Tomasz Smajdor
Continuous Processing Technologies
Stratus EMEA Distributor
Mobile. +48 500 040 601
e-mail: tsmajdor@cptech.pl

Problemy, którym można zapobiec: dlaczego użytkownicy technologii wirtualizacyjnych zapominają o dostępności?

Autor: Dan Kusnetzky, główny analityk
Sponsor: Stratus Technologies Inc.
Tłumaczenie: CPT

WSTĘP

Wiele firm zamierzających zwirtualizować swoje środowiska informatyczne zwraca znacznie większą uwagę na wydajność, konsolidację lub optymalne wykorzystanie serwerów niż na zapobieganie awariom. O dostępności i niezawodności systemu zaczynają myśleć dopiero wtedy, gdy pojawiają się problemy.

W niniejszym artykule zostaną omówione następujące zagadnienia:

- Czym jest wirtualizacja?
- Tworzenie środowiska o wysokiej dostępności
- Metody zwiększania dostępności oparte na sprzęcie i oprogramowaniu
- Odporność na awarie a dostępność
- Jaką dostępność można uznać za wystarczającą?
- Jak podejmować decyzje o zakupie rozwiązań?

Funkcje wysokiej dostępności, podobnie jak zarządzania i bezpieczeństwa, są najwydajniejsze, jeżeli zostały „wtopione" w architekturę środowiska już w trakcie jej tworzenia, a nie gdy są dodawane dopiero po wdrożeniu wszystkich rozwiązań.

CZYM JEST WIRTUALIZACJA?

Wirtualizacja polega na utworzeniu wirtualnego środowiska informatycznego, które umożliwia efektywniejsze niż w przypadku środowiska fizycznego wykorzystanie zasobów. Aplikacje i ich komponenty są „oddzielane" od sprzętu, na którym zostały zainstalowane, za pomocą zaawansowanych urządzeń i oprogramowania. Działają w środowisku, które można określić jako doskonałe. W takich systemach informatycznych prezentowany jest logiczny lub wirtualny widok zasobów fizycznych, który może znacznie różnić się od widoku fizycznego.

Wirtualizacja dokonywana jest zwykle w takich celach, jak zwiększenie wydajności, skalowalności, niezawodności, dostępności lub sprawności systemu, optymalizacja jego wykorzystania bądź stworzenie ujednoliconego środowiska zabezpieczeń lub domeny zarządzania. Czasem chodzi też o to, aby starsze aplikacje mogły funkcjonować nawet w sytuacji, gdy systemy, karty sieciowe bądź karty lub urządzenia pamięci masowej, które stanowią centralną część systemu informatycznego, są przestarzałe lub w ogóle nie występują w środowisku fizycznym.

Document #20090422

Tomasz Smajdor
Continuous Processing Technologies
Stratus EMEA Distributor
Mobile. +48 500 040 601
e-mail: tsmajdor@cptech.pl

Zwirtualizowane środowisko informatyczne powinno działać tak, aby administratorzy i pracownicy mogli wykonywać swoją pracę bez świadomości, gdzie znajdują się używane przez nich zasoby, jaki mają charakter (fizyczny czy wirtualny) oraz jak są ze sobą zestawione.

Firma Kusnetzky Group opracowała model złożony z warstw technologii, które można wykorzystać razem lub oddzielnie w celu utworzenia wirtualnego środowiska informatycznego. W naszym artykule będziemy używać terminologii zastosowanej w tym właśnie modelu, którego opis zamieściliśmy w Dodatku A.

TWORZENIE ŚRODOWISKA O WYSOKIEJ DOSTĘPNOŚCI

Jest wiele sposobów tworzenia środowisk o wysokiej dostępności. W każdym przypadku niezbędny jest pewien poziom nadmiarowości, która polega na zduplikowaniu istotnych zasobów i zorganizowaniu ich pracy w taki sposób, aby system działał nawet wtedy, gdy jeden z komponentów zawiedzie. Celem nadmiarowości jest wyeliminowanie zarówno planowanych (np. związanych z konserwacją), jak i nieplanowanych (np. spowodowanych awarią lub przerwą w dostawie prądu) przestojów w pracy systemu. Wybór zasobów, które zostaną zduplikowane, zależy od budżetu firmy i jej wymagań odnośnie do dostępności, a także od doświadczenia pracowników w planowaniu i wdrażaniu środowisk informatycznych oraz zarządzaniu nimi.

METODY OPARTE NA OPROGRAMOWANIU

Oprogramowanie zwiększające dostępność systemów informatycznych dzieli się na kilka rodzajów w zależności od celu, którym może być niezawodność różnych aspektów funkcjonowania środowiska informatycznego — dostępu do danych, pamięci masowej lub sieci, działania aplikacji, przetwarzania danych w całym stosie aplikacji itp. Rozwiązania takie oferują m.in. firmy Cassatt, Citrix, HP, IBM, Microsoft, Novell, Scalent Systems, Surgient i VMlogix. Zwiększenie dostępności systemu wiąże się jednak oczywiście z pewnymi kosztami. Oprogramowanie, o którym mowa, wymaga odpowiednich zasobów pamięci operacyjnej i masowej oraz odpowiednio mocnych procesorów, a także wykwalifikowanych pracowników, którzy zajmą się jego konfigurowaniem i aktualizowaniem, a w razie potrzeby — rozwiązywaniem problemów z jego działaniem. Przełączanie awaryjne nie następuje natychmiast. Systemy potrzebują trochę czasu, aby rozpoznać awarię oraz wybrać i zastosować najlepsze rozwiązanie. W zależności od wybranych metody i technologii może to potrwać od kilku minut do nawet kilku godzin.

METODY OPARTE NA SPRZĘCIE

Nadmiarowość sprzętową można osiągnąć w różny sposób, np. poprzez zainstalowanie większej liczby niezależnych komputerów, komputera typu *blade* z wieloma procesorami, bądź całego systemu zaprojektowanego pod kątem odporności na awarie. Dell, HP, IBM i inni producenci oferują szereg urządzeń typu *blade*, zasilaczy i innych rozwiązań o przeznaczeniu ogólnym. Na każdym kroku producenci takiego sprzętu zaznaczają, że poszczególne urządzenia można ze sobą łączyć, aby utworzyć środowisko o wysokiej dostępności. W systemach nadmiarowych wykorzystywana jest technologia wirtualizacji, która ułatwia wykrywanie awarii, określanie ich przyczyn oraz wybieranie i realizowanie strategii zapewniającej ciągłość obsługi procesów obliczeniowych.

Komponenty nadmiarowe mogą być używane cały czas do realizacji zadań (jako to tzw. rezerwy gorące), bądź funkcjonować jako urządzenia „zapasowe”, które bezczynnie czekają na przejęcie zadań w przypadku awarii (są to tzw. rezerwy zimne). Trzecie rozwiązanie polega na połączeniu co najmniej trzech systemów w klaster i bazuje na przekonaniu, że nigdy nie nastąpi równoczesna awaria wszystkich, więc zadanie będą na pewno zawsze wykonywane.

Tomasz Smajdor
Continuous Processing Technologies
Stratus EMEA Distributor
Mobile. +48 500 040 601
e-mail: tsmajdor@cptech.pl

W niektórych przypadkach najlepiej jest zastosować systemy zaprojektowane specjalnie pod kątem odporności na uszkodzenia, które pracują nawet wtedy, gdy jeden z komponentów zawiedzie. Systemy takie nie ograniczają się do minimalizacji czasu potrzebnego na przełączanie awaryjne i odtwarzanie danych, ale po prostu eliminują awarie, co stanowi gwarancję realizacji każdego zadania.

ŻADEN SYSTEM NIE PRACUJE W OSAMOTNIENIU

Niezależnie od wybranej metody zwiększania niezawodności środowiska informatycznego należy pamiętać, że każdy system musi mieć dostęp do pamięci masowej zawierającej aplikacje i dane oraz do równie niezawodnej sieci. Jeżeli komponenty te nie zostały uwzględnione w architekturze wysokiej dostępności, to w przypadku awarii systemu pamięci masowej lub sieci nawet najbardziej niezawodny system informatyczny staje się bezużyteczny.

ODPORNOŚĆ NA AWARIE A DOSTĘPNOŚĆ

Spośród wszystkich wymienionych powyżej konfiguracji tylko ostatnia, czyli system odporny na awarie, gwarantuje ciągłość procesów obliczeniowych. Wszystkie pozostałe wymagają oprogramowania, które będzie monitorowało działanie systemu, a w razie potrzeby doprowadzi do przełączenia awaryjnego lub ponownego uruchomienia danej funkcji w innym miejscu. W takim systemie awarie sprzętu są niedostrzegalne dla użytkowników. Praca trwa nawet wtedy, gdy jeden z komponentów systemu ulegnie awarii lub będzie poddawany rutynowej konserwacji.

NIEZAWODNOŚĆ A LICZBA DZIEWIĄTEK PO PRZECINKU

Dostawcy często podają poziom dostępności gwarantowany przez swoje rozwiązania, wyrażony procentowo jako czas pracy systemu bez przestojów. Poniższa tabela pozwala zrozumieć, w jakim stopniu każda dodatkowa cyfra „9” po przecinku w tym wskaźniku zmniejsza ryzyko przestojów.

Wskaźnik czasu pracy bez przestojów w ciągu miesiąca	Wskaźnik czasu przestojów w ciągu miesiąca	Czas przestojów w ciągu miesiąca (w sekundach)	Czas przestojów w ciągu miesiąca (w minutach)	Czas przestojów w ciągu miesiąca (w godzinach)
99,0%	1,0%	26 280,00	438,000	7,300
99,9%	0,1%	2 628,00	43,800	0,730
99,99%	0,01%	262,80	4,380	0,073
99,999%	0,001%	26,28	0,438	0,007
99,9999%	0,0001%	2,63	0,044	0,001

Wskaźnik czasu pracy systemu bez przestojów zależy od wielu czynników, takich jak niezawodność urządzeń, pamięci masowej lub sieci oraz prawdopodobieństwo popełnienia przez człowieka błędu, który spowolni lub zatrzyma procesy obliczeniowe. Jak wynika z powyższej tabeli, klient nabywający rozwiązanie o wskaźniku 99% musi być przygotowany średnio na ponad 7 godzin przestojów w miesiącu. W niektórych przypadkach jest to wprawdzie wystarczający poziom niezawodności, ale coraz częściej jest to zdecydowanie za mało.

Jeżeli jednak dodamy do tego wskaźnika jedną dziewiątkę po przecinku (wyniesie on wówczas 99,9%), to czas przestojów w ciągu miesiąca zmniejszy się do zaledwie trzech kwadransów.

W zależności od typu systemu, czas pracy bez przestojów wynosi zwykle od 95% do 99%.

Tomasz Smajdor
Continuous Processing Technologies
Stratus EMEA Distributor
Mobile. +48 500 040 601
e-mail: tsmajdor@cptech.pl

Producenci rozwiązań klastrowych, które najczęściej są opartymi na migracji klastrami maszyn wirtualnych, deklarują poziom niezawodności od 99,5% do 99,9%. Nawet jednak w przypadku wskaźnika 99,99% użytkownik musi być przygotowany na 4,32 minuty przestojów miesięcznie. W instytucji finansowej taka przerwa w pracy systemów transakcyjnych lub EFT może oznaczać milionowe straty.

Zdaniem firmy Stratus, od dawna oferującej na rynku systemy odporne na awarie, to zdecydowanie za mało dla aplikacji o znaczeniu krytycznym, w przypadku których nie można sobie pozwolić na ***żadne*** przestoje. W takich sytuacjach potrzebna jest niezawodność na poziomie sześciu dziewiątek, tzn. 99,9999%, co oznacza mniej niż 3 sekundy przestojów w miesiącu. Systemy ftServer firmy Stratus gwarantują taki lub bardzo zbliżony wskaźnik już od 2002 r., gdy zostały wprowadzone na rynek. W serwisie internetowym tej firmy (w części Uptime Meter) można znaleźć wskaźniki czasu pracy bez przestojów obliczane codziennie dla różnych urządzeń i systemów operacyjnych uruchamianych na jej serwerach.

JAKĄ DOSTĘPNOŚĆ MOŻNA UZNAĆ ZA WYSTARCZAJĄCĄ?

Jest to pytanie typu „jaką długość ma kawałek sznurka”. Aby na nie odpowiedzieć, trzeba najpierw ten „kawałek sznurka” wybrać i zmierzyć.

Należy więc ocenić zadania i obciążenie systemu informatycznego przy uwzględnieniu wszystkich jego funkcji, a następnie dla każdej z nich określić maksymalny czas przestoju, który nie spowoduje problemów. Niektóre zadania są ważne, ale nie mają znaczenia krytycznego. Funkcja o znaczeniu krytycznym z biznesowego punktu widzenia nie musi mieć takiego znaczenia ze strategicznego punktu widzenia. W większości firm wymagania dotyczące dostępności poszczególnych funkcji są zróżnicowane. Awaria systemów o znaczeniu strategicznym często prowadzi jednak do poważnych problemów.

Warto pamiętać, że niektóre funkcje, np. poczta elektroniczna lub aplikacje ułatwiające współpracę, które kiedyś miały charakter pomocniczy, zyskały znaczenie krytyczne, gdy przedsiębiorstwa przeniosły się do świata Internetu dostępnego przez całą dobę i wszystkie dni w roku.

Każda firma powinna więc uwzględniać w swoich projektach i planach koszty przestojów systemu informatycznego.

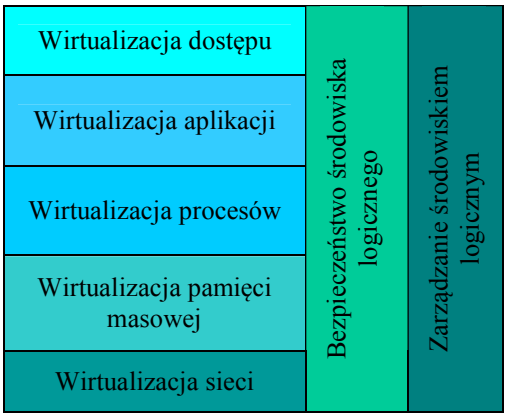
JAK PODEJMOWAĆ DECYZJE O ZAKUPIE ROZWIĄZAŃ?

Podejmowanie decyzji o architekturze ważnych systemów nie jest i nigdy nie było łatwe. Wielu dostawców sprzętu i oprogramowania oraz usług hostingu i usług zarządzanych chciałoby mieć swój udział w rozwiązaniach zwiększających dostępność systemu informatycznego przedsiębiorstwa. Najlepiej jednak znaleźć firmę mającą duże doświadczenie na polu zwiększania dostępności i odporności na awarie. Warto też skontaktować się z jej partnerami, którzy mogą zaoferować wiele cennych wskazówek i porad. Umiejętne planowanie może zapobiec wielu katastrofom.

Tomasz Smajdor
Continuous Processing Technologies
Stratus EMEA Distributor
Mobile. +48 500 040 601
e-mail: tsmajdor@cptech.pl

Dodatek A

MODEL WIRTUALIZACJI OPRACOWANY PRZEZ FIRMĘ KUSNETZKY GROUP



Kusnetzky Group © 2007

Wirtualizacja dostępu — technologia oparta na sprzęcie i oprogramowaniu, która umożliwia uzyskanie dostępu do każdej aplikacji praktycznie z każdego urządzenia w taki sposób, że użytkownicy tych urządzeń nie muszą wiedzieć nawzajem o swoim istnieniu. Warstwa ta może obejmować takie funkcje, jak usługi terminalowe i menedżer prezentacji.

Wirtualizacja aplikacji — technologia oparta na oprogramowaniu, która umożliwia uruchamianie aplikacji na różnych systemach operacyjnych i platformach sprzętowych. Wirtualizacja aplikacji tworzy widok, w którym wszystkie systemy zgodne ze standardami branżowymi działające pod kontrolą tych samych systemów operacyjnych stanowią jedną pulę zasobów. Pulą tą można dynamicznie zarządzać zgodnie z wymaganym poziomem usług oraz dostosowywać ją do planowanych lub nieplanowanych przestojów. Pozwala ona nawet konsolidować zadania na mniejszej liczbie systemów fizycznych (wówczas niektóre z nich można wyłączyć, aby zmniejszyć zużycie prądu i emisję ciepła). Technologia ta umożliwia też równoczesne uruchamianie na tym samym systemie fizycznym wielu aplikacji lub wersji aplikacji, które wcześniej nie były ze sobą kompatybilne.

Wirtualizacja procesów — technologia oparta na sprzęcie i oprogramowaniu, która oddziela fizyczną konfigurację sprzętu od usług systemowych, systemów operacyjnych oraz aplikacji. Dzięki tej technologii jeden system fizyczny może „na zewnątrz" wyglądać jak wiele systemów i odwrotnie. Zwykle technologia ta jest stosowana w celu zmaksymalizowania wydajności, skalowalności, niezawodności, dostępności i sprawności systemu lub skonsolidowania wielu środowisk w jednym systemie. Osiągnięcie każdego z tych celów wymaga nieco innego podejścia do wirtualizacji procesów. Ta warstwa technologii wirtualizacyjnych obsługuje przetwarzanie danych w sieciach obliczeniowych (grid), tworzenie klastrów obrazów jednego systemu, tworzenie klastrów w celu zwiększenia dostępności i ułatwienia przełączania awaryjnego, wirtualizowanie klientów i serwerów oraz partycjonowanie i wirtualizowanie systemów operacyjnych.

Wirtualizacja pamięci masowej — technologia oparta na sprzęcie i oprogramowaniu, która ukrywa miejsce instalacji systemów pamięci masowej oraz typy urządzeń obsługujących aplikacje i dane. Umożliwia współużytkowanie tej samej pamięci masowej przez wiele systemów fizycznych, dzięki czemu nie trzeba jej kupować do każdego systemu odrębnie. Metoda ta umożliwia również wielokrotne replikowanie pamięci masowej w różnych centrach przetwarzania danych w celu przyśpieszenia odtwarzania danych po awarii.

Wirtualizacja sieci — technologia oparta na sprzęcie i oprogramowaniu, która tworzy widok sieci różniący się od widoku fizycznego. Umożliwia ona bezpieczne współużytkowanie tej samej sieci przez różne zasoby. Klient widzi tylko te serwery, do których ma dostęp, a serwer — tylko te urządzenia klienckie, które obsługuje.

Zarządzanie zwirtualizowanymi środowiskami i bezpieczeństwem — dwie warstwy wirtualizacji, które należą do najważniejszych, ponieważ odpowiadają za zarządzanie wszystkimi pozostałymi warstwami oraz zapewnienie im bezpieczeństwa. Ta technologia jest oparta na oprogramowaniu i umożliwia konfigurowanie wielu systemów i zarządzanie nimi w taki sposób, jakby stanowiły jeden zasób obliczeniowy. Bez tej warstwy firmy korzystające z technologii wirtualizacyjnych byłyby narażone na takie same trudności i koszty, co w przypadku systemów fizycznych.

Tomasz Smajdor
Continuous Processing Technologies
Stratus EMEA Distributor
Mobile. +48 500 040 601
e-mail: tsmajdor@cptech.pl

DOSKONAŁOŚĆ, NA KTÓREJ MOŻNA POLEGAĆ

Wiele warstw technologii, które tworzą zwirtualizowane środowisko, zostało niestety zaprojektowanych tak, jakby sprzęt i procesy administracyjne były doskonałe i stuprocentowo niezawodne. Każdy, kto zna systemy informatyczne od strony praktycznej, wie, że jest inaczej. Sprzęt ulega awariom, a personel informatyczny i administracyjny popełnia błędy, które mogą spowolnić, a nawet wstrzymać procesy obliczeniowe. Projektanci systemów, którzy budują zwirtualizowane środowiska informatyczne dla przedsiębiorstw, powinni więc poważnie potraktować problem dostępności. Miejmy nadzieję, że będą to robić częściej niż dotychczas.