

Red Hat Cluster Suite

Cena pewności

PIOTR BARANOWSKI



www.photocase.de

Co powoduje, że system IT jest efektywny? Odpowiedź jest banalna, ale nie wszyscy zdają sobie sprawę, jak ważna. Dostępność – bo o niej mowa – jest najważniejszym parametrem dla wielu menedżerów IT. Linux od około 2 lat zaczyna się oficjalnie pojawiać jako system w serwerowniach korporacji, zastępując inne implementacje usług strukturalnych. Najczęściej są to usługi proste, takie jak FTP, DHCP czy DNS.

Co jednak w sytuacji, gdy uruchamiamy usługi krytyczne? Pojawia się Solaris, AIX czy HPUX. Systemy te dowiodły swojej niezawodności – często restart „dużych” serwerów to wielkie wydarzenie, któremu przygląda się spora grupka administratorów. Jeśli chcielibyśmy, by Linux był systemem do zastosowań krytycznych, samo zaufanie wobec

oprogramowania nie wystarczy. Potrzebna jest technologia, która potrafi wyeliminować awarie lub sprowadzić je do odpowiednio nieszkodliwego poziomu. Taka technologia to właśnie klastry. W tym artykule postaram się przybliżyć wzorcową implementację realizowaną przez Red Hat Cluster Suite.

RHCS to implementacja klastrów High Availability, czyli technologii zapewniania wysokiej dostępności. W odróżnieniu od klastrów Mosix czy Beowulf, zwiększona wydajność nie jest tutaj celem głównym, choć nie została pominięta. Niepotrzebne są również modyfikacje aplikacji czy samego jądra.

Sam klastr składa się funkcjonalnie z trzech warstw. Pierwsza warstwa to dwa serwery realizujące równoważenie obciążenia (load balancing), zwane ruterami LVS. Druga warstwa to praktycznie dowolnie duża farma serwerów roboczych (worker server). Trzecią warstwą jest domena węzłów klastra wysokiej dostępności (HA). W jednej domenie może być aż osiem węzłów udostępniających zasadniczą usługę w trybie HA. Pierwsza warstwa ma za zadanie rozkładać według jednego z zadanych algorytmów obciążenie

na serwery warstwy drugiej. Jako że są dwa serwery realizujące równoważenie obciążenia, w przypadku awarii rutera podstawowego jego działanie przejmuje ruter zapasowy. Co ważne, zastosowana tutaj technologia podtrzymuje sesje klientów. Dzięki takiemu układowi zyskujemy równomierne wykorzystanie systemów roboczych warstwy drugiej, a dzięki redundancji warstwa pierwsza jest dostępna cały czas z 6-sekundowym czasem przełączenia między ruterami LVS.

Dodatkową funkcjonalnością ruterów LVS jest monitorowanie usług udostępnianych przez systemy warstwy drugiej. Takie podejście eliminuje wadliwe serwery i nie dopuszcza do sytuacji, w której ruter LVS przekieruje zapytanie klienta do systemu, który nie działa. Monitorowanie jest dosyć wyrafinowane i sprawdza faktyczny stan aplikacji. Ciekawy jest fakt, że administrator może zdefiniować własne metody weryfikacji serwisów. Przykładowo, serwery HTTP są sprawdzane poprzez wykonanie HTTP GET na zadanym adresie; sprawdzenie polega na określeniu, czy treść zwrócona przez serwer jest zgodna z wzorcem. Test serwera SQL może

AUTOR

Piotr Baranowski – certyfikowany instruktor/egzaminator RHCX/RHCI w Altcom Akademii S.A. Specjalista ds. bezpieczeństwa systemów IT, w szczególności systemów uniksowych. Wykładowca w Altcom Akademii od 2001 roku.

polegać na wykonaniu prostego SELECT na bazie i sprawdzeniu wyniku. Jeżeli takie testy są zbyt skomplikowane, zawsze można się odwołać do nieśmiertelnego ICMP ECHO.

Warstwa druga to serwery realizujące usługi, z którymi łączą się klienci. Często są to Apache, IIS, Samba, NFS, MySQL czy inna dowolna aplikacja. Wstawiając kolejne systemy w warstwie drugiej administrator może określić wagę serwera, opierając się na ilości pamięci czy typie procesora, dzięki czemu mocniejsze maszyny będą otrzymywać więcej zleceń niż słabsze, starsze serwery. Możliwe jest monitorowanie systemów warstwy drugiej i automatyczne strojenie równoważenia obciążenia w oparciu o średnie obciążenie serwerów roboczych. Dzięki temu zasoby systemów warstwy drugiej są wykorzystywane w bardzo efektywny sposób.

Warstwa trzecia to serwery dostarczające najczęściej aplikacji krytyczne, takie jak baza danych czy serwery aplikacyjne. Węzły klastra w domenie awaryjnej posiadają wspólną konfigurację, dzięki czemu zmiany w jednym węźle są widoczne od razu we wszystkich węzłach. Wszystkie węzły muszą mieć współdzielony dostęp do macierzy dyskowej i od razu trzeba zaznaczyć, iż nie jest to technologia dla każdego. Najprostszą implementacją jest przypięcie węzłów do jednego dysku SCSI dwoma oddzielnymi kontrolerami. Nie jest to jednak wspierana konfiguracja i ma ograniczenie ilości węzłów do dwóch. Prawdziwie współdzielona macierz powinna być oparta o technologię Fiber Channel. Wykorzystując przełącznik FC, możliwe jest bezproblemowe przyłączenie kilkunastu systemów do jednego zasobu. Z punktu widzenia węzłów, macierz FC jest współdzieloną przestrzenią, na której wydzielone woluminy pełnią rolę medium wymiany konfiguracji, pozwalając na

wzajemne monitorowanie się węzłów. Przykładowo, jednym z mechanizmów sprawdzania integralności domeny awaryjnej jest cykliczne wstawianie znacznika czasowego przez wszystkie węzły we wspólnym woluminie. Gdy jeden z węzłów nie wstawi znacznika,

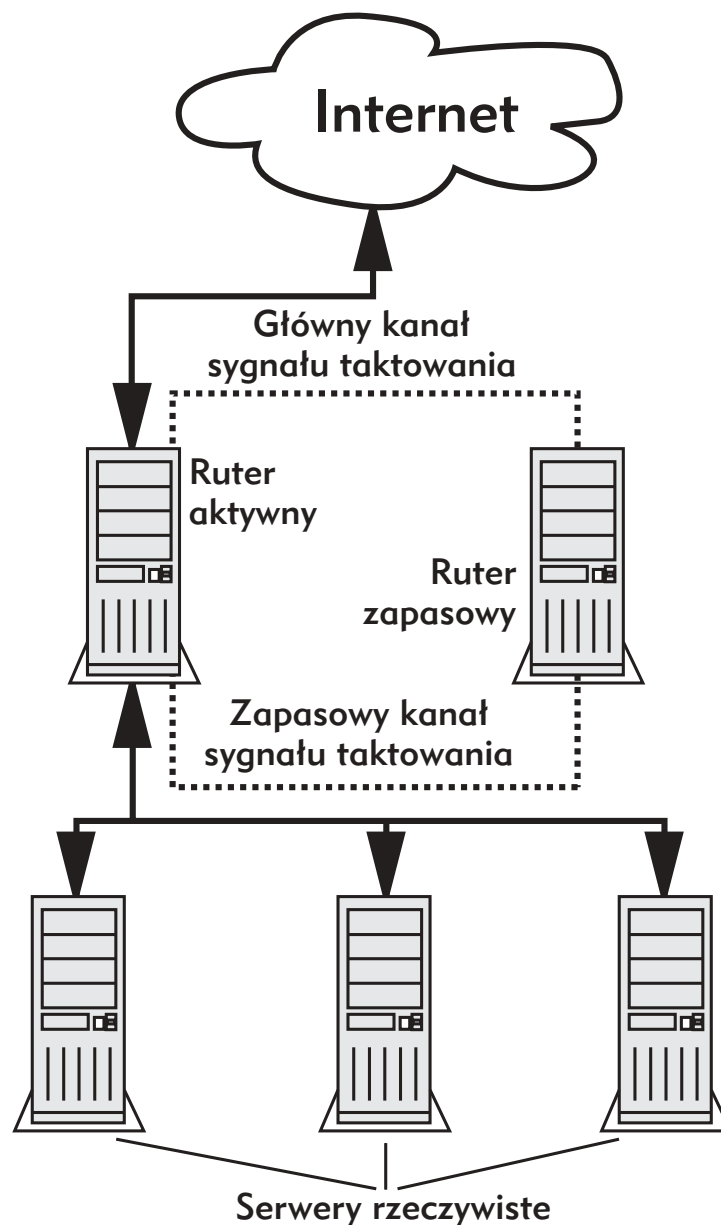
ją stanu swoich sąsiadów. Węzły zgodnie z konfiguracją serwują zadaną aplikację, a gdy wystąpi awaria, przejmują funkcjonalność uszkodzonego węzła. Czas przełączenia jest tak dobrany, aby przerwa w działaniu aplikacji była niezauważalna.

Przedstawiając klastry, najlepiej posłużyć się przykładami żywych implementacji. Dwa serwery pracują jako routery LVS. W warstwie drugiej trzy serwery Apache lub WebSphere łączą się do domeny węzłów serwujących im zawartość bazy DB2 lub Oracle. Taki model daje nadzwyczajną skalowalność. W warstwie drugiej możemy bowiem umieścić praktycznie dowolną ilość serwerów roboczych. Inny model przykładowy to dwa routery LVS, trzy serwery Courier-MTA w warstwie drugiej oraz trzy węzły klastra serwujące macierz dyskową po NFS jako przestrzeń na katalogi domowe użytkowników. Tutaj nie ma praktycznie limitu obsługiwanych klientów.

Przykłady można mnożyć, gdyż w takiej konfiguracji można uruchomić dowolną aplikację. Mechanizmy wykorzystane w Red Hat Cluster Suite są przezroczyste dla aplikacji i dla użytkowników tych aplikacji. Jednocześnie są na tyle skuteczne, aby zapewnić działanie aplikacji na poziomie magicznych pięciu dziewiątek, co jest marzeniem niejednego menedżera IT. Naturalnie nie

jest to technologia dla każdego, ale tam, gdzie pojęcie dostępności jest istotne, koszt zakupu macierzy FC czy zakupu oprogramowania nie jest żadną przeszkodą. Prezentacja opisująca RHCS oraz jego możliwości dostępna jest pod adresem: <http://www.altkom.com.pl/archiwum/konferencje/prezentacje/Linux%20w%20biznesie%20-%20warsztaty.pdf>

Podczas najbliższej konferencji LinuxPro będę prezentował technologie klastrowe w laboratorium składającym się z ośmiu serwerów. Zapraszam serdecznie. ■



ka, oznacza to, że wystąpiła awaria i taki węzeł jest wyłączany z konfiguracji, aby zapobiec ewentualnym szkodom. Węzły monitorują się kilkoma metodami, aby uniknąć fałszywych alarmów przez uszkodzony kanał testujący. Metod jest kilka: test sieciowy, test znacznika czasowego na wspólnej macierzy czy test wykonywany po kablu szeregowym.

Wszystkie te zabiegi mają na celu eliminację niejednoznaczności stanu systemów. Przy poprawnie skonfigurowanym klastrze nie może się zdarzyć sytuacja, gdy węzły nie zna-