

Architektura struktury danych zapewnia 3 natychmiastowe korzyści

Zintegruj i wykorzystaj wszystkie dane
swojej organizacji, aby uzyskać lepsze
wyniki biznesowe

Spis treści

- 3** Współczesne wyzwania związane z danymi
- 5** Korzyści biznesowe ze struktury danych
- 6** Jak IBM wdraża strukturę danych
- 7** Struktura danych w akcji
- 7** Struktura danych to kolejny krok w kierunku transformacji Twojego przedsiębiorstwa

Współczesne wyzwania związane z danymi

Dane są integralnym elementem cyfrowej transformacji przedsiębiorstw.

Jednak w miarę jak organizacje starają się wykorzystywać swoje dane, napotykają na wyzwania wynikające z różnorodności źródeł, typów, struktur, środowisk i platform danych. Problem wielowymiarowych danych staje się jeszcze bardziej złożony, gdy organizacje decydują się na hybrydowe i wielochmurowe architektury. W wielu przedsiębiorstwach dane operacyjne pozostają w dużej mierze odizolowane i ukryte, co prowadzi do gromadzenia się ogromnej ilości tzw. ciemnych danych (z ang. dark data), czyli danych pozornie nieistotnych.

Weźmy za przykład północnoamerykańską firmę energetyczną, która chciała przekształcić się w organizację opartą na danych - stać się firmą, w której analiza wielkich zbiorów danych byłaby łatwo dostępna w wielu jednostkach biznesowych. Szybko wyszło na jaw, że transformacja cyfrowa była utrudniona przez odizolowane dane, niespójne narzędzia i różne poziomy umiejętności, co powodowało krytyczne luki w kompetencjach w zakresie danych.

Problem, z którym się zmierzili, nie był unikalny dla tej firmy; jest to bardzo powszechne zjawisko, gdy środowiska danych przerastają swoje architektury zarządzania danymi.

Owa firma energetyczna wróciła na ścieżkę transformacji cyfrowej dzięki zastosowaniu nowej koncepcji architektury danych, znanej jako struktura danych.

Czym dokładnie jest struktura danych, czym różni się od poprzednich architektur, jakie korzyści może zapewnić i jaka jest rola IBM w jej wdrażaniu? Niniejsze opracowanie rozwieje wszelkie wątpliwości.

Czym jest struktura danych?

W przeszłości organizacje radziły sobie z problemami związanymi z dostępem do danych poprzez integrację typu punkt do punktu lub wprowadzenie koncentratorów danych. Żaden z tych sposobów nie jest odpowiedni, gdy dane są odizolowane i rozproszone w dużej mierze. Integracje typu punkt do punktu powodują wzrost kosztów w przypadku konieczności podłączenia dodatkowego punktu końcowego, co oznacza, że jest to podejście nieskalowalne. Koncentratory danych umożliwiają łatwiejszą integrację aplikacji i źródeł, ale zwiększają koszty i złożoność procesu w celu zachowania jakości i zaufania do danych w obrębie koncentratora.

Struktura danych to kształtująca się architektura, która ma na celu sprostanie wyzwaniom związanym z danymi w obrębie hybrydowego środowiska danych. Podstawową ideą jest osiągnięcie równowagi pomiędzy decentralizacją a globalizacją poprzez działanie w charakterze wirtualnej warstwy komunikacji między punktami końcowymi danych (patrz strona 4).

Dzięki technologiom takim jak automatyzacja i zwiększenie integracji, zarządzanie stowarzyszone oraz aktywacja metadanych architektura struktury danych umożliwia dynamiczną i inteligentną koordynację danych w środowisku rozproszonym, tworząc sieć informacji dostępnych od razu, aby zwiększyć potencjał firmy.

Struktura danych jest niezależna od platform wdrożeniowych, procesów danych, lokalizacji geograficznych i podejścia architektonicznego. Ułatwia korzystanie z danych jako aktywów przedsiębiorstwa. Struktura danych umożliwia sprawne i efektywne łączenie, udostępnianie i zarządzanie różnymi rodzajami danych.

Funkcje i zasady dotyczące struktury danych

Podstawą architektury struktury danych jest platforma zarządzania danymi, która umożliwia pełen zakres zintegrowanych funkcji zarządzania danymi, w tym ich pozyskiwanie, zarządzanie, porządkowanie oraz koordynację.

Jednakże struktura danych wyłania się i rozwija z tradycyjnych koncepcji zarządzania danymi, takich jak DataOps, która skupia się jedynie na ustaleniu praktyk w celu zwiększenia poziomu operacjonalizacji danych. Jest tworzona w oparciu o rozproszoną architekturę i zaawansowaną technologię, które są w stanie sprostać potrzebom wynikającym z ogromnej różnorodności i dystrybucji zasobów informacyjnych.

Strukturę danych można logicznie podzielić na cztery funkcje (lub komponenty):

Źródła danych i użytkownicy

Wiedza, spostrzeżenia i semantyka

- Udostępnianie rynku danych i kształtowanie doświadczenia zakupowego
- Automatyczne wzbogacanie odkrytych zasobów informacyjnych o wiedzę i semantykę, umożliwiając konsumentom znalezienie i zrozumienie danych

Ujednolicone zarządzanie i zgodność z przepisami

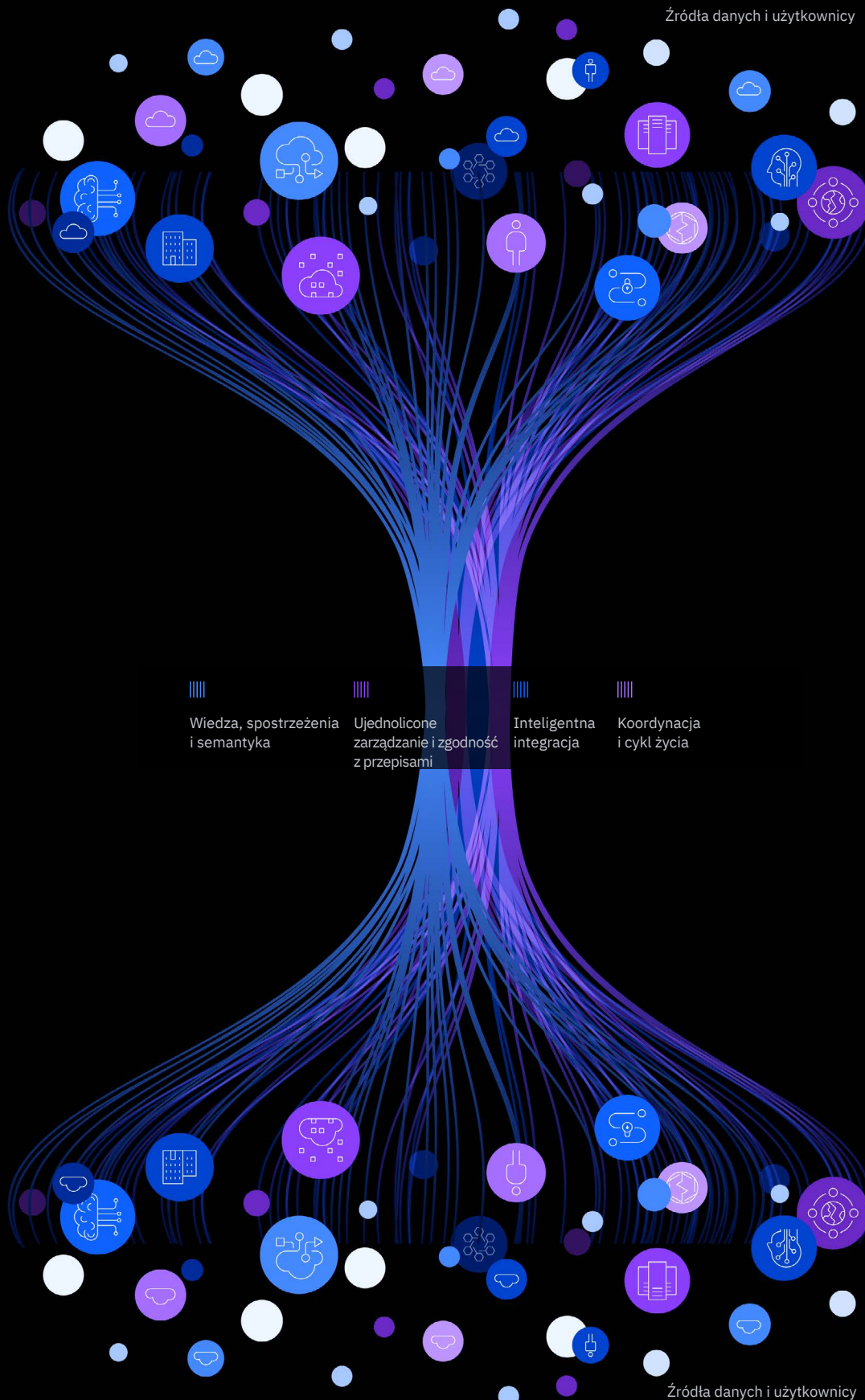
- Lokalne zarządzanie metadanymi i nadzór nad nimi oraz obsługa ujednoliconego widoku globalnego i ułatwianie egzekwowania przepisów
- Automatyczne wdrażanie polityk w zakresie zasobów informacyjnych zgodnie z globalnymi i lokalnymi przepisami
- Zastosowanie zaawansowanych funkcji w celu automatyzacji klasyfikowania i porządkowania zasobów informacyjnych
- Automatyczne tworzenie ścieżek dostępu dla dowolnych skatalogowanych zasobów w celu zwiększenia aktywacji danych

Inteligentna integracja

- Szybsza realizacja zadań inżyniera danych poprzez zautomatyzowany przepływ i tworzenie potoku z różnych źródeł danych rozproszonych
- Samoobsługowe wczytywanie danych i dostęp do nich w przypadku dowolnych danych objętych lokalnymi i globalnymi politykami w zakresie ochrony danych
- Automatyczne określenie najlepszego sposobu działania poprzez zoptymalizowaną dystrybucję obciążeń oraz samodostrajanie i korektę odstępstw schematu

Koordinacja i cykl życia

- Tworzenie, testowanie, obsługa i monitorowanie potoków danych
- Wdrażanie funkcji sztucznej inteligencji do cyklu życia danych w celu automatyzacji zadań, samodostrajania, samonaprawy i wykrywania zmian danych źródłowych, co ułatwia automatyczne aktualizacje



Źródła danych i użytkownicy

Korzyści biznesowe ze struktury danych

Dane nabierają wartości biznesowej tylko wtedy, gdy są skontekstualizowane i dostępne dla każdego użytkownika lub aplikacji w organizacji. Prawidłowo wdrożona struktura danych pomaga zapewnić dostępność tych wartości w całej organizacji w najbardziej efektywny i zautomatyzowany sposób. W związku z tym struktura danych zapewnia trzy kluczowe korzyści:

1. Samoobsługowa konsumpcja danych oraz współpraca z dostępem do nich
2. Automatyzacja nadzoru, ochrony i bezpieczeństwa; możliwa dzięki aktywnym metadansom
3. Automatyzacja zadań inżynierii danych i poprawa integracji danych w hybrydowych zasobach chmurowych

Samoobsługowa konsumpcja danych oraz współpraca z dostępem do nich

Integracja danych z wielu źródeł i analiza większej części ogromnej ilości danych generowanych każdego dnia umożliwiają organizacjom wyciąganie lepszych wniosków i szybsze reagowanie na zmieniające się potrzeby biznesowe. Struktura danych szybko dostarcza dane tam, gdzie są potrzebne. Samoobsługa umożliwia całej organizacji szybsze znalezienie odpowiednich danych i poświęcenie większej ilości czasu na wykorzystanie tych danych w celu uzyskania konkretnych informacji.

Korzyści ze struktury danych w zakresie samoobsługowej konsumpcji danych:

- Użytkownicy biznesowi mają jeden punkt dostępu, aby znaleźć, zrozumieć, ukształtować i wykorzystać dane w całej organizacji.
- Scentralizowany nadzór nad danymi i ich pochodzenie pomaga użytkownikom zrozumieć, co oznaczają dane, skąd pochodzą i jak są powiązane z innymi zasobami.
- Łatwe skalowanie rozbudowanego i konfigurowalnego zarządzania metadansom, dostępne za pośrednictwem interfejsów API
- Samoobsługowy dostęp do zaufanych i nadzorowanych danych umożliwia współpracę z innymi użytkownikami na poziomie pionu operacyjnego.

Badanie całkowitego wpływu ekonomicznego przeprowadzone przez firmę Forrester¹ wykazało, że owe funkcje mogą oznaczać:

mIn USD

5,8 459%

zysków

zwrot z inwestycji

Automatyzacja nadzoru, ochrony i bezpieczeństwa danych; możliwa dzięki aktywnym metadansom

Rozproszona, aktywna warstwa nadzoru nad wszystkimi inicjatywami związanymi z danymi zmniejsza ryzyko w zakresie zgodności z przepisami i regulacjami, zapewniając zaufanie i przejrzystość. Umożliwia automatyczne egzekwowanie przepisów w zakresie każdorazowego dostępu do danych, zapewniając wysoki poziom ochrony danych i zgodności z przepisami.

Wykorzystanie technologii sztucznej inteligencji i uczenia maszynowego pozwala użytkownikom struktury danych zwiększyć poziom automatyzacji, na przykład automatycznie wyodrębniając przepisy dotyczące nadzoru nad danymi w oparciu o język i definicje zawarte w dokumentach regulacyjnych. Dzięki temu organizacje mogą w ciągu kilku minut zastosować się do przepisów nadzoru specyficznych dla danej branży, co pozwala uniknąć kosztownych kar i zapewnić etyczne wykorzystanie danych niezależnie od tego, gdzie się znajdują.

Korzyści ze struktury danych w zakresie nadzorowanej wirtualizacji:

- Większa elastyczność, bezpieczeństwo i wydajność pracy inżynierów danych, analityków wielkich zbiorów danych i analityków biznesowych.
- Wiele globalnych źródeł danych wyświetla się jako jedna baza danych.
- Nowe, wiodące w branży odkrywanie informacji umożliwiających identyfikację osób (z ang. personally identifiable information - PII) oraz elementów danych o znaczeniu newralgicznym jest możliwe na masową skalę.

Owe funkcje mogą oznaczać:

mIn USD

2,4 430%

zysków¹

większa wydajność²

Automatyzacja zadań inżynierii danych i lepsza integracja danych

Zaawansowana inżynieria danych oznacza, że praktycznie każdy proces dostępu do danych lub ich dostarczania jest zautomatyzowany i nie wymaga żmudnego lub podatnego na błędy procesu kodowania. Lepsza integracja poprzez wykorzystanie metadansom w celu optymalizacji dostarczania danych i dostępu do nich.

Korzyści ze struktury danych w zakresie inżynierii i integracji danych:

- Automatycznie optymalizowana integracja danych pomaga przyspieszyć dostarczanie danych.
- Automatyczne równoważenie obciążenia i elastyczne skalowanie oznacza, że zadania mogą być wykonywane w każdym środowisku z wykorzystaniem dowolnego wolumenu danych.
- Wbudowana odporność i automatyzacja narzędzi CI/CD.
- Zautomatyzowany proces przechwytywania zmian w czasie rzeczywistym ułatwia dostarczanie wysokiej jakości danych dla procesów biznesowych.
- Uczenie maszynowe może zautomatyzować i rozbudować niestandardowe procesy pozyskiwania, klasyfikowania i porządkowania danych, co prowadzi do szybszego odnoszenia korzyści.
- Ciągła analiza może być wykonywana automatycznie w czasie rzeczywistym, bez względu na to, gdzie znajdują się dane.

Wyniki czołowego sprzedawcy detalicznego³ są następujące:

60x

szybsze dostarczanie danych

20x

szybsze analizy koszykowe klientów

Jak IBM wdraża strukturę danych

IBM Cloud Pak® for Data umożliwia realizację koncepcji struktury danych. IBM Cloud Pak for Data to platforma Insight, która upraszcza i automatyzuje gromadzenie, organizację i analizę danych oraz przyspiesza wdrażanie sztucznej inteligencji w przedsiębiorstwie.

Dzięki możliwości łączenia danych i uruchamiania obciążeń w dowolnym miejscu oraz budowania, wdrażania i zarządzania sztuczną inteligencją na dużą skalę w środowiskach chmur hybrydowych, IBM Cloud Pak for Data jest narzędziem służącym do cyfrowej transformacji przedsiębiorstwa.

Platforma zapewnia bezproblemową integrację w ramach przedsiębiorstwa hybrydowego dla:

- Usługi w ramach IBM Cloud Pak for Data
- Aplikacje zewnętrzne i źródła danych
- Zaawansowane funkcje zarządzania danymi i nadzoru nad nimi w oparciu o sztuczną inteligencję

Platforma ta udostępnia konsumentom uporządkowane dane są z zachowaniem optymalnej równowagi pomiędzy kosztami, wydajnością i zgodnością z przepisami, umożliwiając także koordynację i optymalizację przetwarzania danych w oparciu o obciążenia, lokalizację danych i przepisy dotyczące danych.

IBM Cloud Pak for Data wykorzystuje następujące funkcje, aby dostarczyć dane gotowe do użytku biznesowego, wymagane w ramach struktury danych. Wszystkie te funkcje odgrywają ważną rolę we wspieraniu architektury struktury danych.

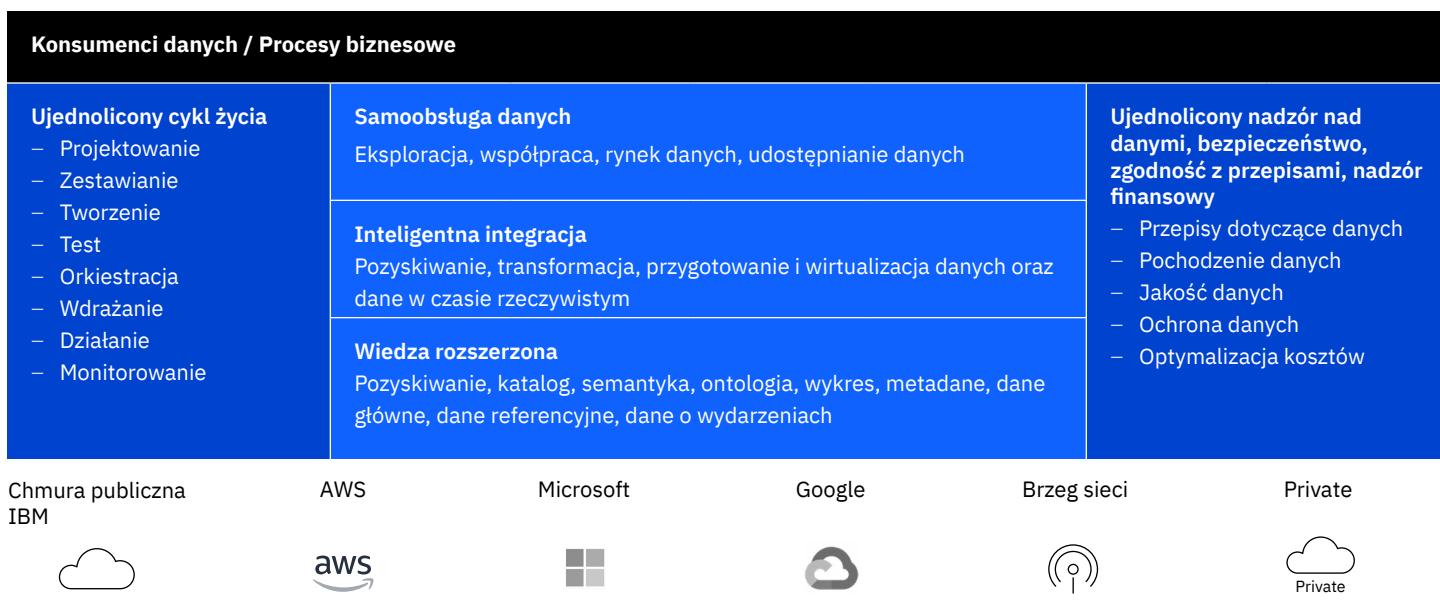
1. Centrum wiedzy oparte na metadanych Ułatwia pozyskiwanie źródeł danych i katalogów, rozbudowuje zasoby danych i przeprowadza analizy, aby przy użyciu sztucznej inteligencji wydobycić wartościową wiedzę i w ten sposób zwiększyć automatyzację. Centrum wiedzy jest wykorzystywane do napędzania rynku za pomocą wyszukiwania semantycznego.

2. Samoobsługowy rynek danych Katalog danych nowej generacji, który pomaga konsumentom danych, takim jak analitycy biznesowi, pobierać dane ze środowiska danych przedsiębiorstwa.

3. Inteligentna integracja Umożliwia konsumpcję danych poprzez wyodrębnianie, wirtualizację, transformację i strumieniowanie danych. Jest ona połączona z centrum wiedzy, aby zautomatyzować integrację danych i wyróżnia się zdolnością wyboru najlepszej strategii integracji w oparciu o obciążenia i przepisy dotyczące danych. Może być również używana do przygotowywania danych w ramach zadań inżynierii danych lub do tworzenia produktów do zarządzania danymi. Umożliwia także publikowanie aktualizacji produktów do zarządzania danymi.

4. Nadzór Kataloguje i porządkuje metadane, definiuje politykę ochrony prywatności danych, porządkuje dane, rejestruje ich pochodzenie oraz wykonuje inne zadania związane z bezpieczeństwem i zgodnością z przepisami. Warstwa ta rozumie format danych (np. ustrukturyzowane a nieustrukturyzowane) oraz ich istotność (np. publiczne a chronione) i stosuje odpowiednie przepisy wobec każdego fragmentu danych i każdego potencjalnego użytkownika. Zamiast ręcznego zastosowania standardów i reguł wobec danych, dzięki możliwościom zintegrowanej platformy można je stosować na poziomie organizacyjnym i propagować w różnych zasobach danych w zależności od potrzeb. Modele analityczne w różnych narzędziach mogą się ze sobą komunikować; precyzyjne egzekwowanie przepisów dotyczących danych można w dużym stopniu zautomatyzować.

5. Ujednolicony rozwój i operacje Umożliwia ujednolicony cykl życia w celu skonfigurowania i uruchomienia wszystkich aspektów platformy danych w produkcji.



Wykres 1. Funkcje platformy IBM Cloud Pak for Data - wszystkie wspierają rozwój sztucznej inteligencji i struktury danych

6. Struktura danych w akcji

Najlepszym sposobem na zrozumienie wartości struktury danych jest poznanie korzyści biznesowych, które organizacje odniosły dzięki wdrożeniu struktury danych na platformie Insight.

Przedsiębiorstwo energetyczne

Na początku niniejszego opracowania, w rozdziale poświęconym wyzwaniom przedstawiono przypadek północnoamerykańskiego przedsiębiorstwa energetycznego przechodzącego transformację cyfrową.

Firma nawiązała współpracę z IBM, który wdrożył architekturę struktury danych w oparciu o platformę IBM Cloud Pak for Data.

Dzięki elastycznej i zintegrowanej strukturze danych firmie udało się wdrożyć szereg ważnych projektów związanych z danymi w różnych jednostkach, w tym:

- eMobility
- Pozyskiwanie dokumentów dotyczących eksploatacji gazu, w tym wyodrębnianie pisma ręcznego
- Segmentacja konsumentów energii elektrycznej i prognozowanie obciążenia
- Zarządzanie zasobami
- Wpływ COVID-19 na obciążenie
- Model ryzyka dla powrotu do pracy

Dzięki strukturze danych owa organizacja może:

- Zapewnić bezpośredni dostęp do danych wielu jednostkom biznesowym za pośrednictwem samoobsługowej platformy Insight.
- Wykorzystać dopasowaną do potrzeb moc obliczeniową, aby wydajnie uruchamiać modele w miliardach rzędów danych.
- Zespołowo opracowywać modele i łatwo je wdrażać w celu wzbogacenia analiz w firmie.

Zmniejszenie ruchu i lepszy nadzór

Duży dostawca usług przemysłowych usprawnił nadzór nad danymi i ułatwił zapewnienie zgodności z przepisami.

Owa organizacja zmagająca się z problemem przenoszenia dużych ilości danych do opartego na chmurze jeziora danych. Miała podwójne wymagania dotyczące zastosowania skutecznej warstwy nadzoru wobec każdorazowego dostępu do danych oraz określenia jakości danych przed umożliwieniem użytkownikom biznesowym dostępu do nich.

Po wdrożeniu struktury danych zbudowanej w oparciu o IBM Cloud Pak for Data organizacja ogromnie usprawniła zarządzanie danymi, zachowanie zgodności z przepisami oraz procesy transformacji danych. Dzięki strukturze danych może zapewnić użytkownikom biznesowym łatwy i zabezpieczony dostęp do setek źródeł danych w jej chmurowym jeziorze danych oraz lokalnych źródeł danych SAP.

Wykorzystuje również branżowe narzędzia regulacyjne do skanowania źródeł danych w poszukiwaniu potencjalnych danych PII, które podlegałyby regulacjom RODO i CCPA. Automatyczne oznaczanie danych pomaga zidentyfikować źródła danych, które wymagają zarządzania w zakresie przetwarzania danych PII zarówno lokalnie, jak i w chmurze.

Dzięki lepszemu dostępowi i szybszej weryfikacji danych PII użytkownicy biznesowi mogą wydobywać istotne dane w celu uzyskania ważnych informacji bez długiego oczekiwania na dostęp do danych lub ryzyka ujawnienia danych chronionych.

Dzięki strukturze danych owa organizacja może:

- Zapewnić właściwy nadzór nad danymi, jednocześnie wykorzystując dane z całej firmy.
- Uzyskać wiarygodne dane i zmniejszyć czas przygotowywania danych.
- Ułatwić przestrzeganie przepisów w zakresie ochrony prywatności, takich jak RODO i CCPA.

Struktura danych to kolejny krok w kierunku transformacji Twojego przedsiębiorstwa

Dzięki strukturze danych zbudowanej w oparciu o technologię IBM Cloud Pak for Data użytkownik może w pełni zautomatyzować pozyskiwanie danych, nadzór nad danymi oraz konsumpcję danych w środowisku hybrydowym i wielochmurowym. Wykorzystanie struktury danych umożliwi szybsze odniesienie korzyści przez użytkowników biznesowych, większą produktywność w zakresie inżynierii i obsługi operacyjnej danych oraz większą dokładność w zakresie nadzoru i zachowania zgodności.

[Na początek wypróbuj IBM Cloud Pak for Data za darmo.](#)

[Zaplanuj bezpłatną, indywidualną konsultację z przedstawicielem IBM ds. danych i sztucznej inteligencji.](#)



© Copyright IBM Corporation 2021

IBM Polska Sp. z o.o.
ul. Krakowiaków 32
02-255 Warszawa
Wyprodukowano w Stanach Zjednoczonych
Maj 2021

IBM, logo IBM i IBM Cloud Pak są znakami towarowymi lub zastrzeżonymi znakami towarowymi International Business Machines Corporation w Stanach Zjednoczonych i/lub innych krajach. Nazwy innych produktów lub usług mogą być znakami towarowymi IBM lub innych podmiotów. Aktualna lista znaków towarowych IBM dostępna jest pod adresem ibm.com/trademark.

Niniejszy dokument jest aktualny na dzień jego pierwszej publikacji i może zostać zmieniony przez IBM w dowolnym momencie. Nie wszystkie produkty są dostępne we wszystkich krajach, w których IBM prowadzi działalność.

Przytoczone dane o wynikach i przykłady wykorzystania przez klientów mają charakter poglądowy. Rzeczywiste wyniki mogą być różne w zależności od konkretnej konfiguracji i warunków użytkowania. Odpowiedzialność za ocenę i weryfikację współdziałania dowolnych innych produktów i programów z produktami i programami IBM ponosi użytkownik. INFORMACJE ZAWARTE W TYM DOKUMENCIE SĄ DOSTARCZANE W STANIE, W JAKIM SIĘ ZNAJDUJĄ („AS IS”), BEZ JAKICHKOLWIEK GWARANCJI (REKOJMIA JEST NINIEJSZYM RÓWNIEŻ WYŁĄCZONA), WYRAŻNYCH LUB DOMNIEMANYCH, A W SZCZEGÓLNOŚCI DOMNIEMANYCH GWARANCJI PRZYDATNOŚCI HANDLOWEJ, GWARANCJI PRZYDATNOŚCI DO OKREŚLONEGO CELU ORAZ GWARANCJI NIENARUSZANIA PRAW OSÓB TRZECICH. Produkty IBM podlegają gwarancjom zgodnym z warunkami umów, na mocy których są dostarczane.

Oświadczenie o sprawdzonych procedurach zabezpieczeń: Zabezpieczenia systemów informatycznych obejmują ochronę systemów i informacji poprzez wykrywanie przypadków nieuprawnionego dostępu do systemu przedsiębiorstwa z zewnątrz i wewnątrz, zapobieganie takim przypadkom oraz reagowanie na nie. Niewłaściwy dostęp może spowodować zmodyfikowanie lub zniszczenie informacji, ich niewłaściwe użycie lub wykorzystanie w niedozwolony sposób. Może również spowodować zniszczenie systemów lub ich niewłaściwe wykorzystanie, w tym do przeprowadzenia ataku na inne podmioty. Żaden system lub produkt informatyczny nie może być uważany za w pełni bezpieczny. Żaden produkt, usługa ani metoda zabezpieczająca nie chroni całkowicie przed nieuprawnionym dostępem do systemu przedsiębiorstwa lub jego niewłaściwym użyciem. Systemy, produkty i usługi IBM zostały zaprojektowane jako część zgodnego z prawem, kompleksowego modelu bezpieczeństwa, w który zostaną włączone dodatkowe procedury operacyjne. Osiągnięcie przez ten model maksymalnej efektywności może wymagać wykorzystania innych systemów, produktów lub usług. IBM NIE GWARANTUJE, ŻE JAKIEKOLWIEK SYSTEMY, PRODUKTY LUB USŁUGI SĄ ZABEZPIECZONE LUB ZABEZPIECZĄ PRZEDSIĘBIORSTWO KLIENTA PRZED SZKODLIWYMI LUB NIEZGODNYMI Z PRAWEM DZIAŁANAMI JAKICHKOLWIEK OSÓB.

Klient odpowiada za zapewnienie zgodności z wszelkimi obowiązującymi przepisami prawa i regulacjami. IBM nie udziela porad prawnych oraz nie deklaruje ani nie gwarantuje, że usługi lub produkty IBM zapewnią zgodność działań przedsiębiorstwa Klienta z przepisami.

- 1 Forrester - Nowa technologia: Badanie całościowego wpływu ekonomicznego przeprowadzone przez firmę Forrester na zlecenie IBM, luty 2020
- 2 Pomiary wydajności zostały przeprowadzone w kontrolowanym środowisku testowym w laboratoriach IBM w Dolinie Krzemowej (Stany Zjednoczone) przy użyciu wirtualizacji danych IBM w odniesieniu do różnych źródeł danych o pojemności 100 TB. Pomiary wykonane w maju 2019 roku i wzrost wydajności są porównywane do federacji IBM.
- 3 Na podstawie doświadczeń klienta firmy IBM.