

— EZD RP —

Wymagania sprzętowe oraz scenariusze implementacji EZD RP

Warszawa 26.05.2023 r.

NASK-PIB
ul. Kolska 12
01-045 Warszawa

nask@nask.pl
+48 22 380 82 00
+48 22 380 82 01

NIP: 521 04 17 157
Regon: 010464542
KRS: 0000012938

BNP Paribas Bank Polska Spółka Akcyjna
z siedzibą w Warszawie
ul. Kasprzaka 2, 01-211 Warszawa
Numer konta:
28 1750 0009 0000 0000 0094 9997

Rekomendowane wymagania sprzętowe i programowe oraz scenariusze implementacji EZD RP (od wersji 15)

Spis treści

1.	Wymagania sprzętowe	2
1.1	Infrastruktura serwerowa przy wdrożeniach dla organizacji do 150 użytkowników	2
1.2	Infrastruktura serwerowa przy wdrożeniach dla organizacji do 500 użytkowników	3
1.3	Infrastruktura serwerowa przy wdrożeniach dla organizacji do 2000 użytkowników:	3
2.	Rekomendowane rozwiązania techniczne	4
3.	Środowisko uruchomieniowe systemu EZD RP	4
4.	Środowisko uruchomieniowe kontenerów Kubernetes i system operacyjny	5
5.	Rekomendowane scenariusze implementacji EZD RP	5
5.1.	Jednowęzłowe środowisko Kubernetes ze zintegrowanymi usługami bazy danych i repozytorium plików	6
5.2.	Jednowęzłowe środowisko Kubernetes połączone z zewnętrznymi usługami baz danych i repozytorium plików	7
5.3.	Wielowęzłowe klastrowe środowisko Kubernetes połączone z zewnętrznymi usługami baz danych, systemów kolejkowych i repozytorium plików – skalowalne i redundantne.	8

1. Wymagania sprzętowe

Do instalacji i prawidłowego działania systemu EZD RP potrzebna jest infrastruktura serwerowa o poniżej podanych parametrach.

1.1 Infrastruktura serwerowa przy wdrożeniach dla organizacji **do 150 użytkowników**

- procesor lub procesory: 16 vCPU;
- pamięć RAM: 32 GB;
- przestrzeń dyskowa: 1 TB (RAW) na szybkich dyskach NVMe/SSD;
- opcjonalnie karta GPU!: 1 x Nvidia; Compute Capability 7.0+; RAM: 10 GB+; 3500+ rdzeni CUDA (np. GeForce RTX 3060 12 GB RAM).

1.2 Infrastruktura serwerowa przy wdrożeniach dla organizacji **do 500 użytkowników**

- procesor lub procesory: 32 vCPU;
- pamięć RAM: 64 GB;
- przestrzeń dyskowa: 3 TB (RAW), rekomendowane co najmniej 1 TB przestrzeni na szybkich dyskach NVMe/SSD na potrzeby obliczeń i bufora danych plus dyski SSD/HDD na potrzeby przechowywania danych;
- opcjonalnie karta GPU¹: 1 x Nvidia; Compute Capability 7.0+; RAM: 10 GB+; 8000+ rdzeni CUDA (np. GeForce RTX 3080 10 GB RAM).

1.3 Infrastruktura serwerowa przy wdrożeniach dla organizacji **do 2000 użytkowników**

- procesor lub procesory: 64 vCPU;
- pamięć RAM: 256 GB;
- przestrzeń dyskowa: 32 TB (RAW), 100% przestrzeni na dyskach NVMe/SSD lub 30% przestrzeni na szybkich dyskach NVMe/SSD na potrzeby obliczeń i bufora danych plus 70% przestrzeni na dyskach SSD/HDD na potrzeby przechowywania danych;
- opcjonalnie karta GPU¹: 1 x Nvidia; Compute Capability 7.0+; RAM: 24 GB+; 10 000+ rdzeni CUDA (np. Ampere A40 48 GB RAM lub GeForce RTX 3090 24 GB).

Uwaga – konfiguracja, wymagania sprzętowe oraz rekomendowana architektura większych instalacji wymaga konsultacji z konsultantami wsparcia technicznego EZD RP.

Powyższe parametry obejmują zapotrzebowanie na moc obliczeniową i przestrzeń dyskową dla:

- aplikacji EZD RP;
- bazy danych i systemów kolejkowych;
- klastrów systemu plików;
- środowiska uruchomieniowego kontenerów Kubernetes.

Zasoby rekomendowane dla wdrożeń EZD RP **nie uwzględniają dodatkowych wymagań sprzętowych** koniecznych do działania środowiska wirtualizacyjnego, systemów bezpieczeństwa oraz infrastruktury sieciowej.

W przypadku wystąpienia niestandardowych obciążeń takich jak np. masowa archiwizacja danych lub równoległe przesyłanie dużej liczby plików, wymienione parametry sprzętowe mogą być niewystarczające.

¹ **użycie kart GPU jest opcjonalne** i rekomendowane w przypadku wykorzystywania niektórych funkcji modułu sztucznej inteligencji (SI) np. rozdzielania korespondencji, proponowania klas JRWA, wyszukiwania podobnych spraw. Zastosowane modele SI wymagają douczenia z użyciem danych własnych instytucji użytkującej EZD RP. Włączenie wymienionych funkcji nie jest wymagane do prawidłowego działania systemu.

2. Rekomendowane rozwiązania techniczne

W celu zapewnienia odpowiedniego poziomu bezpieczeństwa, niezawodności i wydajności zalecana jest wirtualizacja infrastruktury serwerowej oraz sieciowej i zastosowanie wbudowanych w platformy wirtualizacyjne rozwiązań klastrowych, funkcji failover oraz replikacji, a także ulokowanie urządzeń i systemów w dwóch niezależnych serwerowniach.

Zalecane jest, aby bazy danych i systemy kolejkowe oraz repozytorium działały w trybie wysokiej dostępności (failover, replikacja, klastery) i umożliwiały odtworzenie danych oraz przywrócenie działania usług zgodnie ze standardami przyjętymi w instytucji.

W zależności od oczekiwanego poziomu bezpieczeństwa oraz posiadanych kompetencji możliwe jest wykorzystanie darmowego oprogramowania open source, oprogramowania open source ze wsparciem producenta (subskrypcje odpłatne) lub komercyjnych rozwiązań w zakresie:

- relacyjnej bazy danych;
- klastrowego systemu plików lub repozytorium obiektowego (programowego lub sprzętowego);
- wirtualizacji infrastruktury (IaaS);
- środowiska uruchomieniowego kontenerów Kubernetes.

3. Środowisko uruchomieniowe systemu EZD RP

Aplikacja EZD RP została zaprojektowana i przygotowana do implementacji w środowisku uruchomieniowym kontenerów Kubernetes. Obrazy kontenerów wraz z kreatorami ułatwiającymi wdrożenie aplikacji i przeprowadzenie wstępnej konfiguracji są udostępnione w repozytorium. Kreator wspiera również instalacje aktualizacji oraz wdrażanie kolejnych wydań EZD RP.

W zastosowaniach produkcyjnych do prawidłowego działania aplikacji serwerowej EZD RP konieczne są następujące usługi i zasoby (niezależnie od środowiska Kubernetes):

- **relacyjna baza danych** zgodna z PostgreSQL lub MS SQL oraz konto systemowe umożliwiające utworzenie i korzystanie z bazy danych,
- **systemy kolejkowe i buforujące dane:** Redis, Rabbit,
- **usługa składowania plików** w postaci udziału sieciowego NFS lub repozytorium obiektowego (programowego lub sprzętowego) zgodnego z protokołem S3,
- **publikowana nazwa DNS** (publicznie lub wewnętrznie) dla serwera aplikacji EZD RP,
- **certyfikat SSL typu Wildcard** dla domeny w liczbie odpowiadającej uruchamianym instancjom (np. testowa, preprodukcyjna, produkcyjna),
- **usługa wysyłki poczty SMTP** i konto umożliwiające korzystanie z tej usługi.

4. Środowisko uruchomieniowe kontenerów Kubernetes i system operacyjny

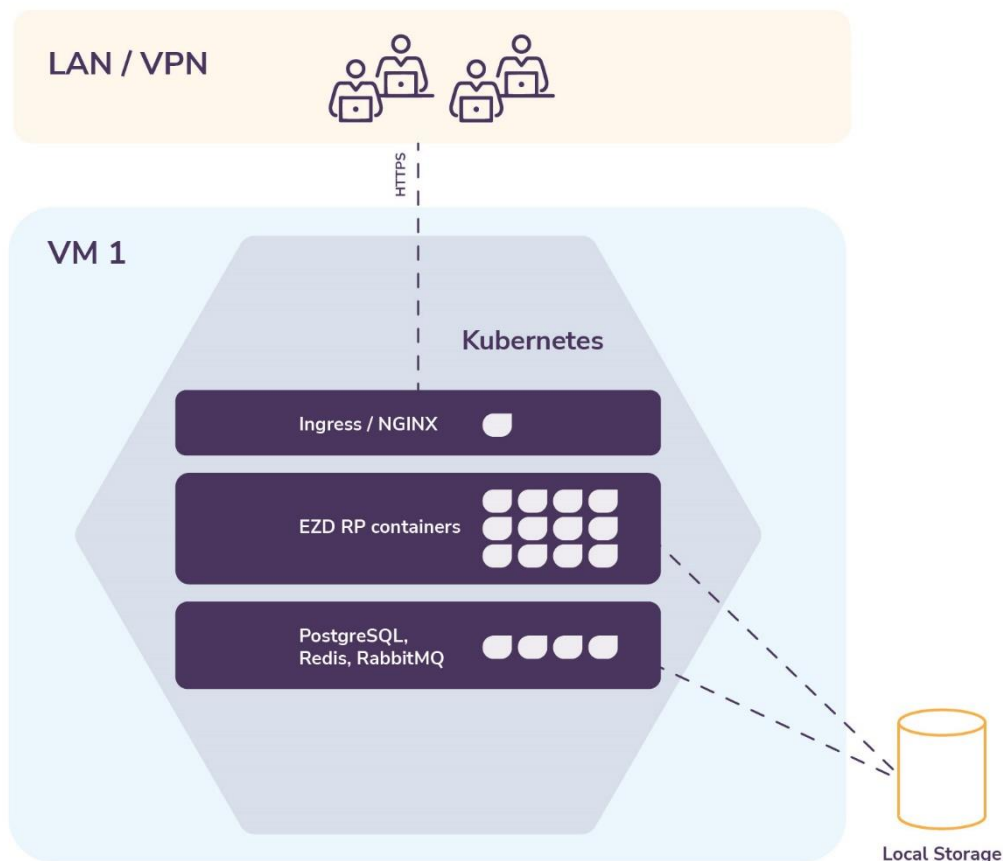
Kubernetes jest środowiskiem pozwalającym uruchamiać aplikacje kontenerowe w sposób niezależny od używanych serwerów wirtualnych lub fizycznych. Dzięki dużej liczbie dostępnych rozwiązań komercyjnych i open source oraz otwartości platformy Kubernetes klaster lub jednowęzłowe środowisko można zainstalować i uruchomić na niemal dowolnym systemie operacyjnym niezależnie od stosowanego systemu wirtualizacji. Popularne, szeroko stosowane i rekomendowane obecnie środowisko to Linux Ubuntu. W przypadku dużych instytucji kilka serwerów Ubuntu będzie zasobem wielowęzłowego klastra Kubernetes. Dla większości zastosowań, zarówno w małej jak i dużej skali rekomendujemy instalację środowiska Kubernetes w oparciu o dystrybucję RKE2 (<https://docs.rke2.io/>).

5. Rekomendowane scenariusze implementacji EZD RP

W zależności od charakteru użycia systemu i planowanych obciążeń środowiska przez użytkowników i/lub zintegrowane systemy rekomendowane są następujące warianty implementacji:

5.1. Jednowęzłowe środowisko Kubernetes ze zintegrowanymi usługami bazy danych i repozytorium plików.

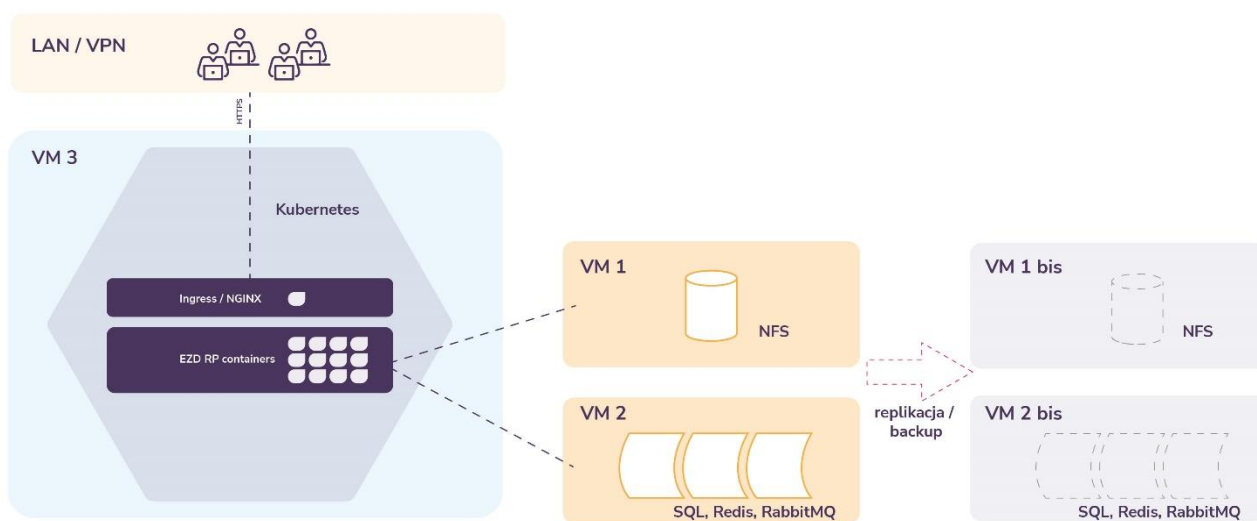
Rozwiązanie to polecane jest do zastosowań testowych, ewaluacyjnych, edukacyjnych, deweloperskich z uwagi na brak dostatecznych zabezpieczeń chroniących przed utratą danych. Opis instalacji znajduje się w artykule: [Instrukcja instalacji aplikacji EZD RP – środowisko testowe](#).



5.2. Jednowęzłowe środowisko Kubernetes połączone z zewnętrznymi usługami baz danych i repozytorium plików.

Wariant rekomendowany do zastosowań produkcyjnych. Można stosować go dla niewielkich obciążeń i maksymalnie 500 użytkowników. Oprócz środowiska Kubernetes należy zaimplementować relacyjne i nierelacyjne bazy danych, systemy kolejkowe i repozytorium plików oraz wskazać aplikacji EZD RP parametry połączenia do uruchomionych usług.

Opis instalacji znajduje się w artykule: [Instrukcja instalacji aplikacji EZD RP – środowisko do 500 użytkowników](#).

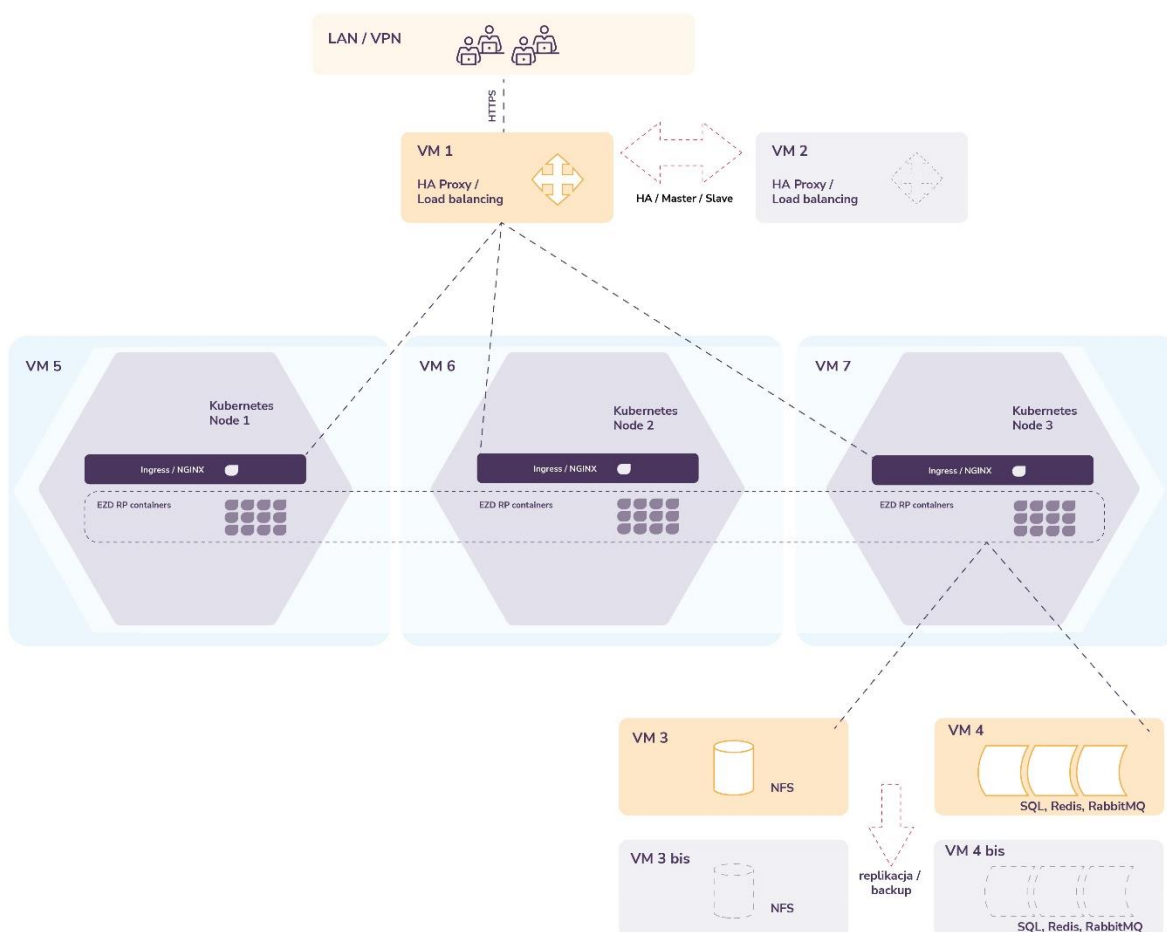


Oczekiwane parametry RPO i RTO determinują sposób implementacji baz danych i repozytorium plików oraz zastosowanie odpowiednio:

- rozwiązań klastrowych active-active lub active-passive;
- replikacji synchronicznej lub asynchronicznej danych i plików;
- określonych interwałów wykonywania kopii zapasowych i retencji backupu.

5.3. Wielowęzłowe klastrowe środowisko Kubernetes połączone z zewnętrznymi usługami baz danych, systemów kolejkowych i repozytorium plików – skalowalne i redundantne.

Wariant implementacji dla zastosowań produkcyjnych dla dużej liczby użytkowników. Oferuje wysoki poziom dostępności systemu i zapewnia jego nieprzerwane działanie w przypadku awarii pojedynczych komponentów. Oprócz środowiska Kubernetes należy zaimplementować relacyjne i nierelacyjne bazy danych, systemy kolejkowe i repozytorium plików oraz wskazać aplikacji EZD RP parametry połączenia do uruchomionych usług.



Oczekiwane parametry RPO i RTO determinują sposób implementacji baz danych i repozytorium plików oraz zastosowanie odpowiednio:

- rozwiązań klastrowych active-active lub active-passive;
- replikacji synchronicznej lub asynchronicznej danych i plików;
- określonych interwałów wykonywania kopii zapasowych i retencji backupu.

5.4. Rekomendacje dotyczące przywracania systemu

Zalecane jest wykonywanie kopii bezpieczeństwa tylko danych perystentych:

- relacyjnej bazy danych (PostgreSQL/MS SQL),
- zawartości systemów kolejkowych i cache (Redis, Rabbit),
- repozytorium plików (udział NFS lub repozytorium obiektowe).

W przypadku wystąpienia awarii wymagającej przywrócenia systemu, należy odtworzyć dane i uruchomić ponownie usługi przechowujące dane persystentne.

Pozostałe elementy środowiska uruchomieniowego EZD RP (maszyny wirtualne, środowisko Kubernetes, HA Proxy, kontenery aplikacji EZD RP) mogą być w przypadku awarii uruchomione lub zainstalowane ponownie i podłączone do odtworzonych już usług przechowywania danych. W celu przyspieszenia instalacji środowiska mogą być użyte narzędzia automatyzujące uruchomienie i instalację, np. szablony maszyn wirtualnych, Terraform, szablony maszyn wirtualnych, Ansible, HelmChart.