

# Profile Skalowania 1.8

Jak skaluje się Vigil Guard Enterprise 1.8 w środowiskach produkcyjnych

## PRZEWIDYWALNE SKALOWANIE WDROŻEŃ

Vigil Guard Enterprise 1.8 skaluje środowiska produkcyjne za pomocą gotowych profili skalowania. Profil jest dobierany i walidowany względem zasobów hosta podczas instalacji lub zatwierdzonej zmiany konfiguracji, a następnie utrwalany.

Dzięki temu wdrożenie działa w powtarzalny sposób po restartach, aktualizacjach i operacjach serwisowych, a konfiguracja uruchomieniowa pozostaje spójna z tym, co zatwierdzono podczas instalacji. Wybór wyższego profilu podnosi zarówno klasę hosta, jak i przepustowość najcięższych etapów analizy.

## KLUCZOWE KORZYŚCI

- » Cztery gotowe profile skalowania dopasowane do klasy hosta
- » Profil dobierany i walidowany względem zasobów hosta podczas instalacji
- » Wyższy profil zwiększa równoległość najcięższych etapów analizy
- » Powtarzalne wdrożenia po restartach, aktualizacjach i serwisie
- » Konfiguracja uruchomieniowa spójna z zatwierdzoną instalacją
- » Dopasowane limity współbieżności i ochrona przed przeciążeniem
- » Aktualizacje zachowują wybrany profil
- » Sygnalizacja, gdy środowisko wymaga większej klasy hosta

<b>4</b>	<b>20-52</b>	<b>30-256 GB</b>	<b>1 host</b>
profile skalowania	rdzenie CPU (zakres klas)	RAM (zakres klas)	wdrożenie jednowęzłowe Linux

Wdrożenie lokalne (on-premise) • Sieć odizolowana (air-gapped) • Jeden serwer Linux

## PROFILE SKALOWANIA

Każde wdrożenie korzysta z jednego z czterech profili skalowania, dobranego do klasy hosta. Profil określa minimalne wymagania serwera oraz docelowe przeznaczenie produkcyjne.

PROFIL	KLASA WDROŻENIA	CPU (MIN.)	RAM (MIN.)	DYSK (ZALECANY)	PRZEZNACZENIE
<b>prod-32-balanced</b>	Wejściowa	20 rdzeni	30 GB	min. 150 GB NVMe SSD	Niski ruch, środowiska wewnętrzne lub kontrolowane pilotaże.
<b>prod-64-balanced</b>	Bazowa (zalecana)	28 rdzeni	64 GB	150-200 GB NVMe SSD	Domyślny wybór dla większości wdrożeń produkcyjnych.
<b>prod-128-balanced</b>	Wzrostowa	36 rdzeni	128 GB	300 GB+ NVMe SSD	Wyższa, utrzymująca się współbieżność i większe obciążenie analityczne.
<b>prod-256-balanced</b>	Duża jednowęzłowa	52 rdzenie	256 GB	500 GB+ NVMe SSD	Największe wdrożenia jednowęzłowe z dużym zapasem zasobów.

## ZALECANY PUNKT STARTOWY

Dla większości nowych wdrożeń produkcyjnych zalecany profil jest **prod-64-balanced** na dedykowanym serwerze Linux. Zapewnia bezpieczny zapas dla warstwy danych, analityki i inferencji modeli, pozostawiając miejsce na wzrost. Profil wejściowy prod-32-balanced jest najmniejszym zgodnym kształtem, a nie domyślnym wyborem dla wdrożeń enterprise.

## CO ZMIENIA WYŻSZY PROFIL

Skalowanie w wersji 1.8 to nie tylko większy serwer. Wraz z profilem rośnie realna zdolność przetwarzania, a system pozostaje stabilny pod obciążeniem.

### Większa równoległość analizy

Wyższy profil zwiększa przepustowość najcięższych etapów detekcji i analizy, poprawiając odporność na ruch szczytowy oraz zapytania unikalne.

### Dopasowane limity współbieżności

Limity przetwarzania równoległego są dobierane do profilu, aby utrzymać przewidywalne zachowanie systemu.

### Ochrona przed przeciążeniem

Mechanizmy dopuszczania ruchu i budżety pracy ograniczają degradację pod obciążeniem i chronią środowisko.

### Zapas dla danych i analityki

Większa klasa hosta zwiększa zapas pamięci, dysku i pamięci podręcznej dla magazynu zdarzeń i obciążeń analitycznych.

## WARSTWY PLATFORMY A SKALOWANIE

Platforma działa jako jeden spójny stos na serwerze Linux. Poszczególne warstwy w różnym stopniu wpływają na dobór klasy hosta.

WARSTWA	ROLA	WPŁYW NA KLASĘ HOSTA
Reverse proxy / brama TLS	Wejście ruchu i terminacja TLS	Lekka, stała ścieżka wejścia.
Rdzeń danych i analityka	Magistrala komunikatów, pamięć podręczna, baza danych analityczna, magazyn wektorowy, kolektor metryk	Wyznacza zapotrzebowanie na pamięć, dysk i wydajność I/O.
Aplikacja i API	Obsługa żądań i panel zarządzania	Umiarkowana pamięć, wrażliwość na opóźnienia użytkownika.
Przetwarzanie i analiza	Równoległe warstwy detekcji	Główne zużycie CPU w trakcie analizy.
Usługi modeli ML	Inferencja modeli detekcji	Skalowane wraz z profilem; główne wąskie gardło wydajności.

## WALIDACJA I POWTARZALNOŚĆ

Każdy profil jest walidowany względem dostępnych zasobów hosta: liczby rdzeni CPU, pamięci RAM i wolnego miejsca na dysku. Jeśli serwer nie spełnia wymagań danego profilu, profil nie zostanie zastosowany. Zatwierdzona konfiguracja jest utrwalana, dzięki czemu kolejne uruchomienia środowiska są powtarzalne, a stan działający nie rozjeżdża się z tym, co zatwierdzono.

## AKTUALIZACJE I ZMIANA PROFILU

Aktualizacja zachowuje wybrany profil skalowania i ponownie dopasowuje konfigurację uruchomieniową do aktualnych zasobów hosta oraz nowej wersji platformy. Dzięki temu aktualizacja nie zmienia charakterystyki działającego środowiska w niekontrolowany sposób.

Zmiana profilu odbywa się przez dostarczone narzędzia administracyjne i jest potwierdzana procedurą weryfikacji wdrożenia. Utrzymuje to konfigurację w zgodzie z procesem instalacji i zachowuje powtarzalność wdrożenia.

### WYDAJNOŚĆ I JEJ UWARUNKOWANIA

Wydajność poszczególnych wdrożeń zależy od charakteru ruchu, udziału zapytań unikalnych, konfiguracji retencji danych, obciążenia analitycznego, parametrów hosta oraz warunków środowiska klienta. Z tego powodu wartości przepustowości i opóźnień nie powinny być traktowane jako uniwersalna gwarancja. Dla większych klas wdrożeniowych końcową ocenę wydajności należy potwierdzić w docelowym środowisku klienta.

## KIEDY PRZEJŚĆ NA WIĘKSZY PROFIL

Następujące sygnały wskazują, że warto zaplanować większą klasę hosta i wyższy profil skalowania:

- Zużycie pamięci często zbliża się do skonfigurowanych limitów usług analitycznych i modeli.
- Rośnie obciążenie analityczne, wolumen zdarzeń lub wymagania dotyczące retencji danych.
- Opóźnienie analizy lub czas odpowiedzi rosną przy normalnym ruchu biznesowym.
- Monitoring raportuje częściową utratę pojemności lub ostrzeżenia o zasobach hosta.
- Zużycie dysku zbliża się do skonfigurowanych progów.

W razie potrzeby przejdź na kolejną, większą klasę hosta i odpowiadający jej profil skalowania zamiast doraźnych zmian konfiguracji.

## OPCJE WDROŻENIA

- **On-premise:** jeden dedykowany serwer Linux, pełna kontrola nad danymi.
- **Air-gapped:** pełna funkcjonalność bez dostępu do internetu.
- **Bez wymogu GPU:** inferencja modeli działa na CPU.

# Skalowanie dopasowane do Twojego wdrożenia

Cztery gotowe profile skalowania pozwalają dopasować Vigil Guard Enterprise do klasy hosta i obciążenia, zachowując powtarzalne i przewidywalne wdrożenia.



## ZALECANY STANDARD: PROD-64-BALANCED

Dedykowany serwer Linux, 28 rdzeni CPU, 64 GB RAM, dysk NVMe SSD min. 150-200 GB. Najbardziej zrównoważony punkt startowy dla wdrożeń produkcyjnych: bezpieczny zapas dla analityki, inferencji modeli oraz wzrostu.

## VIGIL GUARD ENTERPRISE 1.8

Cztery gotowe profile skalowania. Walidacja względem zasobów hosta. Powtarzalne wdrożenia.  
Wdrożenie lokalne, sieć odizolowana, jeden serwer Linux, bez wymogu GPU.

### Zespoły Wdrożeniowe

Powtarzalne, przewidywalne uruchomienia środowiska

### Operacje / SRE

Dobór profilu i kontrolowane aktualizacje

### Architektura / Infrastruktura

Jasne wymagania hosta i planowanie pojemności

[vigilguard.ai](https://vigilguard.ai) | [contact@vigilguard.ai](mailto:contact@vigilguard.ai)