ZESZYTY NAUKOWE POLITECHNIKI RZESZOWSKIEJ 294, Elektrotechnika 35
RUTJEE, z. 35 (2/2016), kwiecień-czerwiec 2016, s. 57-65

Dariusz KOWALSKI[1]
Paweł DYMORA[2]
Mirosław MAZUREK[3]

KLASTRY PRACY AWARYJNEJ W ŚRODOWISKU MICROSOFT WINDOWS SERVER 2012

W artykule poruszono temat klastrów pracy awaryjnej w środowisku Microsoft Windows Server 2012. Klastry tego typu działają w oparciu o tzw. elementy quorum (kworum). W Windows Server elementem quorum może zostać węzeł, dysk „świadek" lub plik współdzielony „świadek". Głównym celem artykułu jest porównanie czasów niedostępności usług świadczonych przez wymienione modele klastrów, w przypadku awarii elementów klastra, świadczących wybrane usługi. Analizie poddano architektury: Node Majority (elementy kworum w postaci węzłów klastra), Node and Disk Majority (elementy kworum w postaci węzłów klastra oraz dysku „świadka"), Node and File Share Majority (elementy kworum w postaci węzłów klastra oraz współdzielonego zasobu) oraz No Majority: Disk Only (element kworum w postaci dysku „świadka").

Słowa kluczowe: failover, HA, wysoka dostępność, serwer, klaster

1. Wprowadzenie

Klastry pracy awaryjnej, zwane także klastrami wysokiej dostępności (ang. High-Availability Clusters – HA) to grupa serwerów pracujących razem, utworzona w celu zapewnienia wysokiej dostępności oraz udostępnianych przez nią aplikacji i usług, widziana przez urządzenia klienckie jako jeden system. W sytuacji gdy poszczególny węzeł (serwer) klastra ulega awarii, jego rola zostaje przejęta przez inny serwer pracujący w klastrze. Proces ten nazwany został

[1] Autor do korespondencji: Dariusz Kowalski, Politechnika Rzeszowska, darkowalski@windowslive.com
[2] Paweł Dymora, Politechnika Rzeszowska, Katedra Energoelektroniki, Elektroenergetyki i Systemów Złożonych, pawel.dymora@prz.edu.pl
[3] Mirosław Mazurek, Politechnika Rzeszowska, Katedra Energoelektroniki, Elektroenergetyki i Systemów Złożonych, miroslaw.mazurek@prz.edu.pl

trybem failover. Celem takiego działania jest osiągnięcie jak najkrótszych czasów przestojów (czasów niedostępności poszczególnej aplikacji bądź usługi).

Usługa klastra pracy awaryjnej dostępna jest w systemach z rodziny Windows Server począwszy od wersji 2000. W Windows Server 2012 usługa ta konfigurowana jest jako funkcja i umożliwia utworzenie wysoko dostępnego magazynu danych wykorzystywanego np. przez maszyny wirtualne Hyper-V. Umożliwia utrzymanie szeregu aplikacji i usług krytycznych przedsiębiorstwa, takich jak m.in. serwer bazy danych, serwer poczty, serwer wydruku, serwer plików, serwer DHCP itd. WS 2012R2 umożliwia utworzenie klastra składającego się maksymalnie z 64 węzłów lub 8000 maszyn wirtualnych (w WS 2008R2 z 16 węzłów lub 1000 maszyn wirtualnych) i ta wartość jest wspólna dla wszystkich wersji WS 2012R2 oraz WS 2012 (Standard, Datacenter).

2. Klastry pracy awaryjnej w środowisku MS Windows Server 2012

Aby klaster pracy awaryjnej pracował poprawnie wymaganym jest osiągnięcie następujących założeń:

- Poszczególne węzły klastra muszą spełniać wymagania stawiane przez system operacyjny Windows Server 2012. Zalecane jest instalowanie identycznej konfiguracji sprzętowej w każdym z węzłów. Architektury procesorów poszczególnych węzłów klastra muszą być ze sobą zgodne tzn., iż nie można łączyć ze sobą węzłów pracujących pod architekturą AMD z węzłami pracującymi pod architekturą Intel;
- Poszczególne węzły klastra muszą pracować pod tą samą wersją systemu operacyjnego Windows (nie można np. łączyć w klaster węzłów pracujących pod kontrolą WS 2012R2 Standard oraz WS 2012R2 Datacenter). Zalecanym jest również, aby węzły te posiadały zainstalowane podobne wersje uaktualnień;
- W przypadku, gdy rozwiązanie wykorzystuje magazyn udostępniony (wolumin CSV), musi on być podłączony do węzłów tego klastra. Woluminy CSV (ang. Cluster Shared Volumes) umożliwiają poszczególnym węzłom klastra jednoczesny dostęp (zapis i odczyt) do jednostki dyskowej LUN (ang. Logical Unit Number), obsługiwanej w WS 2012R2 jako wolumin NTFS lub ReFS;
- Do połączeń węzłów klastra z magazynem danych można wykorzystać następujące interfejsy: Internet SCSI (iSCSI), SAS (ang. Serial Attached SCSI), FC (ang. Fibre Channel), FCoE (ang. Fibre Channel over Internet). Jeśli do połączeń z magazynem wykorzystano iSCSI, każdy z węzłów powinien posiadać co najmniej jeden interfejs wykorzystywany do tego celu. Taki interfejs nie powinien przenosić innego ruchu sieciowego niż ten związany z magazynem danych. Dla lepszej wydajności, zalecanym jest wykorzystywanie przynajmniej 2 interfejsów GigE. W przypadku wykorzystywania kilku połączeń do magazynu danych, należy pamiętać o włączeniu Multipathingu IO;

- Każdy z węzłów powinien mieć zainstalowane identyczne karty sieciowe (obsługujące ten sam protokół IP, taką samą prędkość, transmisję duplex oraz umożliwiające taką samą kontrolę przepływów);

- Zaleca się, aby każdy węzeł klastra posiadał przynajmniej 3 interfejsy sieciowe: jeden dla połączenia z magazynem danych, jeden do połączeń z innymi węzłami klastra oraz jeden do połączeń z zewnętrzną siecią;

- Każdy z serwerów klastra musi być członkiem tej samej domeny Active Directory oraz powinien wykorzystywać ten sam serwer DNS;

- Poszczególne połączenia sieciowe pomiędzy węzłami klastra powinny być redundantne – w celu minimalizacji skutków awarii danych łącz;

- Całościowa konfiguracja ustawień klastra (konfiguracja węzłów, sieci i magazynu) musi przejść pozytywnie wszystkie testy przeprowadzone w „Kreatorze weryfikacji konfiguracji".

Klastry pracy awaryjnej w systemach z rodziny MS Windows Server działają w oparciu o tzw. elementy quorum. Quorum w klastrach pracy awaryjnej rozumiany jest jako minimalny zbiór elementów klastra, które muszą pozostawać aktywne, aby klaster mógł działać. Dzięki niemu poszczególne węzły klastra mogą za pomocą jednego zapytania sprawdzić czy klaster może kontynuować swoją pracę. W Windows Server 2012 elementem quorum może zostać węzeł, dysk „świadek" lub plik współdzielony „świadek". Każdy element pełniący funkcję quorum (za wyjątkiem współdzielonego pliku „świadka") przechowuje kopie konfiguracji klastra. Kopie konfiguracji klastra, przechowywane na kilku elementach quorum są na bieżąco synchronizowane przez działającą usługę klastra. Konieczność stosowania elementów quorum ukazuje się w przypadku problemów z komunikacją między węzłami klastra. Część węzłów może komunikować się między sobą poprzez funkcjonującą część sieci, jednocześnie nie będąc w stanie komunikować się z węzłami pracującymi w drugiej odrębnej części. Fakt ten prowadzi do sytuacji, w której część węzłów musi przestać pracować jako węzły klastra. Aby uniknąć problemów pojawiających się podczas podzielenia klastra, oprogramowanie wymaga od powstałych części klastra, aby wykorzystywały algorytm głosowania (ang. voting algorithm), w celu określenia w danym czasie czy posiadają dostęp do quorum. Ponieważ dany klaster posiada specyficzny zestaw węzłów oraz specyficzną konfigurację quorum, klaster taki wie ile węzłów nadal „głosuje" (kontynuuje swoją pracę). Jeśli liczba ta spadnie poniżej większości, klaster przestaje działać. Węzły nadal nasłuchują obecności pozostałych węzłów, w przypadku, gdy te pojawią się ponownie w sieci, jednak klaster nie odbuduje się ponownie dopóki warunki quorum nie zostaną spełnione. Przykładowo jeśli klaster w konfiguracji Node Majority składa się z 5 węzłów i węzły 1, 2, 3 nie będą mogły się skomunikować z węzłami 4, 5 to w takiej sytuacji węzły 4 oraz 5 jako mniejszość przestają działać w klastrze. Dalej jeśli węzeł 3 również straci komunikacje z węzłami 1 oraz 2 to pozostała część (węzły 1, 2) będzie stanowić mniejszość i cały klaster przestanie działać. W takiej sytuacji

wszystkie węzły przestaną działać w klastrze, jednak będą oczekiwać na wznowienie komunikacji i w sytuacji, gdy sieć ponownie zacznie pracować, klaster może się odbudować i zacząć pracować od nowa.

W systemach z rodziny MS Windows Server 2012 istnieją 4 modele klastra wykorzystujące quorum:

- *Node Majority* – w modelu tym klaster pracuje do momentu, gdy liczba uszkodzonych węzłów jest mniejsza niż działających. Przykładowo, jeśli klaster składa się z 5 węzłów, trzy z nich muszą pozostawać aktywne aby klaster działał prawidłowo. W sytuacji, gdy więcej niż dwa węzły ulegną uszkodzeniu, cały klaster przestanie pracować;

- *Node and Disk Majority* – model wykorzystujący tzw. dysk Quorum (dysk „świadek"). W konfiguracji tej liczba pracujących węzłów nie może być mniejsza niż dwa. Model ten zalecany jest w sytuacjach, gdzie wszystkie węzły klastra mogą korzystać z tych samych zasobów (np. z tej samej macierzy dyskowej). Klaster pracuje prawidłowo, dopóki większość elementów quorum (w tym serwery oraz dysk współdzielony) pozostaje aktywne.

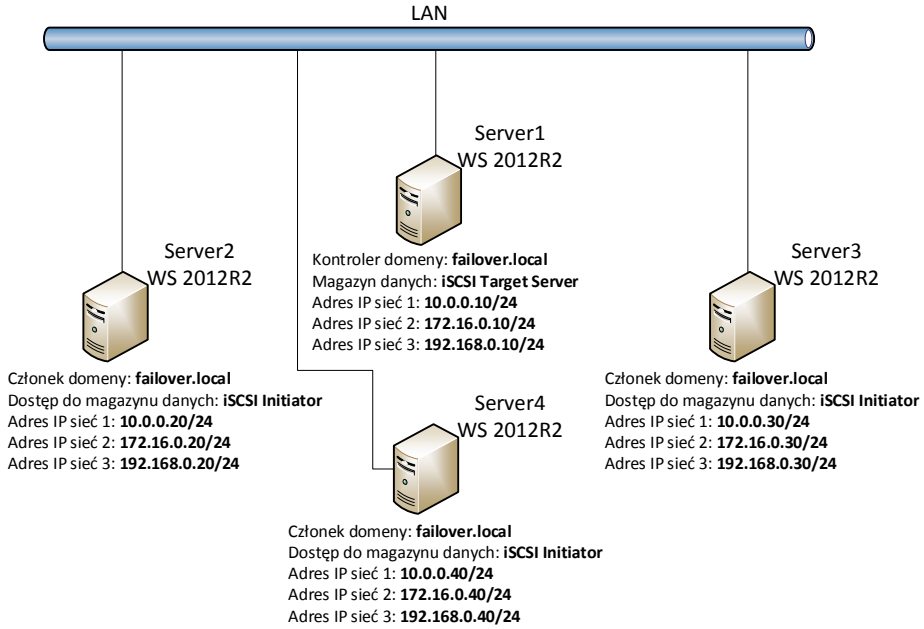
- *Node and File Share Majority* – model bazujący na modelu Node and Disk Majority. Od poprzednika odróżnia go sposób zapisu quorum – zamiast dysku „świadka" wykorzystywany jest udział sieciowy, na którym zostaje ono zapisane. Do zalet modelu należy to, iż pracuje podobnie jak model Node and Disk Majority. Konfiguracja ta zalecana jest w głównej mierze dla rozproszonych geograficznie klastrów.

- *No Majority: Disk Only* – model rzadko wykorzystywany, zaprojektowany głównie w celu przeprowadzenia testów aplikacji i procesów w Windows Server 2012. Klaster zbudowany w oparciu o ten model działa do momentu, w którym dostęp do dysku quorum ulegnie awarii. Nie ma tutaj ograniczenia dotyczącego liczby pozostałych aktywnych węzłów. Konfiguracja tego typu nie jest zalecana w środowisku produkcyjnym.

3. Model klastra Failover

W artykule przetestowano działanie 4 dostępnych modeli klastra pracy awaryjnej w MS Windows Server 2012. Wszystkie testy polegały na sprawdzaniu dostępności węzła (poleceniem `ping`) o adresie IP przypisanym do uruchomionej usługi klastra (serwera DHCP), a następnie zasymulowaniu awarii węzła klastra odpowiedzialnego w danej chwili za udostępnianie tejże usługi. Dla takiego scenariusza dokonano pomiarów czasu niedostępności usługi. Ponadto dla każdego z badanych modeli klastrów wykonanych zostało co najmniej 10 symulacji awarii.

Na Rys. 1 przedstawiono schemat testowanej konfiguracji. Serwer1 jest kontrolerem domeny failover.local. Udostępnia on poprzez oprogramowanie iSCSI Target Server magazyn danych dla serwerów Serwer2, Serwer3 oraz Serwer4. Serwery Serwer2, Serwer3 oraz Serwer4 są serwerami członkowskimi domeny failover.local. Poprzez oprogramowanie iSCSI Initiator korzystają z udostępnionego na Serwer1 magazynu danych. Na serwerach Serwer2, Serwer3 oraz Serwer4 została uruchomiona rola klastra pracy awaryjnej, która następnie poprzez wybraną usługę tego klastra (DHCP) została przetestowana. Sieć1 to sieć wykorzystywana do komunikacji z magazynem danych, sieć2 to sieć wykorzystywana do komunikacji pomiędzy węzłami klastra (tzw. Heartbeat), natomiast sieć3 to sieć wykorzystywana do komunikacji z zewnętrzną siecią. Wszystkie serwery przedstawione w artykule działają pod kontrolą systemu MS Windows Server 2012R2 Datacenter.



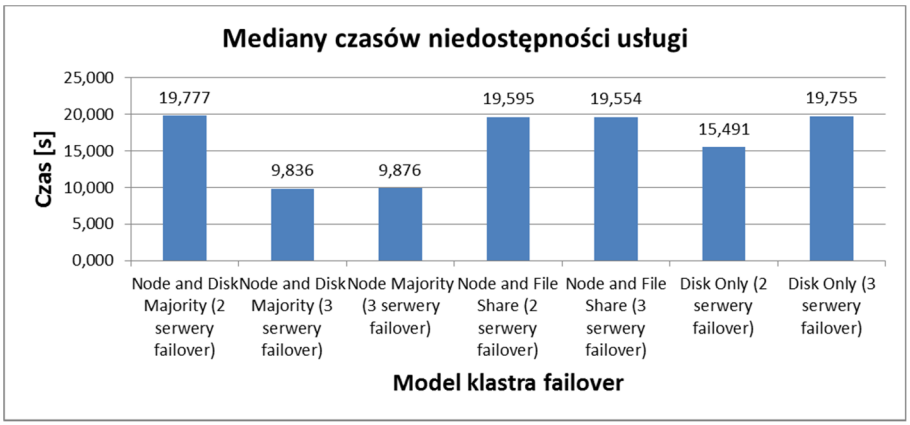
Rys. 1. Schemat testowanej sieci

Fig. 1. Diagram of tested network

4. Analiza niedostępności usług badanych modeli klastra

W niniejszym rozdziale zaprezentowano analizę czasów niedostępności usług dla analizowanych modeli klastra pracy awaryjnej. Dla wszystkich badanych przypadków wyznaczono mediany czasów niedostępności usług klastra

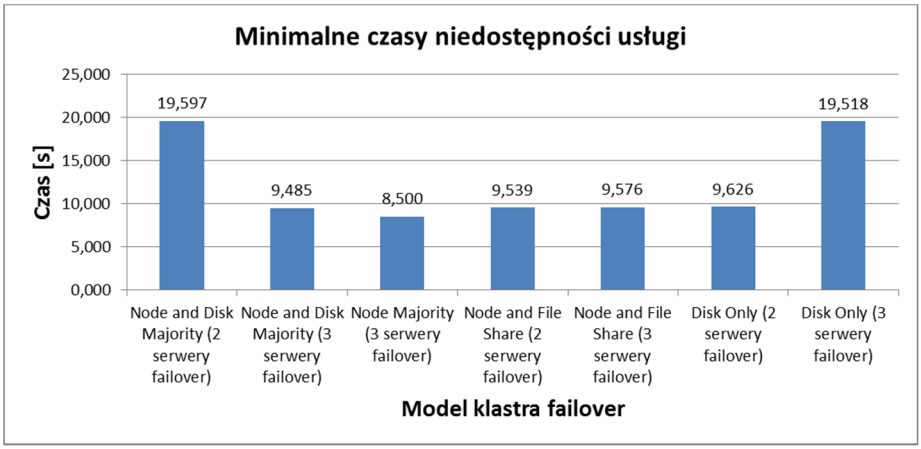
(Rys. 2), minimalny czas niedostępności usługi (Rys. 3), maksymalny czas niedostępności usług klastra (Rys. 4) oraz średniej niedostępności usług klastra dla wszystkich badanych przypadków (Rys. 5).



Rys. 2. Mediana czasów niedostępności usług klastra

Fig. 2. The median duration of unavailability of services cluster

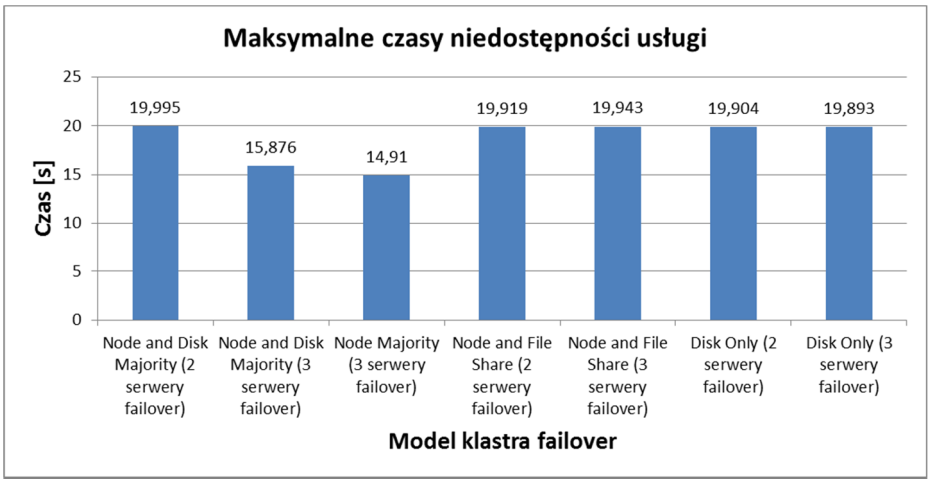
Rys. 3 prezentuje minimalny czas niedostępności usług klastra dla wszystkich badanych przypadków zrealizowany spośród określonej liczby prób.



Rys. 3. Minimalny czas niedostępności usług klastra

Fig. 3. The minimum duration of unavailability of services cluster

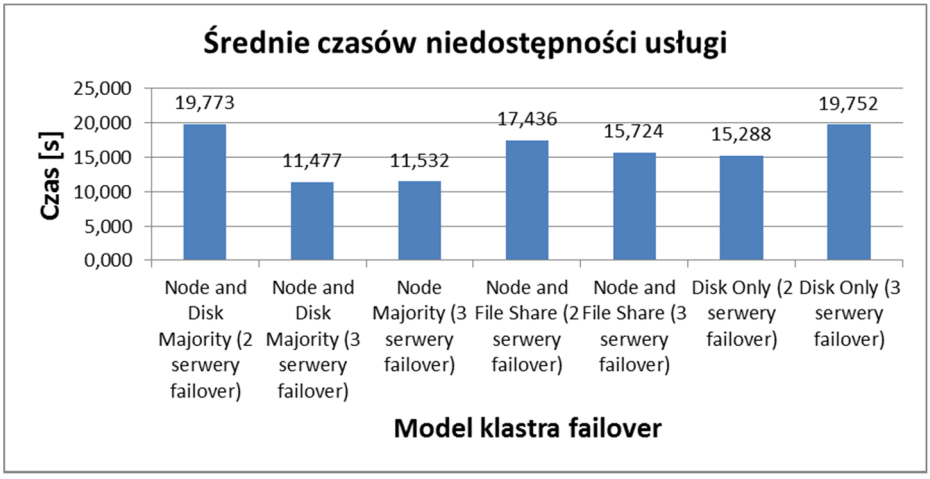
Na Rys. 4 przedstawiono histogram maksymalnych niedostępności usług klastra dla wszystkich badanych przypadków.



Rys. 4. Maksymalne czasy niedostępności usług klastra

Fig. 4. The maximum duration of unavailability of services cluster

Na Rys. 5 przedstawiono histogram średniej niedostępności usług klastra dla wszystkich badanych przypadków.



Rys. 5. Średnie czasy niedostępności usług klastra

Fig. 5. The average duration of unavailability of services cluster

Podczas analizy histogramu przedstawionego na Rys. 3 widać, iż minimalny czas niedostępności danej usługi w środowisku MS Windows Server 2012R2 (dla każdego z badanych modeli) jest nie mniejszy niż 8,500 sekundy. Maksymalny czas niedostępności badanej usługi, co pokazano na Rys. 4 wynosi 19,995 sekundy. Analiza histogramów przedstawionych na Rys. 2 oraz 5 wskazuje podobieństwa w czasach niedostępności usług między modelami klastrów:

- *Node and Disk Majority* w konfiguracji 2 serwerów failover (mediana: 19,777s, średnia: 19,773s);
- *Node and File Share Majority* w konfiguracji 2 serwerów failover (mediana: 19,595s, średnia: 17,436s);
- *Node and File Share Majority* w konfiguracji 3 serwerów failover (mediana: 19,554s, średnia: 15,724s);
- *Disk Only* w konfiguracji 3 serwerów failover (mediana: 19,755s, średnia: 19,752s);

oraz między modelami:

- *Node and Disk Majority* w konfiguracji 3 serwerów failover (mediana: 9,836s, średnia: 11,477s);
- *Node Majority* w konfiguracji 3 serwerów failover (mediana: 9,876s, średnia: 11,532s);

Histogramy przedstawione na Rys. 2 oraz 5 pokazują, iż modele klastra wykorzystujące dysk lub udział sieciowy, które mają decydujący głos podczas „głosowania" w trakcie awarii jednego z węzłów klastra są wyraźnie mniej sprawne pod względem HA, aniżeli modele, w których decydujący głos podczas awarii w klastrze mają węzły mające możliwość świadczenia danej usługi, udostępnianej przez klaster.

5. Podsumowanie

Problem zapewniania wysokiej dostępności usług oraz wysokiego poziomu niezawodności i odporności na uszkodzenia całego środowiska informatycznego nazywanego często infrastrukturą krytyczną jest aktualnym tematem badawczym jak również ważnym zagadnieniem z jakim spotykają się na co dzień administratorzy systemów tej klasy. Wybór odpowiedniej technologii HA umożliwia osiągnięcie minimalnych czasów przestoju usługi w przypadku awarii oraz determinuje poczucie bezpieczeństwa i niezawodności usługi przez klientów.

W artykule przetestowano działanie podstawowych modeli klastra pracy awaryjnej w środowisku MS Windows Server 2012R2. Przesymulowano kilka

scenariuszy awarii jednego z węzłów klastra, który był odpowiedzialny w danej chwili za udostępnianie usług i dokonano pomiarów czasu ich niedostępności. Z przeprowadzonych badań wynika, iż najkorzystniejszymi pod względem HA modelami pracy awaryjnej w środowisku MS Windows 2012R2 są modele: Node and Disk Majority (z co najmniej 3 serwerami failover) oraz Node Majority.

Literatura

[1] Stanek W. R.: Vademecum Administratora Windows Server 2012 R2, APN Promise, 2014.

[2] Mackin J. C., Thomas O.: Egzamin 70-412 Konfigurowanie zaawansowanych usług Windows Server 2012 R2, Microsoft Press, 2014, s. 19-58.

[3] Wołk K., Biblia Windows Server 2012 Podręcznik administratora, Psychoskok, 2012.

[4] https://technet.microsoft.com/en-us/library/cc731739.aspx [dostęp: 5 marzec 2016 r.].

[5] http://resources.intenseschool.com/windows-server-2012-failover-clustering-part-2/ [dostęp: 5 marzec 2016 r.].

FAILOVER CLUSTERING IN MICROSOFT WINDOWS SERVER 2012

S u m m a r y

The article is all about failover clusters in Microsoft Windows Server 2012. Failover clusters called as well High-Availability Clusters creates a group of servers working together to provide high availability of provided by the cluster services and applications. Client devices see the cluster as a single system. Clusters of this type - in the family of Microsoft Windows Server - are based on element quorum. Quorum in failover clusters is considered as a parts of cluster witch has to be active to allow the cluster to work. Thanks to this each individual node of the cluster can check - by the single query - if the whole cluster can be active. In the MS Windows Server Environment component of a quorum may be for example: node, disk quorum, shared file quorum. The clusters models discussed in the article - provided by MS Windows Server 2012 - include: Node Majority, Node and Disk Majority, Node and File Share Majority and No Majority: Disk Only. The main purpose of the article is to compare the unavailability time of the services provided by these models of clustering in the case of cluster component failure.

Keywords: failover, HA, high availability, server, cluster

DOI: 10.7862/re.2016.10

Tekst złożono w redakcji: maj 2016
Przyjęto do druku: czerwiec 2016