

Projektowanie infrastruktury dla obciążeń intensywnie przetwarzających dane

Jak platformy SaaS i API skalują się w chmurze



Skalowanie obciążeń intensywnie przetwarzających dane

W tym materiale wyjaśniamy, jak skalować systemy przetwarzające duże ilości danych. Znajdziesz w nim:

1

Typowe wyzwania pojawiające się wraz z rozwojem systemów

2

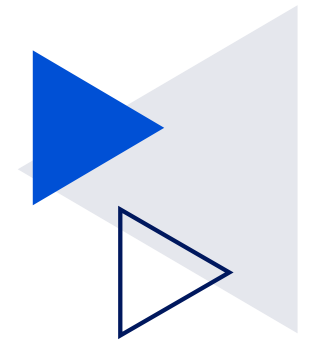
Checklistę ułatwiającą podejmowanie decyzji dotyczących infrastruktury

3

Przykłady wdrożeń, dzięki którym firmy osiągnęły wydajność i zachowały kontrolę nad środowiskiem

Dowiedz się, jak skalować platformy stabilnie, bez zbędnej złożoności i wysokich kosztów.





1. Dowiedz się, jak skalować platformy stabilnie, bez zbędnej złożoności i wysokich kosztów.

Wraz ze wzrostem wolumenu danych, przepustowości i liczby równoległych operacji zespoły coraz częściej mierzą się z wyzwaniami, takimi jak:



Ograniczenia przestrzeni dyskowej, gdy zbiory danych osiągają i przekraczają terabajty



Trudności w przewidywaniu kosztów związanych ze wzrostem ilości danych, operacjami I/O i ruchem wychodzącym



Trudności w utrzymaniu niskich opóźnień podczas nagłych wzrostów ruchu przy przetwarzaniu danych, obsłudze zapytań i wywołań API



Rosnące obciążenie operacyjne wynikające z zarządzania rozproszonymi bazami danych i pipeline'ami



Brak możliwości niezależnego skalowania przestrzeni dyskowej, baz danych i mocy obliczeniowej, co prowadzi do nieefektywnego wykorzystania zasobów



Wymagania dotyczące rezydencji danych i zgodności z przepisami, które zwiększają złożoność architektury systemu



Wąskie gardła sieciowe spowodowane intensywnym ruchem „east-west” między usługami

Skuteczne rozwiązanie tych problemów wymaga infrastruktury, która daje pełną kontrolę nad przechowywaniem, przesyłaniem, przetwarzaniem i skalowaniem danych. Dzięki temu wydajność i koszty pozostają przewidywalne nawet w przypadku intensywnego rozwoju platformy.

2. Checklista ułatwiająca podejmowanie decyzji dotyczących infrastruktury

Zanim wybierzesz platformę chmurową lub zdecydujesz się na konkretny model architektury, sprawdź, jak Twoje obciążenia działają w środowisku produkcyjnym. Duże zbiory danych, intensywne operacje I/O i wysoka liczba zapytań mocno obciążają przestrzeń dyskową, sieć i rozproszone zasoby obliczeniowe.

Poniższa checklista pomaga określić kluczowe parametry techniczne, które warto wziąć pod uwagę przy wyborze dostawcy chmury dopasowanej do potrzeb Twojej firmy.



Przestrzeń dyskowa

- ▶ Wymagany model przechowywania: storage obiektowy, blokowy, oparty na plikach lub hybrydowy
- ▶ Maksymalne zapotrzebowanie na IOPS
- ▶ Wymagania dotyczące NVMe dla obciążeń o niskim opóźnieniu
- ▶ Główny czynnik wpływający na wydajność: przepustowość vs losowe operacje odczytu/zapisu
- ▶ Tempo wzrostu ilości danych (GB/TB/PB)
- ▶ Model replikacji i trwałości danych: jedna strefa, multi-AZ lub wiele regionów



Moc obliczeniowa i przetwarzanie danych

- ▶ Profil obciążenia: zależny od CPU, pamięci lub wspierany przez GPU
- ▶ Tryb przetwarzania danych: wsadowy, strumieniowy lub w czasie rzeczywistym
- ▶ Wymagania dotyczące przetwarzania rozproszonego: klastry Spark, Dask, ClickHouse lub inne
- ▶ Wzorzec wykorzystania zasobów: stałe, skokowe lub oparte na zdarzeniach
- ▶ Konteneryzacja i autoskalowanie



Sieć

- ▶ Poziom wrażliwości na opóźnienia
- ▶ Wymagania dotyczące przepustowości „east-west” między usługami lub węzłami
- ▶ Wymagania dotyczące architektury „multi-zone” lub „multi-region”
- ▶ Wymagania dotyczące prywatnej sieci
- ▶ Spodziewana wielkość ruchu wychodzącego



Skalowanie

- ▶ Model skalowania: poziome (stateless), pionowe (throughput) lub hybrydowe
- ▶ Spodziewane wzrosty ruchu lub obciążenia: codzienne, sezonowe lub związane z określonymi wydarzeniami
- ▶ Wymagania dotyczące autoskalowania, shardingu lub rozbudowy klastra
- ▶ Przystosowanie do konteneryzacji



Koszty i zarządzanie zasobami

- ▶ Przewidywalność wykorzystania zasobów: stabilne vs wysoce zmienne
- ▶ Główne czynniki kosztowe: przechowywanie, obliczenia, ruch wychodzący lub skalowanie bazy danych
- ▶ Preferowany model rozliczeń: opłata miesięczna vs płatność za zużycie
- ▶ Wpływ rosnącej ilości danych i replikacji w object storage na koszty



Zgodność i bezpieczeństwo

- ▶ Wymagania dotyczące rezydencji danych i suwerenności
- ▶ Ograniczenia regulacyjne dotyczące lokalizacji danych
- ▶ Wymagane certyfikaty
- ▶ Wymagania dotyczące szyfrowania, izolacji sieci i integracji IAM
- ▶ Wymagania SLA, oczekiwania klientów oraz regulacje dotyczące prywatności i zgodności



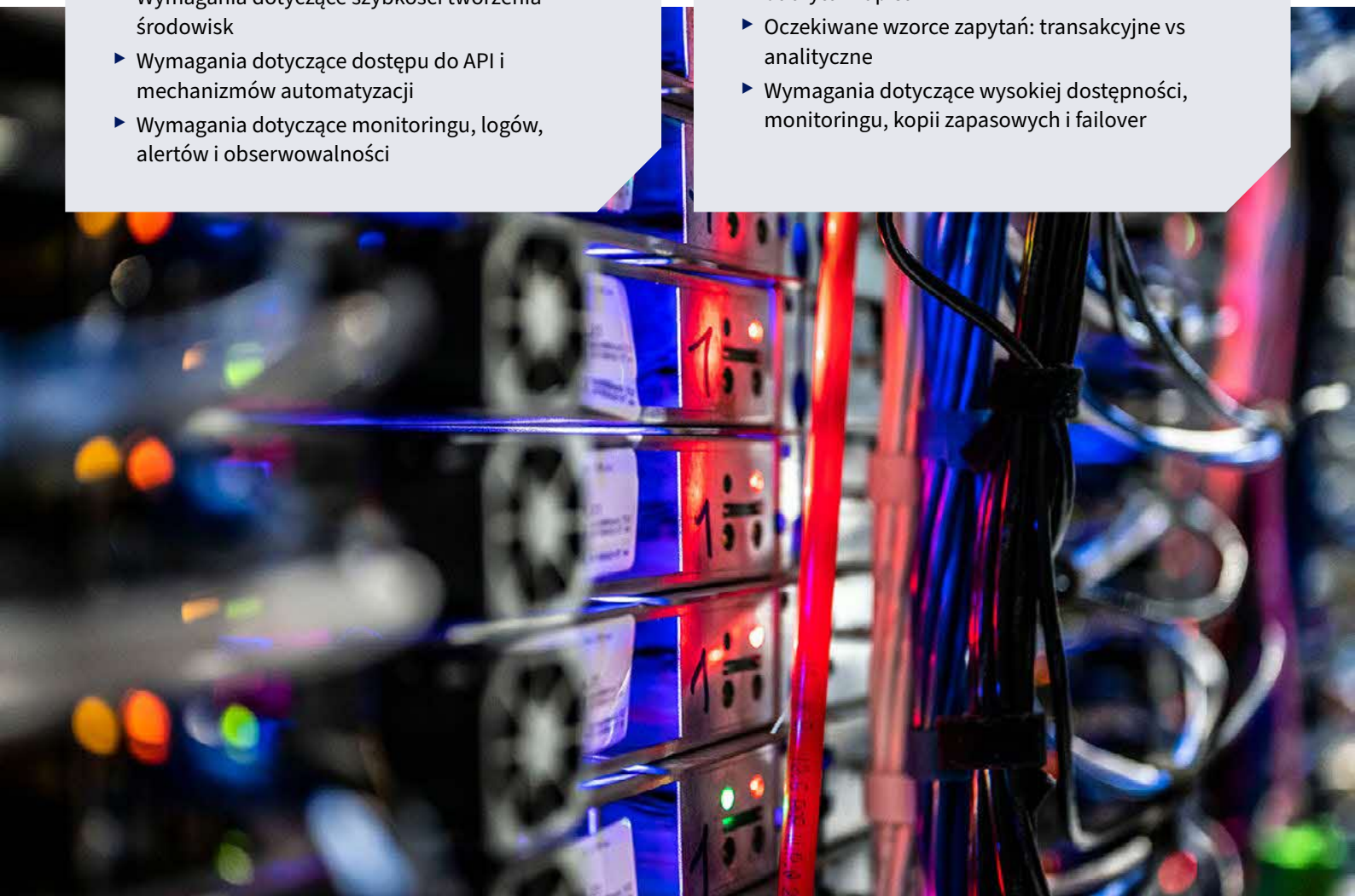
Automatyzacja i wdrażanie

- ▶ Wymagania dotyczące narzędzi Infrastructure as Code
- ▶ Wymagania dotyczące integracji CI/CD
- ▶ Wymagania dotyczące szybkości tworzenia środowisk
- ▶ Wymagania dotyczące dostępu do API i mechanizmów automatyzacji
- ▶ Wymagania dotyczące monitoringu, logów, alertów i obserwowalności



Bazy danych

- ▶ Model bazy danych: dokumentowa (MongoDB) vs relacyjna (PostgreSQL)
- ▶ Wymagania dotyczące równoczesnych operacji odczytu i zapisu
- ▶ Oczekiwane wzorce zapytań: transakcyjne vs analityczne
- ▶ Wymagania dotyczące wysokiej dostępności, monitoringu, kopii zapasowych i failover



Korzyści dla inżynierów

Dobrze zaprojektowane platformy, uwzględniające sposób działania obecnych i przyszłych obciążeń w środowisku produkcyjnym, umożliwiają zwiększenie niezawodności, wydajności i efektywności kosztowej.

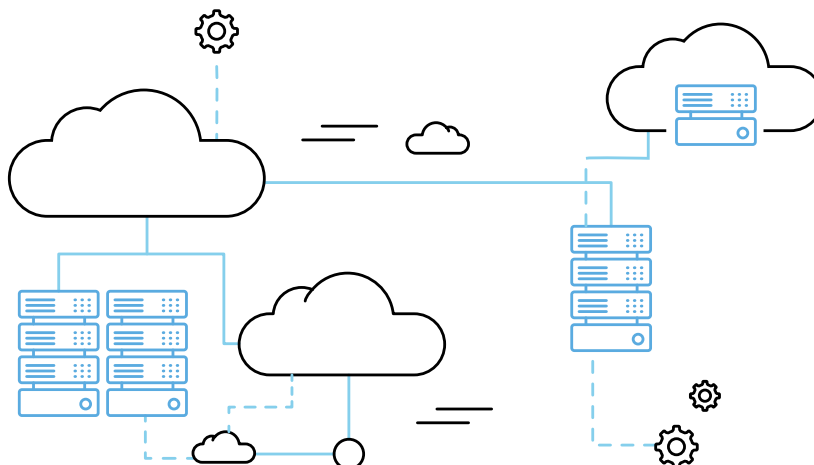
Pozwalają także utrzymać stabilny poziom usług, niezależnie skalować zasoby oraz skutecznie zarządzać usługami i kosztami.

Cecha

- ✓ Stabilna przepustowość podczas ingestu danych
- ✓ Szybsze wykonywanie zapytań i przetwarzanie danych
- ✓ Niezależne skalowanie zasobów
- ✓ Mniejsze obciążenie operacyjne
- ✓ Przewidywalne koszty chmury
- ✓ Kontrola danych i zgodność
- ✓ Otwarta architektura

Korzyści Operacyjne

- ✓ Brak utraty danych, stabilna wydajność
- ✓ Niższe opóźnienia API i krótszy czas analiz
- ✓ Przechyżnia dyskowa, bazy danych i zasoby obliczeniowe skalowane osobno.
- ✓ Automatyzacja (aktualizacje, wysoka dostępność, kopie zapasowe)
- ✓ Brak niespodziewanych opłat za ruch wychodzący lub IOPS
- ✓ Jasne zasady dotyczące lokalizacji i rezydencji danych
- ✓ Brak vendor lock-in, łatwiejsze przenoszenie usług i danych



Projektowanie infrastruktury dla obciążeń AI

Wdrażanie rozwiązań AI powoduje szybki wzrost obciążeń, które wymagają dużej mocy obliczeniowej i znacznej przestrzeni dyskowej. **W związku z tym zespoły potrzebują infrastruktury, która zapewni odpowiednią wydajność, elastyczne modele wdrażania i spójność operacyjną.**

Jednym ze sprawdzonych podejść jest ujednolicenie technologii używanych w środowiskach hybrydowych.

Doskonale sprawdzają się tu platformy takie jak OVHcloud łączące serwery on premises, bare metal i instancje chmurowe, często oparte na wydajnej architekturze AMD EPYC.

Dzięki temu zespoły mogą dobrać **odpowiednią infrastrukturę** do każdego rodzaju obciążenia, korzystać ze znanych narzędzi i zachować wysoką wydajność nawet przy wdrożeniach na dużą skalę.

3. Przykłady wdrożeń: jak firmy skalują usługi SaaS intensywnie przetwarzające dane i API obsługujące duży ruch

Platformy SaaS i API przetwarzające duże ilości danych mają różne wymagania infrastrukturalne, w zależności od ruchu, wolumenu danych i rodzaju obciążeń.

Poniższe przykłady pokazują, jak rozwijające się firmy zachowują wydajność, zwinnie skalują systemy i kontrolują koszty.

Jeśli Twoim wyzwaniem jest...

Zgodność z regulacjami i wysoka dostępność w wielu strefach (multi-AZ)

Przetwarzanie i przechowywanie dużych zbiorów danych

Sprawdź...

iATROS
MapTiler

iATROS: bezpieczna i zgodna z regulacjami platforma e-zdrowia

iATROS oferuje cyfrową platformę medyczną, która zbiera i analizuje dane pacjentów w celu monitorowania ich stanu zdrowia. Aby spełnić rygorystyczne wymagania regulacyjne, w tym RODO i standardy branżowe, oraz zminimalizować

opóźnienia dla użytkowników z różnych regionów, zespół przebudował infrastrukturę, korzystając z bezpiecznej wieloklastrowej chmury o wysokiej dostępności.

Wyzwanie

- ▶ Migracja z dotychczasowego środowiska, które nie spełniało w pełni europejskich wymogów ochrony danych
- ▶ Zapewnienie wysokiej dostępności i niskich opóźnień dla użytkowników z różnych regionów
- ▶ Spełnienie rygorystycznych wymogów zgodności (RODO, ISO, standardy sektora zdrowia)
- ▶ Stworzenie infrastruktury odpornej na awarie

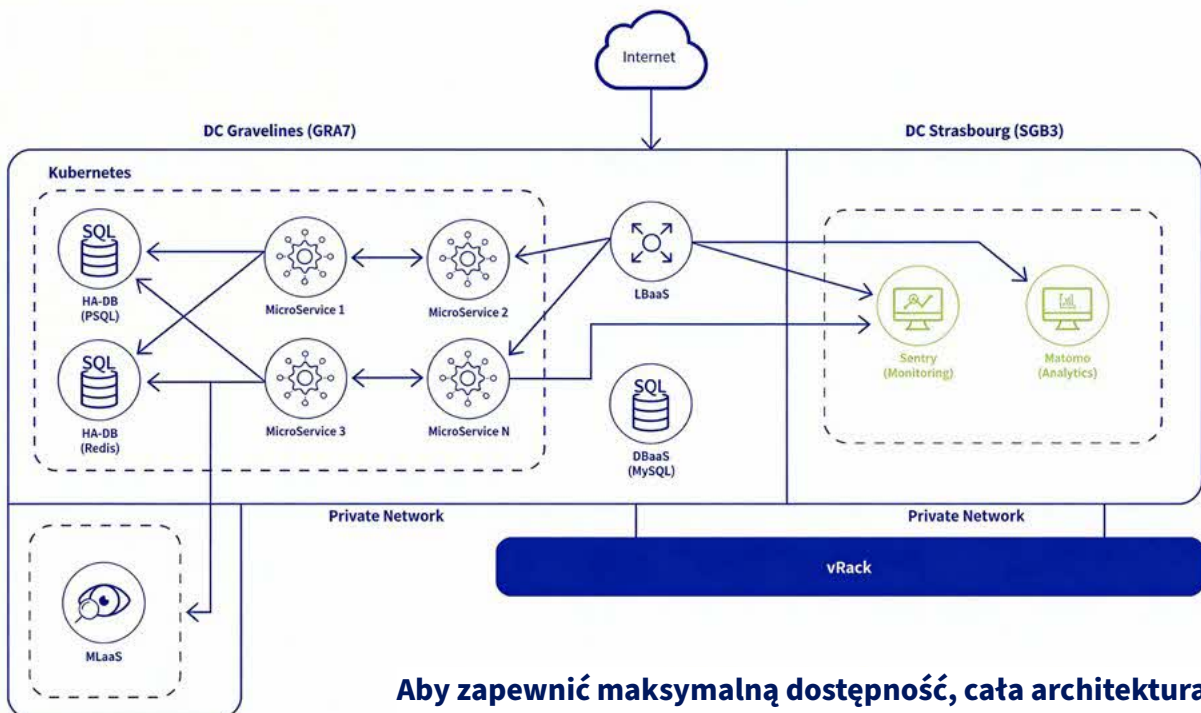
Rozwiązanie

- ▶ Migracja obciążeń do europejskich centrów danych OVHcloud w celu zapewnienia zgodności z RODO
- ▶ Wdrożenie klastrów PostgreSQL o wysokiej dostępności w wielu lokalizacjach (AZ)
- ▶ Wykorzystanie prywatnej sieci vRack do szybkiej komunikacji między systemami („east-west”)
- ▶ Użycie zarządzanych baz danych i skalowalnych instancji obliczeniowych w celu uzyskania stabilnej wydajności

Korzyści

- ▶ Redukcja zużycia zasobów i kosztów o około 20% w porównaniu z poprzednim rozwiązaniem
- ▶ Znaczne zmniejszenie opóźnień, niezależnie od lokalizacji użytkownika
- ▶ Bezpieczne przechowywanie danych, zgodność z RODO i wysoka dostępność

[Dowiedz się więcej](#)



Aby zapewnić maksymalną dostępność, cała architektura jest zdublowana. Oba centra danych monitorują się wzajemnie.

MapTiler: generowanie map satelitarnych na dużą skalę dzięki infrastrukturze z nieograniczoną liczbą instancji cloud

MapTiler to szwajcarska firma, która tworzy mapy oraz spersonalizowane rozwiązania kartograficzne wykorzystywane w różnych branżach, takich jak logistyka, nieruchomości, obrona i turystyka.

MapTiler przetwarza duże ilości danych satelitarnych i obsługuje setki milionów wyświetleń map miesięcznie, dlatego potrzebował chmury, która skaluje się bez wzrostu kosztów i spadku wydajności.

Wyzwanie

- ▶ Szybko rosnące zapotrzebowanie na przestrzeń dyskową w związku ze stałym napływem danych satelitarnych
- ▶ Niezawodna obsługa około 400 milionów zapytań dziennie
- ▶ Eliminacja nieprzewidywalnych kosztów wynikających ze zmiennej ilości danych
- ▶ Skrócenie czasu przetwarzania z wielu lat do kilku tygodni dzięki nieograniczonej mocy obliczeniowej

Rozwiązanie

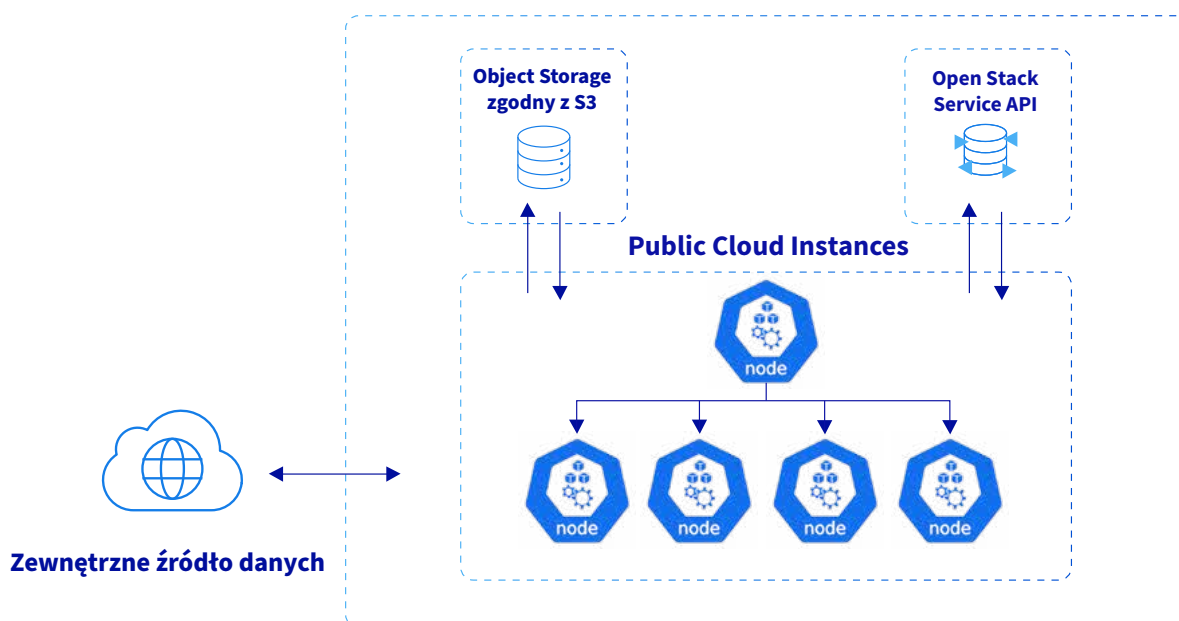
- ▶ Migracja do Public Cloud od OVHcloud ze skalowalnym object storage kompatybilnym z S3
- ▶ Wykorzystanie nieograniczonej liczby instancji cloud do równoległego przetwarzania danych
- ▶ Stały, przewidywalny model kosztów przestrzeni dyskowej i mocy obliczeniowej
- ▶ Uproszczony proces generowania map satelitarnych z wykorzystaniem efektywnej kosztowo infrastruktury

Korzyści

- ▶ Skrócenie czasu przetwarzania danych satelitarnych z ponad 18 lat do kilku tygodni
- ▶ Przewidywalne miesięczne koszty niezależnie od zmiennych obciążeń
- ▶ Brak ograniczeń mocy obliczeniowej, umożliwiając szybszą produkcję map

[Dowiedz się więcej](#)

Infrastruktura chmurowa MapTiler w OVHcloud



*S3 jest zarejestrowanym znakiem towarowym Amazon Technologies, Inc. Usługi OVHcloud nie są sponsorowane ani zatwierdzone przez Amazon Technologies, Inc. i nie są z nią w żaden sposób powiązane.

Skuteczne skalowanie obciążeń intensywnie przetwarzających dane

Wzrost ilości danych, liczby użytkowników i usług zwiększa złożoność utrzymania systemów. Bez dobrze zaprojektowanej architektury system traci wydajność, generuje coraz wyższe koszty i wymaga większego nakładu pracy operacyjnej.

Dlatego warto pomyśleć o infrastrukturze, która umożliwia niezależne skalowanie przestrzeni dyskowej, mocy obliczeniowej, baz danych i sieci. Dzięki temu zespół nie musi zajmować się bieżącym utrzymaniem infrastruktury - obsługa przestrzeni dyskowej, orkiestracja i monitoring są zautomatyzowane w ramach zarządzanych usług.

Wybierając długoterminową skalowalność, zachowujesz kontrolę nawet wtedy, gdy:

- ▶ ilość danych rośnie 10- lub 100-krotnie
- ▶ przybywa usług i pipeline'ów,
- ▶ zwiększa się częstotliwość awarii

Elastyczna chmura upraszcza rozwój systemów i pozwala realizować ambitne projekty bez utraty stabilności.

Skaluj usługi SaaS intensywnie przetwarzające dane i API obsługujące duży ruch. Sprawdź, na czym polega przewaga OVHcloud.

Chcesz dowiedzieć się więcej?

Umów się na rozmowę
z architektem rozwiązań:

[Proszę o kontakt](#)

Odkryj skalowalną chmurę
dla rozwijających się firm:

[Dowiedz się więcej](#)