

Rekomendowane wymagania sprzętowe i programowe oraz scenariusze implementacji systemu EZD RP



EZD^{RP}

wersja 1.03 z dnia 11 października 2022 r.





Wymagania sprzętowe

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

1. Infrastruktura serwerowa przy wdrożeniach dla organizacji **do 2000 użytkowników**:
 - procesor lub procesory: 102 vCPU,
 - pamięć RAM: 1024 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 GPU*: 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).
2. Infrastruktura serwerowa przy wdrożeniach dla organizacji **do 300 użytkowników**:
 - procesor lub procesory: 32 vCPU,
 - pamięć RAM: 96 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 GPU*: 1 x Nvidia; Compute Capability 7.0+; RAM: 10 GB+; 8000+ rdzeni CUDA (np. GeForce RTX 3080 10 GB RAM).
3. Infrastruktura serwerowa przy wdrożeniach dla organizacji **do 50 użytkowników**:
 - procesor lub procesory: 16 vCPU,
 - pamięć RAM: 48 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).

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

- aplikacji EZD RP (wraz z instancją testową, edukacyjną i preprodukcyjną),
- relacyjnych i nierelacyjnych baz danych i systemów kolejkowych,
- klastrowego systemu plików,
- wirtualizacji infrastruktury (IaaS),
- środowiska uruchomieniowego kontenerów Kubernetes.

* zastosowanie kart GPU jest opcjonalne, rekomendowane w przypadku wykorzystywania niektórych funkcji modułu sztucznej inteligencji (SI) np. rozdzielania korespondencji, proponowania klas JRWA, znajdowania podobnych spraw (zastosowane modele SI wymagają douczenia z użyciem danych własnych instytucji użytkującej system EZD RP; włączenie wymienionych funkcji nie jest wymagane do prawidłowego działania systemu).

Zasoby rekomendowane dla wdrożeń EZD RP **nie uwzględniają** dodatkowych wymagań sprzętowych koniecznych do działania systemów bezpieczeństwa oraz urządzeń infrastruktury sieciowej. W przypadku wystąpienia niestandardowych obciążeń, typu masowa archiwizacja, czy równoległe przesyłanie dużej liczby plików, wymienione parametry sprzętowe mogą nie być wystarczające.

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, 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, klaster) i umożliwiały odtworzenie danych oraz przywrócenie działania usług zgodnie ze standardami obowiązującymi 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.

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 zostaną udostępnione w repozytorium. Przygotowany zostanie również kreator pozwalający na automatyzację instalacji 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, MS SQL lub Oracle oraz konto systemowe umożliwiające utworzenie i korzystanie z bazy danych,
- **nierelacyjne bazy danych i systemy kolejkowe** MongoDB, 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, edukacyjna, produkcyjna),
- **usługa wysyłki poczty SMTP** i konto umożliwiające korzystanie z tej usługi.

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 klastr lub jednowęzłowe środowisko można zainstalować i uruchomić na niemal dowolnym systemie operacyjnym niezależnie od stosowanego systemu wirtualizacji.

Popularne, szeroko stosowane i rekomendowane na chwilę obecną środowisko to OS Linux Ubuntu 20.x. W przypadku dużych instytucji kilka serwerów Ubuntu będzie zasobem wielowęzłowego klastra Kubernetes.

Dla instalacji obsługującej pojedynczą instytucję i maksymalnie 1000 użytkowników polecamy użycie odmiany środowiska Kubernetes **k3s** (Lightweight Kubernetes) w połączeniu z systemem Rancher, który pozwala na uruchamianie i zarządzanie środowiskiem Kubernetes. Więcej informacji na temat tych rozwiązań można znaleźć na stronach:

<https://kubernetes.io/pl/docs/concepts/overview/what-is-kubernetes/>

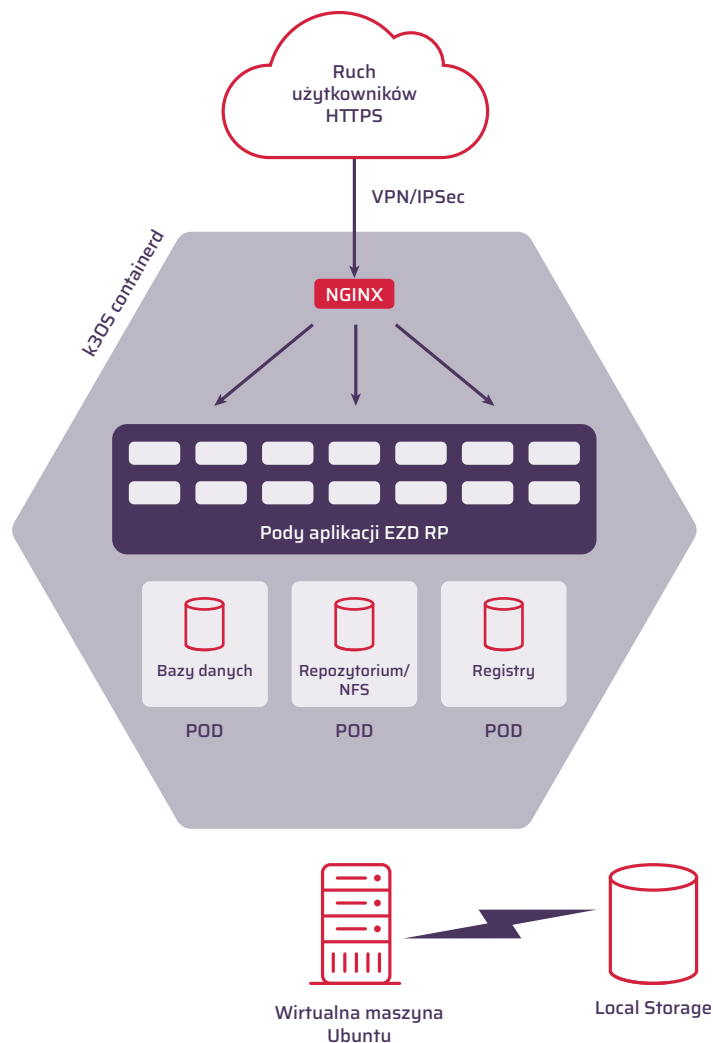
<https://rancher.com/docs/k3s/latest/en/>

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:

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

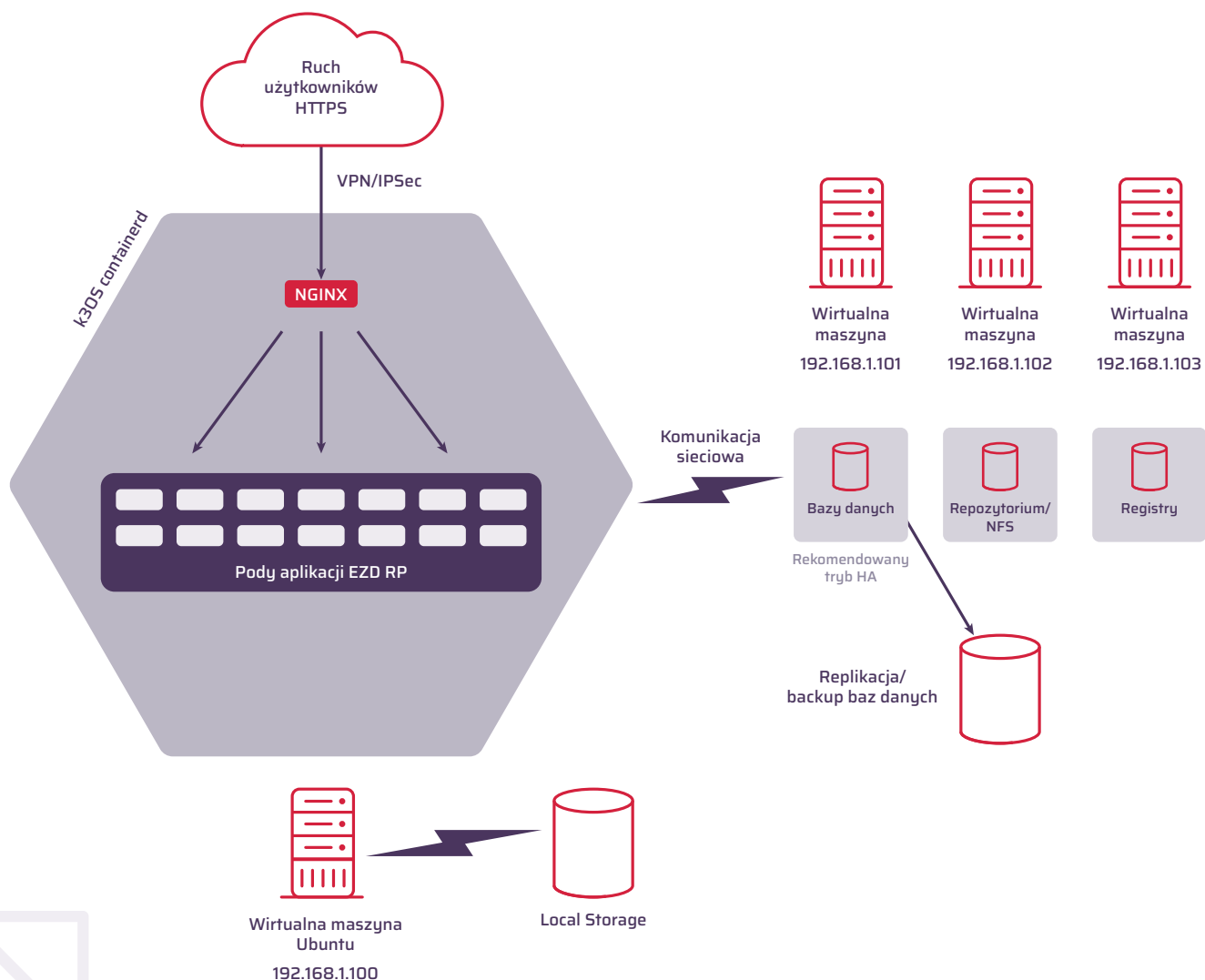
Rozwiązanie to polecane jest do zastosowań testowych, ewaluacyjnych, edukacyjnych, deweloperskich z uwagi na brak dostatecznych zabezpieczeń chroniących przed utratą danych.



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

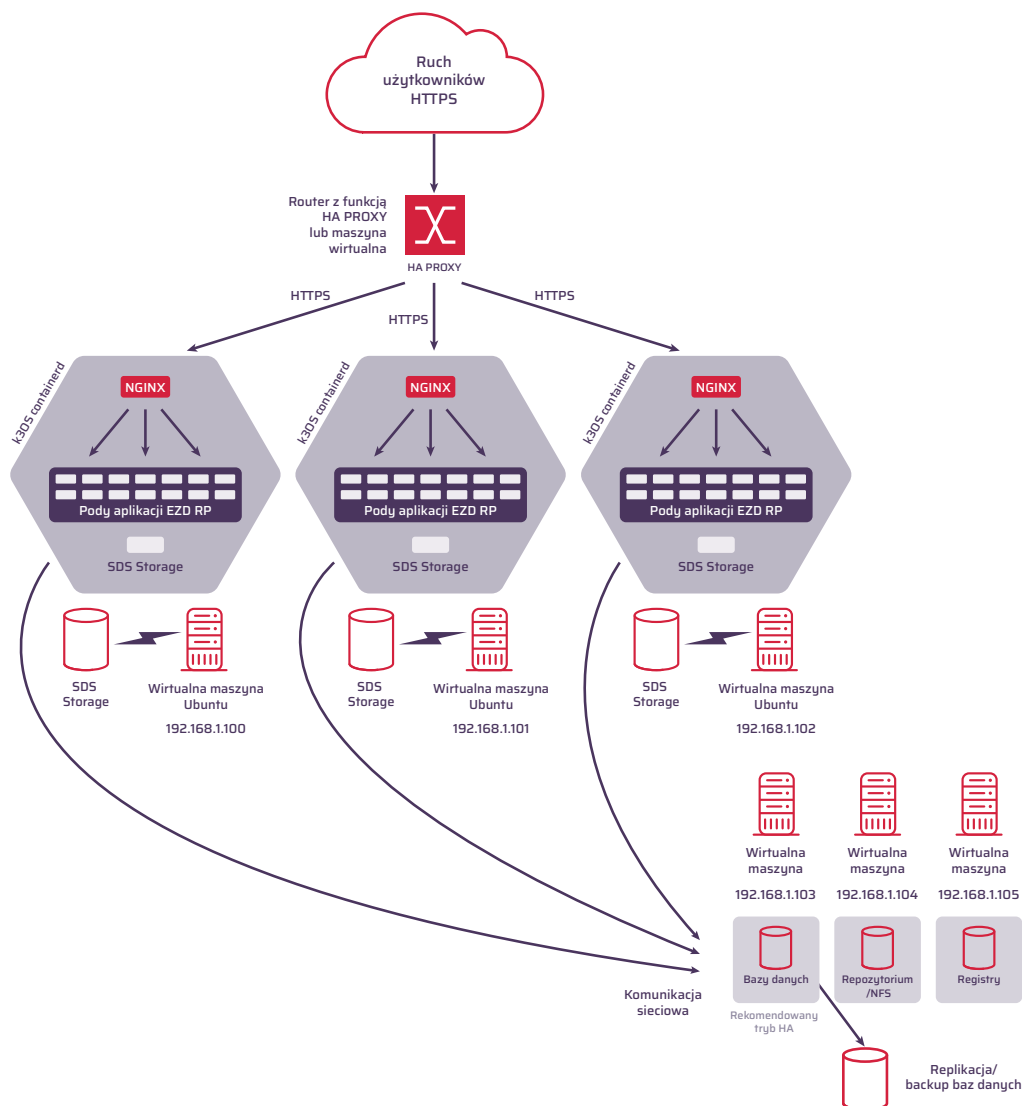
- ❑ rozwiązań klastrowych active-active lub active-passive,
- ❑ replikacji synchronicznej lub asynchronicznej danych i plików,
- ❑ określonych interwałów wykonywania kopii zapasowych i retencji backupu.



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

Wariant implementacji polecany 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.





NASK – Państwowy Instytut Badawczy
ul. Kolska 12, 01-045 Warszawa

Oddział NASK w Białymstoku
Łukowska 2, 15-373 Białystok

✉ ezdrp@nask.pl

🌐 ezdrp.gov.pl

NASK

govtech
Polska



Fundusze Europejskie
Polska Cyfrowa



Rzeczpospolita Polska



EZD^{RP}

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego

