

Witamy w słowniku Generation Data – naszej alfabetycznej liście definicji pojęć dotyczących danych

ACID Test

ACID - akronim od słów: atomicity (niepodzielność), consistency (spójność), isolation (izolacja) i durability (trwałość). Rodzaj testu, w którym te cztery atrybuty gwarantują poprawność przetwarzania transakcji w bazie danych.

Administrator bazy danych (DBA)

Osoba, często certyfikowana, odpowiedzialna za wspieranie i utrzymanie integralności struktury i zawartości bazy danych.

Agregacja danych

Akt polegający na zbieraniu danych z wielu źródeł w celu raportowania lub analizy.

Algorytm

Formuła matematyczna lub process statystyczny prowadzony przez oprogramowanie w celu dokonania analizy danych. Zwykle składa się z wielu kroków i może być wykorzystany do zautomatyzowania procesu przetwarzania danych lub rozwiązywania problemów.

Amazon Web Services

Zbiór usług w chmurze oferowanych przez Amazon, aby pomóc firmom w przeprowadzaniu operacji obliczeniowych na dużą skalę (takich jak projekty dużych zbiorów danych) bez konieczności inwestowania we własne farmy serwerów i magazyny danych. Przestrzeń dyskowa, moc obliczeniowa i operacje na oprogramowaniu są wynajmowane, zamiast być kupowane i instalowane od zera.

Analityka

Proces gromadzenia, przetwarzania i analizowania surowych danych z masy danych w celu generowania informacji, które wspierają podejmowanie decyzji na podstawie faktów. W wielu

przypadkach wiąże się to z analizą dokonywaną przez oprogramowanie przy wykorzystaniu algorytmów.

Analiza eksploracyjna

Znajdowanie wzorców w danych bez standardowych procedur lub metod. Jest to sposób na odkrywanie danych i znajdowanie głównych cech zestawów danych.

Analiza lokalizacji

Analiza lokalizacji zapewnia mapowanie i analizę opartą na mapach w systemach biznesowych przedsiębiorstw i magazynach danych. Pozwala powiązać informacje geoprzestrzenne ze zbiorami danych.

Analiza porównawcza

Prowadzona krok po kroku procedura porównań i obliczeń, której celem jest wykrycie wzorców w bardzo dużych zestawach danych.

Analitka predykcyjna

Korzystanie z funkcji statystycznych w jednym lub większej liczbie zestawów danych do przewidywania trendów lub przyszłych zdarzeń.

Analiza przestrzenna

Odnosi się do analizy danych przestrzennych, takich jak dane geograficzne lub dane topologiczne, w celu zidentyfikowania i zrozumienia wzorców i prawidłowości w danych rozproszonych w przestrzeni geograficznej.

Analiza przyczyn

Proces określania głównej przyczyny zdarzenia lub problemu.

Analiza sentymentów

Zastosowanie funkcji statystycznych do komentarzy, które ludzie czynią w Internecie lub za pośrednictwem sieci społecznościowych, w celu ustalenia, co sądzą o produkcie lub firmie.

Analiza skupień

Proces identyfikowania obiektów, które są do siebie podobne i grupowania ich w celu zrozumienia różnic, a także podobieństw w danych.

Analiza tekstu

Zastosowanie technik statystycznych, językowych i uczenia maszynowego w źródłach tekstowych w celu uzyskania znaczenia lub wglądu.

Anonimizacja danych

Usuwanie powiązań pomiędzy ludźmi znajdującymi się w bazie danych i wpisami ich dotyczącymi w celu uniknięcia możliwości wykrycia źródła danych.

Architektura i projektowanie danych

Struktura danych korporacyjnych. Rzeczywista struktura lub konstrukcja różni się w zależności od ostatecznego wymaganego rezultatu końcowego. Architektura danych składa się z trzech etapów: reprezentacja koncepcyjna podmiotów biznesowych, logiczna reprezentacja relacji między tymi jednostkami oraz fizyczna konstrukcja systemu do obsługi funkcjonalności.



Badacz danych

Ekspert w wydobywaniu informacji i wartości z danych. Zwykle ktoś, kto ma umiejętności w zakresie analityki, informatyki, matematyki, statystyki, kreatywności, wizualizacji i komunikacji danych, a także biznesu i strategii.

Baza danych

Cyfrowy zbiór danych i struktura, wokół której są zorganizowane dane. Dane są zazwyczaj wprowadzane i dostępne za pośrednictwem systemu zarządzania bazą danych (DBMS).

Bazy danych XML

Bazy danych XML umożliwiają przechowywanie danych w formacie XML. Bazy danych XML są często połączone z bazami danych zorientowanymi na dokumenty. Dane przechowywane w bazie danych XML mogą być wyszukiwane, eksportowane i zapisywane w dowolnym wymaganym formacie.

Biometria

Wykorzystanie analiz i technologii do identyfikacji osób na podstawie jednej lub wielu cech fizycznych, takich jak rozpoznawanie odcisków palców, rozpoznawanie twarzy, rozpoznawanie tęczy i tęczy itp.

Call Detail Record (CDR)

Informacje CDR zawierają dane gromadzone przez firmę telekomunikacyjną na temat połączeń telefonicznych, takie jak czas i długość połączenia. Dane te można wykorzystać w dowolnej liczbie analitycznych zastosowań.

Centrum danych

Miejsce, które mieści dużą liczbę serwerów i urządzeń do przechowywania danych. Centra danych mogą należeć do jednej organizacji lub sprzedawać swoje usługi wielu organizacjom.

Chmura obliczeniowa

Szeroki termin odnoszący się do dowolnej aplikacji internetowej działającej na zdalnych serwerach, a nie lokalnie. Dane przechowywane „w chmurze” są zazwyczaj dostępne przez Internet i można się do nich dostać z dowolnego miejsca na świecie.

Cold data storage

Przechowywanie starych danych, które są rzadko używane, na serwerach o niskiej mocy. Pobieranie takich danych trwa dłużej.

Czas oczekiwania

Wszelkie opóźnienia w odpowiedzi lub dostarczeniu danych z jednego punktu do drugiego.

Czyszczenie danych

Proces przeglądania i weryfikacji danych w celu usuwania duplikatów, poprawiania błędów i zapewniania spójności.

Dane częściowo ustrukturyzowane

Dane, które nie są ustrukturyzowane przez formalny model danych, ale zapewniają inne sposoby opisu danych i hierarchii.

Dane generowane maszynowo

Wszelkie dane tworzone automatycznie z procesu komputerowego, aplikacji lub innego źródła innego niż człowiek.



Dane lokalizacji

Dane opisujące lokalizację geograficzną.

Dane publiczne

Informacje publiczne lub zestawy danych, które zostały utworzone przy pomocy środków publicznych.

Dane referencyjne

Dane opisujące obiekt i jego właściwości. Obiekt może być fizyczny lub wirtualny.

Dane transakcyjne

Dane, które zmieniają się nieprzewidywalnie. Przykłady obejmują dane dotyczące zobowiązań i należności lub dane o wysyłkach produktów.

Dane uboczne

Dane, które osoba tworzy jako produkt uboczny wspólnej działalności - na przykład dziennik połączeń komórkowych lub historia wyszukiwania w Internecie.

Dane ustrukturyzowane vs Dane nieustrukturyzowane

Dane ustrukturyzowane to wszystko, co może zostać umieszczone w tabeli i zorganizowane w taki sposób, że odnosi się do innych danych w tej samej tabeli. Nieustrukturyzowane dane to wszystko, co nie może być w taki sposób potraktowane - na przykład wiadomości e-mail, posty w mediach społecznościowych i nagrane wypowiedzi ludzkie.

Dane w czasie rzeczywistym

Dane tworzone, przetwarzane, przechowywane, analizowane i wizualizowane w ciągu milisekund.

Dane wyszukiwania

Zagregowane dane o wyszukiwanych hasłach w określonym czasie.

Dark Data

Wszystkie dane gromadzone i przetwarzane przez przedsiębiorstwa, które nie są wykorzystywane do żadnych istotnych celów. Nazywa się je ciemnymi, ponieważ nie są używane i badane. Obejmują one kanały społecznościowe, dzienniki call center, notatki ze spotkań itp.

Data Lake

Repozytorium pamięci masowej, które może przechowywać ogromną ilość surowych danych w oryginalnym formacie. Data Lake używa płaskiej architektury do przechowywania danych, w przeciwieństwie do hierarchicznej magazynów, które przechowują dane w plikach lub folderach.

Data Science

Dyscyplina obejmująca statystyki, wizualizację danych, programowanie komputerowe, eksplorację danych, uczenie maszynowe i inżynierię baz danych w celu rozwiązywania złożonych problemów.

Depozytariusz danych

Osoba odpowiedzialna za strukturę bazy danych i środowisko techniczne, w tym za przechowywanie danych.

Deska rozdzielcza

Graficzna reprezentacja analiz przeprowadzonych przez algorytmy.

Projekt został zrealizowany przy wsparciu finansowym Komisji Europejskiej. Niniejsza publikacja odzwierciedla jedynie poglądy autora, a Komisja nie ponosi odpowiedzialności za jakiegokolwiek wykorzystanie zawartych w niej informacji.



Co-funded by the
Erasmus+ Programme
of the European Union

Dostęp do danych

Akt lub metoda przeglądania lub odzyskiwania zmagazynowanych

Ekonometria

Zastosowanie teorii statystycznych i matematycznych w ekonomii do testowania hipotez i prognozowania przyszłych trendów. Ekonometria korzysta z modeli ekonomicznych, testuje je poprzez próby statystyczne, a następnie porównuje wyniki z rzeczywistymi przykładami. Można ją podzielić na dwie główne kategorie: teoretyczną i stosowaną.

Eksploracja danych

Proces uzyskiwania wzorców lub wiedzy z dużych zbiorów danych. Celem jest przekształcenie danych w bardziej zrozumiałą i spójny zestaw informacji.

ETL (Extract, transform and load)

Proces wykorzystywany w magazynach danych do przygotowania danych do wykorzystania w raportach lub analizach. Akronim oznaczający wyodrębnianie, przekształcanie i ładowanie. Odnosi się do procesu „wydobycia” surowych danych, „przekształcania” przez czyszczenie/wzbogacanie danych w celu „przydatności do użycia” i „ładowania” do odpowiedniego repozytorium do przyszłego wykorzystania.

Hadoop

Apache Hadoop to jedna z najczęściej używanych platform programistycznych w big data. Jest to zbiór programów, które umożliwiają przechowywanie, pobieranie i analizę bardzo dużych zestawów danych przy użyciu rozproszonego sprzętu (umożliwiając rozłożenie danych na wiele mniejszych urządzeń pamięci masowej zamiast na jednym bardzo dużym).

Integracja danych

Proces łączenia danych z różnych źródeł i prezentowania ich w jednym widoku.

Integralność danych

Miara zaufania do organizacji pod kątem dokładności, kompletności, terminowości i ważności danych.

Internet rzeczy

Urządzenia podłączone do Internetu w dowolnym czasie i dowolnym miejscu za pomocą czujników, które zbierają, analizują i przesyłają dane w celu zwiększenia ich przydatności.

Jakość danych

Miara danych określająca ich przydatność do podejmowania decyzji, planowania lub operacji.

Kanał danych

Sposób, w jaki osoba może otrzymać strumień danych. Przykłady takich kanałów obejmują RSS lub Twitter.

Magazyn danych

Miejsce do przechowywania danych do celów raportowania i analizy.

MapReduce

Procedura oprogramowania polegająca na rozbiciu analizy na części, które można rozdzielić na różne komputery w różnych lokalizacjach. Najpierw dystrybuje analizę (map), a następnie zbiera wyniki z powrotem do jednego raportu (reduce). Kilka firm, w tym Google i Apache (jako część platformy Hadoop), udostępnia narzędzia MapReduce.

Projekt został zrealizowany przy wsparciu finansowym Komisji Europejskiej. Niniejsza publikacja odzwierciedla jedynie poglądy autora, a Komisja nie ponosi odpowiedzialności za jakiegokolwiek wykorzystanie zawartych w niej informacji.



Co-funded by the
Erasmus+ Programme
of the European Union

Mashup

Metoda łączenia różnych zestawów danych w jedną aplikację w celu poprawy wyników. Przykładem może być łączenie ofert pracy z danymi demograficznymi.

Metadane

Dane, które służą do zapewnienia kontekstu lub dodatkowych informacji o innych danych. Na przykład informacje o tytule, temacie, autorze, kroju czcionki, ulepszeniach i rozmiarze pliku danych dokumentu stanowią metadane dotyczące niniejszego dokumentu. Mogą one również opisywać warunki, w jakich dane zgromadzone w bazie danych zostały pozyskane, ich dokładność, datę, godzinę, metodę kompilacji i przetwarzania itp.

Migracja danych

Proces przenoszenia danych między różnymi typami lub formatami pamięci lub między różnymi systemami komputerowymi.

Model danych, modelowanie danych

Model danych definiuje strukturę danych w celu komunikowania się między osobami funkcjonalnymi i technicznymi aby pokazać dane potrzebne do procesów biznesowych lub w celu przedstawienia planu opracowania sposobu przechowywania danych i uzyskiwania do nich dostępu między członkami zespołu tworzącego aplikacje.

Nauczanie maszynowe

Zastosowanie algorytmów pozwalających komputerowi analizować dane w celu „uczenia się”, jakie działania należy podjąć, gdy wystąpi określony wzorzec lub zdarzenie.

NoSQL

Systemy zarządzania bazami danych, które nie używają (lub nie tylko) tabel relacyjnych zwykle używanych w tradycyjnych systemach baz danych. Nazwa odnosi się do systemów przechowywania i pobierania danych, które są zaprojektowane do obsługi dużych ilości danych, ale bez tabelarycznej kategoryzacji (lub schematów).

Obiektowe bazy danych

Przechowują dane w postaci obiektów używanych przez programowanie obiektowe. Różnią się one od relacyjnych lub graficznych baz danych i większość z nich oferuje język zapytań, który pozwala na znalezienie obiektu przy deklaratywnym podejściu do programowania.

Objętość

Ilość danych, od megabajtów do brontobajtów.

Ochrona danych

Praktyka ochrony danych przed zniszczeniem lub nieuprawnionym dostępem.

Operacyjne bazy danych

Operacyjne bazy danych przeprowadzają regularne operacje organizacji i są na ogół bardzo ważne dla firmy. Zasadniczo korzystają z przetwarzania transakcji online, które pozwala im wprowadzać, zbierać i pobierać określone informacje o firmie.

Oprogramowanie jako usługa (SAAS)

Rosnąca tendencja producentów oprogramowania do dostarczania swoich programów w chmurze - co oznacza, że użytkownicy płacą za czas spędzony na korzystaniu z nich (lub za ilość danych, do których mają dostęp), a nie za bezpośrednie kupowanie oprogramowania.



Parsowanie

Podział danych, takich jak ciąg, na mniejsze części do analizy.

Plik dziennika

Plik dziennika jest zdefiniowany jako plik, który prowadzi rejestr zdarzeń, procesów, wiadomości i komunikacji między różnymi aplikacjami komunikacyjnymi a systemem operacyjnym. Pliki dziennika są obecne w wykonywalnym oprogramowaniu, systemach operacyjnych i programach, w których zapisywane są wszystkie komunikaty i szczegóły procesu.

Profilowanie danych

Proces zbierania statystyk i informacji o danych w istniejącym źródle.

Przechowywanie

Wszelkie sposoby trwałego przechowywania danych.

Przetwarzanie języka naturalnego

Algorytmy programowe zaprojektowane w celu umożliwienia komputerom dokładniejszego rozumienia codziennej ludzkiej mowy, umożliwiając nam bardziej naturalną i wydajną interakcję z nimi.

Przetwarzanie rozproszone

Wykonanie procesu na wielu komputerach połączonych siecią komputerową.

Przetwarzanie wsadowe

Standardowa strategia obliczeniowa polegająca na przetwarzaniu danych w dużych zestawach. Ta praktyka staje się niezbędna w przypadku pracy, która nie jest wrażliwa na czas i działa na bardzo dużych zestawach danych. Proces jest zaplanowany, a później wyniki są pobierane przez system.

Punkt danych

Pojedynczy element na wykresie lub wykresie.

R

Popularne środowisko oprogramowania typu open source wykorzystywane do analiz. R zapewnia szeroką gamę technik statystycznych (modelowanie liniowe i nieliniowe, klasyczne testy statystyczne, analizę szeregów czasowych, klasyfikację, grupowanie) i technik graficznych i jest wysoce rozszerzalny.

Replikacja danych

Proces udostępniania informacji w celu zapewnienia spójności między redundantnymi źródłami.

Repozytorium danych

Lokalizacja trwale przechowywanych danych.

Rozproszony system plików

System przechowywania danych zaprojektowany do przechowywania dużych ilości danych na wielu urządzeniach pamięci (często serwerach opartych na chmurze), aby zmniejszyć koszty i złożoność przechowywania dużych ilości danych.

Równoważenie obciążenia

Proces dystrybucji obciążenia w sieci komputerowej lub klastrze komputerowym w celu optymalizacji wydajności.



Schemat

Struktura definiująca organizację danych w systemie bazy danych.

Serwer

Komputer fizyczny lub wirtualny, który obsługuje żądania aplikacji i dostarcza je przez sieć.

Silnik rekomendacji

Algorytm analizujący zakupy i działania klienta w witrynie e-commerce, a następnie wykorzystujący te dane do rekomendowania produktów uzupełniających.

Skalowalność

Zdolność systemu lub procesu do utrzymania akceptowalnego poziomu wydajności wraz ze wzrostem obciążenia lub zakresu.

Structured Query Language (SQL)

Język programowania zaprojektowany specjalnie do zarządzania danymi w systemie relacyjnych baz danych i ich wyszukiwania.

Struktura danych

Konkretny sposób przechowywania i organizowania danych.

System zarządzania bazą danych (DBMS)

Oprogramowanie, które gromadzi i zapewnia dostęp do danych w ustrukturyzowanej formie.

System zastany

Nieaktualne systemy komputerowe, języki programowania lub aplikacje, które są używane zamiast dostępnych uaktualnionych wersji. Systemy zastane są również kojarzone z terminologią lub procesami, które nie mają już zastosowania do bieżących kontekstów lub treści, co powoduje zamieszanie.

Sztuczna inteligencja

Proces opracowywania inteligentnych maszyn i oprogramowania, które mogą postrzegać środowisko i podejmować odpowiednie działania w razie potrzeby, a nawet uczyć się na podstawie tych działań.

Tag

Tag jest informacją opisującą dane lub treść, do której jest przypisany. Tagi to niehierarchiczne słowa kluczowe używane do zakładek internetowych, obrazów cyfrowych, filmów, plików itd.

Taksonomia

Taksonomia odnosi się do klasyfikacji danych zgodnie z wcześniej ustalonym systemem z wynikowym katalogiem. Zapewnia koncepcyjne ramy dla łatwego dostępu i wyszukiwania.

Wizualizacja danych

Wizualna projekcja danych zaprojektowana w celu lepszego zrozumienia znaczenia lub skuteczniejszego przekazywania informacji.

Zapytanie

Żądanie danych lub informacji z tabeli bazy danych lub kombinacji tabel. Dane te mogą być generowane jako wyniki zwracane przez Structured Query Language (SQL) lub jako obrazki, wykresy lub złożone wyniki, np. analizy trendów z narzędzi do eksploracji danych. SQL jest najbardziej znanym i najczęściej używanym językiem zapytań.



Zarządca danych

Osoba odpowiedzialna za dane przechowywane w polu danych

Zarządzanie danymi

Zestaw procesów lub reguł zapewniających integralność danych i spełniających najlepsze praktyki zarządzania danymi.

Zbieranie danych

Każdy proces, który przechwytuje dowolny typ danych.

Zbiór danych

Zbiór danych, zwykle w formie tabelarycznej.

Zgodność danych prawnych

W terminologii prawnej słowo „Zgodność” odnosi się do aktu przestrzegania prawa kraju. W kategoriach biznesowych, w przypadku każdej organizacji, zgodność oznacza ściśle przestrzeganie przepisów, wytycznych i specyfikacji, które są istotne dla cyklu życia jednostki biznesowej. Zgodność danych prawnych jest powszechnie stosowana w kontekście rozwiązań chmurowych, w których dane są przechowywane w innym kraju lub kontynencie. Przechowywanie danych na serwerze lub centrum danych zlokalizowanym w obcym kraju musi być zgodne z przepisami bezpieczeństwa danych danego kraju.

Źródło danych

Każdy dostawca danych - na przykład baza danych lub strumień danych.

