Wysoka dostępność: usługa, która nie ma jednego serwera

Poziom: podstawy

Dotyczy: docker01, klaster hv01 i hv02, serwerownia Zielonej w Pamiątkowie

Czas: ~15 min

Ta strona otwiera książkę o tym, jak sprawić, żeby awaria jednego serwera nie zatrzymywała usługi. Zaczynamy od komputerów z lat siedemdziesiątych, które pierwsze wzięły to na serio, a kończymy na piątkowym popołudniu w Zielonej, gdy `docker01` przestał wstawać. Na końcu pokazujemy, jak ułożona jest książka.

Skróty użyte na tej stronie (4)

Skrót	Co znaczy
RPO	Recovery Point Objective: ile danych wolno stracić przy awarii
RTO	Recovery Time Objective: jak długo wolno nie działać przy awarii
VAX	Virtual Address eXtension: rodzina minikomputerów firmy Digital Equipment Corporation z lat 70. i 80.
VRRP	Virtual Router Redundancy Protocol: protokół, w którym kilka hostów uzgadnia, który z nich trzyma wspólny adres IP (opisany w RFC 5798)

Skąd się wzięło

> **Pojęcie: wysoka dostępność (HA, High Availability).** Takie zaprojektowanie usługi, żeby awaria pojedynczego elementu (serwera, dysku, zasilacza, kabla) nie zatrzymywała jej albo zatrzymywała na czas, który właściciel usługi z góry zaakceptował. Nie oznacza usługi, która nigdy się nie psuje, tylko taką, która psuje się w kontrolowany sposób.

Firma Tandem Computers powstała w 1974 roku, a pierwszy system z linii NonStop trafił do Citibanku w maju 1976 roku [1]. Pomysł był prosty: nie ma w nim jednego wspólnego elementu, bo każdy procesor ma własną pamięć, a procesory ustalają sprawy, wymieniając komunikaty. Po awarii jednego pracę przejmuje drugi [1].

W 1985 roku Jim Gray, jeden z inżynierów Tandema, opublikował raport „Why Do Computers Stop and What Can Be Done About It?” [2]. Z danych o ponad tysiącu lat pracy systemów wynikało, że sprzęt psuł się rzadziej niż ludzie i oprogramowanie, a największą część przestojów powodowały błędy administracji i błędy programów. To wniosek, do którego wracamy w tej książce: drugi serwer nie leczy złego `dnf update`.

Równolegle Digital Equipment Corporation zapowiedziało w maju 1983 roku VAXcluster, a pierwsza wersja z VAX/VMS 4.0 wyszła w 1984 roku [3]. Każdy komputer miał własny system, a dyski i blokady były wspólne dla całego klastra, więc użytkownik widział jeden duży komputer, a administrator mógł wyłączać jego części po kolei [3]. Ten wzorzec widzimy dziś w innych odmianach we Windows Failover Clustering, w Pacemakerze i w Kubernetesie.

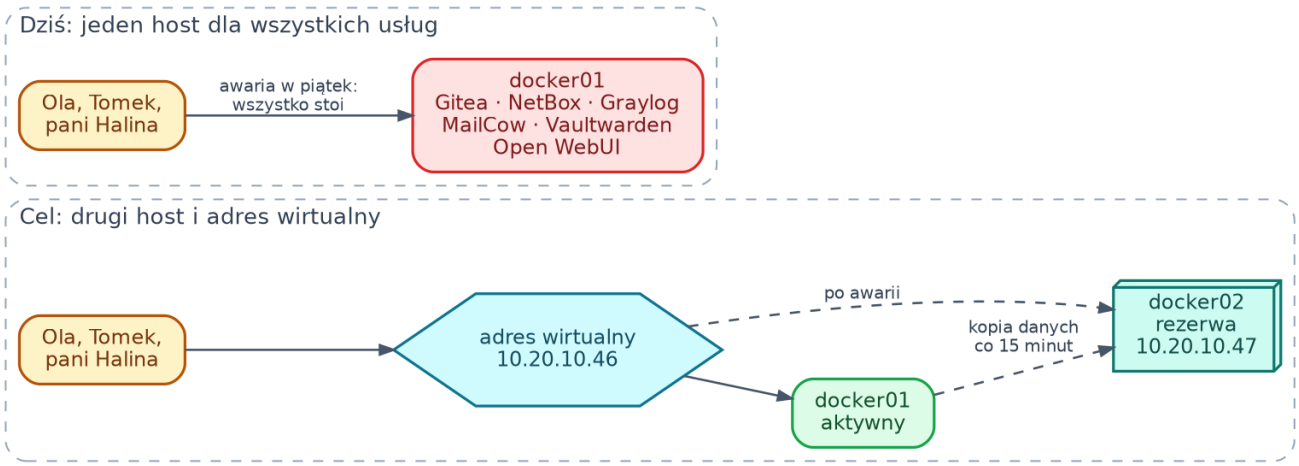
> **Pojęcie: pięć dziewiątek.** Dostępność 99,999 procent, czyli około 5 minut 15 sekund przestoju w roku. Nazwa wywodzi się z telekomunikacji, gdzie centrale telefoniczne miały działać prawie bez przerw [4].

Dokładnej daty powstania nazwy nie znamy. Najstarsze użycie „five nines” w sensie niezawodności, jakie znaleźliśmy, pochodzi z 1992 roku [5], więc nie łączymy jej z Tandemem ani z konkretnym rokiem lat siedemdziesiątych.

Jak to działa w jednym obrazku

Zamiast jednej maszyny, która robi wszystko, mamy dwie, które znają się nawzajem. Aplikacja ma jeden adres, pod którym użytkownicy ją znajdują, a ten adres przechodzi z maszyny na maszynę. Dane są na tyle świeże na obu, żeby przejęcie nie kosztowało nas pracy z całego dnia.

W Zielonej ten adres nazywamy adresem wirtualnym, a przenosi go protokół VRRP. Szczegóły są na kolejnych stronach, tu wystarczy zapamiętać trzy elementy: dwa hosty, adres, który skacze, i dane skopiowane z jednego na drugi.



Rysunek 1. Docker01 dziś i po zmianie. Dziś wszystkie aplikacje stoją na jednym hoście, a docelowo drugi host `docker02` ma świeżą kopię danych i przejmuje adres wirtualny `10.20.10.46`, gdy `docker01` przestaje odpowiadać. (kliknięcie otwiera pełny rozmiar)

Dlaczego Zielona

> **Z ŻYCIA WZIĘTE**
>
> Piątek, 15:20. Tomek zakończył na `docker01` aktualizację pakietów i zrestartował maszynę, żeby załadować nowe jądro. Klaster Hyper-V pokazuje, że maszyna jest uruchomiona, ale do konsoli wirtualnej nie dociera nic poza monitem rozruchowym, który prosi o naprawę systemu plików. Gitea, Vaultwarden i poczta z mailcow stoją, a Marek z magazynu w Swarzędzu dzwoni, że nie otwiera się jedno hasło potrzebne do drukarki etykiet.

Maszyna wirtualna dla Hyper-V żyje, więc klaster nie ma czego przenosić. Uszkodzony system plików gościa przechodzi razem z nią na drugi węzeł, jeśli zajdzie potrzeba. Odtworzenie z kopii restic, wzięcie haseł z kopii poza Vaultwardenem i start stosów po kolei zajmuje Tomkowi i Oli do późnego wieczoru, a poczta wraca dopiero po kolacji.

W poniedziałek pan Zbigniew pyta o jedno: skoro mamy klaster, to po co jeszcze drugi serwer. Odpowiedź ma dwie części. Klaster chroni przed utratą sprzętu, ale nie przed uszkodzonym systemem, złą aktualizacją czy błędem człowieka, a my chcemy, żeby w takim piątku usługa wróciła w kilkanaście minut, a nie w kilka godzin. Za tym idzie koszt, więc ta książka uczy też, kiedy drugi serwer się opłaca, a kiedy wystarczy dobra kopia.

Mapa tej książki

Książka idzie od teorii poziomów rezerwy przez przenoszenie kontenerów i replikację danych do klastrów Windows i Linux, równoważenia i orkiestracji, a kończy się macierzą wszystkich usług Zielonej.

Strona	Na jakie pytanie odpowiada
Wysoka dostępność: pojęcia, liczenie dostępności i poziomy rezerwy	Ile minut w roku to 99,9 procent, czym są RTO i RPO i jak dobieramy poziom do usługi
Przenoszenie kontenerów z docker01 na docker02: zimna i ciepła rezerwa, przełączenie	Jak z jednego hosta zrobić dwa i jak przełączyć usługę adresem wirtualnym
Replikacja danych między hostami: DRBD, GlusterFS, Ceph, ZFS i replikacja w aplikacji	Którą metodę kopiowania danych wybieramy i kiedy
Wzorce projektowania wysokiej dostępności: jak dobrać sposób do rodzaju usługi	Jak rodzaj usługi (plik, baza, usługa bezstanowa, maszyna wirtualna) wyznacza sposób utrzymania wysokiej dostępności i jakie zasady są wspólne dla wszystkich
Klaster Windows dla własnych usług: rola, zasób, zależności, kworum i testy przełączenia	Jak zrobić z własnej usługi na Windows rolę klastra, która przechodzi między węzłami
Klaster Windows od zera: budowa, sieć, świadek i S2D	Jak zbudować klaster Windows krok po kroku, od sieci po wspólny magazyn
Klaster Windows w utrzymaniu: aktualizacje CAU, monitoring i problemy	Jak utrzymujemy działający klaster Windows i co robimy, gdy coś w nim zawodzi
Pacemaker na Linuksie: własna usługa z adresem wirtualnym	Jak Pacemaker pilnuje usługi, przenosi ją na drugi węzeł i odcina węzeł, który przestał odpowiadać
Pacemaker w utrzymaniu: zarządzanie, testy i aktualizacje	Jak na co dzień zarządzamy klastrem Pacemaker, testujemy go i aktualizujemy
Proxmox VE: klaster trzech węzłów i HA	Jak zbudować klaster Proxmox VE i włączyć w nim wysoką dostępność maszyn
Proxmox VE w utrzymaniu: kopie PBS, aktualizacje, użytkownicy	Jak utrzymujemy klaster Proxmox VE, kopie i konta
Proxmox VE a Hyper-V z Failover Cluster: porównanie i przenosiny	Czym Proxmox VE różni się od Hyper-V z Failover Cluster i jak przenieść maszyny
Inne hypervisory z HA: oVirt, Oracle Linux Virtualization Manager i XCP-ng	Jak wysoką dostępność robią pozostałe platformy wirtualizacji
Równoważenie i adres wirtualny: keepalived i HAProxy na parze lb01 i lb02	Jak działają VRRP i kontrole stanu oraz jak ustawić je bez split brain
Docker Swarm: klaster z trzema menedżerami, na którym przenosimy kolorowy nginx	Co daje Swarm w 2026 roku i dlaczego nie jest naszym pierwszym wyborem

Strona	Na jakie pytanie odpowiada
k3s z trzema węzłami serwerowymi: etcd, adres wirtualny API i Longhorn	Jak zbudować klaster Kubernetesa, który przeżyje awarię węzła
Porównanie mechanizmów HA: Windows, Linux i hypervisory	Czym różnią się mechanizmy wysokiej dostępności na różnych systemach i kiedy który wybieramy
Macierz wysokiej dostępności usług Zielonej: poziom rezerwy każdej usługi, zależności i kolejność prac	Jaki poziom rezerwy ma każda usługa Zielonej, co od czego zależy i w jakiej kolejności to poprawiamy
Perforce i Lore w klastrze: cztery drogi, rekomendacja dla Zielonej i test przełączenia (dodatek: przykład praktyczny)	Jak zbudować klaster pod Perforce i Lore, żeby awaria węzła nie zatrzymywała zespołu
Perforce i Lore: porównanie dróg do wysokiej dostępności na Windowsie (dodatek: przykład praktyczny)	Które drogi do wysokiej dostępności Perforce i Lore porównaliśmy i którą wybieramy na start
Perforce i Lore na Linuksie: odpowiedniki dróg z Windowsa, pułapka wielkości liter i rekomendacja dla Zielonej (dodatek: przykład praktyczny)	Jak to samo zrobić na Linuksie i dlaczego wielkość liter jest pułapką przy przenosinach

Do kopii zapasowych i odtwarzania sięgamy do Kopia całego docker01: /docker/apps, /docker/compose i bazy w kontenerach, a o RPO i RTO baz piszemy w RPO i RTO: dwie liczby, od których zaczyna się każdy plan backupu.

Powiązane: Jak trzymamy aplikacje w Dockerze: konwencje Zielonej · S2D: wymiana dysku bez przestoju, gdy pula jest prawie pełna

Źródła:

1. en.wikipedia.org: Tandem Computers
2. Jim Gray, Why Do Computers Stop and What Can Be Done About It?, Tandem Technical Report 85.7, czerwiec 1985
3. en.wikipedia.org: VMScluster
4. networkworld.com: Five nines by the book
5. listserv.linguistlist.org: wątek listy ADS-L o pochodzeniu „five nines”

Revision #15
Created 2026-10-03 12:08:40 UTC by Tomasz Konieczny
Updated 2026-10-26 14:20:44 UTC by Tomasz Konieczny