

Jak uchronić się przed potopem cyfrowych danych

Tempo przyrostu ilości cyfrowych danych zwiększa się z roku na rok. Eksperti z International Data Corporation (IDC) prognozują, że do 2020 roku objętość cyfrowego wszechświata zwiększy się 67-krotnie, a w samym tylko 2010 roku świat zaleje 1,2 zettabajtów informacji elektronicznych. Co zrobić, by nie utonąć w potopie elektronicznych informacji oraz jak efektywnie przechowywać dane?

Cyfrowy wszechświat stale rozszerza swoje granice, a specjaliści alarmują, że w kolejnych latach możemy spodziewać się jeszcze większego tempa przyrostu informacji cyfrowych. Wg prognoz IDC 2020 roku ilość danych cyfrowych wzrośnie aż 67-krotnie. Co ciekawe, jedynie połowa cyfrowego wszechświata jest wynikiem bezpośrednich działań ludzi – wysyłania maili, zakładania stron internetowych czy też rozmów telefonicznych w technologii VoIP. Drugą połowę stanowi tzw. cyfrowy cień – pozostałość, ślad podejmowanych przez nas działań zapisany w rejestrach rozmów, historii przeglądanych stron czy nagraniach kamer monitoringu.

Za tak dynamicznym przyrostem ilości danych stoi cały szereg zjawisk - od rosnącej popularności urządzeń mobilnych, dynamicznej ekspansji internetu i cyfryzacji coraz większej liczby dziedzin życia, po procesy globalizacji, ciągły rozwój społeczeństwa informacyjnego oraz gospodarki opartej na wiedzy, w której informacje – obok ludzi – stanowią najcenniejsze aktywa firm. Jednak tak szybki przyrost ilości cyfrowych danych niesie ze sobą także pewne wyzwania - chcąc uniknąć sytuacji, w której wielkość cyfrowego świata przewyższy możliwości nośników informacji, nieodzowne stanie się wprowadzenie odpowiedniej polityki zarządzania danymi, zarówno w wymiarze korporacyjnym, jak i w przypadku użytkowników domowych.

Nieaktywne dane zabierają miejsce

Istotnym problemem współczesnego storage’u jest fakt, iż znaczna część przechowywanych danych - niezależnie czy na firmowych serwerach, czy na dyskach twardych komputerów domowych – to informacje nieaktywne, niewykorzystywane przez użytkowników. Dobrym rozwiązaniem problemu niekontrolowanego przyrostu zbędnych plików jest wdrożenie systemu zarządzania danymi – profesjonalnego w przypadku firmy, racjonalnej polityki optymalizującej wykorzystanie zasobów dyskowych w odniesieniu do użytkowników domowych.

W przypadku przedsiębiorstw pierwszym i podstawowym krokiem wiodącym do optymalizacji storage'u jest określenie kluczowych z punktu widzenia firmy danych i wymagań odnośnie ich przechowywania (np. z uwagi na uregulowania prawne czy poziom poufności). Istotne jest także zapewnienie niezakłóconej pracy aplikacji roboczych oraz wdrożenie inteligentnych rozwiązań zapisu i magazynowania danych. Ważną rolę odgrywają także technologie pozwalające na maksymalne wykorzystanie pojemności posiadanych dysków – mowa tu m.in. o technologiach deduplikacji i archiwizacji. Nieodzowne z punktu widzenia przechowywania danych w firmach są również procedury backupu oraz zabezpieczania danych.

Powyższe zasady znajdują swoje zastosowanie także w przypadku magazynowania danych przez użytkowników domowych, na komputerach których również znajdują się cenne cyfrowe dane.

Nieuzasadniona duplikacja

Kolejną kwestią, z którą przyjdzie nam się zmierzyć w obliczu dynamicznego przyrostu danych cyfrowych jest nieuzasadniona duplikacja. O ile planowa i celowa replikacja danych jest jednym z podstawowych narzędzi backupu, o tyle niczym nieuzasadniona duplikacja stanowi istotny problem zarówno dla centrów danych, jak i dla pojedynczych serwerów storage'owych czy dysków twardych. Duplikacja to zjawisko występowania jednej lub wielu kopii tej samej porcji danych w obrębie jednego nośnika lub na różnych nośnikach. Z duplikacją danych mamy do czynienia nie tylko w firmach, ale także i na prywatnych, domowych komputerach. Często ten sam plik, zdjęcie, utwór czy film występują na kilku komputerach, urządzeniach cyfrowych czy nośnikach w obrębie jednego gospodarstwa domowego. Dobrym rozwiązaniem, eliminującym multiplikowane porcje danych, jest stworzenie małego, domowego centrum danych. Za centrum tego typu posłużyć może niewielki serwer storage'owy, do którego dostęp uzyskują wszyscy domownicy. Wyposażony w wyjście HDMI – tzw. Home Server -może on stanowić centrum multimedialne. Podobne funkcje może pełnić dysk sieciowy NAS.

NAS (Network Attached Storage) to technologia umożliwiająca podłączenie zasobów zgromadzonych na dyskach bezpośrednio do sieci komputerowej. Jest to macierz dyskowa (lub pojedynczy dysk twardy) podłączony bezpośrednio do sieci lokalnej. W ramach infrastruktury serwer-klient pełni rolę serwera, posiada procesor i okrojony wersję systemu operacyjnego. Zasoby serwerów NAS są udostępniane uprawnionym użytkownikom. Rozwiązanie to umożliwia łatwy dostęp do danych zmagazynowanych w jednym miejscu z różnych punktów sieci. Silną stroną dysków sieciowych NAS jest możliwość stosowania ich w heterogenicznych środowiskach sieciowych bazujących na różnych rozwiązaniach klienckich, dzięki czemu dane dostępne są niezależnie od zainstalowanego systemu operacyjnego.

Ponadto publiczny adres IP czy też usługi VPN oferuje użytkownikowi dostęp do zgromadzonych zasobów z dowolnego miejsca, także za pomocą urządzeń mobilnych.

Odpowiedzią na nieuzasadnioną duplikację są procedury deduplikacji. Polegają one na eliminacji powtarzających się informacji poprzez zastąpienie kopii odnośnikami do oryginalnej, źródłowej porcji danych. Deduplikacja to operacja wykonywana automatycznie przez oprogramowanie, najczęściej w macierzy dyskowej, dzięki której możliwe jest wyeliminowanie powtarzających się danych (ich duplikatów). Umożliwia ona przechowywanie jedynie unikatowych plików lub bloków z danymi, generując tym samym znaczne oszczędności przestrzeni dyskowej.

Wyróżnia się dwa rodzaje deduplikacji: na poziomie systemu plików i na poziomie bloków dyskowych. Deduplikacja na poziomie bloków dyskowych gwarantuje lepsze rezultaty, gdyż nie zależy od ilości lub rodzaju plików ani od systemu operacyjnego, na którym bazuje dany system informatyczny.

Funkcje deduplikacji oferowane są przez niektóre aplikacje odpowiedzialne za backup. Oprogramowanie to często pozwala ponadto na jednoczesną na kompresję danych, co oczywiście daje dodatkowy zysk. Przy szybko rosnącej ilości przechowywanych informacji funkcje te zaczynają nabierać szczególnie istotnego znaczenia - większa ilość przechowywanych danych archiwizuje się dłużej, a czasu na backup pozostaje tyle samo. Tym bardziej, że liczne regulacje coraz częściej zmuszają firmy do archiwizowania danych nawet przez dziesiątki lat.

Archiwizacja

Gdy mamy do czynienia z danymi, z których na co dzień nie korzystamy, ale z różnych względów uznajemy je za cenne i nie chcemy się z nimi rozstać, z pomocą przychodzi nam procedury archiwizacji. Statystyki pokazują, że z 80% stworzonych danych firma nigdy już nie korzysta bądź korzysta bardzo rzadko. Archiwizacja umożliwia przeniesienie owych danych na wolniejsze, a przez to tańsze od dysków twardych nośniki. Archiwizacja to proces przenoszenia danych z systemów komputerowych na inne nośniki w celu zredukowania ich ilości. Archiwizacja może przybierać formę kopii analogowych (wydruki dokumentów), kopiowania danych na nośniki wymienne (np. płyty CD-R, płyty DVD-R czy Blu-ray), kopiowania plików na inny dysk tego samego komputera (np. RAID), kopiowania na taśmę magnetyczną, kopiowania na inny komputer (np. kopia zwierciadlana, serwer plików) lub zautomatyzowanej archiwizacji online (np. Przeklej.pl, Plikus.pl).

Standardowe sposoby archiwizacji to metody ręczne i półautomatyczne, wymagające od użytkownika nakładów pracy, czasu i pieniędzy. Archiwizacja w obrębie sieci lokalnej wiąże się z koniecznością zakupu oraz regularnego unowocześniania sprzętu, ponoszenia kosztów związanych z eksploatacją infrastruktury. Ponadto wymaga wygospodarowania bezpiecznej przestrzeni na składowanie wypełnionych nośników, co niesie ze sobą kolejne koszty (instalacja alarmu, zatrudnienie dodatkowej obsługi). Doskonałe rozwiązanie archiwizacyjne powinno umożliwiać nieprzerwany dostęp do zarchiwizowanych danych, stąd najczęściej stosowanymi rozwiązaniami archiwizacyjnymi są dyski twarde, dyski zewnętrzne i serwery storage'owe czy biblioteki taśmowe. Do celów archiwizacji nadają się także nośniki optyczne (DVD, Blu-ray) – ze względu na swoją stosunkowo dużą pojemność i relatywnie niski koszt

stanowią dobre rozwiązanie zarówno dla firm, jak i dla zaawansowanych użytkowników domowych. Niestety, czytniki tego typu utrudniają dostęp do zarchiwizowanych danych oraz pozostawiają wiele do życzenia, jeśli chodzi o bezpieczeństwo zarchiwizowanych informacji z uwagi na ograniczone możliwości kontroli dostępu.

Coraz więcej firm decyduje się na usługi archiwizacji online. Outsourcing procedur archiwizacyjnych to wygodne rozwiązanie, zdejmujące z barków firmy lub użytkownika konieczność inwestowania w urządzenia i infrastrukturę magazynowania danych. Sceptycy archiwizacji online za jej główną słabość uznają długi - zależny od przepustowości łącza internetowego- upload danych.

Jak przechowywać, to z głową... w chmurze

Coraz większą popularnością cieszą się rozwiązania tzw. cloud storage – przechowywania danych w chmurze. Cloud storage stanowi model przechowywania danych online, w którym informacje magazynowane są na wirtualnych serwerach hostowanych przez dostawców usług. W modelu tym użytkownik płaci tylko za powierzchnię dyskową, z której rzeczywiście korzysta oraz za faktycznie wykorzystywane zasoby. Przechowywanie danych w modelu cloud niesie ze sobą liczne korzyści. Cloud storage nie wymaga od użytkowników ponoszenia nakładów na lokalną infrastrukturę fizyczną, zdejmując z niego także ciężar przeprowadzania backupów, replikacji, konserwacji urządzeń. Z drugiej jednak strony istnieją pewne wątpliwości odnośnie bezpieczeństwa danych (szczególnie tych wrażliwych) magazynowanych w chmurze. Specjaliści zwracają także uwagę na fakt, iż dostęp do cyfrowych informacji przechowywanych w ten sposób jest uzależniony od dostępności i jakości połączenia internetowego.

Specjaliści z IDC prognozują, że technologia cloud storage będzie zyskiwać coraz większy udział w rynku. Obecnie szeroko pojęte usługi cloud storage stanowią 9% rynku rozwiązań cloud computing, którego całkowita wartość szacowana jest na ok. 17,4 mld USD. W 2013 wartość rynku cloud storage ma wzrosnąć do 6,2 mld USD. Polscy użytkownicy nadal z ograniczonym zaufaniem podchodzą do technologii cloud storage – dość niechętnie godzą się z myślą, że ich cenne dane przechowywane będą z dala od nich.

Bezpieczeństwo danych - backup

Wszystkie technologie przechowywania danych mogą być narzędziem backupu. Kopie zapasowe można przechowywać w chmurze, na płytach, na dyskach twardej, dyskach zewnętrznych, na serwerach własnych lub wynajętych od zewnętrznej firmy. Tworzenie zapasowych kopii bezpieczeństwa ma na celu umożliwienie odtworzenia danych utraconych w przypadku ich utraty bądź uszkodzenia. Choć pierwotnie termin backup odwoływał się do zautomatyzowanych i usystematyzowanych form tworzenia kopii zapasowych, obecnie pojęciem tym określamy wszelkie, nawet amatorskie, sposoby sporządzania kopii bezpieczeństwa plików. Kopie bezpieczeństwa najczęściej przechowywane są na dyskach HDD lub nośnikach taśmowych.

Wiele programów, np. edytorów tekstów, automatycznie tworzy kopie opracowywanych plików, przez co umożliwia odtworzenie ich zawartości w sytuacji awarii komputera lub niespodziewanego odcięcia dopływu prądu. Mechanizmy backupu są standardem w przypadku środowisk serwerowych, gdzie od bezpieczeństwa danych zależy może funkcjonowanie całej firmy.

Wyróżniamy kilka rodzajów backupu: całościowy, przyrostowy, różnicowy, lokalny oraz sieciowy. Każdy z nich ma swoje mocne i słabe strony, decyzja o wybranej metodzie tworzenia kopii zapasowych zawsze należy do użytkownika.

Backupowi całościowemu (full backup) poddawane są wszystkie zapisane na nośniku dane, a bit „archive” plików ustawiany jest w stan „0”. Backup całościowy zdecydowanie ułatwia wyszukiwanie danych – wszystkie znajdują się na jednym nośniku, a w przypadku awarii odtworzenie systemu nie zajmuje zbyt wiele czasu. Z drugiej jednak strony full backup nie pozwala na efektywne wykorzystanie nośników – permanentnie tworzone są kopie bezpieczeństwa rzadko modyfikowanych danych. Kolejną komplikacją może być długi czas wykonywania operacji.

Drugą techniką tworzenia kopii bezpieczeństwa jest tzw. backup przyrostowy (incremental backup). W przeciwieństwie do swojego całościowego odpowiednika, backup przyrostowy dokonuje zapisu jedynie nowopowstałych danych lub tych, które uległy zmianie od czasu wcześniejszego backupu. Backupowane są pliki, które bit „archive” mają ustawione w stan „1”, a po tej operacji bit „archive” jest przestawiany w stan „0”. Silną stroną backupu przyrostowego jest krótki czas jego przeprowadzania oraz efektywne wykorzystanie nośników. Backup przyrostowy nie jest jednak metodą idealną, poza długim czasem odtwarzania systemu jego podstawową słabością jest trudność wyszukiwania danych - do odnalezienia zbioru są potrzebne wszystkie nośniki z backupami przyrostowymi oraz ostatni nośnik z backupem całościowym.

Kolejnym sposobem tworzenia kopii zapasowych jest backup różnicowy (differential backup), w którym zapisywane są te dane, które uległy zmianie od czasu ostatniego całościowego lub przyrostowego backupu. Backupowi poddawane są pliki, które bit „archive” mają ustawione w stan „1”. Po tej operacji nie ulega on zmianie. Ten typ backupu umożliwia łatwe wyszukiwanie potrzebnych w danej chwili danych, odtworzenie systemu przeprowadza się stosunkowo szybko. Również sama procedura backupu trwa krócej niż w przypadku backupu całościowego. Także i ten sposób tworzenia kopii zapasowych nie ustrzegł się słabości, którymi są nieefektywne wykorzystanie nośników oraz nadmiarowość backupów – nawet te dane które nie uległy zmianie, są cały czas backupowane. Backup różnicowy ustępuje przyrostowemu pod względem czasu wykonywania operacji.

Spektrum technik tworzenia kopii bezpieczeństwa zamykają backup lokalny oraz backup sieciowy. Backup lokalny to rozwiązanie proste w instalacji i konfiguracji. Gwarantuje szybki transfer danych i krótki czas trwania procedury backupu. Niestety wymaga dość dużego udziału ze strony użytkownika, co zwiększa prawdopodobieństwo wystąpienia błędu ludzkiego. Jest ponadto stosunkowo drogi.

Backup sieciowy z kolei umożliwia centralne zarządzanie, intuicyjną automatyzację. Jest tańszy w administrowaniu aniżeli jego lokalny odpowiednik. Słabością backupu sieciowego jest duże obciążenie sieci, wolny transfer danych oraz długi czas potrzebny do przeprowadzenia backupu.

Ważne przy tworzeniu kopii bezpieczeństwa jest, by pierwotne dane i ich zbackupowane odpowiedniki nie znajdowały się w tym samym miejscu. Podczas ostatniej powodzi wiele firm na własnej skórze odczuło jak bolesnym błędem może być lokalizacja wszystkich danych w jednym miejscu. Podobnie w przypadku użytkowników domowych – przechowywanie kopii bezpieczeństwa danych razem lub w pobliżu danych pierwotnych odbiera sens backupowi.

Niestety nadal bardzo często – potwierdzając słowa przysłowia „mądry Polak po szkodzie” – o tym, jak istotny jest backup dowiadujemy się dopiero po utracie danych. Okazuje się wtedy, że przechowywane na komputerze dane mają w istocie większą wartość niż samo urządzenie.

Replikacja

Jednym z narzędzi backupu jest replikacja - proces powielania danych, tworzenia ich kopii na różnych nośnikach, np. serwerach. Wyróżniamy trzy podstawowe rodzaje replikacji danych: migawkową, transakcyjną oraz dwukierunkową. Replikacja migawkowa polega na powieleniu i rozprowadzeniu na różnych nośnikach danych zapisanych w określonym momencie. Ten typ replikacji znajduje swoje zastosowanie głównie w przypadku danych poddawanych rzadkim, ale znacznym modyfikacjom. Za niedogodność możemy uznać fakt, iż dane zapisane pomiędzy kolejnymi migawkami nie są replikowane. W przypadku replikacji transakcyjnej dane rozprowadzane są na podstawie logów transakcji tylko wtedy, gdy odbywa się synchronizacja. Replikacja dwukierunkowa polega zaś na dwukierunkowej wymianie danych – od serwera głównego, jak i od klientów. Słabością tego rodzaju replikacji jest niebezpieczeństwo zaistnienia konfliktu w czasie synchronizacji.

Podsumowanie

Prawie 500 miliardów gigabajtów informacji cyfrowych wygenerowanych w 2008 roku oraz prognoza mówiąca o 1,2 zettabajta danych, które mają powstać w 2010 roku, nie pozostawiają złudzeń. Tak dynamiczny przyrost ilości informacji elektronicznych wymusi na firmach i użytkownikach domowych zmianę podejścia do kwestii storage’u. Zamiast inwestować w kolejne nośniki, coraz więcej osób zacznie zadawać sobie pytanie, co zrobić, by lepiej wykorzystać te, które już posiadają.

O firmie ACTION S.A.:

ACTION S.A. jest jednym z największych dystrybutorów oraz producentów sprzętu komputerowego w Polsce. Spółka - założona w 1991 roku - jest notowana na Giełdzie

Papierów Wartościowych w Warszawie od 2006 r. oraz jest członkiem Stowarzyszenia Emitentów Giełdowych.

Przez dziennik "Parkiet" uznana za jednego z najlepszych debiutantów giełdowych w 2006 roku, do dziś jest jedną z dynamicznie rozwijających się i cieszących się zaufaniem inwestorów spółek z branży IT notowanych na GPW.

ACTION S.A. sprzedaje produkty największych dostawców na świecie (w ofercie ACTION znajdują się produkty 250 producentów). Jest również producentem i dostawcą produktów pod własnymi markami: Actina i ActiveJet. Klientami ACTION S.A. są sprzedawcy detaliczni oraz odbiorcy korporacyjni. Spółka posiada od lat wiodącą pozycję w dziedzinie dostaw do przetargów w sektorze publicznym.

O serwerach Actina Solar:

W 2006 roku firma ACTION S.A. uruchomiła linię produkcji serwerów Actina Solar.

W ofercie serwerów Actina Solar znajdują się zarówno rozwiązania dla małych i średnich przedsiębiorstw, jak i dla instytucji wymagających ogromnych mocy obliczeniowych. Actina Solar posiada na swym koncie prestiżowe wdrożenia klastrów obliczeniowych w instytucjach naukowych (między innymi wdrożenie klastra Galera w Centrum Informatycznym Trójmiejskiej Akademickiej Sieci Komputerowej – TASK).

W ramach każdej z linii serwerów oferowana jest pełna elastyczność w doborze konfiguracji. Do dyspozycji Klientów oddany został konfigurator on-line. Serwery Actina Solar są objęte 3-letnią gwarancją oraz serwisem on-site.

Więcej informacji o marce Actina Solar na: www.actina.pl.

Dodatkowych informacji udzielają:

Adam Kozłowski
Monday PR
E-mail: adam.kozlowski@mondaypr.pl
Tel.: (22) 487-84-21
kom: 728 926 196

Aleksandra Bojanowska
Monday PR
E-mail: aleksandra.bojanowska@mondaypr.pl
Tel.: (22) 487-84-21

Marcin Bogusz
ACTION S.A.
E-mail: marcin.bogusz@action.pl
Tel.: (22) 333-90-18