Xopero praktyki: Pojęcia teoretyczne

Wojciech Błajda

12.09.2023 - DATA2

# Część 1: 12.09.2023

b

1. VSS – Jest to usługa Microsoft zezwalająca na wykonywanie kopii zapasowej dysku bez blokowania jego funkcjonalności. Taka kopia zapasowa pozwala przywrócić dowolny plik, folder, lub nawet wolumin.
2. Deduplikacja - Polega na eliminacji powtarzających się danych w zbiorze. Proces ten wykorzystywany jest przy tworzeniu kopii zapasowych, gdyż oszczędza on miejsce.
3. Kompresja – Jest to proces mający na celu zmniejszenie wielkości danych, z czego dane musi dać się przywrócić do stanu pierwotnego.
   1. Rodzaje kompresji
      1. Kompresja stratna – Po dekompresji, dane różnią się między sobą
      2. Kompresja bezstratna – Po dekompresji dane są identyczne
   2. Stopnie kompresji - Określają wielkość skompresowanego pliku w stosunku do jego oryginalnej wielkości. Im większy stopień, tym większa kompresja.
4. Szyfrowanie (typy)
   1. Szyfrowanie symetryczne – Są to szyfry, do których zaszyfrowania i odszyfrowania wykorzystuje się ten sam klucz
      1. Algorytm AES (Advanced Encryption Standard) - Bazuje na sieci substytucji permutacji, Długość klucza wynosi 128, 192 lub 256 bitów. Jest on odporny na krypto analizę różnicową i\oraz liniową, i cechuje się dużą prędkością działania. W 2001 roku został przyjęty jako standard.
      2. Tryby – ECB, CBC
         1. ECB - Każdy blok danych szyfrowany jest oddzielnie co pozwala także na wielowątkowość
         2. CBC – Na każdym bloku danych wykonywana jest operacja XOR z poprzednim zaszyfrowanym blokiem. Następnie ostateczny wynik działania szyfrowany jest w klasyczny sposób
   2. Szyfrowanie asymetryczne – Do zaszyfrowania oraz odszyfrowania informacji wykorzystuje się 2 różne klucze
   3. Klucz domyślny i klucz użytkownika w kontekście naszej aplikacji
5. Retencja plików - Są to dane dotyczące połączeń przechowywane przez 12 miesięcy, w celach bezpieczeństwa publicznego. Z danych tych można ustalić czas, miejsce oraz długość połączenia, łącznie z urządzeniami, które wykonywały takowe połączenie.
6. Kopie zapasowe
   1. Pełna - Jest to kopia 1:1 wszystkich danych znajdujących się na dysku.
   2. Przyrostowa – Jest to kopia tylko tych danych które zmieniły się od ostatniej pełnej kopii dysku. Do jej przywrócenia, potrzebna jest pełna kopia dysku wraz ze wszystkimi przeszłymi kopiami przyrostowymi.
   3. Różnicowa - Działa jak kopia przyrostowa, z tym, że zapisuje wszystkie zmiany od ostatniej pełnej kopii dysku, a do jej przywrócenia potrzebna jest ostatnia kopia różnicowa.
7. Wykonywanie kopii zapasowych na prawach użytkownika
8. Zmienne środowiskowe - Zmienna środowiskowa to nazwana wartość, którą system posługuje się w celu zarządzania systemem lub przechowywania informacji.
9. Harmonogram (systemowy i obecny w naszej aplikacji)
10. Pliki ukryte – ukryte pliki są domyślnie niewidoczne z poziomu eksploratora plików lub powłoki systemu. Mogą zostać wyświetlone z pomocą odpowiednich ustawień / argumentów. W systemie linux, ukryte pliki oznacza się prefiksem “.”, a w systemie Windows “~$”.
11. Filtry plików - W dziedzinie backupów, filtr plików określa jakie rodzaje plików mają zostać uwzględnione przy tworzeniu kopii zapasowej, a jakie pominięte.
12. Usługa systemowa (poziomy/rodzaje uprawnień)
13. Active Directory (podstawowe informacje) - Jest to usługa Windows Server, która ułatwia zarządzanie i monitorowanie zasobów sieciowych. Active Directory zapewnia skalowalność, bezpieczeństwo oraz integrację z różnymi innymi usługami Microsoft.
14. Wirtualizacja – W przeciwieństwie do emulacji, wykorzystuje tylko i wyłącznie zasoby platformy, dzięki czemu może wykonywać operacje prosto na zasobach sprzętowych komputera
    1. Rodzaje wirtualizatorów
       1. Pełna wirtualizacja – Wirtualny system operacyjny działa tak, jakby był natywnie zainstalowany na urządzeniu. Odwołania wirtualnego systemu są emulowane przez program. Dzięki temu można zainstalować niemal dowolny system operacyjny bez konfliktu z innymi wirtualnymi maszynami, za to jego działanie jest spowolnione.
       2. Parawirtualizacja – Jest to jedna lub więcej maszyn wirtualnych działających obok systemu hostującego, z czego jeden z systemów odpowiedzialny jest za przyznawanie zasobów komputera. Zaletą jest bardzo dobra wydajność maszyny wirtualnej, za to wirtualny system musi być specjalnie przystosowany do tego typu działania.
15. Redundancja – Biorąc pod uwagę dane, redundancja to użycie większej ilości danych niż jest to wymagane do zapisania pliku. Ma to swoje zastosowania. Dodatkowe dane dopisywane do pliku mogą służyć jego odtworzeniu w razie uszkodzenia. W kontekście programowania, redundancja jest to niewykorzystywany kod, kod który nie ma wpływu na wynik programu, oraz powtarzający się kod, który można skrócić (np. funkcją).
16. Chmura – Jest to sieć serwerów nadająca się do przechowywania plików / danych. Do chmury łatwy dostęp uzyskać można z każdego urządzenia z dostępem do Internetu, z dowolnego miejsca.

# Część 2: DATA

1. NAS – NAS (Network Attached Storage) to modularna sieciowa pamięć masowa. Może składać się z wielu dysków i przechowywać duże ilości danych, lecz jest on trudny do skalowalna oraz wymaga dobrego połączenia z siecią.
2. Rodzaje serwerów NAS
   1. QNAP – Jest to innowacyjna firma zajmująca się m.in. produkcją dysków NAS, które zapewniają wysoką wydajność
   2. Synology – Także zajmuje się wytwarzaniem dysków NAS i skupia się na ochronie danych użytkownika
   3. Asustor – Asustor specjalizuje się w integrowaniu odpowiedniego oprogramowania z aplikacjami i sprzętem
   4. Netgear – Balansuje on pomiędzy dobrą jakością swoich dysków a porównywalnie niską ceną
3. Mono – Jest to zestaw narzędzie open source pozwalających na uruchomienie programów stworzonych dla .NET Framework (w tym kompilator C# oraz środowisko CLR) na różnych systemach operacyjnych
4. Systemy wersjonowania – Pozwalają na śledzenie przebiegu pracy nad projektem oraz pomaga w udostępnianiu plików między osobami.
   1. CVS – Jest to najstarsza technologia wersjonowania oraz ta najwolniejsza. Wszystkie zmiany zapisywane są do głównego projektu, co może powodować konflikty, gdy wiele osób pracuje nad oprogramowaniem jednocześnie.
   2. SVN – Jest szybsze od CVS, lecz opiera się na podobnych zasadach. Oraz jest wspierany przez wiele systemów.
   3. Git – Jest on ciągle wspierany przez społeczność i jest szybszy od CVS oraz SVN. Użytkownik pobiera kopię całego projektu na swoje urządzenie, dzięki czemu może pracować offline i opublikować zmiany po połączeniu z siecią. Dzięki możliwości tworzenia branchów, wiele osób może pracować nad innymi aspektami programu, a następnie połączyć je z powrotem w jeden projekt.
5. Docker – Jest to program do uruchamiania “Kontenerów”. Oznacza to, że tworzony jest nowy odizolowany system operacyjny z gotową do działania aplikacją. Jest to wygodnym rozwiązaniem i zużywa mniej zasobów niż wirtualizacja. Kontenery mogą działać osobno niezależnie od siebie.
6. Kubernetes – Stworzony przez inżynierów Google oraz wdrożony na rynek w 2015 roku, jest on oprogramowaniem służących do zarządzania aplikacjami wielokontenerowymi. Kubernetes jest open source, a co za tym idzie może być używany przez każdego.
7. Strony wspierające wersjonowanie w chmurze (GitHub i inne)
8. Klastry – Jest to zbiór jednostek komputerowych które współpracują ze sobą w celu zwiększenia mocy obliczeniowej, tworząc pojedyncze środowisko pracy. Są klastry wydajnościowe, gdzie każdy komputer wykonuje swoje zadanie, klastry niezawodnościowe, gdzie w razie awarii jednego komputera inny przejmuje kontrolę nad jego zadaniem oraz klastry równoważenia obciążenia, gdzie każdy komputer wykonuje swoje zadanie, ale może także dzielić to zadanie z innymi węzłami.
9. Centra danych – Centrum danych to pomieszczenie lub budynek przeznaczony do masowego przechowywania serwerów firmy, w których znajdują się dane, systemy, strony, oraz informacje potrzebne firmie. Takie centra muszą być odpowiednio chronione (fizycznie i systemowo) by zapobiec uszkodzeniu danych.
10. RAID – Jest to współpraca 2 lub więcej dysków w celu uzyskania zwiększonej wydajności oraz niezawodności
    1. RAID 0 – Dane rozdzielane są na wiele dysków, zmniejszając czas zapisu i odczytu. System ten ma tą wadę, że dane zastają utracone po uszkodzeniu dowolnego z dysków.
    2. RAID 1 – Polega na mirroringu, gdzie połowa dysków zapisuje dane, a druga połowa ma ich dokładną kopię. Wadą tego rozwiązania jest fakt, że 50% pojemności macierzy jest zawsze zajęte.
    3. RAID 10 – Jest to połączenie metod RAID 1 i RAID 0. Dane są zapisywane na połowie dysków oraz mirrorowane. Mimo tego, jeżeli awarii ulegnie dysk wraz z jego lustrzaną kopią, dane zostaną utracone.
    4. RAID 4 – Na kilku dyskach zapisywane są dane, a na ostatnim ich sumy kontrolne, przez co w razie awarii jednego z dysku może być podjęta jego rekonstrukcja.
    5. RAID 5 – Dane są zapisywane na wszystkich dyskach, przy czym każdy z dysków zawiera sumę kontrolną 1 innego dysku. Ta metoda jest odporna na awarię tylko 1 dysku, przy czym wymagana jest późniejsza odbudowa sum kontrolnych.
11. Systemy plików
    1. FAT – Stosowany w systemach DOS i Windows, stworzony dla przenośników danych. Opiera się on na tabeli alokacji danych. Wadą tej metody jest tzw. fragmentacja plików, zmniejszając wydajność.
    2. NTFS – System plików wykorzystywany przez systemy od Windows NT po współczesny Windows 11 oraz przez wiele innych systemów, takich jak Linux czy MacOS. System ten wprowadza obsługę metadanych oraz listę kontroli dostępu. NTFS jest znacznie szybszy od FAT a także może przechowywać znacznie większe pliki, bez konieczności dzielenia ich na mniejsze części. NTFS także podlega fragmentacji, a defragmentacja dużych dysków może trwać nawet do kilku godzin.
    3. EXT - Początkowy system plików używany we wczesnym developmencie Linuxa. Bazowany na systemie plików systemu operacyjnego Minix, zezwalał na pliki do 64 MB a także partycje do 2 GB. System ten podlegał dużej fragmentacji i szybko został zastąpiony przez EXT2, który miał system spowalniający jej proces.
12. Rejestr systemowy Windows – Rejestr systemu Windows przechowuje różne dane, np. O profilach użytkowników, zainstalowanych programach, folderach i ikonach, sterownikach i używanych portach.
13. Foldery systemowe
    1. Windows
       1. Users – Katalog zawierający pliki oraz katalogi użytkowników.
       2. Program files – Zawiera pliki zainstalowanych programów
       3. Windows – Wszystkie pliki niezbędne do poprawnego działania systemu
    2. Linux
       1. Bin – Podstawowe pliki wykonywalne narzędzi systemowych.
       2. Boot – Pliki potrzebne do uruchomienia systemu.
       3. Dev – Pliki odnoszące się do urządzeń.
       4. Etc – Pliki konfiguracyjne i ustawienia systemowe.
       5. Home – Ustawienia każdego użytkownika oraz jego pliki
       6. Lib – Dzielone biblioteki systemowe.
       7. Meida - Dostęp do nośników wymiennych
       8. Mnt – Miejsce montowania dysków
       9. Proc – Wirtualny katalog przechowujący dane o aktualnych procesach
       10. Root – Ustawienia administratora systemu.
       11. Sbin – Pliki wykonywalne tylko dla administratora
       12. Tmp – Pliki tymczasowe
       13. Usr – Dodatkowe programy dla zwyczajnych użytkowników.
       14. Var – Dane systemowe które często zostają zmieniane
14. Amazon Web Service – Jest to publicznie udostępniona platforma chmurowa oraz płatny hosting internetowy, firmy amazon. Udostępnia moc obliczeniową dla baz danych, stron internetowych, etc.
    1. S3 – Jest to prosta do założenia usługa przechowywania masowych danych w chmurze. Ma nieograniczoną ilość miejsca do wykorzystania i płaci się tylko za ilość danych które się fizycznie na niej przechowuje. Ponadto, platforma ta ma swoje SDK, API, CLI przez co można zintegrować ją w różnych oprogramowaniach.
    2. EC2 – Zapewnia skalowalną moc obliczeniową w chmurze, którą można określać dynamicznie w zależności od aktualnych potrzeb. Usługa ta zapewnia wirtualne środowiska, konfigurację procesora, woluminy pamięci tymczasowej i trwałej, statyczne adresy IPv4, etc.
15. Azure – Jest to chmura obliczeniowa firmy Microsoft. Obsługuje wiele języków programowania, w tym C# oraz Java. Płatność ponosi się nie za możliwość korzystania z chmury, a za ilość zużytych zasobów.
16. OpenStack – Jest to chmura, którą można zainstalować na swoim urządzeniu. Z pomocą tej usługi można tworzyć chmury publiczne i prywatne. Projekt ten jest open source i jest całkowicie napisany w języku Python. OpenStack składa się z modułów które umożliwiają zarządzanie użytkownikami, obrazami maszyn wirtualnych, oraz magazynami danych. Zawiera też moduł Horizon, czyli webowy interface. użytkownika.

# Część 3: DATA

1. Linki Symboliczne
2. Specjalne pliki systemowe (pagefile.sys, hiberfil.sys, swapfile.sys)
3. Rejestr zdarzeń Windows i logowanie zdarzeń w kontekście programowania
4. Grupy robocze Windows
5. IPv4, IPv6
6. MSI, EXE
7. Skrypty .BAT, skrypty .sh
8. Ograniczenia systemów w kontekście ograniczeń dotyczących ścieżek do plików
9. Typy plików a rozszerzenia
10. DLL a EXE
11. Procesy a wątki
12. Wielowątkowość
13. Grafika wektorowa a rastrowa
14. BitLocker
15. HDD a SSD
16. UEFI a BIOS