określonych tematach do interlokutora,**
- **blokada zbędnych dyskusji oraz niepożądanych tematów,**
- **blokada niepożądanych informacji — cenzura w celu eliminacji np. fake news.**

— Bezpieczne łączenie modeli, łańcuchów i innych usług: elastyczne i bezpiecznie łączenie LLM z innymi usługami/narzędziami.

— Sterowanie dialogiem z LLM:

- **realizacja zdefiniowanych ścieżek konwersacji - projektowanie interakcji zgodnie z najlepszą praktyką**
- **projektowania konwersacji i egzekwowanie standardowych procedur operacyjnych (np. uwierzytelniania, wsparcia).**

Funkcje Guardrails cd

Programowe bariery

Funkcje Guardrails cd

Statyczne testowanie bezpieczeństwa aplikacji (SAST)

- SAST firmy GuardRails zapewnia analizę kodu źródłowego oprogramowania, w celu wykrycia luk w zabezpieczeniach i zapewnienia jakości w pełnym cyklu życia oprogramowania (SDLC).

Dynamiczne testy bezpieczeństwa aplikacji (DAST)

— Wykrycie luk w zabezpieczeniach w czasie wykonywania (run-time)

- uzupełnia statyczne testy bezpieczeństwa aplikacji (SAST). SAST analizuje kod źródłowy aplikacji, DAST identyfikuje luki, w czasie pracy, zapewniając realistyczny obraz podatności aplikacji.

— Symulacja rzeczywistych ataków

- umożliwia wykrycie i usunięcie luk w zabezpieczeniach, które mogły zostać przeoczone podczas procesu rozwoju.

GuardRails wykrywa luki w zabezpieczeniach w zależnościach aplikacji, wykorzystując technologie Software Composition Analysis (SCA) i Software Bill of Materials (SBOM).

- Software Composition Analysis (SCA) to podejście do testowania bezpieczeństwa, które koncentruje się na identyfikacji i zarządzaniu lukami w komponentach open source i bibliotekach innych firm wykorzystywanych w łańcuchu dostaw oprogramowania. Zabezpiecza tworzenie oprogramowania poprzez zapewnienie kompleksowej analizy zależności, wykrywanie luk w zabezpieczeniach i ocenę ryzyka. Pomaga znaleźć luki w komponentach innych firm w tym open-source) w celu proaktywnej ich eliminacji.

Zintegrowana zarządzanie cyklem życia z securiti.ai

Centrum dowodzenia danymi >

Bezpieczeństwo danych + AI
Zespołów

Dane+sztuczna inteligencja
Zespołów

Nadzór nad danymi
Zespołów

Prywatność danych
Zespołów

Zabezpieczyć Dane+sztuczna inteligencja w dowolnym miejscu →

Zarządzanie stanem bezpieczeństwa danych
Zabezpiecz poufne dane w dowolnym miejscu, od hybrydowego środowiska wielochmurowego po SaaS

Bezpieczeństwo i zarządzanie sztuczną inteligencją
Ustanowienie mechanizmów kontroli w celu bezpiecznego wdrażania technologii sztucznej inteligencji, w tym GenAI

Zabezpieczenia dla pilotów AI w SaaS
Odblokuj największe przeszkody na drodze do bezpiecznego wdrożenia rozwiązań Copilot opartych na sztucznej inteligencji, takich jak Microsoft 365 Copilot

Dostęp do danych, inteligencja i zarządzanie
Monitorowanie dostępu użytkowników do danych i wymuszanie kontroli najmniejszych uprawnień

Odnajdywanie i klasyfikacja danych
Odkrywanie zasobów w tle i natywnych dla chmury oraz dokładne klasyfikowanie danych

Zarządzanie zgodnością
Ocena i poprawa zgodności z ramami najlepszych praktyk w zakresie zabezpieczeń

Analiza wpływu naruszeń
Analizuj wpływ naruszeń i automatyzuj powiadomienia dla osób, których dotyczy naruszenie

Zarządzanie przepływem danych
Informacje o pochodzeniu danych i zabezpieczanie danych przesyłanych strumieniowo w czasie rzeczywistym

HPE | AMD

Więcej: https://securiti.ai/press-release/securiti-partners-with-hewlett-packard-enterprise-to-build-sage-enterprise-ai-systems/

Dziękuję

Piotr NOGAŚ
E: piotr.nogas@hpe.com
M:+48 601 890 046

Zapraszamy do HPE CEC: https://www.hpe.com/h22166/event/eventpage.aspx?eventid=MgA4ADAANwA0ADcA&cc=pl&lang=pl#Wybierz-si%C4%99-w-podr%C3%B3%C5%BC-po-technologiach-HPE!

HPE | AMD

© 2025 Hewlett Packard Enterprise Development LP