

Rozdział 1

Rozdział 1

Struktura

jednostki 1.0

Cel 1.1

Wprowadzenie 1.2 Przetwarzanie w

chmurze w skrócie 1.2.1 Wizja

przetwarzania w chmurze

1.2.2 Definicja chmury

1.2.3 Bliższe spojrzenie 1.2.4 Model referencyjny
przetwarzania w chmurze 1.2.5

Charakterystyka i korzyści

1.2.6 Wyzwania przyszłość 1.3

Rozwój historyczny 1.3.1

Systemy rozproszone

1.3.2 Wirtualizacja 1.3.3 Obliczenia

zorientowane na usługi 1.3.4

Obliczenia zorientowane na użyteczność 1.4

Budowa środowisk przetwarzania

w chmurze 1.4.1 Rozwój aplikacji 1.4.2 Rozwój

infrastruktury i systemów 1.4.3 Platformy

obliczeniowe i technologie 1.4.3.1 Usługi sieciowe Amazon (AWS)

1.4.3.2 Google AppEngine

1.4.3.3 Microsoft Azure

1.4.3.4 Hadoop

1.4.3.5 Force.com i Salesforce.com 1.4.3.6

Manjrasoft Aneka

1.5

Podsumowanie 1.6

Pytania kontrolne 1.7 Odniesienia do dalszej lektury

Cel 1.0

W tym rozdziale zapoznasz się z poniższymi koncepcjami

- Co to jest przetwarzanie w chmurze?
- Jakie są cechy i zalety przetwarzania w chmurze?
- To wyzwania.
- Historyczny rozwój technologii w kierunku rozwoju przetwarzania w chmurze
- Rodzaje modeli przetwarzania w chmurze.
- Różne rodzaje usług w chmurze obliczeniowej.
- Rozwój aplikacji oraz technologie rozwoju infrastruktury i systemu o chmurze obliczeniowej.
- Przegląd różnych zestawów dostawców usług w chmurze.

1.1 Wprowadzenie

Historycznie rzecz biorąc, moc obliczeniowa była rzadkim i kosztownym narzędziem. Obecnie, wraz z pojawieniem się przetwarzania w chmurze, jest ono dostępne w dużych ilościach i niedrogie, co powoduje głęboką zmianę paradygmatu — przejście od obliczeń niedoborów do obliczeń obfitości. Ta rewolucja obliczeniowa przyspiesza utworowanie produktów, usług i modeli biznesowych oraz zakłada obecny przemysł technologii informacyjno-komunikacyjnych (ICT). W ten sam sposób dostarczała usługi do wodociągów, energii elektrycznej, gazu, telefonii i innych urządzeń. Cloud Computing oferuje przetwarzanie na żądanie, przechowywanie, oprogramowanie i inne usługi IT z płatnościami naliczanymi na podstawie wykorzystania. Cloud Computing pomaga na nowo wymyślać i przekształcać partnerstwa technologiczne w celu poprawy marketingu, uproszczenia i zwiększenia bezpieczeństwa oraz zwiększenia zainteresowania interesariuszy i doświadczenia konsumentów przy jednoczesnej redukcji kosztów. Dzięki przetwarzaniu w chmurze nie trzeba nadmiernie alokować zasobów, aby zarządzać potencjalnymi szczytowymi poziomami operacji biznesowych. W takim razie masz zasoby, których naprawdę potrzebujesz. Zasoby te można skalować w celu natychmiastowego zwiększania i zmniejszania możliwości w miarę zmieniających się potrzeb biznesowych. W tym rozdziale przedstawiono krótkie podsumowanie trendu w chmurze obliczeniowej, opisując jego wizję, omawiając kluczowe cechy i analizując postęp techniczny, który to umożliwił. W rozdziale tym przedstawiono także niektóre kluczowe technologie przetwarzania w chmurze oraz pewne spostrzeżenia na temat środowisk przetwarzania w chmurze.

1.2 Przetwarzanie w chmurze w skrócie

Pojęcie przetwarzania w „chmurze” sięga początków obliczeń użytkowych, terminu zaproponowanego publicznie w 1961 roku przez informatyka Johna McCarthy’ego:

„Jeśli komputery tego rodzaju, które zalecałem, staną się komputerami przyszłości, wówczas komputery mogą pewnego dnia zostać zorganizowane jako narzędzie publiczne, tak jak system telefoniczny jest narzędziem publicznym... Narzędzie komputerowe może stać się podstawą nowego i ważnego przemysłu.”

Główny naukowiec Sieci Agencji Zaawansowanych Projektów Badawczych (ARPANET), Leonard Kleinrock powiedział w 1969 roku:

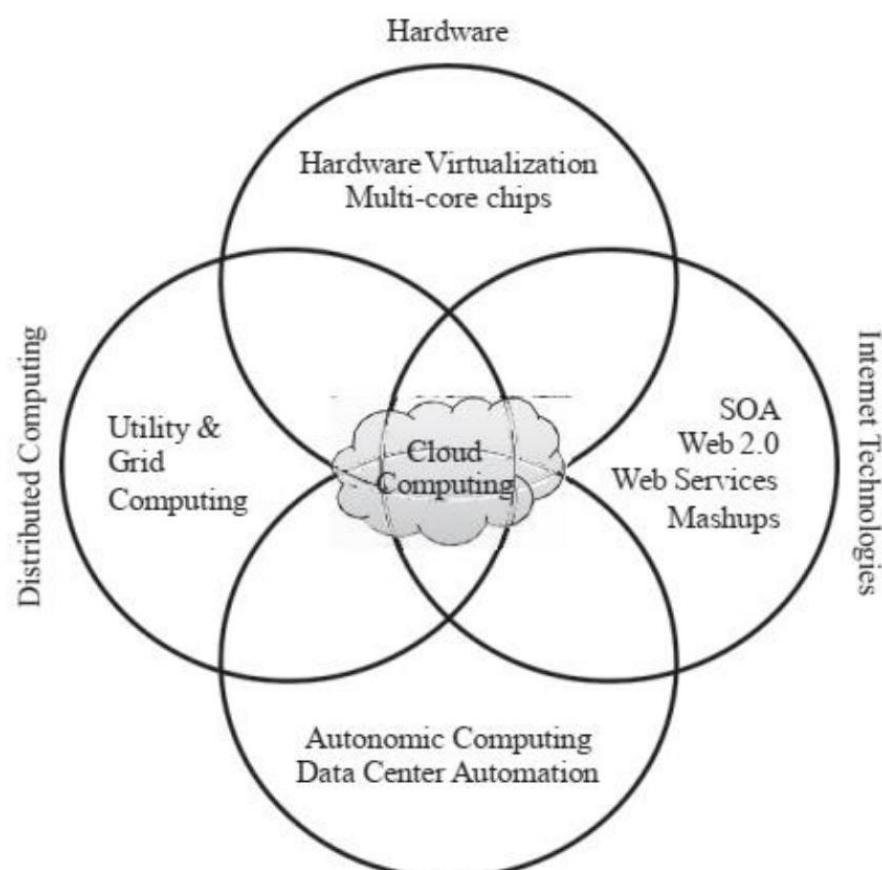
„w chwili obecnej sieci komputerowe są wciąż w powiązach, ale w miarę jak dorosną i staną się wyrafinowane, prawdopodobnie będziemy świadkami rozprzestrzeniania się „urządzeń komputerowych”, które podobnie jak obecne przedsiębiorstwa elektryczne i telefoniczne będą obsługiwać indywidualne domy i biura na całym świecie kraju.”

Ta wizja narzędzia obliczeniowego nabiera kształtu w branży przetwarzania w chmurze w XXI wieku. Świadczenie usług obliczeniowych jest łatwo dostępne na żądanie, podobnie jak inne usługi użyteczności publicznej, takie jak woda, prąd, telefon i gaz, które są dostępne w dzisiejszym społeczeństwie. Podobnie użytkownicy (konsumenci) muszą płacić usługodawcom tylko wtedy, gdy mają oni dostęp do zasobów obliczeniowych. Zamiast utrzymywać własne systemy obliczeniowe lub centra danych, klient może dzierżawić dostęp do aplikacji i pamięci masowej od dostawców usług w chmurze. Zaletą korzystania z usług przetwarzania w chmurze jest to, że organizacje mogą uniknąć początkowych kosztów i trudności związanych z prowadzeniem i zarządzaniem własną infrastrukturą IT oraz płacić za korzystanie z niej. Dostawcy usług w chmurze mogą skorzystać z dużych korzyści skali, oferując te same usługi szerokiej gamie klientów.

W takim przypadku konsumenci mogą uzyskać dostęp do usług zgodnie ze swoimi wymaganiami, wiedząc, gdzie hostowane są wszystkie ich usługi. Modele te można nazwać przetwarzaniem użytkowym, czyli przetwarzaniem w chmurze. Ponieważ przetwarzanie w chmurze nazywane jest przetwarzaniem użytkowym, ponieważ użytkownicy mogą uzyskać dostęp do

infrastruktura jako „chmura” jako aplikacja jako usługi z dowolnego miejsca na świecie. Stąd chmurę obliczeniową można zdefiniować jako nowy dynamiczny model świadczenia usług obliczeniowych, który poprawia wykorzystanie zasobów fizycznych, a centra danych coraz częściej wykorzystują wirtualizację i konwergencję do obsługi wielu różnych systemów działających jednocześnie na platformach serwerowych. Wyniki uzyskane przy różnych schematach rozmieszczenia maszyn wirtualnych będą się znacznie różnić.

Obserwując postęp w kilku technologiach, możemy prześledzić tzw. chmurę obliczeniową (wirtualizacja, chipy wielordzeniowe), szczególnie w sprzęcie; Internet (usługi internetowe, architektury zorientowane na usługi, Web 2.0), przetwarzanie rozproszone (klastry, gridy) i przetwarzanie autonomiczne, automatyzacja centrum danych). Zbieżność rysunku 1.1 ukazuje obszary technologii, które ewoluowały i doprowadziły do pojawienia się przetwarzania w chmurze. Każda z tych technologii na wcześniejszym etapie rozwoju uznawano za spekulację; jednakże później poświęcono im znaczną uwagę. Zakazano ich w środowisku akademickim i dużym przedsiębiorstwom. Dlatego nastąpił proces specyfikacji i standaryzacji, który zaowocował dojrzałością i szerokim przyjęciem. Rozwój chmury obliczeniowej jest ściśle powiązany z dojrzałością tych technologii.



RYSUNEK 1.1. Konwergencja różnych osiągnięć prowadząca do pojawienia się przetwarzania w chmurze

1.2.1 Wizja przetwarzania w chmurze

Wirtualne świadczenie usług w chmurze to sprzęt, środowisko wykonawcze i zasoby udostępniane użytkownikowi w zamian za płatność. Z tych elementów można korzystać tak długo, jak użytkownik, bez wymogu zaangażowania z góry. Cała kolekcja urządzeń komputerowych zostaje przekształcona w zestaw narzędzi, który można dostarczyć i skomponować w ciągu kilku godzin, a nie dni, w celu wdrożenia urządzeń bez ponoszenia kosztów konserwacji. Długoterminowa wizja komputera w chmurze jest taka, że usługi IT są sprzedawane bez technologii i jako narzędzi na otwartym rynku jako bariery dla zasad.

Możemy mieć nadzieję, że w niedalekiej przyszłości uda się zidentyfikować rozwiązanie, które jednoznacznie zaspokoi nasze potrzeby, wprowadzając naszą aplikację na globalny cyfrowy rynek usług w chmurze obliczeniowej. Rynek ten umożliwi automatyzację procesu odkrywania i integracji z istniejącymi systemami oprogramowania. Dostępne usługi cyfrowej platformy handlowej w chmurze umożliwią także usługodawcom zwiększenie przychodów. Usługa w chmurze może również oznaczać obsługę klienta konkurenta mającej na celu wywiązywanie się ze zobowiązań konsumenckich.

Dane firmowe i osobowe są wszędzie dostępne w ustrukturyzowanych formatach, co ułatwia nam dostęp i komunikację na jeszcze większym poziomie. Bezpieczeństwo i stabilność przetwarzania w chmurze będzie w dalszym ciągu poprawiane, czyniąc je jeszcze bezpieczniejszym dzięki szerokiej gamie technik. Zamiast koncentrować się na usługach i aplikacjach, na jakie pozwalają, nie uważamy „chmury” za

najbardziej odpowiednią technologię. Połączenie urządzeń do noszenia i noszenia własnego urządzenia (BYOD) z technologią chmury z Internetem rzeczy (IOT) stałoby się powszechną koniecznością w życiu osobistym i zawodowym, w związku z czym technologia chmury jest pomijana jako czynnik umożliwiający.

BYOD



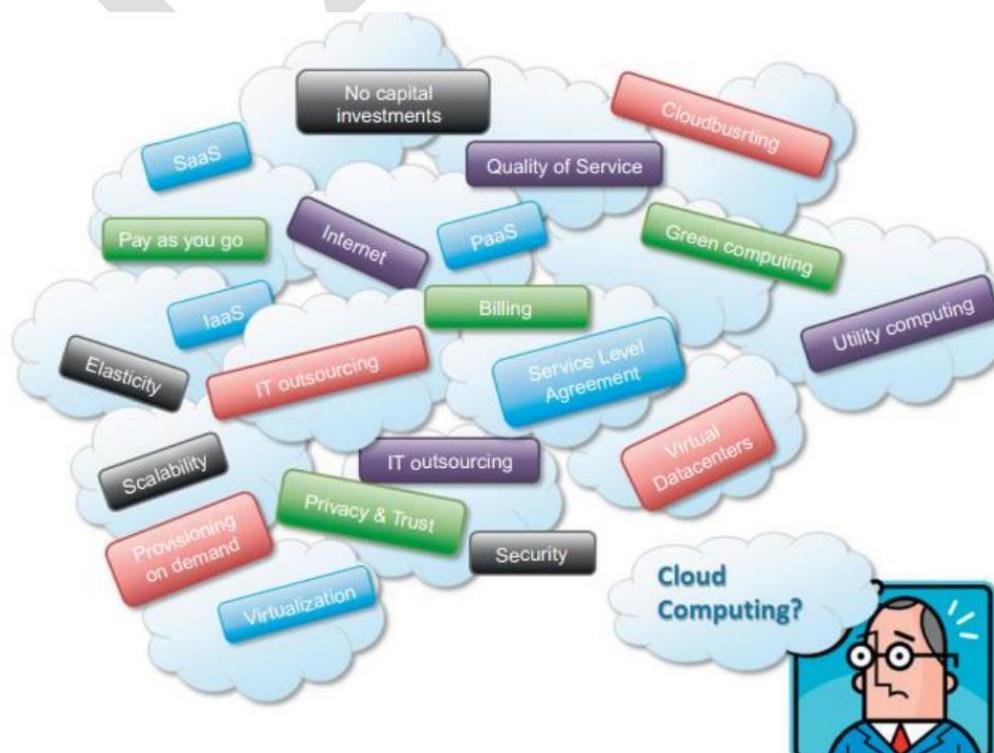
Rysunek 1.2. Wizja przetwarzania w chmurze.

(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming” autorstwa Rajkumara Buyyi)

1.2.2 Definicja chmury

Dość niedawne motto w branży IT „cloud computing”, które powstało po wielu dziesięcioleciach innowacji w zakresie wirtualizacji, obliczeń użytkowych, obliczeń rozproszonych, usług sieciowych i oprogramowania. Chmura tworzy środowisko IT opracowane w celu zdalnego udostępniania mierzonych i skalowalnych zasobów. Ewoluował jako nowoczesny model wymiany informacji i usług internetowych. Zapewnia to bezpieczniejsze, elastyczne i skalowalne usługi dla konsumentów. Jest stosowana jako architektura zorientowana na usługi, która ogranicza ilość informacji obciążających użytkownika końcowego.

Rysunek 1.3 ilustruje różnorodność terminów używanych w obecnych definicjach przetwarzania w chmurze.



RYSUNEK 1.3 Technologie, koncepcje i pomysły przetwarzania w chmurze.

(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming” autorstwa Rajkumara Buyyi)

Internet odgrywa znaczącą rolę w przetwarzaniu w chmurze, ponieważ stanowi środek transportu usług w chmurze, który może być dostarczany i dostępny dla konsumenta w chmurze. Zgodnie z definicją podaną przez Armbrusta

Przetwarzanie w chmurze odnosi się zarówno do aplikacji dostarczanych jako usługi przez Internet, jak i do sprzętu i oprogramowania systemowego w centrach danych, które świadczą te usługi

Powyższa definicja wskazuje na przetwarzanie w chmurze, które obejmuje cały stos, od podstawowego sprzętu po oprogramowanie wysokiego poziomu jako usługę. Wprowadzono koncepcję wszystkiego jako usługi zwanej XaaS, gdzie różne części systemu, takie jak infrastruktura IT, platforma programistyczna dla aplikacji, pamięć masowa, bazy danych itd., mogą być dostarczane jako usługi konsumentom w chmurze, a konsumenti muszą płacić za usługi, czego chcą. Te nowe paradygmaty technologii dotyczą nie tylko rozwoju oprogramowania, ale także sposobu, w jaki użytkownik może wdrożyć aplikację, udostępnić aplikację i zaprojektować infrastrukturę IT, a także sposobu, w jaki firmy alokują koszty na potrzeby IT. Takie podejście zachęca do korzystania z chmury obliczeniowej z globalnego punktu widzenia, w którym jeden użytkownik może przesyłać dokumenty do chmury, a z drugiej strony właściciel firmy chce wdrożyć całą infrastrukturę w chmurze publicznej. Zgodnie z definicją zaproponowaną przez Amerykański Narodowy Instytut Standardów i Technologii (NIST):

Przetwarzanie w chmurze to model umożliwiający wszechobecny, wygodny dostęp sieciowy na żądanie do współdzielonej puli konfigurowalnych zasobów obliczeniowych (np. sieci, serwerów, pamięci masowej, aplikacji i usług), które można szybko udostępnić i udostępnić przy minimalnym wysiłku w zakresie zarządzania lub interakcja z usługodawcą.

Innym podejściem do przetwarzania w chmurze jest „przetwarzanie użytkowe”, w przypadku którego przetwarzanie mogłoby skupiać się głównie na dostarczaniu usług w oparciu o model cenowy, który nazywa się strategią „pay-per-use”. Przetwarzanie w chmurze sprawia, że wszystkie zasoby działają w trybie online, takie jak pamięć masowa, można dzierżawić sprzęt wirtualny lub można wykorzystać zasoby do tworzenia aplikacji, a użytkownicy muszą płacić zgodnie ze swoją wolą, bez lub z minimalną kwotą kosztów początkowych. Wszystkie powyższe operacje są wykonywane, a użytkownik musi zapłacić rachunek, po prostu wprowadzając dane karty kredytowej i uzyskując dostęp do tych usług za pośrednictwem przeglądarek internetowych. Zdaniem George'a Reese'a

Zdefiniował trzy kryteria określające, czy dana usługa jest usługą w chmurze:

- Usługa jest dostępna za pośrednictwem przeglądarki internetowej (niezastrzeżonej) lub interfejsu API usług sieciowych.
- Na początek wymagane są zerowe wydatki kapitałowe.
- Płacisz tylko za to, z czego korzystasz.

Wielu dostawców usług w chmurze świadczy usługi w chmurze użytkownikom bezpłatnie, ale niektóre usługi klasy korporacyjnej mogą być świadczone przez dostawców usług w chmurze w oparciu o określone schematy cenowe, w ramach których użytkownicy muszą wykupić subskrypcję u dostawcy usług, dla którego zdefiniowano umowę o gwarantowanym poziomie usług (SLA). w oparciu o parametry jakościowe pomiędzy dostawcami usług w chmurze a użytkownikami, a dostawcy usług w chmurze muszą świadczyć usługi zgodnie z umową o poziomie usług (SLA)

RajKumar Buyya zdefiniował przetwarzanie w chmurze w oparciu o naturę przetwarzania użytkowego

Chmura to rodzaj równoległego i rozproszonego systemu składającego się ze zbioru wzajemnie połączonych i zwirtualizowanych komputerów, które są dynamicznie udostępniane i prezentowane jako jeden lub więcej ujednoliconych zasobów obliczeniowych w oparciu o umowy o poziomie usług ustalane w drodze negocjacji między dostawcą usług a konsumentami.

1.2.3 Bliższe spojrzenie

Przetwarzanie w chmurze jest przydatne w rządach, przedsiębiorstwach, instytucjach publicznych i prywatnych oraz organizacjach badawczych, które tworzą bardziej efektywne i dostosowane do popytu systemy usług obliczeniowych. Wydaje się, że istnieje wiele konkretnych przykładów demonstrujących pojawiające się zastosowania chmury obliczeniowej zarówno w istniejących firmach, jak i start-upach. Takie przypadki mają na celu zilustrowanie propozycji wartości opłacalnych rozwiązań przetwarzania w chmurze oraz korzyści, jakie przedsiębiorstwa uzyskały dzięki tym usługom.

New York Times: Jeden z najbardziej znanych przykładów zaangażowania w przetwarzanie w chmurze pochodzi z New York Times. „The New York Times” zebrał dużą liczbę zeskanowanych w wysokiej rozdzielcości zdjęć historycznych gazet z lat 1851–1922. Chcia przetworzyć ten zestaw obrazów w osobne artykuły w formacie PDF. Korzystając ze 100 instancji EC2, mogą zakończyć przetwarzanie w ciągu 24 godzin, co daje całkowity koszt 890 USD (czas obliczeń EC2 to 240 USD, transfer danych i wykorzystanie pamięci S3 to 650 USD, przechowywanie i przesyłanie obrazu źródłowego 4,0 TB i 1,5 TB danych wyjściowych [Cloud Computing: wersja nieedytowana s. 6](#)

PDF). Derek Gottfrid zauważał: „Właściwie zadziało to tak dobrze, że uruchomiliśmy go dwukrotnie, ponieważ po zakończeniu znaleźliśmy błąd w pliku PDF”.

„The New York Times” miał możliwość wykorzystania 100 serwerów przez 24 godziny po niskim standardowym koszcie wynoszącym dziesięć centów za godzinę za każdy serwer. W przypadku gdyby New York Times kupił do tego celu choćby pojedynczy serwer, prawdopodobny wydatek przekroczyłby 890 dolarów za sam sprzęt, a firma musiałaby również pomyśleć o kosztach administracji, zasilania i chłodzenia. Podobnie obsługa przejąłaby kontrolę na ponad ćwierć roku z jednym serweryem.

Zakładając, że „New York Times” kupił cztery serwery, jak rozwijał Derek Gottfrid, obliczenia i tak zajęłyby prawie miesiąc. Szybki czas realizacji (wystarczająco szybki, aby dwukrotnie przeprowadzić działalność) i nieskończenie niższy koszt zdecydowanie odzwierciedlały dominującą ocenę usług w chmurze.

Washington Post: W ramach powiązanego, ale nowszego wydarzenia, Washington Post był w stanie w ciągu zaledwie jednego dnia przekształcić 17 481 stron zeskanowanych obrazów dokumentów w bazę danych z możliwością przeszukiwania, korzystając z usługi Amazon EC2. 19 marca o godzinie 10:00 opublikowano publicznie oficjalny harmonogram Hillary Clinton na lata 1993-2001 w postaci dużej liczby zeskanowanych fotografii (w formacie PDF, ale bez możliwości przeszukiwania). Programista Washington Post, Peter Harkins, wykorzystał 200 instancji Amazon EC2 do przeprowadzenia OCR (optycznego rozpoznawania znaków) na zeskanowanych plikach w celu utworzenia tekstu z możliwością przeszukiwania. Wykorzystałem 1407 godzin czasu pracy maszyny wirtualnej, a całkowity koszt wyniósł 144,62 USD. Uważamy, że jest to pozytywny dowód koncepcji.

DISA : Federal Computer Week wspomniał, że Agencja ds. Systemów Informacji Obronnej (DISA) w porównaniu kosztów użytkowania Amazon EC2 z serwerami utrzymywanyymi wewnętrznie: „W najnowszym wydaniu Agencja ds. Systemów Informacji Obronnej porównała cenę rosnącej prosta aplikacja znana jako Tech Early Bird na wewnętrznych serwerach i oprogramowaniu o wartości 30 000 USD z kosztami rozwijania takiej samej aplikacji przy użyciu Amazon Elastic

Oblicz chmurę z Amazon Web Services. Amazon pobierał od dostawcy opłatę w wysokości 10 centów za godzinę, a firma DISA zapłaciła łącznie 5 dolarów za rozszerzenie oprogramowania, które odpowiadało ogólnej wydajności aplikacji wewnętrznej.

SmugMug : SmugMug, witryna internetowa do publikowania i hostowania zdjęć, taka jak Flickr, , przechowuje znaczne poziom informacji o zdjęciach w usłudze przechowywania w chmurze Amazon S3. W 2006 roku udało im się zaoszczędzić „500 000 dolarów na przygotowanych wydatkach na dyski twarde w 2006 roku i zmniejszyć o połowę koszty macierzy dyskowej” dzięki zastosowaniu Amazon S3. Zdaniem dyrektora generalnego SmugMug, dzięki zastosowaniu S3 mogą „z łatwością zaoszczędzić znacznie ponad 1 milion dolarów” w następnym roku. Dyrektor generalny wiedział, że obecna stopa wzrostu firmy w momencie artykułu wymaga zakupu nowego sprzętu o wartości około 80 000 dolarów, a regularne koszty rosną jeszcze bardziej znacznie po umieszczeniu „zasilania, chłodzenia, przestrzeni w centrum informacyjnym oraz siły roboczej, która musiała nimi zarządzać.” Z drugiej strony Amazon S3 kosztuje około 23 000 dolarów miesięcznie za równoważną pamięć masową, która obejmuje wszystko (zasilanie, konserwacja, chłodzenie itp. są wliczane w koszt pamięci masowej).

Eli Lilly : Eli Lilly, jedna z największych firm farmaceutycznych, musi wykorzystać pamięć masową i chmury obliczeniowe firmy Amazon, aby zapewnić wysokowydajne przetwarzanie na żądanie do celów badawczych.

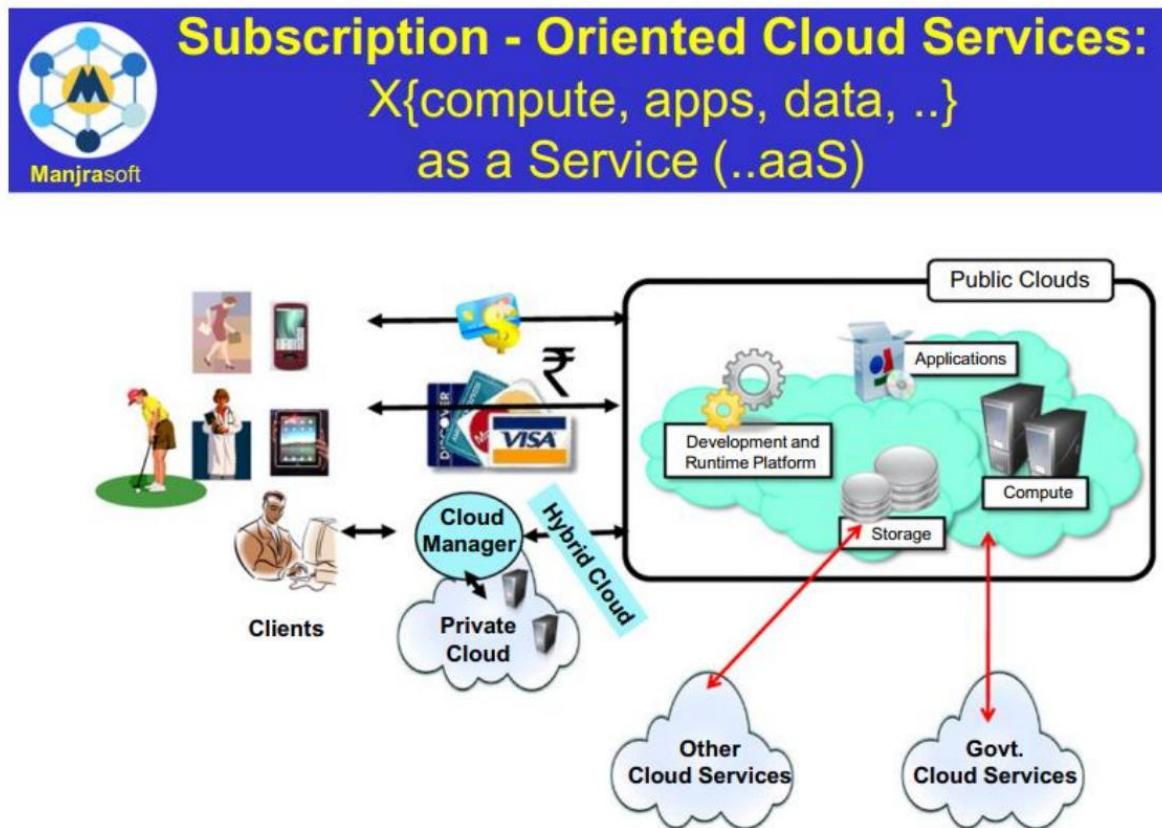
John Foley podkreśla: „Przyzwyczało się, że Eli Lilly potrzebowała siedmiu i pół tygodnia na wewnętrzne wdrożenie serwera”, podczas gdy Amazon może udostępnić serwer wirtualny w 3 minuty. Co więcej, „64-węzłowy klaster Linux mógłby być online w 5 minut (w porównaniu do 90 dni wewnętrznie).” Dostawcy rozwijają chmurowych Amazon nie tylko zapewniają skalowanie na żądanie i rozliczenia oparte na wykorzystaniu, ale umożliwiają firmie Eli Lilly reagowanie ze znacznie większą elastycznością, eliminując czasochłonne funkcje wdrażania i pozyskiwania produktów.

Best Buy's Giftag: Best Buy's Giftag to nowa internetowa usługa list życzeń obsługiwana przez Google App Engine. W wywiadzie wideo programiści zasugerowali, że zaczynają budować platformę z inną technologią i przenieśli się na Google App Engine ze względu na wyjątkową szybkość rozwoju i zalety skalowania. Jak to wymownie stwierdził jeden z programistów, „duża część pracy, której nikt z nas nawet nie musi wykonać, została już za nas wykonana”. Twórcy pochwaliли także projekt App Engine umożliwiający łatwe skalowanie; Aplikacje internetowe oparte na App Engine dziedziczą najlepsze w swojej klasie technologie i wiedzę Google w zakresie prowadzenia witryn internetowych na dużą skalę. Pod koniec dnia App Engine pomaga programistom skoncentrować się na tworzeniu oddzielnych funkcji specyficznych dla danej witryny: „Nie martwiąc się, że aspekty operacyjne aplikacji znikną, zawsze możesz stworzyć doskonały kod lub lepiej go ocenić.

TC3: TC3 (Total Claims Capture & Control) to firma świadcząca usługi opieki zdrowotnej, oferująca rozwiązania do zarządzania roszczeniami. TC3 korzysta teraz z usług chmurowych Amazon, aby umożliwić skalowanie zasobów na żądanie i obniżyć koszty infrastruktury. CTO TC3 zauważa: „Korzystamy z Amazon S3, EC2 i SQS, aby umożliwić wzrost i redukcję naszych możliwości przetwarzania roszczeń w zakresie wymaganym do spełnienia naszych umów o poziomie usług (SLA). Są chwile, gdy potrzebujemy ogromnych ilości zasobów obliczeniowych, które znacznie przekraczają możliwości naszych maszyn i kiedy te [Cloud Computing: Unedited Version str. 7](#)

warunki miały miejsce w przeszłości, naszą naturalną reakcją było wskazanie dostawcy sprzętu w celu uzyskania wyceny. Teraz, korzystając z produktów AWS, możemy radykalnie skrócić czas przetwarzania z tygodni lub miesięcy do dni lub godzin i zapłacić znacznie mniej niż samodzielne zakupy, zakwaterowanie i konserwacja serwerów. Kolejną szczególną cechą działalności TC3 jest że ponieważ świadczą usługi związane ze zdrowiem w USA, są zobowiązani do przestrzegania ustawy HIPPA (ustawy o przenośności i odpowiedzialności w ubezpieczeniach zdrowotnych). Zgodność z przepisami to jedna z głównych przeszkód stojących przed korporacyjnym przyjęciem infrastruktury chmurowej – istotny jest fakt, że TC3 jest w stanie przestrzegać HIPPA na platformie Amazon.

Jak to wszystko stało się możliwe? w tych samych usługach IT na żądanie, takich jak moc obliczeniowa, pamięć masowa i zapewnienie środowisk wykonawczych do tworzenia aplikacji na zasadzie płatności zgodnie z rzeczywistym użyciem. Przetwarzanie w chmurze nie tylko zapewnia możliwość łatwego dostępu do usług IT zgodnie z zapotrzebowaniem, ale także zapewnia nowe pomysły dotyczące usług IT i zasobów jako narzędzi użyteczności publicznej. Rysunek 1.4 przedstawia widok z lotu ptaka na przetwarzanie w chmurze



RYSUNEK 1.4 Widok z lotu ptaka na przetwarzanie w chmurze

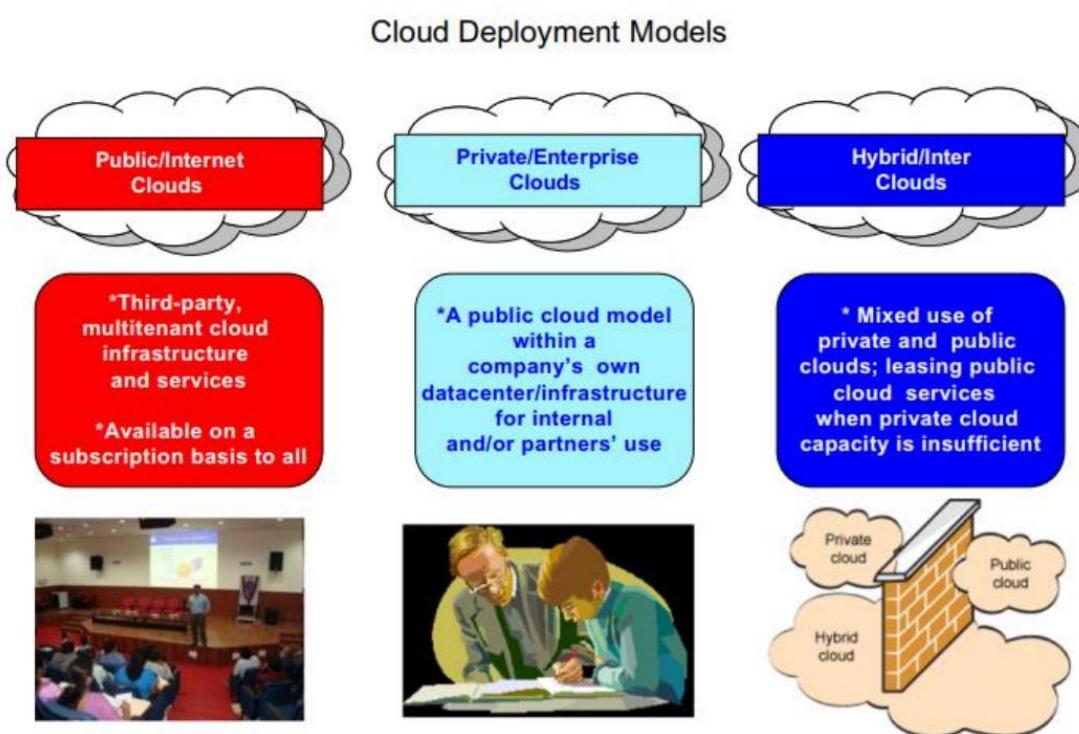
(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming” autorstwa Rajkumara Buyyi)

Istnieją trzy modele wdrażania usług środowiska przetwarzania w chmurze: chmury publiczne, prywatne i hybrydowe (patrz rysunek 1.5). Chmura publiczna to jeden z najpowszechniejszych modeli wdrażania, w ramach którego usługi obliczeniowe oferowane są przez dostawców zewnętrznych, do których konsument może uzyskać dostęp i zakupić zasoby z chmury publicznej za pośrednictwem publicznego Internetu.

Mogą być one bezpłatne lub dostępne na żądanie, co oznacza, że konsumenti płacą za cykle procesora, pamięć masową lub przepustowość za każde użycie. Chmury publiczne uchronią firmy przed kosztownymi zakupami, zarządzaniem i konserwacją na miejscu infrastruktury sprzętowej i aplikacyjnej — za całe zarządzanie i konserwację systemu odpowiada dostawca usług w chmurze.

Chmury publiczne można także wdrażać szybciej niż infrastruktury lokalne, a platforma jest niemal stale skalowalna. Chociaż wdrożenia chmur publicznych stwarzają problemy związane z bezpieczeństwem, chmura publiczna może być tak samo bezpieczna, jak najskuteczniej obsługiwane wdrożenie chmury prywatnej, jeśli zostanie prawidłowo wdrożona. Chmura prywatna to zasadniczo usługa w chmurze świadczona przez jedną organizację. Korzystając z chmury prywatnej, można doświadczać zalet chmury obliczeniowej bez konieczności dzielenia się zasobami z innymi organizacjami. W organizacji może znajdować się chmura prywatna lub być zdalnie kontrolowana przez stronę trzecią i dostępna za pośrednictwem Internetu (ale nie jest ona współdzielona z innymi, w przeciwieństwie do chmury publicznej). Chmura prywatna łączy w sobie kilka zalet przetwarzania w chmurze — w tym elastyczność, skalowalność i łatwe świadczenie usług — z kontrolą na miejscu, bezpieczeństwem i dostosowywaniem zasobów. Wiele firm wybiera chmurę prywatną zamiast chmury publicznej (usługi przetwarzania w chmurze świadczone przez wielu klientów infrastrukturę), ponieważ chmura prywatna to prostszy (lub jedyny sposób) sposób na spełnienie wymagań dotyczących zgodności z przepisami. Inni wolą chmurę prywatną, ponieważ ich zadania dotyczą informacji poufnych, własności intelektualnej i danych osobowych (PII), dokumentacji medycznej, danych finansowych i innych danych wrażliwych. Chmura hybrydowa to infrastruktura zawierająca połączenia między chmurą użytkownika (zwykle nazywaną „chmurą prywatną”) a chmurą strony trzeciej (zwykle nazywaną „chmurą publiczną”).

Chociaż obszary prywatne i publiczne chmury hybrydowej są połączone, pozostają one wyjątkowe. Dzięki temu chmura hybrydowa może jednocześnie oferować zalety kilku modeli wdrożeń. Zaawansowanie chmur hybrydowych jest zupełnie inne. Na przykład niektóre chmury hybrydowe łączą tylko lokalną chmurę z chmurą publiczną. Zespoły operacyjne i aplikacyjne są odpowiedzialne za wszystkie trudności nieodłącznie związane z dwiema różnymi infrastrukturami.



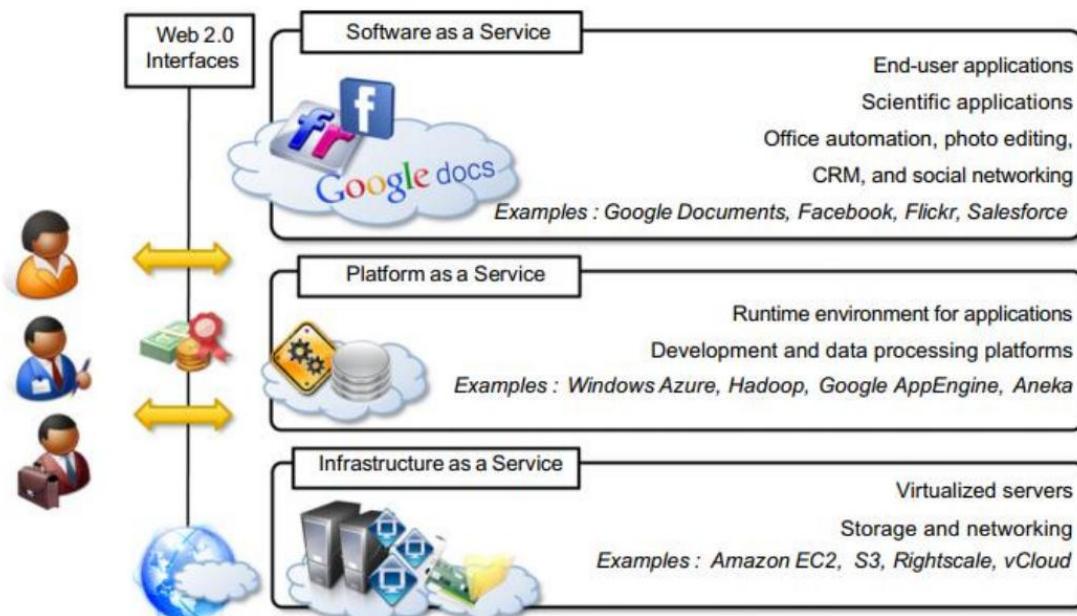
RYSUNEK 1.5 Główne modele wdrażania przetwarzania w chmurze.

Odniesienie z „Opanowania podstaw przetwarzania w chmurze i programowania aplikacji” autorstwa Rajkumara Buyya)

1.2.4 Model referencyjny przetwarzania w chmurze

Modelem charakteryzującym i standaryzującym funkcje środowiska przetwarzania w chmurze jest model referencyjny chmury. Jest to podstawowy punkt odniesienia dla rozwoju przetwarzania w chmurze. Rosnąca popularność przetwarzania w chmurze rozszerzyła definicje różnych architektur przetwarzania w chmurze. Środowisko chmurowe charakteryzuje się szeroką gamą dostawców i wieloma definicjami ofert, co sprawia, że ocena ich usług jest bardzo trudna. Sposób, w jaki chmura funkcjonuje i wchodzi w interakcję z innymi technologiami, może być nieco mylący ze względu na taką złożoność jej wdrożenia.

Aby wykorzystać potencjał chmury obliczeniowej, wymagany jest standardowy model referencyjny chmury dla architektów, inżynierów oprogramowania, ekspertów ds. bezpieczeństwa i przedsiębiorstw. Ten krajobraz chmur jest kontrolowany przez model referencyjny chmury. Rysunek 1.6 przedstawia różnych dostawców usług chmurowych i ich innowacje w modelach usług chmurowych dostępnych na rynku.



RYSUNEK 1.6 Model referencyjny przetwarzania w chmurze.

Odniesienie z „Opanowania podstaw przetwarzania w chmurze i programowania aplikacji” autorstwa Rajkumara Buyya)

Przetwarzanie w chmurze to wszechstronne określenie wszystkich zasobów hostowanych w Internecie. Usługi te są podzielone na trzy główne kategorie: infrastruktura jako usługa (IaaS), platforma jako usługa (PaaS) i oprogramowanie jako usługa (SaaS). Kategorie te są ze sobą powiązane, jak pokazano na rysunku 1.6, który przedstawia organiczny obraz przetwarzania w chmurze. Model organizuje szeroką gamę usług przetwarzania w chmurze w widoku warstwowym od podstawy do szczytu stosu obliczeniowego.

U podstaw stosu infrastruktura jako usługa (IaaS) jest najpopularniejszym modelem usług przetwarzania w chmurze, oferującym podstawową infrastrukturę serwerów wirtualnych, sieci, systemów operacyjnych i dysków pamięci masowej. Zapewnia to elastyczność, niezawodność i skalowalność, których wiele firm szuka w chmurze, i eliminuje potrzebę stosowania sprzętu biurowego. To sprawia, że jest to doskonały sposób na promowanie rozwoju biznesu dla MŚP poszukujących opłacalnego sposobu IT. IaaS to całkowicie zewnętrzna usługa płatna, która może być uruchamiana w infrastrukturze publicznej, prywatnej lub hybrydowej.

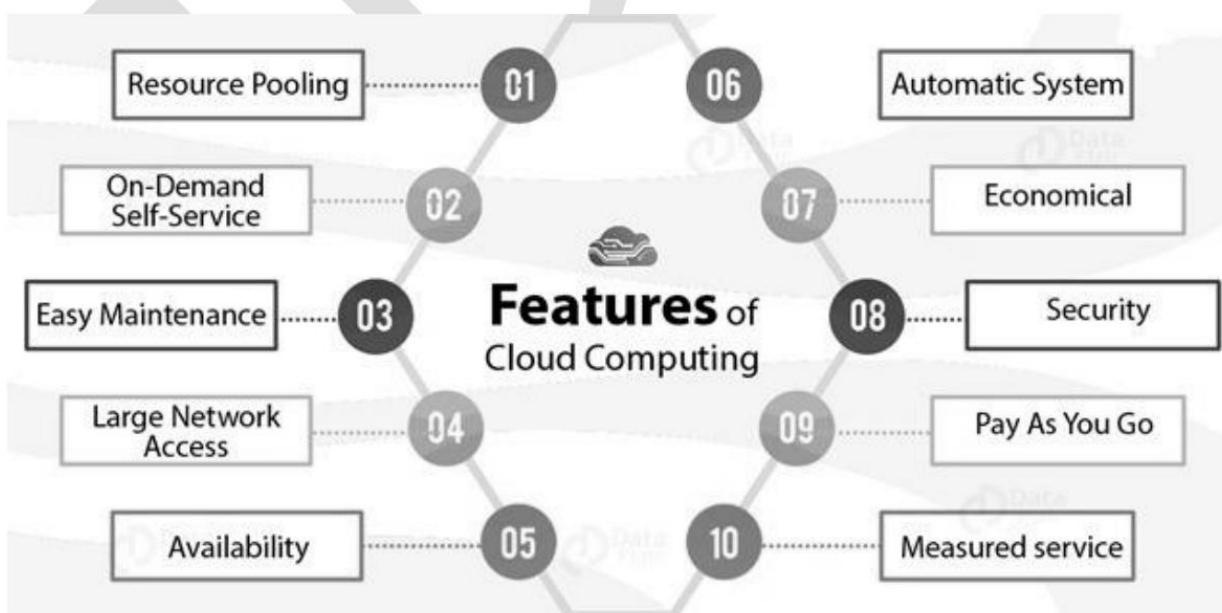
Następnym krokiem w stosie są rozwiązania typu platforma jako usługa (PaaS). Dostawcy usług w chmurze wdrażają ramy oprogramowania i infrastruktury, ale firmy mogą opracowywać i uruchamiać własne aplikacje. Aplikacje internetowe można łatwo i szybko tworzyć za pośrednictwem PaaS, korzystając z elastyczności i niezawodności usługi, która je obsługuje. Rozwiązania PaaS są skalowalne i odpowiednie, jeśli nad jednym projektem pracuje wielu programistów. Jest to również przydatne w przypadku korzystania z ustalonego źródła danych (takiego jak narzędzie CRM).

Szczyt stosu, oprogramowanie jako usługa (SaaS) To rozwiązanie w chmurze obejmuje wdrażanie oprogramowania internetowego dla różnych firm, płacących w ramach subskrypcji lub modelu płatności za użycie. Jest to ważne narzędzie dla CRM i aplikacji wymagających dużego dostępu do Internetu lub urządzenia mobilnego – takie jak oprogramowanie do mobilnego zarządzania sprzedażą. SaaS jest zarządzany z centralnej lokalizacji, dzięki czemu firmy nie muszą martwić się o własne utrzymanie i jest idealny do projektów krótkoterminowych.

Dużą różnicę w kontroli między PaaS i IaaS mają użytkownicy. Zasadniczo PaaS umożliwia dostawcom zarządzanie wszystkim, czego wymaga IaaS w celu lepszego zarządzania klientami. Ogólnie, firmy posiadające pakiet oprogramowania lub aplikację mają już konkretny cel i należy zdecydować się na instalację i uruchomienie go w chmurze IaaS, a nie PaaS.

1.2.5 Charakterystyka i korzyści

Ponieważ usługi w chmurze są dojrzałe zarówno pod względem komercyjnym, jak i technologicznym, firmom łatwiej będzie zmaksymalizować potencjalne korzyści. Jednak równie ważne jest, aby wiedzieć, czym jest przetwarzanie w chmurze i do czego służy.



RYSUNEK 1. 7 Cechy Cloud Computing

Poniżej przedstawiono cechy przetwarzania w chmurze:

1. Łączenie zasobów

Oznacza to, że dostawca chmury zastosował model multileaner w celu dostarczania zasobów obliczeniowych różnym klientom. Istnieją różne przydzielane i ponownie przydzielane zasoby fizyczne i wirtualne, które zależą od zapotrzebowania klientów. Ogólnie rzecz biorąc, klient [Cloud Computing: Unedited Version str. 10](#)

nie ma kontroli ani informacji o lokalizacji dostarczonych zasobów, ale może wybrać lokalizację na wyższym poziomie abstrakcji.

2. Samoobsługa na żądanie

Jest to jedna z głównych i przydatnych zalet Cloud Computing, ponieważ użytkownik może na bieżąco śledzić czas pracy serwera, jego wydajność i pamięć sieciową. Za pomocą tej funkcji użytkownik może również monitorować funkcje obliczeniowe.

3. Łatwa konserwacja

Serwerami można łatwo zarządzać, a przestoje są krótkie i nie występują przestoje, z wyjątkiem niektórych przypadków. Cloud Computing za każdym razem oferuje aktualizację, która coraz bardziej ją ulepsza. Aktualizacje są bardziej przyjazne dla systemu i działają z poprawionymi błędami szybciej niż starsze.

4. Duży dostęp do sieci

Użytkownik może skorzystać z urządzenia i połączenia internetowego, aby uzyskać dostęp do danych w chmurze lub przesyłać je do chmury z dowolnego miejsca. Dostęp do takich możliwości można uzyskać poprzez sieć i Internet.

5. Dostępność

Możliwości chmury można zmieniać i rozszerzać w zależności od sposobu użytkowania. Ta recenzja pomaga konsumentowi w zakupie dodatkowego miejsca w chmurze za bardzo niską cenę, jeśli to konieczne.

6. System automatyczny

Przetwarzanie w chmurze automatycznie analizuje wymagane dane i obsługuje określony poziom usług w zakresie możliwości pomiarowych. Możliwe jest śledzenie, zarządzanie i raportowanie użycia. Zapewnia zarówno gospodarzowi, jak i klientowi odpowiedzialność.

7. Ekonomiczny

Jest to inwestycja jednorazowa, ponieważ firma (host) zobowiązana jest do zakupu przestrzeni dyskowej, którą można udostępnić wielu firmom, co oszczędza hostowi miesięcznych lub rocznych kosztów. Tylko kwota wydana na podstawowe utrzymanie i pewne dodatkowe koszty są duże mniej.

8. Bezpieczeństwo

Cloud Security to jedna z najlepszych funkcji przetwarzania w chmurze. Zapewnia migawkę przechowywanych danych, dzięki czemu nawet w przypadku uszkodzenia jednego z serwerów dane nie mogą zostać utracone. Informacje są przechowywane na urządzeniach pamięci masowej, których żadna inna osoba nie może zhakować ani wykorzystać. Usługa magazynowania jest szybka i niezawodna.

9. Płać według zużycia

Użytkownicy muszą jedynie zapłacić za usługę lub miejsce w chmurze obliczeniowej. Nie ma obowiązku uiszczenia żadnych ukrytych ani dodatkowych opłat. Usługa jest ekonomiczna, a przestrzeń często przydzielana jest bezpłatnie.

10. Usługa zmierzona

Zasoby Cloud Computing, których firma używa do monitorowania i rejestrowania. To wykorzystanie zasobów jest analizowane pod kątem możliwości opłat za użycie. Oznacza to, że dostawca usług może mierzyć i raportować wykorzystanie zasobów albo na instancjach serwerów wirtualnych działających w chmurze. Otrzymasz wynagrodzenie za modele w zależności od rzeczywistego zużycia firmy produkcyjnej.

1.2.6 Wyzwania przed nami

Wszystko ma zalety i wyzwania. Widzieliśmy wiele funkcji chmury i nadszedł czas, aby zidentyfikować wyzwania związane z przetwarzaniem w chmurze za pomocą wskazówek i technik, które możesz zidentyfikować samodzielnie. Zaczniemy zatem badać ryzyko i wyzwania związane z przetwarzaniem w chmurze. Prawie wszystkie firmy korzystają z przetwarzania w chmurze, ponieważ firmy muszą przechowywać dane. Firmy generują i przechowują ogromne ilości danych. Dlatego borykają się z wieloma problemami związanymi z bezpieczeństwem. Firmy obejmowałyby zakłady, aby usprawnić i zoptymalizować proces oraz usprawnić zarządzanie przetwarzaniem w chmurze.

Oto lista wszystkich zagrożeń i wyzwań związanych z przetwarzaniem w chmurze:

1. Bezpieczeństwo i prywatność
2. Interoperacyjność i przenośność
3. Niezawodny i elastyczny
4. Koszt
5. Przestój
6. Brak zasobów
7. Postępowanie w środowiskach wielochmurowych
8. Migracja do chmury
9. Blokada dostawcy
10. Prywatność i kwestie prawne

1. Bezpieczeństwo i prywatność w chmurze

Magazyn danych w chmurze musi być bezpieczny i poufny. Klienci są więc uzależnieni od dostawcy usług w chmurze. Innymi słowy, dostawca chmury musi podjąć środki bezpieczeństwa niezbędne do zabezpieczenia danych klientów. Zabezpieczenia są również obowiązkiem klienta, ponieważ musi on posiadać dobry hasło, nie udostępniać hasła innym osobom i regularnie aktualizować nasze hasło. Jeśli dane znajdują się poza zaporą sieciową, mogą wystąpić pewne problemy, które dostawca chmury może wyeliminować. Hakowanie i złośliwe oprogramowanie są również jednym z największych problemów, ponieważ mogą dotknąć wielu klientów. Może to spowodować utratę danych; zaszyfrowany system plików i kilka innych problemów może zostać zakłóconych.

2. Interoperacyjność i przenośność

Klientowi świadczone są usługi migracji do i z chmury. Nie należy zezwalać na okres kaukcji, ponieważ może to stanowić przeszkodę dla klientów. Chmura będzie w stanie zasilać obiekty lokalowe. Dzalny dostęp jest jedną z przeszkód w chmurze, uniemożliwiającą dostawcy usług w chmurze dostęp do chmury z dowolnego miejsca.

3. Niezawodny i elastyczny

Niezawodność i elastyczność to rzeczywiście trudne zadanie dla klientów chmury, które może wyeliminować wyciek danych dostarczanych do chmury i zapewnić klientom wiarygodność. Aby sprostać temu wyzwaniu, należy monitorować usługi stron trzecich oraz nadzorować wydajność, solidność i zależność firm.

4. Koszt

Przetwarzanie w chmurze jest niedrogie, ale czasami dostosowanie chmury do wymagań klientów może być kosztowne. Ponadto może utrudniać pracę małej firmy, zmieniając chmurę, ponieważ popyt może czasami kosztować więcej. Ponadto czasami przesyłanie danych z chmury do siedziby jest kosztowne.

5. Przestój

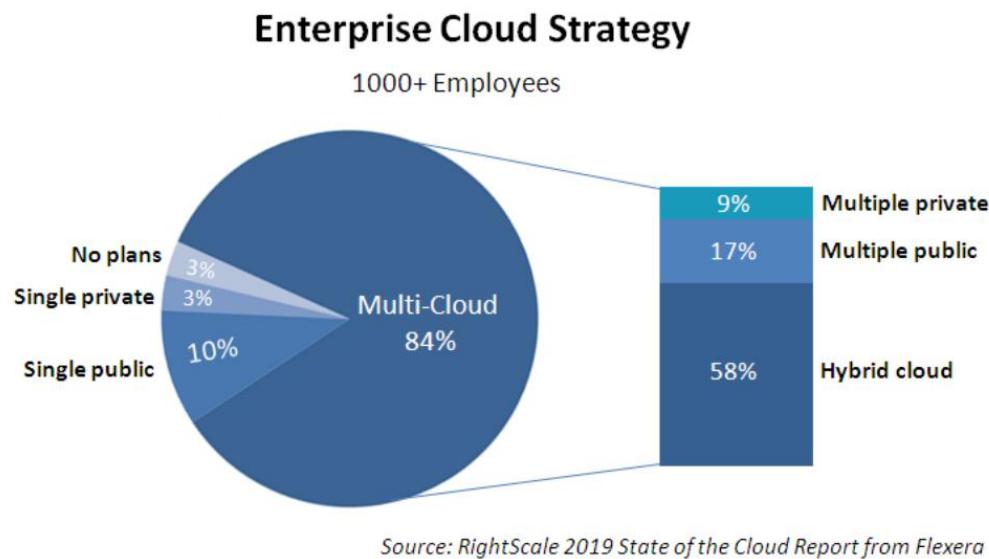
Przestoje to najpopularniejsze wyzwanie związane z przetwarzaniem w chmurze, ponieważ żaden dostawca usług w chmurze nie gwarantuje platformy wolnej od przestojów. Połączenie internetowe również odgrywa ważną rolę, ponieważ może stanowić problem, jeśli firma ma niezaufane połączenie internetowe, ponieważ grozi jej przestój.

6. Brak zasobów

Branża chmurowa również boryka się z brakiem zasobów i wiedzy specjalistycznej, a wiele firm ma nadzieję przezwyciężyć ten problem poprzez zatrudnienie nowych, bardziej doświadczonych pracowników. Pracownicy ci nie tylko pomogą stawić czoła wyzwaniom biznesowym, ale także przeszkołę istniejących pracowników, aby przynieśli korzyści firmie. Obecnie wielu pracowników IT pracuje nad doskonaleniem umiejętności przetwarzania w chmurze, a dyrektorowi naczelnemu jest to trudne, ponieważ pracownicy są słabo wykwalifikowani. Twierdzi, że w przedsiębiorstwach ważniejsi byliby pracownicy mający kontakt z najnowszymi innowacjami i powiązanymi technologiami.

7. Postępowanie w środowiskach wielochmurowych

Dziś nawet jedna chmura nie obsługuje pełnych przedsiębiorstw. Jak wynika z raportu RightScale, prawie 84 procent przedsiębiorstw stosuje podejście wielochmurowe, a 58 procent stosuje podejście oparte na chmurze hybrydowej połączone z chmurami publicznymi i prywatnymi. Ponadto organizacje korzystają z pięciu różnych chmur publicznych i prywatnych.



RYSUNEK 1. 8 Rewelacja raportu RightScale 2019

Zespoły zajmujące się infrastrukturą IT mają większe trudności z długoterminowym przewidywaniem przyszłości technologii chmury obliczeniowej. Specjalisci zaproponowali także najlepsze strategie rozwiązania tego problemu, takie jak ponowne przemyślenie procesów, szkolenie personelu, narzędzia, aktywne zarządzanie relacjami z dostawcami oraz badania.

8. Migracja do chmury

O ile wydanie nowej aplikacji w chmurze jest bardzo proste, o tyle przeniesienie istniejącej aplikacji do środowiska przetwarzania w chmurze jest trudniejsze. Z raportu wynika, że 62% stwierdziło, że ich projekty migracji do chmury są trudniejsze, niż się spodziewali. Ponadto 64% projektów migracyjnych trwało dłużej niż oczekiwano, a 55% przekroczyło budżet. W szczególności organizacje, które migrują swoje aplikacje do chmury, zgłosiły przestoje migracji (37%), problemy z synchronizacją danych przed cięciami (40%), problemy z dobrze działającymi narzędziami do migracji (40%), powolną migrację danych (44%), bezpieczeństwo problemy z konfiguracją (40%) i czasochłonne rozwiązywanie problemów (47%). Aby rozwiązać te problemy, prawie 42% ekspertów IT stwierdziło, że chciałoby zwiększyć swój budżet i że około 45% z nich chciałoby pracować u wewnętrznego specjalisty, 50% chciałoby wydłużyć czas realizacji projektu, 56 % chciałoby więcej testów przedmigracyjnych.

9. Blokada dostawcy

Problem związany z przetwarzaniem w chmurze związanym z zależnością od dostawcy polega na tym, że klienci są zależni (tj. zamknięci) na wdrożeniu jednego dostawcy chmury i nie przechodzą na innego dostawcę bez znaczących kosztów, ograniczeń regulacyjnych lub niezgodności technologicznych w przyszłości. Sytuację blokady można zaobserwować w aplikacjach dla określonych platform chmurowych, takich jak Amazon EC2, Microsoft Azure, których nie można łatwo przenieść na żadną inną platformę chmurową i których użytkownicy są podatni na zmiany wprowadzane przez dostawców w celu dalszego potwierdzenia obiektywności programista. W rzeczywistości kwestia blokady pojawiła się, gdy na przykład firma decyduje się na modyfikację dostawców usług w chmurze (lub być może na integrację usług od różnych dostawców), ale nie może przenosić aplikacji lub danych pomiędzy różnymi usługami w chmurze, ponieważ semantyka dostawców usług w chmurze zasoby i usługi nie odpowiadają sobie. Ta heterogeniczność semantyki chmury i interfejsów API powoduje niezgodność technologiczną, co z kolei prowadzi do kwestionowania interoperacyjności i przenośności. To sprawia, że interoperacyjność, współpraca, przenoszenie, obsługa i konserwacja danych i usług są bardzo skomplikowane i trudne. Z tych powodów z punktu widzenia przedsiębiorstwa ważne jest zachowanie elastyczności w zmianie dostawców zgodnie z potrzebami biznesowymi lub nawet utrzymywanie we własnym zakresie niektórych komponentów, które są mniej krytyczne dla bezpieczeństwa ze względu na ryzyko. Kwestia uzależnienia od dostawców uniemożliwi interoperacyjność i przenośność między dostawcami usług w chmurze. Jest to sposób, w jaki dostawcy i klienci usług w chmurze mogą stać się bardziej konkurencyjni.

10. Prywatność i kwestie prawne

Najwyraźniej głównym problemem dotyczącym prywatności/bezpieczeństwa danych w chmurze jest „naruszenie danych”.

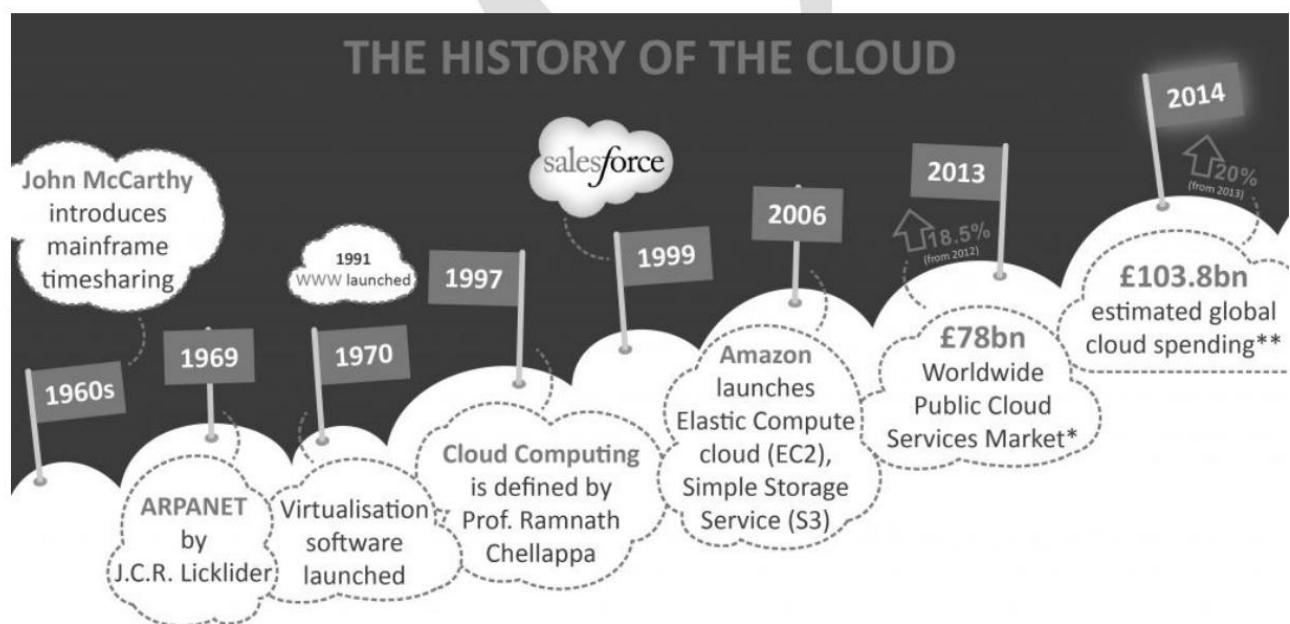
Naruszenie danych można ogólnie zdefiniować jako utratę elektronicznie zaszyfrowanych danych osobowych. Naruszenie informacji może prowadzić do licznych strat zarówno dla dostawcy, jak i klienta; kradzież tożsamości, oszustwo związane z kartą debetową/kredytową klienta, utrata wiarygodności, przyszłe ściganie i tak dalej. W przypadku naruszenia danych prawo amerykańskie wymaga powiadomienia osób, których dane dotyczą, o naruszeniu danych. Prawie każdy stan w USA musi obecnie zgłaszać osobom dotkniętym naruszenia danych. Problemy pojawiają się, gdy dane podlegają kilku jurysdykcji, a przepisy dotyczące prywatności danych różnią się. Na przykład dyrektywa Unii Europejskiej o ochronie danych wyraźnie stwierdza, że „dane mogą opuścić UE jedynie wtedy, gdy trafią do kraju o „dodatkowym poziomie bezpieczeństwa”. Zasada ta, choć prosta do wdrożenia, ogranicza przepływ danych, a tym samym zmniejsza pojemność danych. Regulacje UE można egzekwować.

1.3 Wydarzenia historyczne

Żadna najnowocześniejsza technologia nie jest chmurą obliczeniową. Nastąpił rozwój przetwarzania w chmurze w różnych fazach, w tym przetwarzania sieciowego, przetwarzania użytkowego, świadczenia usług aplikacji i oprogramowania jako usługi itp. Jednak ogólna (cała) koncepcja udostępniania zasobów obliczeniowych za pośrednictwem sieci globalnej rozpoczęła się w latach sześćdziesiątych XX wieku. Prognozuje się, że do 2020 roku wartość rynku przetwarzania w chmurze przekroczy 241 miliardów dolarów. Ale historia przetwarzania w chmurze pokazuje, jak tam dotarliśmy i od czego wszystko się zaczęło. Przetwarzanie w chmurze ma niezbyt długą historię. Pierwsza witryna internetowa poświęcona chmurze dla przedsiębiorstw i konsumentów została uruchomiona w 1999 r. (Salesforce.com i Google). Przetwarzanie w chmurze jest bezpośrednio powiązane z rozwojem Internetu i rozwojem technologii korporacyjnej, ponieważ przetwarzanie w chmurze jest odpowiedzią na problem, w jaki sposób Internet może ulepszyć technologię korporacyjną. Technologia biznesowa ma bogatą i interesującą przeszłość, prawie tak długą jak same przedsiębiorstwa, ale rozwój, który wpłynął najbardziej bezpośrednio na przetwarzanie w chmurze, rozpoczyna się wraz z pojawieniem się komputerów jako dostawców rzeczywistych rozwiązań biznesowych.

Historia przetwarzania w chmurze

Przetwarzanie w chmurze to obecnie jedna z najbardziej przełomowych technologii. Następnie krótka historia przetwarzania w chmurze.



RYSUNEK 1. 9 Historia przetwarzania w chmurze [*Gartner, **Constellation Research]

WCZESNE LATA 60. WIEKU

Informatyk John McCarthy ma koncepcję podziału czasu, która pozwala organizacji na jednoczesne korzystanie z drogiego komputera typu mainframe. Maszyna ta jest opisywana jako znaczący wkład w rozwój Internetu i lider w dziedzinie przetwarzania w chmurze.

W 1969r

JCR Licklider, odpowiedzialny za utworzenie Agencji Zaawansowanych Projektów Badawczych (ARPANET), zaproponował ideę „Miedzygalaktycznej Sieci Komputerowej” lub „Sieci Galaktycznej” (termin sieci komputerowej podobny do dzisiejszego Internetu). Jego wizją było połączenie wszystkich na całym świecie i dostęp do programów i danych z dowolnego miejsca.

W 1970 r

Wykorzystanie narzędzi takich jak VMware do wirtualizacji. W oddzielnym środowisku można jednocześnie uruchomić więcej niż jeden system operacyjny. W innym systemie operacyjnym było to możliwe

obsługiwać zupełnie inny komputer (maszynę wirtualną).

W 1997

Wydaje się, że profesor Ramnath Chellappa w Dallas w 1997 r. jest pierwszą znaną definicją „przetwarzania w chmurze” – „paradygmatu, w którym granice przetwarzania są definiowane wyłącznie na podstawie ograniczeń ekonomicznych, a nie technicznych”.

W 1999

Salesforce.com powstał w 1999 roku jako pionier w dostarczaniu aplikacji klienckich za pośrednictwem swojej prostej strony internetowej. Ta firma usługowa była w stanie dostarczać aplikacje za pośrednictwem Internetu zarówno dla specjalistycznych, jak i głównych firm zajmujących się oprogramowaniem.

W 2003

To pierwsze publiczne wydanie Xen to system oprogramowania, który umożliwia jednoczesne uruchamianie wielu wirtualnych systemów operacyjnych gościa na jednej maszynie, zwanej również monitorem maszyny wirtualnej (VMM) jako hiperwizorem.

W 2006 ROKU

Usługa chmurowa Amazon została uruchomiona w 2006 r. Po pierwsze, Elastic Compute Cloud (EC2) umożliwiła użytkownikom korzystanie z własnych aplikacji w chmurze i dostęp do komputerów. Następnie wypuszczono usługę Simple Storage Service (S3). Objęło to model użytkownika na bieżąco i stało się standardową procedurą zarówno dla użytkowników, jak i całej branży.

W 2013

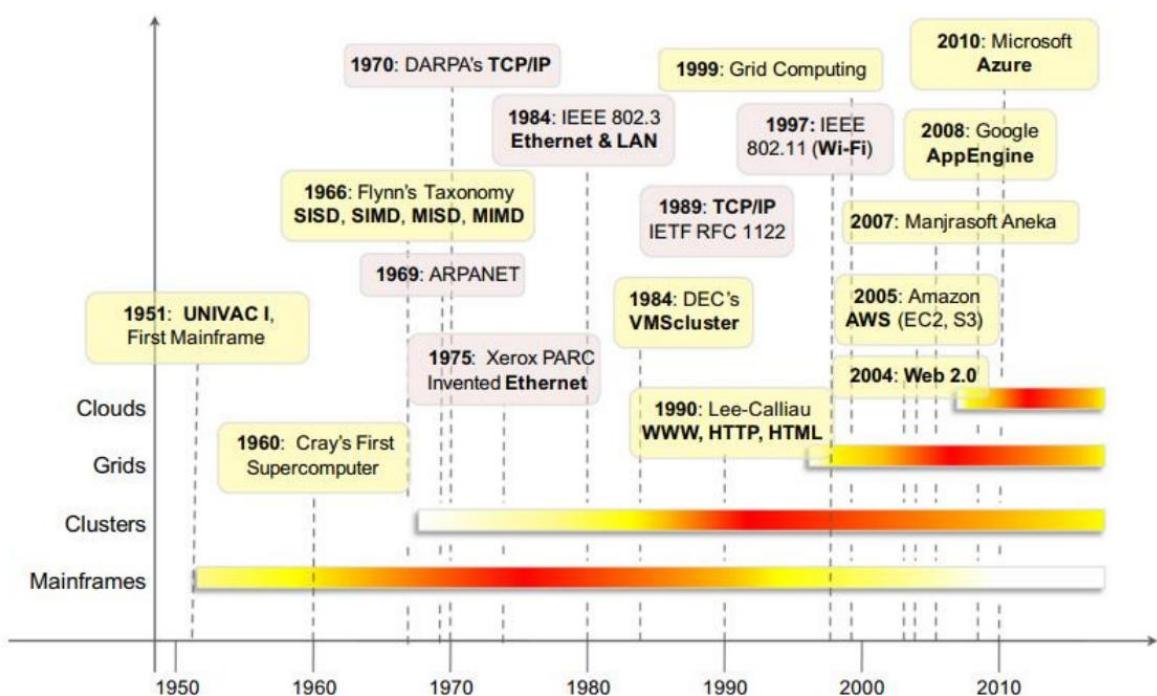
Łączna wartość światowego rynku usług chmury publicznej o wartości 78 miliardów funtów wzrosła w 2012 roku o 18,5%, a IaaS stał się jedną z najszybciej rozwijających się usług na rynku.

W 2014

Szacuje się, że globalne wydatki przedsiębiorstw na technologie i usługi związane z chmurą w 2014 r. wyniosą 103,8 miliarda funtów, co oznacza wzrost o 20% w porównaniu z 2013 r. (Constellation Research).

Rysunek przedstawia analizę rozwoju technologii rozproszonych w chmurze obliczeniowej.

Śledząc rozwój wydarzeń historycznych, dokonamy krótkiego przeglądu pięciu kluczowych technologii, które odegrały znaczącą rolę w przetwarzaniu w chmurze. Są to systemy rozproszone, wirtualizacja, Web 2.0, zorientowanie na usługi i przetwarzanie użytkowe.



RYSUNEK 1.10: Ewolucja technologii przetwarzania rozproszonego, lata 50. – 2010. XX w.

Odniesienie z „Opanowania podstaw przetwarzania w chmurze i programowania aplikacji” autorstwa Rajkumara Buyya)

Przetwarzanie rozproszone to koncepcja komputera, która w większości przypadków odnosi się do wielu systemów komputerowych pracujących nad jednym problemem. Pojedynczy problem w przetwarzaniu rozproszonym został uszkodzony

na wiele części, a każdą część rozwiązuje różne komputery. Gdy komputery są ze sobą połączone, mogą komunikować się ze sobą w celu rozwiązania problemu. Jeśli zostanie wykonany prawidłowo, komputer będzie funkcjonował jako pojedyncza jednostka.

Ostatecznym celem przetwarzania rozproszonego jest poprawa ogólnej wydajności poprzez opłacalne, przejrzyste i bezpieczne połączenia pomiędzy użytkownikami i zasobami IT. Zapewnia także tolerancję na defekty i zapewnia dostęp do zasobów w przypadku awarii jednego komponentu.

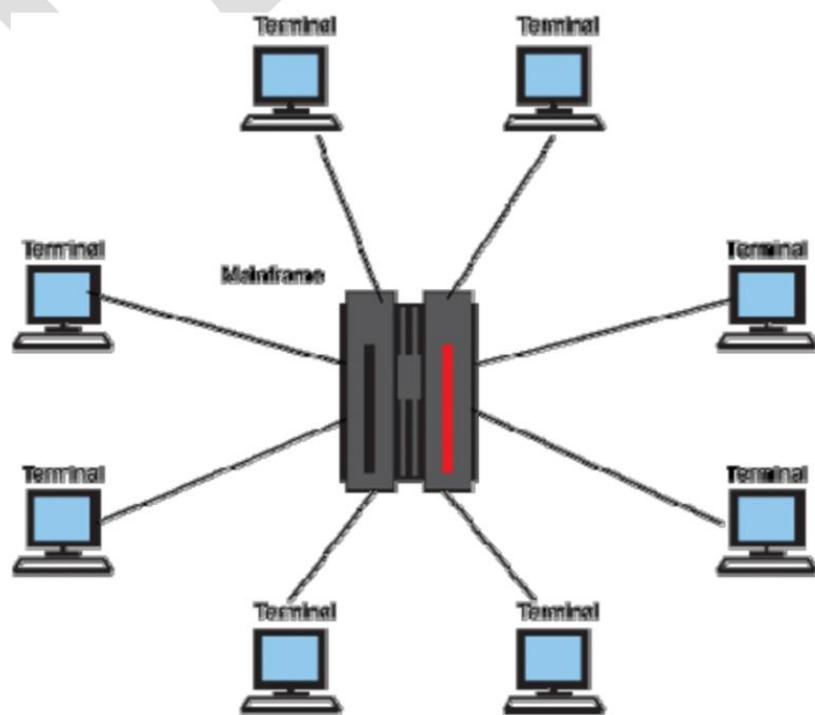
Naprawdę nie ma nic specjalnego w dystrybucji zasobów w sieci komputerowej. Zaczęło się od użycia terminali typu mainframe, następnie przeniosło się na minikomputery i jest teraz możliwe w komputerach osobistych i architekturze serwerów klienckich z kilkoma warstwami.

Rozproszona architektura komputera składa się z szeregu bardzo lekkich komputerów klienckich zainstalowanych z jednym lub kilkoma dedykowanymi serwerami do zarządzania komputerem. Agenci klienta zwykle rozpoznają, kiedy komputer jest bezczynny, dzięki czemu serwer zarządzania jest powiadamiany, że komputer nie jest używany lub że jest dostępny. Następnie agent prosi o paczkę. Kiedy ten pakiet aplikacji jest dostarczany z serwera zarządzania do klienta i ma wolne cykle procesora, oprogramowanie uruchamia aplikację i zwraca wyniki do serwera zarządzania. Gdy użytkownik wróci, serwer zarządzania zwróci zasoby użyte do wykonania szeregu zadań pod nieobecność użytkownika.

Systemy rozproszone charakteryzują się heterogenicznością, otwartością, skalowalnością, przejrzystością, współbieżnością, ciągłą dostępnością i niezależnymi awariami. Charakteryzują one w pewnym stopniu chmury, szczególnie pod względem skalowalności, współbieżności i ciągłej dostępności.

Przetwarzanie w chmurze przyczyniło się do powstania trzech głównych kamieni milowych: komputerów mainframe, przetwarzania klastrowego i przetwarzania sieciowego.

Komputery mainframe: Mainframe to potężny komputer, który często służy jako główne repozytorium danych dla infrastruktury IT organizacji. Łączy się z użytkownikami za pośrednictwem mniej wydajnych urządzeń, takich jak stacje robocze czy terminale. Łatwiej jest zarządzać, aktualizować i chronić integralność danych poprzez centralizację danych w jednym repozytorium na komputerze mainframe. Komputery mainframe są zwykle używane w procesach na dużą skalę, które wymagają większej dostępności i bezpieczeństwa niż mniejsze maszyny. Komputery typu mainframe lub komputery mainframe to przede wszystkim maszyny do podstawowych celów, używane przez duże organizacje; masowe przetwarzanie danych, na przykład spisy ludności, statystyki branżowe i konsumenckie, planowanie zasobów przedsiębiorstwa i przetwarzanie transakcji. Pod koniec lat pięćdziesiątych komputery typu mainframe miały jedynie podstawowy interfejs interaktywny, wykorzystujący karty dziurkowane, taśmę papierową lub taśmę magnetyczną do transmisji danych i programów. Pracowali w trybie wsadowym, obsługując funkcje zaplecza biurowego, takie jak płace i rozliczenia z klientami, głównie w oparciu o powtarzalne taśmy i operacje łączenia, po których następował druk liniowy do ciągłego, stacjonarnego, wstępnie zadrukowanego. Wprowadzenie cyfrowych interfejsów użytkownika służących niemal wyłącznie do wykonywania aplikacji (np. rezerwacji linii lotniczych), a nie do tworzenia oprogramowania. Maszyny do pisania i dalekopisy były standardowymi konsolami sterującymi operatorów sieci na początku lat 70., chociaż w dużej mierze zastąpiono je klawiaturami.



RYSUNEK 1.11 Komputery mainframe

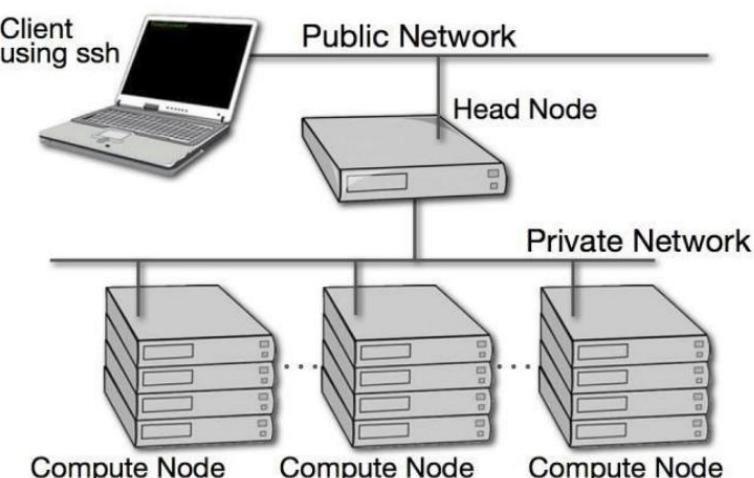
Przetwarzanie klastrowe: podejście do klastrowania komputerów zazwyczaj łączy jakiś komputer

węzły (komputer osobisty używany jako serwer) gotowe do pobrania poprzez szybką lokalną sieć strefową (LAN). Aktywność węzła obliczeniowego koordynowana jest przez oprogramowanie „oprogramowanie pośrednie klastra” – warstwę oprogramowania umieszczoną przed węzłami, która umożliwia użytkownikom dostęp do klastra jako całości za pomocą koncepcji obrazu pojedynczego systemu. Klaster to rodzaj systemu komputerowego, który jest równoległy lub rozproszony i który składa się ze zbioru połączonych ze sobą niezależnych komputerów, współpracujących jako wysoce skonsolidowane narzędzie obliczeniowe, integrujące oprogramowanie i sieć z niezależnymi komputerami w jednym systemie. Klastry są zwykle używane w celu zapewnienia większej mocy obliczeniowej niż może zapewnić pojedynczy komputer w celu zapewnienia wysokiej dostępności, większej niezawodności lub wysokiej wydajności obliczeniowej. W porównaniu z innymi technologiami technika klastrowa jest ekonomiczna pod względem mocy i prędkości przetwarzania, ponieważ wykorzystuje sprzęt i oprogramowanie z półki w porównaniu z komputerami typu mainframe, które korzystają z własnego sprzętu i komponentów oprogramowania budowanych na zamówienie. Wiele komputerów w klastrze współpracuje ze sobą, aby zapewnić ujednolicone i szybsze przetwarzanie. Klaster można uaktualnić do wyższej specyfikacji lub rozszerzyć poprzez dodanie dodatkowych węzłów w przeciwieństwie do komputera typu mainframe. Redundantne maszyny przejmujące przetwarzanie w sposób ciągły minimalizują awarie pojedynczych komponentów. W przypadku aplikacji mainframe ten rodzaj redundancji nie jest dostępny.

PVM i MPI to dwie metody najczęściej stosowane w komunikacji klastrowej.

PVM to równoległa maszyna wirtualna. W Oak Ridge National Laboratory opracowano PVM około 1989 roku. Jest on instalowany bezpośrednio na każdym węźle i zapewnia zestaw bibliotek, które przekształcają węzeł w „równoległą maszynę wirtualną”. Oferuje środowisko wykonawcze do kontroli zarządzania zasobami i zadaniami, raportowania błędów i przekazywania komunikatów. C, C++ lub Fortran mogą używać PVM w programach użytkownika.

MPI to interfejs przekazywania komunikatów. W latach 90-tych utworzono i zastąpiono PVM. Podstawą projektowania MPI są różne dostępne wówczas na rynku systemy. Do implementacji zazwyczaj wykorzystuje połączenia TCP/IP i gniazda. Obecnie najczęściej używany system komunikacji pozwala na równoległe planowanie w językach C, Fortran, Python itp.



RYSUNEK 1.12 Obliczenia klastrowe

Obliczenia siatkowe: Jest to architektura procesora, która łączy zasoby komputerowe z różnych dziedzin, aby osiągnąć główny cel. Komputery sieciowe mogą współpracować przy zadaniu roboczym w przetwarzaniu siatkowym i dlatego działają jako superkomputer. Ogólnie rzecz biorąc, siatka spełnia kilka zadań w sieci, ale może również działać w określonych zastosowaniach. Ogólnie rzecz biorąc, siatka obsługuje różne zadania sieciowe, ale może również działać w określonych zastosowaniach. Ma na celu rozwiązywanie problemów, które są zbyt duże dla superkomputera i zachowanie możliwości obsługi kilku małych problemów. Siatki komputerowe obejmują sieć wielu użytkowników, która spełnia wymagania dotyczące nieciągłego przetwarzania informacji.

Siatka jest połączona z klastrem komputerowym, który działa na systemie operacyjnym, Linuxie lub darmowym oprogramowaniu, wykorzystując równolegle węzły. Klaster może mieć różną wielkość, od małych do wielu sieci. Technologia ta jest wykorzystywana za pośrednictwem kilku zasobów obliczeniowych w różnorodnych zastosowaniach, takich jak zadania matematyczne, badawcze lub edukacyjne. Jest często używany w analizie strukturalnej, a także w usługach internetowych, takich jak bankowość bankomatowa, infrastruktura back-office oraz badania naukowe i marketingowe. Przetwarzanie siatkowe składa się z aplikacji używanych w środowisku sieci równoległej w celu rozwiązywania problemów obliczeniowych. Łączy każdy komputer PC i łączy informacje w aplikację obliczeniową.

Siatki dysponują szeregiem zasobów, czy to za pośrednictwem sieci, czy też otwartych standardów z jasnymi wytycznymi umożliwiającymi osiągnięcie wspólnych celów i zadań, w oparciu o różne struktury oprogramowania i sprzęt, języki komputerowe i frameworki.

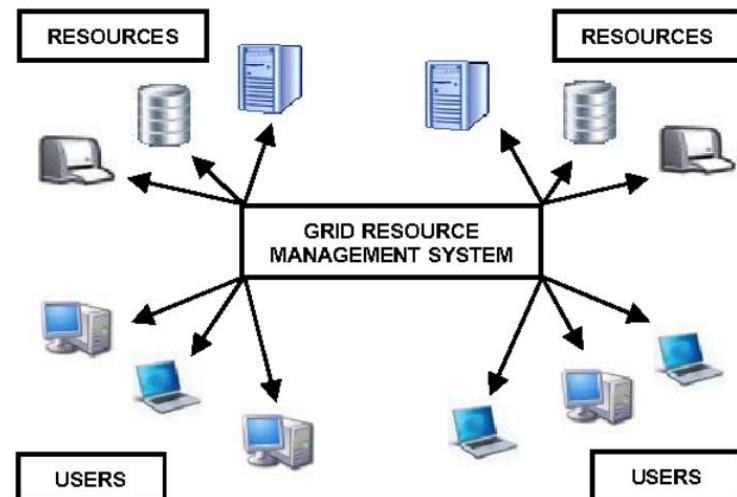
Ogólnie operacje na siatce dzielą się na dwie kategorie:

Data Grid: system obsługujący duże rozproszone zbiory danych wykorzystywane do kontrolowania danych i udostępniania użytkowników. Buduje środowiska wirtualne, które ułatwiają rozproszone i zorganizowane badania. Przykładem siatki danych jest Centrum trzęsień ziemi w południowej Kalifornii, które wykorzystuje środkową platformę oprogramowania do tworzenia biblioteki cyfrowej, rozproszonego systemu plików i ciągłego archiwum.

Siatki oczyszczania procesora: system oczyszczania cyklu, który w razie potrzeby przenosi projekty z jednego komputera na drugi. Poszukiwanie obliczeń inteligencji pozaziemskiej, obejmujące ponad 3 miliony komputerów, reprezentują znajomą siatkę oczyszczania procesora. Wykrywanie sygnałów radiowych w ramach poszukiwań inteligencji pozaziemskiej (SETI) to jedno z najbardziej eksytujących zastosowań radioastronomii. Pod koniec lat pięćdziesiątych pierwszy zespół SETI używał anteny radioastronomicznej. Kilka lat później utworzono finansowany ze środków prywatnych Instytut SETI, aby przeprowadzić więcej poszukiwań za pomocą kilku amerykańskich radioteleskopów. Dziś, we współpracy z inżynierami radioastronomii oraz badaczami z różnych obserwatoriów i uniwersytetów, Instytut SETI ponownie tworzy własną kolekcję środków prywatnych. Ogromna moc obliczeniowa SETI doprowadziła do powstania unikalnej koncepcji obliczeń siatkowych, która została obecnie rozszerzona na wiele zastosowań.

SETI@home to eksperyment naukowy wykorzystujący komputery podłączone do Internetu do pobierania i analizowania danych radioteleskopu programu SETI. Wolny program wykorzystuje moc milionów komputerów i wykorzystuje moc nieaktywnego komputera, aby działać w tle. Ponad dwa miliony lat łącznego czasu przetwarzania miało miejsce ponad 5,2 miliona uczestników.

Obliczenia siatkowe są wykorzystywane w biologii, medycynie, naukach o Ziemi, fizyce, astronomii, chemii i matematyce. Berkeley Open Infrastructure for Network Computing (BOINC) to bezpłatne oprogramowanie typu open source do komputerów stacjonarnych i komputerów stacjonarnych. Korzystając z platformy BOINC, użytkownicy mogą dzielić zadania pomiędzy kilka projektów przetwarzania gridowego i decydować o przydzieleniu im tylko jednego procenta czasu procesora.



RYSUNEK 1.13 Środowisko obliczeń siatkowych

1.3.2 Wirtualizacja

Wirtualizacja to proces, który zwiększa efektywność wykorzystania fizycznego sprzętu komputerowego i stanowi podstawę przetwarzania w chmurze. Wirtualizacja wykorzystuje oprogramowanie do utworzenia warstwy abstrakcji na sprzęcie komputerowym, umożliwiając wielu komputerom wirtualnym, zwykle nazywanym maszynami wirtualnymi, rozdzielenie elementów sprzętowych z jednego komputera — procesorów, pamięci, pamięci masowej i innych. Każda maszyna wirtualna ma własny system operacyjny i działa jak komputer autonomiczny, biorąc pod uwagę fakt, że działa tylko na części podstawowego sprzętu komputerowego.

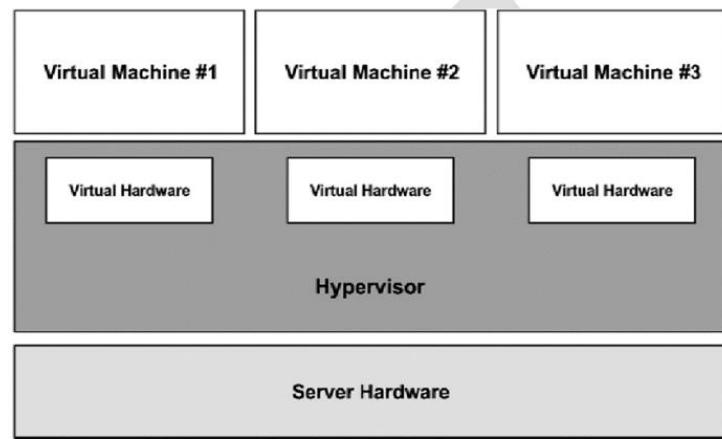
Wirtualizacja ułatwia zatem znacznie bardziej efektywne wykorzystanie fizycznego sprzętu komputerowego, umożliwiając w ten sposób większy zwrot z inwestycji sprzętowej organizacji.

Wirtualizacja jest dziś powszechną praktyką w architekturze IT dla firm. Jest to również technologia napędzająca działalność w zakresie przetwarzania w chmurze. Wirtualizacja umożliwia dostawcom usług w chmurze świadczenie usług klientom za pomocą ich własnego fizycznego sprzętu komputerowego, a użytkownikom chmury zakup tylko tych zasobów komputerowych, których potrzebują, kiedy ich potrzebują, oraz ekonomiczne skalowanie ich w miarę wzrostu obciążenia.

Wirtualizacja polega na utworzeniu wirtualnej platformy, obejmującej wirtualny sprzęt komputerowy, wirtualne urządzenia pamięci masowej i wirtualne sieci komputerowe.

Do wirtualizacji sprzętu służy oprogramowanie zwane hypervisorem. Za pomocą hypervisor maszyny wirtualnej oprogramowanie jest włączane do komponentu sprzętowego serwera. Rolą hypervisor jest kontrolowanie sprzętu fizycznego współdzielonego między klientem a dostawcą. Wirtualizacji sprzętu można dokonać za pomocą programu Virtual Machine Monitor (VMM) w celu usunięcia sprzętu fizycznego. Istnieje kilka rozszerzeń procesów, które pomagają przyspieszyć działania wirtualizacyjne i zwiększyć wydajność hypervisorów. Kiedy taka wirtualizacja jest wykonywana dla platformy serwerowej, nazywa się to socjalizacją serwera.

Hypervisor tworzy abstrakcyjną warstwę pomiędzy oprogramowaniem a używanym sprzętem. Po zainstalowaniu hypervisorów powstają wirtualne reprezentacje, takie jak wirtualne procesory. Po instalacji nie możemy używać procesorów fizycznych. Istnieje kilka popularnych hypervisorów, w tym VMware vSphere i Hyper-V oparte na ESXi.



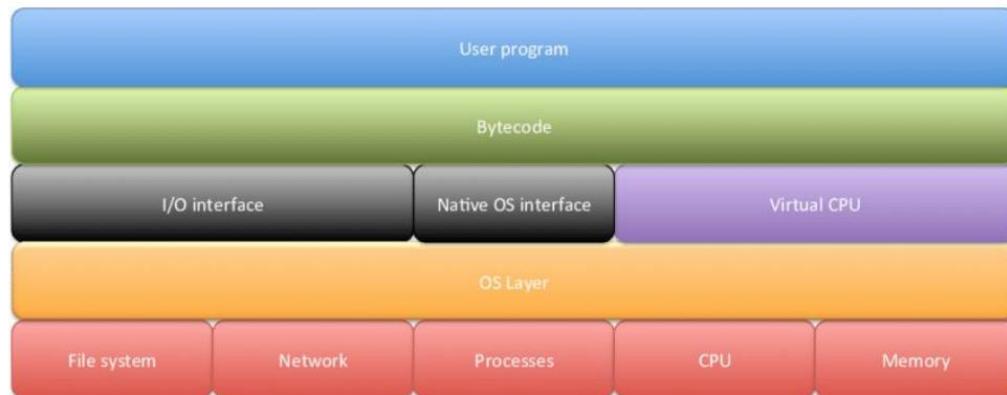
RYSUNEK 1.14 Wirtualizacja sprzętowa

Instancje na maszynach wirtualnych są zazwyczaj reprezentowane przez jeden lub więcej danych, które można łatwo przenosić w strukturach fizycznych. Ponadto są również autonomiczne, ponieważ nie mają innych zależności do użytku niż menedżer maszyny wirtualnej.

Maszyna wirtualna procesu, czasami nazywana maszyną wirtualną aplikacji, działa w systemie operacyjnym hosta jako wspólna aplikacja obsługująca pojedynczy proces. Tworzony jest na początku i na końcu procesu. Jego celem jest zapewnienie niezależnego od platformy środowiska programistycznego, które wyodrębnia informacje o sprzęcie lub systemie operacyjnym leżącym u podstaw programu i umożliwia jego działanie na dowolnej platformie w ten sam sposób. Na przykład oprogramowanie Wine dla systemu Linux pomaga w uruchomieniu systemu Windows.

Abstrakcją wysokiego poziomu procesu VM jest język programowania wysokiego poziomu (w porównaniu z abstrakcją ISA niskiego poziomu systemu VM). Maszyny wirtualne procesów są implementowane za pomocą interpretera; Kompilacja just-in-time osiąga wydajność porównywalną do skompilowanych języków programowania.

Język programowania Java wprowadzony wraz z wirtualną maszyną Java stał się popularny w tej formie maszyn wirtualnych. Innym przykładem jest system .NET, który działa na maszynie wirtualnej zwanej środowiskiem uruchomieniowym języka wspólnego.



RYSUNEK 1.15. Projektowanie maszyny wirtualnej procesu

Odniesienie z „Opanowania podstaw przetwarzania w chmurze i programowania aplikacji” autorstwa Rajkumara Buyya)

Sieć 2.0

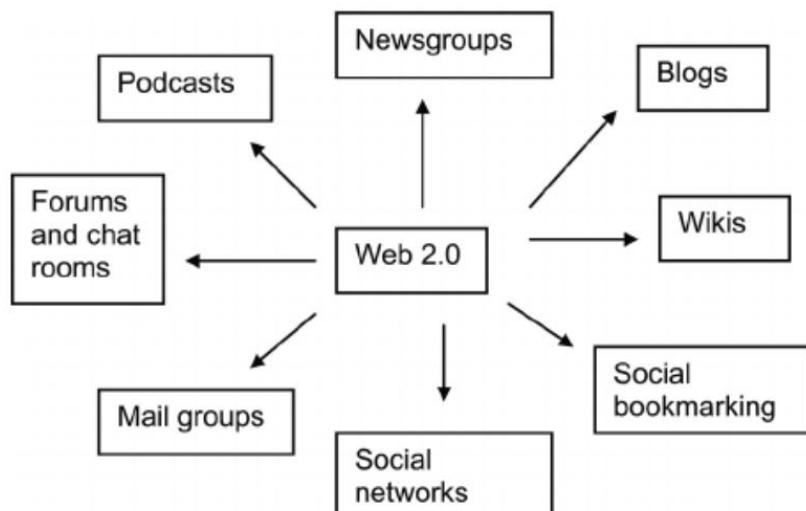
„Witryny internetowe, które kładą nacisk na treści generowane przez użytkowników, przyjazność dla użytkownika, kulturę partycypacyjną i interoperacyjność dla użytkowników końcowych” lub strony internetowe partycypacyjne, partycypacyjne/aktywistyczne i społecznościowe. Web 2.0 to nowa koncepcja, która po raz pierwszy została powszechnie użyta w 1999 roku, około 20 lat temu. Po raz pierwszy został wymyślony przez Darcy'ego DiNucci, a później spopularyzowany podczas konferencji zorganizowanej w 2004 roku przez Tim'a O'Reilly'ego i Dale'a Dougherty'ego. Należy pamiętać, że frameworki Web 2.0 zajmują się jedynie projektowaniem i użytkowaniem stron internetowych, bez narzucania projektantom wymagań technicznych.

Web 2.0 to termin używany do określenia szeregu witryn internetowych i aplikacji, które pozwalają każdemu tworzyć lub udostępniać informacje lub materiały utworzone w Internecie. Jedną z kluczowych cech tej technologii jest możliwość tworzenia, udostępniania i komunikowania się. Web 2.0 różni się od innych rodzajów witryn internetowych tym, że nie wymaga udziału żadnych umiejętności w zakresie projektowania stron internetowych ani publikowania oraz ułatwia ludziom tworzenie, publikowanie i rozpowszechnianie dzieł na świecie. Projekt pozwala na proste i popularne dzielenie się wiedzą przez małą społeczność lub znacznie szerszą publiczność.

Uczelnia będzie wykorzystywać te narzędzia do komunikacji ze studentami, pracownikami i ogólnie społecznością uniwersytecką. Może to być także dobry sposób na komunikację i interakcję dla uczniów i współpracowników.

Reprezentuje ewolucję sieci WWW; aplikacje internetowe, które umożliwiają interaktywne udostępnianie danych, projektowanie zorientowane na użytkownika i współpracę na całym świecie. Web 2.0 to zbiorowa koncepcja technologii internetowych, obejmująca blogi i witryny wiki, platformy sieciowe online, podcasty, sieci społecznościowe, społecznościowe witryny bukmacherskie i kanały RSS (Really Simple Syndication).

Główna koncepcja Web 2.0 jest usprawnienie łączności aplikacji internetowych i umożliwienie użytkownikom łatwego i wydajnego dostępu do sieci. Usługi przetwarzania w chmurze to zasadniczo aplikacje internetowe, które zapewniają usługi obliczeniowe w Internecie na żądanie. W rezultacie Cloud Computing wykorzystuje metodologię Web 2.0. Uważa się, że Cloud Computing zapewnia główną infrastrukturę Web 2.0; ułatwia to i jest ulepszane przez strukturę Web 2.0. Pod Web 2.0 znajduje się zestaw technologii sieciowych. Niedawno pojawiły się lub przeszły na nowy etap produkcyjny RIA (Rich Internet Applications). Jedna z nich to najwybitniejsza technologia WWW i quasi-standard AJAX (Asynchronous JavaScript and XML). Inne technologie, takie jak RSS (Really Simple Syndication), Widgety (wtykowe komponenty modułowe) i usługi internetowe (np. SOAP, REST).



RYSUNEK 1.15 Elementy sieci społecznościowej (Web 2.0)

1.3.3 Obliczenia zorientowane na usługi

Paradygmatem obliczeniowym, który wykorzystuje usługi jako podstawowy element tworzenia aplikacji/rozwiązań, jest przetwarzanie zorientowane na usługi (SOC). Usługi to samoopisowe komponenty specyficzne dla platformy komputerowej, które umożliwiają łatwe i ekonomiczne tworzenie rozproszonych aplikacji. Usługi wykonują funkcje, od prostych żądań po złożone procesy biznesowe. Usługi pozwalają organizacjom, korzystając z powszechnych języków i protokołów XML, wyświetlać programowanie podstawowych umiejętności w Internecie lub wewnątrz sieci oraz wykonywać je za pośrednictwem interfejsu samoopisu o otwartym standardzie.

Ponieważ usługi zapewniają jednolitych i wszechobecnych dystrybutorów informacji dla szerokiej gamy urządzeń komputerowych (np. komputerów przenośnych, PDA, telefonów komórkowych lub sprzętu), a także platform oprogramowania (np. UNIX i Windows), stanowią one kolejny ważny krok w technologii przetwarzania rozprozonego. Usługi świadczone są przez usługodawców – organizacje, które zapewniają realizację usługi, udostępniają swoje opisy usług oraz związane z nimi wsparcie techniczne i biznesowe. Ponieważ mogą być dostępne różne usługi

Firmy i komunikacja internetowa zapewniają scentralizowaną sieć sieciową do integracji i współpracy aplikacji wewnątrz i między firmami. Klientami usług mogą być aplikacje innych firm, niezależnie od tego, czy są to aplikacje zewnętrzne, procesy czy klienci/użytkownicy. Mogą to być aplikacje zewnętrzne.

W związku z tym, aby spełnić te wymagania, usługi powinny być:

- Neutralność technologiczna: muszą być niewidoczne dzięki znormalizowanym technologiom o najniższym wspólnym mianowniku, które są dostępne dla prawie wszystkich środowisk IT. Oznacza to, że mechanizmy wywoływania (protokoły, opisy i mechanizmy wykrywania) powinny być zgodne z powszechnie przyjętymi standardami.
- Luźno powiązane: żadna strona klienta ani serwisu nie potrzebuje wiedzy ani żadnych wewnętrznych struktur ani konwencji (kontekstu).

Przejrzystość lokalizacji wsparcia: Definicje usług i informacje o lokalizacji powinny być zapisane w repozytorium, takim jak UDDI, i dostępne dla szerokiego grona klientów, którzy będą mogli zlokalizować usługi i wywołać je niezależnie od ich lokalizacji.

Interakcje z usługami sieciowymi odbywają się z wykorzystaniem języka opisu usług sieciowych (WSDL) jako powszechny standard (XML) podczas wywoływania protokołu SOAP (Simple Object Access Protocol) zawierającego dane XML i opisy usług sieciowych. WSDL jest używany do publikowania usług sieciowych, typów portów (pojęciowy opis procedury i wymiany komunikatów) oraz do wiązania portów i adresów (materialna koncepcja tego, które protokoły pakowania i transportu, na przykład SOAP, są używane do łączą dwa punkty końcowe konwersacji). Standard UDDI to usługa katalogowa zawierająca publikacje usług i umożliwiająca klientom wyszukiwanie i poznawanie kandydatów na usługi.

Koncepcja oprogramowania jako usługi, propagowana przez przetwarzanie zorientowane na usługi (SOC), była pionierska i po raz pierwszy pojawiła się w modelu oprogramowania ASP (Application Service Provider). Dostawca usług aplikacji (ASP) to podmiot, który wdraża, hostuje i obsługuje dostęp strony trzeciej. Aplikacja w pakiecie i zapewnia klientom usługi i rozwiązania oparte na oprogramowaniu z centralnego centrum danych za pośrednictwem szerokiej sieci. Subskrypcja lub

wnioski o wypożyczenie są dostarczane za pośrednictwem sieci. Zasadniczo platformy ASP zapewniły przedsiębiorstwom możliwość outsourcingu dowolnej lub wszystkich części swoich potrzeb informatycznych.

ASP ponosi odpowiedzialność za zarządzanie aplikacją w swojej infrastrukturze, wykorzystując Internet jako połączenie pomiędzy każdym klientem a kluczową aplikacją za pośrednictwem centralnie hostowanej aplikacji Intent. Oznacza to dla organizacji, że ASP będzie przechowywać i gwarantować dostępność programu i danych, gdy tylko będzie to konieczne, łącznie z powiązaną infrastrukturą i danymi klientów.

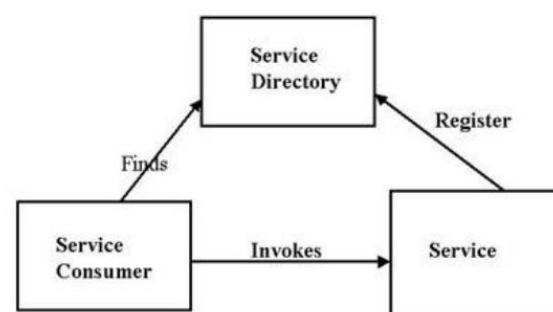
Chociaż model ASP po raz pierwszy wprowadził oprogramowanie jako definicję usługi, nie był w stanie zapewnić pełnych aplikacji z możliwością dostosowania ze względu na liczne nieodłączne ograniczenia, takie jak niemożność projektowania niezwykle interaktywnych aplikacji. Rezultatem były monolityczne architektury i bardzo podatna na ataki integracja zastosowań opartych na zasadzie ścisłego połączenia w architekturach specyficznych dla klienta.

Jesteśmy dziś w trakcie kolejnego znaczącego rozwoju architektury oprogramowania jako usługi dla asynchronicznych, luźno powiązanych interakcji opartych na standardach XML, których celem jest ułatwienie aplikacjom dostępu i komunikacji przez Internet. Model SOC umożliwia rozszerzenie koncepcji oprogramowania jako usługi na wykorzystanie skomplikowanych procesów biznesowych i transakcji w formie usługi oraz umożliwienie tworzenia aplikacji w locie i replikowania usług pomiędzy wszystkimi i przez wszystkich. Wiele ASP preferuje raczej infrastrukturę cyfrową i modele biznesowe podobne do dostawców usług w chmurze pod względem względnych zalet technologii internetowej.

Na atrybuty funkcjonalne i niefunkcjonalne składają się usługi sieciowe. Jakość usług (QoS) to tzw. atrybuty niefunkcjonalne. QoS definiuje się jako zbiór niefunkcjonalnych cech jednostek wykorzystywanych do przemieszczania się z repozytorium usług sieciowych do konsumentów, którzy polegają na zdolności usługi sieciowej do zaspokojenia określonych lub domniemanych potrzeb w sposób kompleksowy, zgodnie z definicją jakości ISO 8402. Przykłady funkcji QoS obejmują wydajność, niezawodność, bezpieczeństwo, dostępność, użyteczność, wykrywanie, adaptację i możliwość komponowania. Umowa SLA określająca minimalne (lub akceptowalny zakres) wartości atrybutów QoS, które należy spełnić przy wywoływaniu usługi, ustanawia wymagania dotyczące QoS pomiędzy klientami a dostawcami.

Czym jest architektura zorientowana na usługi?

Architektura zorientowana na usługi, czyli SOA, pozwala nam wszystkim rozumieć ją jako architekturę zorientowaną na usługi. Usługi to dyskretnie komponenty oprogramowania wdrażane przy użyciu dobrze zdefiniowanych standardów interfejsów. Usługa jest dostarczana do katalogu lub rejestru do czasu jej utworzenia i sprawdzenia, aby umożliwić innym programistom dostęp do usługi. Rejestr udostępnia także repozytorium zawierające informacje na temat opublikowanej usługi, na przykład dotyczące sposobu utworzenia interfejsu, wymaganego poziomu usług, sposobu zachowania uprawnień itp.



RYSUNEK 1.16 Architektura zorientowana na usługi

Korzyści SOA

Usługi SOA pozwalają na elastyczność biznesu. Integrując istniejące usługi, programiści mogą szybko tworzyć aplikacje.

Usługi są odrębnymi jednostkami i można je wywoływać bez znajomości platformy lub języka programowania w czasie wykonywania.

Usługi są zgodne z szeregiem standardów – Web Services Opis Language (WSDL),

Reprezentacyjny transfer stanu (REST) lub prosty protokół dostępu do obiektu (SOAP) – co ułatwia ich integrację zarówno z istniejącymi, jak i nowymi aplikacjami. Uzupełnieniem usług SOAP są poniższe standardy. MYDŁO.

Bezpieczeństwo dzięki jakości usług (QoS). Niektóre elementy QoS obejmują uwierzytelnianie i autoryzację, niezawodne i spójne przesyłanie wiadomości, zasady uprawnień itp. Nie ma wzajemnej zależności pomiędzy komponentami usług.

Wyzwania związane z SOA i przetwarzaniem w chmurze Zależność sieciowa obu tych technologii jest jednym z głównych wyzwań. Ponadto uzależnienie od dostawcy chmury, umów i umów o gwarantowanym poziomie usług to wyzwania specyficzne dla przetwarzania w chmurze.

Jednym z wyzwań stojących dziś przed architekturą SOA są żądania ulepszenia lub zmiany usług świadczonych przez dostawców usług SOA.

Czy Cloud Computing konkuuuje z SOA?

Niektórzy postrzegają przetwarzanie w chmurze jako potomka SOA. Nie byłoby to do końca nieprawdą, gdyż zasady wytycznych serwisowych dotyczą zarówno chmury obliczeniowej, jak i SOA. Poniższa ilustracja pokazuje, jak usługi Cloud Computing pokrywają się z architekturą SOA-

Przetwarzanie w chmurze	Nakładanie	SOA za pośrednictwem usług sieciowych
<ul style="list-style-type: none"> • Oprogramowanie jako usługa (SaaS) • Obliczenia użytkowe • Terabajty na żądanie • Dane rozpowszechniane w a Chmura • Platforma jako usługa • Ewolujące standardy Różne warstwy Stos 	<ul style="list-style-type: none"> się • Warstwa aplikacji Komponenty/usługi • Zależność od sieci • Szeroki obszar chmury/IP Sieć (WAN)-obsługiwana usługa Modły • Wykorzystanie rozproszone Zasoby oprogramowania • Producent/Konsument Model 	<ul style="list-style-type: none"> • System Systemów Fokus Integracyjny • Spójność jazdy Integracja • Aplikacja korporacyjna Integracja (EAI) • W miarę dojrzały Wdrażanie standardów (REST, SOAP, WSDL, UDDI itp.)

Bardzo ważne jest, aby zdać sobie sprawę, że choć przetwarzanie w chmurze pokrywa się z SOA, skupiają się one na różnych projektach wdrożeniowych. W celu wymiany informacji pomiędzy systemami i siecią systemów wykorzystuje się przede wszystkim wdrożenia SOA. Z drugiej strony przetwarzanie w chmurze ma na celu wykorzystanie sieci w całym zakresie funkcji IT.

SOA nie nadaje się do przetwarzania w chmurze, właściwie są to działania dodatkowe. Dostawcy potrzebują bardzo dobrej architektury zorientowanej na usługi, aby móc skutecznie świadczyć usługi w chmurze.

Istnieje wiele wspólnych cech SOA i przetwarzania w chmurze, jednak nie są one i mogą współistnieć. Wydaje się, że jeśli chodzi o wymagania dotyczące świadczenia usług cyfrowych, SOA dojrzała. Cloud Computing i jego usługi są nowością, podobnie jak wielu dostawców, takich jak chmury publiczne, społecznościowe, hybrydowe i prywatne, ze swoimi ofertami. One również rosną.

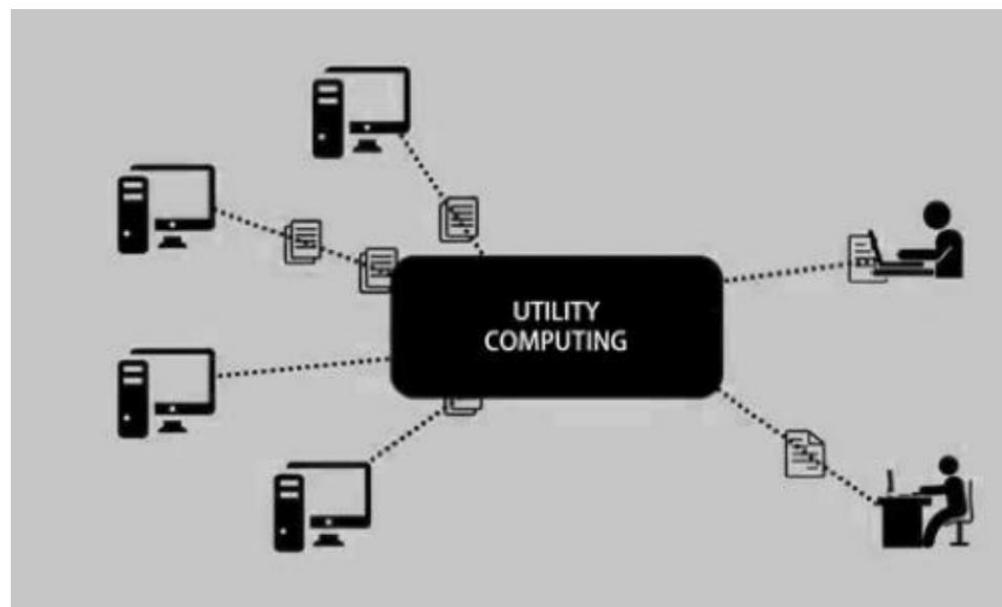
1.3.4 Obliczenia zorientowane na użytkownictwo

Koncepcja Utility Computing odnosi się do narzędzi i modeli biznesowych, które zapewniają swoim klientom dostawcę usług i pobierają opłaty za zużycie. Przykładami takich usług IT są moc obliczeniowa, pamięć masowa czy aplikacje. W tym scenariuszu klientem będą pojedyncze oddziały firmy jako dostawca usług w centrum danych firmy.

Pojęcie narzędzia odnosi się do usług użytkownictwa publicznego oferowanych przez dostawcę mediów, takich jak energia elektryczna, telefon, woda i gaz. W przypadku energii elektrycznej lub telefonu, gdzie odbiorca otrzymuje energię obliczeniową, moc obliczeniowa jest mierzona i opłacana na podstawie wspólnej sieci komputerowej.

Pojęcie narzędzia odnosi się do usług użyteczności publicznej oferowanych przez dostawcę mediów, takich jak energia elektryczna, telefon, woda i gaz. W przypadku energii elektrycznej lub telefonu, gdzie odbiorca otrzymuje energię obliczeniową, moc obliczeniowa jest mierzona i opłacana na podstawie wspólnej sieci komputerowej.

Obliczenia użytkowe są bardzo analogiczne do wirtualizacji, w związku z czym całkowita objętość pamięci masowej sieciowej i moc obliczeniowa dostępna dla klientów są znacznie większe niż w przypadku pojedynczego komputera. Aby umożliwić tego typu serwer WWW, często wykorzystuje się kilka serwerów zaplecza sieciowego. Dedykowane serwery WWW mogą być używane w jawnie budowanych i dzierżawionych typach klastrów dla użytkowników końcowych. Przetwarzanie rozproszone to podejście stosowane do pojedynczych „obliczeń” na wielu serwerach internetowych.



RYSUNEK 1.17 Technologia przetwarzania w chmurze – przetwarzanie użytkowe
Właściwości obliczeń użytkowych

Chociaż znaczenia obliczeń użytkowych są różne, zwykle zawierają one pięć następujących cech.

Skalowalność

Obliczenia użytkowe zapewniają dostępność odpowiednich zasobów informatycznych w każdej sytuacji. Zwiększyły popyt na usługi nie ucierpi na ich jakości (np. czasie reakcji).

Cena popytu

Do tej pory firmy musiały kupować własną moc obliczeniową, taką jak sprzęt i oprogramowanie. Za tę infrastrukturę informatyczną trzeba zapłacić wcześniej, niezależnie od jej przyszłego wykorzystania. Na przykład to, czy dostawcy technologii będą mogli uzyskać dostęp do tego łącza, zależy od liczby procesorów, które klient udostępnił w ramach stawki leasingu swoich serwerów. Jeżeli w przedsiębiorstwie faktycznie można zmierzyć wydajność komputera niezbędną do obsługi poszczególnych sekcji, koszty IT można przypisać przede wszystkim każdej pojedynczej jednostce według kosztów wewnętrznych. Możliwe są dodatkowe formy połączenia przy wykorzystaniu kosztów IT.

Standaryzowane usługi w zakresie obliczeń użytkowych

Zbiór standardowych usług jest dostępny u dostawcy usług obliczeniowych. Umowy te mogą różnić się poziomem usług (Umowa Jakościowa i Cena IT). Konsument nie ma żadnego wpływu na infrastrukturę, taką jak platforma serwerowa

Obliczenia użytkowe i wirtualizacja

Technologie wirtualizacji można wykorzystać do udostępniania sieci i innych zasobów we wspólnej puli maszyn. Zamiast dostępnych zasobów fizycznych, sieć dzieli się na zasoby logiczne. Do aplikacji nie są przypisane żadne z góry określone serwery ani magazyny inne niż wolny serwer lub pula pamięci.

Automatyzacja

Powtarzające się czynności związane z zarządzaniem, takie jak konfigurowanie nowych serwerów lub pobieranie aktualizacji, mogą zostać zautomatyzowane. Ponadto należy wziąć pod uwagę alokację zasobów do usług i zarządzanie usługami IT, które mają zostać zoptymalizowane, a także umowy dotyczące standardów usług i koszty operacyjne zasobów IT.

Zalety obliczeń użytkowych

Obliczenia użytkowe obniżają koszty IT pomimo elastyczności istniejących zasobów. W rzeczywistości wydatki są jasne i można je przypisać bezpośrednio do różnych działów organizacji. Mniej osób potrzeba do działań operacyjnych w działach IT.

Firmy są bardziej elastyczne, ponieważ ich zasoby IT szybciej i łatwiej dostosowują się do zmieniającego się zapotrzebowania. Podsumowując, cały system informatyczny jest prostszy w obsłudze, ponieważ aplikacje nie mogą już korzystać z określonej infrastruktury IT dla dowolnego programu.

1.4 Budowa środowisk przetwarzania w chmurze

Środowisko przetwarzania w chmurze Tworzenie aplikacji odbywa się za pośrednictwem platform i aplikacji frameworkowych, które zapewniają różne rodzaje usług, od infrastruktury typu bare metal po aplikacje niestandardowe służące określonym celom.

1.4.1 Tworzenie aplikacji

Potężny model obliczeniowy, który umożliwia użytkownikom korzystanie z aplikacji na żądanie, zapewnia chmura obliczeniowa. Jedną z najkorzystniejszych klas aplikacji posiadających tę cechę są aplikacje webowe. Na ich wydajność największy wpływ ma szeroka gama aplikacji korzystających z różnych usług w chmurze, które mogą generować obciążenia dla konkretnych wymagań użytkowników.

Szybkie rozpowszechnienie Web 2.0 ułatwiło kilka czynników. Po pierwsze, Web 2.0 opiera się na różnorodnych osiągnięciach i postępach technologicznych, które umożliwiają użytkownikom łatwe tworzenie bogatych i złożonych aplikacji, w tym aplikacji dla przedsiębiorstw, wykorzystując obecnie Internet jako główne narzędzie i platformę interakcji z użytkownikiem. Takie aplikacje charakteryzują się znacznymi złożonymi procesami powodowanymi przez interakcję użytkownika i interakcję pomiędzy wieloma etapami za frontem sieci Web. Są to aplikacje najbardziej wrażliwe na niewłaściwą wielkość wdrożenia infrastruktury i usług oraz zmienność obciążenia.

Aplikacje wymagające dużych zasobów to kolejna klasa aplikacji, które mogą potencjalnie przynieść znaczne korzyści dzięki wykorzystaniu chmury obliczeniowej. Aplikacje te mogą wymagać obliczeń lub intensywnego przetwarzania danych. W obu przypadkach są znaczące zasoby

Wymagane w rozsądny terminie do zakończenia realizacji. Należy zaznaczyć, że tak ogromne ilości zasobów nie są potrzebne stale i przez długi czas. Zastosowania naukowe, na przykład mogą wymagać ogromnej mocy obliczeniowej do przeprowadzania od czasu do czasu testów na dużą skalę, w związku z czym nie można kupić infrastruktury obsługującej je. Rozwiązaniem w tym przypadku jest chmura obliczeniowa. Aplikacje wymagające dużych zasobów nie współpracują i charakteryzują się głównie przetwarzaniem wsadowym.

1.4.2 Rozwój infrastruktury i systemu

1.4.3 Platformy i technologie komputerowe

Tworzenie aplikacji w chmurze obejmuje wykorzystanie platform i frameworków oferujących różne usługi, od infrastruktury typu bare metal po spersonalizowane aplikacje służące konkretnym celom.

1.4.3.1 Usługi sieciowe Amazon (AWS)

Amazon Web Services (AWS) to platforma przetwarzania w chmurze oferująca między innymi takie funkcjonalności jak przechowywanie baz danych, dostarczanie treści i bezpieczna infrastruktura IT dla firm. Jest znana ze swoich usług na żądanie, a mianowicie Elastic Compute Cloud (EC2) i Simple Storage Service (S3). Amazon EC2 i Amazon S3 to niezbędne narzędzia do

zrozumiesz, jeśli chcesz w pełni wykorzystać chmurę AWS.

Amazon EC2 to oprogramowanie do uruchamiania serwerów w chmurze, które jest skrótem od Elastic Cloud compute. Amazon uruchomił EC2 w 2006 roku, ponieważ umożliwił firmom szybkie i łatwe przeniesienie serwerów do chmury, zamiast kupować, konfigurować i zarządzać własnymi serwerami lokalnie.

Chociaż instancje serwerów Amazon EC2 mogą również mieć instancje EC2 typu bare-metal, większość instancji serwerów Amazon EC2 to maszyny wirtualne umieszczone w infrastrukturze Amazon. Serwer jest obsługiwany przez dostawcę chmury i nie musisz konfigurować ani konserwować sprzętu.) Dostępna jest ogromna liczba instancji EC2 w różnych cenach; ogólnie rzecz biorąc, im więcej mocy obliczeniowej używasz, tym wyższej instancji EC2 potrzebujesz. (Instancje chmury Bare Metal umożliwiają hostowanie obciążenia roboczego na komputerze fizycznym, a nie na maszynie wirtualnej. W niektórych przykładach Amazon EC2 zoptymalizowano pod kątem użycia różne typy aplikacji, takie jak przetwarzanie równoległe procesorów graficznych obsługujących duże zbiory danych.

EC2 oferuje takie funkcjonalności jak autoskalowanie, które automatyzuje proces zwiększenia lub zmniejszania zasobów obliczeniowych dostępnych dla danego obciążenia, nie tylko po to, aby ułatwić i przyspieszyć wdrożenie serwera. Automatyczne skalowanie pomaga w ten sposób zoptymalizować koszty i wydajność, szczególnie w warunkach pracy ze znacznymi różnicami w objętości.

Amazon S3 to usługa przechowywania danych działająca w chmurze AWS (jak jej pełna nazwa, Simple Storage Service). Umożliwia użytkownikom przechowywanie praktycznie każdej formy danych w chmurze i dostęp do nich za pośrednictwem interfejsu internetowego, interfejsu wiersza poleceń AWS lub API AWS. Musisz zbudować coś, co Amazon nazwało „wiadrem”, czyli konkretnym obiektem, którego używasz do przechowywania i pobierania danych w celu korzystania z S3. Jeśli chcesz, możesz ustawić wiele wiader.

Amazon S3 to system przechowywania obiektów, który szczególnie dobrze sprawdza się w przypadku masowego, nierównego lub bardzo dynamicznego przechowywania danych.

1.4.3.2 Google AppEngine

Google AppEngine (GAE) to usługa przetwarzania w chmurze (należąca do kategorii platformy jako usługi (PaaS)) służąca do tworzenia i hostowania aplikacji internetowych w centrach danych Google. Aplikacje internetowe GAE działają w trybie piaskownicy i działają na wielu serwerach redundancyjnych, co pozwala na skalowanie zasobów zgodnie z aktualnie istniejącymi wymaganiami dotyczącymi ruchu. App Engine przydziela serwerom dodatkowe zasoby, aby mogły obsługiwać zwiększone obciążenie.

Google App Engine to platforma Google dla programistów i firm, umożliwiająca tworzenie i uruchamianie aplikacji przy użyciu zaawansowanej infrastruktury Google. Aplikacje te muszą być napisane w jednym z kilku obsługiwanych języków, a mianowicie Java, Python, PHP i Go. Wymaga to również użycia języka zapytań Google, a używaną bazą danych jest Google Big Table. Aplikacje muszą być zgodne z tymi normami, zatem aplikacje muszą być albo opracowane zgodnie z GAE, albo zmodyfikowane w celu zapewnienia zgodności.

GAE to platforma do uruchamiania i hostowania aplikacji internetowych, zarówno na urządzeniach mobilnych, jak i w Internecie. Bez tej kompleksowej funkcji programiści powinni być odpowiedzialni za tworzenie własnych serwerów, oprogramowania baz danych i interfejsów API, które sprawią, że wszyscy będą poprawnie współpracować. GAE odciąga programistów, dzięki czemu mogą skoncentrować się na interfejsie aplikacji i funkcjach poprawiających wygodę użytkownika.

1.4.3.3 Microsoft Azure

Microsoft Azure to platforma jako usługa (PaaS) służąca do tworzenia aplikacji i zarządzania nimi do korzystania z produktów Microsoft oraz w centrach danych. To kompletny pakiet produktów chmurowych, który pozwala użytkownikom tworzyć aplikacje klasy biznesowej bez konieczności rozwijania własnej infrastruktury.

Na platformie Azure Cloud dostępne są trzy produkty zorientowane na chmurę: kontroler Windows Azure, SQL Azure i Azure App Fabric. Obejmuje to infrastrukturę hostującą aplikację.

Na platformie Azure rolą usługi w chmurze to zestaw platform wirtualnych, które współpracują ze sobą

realizować podstawowe zadania, które są zarządzane, równoważone z obciążeniem i oparte na platformie jako usłudze. Role usług w chmurze są kontrolowane przez kontroler sieci szkieletowej platformy Azure i zapewniają idealne połączenie skalalności, kontroli i dostosowywania.

Rola internetowa to rola usługi Azure Cloud, która jest skonfigurowana i dostosowana do obsługi aplikacji internetowych tworzonych w językach i technologiach programowania Internet Information (IIS), takich jak ASP.NET, PHP, Windows Communication Foundation i Fast CGI.

Rola sieciowa to rola usługi Azure Cloud, która jest skonfigurowana i dostosowana do obsługi aplikacji internetowych tworzonych w językach i technologiach programowania Internet Information (IIS), takich jak ASP.NET, PHP, Windows Communication Foundation i Fast CGI.

Rola procesu roboczego to dowolna rola platformy Azure, która działa na aplikacjach i usługach, które zwykle nie wymagają usług IIS. Usługi IIS nie są domyślnie włączone w rolach roboczych. Wykorzystuje się je głównie do obsługi internetowych procesów w tle oraz do wykonywania zadań, takich jak automatyczne kompresowanie przesyłanych obrazów, uruchamianie skryptów, pobieranie nowych wiadomości z kolejki i przetwarzanie i nie tylko, gdy coś zmienia bazę danych.

Rola maszyny wirtualnej: Rola maszyny wirtualnej to typ roli platformy Azure, który obsługuje automatyczne zarządzanie już zainstalowanymi pakietami usług, poprawkami, aktualizacjami i aplikacjami dla systemu Windows Azure.

Zasadnicza różnica polega na tym, że:

Rola internetowa automatycznie wdraża i hostuje aplikację za pośrednictwem usług IIS. Rola robocza nie korzysta z usług IIS i uruchamia program niezależnie. Obie role można obsługiwać podobnie i można je uruchamiać w tych samych wystąpieniach platformy Azure, jeśli są wdrażane i dostarczane za pośrednictwem platformy usług Azure.

W niektórych przypadkach instancje ról sieciowych i ról roboczych współpracują ze sobą i są również używane jednocześnie przez aplikację. Na przykład przykładowa rola internetowa może akceptować aplikacje od użytkowników, a następnie przekazywać je przykładowej roli procesu roboczego bazy danych.

1.4.3.4 Hadoop

Apache Hadoop to platforma oprogramowania typu open source do przechowywania i przetwarzania na dużą skalę zbiorów danych w klastrach sprzętu towarowego. Hadoop to najwyższej klasy projekt Apache stworzony i obsługiwany przez globalną społeczność współpracowników i użytkowników. Jest objęty licencją Apache 2.0.

Dwie fazy funkcji MapReduce, Map i Redukcja. Zadania mapy dotyczą dzielenia danych i mapowania danych, natomiast zadania Zmniejsz tasowanie i redukcję danych.

Hadoop może uruchamiać programy MapReduce w różnych językach, takich jak Java, Ruby, Python i C++. Program MapReduce ma charakter równoległy i dlatego jest bardzo przydatny do analiz danych na dużą skalę za pośrednictwem wielu maszyn klastrowych.

Dane wejściowe dla każdej fazy to pary klucz-wartość. Dodatkowo każdy programista musi określić dwie funkcje: funkcję mapowania i funkcję redukcji.

1.4.3.5 Force.com i Salesforce.com

Aby zrozumieć różnicę między Salesforce.com i Force.com, należy zrozumieć podstawowe pojęcia dotyczące przetwarzania w chmurze.

Salesforce to firma, a salesforce.com to aplikacja zbudowana w oparciu o oprogramowanie jako usługę (SaaS) do zarządzania relacjami z klientami (CRM). Platforma force.com pomaga programistom i użytkownikom biznesowym w tworzeniu skutecznych aplikacji biznesowych.

Salesforce to produkt SaaS obejmujący funkcje Out of Box (OOB) wbudowane w system CRM w celu automatyzacji sprzedaży, marketingu, automatyzacji usług itp. Niektóre przykłady SaaS to Dropbox, Google Apps i GoToMeeting, które odnoszą się do pobierania oprogramowania z komputera do chmury.

Force.com to produkt typu PaaS (platforma jako usługa); zawiera framework umożliwiający tworzenie aplikacji. Zawiera środowisko programistyczne. Force.com pomaga dostosować interfejs użytkownika, funkcjonalność i logikę biznesową.

Mówiąc najprościej, funkcjonalność Salesforce.com zapisuje kontakty, wiadomości tekstowe, połączenia i inne standardowe funkcje w aplikacji na iPhone'a. W force.com aplikacje są tworzone i obsługiwane. Salesforce.com działa na stronie force.com, tak jak dialer iPhone'a działa w systemie operacyjnym iPhone'a.

1.4.3.6 Manjrasoft Aneka

MANJRASOFT spółka z ograniczoną odpowiedzialnością Ltd. jest jedną z organizacji zajmujących się technologią przetwarzania w chmurze, opracowując oprogramowanie kompatybilne z sieciami rozproszonymi na wielu serwerach.

- Twórz skalowalne, konfigurowalne elementy składowe niezbędne do przetwarzania w chmurze platformy.
- Twórz oprogramowanie przyspieszające aplikacje przeznaczone do pracy w sieci wielordzeniowej komputery.
- Dostarczanie rozwiązań dotyczących jakości usług (QoS) i umów poziomu usług (SLA) opartych na umowie o poziomie usług (SLA), które umożliwiają planowanie, wysyłanie, ustalanie cen aplikacji i usług księgowych, środowisk biznesowych i/lub publicznych sieci komputerowych.
- Rozwój aplikacji poprzez umożliwienie szybkiego generowania starszych i nowych aplikacji z wykorzystaniem innowacyjnych równoległych i rozproszonych modeli programowania.
- Zdolność organizacji do korzystania z zasobów obliczeniowych. Biznes w celu przyspieszenia aplikacji wymagających intensywnego wykonywania obliczeń lub danych

1.5 Podsumowanie

W tym rozdziale zbadaliśmy cel oraz zalety i wyzwania związane z przetwarzaniem w chmurze. W wyniku rozwoju i integracji wielu wspierających ją modeli i technologii, w szczególności przetwarzania rozproszonego, technologii przetwarzania w chmurze Web 2.0, wirtualizacji, usług zorientowanych na usługi i Utility Computing. Badamy różne definicje, znaczenia i wdrożenia tego pojęcia. Jedynie dynamiczne świadczenie usług IT (niezależnie od tego, czy jest to infrastruktura wirtualna, środowiska uruchomieniowe czy usługi aplikacyjne) oraz wdrożenie modelu kosztów opartego na użyteczności w celu wyceny takich usług jest elementem wspólnym wszystkich różnych poglądów na przetwarzanie w chmurze. Architektura ta jest stosowana w całym stosie obliczeniowym i umożliwia dynamiczne udostępnianie zasobów IT i zasobów wykonawczych w kontekście platform hostowanych w chmurze w celu tworzenia skalowalnych aplikacji i ich usług. Koncepcja ta reprezentuje metodę referencyjną przetwarzania w chmurze. Model ten definiuje trzy ważne komponenty branży przetwarzania w chmurze i oferowanych tam usług: infrastruktura jako usługa (IaaS), platforma jako usługa (PaaS) i oprogramowanie jako usługa (SaaS). Komponenty te wyraźnie odwzorowują szerokie kategorie różnych typów usług przetwarzania w chmurze.

1.6 Pytania przeglądowe

1. Jaka jest innowacyjna cecha przetwarzania w chmurze?
2. Jakie technologie wspierane są przez przetwarzanie w chmurze?
3. Podaj krótką charakterystykę systemu rozproszonego.
4. Zdefiniować przetwarzanie w chmurze i zidentyfikować główne cechy przetwarzania w chmurze.
5. Jakie są najważniejsze technologie rozproszone, które przyczyniły się do powstania chmury przetwarzanie danych?
6. Co to jest wirtualizacja?
7. Wyjaśnij największą rewolucję wprowadzoną przez Web 2.0
8. Podaj przykłady zastosowań Web 2.0.
9. Opisz główne cechy orientacji usługowej.
10. Krótko podsumuj model referencyjny przetwarzania w chmurze.
11. Jaka jest główna zaleta przetwarzania w chmurze?
12. Wyjaśnij różne typy modeli w chmurze obliczeniowej
13. Wyjaśnij trzy usługi chmurowe w chmurze obliczeniowej.
14. Co to są usługi internetowe? Wyjaśnij różne rodzaje usług internetowych.

1.7 Odniesienie do dalszej lektury

1. Opanowanie podstaw przetwarzania w chmurze i programowania aplikacji Rajkumar Buyya, Christian Vecchiola, S. Publikacje Thamarai Selvi MK ISBN: 978-0-12-411454-8
2. Koncepcje, technologia i architektura przetwarzania w chmurze Thomas Erl, Zaigham Mahmood, i Ricardo Puttiniego , seria technologii usług Prentice Hall ISBN-10: 9780133387520 ISBN-13 : 978-0133387520
3. Przetwarzanie rozproszone i przetwarzanie w chmurze: od przetwarzania równoległego do Internetu rzeczy, wydanie pierwsze, Kai Hwang Jack Dongarra Geoffrey Fox ISBN-10: 9789381269237 ISBN-13: 978-9381269237
4. <https://www.geeksforgeeks.org/cloud-computing/>
5. https://en.wikipedia.org/wiki/Cloud_computing
6. <https://aws.amazon.com/what-is-cloud-computing/>
7. http://www.manjrasoft.com/aneka_architecture.html
8. https://en.wikipedia.org/wiki/Microsoft_Azure
9. https://en.wikipedia.org/wiki/Apache_Hadoop

Rozdział 1

Rozdział 2

Struktura jednostki

2.0 Cel 2.1 Ery

obliczeń 2.2 Obliczenia

równoległe a rozproszone 2.3 Elementy

obliczeń równoległych 2.3.1 Co to jest

przetwarzanie równoległe?

2.3.2 Architektury sprzętowe do przetwarzania równoległego

2.3.2.1 Systemy z jedną instrukcją i pojedynczymi danymi

(SISD) 2.3.2.2 Systemy z jedną instrukcją i wieloma danymi (SIMD)

2.3.2.3 Systemy z wieloma instrukcjami i pojedynczymi danymi

(MISD) 2.3.2.4 Systemy z wieloma instrukcjami, wielokrotne Systemy

-data (MIMD) 2.3.1 Podejście do programowania

równoległego 2.3.2 Poziomy

równoległości 2.3.3 Prawa ostrożności

2.4 Elementy przetwarzania rozproszonego

2.4.1 Ogólne pojęcia i definicje 2.4.2

Komponenty systemu rozproszonego 2.4.3 Style

architektoniczne dla obliczeń rozproszonych 2.4.3.1

Komponenty i złącza 2.4.3.2 Style

architektury oprogramowania 2.4.3.3

Style architektury systemu 2.4.4 Modele

komunikacji międzyprocesowej 2.4.4.1 Komunikacja

oparta na komunikatach 2.4.4.2 Modele

komunikacji opartej na komunikatach Technologie

2.5 przetwarzania rozproszonego 2.5.1 Zdalne

wywoływanie procedur 2.5.2

Struktury obiektów rozproszonych 2.5.2.1

Przykłady struktur obiektów rozproszonych 2.5.3

Obliczenia zorientowane na usługi

2.5.3.1 Co to jest serwis?

2.5.3.2 Architektura zorientowana na usługi (SOA)

2.5.3.3 Usługi sieciowe

2.5.3.4 Orientacja na usługi i przetwarzanie w chmurze

2.6 Podsumowanie

2.7 Pytania kontrolne 2.8

Odniesienia do dalszej lektury

2.0 Cel

Komponenty obliczeniowe (sprzęt, oprogramowanie, infrastruktura), które umożliwiają świadczenie usług przetwarzania w chmurze, odnoszą się do systemu chmury lub technologii przetwarzania w chmurze.

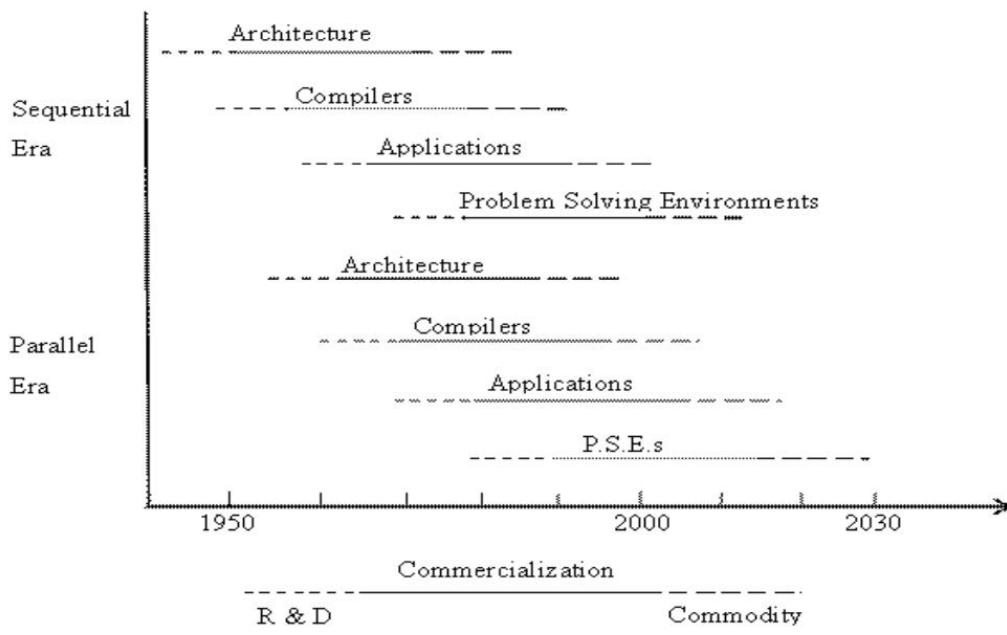
Konsumenci mogą zdobywać nowe umiejętności bez inwestowania w nowy sprzęt lub oprogramowanie za pośrednictwem chmury publicznej. Zamiast tego płacą abonament swojemu dostawcy usług w chmurze lub płacą tylko za swoje zasoby. Te zasoby IT są własnością dostawców usług i są przez nich zarządzane za pośrednictwem Internetu.

W tym rozdziale przedstawiono podstawowe zasady i modele przetwarzania równoległego i rozproszonego, które stanowią podstawę do budowy systemów i frameworków przetwarzania w chmurze.

2.1 Epoki informatyki

Dwie najważniejsze ery informatyki to sekwencyjna i równoległa. W ostatniej dekadzie wysokowydajne komputerowe wyszukiwanie maszyn równoległych stało się ważną konkurencją maszyn wektorowych. Rysunek 2.1 przedstawia stuletni przegląd rozwoju ery komputerowej. W tych okresach tworzone są cztery główne elementy obliczeniowe, takie jak architektury, kompilatory, aplikacje i środowiska rozwiązywania problemów.

Era informatyki rozpoczyna się wraz z rozwojem sprzętu, następnie systemów oprogramowania (zwłaszcza w obszarze kompilatorów i systemów operacyjnych), aplikacji, a wraz z rosnącym środowiskiem rozwiązywania problemów osiąga poziom nasycenia. Każdy element obliczeniowy podlega trzem etapom: badawczo-rozwojowemu, komercjalizacji i towarowemu.



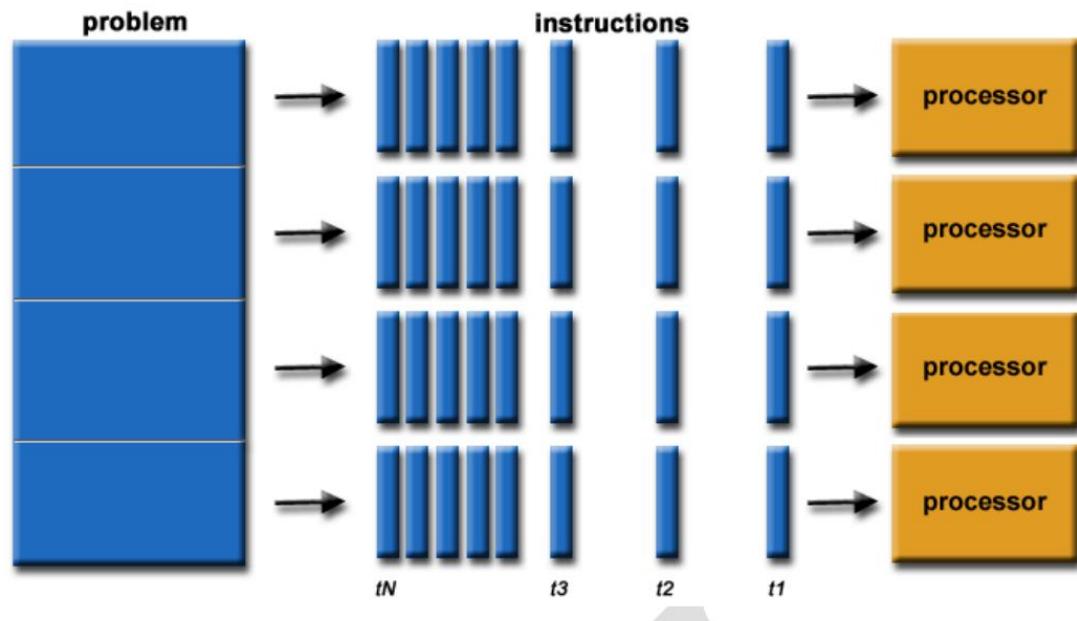
2.2 Obliczenia równoległe a rozproszone

Terminy obliczenia równoległe i obliczenia rozproszone, mimo że są to nieco różne pojęcia, są często używane zamiennie. Termin równoległy oznacza system ściśle powiązany, natomiast rozproszony odnosi się do szerszej klasy systemów, w tym ściśle powiązanych.

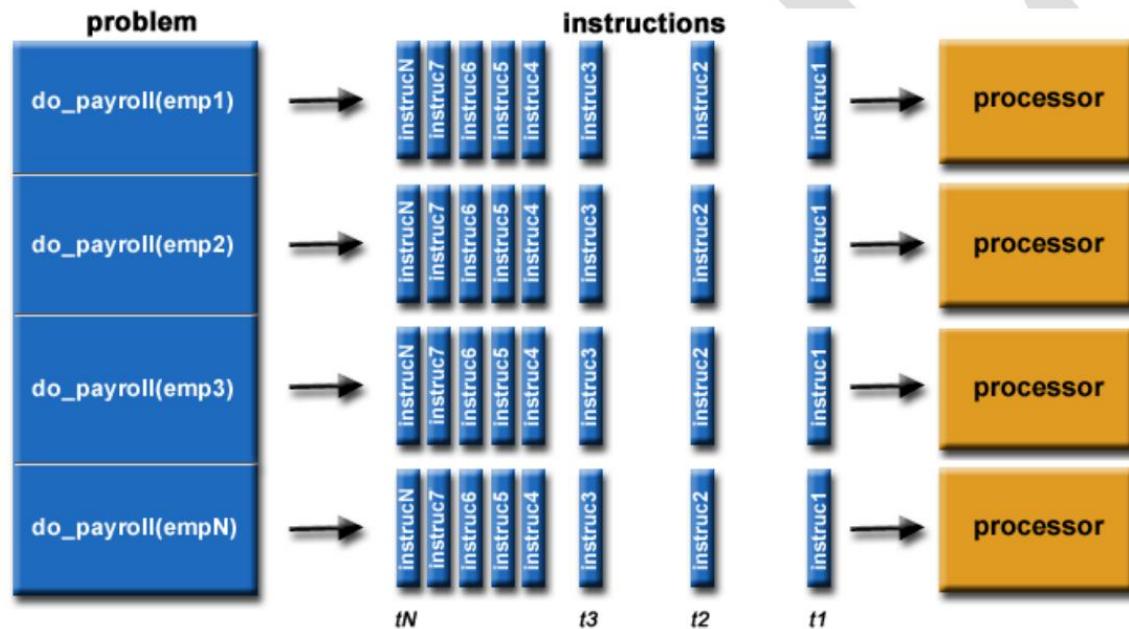
Jednoczesne wykorzystanie kilku zasobów komputera w celu rozwiązywania problemu obliczeniowego to przetwarzanie równoległe:

- Problem dzieli się na odrębne części, które można rozwiązać jednocześnie
- Szereg instrukcji dla każdej części jest dalej podzielonych

- Instrukcje na różnych procesorach z każdej części działają jednocześnie
- Stosowany jest ogólny mechanizm kontroli/koordynacji



Na przykład:



RYSUNEK 2.2 Przetwarzanie sekwencyjne i równoległe

Problem w obliczeniach powinien brzmieć:

Problem w obliczeniach powinien brzmieć:

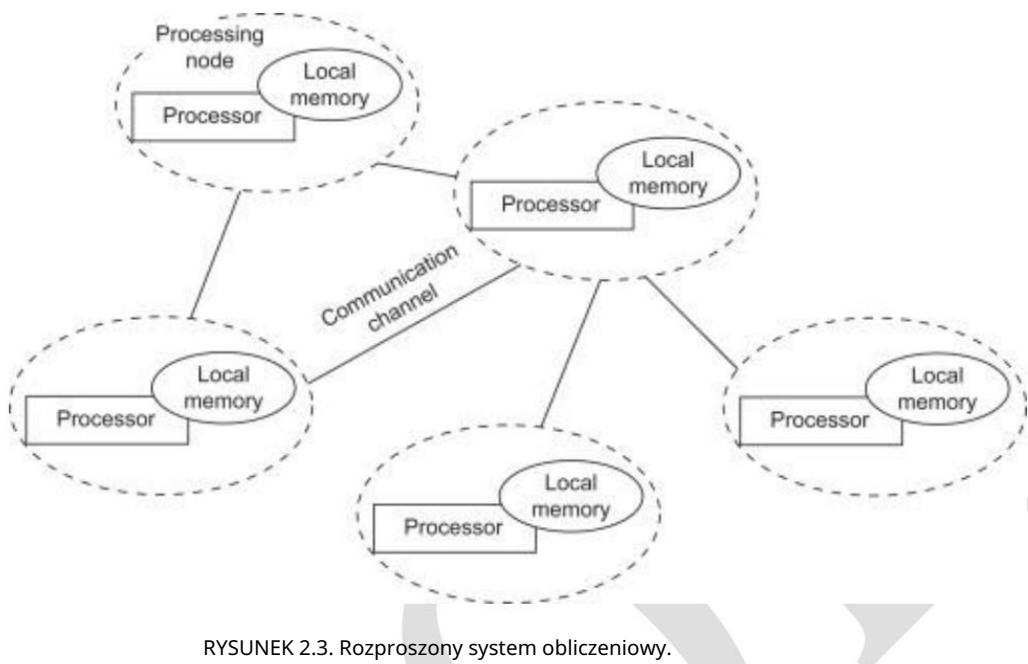
- Być podzielonym na odrębne części pracy, które można rozwiązać w tym samym czasie; • W dowolnym momencie wykonaj wiele instrukcji programu; • Można rozwiązać problem przy użyciu wielu zasobów obliczeniowych w krótszym czasie niż w przypadku jednego zasobu obliczeniowego.

Zazwyczaj zasoby obliczeniowe to:

- Jeden komputer z kilkoma procesorami/rdzeniami
- Losowa liczba tych komputerów podłączonych do sieci

Początkowo tylko niektóre architektury były rozważane jako systemy równoległe. Zawierały wiele procesorów z tą samą pamięcią fizyczną i jednym komputerem. Z biegiem czasu ograniczenia te zostały złagodzone i systemy równoległe obejmują obecnie wszystkie architektury, niezależnie od tego, czy są one fizycznie obecne, czy oparte na koncepcji pamięci współdzielonej, niezależnie od tego, czy obsługa bibliotek, określony sprzęt i bardzo wydajna infrastruktura sieciowa są fizycznie obecne, czy utworzone. Na przykład klaszter węzłów połączonych przez InfiniBand można uznać za system równoległy i skonfigurować go z dystrybucyjnym systemem pamięci współdzielonej

Obliczenia rozproszone są obliczane przez rozproszone, niezależne komputery komunikujące się wyłącznie za pośrednictwem sieci (rysunek). Rozproszone systemy obliczeniowe są zwykle traktowane inaczej niż równoległe systemy obliczeniowe lub systemy pamięci współdzielonej, w których wiele komputerów korzysta ze wspólnej puli pamięci używanej do wzajemnej komunikacji procesorów. Systemy pamięci wykorzystywały wiele komputerów w celu rozwiązania typowego problemu, przetwarzania danych między połączonymi komputerami (węzłami) i komunikowania się między węzłami poprzez przekazywanie komunikatów.



RYSUNEK 2.3. Rozproszony system obliczeniowy.

Przetwarzanie rozproszone ogranicza się do programów znajdujących się na ograniczonym geograficznie obszarze, których komponenty są współdzielone między komputerami. Obie szersze definicje obejmują typowe zadania i komponenty programu. Przetwarzanie rozproszone w najszerszym znaczeniu oznacza, że coś jest współdzielone pomiędzy wieloma systemami, co może mieć również miejsce w różnych lokalizacjach.

Przykłady systemów rozproszonych/zastosowań przetwarzania rozproszonego:

- Intranety, Internet, WWW, poczta elektroniczna.
- Sieci telekomunikacyjne: sieci telefoniczne i sieci komórkowe.
- Sieć komputerów oddziałowych -System informatyczny obsługujący automatyczne przetwarzanie zamówień,
- Sterowanie procesami w czasie rzeczywistym: systemy sterowania statkiem powietrznym,
- Bankowość elektroniczna,
- Systemy rezerwacji linii lotniczych,
- Sieci sensorowe,
- Mobilne i wszechobecne systemy komputerowe.

2.3 Elementy obliczeń równoległych

To wykładniczy wzrost mocy obliczeniowej. W 1965 roku współzałożyciel Intelu, Gordon Moore, zauważył, że liczba tranzystorów w jednocalowym chipie podwaja się rocznie, a koszt spada o około połowę. Teraz trwa 18 miesięcy i jest coraz dłuższy. Krzem osiąga granicę wydajności w coraz większej liczbie zastosowań wymagających zwiększonej prędkości, zmniejszonych opóźnień i wykrywania światła. Aby przezwyciężyć to ograniczenie, możliwym rozwiązaniem jest połączenie kilku procesorów w celu wspólnego rozwiązywania problemów stanowiących „wielkie wyzwanie”. Początkowe kroki w kierunku obliczeń równoległych prowadzą do wzrostu. Obejmuje technologię, architekturę i systemy do wielu równoległych działań. W tym podrozdziale dokonano jego właściwej charakterystyki, która obejmuje równoległość działania wielu procesorów koordynujących się ze sobą w ramach jednego komputera.

2.3.1 Co to jest przetwarzanie równoległe?

Przetwarzanie równoległe to sposób zarządzania różnymi częściami ogólnego zadania podczas porównywania dwóch lub większej liczby procesorów. Procesory. Podział różnych części zadania na kilka procesorów może skrócić czas działania programu. Przetwarzanie równoległe może wykonywać maszyna z więcej niż jednym procesorem lub procesory wielordzeniowe, które są powszechnie spotykane w dzisiejszych komputerach. W równoległej maszynie obliczeniowej koncepcja znana jako dziel i zwyciężaj. Dziel i zwyciężaj to elegancki sposób rozwiązywania problemu. Dzielisz problemy na mniejsze problemy tego samego typu można rozwiązywać indywidualnie, a częściowe wyniki łączyć w całościowe rozwiązanie. Podejście to służy do dzielenia problemu na coraz mniejsze problemy, aż do momentu, gdy każdy problem zostanie łatwo rozwiązany. Programowanie równoległe nazywane jest programowaniem systemów wieloprocesorowych przy użyciu techniki dziel i zwyciężaj

Intensywne problemy obliczeniowe i aplikacje wymagają dodatkowej mocy obliczeniowej niż kiedykolwiek wcześniej. Chociaż prędkość procesora rośnie, tradycyjne komputery sekwencyjne nie zapewniają mocy potrzebnej do rozwiązania tych problemów. W komputerach równoległych, czyli obszarze, w którym wiele procesorów jednocześnie podejmuje problemy, wiele z tych problemów jest potencjalnie rozwiązywanych.

Na rozwój przetwarzania równoległego wpływa kilka czynników. Wśród nich wyróżniają się następujące:

1. W wielu dziedzinach nauki i inżynierii przetwarzanie równoległe było uważane za „informatykę najwyższej klasy” służącą do modelowania problemów trudnych do rozwiązania: W dziedzinach takich jak
 - Atmosfera, Ziemia, Środowisko
 - Fizyka – stosowana, jądrowa, cząsteczkowa, skondensowana, wysokie ciśnienie, synteza termojądrowa, fotonika
 - Biologia, biotechnologia, genetyka
 - Chemia, nauki molekularne
 - Geologia, sejsmologia
 - Inżynieria mechaniczna – od protetyki po statki kosmiczne
 - Elektrotechnika, projektowanie obwodów, mikroelektronika
 - Informatyka, matematyka
 - Obrona, broń
2. Architektury sekwencyjne są fizycznie ograniczone przez prędkość światła i prawa termodynamiki. Punkt nasycenia (brak wzrostu pionowego) osiągany jest przy prędkości, z jaką mogą działać sekwencyjne procesory. Dlatego alternatywnym sposobem na osiągnięcie dużej szybkości obliczeniowej jest podłączenie kilku procesorów (możliwość poziomej rozbudowy).
3. Ulepszenia sprzętu potokowego, superskali itp. nie są skalowalne i wymagają zaawansowanej technologii kompilatora. Opracowanie tej technologii kompilatora jest trudne
4. Kolejną próbą poprawy wydajności było przetwarzanie wektorów poprzez wykonywanie więcej niż jednego zadania na raz. W tym przypadku wprowadzono możliwość dodawania (lub odejmowania, mnożenia lub innego manipulowania) dwiema tablicami numerycznymi do urządzeń. Był to przydatne, gdy w niektórych zastosowaniach inżynierskich dane naturalnie pojawiały się w postaci wektorów lub macierzy. Przetwarzanie wektorów nie było tak przydatne w aplikacjach z mniej dobrze sformułowanymi danymi.
5. Rzeczywiście prowadzone są szeroko zakrojone prace badawczo-rozwojowe nad narzędziami i środowiskami programistycznymi, a technologia przetwarzania równoległego jest dojrzała i nadaje się do komercyjnego wykorzystania.
6. Niezbędny postęp technologii sieciowych toruje drogę heterogeniczności przetwarzanie danych.

2.3.2 Architektury sprzętowe do przetwarzania równoległego

Komputery równoległe w jakiś sposób podkreślają równoległe przetwarzanie operacji. Wszystkie podstawowe koncepcje przetwarzania równoległego i obliczeń zostały określone w poprzednim rozdziale. Komputery równoległe można rozróżnić na podstawie strumieni danych i instrukcji organizacji komputerowych. Można je również klasyfikować ze względu na strukturę komputera, na przykład wiele procesorów z oddzielną pamięcią lub globalną pamięcią współdzieloną. W programie o nazwie grain

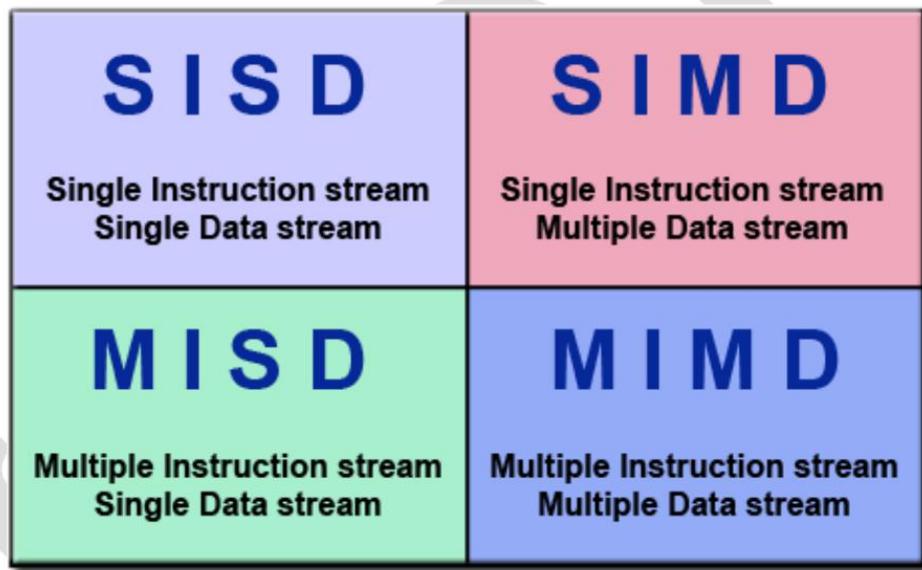
rozmiar, równoległe poziomy przetwarzania można również zdefiniować na podstawie rozmiaru instrukcji. Ale komputery połączone równolegle można klasyfikować według różnych kryteriów

Ustalono następującą klasyfikację komputerów równoległych:

- 1) Klasyfikacja na podstawie instrukcji i strumieni danych
- 2) Klasyfikacja na podstawie budowy komputerów
- 3) Klasyfikacja na podstawie sposobu dostępu do pamięci
- 4) Klasyfikacja na podstawie wielkości ziaren

Klasyczna taksonomia Flynnna

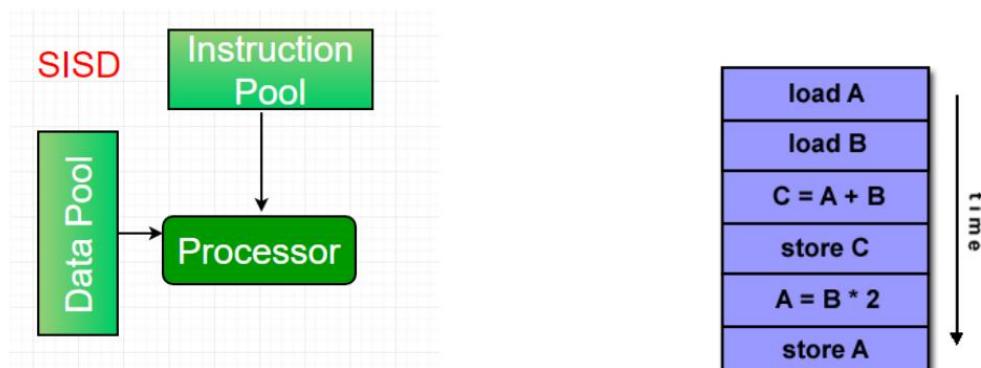
- Komputery równoległe są klasyfikowane na różne sposoby.
- Taksonomia Flynnna jest od tego czasu jedną z najczęściej stosowanych klasyfikacji 1966.
- Taksonomia Flynnna definiuje architekturę komputerów wieloprocesorowych zgodnie z kategoryzacją dwóch odrębnych aspektów instrukcji i strumienia danych. Każdy z tych wymiarów może zawierać tylko stan pojedynczy lub wielokrotny jednego rodzaju.
- Poniższa macierz opisuje cztery możliwe klasyfikacje Flynnna:



RYSUNEK 2.4 • Taksonomia Flynnna

2.3.2.1 Systemy z jedną instrukcją i pojedynczymi danymi (SISD).

System obliczeniowy SISD to jednoprosesor, który może wykonać pojedynczą instrukcję na pojedynczym strumieniu danych. SISD przetwarza instrukcje maszynowe sekwencyjnie. Komputery stosujące ten model są powszechnie określane jako komputery sekwencyjne. Architektura SISD jest stosowana w większości konwencjonalnych komputerów. Wszystkie instrukcje przetwarzania i dane powinny być przechowywane w pamięci podstawowej. W zależności od szybkości, z jaką komputer może przesyłać informacje wewnętrznie, prędkość jest ograniczona w elemencie przetwarzającym modelu SISD. Stacje robocze IBM PC są dominującymi reprezentatywnymi systemami SISD.



RYSUNEK 2.5: Architektura z pojedynczą instrukcją i pojedynczymi danymi (SISD).

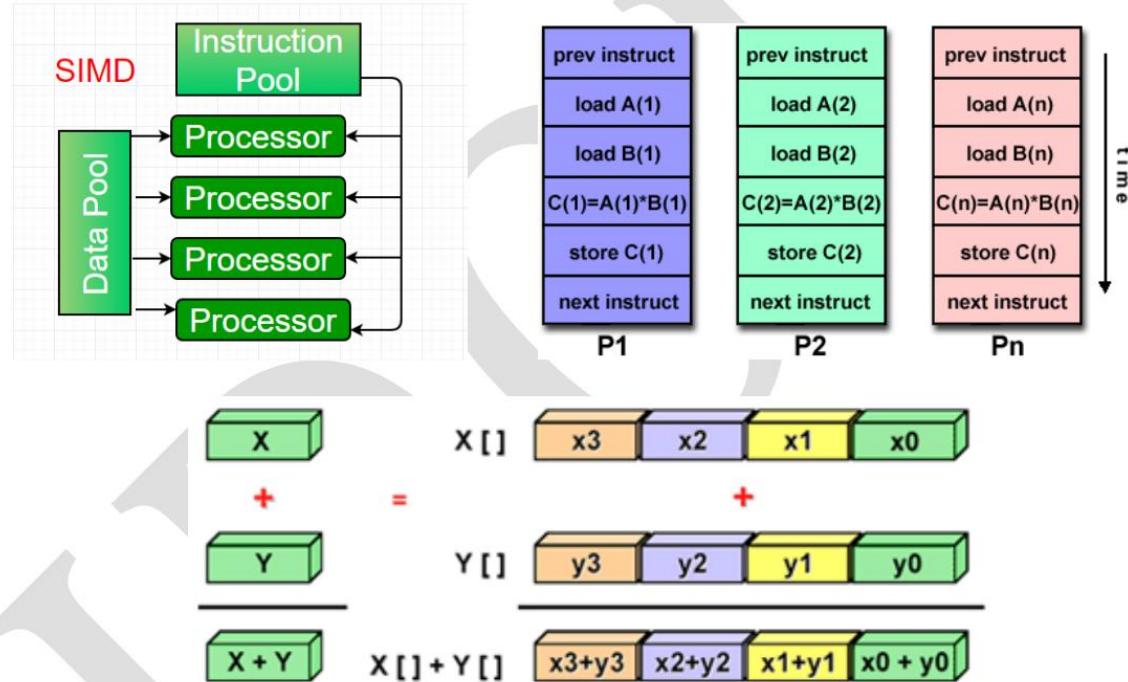
2.3.2.2 Systemy z jedną instrukcją i wieloma danymi (SIMD).

System SIMD to system wieloprocesorowy, który może wykonywać te same instrukcje na wszystkich procesorach, ale działa na wielu strumieniach danych. Maszyny oparte na SIMD idealnie nadają się do obliczeń naukowych, ponieważ obejmują wiele operacji wektorowych i macierzowych. Dane można podzielić na wiele zestawów (N-zestawów dla systemów N PE), dzięki czemu informacja może zostać przeniesiona do wszystkich elementów przetwarzających (PE). Każdy PE może przetwarzać ten sam zestaw danych. Idealnie nadaje się do złożonych problemów o wysokim stopniu regularności, takich jak przetwarzanie grafiki/obrazu. Większość nowoczesnych komputerów, szczególnie tych z procesorami graficznymi (GPU), wykorzystuje instrukcje i jednostki wykonawcze SIMD. Dominującym reprezentatywnym systemem SIMD jest maszyna do przetwarzania wektorów Cray'a.

Przykłady:

Macierze procesorowe: myślące maszyny CM-2, MasPar MP-1 i MP-2, ILLIAC IV

Rurociągi wektorowe: IBM 9000, Cray X-MP, Y-MP i C90, Fujitsu VP, NEC SX-2, Hitachi S820, ETA10



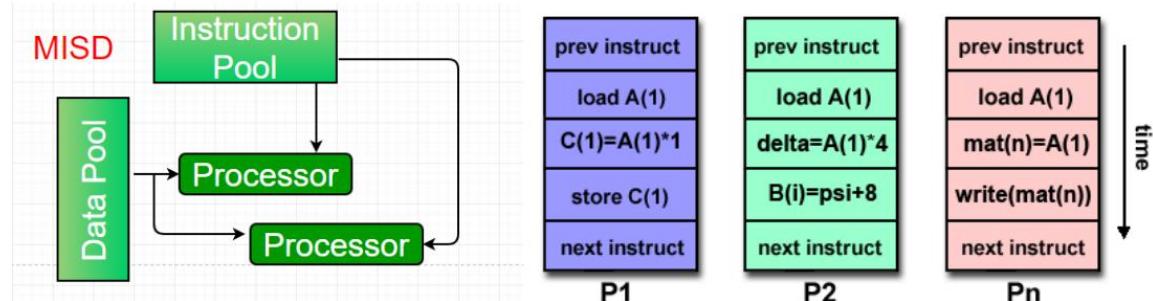
RYSUNEK 2.6: Architektura z pojedynczą instrukcją i wieloma danymi (SIMD).

2.3.2.3 Systemy z wieloma instrukcjami i pojedynczymi danymi (MISD).

MISD to system wieloprocesorowy, który wykonuje różne instrukcje na różnych PE, ale wszystkie działają w tym samym zbiorze danych.

Wiele instrukcji: każda jednostka przetwarzająca działa niezależnie na danych w ramach oddzielnych strumieni instrukcji.

Pojedyncze dane: pojedynczy strumień danych jest wprowadzany do wielu jednostek przetwarzających.



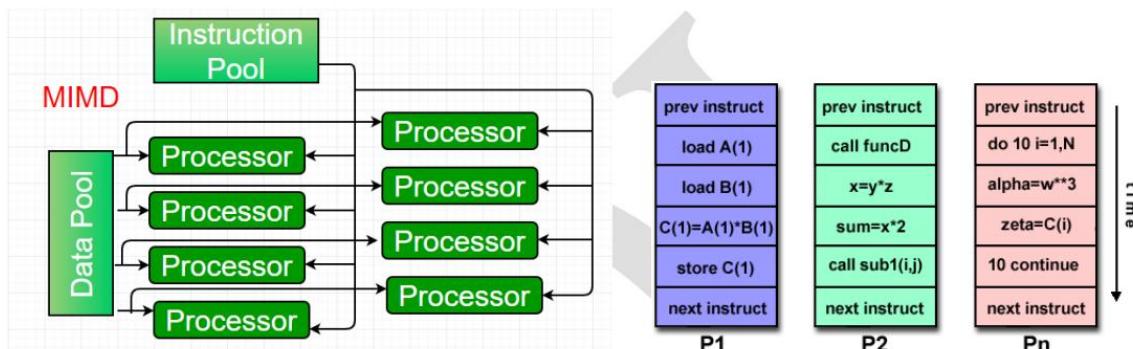
RYSUNEK 2.7 Architektura z wieloma instrukcjami i pojedynczymi danymi (MISD).

Przykład $Z = \sin(x) + \cos(x) + \tan(x)$

Na tym samym zbiorze danych system wykonuje różne operacje. W przypadku większości zastosować maszyny zaprojektowane przy użyciu MISD nie są przydatne, niektóre zostały zaprojektowane, ale żadna z nich nie jest dostępna na rynku.

2.3.2.4 Systemy z wieloma instrukcjami i wieloma danymi (MIMD).

MIMD to system wieloprocesorowy, który może wykonywać wiele instrukcji na wielu zestawach danych. Każdy PE w modelu z MIMD ma osobne instrukcje i strumienie danych, więc na maszynach zbudowanych przy użyciu tego modelu można używać dowolnego typu aplikacji. W porównaniu do SIMD i MISD, PE mogą działać synchronicznie lub asynchronicznie, deterministycznie lub niedeterministycznie w komputerach MIMD. Obecnie najpopularniejszy typ komputera równoległego – do tej kategorii zalicza się większość współczesnych superkomputerów. Przykłady: większość obecnych superkomputerów, połączone w sieć równolegle klastry i „siatki komputerowe”, wieloprocesorowe komputery SMP, wielordzeniowe komputery PC.



RYSUNEK 2.8 Architektura wielu instrukcji i wielu danych (MIMD).

Maszyny MIMD dzieli się ogólnie na MIMD z pamięcią współdzieloną i MIMD z pamięcią rozproszoną w zależności od sposobu, w jaki PE są podłączone do pamięci głównej.

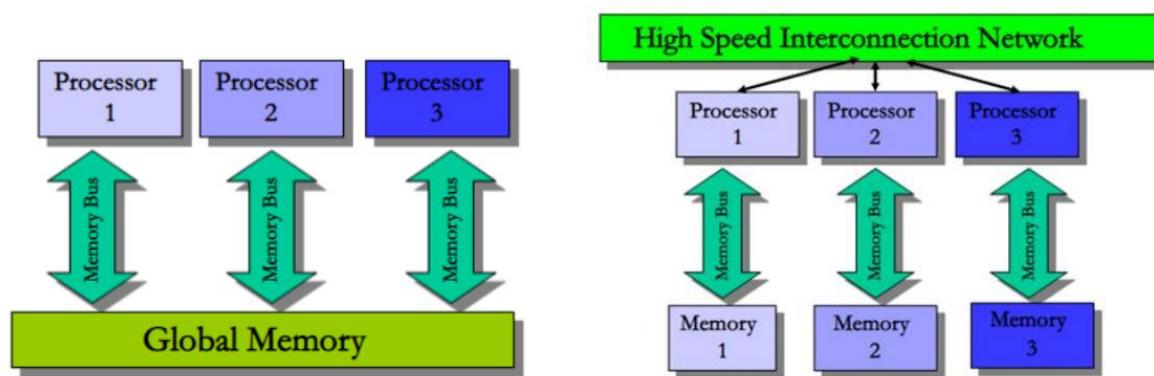
Maszyny MIMD z pamięcią współdzieloną

Wszystkie PE są podłączone do jednej pamięci globalnej we współdzielonym modelu MIMD (ściśle powiązane systemy wieloprocesorowe) i wszyscy mają do niej dostęp. Komunikacja pomiędzy PE w tym modelu odbywa się za pomocą pamięci współdzielonej, zmiany danych zapisanych przez jednego PE w pamięci globalnej są widoczne dla wszystkich pozostałych PE. Dominującymi systemami pamięci współdzielonej dla Silicon Graphics i Sun/IBM (Symmetric Multi-Processing) są systemy pamięci współdzielonej.

Maszyny MIMD z pamięcią rozproszoną

Wszystkie PE mają pamięć lokalną na maszynach MIMD z pamięcią rozproszoną (luźne systemy wieloprocesorowe). W tym modelu komunikacja między PE odbywa się za pośrednictwem sieci połączeń (kanal komunikacji międzyprocesowej lub IPC). Sieć łączącą PE można skonfigurować w formie drzewa, siatki lub w razie potrzeby.

Architektura pamięci współdzielonej MIMD jest łatwiejsza do zaprojektowania, ale mniej odporna na awarie i trudniejsza w rozbudowie w porównaniu z modelem pamięci rozproszonej MIMD. Wspólne awarie MIMD wpływają na cały system, ale nie na model rozproszony, w którym można łatwo odizolować każdy PE. Dla porównania, architektury pamięci współdzielonej MIMD są mniej podatne na skalowanie, ponieważ wprowadzenie większej liczby PE powoduje konflikt pamięci. Inaczej jest w przypadku pamięci rozproszonej, w której każdy PE ma własną pamięć. Dzięki realistycznym efektom i specyfikacjom konsumenckim rozproszona architektura pamięci MIMD jest lepsza od innych.



RYSUNEK 2.9 Architektura MIMD pamięci współdzielonej (po lewej) i rozproszonej (po prawej).

2.3.3 Podejścia do programowania równoległego

Ogólnie rzecz biorąc, program sekwencyjny zawsze uruchamia tę samą sekwencję instrukcji z tymi samymi danymi wejściowymi i zawsze generuje te same wyniki, ponieważ programy muszą być reprezentowane przez podzielenie pracy na kilka części działających na różnych procesorach. Uszkodzony program jest programem równoległym.

Dostępne są różne metody programowania równoległego. Najważniejszy z nich

Czy:

- Równoległość danych
- Równoległość procesów
- Model rolnika i pracownika

Każdy z trzech modeli może zostać użyty do równoległości na poziomie zadania. W przypadku równoległości danych, Dziel i zwyciężaj jest wielorozgałęzionym algorytmem projektowania opartym na rekurencji. Algorytm dziel i zwyciężaj polega na powtarzanym dzieleniu danych na dwa lub więcej podobnych lub powiązanych danych, a do przetwarzania każdego zestawu danych dla różnych PE używane są te same instrukcje. Jest to bardzo przydatne podejście do przetwarzania maszynowego w oparciu o model SIMD.

W przypadku równoległości procesów istnieje wiele (ale oddzielnych) operacji w ramach jednego działania, które można wykonać na kilku procesorach. W modelu rolnika i pracownika główne (główne) obliczenia powodują wiele problemów podrzędnych, które są uruchamiane przez urządzenie podrzędne w celu wykonania. Jedyna komunikacja pomiędzy obliczeniami mastera i slave polega na uruchomieniu mastera obliczenia dla urządzeń podrzędnych i zwrócenie wyniku obliczeń dla urządzeń podrzędnych do urządzenia master.

2.3.4 Poziomy równoległości

Równoległość na poziomie bitowym: w tej równoległości koncentruje się na podwojeniu rozmiaru słowa procesora. Zwiększoną równoległość poziomów bitowych oznacza, że operacje arytmetyczne na dużych liczbach są wykonywane szybciej. Na przykład procesor 8-bitowy potrzebuje 2 cykli, aby wykonać 16-bitowe dodanie, podczas gdy 16-bitowy to pojedynczy cykl. Wydaje się, że wraz z pojawieniem się procesorów 64-bitowych ten stopień równoległości się skończył.

Równoległość na poziomie instrukcji (ILP): Ta forma równoległości ma na celu wykorzystanie możliwego nakładania się instrukcji w programie komputerowym. Na każdym sprzęcie procesora implementowana i stosowana jest większość typów ILP:

Potok instrukcji: Wykonuj różne etapy w tym samym cyklu różnych niezależnych instrukcji i wykorzystuj wszystkie niewykorzystane zasoby.

Równoległość zadań: Równoległość zadań polega na podzieleniu zadania na podzadania, a następnie przydzieleniu każdego z nich do wykonania. Podzadania są realizowane równolegle przez procesory.

Wykonanie poza kolejnością: Instrukcje bez naruszania zależności danych mogą zostać wykonane, jeśli mimo że poprzednie instrukcje są nadal wykonywane, dostępna jest jednostka.

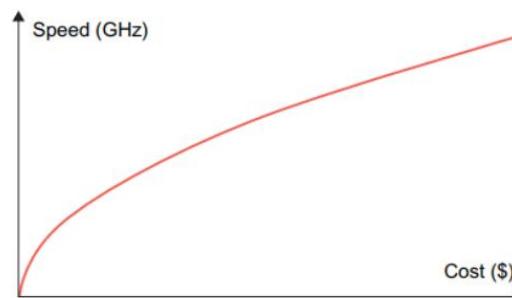
2.3.5 Prawa ostrożności

Już teraz, gdy zaimplementowaliśmy pewne podstawowe elementy obliczeń równoległych w architekturze i modelach, możemy uwzględnić część wiedzy zdobytej podczas projektowania i wdrażania tych systemów. Istnieją zasady, które mogą pomóc nam zrozumieć, w jakim stopniu równoległość pomoże aplikacji lub systemowi oprogramowania.

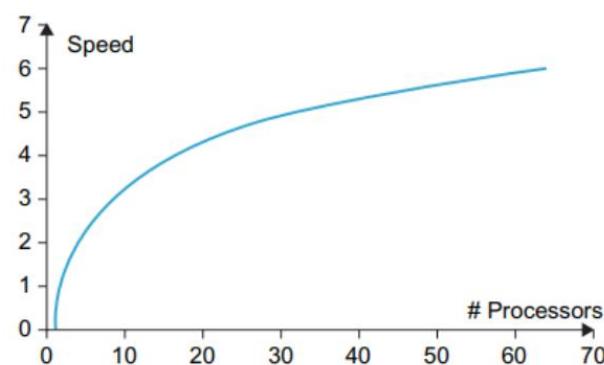
Równoległość jest stosowana w wielu czynnościach łącznie, aby maszyna mogła zmaksymalizować swoją wydajność lub prędkość. Szczególnie warto o tym pamiętać. Ale relacje, które zarządzają wzrostem, nie są prowadzone w tempie liniowym. Na przykład użytkownik zamierza zwiększyć prędkość dla danego procesora maksymalnie n razy. Jest to rozwiązanie optymalne, ale rzadko spotykane ze względu na komunikację napowietrzną.

Należy wziąć pod uwagę dwie ważne wskazówki:

- Szybkość obliczeń jest proporcjonalna do kosztów pierwiastka kwadratowego systemu; nigdy nie wzrasta liniowo. Im szybszy jest system, tym droższa będzie jego prędkość (rysunek).
- Prędkość wzrasta wraz z logarytmem liczby procesorów (tj. $y = k * \log(N)$) komputera równoległego. Rysunek ilustruje tę koncepcję.



RYSUNEK 2.10 Koszt w funkcji szybkości



RYSUNEK 2.11 Liczba procesorów a prędkość.

2.4 Elementy przetwarzania rozproszonego

W tej części poszerzamy te zasady i omawiamy, w jaki sposób można osiągnąć różne zadania, wykorzystując systemy składające się z wielu heterogenicznych systemów komputerowych. Dotyczą one tzw. przetwarzania rozproszonego, a dokładniej, z punktu widzenia projektanta oprogramowania, przedstawiają najważniejsze wytyczne i zasady wdrażania rozproszonych systemów obliczeniowych.

2.4.1 Ogólne pojęcia i definicje

Obliczenia rozproszone badają modele, architektury i algorytmy stosowane w projektowaniu systemów rozproszonych i zarządzaniu nimi. Używamy tej jako ogólnej definicji systemu rozprozonego zaproponowanej przez Tanenbauma

System rozproszony to zbiór niezależnych komputerów, który w oczach jego użytkowników stanowi jeden spójny system

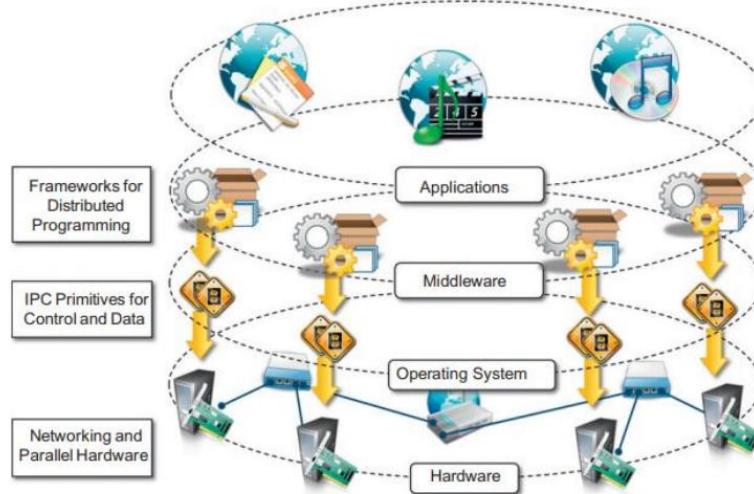
Jest to zdecydowanie idealna forma systemu rozprozonego, która całkowicie ukrywa przed użytkownikiem „szczegóły implementacyjne” tworzenia potężnego systemu z wielu bardziej podstawowych systemów. W tej sekcji koncentrujemy się na modelach architektonicznych, które wykorzystują i prezentują spójny system wykorzystania niezależnych komputerów. Podstawowym krokiem we wszystkich rozproszonych architekturach komputerów jest koncepcja komunikacji komputerowej. System rozproszony to aplikacja, która zbiera protokoły w celu koordynowania kilku procesów działania sieci komunikacyjnej w taki sposób, aby wszystkie komponenty współpracowały w celu wykonania jednego lub kilku podobnych zadań. Współpracujące komputery mogą sterować zarówno zdalnymi, jak i lokalnymi zasobami w systemie rozproszonym za pośrednictwem sieci komunikacyjnej. Wielorakie istnienie Poszczególne komputery w sieci rozproszonej są niewidoczne dla użytkownika. Użytkownik nie wie, że praca jest wykonywana w odległych obszarach, na różnych maszynach. Definicja Coulouris systemu rozprozonego

System rozproszony to taki, w którym komponenty znajdujące się na komputerach połączonych w sieć komunikują się i koordynują swoje działania jedynie poprzez przekazywanie komunikatów.

Komponenty systemu rozprozonego komunikują się za pomocą pewnego rodzaju przekazywania komunikatów, jak zdefiniowano w powyższym opisie. Termin ten obejmuje kilka modeli komunikacji.

2.4.2 Elementy systemu rozprozonego

Prawie wszystkie duże systemy komputerowe są obecnie rozproszone. Systemy są rozproszone. System rozproszony to „zestaw niezależnych maszyn, które przedstawiają użytkownikowi jeden spójny system”. Przetwarzanie informacji jest rozproszone na kilku komputerach, a nie ograniczane do jednego komputera. Przegląd różnych warstw zaangażowanych w świadczenie usług systemu rozprozonego przedstawiono na rysunku 2.12.



RYSUNEK 2.12 Warstwowy widok systemu rozprozonego

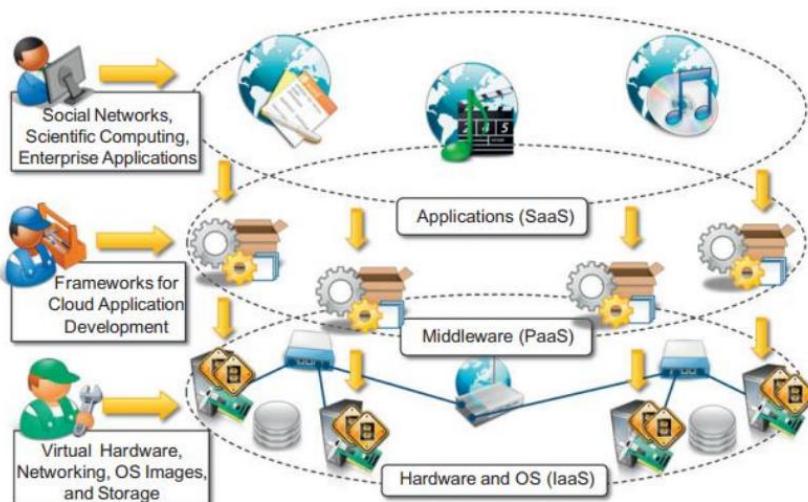
Odniesienie do „Opanowania podstaw przetwarzania w chmurze i programowania aplikacji” autorstwa Rajkumar Buyya)

Na najniższym poziomie infrastruktura fizyczna to sprzęt komputerowy i sieciowy, który jest wyraźnie nadzorowany przez system operacyjny zapewniający podstawowe usługi komunikacji międzyprocesowej (IPC), planowanie i zarządzanie procesem, a także zarządzanie zasobami w systemach plików i systemy lokalne. W połączeniu te dwie pozostałe warstwy stają się podstawą dla specjalistycznego oprogramowania do konwersji wielu komputerów w sieci na system rozproszony

Implementacja dobrze znanych zasad systemów operacyjnych i wielu innych na poziomie sprzętu i sieci umożliwia łatwą integrację heterogenicznych komponentów i ich struktury w spójną, ujednoliczoną strukturę. Na przykład łączność w sieci między różnymi urządzeniami jest regulowana standardami, co pozwala na płynną interakcję. Usługi IPC na poziomie systemu operacyjnego zostały wprowadzone wraz z wprowadzeniem standardowych protokołów komunikacyjnych, takich jak TCP/IP, protokół datagramów użytkownika (UDP) i innych.

Warstwa oprogramowania pośredniego wykorzystuje takie usługi do opracowywania i wdrażania aplikacji rozproszonych w spójnym środowisku. Korzystając z usług udostępnianych przez system operacyjny, oprogramowanie pośrednie tworzy własne protokoły, formaty danych i język programowania lub struktury w celu tworzenia aplikacji rozproszonych. Warstwa ta oferuje wsparcie dla paradygmatów programowania dla systemów rozproszonych. Wszystkie stanowią interfejs całkowicie niezależny od bazowego systemu operacyjnego i obejmują wszystkie heterogeniczności niższych warstw.

Aplikacje i usługi zaprojektowane i zbudowane do użytku z oprogramowaniem pośredniczącym odzwierciedlają górną część stosu systemu rozprozonego. Można je stosować z wielu powodów. Czasami mogą przeglądać swoją funkcjonalność poprzez przeglądarkę internetową w formie interfejsów graficznych (GUI). Na przykład wykorzystanie technologii sieciowej jest wysoce preferowane w kontekście systemu przetwarzania w chmurze nie tylko w przypadku aplikacji interfejsowych rozproszonych aplikacji z konsumentami, ale także w przypadku usług platformowych do tworzenia systemów rozproszonych. Doskonałym przykładem jest dostawca IaaS, na przykład Amazon Web Services (AWS), który zapewnia możliwości tworzenia maszyn wirtualnych, organizuje je w klaster i na nim wdraża aplikacje i systemy. Rysunek przedstawia przykład kontekstualizacji ogólnej architektury referencyjnej systemu rozprozonego w chmurze obliczeniowej.



RYSUNEK 2.13 Rozproszony system przetwarzania w chmurze.

Odniesienie do „Opanowania podstaw przetwarzania w chmurze i programowania aplikacji” autorstwa Rajkumar Buyya)

2.4.2 Style architektoniczne w przetwarzaniu rozproszonym

Systemy rozproszone to także złożone komponenty oprogramowania, które z założenia są dystrybuowane na wiele urządzeń. Ważne jest, aby procesy te były odpowiednio zorganizowane, aby przewyciążyć ich złożoność. Istnieją różne sposoby sprawdzenia, jak zorganizowany jest system rozproszony, ale łatwo jest rozróżnić organizację logiczną od faktycznego fizycznego zaawansowania komponentów oprogramowania.

Struktura systemów rozproszonych opiera się głównie na komponentach oprogramowania sieci. Te architektury oprogramowania informują nas o strukturze i interakcjach różnych komponentów programu.

W tym rozdziale skoncentrujemy się najpierw na niektórych powszechnych podejściach do organizacji systemów komputerowych.

Aby stworzyć efektywną sieć rozproszoną, należy zamontować moduły oprogramowania na określonych komputerach i na nich umieścić. Istnieje wiele różnych możliwości wyboru.
Czasami nazywana architekturą urządzenia ostateczną instancją architektury oprogramowania.

W tym rozdziale przyjrzymy się tradycyjnym architekturom centralnym, w których większość komponentów oprogramowania (a tym samym funkcji) jest implementowana przez pojedynczy serwer, podczas gdy zdalni klienci mogą uzyskać dostęp do tego serwera za pomocą prostych metod komunikacji. Co więcej, systemy zdecentralizowane, w których komputery pełnią mniej więcej te same role, nazywamy organizacjami hybrydowymi.

Style architektoniczne

Pierwotnie znajdujemy logiczne rozmieszczenie systemów rozproszonych w moduły oprogramowania, zwane także architekturą oprogramowania. Prace związane z architekturą komputerową ewoluowały dramatycznie, a projektowanie lub wdrażanie architektur jest obecnie powszechnie uznawane za niezbędną do rozwoju dużych systemów.

Idea stylu architektonicznego jest ważna dla naszej dyskusji. Styl taki formułuje się w komponentach, połączeniach pomiędzy komponentami, wymianie danych pomiędzy komponentami oraz konfiguracji tych elementów razem w jeden system. Część to modułowa jednostka z dobrze zdefiniowanymi interfejsami, którą można wymienić w swoim otoczeniu. Jak omówiono poniżej, kluczową kwestią dotyczącą komponentu systemu rozproszonego jest to, że komponent można zastąpić, jeśli znane są jego interfejsy. Dużo bardziej złożonym terminem jest łącznik, zwykle definiowany jako mechanizm komunikacji, koordynacji lub współpracy pomiędzy elementami. Na przykład (zdalne) wywołania procedur, przekazywanie komunikatów lub dane przesypane strumieniowo mogą być generowane przez łącznik.

Style architektoniczne są podzielone na dwie główne klasy:

- Style architektury oprogramowania
- Style architektoniczne systemu

Pierwsza klasa dotyczy logicznej struktury oprogramowania; druga klasa obejmuje wszystkie typy reprezentujące fizyczną strukturę systemów oprogramowania reprezentowanych przez ich główne komponenty.

2.4.3.1 Komponent i łączca

Wizje komponentów i konektorów opisują modele składające się z elementów o określonej obecności w czasie, takich jak procesy, obiekty, klienci, serwery i przechowywanie danych. Ponadto modele komponentów i łączcy jako komponenty zapewniają mechanizmy interakcji, takie jak łącza i protokoły komunikacyjne, przepływy informacji i dostęp do współdzielonej pamięci. Interakcje te są również prowadzone w złożonej infrastrukturze, takiej jak systemy oprogramowania pośredniego, kanały komunikacji i planiści procesów. Komponent jest jednostką behawioralną. Opis komponentu określa, co zadanie może i musi wykonać. łącznik wskazuje, że jeden komponent jest zwykle połączony relacjami, takimi jak przepływ danych lub przepływ sterowania. Złącze jest mechanizmem.

2.4.3.2 Style architektury oprogramowania

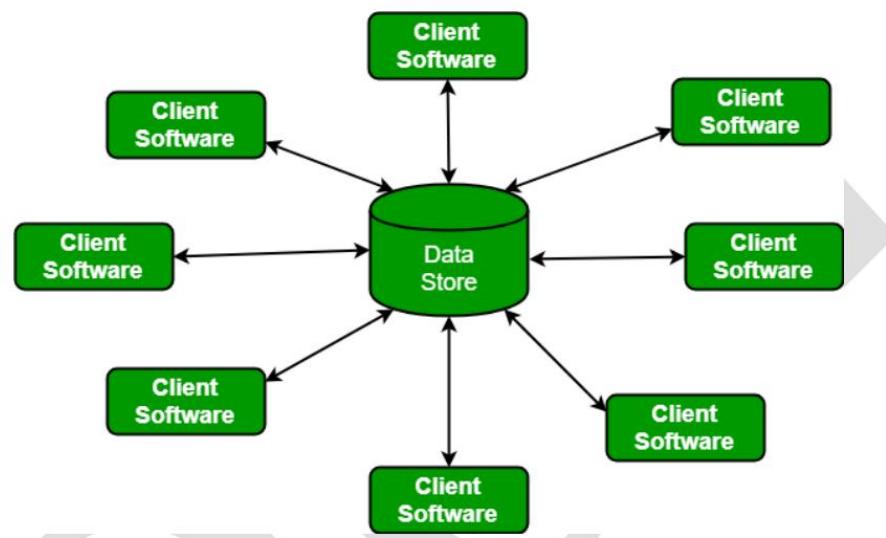
Style i wzorce w architekturze oprogramowania definiują sposób organizacji komponentów systemu w celu budowania kompletnego systemu i spełnienia wymagań klienta. Liczba z

style i wzorce architektury oprogramowania są dostępne w branży oprogramowania, dlatego konieczne jest zrozumienie specjalnego projektu projektu.

Modele te stanowią podstawę, na której logicznie budowane są systemy rozproszone i są omówione w kolejnych sekcjach.

Architektury skoncentrowane na danych

W centrum tej architektury znajduje się magazyn danych, do którego często można uzyskać dostęp za pośrednictwem innych komponentów, które aktualizują, dodają, usuwają lub modyfikują dane znajdujące się w magazynie. Rysunek 2.1.4 przedstawia typowy styl skoncentrowany na danych. Dostęp do centralnego repozytorium uzyskuje się za pośrednictwem oprogramowania klienckiego. Odmiana tej metody służy do zamiany repozytorium w tablicę, ponieważ dane klienta lub dane o zainteresowaniach klienta zmieniają powiadomienia klienta. Ułatwia to integralność z tą architekturą skoncentrowaną na danych. Pozwala to na zmiany w istniejących komponentach i dodawanie nowych komponentów klienta do architektury bez zgody i obaw innych klientów. Klienci mogą wykorzystywać mechanizmy tablicowe do przesyłania danych.



RYSUNEK : 2.14 Typowy styl zorientowany na dane

Architektura repozytorium składa się z centralnej (często bazy danych) struktury danych i niezależnego zbioru komponentów funkcjonujących w centralnej strukturze danych. Na przykład architektury tablicowe, w których tablice służą jako centra komunikacji źródeł wiedzy i repozytoriów dla wielu aplikacji, obejmują architektury repozytoriów. Repozytoria, w tym tworzenie oprogramowania, CAD, są ważne w integracji danych i są wdrażane w różnych aplikacjach.

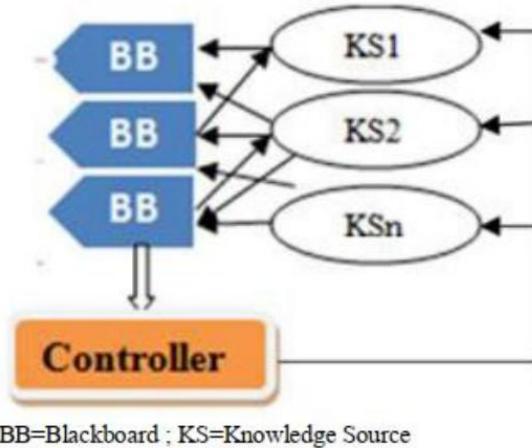
W stylu tablicy główne elementy pokazano na rysunku 1. Problem został rozwiązany na podstawie kilku źródeł wiedzy. Problem rozwiązuje każde źródło informacji, a jego rozwiązanie, częściowe rozwiązanie lub sugestię zapisuje się na tablicy.

Mniej więcej w tym samym czasie każde inne źródło wiedzy modyfikuje lub rozszerza rozwiązanie podane przez poprzednie źródło wiedzy lub pisze, aby samo rozwiązać problem.

Powłoka kontrolna służy do porządkowania i monitorowania działań źródeł informacji, tak aby nie stworzyły one bałaganu, który mógłby odbiegać od bieżącego przebiegu projektu.

Jest to zarządzanie, monitorowanie i kontrola wszystkich działań prowadzonych podczas sesji rozwiązywania problemów za pośrednictwem powłoki kontrolnej.

Skalowalność to jedna z zalet tego projektu architektonicznego. Źródło wiedzy można łatwo dodać lub usunąć z programu w zależności od potrzeb. Źródła wiedzy są niezależne, a zatem możliwe do jednoczesnego wykorzystania pod warunkiem ograniczenia elementu kontrolnego. Problemem w tej architekturze jest to, że nie wiadomo z góry, kiedy zatrzymać proces znajdowania rozwiązania, ponieważ zawsze możliwe jest dalsze udoskonalanie. Para synchronizatorów. Trudno jest dotrzeć do wielu źródeł wiedzy



RYSUNEK 2.15 . Architektura tablicowa

Architektury przepływu danych

Architektura przepływu danych konwertuje dane wejściowe na dane wyjściowe poprzez zbiór elementów obliczeniowych lub zwodniczych. Jest to architektura komputera, która nie ma licznika programów, więc wykonanie jest nieprzewidywalne, co oznacza, że zachowania są nieokreślone. Przepływ danych to aspekt modelu komputerowego Von-neumanna składający się z pojedynczego licznika programu, wykonywania sekwencyjnego i przepływu sterowania, który definiuje kolejność pobierania, wykonywania i zatwierdzania.

Ta architektura została pomyślnie zastosowana.

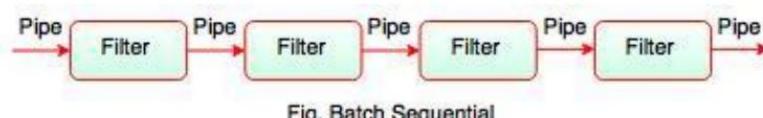
- Architektura przepływu danych eliminuje czas programowania i umożliwia szybkie przełączanie od projektu po realizację.
- Ma na celu przede wszystkim osiągnięcie możliwości ponownego wykorzystania i modyfikacji.
- Dzięki architekturze przepływu danych dane mogą przepływać bez cykli topologię wykresu lub strukturę liniową.

Moduły są realizowane w dwóch różnych typach:

1. Partia sekwencyjna
2. Rura i filtr

Partia sekwencyjna

- Kompilacja sekwencyjna wsadowa w 1970 r. była uważana za proces sekwencyjny.
- W trybie Batch sequential systemy z oddzielnymi programami są uruchamiane sekwencyjnie, a dane są przesyłane z jednego programu do drugiego w formie agregacji.
- Jest to typowy paradygmat przetwarzania danych.



RYSUNEK 2.16 Sekwencja wsadowa

- Powyższy diagram przedstawia przepływ architektury sekwencyjnej wsadowej. Oferuje prostszy podział na podsystemy, a każdy podsystem może być niezależnym programem, który pracuje na danych wejściowych i generuje dane wyjściowe. • Największą wadą sekwencyjnych architektur wsadowych jest brak współbieżności i interaktywnego interfejsu. Zapewnia duże opóźnienia i niską przepustowość.

Rura i filtr

- Potok to łącznik przesyłający dane z jednego filtra do drugiego.
- Potok to kierunkowy strumień danych, który bufor danych implementuje w celu przechowywania wszystkich danych, zanim następny filtr zdąży je przetworzyć.
- Przenosi dane z jednego źródła danych do jednego ujścia danych.
- Bezstanowy strumień danych to potoki.

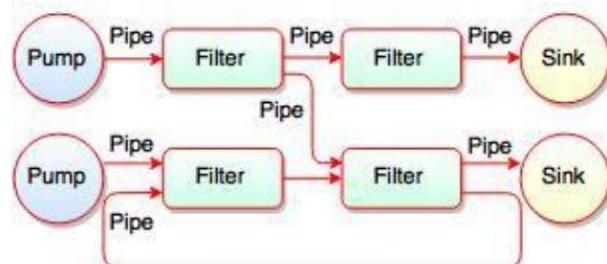


Fig. Pipes and Filters

RYSUNEK 2.17 Rura i filtr

- Powyższy rysunek przedstawia kolejność filtru rurowego. Wszystkie filtry to procesy działające współbieżnie, co oznacza, że mogą działać jako oddzielne wątki lub współprogramy lub być w całości zlokalizowane na różnych komputerach.
- Każda rura posiada przyłącze filtrujące i pełni swoją rolę w działaniu filtra. Filtry są wytrzymałe, a rury można dodawać i usuwać w czasie pracy.
- Filtr odczytuje dane z potoków wejściowych, wykonuje na nich swoją funkcję i umieszcza wynik na wszystkich potokach wyjściowych. Jeśli rury wejściowe są za mało danych, filtr tylko na nie czeka.

Filtr

- Filtr jest elementem.
- Interfejsy służą do przepływu różnych wejść i wypływanego różnych wyjść.
- Przetwarza i udoskonala wprowadzane dane.
- Niezależne byty to filtry.
- Istnieją dwa sposoby tworzenia filtru:
 1. Aktywny filtr
 2. Filtr pasywny
- Aktywny filtr tworzy przepływ danych potoków.
- Przepływ danych w rurach jest sterowany przez filtr pasywny.
- Filtr nie ma wspólnego stanu z innymi filtrami.
- Tożsamość filtrów wejściowych i końcowych jest niejasna.
- Do filtrów zastosowano osobne gwinty. Mogą to być wątki lub współprogramy sprzętu lub oprogramowania.

Zalety rur i filtrów

- Filtr rurowy zapewnia wysoką przepustowość i nadmierne przetwarzanie danych efektywność.
- Umożliwia prostszą konserwację systemu i zapewnia możliwość ponownego użycia.
- Ma niską łączność i elastyczność w trybie sekwencyjnym i równoległym wykonanie.

Wady rur i filtrów

- Nie można uzyskać dynamicznych interakcji za pomocą rur i filtrów. • Do transmisji danych w formacie ASCII potrzebny jest niski wspólny mianownik.
- Architektura filtra rurowego może być trudna do dynamicznej konfiguracji.

Architektury maszyn wirtualnych

Architektura maszyny wirtualnej odnosi się do ustrukturyzowanej specyfikacji interfejsu systemu, obejmującej logiczne zachowanie zasobów obsługiwanych przez interfejs. Implementacja definiuje rzeczywistą implementację architektury. Poziomy abstrakcji to warstwy projektowe, zarówno sprzętowe, jak i programowe, z których każda jest powiązana z innym interfejsem lub architekturą. W systemach korzystających z tej konstrukcji ogólny interfejs wygląda następująco: oprogramowanie (lub aplikacja) określa swoje funkcje i stan, zgodnie z interpretacją silnika maszyny wirtualnej, w formacie abstrakcyjnym.

Wdrożenie opiera się na zrozumieniu programu. Silnik zachowuje wewnętrzną strukturę stanu programu. Bardzo częstymi przykładami w tej grupie są systemy oparte na regułach, interpreterы и procesоры языка польских. Najprostszym rodzajem sztucznej inteligencji są systemy oparte na regułach (znane również jako systemy produkcyjne lub systemy eksperckie). Program oparty na regułach wymaga reguł reprezentowania wiedzy za pomocą wiedzy zakodowanej w systemie. Koncepcje systemu opartego na regułach zależą prawie całkowicie od systemów ekspertowych, które naśladują ludzkie rozumowanie eksperckie przy rozwiązywaniu kwestii wymagających dużej wiedzy. Zamiast opisywać wiedzę jako zbiór prawdziwych faktów w sposób statyczny i deklaratywny, struktura oparta na regułach przedstawia wiedzę jako zbiór praw mówiących, co należy robić, a czego nie. Domena sieciowa zapewnia kolejne fascynujące zastosowanie systemów opartych na regułach: systemy wykrywania włamań do sieci (NIDS) również opierają się na zestawie reguł służących do klasyfikowania podejrzanych zachowań związanych z potencjalnymi włamaniemi do urządzeń komputerowych.

Styl tłumacza

Interpreter to styl architektoniczny idealny do zastosowań, w których nie można użyć najbardziej odpowiedniego języka lub maszyny do wykonania rozwiązania. Styl składa się z kilku części, którymi są program, który próbujemy uruchomić, interpreter, który próbujemy zinterpretować, bieżący stan programu oraz interpreter i część pamięci, która będzie przenosić program, aktualny stan programu i jego bieżący stan. Wezwania do procedur komunikacji między elementami i bezpośredniego dostępu do pamięci są łącznikiem stylu architektonicznego interpretera.

Cztery kompozycje tłumacza:

- Tłumacz silnikowy: praca tłumacza została zakończona
- Obszar przechowywania danych: zawiera pseudokod
- Pole magazynu danych: Raportuje bieżący stan silnika interpretera
- zewnętrzna struktura danych: śledzi rozwój interpretowanego kodu źródłowego

Wejście: wejście do części interpretowanego programu jest przekazywane do stanu programu, w którym program czyta interpreter

Wyjście: Wyjście Programu jest umieszczane w stanie programu, w którym dane są interpretowane przez część systemu interfejsu

Model ten jest całkiem przydatny w projektowaniu maszyn wirtualnych dla programowania wysokiego poziomu (Java, C#) i języków skryptowych (Awk, PERL i tak dalej).

- Przenośność i elastyczność aplikacji na różnych platformach
- Wirtualizacja. Można na nim wykonać kod maszynowy dla jednej architektury sprzętowej inny za pośrednictwem maszyny wirtualnej.
- Zachowanie systemu określone przez niestandardowy język lub strukturę danych; ułatwia tworzenie i zrozumienie oprogramowania.
- Wspiera dynamiczne zmiany (Efektywność)
- Zwykle tłumacz musi jedynie przetłumaczyć kod na a

- Reprezentacja pośrednia (lub brak tłumaczenia), dzięki czemu testowanie zmian zajmuje znacznie mniej czasu.

Interpreter lub maszyna wirtualna nie musi przestrzegać wszystkich instrukcji przetwarzanego kodu źródłowego. Może w szczególności odmówić wykonania kodu naruszającego jakiekolwiek ograniczenia bezpieczeństwa, w ramach których funkcjonuje. Na przykład. Interpreter JS to interpreter JavaScript działający w trybie piaskownicy w JavaScript. Dowolny kod JavaScript można wykonać linia po linii. Wydajność głównego środowiska JavaScript jest całkowicie izolowana. W instancjach JS-Interpretera dostępny jest wielowątkowy konkurencyjny JavaScript bez użycia procesów roboczych.

Architektury połączeń i zwrotów

Najczęściej używanym wzorcem w systemach komputerowych były architektury połączeń i zwrotów styl. Mechanizm wywołania lub wywołania funkcji obejmuje programy główne i podprogramy, zdalne wywołania procedur, systemy obiektowe i systemy warstwowe. Wszystkie podlegają wezwaniu i powracają w stylu architektury.

Styl z góry na dół.

Podejście odgórne polega zasadniczo na rozbiciu systemów w celu uzyskania szczegółów na temat jego składowych podstruktur w sposób inżynierii odwrotnej (znany również jako projektowanie krokowe i rafinacja krokowa, a w niektórych przypadkach stosowany w sposób dekompozycji).

W podejściu odgórny dokonuje się przeglądu systemu i definiuje wszystkie podsystemy pierwszego poziomu, ale nie jest to kompleksowe. Każdy podsystem jest następnie modyfikowany, często na różnych innych poziomach podsystemu, aż do zredukowania pełnej specyfikacji do mniejszych elementów. „Czarne skrzynki” pomagają zdefiniować układ od góry do dołu, którym łatwiej jest manipulować.

Niemniej jednak czarne skrzynki nie mogły wyjaśnić ani być wystarczająco precyzyjne, aby skutecznie zweryfikować model. Szeroki obraz zaczyna się od podejścia odgórnego. Dzieli się na mniejsze kawałki.

Podejście odgórne polega na podziale problemu pomiędzy zadaniami i podziale zadań na mniejsze podzadania. W tym podejściu najpierw opracowujemy moduł główny, a następnie opracowujemy moduły kolejnego etapu. Proces ten trwa aż do utworzenia wszystkich modułów.

Styl obiektowy.

Program zorientowany obiektowo, zamiast działań i logiki, jest paradigmatem języka programowania zorganizowanym wokół obiektów i danych. Aby pobrać dane, przetworzyć je i wygenerować wyniki, organizuje się tradycyjny program postępowania. Program został centralizowany pod względem logicznym, a nie danych. Koncentrują się na programowaniu obiektowym na obiektach i ich manipulacji, a nie na logice, która je tworzy.

Pierwszą fazą OOP jest modelowanie danych, które obejmuje definiowanie, manipulowanie i relacje obejmujące wszystkie obiekty. Modelowanie danych to faza planowania wymagająca ogromnej uwagi. Mamy metodę tworzenia tych obiektów po zidentyfikowaniu każdego obiektu biorącego udział w programie. Jest to znane jako mechanizm klasowy. Klasa obejmuje dane lub właściwości oraz logiczną sekwencję metod manipulowania danymi.

Każdy sposób jest odrębny i nie należy powtarzać uzasadnienia ustalonego już w przypadku innych metod.

Style architektoniczne oparte na niezależnych komponentach

Szereg niezależnych procesów/obiektów komunikujących się za pomocą komunikatorów stanowi część niezależnej architektury komponentowej. Wiadomości mogą być przesyłane w paradigmacie publikowania/subskrybowania dla danego lub nienazwanego uczestnika.

Komponenty zazwyczaj nie kontrolują się nawzajem, wysyłając dane. Można to zmienić, ponieważ komponenty są izolowane.

Przykłady: Podsystemami tego typu są systemy zdarzeń i proces komunikacji.

Systemy eventowe

Ten paradygmat oddziela implementację komponentu od znajomości nazw i lokalizacji komponentów. Wzór wydawcy/subskrybenta, gdzie:

Wydawcy: reklamuj dane, którymi chcesz się podzielić z innymi

Abonent(y): Otrzymanie zainteresowania rejestrem opublikowanych danych.

Do komunikacji między komponentami używany jest menedżer komunikatów. Wydawcy wysyłają wiadomości do menedżera, który rozprowadza je wśród subskrybentów.

Proces komunikacji

Architektoniczny typ procesu komunikacji jest również znany jako architektura klient-serwer.

Klient: rozpoczyna wywołanie serwera żądające jakieś usługi.

Serwer: dostarcza dane klienta.

Zwraca dostęp do danych, gdy serwer działa synchronicznie

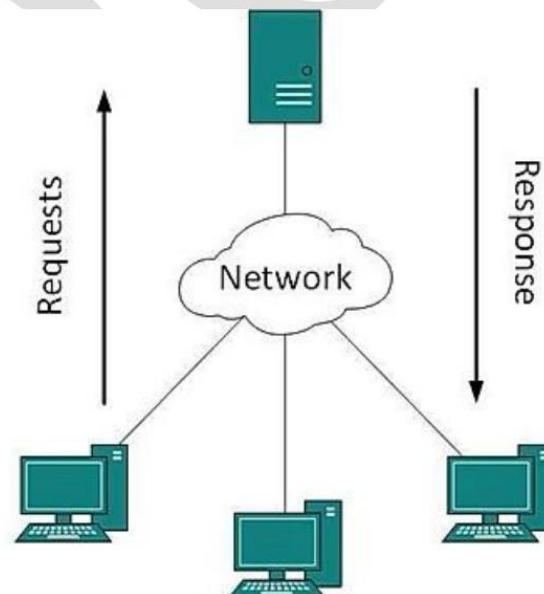
2.4.3.3 Style architektoniczne systemu

Klient-serwer i peer-to-peer (P2P) to dwie kluczowe architektury na poziomie systemu, z których korzystamy obecnie. Na co dzień korzystamy z tych dwóch rodzajów usług, jednak różnica między nimi jest często błędnie interpretowana.

Architektura serwera klienta

W architekturze klienta-serwera znajdują się dwa główne komponenty. Serwer i klient.

Serwer jest miejscem wszystkich transmisji, transmisji i przetwarzania danych, podczas gdy klient może uzyskać dostęp do usług i zasobów serwera zdalnego. Serwer umożliwia klientom wysyłanie żądań, a serwer odpowie. Ogólnie rzecz biorąc, strona zdalna jest zarządzana tylko przez komputer. Aby jednak zachować bezpieczeństwo, stosujemy kilka technik równoważenia obciążenia serwery.



RYSUNEK 2.18 Style architektoniczne klienta/serwera

Architektura Client Server to standardowa funkcja projektowa ze skoncentrowaną bazą danych zabezpieczeń. Ta baza danych zawiera informacje dotyczące bezpieczeństwa, takie jak dane uwierzytelniające i dane dostępu. W przypadku braku kluczy bezpieczeństwa użytkownicy nie mogą zalogować się na serwer. Dlatego architektura ta staje się nieco bardziej stabilna i bezpieczna niż Peer to Peer. Stabilność wynika z faktu, że baza danych zabezpieczeń może zapewnić bardziej efektywne wykorzystanie zasobów. Z drugiej jednak strony system może ulec awarii, ponieważ w określonym czasie serwer może wykonać tylko niewielką ilość pracy.

Zalety:

- Łatwiejszy w budowie i utrzymaniu

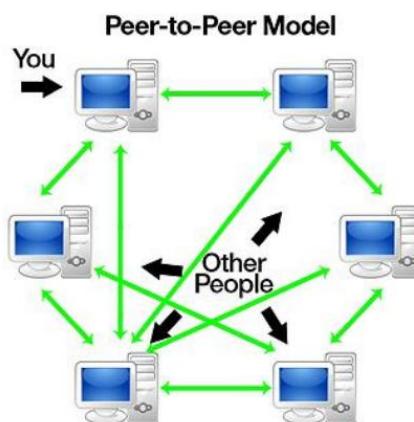
- Lepsze bezpieczeństwo
- Stabilne

Niedogodności:

- Pojedynczy punkt awarii.
- Mniej skalowalna

Każdy z każdym (P2P)

W systemie rozproszonym nie ma centralnego sterowania za peer to peer. Podstawową ideą jest to, że w pewnym momencie każdy węzeł może być klientem lub serwerem. Jeśli węzeł o coś zapyta, można to nazwać klientem, a jeśli coś przyjdzie z węzła, można to nazwać serwerem. Zwykle każdy węzeł nazywany jest równorzędnym.



RYSUNEK 2.19 Peer to Peer (P2P)

Każdy nowy węzeł najpierw dołączy do tej sieci. Po dołączeniu mogą poprosić o usługę lub ją świadczyć. Faza inicjacji węzła (dołączenia do węzła) może się różnić w zależności od implementacji sieci. Nowy węzeł może dowiedzieć się, co zapewniają inne węzły, na dwa sposoby.

Scentralizowany serwer wyszukiwania — nowy węzeł musi zarejestrować i wspomnieć o usługach w sieci za pomocą scentralizowanego serwera wyszukiwania. Zatem wystarczy skontaktować się ze scentralizowanym systemem wyszukiwania, gdy zajdzie taka potrzeba, a system przekieruje Cię do odpowiedniego usługodawcy.

System zdecentralizowany — węzeł poszukujący określonych usług będzie rozwijał i wysyłał żądania do innych węzłów w sieci, aby dostawca usług mógł odpowiedzieć.

Porównanie serwera klienckiego i architektury peer to peer

PODSTAWA DLA PORÓWNANIE	KLIENT-SERWER	PEER-TO-PEER
Podstawowy	Jest konkretny serwer i konkretnych klientów podłączony do serwera	Klienci i serwer nie są wybitny; każdy węzeł działa jako klient i serwer.
Praca	Żądanie klienta do serwisu i serwer odpowiedź serwisem.	Każdy węzeł może zażądać usług i może również świadczyć usługi.
Centrum	Dzielenie się informacjami.	Łączność.
Dane	Dane są przechowywane w scentralizowany serwer.	Każdy peer ma swoje własne dane.
serwer	Gdy kilku klientów zapytanie o usługi jednocześnie serwer może wpalić w pułapkę.	Ponieważ usługi świadczone są przez kilka serwerów rozproszonych w systemie peer-to-peer, serwer nie stanowi wąskiego gardła.
Koszt	Klient-serwer jest drogie w realizacji.	Wdrożenie rozwiązań typu peer-to-peer jest tańsze.

Stabilność	Klient-Serwer to coś więcej stabilny i skalowalny.	Peer-to-Peer ucierpi, jeśli liczba peerów w systemie wzrośnie.
------------	--	--

2.4.4 Modele komunikacji międzyprocesowej

System rozproszony to zbiór komputerów, które dla swoich użytkowników zachowują się jak spójna sieć. Ważną rzeczą jest to, że różnice między różnymi komputerami i sposobem ich interakcji są często ukryte przed użytkownikami. Daje to użytkownikowi pojedynczy obraz systemu. System operacyjny ukrywa wszystkie szczegóły komunikacji pomiędzy procesami użytkownika. Użytkownik nie wie, że istnieje wiele systemów. Komunikacja międzyprocesowa zwana IPC odbywa się za pomocą różnych mechanizmów w systemach rozproszonych i dla różnych systemów mechanizmy te mogą się różnić. Innym istotnym aspektem jest to, że użytkownicy i aplikacje mogą komunikować się w spójny i jednolity sposób z systemem rozproszonym

Komunikacja pomiędzy różnymi procesami jest istotą wszystkich systemów rozproszonych i ważne jest zrozumienie, w jaki sposób procesy mogą dzielić się informacjami na różnych maszynach. W celu wymiany danych pomiędzy dwiema aplikacjami i procesami, komunikacja między procesami lub IPC, jak sama nazwa wskazuje. Procesy mogą znajdować się na tym samym komputerze lub w innym miejscu. Komunikacja w systemach rozproszonych często opiera się na przesyłaniu komunikatów niskiego poziomu, zapewnianym przez sieć bazową. Trudno jest komunikować się poprzez przekazywanie komunikatów niż prymitywną komunikacją opartą na pamięci współdzielonej dostępnej na platformach nierożproszonych

Komunikacja międzyprocesowa (IPC) to metoda komunikacji i synchronizacji systemów. Komunikację między takimi procesami można uznać za metodę współpracy między nimi. Te trzy metody pozwalają procesom komunikować się między sobą: pamięć współdzielona, zdalne wywoływanie procedur (RPC) i przekazywanie komunikatów. W systemach rozproszonych dużą popularnością cieszą się IPC z gniazdami. Krótko mówiąc, adres IP i numer portu to para gniazd. Każdy z nich wymaga gniazda, aby dwa procesy mogły się komunikować.

Jeśli demon serwera działa na hoście, nasłuchuje na swoim porcie i zarządza wszystkimi żądaniami klientów wysyłanymi do portu klienta (gniazda serwera). Aby wysłać wiadomość, klient musi znać adres IP i port serwera (gniazdo serwera). Gdy klient rozpoczęcie komunikację z serwerami i zostanie zwolniona po zakończeniu komunikacji, jądra systemu operacyjnego udostępniają również port klienta.

Chociaż komunikacja za pomocą gniazd jest popularna i skuteczna, uważa się ją za mało popularną, ponieważ gniazda umożliwiają przesyłanie nieuszkodzonych strumieni bajtów jedynie pomiędzy procesami. Dane przesyłane jako strumień bajtów są organizowane przez aplikacje klienckie i serwerowe.

2.4.4.1 Komunikacja oparta na komunikatach

Abstrakcja komunikatów jest niezbędna w rozwoju modeli i technologii umożliwiających przetwarzanie rozproszone. System rozproszony to system, w którym komponenty znajdują się w komunikacji sieciowej i jedynie poprzez przesyłanie komunikatów koordynują swoje funkcje.

W ramach tej wiadomości identyfikowane są wszelkie poufne dane przekazywane od jednej osoby do drugiej. Obejmuje dowolny typ reprezentacji danych z ograniczeniami dotyczącymi rozmiaru i czasu podczas wywoływania zdalnego procesu, sekwencji instancji obiektu lub wspólnego komunikatu. Właśnie dlatego „model komunikacji oparty na komunikatach”, który opiera się na abstrakcji strumienia danych, może odnieść korzyści dzięki odwoływaniu się do różnych modeli komunikacji między procesami.

Pomimo abstrakcji, jaką pokazano programistom w programowaniu koordynacji wspólnych komponentów, różne modele programowania rozproszonego wykorzystują ten typ komunikacji. Poniżej kilka głównych rozproszonych modeli programowania wykorzystujących szablony wiadomości,

Chociaż komunikacja za pomocą gniazd jest popularna i skuteczna, uważa się ją za mało popularną, ponieważ gniazda umożliwiają przesyłanie nieuszkodzonych strumieni bajtów jedynie pomiędzy procesami. Dane przesyłane jako strumień bajtów są organizowane przez aplikacje klienckie i serwerowe.

Przekazywanie wiadomości

Zasada przekazu jest w tym modelu realizowana jako główna abstrakcja modelu. Jednostki wymieniające jawnie zakodowane dane i informacje w formie wiadomości. Struktura i treść przekazu różnią się lub różnią w zależności od modelu. Message Passing Interface i OpenMP są znaczącymi przykładami tego typu modelu.

Zdalne wywołanie procedury

Model ten bada klucze wywołania procedury poza granicami pojedynczego procesu, sugerując wykonanie systemu w procesach zdalnych. Zawiera główny klient-serwer. Zdalny proces utrzymuje komponent serwera, umożliwiając procesom klienckim odwoływanie się do procesów i zwraca wynik wykonania. Komunikaty generowane przez implementację zdalnego wywołania procedury (RPC) samodzielnie zbierają informacje o metodzie i wykonują wymagane dla niej argumenty, a także zwracają wartości. Użycie komunikatów nazywa się zestawianiem argumentów i zwracanych wartości.

Rozproszone obiekty

Jest to implementacja zorientowanego obiektowo modelu zdalnego wywołania procedury (RPC), rozumianego w kontekście metod zdalnego wywoływania, które są rozszerzane poprzez obiekty. Każdy proces przypisuje szereg interfejsów, do których można uzyskać zdalny dostęp. Proces klienta może żądać i wywoływać metody dostępne za pośrednictwem tych interfejsów. Standardowa infrastruktura uruchomieniowa przekształca wywołanie metody lokalnej w zdalne wywołanie żądania i zbiera wyniki wykonania. Interakcja pomiędzy osobą wywołującą a procesem zdalnym odbywa się za pośrednictwem wiadomości. Model ten jest z założenia bezstanowy, złożoność zarządzania stanem obiektów i czasem życia ilustrują rozproszone modele obiektowe. Architektura Common Object Request Broker Architecture (CORBA), Component Object Model (COM, DCOM i COM+), Java Remote Method Invocation (RMI) i .NET Remoting to tylko niektóre z najważniejszych przykładów infrastruktury obiektów rozproszonych.

Aktywne obiekty

Modele programistyczne oparte na obiektach aktywnych, jakkolwiek dostępnych, z definicji zawierają istnienie instancji, niezależnie od tego, czy są one agentami obiektów. Oznacza to, że obiekty mają specjalny wątek kontrolny, który pozwala im pokazywać swoją aktywność. Takie modele często ręcznie wykorzystują komunikaty do wykonywania funkcji, a komunikat jest powiązany z bardziej skomplikowaną semantyką.

Usługi internetowe

Technologia usług internetowych stanowi alternatywę dla platformy RPC za pośrednictwem protokołu HTTP, umożliwiając interakcję ustalonych komponentów z różnymi technologiami. Usługa internetowa jest udostępniana jako obiekt zdalny przechowywany na serwerze WWW, a wywołania systemu są konwertowane na żądania HTTP spakowane przy użyciu określonego protokołu. Należy pamiętać, że pojęcie komunikatu jest podstawową abstrakcją komunikacji między procesami i jest używane w sposób ukryty lub jawnym.

2.4.4.2 Modele komunikacji opartej na komunikatach

Model komunikatu punkt-punkt

Oprogramowanie lub aplikacja jest projektowana od punktu do punktu (PTP) w oparciu o ideę kolejek wiadomości, nadawców i odbiorców. Wiadomość ta jest wysyłana do określonej kolejki, a klienci otrzymują wiadomości z kolejek skonfigurowanych do przechowywania ich wiadomości. Wszystkie wysłane do nich wiadomości są przechowywane do momentu ich wykorzystania lub wyjaśnienia ważności.

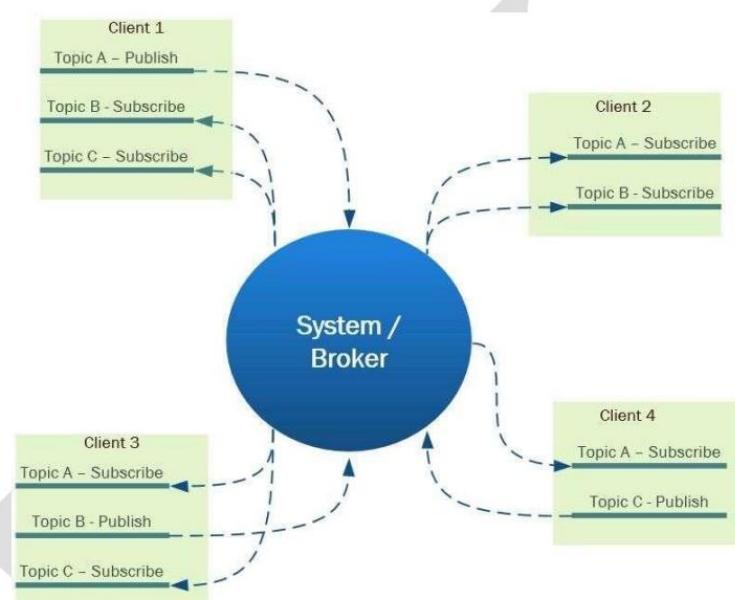
Model wiadomości typu „publikuj i subskrybuj”.

Publikuj-subskrybuj to usługa wiadomości. Opisuje szczególny rodzaj komunikacji pomiędzy komponentami lub modułami oprogramowania. Nazwa została wybrana tak, aby reprezentować najważniejsze cechy tego modelu komunikacji.

Moduły oprogramowania współpracują ze sobą bezpośrednio w prostych interakcjach, korzystając z mechanizmów i mediów rozpoznawanych przez wszystkie strony. Aby potrzeby komunikacyjne stały się bardziej złożone lub wymagające, opracowano inne systemy komunikacji. Publish-subscription to jedna z takich subskrypcji i tylko jedna z wielu.

Podstawowe pomysły na publikację i subskrypcję

- Niekoniecznie istnieją komponenty oprogramowania, które wiedzą, z kim wchodzą w interakcję.
- Producenci danych publikują dane w całej sieci.
- Konsumenti danych subskrybują system i otrzymują z niego dane jako całość.
- Informacje są nazywane w sposób umożliwiający zdefiniowanie dostępnych informacji moduły oprogramowania. Czasami ta etykieta nazywana jest tematem.



RYSUNEK 2.20 Model wiadomości typu „publikuj i subskrybij”.

Centralny moduł oprogramowania zapewnia zarządzanie i dopasowanie wszystkich danych, publikacji i subskrypcji. Powszechnie określa się go mianem „brokera”. Często brokerzy to sieć współpracujących modułów oprogramowania, a moduły oprogramowania korzystające z usług brokerów nazywane są klientami.

Klienci, którzy publikują i subskrybują, „rejestrują się” u brokera w celu zarządzania ścieżkami komunikacyjnymi, klientami i innymi czynnościami porządkowymi w celu uwierzytelnienia.

Dostarczanie wiadomości do subskrybentów filtrowane według treści, a nie tematu. Można tego użyć zamiast lub z tematem. Tylko kilka systemów publikowania i subskrybowania wdrożyło to.

Dane mogą być „trwałe”, ponieważ abonenci, którzy zarejestrują się w sieci po ostatniej publikacji danych, będą mieli do dyspozycji ostatnio opublikowane dane na dany temat.

Model wiadomości żądanie-odpowiedź

Model przesyłania wiadomości na żądanie różni się od tradycyjnego modelu pub/sub lub P2P, który publikuje wiadomość w temacie lub kolejce i umożliwia klientom otrzymanie wiadomości bez podawania odpowiedzi.

Wiadomości z żądaniem odpowiedzi mogą być wykorzystywane, gdy klient wysyła żadaną wiadomość w celu uzyskania informacji ze zdalnej aplikacji klienckiej lub w celu przeprowadzenia czynności przetwarzania. Gdy aplikacja kliencka odbierze komunikat żądania, otrzymuje niezbędne informacje

informacji lub wykonuje żądaną czynność. Informacje te są następnie uwzględniane w wiadomości zwrotnej lub w odpowiedzi na żądanie przesyłane jest potwierdzenie wykonania zadania.

2.5 Technologie przetwarzania rozproszonego

W tej części wdrażamy odpowiednie technologie, które pozwalają na realistyczną realizację modeli interakcji opierających się głównie na komunikacji opartej na komunikatach. Takie systemy obejmują zdalne wywoływanie procedur (RPC), struktury obiektów rozproszonych i przetwarzanie zorientowane na usługi.

2.5.1 Zdalne wywołanie procedury

Zdalne wywołanie procedury (RPC) to protokół, którego program może używać do żądania usługi na innym komputerze w sieci bez konieczności podawania szczegółów sieci.

RPC jest używany w systemach zdalnych do wywoływania innych procesów, takich jak system lokalny. Czasami wywołanie procedury nazywane jest także wywołaniem funkcji lub wywołaniem podprogramu.

Model klient-serwer jest wykorzystywany przez RPC. Program, o który prosisz, to klient, a usługodawca to serwer. Podobnie jak ciągłe lub lokalne wywołanie procedury, RPC jest operacją synchroniczną, która wymaga zawieszenia programu żądającego do czasu zwrócenia wyniku procedury zdalnej. Niemniej jednak wiele wywołań RPC można wykonywać jednocześnie, korzystając z lekkich procesów lub wątków współdzielących tę samą przestrzeń.

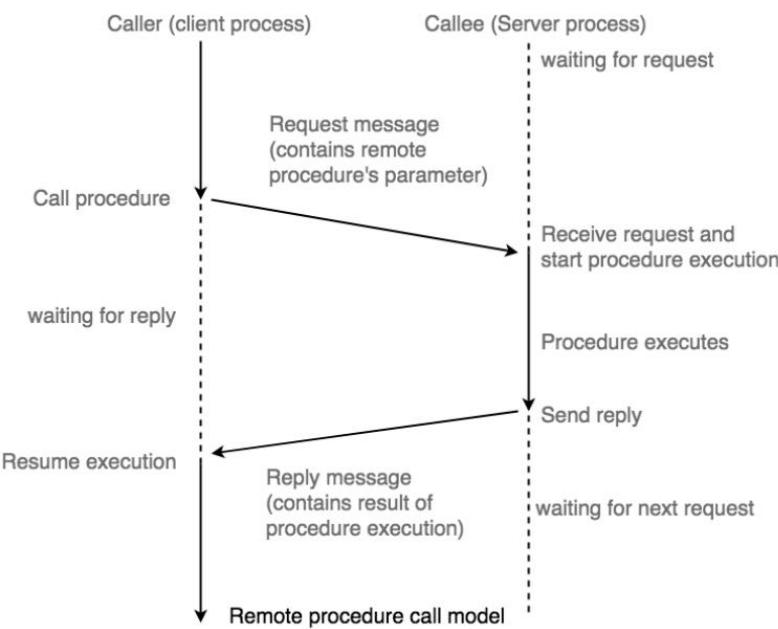
W oprogramowaniu do zdalnego wywoływania procedur: język definicji interfejsu (IDL), język specyfikacji używany do opisu interfejsu programowania aplikacji (API) komponentu oprogramowania. IDL zapewnia w tym przypadku most pomiędzy dwoma końcami połączenia, które mogą być połączone różnymi językami komputerowymi i systemami operacyjnymi (OS).

Procedura komunikatu RPC

Kiedy instrukcje programu korzystające ze struktury RPC są komplikowane do programu wykonywalnego, skompilowany kod zawiera fragment reprezentujący kod procedury zdalnej. Odcinek odbiera żądanie i przesyła je do klienckiego programu wykonawczego na komputerze lokalnym, gdy program jest uruchomiony i następuje wywołanie. Po pierwszym wywołaniu klienta pośredniczącego, kontaktuje się on z serwerem nazw, aby określić, gdzie serwer się znajduje.

Program Client Runtime wie, jak adresować zdalny komputer i aplikację serwera, i wysyła przez sieć komunikat z żądaniem zdalnej procedury. Serwer ma również program wykonawczy i łączy ten interfejs z samą zdalną procedurą. Protokoły żądania odpowiedzi zwracane w ten sam sposób

Podczas wykonywania zdalnego wywołania procedury



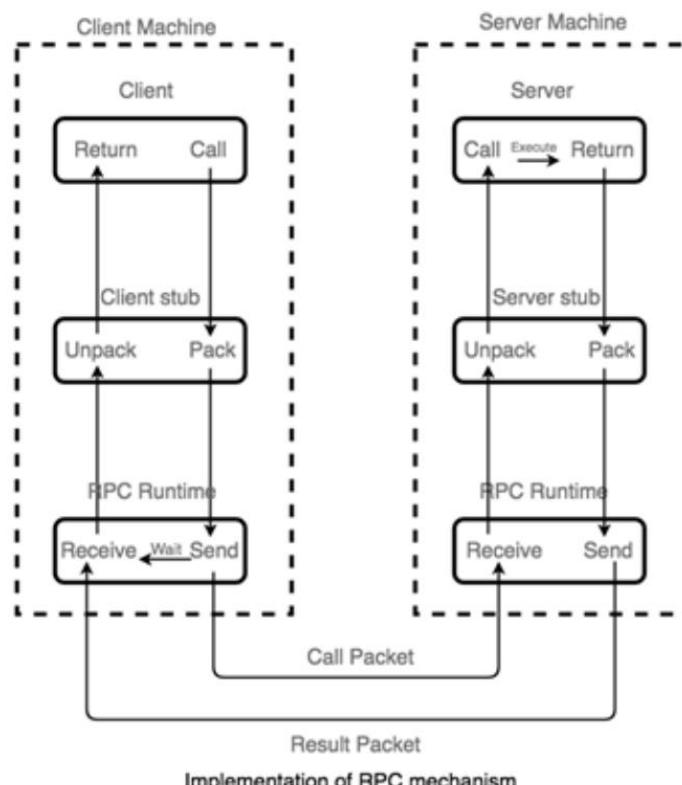
RYSUNEK 2.21 Zdalne wywołanie procedury

1. Środowisko wywołujące zostaje zakończone, parametry proceduralne są przesyłane przez sieć i procedura jest wykonywana w środowisku.

2. Po zakończeniu procedury i wygenerowaniu wyników jej wyniki są zwracane do środowiska wywołującego, gdzie wznowiane jest wykonywanie, tak jak w przypadku zwykłego wywołania procedury.

UWAGA: RPC jest szczególnie odpowiedni do interakcji klient-serwer (np. zapytanie-odpowiedź) pomiędzy osobą wywołującą a odbiorcą. W koncepcji klient i serwer nie są uruchamiane jednocześnie. Zamiast tego przeskakuje tam i z powrotem od dzwoniącego do wywoływanego.

Działanie RPC



RYSUNEK 2.22 Działanie zdalnego wywołania procedury

W RPC będą następujące kroki:

1. Klient wywołuje procedurę pośredniczącą klienta, która zwykle przekazuje parametry. Klienten znajduje się we własnym obszarze adresowym klienta.
2. Klient stubbuje (pakuje) parametry w wiadomości. Marshalling polega na konwersji reprezentacji parametrów na standardowy format i skopiowaniu każdego parametru do komunikatu.
3. Stub klienta przesyła wiadomość do warstwy transportowej i przesyła ją do warstwy transportowej zdalny serwer.
4. Na serwera warstwa transportowa przesyła wiadomość do odcinka pośredniczącego serwera w celu rozmieszczenia (rozpakowania) parametrów i wykorzystuje standardową metodę wywołania procedury do wywołania żądanej procedury procesu.
5. Po zakończeniu procedury serwera wraca ona do serwera pośredniczącego (na przykład poprzez powrót normalnego wywołania procedury). Następnie serwer pośredniczy przesyła wiadomość do warstwy transportowej.
6. Warstwa transportowa zwraca wynikowy komunikat do warstwy transportowej klienta, która zwraca komunikat do odcinka pośredniczącego klienta.
7. Klient zatrzymuje zwracane parametry i zwraca wykonanie wywołującemu.

2.5.2 Rozproszone struktury obiektowe

Najbardziej znanymi sposobami tworzenia systemów lub frameworków rozproszonych są systemy klient-serwer. Rozszerzeniem tego modelu klient-serwer jest struktura obiektów rozproszonych. Jest to biblioteka, w której można tworzyć aplikacje rozproszone przy użyciu programowania obiektowego. Obiekty rozproszone w przetwarzaniu rozproszonym to obiekty rozproszone w różnych przestrzeniach adresowych, w różnych procesach na tym samym komputerze lub nawet na wielu komputerach podłączonych do sieci. Jednakże działają wokół siebie poprzez udostępnianie danych i metody wywoływania. Obejmuje to również przezroczystość lokalizacji, w której odległe obiekty wyglądają tak samo jak obiekty lokalne. W przypadku zdalnego wywoływania metod, zwykle przekazywania komunikatów, główną metodą rozproszonej komunikacji obiektów jest wysłanie komunikatu do innego obiektu na zdalnej maszynie lub procesie w celu wykonania jakiegoś zadania. Wyniki są zwracane do wywoływanego obiektu.

Metoda zdalnego wywoływania procedur (RPC) stosuje się do środowisk rozproszonych, typową abstrakcję programistyczną wywołania procedury, umożliwiając procesowi wywołującemu wywoływanie zdalnego węzła jako lokalnego.

Zdalne wywołanie metody (RMI) przypomina RPC dla obiektów rozproszonych, ale ma dodatkowe korzyści w zakresie stosowania koncepcji programowania obiektowego dla systemów rozproszonych oraz rozszerza na rozproszone środowisko globalne koncepcję odniesienia do obiektu i umożliwia wykorzystanie odwołań do obiektów, takich jak Parametr, w zdalnym wywołaniu.

Zdalne wywołanie procedury — klient wywołuje procedury w innym programie serwera
Zdalne wywołanie metody (RMI) — obiekt może wywołać metody obiektowe w innym procesie

Powiadomienie o zdarzeniu — Obiekty otrzymują powiadomienia o zdarzeniach w innych obiektach, w których się zarejestrowały

2.5.2.1 Przykłady rozproszonych frameworków obiektowych

Oprogramowanie rozproszonego środowiska programowania (DPE) może być rozwijane i zarządzane przez programistów rozproszonych po całym świecie. Praktycznie wspiera rozproszone przetwarzanie obiektowe, takie jak Internet, w systemie rozproszonym. Badania mają na celu opracowanie środowiska programistycznego wspierającego efektywne programowanie w środowisku rozproszonym. System korzystający z obiektów rozproszonych zapewnia rozproszonemu i równoległe środowisku programowania elastyczne i skalowalne programowanie. Wiele z nich to obsługiwane systemy przetwarzania obiektów rozproszonych, takie jak CORBA, DCOM i Java

Architektura brokera żądań obiektów wspólnych (CORBA)

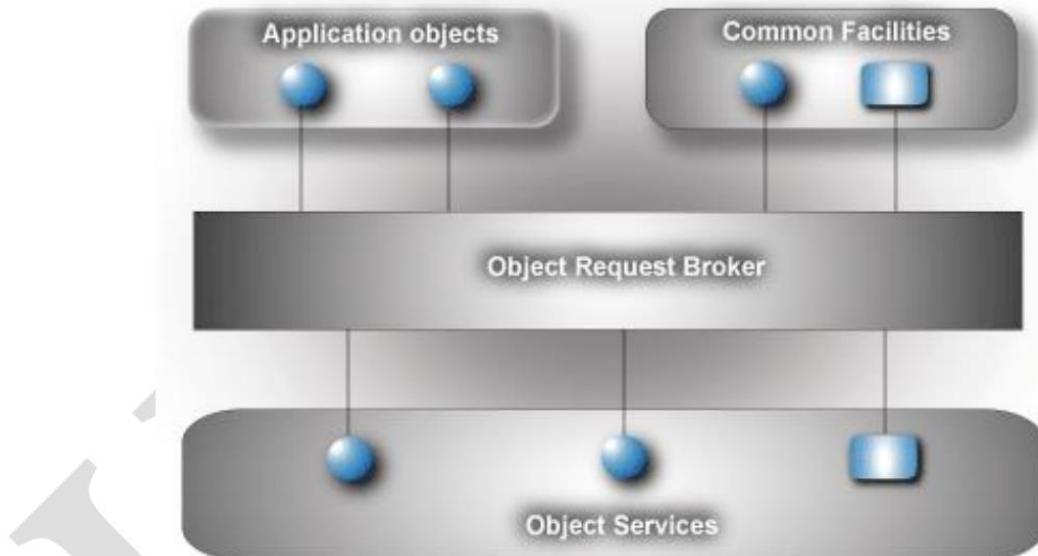
Common Object Request Broker Architecture (CORBA) to konsorcjum ponad 800 firm obsługujące najsłynniejsze oprogramowanie pośredniczące. Z wyjątkiem Microsoftu, w tym konsorcjum znajduje się większość firm komputerowych, które posiadają własne rozwiązania rozproszone

Broker obiektów Component Object Model (DCOM). Magistrala obiektowa CORBA ustawa i definiuje interoperacyjność komponentów obiektu. Magistrala obiektowa to broker żądań obiektów (ORB). CORBA ma na celu wykrywanie i współdziałanie komponentów obiektów w magistrali obiektowej. CORBA obsługuje przezroczyste odniesienia do obiektów poprzez interfejsy obiektowe pomiędzy rozproszonymi obiekty.

CORBA to zasadniczo specyfikacja projektowa Brokeru żądań obiektów (ORB), który zapewnia mechanizm ORB umożliwiający komunikację między rozproszonymi obiektami, lokalnie lub na urządzeniach zdalnych, napisanych w różnych językach lub w różnych lokalizacjach sieciowych.

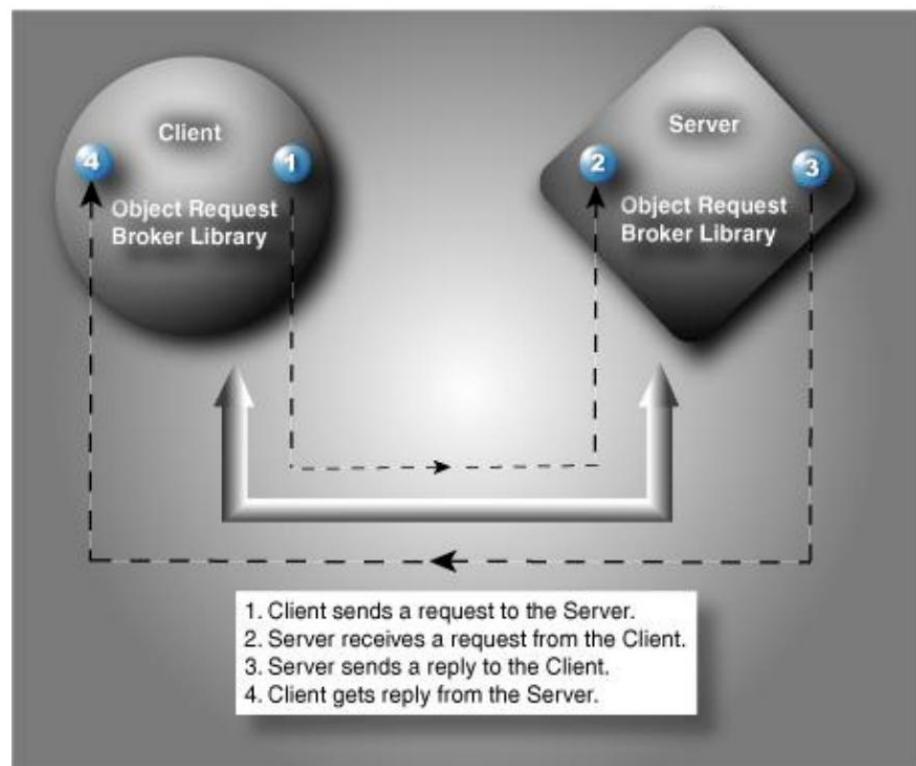
Język definicji interfejsu CORBA, w skrócie IDL, umożliwia rozwój języka, rozwój interfejsu niezależnego od lokalizacji i dystrybucję rozproszonych obiektów. Komponenty aplikacji mogą komunikować się ze sobą poprzez CORBA, niezależnie od tego gdzie się znajdują i kto je zaprojektował. CORBA zapewnia przejrzystość lokalizacji w celu realizacji tych żądań.

CORBA jest zwykle opisywana jako „magistrala oprogramowania”, ponieważ lokalizowanie obiektów i dostęp do nich odbywa się poprzez interfejs komunikacyjny oprogramowania. Poniższa ilustracja przedstawia główne elementy implementacji CORBA.



RYSUNEK 2.23 Architektura brokeru żądań obiektów wspólnych (CORBA)

Dobrze zdefiniowany interfejs obiektowy zapewnia transmisję danych od klienta do serwera. Broker żądań obiektów (ORB) ustala lokalizację obiektu docelowego, wysyła do niego żądanie i zwraca obiekowi wywołującemu dowolną odpowiedź. Dzięki tej technologii zorientowanej obiekto programiści mogą korzystać w czasie wykonywania z takich cech, jak starsze rozwiązania, enkapsulacja, polimorfizm i wiązanie dynamiczne. Funkcje te pozwalają na modyfikację, modyfikację i ponowne wykorzystanie aplikacji przy minimalnych zmianach interfejsu nadzorowanego. Poniższa ilustracja pokazuje, jak klient przesyła żądanie za pośrednictwem ORB do serwera:



RYSUNEK 2.24 działanie architektury brokera żądań obiektów Common (CORBA)

Język definicji interfejsu (IDL)

Język definicji interfejsu jest kluczowym filarem standardów CORBA. IDL to OMG w definicji interfejsów API neutralnych językowo i zapewnia niezależny od platformy zestaw rozproszonych interfejsów obiektowych. Dane i operacje ustalane pod kątem interfejsu klient-serwer zaczynają zapewniać spójne podejście między środowiskami CORBA i klientami w środowiskach heterogenicznych. Mechanizm ten nazywa się IDL i jest używany przez CORBA do opisu interfejsów obiektowych.

W przypadku aplikacji IDL definiuje język programowania jako moduły, interfejsy i operacje, a nie je. Różne języki programowania, takie jak Ada, C++, C# i Java, zapewniają standardowe mapowanie IDL w implementacji interfejsu.

Kompilator IDL tworzy kod pośredniczący i szkieletowy do organizowania i unmarshalowania parametrów ze strumienia sieciowego do instancji pamięci zaimplementowanego języka itp. Stub jest klienckim proxy dla odniesienia do obiektu, do którego uzyskuje się dostęp od serwera i jest proxy dla klienta służącym. Specyficzne dla języka kody pośredniczące i szkielety IDL mogą komunikować się ze szkieletem w danym języku. Kod pośredniczący jest powiązany z kodem klienta, a kod szkieletowy jest połączony z implementacją obiektu i komunikuje się w celu realizacji zdalnych operacji z systemem wykonawczym ORB.

IIOP (Internet Inter-ORB Protocol) to protokół umożliwiający rozproszonym programom komunikację w Internecie w różnych językach programowania. Architektura Common Object Request Broker Architecture (CORBA) jest kluczowym elementem strategicznego standardu branżowego. Korzystając z CORBA IIOP i powiązanych procedur, firma może opracować programy, które będą w stanie komunikować się, gdziekolwiek się znajdują i bez konieczności rozumienia czegokolwiek na temat programu poza jego własną usługą lub jego nazwą, z istniejącymi lub przyszłymi programami własnej firmy lub innymi.

Rozproszony model obiektowy komponentów (DCOM/COM)

Model obiektowy komponentów rozproszonych (DCOM) to zastrzeżona technologia komunikacji firmy Microsoft między składnikami oprogramowania rozproszonymi po komputerach w sieci. DCOM to rozproszony model obiektowy komponentów. Model obiektowy komponentów rozproszonych to technologia rozbudowy sieci oparta na modelu obiektowym komponentów (COM), która umożliwia komunikację między procesami w całej sieci. Zarządzając szczegółami protokołu sieciowego niskiego poziomu, DCOM wspiera komunikację pomiędzy obiektami w sieci. Umożliwia to współpracę wielu procesów w celu osiągnięcia jednego zadania przy użyciu programów rozproszonych.

Zdalne wywołanie metody Java (RMI)

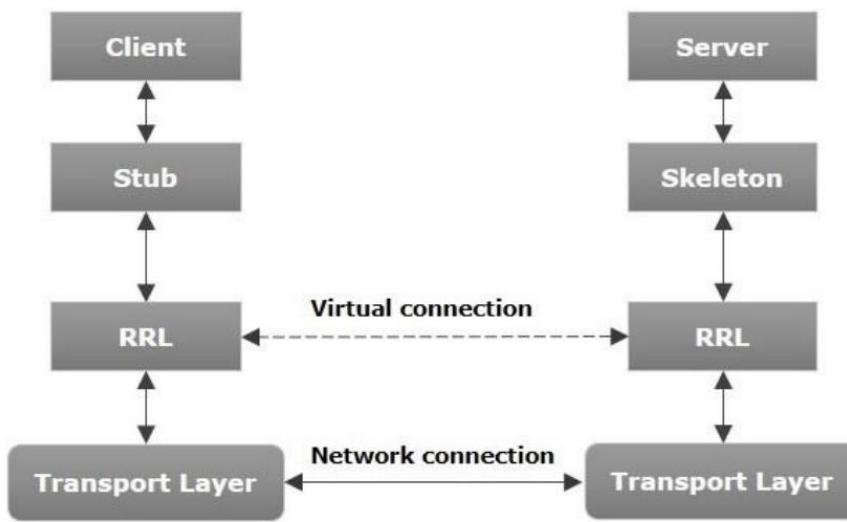
RMI oznacza zdalne wywołanie metody. Mechanizm umożliwiający dostęp/wywołanie obiektu w jednym programie (JVM) na innej maszynie JVM. Umożliwia zdальną komunikację pomiędzy programami w języku Java, RMI służy do tworzenia aplikacji rozproszonych.

W aplikacji RMI tworzymy dwa programy: program serwera (rezydujący na serwerze) i program kliencki (rezydujący na kliencie).

Program serwera tworzy obiekt zdalny i udostępnia klientowi referencję do tego obiektu (za pomocą rejestracji).

Program kliencki żąda obiektów zdalnych i próbuje wywołać swoje metody na serwerze.

Poniższy diagram przedstawia architekturę aplikacji RMI.



RYSUNEK 2.25 Zdalne wywołanie metody Java (RMI)

Warstwa transportowa – za pomocą tej warstwy klient łączy się z serwerem. Ten połączanie jest utrzymywane z istniejącym połączeniem i tworzone są także nowe połączenia.

Odgałęzienie — odgałęzienie jest serwerem proxy zdalnego obiektu klienta. Znajduje się ono w systemie klienta; służy jako brama klienta.

Szkielet — jest to obiekt po stronie serwera. Aby przekazać żądanie do pilota obiektu, Stub wchodzi w interakcję ze szkieletem.

RRL (Remote Reference Layer) — jest to warstwa zarządzająca klientem odniesienie do obiektu zdalnego.

Poniższe punkty podsumowują działanie programu RMI.

Za każdym razem, gdy klient wysyła żądanie do obiektu zdalnego, kod pośredniczy odbiera żądanie do RRL.

Jeśli RRL od klienta otrzyma żądanie, używa metody zwanej invoke () z obiektu RemoteRef.

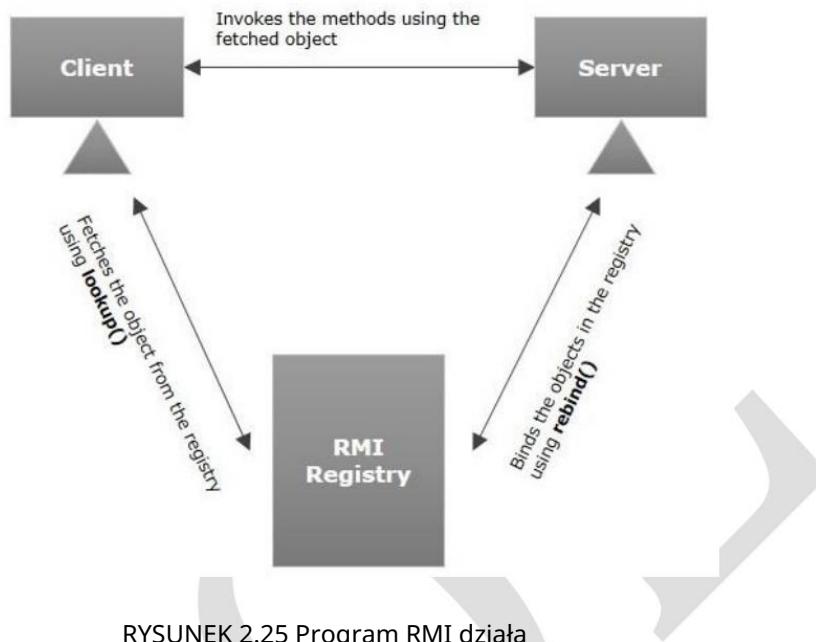
Żądanie jest przekazywane po stronie serwera do RRL. RRL serwera przekazuje klienta do szkieletu serwera, który ostatecznie wywołuje obiekt na serwerze. Wyniki przekazywane są klientowi

Gdy klient wywołuje metodę obsługującą parametry obiektu zdalnego, parametry te powinny zostać dołączone do komunikatu przed przesaniem ich przez sieć. Mogą to być parametry lub obiekty typu pierwotnego. Jeśli używany jest typ pierwotny, parametry są składane i dołączany jest nagłówk. Jeśli parametr jest obiektem, jest serializowany. Metodę tę nazywa się marshallingiem.

Spakowane parametry są rozdzielane po stronie serwera i wywoływana jest odpowiednia metoda. Metodę tę nazywa się unmarshallingiem.

Rejestr RMI to przestrzeń nazw zawierająca wszystkie obiekty serwera. Serwer rejestruje ten obiekt w rejestrze RMI (przy użyciu metod bind () lub Rebind ()) (metod) za każdym razem, gdy obiekt jest tworzony. Są one rejestrowane przy użyciu jednej nazwy znanej jako nazwa powiązania.

Klient wymaga odniesienia do tego obiektu, aby wywołać obiekt zdalny. Klient musi następnie pobrać obiekt z rejestrów według jego nazwy powiązania (przy użyciu metody wyszukiwania ()) .



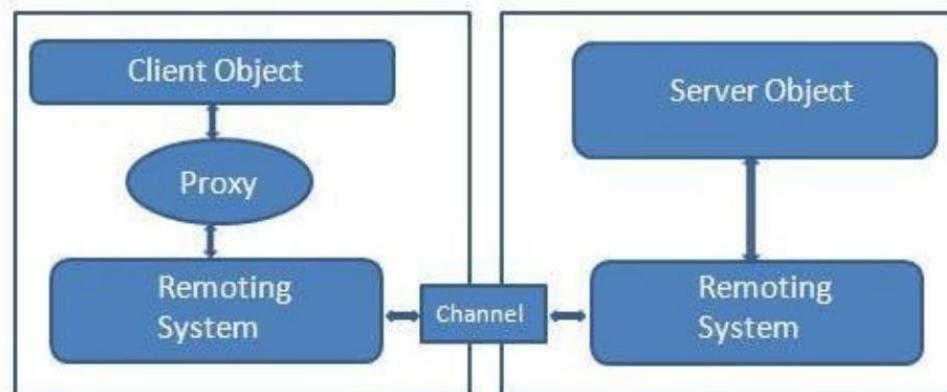
RYSUNEK 2.25 Program RMI działa

Komunikacja zdalna .NET

The. Zdalny system NET oferuje komunikację międzyprocesową pomiędzy domenami aplikacji poprzez wykorzystanie Remoting Framework. Programy mogą być instalowane na tym samym komputerze lub na różnych komputerach w tej samej sieci. Dzięki zastosowaniu formaterów binarnych lub SOAP w strumieniu danych .NET Remoting ułatwia komunikację obiektów rozproszonych za pośrednictwem kanałów TCP i HTTP.

Trzy główne komponenty Remoting Framework to:

1. Obiekt zdalny 2.
- Aplikacja zdalnego nasłuchiwanego – (żądania obiektu zdalnego)
3. Zdalna aplikacja kliencka-(wysyła żądania obiektów zdalnych)



RYSUNEK 2.26 Struktura komunikacji zdalnej .NET

Obiekt zdalny jest zaimplementowany w klasie MarshalByRefObject.

Podstawowy przebieg pracy w .Net Remoting można zobaczyć na powyższym rysunku. Ponadto, jeśli klient wywołuje metodę Remote, nie wywołuje bezpośrednio metod. Obiekt zdalny otrzymuje proxy i służy do wywołania metody obiektu zdalnego. Wiadomość jest szyfrowana za pomocą odpowiedniego narzędzia formatującego (formatującego binarnie lub formatującego SOAP) w pliku konfiguracyjnym, gdy serwer proxy odbierze wywołanie procesowe z serwera, a następnie wywołanie zostanie wysłane do serwera przy użyciu wybranego kanału (TcpChannel lub HttpChannel). Kanał po stronie serwera przyjmuje żądanie od proxy i wysyła je do serwera w systemie Remoting, gdzie znajdują się metody obiektu zdalnego i są wywoływane metody obiektu zdalnego. Po wykonaniu procedury zdalnej wynik każdego połączenia jest zwracany klientowi w ten sam sposób. Musi zostać wygenerowany i zainicjowany w procesie znany jako Aktywacja, zanim będzie można uzyskać dostęp do instancji obiektu typu Remotable. Aktywacja jest klasyfikowana jako obiekty aktywowane przez klienta i obiekty aktywowane przez serwer w dwóch typach.

2.5.3.2 Architektura zorientowana na usługi (SOA)

Architektura zorientowana na usługi (SOA) to styl oprogramowania, w którym usługi za pośrednictwem połączonego protokołu komunikacyjnego z innymi komponentami są dystrybuowane przez komponenty aplikacji. Zasady są oddzielone od producentów i innych. Większość usług komunikuje się między sobą w architekturze zorientowanej na usługi: poprzez transmisję danych lub za pośrednictwem dwóch lub więcej usług koordynujących działanie. To tylko jedno określenie architektury usług.

Charakterystyka architektury zorientowanej na usługi

Wartość biznesowa

Cele strategiczne

Wewnętrzna interoperacyjność

Usługi wspólne

Elastyczność

Ewolucyjne udoskonalenie

Obie te podstawowe zasady można przedstawić poprzez starszy paradygmat aplikacji rozproszonych, do architektury zorientowanej na usługi i związanej z chmurą (która również jest uważana za odgałęzienie architektury zorientowanej na usługi).

Wzorce architektury zorientowanej na usługi



RYSUNEK 2.27 Architektura zorientowana na usługi

Każdy z elementów Architektury Zorientowanej na Usługi składa się z trzech ролей: dostawcy usług; broker usług, rejestr usług, repozytorium usług i obsługa klienta/zleceniodawcy.

Zgodnie z rejestrem usług usługodawca jest odpowiedzialny za określenie, czy i w jaki sposób usługi są świadczone, takie jak bezpieczeństwo, dostępność, koszty i inne.

Rola ta decyduje również o rodzaju usługi i ewentualnych umowach handlowych.

Broker usług udostępnia żądającemu informacje o usłudze. Determinowany jest ten, kto wdraża zakres brokera.

Obiekt żądający usługi lokalizuje, a następnie dodaje wpisy do rejestru brokerów. Możesz uzyskać dostęp do wielu usług lub nie; zależy to od możliwości wnioskodawcy.

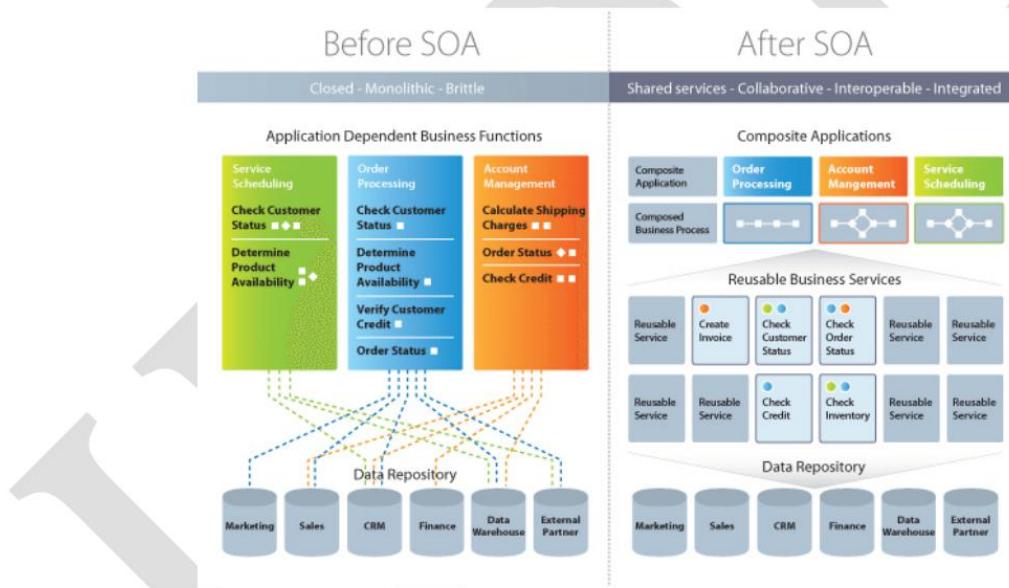
Wdrażanie architektury zorientowanej na usługi

Istnieje wiele różnych technologii, które można zastosować przy wdrażaniu architektury zorientowanej na usługi (SOA), w zależności od ostatecznego celu i tego, co próbujesz osiągnąć.

Architektura zorientowana na usługi jest zwykle wdrażana za pomocą usług sieciowych, które udostępniają „funkcjonalne elementy składowe za pośrednictwem standardowych protokołów internetowych”. SOAP, czyli Simple Object Access Protocol, jest przykładem standardu usług internetowych. Krótko mówiąc, SOAP to specyfikacja protokołu przesyłania wiadomości służąca do standaryzowanego udostępniania informacji w implementacji usług sieciowych w sieci komputerowej. Chociaż SOAP początkowo nie został dobrze przyjęty, od 2003 roku zyskał na popularności i jest szerzej stosowany i akceptowany. Jini, COBRA lub REST to inne opcje wdrożenia architektury zorientowanej na usługi.

Należy pamiętać, że architektury można stosować na różne sposoby, w tym do przesyłania wiadomości, takie jak ActiveMQ, Apache Thrift i SORCER, „niezależnie od konkretnych technologii”.

Dlaczego architektura zorientowana na usługi jest ważna



RYSUNEK 2.28 Przed i po architekturze zorientowanej na usługi

Architektura zorientowana na usługi ma wiele zalet, szczególnie w biznesie internetowym.

Tutaj szybko omówimy niektóre z tych zalet: Aby utworzyć kod nadający się do ponownego użycia, użyj architektury zorientowanej na usługi: jest to nie tylko czasochłonne, ale także nie jest konieczne wymyślanie koła kodowania za każdym razem, gdy potrzebna jest nowa usługa lub proces. SOA pozwala również na użycie tych języków kodowania, ponieważ wszystko działa poprzez centralny interfejs.

Korzystanie z architektury zorientowanej na usługi w celu ułatwienia interakcji: W architekturze zorientowanej na usługi generowany jest wspólny sposób komunikacji, który umożliwia niezależne działanie różnych systemów i platform. Dzięki temu połączeniu architektura zorientowana na usługi może również obejść zapory sieciowe, które umożliwiają „firmom współdzielenie usług ważnych operacyjnie”.

Korzystanie ze skalowalności Architektury Zorientowanej na Usługi: istotna jest możliwość skalowania biznesu do wymagań klienta, jednak można zapobiec pewnym zależnościom. Zastosowanie architektury zorientowanej na usługi ogranicza interakcję między klientami, co ułatwia skalowanie.

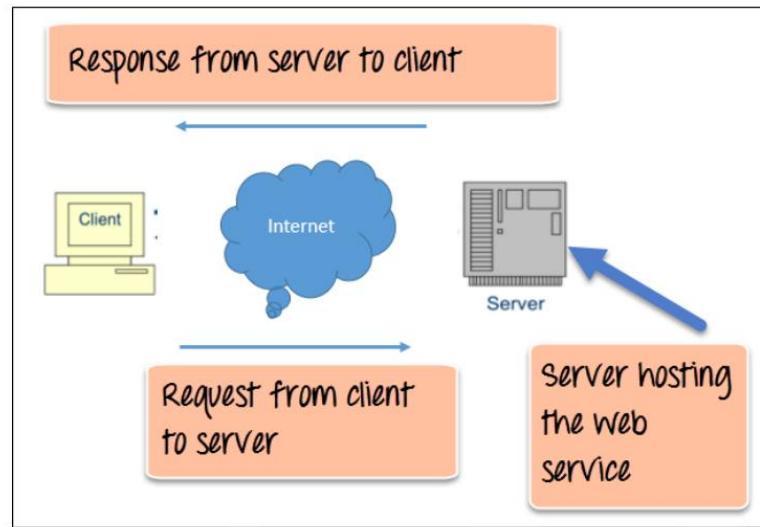
Korzystanie z architektury zorientowanej na usługi w celu zmniejszenia kosztów: dzięki architekturze zorientowanej na usługi możliwe jest zmniejszenie kosztów przy jednoczesnym „utrzymaniu pożąданiej wydajności”. Firmy mogą ograniczyć ilość analiz potrzebnych do tworzenia niestandardowych rozwiązań przy użyciu architektury zorientowanej na usługi.

2.5.3.3 Usługi sieciowe

Usługa sieciowa to ustrukturyzowana metoda dystrybucji komunikacji klient-serwer w sieci WWW. Usługa internetowa to moduł oprogramowania, który wykonuje różnorodne zadania.

Można wyszukiwać usługi sieciowe w sieci i odpowiednio je wywoływać.

Usługa internetowa po wywołaniu zapewni klientowi funkcje, które wywołuje ta usługa internetowa.



RYSUNEK 2.29 Diagram architektury usług sieciowych

Powyższy diagram daje bardzo jasny obraz wewnętrznego działania usługi internetowej. Klient na żądanie wykona serię wywołań usług internetowych do serwera w celu hostowania bieżącej usługi internetowej.

Aplikacje te są renderowane poprzez tak zwane zdalne wywoływanie procedur.

Zdalne wywołanie procedury (RPC) to wywoływanie wykonywane przy użyciu procedur usługi hostingu usług internetowych. Amazon świadczy usługę internetową dla produktów sprzedawanych online, na przykład za pośrednictwem witryny amazon.com. Frontend i warstwa prezentacji mogą być w Net lub Java, ale usługa internetowa będzie współdziałać w dowolnym języku programowania.

Dane przesypane pomiędzy klientem a serwerem są podstawowym składnikiem usługi sieciowej, czyli XML. XML jest odpowiednikiem HTML i językiem pośrednim, który wiele języków programowania jest łatwy do zrozumienia i który porozumiewa się w formacie XML tylko wtedy, gdy aplikacje komunikują się ze sobą. Zapewnia to interfejs aplikacji can do wzajemnej interakcji w różnych językach programowania. Usługi internetowe korzystają z protokołu SOAP (Simple Object Access Protocol) do przesyłania danych XML pomiędzy aplikacjami. Dane przesypane są poprzez standardowy protokół HTTP. Dane przesypane do programu z serwera WWW nazywane są SOAP. Wiadomość z SOAP to po prostu XML. Aplikacja kliencka wywołująca usługę sieciową może zostać napisana w dowolnym języku programowania, ponieważ ten dokument jest napisany w formacie XML.

Dlaczego potrzebujesz usługi internetowej?

Systemy oprogramowania na co dzień korzystają z szerokiej gamy internetowych narzędzi programistycznych. Można zbudować kilka aplikacji w Javie, inne w Net, inne w Angular JS, Node.js itp. Te

aplikacje heterogeniczne najczęściej wymagają pewnego rodzaju komunikacji między nimi.

Ponieważ są one zbudowane w różnych językach programowania, bardzo trudno jest zapewnić efektywną komunikację pomiędzy aplikacjami.

Tutaj oferowane są usługi internetowe. Usługi internetowe zapewniają wspólną platformę, która umożliwia komunikację między wieloma aplikacjami opartymi na różnych językach programowania.

Rodzaj usługi internetowej

Dostępne są głównie dwa rodzaje usług internetowych.

1. Usługi sieciowe SOAP.
2. RESTful usługi sieciowe.

Aby usługa internetowa była w pełni funkcjonalna, należy wdrożyć pewne elementy. Niezależnie od tego, jaki język programowania jest używany do programowania usługi internetowej, komponenty te muszą być obecne.

Przyjrzyjmy się bliżej tym elementom

SOAP jest uważany za niezależny protokół komunikatów do transportu. SOAP opiera się na przesyłaniu danych XML za pomocą komunikatów SOAP. Każda wiadomość ma dokument zwany dokumentem XML. Tylko struktura dokumentu XML jest zgodna z pewnym wzorcem, ale treść nie jest zgodna z nim. Najlepszym składnikiem usług internetowych i protokołu SOAP jest to, że są one dostarczane za pośrednictwem protokołu HTTP, standardowego protokołu sieciowego.

To jest przesłanie SOAP

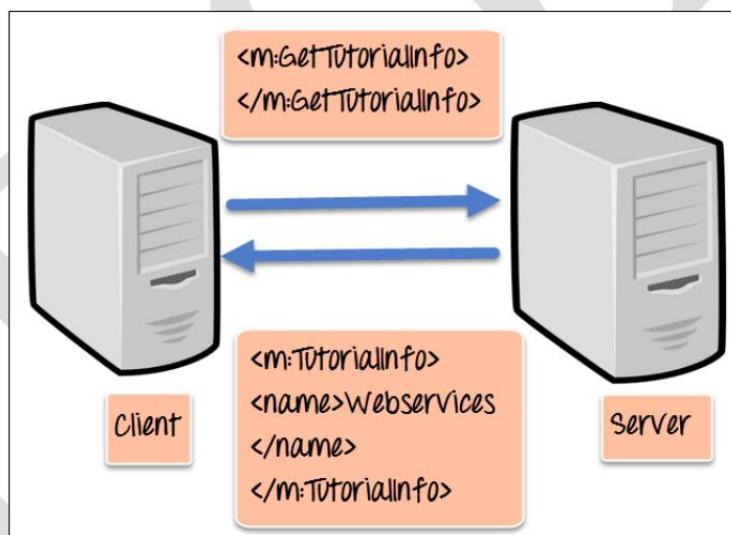
W każdym dokumencie SOAP potrzebny jest element główny zwany <Envelope>. Pierwszym elementem dokumentu XML jest element główny.

Koperta podzielona jest kolejno na 2 części. Pierwszy to nagłówek, drugi to treść.

Nagłówek zawiera dane routingowe, czyli informację, do której powinien zostać wysłany dokument XML.

Prawdziwy przekaz znajduje się w treści.

Prosty przykład komunikacji poprzez SOAP przedstawiono na poniższym schemacie.



RYSUNEK 2.30 WSDL (język opisu usług sieciowych)

Jeśli zostanie znaleziony, usługa internetowa nie będzie używana. Klient wywołujący usługę internetową powinien znać lokalizację usługi internetowej.

Po drugie, aplikacja kliencka chce dowiedzieć się, co robi usługa internetowa, aby wywołać właściwą usługę internetową. Osiąga się to za pomocą języka WSDL, znanego jako język opisu usług sieciowych. Plik WSDL to kolejny plik XML, który głównie informuje usługę internetową, co robi jej aplikacja kliencka. Aplikacje klienckie będą rozumieć lokalizację i sposób korzystania z usług sieciowych przy użyciu dokumentu WSDL.

Przykład usługi internetowej

Poniżej podano przykład pliku WSDL.

```
<definicje>
<message name="TutorialRequest">
  <part name="TutorialID" type="xsd:string"/>
</wiadomość>

<message name="TutorialResponse">
  <part name="Nazwa samouczka" type="xsd:string"/>
```

```

</wiadomość>

<portType name="Tutorial_PortType">
    <nazwa operacji="Samouczek">
        <input message="tns:TutorialRequest"/>
        <output message="tns:TutorialResponse"/>
    </operacja>
</portType>

<binding name="Tutorial_Binding" type="tns:Tutorial_PortType">
    <soap:binding style="rpc"
        transport="http://schemas.xmlsoap.org/soap/http"/>
    <nazwa operacji="Samouczek">
        <soap:operacja mydłoAction="Samouczek"/>
        <wejście>
            <mydło:ciało
                encodingStyle="http://schemas.xmlsoap.org/soap/encoding/"
                namespace="urna:przykłady:usługa samouczka"
                użyj="zakodowany"/>
        </wejście>
        <wyjście>
            <mydło:ciało
                encodingStyle="http://schemas.xmlsoap.org/soap/encoding/"
                namespace="urna:przykłady:usługa samouczka"
                użyj="zakodowany"/>
        </wyjście>
    </operacja>
    </wiązanie>
</definicje>

```

Główne aspekty powyższej deklaracji WSDL są następujące;

< komunikat > – Parametr opisu komunikatu WSDL używany jest do opisu różnych elementów danych w ramach operacji usługi internetowej. W tym przykładzie istnieją dwie wiadomości, jedna to „TutorialRequest”, a druga to operacja „TutorialResponse”, które mogą być wymieniane pomiędzy usługą internetową a aplikacją kliencką. TutorialRequest zawiera element w postaci ciągu „TutorialID”. Podobnie element o nazwie „TutorialName”, będący jednocześnie ciągiem formularza, znajduje się w TutorialResponse.

< portType > — w rzeczywistości definiuje działanie usługi sieciowej, zwane w naszym przypadku samouczkiem. Ta procedura spowoduje otrzymanie 2 komunikatów, jeden na wejściu, drugi na wyjściu.

< wiązanie > – Używany protokół zawiera ten element. I opisujemy to w naszym przypadku, używając http (http://schemas.xmlsoap.org/soap/http). Określono dodatkowe szczegóły dotyczące treści operacji, w tym przestrzeń nazw i kodowanie wiadomości.

Uniwersalny opis, wykrywanie i integracja (UDDI)

UDDI to standard, w którym opisywane, publikowane i odkrywane są usługi sieciowe oferowane przez konkretnego dostawcę. Zawiera specyfikację hostingu treści usług internetowych.

W poprzednim temacie omówiliśmy WSDL i sposób, w jaki dostarcza on szczegółów na temat rzeczywistych działań usługi internetowej. Jednak w jaki sposób aplikacja kliencka może uwzględnić plik WSDL w celu rozpoznawania różnych operacji internetowych. UDDI zapewnia rozwiązanie i serwer, który może hostować pliki WSDL. Oznacza to, że aplikacja kliencka ma pełny dostęp do UDDI, czyli bazy danych zawierającej wszystkie pliki WSDL.

Tak jak książka telefoniczna zawiera nazwisko, adres i numer telefonu określonej osoby, tak rejestr UDDI zawiera powiązane informacje o usługach internetowych. Dlatego użytkownik-programista wie, gdzie go znaleźć.

Teraz już rozumiemy, po co powstały usługi sieciowe, które miały zapewnić platformę do wzajemnej komunikacji za pomocą różnych aplikacji.

Porozmawiajmy jednak o innych zaletach, które uzasadnią znaczenie usług sieciowych.

Udostępnianie funkcjonalności biznesowej w sieci — serwer WWW to jednostka zarządzanego kodu, która oferuje aplikacjom klienckim lub użytkownikom końcowym pewien rodzaj funkcjonalności. Protokół HTTP umożliwia wywołanie tej funkcjonalności, dzięki czemu można ją wywołać również przez Internet. Obydwa programy są już dostępne w Internecie, co zwiększa użyteczność serwisów internetowych. Zapewnia, że usługa internetowa będzie dostępna w dowolnym miejscu sieci WWW i będzie w stanie zapewnić wymaganą funkcjonalność.

Interoperacyjność pomiędzy aplikacjami – usługami internetowymi umożliwia różnym aplikacjom komunikację ze sobą oraz współdzielenie danych i usług. Można ze sobą rozmawiać na każdy temat. Możesz teraz pisać ogólny kod, który będzie zrozumiały dla wszystkich aplikacji, zamiast pisać konkretny kod, który zrozumieją tylko określone aplikacje.

Standardowy protokół, który każdy rozumie — usługi sieciowe korzystają ze standardowego protokołu branżowego do komunikacji, który jest zrozumiały dla każdego. Wszystkie cztery warstwy (usługa transportowa, przesyłanie wiadomości XML, opis usługi i warstwy wykrywania usług) korzystają z dobrze zdefiniowanych protokołów stosu sieciowego usług sieciowych.

Obniżenie kosztów komunikacji — dostawcy usług internetowych korzystają z protokołu SOAP poprzez protokół HTTP, aby wdrażać swoje usługi internetowe przy użyciu istniejącego, taniego Internetu.

2.5.3.4 Orientacja na usługi i przetwarzanie w chmurze

Orientacja na usługi to wbudowane podejście architektoniczne, które wykorzystuje zautomatyzowane zasoby oprogramowania do włączania procesów biznesowych. Takie usługi biznesowe składają się ze zbioru luźno powiązanych komponentów zaprojektowanych w celu zmniejszenia zależności i zaprojektowanych w celu wspierania dobrze określonej funkcji biznesowej. Tworzenie modułowych systemów usług biznesowych przyczynia się do tworzenia bardziej wszechstronnych i efektywnych systemów informatycznych.

Systemy zaprojektowane z myślą o integracji orientacji na usługi pozwalają przedsiębiorstwom wykorzystać istniejące zasoby i łatwo zarządzać nieuniknionymi zmianami, których doświadcza dynamiczna firma. Istnieją również okoliczności, w których konieczne jest połączenie kilku usług. Oznacza to, że te połączone obciążenia będą działać z mniejszymi opóźnieniami niż w przypadku luźno powiązanych części.

Środowiska chmury hybrydowej stają się ważne, ponieważ organizacje nieustannie odkrywają siebie na nowo i stają się bardziej konkurencyjne, aby reagować na zmiany. Dział IT musi znajdować się na czele strategii biznesowej opartej na innowacjach i transformacji. Organizacje rozumieją, że dla wszystkich rodzajów obciążzeń trudno jest znaleźć jedno najlepsze podejście do obliczeń IT. Tym samym najbardziej realistycznym rozwiązaniem jest system chmury hybrydowej.

Wysoki stopień elastyczności i modułowości, aby infrastruktura chmurowa działała w świecie rzeczywistym. Aby obsługiwać szereg obciążzeń i usług biznesowych, należy zaprojektować chmurę. Można stwierdzić, kiedy usługa zostanie zmodernizowana, a kiedy można ją obniżyć.

W szczególności to podejście do projektowania architektury oparte na usługach obsługuje kluczowe cechy chmury, takie jak elastyczność, samoobsługę, interfejsy oparte na standardach i elastyczność płatności zgodnie z rzeczywistym użyciem. Połączenie podejścia zorientowanego na usługi z usługami w chmurze umożliwia przedsiębiorstwom obniżenie kosztów i poprawę elastyczności w biznesie. Skalowalność i elastyczność systemów chmur publicznych i prywatnych są wymienne i luźno mieszane.

2.6 Podsumowanie

W tym rozdziale przedstawiliśmy przetwarzanie równoległe i rozproszone jako ramy, na których można właściwie opisać przetwarzanie w chmurze. Rozwiążanie głównego problemu wyłoniło się z przetwarzania równoległego i rozproszonego poprzez wykorzystanie najpierw kilku komponentów przetwarzających, a następnie wielu węzłów komputerowych w sieci.

2.7 Pytania przeglądowe

1. Rozróżnij przetwarzanie równoległe i rozproszone.
2. Co to jest architektura SIMD?
3. Wyjaśnij główne kategorie systemów obliczeń równoległych.
4. Wyjaśnij różne poziomy równoległości, jakie można uzyskać w obliczeniach system
5. Co to jest system rozproszony? Jakie elementy ją charakteryzują?
6. Co to jest styl architektoniczny i jak radzi sobie z systemem rozproszonym?
7. Wymień najważniejsze style architektury oprogramowania.
8. Jakie są podstawowe style architektoniczne systemu?
9. Opisz najważniejszy model komunikacji opartej na komunikatach.
10. Omów RPC i sposób, w jaki umożliwia komunikację między procesami.
11. Co to jest CORBA?
12. Co to jest przetwarzanie zorientowane na usługi?
13. Czym jest chmura obliczeniowa zorientowana rynkowo?

2.8 Odsyłacz do dalszej lektury

1. Opanowanie podstaw przetwarzania w chmurze i programowania aplikacji Rajkumar Buyya, Christian Vecchiola, S. Publikacje Thamarai Selvi MK ISBN: 978-0-12-411454-8
2. Koncepcje, technologia i architektura przetwarzania w chmurze Thomas Erl, Zaigham The Prentice Hall Mahmooda i Ricardo Puttiniego , Service Technology Series ISBN-10: 9780133387520 ISBN-13: 978-0133387520
3. Przetwarzanie rozproszone i przetwarzanie w chmurze: od przetwarzania równoległego do Internetu rzeczy, wydanie pierwsze, Kai Hwang Jack Dongarra Geoffrey Fox ISBN-10: 9789381269237 ISBN-13: 978-9381269237

Rozdział 2

Rozdział 3

Struktura

jednostki 3.0

Cel 3.1

Wprowadzenie 3.2 Główne komponenty środowiska wirtualizacji 3.2.1 Charakterystyka

wirtualizacji 3.3 Taksonomia technik wirtualizacji

3.3.1 Wirtualizacja wykonania

3.3.2 Model referencyjny maszyny

3.3.2.1 Architektura zestawu instrukcji (ISA)

3.3.2.2 Interfejs binarny aplikacji

3.4 Pierścienie bezpieczeństwa i tryb uprzywilejowany

3.4.1 Pierścień 0 (najbardziej uprzywilejowany) i 3 (najmniej uprzywilejowany)

3.4.2 Pierścienie 1 i 2

3.5 Wirtualizacja na poziomie sprzętu

3.6 Hiperwizor

3.6.1 Hiperwizor typu 1

3.6.2 Hiperwizor typu 2

3.6.3 Wybór odpowiedniego hiperwizora

3.6.7 Model referencyjny hiperwizora

3.7 Techniki wirtualizacji sprzętowej 3.7.1

Zalety sprzętu Wirtualizacja wspomagana 3.8 Wirtualizacja

pełna 3.9 Parawirtualizacja

3.10 Wirtualizacja na

poziomie języka programowania 3.10.1 Wirtualizacja

na poziomie aplikacji 3.11 Inne typy

wirtualizacji 3.11.1 Wirtualizacja

pamięci masowej 3.11.2

Wirtualizacja sieci 3.11.3

Wirtualizacja komputerów

stacjonarnych 3.11.4 Wirtualizacja serwerów aplikacji

3.12 Wirtualizacja i przetwarzanie w chmurze

3.12.1 Plusy i minusy wirtualizacji

3.12.1.1 Zalety wirtualizacji 3.12.1.2 Wady

wirtualizacji 3.13 Przykłady technologii

3.13.1 Xen: parawirtualizacja

3.13.2 VMware: pełna wirtualizacja

3.13.3 Pełna wirtualizacja i tłumaczenie

binarne 3.13.4 Rozwiązania wirtualizacyjne 3.13.5

Użytkownik końcowy (komputer

stacjonarny) wirtualizacja 3.13.6 Wirtualizacja

serwerów

3.14 Microsoft Hyper-V

3.14.1 Architektura

3.15

Podsumowanie 3.16

Pytania przeglądowe 3.17 Odniesienia do dalszej lektury

Cel 3.0

Wirtualizacja wyodrębnia sprzęt, który może współużytkować wspólne zasoby z wieloma obciążeniami. Różnorodne obciążenia można kolokować na współdzielonym, zwirtualizowanym sprzęcie, zachowując jednocześnie pełną izolację, swobodnie migrując pomiędzy infrastrukturami i skalując, jeśli jest to wymagane.

Firmy generują znaczne aktywa i wydajność dzięki wirtualizacji, ponieważ skutkuje to lepszym wykorzystaniem i konsolidacją serwerów, dynamicznym przydzielaniem zasobów i zarządzaniem nimi, izolacją obciążeń roboczych, bezpieczeństwem i automatyzacją. Wirtualizacja umożliwia samodzielne świadczenie usług na żądanie i orkiestrację zasobów definiowanych programowo, które są dostępne lokalnie lub zdalnie, w dowolnym miejscu w chmurze hybrydowej, zgodnie z konkretnymi potrzebami biznesowymi.

3.1 Wprowadzenie

Wirtualizacja chmury sprawia, że system operacyjny serwera i urządzenia pamięci masowej stają się platformą wirtualną. Umożliwi to użytkownikowi współdzielenie pojedynczej instancji zasobu fizycznego lub aplikacji z kilkoma użytkownikami poprzez udostępnienie wielu maszyn. Wirtualizacja w chmurze zarządza także pracą poprzez transformację, skalowalność, ekonomiczność i wydajność tradycyjnego przetwarzania.

Wirtualizacja przetwarzania w chmurze szybko integruje kluczową metodę obliczeniową. Jedną z kluczowych cech wirtualizacji jest to, że umożliwia wielu klientom i firmom współdzielenie aplikacji.

Środowisko wirtualizacji można również nazwać usługami i aplikacjami opartymi na chmurze. Albo publiczne, albo prywatne to środowisko. Klient może zmaksymalizować zasoby poprzez wirtualizację i zmniejszyć potrzebny system fizyczny.

Ostatnio, na skutek splotu kilku zjawisk, coraz większym zainteresowaniem cieszy się technologia wirtualizacji:

Zwiększoną wydajność i pojemność obliczeniową:

Unikalne korporacyjne centrum danych w większości przypadków nie jest w stanie konkurować pod względem bezpieczeństwa, wydajności, szybkości i opłacalności z siecią centrów danych udostępnianą przez usługodawcę. Ponieważ większość usług jest dostępna na żądanie, w krótkim czasie użytkownicy mogą mieć także duże ilości zasobów obliczeniowych, z niezwykłą łatwością i elastycznością oraz bez kosztownych inwestycji.

Z kolei usługi chmurowe oferują możliwość zwolnienia pamięci i mocy obliczeniowej na poszczególnych komputerach poprzez zdalny hosting platform, oprogramowania i baz danych. Oczywistym rezultatem jest w rzeczywistości znaczna poprawa wydajności.

Niewykorzystane zasoby sprzętu i oprogramowania.

Niedostateczne wykorzystanie sprzętu i oprogramowania jest spowodowane zwiększoną mocą obliczeniową i wydajnością oraz ograniczonym lub rzadkim wykorzystaniem zasobów. Systemy komputerowe stały się dziś tak potężne, że w niektórych przypadkach te, które stanowią jedynie ułamek ich możliwości, są wykorzystywane przez aplikację lub system. Ponadto, biorąc pod uwagę infrastrukturę informatyczną firmy, wiele systemów komputerowych jest wykorzystywanych tylko częściowo, podczas gdy można z nich korzystać bez przerwy, 24 godziny na dobę, 7 dni w tygodniu, 365 dni w tygodniu. Na przykład,

komputery stacjonarne przeznaczone głównie do zadań automatyzacji biura i wykorzystywane przez personel administracyjny są wykorzystywane wyłącznie w godzinach pracy. Efektywność infrastruktury IT można zwiększyć, wykorzystując te zasoby do innych celów. Aby świadczyć taką usługę w sposób przejrzysty, potrzebne jest całkowicie odrębne środowisko, które można osiągnąć poprzez wirtualizację.

Brak miejsca.

Centra danych stale się rozwijają, co wiąże się z koniecznością rozbudowy infrastruktury, takiej jak pamięć masowa czy moc obliczeniowa. Organizacje takie jak Google i Microsoft rozbudowują swoją infrastrukturę, budując centra danych na wzór stadionu piłkarskiego, na których znajdują się tysiące węzłów. Chociaż jest to wykonalne dla dużych graczy z branży IT, firmy często nie są w stanie zbudować dodatkowego centrum danych, aby pomieścić dodatkową pojemność zasobów. Wraz z tą sytuacją niewykorzystane zasoby sprzętowe, które doprowadziły do rozpowszechnienia konsolidacji serwerów, stanowiącej podstawę wirtualizacji, są wykorzystywane w technice.

Inicjatywy proekologiczne

Wirtualizacja to podstawa technologia wdrażania infrastruktury opartej na chmurze, umożliwiającej jednocześnie uruchamianie wielu obrazów systemów operacyjnych na jednym serwerze fizycznym. Jako czynnik umożliwiający konsolidację, wirtualizacja serwerów zmniejsza całkowity rozmiar serwera fizycznego, co wiąże się z korzyściami ekologicznymi.

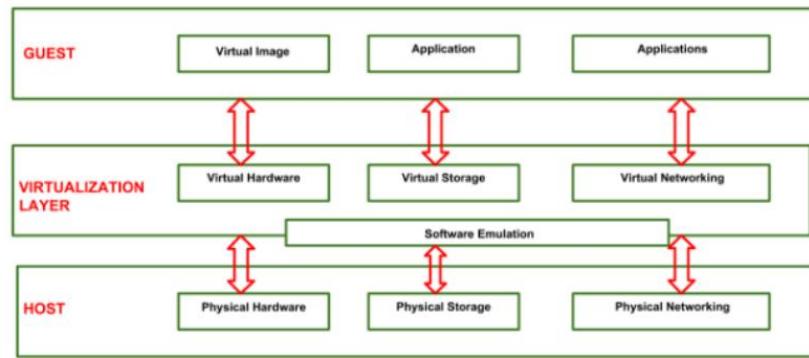
Z punktu widzenia zasoboszczędności wymaganych jest mniej obciążen, co aktywnie zmniejsza przestrzeń w centrum danych i ostateczny ślad e-odpadów. Z punktu widzenia efektywności energetycznej centrum danych będzie zużywać mniej energii elektrycznej przy mniejszej liczbie sprzętu fizycznego. Chłodzenie w centrach danych jest głównym wymogiem i może pomóc w przypadku wysokiego zużycia energii. Dzięki metodom darmowego chłodzenia, takim jak wykorzystanie powietrza i wody w porównaniu z klimatyzacją i chłodzeniem, centra danych mogą obniżyć koszty chłodzenia. Menedżerowie centrów danych mogą zaoszczędzić na kosztach energii elektrycznej dzięki panelom słonecznym, regulatorom temperatury i panelom wykorzystującym energię wiatrową.

Wzrost kosztów administracyjnych.

Rośnie zużycie energii i koszty chłodzenia, podobnie jak koszty urządzeń IT. Ponadto zwiększone zapotrzebowanie na dodatkową pojemność, która przekształca się w większą liczbę serwerów w centrum danych, prowadzi do znacznego wzrostu kosztów administracyjnych. Nie wszystkie komputery — zwłaszcza serwery — będą działać niezależnie, ale wymagają opieki i uwagi administratora systemu. Monitorowanie sprzętu, wymiana wadliwego sprzętu, instalacja i aktualizacje serwerów, monitorowanie zasobów serwera i tworzenie kopii zapasowych należą do typowych zadań administracyjnych systemu. Operacje te są czasochłonne, a im więcej serwerów do obsługi, tym wyższe koszty administracyjne. Im większe są wydatki administracyjne, tym wirtualizacja może przyczynić się do zmniejszenia liczby serwerów wymaganych do danego obciążenia i zmniejszenia kosztów personelu administracyjnego.

3.2 Główne składniki środowiska wirtualizacji

Wirtualizacja to sposób na utworzenie „wersji wirtualnej” maszyny fizycznej. Za pomocą monitora maszyny wirtualnej osiąga się wirtualizację. Umożliwia działanie kilku maszyn wirtualnych na jednym urządzeniu fizycznym. Bez żadnych zmian obserwowanych w maszynach wirtualnych można go łatwo przenieść ze sprzętu na inny. W chmurze obliczeniowej szeroko stosowana jest wirtualizacja. Wirtualizacja pomaga uruchamiać wiele systemów operacyjnych i aplikacji na tych samych komponentach sprzętowych na każdym z nich.



Postać: 3.1 Referencyjny model wirtualizacji.

(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming” autorstwa Rajkumara Buyyi)

W środowisku zwirtualizowanym do tej kategorii zaliczają się trzy główne komponenty:

1. GOŚĆ:

Jak zwykle, gość oznacza komponent systemu współpracujący z warstwą wirtualizacji, a nie z maszyną hosta. Zwykle gościom prezentowany jest jeden lub więcej plików definicji dysku wirtualnego i maszyny wirtualnej. Aplikacja hosta, która wygląda i zarządza każdą maszyną wirtualną jako odrębną aplikacją, jest centralnie obsługiwana przez maszyny wirtualne.

2. Hosty:

Host to oryginalne środowisko, w którym ma być zarządzany gość. Każdy gospodarz korzysta ze wspólnych zasobów, które gospodarz udostępnia każdemu gościowi. System operacyjny działa jako host i zarządza fizycznym zarządzaniem zasobami oraz usługą urządzenia.

3. Warstwa wirtualizacji

Warstwa wirtualizacji zapewnia odtworzenie tego samego lub innego środowiska, w którym działa gość. Jest to dodatkowa warstwa abstrakcji pomiędzy sprzętem, systemem obliczeniowym i aplikacją działającą w sieci oraz pamięcią masową. Zwykle pomaga obsługa jednego systemu operacyjnego na maszynę, co w porównaniu z wirtualizacją jest bardzo mało elastyczne.

3.2.1 Charakterystyka wirtualizacji

1. Zwiększone bezpieczeństwo –

Możliwość w pełni przejrzystego zarządzania wykonywaniem programu gościa stwarza nowe możliwości zapewnienia bezpiecznego, kontrolowanego środowiska wykonywania. Wszystkie programy gościnne działają zwykle na maszynie wirtualnej, tłumacząc je i wykorzystując w programie hosta.

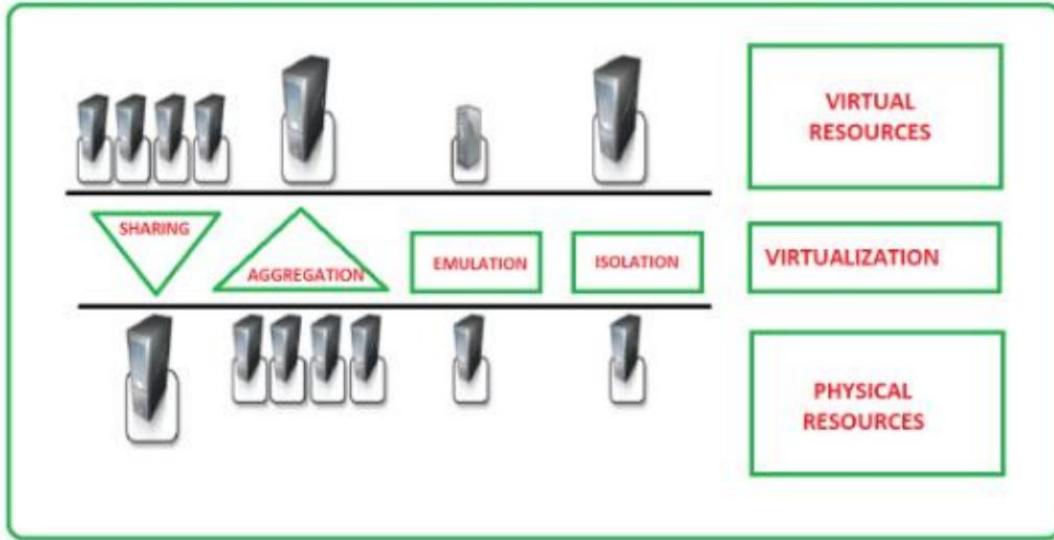
Menedżer maszyny wirtualnej może zarządzać i filtrować aktywność programów gościnnych, aby zapobiec wykonywaniu szkodliwych operacji. Zasoby ujawnione przez gospodarza można następnie ukryć lub po prostu chronić przed gościem.

Przykład-1: W środowisku piaskownicy Cuckoo można ocenić niezaufany kod. W kontekście piaskownicy instrukcje mogą być filtrowane i blokowane w izolowanym środowisku wykonawczym przed tłumaczeniem i wykonaniem w rzeczywistym środowisku wykonawczym.

Przykład 2: Wyrażenie wirtualnej maszyny Java (JVM) w trybie piaskownicy oznacza konkretną konfigurację maszyny JVM, w której instrukcje uważane za potencjalnie szkodliwe mogą być blokowane za pomocą zasad bezpieczeństwa.

2. Zarządzane wykonaniem –

W szczególności najważniejszymi funkcjami są udostępnianie, agregacja, emulacja i izolacja.



Rysunek: 3.2 Funkcje udostępniane przez wykonanie zarządzane
(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming”
autorstwa Rajkumara Buyyi)

3. Dzielenie się –

Wirtualizacja umożliwia utworzenie oddzielnego środowiska obliczeniowego na tym samym gościu. Ta wspólna funkcja zmniejsza liczbę aktywnych serwerów i zmniejsza zużycie energii.

4. Agregacja –

Zasób fizyczny może być nie tylko współdzielony pomiędzy kilkoma gośćmi, ale wirtualizacja umożliwia także agregację. Grupę pojedynczych hostów można połączyć i przedstawić jako pojedynczy host wirtualny. Funkcjonalność ta jest realizowana za pomocą oprogramowania Cluster Management Software, które wykorzystuje i reprezentuje zasoby fizyczne jednolitej grupy maszyn.

5. Emulacja –

W warstwie wirtualizacji, która zasadniczo jest programem, programy gościa są wykonywane w środowisku. Można również emulować zupełnie inne środowisko w odniesieniu do hosta, dzięki czemu można realizować programy gościa wymagające pewnych funkcji, których nie ma na gościu fizycznym.

6. Izolacja –

Wirtualizacja pozwala gościom zapewnić całkowicie oddzielne środowisko, w którym są wykonywane — jeśli są to systemy operacyjne, aplikacje lub inne podmioty. Program gościa działa poprzez warstwę abstrakcji, która oferuje dostęp do podstawowych zasobów. Maszyna wirtualna jest w stanie filtrować działania gościa i zapobiegać niebezpiecznym operacjom przeciwko hostowi.

Oprócz tych funkcji dostrajanie wydajności to kolejna ważna funkcja, która umożliwia wirtualizację. Ta funkcja stała się dostępna dzięki znacznemu postępowi w oprogramowaniu i sprzęcie wspierającym wirtualizację. Dzięki precyzyjnemu dostosowaniu właściwości zasobów dostępnych w środowisku wirtualnym łatwiej jest kontrolować wydajność gości. Oferuje środki do skutecznego wdrożenia infrastruktury jakości usług (QoS).

7. Przenośność –

W zależności od konkretnego rodzaju wirtualizacji koncepcja przenośności ma różne zastosowania.

W przypadku wirtualizacji sprzętowej gość jest pakowany w obraz wirtualny, który w wielu przypadkach można bezpiecznie przenosić i uruchamiać na różnych maszynach wirtualnych. W przypadku wirtualizacji poziomu programowania, realizowanej w środowisku wykonawczym JVM lub .NET, kod binarny komponentów aplikacji (słoiki lub zespoły) może działać na odpowiedniej maszynie wirtualnej bez rekompilacji.

3.3 Taksonomia technik wirtualizacji

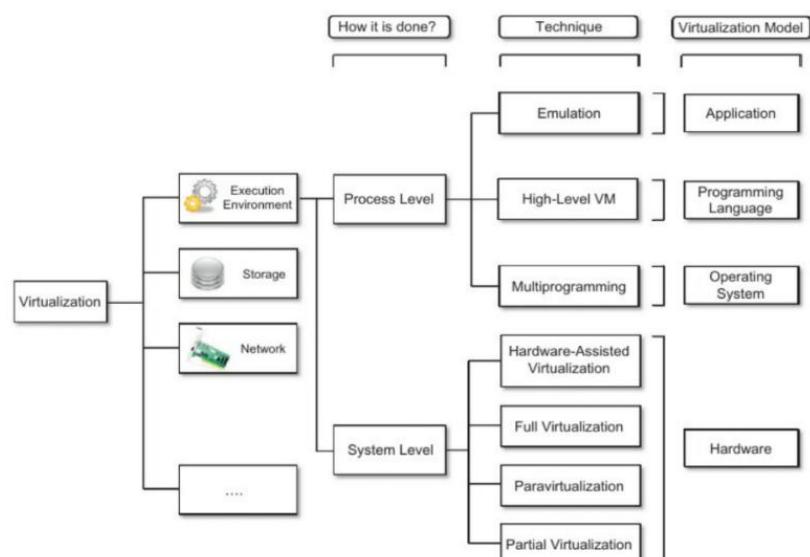
Wirtualizacja obejmuje szeroką gamę technik emulacji stosowanych w różnych obszarach obliczeniowych. Klasyfikacja takich metod pozwala je zrozumieć i zastosować.

Usługa lub podmiot jest dyskryminowany na podstawie pierwszej naśladowanej klasyfikacji. Wirtualizacja służy przede wszystkim do emulacji w środowiskach wykonawczych, pamięci masowej i sieciowych. Najstarszym, popularnym i rozwiniętym obszarem tych kategorii jest wirtualizacja wykonawcza. Wymaga zatem dalszych badań i klasyfikacji. Techniki wirtualizacji możemy szczegółowo podzielić, badając typ hosta, jakiego wymagają, na dwie główne kategorie.

Techniki wirtualizacji możemy szczegółowo podzielić, badając typ hosta, jakiego wymagają, na dwie główne kategorie.

Techniki na poziomie procesu są wdrażane jako dodatek do istniejącego systemu operacyjnego z pełną kontrolą sprzętu.

Technika poziomów systemowych jest wykonywana bezpośrednio na sprzęcie i nie wymaga wsparcia ze strony istniejącego systemu operacyjnego lub wymaga ograniczonego wsparcia. W tych dwóch kategoriach możemy wyróżnić różne metody zapewniania gościom innego wirtualnego środowiska obliczeniowego: goły sprzęt, zasoby systemów operacyjnych, język programowania niskiego poziomu i biblioteki aplikacji.



RYSUNEK 3.3 Taksonomia technik wirtualizacji.
(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming” autorstwa Rajkumara Buyyi)

3.3.1 Wirtualizacja wykonania

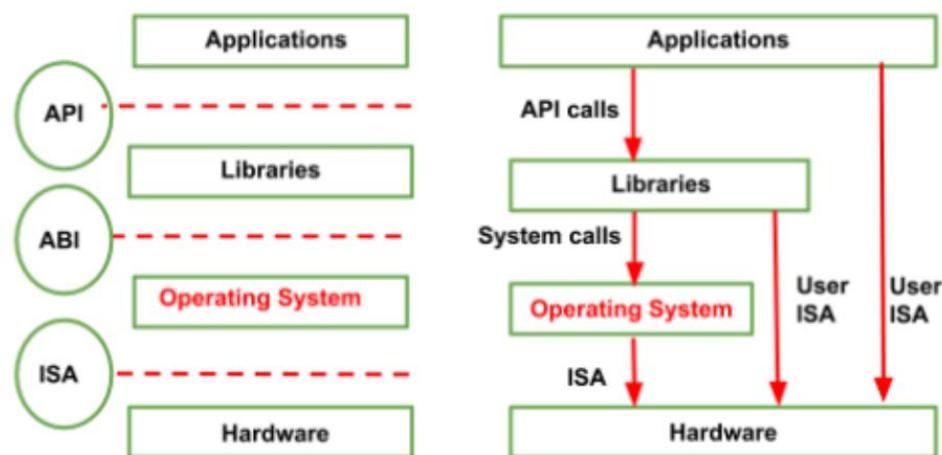
Wirtualizacja wykonania obejmuje wszystkie metody imitowania środowiska wykonawczego oddzielnego od hosta warstwy wirtualizacji. Wszystkie te techniki skupią się na wspieraniu wykonywania programu, niezależnie od tego, czy jest to system operacyjny, specyfikacja programu binarnego skompilowana na podstawie modelu, czy zastosowanie abstrakcyjnego modelu maszyny. Dlatego system operacyjny, aplikacja i biblioteki mogą być bezpośrednio lub dynamicznie połączone z obrazem aplikacji na wierzchu sprzętu.

3.3.2 Model referencyjny maszyny

Jeśli środowisko wykonawcze jest zwirtualizowane na poziomach innych niż stos obliczeniowy, należy opracować strukturę odniesienia, która definiuje interfejsy na poziomie abstrakcyjnym, a ten poziom abstrakcji maskuje szczegóły implementacji. Sugeruje to, że techniki wirtualizacji mogą zastąpić każdą warstwę i przechwycić wywołania do niej. Z tego powodu wyraźne oddzielenie warstw może uprościć ich

implementacja, w której wystarczy emulować interfejsy i reagować na warstwę podrzędną.

W warstwie bazowej deklarowany lub demonstrowany jest model sprzętu zgodnie z architekturą, tj. architekturą zestawu instrukcji (ISA).



RYSUNEK 3.4 Model referencyjny maszyny

(Odniesienie z „Opanowanie podstaw i aplikacji przetwarzania w chmurze Programowanie” Rajkumara Buyyi)

3.3.2.1 Architektura zestawu instrukcji (ISA)

Zestaw instrukcji, znany jako ISA, jest składnikiem komputera związanym z programowaniem, które zasadniczo jest językiem maszynowym. Zestaw instrukcji dostarcza procesorowi instrukcje, które mówią mu, co ma robić. Zestaw instrukcji składa się z trybów adresowania, instrukcji, natywnych typów danych, rejestrów, architektury pamięci, obsługi przerwań i wyjątków oraz zewnętrznych wejść/wyjść.

Przykładem zestawu instrukcji jest zestaw instrukcji x86, powszechnie stosowany obecnie w komputerach. Dzięki wciąż dość zróżnicowanej konstrukcji wewnętrznej, różne procesory komputerowe mogą korzystać z prawie tego samego zestawu instrukcji. Obydwa procesory Intel Pentium i AMD Athlon korzystają z niemal tego samego zestawu instrukcji x86. Zestaw instrukcji może być wbudowany w sprzęt procesora lub emulowany przez interpreter za pomocą oprogramowania. Konstrukcja sprzętu do uruchamiania programów jest wydajniejsza i szybsza niż emulowana wersja programu.

3.3.2.2 Interfejs binarny aplikacji

ABI to binarny interfejs aplikacji. Kod binarny ABI definiuje sposób wywoływania funkcji, sposób przekazywania parametrów między obiektem wywołującym a obiektem wywoływanym, w jaki sposób zwarcane wartości są przekazywane obiektem wywołującym, w jaki sposób wdrażane są biblioteki i jak programy są ładowane do pamięci. W ten sposób linker stosuje ABI: ABI to zasady działania niepowiązanego kodu w połączeniu. ABI reguluje także współistnienie procesów w tym samym systemie. Na przykład ABI może określić w systemie UNIX, w jaki sposób wykonywane są sygnały, w jaki sposób proces wywołuje wywołania systemowe, jaka jest używana końcówka i tworzona są stosy. ABI to zbiór reguł implementowanych w określonej architekturze przez system operacyjny.

Jądro, łańcuch narzędzi i trojka architektury definiują ABI. Wszyscy muszą się na to zgodzić. Architektury zazwyczaj projektują preferowany lub standaryzowany ABI, a systemy operacyjne mniej więcej przestrzegają tej standaryzacji. Informacje takie są zwykle dokumentowane w podręczniku referencyjnym architektury. Na przykład x86-64,

3.4 Pierścienie bezpieczeństwa i tryb uprzywilejowany

Procesor działa głównie na dwóch poziomach uprawnień:

Tryb użytkownika: Dostęp do pamięci jest w tym trybie ograniczony do pewnego stopnia, natomiast dostęp urządzeń peryferyjnych jest zabroniony.

Tryb jądra: Procesory posiadają instrukcje dotyczące zarządzania i uzyskiwania dostępu do pamięci w tym trybie, a także instrukcje dotyczące dostępu do urządzeń peryferyjnych, takich jak dyski i karty sieciowe. Procesor automatycznie przełącza się z jednego działającego programu na inny działający program. To warstwowe podejście upraszcza rozbudowę i zastosowania systemu komputerowego. Ten

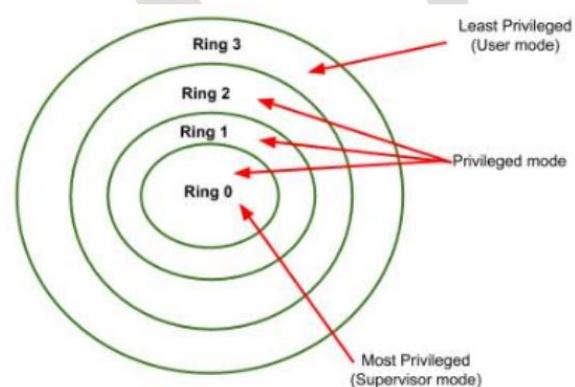
podejście warstwowe upraszcza stosowanie wielozadaniowości i współistnienie wielu wykonarń.

Pierwszą z nich można wykonać w instrukcjach uprzywilejowanych i nieuprzywilejowanych. Instrukcje, których można użyć podczas przerwania innego zadania, można nazwać nieuprzywilejowanymi Instrukcją. Nazywa się go również, ponieważ nie jest dostępny za pośrednictwem współdzielonych zasobów. Ex- zawiera wszystkie punkty stałe oraz instrukcje zmiennoprzecinkowe i arytmetyczne. Instrukcje wykonywane z określonymi ograniczeniami i powszechnie używane do wrażliwych operacji (ujawnianie kontroli wrażliwych na zachowanie lub wrażliwych na zmiany) są znane jako instrukcje uprzywilejowane.

System operacyjny zarządza zasobami komputera, takimi jak czas przetwarzania procesora i dostęp do pamięci. Na komputerach często uruchamianych jest jednocześnie kilka procesów programowych, co wymaga różnych poziomów dostępu do zasobów i sprzętu.

Procesy realizowane są w warstwowych „pierścieniach” z różnymi prawami dostępu do zasobów w każdym pierścieniu. Pierścień centralny ma najwyższe uprawnienia, a dostęp jest ograniczany w każdej kolejnej warstwie. Typowa implementacja pierścienia zabezpieczającego procesor x86 (typowy typ procesora) ma cztery pierścienie, od 0 do 3

Model warstwowy ma dwie główne zalety. Przede wszystkim chroni przed awariami systemu. Błędy można zwykle odzyskać w wyższych pierścieniach (z mniejszym dostępem). Ponieważ Ring 0 ma bezpośredni dostęp do pamięci i procesora, można go ponownie uruchomić bez utraty danych lub błędu procesora w procesie awarii pierścienia zewnętrznego. Po drugie, zapewnia większe bezpieczeństwo. Procedura wymaga pozwolenia systemu operacyjnego na wykonanie instrukcji wymagających większego dostępu do zasobów. Następnie system operacyjny może zdecydować, czy uwzględni żądanie. Ten proces selekcji pomaga zapobiegać niepożdanemu lub złośliwemu zachowaniu systemu.



RYSUNEK 3.5 Pierścienie bezpieczeństwa i tryby uprawnień

(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming” autorstwa Rajkumara Buyyi)

3.4.1 Pierścień 0 (najbardziej uprzywilejowany) i 3 (najmniej uprzywilejowany)

Dostępny jest pierścień jądra 0, który jest podstawowym elementem większości systemów operacyjnych i ma dostęp do wszystkiego. Mówiąc, że działa kod działający w trybie jądra. Procesy działające w jądrze mogą znacznie wpływać na cały system; jeśli coś się nie powiedzie, prawdopodobnie nastąpi zamknięcie systemu. Pierścień ten ma bezpośredni dostęp do procesora i pamięci systemowej, co oznacza, że istnieją instrukcje wymagające użycia któregokolwiek z nich.

Pierścień 3, najmniej uprzywilejowany pierścień, jest dostępny dla procesów użytkownika w trybie użytkownika. Jest to lokalizacja większości aplikacji działających na komputerze. Pierścień ten nie ma bezpośredniego dostępu do procesora ani pamięci i dlatego musi przekazywać instrukcje do pierścienia 0.

3.4.2 Pierścienie 1 i 2

Istnieją specjalne uprawnienia dla pierścieni 1 i 2, które nie istnieją w pierścieniu 3 (tryb użytkownika). Pierścień 1 służy do interakcji ze sprzętem podłączonym do komputera i kontrolowania go. Odtwarzanie utworu przez głośniki lub słuchawki albo wyświetlanie wideo na monitorze to przykłady wykorzystania tej funkcji

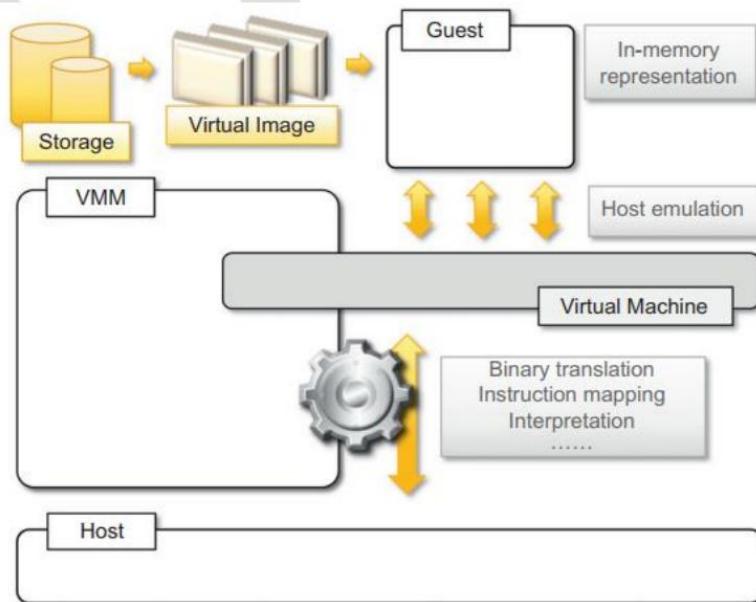
pierścień. Pierścień 2 jest używany do instrukcji, które wchodzą w interakcję z pamięcią systemową, ładują lub zapisują pliki. Tego typu uprawnienia nazywane są uprawnieniami wejściowymi i wyjściowymi, ponieważ dane są przenoszone do lub z pamięci roboczej (RAM). Na przykład znajduje się w pierścieniu 2, aby załadować dokument przechowywany w programie Word. Przeglądanie i edycja dokumentów, czyli warstwa aplikacji, będzie objęta pierścieniem 3.

W środowisku hypervisor oczekuje się, że kod systemu operacyjnego gościa zostanie uruchomiony w użytkowniku, aby uniemożliwić mu bezpośredni dostęp do stanu systemu operacyjnego. Po zaimplementowaniu instrukcji nieuprzywilejowanych nie jest już możliwe całkowite odizolowanie systemu gościa. Rozróżnienie pomiędzy trybem użytkownika i nadzorcy pozwala nam zrozumieć rolę hypervisor. Hiperwizor koncepcyjnie znajduje się nad nadzorcą i używany jest przedrostek hyper-. Hypervisory faktycznie działają w trybie nadzorcy, a projektowanie menedżerów maszyn wirtualnych wiąże się z wyzwaniami związanymi z podziałem na instrukcje uprzywilejowane i nieuprzywilejowane. Oczekuje się, że wszystkie wrażliwe instrukcje będą wykonywane w trybie uprzywilejowanym, który wymaga trybu nadzorcy, aby zapobiec pułapkom. Bez tego założenia nie można w pełni emulować i zarządzać stanem procesora w systemach operacyjnych gościa. Niestety nie jest tak w przypadku oryginalnego ISA, który pozwala na 17 wrażliwych instrukcji trybu użytkownika. Zapobiega to separacji i zmianom wielu systemów operacyjnych zarządzanych przez pojedynczy system hypervisor. Ostatnie wdrożenia ISA (Intel VT, AMD Pacifica) rozwiązały ten problem, przekształcając takie instrukcje na instrukcje uprzywilejowane.

3.5 Wirtualizacja na poziomie sprzętowym

Wirtualizacja na poziomie sprzętowym to technika wirtualizacji zapewniająca technikę umożliwiającą wykonanie abstrakcyjnego środowiska wykonawczego sprzętu komputerowego, w którym można uruchomić system operacyjny gościa. Gość jest definiowany w tym modelu poprzez system operacyjny, a host poprzez komputer fizyczny

Sprzęt, emulacja maszyny wirtualnej i menedżer maszyny wirtualnej hypervisor (patrz rysunek 3.6). Hiperwizor to zwykle oprogramowanie/program sprzętowy, który umożliwia wyodrębnienie podstawowego sprzętu fizycznego. Wirtualizacja na poziomie sprzętu, która reprezentuje interfejs sprzętowy systemu, jest również nazywana wirtualizacją systemu, ponieważ ISA zapewnia maszyny wirtualne. Oznacza to, że wyróżnia się maszyny wirtualne, które wystawiają ABI na działanie praktycznie różnych procesów.



RYSUNEK 3.6 Model referencyjny wirtualizacji sprzętu.

(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming” autorstwa Rajkumara Buyyi)

3.6 Hypervisory

Hypervisor to kluczowy element oprogramowania umożliwiający wirtualizację. Abstrahuje od rzeczywistego sprzętu komputerów-gostów i używanego przez nich systemu operacyjnego.

Hypervisory tworzą wirtualną warstwę procesora/procesora, pamięci RAM i innych zasobów fizycznych, która oddziela Cię od tworzonych urządzeń wirtualnych.

Hypervisor, na którym instalujemy maszynę, nazywany jest maszyną hosta, w porównaniu z działającymi na niej wirtualnymi maszynami-gosćmi. Hypervisory emulują zasoby dostępne dla komputerów gościnnych. Niezależnie od tego, który system operacyjny uruchamiasz z rzeczywistego sprzętu, uważa się, że dostępnny jest prawdziwy sprzęt fizyczny. Z punktu widzenia maszyny wirtualnej środowisko fizyczne i wirtualne nie różnią się niczym. W środowisku wirtualnym maszyny-Gości nie wiedzą, że utworzył je hiperwizor. Lub udostępnij dostępną moc obliczeniową. Maszyny wirtualne działają jednocześnie na sprzęcie, który je zasila, dlatego są w pełni zależne od ich stabilności.

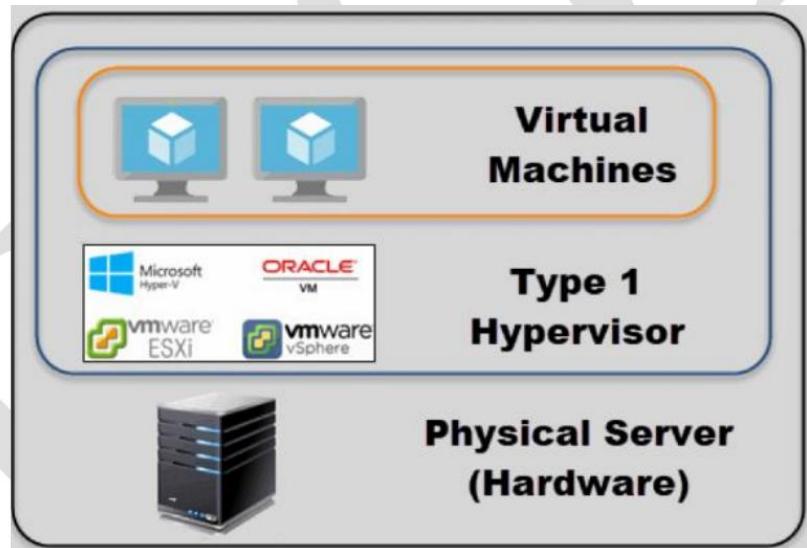
- Hypervisor typu 1 (zwany także bare metal lub natywny)
- Hypervisor typu 2 (znany również jako hostowany hypervisor)

3.6.1 Hiperwizor typu 1

Hypervisor typu bare-metal (typ 1) to warstwa oprogramowania instalowana bezpośrednio nad serwerem fizycznym i leżącym na nim sprzętem. Przykłady hiperwizorów typu 1 obejmują hiperwizor VMware ESXi, Citrix XenServer i Microsoft Hyper-V.

Nie ma oprogramowania pośredniego ani systemu operacyjnego, dlatego nazwa to hypervisor typu bare-metal. Hiperwizor typu 1, który nie działa w systemie Windows ani w żadnym innym systemie operacyjnym, dlatego udowodniono, że zapewnia doskonałą wydajność i stabilność.

Hypervisory typu 1 same w sobie są bardzo prostym systemem operacyjnym, na którym można obsługiwać maszyny wirtualne. Maszyna fizyczna hypervisor jest używana wyłącznie do celów wirtualizacji serwera. Do niczego innego nie możesz go użyć. W środowiskach korporacyjnych najczęściej spotykane są hypervisory typu 1.



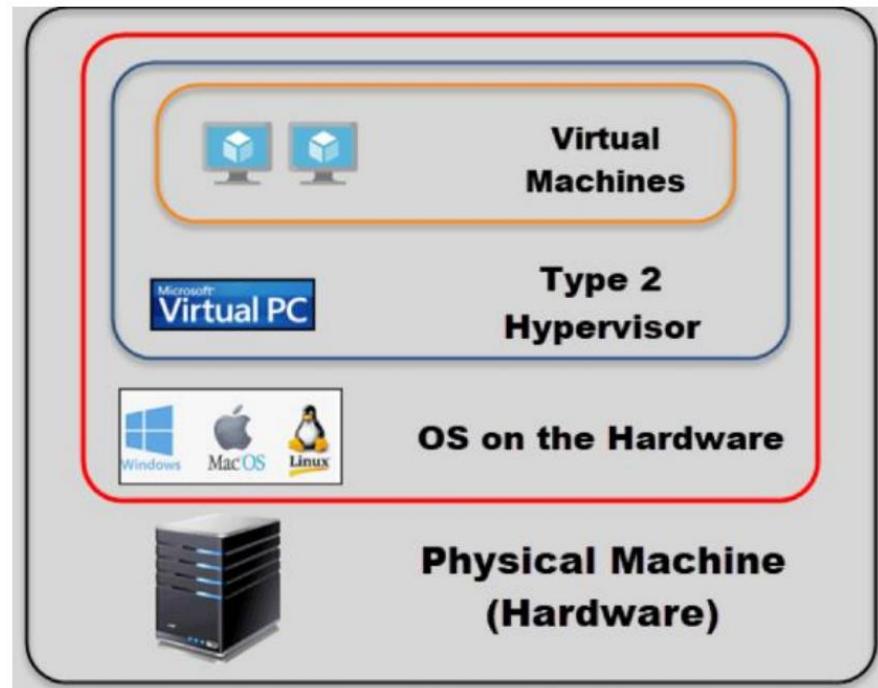
RYSUNEK 3.7 Hiperwizor typu 1

3.6.2 Hiperwizor typu 2

Ten typ hypervisorów działa w systemie operacyjnym hosta fizycznego. .Przykładami hypervisorów typu 2 są VMware Player lub Parallels Desktop.

Dlatego nazywamy hypervisorami typu 2 – hypervisorami hostowanymi. W przeciwieństwie do hiperwizorów typu 1, które działają bezpośrednio na sprzęcie, hiperwizory hostowane mają jedną podstawową warstwę oprogramowania. Tutaj mamy co następuje:

- Maszyna fizyczna. •
- Zainstalowany sprzętowy system operacyjny (Windows, Linux, macOS). •
- Oprogramowanie dla hypervisorów typu 2 w tym systemie operacyjnym. •
- Bieżące instancje wirtualnych maszyn gościnnych.



RYSUNEK 3.8 Hiperwizor typu 2

3.6.3 Wybór odpowiedniego hypervisora

Hiperwizory typu 1 oferują znacznie lepszą wydajność niż typ 2. Stanowią one logiczny wybór w przypadku aplikacji i obciążen o znaczeniu krytycznym, ponieważ nie mają warstwy środkowej. Nie oznacza to jednak, że hostowane hiperwizory, które hostują, nie mają dla siebie miejsca. Są znacznie łatwiejsze w konfiguracji, więc jest to dobry wybór, jeśli, powiedzmy, będziesz musiał szybko wdrożyć środowisko testu. Jednym z najlepszych sposobów sprawdzenia, który hypervisor spełnia Twoje potrzeby, jest porównanie ich wskaźników wydajności. Przed wybrianiem odpowiedniego hypervisora należy sprawdzić następujące czynniki: Obejmują one obciążenie procesora, maksymalną pamięć hosta i gościa oraz obsługę procesorów wirtualnych

1. Zrozum swoje potrzeby: centrum danych (i swoją pracę) dla firmy i jej aplikacji. Oprócz wymagań Twojej firmy, Ty (i Twój personel IT) macie również własne wymagania.

- A. Elastyczność
- B. Skalowalność
- C. Użyteczność
- D. Dostępność
- mi. Niezawodność
- F. Efektywność
- G. Niezawodne wsparcie

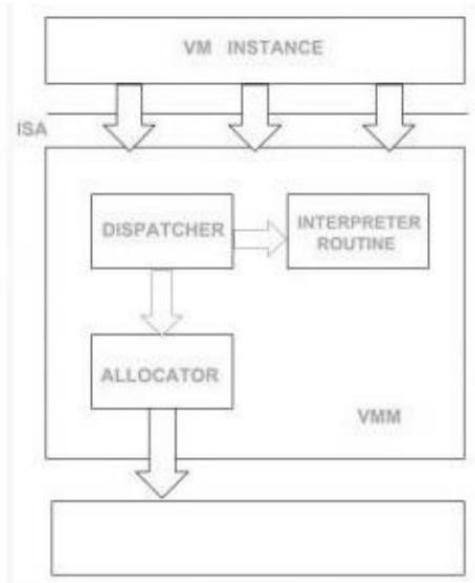
2. Koszt hypervisora: Dla wielu kupujących hypervisor jest najtrudniejszą rzeczą w wyborze właściwej równowagi pomiędzy kosztem a funkcjonalnością. Chociaż niektóre rozwiązania podstawowe są bezpłatne lub praktycznie bezpłatne, ceny na drugim końcu rynku mogą być oszałamiające. Ramy licencjonowania również są różne, dlatego ważne jest, aby dokładnie wiedzieć, na co trafiają Twoje pieniądze.

3. Wydajność maszyny wirtualnej: Systemy wirtualne powinny osiągać lub przekraczać, przynajmniej w odniesieniu do aplikacji serwerowych, wydajność swoich fizycznych odpowiedników. Wszystko, co wykracza poza ten punkt odniesienia, to zysk.

4. Ekosystem: Rola, jaką może pełnić hiperwizor ds. ekosystemów w podejmowaniu decyzji, czy rozwiązanie jest opłacalne, czy nie, jest kusząca do zignorowania – tj. dostępność dokumentacji, wsparcia, szkoleń, rozwoju i usług doradczych.

5. Przetestuj sam: Możesz uzyskać podstawowe doświadczenia z istniejącego komputera stacjonarnego lub laptopa. Aby zbudować ładne wirtualne środowisko do nauki i testowania, możesz uruchomić VMware vSphere i Microsoft Hyper-V na VMware Workstation lub VMware Fusion.

3.6.7 Model referencyjny hiperwizora



RYSUNEK 3.9 Architektura referencyjna hypervisor.

(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming” autorstwa Rajkumara Buyyi)

Istnieją 3 główne moduły koordynujące emulację podstawowego sprzętu:

1. Dyspozytor
2. Alokator
3. Tłumacz

Dyspozytor:

Dyspozytor pełni rolę punktu wejścia monitora, przekierowując instrukcje instancji maszyny wirtualnej do jednego z dwóch pozostałych modułów.

ALOKATOR:

Alokator jest odpowiedzialny za podjęcie decyzji o przydzieleniu zasobów systemowych instancji maszyny wirtualnej. Wskazuje, że dyspozytor wywołuje alokatora za każdym razem, gdy maszyna wirtualna próbuje wykonać instrukcje modyfikujące zasoby maszyny powiązane z maszyną wirtualną.

Interpretator:

Moduł interpretera składa się z procedur. Są one wykonywane, gdy maszyna wirtualna wykonuje uprzywilejowaną instrukcję.

Wymagania Popeka i Goldberga to szereg warunków, które są wystarczające, aby architektura komputera skutecznie wspierała wirtualizację systemu. W artykule z 1974 r. zatytułowanym „Formal Results for Virtualizable Third Generation Architectures” przedstawionym przez Geralda J. Popeka i Roberta P. Goldberga. Chociaż uwzględniono założenia upraszczające, wymagania pozostają użytecznym sposobem ustalenia, czy architektura komputera obsługuje wydajne wirtualizacji i zawiera wytyczne dotyczące projektowania zwirtualizowanych architektur komputerów.

Maszyny systemu wirtualnego mogą wirtualizować całą gamę zasobów sprzętowych, takich jak procesory, zasoby pamięci i pamięci masowej oraz urządzenia peryferyjne. Monitor maszyny wirtualnej to komponent oprogramowania zapewniający abstrakcję maszyny wirtualnej, zwany także hiperwizorem. Analizując środowisko stworzone przez VMM, istnieją trzy następujące właściwości: -

Równoważność/wierność

W przypadku VMM uruchomiony program powinien zachowywać się zasadniczo tak samo, jak w przypadku działania bezpośrednio na równoważnej maszynie.

Kontrola zasobów / Bezpieczeństwo

Zwirtualizowane zasoby muszą być w pełni kontrolowane przez VMM.

Wydajność/Wydajność

Bez interwencji VMM należy wykonać statystycznie dominującą część instrukcji maszynowych.

VMM musi zawierać wszystkie trzy właściwości w terminologii Popeka i Goldberga.

Zwykle zakłada się, że VMM spełniają właściwości równoważności i kontroli zasobów, znane również jako efektywne VMM. Cechy, jakie muszą posiadać Popek i Goldberg, aby wykonać VMM posiadające powyższe właściwości, są opisane w architekturze zestawu instrukcji (ISA). Model ten składa się z procesora systemowego lub trybu użytkownika, który ma dostęp do liniowej, spójnej adresowej pamięci. Zakłada się, że podzbiór zestawu instrukcji jest dostępny tylko w trybie systemowym i adresowana jest pamięć związana z rejestrów relokacji. Przerwania i wejścia/wyjścia nie są modelowane.

Popek i Goldberg dzielą instrukcje ISA na trzy różne grupy, aby wyprowadzić swoje twierdzenia o wirtualizacji, które dają wystarczające (ale nie konieczne) warunki wirtualizacji:

Uprzywilejowane instrukcje

Te, które zatrzymują się w trybie użytkownika, gdy procesor jest w trybie systemowym, nie powodują pułapek (tryb nadzorczy).

Kontroluj wrażliwe instrukcje

Ci, którzy próbują zmienić konfigurację zasobów systemowych.

Instrukcje dotyczące zachowania

Zachowanie tych, których wyniki zależą od konfiguracji zasobu (zawartość rejestrów relokacji lub tryb procesora).

Taki może być zatem główny wynik analiz Popeka i Goldberga.

Twierdzenie 1. Efektywny VMM można zbudować na dowolnym konwencjonalnym komputerze trzeciej generacji, jeśli wrażliwy zestaw instrukcji jest podzbiorem uprzywilejowanych instrukcji dla tego komputera.

Twierdzenie stwierdza intuicyjnie, że wszystkie instrukcje, które mogą wpływać na prawidłowe działanie VMM (instrukcje wrażliwe), zawsze przechwytyują i przekazują kontrolę do VMM, są wystarczające do zbudowania VMM. Zapewnia to właściwość kontroli zasobów.

Zamiast tego powinny być wykonywane natywne (tzn. wydajne) instrukcje nieuprzywilejowane. Konieczne jest także zachowanie własności równoważności.

Twierdzenie to zapewnia również prostą metodę dla VMM, znaną ostatnio jako klasyczna wirtualizacja zwana wirtualizacją typu pułapka i emulacji: jedyne, co wrażliwe instrukcje VMM muszą zrobić, to przechwytywać i emulować każdą z nich.

Powiązanym problemem jest to, że uzyskuje się wystarczające warunki dla wirtualizacji rekurencyjnej, tj. warunki, w których można utworzyć VMM, który może pracować z własną kopią. Następujące warunki (wystarczające) prezentują Popek i Goldberg.

Twierdzenie 2. Konwencjonalny komputer trzeciej generacji można rekurencyjnie wirtualizować, jeśli:

- Można go wirtualizować i
- Można dla niego skonstruować VMM bez żadnych zależności czasowych.

Istnieją architektury, które nie spełniają tych warunków, np. x86 niewspomagane sprzętowo, dlatego nie można ich normalnie wirtualizować. Jednak architektury (w x86

przypadek, poziom procesora/MMU) można nadal w pełni zwirtualizować przy użyciu różnych technik, takich jak translacja binarna, która zastępuje wrażliwe instrukcje, które nie generują pułapek, czasami określane jako instrukcje krytyczne. Jednak to dodatkowe przetwarzanie zmniejsza teoretyczną wydajność VMM, a pułapki sprzętowe są opłacalne.

Porównywalną wydajność można osiągnąć w płynnie dostrojonym systemie tłumaczenia binarnego, który pozwala na przechwytywanie jedynie wrażliwych instrukcji w odniesieniu do wspomagania sprzętowego x86 pierwszej generacji.

Twierdzenie 3. Hybrydowy VMM można skonstruować dla dowolnej maszyny trzeciej generacji, w której zestaw instrukcji wrażliwych dla użytkownika stanowi podzbiór zbioru instrukcji uprzywilejowanych:

Można zbudować hybrydowy VMM trzeciej generacji, w którym instrukcje wrażliwe dla użytkownika stanowią część zestawu instrukcji uprzywilejowanych.

– Więcej instrukcji będzie interpretowanych w HVM zamiast bezpośrednio wykonywanych.

3.7 Techniki wirtualizacji sprzętu

Wirtualizacja wspomagana sprzętowo, pierwszy system operacyjny maszyny wirtualnej (VM/370 w 1972 r.), został wprowadzony na IBM System/370. Pod koniec lat 70-tych zapomniano o wirtualizacji, ale rozwój serwerów x86 na nowo rozjaśnił zainteresowanie wirtualizacją ze względu na wymagania konsolidacji serwerów. Wirtualizacja umożliwiła zastąpienie wielu niewykorzystanych serwerów dedykowanych jednym serwerem.

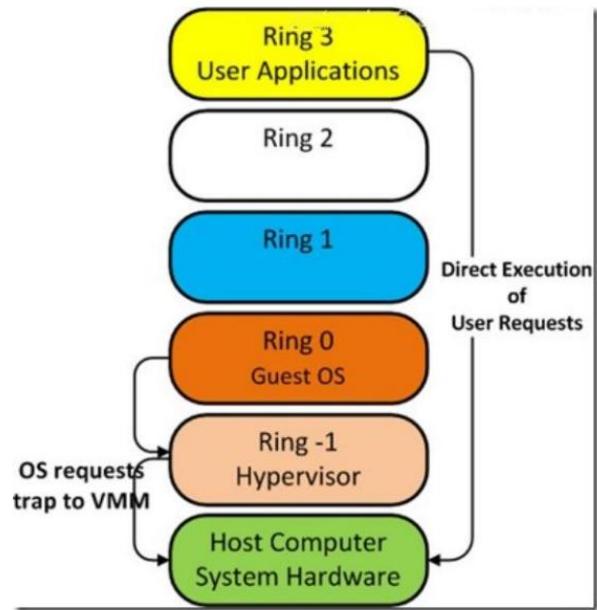
Architektura x86 nie spełniała jednak kryteriów Goldberga i Popeka dla „klasycznej wirtualizacji”. Aby zrekompensować te ograniczenia: Wirtualizacja architektury x86 została przeprowadzona dwiema metodami: wirtualizacją pełną lub parawirtualizacją.

Iluzja fizycznego sprzętu jest tworzona, aby osiągnąć cel polegający na niezależnym zarządzaniu systemem operacyjnym od sprzętu, ale w celu osiągnięcia pewnej wydajności i złożoności.

Intel i AMD wprowadziły nowe technologie wirtualizacji, szereg nowych instrukcji i – co najważniejsze – nowy poziom uprawnień. Hiperwizor jest teraz obecny w „Pierścieniu -1”, dzięki czemu system operacyjny gościa może działać w pierścieniu 0.

Wirtualizacja sprzętu wykorzystuje funkcjonalność wirtualizacji wbudowaną w najnowszą generację procesorów Intel i AMD. Technologie te, zwane odpowiednio Intel VT i AMD-V, oferują ulepszenia potrzebne do uruchomienia niezmodyfikowanych maszyn wirtualnych bez narzutu związanego z pełną emulacją wirtualizacji procesora. Te nowe procesory obejmują dodatkowy tryb uprawnień poniżej pierścienia 0, w którym hiperwizor zasadniczo może działać, pozostawiając pierścień 0 dla niezmodyfikowanych systemów operacyjnych gościa.

VMM może wydajnie wirtualizować całą instrukcję X86 za pomocą wirtualizacji wspomaganej sprzętowo, korzystając z klasycznie używanego modelu sprzętowej pułapki i emulacji, a nie oprogramowania, obsługując te wrażliwe instrukcje. Dostęp do procesora może być dostępny w Ring 1 i systemach gościnych za pośrednictwem hiperwizorów obsługujących tę technologię w taki sam sposób, jak w przypadku pracy na hoście fizycznym. Umożliwia to wirtualizację systemów gościa bez żadnych zmian.



RYSUNEK 3.10 Nowy poziom uprawnień w architekturze x86

3.7.1 Zalety wirtualizacji wspomaganej sprzętowo

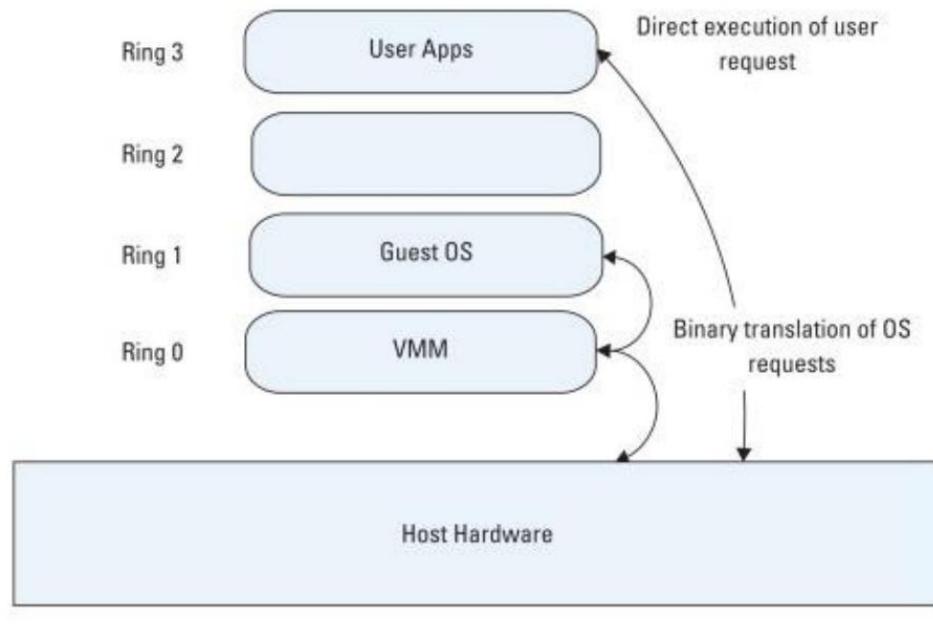
Wirtualizacja wspomagana sprzętowo zmienia dostęp do systemu operacyjnego. Systemy operacyjne x86 mają bezpośredni dostęp do uruchomionych zasobów systemowych. VMM emuluje niezbędny sprzęt w systemie operacyjnym za pomocą wirtualizacji oprogramowania. System operacyjny zapewnia bezpośredni dostęp do zasobów bez emulacji lub modyfikacji za pomocą wirtualizacji wspomaganej sprzętowo, co poprawia ogólną wydajność.

Oznacza to, że jądra systemu operacyjnego nie muszą być modyfikowane i mogą działać bez zmian (jak w przypadku wirtualizacji par). Hiperwizor nie musi jednocześnie brać udziału w nieefektywnym binarnym tłumaczeniu wrażliwych instrukcji. Tym samym nie tylko spełnia kryteria Popeka i Goldberga (pełnej wirtualizacji), ale także poprawia swoją wydajność, ponieważ instrukcje są teraz przechwytywane i emulowane bezpośrednio w sprzęcie.

3.8 Pełna wirtualizacja

Pełna wirtualizacja to technika wirtualizacji VME, która całkowicie symuluje podstawowy sprzęt. Każde oprogramowanie, które można uruchomić na sprzęcie fizycznym, można uruchomić w tego typu środowisku na maszynie wirtualnej, a na każdej maszynie wirtualnej można uruchomić dowolny system operacyjny obsługiwany przez podstawowy sprzęt. Użytkownicy mogą jednocześnie uruchamiać kilka różnych systemów gościa. Maszyna wirtualna symuluje wystarczający sprzęt, aby niezmodyfikowany system operacyjny gościa mógł działać w izolacji w pełnym oprogramowaniu do wirtualizacji. W wielu sytuacjach jest to szczególnie pomocne. Na przykład nowy, eksperymentalny kod można uruchomić w fazie rozwoju systemu operacyjnego na osobnej maszynie wirtualnej jednocześnie ze starszymi wersjami. Hiperwizor zapewnia każdej maszynie wirtualnej, w tym wirtualny BIOS, urządzenia wirtualne i zarządzanie pamięcią wirtualną, wszystkie usługi systemu fizycznego. System operacyjny gościa jest całkowicie odłączony od warstwy wirtualizacji od podstawowego sprzętu.

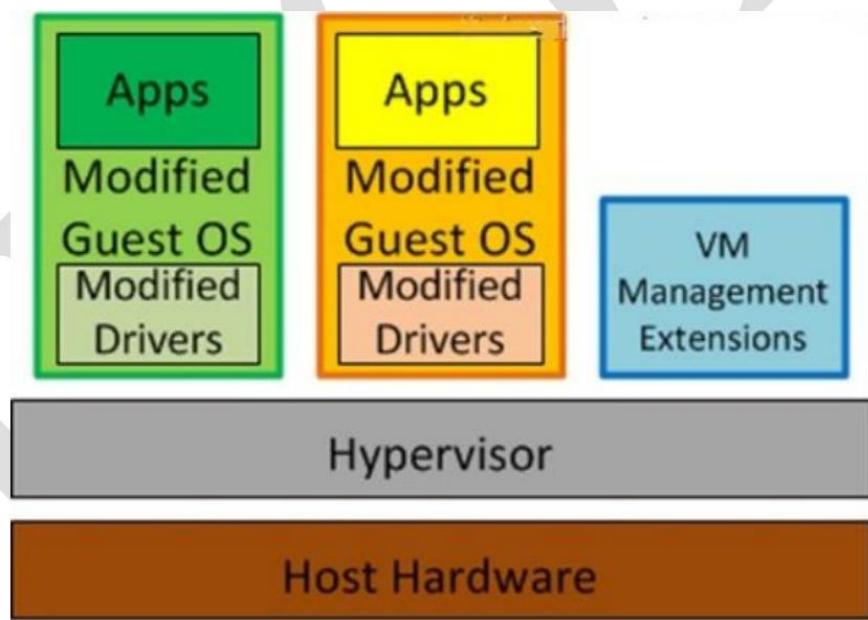
Pełną wirtualizację osiąga się poprzez zastosowanie kombinacji wykonania binarnego i bezpośredniego. Fizyczny procesor wykonuje instrukcje z natywną szybkością, niewrażliwe na pełne hiperwizory wirtualizacji, tłumaczy instrukcje systemu operacyjnego i jest buforowany do wykorzystania w przeszłości, a instrukcje na poziomie użytkownika są wykonywane z natywną szybkością bez zmian. Pełna wirtualizacja zapewnia optymalną izolację i bezpieczeństwo maszyn wirtualnych, ułatwiając migrację i przenośność, ponieważ sprzęt wirtualny i natywny jest używany przez tę samą instancję systemu gościa. Koncepcję pełnej wirtualizacji pokazano na rysunku



RYSUNEK 3.11 Pełna wirtualizacja

3.9 Parawirtualizacja

Parawirtualizacja to kolejne podejście do wizualizacji serwerów, w którym parawirtualizacja to cienka warstwa, która nie imituje kompletnego środowiska sprzętowego; dba o to, aby wszystkie systemy-gosci współdzieliły swoje zasoby systemowe i dobrze współpracowały ze sobą. „Para” to angielski przyrostek pochodzenia greckiego, który oznacza „obok”, „z” lub „obok”.

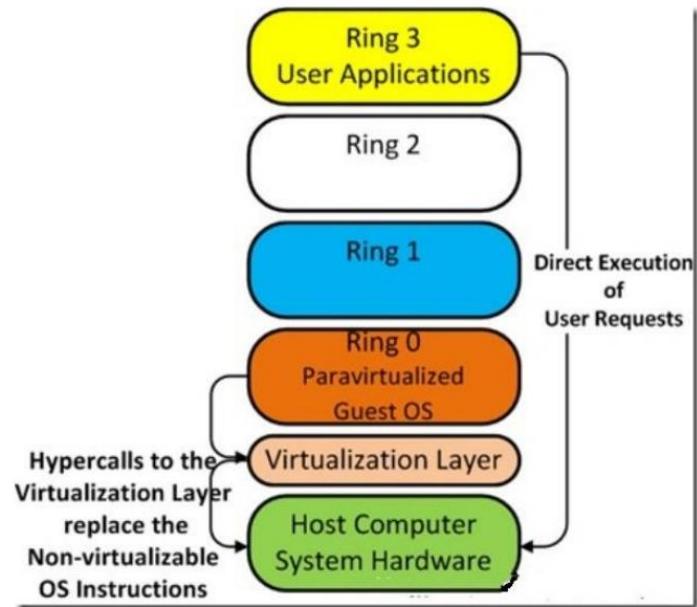


RYSUNEK 3.12 Parawirtualizacja

Jądro systemu operacyjnego gościa zostanie zmienione tak, aby działało na hiperwizorze w ramach parawirtualizacji. Zwykle wymaga to zastąpienia w pierścieniu 0 procesora operacji uprzywilejowanych poprzez wywołanie hiperwizora (zwanego hiperwywołaniami). Z kolei hiperwizor wykonuje zadanie w imieniu jądra gościa i oferuje interfejsy hiperwywołań do innych kluczowych operacji jądra, takich jak zarządzanie pamięcią, obsługa przerwań i utrzymywanie czasu.

Parawirtualizacja próbuje rozwiązać wszystkie problemy związane z wirtualizacją, umożliwiając systemom operacyjnym gościa bezpośredni dostęp do podległego sprzętu, a tym samym usprawniając komunikację między systemem operacyjnym gościa a hiperwizorem. Ponieważ zawiera modyfikacje systemu operacyjnego, parawirtualizacja jest czasami nazywana wirtualizacją wspomaganą przez system operacyjny.

Parawirtualizacja, w której system gościa „wie”, w jaki sposób jest wirtualizowany, różni się od pełnej wirtualizacji, w której niezmodyfikowany system operacyjny nie wie, że jest zwirtualizowany, a wrażliwe wywołania systemu operacyjnego są uwięzione przez translację binarną.



RYSUNEK 3.13 Hiperwywołania do wirtualizacji w parawirtualizacji

Zalety parawirtualizacji

Takie podejście ma dwie zalety:

Zdolność jądra gościa do bezpośredniej komunikacji z hiperwizorem prowadzi do wyższego poziomu wydajności. Jak pamiętasz, pełna wirtualizacja polega na umieszczeniu pełnej warstwy emulacji sprzętowej pomiędzy systemem gościa a sprzętem fizycznym. Cienka warstwa oprogramowania parawirtualizacji działa bardziej jak zwirtualizowany serwer, taki jak kontroler ruchu lotniczego, który zapewnia systemowi gościa dostęp do fizycznych zasobów sprzętowych, podczas gdy wszystkie inne systemy operacyjne gościa przestają jednocześnie uzyskiwać dostęp do tych samych zasobów. Wartość, jaką oferuje parawirtualizacja, polega na wirtualizacji o niższym narzucie, ale przewaga wydajności w porównaniu z pełną wirtualizacją zależy od obciążenia; metoda ta jest ogólnie znacznie bardziej wydajna niż konwencjonalne wirtualizacje z emulacją sprzętu;

Drugą zaletą podejścia parawirtualizacji w porównaniu z pełną wirtualizacją jest to, że parawirtualizacja nie ogranicza się do sterowników urządzeń zawartych w oprogramowaniu do wirtualizacji. Zamiast tego wykorzystuje sterowniki urządzeń, zwane gościćmi uprzywilejowanymi, w jednym z naszych systemów operacyjnych gościa. Jeśli nie zagłębisz się tutaj zbytnio w tę architekturę, musisz po prostu powiedzieć, że jest to zaleta, ponieważ daje organizjom możliwość korzystania ze wszystkich możliwości sprzętowych serwera, zamiast ograniczać się do sprzętu, dla którego sterowniki są dostępne w całości w programach do wirtualizacji.

Ograniczenia parawirtualizacji

W przypadku parawirtualizacji systemy operacyjne gościa muszą zostać zmienione, aby współdziałały z interfejsami parawirtualizacji. Zwykle ogranicza to obsługę systemów operacyjnych typu open source, takich jak Linux, które można w sposób otwarty modyfikować, oraz systemów operacyjnych zastrzeżonych, których właściciele zgadzają się na tworzenie kodów dla konkretnego hypervisor. Ponieważ parawirtualizacja nie obsługuje niezmodyfikowanego systemu operacyjnego (np. rodziny Windows), nie jest kompatybilna i przenośna;

Parawirtualizacja może również spowodować poważne problemy ze wsparciem produkcyjnym i konserwacją, ponieważ potrzebne są głębokie poprawki w jądrze systemu operacyjnego.

3.9.1.2 Częściowa wirtualizacja

W informatyce częściowa wirtualizacja to technika wirtualizacji zastosowana do wdrożenia środowiska maszyny wirtualnej: zapewniającej „częściową symulację podstawowego sprzętu”. Większość, choć nie wszystkie, funkcjonalności sprzętu jest symulowana, co skutkuje maszynami wirtualnymi, które mogą obsługiwać określone lub całe oprogramowanie bez modyfikacji. Ogólnie oznacza to, że całe systemy operacyjne „nie mogłyby”, ale wiele aplikacji można uruchomić, uruchomić na maszynie wirtualnej. To oznaka pełnej wirtualizacji.

Kluczem do częściowej wirtualizacji jest „wirtualizacja przestrzeni adresowej”, w której każda maszyna wirtualna składa się z odrębnej przestrzeni adresowej. Możliwość ta wymaga sprzętu do relokacji i dostosowania częściową wirtualizację w innych praktycznych przykładach.

Główne historyczne punktem zwrotnym na drodze do pełnej wirtualizacji była wirtualizacja częściowa. Wykorzystano go w systemie CTSS z współdzieleniem czasu pierwszej generacji oraz w eksperymentalnym systemie przywoławczym IBM M44/44X. Pod pojęciem tym można zdefiniować dowolny system operacyjny z wydzielonymi przestrzeniami adresowymi dla niezależnych użytkowników lub procesów, do których zalicza się wiele, które obecnie nie spełniają kryteriów systemów maszyn wirtualnych. Doświadczenie i ograniczenia częściowej wirtualizacji doprowadziły do powstania pierwszego systemu pełnej wirtualizacji

Częściowa wirtualizacja jest znacznie łatwiejsza niż pełna wirtualizacja. Często udostępnia przydatne, mocne maszyny wirtualne obsługujące główne aplikacje. Jego wadą są sytuacje, w których wymagana jest kompatybilność wstępna lub przenośność (w przeciwieństwie do pełnej wirtualizacji). Jeśli nie są symulowane pewne funkcje sprzętu, całe oprogramowanie korzystające z takich funkcji nie będzie działać. Ponadto funkcje używane w konkretnym zastosowaniu mogą być trudne do dokładnego przewidzenia.

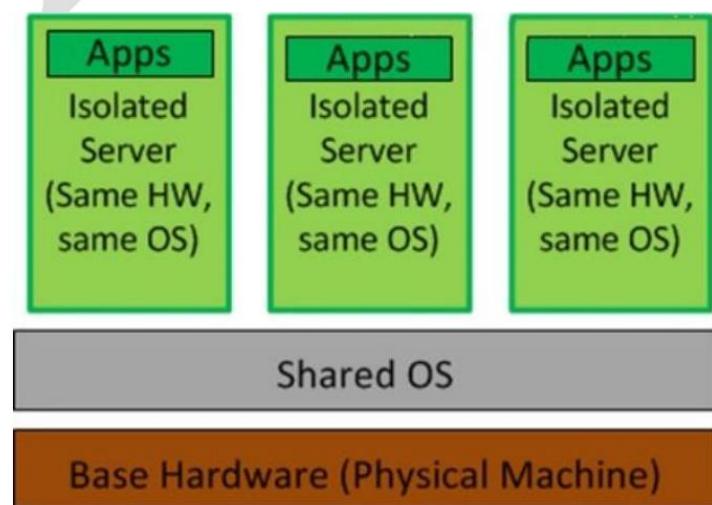
Częściowa wirtualizacja okazała się niezwykle skuteczna w przypadku współużytkowania komputera przez wielu użytkowników zasoby.

3.9.2 Wirtualizacja na poziomie systemu operacyjnego

Wirtualizacja na poziomie systemu operacyjnego (wirtualizacja systemu operacyjnego) to technologia wirtualizacji serwerów, która obejmuje modyfikowanie systemu operacyjnego w celu umożliwienia jednoczesnego uruchamiania różnych aplikacji na jednym komputerze przez różnych użytkowników. Choć działają na tym samym komputerze, wirtualne systemy operacyjne nie kolidują ze sobą niezależnie. Standardowy system operacyjny jest zmieniany i dostosowywany do pracy niezależnych systemów. Ten wirtualny system jest zaprojektowany tak, aby spełniać polecenia użytkownika, który może jednocześnie uruchamiać różne aplikacje na maszynie. Wirtualny system operacyjny przetwarza każde żądanie użytkownika indywidualnie. Zaletą wirtualizacji na poziomie systemu operacyjnego jest to, że dostępność aplikacji będzie miała minimalny wpływ nawet podczas aktualizacji systemu i poprawek bezpieczeństwa. Wirtualizacja systemu operacyjnego umożliwia przeniesienie kluczowych aplikacji do innych wirtualnych systemów operacyjnych, dzięki czemu wydajność może być kontynuowana.

Metodologia, w której jądro systemu operacyjnego umożliwia wykorzystanie kilku izolowanych instancji przestrzeni użytkownika w tego rodzaju wirtualizacji serwerów. Instancje działają na bazie poprzedniego systemu operacyjnego hosta i zawierają zestaw bibliotek, z którymi aplikacje współpracują, ilustrując sposób ich działania na maszynie przeznaczonej do ich użytku.

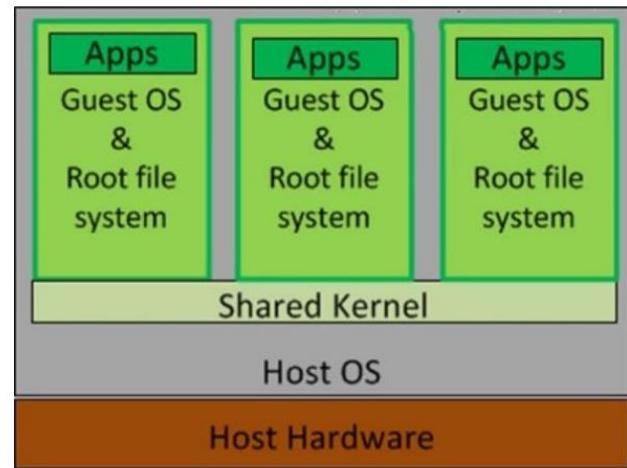
Kontenery, wirtualne serwery prywatne lub środowiska wirtualne nazywane są instancjami.



RYSUNEK 3.14 Wirtualizacja na poziomie systemu operacyjnego

Osiągnięto system hosta z pojedynczym jądrem systemu operacyjnego i kontrolą funkcjonalności gościa na poziomie systemu operacyjnego wirtualizującego. W tej wirtualizacji współdzielonego jądra każdy wirtualny system gościa ma własny główny system plików.

Wirtualizacja hosta, w której hiperwizor (kontener) ma bardzo ograniczoną funkcjonalność, która zależy od systemu operacyjnego hosta w zakresie planowania procesora i zarządzania pamięcią. Ta metoda, która wykorzystuje wirtualizację na poziomie systemu operacyjnego, nie obejmuje nawet zastosowania prawdziwego hiperwizora, ale jest składnik systemu operacyjnego, który wykonuje wszystkie zadania hypervisorów.



RYSUNEK 3.15 Wirtualizacja na poziomie systemu operacyjnego 2

Oznacza to, że wirtualizacja systemu operacyjnego opiera się na utworzeniu na pojedynczym serwerze fizycznym izolowanych kontenerów lub partycji, w tym na wykorzystaniu instancji systemu operacyjnego do niezależnego działania z innych partycji, w każdym środowisku aplikacji gościa. Technika ta instaluje warstwę oprogramowania do wirtualizacji w systemie operacyjnym, a system dla wszystkich gości działa na tej warstwie przy użyciu tego samego systemu operacyjnego, co system operacyjny hosta, ale każdy gość ma własne zasoby i działa w całkowitej izolacji od gości.

Mogą argumentować, że nie jest to wirtualizacja w ścisłym tego słowa znaczeniu; jest to raczej technika, która umożliwia jedynie konsolidację maszyny.

3.10 Wirtualizacja na poziomie języka programowania

3.10.1 Wirtualizacja na poziomie aplikacji

Wirtualizacja aplikacji to mechanizm, który oszukuje standardową aplikację, wierząc, że wchodzi w bezpośrednią interakcję z funkcjonalnością systemu operacyjnego, podczas gdy w rzeczywistości tak nie jest.

Wymaga to warstwy wirtualizacji umieszczonej pomiędzy aplikacją a systemem operacyjnym. Ta warstwa lub system musi wirtualnie uruchamiać podzbiory aplikacji bez wpływu na podstawowy system operacyjny. Warstwa wirtualizacji zastępuje część środowiska wykonawczego zwykle zapewnianą przez system operacyjny, w przejrzysty sposób przekierowując pliki do jednego pliku wykonywalnego i wprowadzając zmiany w dzienniku rejestru.

Dzięki przekierowaniu procesów aplikacji do jednego pliku, a nie kilku rozproszonych po całym systemie operacyjnym, aplikacja działa łatwo na innym urządzeniu, a aplikacje, które wcześniej były niekompatybilne, mogą teraz działać obok siebie.

Wirtualizacja pulpitu jest używana w połączeniu z wirtualizacją aplikacji — oddzieleniem od systemu użytkownika końcowego, który uzyskuje dostęp do fizycznego środowiska pulpitu i powiązanej z nim aplikacji.

Korzyści z wirtualizacji aplikacji

- Umożliwia korzystanie ze starszych aplikacji (np. platform systemów operacyjnych, takich jak Windows 7 lub XP, które są przeznaczone do programowania). koniec można uruchomić.
- Umożliwia działanie na wielu platformach (np. z systemami iOS, Android, macOS i Chrome OS).
- Zapobiega konfliktom z innymi zwirtualizowanymi aplikacjami
- Umożliwia użytkownikom obsługę wielu instancji aplikacji — jeśli nie są one zwirtualizowane, kilka aplikacji może wykryć nowe instancje i uniemożliwić ich uruchomienie

Ograniczenia wirtualizacji aplikacji

- Trudno jest wirtualizować wszystkie programy komputerowe. Przykładami są aplikacje wymagające sterownika systemowego (rodzaj integracji z systemem operacyjnym) i aplikacje 16-bitowe do działania w przestrzeni pamięci współdzielonej.
- Oprogramowanie antywirusowe i programy wymagające dużej integracji z systemem operacyjnym są trudne do wirtualizacji, takie jak WindowBlinds lub StyleXP.
- Wirtualizacja aplikacji podlega poważnym wadom licencyjnym w licencjonowaniu oprogramowania, zwłaszcza dlatego, że musi prawidłowo licencjonować zarówno oprogramowanie do wirtualizacji aplikacji, jak i aplikacje zwirtualizowane.
- Podczas gdy wirtualizacja aplikacji może rozwiązać problemy występujące pomiędzy starymi aplikacjami a nowszymi systemami operacyjnymi w zakresie zgodności plików i rejestru, aplikacje, które nie obsługują sterów prawidłowo, nie będą działać z systemem Windows Vista, ponieważ nadal przydziela on pamięć, niezależnie od sposobu zwirtualizowany. Dlatego nawet jeśli program jest zwirtualizowany, mogą być wymagane określone poprawki dotyczące zgodności aplikacji (podkładki).

3.11 Inne typy wirtualizacji

Wiele form wirtualizacji ma abstrakcyjne środowisko, w którym można z nimi współdziałać w sposób inny niż wirtualizacja. Obejmują one przechowywanie, tworzenie sieci i interakcję pomiędzy klientem a serwerem.

3.11.1 Wirtualizacja pamięci masowej

Pamięć masowa to kolejna część koncepcji obliczeń wirtualizacyjnych. Definicja wirtualizacji pamięci masowej brzmi: Wirtualizacja pamięci masowej odnosi się do metody abstrakcji fizycznej pamięci masowej.

Chociaż funkcjonalność RAID jest na poziomie podstawowym, słowo „wirtualizacja pamięci masowej” zazwyczaj obejmuje dodatkowe koncepcje, takie jak migracja danych i buforowanie. Wirtualizacja pamięci masowej jest trudna do opisania w konkretny sposób, ponieważ możliwe jest posiadanie wielu różnych funkcjonalności. Zazwyczaj jest ona podawana jako funkcja:

- Oparty na hoście ze specjalnymi sterownikami urządzeń
- Kontrolery macierzowe
- Przełączniki sieciowe
- Samodzielne urządzenia sieciowe

Pod tym względem każdy sprzedawca ma inne podejście. Podstawowym sposobem klasyfikacji wirtualizacji pamięci masowej jest podział na wirtualizację wewnętrzpasową lub pozapasmową. Wewnętrzpasowe (często nazywane symetrycznym) pomiędzy hostem a urządzeniem magazynującym umożliwia buforowanie. Wirtualizacja pozapasmowa (często nazywana asymetryczną) wykorzystuje sterowniki oparte na hoście, które najpierw sprawdzają metadane (wskażając lokalizację pliku), a następnie umożliwiają hostowi dostęp do pliku bezpośrednio z zapisanej lokalizacji. Ta metoda nie wymaga buforowania na poziomie wirtualizacji.

Ogólne zalety wirtualizacji pamięci masowej obejmują:

- Migracja – dane można szybko przenosić przez lokalizacje przechowywania bez zakłócania bieżącego dostępu większości technologii do partycji wirtualnej.
- Wykorzystanie – Można zarządzać wykorzystaniem urządzeń pamięci masowej w celu ciągłego adresowania, w taki sam sposób, jak w przypadku wirtualizacji serwerów.
- Zarządzanie – większość hostów może używać pamięci masowej do centralnego zarządzania urządzeniem fizycznym.

Niektóre z wad obejmują:

- Brak standardów i interoperacyjności – wirtualizacja pamięci masowej to termin, a nie standard. Oznacza to również, że dostawcy nie współpracują łatwo.
 - Metadane – metadane dotyczące przechowywania i zarządzanie są niezbędne do funkcjonowania niezawodnego systemu, ponieważ istnieje korelacja pomiędzy lokalizacją logiczną i fizyczną.
- Wycofywanie — mapowanie wycofywania zwirtualizowanej infrastruktury z sieci z lokalizacji cyfrowych do fizycznych jest często mniej niż trywialne.

3.11.2 Wirtualizacja sieci

W dziedzinie informatyki wirtualizacja sieci integruje sprzęt, oprogramowanie i infrastrukturę sieciową z funkcjonalnościami sieciowymi w jedną wirtualną jednostkę administracyjną sieci. Wirtualizacja sieci wymaga wirtualizacji platformy, często połączonej z wirtualizacją zasobów. Wirtualizacja sieci jest definiowana jako zewnętrzna, połączona w jednostkę wirtualną lub wewnętrzną, co zapewnia kontenerom oprogramowania funkcjonalność podobną do sieci w jednym systemie.

Zgodnie z wewnętrzną definicją tego pojęcia, wirtualizacja pulpu i serwera zapewnia łączność sieciową zarówno hostowi i gościowi, jak i kilku gościom. Przełączniki wirtualne są rozpoznawane po stronie serwera jako część stasu wirtualizacji. Niemniej jednak, zewnętrzna koncepcja wirtualizacji sieci jest prawdopodobnie najczęściej używaną wersją. Wirtualne sieci prywatne (VPN), w których większość przedsiębiorstw obsługuje VPN, są od lat standardowymi komponentami zestawu narzędzi. Kolejną szeroko stosowaną definicją wirtualizacji sieci są wirtualne sieci LAN (VLAN). Sieci należy organizować wyłącznie według granic regionalnych w związku z rozwojem sieci, ponieważ 10-gigabitowy Ethernet to krótki czas.

Ogólne zalety wirtualizacji sieci obejmują:

Personalizacja dostępu – administratorzy mogą łatwo dostosować opcje dostępu i sieci, w tym ograniczanie przepustowości i jakość usług.

Konsolidacja — sieci wirtualne można połączyć w jedną sieć wirtualną, aby ogólnie uprościć zarządzanie.

Podobnie jak wirtualizacja serwerów, wirtualizacja sieci zwiększa złożoność, zwiększa wydajność i wymaga od administratorów większych umiejętności.

3.11.3 Wirtualizacja pulpu

Wirtualizacja komputerów stacjonarnych to paradymat oprogramowania, który przejmuje konwencjonalny model cienkiego klienta z chmury, ale ma na celu zapewnienie administratorom i użytkownikom końcowym tego, co najlepsze na świecie: hostingu i centralnego zarządzania w centrum danych maszyn wirtualnych, zapewniając jednocześnie użytkownikom końcowym najpełniejsze Doświadczenie na komputerze stacjonarnym.

Hostowana aplikacja Wirtualizacja jest identyczna z wirtualizacją hostowanego pulpu, która rozszerza doświadczenie użytkownika na cały pulpit. Usługi terminalowe firmy Microsoft, XenDesktop firmy Citrix i VDI firmy VMware są produktami komercyjnymi.

Zalety wirtualizacji komputerów stacjonarnych obejmują większość zastosowań wirtualizacyjnych, a także:

- Wysoka dostępność – przestoje związane z replikacją sieci i ustawieniami odporności na awarie może być zredukowany.
- Wydłużone cykle odświeżania – Większa pojemność serwerów i mniejsze wymagania dotyczące komputerów klienckich zwiększa ich żywotność. • Wiele komputerów stacjonarnych – użytkownicy mogą kontrolować wiele komputerów stacjonarnych z tego samego komputera klienckiego, który może wykonywać różne zadania.

Wirtualizacja komputerów stacjonarnych ma również wady, które są zbliżone do wirtualizacji serwerów. Dodatkową wadą jest to, że konsumenti muszą być połączeni z siecią, aby uzyskać dostęp do swoich wirtualnych pulpitów. W przypadku prac offline jest to kłopotliwe, a także zwiększa zapotrzebowanie na sieć w miejscu pracy.

3.11.4 Wirtualizacja serwera aplikacji

Wirtualizacja serwera aplikacji opisuje wybór za pomocą technik równoważenia obciążenia i architektury o wysokiej wartości dla usług hostowanych na serwerze aplikacji, które oferują te same usługi, co pojedynczy wirtualny serwer aplikacji. Jest to specjalny rodzaj wirtualizacji, którego celem jest zwiększenie wydajności wirtualnej pamięci masowej w porównaniu z emulacją innego środowiska. To inna wirtualizacja.

3.12 Wirtualizacja i przetwarzanie w chmurze

Przetwarzanie w chmurze, które jest elastyczne, skalowalne i często zmniejsza koszty i złożoność aplikacji, jest dziś jedną z najlepiej brzmiących i ekskluzywnych technologii.

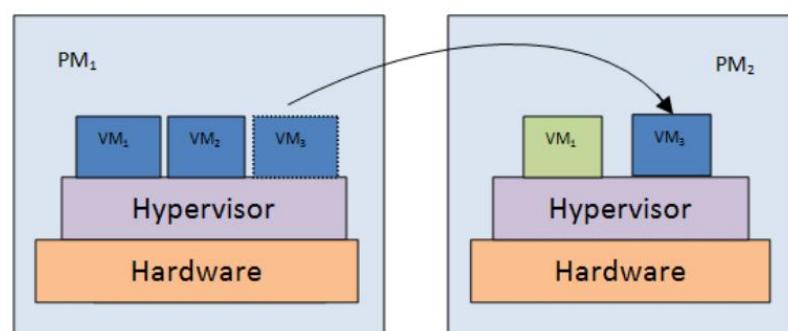
Wirtualizacja jest podstawową technologią wykorzystywaną w chmurze obliczeniowej. Wirtualizacja jest elementem przetwarzania w chmurze. W oparciu o wirtualizację w chmurze obciążenia można szybko wdrażać i skalować poprzez szybkie udostępnianie maszyn wirtualnych i fizycznych.

Chmury są postrzegane jako pula zwirtualizowanych zasobów, które są łatwe w użyciu i dostępne. Oprogramowanie jako model usługi, platforma jako model usług i infrastruktura jako model usług to trzy modele usług w chmurze. Oprogramowanie to model usług rozwoju usług, który gwarantuje, że konsumenti płacą za to, że nie są właścicielami aplikacji, z których korzystają. Platforma jako model usług zapewnia użytkownikom platformę do odpowiedniego tworzenia i konfigurowania oprogramowania. Model infrastruktury jako usługi to samozarządzający się model służący do kontrolowania i monitorowania zdalnych danych.

Dostawca chmury zarządza usługami sieciowymi, magazynowymi i obliczeniowymi. Przetwarzanie w chmurze zasadniczo oferuje dostęp do zasobów niezbędnych do prowadzenia różnych działań przy rosnących wymaganiach użytkowników. Ideą chmury obliczeniowej jest umożliwienie firmom zwiększenia wydajności sprzętu komputerowego, wykorzystania zasobów i elastyczności.

Wirtualizacja jest obecnie najodpowiedniejszą technologią. W przypadku przetwarzania w chmurze wirtualizacja odgrywa kluczową rolę, ponieważ umożliwia stopień dostosowania, ochrony, izolacji i zarządzania niezbędnego do świadczenia usług na żądanie.

Metoda konsolidacji ujawnia przypisane maszyny wirtualne do serwera fizycznego i wykorzystuje parametry takie jak minimalna (min) wymagana ilość zasobów dla każdej maszyny wirtualnej i maksymalny dopuszczalny zasób (maks.) przez każdą aplikację maszyny wirtualnej w celu określenia ilości zasobów maszyny wirtualnej. Te parametry wirtualizacji stanowią ważne ramy gwarantujące intelligentną dystrybucję zasobów pomiędzy różnymi aplikacjami (zwłaszcza gdy dostępny jest zróżnicowany zakres aplikacji maszyn wirtualnych w różnych preferencjach i zasobach). Metoda konsolidacji ma na celu rozmieszczenie maszyn wirtualnych w celu rozmieszczenia kombinacji aplikacji maszyn wirtualnych na każdym serwerze fizycznym, biorąc pod uwagę alokację zasobów dla każdej maszyny wirtualnej w oparciu o różne priorytety i powinowactwa zasobów. Skuteczność jednolitego programu ma korzystny wpływ na szczegółowość dystrybucji zasobów. Proces ten często nazywany jest konsolidacją serwerów, natomiast migracja maszyn wirtualnych polega na przeniesieniu instancji maszyn wirtualnych



RYSUNEK 3.16 Migracja maszyny wirtualnej

Live Migration przenosi działającą maszynę wirtualną z serwera fizycznego na inny, bez zakłócania dostępności maszyny wirtualnej dla użytkowników. Migracja na żywo. Celem migracji na żywo maszyny wirtualnej jest umożliwienie konserwacji lub aktualizacji bez przerwy na maszynie wirtualnej podczas migracji na maszynę wirtualną. Często nazywana jest płynną migracją na żywo, gdy użytkownik końcowy nie zauważa przestojów podczas cyklu migracji.

3.12.1 Plusy i minusy wirtualizacji

Większość firm pragnie unowocześniać swoje rozwiązania na maszynach wirtualnych wraz z rozwojem wirtualizacji. Niemniej jednak ważne jest rozważenie zalet i wad wirtualizacji przed wprowadzeniem jakichkolwiek ulepszeń. Zalety i wady systemów fizycznych i wirtualnych, z których każdy ma czas i lokalizację.

Technologia wirtualna ma oczywiste zalety, ma jednak kilka wad. Aby zdecydować, w jaki sposób najlepiej odpowiada ona wymaganiom firmy, podsumowaliśmy zalety i wady wirtualizacji.

3.12.1.1 Zalety wirtualizacji

Skalowalność

Maszyna wirtualna jest tak samo skalowalna jak każde inne rozwiązanie. Jedną z kluczowych zalet wirtualizacji jest możliwość integracji kilku systemów. Oferuje niewiarygodną elastyczność w porównaniu z systemem fizycznym i gołym metałem, co nie jest możliwe. Eластyczność ta ma bezpośredni wpływ na to, jak firmy mogą rozwijać się szybko i efektywnie. Wirtualizacja umożliwia migrację danych, aktualizację i natychmiastową poprawę wydajności do nowych maszyn wirtualnych w krótkim czasie.

Konsolidacja serwerów

Konstrukcja maszyn wirtualnych zastąpi prawie 10:1 maszyny fizyczne. Eliminuje to potrzebę stosowania komputerów fizycznych, zapewniając jednocześnie wydajne działanie systemów i specyfikacji. Taka konsolidacja zmniejsza koszty i wymaganą przestrzeń fizyczną dla systemów komputerowych.

Większa niezawodność systemu

Jednym z powodów zastosowania wirtualizacji jest jej zdolność do zapobiegania awariom systemu. Uszkodzenie pamięci spowodowane przez sterowniki systemowe i tym podobne to najczęstsze awarie uniemożliwiające maszynę wirtualną. Takie systemy opisują architekturę DMA w celu poprawy izolacji wejść/wyjść. Zapewnia zwiększone bezpieczeństwo i niezawodność.

Wirtualne stacje robocze

Wirtualizacja zapewnia globalną wszechstronność, umożliwiając uruchamianie wielu systemów na jednym komputerze w celu zdalnej obsługi systemów. Maszyna wirtualna zmniejsza także wielkość całego sprzętu i pulpitu.

3.12.1.2 Wady wirtualizacji

Programy wymagające sprzętu fizycznego

Wirtualizacja nie sprawdza się dobrze w przypadku aplikacji wymagających sprzętu fizycznego. Przykładem jest coś wykorzystującego klucz sprzętowy lub inny podłączony sprzęt. Ponieważ program musi być elementem fizycznym, wirtualizacja spowodowałaby większy ból głowy niż pozostawanie w systemie fizycznym.

Jakość wydajności może się obniżyć

Jeśli uruchomisz aplikację, która wykorzystuje pamięć RAM lub procesor, wirtualizacja może spowodować opóźnienie wydajności. Maszyna wirtualna działa warstwowo w swoich systemach hostingowych, więc każda operacja o ekstremalnej wydajności będzie miała zmniejszoną wydajność, jeśli nie użyjesz jednej aplikacji lub serwera. Wadą wirtualizacji jest to, że wiele aplikacji może działać na małych serwerach fizycznych, dlatego trudno jest umieścić jednego hosta na jednym serwerze.

Testowanie jest krytyczne

Celem IT jest tak naprawdę osiągnięcie celów Twojej firmy, tak abyś nie miał nieprzetestowanej platformy oprogramowania dla swojej firmy. Jest to szczególnie prawdziwe w przypadku wirtualizacji, ponieważ nie działa tak, jakby można było ją wyłączyć i włączyć ponownie. System, który działa sprawnie, może już wykonywać wirtualizację, co prowadzi do błędów i potencjalnej straty czasu i kosztów.

Przed przełączeniem na maszynę wirtualną zawsze sprawdź.

Nieoczekiwane wydatki

Początkowo mogłoby się wydawać, że wirtualizacja pozwala zaoszczędzić trochę pieniędzy. Jest to jednak proces, który należy zakończyć i wykonać poprawnie po raz pierwszy. Wydasz więcej, niż początkowo planowałeś, aby uwzględnić tą dbałość o czas i szczegóły. Zanim podejmiesz decyzję, przejrzyj narzędzia i systemy zarządzania, które mogą być potrzebne do przejścia na maszynę wirtualną.

Dane mogą być zagrożone

Twoje dane są hostowane w sieci innej firmy, podczas gdy działają na instancjach wirtualnych na współdzielonych zasobach sprzętowych. Może to narazić dane na zagrożenia lub nieautoryzowany dostęp. Problemem jest, jeśli rozwiązanie bezpieczeństwa Twojego usługodawcy nie zabezpiecza Twojej wirtualnej instancji i danych. Szczególnie w przypadku wirtualizacji pamięci masowej jest to prawda.

Szybka skalowalność jest wyzwaniem

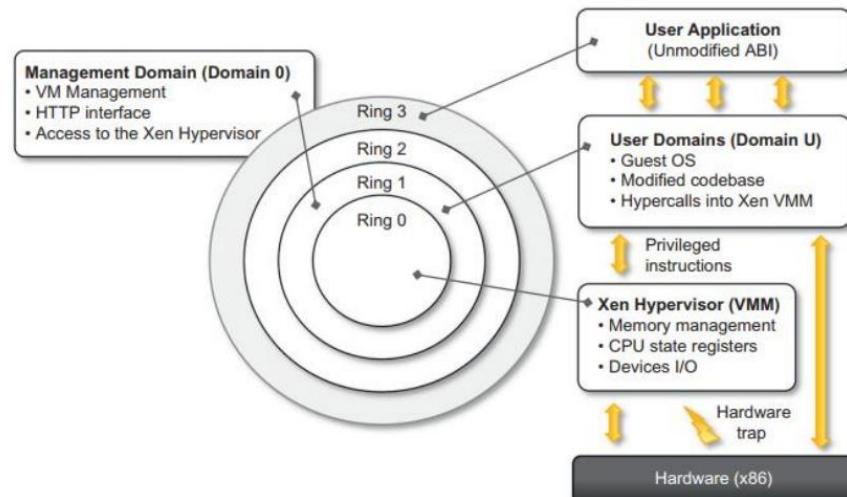
Skalowanie na wirtualizacji jest ograniczone czasowo, dlatego nie ma znaczenia, czy ma zostać osiągnięte w tak krótkim czasie. Dzięki fizycznej konfiguracji można łatwo zbudować nowy sprzęt i nową skalę, nawet jeśli wiąże się to z pewnymi początkowymi problemami. Wirtualizacja może być żmudnym zadaniem, ponieważ zapewnia całe niezbędne oprogramowanie, ochronę, odpowiednią pamięć masową i dostępność zasobów. Zajmuje to więcej czasu, niż można by zaplanować, ponieważ w sprawę zaangażowany jest zewnętrzny dostawca. Jednakże dalszymi problemami w zarządzaniu są także dodatkowe koszty związane ze wzrostem wykorzystania kosztów Niezamierzony rozrost serwerów

Niezamierzony zwiększenie zasięgu serwerów jest dużym problemem zarówno dla administratorów, jak i użytkowników. Niektóre z problemów zgłaszanych przez osoby w dziale obsługi dotyczą rozszerzeń aplikacji. Instalacja serwera fizycznego wymaga czasu i zasobów, natomiast w ciągu kilku minut można zbudować serwer wirtualny. Użytkownicy za każdym razem budują nowe serwery, zamiast ponownie używać tego samego serwera wirtualnego, ponieważ pozwala im to zacząć od nowa. Administrator serwera, który ma pięć lub sześć serwerów, ma do obsługi 20 serwerów wirtualnych. Może to prowadzić do dużych trudności w płynnym działaniu, a wymuszone zakończenie działania niektórych serwerów może również skutkować utratą danych.

3.13 Przykłady technologii

3.13.1 Xen: parawirtualizacja

Xen to parawirtualny hiperwizor typu open source. Najbardziej popularną aplikacją jest parawirtualizacja. Aby zapewnić pełną zgodność, Xen został rozszerzony o wirtualizację wspomaganą sprzętowo. Pozwala to na wysoką wydajność w systemie operacyjnym gościa. Można to prawdopodobnie osiągnąć poprzez wyeliminowanie utraty wydajności podczas wykonywania instrukcji wymagających znacznego nadzoru i zmianę części systemu operacyjnego gościa Xen pod kątem wykonywania tych instrukcji. Obsługuje to szczególnie architekturę x86, najpowszechniejszą architekturę na komputerach towarowych i serwerowych.



RYSUNEK 3.16 Architektura Xen i zarządzanie systemem operacyjnym gościa.

(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming” autorstwa Rajkumara Buyyi)

Rysunek powyżej definiuje Xen i jego mapowanie na klasyczny paradygmat uprawnień x86. Hiperwizor Xen obsługuje system oparty na Xen, który działa w najbardziej komfortowym trybie i zachowuje dostęp systemu operacyjnego gościa do niezbędnego sprzętu. System operacyjny gościa działający pomiędzy domenami, reprezentujący instancje maszyn wirtualnych.

Jednakże pewien program sterujący z uprzywilejowanym dostępem do hosta i zarządzaniem wszystkimi innymi systemami operacyjnymi gościa działa w określonej domenie zwanej Domeną 0. Jest to jedyny program ładowany po tym, jak w pełni uruchomiony menedżer maszyny wirtualnej hostuje serwer HTTP, który dostarcza żądania do zbuduj, skonfiguruj i zakończ maszynę wirtualną. Komponent udostępnia pierwszą wersję rozwiązania IaaS (Infrastructure-as-a-Service), wspólnego menedżera maszyn wirtualnych (VMM). Program jest niezbędny dla komputerów pracujących w chmurze.

Różne implementacje x86 umożliwiają cztery różne poziomy bezpieczeństwa, zwane pierścieniami, tj. Pierścień 0, Pierścień 1, Pierścień 2 i Pierścień 3.

Tutaj Pierścień 0 jest poziomem najbardziej uprzywilejowanym, a Pierścień 3 jest poziomem mniej uprzywilejowanym. Prawie każdy system operacyjny, z wyjątkiem OS/2, wykorzystuje tylko dwa różne poziomy, tj. pierścień 3 dla programu użytkownika i systemu operacyjnego nieuprzywilejowanego, pierścień 0 dla kodu jądra i. Daje Xenowi możliwość osiągnięcia parawirtualizacji. Umożliwia to zarządzanie interfejsem binarnym aplikacji (ABI) w niezmienionej postaci, a tym samym przejście z punktu widzenia aplikacji na rozwiązania zwirtualizowane Xen.

Struktura zestawu instrukcji x86 umożliwia wykonanie kodu w pierścieniu 3 w celu przejścia do pierścienia 0 (tryb jądra). Taka operacja jest wykonywana na poziomie sprzętowym i dlatego może prowadzić do TRAP lub cichej awarii w zwirtualizowanym systemie, uniemożliwiając w ten sposób ogólne działanie systemu-gosia w pierścieniu 1.

Teoretycznie warunek ten istnieje poprzez podzbior wywołań systemowych. Wdrożenie systemu operacyjnego wymaga modyfikacji, a wszystkie krytyczne wywołania systemowe wymagają ponownej implementacji za pomocą hiperwywołań, aby wyeliminować tę sytuację. W tym przypadku hiperwywołania to specjalne wywołania udostępniane przez interfejs maszyny wirtualnej Xen (VM), a hiperwizor Xen wydaje się uzyskiwać, zarządzać i zwracać kontrolę za pomocą dostarczonego modułu obsługi do systemu operacyjnego gościa.

Parawirtualizacja wymaga przejścia na podstawowy kod systemu operacyjnego, tak aby w środowisku opartym na xen żaden system operacyjny gościa nie był dostępny dla wszystkich systemów operacyjnych. Ten warunek ma na celu zapobieganie swobodnej wirtualizacji wspomaganej sprzętowo, która wymaga, aby hypervisor działał w Ring 1, a system operacyjny gościa w Ring 0. W ten sposób Xen wykazuje pewne wady w odniesieniu do starszego sprzętu i starszego systemu operacyjnego.

Parawirtualizacja wymaga przejścia na podstawowy kod systemu operacyjnego, tak aby w środowisku opartym na xen żaden system operacyjny gościa nie był dostępny dla wszystkich systemów operacyjnych. Ten warunek ma na celu zapobieganie swobodnej wirtualizacji wspomaganej sprzętowo, która wymaga, aby hypervisor działał w Ring 1, a system operacyjny gościa w Ring 0. W ten sposób Xen wykazuje pewne wady w odniesieniu do starszego sprzętu i starszego systemu operacyjnego.

W rzeczywistości nie można ich w odpowiedzialny sposób zmienić, aby uruchamiały pierścień 1, ponieważ ich baza kodowa jest nieosiągalna, a podstawowy sprzęt nie obsługuje obecnie uruchamiania go w trybie bardziej uprzywilejowanym niż pierścień 0. System operacyjny typu open source, taki jak Linux, można łatwo zaktualizować ponieważ ma otwarty kod, a Xen zapewnia pełną obsługę wirtualizacji, podczas gdy komponenty systemu Windows zasadniczo nie są zgodne z Xen, dopóki nie będzie dostępna wirtualizacja wspomagana sprzętowo.

Wraz z wprowadzeniem nowych wersji systemu operacyjnego problem został rozwiązany, a nowy sprzęt będzie obsługiwał wirtualizację x86.

3.13.2 VMware: pełna wirtualizacja

W pełnej wirtualizacji główny sprzęt jest duplikowany i udostępniany systemowi operacyjnemu gościa, który nie jest świadomy abstrakcji i żadnych kryteriów do zmiany. Technologia VMware opiera się na pełnej wirtualizacji. Firma VMware wdraża pełną wirtualizację w środowisku stacjonarnym przy użyciu hiperwizora typu II lub w środowisku serwerowym przy użyciu hiperwizora typu I. We wszystkich przypadkach może wystąpić pełna wirtualizacja niewrażliwych instrukcji.

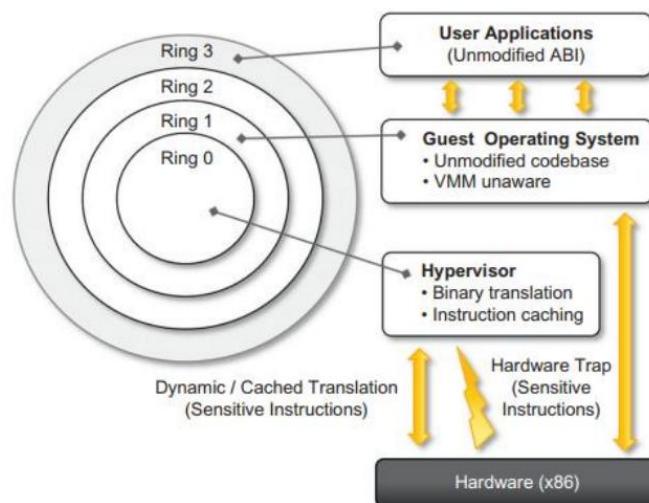
można osiągnąć bezpośrednio, oraz tłumaczenie binarne wrażliwych instrukcji lub pułapek sprzętowych, co pozwala na wirtualizację architektury takiej jak x86.

3.13.3 Pełna wirtualizacja i tłumaczenie binarne

VMware jest często używany, ponieważ architektury x86 są zasadniczo zwirtualizowane i uruchamiają hiperwizory w niezmienionej formie na górze. Pełna wirtualizacja jest możliwa poprzez wdrożenie wirtualizacji wspomaganej sprzętowo poprzez obsługę sprzętu. Niemniej jednak wcześniej systemy gościa x86 można było wdrażać jedynie z dynamiczną translacją binarną bez modyfikacji w środowisku zwirtualizowanym.

Ponieważ instrukcja wrażliwa nie jest instrukcją uprzywilejowaną, pierwsze twierdzenie o wirtualizacji nie jest spełnione przy projektowaniu architektury x86. Ze względu na tę konkretną czynność instrukcje te nie są zaimplementowane w Pierścieniu 0, jak to zwykle bywa w środowisku wirtualizacji, w którym system operacyjny gościa jest uruchamiany w Pierścieniu 1. Zasadniczo generowana jest pułapka i wykorzystywany jest proces, w którym rozwiązanie dla x86 wyróżnia się. W przypadku dynamicznej translacji binarnej pułapka jest konwertowana na serię instrukcji, które określają ten sam cel bez robienia wyjątków. Dlatego też odpowiednie instrukcje są zapisywane w celu poprawy wydajności, stąd tłumaczenie nie jest już konieczne przy kolejnych spotkaniach z tymi samymi instrukcjami. Rysunek przedstawiający to znajduje się poniżej.

Główną zaletą tej metody jest to, że goście mogą działać w niezmodyfikowanej formie w środowisku zwirtualizowanym, niezbędnym komponencie systemu operacyjnego, który nie posiada kodu źródłowego. W przypadku pełnej wirtualizacji tłumaczenie binarne jest przenośne. Oprócz tłumaczenia instrukcji w czasie wykonywania, dodatkowe obciążenie nie występuje w innych formach, takich jak parawirtualizacja lub wirtualizacja wspomagana sprzętowo. Dla porównania, tłumaczenie binarne odbywa się tylko w podzbiorze zestawów instrukcji, podczas gdy inne są wykonywane bezpośrednio na głównym sprzęcie. To w jakiś sposób zmniejsza wpływ na wydajność tłumaczenia binarnego.



RYSUNEK 3.17 Model referencyjny pełnej wirtualizacji.

(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming” autorstwa Rajkumara Buyyi)

Zalety tłumaczenia binarnego

- Ta metoda wirtualizacji zapewnia maszynom wirtualnym najlepszą izolację i bezpieczeństwo.
- Wiele systemów gościa będzie faktycznie działać jednocześnie na tym samym sprzęcie w bardzo krótkim czasie izolowany sposób.
- Jest to stosowane tylko bez wsparcia sprzętowego lub systemu operacyjnego w wirtualizacji wrażliwych instrukcji i uprzywilejowanych instrukcji.

Wady tłumaczenia binarnego

- Jest to czasochłonne w czasie wykonywania.
- Osiąga wysoką wydajność narzutową.

- Pamięć podrzeczna kodu służy do przechowywania najczęściej używanych tłumaczonych instrukcji poprawia wydajność, ale zwiększa wykorzystanie pamięci wraz z kosztami sprzętu. • W architekturze x86 wydajność pełnej wirtualizacji wynosi 80-95% maszyny hosta.

3.13.5 Rozwiązań wirtualizacyjnych

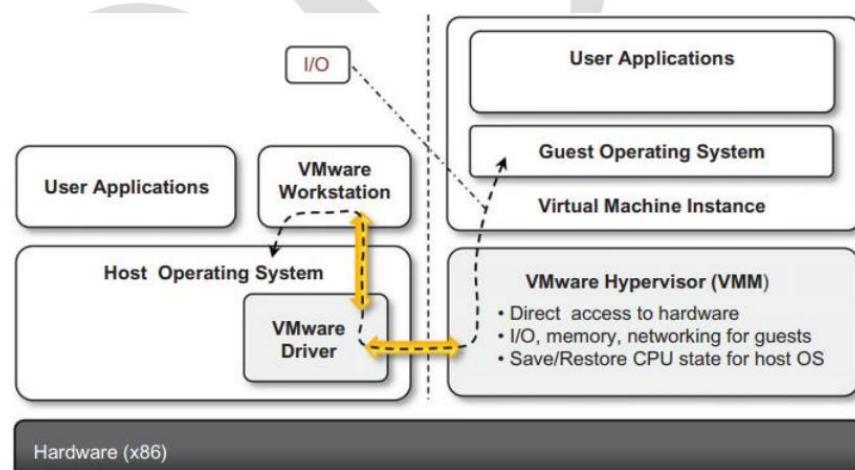
VMware jest pionierem w dziedzinie rozwiązań do wirtualizacji i infrastruktury chmurowej, które pozwalają nam ponad 350 000 przedsiębiorstwom i klientom odnieść sukces w erze chmury. VMware upraszcza złożoność całego centrum danych i umożliwia klientom korzystającym z rozwiązań definiowanych programowo dla centrów danych przekształcenie się w hybrydową chmurę obliczeniową i mobilną przestrzeń roboczą.

3.13.5 Wirtualizacja użytkownika końcowego (komputera stacjonarnego).

Technologie wirtualizacji komputerów stacjonarnych i aplikacji VMware zapewniają działowi IT usprawnioną metodę dostarczania, zabezpieczania i utrzymywania komputerów stacjonarnych i aplikacji z systemami Windows i Linux na miejscu lub w chmurze, redukując koszty i zapewniając użytkownikom końcowym możliwość działania w dowolnym miejscu i czasie. VMware Workstation umożliwia użytkownikom jednocześnie uruchamianie różnych systemów operacyjnych na tym samym komputerze PC z systemem Windows lub Linux. Twórz prawdziwe maszyny wirtualne z systemem Linux i Windows, a także inne środowiska komputerów stacjonarnych, serwerów i tabletów, wraz z konfiguracją sieci wirtualnej i symulacją sieci, do wykorzystania przy opracowywaniu kodu, architektur rozwiązań i testowaniu aplikacji oraz demonstracjach produktów i nie tylko.

VMware Fusion umożliwia użytkownikom komputerów Mac uruchamianie systemu Windows na komputerze Mac wraz z setkami kilku innych systemów operacyjnych obok siebie, bez konieczności ponownego uruchamiania. Fusion jest wystarczająco łatwy dla użytkowników domowych i wystarczająco wydajny dla ekspertów IT, programistów i firm. Oprócz skonfigurowania niezależnego środowiska komputerowego te dwa produkty umożliwiają systemowi operacyjnemu gościa wykorzystanie zasobów komputera hosta (urządzenia USB, udostępnianie folderów i integracja z graficznym interfejsem użytkownika (GUI) systemu operacyjnego hosta).

Rysunek przedstawia opis architektury systemów.



RYSUNEK 3.18 Architektura stacji roboczej VMware.

(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming” autorstwa Rajkumara Buyyi)

Aplikacja zainstalowana w systemie operacyjnym gościa tworzy środowisko wirtualizacji, które umożliwia tym systemom operacyjnym pełną wirtualizację sprzętu, na którym się opiera. Odbywa się to poprzez instalację w systemie operacyjnym hosta specjalnego sterownika, który zapewnia dwie główne usługi:

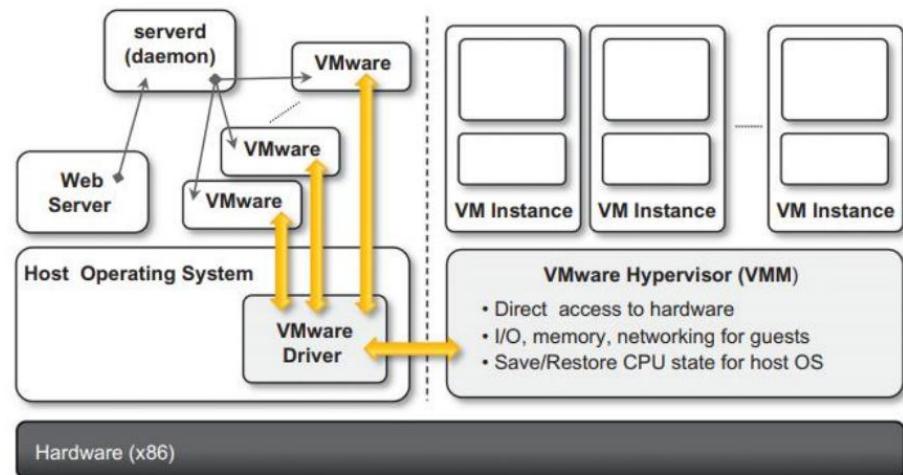
- Wykorzystuje menedżera maszyn wirtualnych, z którego można korzystać w trybie uprzywilejowanym.
- Aplikacja VMware oferuje paski do przetwarzania określonych żądań wejścia/wyjścia poprzez późniejsze przekazywanie tych żądań za pośrednictwem wywołań systemowych do systemu operacyjnego hosta.

Architektura ta, znana również jako Hosted Virtual Machine Architecture, może zarówno oddzielać instancje maszyn wirtualnych w przestrzeni pamięci aplikacji, jak i zapewniać przyzwoitą wydajność, ponieważ zaangażowanie aplikacji VMware jest niezbędne tylko w przypadku instrukcji, na przykład urządzeń we/wy, które wymagają translacji binarnej. Menedżer maszyny wirtualnej zarządza procesorem i MMU oraz zmienia funkcjonalność operacyjną procesora i MMU w systemie operacyjnym hosta. Obrazy maszyn wirtualnych są przechowywane w katalogu systemu plików hosta i że zarówno VMware, jak i VMware Fusion umożliwiają tworzenie, uruchamianie, tworzenie migawek i cofanie działań operacyjnych nowych obrazów poprzez przywrócenie poprzedniego stanu maszyny wirtualnej.

VMware Player, VMware ACE i VMware ThinApp to dodatkowe technologie istotne z punktu widzenia wirtualizacji środowisk obliczeniowych użytkowników końcowych. VMware Player to ograniczona wersja VMware Workstation, która umożliwia tworzenie i emulację maszyn wirtualnych w środowisku operacyjnym takim jak Windows lub Linux. VMware ACE jest taki sam jako VMware Workstation do opracowywania maszyn wirtualnych objętych polityką w celu zapewnienia bezpiecznego wdrażania klienckich środowisk wirtualnych na komputerach użytkowników końcowych. VMware ThinApp to rozwiązanie do wirtualizacji aplikacji. Oferuje niezależne środowisko programistyczne, które zapobiega zmianom wynikającym z wersji i niekompatybilnych aplikacji. Identyfikuje zmiany w środowisku operacyjnym poprzez zainstalowanie określonej aplikacji i przechowuje je w pakiecie, który można uruchomić za pomocą VMware ThinApp wraz z aplikacją binarną.

3.13.6 Wirtualizacja serwerów

GSX Server to zwirtualizowany system serwerów Windows i Linux opracowany i dystrybuowany przez VMware, spółkę zależną EMC Corporation. Program promuje zdalne zarządzanie, udostępnianie i standaryzację aplikacji. Rysunek przedstawia architekturę serwera VMware GSX.



RYSUNEK 3.19 Architektura serwera VMware GSX.

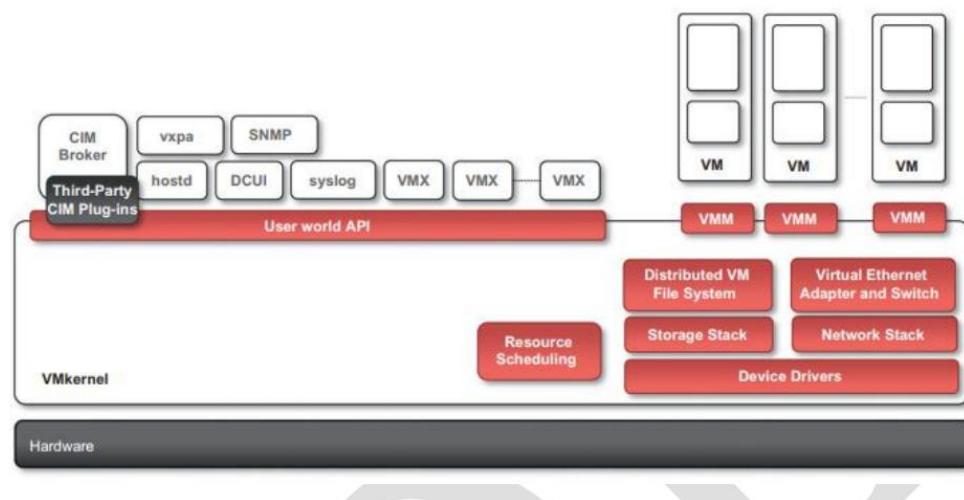
(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming” autorstwa Rajkumara Buyyi)

Serwer VMware GSX przekształca komputery w zbiór maszyn wirtualnych. Systemy operacyjne i frameworki oddzielają różne maszyny wirtualne na jednym urządzeniu sprzętowym. VMware GSX Server oferuje szeroką obsługę sprzętu odziedziczonego dla urządzenia z hosta. Niezawodna architektura i możliwości integracji produktu VMware GSX Server sprawiają, że środowiska hostów Windows i Linux są proste w obsłudze i zarządzaniu. Program hosta dla VMware GSX Server pomaga wdrażać, monitorować i kontrolować aplikację oraz wiele serwerów na maszynach wirtualnych działających zdalnie.

Architektura jest przeznaczona przede wszystkim do wirtualizacji serwerów WWW. Serwer zwany procesem demonem monitoruje i zarządza aplikacjami dla VMware. Następnie sterownik VMware w systemie operacyjnym hosta łączy te programy zinstancjami maszyny wirtualnej.

VMM służy do obsługi instancji maszyn wirtualnych, jak zdefiniowano wcześniej. Żądania użytkowników dotyczące zarządzania i udostępniania maszyn wirtualnych są przekierowywane z serwera WWW za pośrednictwem serwera za pośrednictwem VMM.

Metodę opartą na hypervisorze zademonstrowano w VMware ESX Server i jego ulepszonej edycji VMWare ESXi Server. Każdy z nich można zamontować na serwerach typu bare metal i zapewniać usługi zarządzania maszynami wirtualnymi. Te dwa produkty oferują podobne usługi, ale różnią się architektura wewnętrzna, szczególnie w organizacji jądra hypervisorza. Zaktualizowana wersja systemu operacyjnego Linux, która umożliwia dostęp do hypervisorza za pośrednictwem konsoli serwisowej, została zaimplementowana przez VMware ESX. VMware ESXi wprowadza niewiarygodnie cienką warstwę systemu operacyjnego i zastępuje konsolę usług interfejsami i narzędziami zdalnego monitorowania, znacznie zmniejszając w ten sposób rozmiar kodu hypervisorza i zużycie pamięci. Rysunek przedstawia architekturę VMware ESXi.



RYSUNEK 3.20 Architektura serwera VMware ESXi

(Odniesienie do „Mastering Cloud Computing Foundations and Applications Programming” autorstwa Rajkumara Buyyi)

3.14 Microsoft Hyper-V

Hyper-V to hypervisor, oprogramowanie do wirtualizacji firmy Microsoft, które może budować i hostować korporacyjne maszyny wirtualne na systemach x86-64, takich jak komputery stacjonarne i serwery. Zasadniczo hypervisor to oprogramowanie, które umożliwia obsługę kilku serwerów wirtualnych (maszyn gościnnych) na serwerze fizycznym (hostie). Serwer Hyper-V można skonfigurować tak, aby udostępniał maszyny wirtualne jednej lub większej liczbie sieci. W systemie Windows Server 2008 po raz pierwszy uruchomiono funkcję Hyper-V

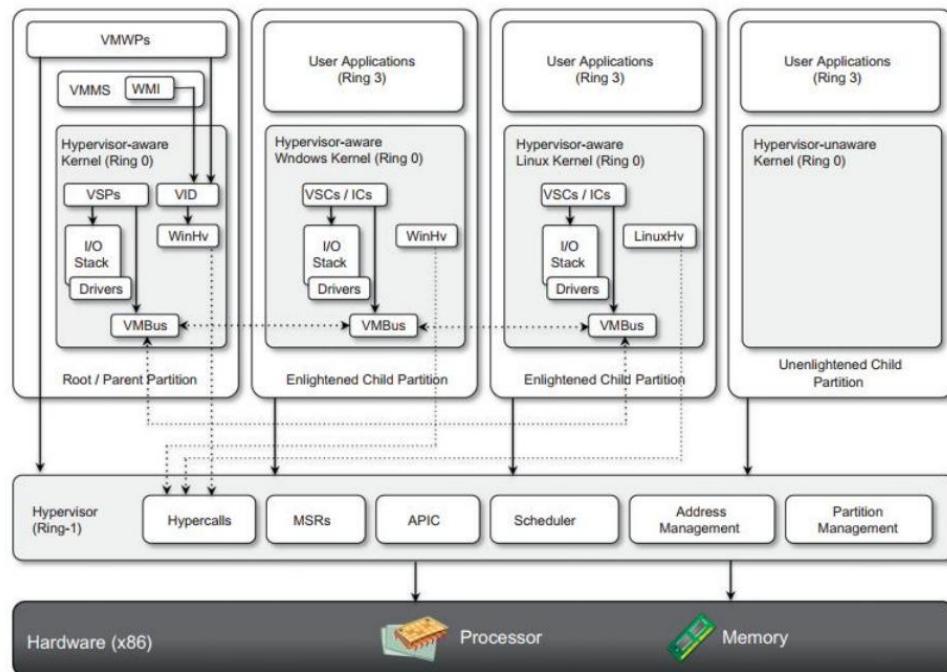
3.14.1 Architektura

Jeśli chodzi o partycję, Hyper-V implementuje izolację maszyny wirtualnej. Partycja obsługiwana przez hypervisorza jest logiczną jednostką izolującą, w której pracuje każde urządzenie operacyjne gościa. W instancji hypervisorza musi istnieć co najmniej jedna partycja nadziedzona z włączonym systemem Windows Server (2008 i nowsze). W partycji nadziedznej działa oprogramowanie wirtualizacyjne i ma bezpośredni dostęp do urządzeń sprzętowych. Partycja nadziedzona tworzy partycje podrzędne zawierające systemy operacyjne gościa. Partycja nadziedzona korzysta z interfejsu API hypercall, platformy aplikacji, do tworzenia partycji podrzędnych udostępnianych funkcji Hyper-V.

Partycja podrzędna nie ma dostępu ani nie ma faktycznych przerwań do procesora fizycznego. Zamiast tego procesor ma widok wirtualny i działa pod wirtualnym adresem gościa, który w rzeczywistości może nie być całą wirtualną przestrzenią adresową, w zależności od konfiguracji hipervizora. Funkcja Hyper-V pokaże tylko podzbiór procesorów na każdej partycji, w zależności od konfiguracji maszyny wirtualnej. Hipervizor zarządza przerwaniami procesora za pomocą logicznego kontrolera przerwań syntetycznych (SyNIC) dla odpowiedniej partycji.

Partycje podrzędne zapewniają wirtualny widok zasobów maszyn wirtualnych, a nie bezpośredni dostęp do źródeł sprzętowych. Każde żądanie urządzenia wirtualnego jest wysyłane do urządzeń partycji nadziedznej, które obsługują żądania, za pośrednictwem VMBus. VMBus jest kanałem logicznym umożliwiającym komunikację pomiędzy sekcjami. Odpowiedź jest również przekazywana za pośrednictwem VMBus. Jeśli urządzenia partycji nadziedznej są także urządzeniami wirtualnymi, są one kierowane do partycji nadziedznej, gdzie uzyskiwany jest dostęp do urządzeń fizycznych. Partycje nadziedzne obsługują dostawcę usług wirtualizacji (VSP), który łączy się z VMBus i obsługuje żądania dostępu do systemu z partycji podrzędnych. Urządzenia partycji wirtualnych dziecka uruchamiają plik

Klient usługi wirtualizacji (VSC) wewnętrznie, przekierowujący program do dostawców VSP na elemencie nadzorowanym VMBus. Dla systemu gościa cały ten proces jest przejrzysty.



RYSUNEK 3.17 Architektura Microsoft Hyper-V

(Odniesienie z „Opanowanie podstaw i aplikacji przetwarzania w chmurze Programowanie” Rajkumara Buyyi)

3.15 Podsumowanie

Wirtualizacja stanowi podstawę wielu technologii i koncepcji. Popularnym źródłem wirtualizacji jest możliwość zademonstrowania, przy użyciu pewnego rodzaju warstwy emulacji lub abstrakcji, danego środowiska wykonawczego, niezależnie od tego, czy jest to oprogramowanie, obiekt pamięci masowej, połączenie sieciowe czy zdalny pulpit. W budowaniu usług i infrastruktury chmurowej wszystkie te zasady odgrywają zasadniczą rolę, gdzie sprzęt, infrastruktura informatyczna, aplikacje i usługi są udostępniane na żądanie za pośrednictwem Internetu lub zwykle poprzez połączenie sieciowe.

3.16 Pytania przeglądowe

1. Zdefiniuj wirtualizację. Jakie są zalety wirtualizacji?
2. Jaka jest charakterystyka środowisk zwirtualizowanych?
3. Opisać klasyfikację lub taksonomię wirtualizacji na różnych poziomach.
4. Omów model referencyjny maszyny wirtualizacji wykonawczej.
5. Jakie są techniki wirtualizacji sprzętu?
6. Wymień i omów różne typy wirtualizacji.
7. Jakie są korzyści z wirtualizacji w kontekście przetwarzania w chmurze?
8. Jakie są wady wirtualizacji?
9. Czym jest Xen? Omów jego elementy do wirtualizacji.
10. Omów model referencyjny pełnej wirtualizacji.
11. Omów referencyjny model parawirtualizacji.
12. Omów architekturę Hyper-V. Omów jego zastosowanie w chmurze obliczeniowej

3.17 Odniesienie do dalszej lektury

1. Opanowanie podstaw przetwarzania w chmurze i programowania aplikacji Rajkumar Buyya, Christian Vecchiola, S. Publikacje Thamarai Selvi MK ISBN: 978-0-12-411454-8
2. Koncepcje, technologia i architektura przetwarzania w chmurze Thomas Erl, Zaigham Mahmooda i Ricardo Puttiniego, Seria technologii serwisowych Prentice Hall ISBN-10 : 9780133387520 ISBN-13 : 978-0133387520

3. Przetwarzanie rozproszone i przetwarzanie w chmurze: od przetwarzania równoległego do Internetu rzeczy, wydanie pierwsze, Kai Hwang Jack Dongarra Geoffrey Fox ISBN-10: 9789381269237 ISBN-13: 978-9381269237

Idol

Rozdział 2

Rozdział 1

2.1.0. Cel

2.1.1. Wprowadzenie do architektury przetwarzania w chmurze

 2.1.1.1. Architektura

 2.1.2. Podstawowe pojęcia i modele

 2.1.2.1. Role i granice

 2.1.2.1.1. Dostawca chmury

 2.1.2.1.2. Konsument chmury

 2.1.2.1.2.1. Właściciel usługi w chmurze

 2.1.2.1.4. Administrator zasobów chmury

 2.1.4. Charakterystyka chmur

 2.1.5. Modele dostarczania w chmurze.

 2.1.5.1. Użycie na żądanie

 2.1.5.2. Wszechobecny dostęp

 2.1.5.2.1. Wielodostępność

 2.1.5.4. Elastyczność

 2.1.5.5. Zmierzone użycie

 2.1.5.6. Odporność

 2.1.6. Modele wdrażania w chmurze

 2.1.6.1. Infrastruktura jako usługa (IaaS)

 2.1.6.2. Platforma jako usługa (PaaS)

 2.1.6.2.1. Oprogramowanie jako usługa (SaaS)

 2.1.7. Ekonomika chmury

 2.1.7.1. Chmury publiczne

 2.1.7.2. Chmury społeczności

 2.1.7.2.1. Chmury prywatne

 2.1.7.4. Chmury hybrydowe

 2.1.8. Otwarte wyzwania.

2.1.0. Cel

Po przejściu tego rozdziału będziesz mógł:

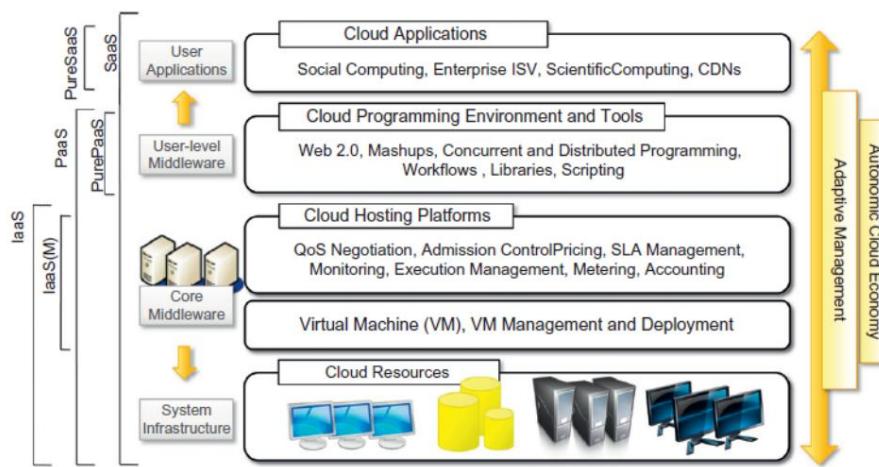
- Architektura przetwarzania w chmurze
- Wstęp
- Podstawowe pojęcia i modele
- Role i granice
- Charakterystyka chmur
- Modele dostarczania w chmurze
- Modele wdrażania w chmurze
- Ekonomika chmury
- Otwarte wyzwania.

2.1.1. Prolog do architektury przetwarzania w chmurze:

Przetwarzanie rozproszone wspiera każdą administrację IT, która zostanie pochłonięta jako narzędzie i przeniesiona przez system, najwyraźniej przez Internet. Taki obraz obejmuje bardzo różne perspektywy: ramy, etapy zaawansowania, zastosowanie i administrację.

2.1.1.1. Inżynieria

Można sobie wyobrazić uporządkowanie wszystkich solidnych podstawa przetwarzania rozproszonego w warstwowy widok obejmujący cały stos (patrz rysunek 2.1.1), od maszyn sprzętowych po struktury programistyczne. Zasoby w chmurze są przystosowane do elastycznego „pobierania rejestracji” wymaganego do oferowania rodzajów pomocy. Często warstwa ta jest realizowana przy wykorzystaniu centrum danych, w którym ułożone są razem setki i tysiące koncentratorów. Struktury chmurowe mają zwykle heterogeniczny charakter pod względem szczegółów, które często są tworzone przez różne zasoby, podobne do pęczków, a nawet zorganizowanych komputerów PC. Ponadto podstawą mogą być struktury baz danych i inne narzędzia służące do administrowania gromadzeniem zapasów.



Rysunek 2.1.1 Architektura przetwarzania w chmurze

(Odniesienie : Cloud Computing (konceptje, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

Fizyczna struktura jest realizowana przez centralne oprogramowanie pośredniczące, którego przeznaczeniem jest elastyczne dostosowywanie warunków działania aplikacji i najlepsze wykorzystanie zasobów. W absolutie

na samym dole stosu innowacje wirtualizacyjne zwykle zapewniają dostosowywanie warunków środowiska wykonawczego, izolowanie aplikacji, piaskownicę i charakter administracji.

Na tym poziomie najczęściej wykorzystywana jest wirtualizacja sprzętu. Hypervisory realizują pulę właściwości i odkrywają rozproszoną strukturę w postaci wielu maszyn wirtualnych. Pochłaniając innowacje w zakresie maszyn wirtualnych, prawdopodobnie nastąpi niewiarygodny podział zasobów sprzętowych, takich jak procesor i pamięć, oraz wirtualizacja jawnych gadżetów, spełniając w ten sposób potrzeby klientów i aplikacji. Celowi temu regularnie towarzyszą podejścia do wirtualizacji pojemności i systemu, które umożliwiają całkowitą wirtualizację i kontrolę platformy. niezawodny z dokładną pomocą oferowaną kompletnym klientom, regularnie stosowane są inne procedury wirtualizacji; na przykład wirtualizacja na poziomie rozwoju pomaga w tworzeniu przenośnych warunków wykonawczych, w których aplikacje są często uruchamiane i kontrolowane. Sytuacja ta w większości sugeruje, że aplikacje udostępniane w chmurze można tworzyć z wykorzystaniem wybranej innowacji lub języka programowania na wzór Java, .NET czy Python. w tym przypadku klient nie musi montować ramy z odsłoniętego metalu. Podstawa kadry kierowniczej jest główną motywacją stojącą za oprogramowaniem pośredniczącym centrum, które leży u podstaw umiejętności takich jak kompromis w zakresie standardów administracyjnych, kontrola afirmacji, wykonywanie zarządu i sprawdzanie, księgowość i pobieranie opłat.

2.1.2. Podstawowe idee i modele

Połączenie etapów i zasobów obsługujących chmurę jest zazwyczaj delegowane w ramach rozwiązania Infrastructure-as-a-Service (IaaS). podzielimy różne instancje IaaS na dwie klasyfikacje: różne przenoszą zarówno warstwę administracyjną, jak i następnie strukturę fizyczną; inne przekazują tylko warstwę administracyjną (IaaS (M)). w tym kolejnym przypadku warstwa administracyjna jest zwykle połączona z różnymi rozwiązaniami IaaS, które zapewniają fizyczne podstawy i zwiększa ich wartość.

Rozwiązania IaaS są odpowiednie do planowania podstaw ramowych, jednakże oferują ograniczone rodzaje pomocy przy tworzeniu aplikacji. Taką pomoc zapewniają warunki i urządzenia programistyczne w chmurze, które tworzą warstwę handlową, oferując klientom etap udoskonalania aplikacji. Zakres instrumentów obejmuje interfejsy internetowe, urządzenia linii zamówień oraz struktury do programowania jednoczesnego i obiegowego.

W tej sytuacji klienci tworzą swoje aplikacje zdecydowanie dla chmury, korzystając z interfejsu API dostępnego na poziomie oprogramowania pośredniczącego. Dlatego strategia ta jest inaczej nazywana Platformą jako usługą (PaaS) ze względu na to, że biuro oferowane klientowi jest etapem doskonalenia, a nie fundamentem. Ustalenia PaaS dotyczące maksymalnego udziału obejmują również ramy, które stanowią główny aspekt administracji udostępniany klientom. W przypadku Unfaultered PaaS oferowane jest tylko oprogramowanie pośrednie na poziomie klienta, które należy uzupełnić o fundament wirtualny lub fizyczny.

Góra warstwa modelu referencyjnego pokazanego na rysunku 2.1.1 rozciąga się na biura wprowadzone na poziomie aplikacji. Są one powszechnie określane jako oprogramowanie jako usługa (SaaS). Ogólnie rzeczą biorąc, są to aplikacje internetowe, które z szacunkiem odnoszą się do chmury i oferują wsparcie klientom końcowym. Siła chmury przekazywana przez cele IaaS i PaaS pozwala samorządny sprzedawcom oprogramowania na przenoszenie swoich biur aplikacji przez Internet.

Dodatkowe aplikacje wyruszające do tej warstwy to te, które jednoznacznie wpływają na Internet ze względu na swoje centralne funkcjonalności, które ufają chmurze, aby wytrzymać większą liczbę klientów; taka jest sytuacja w przypadku wejść do gier i, ogólnie rzecz biorąc, witryn służących do komunikacji międzyludzkiej.

Tabela 2.1.1 przedstawia atrybuty trzech głównych klasyfikacji stosowanych do klasyfikacji rozwiązań potwierdzających zachmurzenie. W załączonym segmencie celowo celujemy w te cechy wraz z pewnymi wskazówkami dotyczącymi rozsądnego ich wykorzystania.

Charakterystyka kategorii		Dostawcy i produkty typu produktu	
Aplikacja SaaS	Konsumenci są dostarczani z Aplikacją SaaS, która jest dostępna o każdej porze i z dowolnego miejsca	Sprzedaż internetowa Forcce.com (CRM) aplikacji i Clarizen.com(usługi (zarządzanie projektami internetowymi) 2.0) aplikacje Google	
PaaS	Konsumenci otrzymują etap pojawiających się aplikacji hostowanych w dudu	Programowanie API Google AppEngine i frameworków Microsoft Azure Manjrasoft Aneka Data Deployment Synapse systemy	
IaaS/HaaS	Konsumenci otrzymują zwirtualizowany sprzęt i pamięć masową, na których mogą budować swoją infrastrukturę.	Zarządzanie maszynami wirtualnymi Zarządzanie infrastrukturą pamięci masowej Zarządzanie siecią	Amazon EC2 i S2 GoGrid Nirvanix

2.1.2.1. Zadania i punkty odcięcia

Stowarzyszenia i ludzie mogą akceptować różne rodzaje, wstępnie scharakteryzowane stanowiska pracy, w zależności od tego, jak identyfikują się z chmurą i wprowadzonymi do niej zasobami IT lub potencjalnie z nią współpracującymi. Wszystkie przyszłe elementy wnoszą wkład i przekazują odpowiedzialność związaną z działalnością w chmurze. towarzyszące segmenty przedstawiają te części i rozpoznają ich główne skojarzenia.

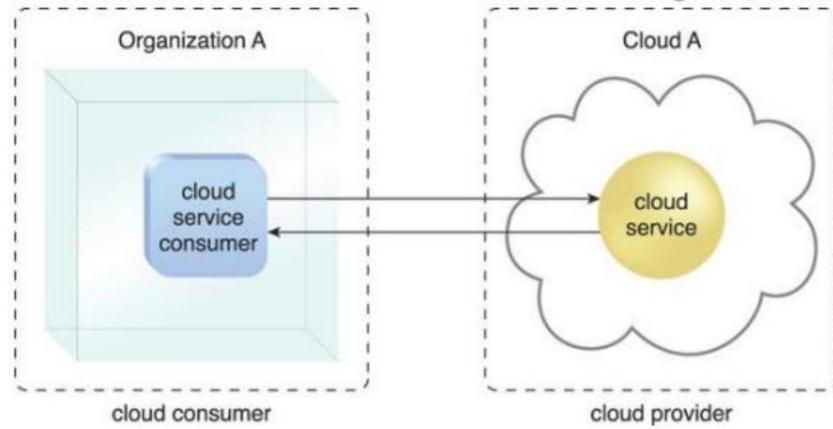
2.1.2.1.1. Dostawca chmury

Stowarzyszeniem udostępniającym zasoby IT w chmurze jest dostawca chmury. Podejmując pracę dostawcy usług w chmurze, firma naraża się na ryzyko, że administracja chmury będzie dostępna dla klientów korzystających z usługi w chmurze, zgodnie z zatwierdzoną umową SLA. Dostawcy chmury powierza się dodatkowo wszelkie podstawowe obowiązki administracyjne i regulacyjne, aby zapewnić ciągłość procedury ogólnych podstaw chmury. Dostawcy usług w chmurze zwykle są właścicielami nieruchomości IT udostępnianych do wynajęcia przez nabywców usług w chmurze; tak czy inaczej, niektórzy dostawcy usług w chmurze dodatkowo „wymieniają” zasoby IT wynajmowane od innych dostawców usług w chmurze.

2.1.2.1.2. Konsument chmury

Nabywca chmury to podmiot (lub osoba), który ma oficjalną umowę lub sposób działania z dostawcą chmury w celu korzystania z właściwości IT udostępnionych przez dostawcę chmury.

Mówiąc wprost, nabywca chmury korzysta z narzędzia do administrowania chmurą, aby uzyskać dostęp do administracji chmury (rysunek).

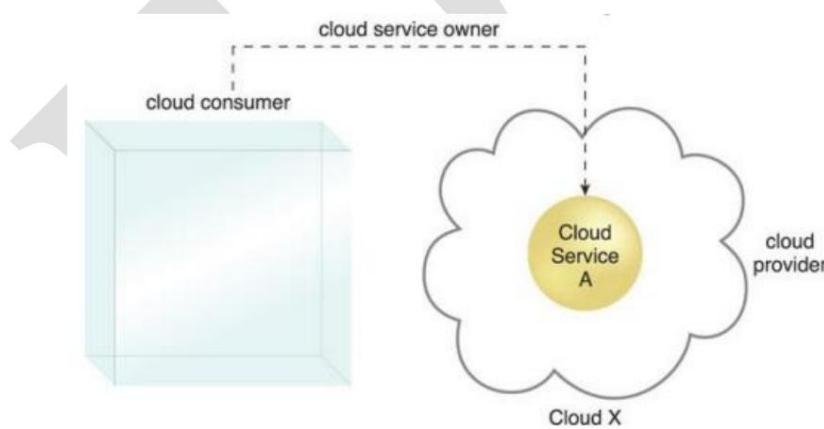


Rysunek: Konsument chmury (organizacja A) współpracuje z usługą chmury od dostawcy chmury (będącego właścicielem chmury A). W Grupie A klient korzystający z usługi w chmurze jest wykorzystywany do uzyskiwania dostępu do usługi w chmurze.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

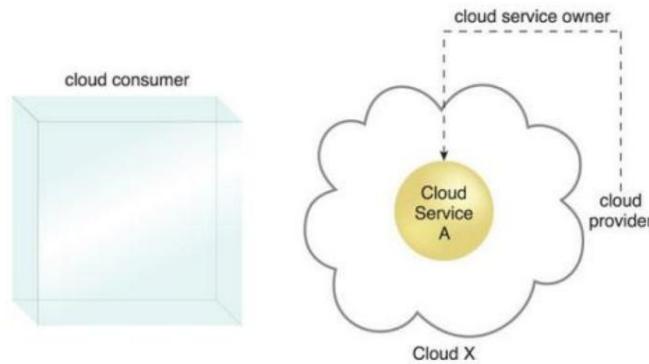
2.1.2.1.2.1. Właściciel usługi w chmurze

Osoba lub podmiot, który autorytatywnie twierdzi, że jest administratorem chmury, nazywany jest właścicielem administracji chmurą. Właścicielem administracji chmury może być kupujący w chmurze lub dostawca chmury, który rości sobie prawa do chmury, w której znajduje się administracja chmury. Na przykład dodatkowo klient Cloud X lub dostawca chmury Cloud X może posiadać Usługę Cloud A (rysunki 2.1.2 i 2.1.2.1).



Rysunek 2.1.2. Konsument chmury może być właścicielem usługi w chmurze, gdy wdraża własną usługę w chmurze.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

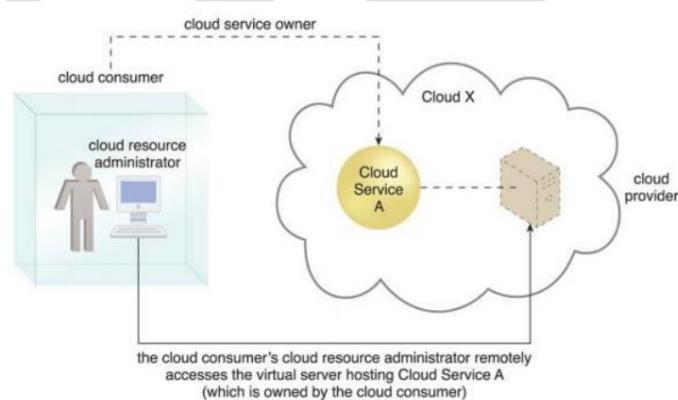


Rysunek 2.1.2.1. Dostawca chmury staje się właścicielem usługi w chmurze, jeśli wdroży własną chmurę usługi, z której zazwyczaj mogą korzystać inni konsumenci chmury.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

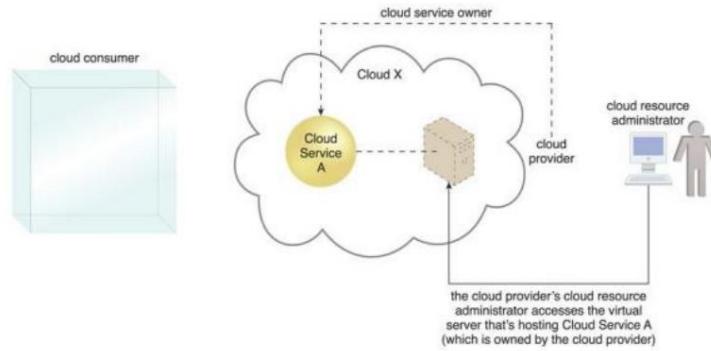
2.1.2.1.4. Administrator zasobów chmury

Dyrektor ds. zasobów w chmurze to osoba lub podmiot odpowiedzialny za opiekę nad zasobami IT opartymi na chmurze (licząc administrację chmurą). Nadzorcą zasobów w chmurze może być (lub mieć miejsce) klient chmury lub dostawca chmury w chmurze, w której istnieje administracja chmury. W innym przypadku może to być (lub mieć miejsce) podmiot zewnętrzny, któremu zlecono koordynację zasób IT oparty na chmurze. Na przykład właściciel biura w chmurze może wyznaczyć kierownika zasobów w chmurze, który będzie nadzorował administrację chmurą (rysunki 2.1.4 i 2.1.5).



Rysunek 2.1.4. Administrator zasobów chmury może pracować w organizacji konsumenckiej chmury i administrować zdalnie dostępnymi zasobami IT należącymi do konsumenta chmury.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

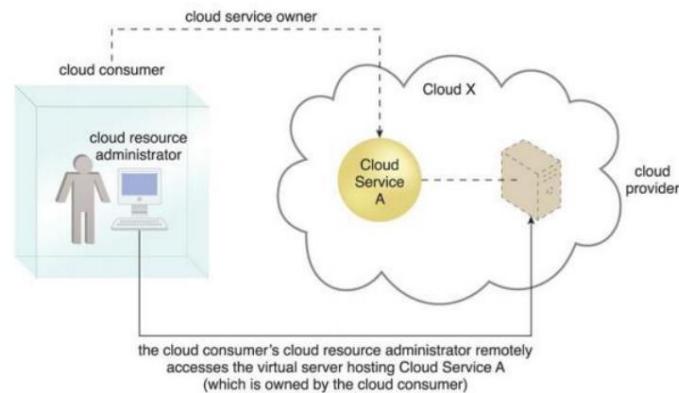


Rysunek 2.1.5. Administrator zasobów chmury może współpracować z organizacją dostawcy usług w chmurze, którym może administrować dostępnymi wewnętrznie i zewnętrznie zasobami IT dostawcy usług w chmurze.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

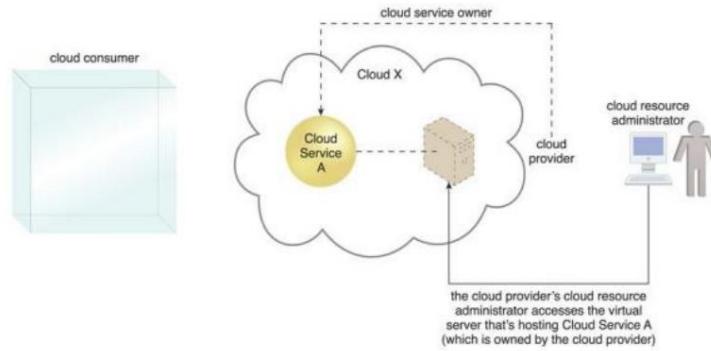
2.1.2.1.4. Administrator zasobów chmury

Dyrektor ds. zasobów w chmurze to osoba lub podmiot odpowiedzialny za opiekę nad zasobami IT opartymi na chmurze (licząc administrację chmurą). Nadzorcą zasobów w chmurze może być (lub mieć miejsce) klient chmury lub dostawca chmury w chmurze, w której istnieje administracja chmury. W innym przypadku może to być (lub mieć miejsce) podmiot zewnętrzny, któremu zlecono koordynację zasób IT oparty na chmurze. Na przykład właściciel biura w chmurze może wyznaczyć kierownika zasobów w chmurze, który będzie nadzorował administrację chmurą (rysunki 2.1.4 i 2.1.5).



Rysunek 2.1.4. Administrator zasobów chmury może pracować w organizacji konsumenckiej chmury i administrować zdalnie dostępnymi zasobami IT należącymi do konsumenta chmury.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)



Rysunek 2.1.5. Administrator zasobów chmury może współpracować z organizacją dostawcy usług w chmurze, którym może administrować dostępnymi wewnętrznie i zewnętrznie zasobami IT dostawcy usług w chmurze.

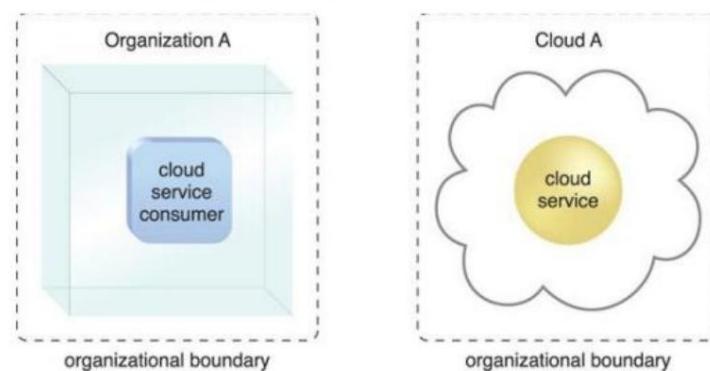
(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

Wyjaśnienie, że dyrektor ds. zasobów w chmurze nie jest określane mianem „menedżera administracji chmury”, jest takie, że na tym stanowisku może być odpowiedzialne za kierowanie zasobami IT w chmurze, które nie istnieją jako administratorzy chmury. Na przykład, jeśli prezes zasobów w chmurze pasuje do dostawcy chmury (lub jest przez niego skurczony), zasoby IT, które nie są udostępniane zdalnie, mogą być kontrolowane przez to stanowisko (a tego rodzaju zasoby IT nie są delegowaną administracją w chmurze).

2.1.4. Limit

2.1.4.1. Granica hierarchiczna

Limit hierarchiczny oznacza fizyczne obrzeża, które warunkują wiele kapitałów IT posiadanych i kontrolowanych przez stowarzyszenie. Autorytatywny limit nie pokazuje limitu rzeczywistego powiązania, a jedynie hierarchiczne rozmieszczenie właściwości IT i zasobów IT. Ponadto mgły mają miarodajny limit (rysunek 2.1.6.).

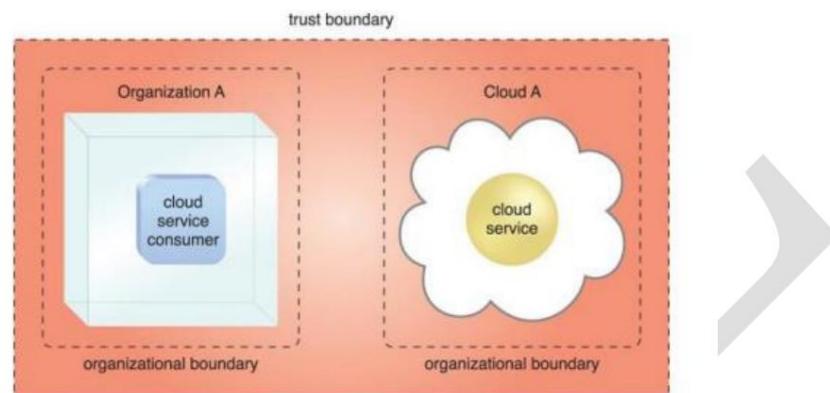


Rysunek 2.1.6. Granice organizacyjne konsumenta chmury (po lewej) i dostawcy chmury (po prawej), reprezentowane przez linię przerywaną.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.1.4.2. Granica zaufania

W momencie, gdy stowarzyszenie przejmuje zadanie klienta w chmurze, aby uzyskać dostęp do zasobów IT w chmurze, musi rozszerzyć swoje zaufanie poza fizyczne granice powiązania, aby objąć części warunków chmury. Limit zaufania to spójny margines, który zwykle wykracza poza fizyczne limity i określa stopień zaufania do zasobów IT (rysunek 2.1.7). Podczas badania sytuacji w chmurze limit zaufania jest najczęściej kojarzony z zaufaniem udzielonym przez stowarzyszenie działające jako nabywca chmury.



Rysunek 2.1.7. Rozszerzona granica zaufania obejmuje granice organizacyjne dostawca chmury i konsument chmury.

(Odniesienie : Cloud Computing (koncept, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.1.5. Charakterystyka chmur

Środowisko IT wymaga starannego zestawienia cech, aby w rzeczywisty sposób umożliwić zdalne udostępnianie dostępnych i szacowanych kapitałów IT. Te cechy muszą istnieć w stopniu wyrazistym, aby klimat IT można było uznać za opłacalną chmurę.

Twarzyszących sześć dokładnych atrybutów jest normalnych dla sytuacji w chmurze:

- wykorzystanie na żądanie •
- wszechobecny
- dostęp • wielodostępność (i łączenie zasobów) •
- elastyczność • szacowane

wykorzystanie • siła Dostawcy chmury i klienci chmury mogą niezależnie i razem określić ilościowo te atrybuty, aby je zmierzyć wartość zaangażowania danego etapu chmury. Pomimo faktu, że administracja oparta na chmurze i zasoby IT będą uzyskiwać i wykazywać określone cechy w różnym stopniu, ogólnie rzecz biorąc, im lepiej zostaną wzmocnione i zastosowane, tym lepsza będzie wynikająca z tego propozycja oceny.

2.1.5.1. Użycie na żądanie

Klient w chmurze może oddzielnie uzyskać dostęp do właściwości IT w chmurze, dając mu możliwość samodzielnego zarządzania tymi właściwościami IT. Po rozwiązaniu problemu można zaprogramować wykorzystanie własnych właściwości IT, co nie wymaga dodatkowego zaangażowania ze strony klienta lub dostawcy chmury. Powoduje to sytuację wykorzystania na żądanie. Inaczej nazywany „samodzielnym wykorzystaniem na żądanie”, ten znak towarowy umożliwia administrację i wykorzystanie efektów świetlnych rozpoczętających się w zwyczajowych mgłach.

2.1.5.2. Uniwersalny dostęp

Powszechny dostęp oznacza, że administracja w chmurze może być szeroko dostępna.

Ustanowienie wszechobecnego dostępu dla administratorów chmury może wymagać ustalenia zakresu systemów, konwencji transportowych, ograniczeń i poprawy dobrostanu. Aby umożliwić taki stopień dostępu, zwykle konieczne jest dostosowanie projektu administracji chmurą do konkretnych wymagań wstępnych różnych klientów administracji chmurą.

2.1.5.2.1. Wielodostępność (i łączenie zasobów)

Jakość pakietu produktów, która umożliwia programowi obsługę różnych klientów (użytkowników), przy czym każdy z nich jest oddalony od drugiego, nazywana jest wielodostępnością. Dostawca chmury łączy swoje właściwości IT, aby umożliwić różnym chmurom obsługę klientów poprzez wykorzystanie imitacji wielodostępnych, które rutynowo opierają się na wykorzystaniu postępów w zakresie wirtualizacji.

Dzięki wykorzystaniu innowacji związanych z wielodostępnością właściwości IT można animować i ponownie przydzielać, renderując żądania klientów do administrowania chmurą.

2.1.5.4. Wszechstronność

Wszechstronność to zaprogramowana skłonność chmury do jednoznacznego skalowania właściwości IT zgodnie z wymaganiami warunków wykonawczych lub dostosowanymi przez klienta chmury lub dostawcę chmury. Powszechnie szacuje się, że wszechstronność stanowi główny opór w uznaniu przetwarzania rozproszonego, przede wszystkim ze względu na sposób, w jaki jest ona ściśle powiązana z przewagą w postaci skróconych aktywów i kosztów porównawczych. Dostawcy rozwiązań chmurowych posiadający ogromne właściwości IT mogą zaoferować najbardziej godny uwagi zakres wszechstronności

2.1.5.5. Szacowane użycie

Znak towarowy celowego wykorzystania oznacza zdolność etapu chmury do zapobiegania korzystaniu z zasobów IT, w większości przez klientów chmury. Na podstawie szacunków dostawca usług w chmurze może obciążyć klienta w chmurze wyłącznie za faktycznie wykorzystane właściwości IT, a także za okres, w którym wybrano dostęp do właściwości IT. W tej wyjątkowej sytuacji szacowane wykorzystanie jest ściśle powiązane ze znakiem towarowym na żądanie.

2.1.5.6. Wytrzymałość

Silne przetwarzanie to rodzaj przełączania awaryjnego, który eliminuje nadmierne wykorzystanie właściwości IT w różnych miejscach fizycznych. Właściwości IT można wstępnie zaplanować w taki sposób, że w przypadku ich braku, zgoda zostanie przekazana na dodatkową, nadmiarową aplikację. W przetwarzaniu rozproszonym atrybut elastyczności może odnosić się do powtarzalnych właściwości IT w podobnej chmurze (jednak w różnych obszarach fizycznych) lub nad różnymi mgłami. Klienci w chmurze mogą

rozwijać zarówno niezawodność, jak i dostępność swoich aplikacji, wykorzystując siłę rozwiązań IT opartych na chmurze.

2.1.6. Model transportu w chmurze

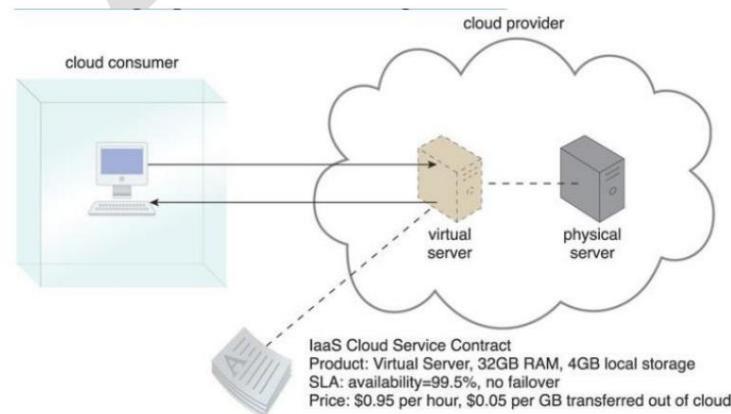
Model przenoszenia w chmurze oznacza przypisaną, wstępnie powiązaną mieszkankę zasobów IT dostępnych przez dostawcę chmury. Trzy popularne modele transportu w chmurze okazały się wszechstronnie postrzegane i honorowe:

- Infrastruktura jako usługa (IaaS)
- Platforma jako usługa (PaaS)
- Oprogramowanie jako usługa (SaaS)

2.1.6.1. Struktura jako usługa (IaaS)

Model obiegu IaaS zakłada niezależny klimat IT składający się z podstawowych zasobów IT, które zostaną odzyskane i zrealizowane za pomocą interfejsów i urządzeń opartych na administracji chmurą. Klimat ten może obejmować sprzęt, organizację, sieć, ramy robocze i inne „proste” zasoby IT. W odróżnieniu od konwencjonalnych rozwiązań ułatwiających lub redystrybuujących czynniki środowiskowe, w przypadku IaaS zasoby IT są zwykle wirtualizowane i umieszczane w opakowaniach, które kompresują z wyprzedzeniem wspinaczkę w czasie wykonywania i dostosowywanie środowiska. Szeroko użyteczną użytecznością domeny IaaS jest elastyczne udostępnianie klientom chmury przy podwyższonym poziomie kontroli i odpowiedzialności za nią.

projektowanie i użytkowanie. Zasoby IT udostępniane przez IaaS w zasadzie nie są z góry ustalone, co nakłada oficjalne obowiązki bezpośrednio na kupującego w chmurze. Model ten jest konsekwentnie wykorzystywany przez nabywców usług w chmurze, którzy potrzebują znacznego poziomu kontroli nad warunkami opartymi na chmurze, które będą zapewniać. Gdziekolwiek dostawcy usług w chmurze będą kontraktować wkład IaaS od innych dostawców usług w chmurze, aby skalować własne środowisko chmurowe. rodzaje i marki zasobów IT udostępnianych przez elementy IaaS oferowane przez różnych dostawców usług w chmurze mogą się zmieniać. Zasoby IT dostępne w warunkach IaaS są w większości oferowane jako nowo utworzone wirtualne wystąpienia. Najważniejszym i pierwszym zasobem IT w standardowym rozwiązań IaaS jest serwer wirtualny. Serwery wirtualne wynajmuje się, wskazując zapotrzebowanie na sprzęt serwerowy, takie jak limit procesora, pamięć i lokalna przestrzeń do przechowywania, jak pokazano na rysunku



Rysunek 2.1.8. Klient w chmurze korzysta z serwera wirtualnego w atmosferze IaaS. Konsument chmury otrzymuje od dostawcy chmury szereg gwarancji umownych dotyczących fizjonomii, takich jak pojemność, wydajność i dostępność.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

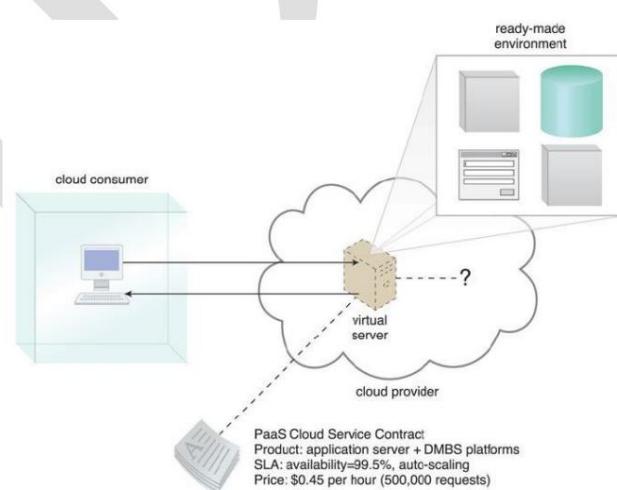
2.1.6.2. Etap jako usługa (PaaS)

Model przenoszenia PaaS mówi o wstępnie skategoryzowanym stanie „gotowym do użycia”, zwykle związanym z wcześniej wysłanymi i uporządkowanymi zasobami IT. W szczególności PaaS podlega zastosowaniu obszaru podpowiedzi, w którym znajduje się wiele wstępnie przygotowanych elementów i instrumentów używanych do wspomagania całego cyklu życia aplikacji niestandardowych.

Nawet motyw, które nabywca chmury zastosowałby i umieściłby nieruchomości w domenie PaaS, obejmują:

- Nabywca chmury musi połączyć warunki lokalne z chmurą ze względu na wszechstronność i ze względów finansowych.
- Klient korzystający z chmury wykorzystuje warunek natychmiastowy, aby całkowicie zastąpić warunek lokalny.
- Konsument chmury chce udać się do sprzedawcy chmury i zapewnia pełną administrację osobistą chmurą dostępną dla dodatkowych nabywców chmury zewnętrznej.

Zatrudniając uprzywilejowaną fazę natychmiastową, kupujący w chmurze oszczędza wiarygodną masę lokalizacji i utrzymuje odsłonięte podstawowe zasoby IT zapewniane przez model IaaS. Z drugiej strony klientowi chmury przyznaje się niższy poziom władzy nad podstawowymi zasobami IT, które hostują i aranżują scenę (rysunek 4.12).



Rysunek 2.1.9. Konsument chmury uzyskuje dostęp do gotowego środowiska PaaS. Znak zapytania wskazuje, że konsument chmury jest celowo chroniony przed szczegółami implementacji platformy.

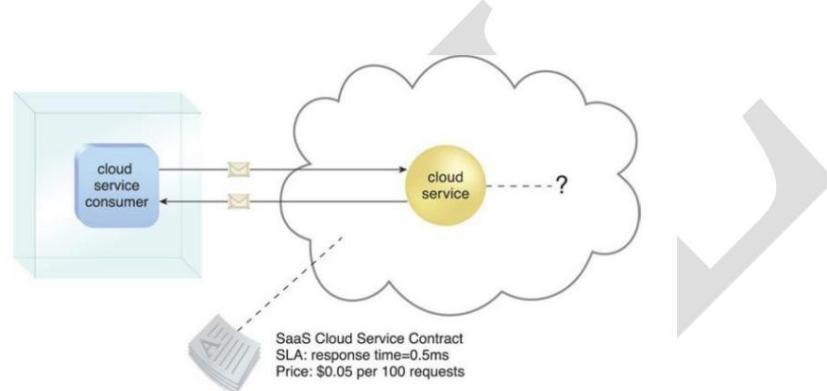
(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

Produkty PaaS są dostępne z różnymi stosami programistycznymi. Na przykład Google App Engine oferuje środowisko oparte na Javie i Pythonie.

2.1.6.2.1. Programowanie jako usługa (SaaS)

Program produktu umiejscowiony jako typowa administracja w chmurze i udostępniony jako „przedmiot” lub ogólna wartość, nawiązuje do standardowego profilu oferty SaaS. Model przenoszenia SaaS zwykle zapewnia szeroko dostępną administrację chmurą z możliwością ponownego uzupełnienia (często za pieniądze) różnym klientom chmury. Kompletne dochodowe centrum występuje ok

Sprawy SaaS, które zostaną pożyczone i zastosowane dla różnych dysków i dochodów liczne terminy



Rysunek 2.1.10. Klient korzystający z usługi w chmurze uzyskuje dostęp do umowy w sprawie chmury, ale nie do żadnych podstawowych zasobów IT ani szczegółów aplikacji.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

Nabywca chmury zwykle może uzyskać ograniczone, autorytatywne polecenie wykraczające poza praktykę SaaS. Najczęściej jest ona zapewniana przez dostawcę chmury, jednak często jest ona formalnie zgłoszana przez dowolny element, który oczekuje stanowiska właściciela administracji chmury. na przykład przedsiębiorstwo kupujące chmurę, korzystające z domeny PaaS i zarządzające nią, może stworzyć administrację chmurą, którą zdecyduje się przekazać w stanie równoważnym z ofertą SaaS. Identyczne stwarzyszenie w tym momencie odpowiednio akceptuje pracę dostawcy chmury, biorąc pod uwagę fakt, że administracja chmurą oparta na SaaS jest dostępna dla różnych stwarzyszeń, które korzystają z tej administracji chmurą jako nabywcy chmury.

2.1.7. Modele wdrażania w chmurze

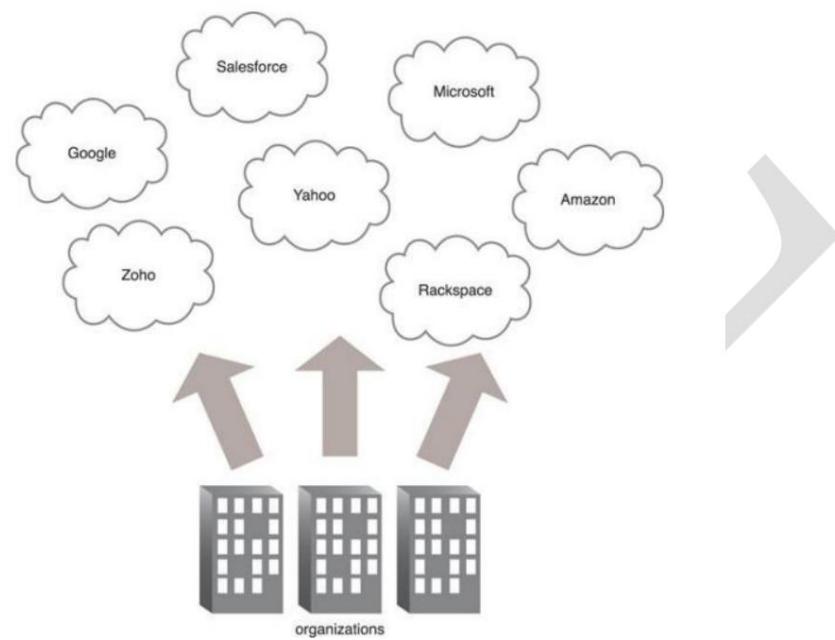
Klasyczna procedura w chmurze odnosi się do wybranego rodzaju stanu chmury, zasadniczo udokumentowanego przez własność, zakres i dostęp. Istnieją cztery modele stałego przesyłania w chmurze:

- Chmura publiczna

- Chmura społecznościowa
- Chmura prywatna
- Chmura hybrydowa. Powstałe obszary przedstawiają każdy z nich.

2.1.7.1. Otwórz chmury

Otwarta chmura może być swobodnie dostępny środowiskiem chmurowym, którego właścicielem jest zewnętrzny dostawca usług w chmurze. Właściwości IT w odsłoniętych mgłach są powszechnie zapewniane za pośrednictwem niedawno scharakteryzowanych replik do transportu w chmurze i zazwyczaj są oferowane nabywców chmury po określonej cenie lub są sprzedawane na różne sposoby. żywiciel chmury jest odpowiedzialny za organizację i bieżące wsparcie całej chmury populacji i jej właściwości IT. Znacząca liczba okoliczności i projektów analizowanych w nadchodzących sekcjach obejmuje otwarte mgły, a co za tym idzie, powiązanie dostawców i nabywców aktywów IT poprzez otwarte mgły. Rysunek 2.1.10 przedstawia połowową perspektywę na ogólną scenę otwartych chmur, przedstawiającą niektórych głównych sprzedawców w centrum handlowym.

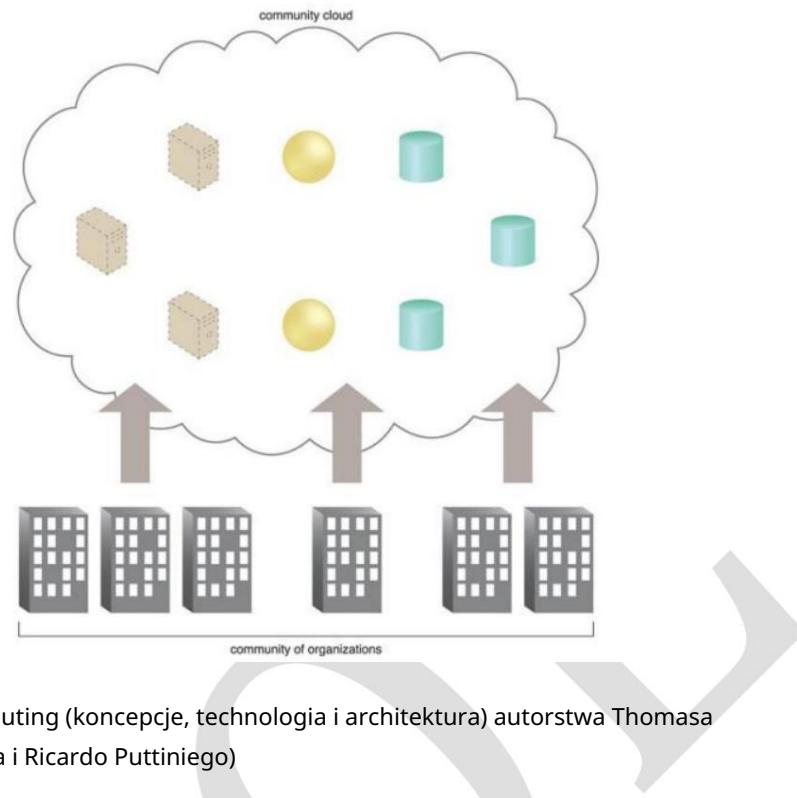


Rysunek 2.1.11. Organizacje zachowują się jak konsumenty chmury, uzyskując dostęp do usług w chmurze i zasobów IT udostępnianych przez różnych dostawców usług w chmurze.

(Odniesienie : Cloud Computing (konsepcje, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.1.7.2. Chmury sieciowe

Chmura grupowa osób przypomina chmurę sieciową, z tą różnicą, że dostęp do niej jest ograniczony do określonej sieci nabywców chmury. Otwarta chmura może być wspólnie kontrolowana przez otwartych partnerów lub zewnętrznego dostawcę chmury, który zapewnia otwartą chmurę z ograniczonym dostępem. Ogólnie rzecz biorąc, partnerzy nabywcy chmury sieciowej wspólnie ponoszą odpowiedzialność za charakterystykę i rozwój chmury sieciowej (rysunek 2.1.11).



(Odniesienie : Cloud Computing (konceptje, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

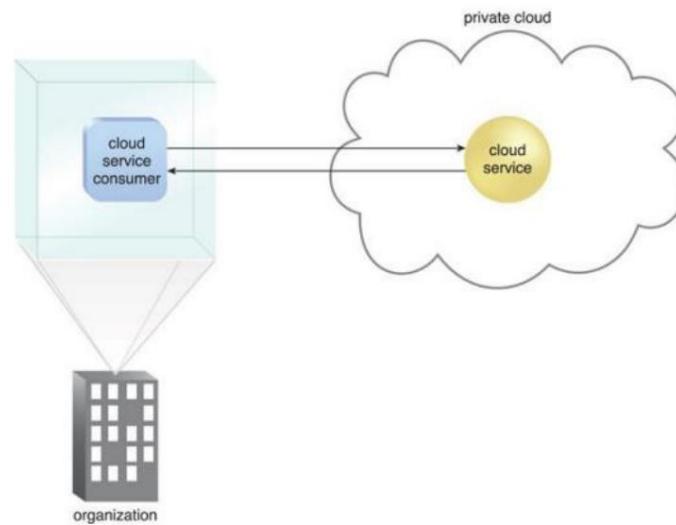
2.1.7.2.1. Chmury prywatne

Jedno stowarzyszenie zgłasza roszczenia do pojedynczej chmury. Prywatne mgły umożliwiają partnerstwu wykorzystanie innowacji w zakresie przetwarzania rozproszonego jako metody koncentracji dostępu do zasobów dla różnych części, obszarów lub oddziałów stowarzyszenia. W momencie, gdy pojedyncza chmura istnieje jako sytuacja kontrolowana, problemy przedstawione w segmencie Ryzyka i wyzwania z rozdziału 2.1 nie będą powszechnie stosowane

Korzystanie z pojedynczej chmury może zmienić sposób charakteryzowania i stosowania limitów autorytatywnych i uznaniowych. Specyficzna organizacja poszczególnych warunków w chmurze może również być regulowana przez personel wewnętrzny lub ponownie przywłaszczony.

W przypadku pojedynczej chmury identycznym stowarzyszeniem jest zarówno nabywca chmury, jak i dostawca chmury (rysunek 2.1.12.1). aby oddzielić te zadania:

- niewątpliwie autorytatywny segment zwykle przyjmuje odpowiedzialność za udostępnienie chmury (i tym samym oczekuje pracy jako dostawca chmury)
- Oddziały proszące o dostęp do chmury prywatnej oczekują pracy jako klient chmury.



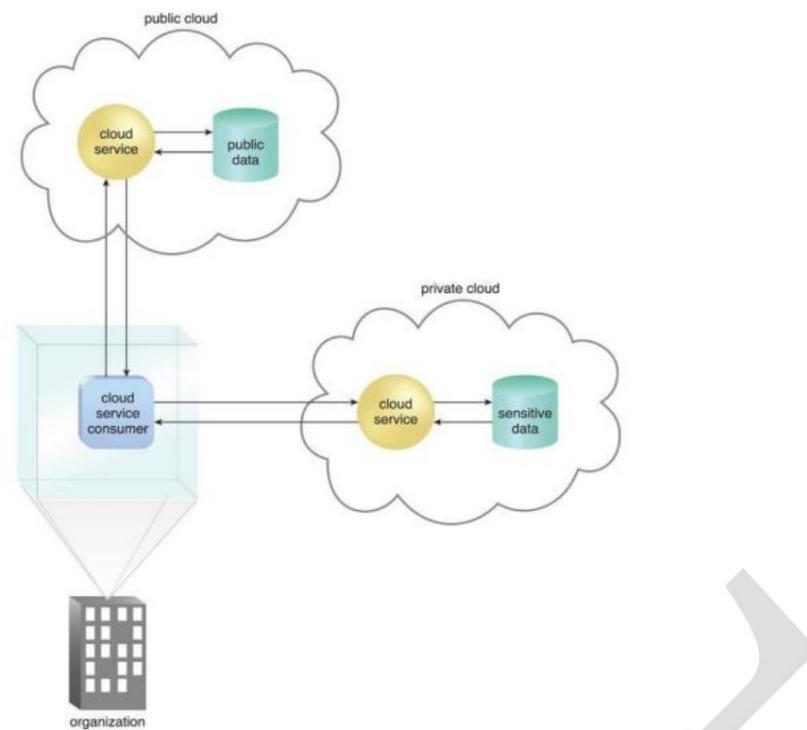
Rysunek 2.1.12.1. Konsument usług w chmurze w środowisku lokalnym organizacji uzyskuje dostęp do usługi w chmurze hostowanej w chmurze prywatnej równoważnej organizacji za pośrednictwem wirtualnej sieci prywatnej.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

Bardzo ważne jest dokładne wykorzystanie powinowactwa „w oparciu o rozsądek” i „oparte na chmurze” w ustawieniach pojedynczej chmury. jednak chmura prywatna może oczywiście znajdować się na terenie stowarzyszenia. Szacuje się, że zasoby IT sił zbrojnych są jeszcze „oparte na chmurze”, o ile są zdalnie udostępniane klientom korzystającym z usług w chmurze. Zasoby IT udostępniane poza chmurą prywatną przez segmenty działające jako klienci chmury są w tym sensie uważane za „lokalne” w odniesieniu do zasobów IT opartych na chmurze prywatnej.

2.1.7.4. Pół na pół Chmury

Chmura pół na pół może być środowiskiem chmurowym obejmującym co najmniej dwa różne modele rozmieszczenia chmur. na przykład klient w chmurze może chcieć zarządzać usługami w chmurze udostępnianie drażliwych informacji w pojedynczej chmurze i dodatkowe, mniej delikatne administrowanie chmurą w chmurze otwartej. Skutkiem tego połączenia może być model organizacji pół na pół (rysunek 2.1.12.1).



Rysunek 2.1.14. placówka stosująca architekturę chmury hybrydowej, która wykorzystuje zarówno osobę fizyczną i chmura publiczna.

(Odniesienie : Cloud Computing (konceptje, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

Struktury mieszane są regularnie łączone i testowane, aby tworzyć i postępować z wdzięcznością za możliwy kontrast w powietrzu chmur, a tym samym za niezaprzecjalną rzeczywistość, że odpowiedzialność organizacji jest tradycyjnie podzielona pomiędzy stowarzyszenie dostawców chmury prywatnej i dostawcę chmury otwartej.

2.1.8. Sprawy finansowe chmury

Głównymi czynnikami napędzającymi przetwarzanie rozproszone są ekonomia skali i prostota przenoszenia oprogramowania i jego działania. Rzeczywiście, najbardziej korzystna pozycja tego cudu ma charakter pieniężny: model „płać więcej tylko w miarę powstawania kosztów” oferowany przez dostawców usług w chmurze. w szczególności zezwolenia na przetwarzanie rozproszone:

- Redukcja wydatków kapitałowych związanych ze strukturą IT
- Eliminacja kosztów dewaluacji lub kosztów życia związanych z zasobami kapitału IT
- Zastąpienie pozwoleń na programowanie członkostwem
- Obniżenie kosztów utrzymania i regulacji aktywów IT

Wydatek kapitału polega na tym, że wydatek dotyczy zakupu korzyści, która jest cenna podczas tworzenia przedmiotów lub świadczenia usług administracyjnych. Wydatki inwestycyjne są jednorazowe

koszty, które są zwykle płacone bezpośrednio i które w przyszłości przyczynią się do uzyskania korzyści. Podstawą IT i zgodnie z tym produktem są zasoby kapitałowe, ponieważ firmy oczekują od nich prowadzenia swojej działalności. w chwili obecnej nie ma znaczenia, czy mówi się tam o najważniejszym biznesie danego przedsięwzięcia, biorąc pod uwagę fakt, że firma z pewnością będzie miała biuro IT, które będzie przyzwyczajone do robotyzacji znacznej części wykonywanych ćwiczeń wewnętrznych przedsięwzięcia: finanse, relacje z klientami, zarząd, organizowanie aktywów, śledzenie i magazynowanie towarów i inne. W związku z tym aktywa IT stanowią wydatek kapitałowy dla każdego przedsięwzięcia. akceptowalną praktyką jest utrzymywanie niskich kosztów kapitałowych, ponieważ wiążą się one z kosztami, które po pewnym czasie przyniosą korzyści; bardzo to, że ponieważ są utożsamiane z rzeczami materiałnymi, są zależne od dewaluacji powyżej ekstensywności

run, co ostatecznie zmniejsza korzyść przedsiębiorstwa w świetle faktu, że wydatki te są zgodnie z prawem odliczane od dochodów przedsięwzięcia. w przypadku kosztów kapitałowych IT, koszty pogorszenia jakości są określane jako mniej wartościowy sprzęt po pewnym czasie i w ten sposób dojrzewanie elementów programowych, które wymagają zastąpienia ze względu na konieczność wprowadzenia nowych funkcji.

Zanim przetwarzanie rozproszone rozprzestrzeniło się w przedsiębiorstwie, plan finansowy wydany w związku z tym na ramy i oprogramowanie stanowił główny koszt dla średnich i gigantycznych przedsięwzięć. Wiele przedsiębiorstw posiada małe lub średnie centra danych, które wiążą się z niewielkimi kosztami operacyjnymi w zakresie utrzymania, zasilania i chłodzenia. Dodatkowe koszty operacyjne powstają w związku z utrzymaniem działu IT i koncentracją na wspieraniu IT. Ponadto zabezpieczenie prawdopodobnie kosztownych programów powoduje aktywację różnych wydatków. Dzięki przetwarzaniu rozproszonemu wydatki te są całkowicie zmniejszone lub po prostu znikają wraz z ich infiltracją. Jedną zalet modelu przetwarzania rozproszonego jest to, że przenosi wydatki kapitałowe ostatnio przeznaczone na zabezpieczenie sprzętu i programowania na wydatki operacyjne wynikające z dzierżawy fundacji i opłacania składek za korzystanie z produktu.

Wydatki te są regularnie lepiej kontrolowane i przewidywalne w zależności od potrzeb biznesowych i rozwoju przedsięwzięcia. Przetwarzanie rozproszone zapewnia również zmniejszenie kosztów autorytatywnych i wsparcia. ilość funduszy rezerwowych na koszty, jakie przetwarzanie rozproszone może zapewnić w przedsiębiorstwie, zależy od konkretnej sytuacji, w której administracja chmury jest wykorzystywana, oraz sposobu, w jaki ją dodają, aby uzyskać korzyści dla przedsięwzięcia. w przypadku małego startupu możliwe jest całkowite wykorzystanie chmury pod kilkoma kątami, na przykład

- Podstawa IT
- Ulepszanie oprogramowania
- CRM i ERP

W tej sytuacji można całkowicie zrezygnować z wydatków kapitałowych, biorąc pod uwagę fakt, że nie istnieją żadne podstawowe zasoby IT. W przypadku przedsiębiorstw, które właśnie otrzymały hojną ilość zasobów IT, sytuacja jest zupełnie nieoczekiwana. w tym przypadku przetwarzanie rozproszone, zwłaszcza rozwiązania oparte na IaaS, może pomóc w nadzorowaniu zaimprowizowanych kosztów kapitałowych, które powstają w wyniku warunków wstępnych próby uprzywilejowanej w danej chwili. w tym przypadku, dzięki wykorzystaniu przetwarzania rozproszonego, wydatki te często się pojawiają

są to koszty operacyjne, które utrzymują się tak długo, jak długo istnieją ku temu przesłanki. na przykład wynajem fundamentów IT pomaga w jeszcze bardziej produktywny sposób nadzorować najważniejsze obciążenia bez konieczności uruchamiania kosztów kapitałowych. Gdy tylko zwiększone obciążenie nie legitymizuje do wykorzystania dodatkowych aktywów, są one regularnie odprowadzane, a w konsekwencji zanikają utożsamiane z nimi koszty. jest to często najczęściej stosowany model przetwarzania rozproszonego, biorąc pod uwagę fakt, że wiele przedsiębiorstw właśnie uruchomiło biura IT. decyzją wybiorczą jest określenie umiarkowanego postępu w kierunku rozwiązań opartych na chmurze, podczas gdy kapitałowe zasoby IT ulegają degradacji i chcą zostać zastąpione. Pomiędzy tymi dwoma przypadkami istnieje przyzwoita sytuacja, w której przetwarzanie rozproszone może być pomocne w generowaniu korzyści dla przedsiębiorstw. Jeśli chodzi o modele wyceny prezentowane przez przetwarzanie rozproszone, wyróżnimy trzy unikalne metodyki, które otrzymują dostawcy.

Odniesienie

Przetwarzanie w chmurze (koncept, technologia i architektura) autorstwa Thomasa Erla, Zaigham Mahmooda i Ricardo Puttiniego

Rozdział 2

Rozdział 2

2.0. Cel 2.1.

Podstawowe bezpieczeństwo chmury

2.1.1. Poufność 2.1.2.

Uczciwość 2.1.3.

Autentyczność 2.1.2.

Dostępność 2.1.5.

Zagrożenie

2.1.6. Luka 2.1.7. Ryzyko

2.1.8. Kontrole

bezpieczeństwa 2.1.9.

Mechanizmy bezpieczeństwa

2.1.10. Polityka bezpieczeństwa

2.2. Podstawy

2.2.1. Agenci zagrażający

2.2.2. Anonimowy atakujący 2.2.3.

Złośliwy agent usług 2.2.2. Zaufany
atakujący 2.2.5. Złośliwy

Insider 2.3. Czynniki

zagrożenia 2.3.1.

Podsłuch ruchu drogowego 2.3.2.

Złośliwy pośrednik 2.3.3. Odmowa
usługi

2.3.2. Niewystarczające zezwolenie

2.3.5. Atak wirtualizacyjny 2.3.6.

Nakładające się granice zaufania 2.3.7.

Zarządzanie ryzykiem 2.2.

Zagrożenia bezpieczeństwa chmury 2.5. Dodatkowe uwagi 2.5.1.

Propozycja AWS 2.5.2.

Zrozumienie usług internetowych Amazon 2.5.3.

Komponenty i usługi sieciowe AWS 2.5.2. Elastyczne
przetwarzanie w chmurze 2.6.

Platformy przemysłowe i nowe rozwiązania 2.7. Usługi internetowe Amazona 2.7.1.

Więcej o MS Cloud 2.7.2.

Maszyny wirtualne platformy Azure

2.7.3. Element Microsoft Azure 2.7.2.

Kontrola dostępu do MS Cloud

2.8. Silnik aplikacji Google

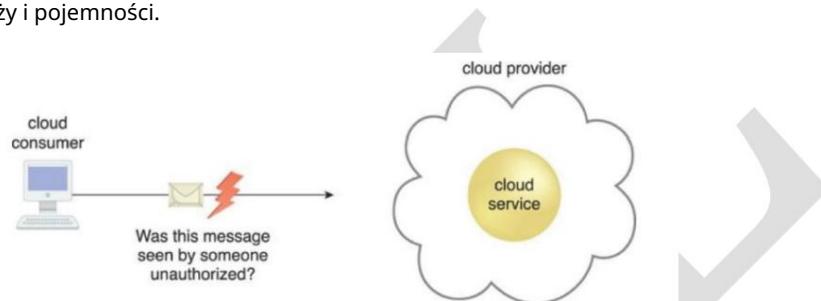
2.9. Microsoft Azure

2.1. Podstawowe bezpieczeństwo chmury

Bezpieczeństwo danych to złożona grupa procedur, udoskonaleń, wytycznych i wystaw, które wspólnie chronią autentyczność struktur i informacji komputerów osobistych oraz dostęp do nich. Okazje związane z bezpieczeństwem IT oczekują zabezpieczenia przed niebezpieczeństwami i przeszkodami, które wynikają zarówno ze szkodliwego celu, jak i przypadkowego błędu klienta.

2.1.1. Tajność

Tajemnica jest cechą udostępniania czegoś wyłącznie zatwierdzonym zgromadzeniom (rysunek 2.1). W sytuacjach związanych z chmurą tajemnica zasadniczo odnosi się do ograniczania dostępu do informacji podczas podróży i pojemości.

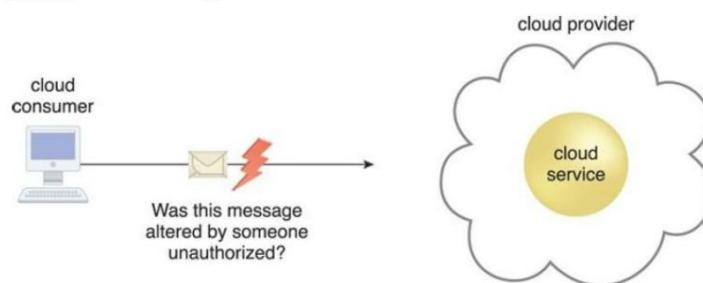


Rysunek 2.1.: Wiadomość wysłana przez konsumenta chmury do usługi w chmurze jest uważana za poufną tylko wtedy, gdy nie ma do niej dostępu ani nie jest czytana przez osobę nieuprawnioną.

(Odniesienie : Cloud Computing (koncepcje, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.1.2. Integralność

Uczciwość to cecha nieorganizowania, która została zmieniona przez niezatwierdzone zgromadzenie (rysunek 2.2). Ważną kwestią, która niepokoi poważanie informacji w chmurze, jest to, czy nabywca chmury może mieć pewność, że informacje, które przekazuje administratorowi chmury, koordynują informacje otrzymane przez tę administrację chmury. Szacunek może rozciągać się na sposób przechowywania, przygotowywania i odzyskiwania informacji przez administrację chmury i zasoby IT oparte na chmurze.



Rysunek 2.2.: Wiadomość wyslaną przez konsumenta chmury do usługi w chmurze uważa się za integralną, jeśli nie została zmieniona.

(Odniesienie : Cloud Computing (koncepcje, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.1.3. Rzeczywistość

Prawdziwość to jakość czegoś, która została podana przez zatwierdzone źródło. Wiedza ta obejmuje brak odwołania, co oznacza zawiedzienie spotkania w celu odrzucenia lub zakwestionowania upoważnienia stowarzyszenia. Weryfikacja niepodważalnych powiązań dostarcza dowodów na to, że powiązania te są w znaczący sposób powiązane z zatwierdzonym źródłem. Weźmy teraz przykład: najbardziej oczekiwany konsument nie będzie mógł dostać się do artykułu, który nie cieszy się dobrą reputacją, po jego otrzymaniu pozbawiony możliwości podobnego wygenerowania zapisu tego wpisu.

2.1.2. Dostępność

Dostępność to cecha otwartości i użyteczności w określonych ramach czasowych. W przeciętnych sytuacjach związanych z chmurą dostępność administracji chmurą może być obowiązkiem wspólnym dla dostawcy chmury i nośnika chmury. Dostępność rozwiązania opartego na chmurze, które dociera do klientów zajmujących się administracją chmurą, jest dodatkowo udostępniana przez nabywcę chmury.

2.1.5. Niebezpieczeństwo

Zagrożenie to potencjalne naruszenie bezpieczeństwa, które może poruszyć strażników próbujących złamać zabezpieczenia i dodatkowo spowodować obrażenia. Zarówno zagrożenia fizyczne, jak i naturalne mają na celu niewłaściwe wykorzystanie wad, zwanych także słabymi punktami. Niebezpieczeństwo, które ma miejsce, kończy się atakiem.

2.1.6. Bezsilność

Bezbronność to wada, która może zostać niewłaściwie wykorzystana albo ze względu na szczegółowość zapewnianą przez nieodpowiednie kontrole bezpieczeństwa, albo na tej podstawie, że obecne panele bezpieczeństwa są przytłoczone atakiem. Luki w zasobach IT mogą mieć wiele przyczyn, w tym nieprawidłową konfigurację, niedociągnięcia w strategii bezpieczeństwa, błędy klientów, defekty sprzętu lub oprogramowania sprzętowego, błędy programistyczne i bezradny projekt zabezpieczeń.

2.1.7. Zaryzykować

Hazard to szansa na nieszczęście lub zgorszenie wynikłe z rozegrania akcji. Zagrożenie jest regularnie oceniane na podstawie poziomu zagrożenia oraz ilości możliwych lub znanych luk w zabezpieczeniach. Dwie miary, które można wykorzystać do określenia szansy na zasób IT, to:

- Prawdopodobieństwo wystąpienia niebezpieczeństwa związanego z nadużyciem luk w zasobie IT
- Pragnienie nieszczęścia w związku z osłabieniem zasobu IT

Informacje na temat hazardu na planszy zostaną omówione w dalszej części tej sekcji.

2.1.8. Kontrola bezpieczeństwa

Kontrole bezpieczeństwa to środki zaradcze stosowane w celu zapobiegania zagrożeniom dla bezpieczeństwa lub reagowania na nie oraz w celu zmniejszenia lub utrzymania strategicznej odległości od zagrożenia. Subtelności na największych możliwościach Techniki stosowania środków zaradczych bezpieczeństwa są zwykle określone w strategii bezpieczeństwa, która zawiera wiele zasad i pracuje nad określeniem, w jaki sposób wykonać strukturę, administrację lub plan bezpieczeństwa w celu największego zapewnienia drażliwych i podstawowych zasobów IT.

2.1.9. Mechanizmy bezpieczeństwa

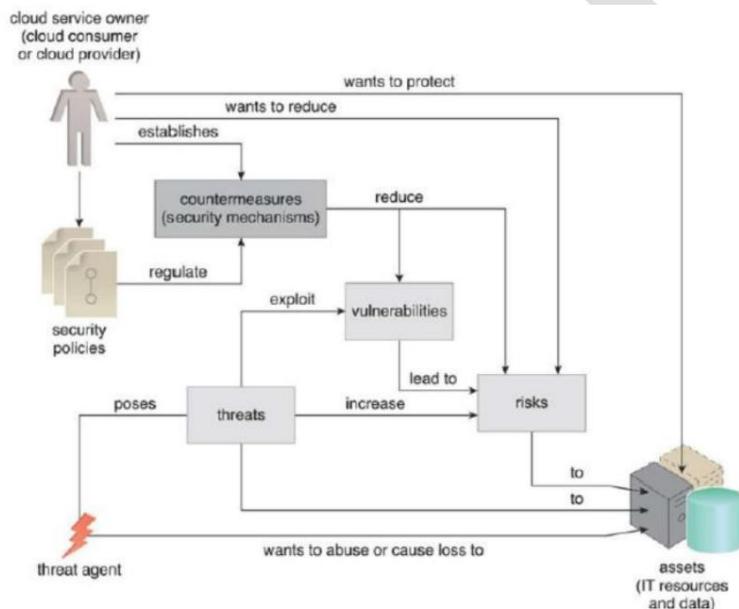
Środki zaradcze są powszechnie przedstawiane w przypadku systemów bezpieczeństwa, które są częściami zawierającymi strzeżoną strukturę zapewniającą zasoby IT, dane i administrację.

2.1.10. Zasady bezpieczeństwa

Strategia bezpieczeństwa określa wiele zasad i wytycznych dotyczących bezpieczeństwa. Regularnie strategie bezpieczeństwa będą dodatkowo charakteryzować sposób, w jaki te wytyczne i wytyczne są wykonywane i wdrażane. Na przykład rozmieszczenie i wykorzystanie środków kontroli i komponentów bezpieczeństwa może być kontrolowane za pomocą ustaleń dotyczących bezpieczeństwa

2.2.1. Niebezpieczni agenci

Specjalista od niebezpieczeństw to element, który reprezentuje zagrożenie, ponieważ nadaje się do przeprowadzenia ataku. Zagrożenia bezpieczeństwa chmury mogą mieć swój początek wewnętrz lub zdalnie, od ludzi lub programów programistycznych. W nadchodzących obszarach przedstawiono porównania operatorów zagrożeń. Rysunek 2.3 przedstawia pracę, którą operator niebezpieczeństw akceptuje, porównywalną do słabych punktów, niebezpieczeństw i niebezpieczeństw oraz osłon zbudowanych przez podejście do bezpieczeństwa i systemy bezpieczeństwa.



Rysunek 2.3.: Jak mechanizmy polityk bezpieczeństwa są wykorzystywane do przeciwdziałania słabym punktom zagrożeń i ryzykom powodowanym przez czynniki zagrażające.

(Odniesienie : Cloud Computing (konceptje, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.2.2. Nieznany napastnik

Nieznany napastnik to niewierzący nabywca administracji chmurą bez uprawnień w chmurze. Regularnie istnieje jako zewnętrzny program programistyczny, który wysyła zorganizowane ataki na poziomie poprzez systemy otwarte. W momencie, gdy tajemniczy agresor ogranicza dane na temat podejść i zabezpieczeń w zakresie bezpieczeństwa, może to ograniczyć ich zdolność do szczegółowego opisywania skutecznych ataków. W ten sposób nieznani napastnicy często uciekają się do takich działań, jak omijanie danych klientów lub sprawdzanie ich kwalifikacji, stosując jednocześnie strategie, które albo gwarantują anonimowość, albo wymagają znaczących aktywów do postawienia w stan oskarżenia.



Rysunek 2.2. Notacja używana w przypadku anonimowego atakującego.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.2.2. Tajemniczy napastnik

Tajemniczy napastnik to nieufny klient administracji chmury, który nie ma uprawnień w chmurze. Zwykle istnieje jako zewnętrzny program programistyczny, który wysyła ataki na poziomie zorganizowanym za pośrednictwem systemów otwartych. W momencie, gdy tajemniczy agresorzy ograniczyli dane dotyczące ustaleń i barier bezpieczeństwa, może to ograniczyć ich zdolność do planowania potężnych ataków. W związku z tym tajemniczy agresorzy regularnie uciekają się do takich działań, jak omijanie danych klientów lub odbieranie ich zaświadczeń, stosując jednocześnie strategie, które albo gwarantują anonimowość, albo wymagają znaczących zasobów do postawienia aktu oskarżenia.



Rysunek 2.5. Notacja używana dla złośliwego agenta usługi.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.2.2. Zwierzonoszy napastnikowi

Uznany sprawca udostępnia zasoby IT w podobnym stanie chmury co nabywca chmury i stara się niewłaściwie wykorzystać prawdziwe akredytacje, aby obrać za cel dostawców usług chmurowych i mieszkańców chmur, świętując zasoby IT (rysunek 2.6). W przeciwieństwie do tajemniczych agresorów (w które nie wierzy się), wierzący napastnicy z reguły wysyłają swoje ataki spoza granic zaufania chmury, posługując się prawdziwymi kwalifikacjami lub wybierając draźliwe i prywatne dane. Zaufani agresorom (inaczej zwani mściwi mieszkańcami) mogą wykorzystywać zasoby IT oparte na chmurze do szerokiego zakresu nadużyć, w tym na przykład włamywania się do słabych formularzy weryfikacyjnych, łamania szyfrowania, spamowania kont e-mail lub przeprowadzania podstawowych ataków., wyrzekając się krucjał administracyjnych.



Rysunek 2.6. Notacja używana w przypadku zaufanego atakującego.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.2.5. Szkodliwy insider

Szkodliwi insiderzy to specjalisci ds. zagrozeń dla ludzi, którzy działają na rzecz dostawcy chmury lub mu odpowiadają. Są to zazwyczaj obecni lub byli pracownicy lub osoby z zewnątrz mające dostęp do siedziby dostawcy usług w chmurze. Ten rodzaj specjalistów ds. zagrożeń niesie ze sobą ogromny potencjał szkód, ponieważ złośliwy insider może uzyskać korzyści regulacyjne w zakresie dotarcia do zasobów IT nabywcy chmury.



Rysunek 2.7. Notacja używana w przypadku ataku pochodzącego ze stacji roboczej. The symbol człowieka jest opcjonalny.

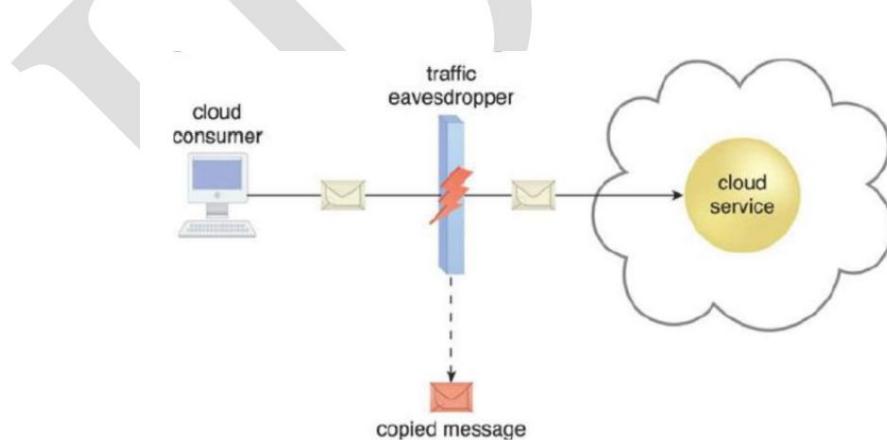
(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.3. Zagrożenia bezpieczeństwa chmury

Ten segment przedstawia kilka typowych zagrożeń i luk w zabezpieczeniach w sytuacjach opartych na chmurze oraz przedstawia zadania wspomnianych wcześniej operatorów zagrożeń.

2.3.1. Podsłuch ruchu drogowego

Podsłuchiwanie ruchu ma miejsce, gdy informacje przenoszone do chmury lub wewnętrznej niej (zwykle od klienta chmury do dostawcy chmury) są nieaktywnie przechwytywane przez mściwego specjalistę ds. pomocy w nieprzymyślanych celach gromadzenia danych (rysunek 2.8). Celem tego ataku jest bezpośrednie wynegocjowanie poufności informacji i, co możliwe, klasyfikacji powiązania pomiędzy nabywcą chmury a dostawcą chmury. Ze względu na niezaangażowaną koncepcję ataku, może on tym skuteczniej pozostać niewykryty przez dłuższe ramy czasowe.

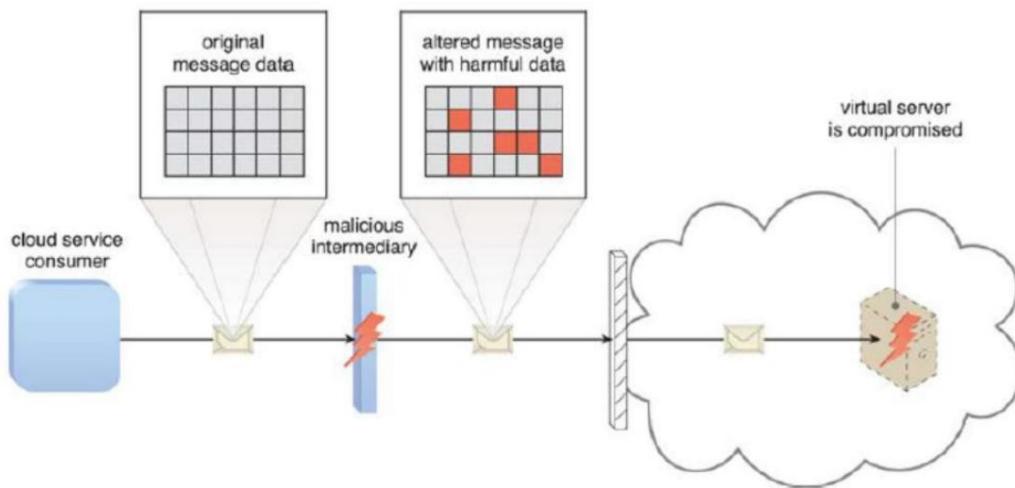


Rysunek 2.8.: Zewnętrznie umieszczeni złośliwi agenci usług przeprowadzają atak polegający na podsłuchiwaniu ruchu, przechwytyując wiadomość wyslaną do usługi w chmurze przez konsumenta usługi w chmurze. Agent usługi tworzy nieautoryzowaną kopię wiadomości, zanim zostanie ona wysłana oryginalną ścieżką do usługi w chmurze.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.3.2. Zły pośrednik

Niebezpieczeństwo złowrogiego mediatora pojawia się, gdy wiadomości są przechwytywane i zmieniane przez zgubnego specjalistę ds. pomocy, co może w ten sposób naruszyć klasyfikację wiadomości i dodatkowo jej uczciwość. Może również umieścić w wiadomości destrukcyjne informacje przed wysłaniem jej do celu. Rysunek 2.9. stanowi typowy przypadek wrogiego ataku mediatora.



Rysunek 2.9: Złośliwy agent usługi przechytuje i modyfikuje wiadomość wyslaną przez konsumenta usługi w chmurze do usługi w chmurze (niepokazanej) hostowanej na serwerze wirtualnym. Ponieważ w wiadomości zawarte są szkodliwe dane, serwer wirtualny jest zagrożony.

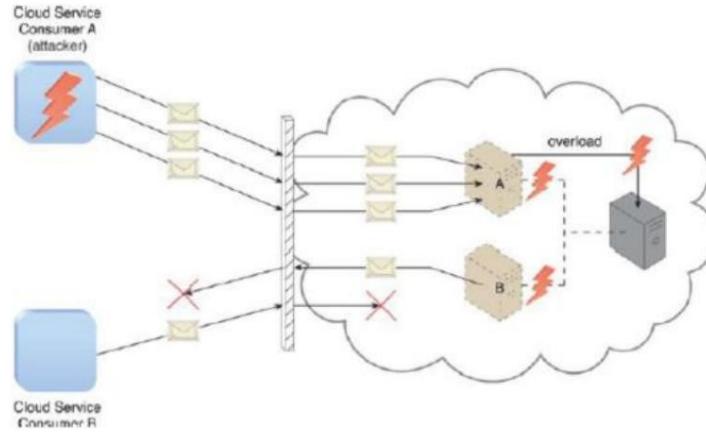
(Odniesienie : Cloud Computing (koncept, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.3.3. Wyrzeczenie się służby

Celem ataku wyparcia się administracji (DoS) jest nadmierne obciążenie zasobów IT do poziomu, w którym nie mogą one prawidłowo działać. Ten rodzaj ataku jest regularnie napędzany na jeden z towarzyszących sposobów:

- Zaległe zadanie stojące przed administracją chmury jest mylnie rozszerzone o podszywające się pod siebie wiadomości lub przeformułowane żądania korespondencji.
- System jest nadmiernie obciążony ruchem, co zmniejsza jego responsywność i utrudnia prezentację.
- Wysyłanych jest wiele żądań administracyjnych w chmurze, a każde z nich ma na celu wykorzystanie nieuzasadnionej ilości pamięci i zasobów obsługi.

Skuteczne ataki DoS powodują obniżenie wartości serwera i rozczarowanie, jak pokazano na rysunku 2.10:



Rysunek 2.10.: Konsument usługi w chmurze A wysyła wiele wiadomości do usługi w chmurze (niepokazanej) hostowanej na serwerze wirtualnym A. To powoduje przeciążenie wydajności bazowego serwera fizycznego, co powoduje awarie serwerów wirtualnych A i B. W rezultacie legalna chmura konsumenti usług, tacy jak konsument usług w chmurze B. nie mogą komunikować się z żadnymi usługami w chmurze hostowanymi na serwerach wirtualnych A i B.

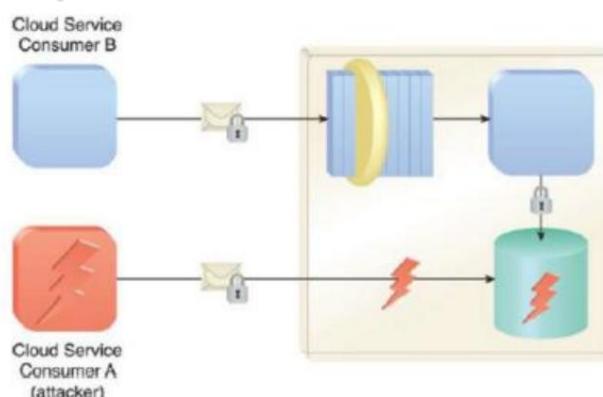
(Odniesienie : Cloud Computing (konceptje, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.3.3. Odmowa świadczenia usług

Celem ataku typu „odmowa administracji” (DoS) jest nadmierne obciążenie zasobów IT do poziomu, w którym nie mogą one działać prawidłowo. Ten typ ataku jest zwykle napędzany na jeden z towarzyszących sposobów:

- Pozostałe zadania administratorów chmury są fałszywie rozszerzane o podszywające się wiadomości lub preformułowane żądania korespondencji.
- System jest nadmiernie obciążony ruchem, co zmniejsza jego responsywność i wyświetlanie informacji dla osób niepełnosprawnych.
- Wysyłanych jest wiele żądań administracyjnych w chmurze, z których każde ma na celu wykorzystanie pamięci górnej i przygotowanie zasobów.

Owocne ataki DoS powodują uszkodzenie serwera, a także rozczarowanie, jak pokazano na rysunku 2.10:



Rysunek 2.11. : Konsument usługi w chmurze A uzyskuje dostęp do bazy danych, która została wdrożona przy założeniu, że dostęp do niej będzie możliwy wyłącznie za pośrednictwem usługi internetowej z opublikowaną umową o świadczenie usług (tak jak konsument B usługi w chmurze)

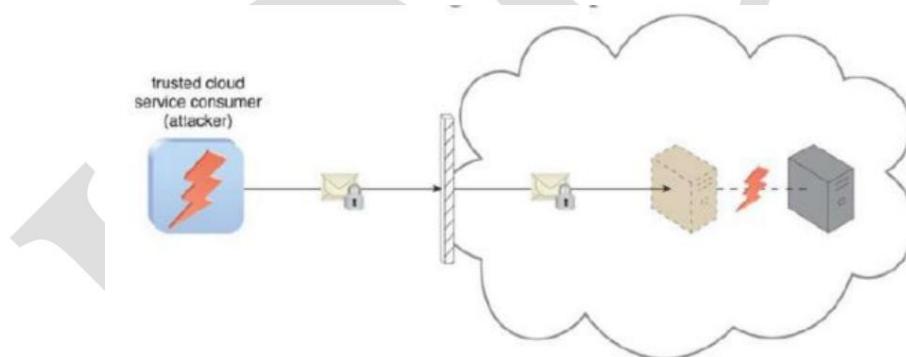
(Odniesienie : Cloud Computing (koncepcje, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

Różne rodzaje tego ataku, zwane słabym potwierdzeniem, mogą wynikać z wykorzystania słabych haseł lub wspólnych zapisów w celu zabezpieczenia zasobów IT. W sytuacjach związanych z chmurą tego rodzaju ataki mogą wywołać godne uwagi skutki w zależności od zakresu zasobów IT i zakresu dostępu do tych zasobów IT, jaki uzyska napastnik.

2.3.5. Atak wirtualizacji

Wirtualizacja zapewnia różnym nabywcom usług w chmurze dostęp do zasobów IT, które mają ten sam podstawowy sprzęt, ale są od siebie spójnie odłączone. Ponieważ dostawcy usług w chmurze przyznają nabywcom usług w chmurze regulacyjny dostęp do zwirtualizowanych zasobów IT (na przykład serwerów wirtualnych), istnieje niezbywalne ryzyko, że nabywcy usług w chmurze mogą wykorzystać to wejście w celu ataku na podstawowe fizyczne zasoby IT. Atak na wirtualizację wykorzystuje luki w zabezpieczeniach na etapie wirtualizacji, ryzykując jej poufność, uczciwość i dostępność. Niebezpieczeństwo to przedstawiono na rysunku 2.12, gdzie domniemany agresor skutecznie przedostaje się do serwera wirtualnego, aby targować się z ukrytym serwerem fizycznym.

W przypadku otwartej mgły, w której pojedynczy fizyczny zasób IT może udostępniać zwirtualizowane zasoby IT różnym nabywcom usług w chmurze, taki atak może mieć godne uwagi konsekwencje.

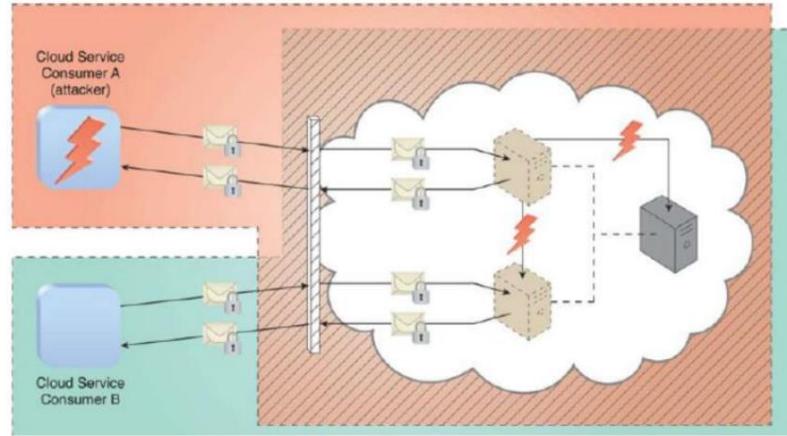


Rysunek 2.12.: Autoryzowany konsument usług w chmurze przeprowadza atak wirtualizacyjny, nadużywając swojego dostępu administracyjnego do serwera wirtualnego w celu wykorzystania podstawowego sprzętu.

(Odniesienie : Cloud Computing (koncepcje, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.3.6. Obejmowanie granic zaufania

W przypadku, gdy fizyczne zasoby IT w chmurze są współdzielone przez różnych klientów zajmujących się administracją chmurą, ci nabywcy administracji chmurą mają pokrywane limity zaufania. Złośliwi klienci administrujący chmurą mogą kierować przekazane zasoby IT w celu negocjowania zakupów w chmurze lub innych zasobów IT, które mają podobny limit zaufania. W rezultacie atak może dotyczyć kilku lub wszystkich innych osób kupujących usługi związane z administracją chmury, a napastnik może wykorzystać wirtualne zasoby IT przeciwko innym osobom, które również mają podobny limit zaufania.

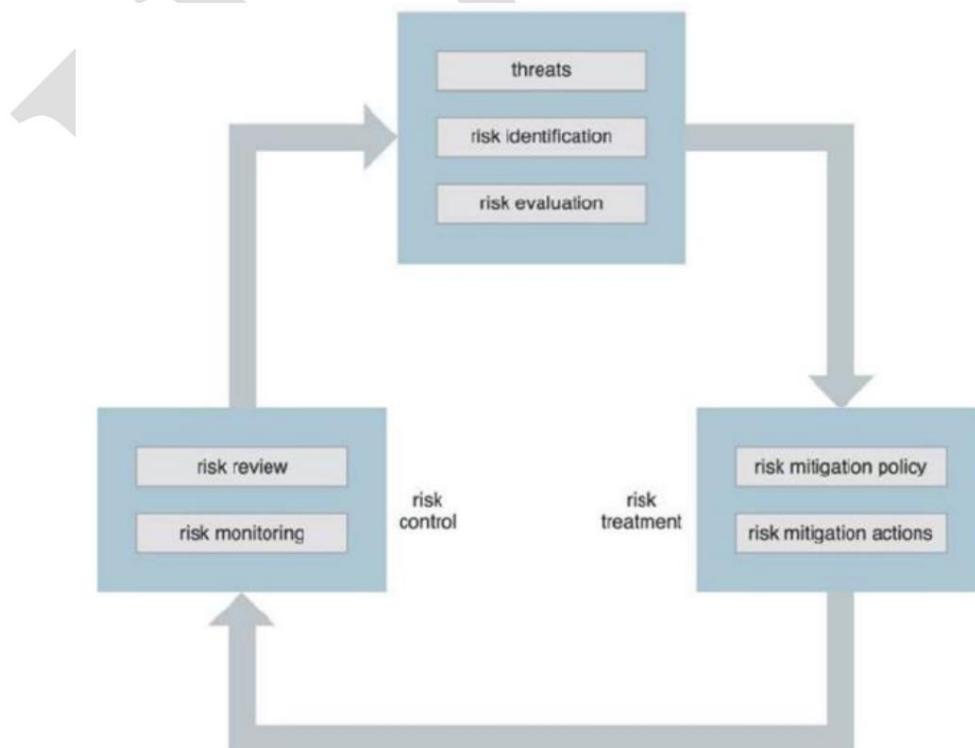


Rysunek 2.13 ilustruje przykład, w którym dwóch konsumentów usług w chmurze współużytkuje serwery wirtualne hostowane przez ten sam serwer fizyczny, w wyniku czego ich granice zaufania nakładają się.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.3.7. Zarządzanie zagrożeniami

Oceniając oczekiwane skutki i trudności związane z odbiorem w chmurze, zachęca się nabywców usług w chmurze do przeprowadzenia konwencjonalnej oceny zagrożenia jako głównego aspektu procedury dotyczącej zagrożenia. Konsekwentnie realizowana procedura służąca podnoszeniu poziomu bezpieczeństwa kluczowego i strategicznego, zarządzanie szansami obejmuje wiele ułatwionych ćwiczeń w zakresie nadzorowania i kontrolowania zagrożeń. Podstawowe ćwiczenia są powszechnie charakteryzowane jako ocena zagrożeń, leczenie przypadków i kontrola zagrożeń (rysunek 2.12).



Rysunek 2.12.: Bieżący proces zarządzania ryzykiem, który można rozpocząć od dowolnego z trzech etapów.

(Odniesienie : Cloud Computing (koncepty, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

- Postępowanie z ryzykiem – Ustalenia i plany łagodzenia są konstruowane na etapie leczenia zagrożeń w taki sposób, aby skutecznie nagradzać zagrożenia wykryte podczas oceny zagrożeń. Niektórych niebezpieczeństw można uniknąć, inne można złagodzić, a innymi można zarządzać poprzez ponowne zawłaszczenie lub nawet włączenie do ochrony, a także opracowanie planów finansowych na wypadek nieszczęścia. Dostawca usług w chmurze sam może zgodzić się na przyjęcie odpowiedzialności jako elementu swoich prawnie wiążących zobowiązań.
- Kontrola ryzyka – etap kontroli zagrożeń utożsamiany jest ze sprawdzaniem szans, procedurą składającą się z trzech etapów, która obejmuje badanie powiązanych okazji, badanie tych okazji w celu podjęcia decyzji o adekwatności wcześniejszych ocen i leków oraz rozpoznanie wszelkich potrzeb w zakresie modyfikacji strategii. W zależności od koncepcji wymaganej kontroli ten etap może zostać wykonany lub udostępniony przez dostawcę chmury.

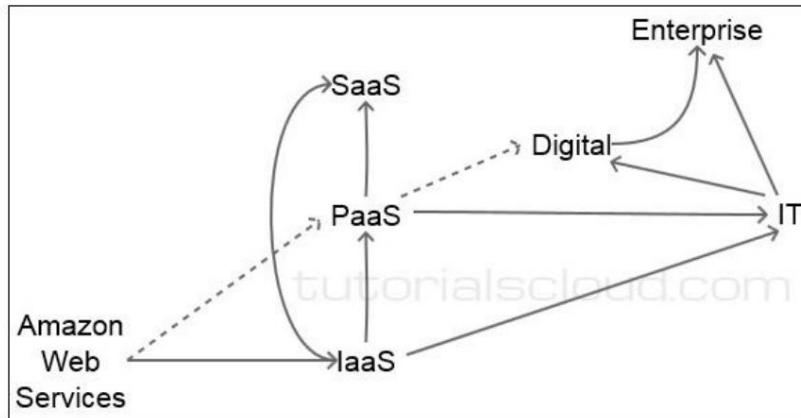
2.2. Platformy mechaniczne i nowe rozwiązania:

Rozwój aplikacji do przetwarzania rozproszonego następuje poprzez wykorzystanie etapów i struktur zapewniających różne rodzaje administracji, od odsłoniętego metalowego fundamentu po adaptowalne aplikacje spełniające wyraźne potrzeby.

2.5. Administracja siecią Amazon (AWS)

Jedną z najbardziej podziwianych witryn o ogromnym ruchu jest Amazon.com, która oferuje ogromną selekcję artykułów, wykorzystując administrację sieciową opartą na frameworku. Organizacja ta rozpoczęła się w 2006 roku od udostępnienia projektantom etapu administracji internetowej w modelu użytkowym. Organizacja ta przedstawia najlepszy przykład administracji internetowej, realizowanej poprzez inżynierię zlokalizowaną w administracji.

Usługa Amazon Web Service jest cennym elementem Amazon.com. Amazon umożliwił stworzenie prywatnych serwerów wirtualnych, które mogą ogólnie działać poprzez „wirtualizację sprzętu” na hiperwizorze Xen. Serwery te można wyposażyć w różnego rodzaju programy użytkowe, które klient może przewidzieć, wraz z szeregiem korzyści pomocniczych, które umożliwiają realizację rozproszonych aplikacji obliczeniowych, a także czynią je solidnymi, aby wytrzymać obliczenia.



Rysunek 2.15.

(Odniesienie : Cloud Computing (konceptje, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

W świetle standardu SOA oraz konwencji przenoszenia SOAP, REST i HTTP, ponadto otwarte - Źródłowy i biznesowy system operacyjny, programy oparte na programach i serwery aplikacji działają przez Amazon Web Service. AWS oferuje różne konfiguracje innowacji rejestrujących w chmurze, które stanowią etap obliczeniowy na żądanie. Administracje te pracują w dwunastu odrębnych obszarach geologicznych, a wśród nich najpopularniejsze to Elastic Compute Cloud (EC2) firmy Amazon i usługa Simple Storage Service (S3) firmy Amazon.

2.5.1. Rekomendacja AWS

AWS ma ogromną ofertę. Po prostu klienci muszą płacić za to, z czego korzystają, co pozwala zaoszczędzić dużo gotówki. AWS ma dodatkowo siedemdziesiąt administracji, w tym pojemność, liczbę, bazę danych, organizację, administrację aplikacjami, wszechstronność, kadrę kierowniczą, urządzenia inżynieryjne i IoT.

2.5.2. Bagatelizowanie usług internetowych Amazon

Jest to największy na świecie sklep internetowy. Przed Amazon.com największym sprzedawcą detalicznym na świecie był Wal-Mart. Według raportu rocznego za rok 2009 oferta netto Amazona wynosi 22,51 miliarda dolarów. Ma ogromną działalność i w tym celu stworzyła kolosalny system ram IT do wsparcia. AWS zasadniczo wykorzystuje naprawdę korzystny biznes systemu Amazon.com, przynosząc ogromny dochód.

AWS ma ogromną moc i wpływ na innowacje w chmurze, tworząc największe centrum komercyjne infrastruktury jako usługi (IaaS).

2.5.3. Segmente i usługi sieciowe AWS

Administracje internetowe Amazona mają towarzyszące segmenty:

- Amazon Elastic Compute Cloud: (EC2; <http://aws.amazon.com/ec2/>) to wbudowane rozwiązanie AWS, które zachęca do administrowania i używania wirtualnych serwerów prywatnych, które mogą działać na platformach opartych na systemach Windows i Linux za pośrednictwem Xen Hypervisor . Do pomocy administracji internetowej Amazon wykorzystywane są różne urządzenia. To są:

- Amazon Simple Queue Service to platforma przesyłania wiadomości i wymiany dla rozproszonych aplikacji internetowych.
 - Usługa Amazon Simple Notification Service służy do dystrybucji wiadomości z aplikacji.
 - Amazon CloudWatch służy do obserwacji chmury EC2, która opiera się na zapewnieniu konsoli lub linii zamówienia perspektywy na używane zasoby.
 - Elastyczne równoważenie obciążenia służy do identyfikowania, czy wystąpienie ma charakter płaski, czy też nie sprawdź, czy ruch jest prawidłowy, czy nie.
 - Usługa Simple Storage firmy Amazon: to gromadzenie i wzmacnianie zapasów online framework, który ma strategię szybkiego przepływu informacji zwaną importem/eksportem AWS.
- Dodatkowe mechanizmy usług sieciowych to:

- Sklep Amazon z elastycznymi blokami
- Prosta baza danych Amazon (DB)
- Usługa relacyjnej bazy danych Amazon
- Amazon Cloudfront

Duża liczba usług i narzędzi wspiera także partnerów Amazon, czyli samą infrastrukturę AWS. To są:

- Internetowy serwis informacyjny Alexa
- Usługi sieciowe Amazon Associates (A2S)
- Amazon DevPay
- Elastyczna redukcja mapy
- Mechaniczny Turek Amazona
- Uwierzytelnianie wieloskładnikowe AWS
- Elastyczna usługa płatności Amazon (FPS)
- Usługa internetowa Amazon dotycząca realizacji zamówień (FWS)
- Wirtualna chmura prywatna Amazon

2.5.2. Elastyczne przetwarzanie w chmurze

Jest to etap serwera wirtualnego, umożliwiający klientom tworzenie i uruchamianie maszyn wirtualnych na serwerze Amazon. Obrazy Amazon Machine Images (AMI) są używane przez EC2 do komunikacji i uruchamiania próbek serwerów do uruchamiania działających frameworków, takich jak: Linux (Red-Hat), Windows i tak dalej, na różnych serwerach. Jak sama nazwa wskazuje, możemy elastycznie dodawać lub usuwać, w razie potrzeby; odtwarzać i dostosowywać obciążenie serwerów. Możemy również znaleźć nasze serwery w różnych strefach na całym świecie, aby zapewnić adaptację w przypadku awarii niekrytycznych.

Termin „wszechstronny” charakteryzuje zdolność do szybkiego zmieniania rozmiaru umiejętności. Wykonanie pomocy może wymagać towarzyszących segmentów:

- Serwer aplikacji (posiadający dużą alokację RAM)
- Moduł równoważenia obciążenia
- Serwer bazy danych
- Zapora sieciowa i przełączniki sieciowe
- Dodatkowa pojemność stojaka

2.6. Silnik Aplikacji Google

Google App Engine umożliwia nam uruchamianie (posiadanie) własnych aplikacji internetowych na fundamencie Google. Tak czy inaczej, w żaden sposób, w żadnej postaci ani w żadnej formie nie jest to „wynajmowanie kawałka serwera” ułatwiającego administrację. Dzięki App Engine Twoja aplikacja nie jest obsługiwana na pojedynczym serwerze. Nie ma serwerów, które trzeba by nadązać: po prostu przesyłasz swoją aplikację, a ona jest gotowa do obsługi Twoich klientów. Podobnie jak dostosowywanie żądania wyszukiwania Google może obejmować kilka, a nawet wiele serwerów Google, a wszystkie są całkowicie zakryte i wykonywane w ciągu krótkiej sekundy, aplikacje Google App Engine działają w podobny sposób i na podobnym framework. To nowatorska część metodologii Google. Tak naprawdę oddajesz część kontroli Google, a mimo to otrzymujesz rekompensatę w postaci całkowitego uwolnienia się od fundamentów, ograniczenia liczby kadry kierowniczej i obciążenia dostosowywaniem zadań, które przedsiębiorstwo zwykle musi nadzorować, niezależnie od tego, czy mają one charakter samoobsługowy, czy też ułatwiają korzystanie z PaaS innej osoby lub IaaS.

Możesz zdecydować o udostępnieniu swojej aplikacji światu lub ograniczyć dostęp do poszczególnych osób ze swojego stowarzyszenia. Google App Engine wspiera aplikacje napisane w kilku programach językach narodowych:

Dzięki warunkowi środowiska wykonawczego Java App Engine możesz tworzyć aplikacje, korzystając ze standardowych wynalazków Java, włączając JVM, serwlety Java i projekt programu Java język — lub inny język wykorzystujący mediator lub kompilator oparty na JVM, na przykład JavaScript lub Ruby. Silnik aplikacji zawiera również zatwierdzony wykonawczy język Python, który zawiera szybki tłumacz języka Python i standardową bibliotekę języka Python. Sytuacje środowiska wykonawczego Java i Python są opracowywane w celu zagwarantowania, że aplikacja będzie działać szybko, bezpiecznie i bez impedancji z różnych aplikacji w środowisku.

Podobnie jak w przypadku większości administracji obsługujących chmurę, w przypadku App Engine otrzymujesz jedynie rekompensatę za to, czego używasz. Google nie wymaga żadnych kosztów konfiguracyjnych ani powtarzalnych opłat. Podobnie jak w przypadku AWS firmy Amazon, zasoby, na przykład gromadzenie zapasów i przepustowość danych, szacowane są w gigabajtach.

Uruchomienie silnika aplikacji nic nie kosztuje. Wszystkie aplikacje mogą wykorzystywać około 500 MB pojemności oraz wystarczającą ilość procesora i transmisji danych, aby pomóc w sprawnej części aplikacji około 5 milionów wejść na stronę miesięcznie, całkowicie za darmo. W momencie, gdy umożliwisz pobieranie opłat za swoją aplikację, Twoje punkty odcienia bezpłatnych opłat zostaną podniesione, a Ty otrzymasz jedynie rekompensatę za zasoby, których używasz powyżej bezpłatnych poziomów.

Projektanci aplikacji wykorzystują innowacje w zakresie niewyczerpanej pojemności, na przykład system plików Google (GFS) i Bigtable — platformę do gromadzenia nieustrukturyzowanych informacji w obiegu. Wersja Java stanowi podstawę nietypowych, nieblokujących zapytań wykorzystujących interfejs Twig Object Datastore. Daje to opcję w przeciwnieństwie do wykorzystywania ciągów znaków do równego przygotowywania informacji.

„Dzięki Google App Engine projektanci mogą tworzyć aplikacje internetowe w oparciu o podobną strukturę, z której korzysta Google” – napisał Kevin Gibbs, specjalista Google odpowiedzialny za przedsięwzięcie w oficjalnym blogu Google. „Google. Twig to interfejs pracowitości oparty na Google”. Niskopoziomowy magazyn danych App Engine, który pokonuje wiele przeszkód JDO-GAE, w tym pełną pomoc dla starszych wersji, polimorfizmu i konwencjonalnych rodzajów. Możesz bez większego naciągania projektu zmienić lub rozszerzyć postępowanie Twiga, aktualizując własne systemy lub zastępując je augmentacją koncentruje się na nieskażonej Javie

kod. Application Engine łączy te kwadraty strukturalne i zapewnia dostęp do elastycznych podstaw, które, jak mamy nadzieję, ułatwiają projektantom naturalne skalowanie aplikacji w miarę ich rozwoju.

Google App Engine pojawił się, gdy rosnąca liczba organizacji technologicznych przenosi swoje zadania do chmury; stawia to Google w ścisłej rywalizacji z produktami Amazon Elastic Cloud Computing (EC2) i Simple Storage Service (S3).

Google twierdzi, że jego wizja w przypadku Google App Engine polega na oferowaniu projektantom stopniowo wszechstronnej odpowiedzi, od początku do końca, w zakresie tworzenia i skalowania aplikacji w Internecie. Jej serwery są zaprojektowane tak, aby dostosować stertę ruchu do aplikacji inżynierów, skalując się w celu zaspokojenia potrzeby zalewu ruchu. Application Engine zawiera również interfejsy API do sprawdzania poprawności klienta, aby umożliwić projektantom logowanie się do administracji i poczty elektronicznej w celu nadzorowania korespondencji.

InternetNews.com szczegółowe,

Dzięki wstępemu przeglądowi silnik App Engine firmy Google będzie dostępny dla pierwszych 10 000 projektantów, którzy się zarejestrują, a w przyszłości planuje się zwiększyć tę liczbę. W tym okresie klienci będą ograniczeni do 500 MB pojemności i 10 GB dziennej szybkości transferu i 5 milionów wejść na stronę dziennie – twierdzi organizacja. Inżynierowie będą mieli możliwość zarejestrowania maksymalnie trzech aplikacji.

2.7. Microsoft Azure

daje szeroki asortyment administracji, z których klienci chmury mogą korzystać bez konieczności zakupu własnego sprzętu. Umożliwia szybkie doskonalenie ustaleń i doprowadzenie majątku do tego stanu. Bez konieczności stresu związanego z gromadzeniem fizycznych fundamentów, figura Azure, procesy Azure, system, gromadzenie zapasów i administracja aplikacjami pozwalają klientom skoncentrować się na budowaniu niesamowitych aranżacji.

Usługi Azure

Azure Services pamięta różne administracje ze względu na innowacje w chmurze. To są:

1. Usługi obliczeniowe: obejmują administrację MS Azure, na przykład maszynę wirtualną Azure, witrynę internetową Azure, administrację mobilną i tak dalej.
 2. Usługi danych: obejmują pamięć MS Azure, bazę danych Azure SQL i tak dalej.
 3. Usługi aplikacyjne: obejmują administrację, która powoduje, że klienci tworzą i obsługują aplikacje, na przykład Azure Active Directory, transport usług w celu kojarzenia odpowiednich struktur, przygotowywania ogromnych informacji i tak dalej.
2. Usługi sieciowe: obejmują system wirtualny, system przenoszenia treści i administratora ruchu platformy Azure.

Istnieją inne usługi, takie jak:

- BizTalk
- Duże obliczenia
- Tożsamość

- Wiadomości
- Media
- CDN itp.

2.7.1. Więcej o MS Cloud

Strefę początkową działań innowacyjnych firmy Microsoft w chmurze można znaleźć pod adresem Microsoft.com/cloud. Ma ogromny zakres innowacji w chmurze i niektóre elementy napędzające wykorzystanie Internetu w biznesie. Microsoft Messenger stał się szefem rynku po America Online Instant Messenger (AIM). Stopniowo wraz z rozwojem branży e-biura i reklamy Microsoft uważa, że zapewnia ona najlepsze zrozumienie sieci WWW dla różnego rodzaju gadżetów, na przykład komputerów stacjonarnych, stanowisk pracy, komputerów stacjonarnych, tabletów, zaawansowanych zawieszeń i tak dalej.

2.7.2. Fioletowo-niebieskie maszyny wirtualne

Jest to jedna z ujednolicionych zalet IaaS platformy MS Azure wraz z systemem wirtualnym. Maszyny wirtualne Purplish blue obsługują ulepszenia maszyn wirtualnych Windows Server (lub Linux) w centrum danych MS Azure; gdzie masz nieograniczony nadzór nad konfiguracją maszyny wirtualnej. Maszyna wirtualna Sky Blue ma trzy potencjalne stany:

- Praca
- Zatrzymany • Zatrzymany (Desygnowany)

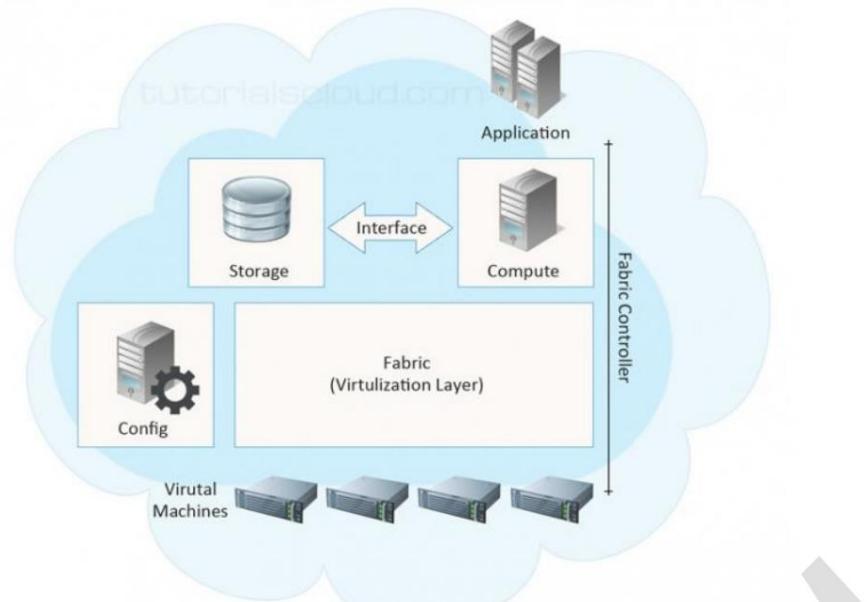
Maszyna wirtualna automatycznie otrzymuje stan zatrzymania (cofnięcia przypisania), gdy zostanie zatrzymana w portalu zarządzania Azure. W przypadku, gdy będziemy musieli zatrzymać go w takim samym podziale, musimy użyć polecenia cmdlet PowerShell z towarzyszącą mu kolejnością:

```
> Stop-AzureVM -Name "az-essential" -ServiceName "az-essential" -StayProvisioned
```

2.7.3. Element Microsoft Azure

Istnieje 6 głównych elementów tworzących system Windows Azure. To są:

- Oblicz
- Składowanie
- Aplikacja
- Tkanina
- VM (maszyny wirtualne)
- Konfiguracja (Konfiguracja)



Rysunek 2.16. - Elementy Microsoft Azure:

(Odniesienie : Cloud Computing (koncepcje, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego)

2.7.2. Kontrola dostępu do MS Cloud

Pozwala aplikacji uwierzyć w charakter innej aplikacji, a strategia ta może spotkać się z dostawcami osobowości, na przykład ADFS, aby uzależnić przekazywane struktury od SOA. Środki wymagane do kontroli dostępu to:

- Klient/klient wysyła żądanie weryfikacji z AC (Access Control)
 - Kontrola dostępu tworzy token w zależności od wytycznych odłożenia dla aplikacji serwerowej
 - Token zostaje oznaczony i wraca do aplikacji klienta
 - Klient przedstawia otrzymany token administracji aplikacja
- Sprawdzanie znaku odbywa się na końcu i przy użyciu tokenu określa, czy aplikacja w chmurze jest dozwolona, czy nie.

Odniesienie: Cloud Computing (koncepcje, technologia i architektura) autorstwa Thomasa Erla, Zaighama Mahmooda i Ricardo Puttiniego

JEDNOSTKA III

ROZDZIAŁ 1
SPECJALISTYCZNE MECHANIZMY CHMURY

Struktura:

3.1.1 Cele 3.1.2

Wprowadzenie 3.1.3

Automatyczny odbiornik skalowania

3.1.3.1 Przypadek DTGOV

 3.1.3.1. Skalowanie w dół

 3.1.3.1. B Skalowanie w

góre 3.1.4 Load Balancer

 3.1.4.1 Jak działa równoważenie obciążenia?

3.1.3.1 Monitor SLA 3.1.6

Monitor płatności za użycie 3.1.7

Monitor audytu 3.1.8 System

przełączania awaryjnego

 3.1.8.1 Systemy przełączania awaryjnego występują w dwóch podstawowych konfiguracjach

 3.1.8.1. Aktywny-Aktywny

 3.1.8.1. B Aktywny-Pasywny

3.1.9 Hypervisor

 3.1.9.1 Hypervisor dzieli się na dwa typy

 3.1.9.1. Typ pierwszy to hypervisor typu bare-metal 3.1.9.1.

 B Typ drugi to hostowany hypervisor działający jako warstwa oprogramowania

3.1.10 Klaster zasobów

 3.1.10.1 Typowe typy klastrów zasobów 3.1.10.1.

 Klaster serwerów 3.1.10.1. B

 Klaster Baz Danych 3.1.10.1. C

 Klaster z dużym zbiorem danych 3.1.10.2

 Istnieją dwa podstawowe typy klastrów zasobów 3.1.10.2 A

 Klaster ze zrównoważonym obciążeniem

 3.1.10.2 B Klaster o wysokiej dostępności (HA)

3.1.11 Broker wielourządzeniowy

3.1.12 Baza danych zarządzania stanem

3.1.13 Przykładowe ćwiczenie z pytaniem

3.1.14 Referencje

3.1.1 Cel:

Kolektyw oferujący inną i tradycyjną architekturę. Mechanizm ten jest agentem usługi, który monitoruje i śledzi komunikację między konsumentami usług w chmurze a usługami w chmurze na potrzeby dynamicznego skalowania.

W ten sposób konsument chmury może dostosować swoją obecną alokację zasobów IT.

3.1.2 Wprowadzenie:

Typowa architektura technologii chmurowej zawiera wiele ruchomych części, które odpowiadają różnym wymaganiom dotyczącym wykorzystania zasobów i rozwiązań IT. Każdy mechanizm omówiony w tym rozdziale spełnia określona funkcję środowiska wykonawczego, wspierając jedną lub więcej cech chmury.

3.1.3 Odbiornik automatycznego skalowania:

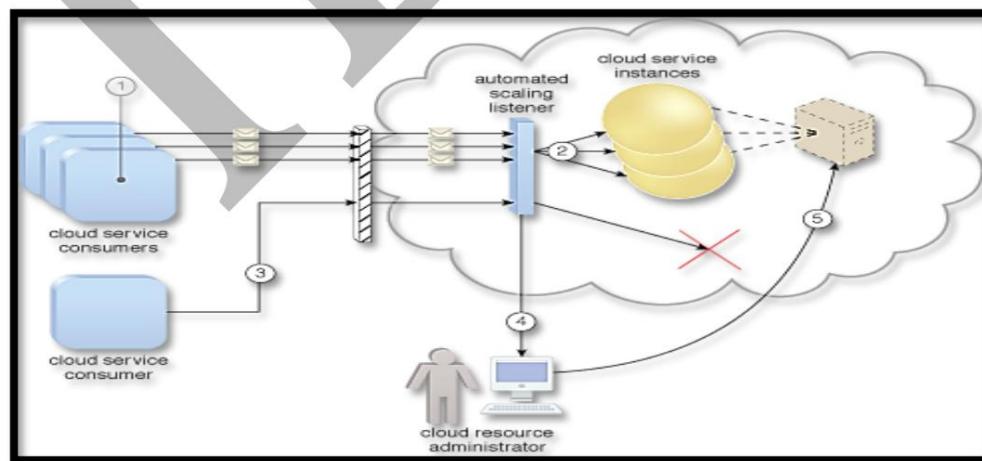
Mechanizm automatycznego nasłuchiwanego skalowania to agent usługi, który monitoruje i śledzi komunikację między konsumentami usług w chmurze a usługami w chmurze na potrzeby dynamicznego skalowania. Zautomatyzowane odbiorniki skalowania są wdrażane w chmurze, zazwyczaj w pobliżu zapory, skąd automatycznie śledzą informacje o stanie obciążenia.

Obciążenie można określić na podstawie liczby żądań generowanych przez konsumentów w chmurze lub na podstawie wymagań dotyczących przetwarzania zaplecza wyzwalanych przez określone typy żądań. Na przykład niewielka ilość przychodzących danych może skutkować dużą ilością przetwarzania.

Zautomatyzowane odbiorniki skalowania mogą zapewniać różne typy odpowiedzi na warunki wału obciążenia, takie jak:

- Automatyczne skalowanie zasobów IT w oparciu o parametry zdefiniowane wcześniej przez konsumenta chmury (powszechnie określane jako automatyczne skalowanie).
- Automatyczne powiadamianie konsumenta chmury, gdy obciążenie przekracza bieżące progi lub spaść poniżej przydzielonych zasobów. W ten sposób konsument chmury może dostosować swoją obecną alokację zasobów IT. (automatyczne powiadomienie)

Różni dostawcy usług w chmurze mają różne nazwy agentów usług, którzy działają jako odbiorniki automatycznego skalowania.



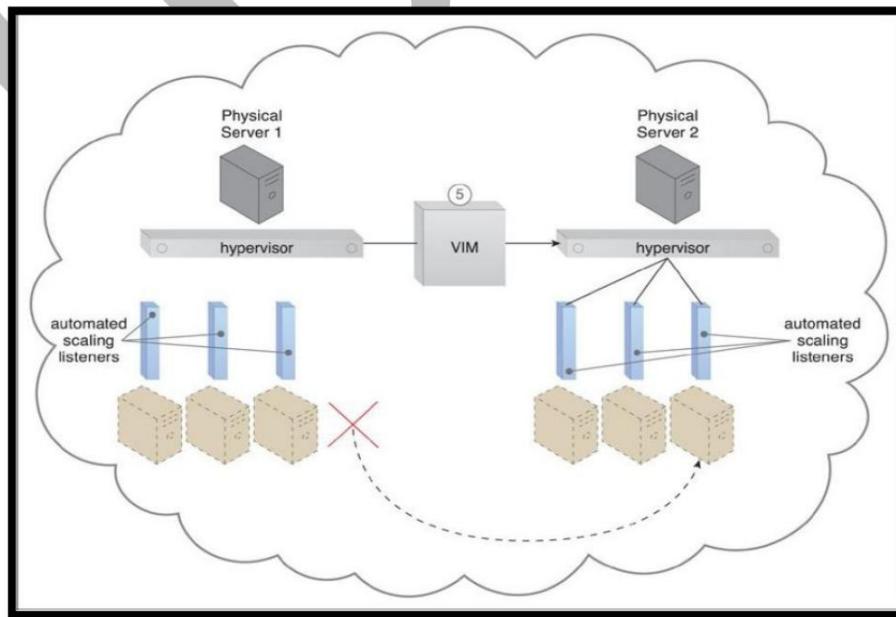
Ryc. 3.1.3

Trzech konsumentów usług w chmurze próbują jednocześnie uzyskać dostęp do jednej usługi w chmurze (1). Zautomatyzowany odbiornik skalowania skaluje się w poziomie i inicjuje tworzenie trzech nadmiarowych instancji usługi (2). Czwarty konsument usług w chmurze próbuje skorzystać z usługi w chmurze (3). Zaprogramowany tak, aby zezwalał na maksymalnie trzy wystąpienia usługi w chmurze, zautomatyzowany odbiornik skalowania odrzuca czwartą próbę i powiadamia konsumenta chmury o przekroczeniu żądanego limitu obciążenia (4). Administrator zasobów chmury konsumenta chmury uzyskuje dostęp do środowiska administracji zdalnej, aby dostosować konfigurację udostępniania i zwiększyć limit nadmiarowych instancji (3.1).

3.1.3.1 Sprawa DTGOV

Platforma wirtualizacji jest skonfigurowana tak, aby automatycznie skalować serwer wirtualny w czasie wykonywania, w następujący sposób:

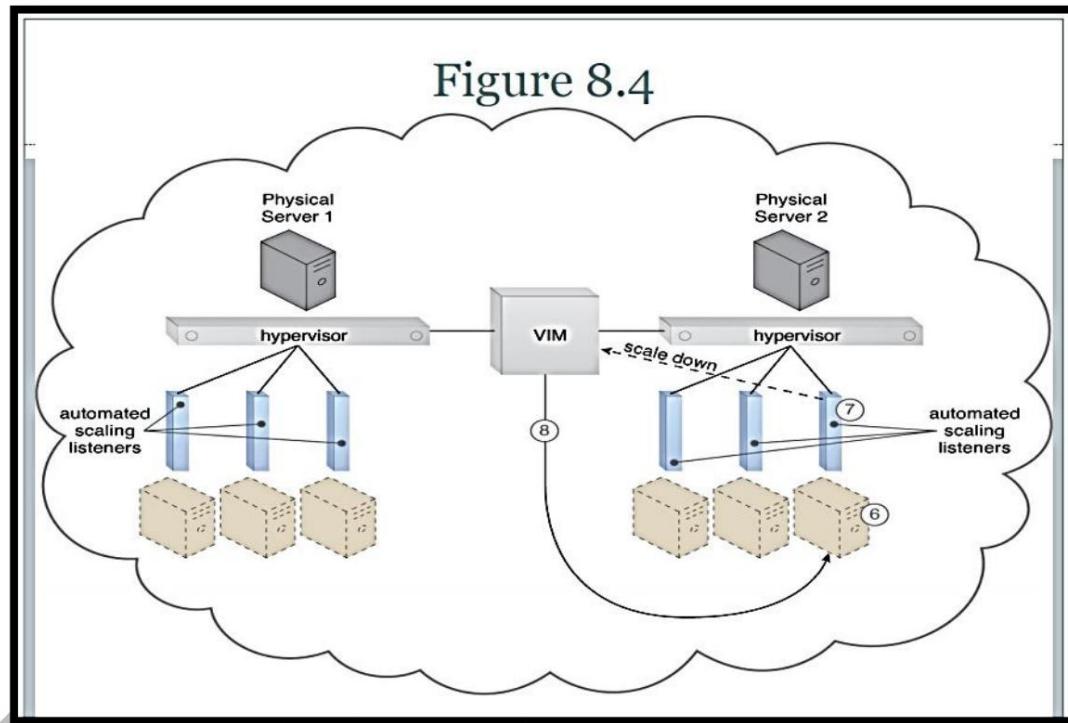
- A. Skalowanie w dół – serwer wirtualny nadal znajduje się na tym samym serwerze hosta fizycznego podczas skalowania do konfiguracji o niższej wydajności.
- B. Skalowanie w górę – pojemność serwera wirtualnego zostaje podwojona na jego oryginalnym serwerze hosta fizycznego. VIM może również na żywo przeprowadzić migrację serwera wirtualnego na inny serwer fizyczny, jeśli pierwotny serwer hosta jest przeciążony. Migracja odbywa się automatycznie w czasie wykonywania i nie wymaga wyłączenia serwera wirtualnego.
- 1. Konsument chmury tworzy i uruchamia serwer wirtualny z 8 rdzeniami procesorów wirtualnych i 16 GB wirtualnej pamięci RAM (1).
- 2. VIM tworzy serwer wirtualny na żądanie konsumenta usługi w chmurze, a odpowiednia maszyna wirtualna jest przydzielana do Serwera Fizycznego 1, aby dołączyć do 3 innych aktywnych maszyn wirtualnych (2).
- 3. Zapotrzebowanie konsumentów na chmurę powoduje wzrost wykorzystania serwerów wirtualnych o ponad 80%. Wydajność procesora przez 60 kolejnych sekund (3).
- 4. Automatyczny detektor skalowania działający na hypervisorze wykrywa potrzebę skalowania w górę i odpowiednio wydaje polecenia VIMowi (4).



Rys. 3.1.3.1

Na rys. 3.1.3.1 VIM stwierdza, że skalowanie serwera wirtualnego w górę na Serwerze Fizycznym 1 nie jest możliwe i przystępuje do migracji na żywo do Serwera Fizycznego 2.

5. Użycie procesora/RAM przez serwer wirtualny pozostaje poniżej 13,1% pojemności przez 60 kolejne sekundy (6).
6. Automatyczny detektor skalowania wykrywa potrzebę skalowania w dół i wydaje polecenie VIM (7), który skaluje serwer wirtualny (8), dopóki pozostaje on aktywny na serwerze fizycznym 2.



Rys. 3.1.3.2

3.1.4 Moduł równoważenia obciążenia

Równoważenie obciążenia w chmurze to proces dystrybucji obciążzeń i zasobów obliczeniowych w środowisku technologii chmury. Pomaga także organizacjom i przedsiębiorstwom zarządzać wymaganiami dotyczącymi obciążenia poprzez alokację zasobów pomiędzy wieloma systemami lub serwerami. Równoważenie obciążenia w chmurze obejmuje również hosting dystrybucji ruchu obciążenia, który odbywa się w Internecie.

Dzięki tej technice równoważenia obciążenia można osiągnąć wysoki poziom wydajności zadań przy niższych kosztach niż tradycyjna lokalna technologia równoważenia obciążenia. Oprócz dystrybucji obciążenia i ruchu, równoważenie obciążenia w technologii chmury może również zapewniać kontrolę stanu aplikacji w chmurze.

Mechanizm równoważenia obciążenia jest agentem wykonawczym, którego logika zasadniczo opiera się na tym założeniu. Moduły równoważenia obciążenia mogą wykonywać szereg wyspecjalizowanych funkcji dystrybucji obciążenia w czasie wykonywania, które obejmują:

- Dystrybucja asymetryczna – większe obciążenia są przydzielane zasobom IT o większych możliwościach przetwarzania

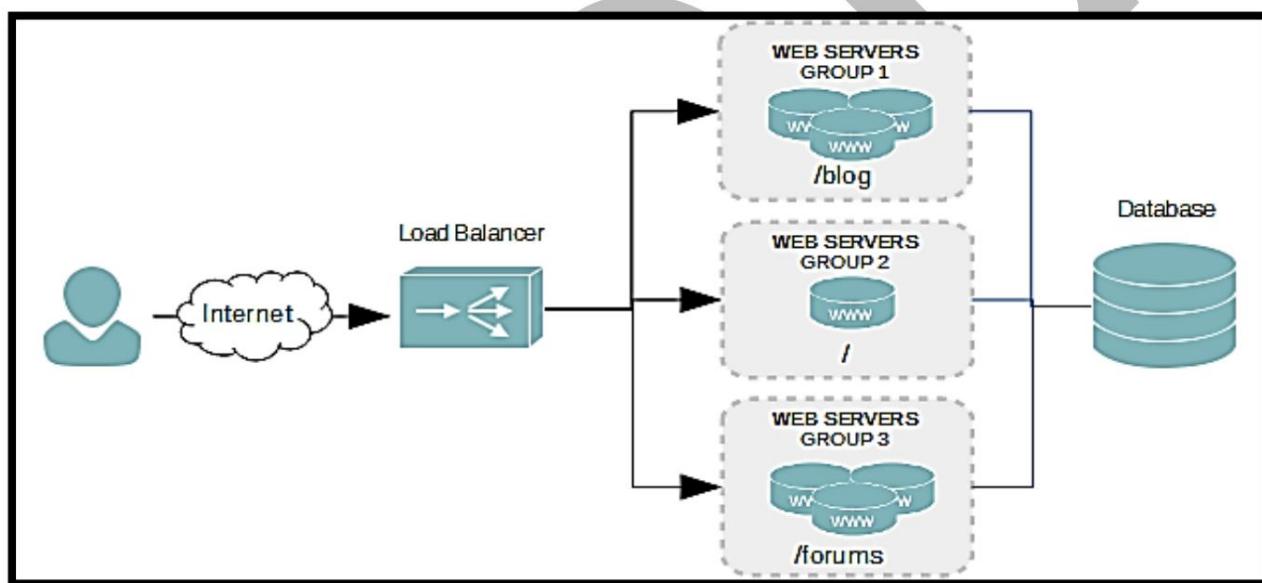
- Priorytety zadań – obciążenia są planowane, umieszczane w kolejce, odrzucane i rozdzielane zgodnie z ich poziomami priorytetów
- Dystrybucja uwzględniająca treść – żądania są dystrybuowane do różnych zasobów IT, np. podyktywane treścią żądania

Wspólne cele stosowania modułów równoważenia obciążenia to:

- Aby zachować sztywność systemu.
- Aby poprawić wydajność systemu.
- Aby chronić przed awariami systemu.

3.1.4.1 Jak działa równoważenie obciążenia?

W tym przypadku obciążenie odnosi się nie tylko do ruchu w witrynie, ale obejmuje również obciążenie procesora, obciążenie sieci i pojemność pamięci każdego serwera. Technika równoważenia obciążenia zapewnia, że każdy system w sieci ma taką samą ilość pracy w dowolnym momencie. Oznacza to, że żaden z nich nie jest ani nadmiernie przeciążony, ani niedostatecznie wykorzystany. [1]



Rys. 3.1.4.1

Mechanizmy równoważenia obciążenia mogą istnieć jako:

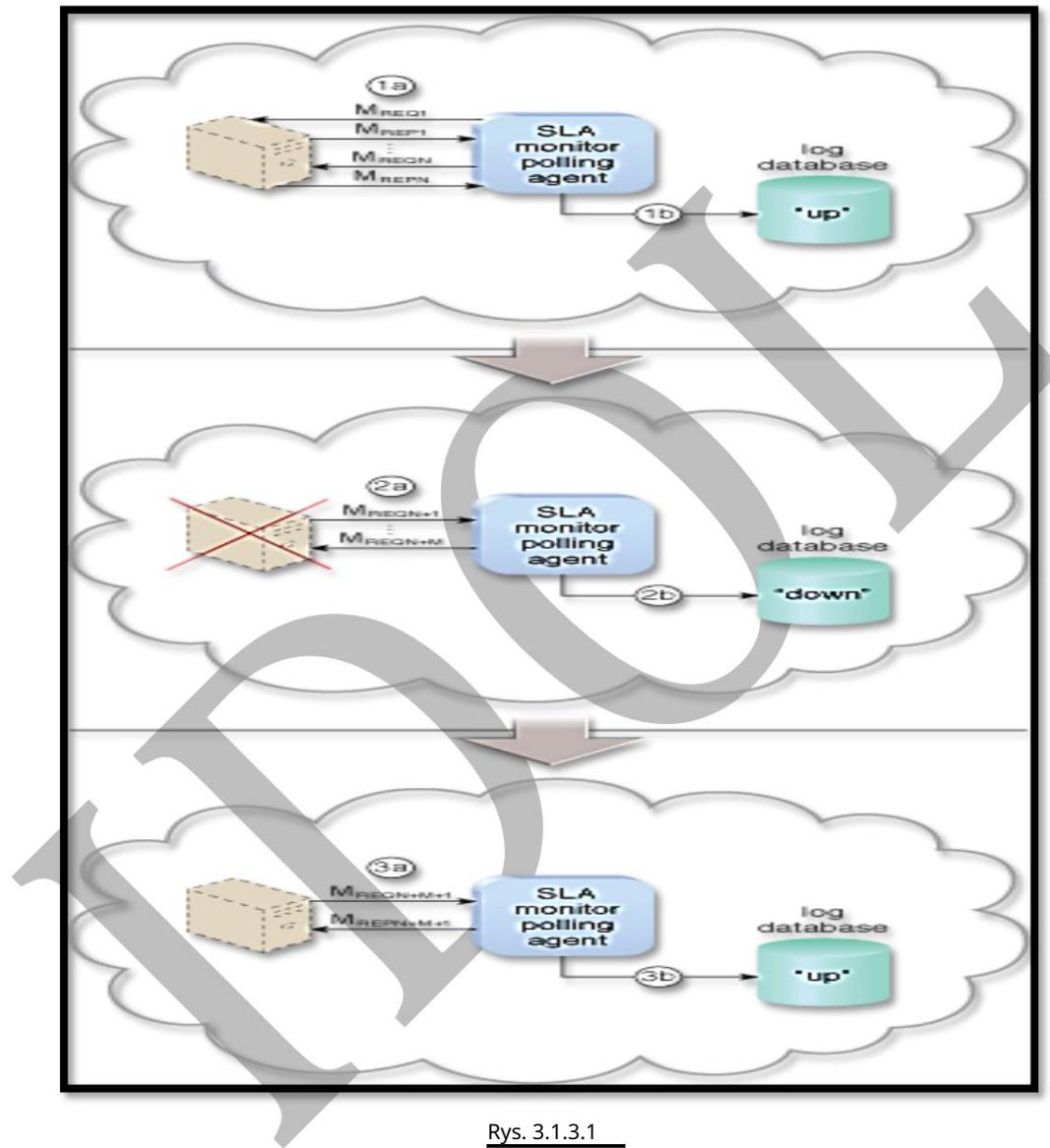
- Wielowarstwowy przełącznik sieciowy
- Dedykowane urządzenie sprzętowe
- Dedykowany system oparty na oprogramowaniu.
- Agent serwisowy

3.1.3.1 Monitor SLA:

Mechanizm systemu zarządzania SLA reprezentuje szereg dostępnych na rynku produktów do zarządzania chmurą, które zapewniają funkcje związane z administracją, gromadzeniem, przechowywaniem, raportowaniem i powiadamianiem o danych SLA w czasie wykonywania.

Wdrożenie systemu zarządzania SLA będzie zazwyczaj obejmować repozytorium używane do przechowywania i pobierania zebranych danych SLA w oparciu o wcześniej zdefiniowane wskaźniki i parametry raportowania. Będzie dalej

polegać na jednym lub większej liczbie mechanizmów monitorowania SLA w celu gromadzenia danych SLA, które można następnie udostępniać w czasie zbliżonym do rzeczywistego portalom użytkowym i administracyjnym w celu zapewnienia bieżącej informacji zwrotnej na temat aktywnych usług w chmurze (rysunek 3.1.3.1). Wskaźniki monitorowane dla poszczególnych usług w chmurze są zgodne z gwarancjami SLA zawartymi w odpowiednich umowach o świadczenie usług w chmurze.



Rys. 3.1.3.1

Rysunek 3.1.3.1 – Konsument usługi w chmurze wchodzi w interakcję z usługą w chmurze (1). Monitor SLA przechwytuje wymieniane wiadomości, ocenia interakcję i zbiera odpowiednie dane wykonawcze w odniesieniu do gwarancji jakości usług określonych w umowie SLA usługi w chmurze (2A). Zebrane dane przechowywane są w repozytorium (2B) będącym częścią systemu zarządzania SLA (3). Można wysyłać zapytania i generować raporty dla zewnętrznego administratora zasobów w chmurze za pośrednictwem portalu użytkownika i administracji (4) lub dla wewnętrznego administratora zasobów w chmurze za pośrednictwem natywnego interfejsu użytkownika systemu zarządzania SLA (3.1).

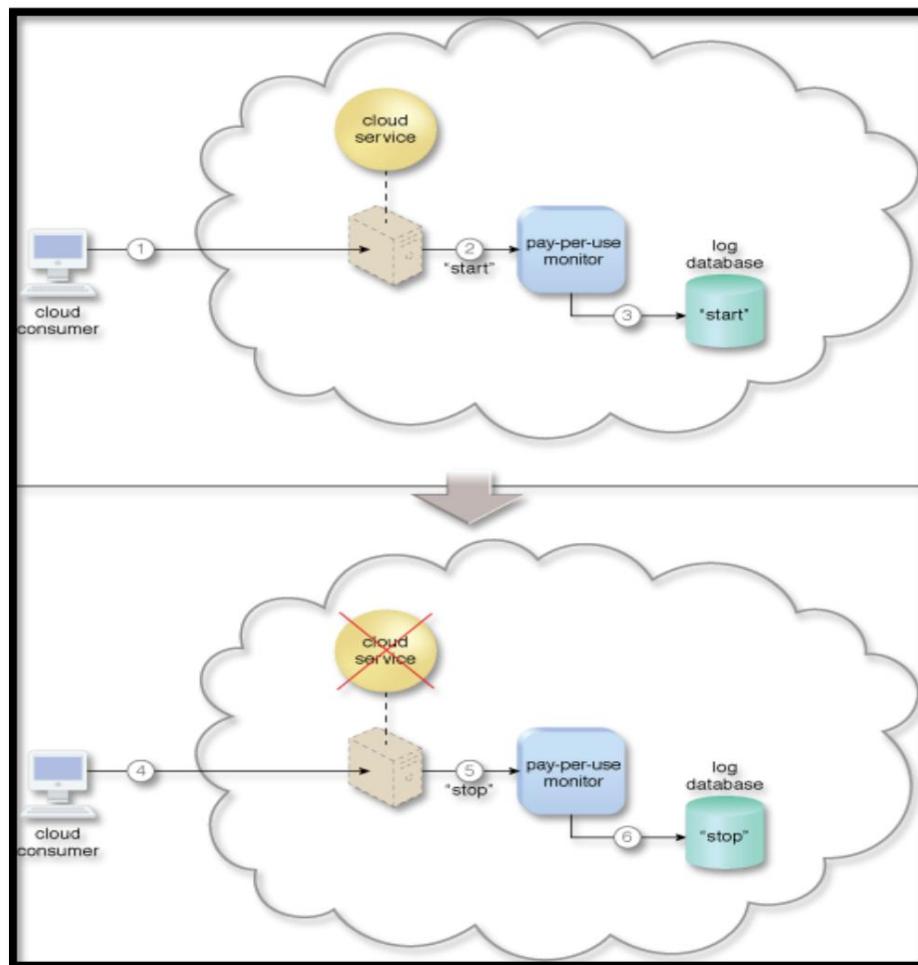
3.1.6 Monitor płatności za użycie:

Mechanizm monitorowania płatności za użycie mierzy wykorzystanie zasobów IT w chmurze zgodnie z predefiniowanymi parametrami cenowymi i generuje dzienniki użytkowania na potrzeby kalkulacji opłat i celów rozliczeniowych. Niektóre typowe zmienne monitorowane to:

- ilość komunikatów żądania/odpowiedź
- ilość przesyłanych danych
- zużycie pasma

Dane zebrane przez monitor płatności za użycie są przetwarzane przez system zarządzania rozliczeniami, który oblicza opłaty za płatności.

Rysunek 3.1.6 przedstawia monitor płatności za użycie zaimplementowany jako agent zasobów używany do określania okresu użytkowania serwera wirtualnego.



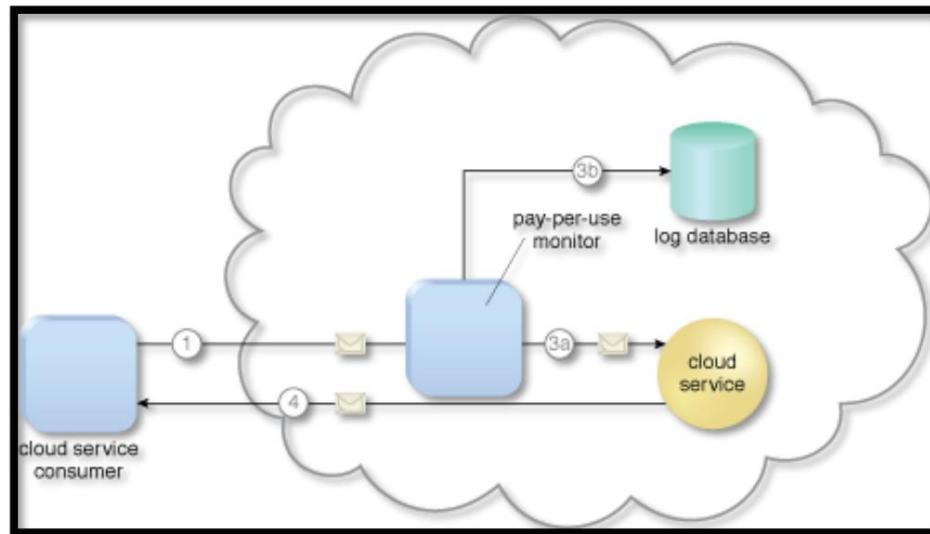
Rys. 3.1.6.1

Rysunek 3.1.6.1 – Konsument chmury żąda utworzenia nowej instancji usługi w chmurze (1).

Tworzona jest instancja zasobu IT, a mechanizm monitorowania płatności za użycie otrzymuje powiadomienie o zdarzeniu „startowym” z oprogramowania zasobu (2). Monitor płatności za użycie przechowuje znacznik czasu wartości w bazie danych dziennika (3). Konsument chmury żąda później zatrzymania instancji usługi w chmurze

(4). Monitor płatności za użycie otrzymuje powiadomienie o zdarzeniu „zatrzymania” z oprogramowania zasobów (3.1) i zapisuje znacznik czasu wartości w bazie danych dziennika (6).

Rysunek 3.1.6 ilustruje monitor płatności za użycie zaprojektowany jako agent monitorowania, który w przejrzysty sposób przechwytuje i analizuje komunikację w czasie wykonywania z usługą w chmurze.

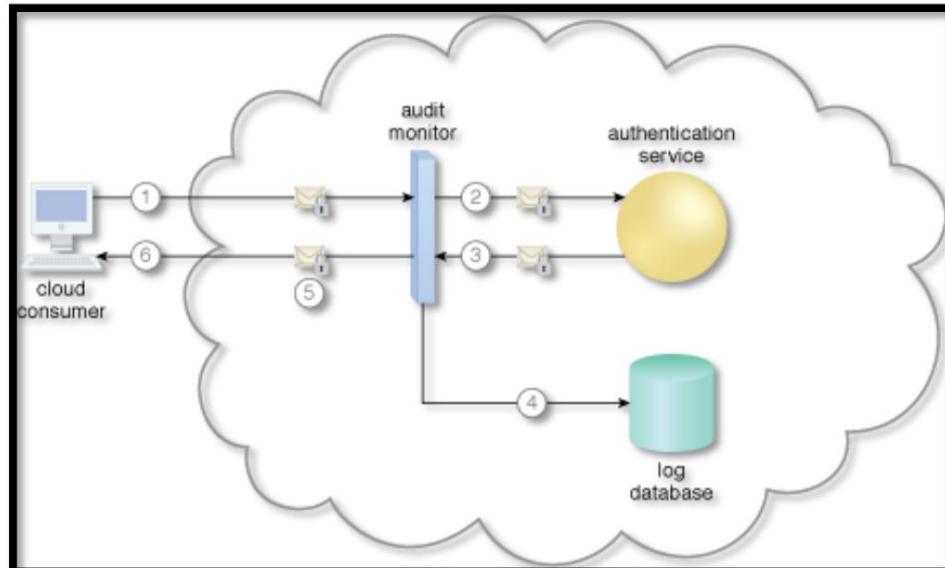


Rys. 3.1.6.2

Rysunek 3.1.6.2 – Konsument usługi w chmurze wysyła komunikat z żądaniem do usługi w chmurze (1). Monitor płatności za użycie przechwytuje wiadomość (2), przekazuje ją do usługi w chmurze (3a) i przechowuje informacje o użytkowaniu zgodnie ze swoimi metrykami monitorowania (3b). Usługa w chmurze przekazuje komunikaty odpowiedzi z powrotem do konsumenta usługi w chmurze, aby świadczył żądaną usługę (4).

3.1.7 Monitor audytu:

Mechanizm monitorowania audytu służy do gromadzenia danych śledzących audit dla sieci i zasobów IT w celu wsparcia lub podyktowanych zobowiązaniemi regulacyjnymi i umownymi. Rysunek przedstawia monitor audytu zaimplementowany jako agent monitorujący, który przechwytuje żądania „logowania” i przechowuje dane uwierzytelniające osoby żądającej, a także zarówno nieudane, jak i udane próby logowania w bazie danych dzienników na potrzeby przyszłych raportów audytowych.



Ryc. 3.1.7

Konsument usługi w chmurze żąda dostępu do usługi w chmurze, wysyłając wiadomość z żądaniem logowania zawierającą dane uwierzytelniające (1). Monitor audytu przechwytuje wiadomość (2) i przekazuje ją do usługi uwierzytelniania (3). Usługa uwierzytelniania przetwarza poświadczenia bezpieczeństwa. Oprócz wyników próby logowania generowany jest komunikat odpowiedzi dla konsumenta usługi w chmurze (4). Monitor audytu przechwytuje wiadomość odpowiedzi i przechowuje wszystkie zebrane szczegóły zdarzeń logowania w bazie danych logów, zgodnie z wymogami polityki audytu organizacji (3.1).

Dostęp został przyznany, a odpowiedź jest wysyłana z powrotem do konsumenta usługi w chmurze (6).

3.1.8 System awaryjny:

Mechanizm systemu Failover służy do zwiększenia niezawodności i dostępności zasobów IT poprzez wykorzystanie uznanej technologii klastrowania w celu zapewnienia redundantnych wdrożeń. System przełączania awaryjnego jest skonfigurowany tak, aby automatycznie przełączał się na nadmiarową lub rezerwową instancję zasobu IT, gdy aktualnie aktywny zasób IT stanie się niedostępny.

Systemy przełączania awaryjnego są powszechnie stosowane w programach o znaczeniu krytycznym lub w usługach wielokrotnego użytku, które mogą spowodować powstanie pojedynczego punktu awarii w wielu aplikacjach. System przełączania awaryjnego może obejmować więcej niż jeden region geograficzny, tak że w każdej lokalizacji znajduje się jedna lub więcej nadmiarowych implementacji tego samego zasobu IT.

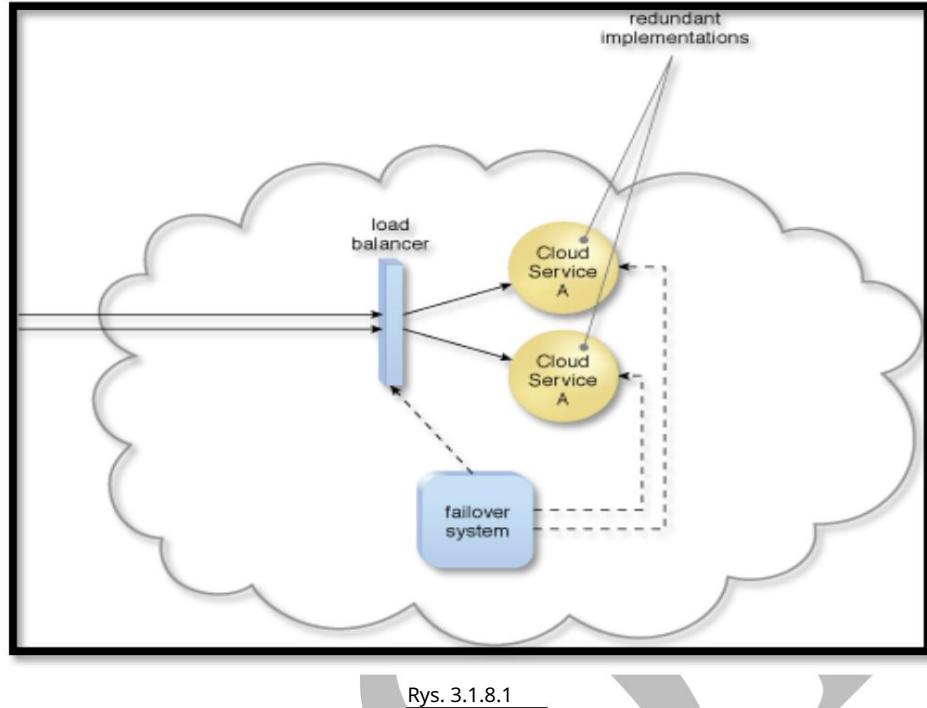
Mechanizm ten może opierać się na mechanizmie replikacji zasobów w celu dostarczenia nadmiarowych instancji zasobów IT, które są aktywnie monitorowane pod kątem wykrywania błędów i stanów niedostępności.

3.1.8.1 Systemy przełączania awaryjnego występują w dwóch podstawowych konfiguracjach:

A. Aktywny-aktywny

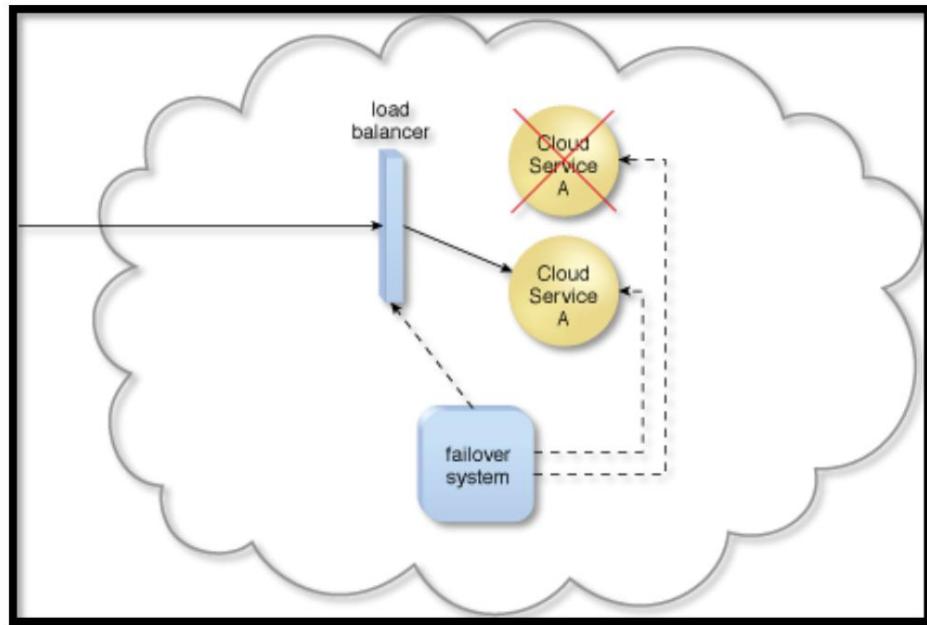
W konfiguracji aktywny-aktywny nadmiarowe implementacje zasobów IT aktywnie i synchronicznie obsługują obciążenie (rysunek 3.1.8.1). Wymagane jest równoważenie obciążenia pomiędzy aktywnymiinstancjami.

Po wykryciu awarii uszkodzona instancja jest usuwana z harmonogramu równoważenia obciążenia (rysunek 3.1.8.2). Niezależnie od tego, który zasób IT pozostanie sprawny w momencie wykrycia awarii, przejmuje przetwarzanie (rysunek 3.1.8.3).



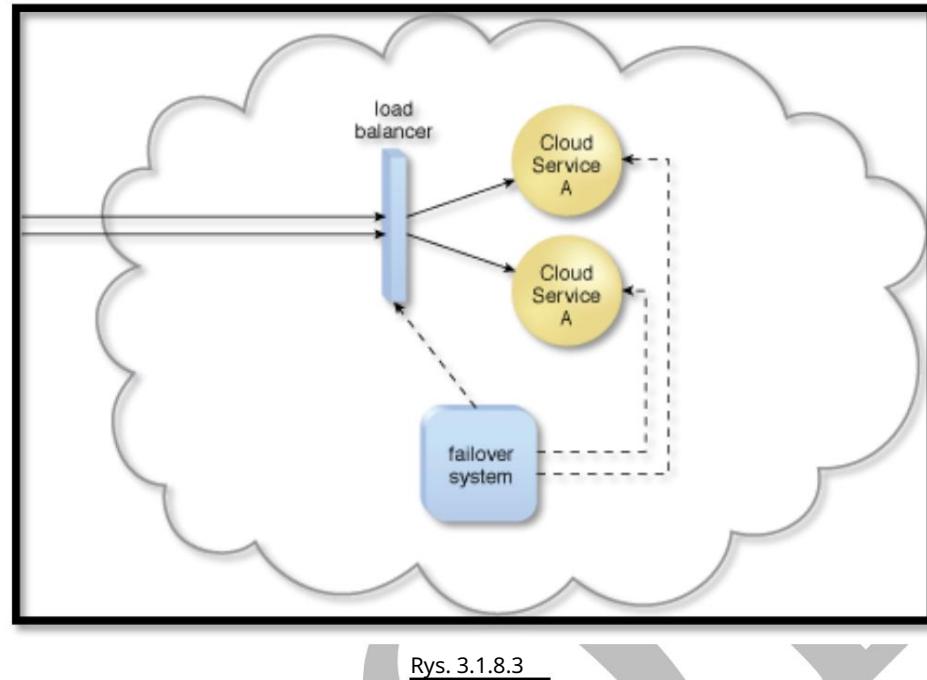
Rys. 3.1.8.1

Rysunek 3.1.8.1 – System przełączania awaryjnego monitoruje stan operacyjny Usługi Przetwarzania w Chmurze A.



Rys. 3.1.8.2

Rysunek 3.1.8.2 – W przypadku wykrycia awarii w jednej implementacji Usługi Przetwarzania w Chmurze A, system przełączania awaryjnego wydaje modułowi równoważenia obciążenia polecenie przełączenia obciążenia do nadmiarowej implementacji Usługi Przetwarzania w Chmurze A.



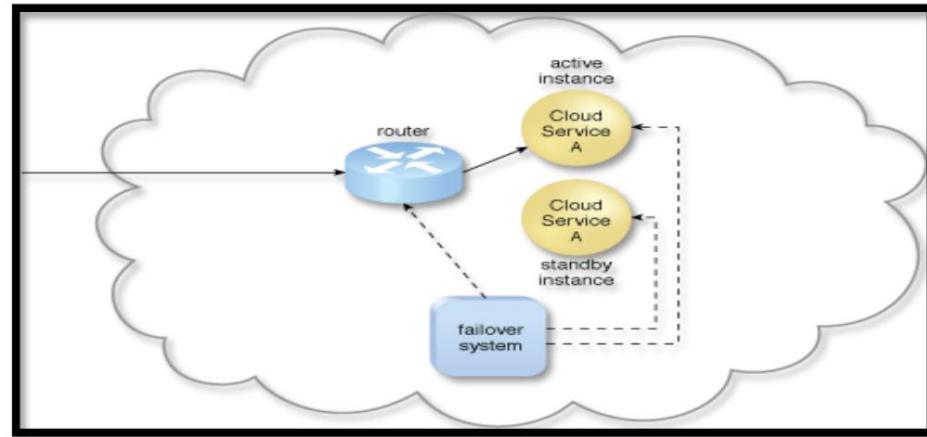
Rys. 3.1.8.3

Rysunek 3.1.8.3 – Usługa w chmurze, która uległa awarii. Implementacja została odtworzona lub zreplikowana do działającej usługi w chmurze. System przełączania awaryjnego nakazuje teraz modułowi równoważenia obciążenia ponowne rozłożenie obciążenia.

B. Aktywno-pasywnie

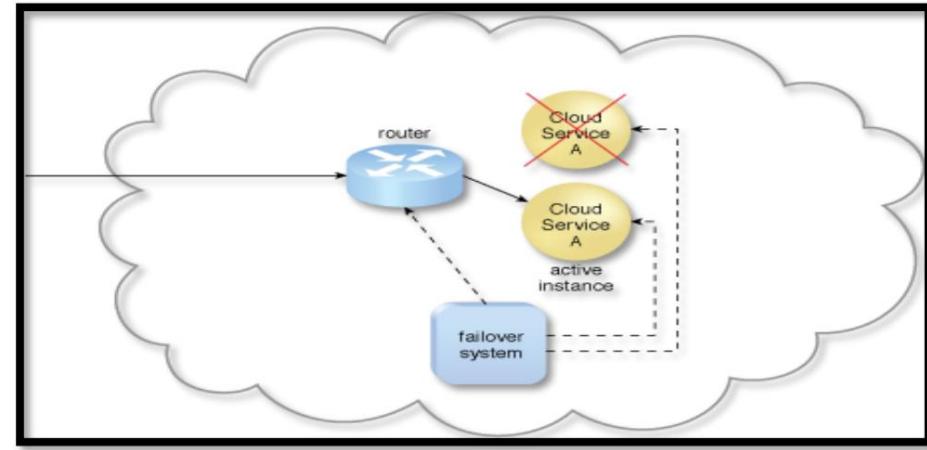
W konfiguracji aktywnej-pasywnej aktywowana jest implementacja rezerwowa lub nieaktywna, która przejmie przetwarzanie od niedostępnego zasobu IT, a odpowiednie obciążenie zostanie przekierowane do instancji przejmującej operację (rysunki 4 do 3.1).

Niektóre systemy przełączania awaryjnego są zaprojektowane do przekierowywania obciążeń do aktywnych zasobów IT, które opierają się na wyspecjalizowanych modułach równoważenia obciążenia, które wykrywają warunki awarii i wykluczają instancje zasobów IT, które uległy awarii, z dystrybucją obciążenia. Ten typ systemu awaryjnego jest odpowiedni dla zasobów IT, które nie wymagają zarządzania stanem wykonania i zapewniają możliwości przetwarzania bezstanowego. W architekturach technologicznych, które zazwyczaj opierają się na technologiach klastrowych i wirtualizacyjnych, wymagane jest również, aby implementacje nadmiarowych lub rezerwowych zasobów IT udostępniały swój stan i kontekst wykonawczy. Złożone zadanie, które zostało wykonane na uszkodzonym zasobie IT, może nadal działać w jednym, jeśli ma redundantne wdrożenia.



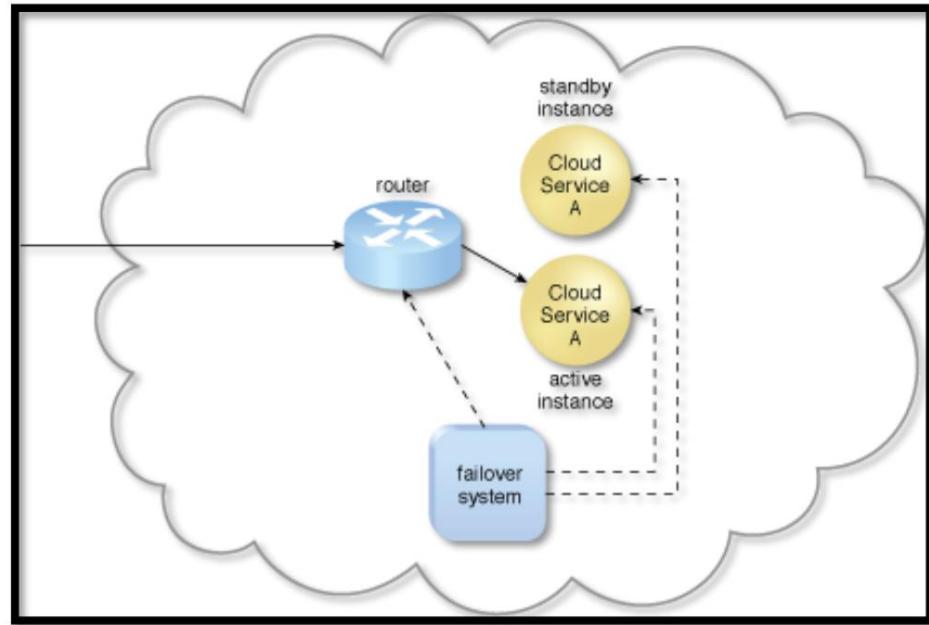
Rys. 3.1.8.4

Rysunek 3.1.8.4 – System przełączania awaryjnego monitoruje status operacyjny Usługi Przetwarzania w Chmurze A. Usługa Przetwarzania w Chmurze Implementacja działająca jako aktywna实施acja odbiera żądanie konsumentów usługi w chmurze.



Rys. 3.1.8.3.1

Rysunek 3.1.8.3.1 – Usługa Przetwarzania w Chmurze Implementacja działająca jako aktywna实施acja napotyka awarię wykrytą przez system przełączania awaryjnego, który następnie aktywuje nieaktywną implementację Usługi Przetwarzania w Chmurze A i przekierowuje na nią obciążenie. Nowo wywołana implementacja usługi w chmurze przyjmuje teraz rolę aktywnej实施acji.



Rys. 3.1.8.6

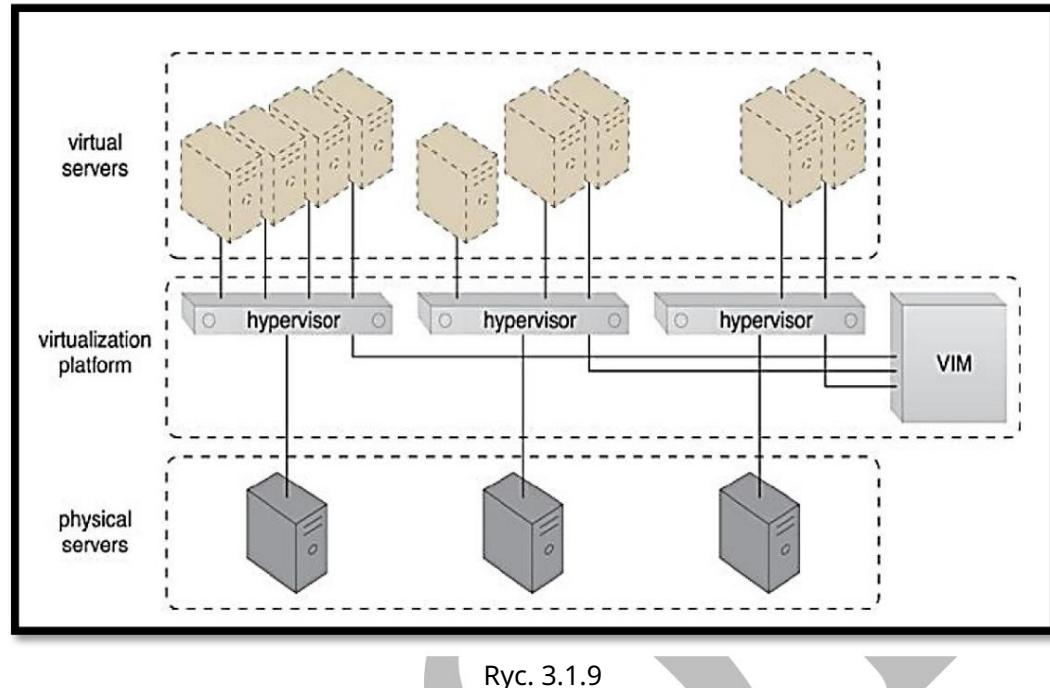
Rysunek 3.1.8.6 – Awaria Usługi Przetwarzania w Chmurze Implementacja jest odzyskiwana lub replikowana do działającej usługi w chmurze i jest teraz ustawiana jako instancja rezerwowa, podczas gdy poprzednio wywołana Usługa Przetwarzania w Chmurze A nadal służy jako instancja aktywna.

3.1.9 Hiperwizor

Hiperwizor to technika wirtualizacji sprzętu, która umożliwia jednocześnie działanie wielu systemów operacyjnych gościa na jednym hostie. System operacyjny gościa współużytkuje sprzęt komputera hosta, tak że każdy system operacyjny wydaje się mieć własny procesor, pamięć i inny sprzęt zasoby.

Hiperwizor jest również nazywany menedżerem maszyny wirtualnej (VMM).

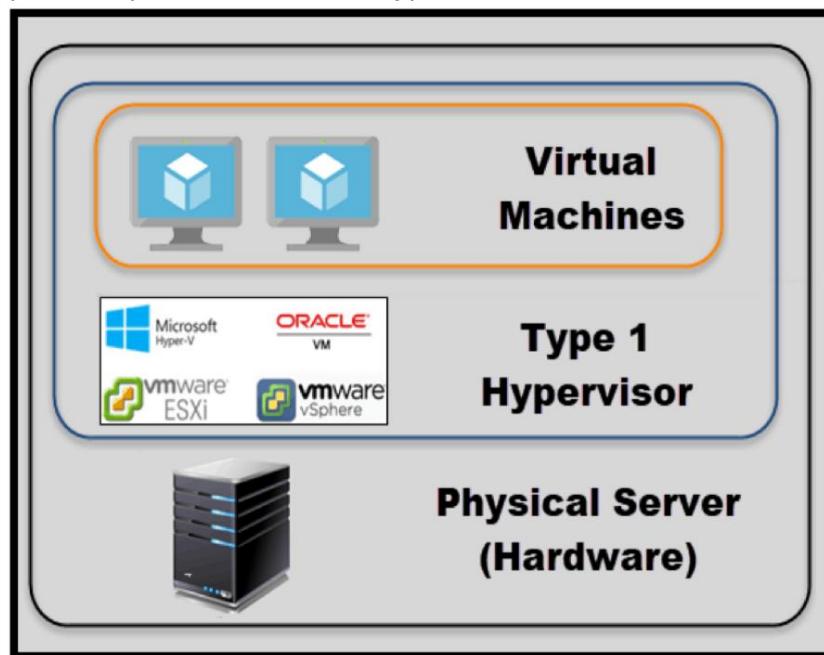
Hiperwizor izoluje systemy operacyjne od głównego komputera-hosta. Zadaniem hypervisor jest zaspokajanie potrzeb systemu operacyjnego gościa i efektywne zarządzanie nim. Każda maszyna wirtualna jest niezależna i nie koliduje ze sobą, chociaż działa na tej samej maszynie hosta.[2]



Ryc. 3.1.9

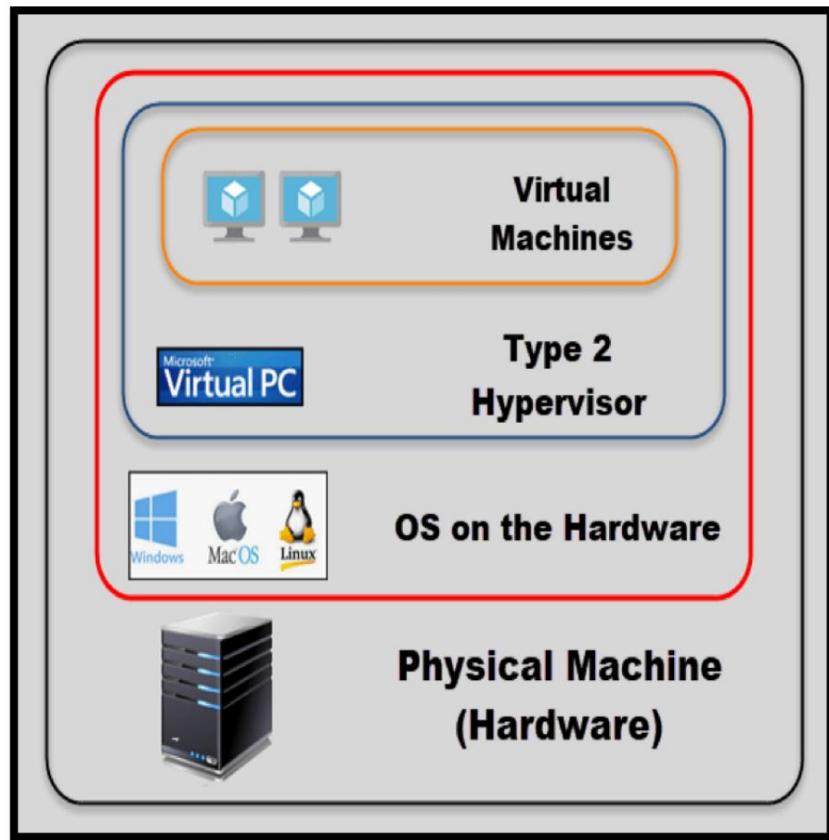
3.1.9.1 Hypervisory dzieli się na dwa typy:

O. Typ pierwszy to hiperwizor typu bare-metal , który jest wdrażany bezpośrednio na sprzęcie systemu hosta, bez żadnych systemów operacyjnych ani oprogramowania. Niektóre przykłady hiperwizorów typu 1 to hiperwizor Microsoft Hyper-V, VMware ESXi, Citrix XenServer.



Rys. 3.1.9.1.A

B. Typ drugi to hostowany hypervisor, który działa jako warstwa oprogramowania w fizycznym systemie operacyjnym. Hiperwizor działa jako osobna druga warstwa na sprzęcie, podczas gdy system operacyjny działa jako trzecia warstwa. Hostowane hypervisory obejmują Parallels Desktop i VMware Player.



Rys. 3.1.9.1.B

3.1.10 Klaster zasobów

Zasoby IT oparte na chmurze, które są zróżnicowane geograficznie, można logicznie łączyć w grupy, aby usprawnić ich alokację i wykorzystanie. Mechanizm klastra zasobów służy do grupowania wielu instancji zasobów IT, tak aby można je było obsługiwać jako pojedynczy zasób IT. Zwiększa to łączną moc obliczeniową, równoważenie obciążenia i dostępność klastrowych zasobów IT.

Architektury klastrów zasobów opierają się na szybkich dedykowanych połączeniach sieciowych, czyli węzłach klastra, pomiędzyinstancjami zasobów IT w celu komunikowania się na temat dystrybucji obciążenia, planowania zadań, udostępniania danych i synchronizacji systemu. Za te działania zwykle odpowiada platforma zarządzania klastrami, która działa jako rozproszone oprogramowanie pośredniczące we wszystkich węzłach klastra. Platforma ta realizuje funkcję koordynacyjną, która pozwala na występowanie rozproszonych zasobów IT jako jednego zasobu IT, a także realizuje zasoby IT wewnętrznych klastra.

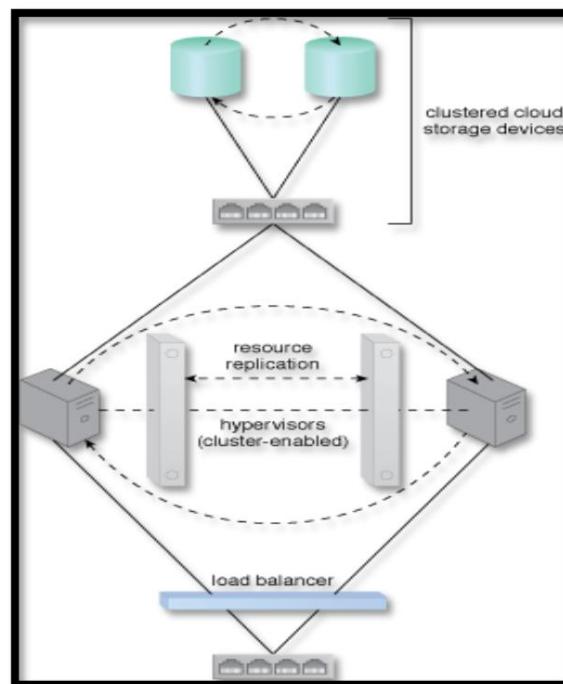
5.10.1 Typowe typy klastrów zasobów obejmują:

- A. Klaster serwerów — serwery fizyczne lub wirtualne łączy się w klastry w celu zwiększenia wydajności i dostępności. Hiperwizory działające na różnych serwerach fizycznych można skonfigurować tak, aby udostępniały stan wykonania serwera wirtualnego (taki jak strony pamięci i stan rejestrów procesora) w celu ustanowienia klastrowych serwerów wirtualnych. W takich konfiguracjach, które zwykle wymagają, aby serwery fizyczne miały dostęp do współdzielonej pamięci, serwery wirtualne mogą migrować na żywo między sobą. W procesie tym platforma wirtualizacyjna wstrzymuje wykonanie danego zadania

serwer wirtualny na jednym serwerze fizycznym i wznawia go na innym serwerze fizycznym. Proces ten jest przezroczysty dla systemu operacyjnego serwera wirtualnego i można go wykorzystać do zwiększenia skalowalności poprzez migrację na żywo serwera wirtualnego działającego na przeciążonym serwerze fizycznym na inny serwer fizyczny o odpowiedniej wydajności.

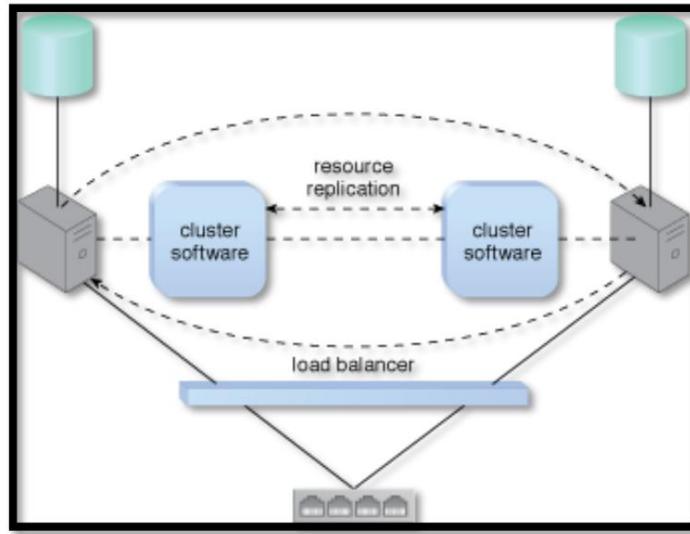
- B. Klaster bazy danych — ten klaster zasobów o wysokiej dostępności, zaprojektowany w celu poprawy dostępności danych, posiada funkcję synchronizacji, która utrzymuje spójność danych przechowywanych na różnych urządzeniach pamięci masowej używanych w klastrze. Nadmiarowa pojemność jest zwykle oparta na systemie przełączania awaryjnego typu aktywny-aktywny lub aktywny-pasywny, którego zadaniem jest utrzymanie warunków synchronizacji.
- C. Klaster dużych zbiorów danych – partycjonowanie i dystrybucja danych są realizowane w taki sposób, aby docelowe zbiory danych mogły być efektywnie partycjonowane bez uszczerobków dla integralności danych i dokładności obliczeń. Każdy węzeł klastra przetwarza obciążenia bez komunikowania się z innymi węzłami tak często, jak w klastrach innych typów.

Wiele klastrów zasobów wymaga, aby węzły klastra miały niemal identyczną moc obliczeniową i charakterystykę, aby uprościć projektowanie i zachować spójność w architekturze klastra zasobów. Węzły klastra w architekturach klastrów o wysokiej dostępności muszą uzyskiwać dostęp do wspólnych zasobów IT pamięci masowej i je współużytkować. Może to wymagać dwóch warstw komunikacji między węzłami — jednej umożliwiającej dostęp do urządzenia pamięci masowej i drugiej umożliwiającej orkiestrację zasobów IT (rysunek 3.1.10.1). Niektóre klastry zasobów są zaprojektowane z luźno powiązanymi zasobami IT, które wymagają jedynie warstwy sieciowej (rysunek 2).



Rys. 3.1.10.1

Rysunek 3.1.10.1 – Równoważenie obciążenia i replikacja zasobów są realizowane za pośrednictwem hiperwizora obsługującego klaster. Dedykowana sieć pamięci masowej służy do łączenia klastrowej pamięci masowej i serwerów w klastrze, które mogą współużytkować wspólne urządzenia pamięci masowej w chmurze. Upraszczają to proces replikacji pamięci masowej, który odbywa się niezależnie w klastrze pamięci masowej.



Rys. 3.1.10.2

Rysunek 3.1.10.2 – Luźno powiązany klaster serwerów zawierający moduł równoważenia obciążenia. Nie ma wspólnego magazynu. Replikacja zasobów służy do replikowania urządzeń pamięci masowej w chmurze za pośrednictwem sieci za pomocą oprogramowania klastra.

3.1.10.2 Istnieją dwa podstawowe typy klastrów zasobów:

- A. Klaster ze zrównoważonym obciążeniem – ten klaster zasobów specjalizuje się w rozdzielaniu obciążień pomiędzy węzłami klastra w celu zwiększenia wydajności zasobów IT przy jednoczesnym zachowaniu centralizacji zarządzania zasobami IT. Zwykle implementuje mechanizm równoważenia obciążenia, który jest wbudowany w platformę zarządzania klastrem lub skonfigurowany jako oddzielnny zasób IT.
- B. Klaster o wysokiej dostępności (HA) – klaster o wysokiej dostępności utrzymuje dostępność systemu w przypadku awarii wielu węzłów i posiada nadmiarowe implementacje większości wszystkich zasobów IT klastra. Implementuje mechanizm systemu przełączania awaryjnego, który monitoruje warunki awarii i automatycznie przekierowuje obciążenie z dala od uszkodzonych węzłów.

3.1.11 Broker wielu urządzeń

Dostęp do poszczególnych usług w chmurze może wymagać różnych typów konsumentów usług w chmurze, a niektórzy z nich mogą być niezgodni z opublikowaną umową o świadczenie usług w chmurze. Różni konsumenti usług w chmurze mogą różnić się urządzeniami hostingowymi i/lub mogą mieć różne rodzaje wymagań komunikacyjnych. Aby przewyściżyć niezgodności między usługą w chmurze a odmiennym konsumentem usług w chmurze, należy utworzyć logikę mapowania w celu przekształcenia (lub przekształcenia) informacji wymienianych w czasie wykonywania.

Mechanizm brokera obsługującego wiele urządzeń ułatwia transformację danych w czasie wykonywania, tak aby usługa w chmurze była dostępna dla szerszej gamy programów i urządzeń konsumenckich usług w chmurze (rysunek 1).

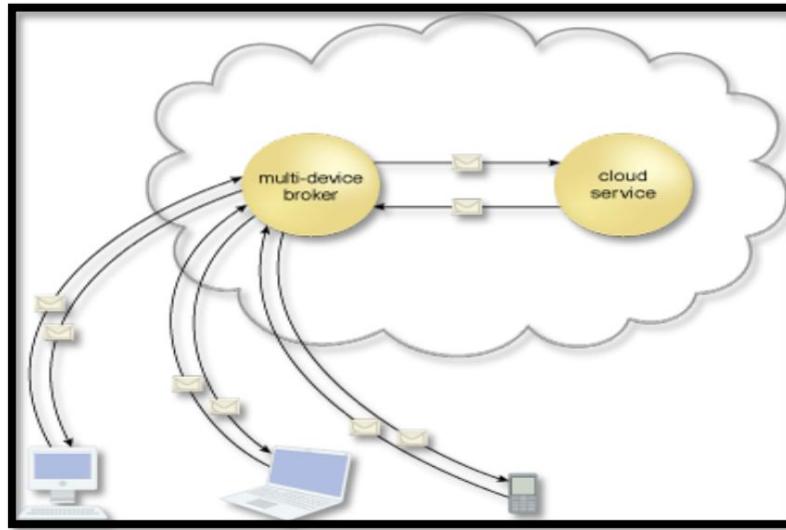
Brokerzy obsługujący wiele urządzeń powszechnie istnieją lub zawierają komponenty bram, takie jak:

- Bramka XML – przesyła i sprawdza dane XML
- Cloud Storage Gateway – przekształca protokoły przechowywania w chmurze i koduje urządzenia pamięci masowej ułatwiając przesyłanie i przechowywanie danych
- Brama urządzeń mobilnych – zmienia protokoły komunikacyjne używane przez urządzenia mobilne

Poziomy, na których można tworzyć logikę transformacji, obejmują:

- protokoły transportowe
- protokoły przesyłania wiadomości
- protokoły urządzeń pamięci masowej
- schematy/modeli danych

Na przykład broker obsługujący wiele urządzeń może zawierać logikę mapowania, która obejmuje protokoły transportu i przesyłania wiadomości dla konsumenta usług w chmurze uzyskującego dostęp do usługi w chmurze za pomocą urządzenia mobilnego.



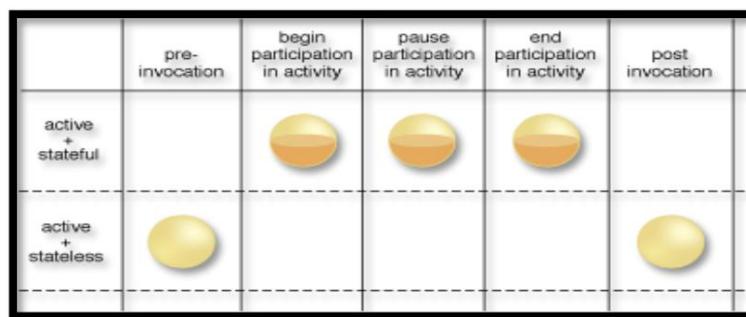
Ryc. 3.1.11

Rysunek 3.1.11 – Broker obsługujący wiele urządzeń zawiera logikę mapowania niezbędną do transformacji wymiany danych między usługą w chmurze a różnymi typami urządzeń konsumenckich usług w chmurze.

3.1.12 Baza danych zarządzania stanem

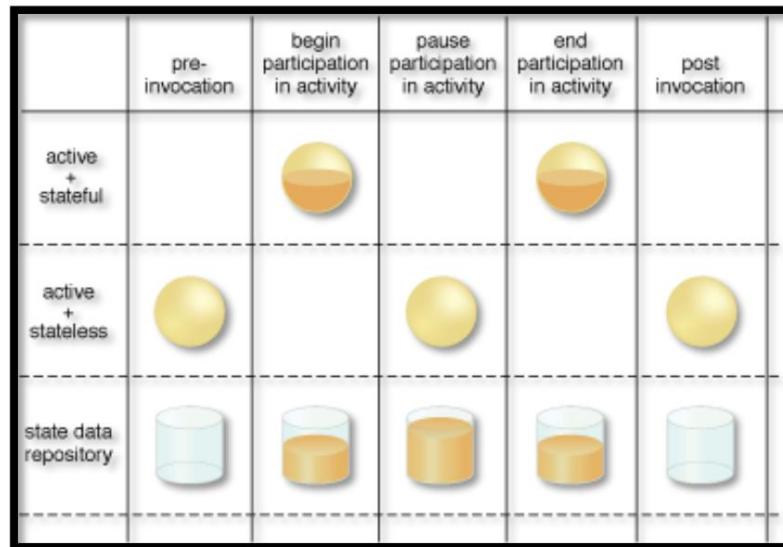
Baza danych zarządzania stanem to urządzenie pamięci masowej używane do tymczasowego utrwalania danych stanu programów.

Alternatywą dla buforowania danych o stanie w pamięci, programy mogą przenosić dane o stanie do bazy danych, aby zmniejszyć ilość zużywanej pamięci wykonawczej (rysunki 3.1.12.1 i 3.1.12.2). Dzięki temu programy i otaczająca je infrastruktura są bardziej skalowalne. Bazy danych zarządzania stanem są powszechnie używane przez usługi w chmurze, szczególnie te zaangażowane w długotrwałe działania wykonawcze.



Rys. 3.1.12.1

Rysunek 3.1.12.1 – W okresie istnienia instancji usługi w chmurze może być wymagane zachowanie stanu i przechowywanie danych stanu w pamięci podręcznej, nawet w stanie bezczynności.



Rys. 3.1.12.2

Rysunek 3.1.12.2 – Odkładając dane stanu do repozytorium stanu, usługa w chmurze może przejść do stanu bezstanowego (lub stanu częściowo bezstanowego), tym samym tymczasowo zwalniając system zasoby.



3.1.13 Przykładowe ćwiczenie z pytaniem

1. Wyjaśnij odbiornik automatycznego skalowania.
2. Omów studium przypadku dotyczące DTGOV. Wyjaśnij szczegółowo.
3. Na czym polega praca z równoważeniem obciążenia? Jak działa równoważenie obciążenia?
4. Wyjaśnij szczegółowo monitor SLA.
- 3.1. Wyjaśnij szczegółowo monitor płatności za użycie.
6. Wyjaśnij szczegółowo monitora audytu.
7. Co to jest system awaryjny? Omów szczegółowo jego rodzaje.
8. Co to jest Hypervisor? Omów szczegółowo jego rodzaje.
9. Czym jest klaster zasobów?
10. Wyjaśnij klaster wspólnego typu i podstawowy typ klastra zasobów.
11. Wyjaśnij szczegółowo brokera wielourządzeniowego.
12. Wyjaśnij szczegółowo bazę danych zarządzania państwem.

BIBLIOGRAFIA: 1.

1. <https://www.znetlive.com/blog/what-is-load-balancing-in-cloud-computing-and-its-zalety/>

2. <https://www.cloudoye.com/kb/general/what-is-hypervisor-in-cloud-computing-and-its-typy/>

3. [https://patterns.arcitura.com/cloud-computing-wzorce/mechanizmy/baza danych_zarzadzania stanem](https://patterns.arcitura.com/cloud-computing-wzorce/mechanizmy/baza-danych_zarzadzania_stanem)

JEDNOSTKA III
ROZDZIAŁ 2
MECHANIZMY ZARZĄDZANIA CHMURĄ I MECHANIZMY BEZPIECZEŃSTWA CHMURY

Struktura jednostki:

- 3.2.1 Cel 3.2.2
- Wprowadzenie
- 3.2.3 System zdalnej administracji
- 3.2.4 System zarządzania zasobami
- 3.2.5 System zarządzania SLA
- 3.2.3.2 System zarządzania rozliczeniami
- 3.2.7 Szyfrowanie
 - 3.2.7.1 Szyfrowanie symetryczne
 - 3.2.7.2 Szyfrowanie asymetryczne
- 3.2.8 Haszowanie
- 3.2.9 Podpis cyfrowy
- 3.2.10 Infrastruktura klucza publicznego (PKI)
- 3.2.11 Zarządzanie tożsamością i dostępem (IAM)
- 3.2.12 Pojedyncze logowanie (SSO)
- 3.2.13 Grupy zabezpieczeń oparte na chmurze
- 3.2.14 Wzmocnione obrazy serwerów wirtualnych
- 3.2.15 Przykładowe ćwiczenie z pytaniem
- 3.2.17 Odniesienia

3.2.1 Cel:

Aby zrozumieć problemy bezpieczeństwa i zidentyfikować odpowiednie techniki bezpieczeństwa używane w obecnym świecie przetwarzania w chmurze. Aby zidentyfikować wyzwania związane z bezpieczeństwem oczekiwanych w przyszłości przetwarzania w chmurze. Zasugerowanie pewnych środków zaradczych na przyszłość wyzwania stojące przed chmurą obliczeniową.

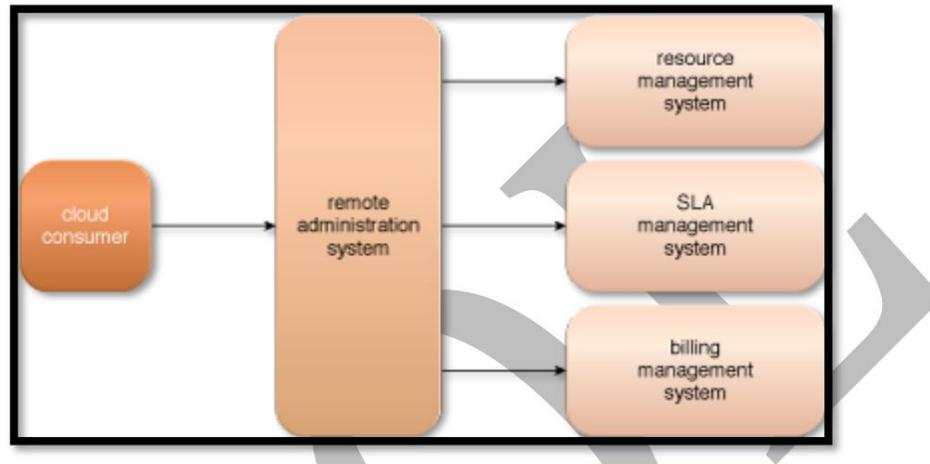
3.2.2 Wprowadzenie:

Mechanizmy zarządzania chmurą to środki, które należy podjąć, aby zapewnić istnienie mechanizmów bezpieczeństwa rozwiązań chmurowych pozwalających stawić czoła atakom i zagrożeniom bezpieczeństwa. Mechanizmy zarządzania chmurą mogą pomóc w ułatwieniu kontroli, zarządzania i ewolucji technologii chmury oraz zasobów IT, które stanowią część platform i rozwiązań chmurowych. Ponieważ zasoby IT oparte na chmurze muszą być konfigurowane, konfigurowane, utrzymywane i monitorowane, istnieją systemy i mechanizmy, które powinny istnieć do zarządzania tymi zadaniami. Omówiono mechanizmy zarządzania tymi systemami, które zazwyczaj zapewniają zintegrowane interfejsy API, które mogą oferować indywidualne produkty, niestandardowe aplikacje, które można łączyć w różne zestawy produktów lub aplikacje wielofunkcyjne.

3.2.3 System zdalnej administracji:

Mechanizm zdalnego zarządzania systemem zapewnia narzędzia i interfejsy użytkownika dla administratorów zewnętrznych zasobów chmurowych, umożliwiające konfigurowanie i administrowanie zasobami IT opartymi na chmurze.

System zdalnej administracji może stworzyć portal umożliwiający dostęp do funkcji administracyjnych i zarządczych różnych systemów bazowych, w tym systemów zarządzania zasobami, zarządzania umowami SLA i systemów zarządzania rozliczeniami (rysunek 1).



Rys. 3.2.3.1

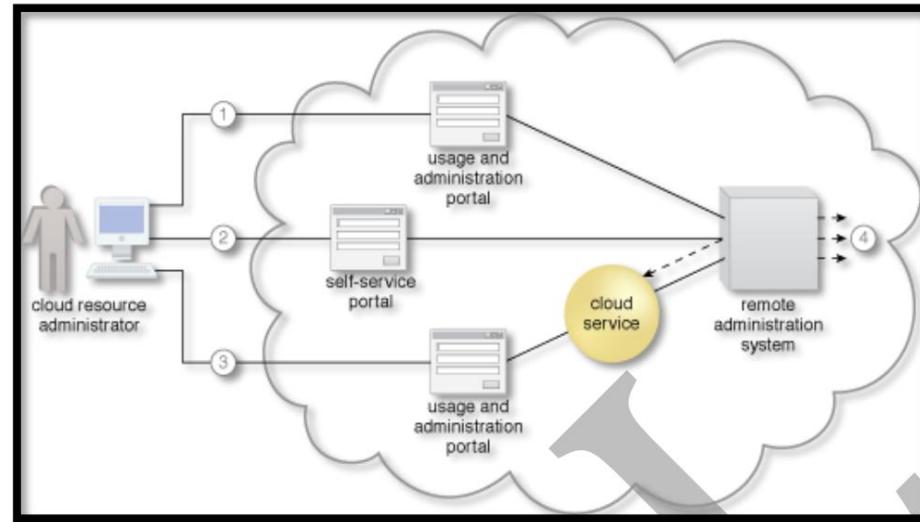
Rysunek 3.2.3.1 – System zdalnej administracji wyodrębnia podstawowe systemy zarządzania, aby udostępnić i skoncentrować kontrolę administracyjną zewnętrznym administratorom zasobów chmury. System zapewnia konfigurowalną konsolę użytkownika, a jednocześnie programowo współpracuje z podstawowymi systemami zarządzania za pośrednictwem ich interfejsów API.

Narzędzia i interfejsy API udostępniane przez system zdalnej administracji są zazwyczaj wykorzystywane przez dostawcę usług w chmurze do opracowywania i dostosowywania portali internetowych, które zapewniają konsumentom chmury różnorodne funkcje administracyjne.

Poniżej przedstawiono dwa podstawowe typy portali tworzonych za pomocą systemu administracji zdalnej:

- Portal użytkowania i administracji — portal ogólnego przeznaczenia, który skoncentrował kontrolę zarządzania różnymi zasobami IT w chmurze i może dodatkowo udostępniać raporty dotyczące wykorzystania zasobów IT.
- Portal samoobsługowy – jest to zasadniczo portal zakupowy, który umożliwia konsumentom chmury przeszukiwanie aktualnej listy usług chmurowych i zasobów IT dostępnych u dostawcy chmury (zwykle w ramach dzierżawy). Konsument chmury przesyła wybrane elementy dostawcy chmury w celu udostępnienia.

Rysunek 3.2.3.2 ilustruje scenariusz obejmujący system zdalnej administracji oraz portale użytkowe, administracyjne i samoobsługowe.



Rys. 3.2.3.2

Rysunek 3.2.3.2 – Administrator zasobów chmury korzysta z portalu użytkowania i administracji, aby skonfigurować już wynajęty serwer wirtualny (niepokazany) w celu przygotowania go do hostingu (1). Administrator zasobów chmury następnie korzysta z portalu samoobsługowego, aby wybrać nową usługę w chmurze i poprosić o udostępnienie jej (2). Administrator zasobów chmury następnie ponownie uzyskuje dostęp do portalu użytkowania i administracji, aby skonfigurować nowo udostępnioną usługę w chmurze hostowaną na serwerze wirtualnym (3). Na wszystkich etapach system zdalnej administracji współdziała z niezbędnymi systemami zarządzania w celu wykonania żądanych działań (4).

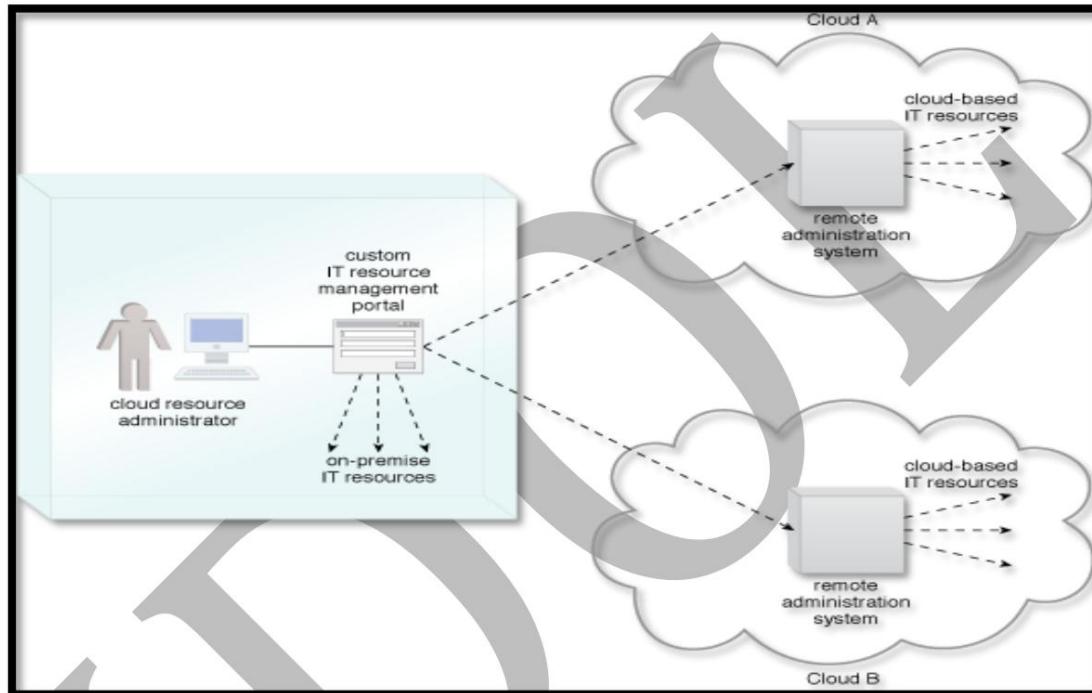
Zależy od:

- rodzaj produktu w chmurze lub model dostarczania chmury, który konsument chmury leasinguje lub z którego korzysta od dostawcy chmury,
 - poziom kontroli dostępu przyznany przez dostawcę chmury konsumentowi chmury oraz
 - dalej, w zależności od tego, które systemy zarządzania stanowią podstawę zdalnej administracji
- interfejsy systemowe,

Zadania, które użytkownicy chmury mogą zwykle wykonywać za pośrednictwem zdalnej konsoli administracyjnej, obejmują:

- konfigurowanie i konfigurowanie usług w chmurze
- udostępnianie i udostępnianie zasobów IT dla usług w chmurze na żądanie
- monitorowanie stanu usług w chmurze, wykorzystania i wydajności
- monitorowanie realizacji QoS i SLA
- zarządzanie kosztami leasingu i opłatami za użytkowanie
- zarządzanie kontami użytkowników, danymi uwierzytelniającymi, autoryzacją i kontrolą dostępu
- śledzenie wewnętrznego i zewnętrznego dostępu do dzierżawionych usług
- planowanie i ocena udostępniania zasobów IT
- planowanie wydajności

Chociaż interfejs użytkownika zapewniany przez system zdalnej administracji będzie zwykle zastrzeżony dla dostawcy chmury, wśród konsumentów chmury preferowane jest korzystanie z systemów zdalnej administracji, które oferują standardowe interfejsy API. Dzięki temu konsument chmury może zainwestować w stworzenie własnego front-endu, mając wcześniej świadomość, że może ponownie wykorzystać tę konsolę, jeśli zdecyduje się przejść do innego dostawcy chmury, który obsługuje ten sam ustandaryzowany interfejs API. Ponadto konsument chmury będzie mógł w większym stopniu korzystać ze standardowych interfejsów API, jeśli będzie zainteresowany dzierżawą i centralnym administrowaniem zasobami IT od wielu dostawców usług w chmurze i/lub zasobami IT znajdującymi się w chmurze i środowiskach lokalnych.



Rys. 3.2.3.4

Rysunek 3.2.3.4 – Standardyzowane interfejsy API publikowane przez systemy zdalnej administracji z różnych chmur umożliwiają konsumentowi chmury opracowanie niestandardowego portalu, który centralizuje pojedynczy portal zarządzania zasobami IT zarówno dla zasobów IT w chmurze, jak i lokalnych.

3.2.4 System zarządzania zasobami:

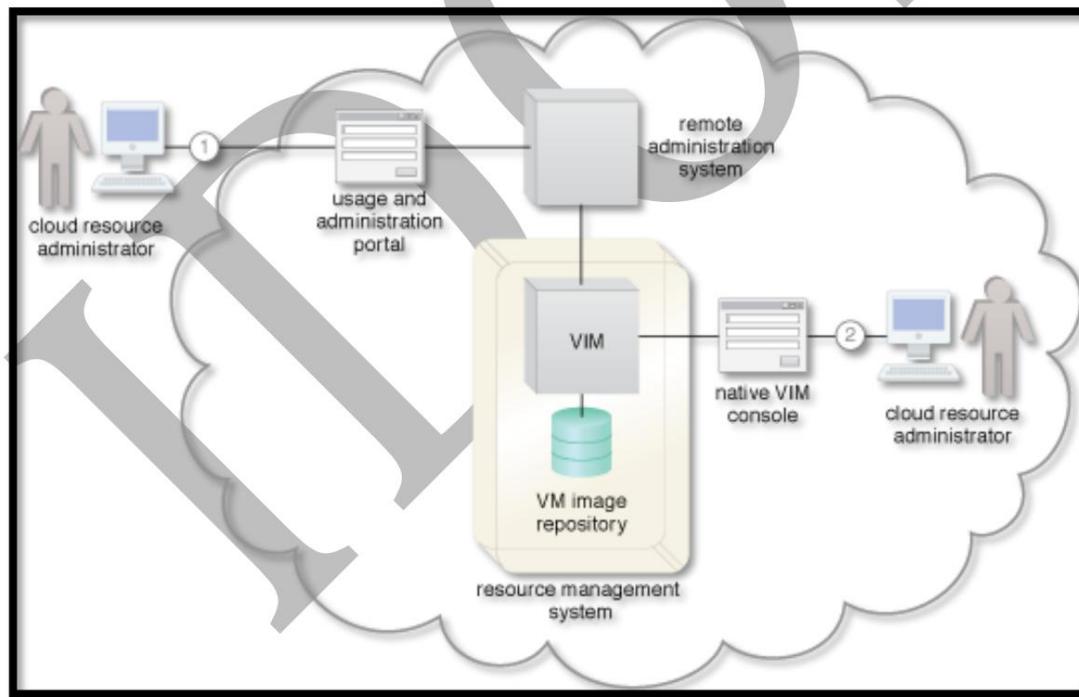
Mechanizm systemu zarządzania zasobami pomaga koordynować zasoby IT w odpowiedzi na działania zarządcze realizowane zarówno przez konsumentów chmury, jak i dostawców chmury (rysunek 1). Sercem tego systemu jest menedżer infrastruktury wirtualnej (VIM), który koordynuje sprzęt serwera, tak aby instancje serwerów wirtualnych mogły być tworzone z najwygodniejszego podstawowego serwera fizycznego. VIM to produkt komercyjny, którego można używać do zarządzania szeregiem wirtualnych zasobów IT na wielu serwerach fizycznych. Na przykład VIM może tworzyć i zarządzać wielomainstancjami hypervisorów na różnych serwerach fizycznych lub przydzielać serwer wirtualny na jednym serwerze fizycznym innemu (lub puli zasobów).

Zadania, które są zazwyczaj zautomatyzowane i realizowane za pośrednictwem systemu zarządzania zasobami, obejmują:

- zarządzanie szablonami wirtualnych zasobów IT, które służą do tworzenia gotowych instancji, takich jak obrazy serwerów wirtualnych
- alokowanie i zwalnianie wirtualnych zasobów IT do dostępnej infrastruktury fizycznej w odpowiedzi na uruchamianie, wstrzymywanie, wznowianie i kończenie instancji wirtualnych zasobów IT
- koordynacja zasobów IT w związku z zaangażowaniem innych mechanizmów, np. replikacja zasobów, moduł równoważenia obciążenia i system przełączania awaryjnego
- egzekwowanie zasad użytkowania i bezpieczeństwa w całym cyklu życia instancji usług w chmurze
- monitorowanie warunków pracy zasobów IT

Dostęp do funkcji systemu zarządzania zasobami mają administratorzy zasobów chmury zatrudnieni przez dostawcę chmury lub konsumenta chmury. Osoby pracujące w imieniu dostawcy usług w chmurze często będą miały bezpośredni dostęp do natywnej konsoli systemu zarządzania zasobami.

Systemy zarządzania zasobami zazwyczaj udostępniają interfejsy API, które umożliwiają dostawcom usług w chmurze tworzenie portali systemów zdalnej administracji, które można dostosować tak, aby selektywnie oferowały kontrolę zarządzania zasobami zewnętrznym administratorom zasobów chmury działającym w imieniu organizacji konsumenckich w chmurze za pośrednictwem portali użytkowania i administracji.

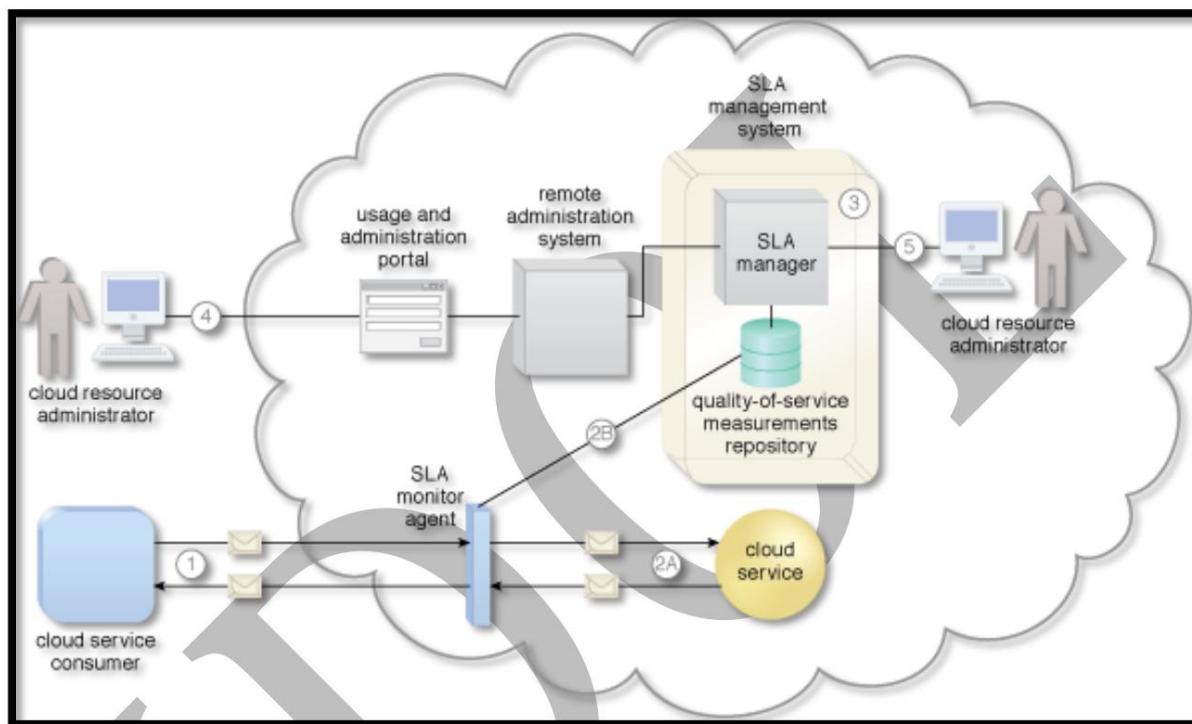


Ryc. 3.2.4

Rysunek 3.2.4 – Administrator zasobów chmury konsumenta chmury uzyskuje dostęp do portalu użytkowania i administracji z zewnątrz w celu administrowania dzierżawionymi zasobami IT (1). Administrator zasobów chmury dostawcy chmury korzysta z natywnego interfejsu użytkownika udostępnianego przez VIM do wykonywania zadań związanych z wewnętrznym zarządzaniem zasobami (2).

3.2.5 System zarządzania SLA:

Mechanizm monitorowania SLA służy do szczegółowej obserwacji wydajności usług w chmurze w celu zapewnienia, że spełniają one umowne wymagania QoS opublikowane w umowach SLA (rysunek 1). Dane zebrane przez monitor SLA są przetwarzane przez system zarządzania SLA w celu zagregowania ich w metryki raportowania SLA. Systemy te mogą proaktywnie naprawiać lub przełączać awaryjnie usługi w chmurze, gdy wystąpią wyjątkowe warunki, na przykład gdy monitor SLA zgłosi usługę w chmurze jako „nie działa”.



Ryc. 3.2.5

Rysunek 3.2.5 – Monitor SLA odpytuje usługę w chmurze, wysyłając komunikaty żądań odpytywania (MREQ1 do MREQN). Monitor odbiera komunikaty odpowiedzi na odpytywanie (M do M), które informują, że usługa działa w każdym cyklu odpytywania (1a). Monitor SLA przechowuje czas „up” – okres wszystkich cykli odpytywania od 1 do N – w bazie danych dziennika (1b). Monitor SLA odpytuje usługę w chmurze, która wysyła komunikaty żądań odpytywania (M do M). Wiadomości z odpowiedzią na odpytywanie nie są odbierane (2a). Komunikaty odpowiedzi nadal przekraczają limit czasu, więc monitor SLA przechowuje czas „przestoju” – okres czasu wszystkich cykli odpytywania N+1 do N+M – w bazie danych logów (2b). Monitor SLA wysyła komunikat żądania odpytywania (M) i odbiera komunikat odpowiedzi odpytywania (M) (3a). Monitor SLA przechowuje czas „up” w bazie danych logów (3b).

3.2.3.2 System zarządzania rozliczeniami:

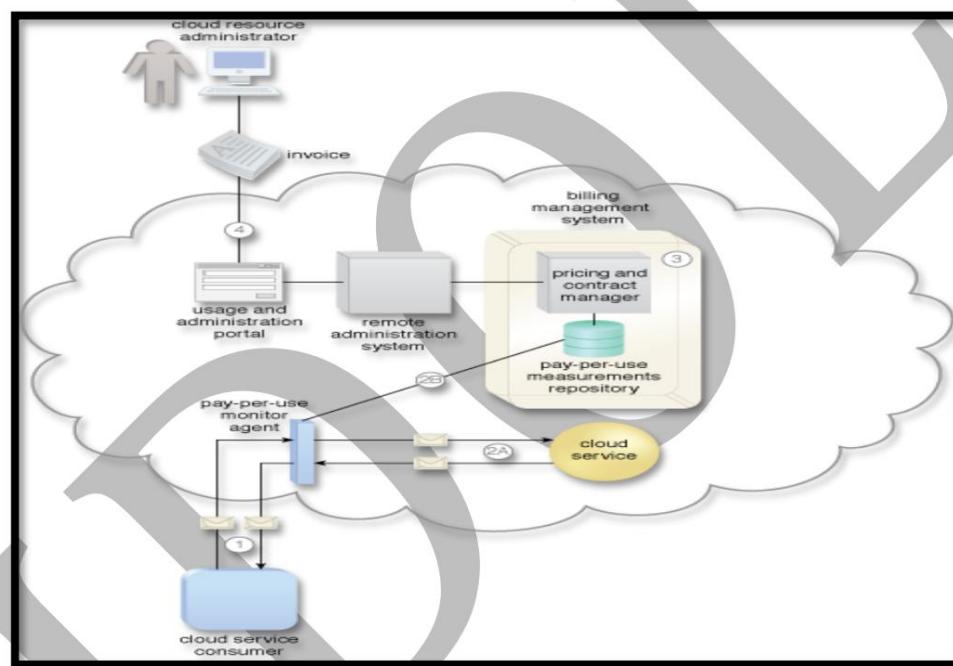
Mechanizm systemu zarządzania billingami jest dedykowany do gromadzenia i przetwarzania danych o użytkowaniu w zakresie księgowości dostawców usług chmurowych i rozliczeń konsumenckich usług chmurowych. W szczególności system zarządzania rozliczeniami opiera się na monitorach płatności za użycie w celu gromadzenia danych o użytkowaniu w czasie wykonywania, które są przechowywane w

repozytorium, z którego następnie korzystają komponenty systemu na potrzeby raportowania i fakturowania (Rysunek 1).

System zarządzania rozliczeniami umożliwia definiowanie różnych polityk cenowych, a także niestandardowych modeli cenowych w oparciu o konsumenta w chmurze i/lub zasoby IT. Modele cenowe mogą się różnić od tradycyjnych modeli płatności za użycie po modele ryczałtowe lub modele płatności za alokację lub ich kombinacje.

Ustalenia dotyczące rozliczeń mogą opierać się na płatnościach przed i po użyciu. Ten ostatni typ może obejmować z góry określone limity lub może zostać skonfigurowany (za obopólną zgodą konsumenta chmury) tak, aby umożliwić nieograniczone korzystanie (a co za tym idzie, brak limitu w późniejszych rozliczeniach).

Ustalane limity mają zazwyczaj formę kwot użytkowania. W przypadku przekroczenia limitów system zarządzania rozliczeniami może zablokować dalsze żądania użycia przesyłane przez konsumentów chmury.

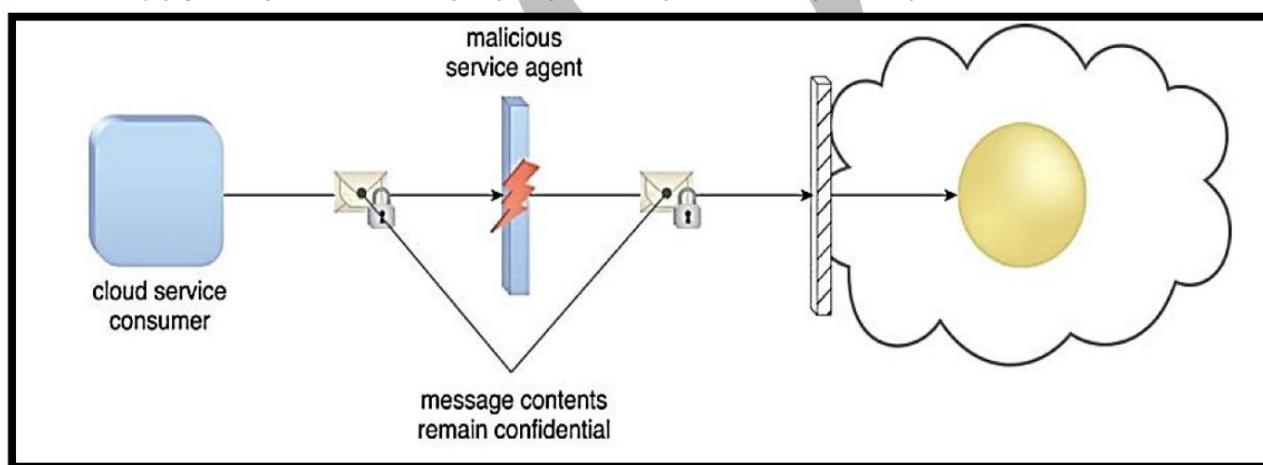


Rys. 3.2.3.2

Rysunek 3.2.3.2 – Konsument usługi w chmurze wymienia wiadomości z usługą w chmurze (1). Monitor płatności za użycie śledzi wykorzystanie i zbiera dane istotne dla rozliczeń (2A), które są przekazywane do repozytorium będącego częścią systemu zarządzania rozliczeniami (2B). System okresowo nalicza skonsolidowane opłaty za korzystanie z usług chmurowych i generuje fakturę dla konsumenta chmury (3). Fakturę można udostępnić konsumentowi chmury za pośrednictwem portalu użytkowania i administracji (4).

3.2.7 Szyfrowanie:

- Dane są domyślnie kodowane w czytelnym formacie znanym jako zwykły tekst. Tekst jawnego przesyłanego przez sieć jest podatny na nieautoryzowany i potencjalnie złośliwy dostęp. • Mechanizm szyfrujący to cyfrowy system kodowania mający na celu zachowanie poufności i integralności danych. Służy do kodowania danych w postaci zwykłego tekstu do chronionego i nieczytelnego formatu.
 - Technologia szyfrowania zwykle opiera się na ustandaryzowanym algorytmie zwany szyfrem, który przekształca oryginalne dane w postaci zwykłego tekstu w dane zaszyfrowane, zwane tekstem zaszyfrowanym.
 - Gdy szyfrowanie jest stosowane do danych w postaci zwykłego tekstu, dane są łączone w parę znaków zwanych kluczem szyfrowania, czyli tajną wiadomością ustaloną i udostępnianą upoważnionym stronom. Klucz szyfrowania służy do odszyfrowania tekstu zaszyfrowanego z powrotem do oryginalnego formatu zwykłego tekstu.
 - Szyfrowanie danych w chmurze to proces przekształcania lub kodowania danych przed ich przeniesieniem do magazynu w chmurze. Zazwyczaj dostawcy usług w chmurze oferują usługi szyfrowania, począwszy od szyfrowanego połączenia po ograniczone szyfrowanie wrażliwych danych, a także udostępniają klucze szyfrujące w celu odszyfrowania danych w razie potrzeby.
 - Tego typu usługi szyfrowania zapobiegają nieautoryzowanemu, swobodnemu dostępowi do systemu lub danych plików bez klucza deszyfrującego, co czyni je skuteczną metodą zabezpieczania danych. Bezpieczeństwo informacji w chmurze powinno być Twoim najwyższym priorytetem. Już samo podjęcie kilku środków zapobiegawczych związanych z szyfrowaniem danych może zwiększyć bezpieczeństwo najbardziej wrażliwych informacji.
- Postępuj zgodnie z tymi wskazówkami dotyczącymi szyfrowania, aby zablokować swoje informacje w chmurze.



Ryc. 3.2.7

Rysunek 3.2.7 Złośliwy agent usługi nie może pobrać danych z zaszyfrowanej wiadomości. Próba odzyskania może ponadto zostać ujawniona konsumentowi usługi w chmurze.

Istnieją dwie popularne formy szyfrowania, znane jako szyfrowanie symetryczne i szyfrowanie asymetryczne.

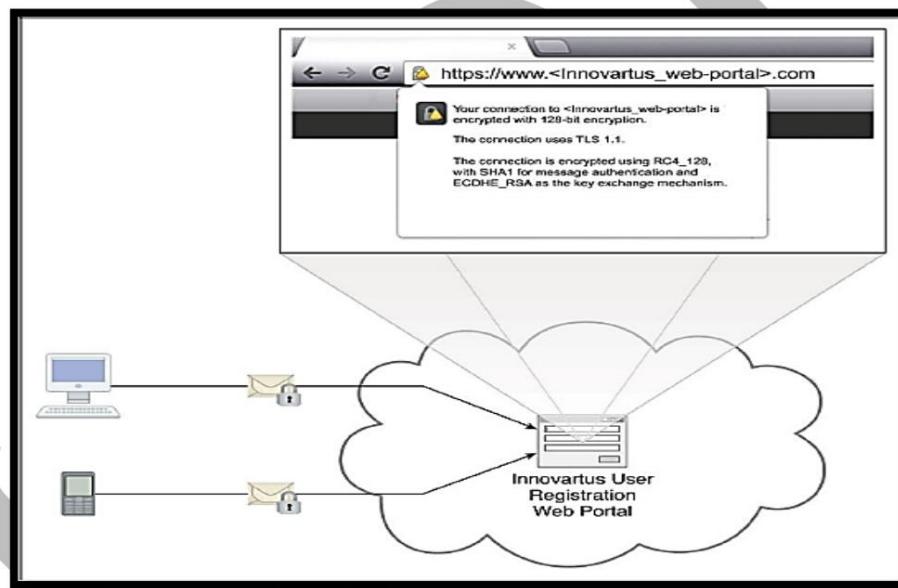
3.2.7.1 Szyfrowanie symetryczne:

- Szyfrowanie symetryczne wykorzystuje ten sam klucz zarówno do szyfrowania, jak i deszyfrowania, przy czym oba są wykonywane przez upoważnione strony, które korzystają z jednego wspólnego klucza. Wiadomości zaszyfrowane przy użyciu określonego klucza, zwane także kryptografią tajnego klucza, można odszyfrować tylko za pomocą szyfrowania

ten sam klucz Należy pamiętać, że szyfrowanie symetryczne nie charakteryzuje się niezaprzeczalnością.

3.2.7.2 Szyfrowanie asymetryczne

- Szyfrowanie asymetryczne opiera się na użyciu dwóch różnych kluczy, mianowicie klucza prywatnego i klucza publicznego. W przypadku szyfrowania asymetrycznego (nazywanego również kryptografią klucza publicznego) klucz prywatny jest znany tylko jego właścielowi, podczas gdy klucz publiczny jest powszechnie dostępny. Dokument zaszyfrowany kluczem prywatnym można poprawnie odszyfrować tylko za pomocą odpowiedniego klucza publicznego.
- I odwrotnie, dokument zaszyfrowany kluczem publicznym można odszyfrować tylko przy użyciu jego odpowiednika w kluczu prywatnym. Szyfrowanie asymetryczne jest prawie zawsze wolniejsze obliczeniowo niż szyfrowanie symetryczne. Szyfrowanie klucza prywatnego zapewnia zatem ochronę integralności oprócz autentyczności i niezaprzeczalności.



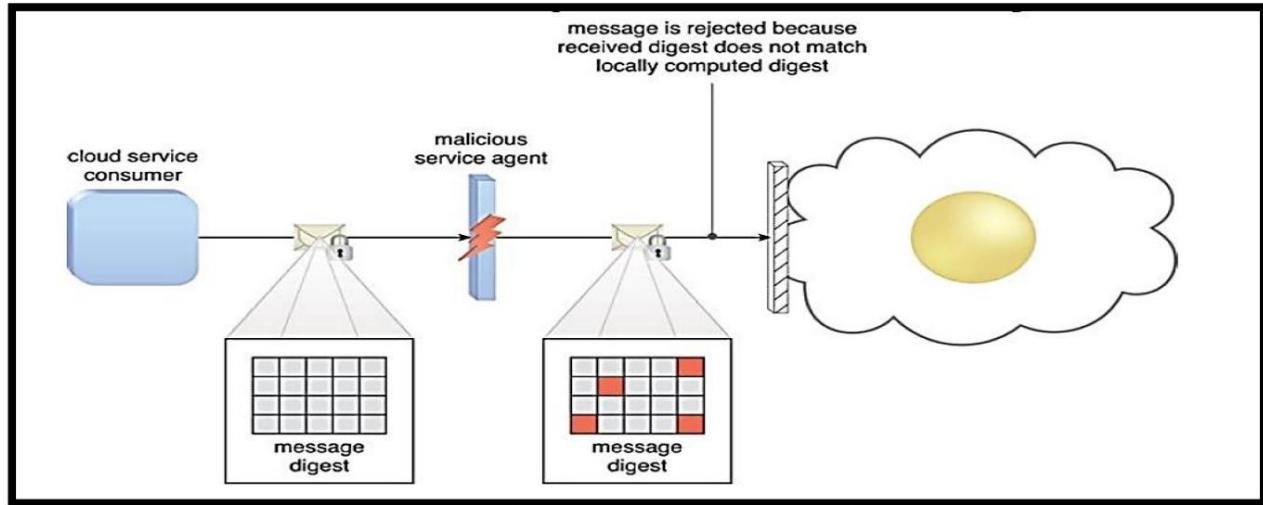
Rys. 3.2.7.2

Rysunek 3.2.7.2 Do kanału komunikacji pomiędzy użytkownikami zewnętrznymi a Portalem Rejestracji Użytkowników Innovartus dodano mechanizm szyfrujący. Zapewnia to poufność wiadomości dzięki wykorzystaniu protokołu HTTPS.

3.2.8 Haszowanie

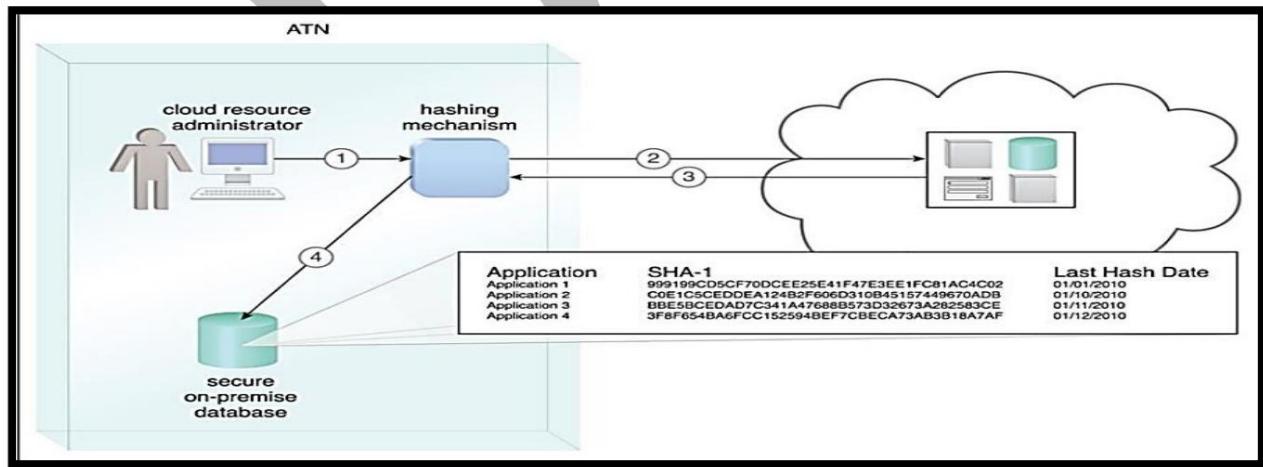
- Mieszanie mechanizmu mieszającego jest stosowane, gdy wymagana jest jednokierunkowa, nieodwracalna forma ochrony danych. Po zastosowaniu haszowania do wiadomości jest ona blokowana i nie jest udostępniany żaden klucz umożliwiający odblokowanie wiadomości. • Powszechnym zastosowaniem tego mechanizmu jest przechowywanie haseł. Technologię mieszania można wykorzystać do uzyskania kodu skrótu lub skrótu wiadomości z wiadomości, która często ma stałą długość i jest mniejsza niż oryginalna wiadomość.

- Nadawca wiadomości może następnie wykorzystać mechanizm mieszający w celu dołączenia podsumowania wiadomości do wiadomości. Odbiorca stosuje tę samą funkcję skrótu do wiadomości, aby sprawdzić, czy utworzony skrót wiadomości jest identyczny z tym, który towarzyszył wiadomości. • Jakakolwiek zmiana oryginalnych danych skutkuje zupełnie innym podsumowaniem wiadomości i wyraźnie wskazuje, że doszło do manipulacji.



Rys. 3.2.8.1

Rysunek 3.2.8. Funkcja hashująca stosowana jest w celu ochrony integralności wiadomości przechwytywanej i modyfikowanej przez złośliwego agenta usługi przed jej przesaniem dalej. Zaporę sieciową można skonfigurować tak, aby ustalić, czy wiadomość została zmieniona, umożliwiając w ten sposób odrzucenie wiadomości, zanim będzie mogła przejść do usługi w chmurze.

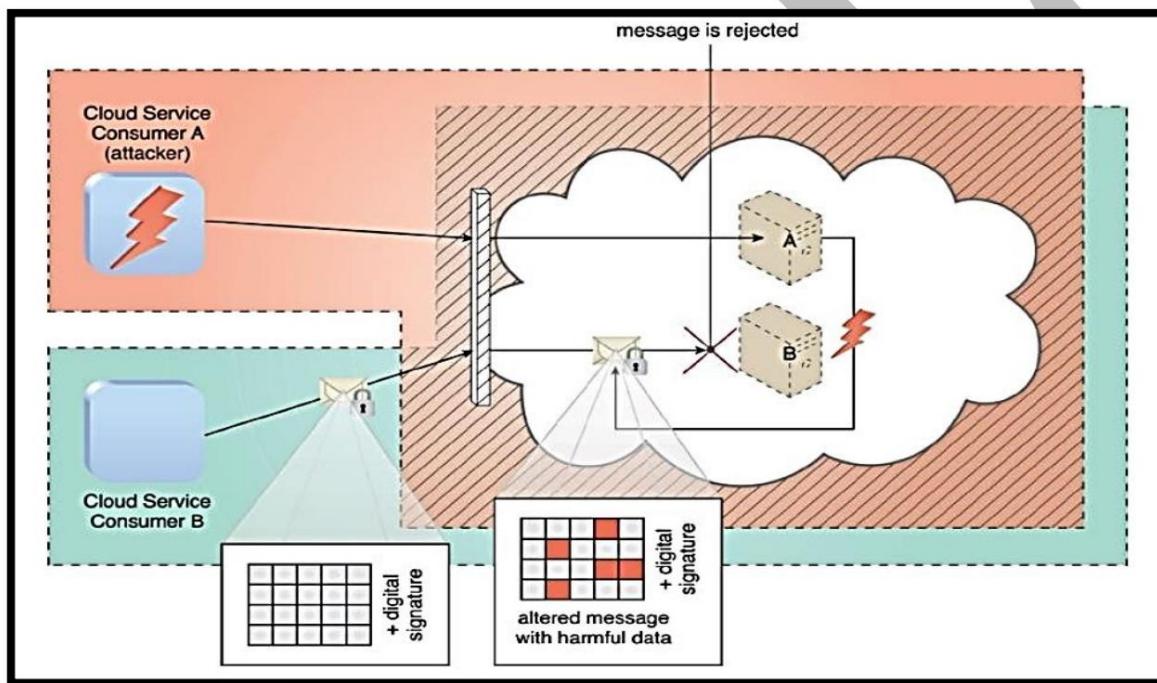


Rys. 3.2.8.2

Rysunek 3.2.8.2 Procedura miesząca jest wywoływana podczas uzyskiwania dostępu do środowiska PaaS (1). Sprawdzane są aplikacje przeniesione do tego środowiska. (2). i obliczane są skróty ich wiadomości. (3). Streszczenia wiadomości są przechowywane w bezpiecznej lokalnej bazie danych. (4) i następuje powiadomienie, jeśli którykolwiek z ich wartości nie jest identyczna z wartością w magazynie.

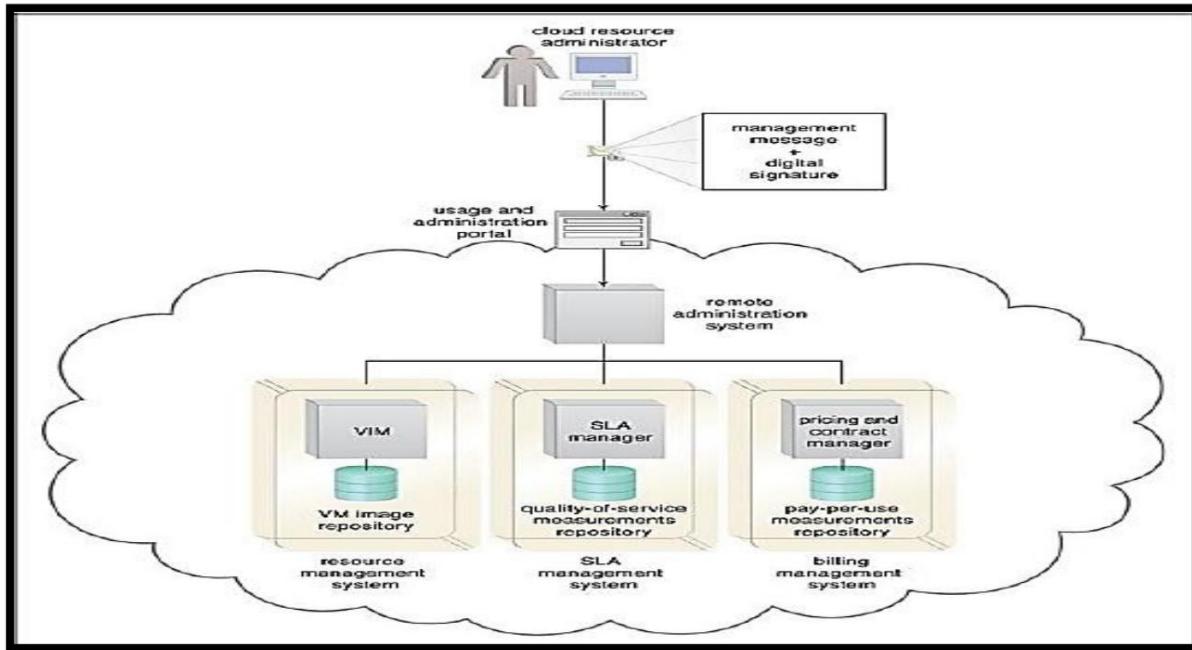
3.2.9 Podpis cyfrowy:

- Mechanizm podpisu cyfrowego jest środkiem zapewniającym autentyczność i integralność danych poprzez uwierzytelnianie i niezaprzecjalność. • Przed transmisją do wiadomości przypisywany jest podpis cyfrowy, który następnie traci ważność, jeśli wiadomość ulegnie późniejszym, nieautoryzowanym modyfikacjom. • Podpis cyfrowy stanowi dowód, że otrzymana wiadomość jest taka sama jak ta stworzony przez jego prawowitego nadawcę.
- Zarówno haszowanie, jak i szyfrowanie asymetryczne biorą udział w tworzeniu podpisu cyfrowego, który zasadniczo istnieje w postaci skrótu wiadomości zaszyfrowanej kluczem prywatnym i dołączonego do oryginalnej wiadomości. Odbiorca weryfikuje ważność podpisu i używa odpowiedniego klucza publicznego do odszyfrowania podpisu cyfrowego, w wyniku czego powstaje skrót wiadomości.



Rys. 3.2.9.1

Rysunek 3.2.9.1 Konsument B usługi w chmurze wysyła wiadomość podpisana cyfrowo, ale zmienioną przez zaufanego atakującego Konsument usługi w chmurze A. Serwer wirtualny B jest skonfigurowany tak, aby weryfikować podpisy cyfrowe przed przetwarzaniem wiadomości przychodzących, nawet jeśli znajdują się one w jego granicach zaufania. Wiadomość zostaje uznana za nielegalną ze względu na nieprawidłowy podpis cyfrowy i dlatego zostaje odrzucona przez Serwer Wirtualny B.

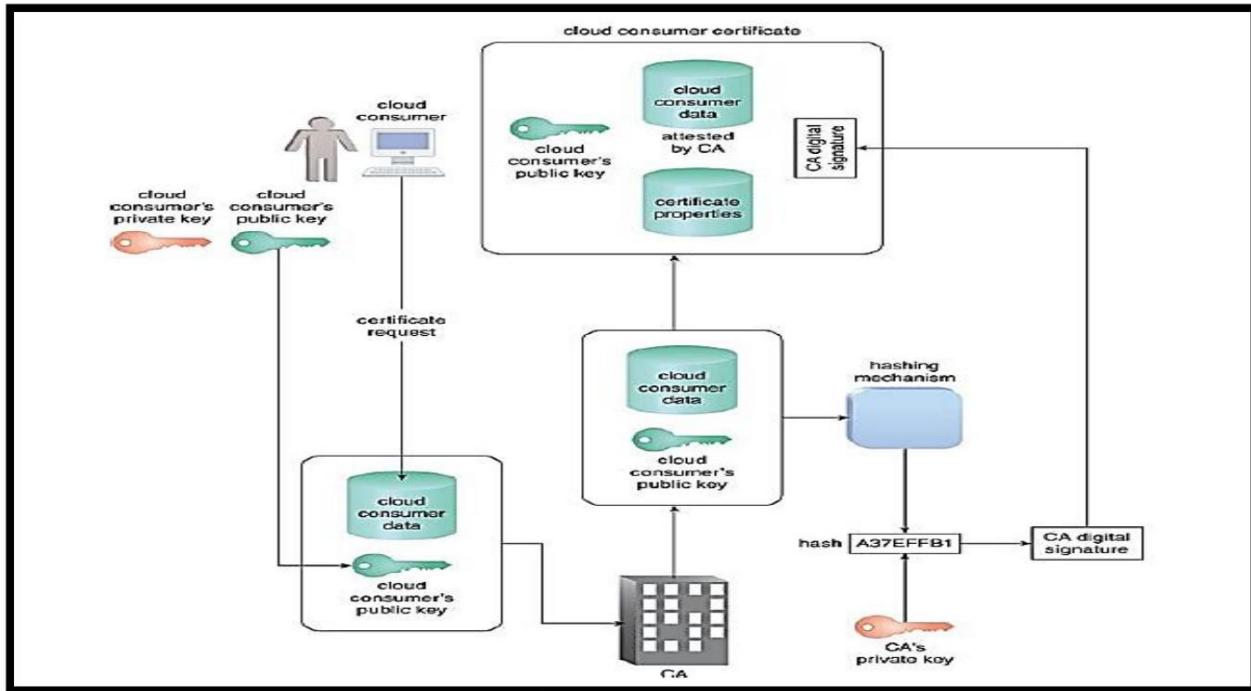


Rys. 3.2.9.2

Rysunek 3.2.9.2 Ilekroć konsument chmury wykonuje czynność zarządczą związaną z zasobami IT udostępnianymi przez DTGOV, program konsumencki usługi w chmurze musi zawierać podpis cyfrowy w żądaniu wiadomości, aby potwierdzić legalność swojego użytkownika.

3.2.10 Infrastruktura klucza publicznego (PKI):

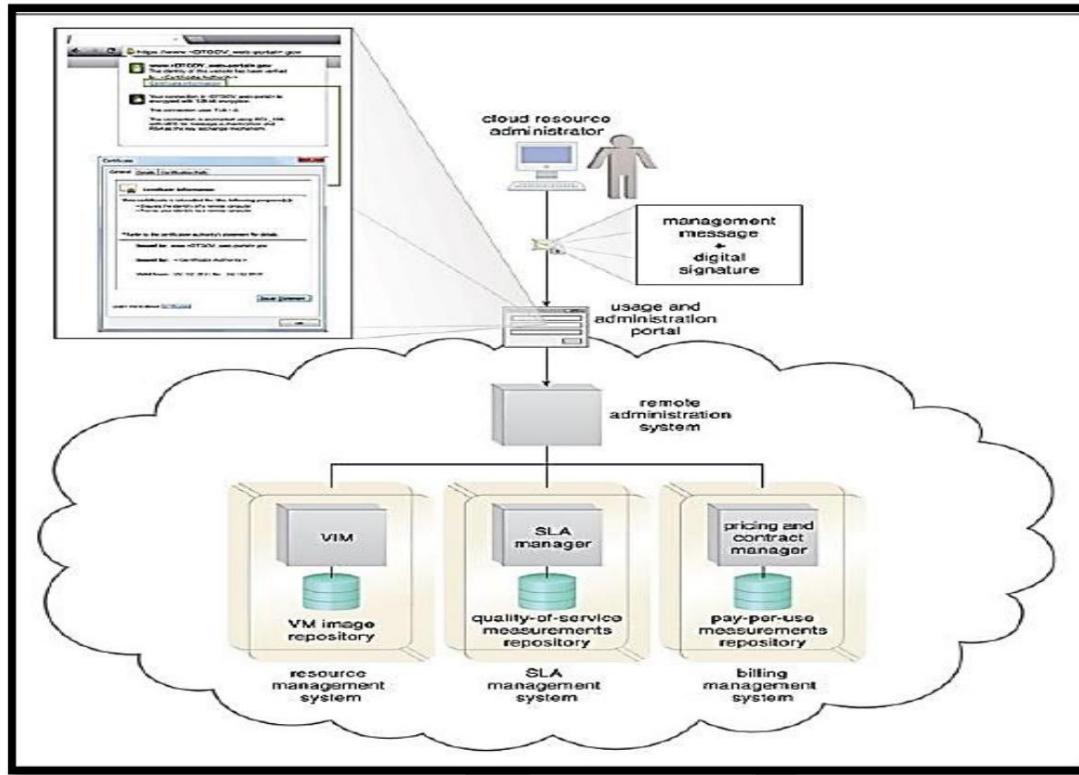
- Infrastruktura klucza publicznego (PKI) Powszechnie podejście do zarządzania wydawaniem kluczy asymetrycznych opiera się na mechanizmie infrastruktury klucza publicznego (PKI), który istnieje jako system protokołów, formatów danych, reguł i praktyk umożliwiających tworzenie systemów na dużą skalę do bezpiecznego korzystania z kryptografii
- System ten służy do kojarzenia kluczy publicznych z odpowiadającymi im właścicielami kluczy (tzw. identyfikacja klucza publicznego), umożliwiając jednocześnie weryfikację ważności klucza.
- PKI opierają się na wykorzystaniu certyfikatów cyfrowych, które są podpisany cyfrowo strukturami danych, które wiążą klucze publiczne z tożsamością właściciela certyfikatu, a także z powiązanymi informacjami, takimi jak okresy ważności. Certyfikaty cyfrowe są zwykle podpisane cyfrowo przez zewnętrzny urząd certyfikacji (CA), jak pokazano na rysunku 7.



Rys. 3.2.10.1

Rysunek 3.2.10.1 Typowe kroki wykonywane podczas generowania certyfikatów przez urząd certyfikacji.

- Infrastruktura klucza publicznego (PKI) Większe organizacje, takie jak Microsoft, mogą działać jako własny urząd certyfikacji i wystawiać certyfikaty swoim klientom i społeczeństwu, ponieważ nawet indywidualni użytkownicy mogą generować certyfikaty, o ile posiadają odpowiednie narzędzia programowe.
- PKI to niezawodna metoda wdrażania szyfrowania asymetrycznego, zarządzania informacjami o tożsamości konsumenta i dostawcy chmury oraz pomagająca w obronie przed złośliwym pośrednikiem i zagrożeniami związanymi z niewystarczającą autoryzacją.
- Mechanizm PKI służy przede wszystkim przeciwdziałaniu zagrożeniu niewystarczającej autoryzacji.



Rys. 3.2.10.2

Rysunek 3.2.10.2 Zewnętrzny administrator zasobów chmury używa certyfikatu cyfrowego, aby uzyskać dostęp do internetowego środowiska zarządzania. Certyfikat cyfrowy DTGOV jest używany w połączeniu HTTPS, a następnie podpisywany przez zaufany urząd certyfikacji.

3.2.11 Zarządzanie tożsamością i dostępem (IAM):

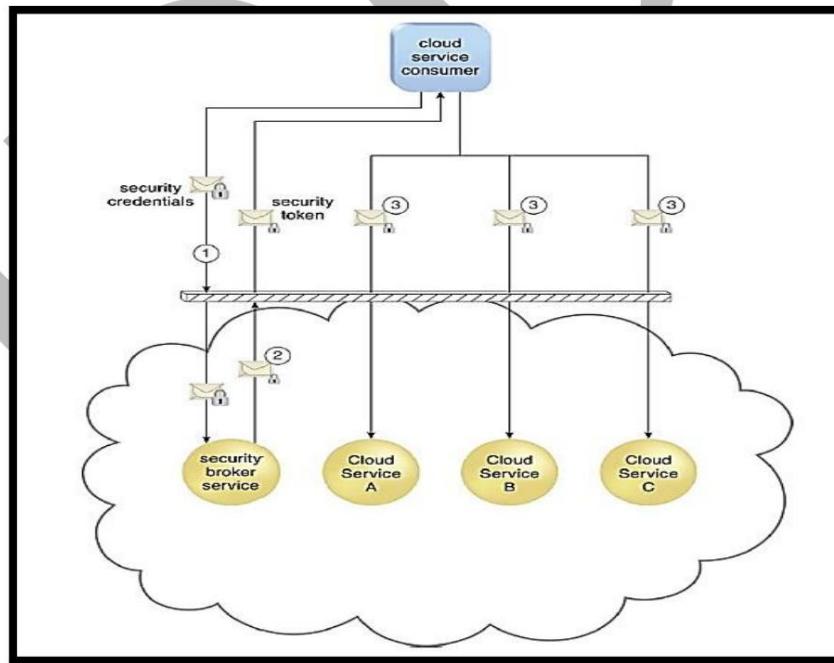
Zarządzanie tożsamością i dostępem (IAM) Mechanizm zarządzania tożsamością i dostępem (IAM) obejmuje komponenty i zasady niezbędne do kontrolowania i śledzenia tożsamości użytkowników oraz uprawnień dostępu do zasobów, środowisk i systemów informatycznych. W szczególności mechanizmy IAM istnieją jako systemy składające się z czterech głównych komponentów:

1. Uwierzytelnianie – kombinacje nazwy użytkownika i hasła pozostają najpopularniejszymi formami uwierzytelniania użytkowników zarządzanymi przez system IAM, który może również obsługiwać podpisy cyfrowe, certyfikaty cyfrowe, sprzęt biometryczny (czytniki linii papilarnych), specjalistyczne oprogramowanie (takie jak programy do analizy głosu), i blokowanie kont użytkowników do zarejestrowanych adresów IP lub MAC.
2. Autoryzacja – Komponent autoryzacji definiuje odpowiednią szczegółowość kontroli dostępu i nadzoruje relacje pomiędzy tożsamościami, prawami kontroli dostępu i dostępnością zasobów IT.
3. Zarządzanie użytkownikami – w odniesieniu do możliwości administracyjnych systemu, program do zarządzania użytkownikami jest odpowiedzialny za tworzenie nowych tożsamości użytkowników i grup dostępu, resetowanie haseł, definiowanie zasad haseł i zarządzanie uprawnieniami.

4. Zarządzanie uprawnieniami – System zarządzania uprawnieniami ustala tożsamość i zasady kontroli dostępu dla zdefiniowanych kont użytkowników, co minimalizuje ryzyko niewystarczającej autoryzacji. Mechanizm IAM służy przede wszystkim do przeciwdziałania zagrożeniom związanym z niewystarczającą autoryzacją, odmową usługi i nakładającymi się granicami zaufania.

3.2.12 Pojedyncze logowanie (SSO):

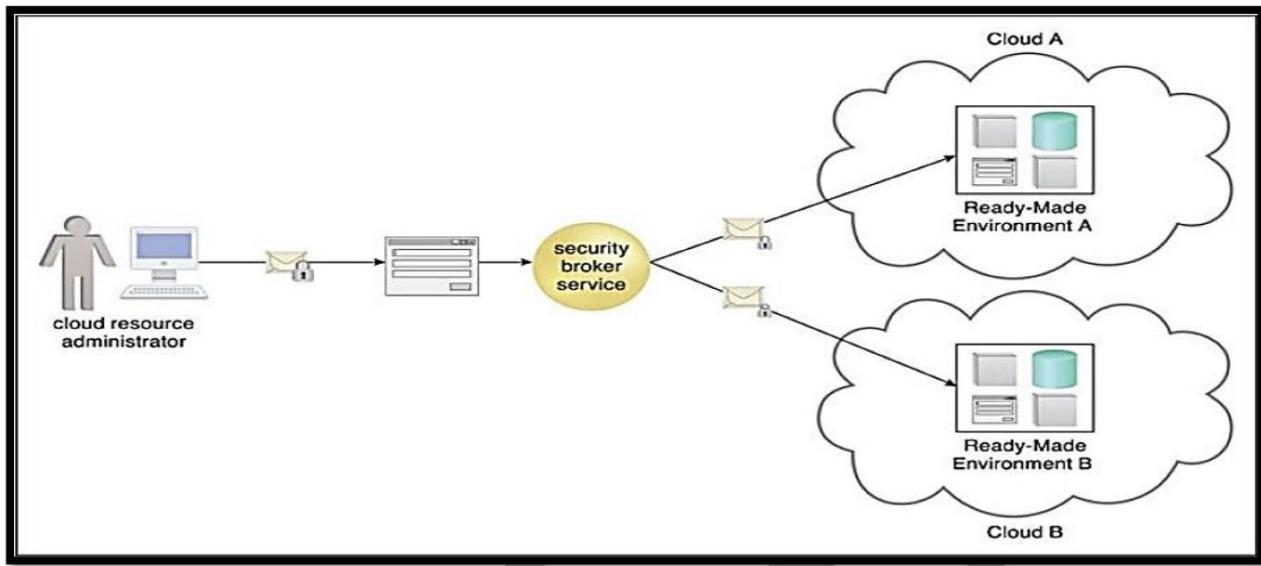
- Pojedyncze logowanie (SSO) Propagowanie informacji o uwierzytelnianiu i autoryzacji dla konsumenta usług w chmurze w wielu usługach w chmurze może być wyzwaniem, zwłaszcza jeśli trzeba wywołać wiele usług w chmurze lub zasobów IT w chmurze w ramach tego samego ogólnego czasu działania działalności.
- Mechanizm pojedynczego logowania (SSO) umożliwia uwierzytelnienie jednego konsumenta usług w chmurze przez brokera bezpieczeństwa, który ustanawia kontekst bezpieczeństwa, który jest trwałym, gdy konsument usług w chmurze uzyskuje dostęp do innych usług w chmurze lub IT opartego na chmurze zasoby.
- W przeciwnym razie konsument usług w chmurze musiałby się ponownie uwierzytelniać przy każdym kolejnym żądaniu. Mechanizm SSO zasadniczo umożliwia wzajemnie niezależnym usługom w chmurze i zasobom IT generowanie i rozpowszechnianie danych uwierzytelniających i autoryzacyjnych w czasie wykonywania.



Rys. 3.2.12.1

Rysunek 3.2.12.1 Konsument usługi w chmurze udostępnia brokerowi bezpieczeństwa dane logowania (1). Broker bezpieczeństwa odpowiada tokenem uwierzytelniającym (komunikat z małym symbolem kłódki) po pomyślnym uwierzytelnieniu, który zawiera informacje o tożsamości konsumenta usługi w chmurze (2), tj.

używany do automatycznego uwierzytelniania konsumenta usług w chmurze w usługach w chmurze A, B i C (3).

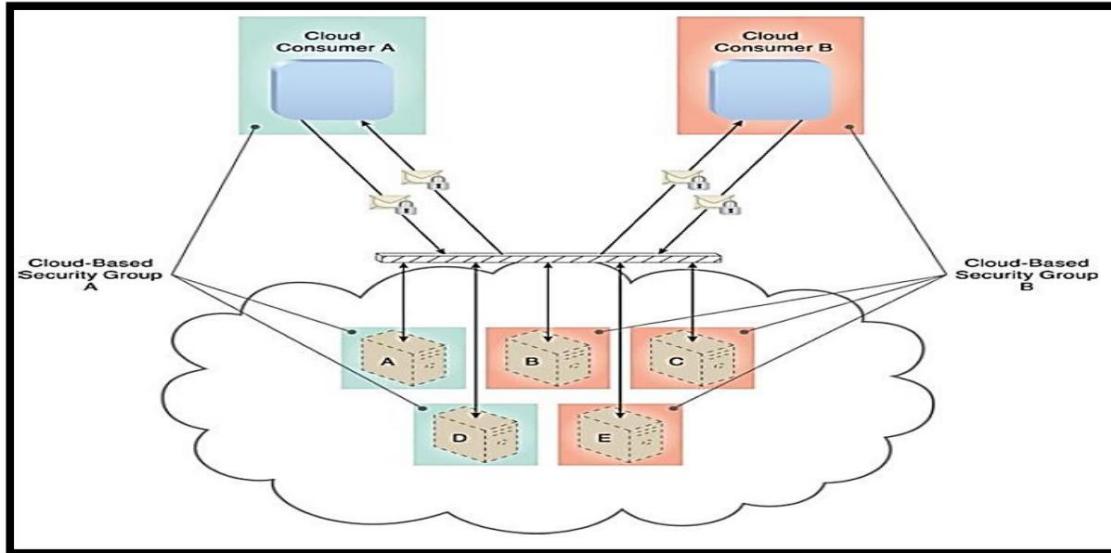


Rys. 3.2.12.2

Rysunek 3.2.12.2. Poświadczenia otrzymane przez brokera bezpieczeństwa są propagowane do gotowych środowisk w dwóch różnych chmurach. Broker bezpieczeństwa jest odpowiedzialny za wybór odpowiednich procedur bezpieczeństwa, za pomocą której można skontaktować się z każdą chmurą.

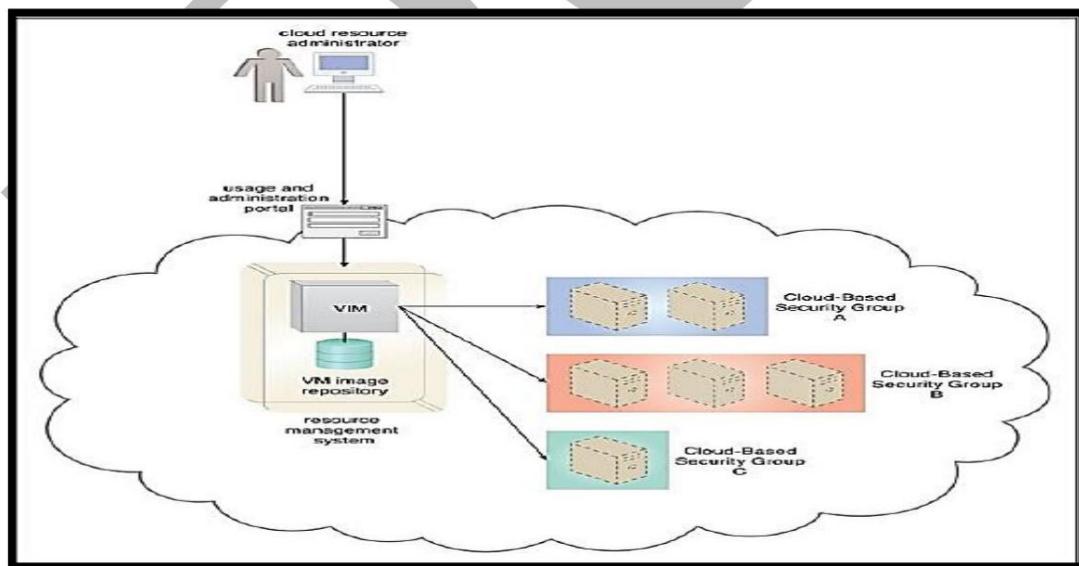
3.2.13 Grupy zabezpieczeń oparte na chmurze:

- Grupy zabezpieczeń oparte na chmurze Segmentacja zasobów chmury to proces, w ramach którego tworzone są oddzielne fizyczne i wirtualne środowiska IT dla różnych użytkowników i grup. Na przykład sieć WAN organizacji można podzielić na partie zgodnie z indywidualnymi wymaganiami bezpieczeństwa sieci.
- Jedną sieć można utworzyć z odporną zaporą ogniową zapewniającą zewnętrzny dostęp do Internetu, a drugą można wdrożyć bez zapory, ponieważ jej użytkownicy są wewnętrzni i nie mogą uzyskać dostępu do Internetu. Segmentacja zasobów umożliwia wirtualizację poprzez przydzielanie maszynom wirtualnym różnych fizycznych zasobów IT.
- Grupy zabezpieczeń oparte na chmurze Proces segmentacji zasobów oparty na chmurze tworzy mechanizmy grup zabezpieczeń oparte na chmurze, które są określone poprzez zasady bezpieczeństwa. Sieci są podzielone na logiczne grupy zabezpieczeń oparte na chmurze, które tworzą logiczne obwody sieci. Wiele serwerów wirtualnych działających na tym samym serwerze fizycznym może stać się członkami różnych logicznych grup zabezpieczeń opartych na chmurze (Rysunek 11).
- Serwery wirtualne można dalej podzielić na grupy publiczno-prywatne, grupy programistyczno-produkcyjne lub dowolne inne oznaczenie skonfigurowane przez administratora zasobów chmury.



Rys. 3.2.13.1

Rysunek 3.2.13.1 Oparta na chmurze grupa zabezpieczeń A obejmuje serwery wirtualne A i D i jest przypisana do Konsumenta chmury A. Oparta na chmurze grupa zabezpieczeń B składa się z serwerów wirtualnych B, C i E i jest przypisana do Konsumenta chmury B. Jeśli Poświadczenia konsumenta usługi w chmurze A zostaną naruszone, osoba atakująca będzie mogła uzyskać dostęp do serwerów wirtualnych w grupie zabezpieczeń opartej na chmurze A i je uszkodzić, chroniąc w ten sposób serwery wirtualne B, C i E.

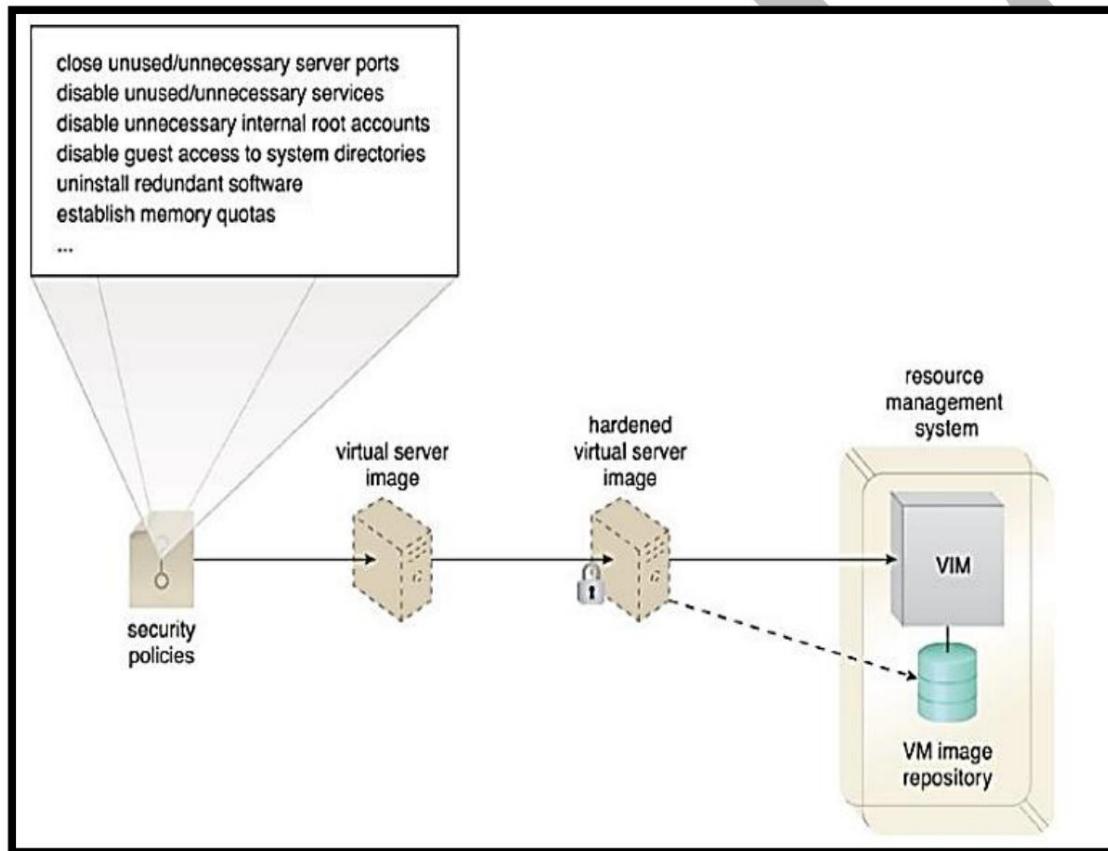


Rys. 3.2.13.2

Rysunek 3.2.13.2 Gdy zewnętrzny administrator zasobów chmury uzyskuje dostęp do portalu internetowego w celu przydzielenia serwera wirtualnego, żądane poświadczania bezpieczeństwa są oceniane i mapowane do wewnętrznej polityki bezpieczeństwa, która przypisuje odpowiednią grupę zabezpieczeń opartą na chmurze do nowego serwera wirtualnego.

3.2.14 Wzmocnione obrazy serwerów wirtualnych

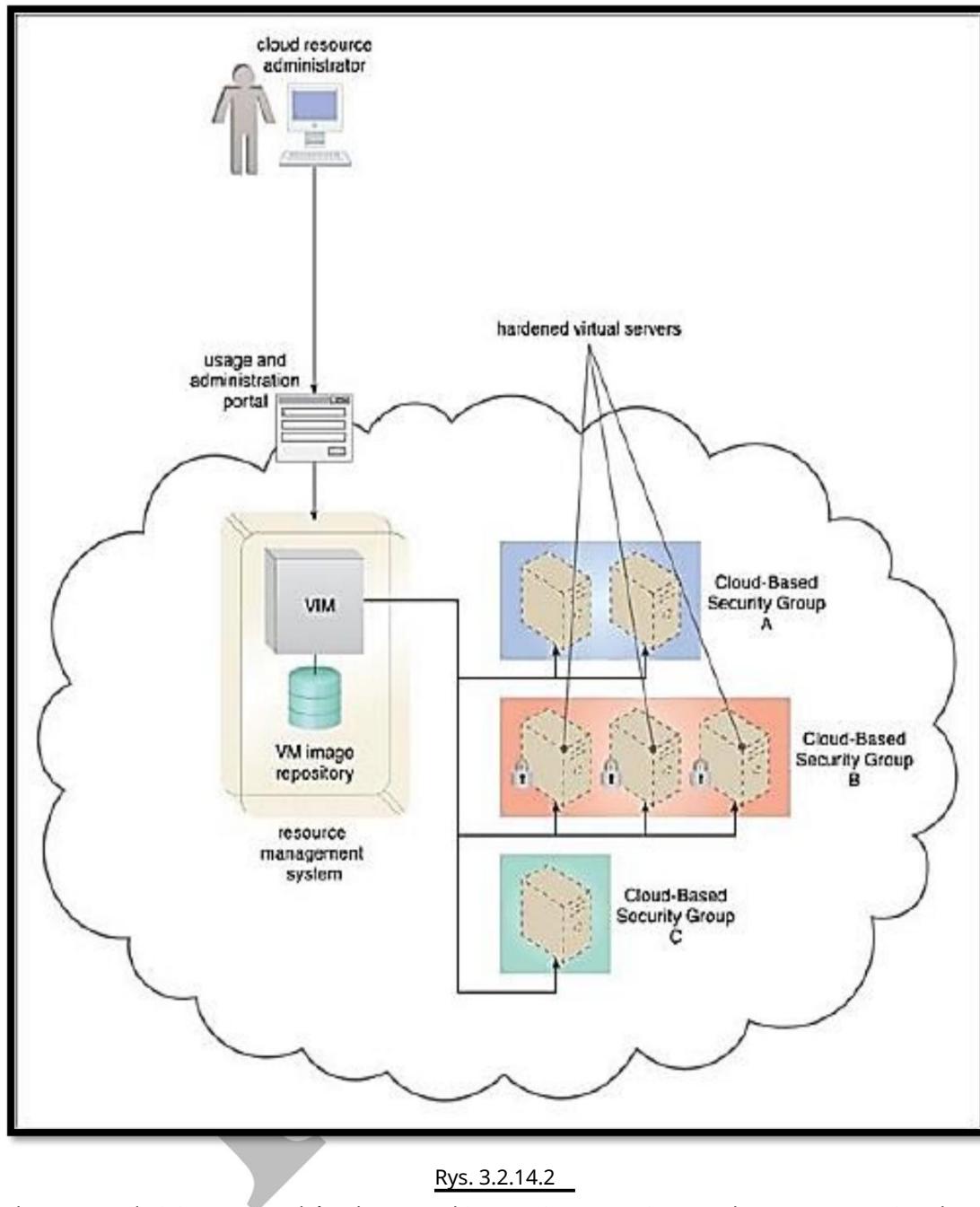
- Wzmocnione obrazy serwerów wirtualnych Jak wspomniano wcześniej, serwer wirtualny jest tworzony na podstawie konfiguracji szablonu zwanej obrazem serwera wirtualnego (lub obrazem maszyny wirtualnej). Hartowanie to proces usuwania niepotrzebnego oprogramowania z systemu w celu ograniczenia potencjalnych luk, które mogą zostać wykorzystane przez atakujących.
- Usuwanie zbędnych programów, zamknięcie niepotrzebnych portów serwerów i wyłączenie nieużywanych usług, wewnętrznych kont root i dostępu dla gości to przykłady wzmocniania zabezpieczeń. • Wzmocniony obraz serwera wirtualnego to szablon do tworzenia instancji usługi wirtualnej, który ma zostało poddane procesowi hartowania (Rys. 13).
- Zwykle powoduje to, że szablon serwera wirtualnego jest znacznie bezpieczniejszy niż szablon oryginalny obraz standardowy.



Rys. 3.2.14.1

Rysunek 3.2.14.1 Dostawca chmury stosuje swoje zasady bezpieczeństwa, aby wzmocnić swoje standardowe obrazy serwerów wirtualnych. Szablon obrazu hartowanego jest zapisywany w repozytorium obrazów maszyny wirtualnej jako część systemu zarządzania zasobami.

- Wzmocnione obrazy serwerów wirtualnych pomagają przeciwdziałać odmowie usługi, co jest niewystarczające autoryzacji i nakładających się granic zaufania.



Rys. 3.2.14.2

Rysunek 3.2.14.1 Administrator zasobów chmury wybiera opcję wzmocnionego obrazu serwera wirtualnego dla serwerów wirtualnych udostępnionych dla grupy zabezpieczeń B opartej na chmurze.

3.2.15 PRZYKŁADOWE ĆWICZENIE Z PYTANIAMI

1. Wyjaśnij system zdalnego zarządzania. Jakie dwa portale tworzą zdalnie system administracyjny?
 2. Jakie zadania najczęściej mogą wykonywać konsumenti chmury za pomocą pilota konsola administracyjna? Wyjaśnić.
 3. Czym jest zarządzanie zasobami? Jakie zadania są zazwyczaj zautomatyzowane i realizowane poprzez system zarządzania zasobami?
 4. Wyjaśnij szczegółowo system zarządzania SLA.
 5. Wyjaśnij szczegółowo system zarządzania rozliczeniami.
- 3.2. Co to jest szyfrowanie? Wyjaśnij jego rodzaje.
7. Co to jest haszowanie? Wyjaśnij za pomocą diagramu.
 8. Wyjaśnij szczegółowo podpis cyfrowy.
 9. Co to jest PKI (infrastruktura klucza publicznego)? Wyjaśnij szczegółowo.
 10. Czym jest zarządzanie tożsamością i dostępem (IAM)? Wyjaśnij jego składniki.
 11. Wyjaśnij szczegółowo logowanie jednokrotne (SSO).
 12. Szczegółowo wyjaśnij grupy zabezpieczeń oparte na chmurze.
 13. Jak wzmocnione obrazy serwerów wirtualnych? Wyjaśnić.

3.2.17 Odniesienia:

1. https://patterns.arcitura.com/cloud-computing-patterns/mechanisms/remote_administration_system
2. Koncepcje, technologia i architektura przetwarzania w chmurze autorstwa Thomasa Erla, Zaighama Mahmooda i Sala Ricardo Puttiniego Prentice'a - 2013.

Jednostka 4

Rozdział 1

Podstawowe architektury chmurowe

Struktura jednostki:

4.1.1 Architektura dystrybucji obciążenia

4.2.1 Architektura łączenia zasobów

4.3.1 Architektura dynamicznej skalowalności 4.4.1

Architektura elastycznej wydajności zasobów 4.5.1

Architektura równoważenia obciążenia usług

4.6.1 Architektura rozrywająca chmury

4.7.1 Architektura udostępniania dysków elastycznych

4.8.1 Architektura nadmiarowej pamięci masowej

Cel:

Aby dowiedzieć się, jak korzystać z usług w chmurze.

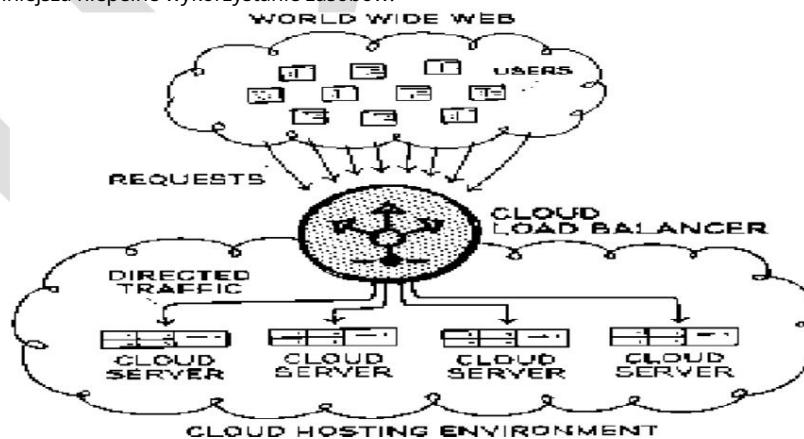
Wstęp:

W tym rozdziale przedstawiono i opisano kilka bardziej powszechnych, podstawowych modeli architektury chmur, z których każdy wyjaśnia typowe zastosowania i cechy współczesnych środowisk opartych na chmurze. Dalej

W rozdziale tym zbadano także zaangażowanie i znaczenie różnych kombinacji mechanizmów przetwarzania w chmurze w odniesieniu do tych architektur.

4.1. Architektura dystrybucji obciążenia

- Zasoby w chmurze można skalować poziomo przy użyciu dodatkowego lub identycznego zasobu i a moduł równoważenia obciążenia, który jest w stanie zapewnić dystrybucję obciążenia pomiędzy zasobami w czasie wykonywania.
- Taka architektura dystrybucji ma podwójną zaletę
 - I. Zmniejsza nadmierne wykorzystanie zasobów.
 - II. Zmniejsza niepełne wykorzystanie zasobów.



Rysunek: Architektura dystrybucji obciążenia

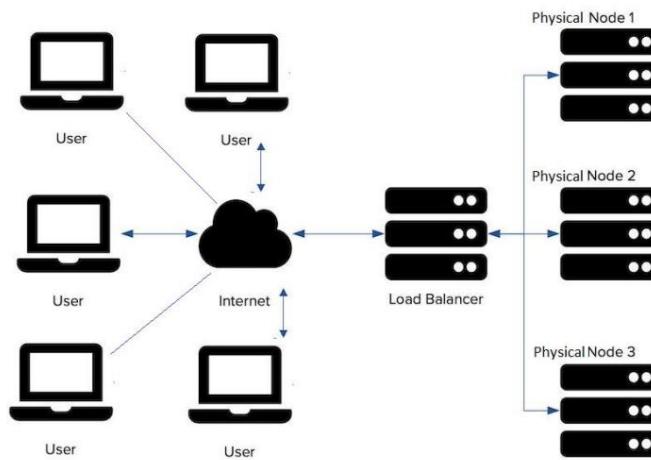
- Dystrybucja obciążenia odbywa się w celu obsługi rozproszonych serwerów wirtualnych, urządzeń pamięci masowej i usługi.
- System równoważenia obciążenia tworzy wyspecjalizowane odmiany, które uwzględniają aspekt obciążenia
 - I. Architektura instancji usług z równoważeniem obciążenia
 - II. Architektura instancji serwerów wirtualnych z równoważeniem obciążenia

iii. Architektura przełączników wirtualnych ze zrównoważonym obciążeniem

- Oprócz wyżej wymienionego mechanizmu częścią mogą być również następujące mechanizmy ta architektura chmurowa:
 - I. Monitor audytu – zasoby przetwarzające dane mogą określić, czy monitorowanie jest konieczne w celu spełnienia wymogów prawnych i regulacyjnych.
 - II. Cloud Usage Monitor – do wykonania środowiska wykonawczego można zaangażować różne monitory śledzenia obciążenia pracą i przetwarzanie danych.
 - iii. Hypervisor — obciążenie między hypervisorami a hostowanymi przez nie serwerami wirtualnymi może wymagać dystrybucji.
 - IV. Obwód sieci logicznej — obwód sieci logicznej izoluje konsumenta chmury granice sieci w odniesieniu do sposobu i miejsca dystrybucji obciążen.
 - v. Klaster zasobów – Klastrowe zasoby IT w trybie aktywny/aktywny są powszechnie używane do obsługi równoważenia obciążenia pomiędzy różnymi węzłami klastra.
 - VI. Replikacja zasobów – ten mechanizm może generować nowe instancje zwirtualizowanych zasobów IT w odpowiedzi na wymagania dotyczące dystrybucji obciążenia w czasie wykonywania.

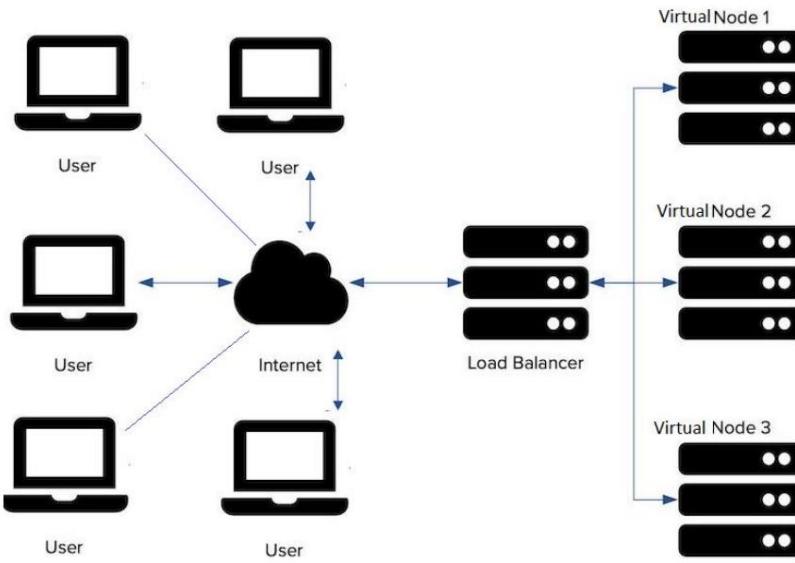
4.2. Architektura łączenia zasobów

- Architektura ta opiera się na wykorzystaniu jednego lub większej liczby zasobów z puli zasobów, w której identyczne zsynchronizowane zasoby są grupowane i obsługiwane przez system.
- Przykłady pul zasobów: 1) Pule serwerów fizycznych: to grupa serwerów fizycznych połączonych w sieć, na których zainstalowano systemy operacyjne oraz inne niezbędne programy i/lub aplikacje i które są gotowe do natychmiastowego użycia.

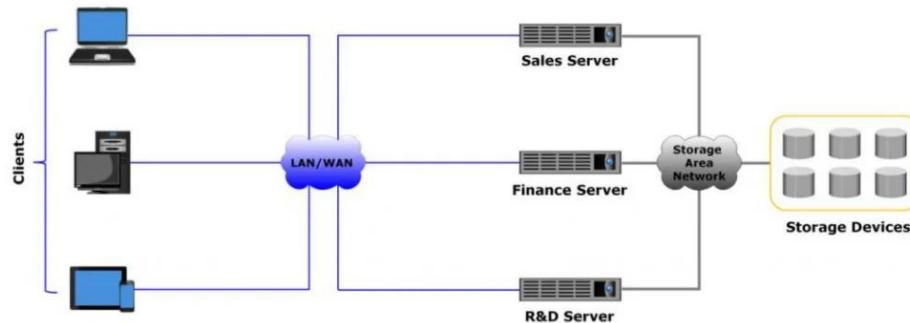


2) Pule serwerów wirtualnych: to grupa serwerów wirtualnych połączonych w sieć, które mają zostać zainstalowane systemów operacyjnych oraz innych niezbędnych programów i/lub aplikacji i są gotowe do natychmiastowego użycia. Zwykle konfiguruje się je przy użyciu jednego z kilku dostępnych szablonów wybranych przez konsumenta chmury podczas udostępniania.

- Na przykład użytkownik chmury może skonfigurować pulę systemów Windows średniej warstwy serwery z 4 GB pamięci RAM lub pula serwerów niższego poziomu Ubuntu z 2 GB pamięci RAM.

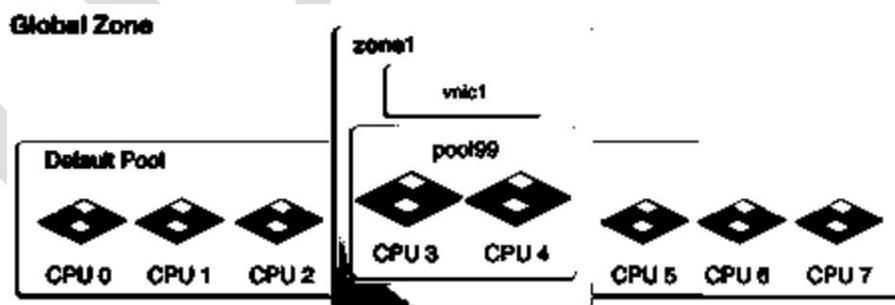


3) Pule pamięci masowej lub pule urządzeń pamięci masowej w chmurze: to grupa struktur pamięci masowej opartych na plikach lub blokach, które zawierają puste i/lub wypełnione urządzenia pamięci masowej w chmurze.

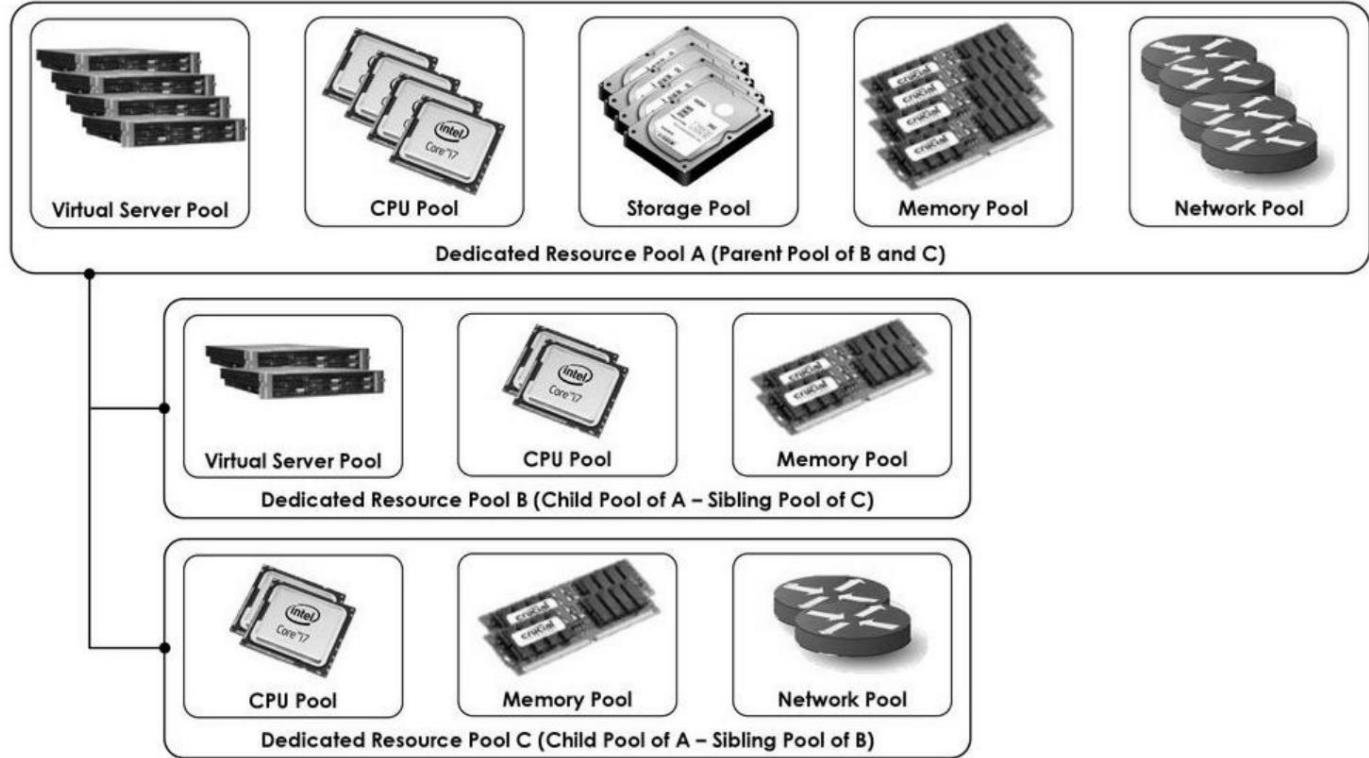


4) Pule sieciowe (lub pule połączeń wzajemnych): to grupa różnych, wstępnie skonfigurowanych urządzeń łączności sieciowej. • Można na przykład utworzyć pulę wirtualnych urządzeń firewall lub fizycznych przełączników sieciowych w celu zapewnienia nadmiarowej łączności, równoważenia obciążenia itp.

5) Pule procesorów: to grupa jednostek przetwarzających gotowa do przydzielenia serwerom wirtualnym i zazwyczaj podzielona na poszczególne rdzenie przetwarzające.



- Pule fizycznej pamięci RAM mogą być wykorzystywane w nowo udostępnianych serweraх fizycznych lub w pionie skalować serwery fizyczne.
- Dla każdego typu zasobu można tworzyć dedykowane pule, a poszczególne pule można grupować do większej puli, w takim przypadku każda pojedyncza pula stanie się podpułą.
- Pule zasobów mogą stać się bardzo złożone i tworzyć wiele pul dla konkretnych konsumentów lub aplikacji w chmurze. • Można ustanowić strukturę hierarchiczną, tworząc pule nadzędne, siostrzane i zagnieżdżone, aby ułatwić organizację różnorodnych wymagań dotyczących łączenia zasobów, jak pokazano na rysunku poniżej.



Rysunek: Inna architektura basenu

Architektura dynamicznej skalowalności to model architektoniczny oparty na systemie predefiniowanych warunków skalowania, które wyzwalają dynamiczną alokację zasobów IT z pul zasobów. Dynamiczna alokacja umożliwia zmienne wykorzystanie w zależności od wahań zapotrzebowania, ponieważ niepotrzebne zasoby IT są skutecznie odzyskiwane bez konieczności ręcznej interakcji.

W odbiorniku automatycznego skalowania skonfigurowano progi obciążenia, które decydują o tym, kiedy do przetwarzania obciążenia należy dodać nowe zasoby IT. Mechanizm ten może być wyposażony w logikę określającą, ile dodatkowych zasobów IT można dynamicznie udostępniać, w oparciu o warunki umowy o świadczenie usług danego konsumenta chmury.

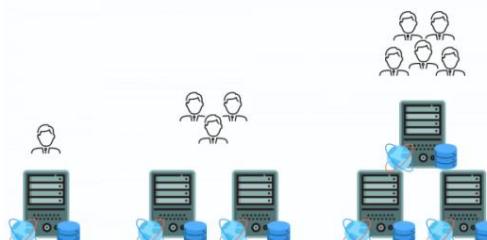
4.3. • Architektura dynamicznej skalowalności

Architektura ta jest modelem opartym na systemie predefiniowanych warunków skalowania puli zasobów, które wyzwalają dynamiczną alokację zasobów chmury z puli.

- Dynamiczna alokacja umożliwia zmienne wykorzystanie określone przez wahania zapotrzebowania na użytkowanie, co skutkuje efektywnym wykorzystaniem zasobów, a niepotrzebne zasoby są skutecznie odzyskiwane bez konieczności ręcznej interakcji.
- Istnieją trzy typy skalowania dynamicznego:

1. Dynamiczne skalowanie poziome – w tym typie instancje zasobów są skalowane w górę i w dół, aby obsłużyć dynamiczne obciążenia podczas wykonywania. Automatyczny odbiornik skalowania monitoruje żądania i sygnalizuje replikację zasobów w celu zainicjowania duplikacji zasobów zgodnie z wymaganiami i uprawnieniami ustalonymi przez administratora.

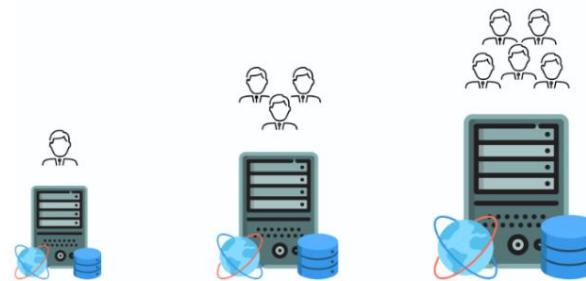
Horizontal Scaling



2. Dynamiczne skalowanie pionowe – w tym typie instancje zasobów są skalowane w górę i w dół gdy zachodzi potrzeba dostosowania mocy przerobowej pojedynczego zasobu. Na przykład:

Serwer wirtualny, który jest przeciążony, może mieć dynamicznie zwiększającą pamięć lub może zostać dodany rdzeń przetwarzający.

Vertical Scaling



3. Relokacja dynamiczna – w tym typie zasób jest relokowany do hosta o większej pojemności.

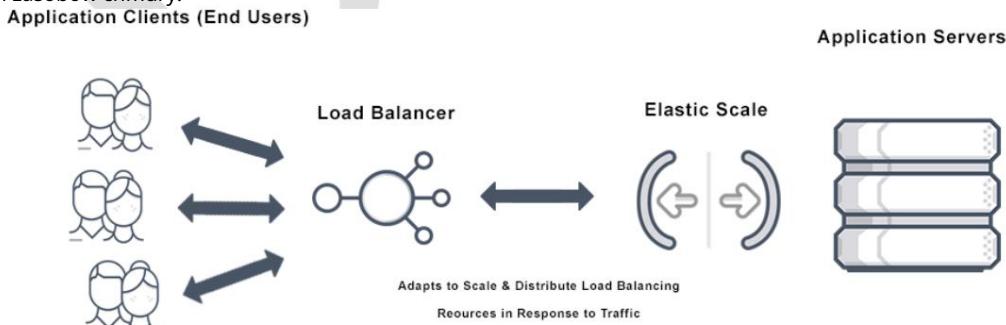
Na przykład może zaistnieć potrzeba przeniesienia pliku z taśmowego urządzenia pamięci masowej SAN o przepustowości we/wy 4 GB na sekundę na inne dyskowe urządzenie pamięci masowej SAN o przepustowości we/wy 8 GB na sekundę.

Architekturę dynamicznej skalowalności można zastosować do szeregu zasobów IT, w tym serwery wirtualne i urządzenia pamięci masowej w chmurze. Oprócz podstawowego odbiornika automatycznego skalowania i mechanizmów replikacji zasobów, dostępne są następujące mechanizmy można również zastosować w tej formie architektury chmury:

- Cloud Usage Monitor – wyspecjalizowane monitory wykorzystania chmury mogą śledzić wykorzystanie czasu działania w odpowiedzi na spowodowane tym dynamiczne wahania architektura.
- Hypervisor – Hypervisor jest wywoływany przez system dynamicznej skalowalności w celu tworzenia lub usuwania instancji serwerów wirtualnych lub w celu samodzielnego skalowania.
- Monitor płatności za użycie – monitor płatności za użycie jest zaangażowany do zbierania informacji o kosztach użytkowania w odpowiedzi na skalowanie zasobów IT.

4.4. • Architektura elastycznej pojemności zasobów

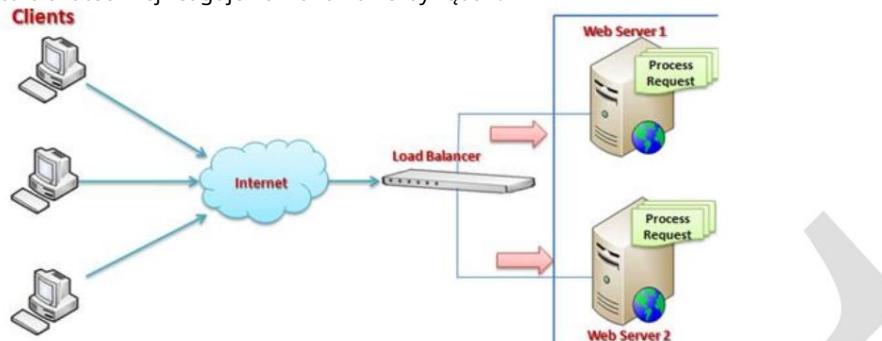
Architektura ta jest związana z dynamicznym udostępnianiem serwerów wirtualnych przy użyciu systemu, który przydziela i odzyskuje procesory i pamięć w natychmiastowej reakcji na zmieniające się wymagania dotyczące przetwarzania hostowanych zasobów chmury.



- Pule zasobów są wykorzystywane przez technologię skalowania, która współdziała z hiperwizorem i/lub VIM-em pobierać i zwracać zasoby procesora i pamięci RAM w czasie wykonywania.
- Przetwarzanie serwera wirtualnego w czasie wykonywania jest monitorowane, dzięki czemu można wykorzystać dodatkową moc obliczeniową z puli zasobów w drodze dynamicznej alokacji, zanim zostaną osiągnięte progi wydajności.
- Serwer wirtualny oraz hostowane na nim aplikacje i zasoby są w odpowiedzi skalowane w pionie. • Ten typ architektury chmurowej można zaprojektować tak, aby działał inteligentny skrypt automatyzacji wysyła żądanie skalowania poprzez VIM zamiast bezpośrednio do hypervisora.
- Serwery wirtualne uczestniczące w systemach elastycznej alokacji zasobów mogą wymagać ponownego uruchomienia aby dynamiczna alokacja zasobów zaczęła obowiązywać.

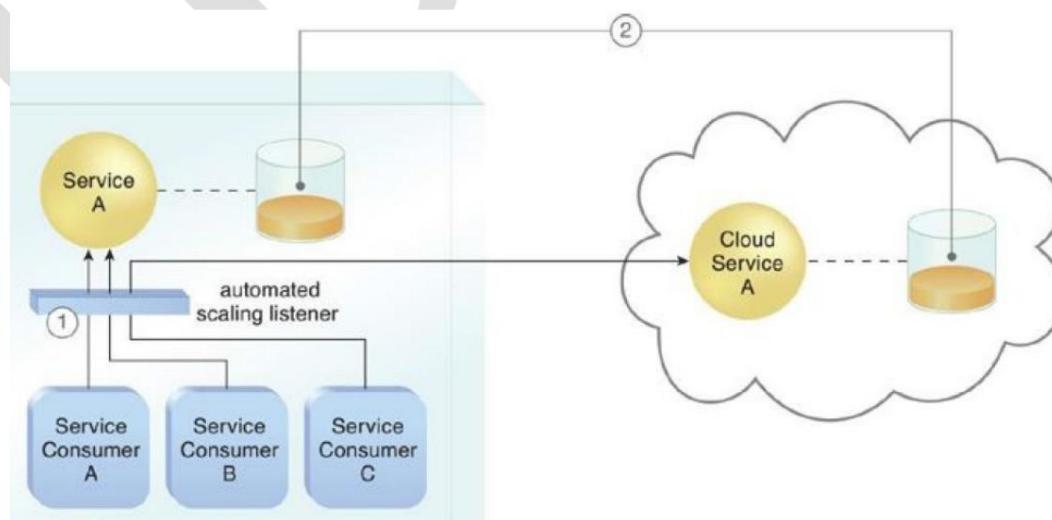
4.5. Architektura równoważenia obciążenia usług

- Ta architektura jest wyspecjalizowanym wariantem architektury dystrybucji obciążenia, która jest dostosowana specjalnie do skalowania wdrożeń usług w chmurze.
- Tworzone są redundantne wdrożenia usług w chmurze, do których dodawany jest system równoważenia obciążenia dynamicznie dystrybuuj obciążenia.
- Zduplikowane implementacje usług w chmurze są zorganizowane w pulę zasobów, podczas gdy moduł równoważenia obciążenia jest umieszczony jako komponent zewnętrzny lub wbudowany, aby umożliwić serwerom hosta samodzielne równoważenie obciążień.
- W zależności od przewidywanego obciążenia i mocy obliczeniowej serwera hosta
W różnych środowiskach można wygenerować wiele instancji każdej implementacji usługi w chmurze w ramach puli zasobów, która skuteczniej reaguje na wahania liczby żądań.



4.6. Architektura rozrywająca chmury

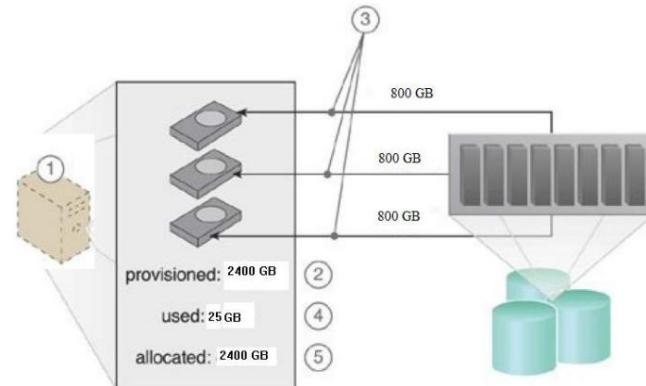
- Ta architektura ustanawia formę dynamicznego skalowania, która skaluje lub „wyrzuca” lokalne zasoby chmury do chmury po osiągnięciu wcześniej zdefiniowanych progów wydajności.
- Odpowiednie zasoby oparte na chmurze są wstępnie wdrożone nadmiarowo, ale pozostają nieaktywne aż do pojawienia się chmur. Gdy nie są już potrzebne, zasoby są zwalniane, a architektura „wkracza” z powrotem do środowiska lokalnego.
- Cloud Bursting to elastyczna, skalowalna architektura, która zapewnia konsumentom chmury możliwość korzystania z zasobów IT opartych na chmurze wyłącznie w celu spełnienia wyższych wymagań w zakresie użytkowania. • Podstawą tego modelu architektonicznego jest automatyczny detektor skalowania mechanizmy replikacji zasobów.
- Zautomatyzowany odbiornik skalowania określa, kiedy przekierować żądania do zasobów w chmurze, a replikacja zasobów służy do utrzymania synchronizacji między zasobami IT lokalnymi i chmurowymi w odniesieniu do informacji o stanie



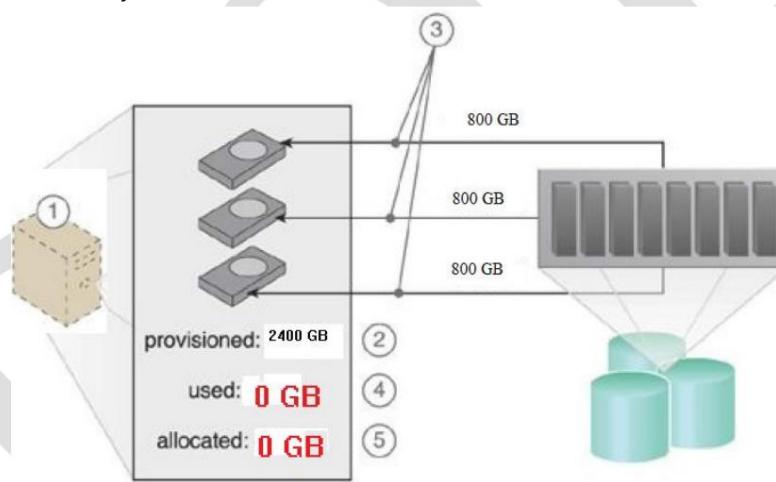
4.7. Architektura elastycznego udostępniania dysków

Konsumenci usług chmurowych są powszechnie obciążani opłatami za przestrzeń dyskową w chmurze na dysku stałym alokację przestrzeni dyskowej, co oznacza, że opłaty są ustalane z góry na podstawie pojemności dysku i nie odpowiadają faktycznemu zużyciu pamięci masowej.

- Chmura udostępnia serwer wirtualny z systemem operacyjnym Windows Server 2019 i trzema dyskami twardymi o pojemności 800 GB. Konsument chmury jest obciążany opłatą za późniejsze wykorzystanie 2400 GB przestrzeni dyskowej instalowanie systemu operacyjnego, mimo że system operacyjny wymaga tylko 25 GB przestrzeni dyskowej.



- Architektura elastycznego udostępniania dysków ustanawia dynamiczne udostępnianie pamięci system, który gwarantuje, że konsument chmury będzie szczegółowo rozliczany za dokładną ilość faktycznie wykorzystywanej przestrzeni dyskowej. System ten wykorzystuje technologię cienkiego udostępniania do dynamicznej alokacji przestrzeni dyskowej i jest dodatkowo wspierany przez monitorowanie wykorzystania w czasie wykonywania w celu gromadzenia dokładnych danych o użytkowaniu do celów rozliczeniowych

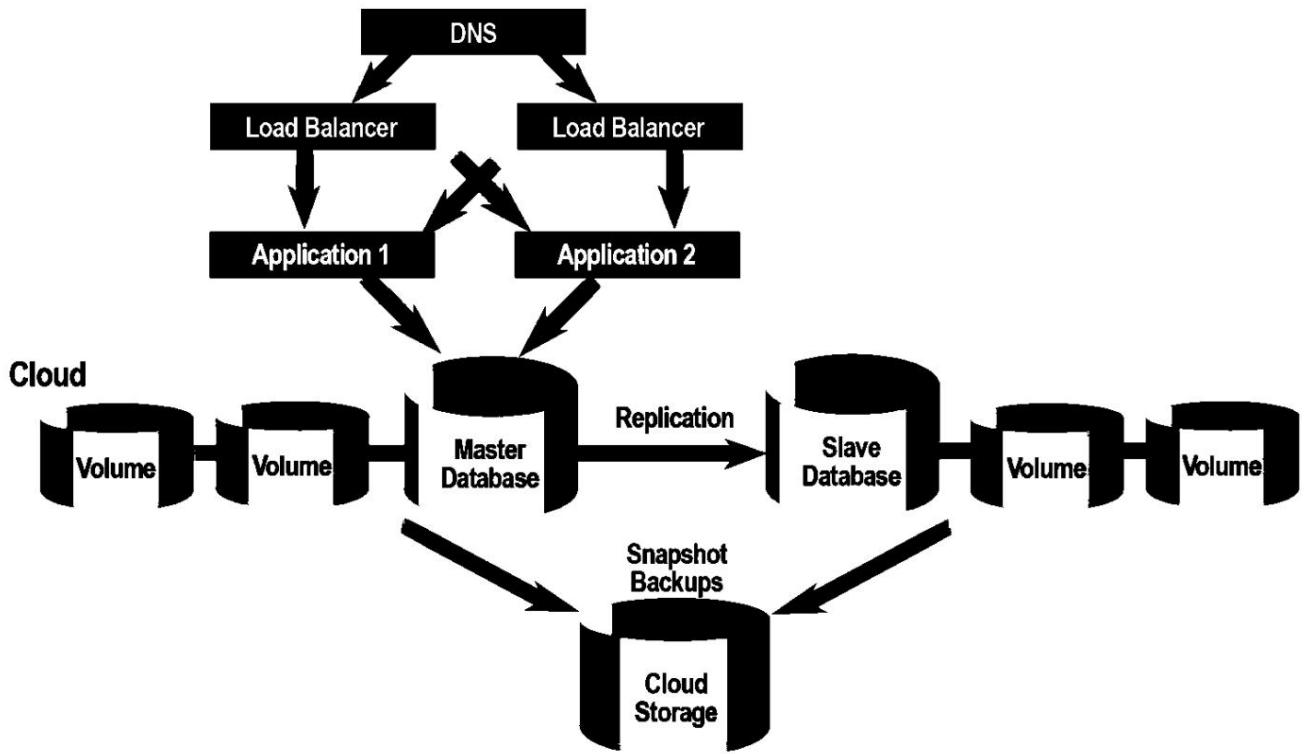


- Oprogramowanie do alokacji elastycznej jest instalowane na serwerach wirtualnych, które przetwarzają dynamiczną alokację pamięci za pośrednictwem hiperwizora, podczas gdy monitor płatności za użycie śledzi i raportuje szczegółowe dane dotyczące wykorzystania dysku związane z rozliczeniami.

4.8. Nadmiarowa architektura pamięci masowej

- Urządzenia do przechowywania w chmurze czasami ulegają awariom i zakłóceniom spowodowanym problemami z łącznością sieciową, awarią kontrolera lub ogólnym sprzętem albo naruszeniami bezpieczeństwa. • Naruszona niezawodność urządzenia pamięci masowej w chmurze może wywołać efekt falowy i spowodować awarię wszystkich usług, aplikacji i komponentów infrastruktury w chmurze zależnych od jego dostępności.

- W architekturze nadmiarowej pamięci masowej dodatkowe zduplikowane urządzenie pamięci masowej w chmurze stanowi część systemu awaryjnego, który synchronizuje dane z danymi znajdującymi się na głównym urządzeniu pamięci masowej w chmurze. • Bramka usług pamięci masowej przekierowuje żądania konsumentów chmury do urządzenia dodatkowego w przypadku awarii urządzenia podstawowego.



- Ta architektura chmury opiera się głównie na systemie replikacji pamięci masowej, który utrzymuje główne urządzenie pamięci masowej w chmurze zsynchronizowane ze zduplikowanymi dodatkowymi urządzeniami pamięci masowej w chmurze.
- Dostawcy usług w chmurze mogą lokalizować dodatkowe urządzenia do przechowywania w chmurze w innym regionie geograficznym niż główne urządzenie do przechowywania w chmurze, zwykle ze względów ekonomicznych. • Lokalizacja dodatkowych urządzeń pamięci masowej w chmurze może narzucać protokół i metodę używaną do synchronizacji, ponieważ niektóre protokoły transportu replikacji mają ograniczenia dotyczące odległości.

Rozdział – 8

Zaawansowane architektury chmurowe

Struktura jednostki:

- 8.1. Architektura klastrowania hiperwizora
- 8.2. Architektura instancji serwerów wirtualnych z równoważeniem obciążenia
- 8.3. Architektura przenoszenia usług bez zakłóceń
- 8.4. Architektura zero przestojów
- 8.5. Architektura równoważenia chmur
- 8.6. Architektura rezerwacji zasobów
- 8.7. Architektura dynamicznego wykrywania i odzyskiwania awarii usług bare-metal
- 8.8. Architektura udostępniania szybkiego udostępniania
- 8.9. Architektura zarządzania obciążeniem pamięci masowej

Cel:

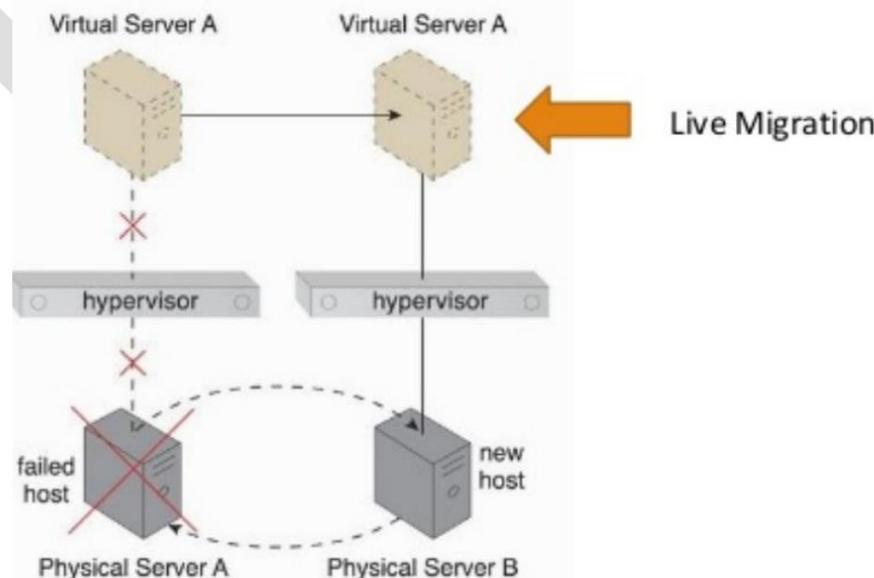
Aby dowiedzieć się, jak korzystać z usług Advance Cloud Services.

Wstęp:

W tym rozdziale przedstawiono odrębne i wyrafinowane warstwy architektury technologii chmury, z których kilka można zbudować w oparciu o bardziej podstawowe środowiska ustanowione przez modele architektoniczne omówione w poprzednim rozdziale.

8.1. Architektura klastrowania hiperwizora

- Hiperwizorzy są odpowiedzialni za tworzenie i hostowanie wielu serwerów wirtualnych.
- Z powodu tej zależności wszelkie awarie wpływające na hiperwizora mogą mieć kaskadowy wpływ na jego serwery wirtualne.
- Architektura klastrowa hypervisorów tworzy klasztro hypervisorów o wysokiej dostępności na wielu serwerach fizycznych.
- Jeśli dany hypervisor lub odpowiadający mu serwer fizyczny stanie się niedostępny, hostowane serwery wirtualne można przenieść na inny serwer fizyczny lub hiperwizor, aby zachować czas działania operacji.

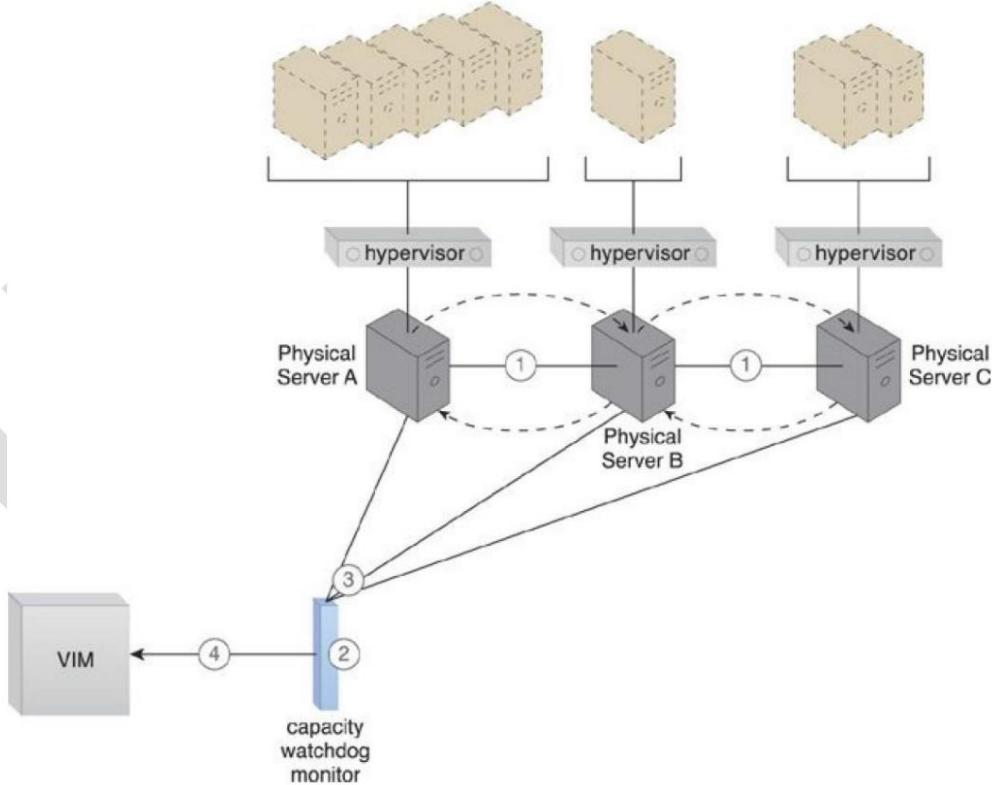


- Klasztro hypervisorów jest kontrolowane przez centralny moduł VIM, który regularnie wysyła komunikaty o pulsie do hypervisorów, aby potwierdzić, że działają.

- Niepotwierdzone komunikaty pulsu powodują, że VIM inicjuje program migracji maszyny wirtualnej na żywo, aby dynamicznie przenieść dotknięte serwery wirtualne do nowego hosta.

8.2. Architektura instancji serwerów wirtualnych z równoważeniem obciążenia

- Czasami utrzymywanie równomiernego obciążenia między serwerami pomiędzy serwerami fizycznymi, których działanie i zarządzanie są odizolowane, co może być najtrudniejszym elementem w chmurze.
- Serwer fizyczny może z łatwością obsługiwać więcej serwerów wirtualnych lub przyjmować większe obciążenia niż sąsiednie serwery fizyczne.
- Zarówno nadmierne, jak i niedostateczne wykorzystanie serwera fizycznego może z czasem znacznie wzrosnąć, co prowadzi do: ciągłe wyzwania związane z wydajnością (w przypadku nadmiernie wykorzystywanych serwerów) i ciągłe marnotrawstwo (w przypadku utraconego potencjału przetwarzania w przypadku niedostatecznie wykorzystanych serwerów).
- Architektura instancji serwerów wirtualnych z równoważeniem obciążenia ustanawia system nadzoru wydajności, który dynamicznie oblicza instancje serwerów wirtualnych i powiązane z nimi obciążenia, przed rozdzieleniem przetwarzania pomiędzy dostępne hosty serwerów fizycznych.
- System nadzorujący pojemność składa się z monitora wykorzystania chmury, monitorującego pojemność program migracji maszyn wirtualnych na żywo i narzędzie do planowania wydajności.
- Monitor wydajności monitoruje wykorzystanie serwera fizycznego i wirtualnego oraz rapportuje wszelkie znaczące wahania planiste wydajności, który jest odpowiedzialny za dynamiczne obliczanie wydajności obliczeniowej serwera fizycznego w stosunku do wymagań dotyczących wydajności serwera wirtualnego.
- Jeśli planista wydajności zdecyduje się przenieść serwer wirtualny na inny host w celu rozłożenia obciążenia, program migracji maszyny wirtualnej na żywo otrzymuje sygnał o konieczności przeniesienia serwera wirtualnego.



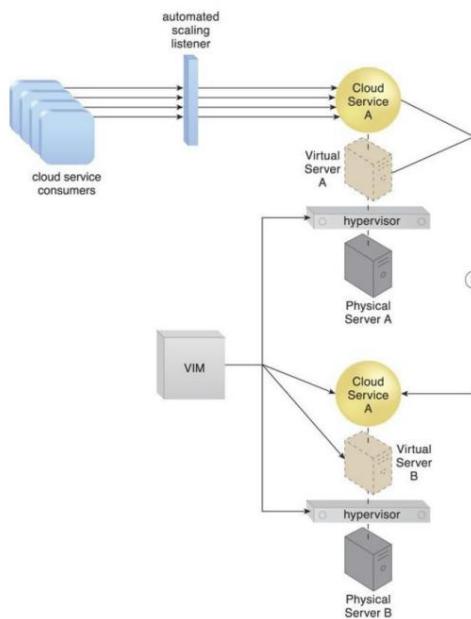
8.3. • Architektura przenoszenia usług bez zakłóceń

Usługa w chmurze może stać się niedostępna z wielu powodów, takich jak:

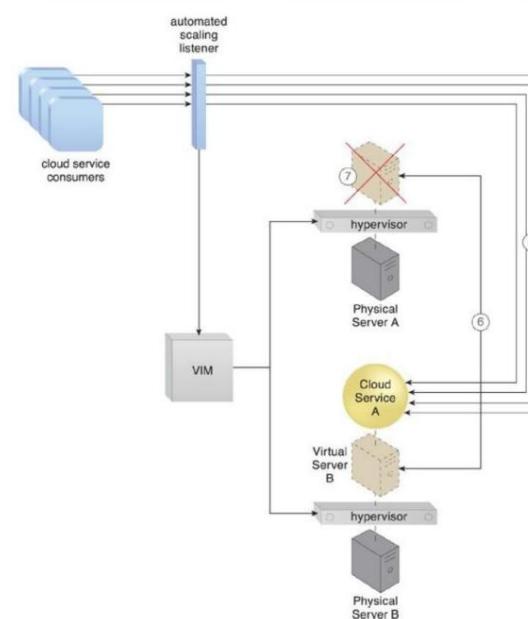
- 1) wymagania dotyczące wykorzystania środowiska wykonawczego, które przekraczają jego możliwości przetwarzania
- 2) aktualizacja konserwacyjna nakazująca tymczasową przerwę w działaniu
- 3) trwała migracja na nowy host serwera fizycznego

- Żądania konsumentów dotyczące usług w chmurze są zwykle odrzucane, jeśli usługa w chmurze staje się niedostępna, co może potencjalnie skutkować powstaniem wyjątkowych warunków.

- Niezakłócająca pracy architektura relokacji usług ustanawia system, w którym wcześniej zdefiniowane zdarzenie inicjuje duplikację lub migrację wdrożenia usługi w chmurze w czasie jej wykonywania, unikając w ten sposób jakichkolwiek zakłóceń.
- Zamiast skalować lub zmniejszać usługi w chmurze za pomocą nadmiarowych implementacji, działalność usług w chmurze można tymczasowo przenieść do innego środowiska hostingowego w czasie wykonywania, dodając zduplikowaną implementację na nowym hoście.
- Podobnie żądania konsumentów dotyczące usług w chmurze mogą być tymczasowo przekierowywane do duplikatów wdrożenia, gdy pierwotna implementacja musi przejść przerwę konserwacyjną. • Przeniesienie wdrożenia usługi w chmurze i wszelkiej działalności związanej z usługą w chmurze może również mieć charakter trwały, aby uwzględnić migrację usług w chmurze do nowych hostów serwerów fizycznych.
- Kluczowym aspektem podstawowej architektury jest wdrożenie nowej usługi w chmurze gwarantujemy, że pomyślnie będziemy odbierać żądania konsumentów dotyczące usług w chmurze i odpowiadając na nie, zanim pierwotna implementacja usługi w chmurze zostanie dezaktywowana lub usunięta.
- Powszechnym podejściem jest migracja maszyny wirtualnej na żywo w celu przeniesienia całej instancji serwera wirtualnego hosting usługi w chmurze.
- Mechanizmy automatycznego nasłuchiwanego skalowania i/lub mechanizmu równoważenia obciążenia można wykorzystać do tymczasowego przekierowania żądań konsumentów usług w chmurze w odpowiedzi na wymagania dotyczące skalowania i rozkładu obciążenia. Każdy mechanizm może skontaktować się z VIMem w celu zainicjowania procesu migracji maszyny wirtualnej na żywo.



Rys. : Przed awarią • Migracja



Rys. : Po awarii

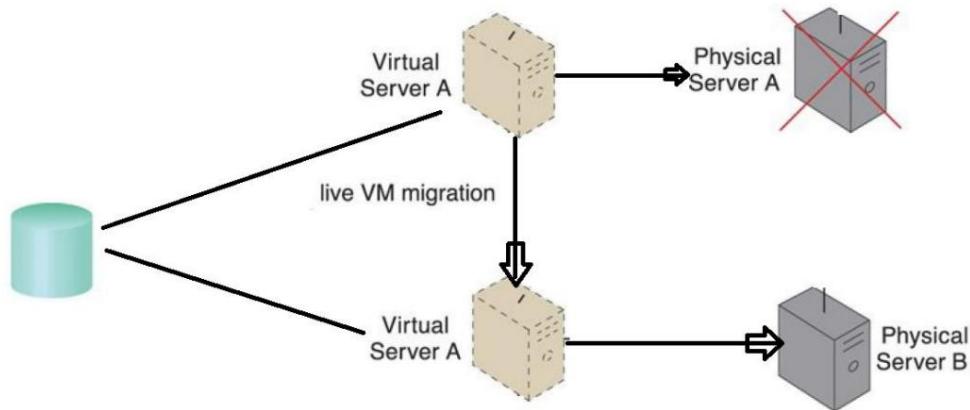
serwera wirtualnego może nastąpić na jeden z dwóch poniższych sposobów, w zależności od lokalizacji dysków serwera wirtualnego i jego konfiguracji:

- Kopia dysków serwera wirtualnego jest tworzona na hoście docelowym, jeśli dyski serwera wirtualnego są przechowywane na lokalnym urządzeniu pamięci masowej lub na niewspółdzielonym zdalnym urządzeniu magazynującym podłączonym do hosta źródłowego. Po utworzeniu kopii obie instancje serwerów wirtualnych są synchronizowane, a pliki serwera wirtualnego są usuwane z hosta źródłowego.
- Kopiowanie dysków serwera wirtualnego nie jest konieczne, jeśli pliki serwera wirtualnego są przechowywane na zdalnym urządzeniu pamięci masowej, które jest współdzielone między hostami źródłowymi i docelowymi. Własność serwera wirtualnego jest po prostu przenoszona ze źródła na docelowy host serwera fizycznego, a stan serwera wirtualnego jest automatycznie synchronizowany.

8.4. • Architektura zero przestojów

Serwer fizyczny w naturalny sposób działa jako pojedynczy punkt awarii dla hostowanych na nim serwerów wirtualnych. W rezultacie awaria serwera fizycznego lub jego naruszenie może mieć wpływ na dostępność któregoś (lub wszystkich) hostowanych serwerów wirtualnych. To sprawia, że udzielenie przez dostawcę usług w chmurze gwarancji zerowych przestojów na rzecz konsumentów usług w chmurze stanowi wyzwanie.

- Architektura zapewniająca zerowy czas przestojów ustanawia wyrafinowany system przełączania awaryjnego, który umożliwia dynamiczne przenoszenie serwerów wirtualnych do różnych hostów serwerów fizycznych w przypadku awarii ich pierwotnego hosta serwera fizycznego



8.5. Architektura równoważenia chmur

Architektura ta ustanawia wyspecjalizowany model architektoniczny, w którym zasoby chmury mogą być równoważone w wielu chmurach. Równoważenie między chmurami żądań konsumentów usług w chmurze może pomóc:

- 1) poprawić wydajność i skalowalność zasobów
- 2) zwiększyć dostępność i niezawodność zasobów
- 3) poprawić równoważenie obciążenia i optymalizację zasobów

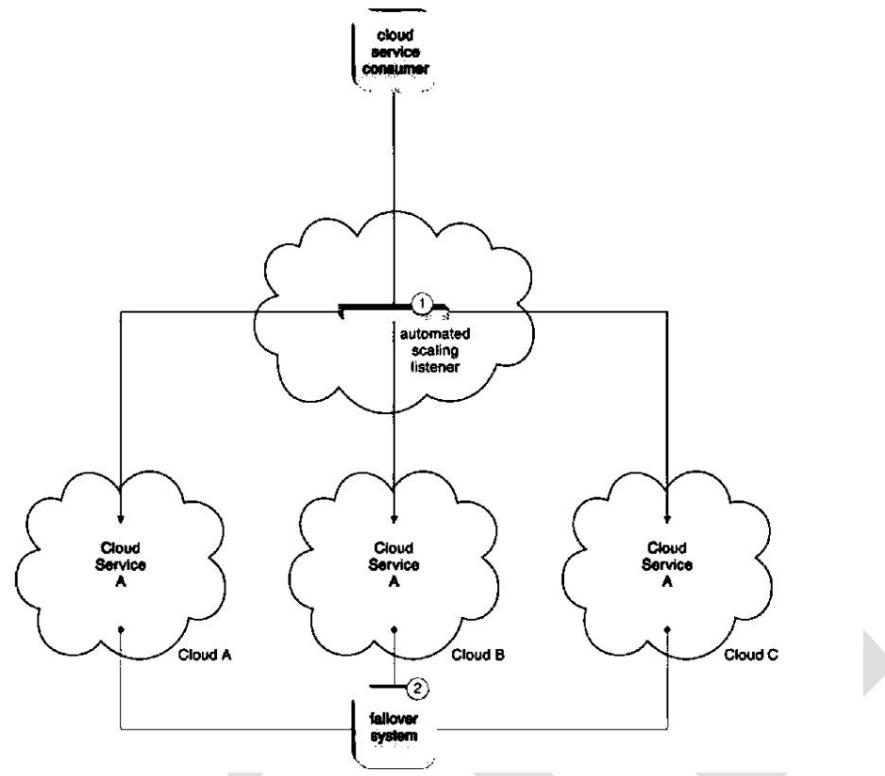
Jego funkcjonalność opiera się przede wszystkim na połączeniu automatycznego słuchacza skalowania i mechanizmów systemu przełączania awaryjnego. Częścią tego kompletnego rozwiązania może być znacznie więcej komponentów i mechanizmów architektury.

Na początek oba mechanizmy wykorzystuje się w następujący sposób:

- o Automatyczny odbiornik skalowania przekierowuje żądania konsumentów usług w chmurze do jednej z kilku implementacji nadmiarowych zasobów IT w oparciu o bieżące wymagania dotyczące skalowania i wydajności.
- o System przełączania awaryjnego zapewnia możliwość przełączania awaryjnego nadmiarowych zasobów IT między chmurami w przypadku awarii zasobu IT lub jego podstawowego środowiska hostingowego. Awarie zasobów IT są ogłoszane, dzięki czemu odbiornik automatycznego skalowania może uniknąć niezamierzonych kierowania żądań konsumentów usług w chmurze do niedostępnych lub niestabilnych zasobów IT.

Aby architektura równoważąca chmurę działała efektywnie, słuchacz zautomatyzowanego skalowania musi być świadomy wszystkich implementacji nadmiarowych zasobów IT w ramach architektury zrównoważonej w chmurze.

Ponadto, jeśli ręczna synchronizacja wdrożeń zasobów IT między chmurami nie jest możliwa, może być konieczne włączenie mechanizmu replikacji zasobów w celu zautomatyzowania synchronizacji.



8.6. Architektura rezerwacji zasobów

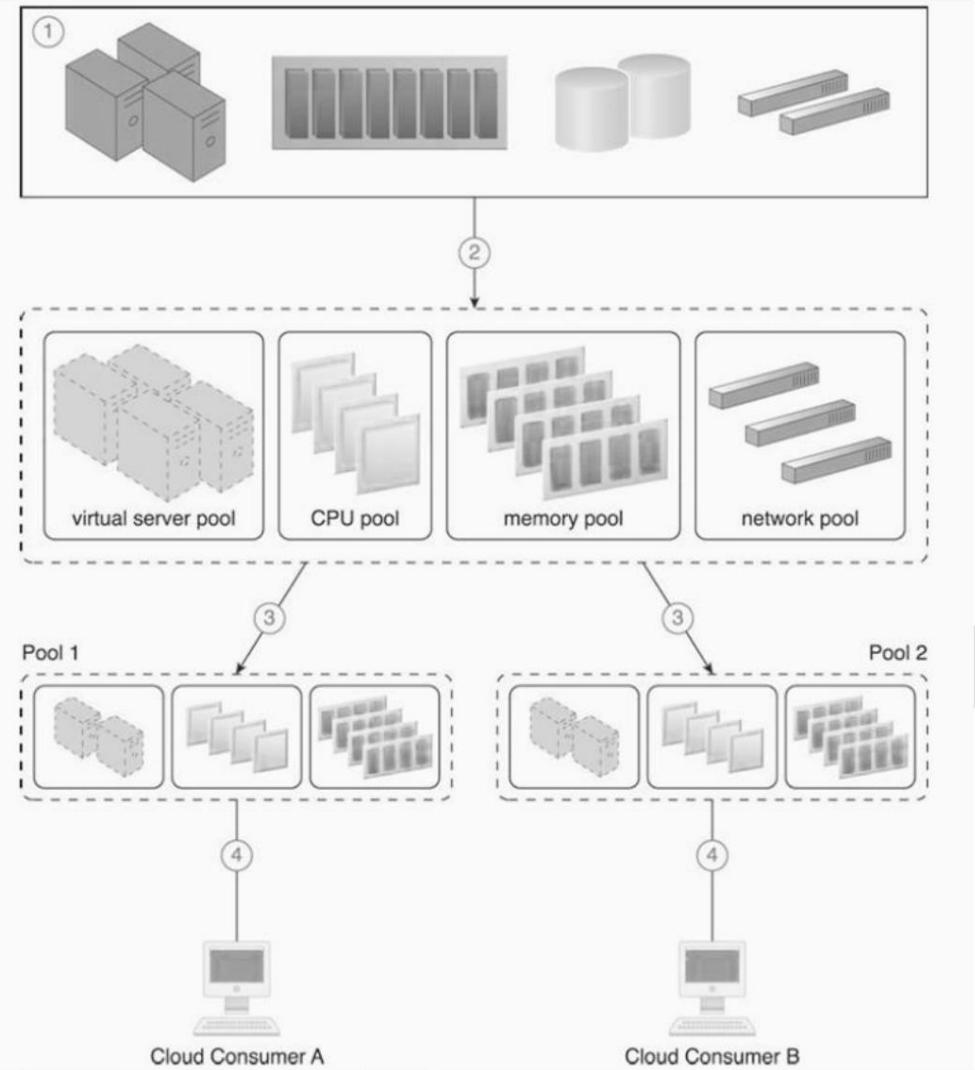
Architektura rezerwacji zasobów ustanawia system, w którym jedno z poniższych jest zarezerwowane wyłącznie dla danego konsumenta chmury

- pojedynczy zasób
- część zasobu
- wiele zasobów

Stworzenie systemu rezerwacji zasobów może wymagać zaangażowania mechanizmu systemu zarządzania zasobami, który służy do określenia progów wykorzystania poszczególnych zasobów i pul zasobów. Rezerwacje blokują ilość zasobów, które każda pula musi przechowywać, a pozostała część zasobów puli jest nadal dostępna do udostępniania i pożyczania. Mechanizm systemu zdalnego administrowania umożliwia także dostosowywanie frontonu, dzięki czemu klienci korzystający z chmury mają kontrolę administracyjną w zakresie zarządzania zarezerwowanymi alokacjami zasobów.

Typami mechanizmów powszechnie zarezerwowanych w tej architekturze są urządzenia pamięci masowej w chmurze i serwery wirtualne. Inne mechanizmy, które mogą być częścią architektury, mogą obejmować:

- o Monitor audytu o
- Monitor wykorzystania chmury o
- Hypervisor o Obwód sieci logicznej
- o Replikacja zasobów



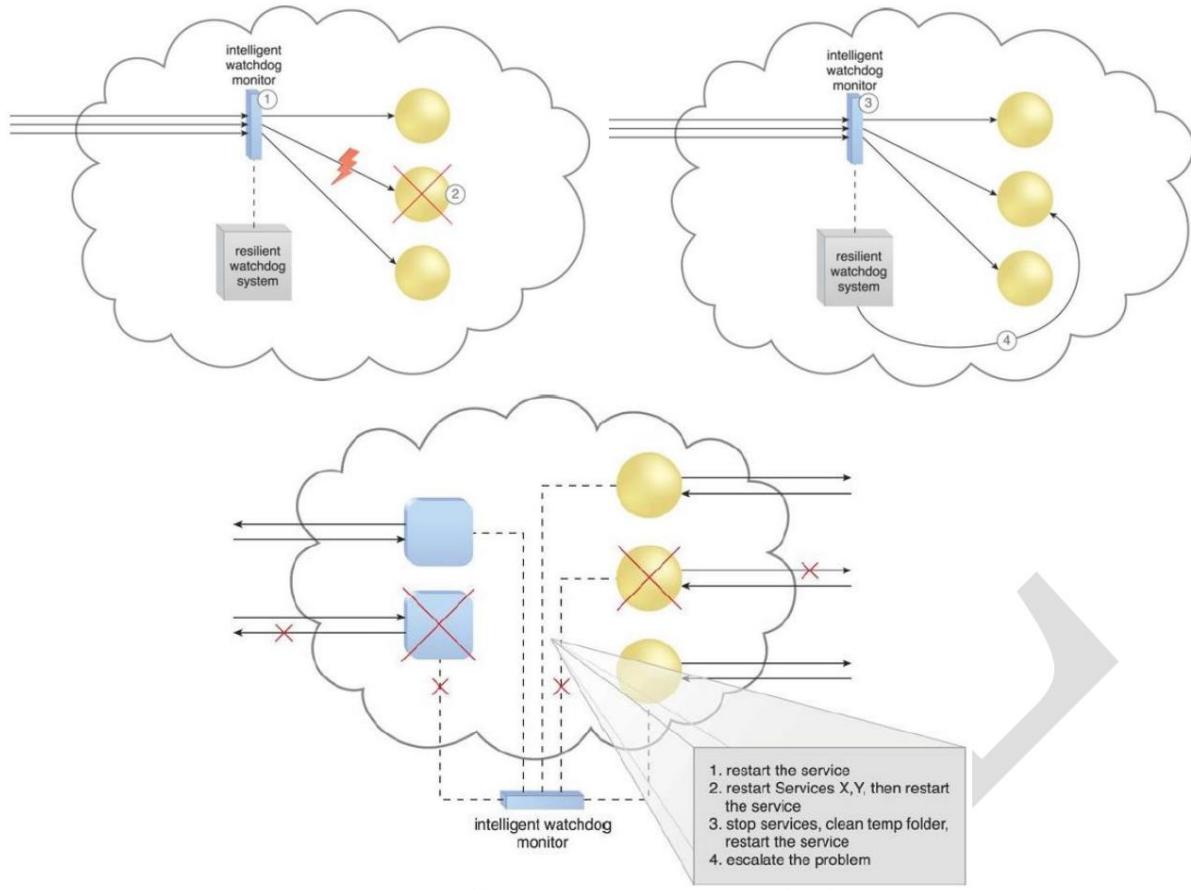
8.7. Architektura dynamicznego wykrywania i odzyskiwania awarii

Architektura ta ustanawia odporny system nadzorujący do monitorowania i reagowania na szeroki zakres wstępnie zdefiniowanych scenariuszy awarii. System ten powiadamia i eskaluje warunki awarii, których nie może sam automatycznie rozwiązać. Opiera się na wyspecjalizowanym monitorze wykorzystania chmury zwany inteligentnym monitorem monitorującym, który aktywnie śledzi zasoby i podejmuje wcześniej zdefiniowane działania w odpowiedzi na wcześniej zdefiniowane wydarzenia.

Odporny system nadzorujący spełnia pięć następujących podstawowych funkcji:

- 1) oglądanie
- 2) podjęcie decyzji o wydarzeniu
- 3) działanie na podstawie zdarzenia
- 4) raportowanie
- 5) eskalacja

Dla każdego zasobu można zdefiniować zasady odtwarzania sekwencyjnego, aby określić kroki, jakie musi podjąć inteligentny monitor monitorujący w przypadku wystąpienia awarii. Na przykład zasady odzyskiwania mogą określać, że przed wysłaniem powiadomienia należy automatycznie przeprowadzić jedną próbę odzyskania danych

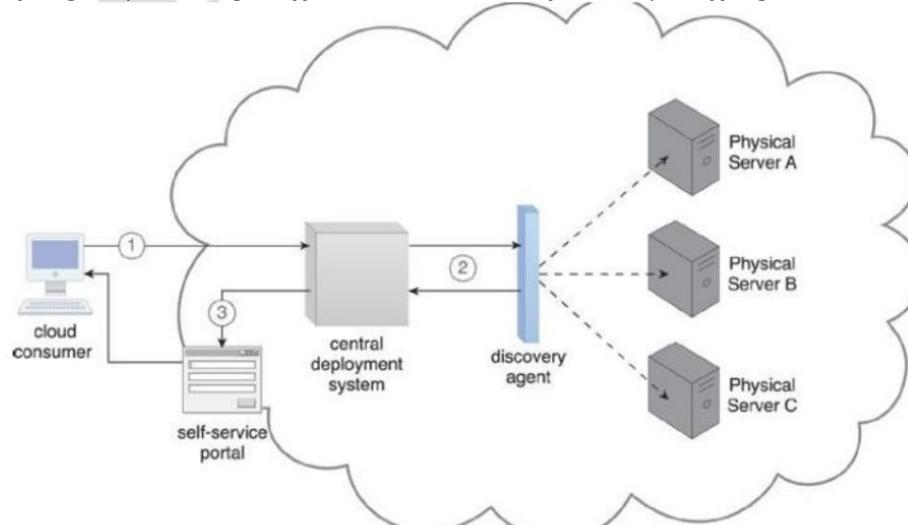


8.8. Architektura udostępniania usług bare-metal

Architektura ta ustanawia system wykorzystujący tę funkcję wraz z wyspecjalizowanymi agentami usług, które służą do zdalnego wykrywania i skutecznego udostępniania całych systemów operacyjnych.

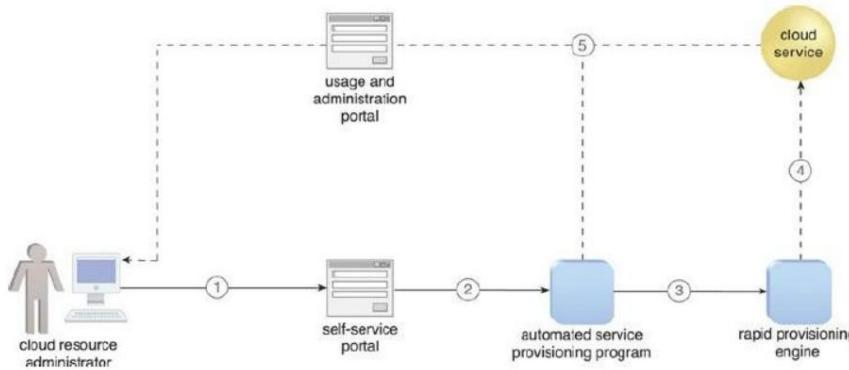
Oprogramowanie do zdalnego zarządzania zintegrowane z pamięcią ROM serwera staje się dostępne po uruchomieniu serwera. Do łączenia się z natywnym interfejsem zdalnego zarządzania serwera fizycznego zwykle używany jest internetowy lub zastrzeżony interfejs użytkownika, taki jak portal udostępniany przez system zdalnej administracji. Adresy IP na platformach IaaS można przekazywać bezpośrednio do konsumentów chmury, aby mogli niezależnie przeprowadzać instalacje systemu operacyjnego bare-metal.

System udostępniania bare-metal zapewnia funkcję automatycznego wdrażania, która umożliwia klientom chmury łączenie się z oprogramowaniem wdrożeniodowozowym i jednocześnie udostępnianie więcej niż jednego serwera lub systemu operacyjnego. Centralny system wdrażania łączy się z serwerami za pośrednictwem ich interfejsów zarządzania i używa tego samego protokołu do przesyłania plików i działania jako agent w pamięci RAM serwera fizycznego. Serwer typu bare-metal staje się wówczas surowym klientem z zainstalowanym agentem zarządzania, a oprogramowanie wdrażające przesyła wymagane pliki konfiguracyjne w celu wdrożenia systemu operacyjnego.



8.9. Architektura szybkiego udostępniania

Architektura szybkiego udostępniania tworzy system, który automatyzuje dostarczanie szerokiego zakresu zasobów, indywidualnie lub zbiorowo. Podstawowa architektura technologii szybkiego udostępniania zasobów może być wyrafinowana i złożona i opiera się na systemie składającym się z programu do automatycznego udostępniania, silnika szybkiego udostępniania oraz skryptów i szablonów do udostępniania na żądanie.



- (1) Administrator zasobów w chmurze żąda nowej usługi w chmurze w ramach samoobsług portalu.
- (2) Portal samoobsługowy przekazuje żądanie do programu automatycznego świadczenia usług zainstalowanego na serwerze wirtualnym.
- (3) który przekazuje niezbędne zadania do wykonania silnikowi szybkiego udostępniania.
- (4) Silnik szybkiego udostępniania ogłasza, kiedy nowa usługa w chmurze będzie gotowa.
- (5) Program zautomatyzowanego świadczenia usług finalizuje i publikuje usługę w chmurze w portalu użytkowania i administracji umożliwiającym dostęp do chmury dla konsumentów.

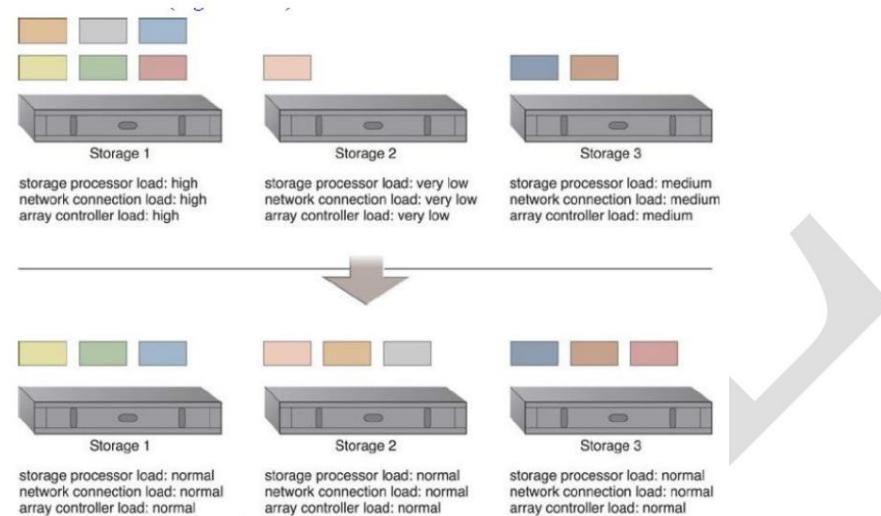
Opis krok po kroku opisuje wewnętrzne działanie mechanizmu szybkiego udostępniania:

1. Konsument chmury składa wniosek o nowy serwer za pośrednictwem portalu samoobsługowego.
2. Menedżer sekwencji przekazuje żądanie do silnika wdrażania w celu przygotowania systemu operacyjnego.
3. Silnik wdrażania używa szablonów serwerów wirtualnych do udostępniania, jeśli żądanie dotyczy serwera wirtualnego. W przeciwnym razie aparat wdrażania wysyła żądanie udostępnienia serwera fizycznego.
4. Predefiniowany obraz dla żadanego typu systemu operacyjnego jest używany do udostępniania systemu operacyjnego, jeśli jest dostępny. W przeciwnym razie wykonywany jest zwykły proces wdrażania w celu zainstalowania systemu operacyjnego.
5. Silnik wdrażania informuje menedżera sekwencji, kiedy system operacyjny jest gotowy gotowy.
6. Menedżer sekwencji aktualizuje i wysyła dzienniki do rejestratora sekwencji składowanie.
7. Menedżer sekwencji żąda, aby silnik wdrażania zastosował operację linię bazową systemu do udostępnionego systemu operacyjnego.
8. Mechanizm wdrażania stosuje żadaną wersję bazową systemu operacyjnego.
9. Silnik wdrażania informuje menedżera sekwencji, że system operacyjny zastosowano poziom bazowy.
10. Menedżer sekwencji aktualizuje i wysyła dzienniki ukończone kroków do rejestrator sekwencji do przechowywania.
11. Menedżer sekwencji żąda, aby silnik wdrażania zainstalował aplikacje.
12. Silnik wdrażania wdraża aplikacje na udostępnionym serwerze.
13. Silnik wdrażania informuje menedżera sekwencji, że aplikacje mają zostać zainstalowany.
14. Menedżer sekwencji aktualizuje i wysyła dzienniki ukończone kroków do rejestrator sekwencji do przechowywania.
15. Menedżer sekwencji żąda, aby silnik wdrażania zastosował linię bazową konfiguracji aplikacji.

16. Silnik wdrażania stosuje linię bazową konfiguracji.
17. Silnik wdrażania informuje menedżera sekwencji, że konfiguracja zastosowano poziom bazowy.
18. Menedżer sekwencji aktualizuje i wysyła dzienniki ukończonych kroków do rejestrator sekwencji do przechowywania.

8.10. Architektura zarządzania obciążeniem pamięci masowej

Architektura ta umożliwia równomierną dystrybucję jednostek LUN pomiędzy dostępnymi urządzeniami pamięci masowej w chmurze, a system pojemności pamięci masowej jest tworzony w celu zapewnienia równomiernego rozłożenia obciążzeń w czasie wykonywania pomiędzy jednostkami LUN.



Łączenie urządzeń pamięci masowej w chmurze w grupę umożliwia równomierną dystrybucję danych LUN pomiędzy dostępnymi hostami pamięci masowej. Skonfigurowano system zarządzania pamięcią masową i umieszczono zautomatyzowany odbiornik skalowania w celu monitorowania i wyrównywania obciążzeń w czasie wykonywania pomiędzy zgrupowanymi urządzeniami pamięci masowej w chmurze.

Słowniczek:

- Inteligentny silnik automatyzacji: Inteligentny silnik automatyzacji automatyzuje administrację zadania poprzez wykonanie skryptów zawierających logikę przepływu pracy.
- LUN: Numer jednostki logicznej (LUN) to dysk logiczny reprezentujący partycję dysku fizycznego prowadzić.
- Bramka usług pamięci masowej: Bramka usług pamięci masowej to komponent, który działa jako zewnętrzny interfejs do usług przechowywania w chmurze i jest w stanie automatycznie przekierowywać żądania konsumentów w chmurze za każdym razem, gdy zmieni się lokalizacja żądanego danych.
- Replikacja pamięci masowej: Replikacja pamięci masowej jest odmianą mechanizmów replikacji zasobów używany do synchronicznej lub asynchronicznej replikacji danych z podstawowego urządzenia pamięci masowej na dodatkowe urządzenie pamięci masowej. Można go używać do replikowania częściowych i całych jednostek LUN.
- Heartbeats: Heartbeats to komunikaty na poziomie systemu wymieniane pomiędzy hypervisorami, hypervisory i serwery wirtualne oraz hypervisory i VIM.
- Migracja maszyn wirtualnych na żywo: Migracja maszyn wirtualnych na żywo to system umożliwiający relokację wirtualną serwerów lub instancji serwerów wirtualnych w czasie wykonywania.
- Migracja jednostek LUN: migracja jednostek LUN to wyspecjalizowany program do przechowywania danych, który służy do przenoszenia jednostek LUN z jednego urządzenia pamięci masowej na drugie bez przerw, zachowując jednocześnie przejrzystość dla chmury konsumenci.

Bibliografia:

Koncepcje, technologia i architektura przetwarzania w chmurze

- Thomas Erl, Zaigham Mahmood i Ricardo Puttini – Prentice Hall – 2013
- Rozdział 11: Podstawowe architektury chmurowe
- Rozdział 12: Zaawansowane architektury chmurowe

Opanowanie podstaw przetwarzania w chmurze i programowania aplikacji

- Rajkumar Buyya, Christian Vecchiola, S. Thamarai Selvi – Elsevier – 2013

Przetwarzanie rozproszone i chmura obliczeniowa, od przetwarzania równoległego po Internet rzeczy

- Kai Hwang, Jack Dongarra, Geoffrey Fox – Wydawnictwo MK – 2012

Jednostka 4

Rozdział 1

Podstawowe architektury chmurowe

Struktura jednostki:

4.1.1 Architektura dystrybucji obciążenia

4.2.1 Architektura łączenia zasobów

4.3.1 Architektura dynamicznej skalowalności 4.4.1

Architektura elastycznej wydajności zasobów 4.5.1

Architektura równoważenia obciążenia usług

4.6.1 Architektura rozrywająca chmury

4.7.1 Architektura udostępniania dysków elastycznych

4.8.1 Architektura nadmiarowej pamięci masowej

Cel:

Aby dowiedzieć się, jak korzystać z usług w chmurze.

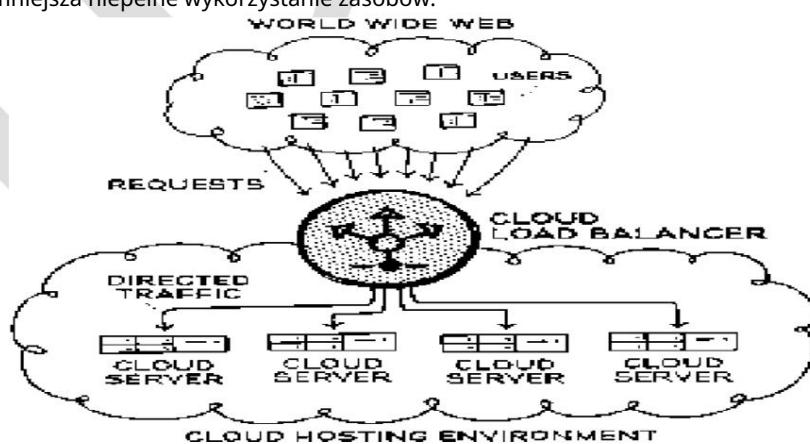
Wstęp:

W tym rozdziale przedstawiono i opisano kilka bardziej powszechnych, podstawowych modeli architektury chmur, z których każdy wyjaśnia typowe zastosowania i cechy współczesnych środowisk opartych na chmurze. Dalej

W rozdziale tym zbadano także zaangażowanie i znaczenie różnych kombinacji mechanizmów przetwarzania w chmurze w odniesieniu do tych architektur.

4.1. Architektura dystrybucji obciążenia

- Zasoby w chmurze można skalować poziomo przy użyciu dodatkowego lub identycznego zasobu i a moduł równoważenia obciążenia, który jest w stanie zapewnić dystrybucję obciążenia pomiędzy zasobami w czasie wykonywania.
- Taka architektura dystrybucji ma podwójną zaletę
 - I. Zmniejsza nadmierne wykorzystanie zasobów.
 - II. Zmniejsza niepełne wykorzystanie zasobów.



Rysunek: Architektura dystrybucji obciążenia

- Dystrybucja obciążenia odbywa się w celu obsługi rozproszonych serwerów wirtualnych, urządzeń pamięci masowej i usługi.
- System równoważenia obciążenia tworzy wyspecjalizowane odmiany, które uwzględniają aspekt obciążenia balansowanie jak:
 - I. Architektura instancji usług z równoważeniem obciążenia

II. Architektura instancji serwerów wirtualnych z równoważeniem obciążenia

iii. Architektura przełączników wirtualnych ze zrównoważonym obciążeniem

- Oprócz wyżej wymienionego mechanizmu częścią mogą być również następujące mechanizmy ta architektura chmurowa:

I. Monitor audytu – zasoby przetwarzające dane mogą określić, czy monitorowanie jest konieczne w celu spełnienia wymogów prawnych i regulacyjnych.

II. Cloud Usage Monitor – do wykonania środowiska wykonawczego można zaangażować różne monitory śledzenie obciążenia pracą i przetwarzanie danych.

iii. Hypervisor — obciążenie między hypervisorami a hostowanymi przez nie serwerami wirtualnymi może wymagać dystrybucji.

IV. Obwód sieci logicznej — obwód sieci logicznej izoluje konsumenta chmury granice sieci w odniesieniu do sposobu i miejsca dystrybucji obciążenia.

v. Klaster zasobów – Klastrowe zasoby IT w trybie aktywny/aktywny są powszechnie używane do obsługi równoważenia obciążenia pomiędzy różnymi węzłami klastra.

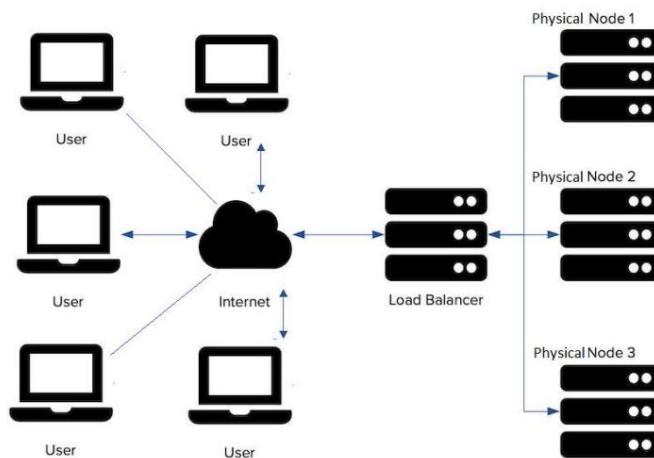
VI. Replikacja zasobów – ten mechanizm może generować nowe instancje zwirtualizowanych zasobów IT w odpowiedzi na wymagania dotyczące dystrybucji obciążenia w czasie wykonywania.

4.2. Architektura łączenia zasobów

- Architektura ta opiera się na wykorzystaniu jednego lub większej liczby zasobów z puli zasobów, w której identyczne zsynchronizowane zasoby są grupowane i obsługiwane przez system.

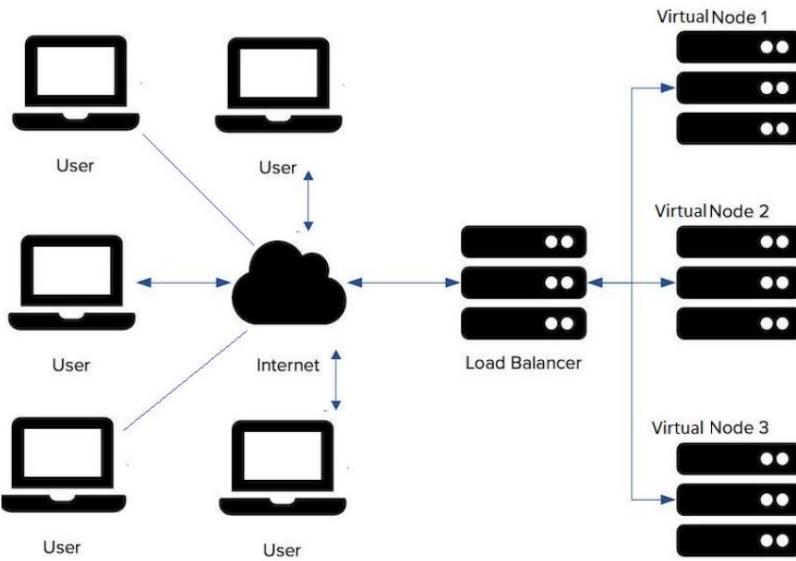
Przykłady pul zasobów: 1) Pule

serwerów fizycznych: to grupa serwerów fizycznych połączonych w sieć, na których zainstalowano systemy operacyjne oraz inne niezbędne programy i/lub aplikacje i które są gotowe do natychmiastowego użycia.

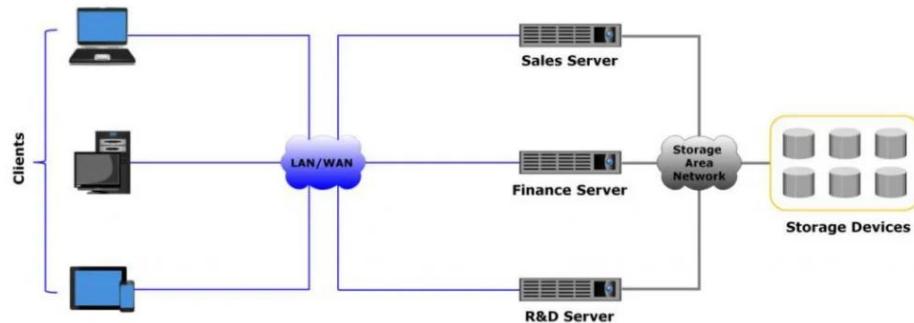


2) Pule serwerów wirtualnych: to grupa serwerów wirtualnych połączonych w sieć, które mają zostać zainstalowane systemów operacyjnych oraz innych niezbędnych programów i/lub aplikacji i są gotowe do natychmiastowego użycia. Zwykle konfiguruje się je przy użyciu jednego z kilku dostępnych szablonów wybranych przez konsumenta chmury podczas udostępniania.

- Na przykład użytkownik chmury może skonfigurować pulę systemów Windows średniej warstwy serwery z 4 GB pamięci RAM lub pula serwerów niższego poziomu Ubuntu z 2 GB pamięci RAM.

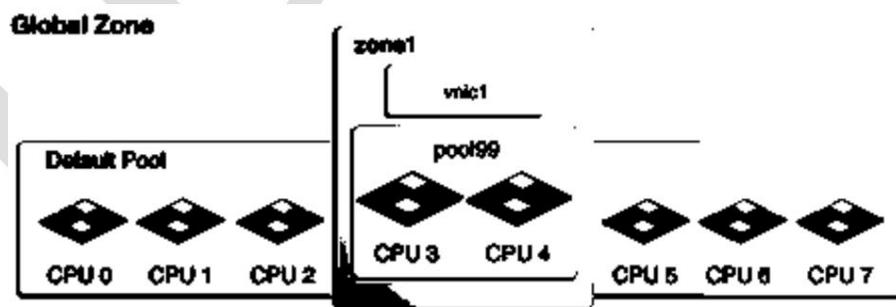


3) Pule pamięci masowej lub pule urządzeń pamięci masowej w chmurze: to grupa struktur pamięci masowej opartych na plikach lub blokach, które zawierają puste i/lub wypełnione urządzenia pamięci masowej w chmurze.

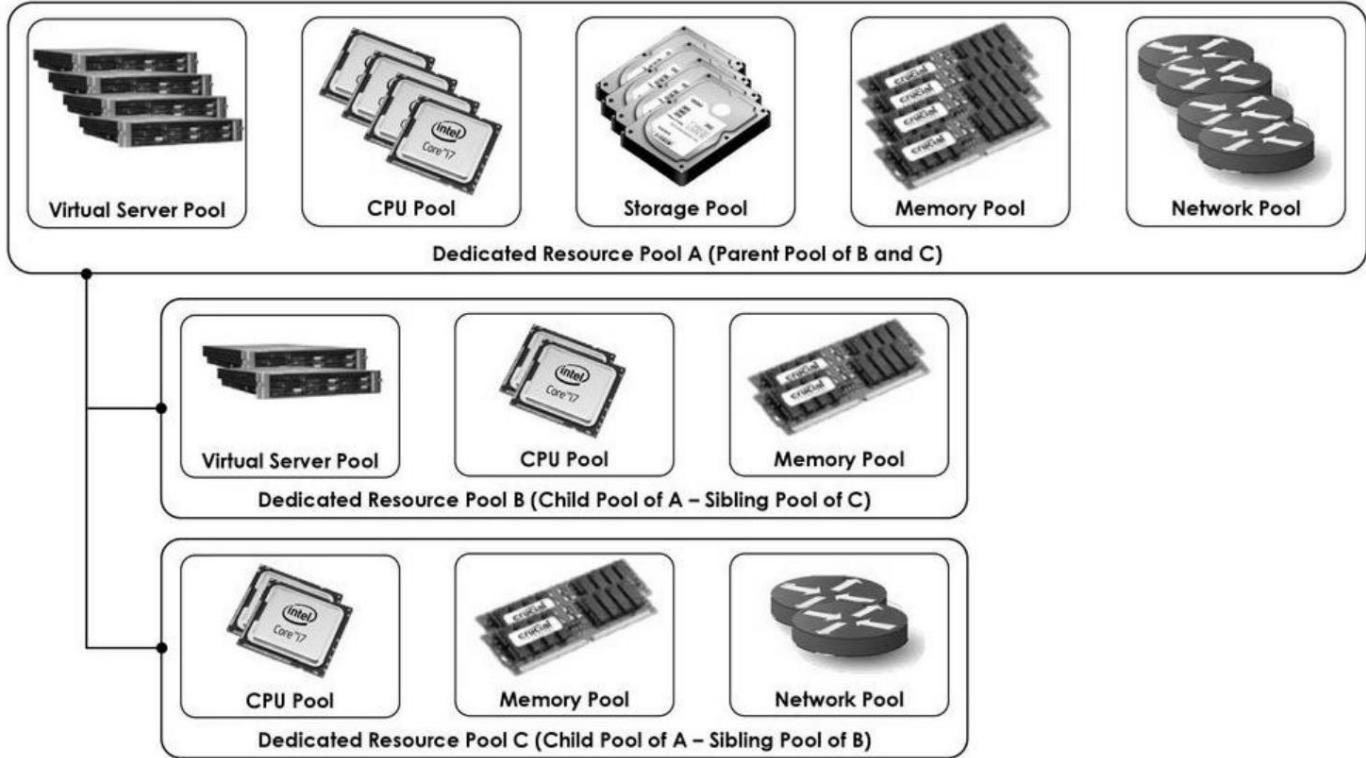


4) Pule sieciowe (lub pule połączeń wzajemnych): to grupa różnych, wstępnie skonfigurowanych urządzeń łączności sieciowej. • Można na przykład utworzyć pulę wirtualnych urządzeń firewall lub fizycznych przełączników sieciowych w celu zapewnienia nadmiarowej łączności, równoważenia obciążenia itp.

5) Pule procesorów: to grupa jednostek przetwarzających gotowa do przydzielenia serwerom wirtualnym i zazwyczaj podzielona na poszczególne rdzenie przetwarzające.



- Pule fizycznej pamięci RAM mogą być wykorzystywane w nowo udostępnianych serweraх fizycznych lub w pionie skalować serwery fizyczne.
- Dla każdego typu zasobu można tworzyć dedykowane pule, a poszczególne pule można grupować do większej puli, w takim przypadku każda pojedyncza pula stanie się podpułą.
- Pule zasobów mogą stać się bardzo złożone i tworzyć wiele pul dla konkretnych konsumentów lub aplikacji w chmurze. • Można ustanowić strukturę hierarchiczną, tworząc pule nadzędne, siostrzane i zagnieżdżone, aby ułatwić organizację różnorodnych wymagań dotyczących łączenia zasobów, jak pokazano na rysunku poniżej.



Rysunek: Inna architektura basenu

Architektura dynamicznej skalowalności to model architektoniczny oparty na systemie predefiniowanych warunków skalowania, które wyzwalają dynamiczną alokację zasobów IT z pul zasobów. Dynamiczna alokacja umożliwia zmienne wykorzystanie w zależności od wahań zapotrzebowania, ponieważ niepotrzebne zasoby IT są skutecznie odzyskiwane bez konieczności ręcznej interakcji.

W odbiorniku automatycznego skalowania skonfigurowano progi obciążenia, które decydują o tym, kiedy do przetwarzania obciążenia należy dodać nowe zasoby IT. Mechanizm ten może być wyposażony w logikę określającą, ile dodatkowych zasobów IT można dynamicznie udostępniać, w oparciu o warunki umowy o świadczenie usług danego konsumenta chmury.

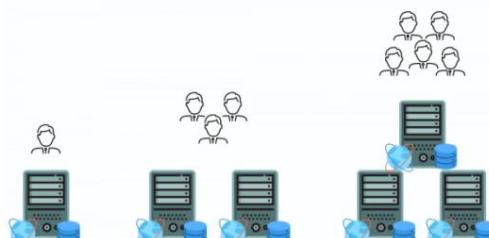
4.3. • Architektura dynamicznej skalowalności

Architektura ta jest modelem opartym na systemie predefiniowanych warunków skalowania puli zasobów, które wyzwalają dynamiczną alokację zasobów chmury z puli.

- Dynamiczna alokacja umożliwia zmienne wykorzystanie określone przez wahania zapotrzebowania na użytkowanie, co skutkuje efektywnym wykorzystaniem zasobów, a niepotrzebne zasoby są skutecznie odzyskiwane bez konieczności ręcznej interakcji.
- Istnieją trzy typy skalowania dynamicznego:

1. Dynamiczne skalowanie poziome – w tym typie instancje zasobów są skalowane w górę i w dół, aby obsłużyć dynamiczne obciążenia podczas wykonywania. Automatyczny odbiornik skalowania monitoruje żądania i sygnalizuje replikację zasobów w celu zainicjowania duplikacji zasobów zgodnie z wymaganiami i uprawnieniami ustalonymi przez administratora.

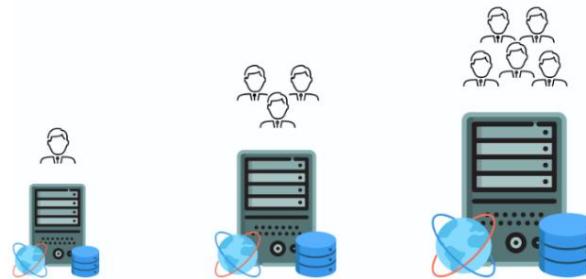
Horizontal Scaling



2. Dynamiczne skalowanie pionowe – w tym typie instancje zasobów są skalowane w górę i w dół gdy zachodzi potrzeba dostosowania mocy przerobowej pojedynczego zasobu. Na przykład:

Serwer wirtualny, który jest przeciążony, może mieć dynamicznie zwiększającą pamięć lub może zostać dodany rdzeń przetwarzający.

Vertical Scaling



3. Relokacja dynamiczna – w tym typie zasób jest relokowany do hosta o większej pojemności.

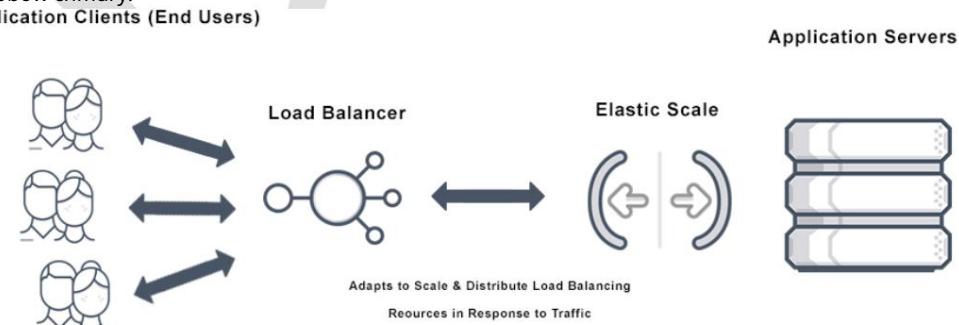
Na przykład może zaistnieć potrzeba przeniesienia pliku z taśmowego urządzenia pamięci masowej SAN o przepustowości we/wy 4 GB na sekundę na inne dyskowe urządzenie pamięci masowej SAN o przepustowości we/wy 8 GB na sekundę.

Architekturę dynamicznej skalowalności można zastosować do szeregu zasobów IT, w tym serwery wirtualne i urządzenia pamięci masowej w chmurze. Oprócz podstawowego odbiornika automatycznego skalowania i mechanizmów replikacji zasobów, dostępne są następujące mechanizmy można również zastosować w tej formie architektury chmury:

- Cloud Usage Monitor – wyspecjalizowane monitory wykorzystania chmury mogą śledzić wykorzystanie czasu działania w odpowiedzi na spowodowane tym dynamiczne wahania architektura.
- Hypervisor – Hypervisor jest wywoływany przez system dynamicznej skalowalności w celu tworzenia lub usuwania instancji serwerów wirtualnych lub w celu samodzielnego skalowania.
- Monitor płatności za użycie – monitor płatności za użycie jest zaangażowany do zbierania informacji o kosztach użytkowania w odpowiedzi na skalowanie zasobów IT.

4.4. • Architektura elastycznej pojemności zasobów

Architektura ta jest związana z dynamicznym udostępnianiem serwerów wirtualnych przy użyciu systemu, który przydziela i odzyskuje procesory i pamięć w natychmiastowej reakcji na zmieniające się wymagania dotyczące przetwarzania hostowanych zasobów chmury.



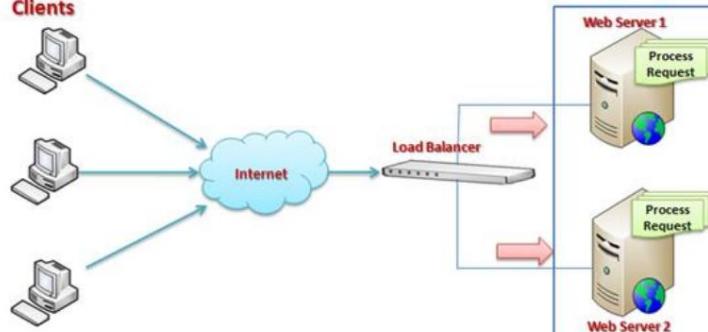
- Pule zasobów są wykorzystywane przez technologię skalowania, która współdziała z hiperwizorem i/lub VIM-em pobierać i zwracać zasoby procesora i pamięci RAM w czasie wykonywania.
- Przetwarzanie serwera wirtualnego w czasie wykonywania jest monitorowane, dzięki czemu można wykorzystać dodatkową moc obliczeniową z puli zasobów w drodze dynamicznej alokacji, zanim zostaną osiągnięte progi wydajności.
- Serwer wirtualny oraz hostowane na nim aplikacje i zasoby są w odpowiedzi skalowane w pionie. • Ten typ architektury chmurowej można zaprojektować tak, aby działał inteligentny skrypt automatyzacji wysyła żądanie skalowania poprzez VIM zamiast bezpośrednio do hypervisor-a.
- Serwery wirtualne uczestniczące w systemach elastycznej alokacji zasobów mogą wymagać ponownego uruchomienia aby dynamiczna alokacja zasobów zaczęła obowiązywać.

4.5. Architektura równoważenia obciążenia usług

Przetwarzanie w chmurze: wersja nieedytowana

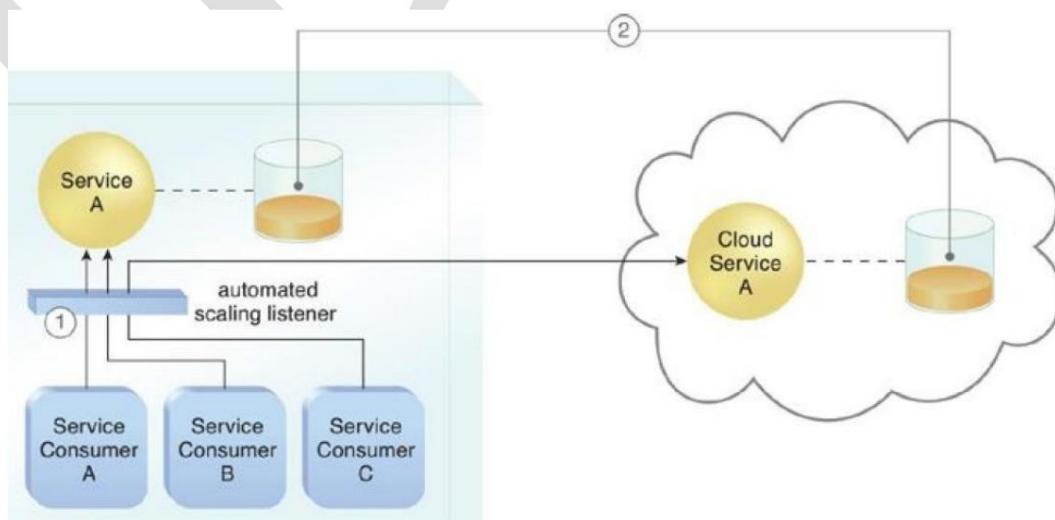
- Ta architektura jest wyspecjalizowanym wariantem architektury dystrybucji obciążenia, która jest dostosowana specjalnie do skalowania wdrożeń usług w chmurze.
- Tworzone są redundantne wdrożenia usług w chmurze, do których dodawany jest system równoważenia obciążenia dynamicznie dystrybuuj obciążenia.
- Zduplikowane implementacje usług w chmurze są zorganizowane w pulę zasobów, podczas gdy moduł równoważenia obciążenia jest umieszczony jako komponent zewnętrzny lub wbudowany, aby umożliwić serwerom hosta samodzielne równoważenie obciążzeń.
- W zależności od przewidywanego obciążenia i mocy obliczeniowej serwera hosta

W różnych środowiskach można wygenerować wiele instancji każdej implementacji usługi w chmurze w ramach puli zasobów, która skuteczniej reaguje na wahania liczby żądań.



4.6. Architektura rozrywająca chmury

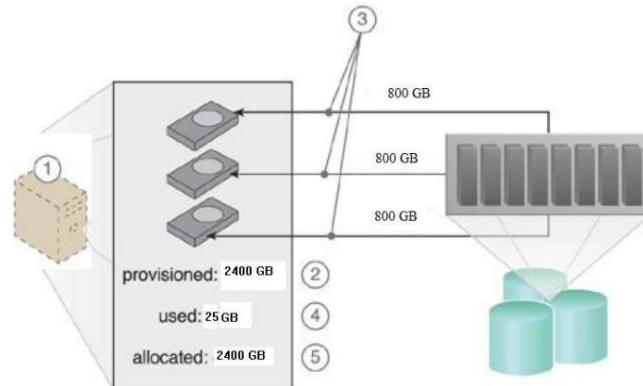
- Ta architektura ustanawia formę dynamicznego skalowania, która skaluje lub „wyrzuca” lokalne zasoby chmury do chmury po osiągnięciu wcześniej zdefiniowanych progów wydajności.
- Odpowiednie zasoby oparte na chmurze są wstępnie wdrożone nadmiarowo, ale pozostają nieaktywne aż do pojawienia się chmur. Gdy nie są już potrzebne, zasoby są zwalniane, a architektura „wkracza” z powrotem do środowiska lokalnego.
- Cloud Bursting to elastyczna, skalowalna architektura, która zapewnia konsumentom chmury możliwość korzystania z zasobów IT opartych na chmurze wyłącznie w celu spełnienia wyższych wymagań w zakresie użytkowania. • Podstawą tego modelu architektonicznego jest automatyczny detektor skalowania mechanizmy replikacji zasobów.
- Zautomatyzowany odbiornik skalowania określa, kiedy przekierować żądania do zasobów w chmurze, a replikacja zasobów służy do utrzymania synchronizacji między zasobami IT lokalnymi i chmurowymi w odniesieniu do informacji o stanie



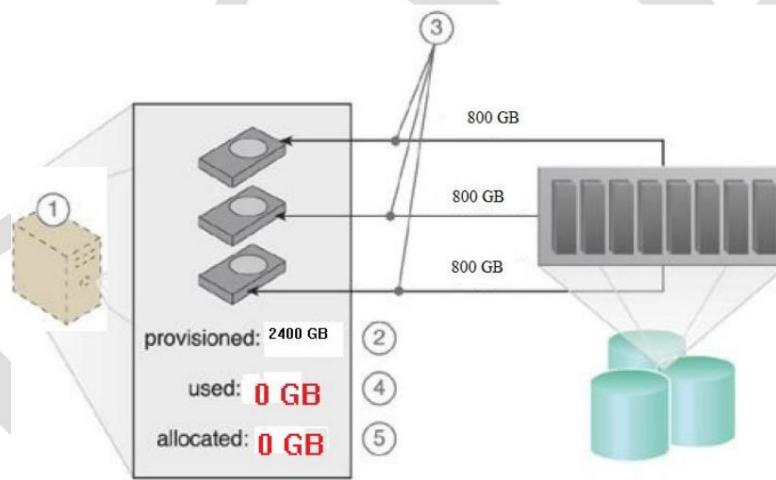
4.7. • Architektura elastycznego udostępniania dysków

Konsumenci usług chmurowych są powszechnie obciążani opłatami za przestrzeń dyskową w chmurze na dysku stałym alokację przestrzeni dyskowej, co oznacza, że opłaty są ustalane z góry na podstawie pojemności dysku i nie odpowiadają faktycznemu zużyciu pamięci masowej.

- Chmura udostępnia serwer wirtualny z systemem operacyjnym Windows Server 2019 i trzema dyskami twardymi o pojemności 800 GB. Konsument chmury jest obciążany opłatą za późniejsze wykorzystanie 2400 GB przestrzeni dyskowej instalowanie systemu operacyjnego, mimo że system operacyjny wymaga tylko 25 GB przestrzeni dyskowej.



- Architektura elastycznego udostępniania dysków ustanawia dynamiczne udostępnianie pamięci system, który gwarantuje, że konsument chmury będzie szczegółowo rozliczany za dokładną ilość faktycznie wykorzystywanej przestrzeni dyskowej. System ten wykorzystuje technologię cienkiego udostępniania do dynamicznej alokacji przestrzeni dyskowej i jest dodatkowo wspierany przez monitorowanie wykorzystania w czasie wykonywania w celu gromadzenia dokładnych danych o użytkowaniu do celów rozliczeniowych

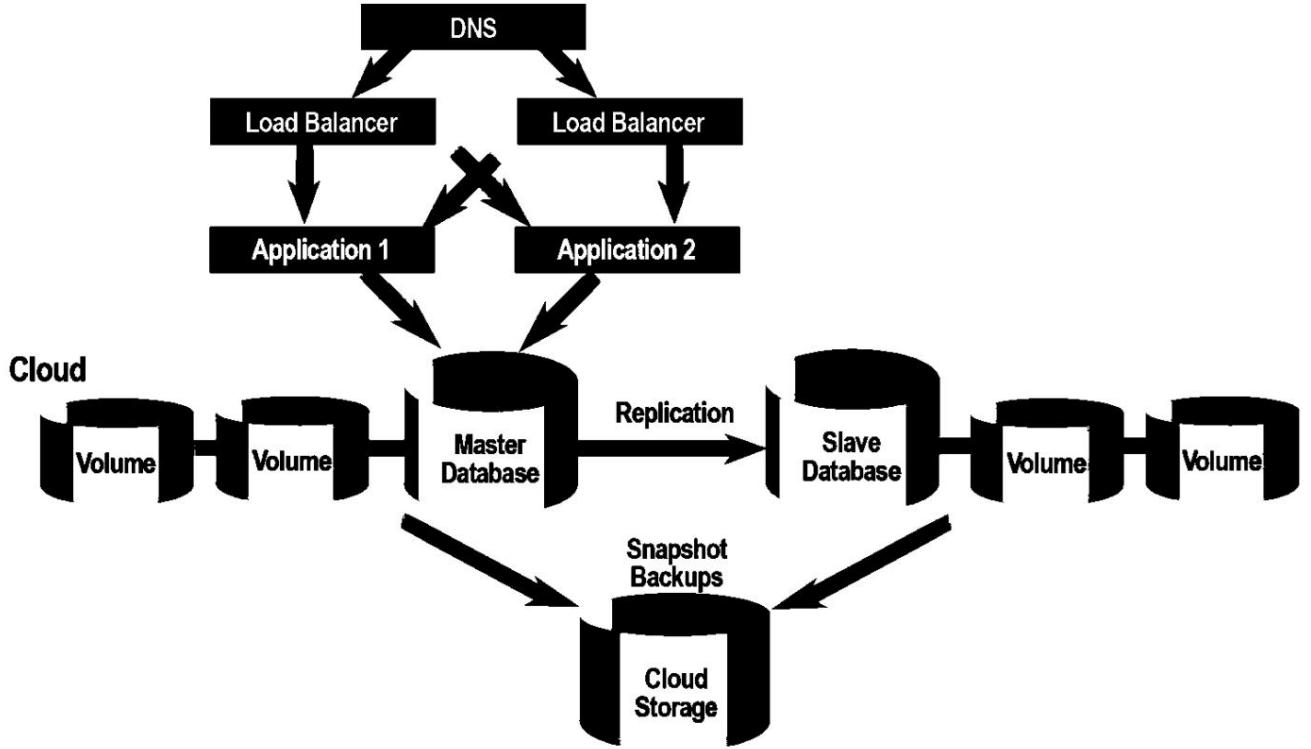


- Oprogramowanie do alokacji elastycznej jest instalowane na serwerach wirtualnych, które przetwarzają dynamiczną alokację pamięci za pośrednictwem hiperwizora, podczas gdy monitor płatności za użycie śledzi i raportuje szczegółowe dane dotyczące wykorzystania dysku związane z rozliczeniami.

4.8. Nadmiarowa architektura pamięci masowej

- Urządzenia do przechowywania w chmurze czasami ulegają awariom i zakłóceniom spowodowanym problemami z łącznością sieciową, awarią kontrolera lub ogólnym sprzętem albo naruszeniami bezpieczeństwa. • Naruszona niezawodność urządzenia pamięci masowej w chmurze może wywołać efekt falowy i spowodować awarię wszystkich usług, aplikacji i komponentów infrastruktury w chmurze zależnych od jego dostępności.

- W architekturze nadmiarowej pamięci masowej dodatkowe zduplikowane urządzenie pamięci masowej w chmurze stanowi część systemu awaryjnego, który synchronizuje dane z danymi znajdującymi się na głównym urządzeniu pamięci masowej w chmurze. • Bramka usług pamięci masowej przekierowuje żądania konsumentów chmury do urządzenia dodatkowego w przypadku awarii urządzenia podstawowego.



- Ta architektura chmury opiera się głównie na systemie replikacji pamięci masowej, który utrzymuje główne urządzenie pamięci masowej w chmurze zsynchronizowane ze zduplikowanymi dodatkowymi urządzeniami pamięci masowej w chmurze.
- Dostawcy usług w chmurze mogą lokalizować dodatkowe urządzenia do przechowywania w chmurze w innym regionie geograficznym niż główne urządzenie do przechowywania w chmurze, zwykle ze względów ekonomicznych. • Lokalizacja dodatkowych urządzeń pamięci masowej w chmurze może narzucać protokół i metodę używaną do synchronizacji, ponieważ niektóre protokoły transportu replikacji mają ograniczenia dotyczące odległości.

Rozdział 2

Zaawansowane architektury chmurowe

Struktura jednostki:

- 4.2.1. Architektura klastrowania hiperwizora
- 4.2.2. Architektura instancji serwerów wirtualnych z równoważeniem obciążenia
- 4.2.3. Architektura przenoszenia usług bez zakłóceń
- 4.2.4. Architektura zero przestojów
- 4.2.5. Architektura równoważenia chmur
- 4.2.6. Architektura rezerwacji zasobów 4.2.7.
- Architektura dynamicznego wykrywania i odzyskiwania awarii 4.2.8.
- Architektura udostępniania usług bare-metal
- 4.2.9. Architektura szybkiego udostępniania
- 4.2.10. Architektura zarządzania obciążeniem pamięci masowej

Cel:

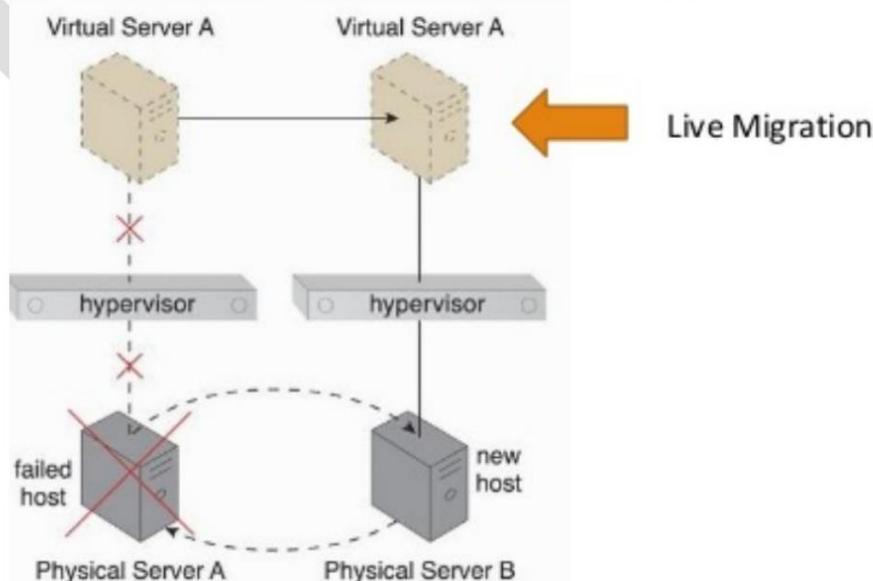
Aby dowiedzieć się, jak korzystać z usług Advance Cloud Services.

Wstęp:

W tym rozdziale przedstawiono odrębne i wyrafinowane warstwy architektury technologii chmury, z których kilka można zbudować w oparciu o bardziej podstawowe środowiska ustanowione przez modele architektoniczne omówione w poprzednim rozdziale.

4.2.1 Architektura klastrowa hiperwizora

- Hiperwizorzy są odpowiedzialni za tworzenie i hostowanie wielu serwerów wirtualnych.
- Z powodu tej zależności wszelkie awarie wpływające na hiperwizora mogą mieć kaskadowy wpływ na jego serwery wirtualne.
- Architektura klastrowa hypervisorów tworzy klaszter hypervisorów o wysokiej dostępności na wielu serwerach fizycznych.
- Jeśli dany hypervisor lub odpowiadający mu serwer fizyczny stanie się niedostępny, hostowane serwery wirtualne można przenieść na inny serwer fizyczny lub hiperwizor, aby zachować czas działania operacje.

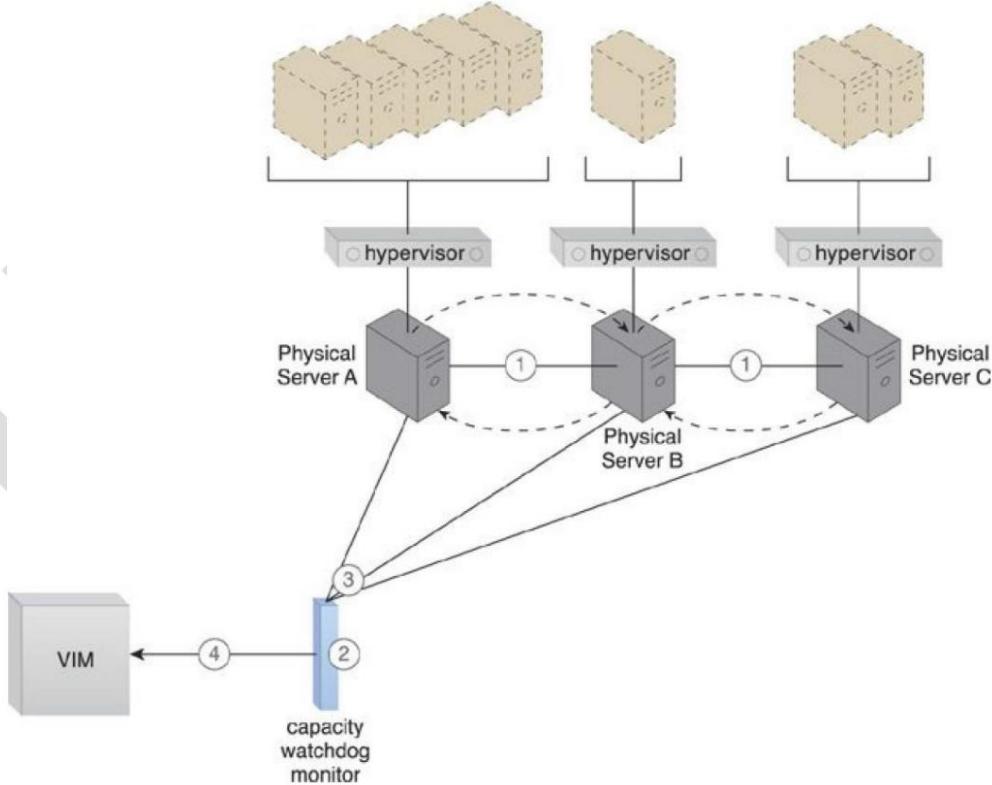


- Klaszter hypervisorów jest kontrolowany przez centralny moduł VIM, który regularnie wysyła komunikaty o pulsie do hypervisorów, aby potwierdzić, że działają.

- Niepotwierdzone komunikaty pulsu powodują, że VIM inicjuje program migracji maszyny wirtualnej na żywo, aby dynamicznie przenieść dotknięte serwery wirtualne do nowego hosta.

4.2.2 Architektura instancji serwerów wirtualnych ze zrównoważonym obciążeniem

- Czasami utrzymywanie równomiernego obciążenia między serwerami fizycznymi, których działanie i zarządzanie są odizolowane, co może być najtrudniejszym elementem w chmurze.
- Serwer fizyczny może z łatwością obsługiwać więcej serwerów wirtualnych lub przyjmować większe obciążenia niż sąsiednie serwery fizyczne.
- Zarówno nadmierne, jak i niedostateczne wykorzystanie serwera fizycznego może z czasem znacznie wzrosnąć, co prowadzi do: ciągłe wyzwania związane z wydajnością (w przypadku nadmiernie wykorzystywanych serwerów) i ciągłe marnotrawstwo (w przypadku utraconego potencjału przetwarzania w przypadku niedostatecznie wykorzystanych serwerów).
- Architektura instancji serwerów wirtualnych z równowagą obciążenia ustanawia system nadzoru wydajności, który dynamicznie oblicza instancje serwerów wirtualnych i powiązane z nimi obciążenia, przed rozdzieleniem przetwarzania pomiędzy dostępne hosty serwerów fizycznych.
- System nadzorujący pojemność składa się z monitora wykorzystania chmury, monitorującego pojemność program migracji maszyn wirtualnych na żywo i narzędzie do planowania wydajności.
- Monitor wydajności monitoruje wykorzystanie serwera fizycznego i wirtualnego oraz raportuje wszelkie znaczące wahania planiste wydajności, który jest odpowiedzialny za dynamiczne obliczanie wydajności obliczeniowej serwera fizycznego w stosunku do wymagań dotyczących wydajności serwera wirtualnego.
- Jeśli planista wydajności zdecyduje się przenieść serwer wirtualny na inny host w celu rozłożenia obciążenia, program migracji maszyny wirtualnej na żywo otrzymuje sygnał o konieczności przeniesienia serwera wirtualnego.



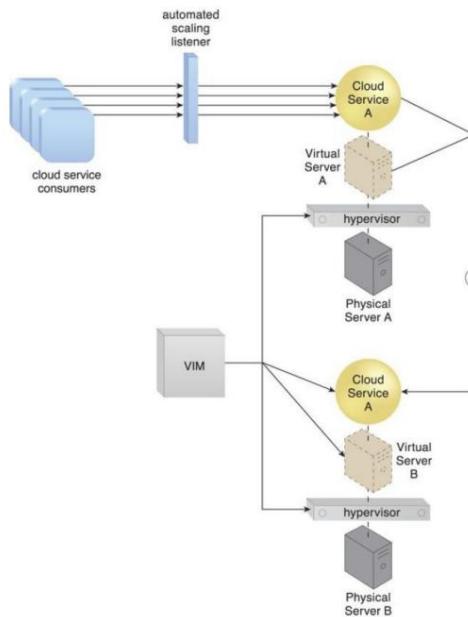
4.2.3 • Architektura przenoszenia usług bez zakłóceń

Usługa w chmurze może stać się niedostępna z wielu powodów, takich jak:

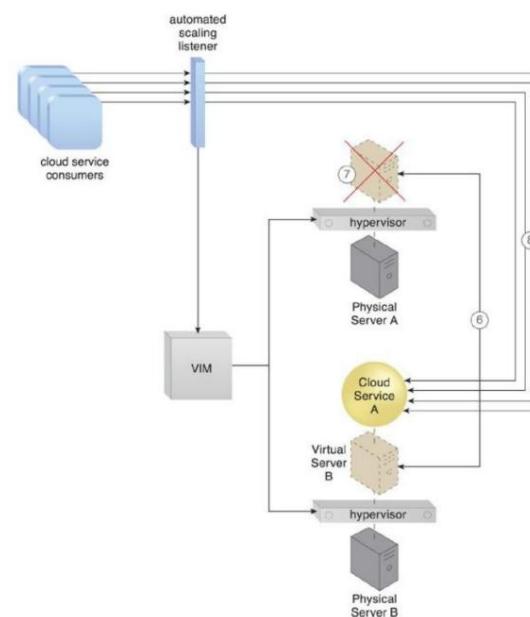
- 1) wymagania dotyczące wykorzystania środowiska wykonawczego, które przekraczają jego możliwości przetwarzania
- 2) aktualizacja konserwacyjna nakazująca tymczasową przerwę w działaniu
- 3) trwała migracja na nowy host serwera fizycznego

- Żądania konsumentów dotyczące usług w chmurze są zwykle odrzucane, jeśli usługa w chmurze staje się niedostępna, co może potencjalnie skutkować powstaniem wyjątkowych warunków.

- Niezakłócająca pracy architektura relokacji usług ustanawia system, w którym wcześniej zdefiniowane zdarzenie inicjuje duplikację lub migrację wdrożenia usługi w chmurze w czasie jej wykonywania, unikając w ten sposób jakichkolwiek zakłóceń.
- Zamiast skalować lub zmniejszać usługi w chmurze za pomocą nadmiarowych implementacji, działalność usług w chmurze można tymczasowo przenieść do innego środowiska hostingowego w czasie wykonywania, dodając zduplikowaną implementację na nowym hoście.
- Podobnie żądania konsumentów dotyczące usług w chmurze mogą być tymczasowo przekierowywane do duplikatów wdrożenia, gdy pierwotna implementacja musi przejść przerwę konserwacyjną. • Przeniesienie wdrożenia usługi w chmurze i wszelkiej działalności związanej z usługą w chmurze może również mieć charakter trwały, aby uwzględnić migrację usług w chmurze do nowych hostów serwerów fizycznych.
- Kluczowym aspektem podstawowej architektury jest wdrożenie nowej usługi w chmurze gwarantujemy, że pomyślnie będziemy odbierać żądania konsumentów dotyczące usług w chmurze i odpowiadając na nie, zanim pierwotna implementacja usługi w chmurze zostanie dezaktywowana lub usunięta.
- Powszechnym podejściem jest migracja maszyny wirtualnej na żywo w celu przeniesienia całej instancji serwera wirtualnego hosting usługi w chmurze.
- Mechanizmy automatycznego nasłuchiwanego skalowania i/lub mechanizmu równoważenia obciążenia można wykorzystać do tymczasowego przekierowania żądań konsumentów usług w chmurze w odpowiedzi na wymagania dotyczące skalowania i rozkładu obciążenia. Każdy mechanizm może skontaktować się z VIMem w celu zainicjowania procesu migracji maszyny wirtualnej na żywo.



Rys. : Przed awarią • Migracja



Rys. : Po awarii

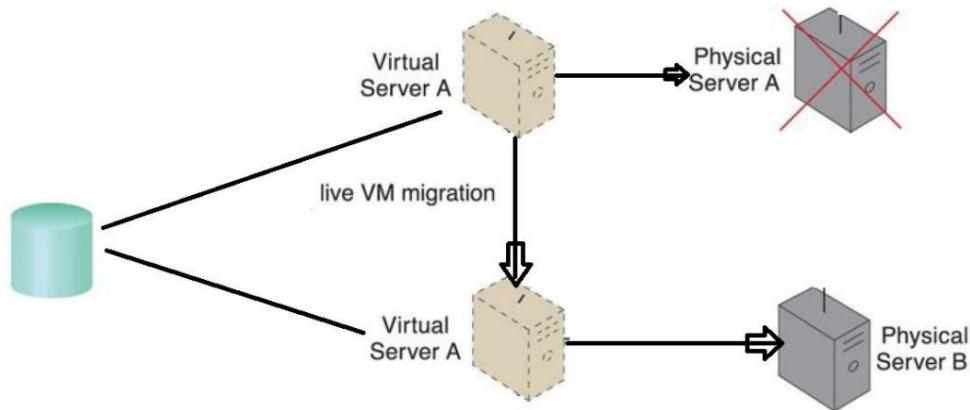
serwera wirtualnego może nastąpić na jeden z dwóch poniższych sposobów, w zależności od lokalizacji dysków serwera wirtualnego i jego konfiguracji:

- Kopia dysków serwera wirtualnego jest tworzona na hoście docelowym, jeśli dyski serwera wirtualnego są przechowywane na lokalnym urządzeniu pamięci masowej lub na niewspółdzielonym zdalnym urządzeniu magazynującym podłączonym do hosta źródłowego. Po utworzeniu kopii obie instancje serwerów wirtualnych są synchronizowane, a pliki serwera wirtualnego są usuwane z hosta źródłowego.
- Kopiowanie dysków serwera wirtualnego nie jest konieczne, jeśli pliki serwera wirtualnego są przechowywane na zdalnym urządzeniu pamięci masowej, które jest współdzielone między hostami źródłowymi i docelowymi. Własność serwera wirtualnego jest po prostu przenoszona ze źródła na docelowy host serwera fizycznego, a stan serwera wirtualnego jest automatycznie synchronizowany.

4.2.4 Architektura z zerowym przestojem

- Serwer fizyczny w naturalny sposób działa jako pojedynczy punkt awarii dla hostowanych na nim serwerów wirtualnych. W rezultacie awaria serwera fizycznego lub jego naruszenie może mieć wpływ na dostępność którychkolwiek (lub wszystkich) hostowanych serwerów wirtualnych. To sprawia, że udzielenie przez dostawcę usług w chmurze gwarancji zerowych przestojów na rzecz konsumentów usług w chmurze stanowi wyzwanie.

- Architektura zapewniająca zerowy czas przestojów ustanawia wyrafinowany system przełączania awaryjnego, który umożliwia dynamiczne przenoszenie serwerów wirtualnych do różnych hostów serwerów fizycznych w przypadku awarii ich pierwotnego hosta serwera fizycznego



4.2.5 Architektura równoważenia chmury

Architektura ta ustanawia wyspecjalizowany model architektoniczny, w którym zasoby chmury mogą być równoważone w wielu chmurach. Równoważenie między chmurami żądań konsumentów usług w chmurze może pomóc:

- 1) poprawić wydajność i skalowalność zasobów
- 2) zwiększyć dostępność i niezawodność zasobów
- 3) poprawić równoważenie obciążenia i optymalizację zasobów

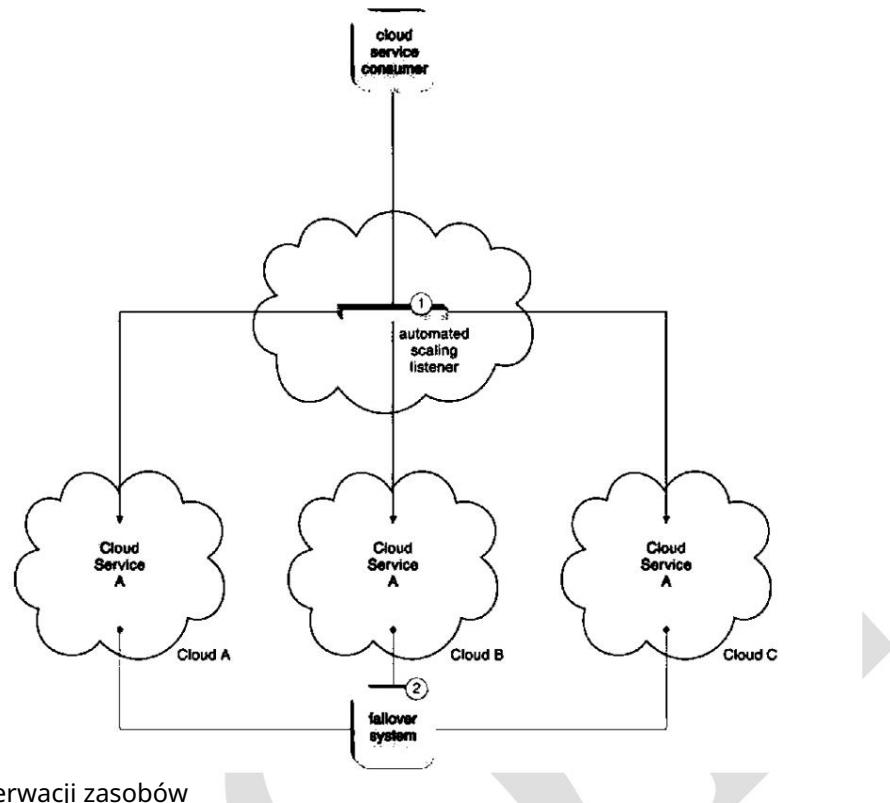
Jego funkcjonalność opiera się przede wszystkim na połączeniu automatycznego słuchacza skalowania i mechanizmów systemu przełączania awaryjnego. Częścią tego kompletnego rozwiązania może być znacznie więcej komponentów i mechanizmów architektury.

Na początek oba mechanizmy wykorzystuje się w następujący sposób:

- o Automatyczny odbiornik skalowania przekierowuje żądania konsumentów usług w chmurze do jednej z kilku implementacji nadmiarowych zasobów IT w oparciu o bieżące wymagania dotyczące skalowania i wydajności.
- o System przełączania awaryjnego zapewnia możliwość przełączania awaryjnego nadmiarowych zasobów IT między chmurami w przypadku awarii zasobu IT lub jego podstawowego środowiska hostingowego. Awarie zasobów IT są ogłoszane, dzięki czemu odbiornik automatycznego skalowania może uniknąć niezamierzonych kierowania żądań konsumentów usług w chmurze do niedostępnych lub niestabilnych zasobów IT.

Aby architektura równoważąca chmurę działała efektywnie, słuchacz zautomatyzowanego skalowania musi być świadomy wszystkich implementacji nadmiarowych zasobów IT w ramach architektury zrównoważonej w chmurze.

Ponadto, jeśli ręczna synchronizacja wdrożeń zasobów IT między chmurami nie jest możliwa, może być konieczne włączenie mechanizmu replikacji zasobów w celu zautomatyzowania synchronizacji.



4.2.6 Architektura rezerwacji zasobów

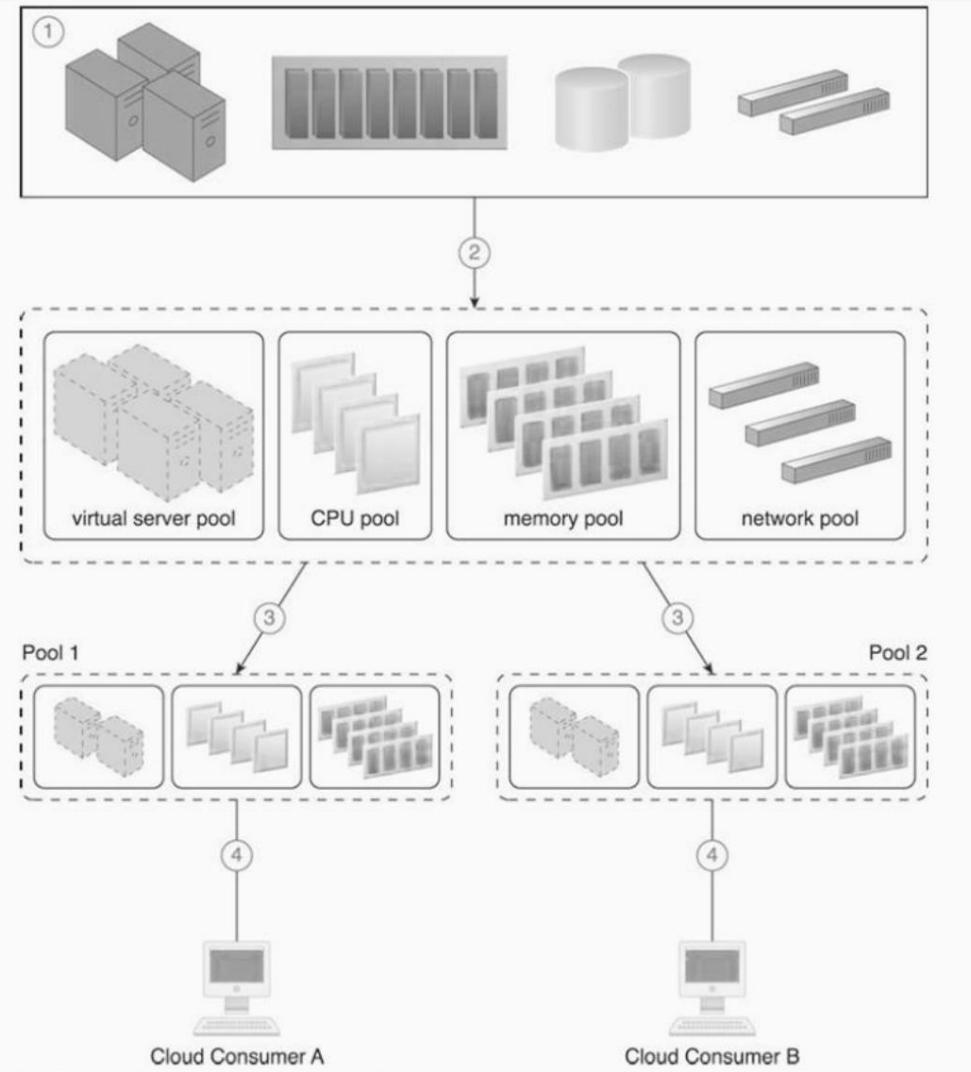
Architektura rezerwacji zasobów ustanawia system, w którym jedno z poniższych jest zarezerwowane wyłącznie dla danego konsumenta chmury

- pojedynczy zasób
- część zasobu
- wiele zasobów

Stworzenie systemu rezerwacji zasobów może wymagać zaangażowania mechanizmu systemu zarządzania zasobami, który służy do określenia progów wykorzystania poszczególnych zasobów i pul zasobów. Rezerwacje blokują ilość zasobów, które każda pula musi przechowywać, a pozostała część zasobów puli jest nadal dostępna do udostępniania i pożyczania. Mechanizm systemu zdalnego administrowania umożliwia także dostosowywanie frontonu, dzięki czemu klienci korzystający z chmury mają kontrolę administracyjną w zakresie zarządzania zarezerwowanymi alokacjami zasobów.

Typami mechanizmów powszechnie zarezerwowanych w tej architekturze są urządzenia pamięci masowej w chmurze i serwery wirtualne. Inne mechanizmy, które mogą być częścią architektury, mogą obejmować:

- o Monitor audytu o
- Monitor wykorzystania chmury o
- Hypervisor o Obwód sieci logicznej
- o Replikacja zasobów



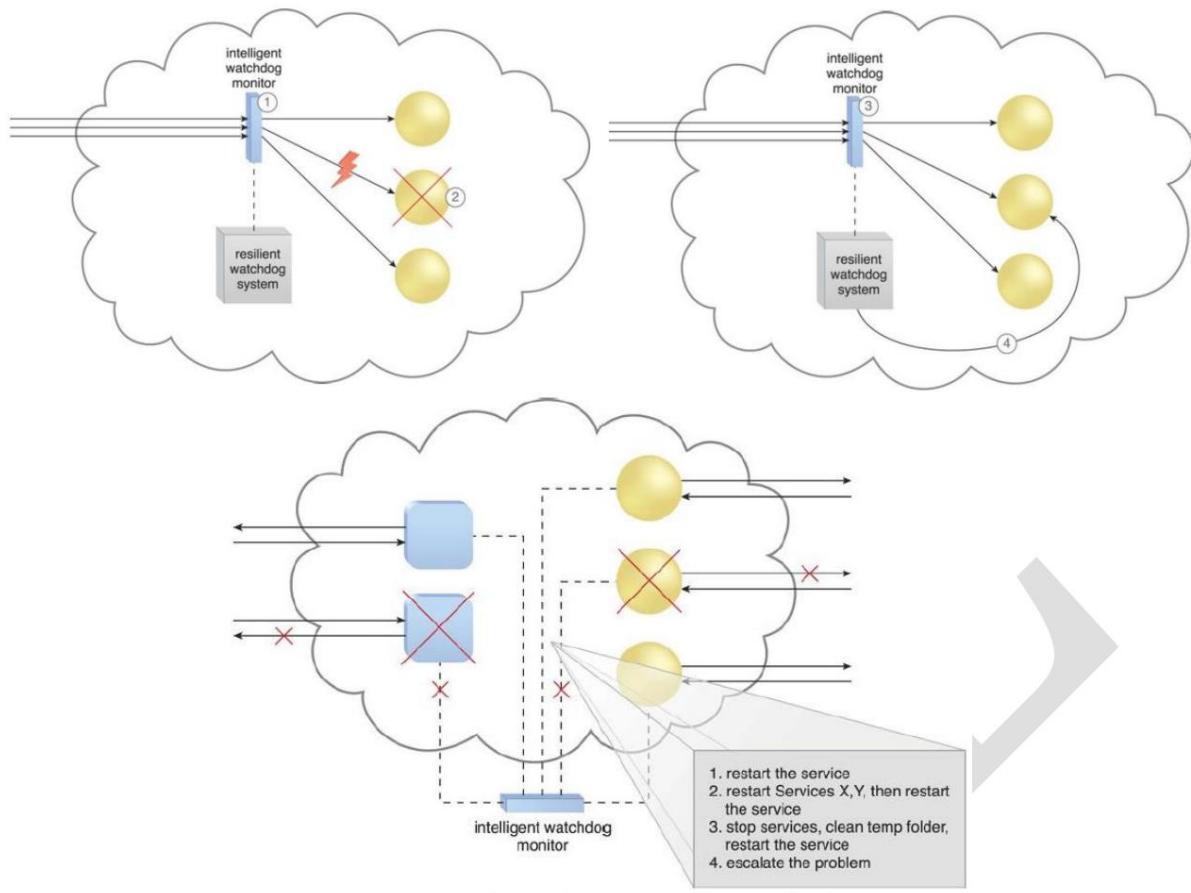
4.2.7 Architektura dynamicznego wykrywania i odzyskiwania awarii

Architektura ta ustanawia odporny system nadzorujący do monitorowania i reagowania na szeroki zakres wstępnie zdefiniowanych scenariuszy awarii. System ten powiadamia i eskaluje warunki awarii, których nie może sam automatycznie rozwiązać. Opiera się na wyspecjalizowanym monitorze wykorzystania chmury zwany inteligentnym monitorem monitorującym, który aktywnie śledzi zasoby i podejmuje wcześniej zdefiniowane działania w odpowiedzi na wcześniej zdefiniowane wydarzenia.

Odporny system nadzorujący spełnia pięć następujących podstawowych funkcji:

- 1) oglądanie
- 2) podjęcie decyzji o wydarzeniu
- 3) działanie na podstawie zdarzenia
- 4) raportowanie
- 5) eskalacja

Dla każdego zasobu można zdefiniować zasady odtwarzania sekwencyjnego, aby określić kroki, jakie musi podjąć inteligentny monitor monitorujący w przypadku wystąpienia awarii. Na przykład zasady odzyskiwania mogą określać, że przed wysłaniem powiadomienia należy automatycznie przeprowadzić jedną próbę odzyskania danych

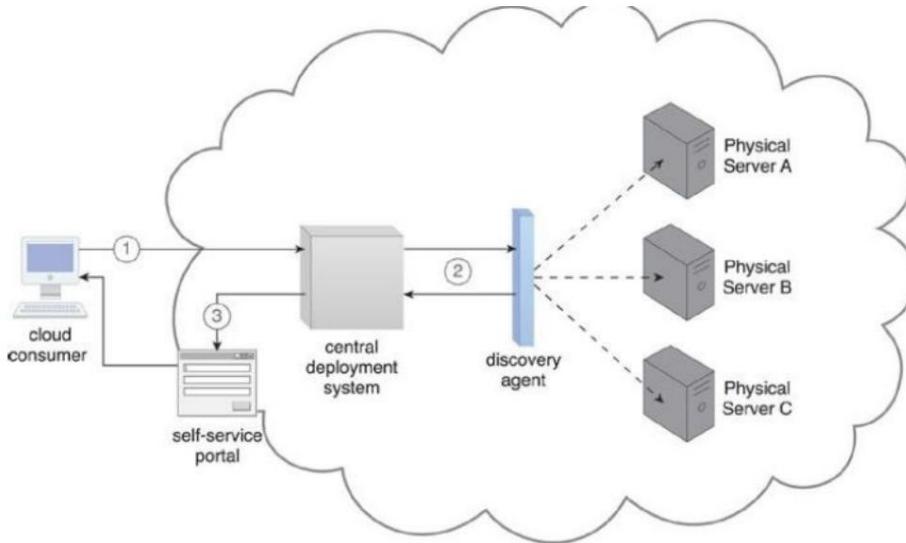


4.2.8 Architektura udostępniania typu bare-metal

Architektura ta ustanawia system wykorzystujący tę funkcję wraz z wyspecjalizowanymi agentami usług, które służą do zdalnego wykrywania i skutecznego udostępniania całych systemów operacyjnych.

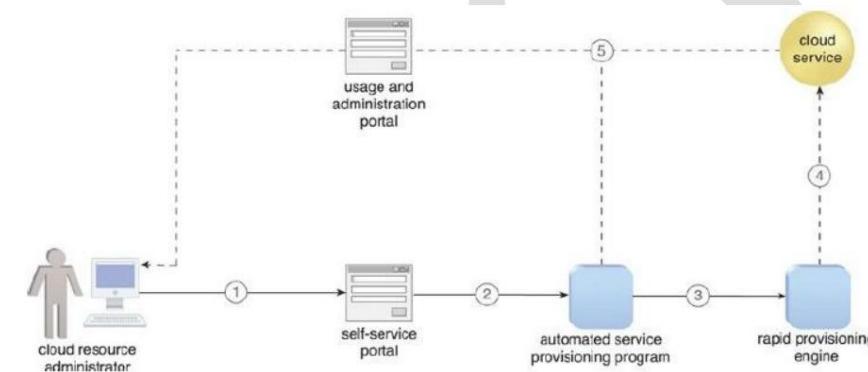
Oprogramowanie do zdalnego zarządzania zintegrowane z pamięcią ROM serwera staje się dostępne po uruchomieniu serwera. Do łączenia się z natywnym interfejsem zdalnego zarządzania serwera fizycznego zwykle używany jest internetowy lub zastrzeżony interfejs użytkownika, taki jak portal udostępniany przez system zdalnej administracji. Adresy IP na platformach IaaS można przekazywać bezpośrednio do konsumentów chmury, aby mogli niezależnie przeprowadzać instalacje systemu operacyjnego bare-metal.

System udostępniania bare-metal zapewnia funkcję automatycznego wdrażania, która umożliwia klientom chmury łączenie się z oprogramowaniem wdrożeniowym i jednocześnie udostępnianie więcej niż jednego serwera lub systemu operacyjnego. Centralny system wdrażania łączy się z serwerami za pośrednictwem ich interfejsów zarządzania i używa tego samego protokołu do przesyłania plików i działania jako agent w pamięci RAM serwera fizycznego. Serwer typu bare-metal staje się wówczas surowym klientem z zainstalowanym agentem zarządzania, a oprogramowanie wdrażające przesyła wymagane pliki konfiguracyjne w celu wdrożenia systemu operacyjnego.



4.2.9 Architektura szybkiego udostępniania

Architektura szybkiego udostępniania tworzy system, który automatyzuje dostarczanie szerokiego zakresu zasobów, indywidualnie lub zbiorowo. Podstawowa architektura technologii szybkiego udostępniania zasobów może być wyrafinowana i złożona i opiera się na systemie składającym się z programu do automatycznego udostępniania, silnika szybkiego udostępniania oraz skryptów i szablonów do udostępniania na żądanie.



- (1) Administrator zasobów w chmurze żąda nowej usługi w chmurze w ramach samoobsług portal.
- (2) Portal samoobsługowy przekazuje żądanie do programu automatycznego świadczenia usług zainstalowanego na serwerze wirtualnym.
- (3) który przekazuje niezbędne zadania do wykonania silnikowi szybkiego udostępniania.
- (4) Silnik szybkiego udostępniania ogłasza, kiedy nowa usługa w chmurze będzie gotowa.
- (5) Program zautomatyzowanego świadczenia usług finalizuje i publikuje usługę w chmurze w portalu użytkowania i administracji umożliwiającym dostęp do chmury dla konsumentów.

Opis krok po kroku opisuje wewnętrzne działanie mechanizmu szybkiego udostępniania:

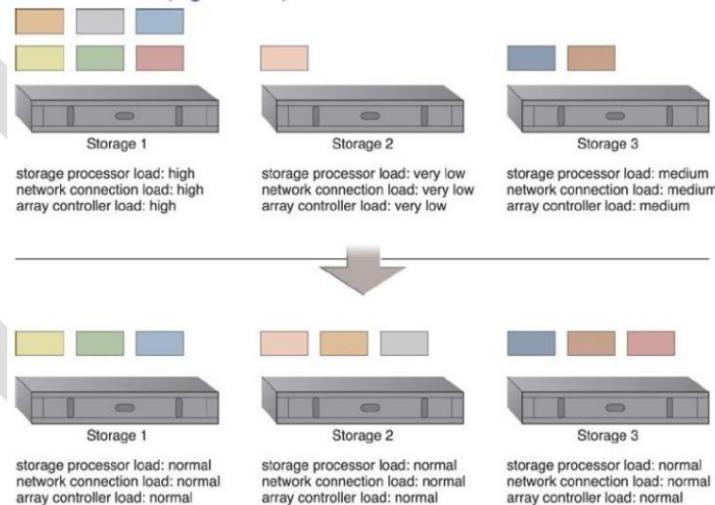
1. Konsument chmury składa wniosek o nowy serwer za pośrednictwem portalu samoobsługowego.
2. Menedżer sekwencji przekazuje żądanie do silnika wdrażania w celu przygotowania systemu operacyjnego.
3. Silnik wdrażania używa szablonów serwerów wirtualnych do udostępniania, jeśli żądanie dotyczy serwera wirtualnego. W przeciwnym razie aparat wdrażania wysyła żądanie udostępnienia serwera fizycznego.
4. Predefiniowany obraz dla żadanego typu systemu operacyjnego jest używany do udostępniania systemu operacyjnego, jeśli jest dostępny. W przeciwnym razie wykonywany jest zwykły proces wdrażania w celu zainstalowania systemu operacyjnego.
5. Silnik wdrażania informuje menedżera sekwencji, kiedy system operacyjny jest gotowy.

6. Menedżer sekwencji aktualizuje i wysyła dzienniki do rejestratora sekwencji składowanie.
7. Menedżer sekwencji żąda, aby silnik wdrażania zastosował operację linię bazową systemu do udostępnionego systemu operacyjnego.
8. Mechanizm wdrażania stosuje żądaną wersję bazową systemu operacyjnego.
9. Silnik wdrażania informuje menedżera sekwencji, że system operacyjny zastosowano poziom bazowy.
10. Menedżer sekwencji aktualizuje i wysyła dzienniki ukończonych kroków do rejestrator sekwencji do przechowywania.
11. Menedżer sekwencji żąda, aby silnik wdrażania zainstalował aplikacje.
12. Silnik wdrażania wdraża aplikacje na udostępnionym serwerze.
13. Silnik wdrażania informuje menedżera sekwencji, że aplikacje mają zostać zainstalowany.
14. Menedżer sekwencji aktualizuje i wysyła dzienniki ukończonych kroków do rejestrator sekwencji do przechowywania.
15. Menedżer sekwencji żąda, aby silnik wdrażania zastosował linię bazową konfiguracji aplikacji.

16. Silnik wdrażania stosuje linię bazową konfiguracji.
17. Silnik wdrażania informuje menedżera sekwencji, że konfiguracja zastosowano poziom bazowy.
18. Menedżer sekwencji aktualizuje i wysyła dzienniki ukończonych kroków do rejestrator sekwencji do przechowywania.

4.2.10 Architektura zarządzania obciążeniem pamięcią masową

Architektura ta umożliwia równomierną dystrybucję jednostek LUN pomiędzy dostępnymi urządzeniami pamięci masowej w chmurze, a system pojemności pamięci masowej jest tworzony w celu zapewnienia równomiernego rozłożenia obciążzeń w czasie wykonywania pomiędzy jednostkami LUN.



Łączenie urządzeń pamięci masowej w grupę umożliwia równomierną dystrybucję danych LUN pomiędzy dostępnymi hostami pamięci masowej. Skonfigurowano system zarządzania pamięcią masową i umieszczono zautomatyzowany odbiornik skalowania w celu monitorowania i wyrównywania obciążzeń w czasie wykonywania pomiędzy zgrupowanymi urządzeniami pamięci masowej w chmurze.

Słowniczek:

- Inteligentny silnik automatyzacji: Inteligentny silnik automatyzacji automatyzuje administrację zadania poprzez wykonanie skryptów zawierających logikę przepływu pracy.
- LUN: Numer jednostki logicznej (LUN) to dysk logiczny reprezentujący partycję dysku fizycznego prowadzić.
- Bramka usług pamięci masowej: Bramka usług pamięci masowej to komponent, który działa jako zewnętrzny interfejs do usług przechowywania w chmurze i jest w stanie automatycznie przekierowywać żądania konsumentów w chmurze za każdym razem, gdy zmieni się lokalizacja żądanego danych.
- Replikacja pamięci masowej: Replikacja pamięci masowej jest odmianą mechanizmów replikacji zasobów używany do synchronicznej lub asynchronicznej replikacji danych z podstawowego urządzenia pamięci masowej na dodatkowe urządzenie pamięci masowej. Można go używać do replikowania częściowych i całych jednostek LUN.
- Heartbeats: Heartbeats to komunikaty na poziomie systemu wymieniane pomiędzy hypervisorami, hypervisory i serwery wirtualne oraz hypervisory i VIM.
- Migracja maszyn wirtualnych na żywo: Migracja maszyn wirtualnych na żywo to system umożliwiający relokację wirtualną serwerów lub instancji serwerów wirtualnych w czasie wykonywania.
- Migracja jednostek LUN: migracja jednostek LUN to wyspecjalizowany program do przechowywania danych, który służy do przenoszenia jednostek LUN z jednego urządzenia pamięci masowej na drugie bez przerw, zachowując jednocześnie przejrzystość dla chmury konsumenci.

Bibliografia:

Koncepcje, technologia i architektura przetwarzania w chmurze

- Thomas Erl, Zaigham Mahmood i Ricardo Puttini – Prentice Hall – 2013
- Rozdział 11: Podstawowe architektury chmurowe
- Rozdział 12: Zaawansowane architektury chmurowe

Opanowanie podstaw przetwarzania w chmurze i programowania aplikacji

- Rajkumar Buyya, Christian Vecchiola, S. Thamarai Selvi – Elsevier – 2013

Przetwarzanie rozproszone i chmura obliczeniowa, od przetwarzania równoległego po Internet rzeczy

- Kai Hwang, Jack Dongarra, Geoffrey Fox – Wydawnictwo MK – 2012

Jednostka 5: Rozdział 1

Uwagi dotyczące modelu dostarczania w chmurze

Struktura jednostki

5.0 Cele 5.1

Wprowadzenie

5.2 Modele dostarczania w chmurze: perspektywa dostawcy chmury

5.2.1 Budowanie środowisk IaaS 5.2.2

Wyposażenie środowisk PaaS 5.2.3

Optymalizacja środowisk SaaS

5.3 Modele dostarczania w chmurze: perspektywa konsumenta chmury

5.3.1 Praca ze środowiskami IaaS 5.3.2 Praca

ze środowiskami PaaS 5.3.3 Praca z usługami

SaaS

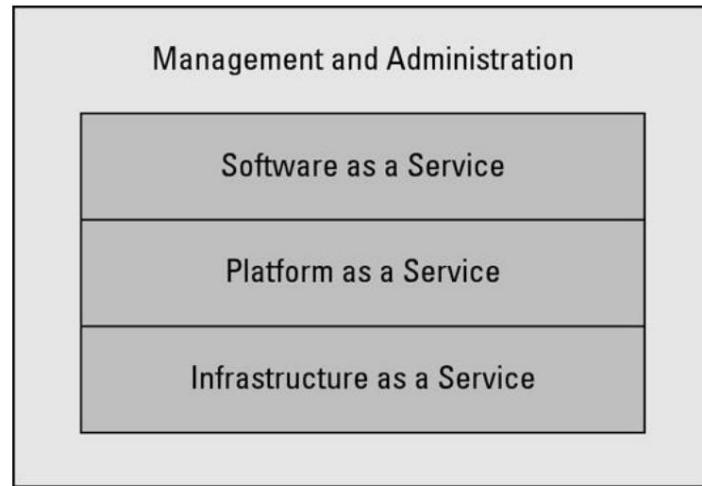
CELE 5.0

- Opisać modele dostarczania w chmurze dla IaaS
- Opisać modele dostarczania w chmurze dla PaaS
- Opisać modele dostarczania w chmurze dla SaaS
- Opisać różne sposoby, w jakie modele dostarczania w chmurze są administrowane i wykorzystywane przez konsumenci chmury
- Praca ze środowiskami IaaS • Praca ze środowiskami PaaS • Praca ze środowiskami SaaS

5.1 WSTĘP Model dostarczania chmury

reprezentuje specyficzną kombinację zasobów IT oferowanych przez dostawcę chmury. Terminologia ta jest zwykle kojarzona z przetwarzaniem w chmurze i często używana do opisania rodzaju zdalnego środowiska i poziomu kontroli.

5.2 Modele dostarczania w chmurze: perspektywa dostawcy chmury W tej sekcji omówiono architekturę i administrację modeli dostarczania w chmurze IaaS, PaaS i SaaS z punktu widzenia dostawcy chmury (rysunek 5.1). Badana jest integracja i zarządzanie tymi środowiskami opartymi na chmurze jako częścią większych środowisk oraz ich powiązanie z różnymi technologiami i kombinacjami mechanizmów chmurowych.

Rysunek 5.1

5.2.1 Budowanie środowisk IaaS

Mechanizmy serwerów wirtualnych i urządzeń pamięci masowej w chmurze reprezentują dwa najbardziej podstawowe zasoby IT dostarczane w ramach standardowej architektury szybkiego udostępniania w środowiskach IaaS. Oferowane są w różnych standardowych konfiguracjach, które są definiowane przez następujące właściwości:

- System operacyjny
- Pojemność pamięci podstawowej
- Zdolność przetwarzania
- Zwirtualizowana pojemność pamięci masowej

Pamięć i pojemność zwirtualizowanej pamięci masowej są zwykle przydzielane w przyrostach co 1 GB, aby uprościć udostępnianie podstawowych fizycznych zasobów IT. W przypadku ograniczania dostępu konsumentów chmury do środowisk zwirtualizowanych oferty IaaS są prewencyjnie składane przez dostawców usług w chmurze za pośrednictwem obrazów serwerów wirtualnych, które przechwytyują wstępnie zdefiniowane konfiguracje. Niektórzy dostawcy usług w chmurze mogą oferować konsumentom usług w chmurze bezpośredni dostęp administracyjny do fizycznych zasobów IT, w takim przypadku w grę może wchodzić architektura udostępniania typu bare-metal.

Można wykonywać migawki serwera wirtualnego w celu rejestracji jego bieżącego stanu, pamięci i konfiguracji zwirtualizowanego środowiska IaaS na potrzeby tworzenia kopii zapasowych i replikacji, zgodnie z wymaganiami skalowania poziomego i pionowego. Na przykład serwer wirtualny może wykorzystać swoją migawkę do ponownej inicjalizacji w innym środowisku hostingowym po zwiększeniu jego pojemności w celu umożliwienia skalowania w pionie. Migawkę można alternatywnie wykorzystać do zduplikowania serwera wirtualnego. Zarządzanie niestandardowymi obrazami serwerów wirtualnych jest istotną funkcją zapewnianą poprzez mechanizm systemu zdalnej administracji. Większość dostawców usług w chmurze obsługuje również opcje importowania i eksportowania niestandardowych obrazów serwerów wirtualnych, zarówno w formatach zastrzeżonych, jak i standardowych.

Centra danych

Dostawcy usług w chmurze mogą oferować zasoby IT oparte na IaaS z wielu geograficznie zróżnicowanych centrów danych, co zapewnia następujące podstawowe korzyści:

- Można połączyć wiele centrów danych w celu zwiększenia odporności. Każde centrum danych jest umieszczone w innej lokalizacji, aby zmniejszyć ryzyko, że pojedyncza awaria zmusi wszystkie centra danych do jednoczesnego przejścia w tryb offline.
- Połączone za pośrednictwem szybkich sieci komunikacyjnych o niskim poziomie opóźnienia, centra danych mogą równoważyć obciążenie, tworzyć kopie zapasowe i replikować zasoby IT oraz zwiększać pojemność pamięci masowej, poprawiając jednocześnie dostępność i niezawodność. Posiadanie wielu centrów danych rozproszonych na większym obszarze dodatkowo zmniejsza opóźnienia sieci.
- Centra danych wdrożone w różnych krajach sprawiają, że dostęp do zasobów IT jest wygodniejszy dla konsumentów chmury, którzy są ograniczeni wymogami prawnymi i regulacyjnymi.

Gdy środowisko IaaS jest wykorzystywane do zapewniania klientom chmury zwirtualizowanych środowisk sieciowych, każdy konsument chmury jest segregowany w środowisku dzierżawcy, które izoluje zasoby IT od reszty chmury za pośrednictwem Internetu. Sieci VLAN i oprogramowanie do kontroli dostępu do sieci wspólnie realizują odpowiednie logiczne obwody sieci.

Skalowalność i niezawodność

W środowiskach IaaS dostawcy usług w chmurze mogą automatycznie udostępniać serwery wirtualne za pośrednictwem dynamicznego skalowania pionowego architektury dynamicznej skalowalności. Można to zrobić za pośrednictwem VIM, o ile fizyczne serwery hosta mają wystarczającą pojemność. VIM może skalować serwery wirtualne przy użyciu replikacji zasobów w ramach architektury puli zasobów, jeśli dany serwer fizyczny nie ma wystarczającej wydajności, aby obsługiwać skalowanie pionowe. Mechanizm równoważenia obciążenia, będący częścią architektury dystrybucji obciążenia, może zostać wykorzystany do rozłożenia obciążenia pomiędzy zasobami IT w puli, aby zakończyć proces skalowania poziomego.

Ręczna skalowalność wymaga od konsumenta chmury interakcji z programem do obsługi i administrowania, aby jawnie zażądać skalowania zasobów IT. Natomiast automatyczna skalowalność wymaga, aby odbiornik automatycznego skalowania monitorował obciążenie i reaktywnie skalował pojemność zasobów. Mechanizm ten zazwyczaj działa jako agent monitorujący, który śledzi wykorzystanie zasobów IT w celu powiadamiania systemu zarządzania zasobami w przypadku przekroczenia wydajności.

Zreplikowane zasoby IT można zorganizować w konfiguracji o wysokiej dostępności, która tworzy system przełączania awaryjnego do wdrożenia za pośrednictwem standardowych funkcji VIM. Alternatywnie klaster zasobów o wysokiej dostępności i wydajności można utworzyć na poziomie serwera fizycznego lub wirtualnego albo na obu poziomach jednocześnie. Architektura wielościeżkowego dostępu do zasobów jest powszechnie stosowana w celu zwiększenia niezawodności poprzez wykorzystanie redundantnych ścieżek dostępu, a niektórzy dostawcy usług w chmurze oferują ponadto udostępnianie dedykowanych zasobów IT poprzez architekturę rezerwacji zasobów.

Monitorowanie

Monitory wykorzystania chmury w środowisku IaaS można wdrożyć przy użyciu VIM lub specjalistycznych narzędzi monitorujących, które bezpośrednio obejmują platformę wirtualizacyjną i/lub łączą się z nią. Kilka typowych możliwości platformy IaaS obejmuje monitorowanie:

- Cykle życia serwerów wirtualnych — rejestrowanie i śledzenie okresów czasu pracy oraz alokacja zasobów IT na potrzeby monitorów typu pay-per-use i celów rozliczeń opartych na czasie.
- Przechowywanie danych — śledzenie i przypisywanie alokacji pojemności do urządzeń pamięci masowej w chmurze na serwerach wirtualnych, w przypadku monitorów płatnych za użycie, które rejestrują wykorzystanie pamięci masowej

dla celów rozliczeniowych.

- Ruch sieciowy – w przypadku monitorów typu pay-per-use, które mierzą przychodzące i wychodzące wykorzystanie sieci oraz monitorów SLA, które śledzą wskaźniki QoS, takie jak czas odpowiedzi i straty w sieci.
- Warunki awarii — w przypadku monitorów SLA śledzących metryki zasobów IT i QoS w celu ostrzegania w przypadku awarii.
- Wyzwalacze zdarzeń — dla monitorów audytu, które ocenają i oceniają zgodność z przepisami wybranych zasobów IT.

Architektury monitorowania w środowiskach IaaS zazwyczaj obejmują agentów usług, którzy komunikują się bezpośrednio z systemami zarządzania backendem.

Bezpieczeństwo

Mechanizmy bezpieczeństwa w chmurze istotne dla zabezpieczenia środowisk IaaS obejmują:

- mechanizmy szyfrowania, haszowania, podpisu cyfrowego i PKI dla ogólnej ochrony danych przenoszenie
- Mechanizmy IAM i SSO umożliwiające dostęp do usług i interfejsów w systemach bezpieczeństwa, które opierają się na możliwościach identyfikacji, uwierzytelniania i autoryzacji użytkowników
- grupy bezpieczeństwa oparte na chmurze do izolowania środowisk wirtualnych za pomocą hypervisorów i segmentów sieci za pomocą oprogramowania do zarządzania siecią
- wzmocnione obrazy serwerów wirtualnych dla serwerów wirtualnych dostępnych wewnętrznie i zewnętrznie środowiska
- różne monitory wykorzystania chmury do śledzenia udostępnionych wirtualnych zasobów IT w celu wykrycia nieprawidłowe wzorce użytkowania.

5.2.2 Wyposażenie środowisk PaaS

Środowiska PaaS zazwyczaj muszą być wyposażone w różnorodne platformy do tworzenia i wdrażania aplikacji, aby dostosować się do różnych modeli programowania, języków i frameworków. Dla każdego stosu programistycznego tworzone jest zwykle osobne, gotowe środowisko, które zawiera oprogramowanie niezbędne do uruchomienia aplikacji opracowanych specjalnie dla danej platformy.

Każdej platformie towarzyszy odpowiedni zestaw SDK i IDE, które można zbudować na zamówienie lub włączyć za pomocą wtyczek IDE dostarczonych przez dostawcę chmury. Zestawy narzędzi IDE mogą lokalnie symuluować środowisko wykonawcze w chmurze w środowisku PaaS i zwykle obejmują serwery aplikacji wykonywalnych. W środowisku programistycznym symulowane są również ograniczenia bezpieczeństwa nieodłącznie związane ze środowiskiem wykonawczym, obejmujące kontrole pod kątem nieautoryzowanych prób uzyskania dostępu do zasobów IT systemu.

Dostawcy usług w chmurze często oferują mechanizm systemu zarządzania zasobami dostosowany do platformy PaaS, dzięki czemu konsumenci chmury mogą tworzyć i kontrolować niestandardowe obrazy serwerów wirtualnych za pomocą gotowych środowisk. Mechanizm ten zapewnia również funkcje specyficzne dla platformy PaaS, takie jak zarządzanie wdrożonymi aplikacjami i konfigurowanie wielu dzierżawców. Dostawcy usług w chmurze polegają ponadto na odmianie architektury szybkiego udostępniania, znanej jako udostępnianie platform, która została zaprojektowana specjalnie w celu udostępniania gotowych środowisk.

Skalowalność i niezawodność

Wymagania dotyczące skalowalności usług i aplikacji w chmurze wdrażanych w środowiskach PaaS są zazwyczaj zaspokajane poprzez architektury dynamicznej skalowalności i dystrybucji obciążenia, które opierają się na wykorzystaniu natywnych odbiorników automatycznego skalowania i modułów równoważenia obciążenia. Architektura łączenia zasobów jest dalej wykorzystywana do udostępniania zasobów IT z pul zasobów udostępnianych wielu klientom chmury.

Dostawcy usług w chmurze mogą oceniać ruch sieciowy i wykorzystanie połączeń po stronie serwera pod kątem obciążenia instancji, określając, w jaki sposób skalować przeciążoną aplikację zgodnie z parametrami i ograniczeniami kosztów dostarczonymi przez konsumenta chmury. Alternatywnie konsumenti chmury mogą skonfigurować projekty aplikacji, aby samodzielnie dostosować włączenie dostępnych mechanizmów.

Niezawodność gotowych środowisk oraz hostowanych usług i aplikacji w chmurze można wspierać za pomocą standardowych mechanizmów systemu przełączania awaryjnego ([Rysunek 142](#)), a także niezakłócającej pracy architektury relokacji usług, aby chronić konsumentów chmury przed warunkami przełączania awaryjnego. Architektura rezerwacji zasobów może również zapewniać wyłączny dostęp do zasobów IT opartych na PaaS. Jak

w połączeniu z innymi zasobami IT gotowe środowiska mogą również obejmować wiele centrów danych i regionów geograficznych, aby jeszcze bardziej zwiększyć dostępność i odporność

Monitorowanie

Specjalistyczne monitory wykorzystania chmury w środowiskach PaaS służą do monitorowania:

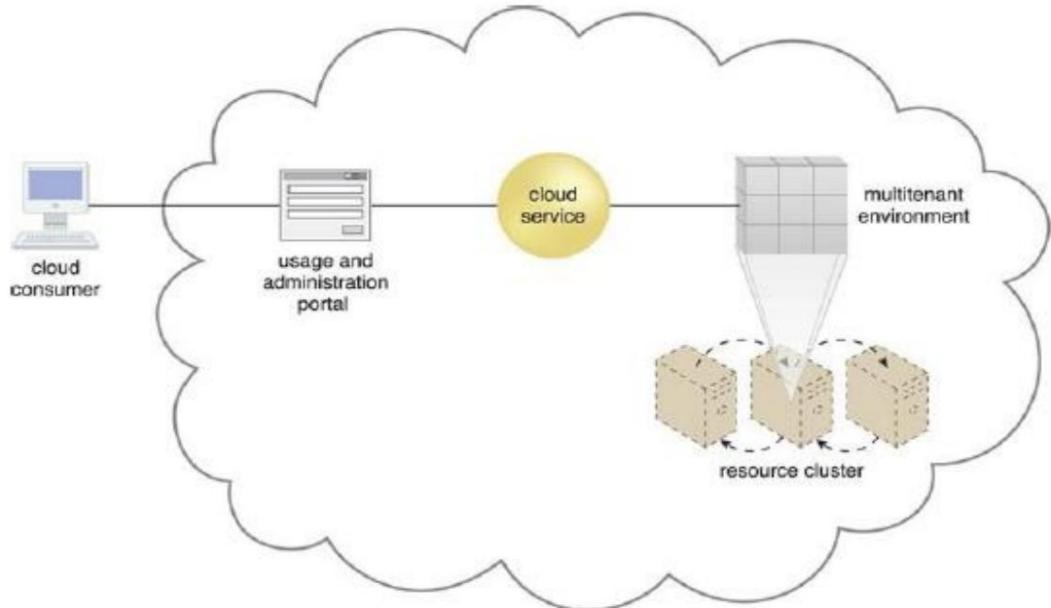
- Gotowe instancje środowiska — zastosowania tych instancji są rejestrowane przez monitory płatności za użycie w celu obliczenia opłat za użytkowanie na podstawie czasu.
- Trwałość danych — ta statystyka jest dostarczana przez monitory płatności za użycie, które rejestrują liczbę obiektów, wielkość zajmowanej pamięci masowej i transakcje w bazie danych w okresie rozliczeniowym.
- Wykorzystanie sieci — przychodzące i wychodzące wykorzystanie sieci jest śledzone za pomocą monitorów płatności za użycie i monitorów SLA, które śledzą metryki QoS związane z siecią.
- Warunki awarii — monitory SLA śledzące wymagane wskaźniki QoS zasobów IT przechwytywanie statystyk niepowodzeń.
- Wyzwalacze zdarzeń — ta metryka jest używana głównie przez monitory audytu, które muszą reagować na określone typy zdarzeń.

Bezpieczeństwo

Środowisko PaaS domyślnie nie powoduje potrzeby stosowania nowych mechanizmów bezpieczeństwa w chmurze poza tymi, które są już zapewnione dla środowisk IaaS.

5.2.3 Optymalizacja środowisk SaaS

We wdrożeniach SaaS architektury usług w chmurze są zazwyczaj oparte na środowiskach wielodostępnych, które umożliwiają i regulują wspólnie dostęp konsumentów do chmury ([rysunek 5.2](#)).



Rysunek 5.2 Usługa w chmurze oparta na SaaS jest hostowana w środowisku wielodostępnym wdrożonym w wysokowydajnym klastrze serwerów wirtualnych. Konsument chmury korzysta z portalu użytkowania i administracji, aby uzyskać dostęp do usługi w chmurze i ją skonfigurować.

Segregacja zasobów IT SaaS zazwyczaj nie występuje na poziomie infrastruktury w środowiskach SaaS, jak ma to miejsce w środowiskach IaaS i PaaS.

Wdrożenia SaaS w dużym stopniu opierają się na funkcjach zapewnianych przez natywne architektury dynamicznej skalowalności i dystrybucji obciążenia, a także na niezakłócającej relokacji usług, aby zapewnić, że warunki przełączania awaryjnego nie będą miały wpływu na dostępność usług w chmurze opartych na SaaS.

Należy jednak pamiętać, że w przeciwieństwie do stosunkowo prostych projektów produktów IaaS i PaaS, każde wdrożenie SaaS będzie wiązać się z unikalnymi wymaganiami dotyczącymi architektury, funkcjonalności i czasu działania. Wymagania te są specyficzne dla charakteru logiki biznesowej, w ramach której zaprogramowana jest usługa w chmurze oparta na SaaS, a także różnych wzorców użytkowania, którym podlegają klienci usług w chmurze.

Weźmy na przykład pod uwagę różnorodność funkcjonalności i wykorzystania następujących uznanych w Internecie Oferty SaaS:

- Wspólne tworzenie i udostępnianie informacji (Wikipedia, Blogger)
- Zarządzanie oparte na współpracy (Zimbra, Google Apps)
- Usługi konferencyjne w zakresie komunikatorów internetowych i komunikacji audio/wideo (Skype, Google Talk)
- Systemy zarządzania przedsiębiorstwem (ERP, CRM, CM)
- Udostępnianie plików i dystrybucja treści (YouTube, Dropbox)
- Oprogramowanie branżowe (inżynieria, bioinformatyka)
- Systemy przesyłania wiadomości (e-mail, poczta głosowa)
- Marketplace aplikacji mobilnych (Android Play Store, Apple App Store)

- Pakiety oprogramowania biurowego (Microsoft Office, Adobe Creative Cloud)
- Wyszukiwarki (Google, Yahoo)
- Media społecznościowe (Twitter, LinkedIn)

Weźmy teraz pod uwagę, że wiele z wcześniej wymienionych usług w chmurze jest oferowanych w jednym lub większej liczbie następujących nośników implementacji:

- Mobilna aplikacja
- Usługa REST
- Serwis internetowy

Każde z tych mediów implementacyjnych SaaS zapewnia internetowe interfejsy API umożliwiające komunikację między konsumentami w chmurze. Przykłady internetowych usług w chmurze opartych na SaaS z internetowymi interfejsami API obejmują:

- Usługi płatności elektronicznych (PayPal)
- Usługi mapowania i wyznaczania tras (Google Maps)
- Narzędzia do publikowania (WordPress)

Implementacje SaaS z obsługą urządzeń mobilnych są powszechnie obsługiwane przez mechanizm brokera wielourządzeniowego, chyba że usługa w chmurze jest przeznaczona wyłącznie do dostępu dla określonych urządzeń mobilnych.

Potencjalnie różnorodny charakter funkcjonalności SaaS, różnorodność technologii wdrażania oraz tendencja do oferowania usługi chmurowej opartej na SaaS w sposób redundantny z wieloma różnymi nośnikami implementacji, powodują, że projektowanie środowisk SaaS jest wysoce wyspecjalizowane. Chociaż nie jest to istotne dla wdrożenia SaaS, specjalistyczne wymagania dotyczące przetwarzania mogą powodować potrzebę włączenia modeli architektonicznych, takich jak:

- Równoważenie obciążenia usług — do dystrybucji obciążenia pomiędzy nadmiarowymi wdrożeniami usług chmurowych opartych na SaaS
- Dynamiczne wykrywanie i odzyskiwanie awarii – w celu ustanowienia systemu, który może automatycznie rozwiązać niektóre warunki awarii bez zakłócania usług wdrożenia SaaS.
- Okno konserwacji pamięci masowej — umożliwiające planowane przestoje konserwacyjne, które nie mają wpływu na dostępność wdrożenia SaaS
- Elastyczna pojemność zasobów/elastyczna pojemność sieci — w celu zapewnienia nieodłącznej elastyczności w architekturze usług w chmurze opartej na SaaS, która umożliwia automatyczne dostosowanie szeregu wymagań dotyczących skalowalności środowiska wykonawczego
- Równoważenie chmury – w celu zapewnienia dużej odporności na wdrażanie SaaS, co może być szczególnie ważne w przypadku usług w chmurze narażonych na ekstremalne jednocienne wykorzystanie

W środowiskach SaaS można używać wyspecjalizowanych monitorów wykorzystania chmury do śledzenia następujących typów metryk:

- Okres subskrypcji najemcy — ten wskaźnik jest używany przez monitory płatności za użycie do rejestracji i śledzenia wykorzystania aplikacji w celu rozliczania czasowego. Ten rodzaj monitorowania zwykle obejmuje licencjonowanie aplikacji i regularne oceny okresów leasingu wykraczających poza godzinne okresy środowisk IaaS i PaaS.

- Wykorzystanie aplikacji — ta metryka, oparta na użytkownikach lub grupach zabezpieczeń, jest używana w przypadku płatności za używać monitorów do rejestrowania i śledzenia użycia aplikacji do celów rozliczeniowych.
- Moduł funkcjonalny aplikacji najemcy — ta metryka jest używana przez monitory płatności za użycie w celu rozliczania opartego na funkcjach. Usługi w chmurze mogą mieć różne poziomy funkcjonalności w zależności od tego, czy konsument chmury korzysta z subskrypcji bezpłatnej, czy płatnej.

5.3 Modele dostarczania w chmurze: perspektywa konsumenta w chmurze

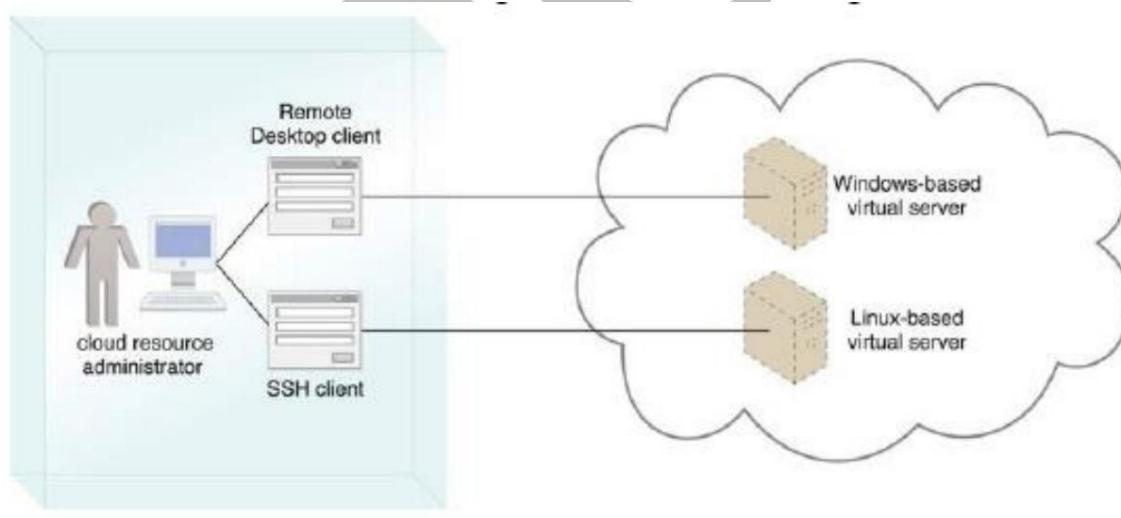
W tej sekcji poruszone różne rozważania dotyczące różnych sposobów administrowania modelami dostarczania usług w chmurze i korzystania z nich przez konsumentów usług w chmurze.

5.3.1 Praca ze środowiskami IaaS

Dostęp do serwerów wirtualnych można uzyskać na poziomie systemu operacyjnego za pośrednictwem zdalnych aplikacji terminalowych. W związku z tym typ używanego oprogramowania klienckiego zależy bezpośrednio od typu systemu operacyjnego działającego na serwerze wirtualnym, przy czym dwie typowe opcje to:

- Klient pulpitu zdalnego (lub połączenia pulpitu zdalnego) — dla systemu Windows środowiskach i przedstawia pulpit z graficznym interfejsem użytkownika systemu Windows
- Klient SSH – dla komputerów Mac i innych środowisk opartych na systemie Linux, zapewniający bezpieczeństwo połączenia kanałowe z tekstowymi kontami powłoki działającymi w systemie operacyjnym serwera

Rysunek 5.3 ilustruje typowy scenariusz użycia serwerów wirtualnych oferowanych jako usługi IaaS po ich utworzeniu za pomocą interfejsów zarządzania



Rysunek 5.3 Administracja zasobami w chmurze wykorzystuje klienta pulpitu zdalnego opartego na systemie Windows do administrowania serwerem wirtualnym opartym na systemie Windows i klientem SSH dla wirtualnego serwera opartego na systemie Linux

serwer.

Urządzenie pamięci masowej w chmurze można podłączyć bezpośrednio do serwerów wirtualnych i uzyskać do niego dostęp za pośrednictwem funkcjonalnego interfejsu serwerów wirtualnych w celu zarządzania przez system operacyjny. Alternatywnie urządzenie pamięci masowej w chmurze można podłączyć do zasobu IT hostowanego poza chmurą, na przykład urządzenia lokalnego za pośrednictwem sieci WAN lub VPN. W takich przypadkach powszechnie stosuje się następujące formaty manipulacji i przesyłania danych przechowywanych w chmurze:

- Sieciowy system plików — systemowy dostęp do pamięci masowej, którego renderowanie plików przypomina organizację folderów w systemach operacyjnych (NFS, CIFS)
- Urządzenia sieciowe do przechowywania danych — dostęp do pamięci masowej oparty na blokach zestawia i formatuje zróżnicowane geograficznie dane w spójne pliki w celu zapewnienia optymalnej transmisji sieciowej (iSCSI, Kanał światłowodowy)
- Zasoby internetowe — dostęp do pamięci obiektowej, za pomocą którego interfejs niezintegrowany z systemem operacyjnym logicznie reprezentuje pliki, do których można uzyskać dostęp za pośrednictwem interfejsu internetowego (Amazon S3)

Uwagi dotyczące udostępniania zasobów IT

Konsumenci chmury mają wysoki stopień kontroli nad tym, w jaki sposób i w jakim stopniu zasoby IT są udostępniane w ramach ich środowisk IaaS.

Na przykład:

- Kontrolowanie funkcji skalowalności (automatyczne skalowanie, równoważenie obciążenia)
- Kontrolowanie cyklu życia wirtualnych zasobów IT (wyłączanie, ponowne uruchamianie, włączanie urządzenia wirtualne)
- Kontrolowanie środowiska sieci wirtualnej i reguł dostępu do sieci (firewall, granice sieci logicznej)
- Tworzenie i wyświetlanie umów o świadczenie usług (warunki konta, użytkowanie warunki)
- Zarządzanie podłączaniem urządzeń pamięci masowej w chmurze
- Zarządzanie wstępna alokacją zasobów IT w chmurze (rezerwacja zasobów)
- Zarządzanie poświadczeniami i hasłami dla administratorów zasobów w chmurze
- Zarządzanie poświadczeniami dla grup zabezpieczeń opartych na chmurze, które uzyskują dostęp do zwirtualizowanych zasobów IT poprzez IAM
- Zarządzanie konfiguracjami związanymi z bezpieczeństwem
- Zarządzanie dostosowanym przechowywaniem obrazów serwerów wirtualnych (importowanie, eksportowanie, tworzenie kopii zapasowych)
- Wybór opcji wysokiej dostępności (przelączanie awaryjne, klastrowanie zasobów IT)
- Wybór i monitorowanie metryk SLA
- Wybór podstawowych konfiguracji oprogramowania (system operacyjny, preinstalowane oprogramowanie dla nowych serwerów wirtualnych) Wybór instancji zasobów IaaS spośród szeregu dostępnych konfiguracji i opcji sprzętowych (możliwości przetwarzania, RAM, pamięć masowa)
- Wybór regionów geograficznych, w których powinny być hostowane zasoby IT oparte na chmurze
- Śledzenie i zarządzanie kosztami

Interfejs zarządzania tego typu zadaniami udostępniania to zazwyczaj portal użytkowania i administracji, ale może być również oferowany za pośrednictwem narzędzi interfejsu wiersza poleceń (CLI), które mogą uprościć wykonywanie wielu skryptowych działań administracyjnych.

Chociaż zazwyczaj preferowana jest standaryzacja prezentacji funkcji administracyjnych i kontroli, czasami uzasadnione może być użycie różnych narzędzi i interfejsów użytkownika. Na przykład można utworzyć skrypt włączający i wyłączający serwery wirtualne co noc za pośrednictwem interfejsu CLI, natomiast dodawanie lub usuwanie pojemności pamięci masowej można łatwiej przeprowadzić za pomocą portalu.

5.3.2 Praca ze środowiskami PaaS

Typowe środowisko IDE PaaS może oferować szeroką gamę narzędzi i zasobów programistycznych, takich jak biblioteki oprogramowania, biblioteki klas, frameworki, interfejsy API i różne możliwości środowiska wykonawczego, które emulują zamierzone środowisko wdrażania oparte na chmurze. Funkcje te pozwalają programistom tworzyć, testować i uruchamiać kod aplikacji w chmurze lub lokalnie (lokalnie), korzystając ze środowiska IDE do emulacji środowiska wdrażania w chmurze. Skompilowane lub ukończone aplikacje są następnie pakowane i przesyłane do chmury, a następnie wdrażane za pośrednictwem gotowych środowisk. Ten proces wdrażania można również kontrolować za pomocą środowiska IDE.

PaaS umożliwia także aplikacjom korzystanie z urządzeń do przechowywania w chmurze jako niezależnych systemów przechowywania danych do przechowywania danych specyficznych dla rozwoju (na przykład w repozytorium dostępnym poza środowiskiem chmury). Ogólnie obsługiwane są struktury baz danych SQL i NoSQL.

Uwagi dotyczące udostępniania zasobów IT

Środowiska PaaS zapewniają mniejszą kontrolę administracyjną niż środowiska IaaS, ale nadal oferują znaczny zakres funkcji zarządzania.

Na przykład:

- Tworzenie i wyświetlanie umów o świadczenie usługi, takich jak warunki konta i warunki użytkowania
- Wybór platformy oprogramowania i frameworków programistycznych dla gotowych środowisk
- Wybór typów instancji, które są najczęściejinstancjami frontendowymi lub backendowymi • Wybór urządzeń do przechowywania danych w chmurze do wykorzystania w gotowych środowiskach
- Kontrolowanie cyklu życia aplikacji opracowanych w ramach PaaS (wdrożenie, uruchomienie, zamknięcie, ponowne uruchomienie i zwolnienie)
- Kontrolowanie wersjonowania wdrożonych aplikacji i modułów
- Konfigurowanie mechanizmów związanych z dostępnością i niezawodnością
- Zarządzanie poświadczaniami dla programistów i administratorów zasobów w chmurze przy użyciu uprawnień
- Zarządzanie ogólnymi ustawieniami zabezpieczeń, takimi jak dostępne porty sieciowe
- Wybór i monitorowanie wskaźników SLA związanych z PaaS
- Zarządzanie i monitorowanie kosztów użytkowania i zasobów IT
- Kontrolowanie funkcji skalowalności, takich jak limity użycia, progi aktywnych instancji oraz konfiguracja i wdrażanie automatycznego odbiornika skalowania i mechanizmów równoważenia obciążenia

5.3.3 Praca z usługami SaaS

Ponieważ usługom w chmurze opartym na SaaS prawie zawsze towarzyszą ulepszone i ogólne interfejsy API, zwykle projektuje się je tak, aby można je było włączyć jako część większych rozwiązań rozproszonych. Typowym przykładem są Mapy Google, które oferują kompleksowy interfejs API umożliwiający włączanie informacji i obrazów map do witryn internetowych i aplikacji internetowych.

Wiele ofert SaaS jest świadczonych bezpłatnie, chociaż te usługi w chmurze często obejmują podprogramy gromadzące dane, które zbierają dane o użytkowaniu na korzyść dostawcy chmury. Podczas korzystania z dowolnego produktu SaaS sponsorowanego przez strony trzecie istnieje uzasadnione prawdopodobieństwo, że jest to forma gromadzenia informacji ogólnych. Zapoznanie się z umową dostawcy usług w chmurze zazwyczaj pomoże rzucić światło na wszelkie dodatkowe działania, do wykonywania których zaprojektowano usługę w chmurze.

Konsumenci chmury korzystający z produktów SaaS dostarczanych przez dostawców usług w chmurze są zwolnieni z obowiązków związanych z wdrażaniem i administrowaniem bazowymi środowiskami hostingowymi.

Opcje dostosowywania są zwykle dostępne dla konsumentów chmury; jednak opcje te są zazwyczaj ograniczone do kontroli wykorzystania w czasie wykonywania wystąpień usług w chmurze, które są generowane specjalnie przez i dla konsumenta chmury.

Na przykład:

- Zarządzanie konfiguracjami związanymi z bezpieczeństwem
- Zarządzanie wybranymi opcjami dostępności i niezawodności
- Zarządzanie kosztami użytkowania
- Zarządzanie kontami użytkowników, profilami i autoryzacją dostępu
- Wybór i monitorowanie umów SLA
- Ustawianie opcji i ograniczeń ręcznej i automatycznej skalowalności.



Jednostka 5: Rozdział 2

Metryki kosztów i modele cenowe

Struktura jednostki

- 5.2..0 Cele
- 5.2..1 Wprowadzenie
- 5.2..2 Wskaźniki kosztów biznesowych
- 5.2..3 Wskaźniki kosztów korzystania z chmury
 - 5.2..3.1 Wykorzystanie sieci
 - 5.2..3.2 Wykorzystanie serwera
 - 5.2..3.4 Korzystanie z Usługi Przetwarzania w Chmurze
- 5.2..4 Rozważania dotyczące zarządzania kosztami
 - 5.2..4.1 Modele cenowe
 - 5.2..4.2 Dodatkowe uwagi
- 5.2..5 Umowy dotyczące poziomu usług (SLA)
- 5.2..6 Wskaźniki jakości usług 5.2..6.1
 - Wskaźniki dostępności usług
 - 5.2..6.2 Wskaźniki niezawodności usług
 - 5.2..6.3 Wskaźniki wydajności usług
 - 5.2..6.4 Wskaźniki skalowalności usług
 - 5.2..6.5 Wskaźniki odporności usług
- 5.2..7 Wytyczne SLA

5.2..0 CELE

W tym rozdziale przedstawiono metryki, formuły i praktyki pomagające konsumentom chmury w przeprowadzaniu dokładnej analizy finansowej planów wdrożenia chmury.

5.2..1 WSTĘP

Obniżanie kosztów operacyjnych i optymalizacja środowisk IT mają kluczowe znaczenie dla zrozumienia i możliwości porównania modeli kosztów leżących u podstaw udostępniania środowisk lokalnych i środowisk opartych na chmurze. Struktury cenowe stosowane w chmurach publicznych są zazwyczaj oparte na modelach płatności za użycie zorientowanych na użyteczność, co pozwala organizacjom uniknąć początkowych inwestycji w infrastrukturę. Modele te należy ocenić pod kątem konsekwencji finansowych infrastruktury lokalnej inwestycji i związanych z nimi zobowiązań w zakresie całkowitego kosztu posiadania.

5.2..2 Wskaźniki kosztów biznesowych

Ta sekcja rozpoczyna się od opisu typowych typów wskaźników używanych do oceny szacunkowych kosztów i wartości biznesowej leasingu zasobów IT w chmurze w porównaniu z zakupem lokalnych zasobów IT.

Koszty początkowe i bieżące

Koszty początkowe są związane z początkowymi inwestycjami, które organizacje muszą poczynić, aby sfinansować zasoby IT, z których zamierzają korzystać. Dotyczy to zarówno kosztów związanych z pozyskaniem zasobów informatycznych, jak i wydatków związanych z ich wdrożeniem i administracją.

- Początkowe koszty zakupu i wdrożenia lokalnych zasobów IT są zazwyczaj wysokie. Przykładowe koszty początkowe w przypadku środowisk lokalnych mogą obejmować sprzęt, oprogramowanie i pracę wymaganą do wdrożenia.
- Początkowe koszty leasingu zasobów IT w chmurze są zwykle niskie. Przykłady kosztów początkowych w przypadku środowisk chmurowych mogą obejmować koszty pracy wymagane do oceny i skonfigurowania środowiska chmurowego.

Koszty bieżące to wydatki wymagane przez organizację do obsługi i utrzymania wykorzystywanych przez nią zasobów IT.

- Bieżące koszty obsługi lokalnych zasobów IT mogą się różnić. Przykłady obejmują opłaty licencyjne, energię elektryczną, ubezpieczenie i robociznę.
- Bieżące koszty obsługi zasobów IT w chmurze również mogą się różnić, ale często przekraczają bieżące koszty lokalnych zasobów IT (szczególnie w dłuższym okresie). Przykładami mogą być opłaty za leasing sprzętu wirtualnego, opłaty za wykorzystanie przepustowości, opłaty licencyjne i robociznę.

Dodatkowe koszty

Aby uzupełnić i rozszerzyć analizę finansową poza obliczenia i porównanie standardowych wskaźników kosztów biznesowych, bieżących i bieżących, można wziąć pod uwagę kilka innych, bardziej wyspecjalizowanych wskaźników kosztów biznesowych.

Na przykład:

- Koszt kapitału – Koszt kapitału to wartość reprezentująca koszt poniesiony w wyniku pozyskania wymaganych funduszy. Na przykład zazwyczaj droższe będzie zgromadzenie początkowej inwestycji w wysokości 150 000 dolarów niż zebranie tej kwoty w ciągu trzech lat. Adekwatność tego kosztu zależy od sposobu, w jaki organizacja zamierza gromadzić potrzebne fundusze. Jeśli koszt kapitału początkowej inwestycji jest wysoki, dodatkowo uzasadnia to leasing zasobów IT oparty na chmurze.
- Koszty utopione — organizacja często dysponuje istniejącymi zasobami IT już opłacony i sprawny. Wcześniejsze inwestycje dokonane w te lokalne zasoby IT nazywane są kosztami utopianymi. Porównując koszty początkowe ze znacznymi kosztami utopianymi, uzasadnienie leasingu zasobów IT w chmurze jako alternatywy może być trudniejsze.
- Koszty integracji – Testowanie integracji to forma testowania wymagana do pomiaru wysiłku wymaganego do zapewnienia kompatybilności i interoperacyjności zasobów IT w środowisku obcym, takim jak nowa platforma chmurowa. W zależności od modelu wdrażania chmury i modelu dostarczania chmury rozważanego przez organizację może zaistnieć potrzeba dalszego przydzielania środków na przeprowadzenie testów integracyjnych i dodatkowej pracy związanej z zapewnieniem interoperacyjności między konsumentami usług w chmurze a usługami w chmurze. Wydatki te nazywane są kosztami integracji. Wysokie koszty integracji mogą sprawić, że leasing zasobów IT w chmurze będzie mniej atrakcyjny.
- Koszty zablokowane — jak wyjaśniono w sekcji Zagrożenia i wyzwania w Rozdziale 3, środowiska chmurowe mogą nakładać ograniczenia w zakresie przenośności. Podczas przeprowadzania analizy metryk przez dłuższy okres czasu może być konieczne wzięcie pod uwagę możliwości posiadania

przejścia od jednego dostawcy usług w chmurze do innego. Ze względu na fakt, że konsumenti usług w chmurze mogą uzależnić się od zastrzeżonych cech środowiska chmury, z tego typu posunięciem wiążą się pewne koszty. Zablokowane koszty mogą jeszcze bardziej zmniejszyć długoterminową wartość biznesową leasingu zasobów IT w chmurze.

5.2.3 Wskaźniki kosztów korzystania z chmury

W poniższych sekcjach opisano zestaw wskaźników kosztów użytkowania służących do obliczania kosztów związanych z pomiarami wykorzystania zasobów IT w chmurze:

- Użycie sieci – ruch sieciowy przychodzący i wychodzący, a także sieć wewnętrz chmury
ruch drogowy
- Użycie serwera – alokacja serwera wirtualnego (i rezerwacja zasobów)
- Urządzenie Cloud Storage – alokacja pojemności
- Usługa Chmurowa – czas trwania subskrypcji, liczba nominowanych użytkowników, liczba transakcji (usług chmurowych i aplikacji opartych na chmurze)

Dla każdego miernika kosztów użytkowania podany jest opis, jednostka miary i częstotliwość pomiarów, a także model dostarczania usług w chmurze najbardziej odpowiedni dla danego miernika. Każdy wskaźnik jest dodatkowo uzupełniony krótkim przykładem.

5.2..3.1 Wykorzystanie sieci

Zdefiniowane jako ilość danych przesyłanych przez połączenie sieciowe, wykorzystanie sieci jest zwykle obliczane na podstawie oddzielnie mierzonych wskaźników ruchu przychodzącego i ruchu wychodzącego w odniesieniu do usług w chmurze lub innych zasobów IT.

Wskaźnik wykorzystania sieci przychodzącej

- Opis – przychodzący ruch sieciowy
- Pomiar - £, przychodzący ruch sieciowy w bajtach
- Częstotliwość – ciągła i kumulacyjna w zdefiniowanym okresie
- Model dostarczania w chmurze - IaaS, PaaS, SaaS
- Przykład – do 1 GB bezpłatnie, 0,001 USD/GB do wersji 5.2. Gruźlica miesięcznie

Wskaźnik wykorzystania sieci wychodzącej

- Opis - wychodzący ruch sieciowy
- Pomiar - £, wychodzący ruch sieciowy w bajtach
- Częstotliwość – ciągła i kumulacyjna w zdefiniowanym okresie
- Model dostarczania w chmurze - IaaS, PaaS, SaaS
- Przykład — do 1 GB bezpłatnie miesięcznie, 0,01 USD/GB od 1 GB do 5,2. TB miesięcznie

Metryki wykorzystania sieci można zastosować do ruchu WAN pomiędzy zasobami IT jednej chmury, które są zlokalizowane w różnych regionach geograficznych, w celu obliczenia kosztów synchronizacji, replikacji danych i powiązanych form przetwarzania. Z drugiej strony wykorzystanie sieci LAN i inny ruch sieciowy pomiędzy zasobami IT znajdującymi się w tym samym centrum danych zazwyczaj nie są śledzone.

Wskaźnik wykorzystania sieci WAN wewnętrz chmury

- Opis - ruch sieciowy pomiędzy geograficznie zróżnicowanymi zasobami IT Chmura
- Pomiar - £, wewnętrzchmurowy ruch WAN w bajtach
- Częstotliwość – ciągła i kumulacyjna w zdefiniowanym okresie

- Model dostarczania w chmurze - IaaS, PaaS, SaaS
- Przykład — do 500 MB bezpłatnie dziennie, a później 0,01 USD/GB, 0,005 USD/GB po 1 TB na miesiąc

Wielu dostawców usług w chmurze nie pobiera opłat za ruch przychodzący, aby zachęcić konsumentów chmury do migracji danych do chmury. Niektóre nie pobierają również opłat za ruch WAN w tej samej chmurze.

Wskaźniki kosztów związane z siecią są określane na podstawie następujących właściwości:

- Użycie statycznego adresu IP - czas przydziału adresu IP (jeśli wymagany jest statyczny adres IP)
- Równoważenie obciążenia sieciowego - ilość ruchu sieciowego z równoważeniem obciążenia (w bajtach)
- Wirtualna zapora sieciowa - ilość ruchu sieciowego przetwarzanego przez zaporę (według czasu alokacji)

5.2..3.2 Wykorzystanie serwera

Alokację serwerów wirtualnych mierzy się za pomocą typowych wskaźników płatności za użycie w środowiskach IaaS i PaaS, które są określone ilościowo na podstawie liczby serwerów wirtualnych i gotowych środowisk.

Ta forma pomiaru wykorzystania serwera dzieli się na metryki alokacji instancji maszyny wirtualnej na żądanie i metryki alokacji instancji zarezerwowanej maszyny wirtualnej .

Pierwszy wskaźnik mierzy opłaty za użytkowanie w ujęciu krótkoterminowym, podczas gdy drugi mierzy opłaty rezerwacyjne z góry za korzystanie z serwerów wirtualnych przez dłuższe okresy. Opłata za rezerwację z góry jest zwykle stosowana w połączeniu ze obniżonymi opłatami za użytkowanie.

Metryka alokacji instancji maszyny wirtualnej na żądanie

- Opis - czas działania instancji serwera wirtualnego
- Pomiar - E, data rozpoczęcia i zakończenia serwera wirtualnego
- Częstotliwość - ciągła i kumulacyjna w zdefiniowanym okresie
- Model dostarczania w chmurze - IaaS, PaaS
- Przykład — mała instancja 0,52 USD/godzinę, średnia instancja 0,20 USD/godzinę, duża instancja 0,50 USD/godzinę instancja

Metryka alokacji instancji zarezerwowanej maszyny wirtualnej

- Opis - koszt początkowej rezerwacji instancji serwera wirtualnego
- Pomiar - E, data rozpoczęcia rezerwacji serwera wirtualnego do daty wygaśnięcia
- Częstotliwość - codziennie, co miesiąc, co rok
- Model dostarczania w chmurze - IaaS, PaaS
- Przykład - 55,52 USD/mała instancja, 55,50 USD/średnia instancja, 245,50 USD/duża instancja

3-Użycie urządzenia magazynującego w chmurze

Opłaty za przechowywanie w chmurze są zazwyczaj naliczane na podstawie ilości miejsca przydzielonego w określonym z góry okresie, mierzonego za pomocą wskaźnika alokacji miejsca na żądanie . Podobnie jak w przypadku wskaźników kosztów opartych na IaaS, opłaty za alokację pamięci masowej na żądanie opierają się zwykle na krótkich odstępach czasu (np. na podstawie stawki godzinowej). Innym powszechnym miernikiem kosztów przechowywania w chmurze są przesłane dane we/wy, które mierzą ilość przesyłanych danych wejściowych i wyjściowych.

Wskaźnik alokacji przestrzeni dyskowej na żądanie

- Opis – czas trwania i wielkość przydziału pamięci na żądanie w bajtach • Pomiar – E, data zwolnienia pamięci/relokacji do daty przydziału pamięci (resetuje się po zmianie wielkości pamięci)

- Częstotliwość – ciągła • Model dostarczania w chmurze – IaaS, PaaS, SaaS • Przykład – 0,01 USD/GB na godzinę (zwykle wyrażane w GB/miesiąc)

Metryka przesyłanych danych we/wy

- Opis – ilość przesyłanych danych I/O • Pomiar – E, dane I/O w bajtach • Częstotliwość – ciągła • Model dostarczania w chmurze – IaaS, PaaS • Przykład – \$0.5.2./TB

5.2..3.4 Wykorzystanie usług w chmurze

Wykorzystanie usług w chmurze w środowiskach SaaS jest zazwyczaj mierzone przy użyciu następujących trzech wskaźników:

Wskaźnik czasu trwania subskrypcji aplikacji • Opis — czas

- trwania subskrypcji korzystania z usług w chmurze • Pomiar — E, data rozpoczęcia subskrypcji do daty wygaśnięcia • Częstotliwość — dzienna, miesięczna, roczna • Model dostarczania w chmurze — SaaS • Przykład — 65,50 USD miesięcznie

Wskaźnik liczby nominowanych użytkowników

- Opis – liczba zarejestrowanych użytkowników z legalnym dostępem • Pomiar – liczba użytkowników
- Częstotliwość – miesięczna, roczna • Model dostarczania w chmurze – SaaS • Przykład – 0,50 USD/dodatkowy użytkownik miesięcznie

Wskaźnik liczby użytkowników transakcji

- Opis – liczba transakcji obsługiwanych przez usługę w chmurze • Pomiar – liczba transakcji (wymiana komunikatów żądanie-odpowiedź) • Częstotliwość – ciągła • Model dostarczania w chmurze – PaaS, SaaS • Przykład – 0,05 USD za 1000 transakcji

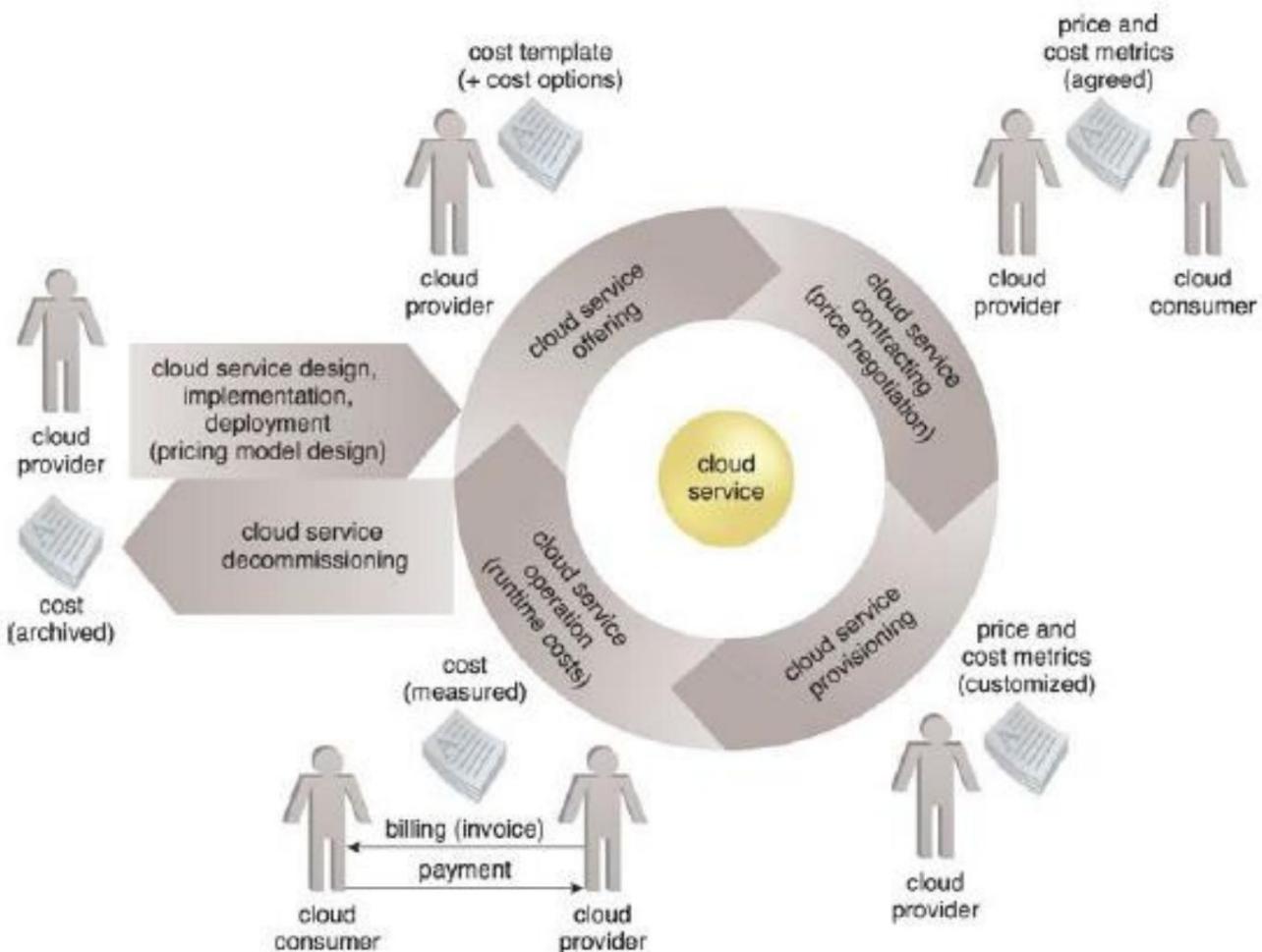
5.2.4 Rozważania dotyczące zarządzania kosztami Zarządzanie kosztami

często koncentruje się wokół faz cyklu życia usług w chmurze, jak następuje:

- Projektowanie i rozwój usług w chmurze — na tym etapie organizacja dostarczająca usługę w chmurze zazwyczaj definiuje podstawowe modele cenowe i szablony kosztów. • Wdrożenie usługi w chmurze — przed i w trakcie wdrażania usługi w chmurze

Określono i wdrożono architekturę zaplecza do pomiaru wykorzystania i gromadzenia danych związanych z rozliczeniami, w tym umiejscowienie monitora płatności za użycie i mechanizmów systemu zarządzania rozliczeniami.

- Umowa o świadczenie usług w chmurze – ta faza składa się z negocjacji pomiędzy konsumentem chmury a dostawcą chmury, których celem jest osiągnięcie wzajemnego porozumienia w sprawie stawek opartych na wskaźnikach kosztów użytkowania.
- Oferta usług w chmurze — ten etap obejmuje konkretną ofertę modeli cenowych usług w chmurze za pomocą szablonów kosztów i wszelkich dostępnych opcji dostosowywania.
- Dostarczanie usług w chmurze — progi dotyczące wykorzystania usług w chmurze i tworzenia instancji mogą zostać narzucone przez dostawcę chmury lub ustalone przez konsumenta chmury. Tak czy inaczej, te i inne opcje udostępniania mogą mieć wpływ na koszty użytkowania i inne opłaty.
- Działanie usługi w chmurze – Jest to faza, podczas której aktywne jest korzystanie z usługi w chmurze generuje dane dotyczące metryk kosztów użytkowania.
- Likwidacja usługi w chmurze — w przypadku tymczasowej lub trwałej dezaktywacji usługi w chmurze mogą zostać zarchiwizowane statystyczne dane dotyczące kosztów.



Rysunek 5.2.1 Typowe etapy cyklu życia usług w chmurze w kontekście zarządzania kosztami.

5.2..4.1 Modele cenowe

Modele cenowe stosowane przez dostawców usług w chmurze są definiowane przy użyciu szablonów określających koszty jednostkowe dla szczegółowego wykorzystania zasobów zgodnie ze wskaźnikami kosztów użytkowania. Na model cenowy mogą wpływać różne czynniki, takie jak:

- Konkurencja rynkowa i wymogi regulacyjne
- Koszty ogólne poniesione podczas projektowania, rozwoju, wdrażania i obsługi chmury usług i innych zasobów IT
- Możliwości redukcji wydatków poprzez współdzielenie zasobów IT i optymalizację centrum danych

Większość głównych dostawców usług w chmurze oferuje usługi w chmurze po stosunkowo stabilnych, konkurencyjnych cenach, mimo że ich własne wydatki mogą być zmienne. Szablon cenowy lub plan cenowy zawiera zestaw standardowych kosztów i metryk określających sposób pomiaru i obliczania opłat za usługi w chmurze. Szablony cen definiują strukturę modelu cenowego poprzez ustalenie różnych jednostek miary, limitów wykorzystania, rabatów i innych skodyfikowanych opłat. Model cenowy może zawierać wiele szablonów cen, których formuła jest określana na podstawie zmiennych takich jak:

- Wskaźniki kosztów i powiązane ceny – są to koszty zależne od rodzaju Alokacja zasobów IT (taką jak alokacja na żądanie lub alokacja zarezerwowana).
- Definicje stawek stałych i zmiennych – stawki stałe opierają się na alokacji zasobów i definiują limity wykorzystania zawarte w cenie stałej, natomiast stawki zmienne są dostosowane do rzeczywistego wykorzystania zasobów.
- Rabaty zbiorcze — w miarę stopniowego zwiększania się stopnia skalowania zasobów IT zużywa się więcej zasobów IT, co prawdopodobnie kwalifikuje konsumenta chmury do wyższych rabatów.
- Opcje dostosowywania kosztów i cen — ta zmienna jest powiązana z opcjami i harmonogramami płatności. Na przykład konsumenti chmury mogą mieć możliwość wyboru miesięcznych, półrocznych lub rocznych rat płatności.

5.2..4.2 Dodatkowe uwagi

- Negocjacje — ceny dostawców usług w chmurze są często otwarte na negocjacje, szczególnie w przypadku klientów, którzy chcą oferować większe wolumeny lub dłuższe okresy. Negocjacje cenowe można czasami przeprowadzić online za pośrednictwem witryny internetowej dostawcy usług w chmurze, przesyłając szacunkowe wielkości wykorzystania wraz z proponowanymi rabatami. Istnieją nawet narzędzia dostępne dla konsumentów chmury, które pomagają w generowaniu dokładnych szacunków wykorzystania zasobów IT w tym celu.
- Opcje płatności — po zakończeniu każdego okresu pomiarowego system zarządzania rozliczeniami dostawcy usług w chmurze oblicza kwotę należną od konsumenta usług w chmurze. Konsumenti korzystający z chmury mają do dyspozycji dwie popularne opcje płatności: przedpłata i płatność po wykonaniu usługi. Dzięki rozliczeniom przedpłaconym klienci korzystający z chmury otrzymują kredyty na wykorzystanie zasobów IT, które można wykorzystać do pokrycia przyszłych rachunków za użytkowanie. W przypadku metody płatności z dołu klienci korzystający z chmury otrzymują rachunki i faktury za każdy okres wykorzystania zasobów IT, który zwykle odbywa się co miesiąc.
- Archiwizacja kosztów — śledząc historyczne informacje rozliczeniowe, zarówno dostawcy usług chmurowych, jak i konsumenti usług chmurowych mogą generować wnikliwe raporty, które pomagają identyfikować trendy w zakresie użytkowania i finansów.

5.2..5 Umowy dotyczące poziomu usług (SLA)

Umowy dotyczące poziomu usług (SLA) są centralnym punktem negocjacji, warunków umów, zobowiązań prawnych oraz wskaźników i pomiarów czasu działania. Umowy SLA formalizują gwarancje oferowane przez dostawców usług w chmurze i odpowiednio wpływają lub określają modele cenowe i warunki płatności. Umowy SLA wyznaczają oczekiwania konsumentów w chmurze i są integralną częścią sposobu, w jaki organizacje budują biznes automatyzacją wykorzystania zasobów IT w chmurze.

Gwarancje udzielone przez dostawcę chmury konsumentowi chmury są często przenoszone na inne osoby, w tym sensie, że te same gwarancje są udzielane przez organizację konsumencką chmury swoim klientom, partnerom biznesowym lub komukolwiek, kto będzie polegał na usługach i rozwiązaniach udostępnianych przez dostawcę chmury. Dlatego kluczowe znaczenie ma zrozumienie umów SLA i powiązanych wskaźników jakości usług oraz dostosowanie wsparcia do wymagań biznesowych konsumenta chmury, przy jednoczesnym zapewnieniu, że gwarancje mogą w rzeczywistości zostać realistycznie spełnione w sposób spójny i niezawodny przez dostawcę usług w chmurze. Ta ostatnia kwestia jest szczególnie istotna dla dostawców usług w chmurze, którzy hostują współdzielone zasoby IT dla dużej liczby konsumentów usług w chmurze, z których każdy będzie miał wystawione własne gwarancje SLA.

5.2..6 Wskaźniki jakości usług

Umowy SLA wystawiane przez dostawców usług w chmurze to czytelne dla człowieka dokumenty opisujące cechy jakości usług (QoS), gwarancje i ograniczenia jednego lub większej liczby zasobów IT opartych na chmurze.

Umowy SLA wykorzystują metryki jakości usług do wyrażania mierzalnych cech QoS.

Na przykład:

- Dostępność – czas pracy, przestoje, czas trwania usługi
- Niezawodność – minimalny czas pomiędzy awariami, gwarantowany odsetek skutecznych reakcji
- Wydajność – gwarancja wydajności, czasu reakcji i czasu dostawy
- Skalowalność – gwarancja wahań wydajności i responsywności
- Odporność – średni czas do przełączenia i powrotu do zdrowia

Systemy zarządzania SLA wykorzystują te metryki do wykonywania okresowych pomiarów weryfikujących zgodność z gwarancjami SLA, a także do gromadzenia danych związanych z SLA do różnego rodzaju analiz statystycznych.

Każdy miernik jakości usługi jest idealnie zdefiniowany przy użyciu następujących cech:

- Wymierny — jednostka miary jest jasno ustalona, bezwzględna i odpowiednia, aby metryka mogła opierać się na pomiarach ilościowych.
- Powtarzalność – metody pomiaru metryki muszą dawać identyczne wyniki, gdy powtórzyć w identycznych warunkach.
- Porównywalne — jednostki miary stosowane w metryce muszą być ustalane i porównywalne. Na przykład metryka jakości usługi nie może mierzyć mniejszych ilości danych w bitach, a większych w bajtach.
- Łatwo dostępne – metryka musi opierać się na niezastrzeżonej, powszechnie znanej formie pomiaru, który może być łatwo uzyskany i zrozumiałej dla konsumentów chmury.

5.2..6.1 Wskaźniki dostępności usług

Wskaźnik wskaźnika dostępności

Ogólną dostępność zasobów IT wyraża się zwykle jako procent czasu sprawności. Na przykład zawsze dostępny zasób IT będzie miał czas sprawności wynoszący 5,2,0%.

- Opis - procent czasu sprawności usługi
- Pomiar - całkowity czas pracy / całkowity czas
- Częstotliwość - tygodniowa, miesięczna, roczna
- Model dostarczania w chmurze - IaaS, PaaS, SaaS
- Przykład - minimum 55,5% czasu sprawności

Wskaźniki dostępności liczone są kumulatywnie, co oznacza, że okresy niedostępności są łączone w celu obliczenia całkowitego czasu przestoju (Tabela 5.2..1)

Availability (%)	Downtime/Week (Seconds)	Downtime/Month (Seconds)	Downtime/Year (Seconds)
99.5	3024	216	158112
99.8	1210	5174	63072
99.9	606	2592	31536
99.95	302	1294	15768
99.99	60.6	259.2	3154
99.999	6.05	25.9	316.6
99.9999	0.605	2.59	31.5

Tabela 5.2..1 Przykładowe wskaźniki dostępności mierzone w sekundach

Wskaźnik czasu trwania przestoju

Ta metryka jakości usług służy do definiowania zarówno maksymalnych, jak i średnich docelowych poziomów usług w zakresie ciągłych przestojów.

- Opis – czas trwania pojedynczej awarii
- Pomiar - data/godzina zakończenia przestoju - data/godzina rozpoczęcia przestoju
- Częstotliwość - na wydarzenie
- Model dostarczania w chmurze - IaaS, PaaS, SaaS
- Przykład - maksymalnie 1 godzina, średnio 15 minut

5.2..6.2 Wskaźniki niezawodności usług

Cecha ściśle związana z dostępnością; niezawodność to prawdopodobieństwo, że zasób IT może wykonywać zamierzoną funkcję w określonych warunkach bez wystąpienia awarii.

Niezawodność koncentruje się na tym, jak często usługa działa zgodnie z oczekiwaniami, co wymaga, aby usługa pozostawała w stanie operacyjnym i dostępnym. Niektóre wskaźniki niezawodności uwzględniają tylko czas działania

błędy i warunki wyjątkowe jako awarie, które są zwykle mierzone tylko wtedy, gdy zasoby IT są dostępne.

Metryka średniego czasu międzyawaryjnego (MTBF).

- Opis - oczekiwany czas pomiędzy kolejnymi awariami usługi
- Pomiar - \bar{E} , czas trwania normalnego okresu eksploatacji / liczba awarii
- Częstotliwość – miesięczna, roczna
- Model dostarczania w chmurze – IaaS, PaaS
- Przykład – średnia z 50 dni

Metryka wskaźnika niezawodności

Ogólna niezawodność jest trudniejsza do zmierzenia i jest zwykle definiowana za pomocą wskaźnika niezawodności, który reprezentuje procent pomyślnych wyników usług.

Ta metryka mierzy wpływ błędów i awarii niekrytycznych, które występują w okresach przestoju.

Na przykład niezawodność zasobu IT wynosi 5,2,0%, jeśli przy każdym wywołaniu działa zgodnie z oczekiwaniemi, ale tylko 80%, jeśli nie działa co piąty raz.

- Opis – procent pomyślnych wyników usługi w ramach wcześniej zdefiniowanych warunków
- Pomiar – całkowita liczba pomyślnych odpowiedzi / całkowita liczba żądań
- Częstotliwość – tygodniowa, miesięczna, roczna
- Model dostarczania w chmurze – SaaS
- Przykład – minimum 55,5%

5.2.6.3 Wskaźniki wydajności usług

Wydajność usługi odnosi się do zdolności zasobu IT do wykonywania swoich funkcji w ramach oczekiwanych parametrów.

Jakość tę mierzy się za pomocą wskaźników wydajności usług, z których każdy koncentruje się na powiązanej, mierzalnej charakterystyce wydajności zasobów IT. W tej sekcji przedstawiono zestaw typowych wskaźników wydajności. Należy pamiętać, że mogą mieć zastosowanie różne metryki, w zależności od rodzaju mierzonych zasobów IT.

Wskaźnik przepustowości sieci

- Opis – mierzalne cechy przepustowości sieci
- Pomiar – szerokość pasma/przepustowość w bitach na sekundę
- Częstotliwość – ciągła
- Model dostarczania w chmurze - IaaS, PaaS, SaaS
- Przykład - 5.2. MB na sekundę

Wskaźnik pojemności urządzenia pamięci masowej

- Opis – mierzalna charakterystyka pojemności urządzenia magazynującego
- Pomiar – wielkość pamięci w GB
- Częstotliwość – ciągła
- Model dostarczania w chmurze - IaaS, PaaS, SaaS
- Przykład – 80 GB przestrzeni dyskowej

Wskaźnik wydajności serwera

- Opis – mierzalna charakterystyka wydajności serwera • Pomiar – liczba procesorów, częstotliwość procesora w GHz, wielkość pamięci RAM w GB, wielkość pamięci masowej w GB
- Częstotliwość – ciągła • Model dostarczania w chmurze – IaaS, PaaS • Przykład – 1 rdzeń przy 1,7 GHz, 16 GB RAM, 80 GB pamięci

Wskaźnik wydajności aplikacji internetowej

- Opis – mierzalna charakterystyka wydajności aplikacji WWW • Pomiar – liczba żądań na minutę • Częstotliwość – ciągła • Model dostarczania w chmurze – SaaS • Przykład – maksymalnie 5.2.0 000 żądań na minutę

Metryka czasu rozpoczęcia instancji •

- Opis – długość czasu wymagana do zainicjowania nowej instancji • Pomiar – data/godzina aktualizacji instancji – data/godzina żądania uruchomienia • Częstotliwość – na zdarzenie
- Model dostarczania w chmurze – IaaS, PaaS
- Przykład – 5 minut maksymalnie, średnio 3 minuty

Wskaźnik czasu reakcji

- Opis – czas wymagany do wykonania operacji synchronicznej • Pomiar – (data/godzina żądania – data/godzina odpowiedzi) / całkowita liczba żądań • Częstotliwość – codziennie, co tydzień, co miesiąc • Model dostarczania w chmurze – SaaS • Przykład – średnia 5 milisekund

Wskaźnik czasu ukończenia

- Opis – czas potrzebny na wykonanie zadania asynchronicznego • Pomiar – (data żądania – data odpowiedzi) / łączna liczba żądań • Częstotliwość – codziennie, co tydzień, co miesiąc • Model dostarczania w chmurze – PaaS, SaaS • Przykład – średnia 1 sekunda

5.2..6.4 Wskaźniki skalowalności usług Wskaźniki

skalowalności usług są powiązane ze zdolnością do elastyczności zasobów IT, która jest powiązana z maksymalną wydajnością, jaką może osiągnąć zasób IT, a także z pomiarami jego zdolności do adaptacji do wahań obciążenia. Na przykład serwer można skalować do maksymalnie 128 rdzeni procesora i 512 GB pamięci RAM lub skalować w poziomie do maksymalnie 16 replikowanych instancji z równoważeniem obciążenia.

Poniższe wskaźniki pomagają określić, czy dynamiczne wymagania dotyczące usług będą zaspokajane proaktywnie czy reaktywnie, a także określić wpływ ręcznych lub zautomatyzowanych procesów alokacji zasobów IT.

Wskaźnik skalowalności pamięci masowej (poziomej).

- Opis – zmiany dopuszczalnej pojemności urządzenia magazynującego w odpowiedzi na jej zwiększenie obciążenia
- Pomiar – wielkość pamięci w GB
- Częstotliwość – ciągła
- Model dostarczania w chmurze - IaaS, PaaS, SaaS
- Przykład – maksymalnie 1000 GB (skalowanie automatyczne)

Wskaźnik skalowalności serwera (poziomy).

- Opis - dopuszczalne zmiany wydajności serwera w odpowiedzi na zwiększone obciążenie
- Pomiar – liczba serwerów wirtualnych w puli zasobów
- Częstotliwość – ciągła
- Model dostarczania w chmurze – IaaS, PaaS
- Przykład - minimum 1 serwer wirtualny, 5.2. maksimum serwera wirtualnego (automatyczne skalowanie)

Wskaźnik skalowalności serwera (w pionie).

- Opis - dopuszczalne wahanie wydajności serwera w odpowiedzi na wahania obciążenia
- Pomiar – liczba procesorów, wielkość pamięci RAM w GB
- Częstotliwość – ciągła
- Model dostarczania w chmurze – IaaS, PaaS
- Przykład – maksymalnie 512 rdzeni, 512 GB pamięci RAM

5.2..6.5 Wskaźniki odporności usług

Zdolność zasobów IT do regeneracji po zakłóceniami operacyjnymi jest często mierzona za pomocą wskaźników odporności usług. Kiedy odporność jest opisana w ramach gwarancji odporności SLA lub w powiązaniu z nią, często opiera się ona na nadmiarowych wdrożeniach i replikacji zasobów w różnych lokalizacjach fizycznych, a także na różnych systemach odzyskiwania po awarii.

Rodzaj modelu dostarczania w chmurze określa sposób wdrażania i pomiaru odporności. Na przykład fizyczne lokalizacje zreplikowanych serwerów wirtualnych, które wdrażają odporne usługi w chmurze, mogą być wyraźnie wyrażone w umowach SLA dla środowisk IaaS, podczas gdy w sposób dorozumiany mogą być wyrażone w przypadku odpowiednich środowisk PaaS i SaaS.

Wskaźniki odporności można zastosować w trzech różnych fazach, aby rozwiązać problem wyzwania i zdarzenia mogące zagrozić regularnemu poziomowi usługi:

- Faza projektowania – wskaźniki mierzące, jak przygotowane systemy i usługi mają sobie poradzić z wyzwaniem.
- Faza operacyjna — wskaźniki mierzące różnicę w poziomach usług przed, w trakcie i po przestoju lub przerwie w świadczeniu usług, które są dalej kwalifikowane na podstawie wskaźników dostępności, niezawodności, wydajności i skalowalności.
- Faza odzyskiwania — wskaźniki mierzące szybkość odzyskiwania zasobów IT po przestoju, np. czas potrzebny na zarejestrowanie przez system awarii i przejście na nowy serwer wirtualny.

Dwie popularne metryki związane z pomiarem odporności są następujące:

Metryka średniego czasu do przełączenia (MTSO).

- Opis – oczekiwany czas zakończenia przejścia z poważnej awarii na replikowaną instancję w innym obszarze geograficznym
- Pomiar - (data/godzina zakończenia przełączenia - data/godzina awarii) / liczba całkowita niepowodzeń
- Częstotliwość – miesięczna, roczna
- Model dostarczania w chmurze - IaaS, PaaS, SaaS
- Przykład - 5.2. średnia minut

Metryka średniego czasu odzyskiwania systemu (MTSR).

- Opis — oczekiwany czas, w którym odporny system wykona pełne odzyskanie danych z a poważna awaria
- Pomiar – (data/czas naprawy – data/godzina awarii) / całkowita liczba awarii
- Częstotliwość – miesięczna, roczna
- Model dostarczania w chmurze - IaaS, PaaS, SaaS
- Przykład – średnia 120 minut

5.2..7 Wytyczne SLA

W tej sekcji przedstawiono szereg najlepszych praktyk i zaleceń dotyczących pracy z umowami SLA, z których większość ma zastosowanie do konsumentów usług w chmurze:

- Mapowanie uzasadnień biznesowych do umów SLA – pomocne może być zidentyfikowanie niezbędnych wymagań QoS dla danego rozwiązania automatyzacji, a następnie konkretne powiązanie ich z gwarancjami wyrażonymi w umowach SLA dla zasobów IT odpowiedzialnych za realizację automatyzacji. Pozwala to uniknąć sytuacji, w których umowy SLA są przypadkowo źle dopasowane lub być może w nieuzasadniony sposób odbiegają od gwarancji w następstwie wykorzystania zasobów IT.
- Praca z umowami SLA w chmurze i lokalnymi — ze względu na rozległą infrastrukturę dostępną do obsługi zasobów IT w chmurach publicznych, gwarancje QoS określone w umowach SLA dla zasobów IT w chmurze są generalnie lepsze niż te zapewniane dla lokalnych zasobów IT. Tę rozbieżność należy zrozumieć, zwłaszcza podczas tworzenia hybrydowych rozwiązań rozproszonych, które wykorzystują zarówno usługi lokalne, jak i usługi w chmurze, lub gdy włączają architektury technologiczne obejmujące wiele środowisk, takie jak rozrywanie chmury.
- Zrozumienie zakresu umowy SLA — środowiska chmurowe składają się z wielu wspierających warstw architektury i infrastruktury, na których rezydują i są zintegrowane zasoby IT. Ważne jest, aby uwzględnić zakres, w jakim obowiązuje dana gwarancja zasobów IT. Na przykład umowa SLA może ograniczać się do wdrożenia zasobów IT, ale nie do podstawowego środowiska hostingowego.
- Zrozumienie zakresu monitorowania SLA — umowy SLA muszą określać, gdzie przeprowadzane jest monitorowanie i gdzie obliczane są pomiary, przede wszystkim w odniesieniu do zapory sieciowej w chmurze. Na przykład monitorowanie w zaporze chmurowej nie zawsze jest korzystne lub istotne z punktu widzenia gwarancji QoS wymaganych przez konsumenta chmury. Nawet najbardziej wydajne zapory ogniowe mają wymierny wpływ na wydajność i mogą dodatkowo powodować:

punkt awarii.

- Dokumentowanie gwarancji z odpowiednią szczegółowością – szablony SLA używane przez dostawców usług w chmurze czasami definiują gwarancje w sposób ogólny. Jeżeli konsument chmury ma szczególne wymagania, do opisu gwarancji należy zastosować odpowiedni poziom szczegółowości.
Na przykład, jeśli replikacja danych musi odbywać się w określonych lokalizacjach geograficznych, należy je określić bezpośrednio w umowie SLA.
- Określenie kar za nieprzestrzeganie zasad — jeśli dostawca usług w chmurze nie jest w stanie dotrzymać gwarancji QoS obiecanych w umowach SLA, odwołanie może zostać formalnie udokumentowane w postaci odszkodowania, kar, zwrotów kosztów lub w inny sposób.
- Uwzględnianie wymagań niemierzalnych — niektórych gwarancji nie można łatwo zmierzyć za pomocą wskaźników jakości usług, mimo to są one istotne dla QoS i dlatego nadal powinny być udokumentowane w ramach umowy SLA. Na przykład konsument chmury może mieć szczególne wymagania dotyczące bezpieczeństwa i prywatności danych hostowanych przez dostawcę chmury, którym można sprostać poprzez zapewnienia zawarte w umowie SLA dotyczącej dzierżawionego urządzenia do przechowywania danych w chmurze.
- Ujawnianie informacji o weryfikacji zgodności i zarządzaniu — dostawcy usług w chmurze są często odpowiedzialni za monitorowanie zasobów IT w celu zapewnienia zgodności z własnymi umowami SLA. W takim przypadku same umowy SLA powinny określać, jakie narzędzia i praktyki są stosowane w celu przeprowadzenia procesu sprawdzania zgodności, oprócz wszelkich audytów prawnych, które mogą mieć miejsce.
- Włączenie określonych wzorów metryk — niektórzy dostawcy usług w chmurze nie wspominają o powszechnych metrykach SLA ani obliczeniach związanych z metrykami w swoich umowach SLA, zamiast tego skupiają się na opisach poziomu usług, które podkreślają wykorzystanie najlepszych praktyk i obsługi klienta. Metryki używane do pomiaru umów SLA powinny stanowić część dokumentu SLA, łącznie z formułami i obliczeniami, na których opierają się te metryki.
- Rozważenie niezależnego monitorowania SLA – chociaż dostawcy usług w chmurze często posiadają zaawansowane systemy zarządzania SLA i monitory SLA, w najlepszym interesie konsumenta chmury może leżeć wynajęcie organizacji zewnętrznej do przeprowadzenia niezależnego monitorowania, szczególnie jeśli istnieją podejrzenia że gwarancje SLA nie zawsze są wywiązywane przez dostawcę chmury (mimo wyników wykazywanych w okresowo wystawianych raportach monitorujących).
- Archiwizacja danych SLA – Statystyki związane z SLA gromadzone przez SLA monitory są zwykle przechowywane i archiwizowane przez dostawcę usług w chmurze na potrzeby przyszłych raportów. Jeżeli dostawca usług w chmurze zamierza przechowywać dane SLA specyficzne dla konsumenta chmury nawet po tym, jak konsument chmury przestanie kontynuować swoje relacje biznesowe z dostawcą chmury, należy to ujawnić. Konsument chmury może mieć wymagania dotyczące prywatności danych, które uniemożliwiają nieuprawnione przechowywanie tego typu informacji. Podobnie w trakcie i po nawiązaniu współpracy przez konsumenta chmury z dostawcą usług w chmurze może on również chcieć zachować kopię historycznych danych związanych z umową SLA. Może to być szczególnie przydatne w przyszłości do porównywania dostawców usług w chmurze.
- Ujawnianie zależności między chmurami — dostawcy usług w chmurze mogą dzierżawić zasoby IT od innych dostawców usług chmurowych, co skutkuje utratą kontroli nad gwarancjami, którymi są w stanie udostępnić konsumentom usług w chmurze. Chociaż dostawca usług w chmurze będzie polegał na zapewnieniach wynikających z umowy SLA dokonane przez innych dostawców usług w chmurze, konsument chmury może chcieć ujawnienia tego faktu

dzierżawione zasoby IT mogą mieć zależności wykraczające poza środowisko chmury organizacja dostawców.

