

**Velkommen til Generation Datas ordliste -
vores alfabetisk liste over termer eller ord, der findes i det specifikke dataemne,
med korte forklaringer.**

Algoritme

En matematisk formel eller statistisk proces kørt af software til at udføre en analyse af data. Det består normalt af flere beregningstrin og kan bruges til automatisk at behandle data eller løse problemer.

Anonymisering

Afbrydelsen af forbindelser mellem mennesker i en database og deres optegnelser for at forhindre opdagelsen af kilden til optegnelserne.

Analytics

Processen med at indsamle, behandle og analysere rådata fra en datamasse for at generere indsigt, der informerer faktabaseret beslutningstagning. I mange tilfælde involverer det softwarebaseret analyse ved hjælp af algoritmer.

Kunstig intelligens

Processen med at udvikle efterretningsmaskiner og software, der kan opfatte miljøet og tage den tilsvarende handling efter behov, og endda lære af disse handlinger.

SURETEST

Står for atomicitet, konsistens, isolation og holdbarhed (ACID) test af data. Disse fire attributter er benchmarks for at sikre gyldigheden af en datatransaktion.

Amazon Web Services

En samling af cloud computing-tjenester, der tilbydes af Amazon for at hjælpe virksomheder med at udføre computere i stor skala (såsom big data-projekter) uden at skulle investere i deres egen server

gårde og datalagringslagre. Opbevaringsplads, processorkraft og softwarefunktioner lejes i stedet for at skulle købes og installeres fra bunden.

Batchbehandling

En standard computerstrategi, der involverer behandling af data i store sæt. Denne praksis bliver afgørende for ikke-tidsfølsomt arbejde, der fungerer på meget store datasæt. Processen er planlagt, og på et senere tidspunkt hentes resultaterne af systemet.

Biometri

Brug af analyse og teknologi til at identificere mennesker ved en eller mange af deres fysiske egenskaber, såsom fingeraftryksgenkendelse, ansigtsgenkendelse, irisgenkendelse osv.

Call Detail Record (CDR) analyse

CDR'er indeholder data, som et telekommunikationsselskab indsamler om telefonopkald, f.eks. Tid og længde på opkaldet. Disse data kan bruges i et vilkårligt antal analytiske applikationer.

Skyen

Et bredt udtryk, der refererer til ethvert internetbaseret program, der kører på eksterne servere, snarere end lokalt. Data gemt "i skyen" er typisk tilgængelige via internettet, uanset hvor i verden ejeren af disse data måtte være.

Klyngeanalyse

Processen med at identificere objekter, der ligner hinanden og gruppere dem for at forstå forskellene såvel som lighederne i dataene.

Opbevaring af kolde data

Lagring af gamle data, der næppe bruges på servere med lav effekt. Det vil tage længere tid at hente dataene.

Sammenlignende analyse

Det sikrer en trinvis procedure for sammenligninger og beregninger for at opdage mønstre inden for meget store datasæt.

Dashboard

En grafisk gengivelse af analyserne udført af algoritmer.

Dataadgang

Handlingen eller metoden til visning eller hentning af lagrede data.

Data aggregering

Handlingen med at indsamle data fra flere kilder med henblik på rapportering eller analyse.

Dataarkitektur og design

Hvordan virksomhedsdata er struktureret. Den faktiske struktur eller design varierer afhængigt af det eventuelle slutresultat, der kræves. Dataarkitektur har tre faser: konceptuel repræsentation af forretningsenheder, den logiske repræsentation af forholdet mellem disse enheder og den fysiske konstruktion af systemet til understøttelse af funktionaliteten.

Database

En digital indsamling af data og strukturen, omkring hvilken dataene er organiseret. Dataene indtastes typisk og åbnes via et databasestyringssystem (DBMS).



Database administrator (DBA)

En person, ofte certificeret, der er ansvarlig for at understøtte og opretholde integriteten af strukturen og indholdet af en database.

Database management system (DBMS)

Software, der indsamler og giver adgang til data i et struktureret format.

Datacenter

Et fysisk anlæg, der huser et stort antal servere og datalagringsenheder. Datacentre kan tilhøre en enkelt organisation eller sælge deres tjenester til mange organisationer.

Datarensning

Processen med at gennemgå og revidere data for at slette dubletter, rette fejl og give konsistens.

Dataindsamling

Enhver proces, der fanger enhver form for data.

Opbevaring af data

En person, der er ansvarlig for databasestrukturen og det tekniske miljø, herunder lagring af data.

Data udstødning

De data, som en person opretter som et biprodukt af en fælles aktivitet - for eksempel en celleopkaldslog eller websøgningshistorik.

Datafeed

Et middel for en person at modtage en strøm af data. Eksempler på datafeedsmekanismer inkluderer RSS eller Twitter.

Datastyring

Et sæt processer eller regler, der sikrer dataintegriteten, og at den bedste praksis for datastyring overholdes.

Dataintegration

Processen med at kombinere data fra forskellige kilder og præsentere dem i en enkelt visning.

Dataintegritet

Det mål for tillid, som en organisation har i nøjagtigheden, fuldstændigheden, aktualiteten og gyldigheden af dataene.

Data Lake

Et lageropbevaringssted, der kan rumme en enorm mængde rådata i dets oprindelige format. Data Lake bruger en flad arkitektur til at gemme data i modsætning til et hierarkisk datalager, der gemmer data i filer eller mapper.

Datamigrering

Processen med at flytte data mellem forskellige lagringstyper eller formater eller mellem forskellige computersystemer.

Data minedrift

Processen med at udlede mønstre eller viden fra store datasæt. Formålet er at forfine data til et mere forståeligt og sammenhængende sæt information.



Datamodel, datamodellering

En datamodel definerer datastrukturen med det formål at kommunikere mellem funktionelle og tekniske personer for at vise data, der er nødvendige for forretningsprocesser, eller for at kommunikere en plan for at udvikle, hvordan data lagres og åbnes blandt applikationsudviklingsholdets medlemmer.

Datapunkt

Et individuelt element på en graf eller et diagram.

Dataprofilering

Processen med at indsamle statistik og information om data i en eksisterende kilde.

Datakvalitet

Mål for data for at bestemme dets værdighed til beslutningstagning, planlægning eller drift.

Datareplikering

Processen med at dele information for at sikre sammenhæng mellem overflødige kilder.

Datalager

Placeringen af permanent lagrede data.

Data videnskabsmand

En ekspert i at udvinde indsigt og værdi fra data. Normalt en person, der har færdigheder inden for analyse, datalogi, matematik, statistik, kreativitet, datavisualisering og kommunikation samt forretning og strategi.

Datavidenskab

En disciplin, der inkorporerer statistik, datavisualisering, computerprogrammering, datamining, maskinindlæring og databaseteknik for at løse komplekse problemer.

Datasikkerhed

Praksis med at beskytte data mod ødelæggelse eller uautoriseret adgang.

Datasæt

En samling af data, typisk i tabelform.

Datakilde

Enhver udbyder af data - for eksempel en database eller en datastrøm.

Datastyrer

En person, der er ansvarlig for data gemt i et datafelt.

Datastruktur

En bestemt måde at lagre og organisere data på.

Datavisualisering

En visuel abstraktion af data designet med det formål at udlede mening eller kommunikere information mere effektivt.

Data varehus

Et sted at gemme data med henblik på rapportering og analyse.



Mørke data

Alle data, der indsamles og behandles af virksomheder, der ikke bruges til meningsfulde formål. Det kaldes mørkt, fordi det er ubrugt og uudforsket. Dette inkluderer feeds fra sociale netværk, logcentraler, mødedokumenter osv.

Distribueret filsystem

Datalagringssystem designet til at gemme store datamængder på tværs af flere lagerenheder (ofte skybaserede handelsservere) for at reducere omkostningerne og kompleksiteten ved lagring af store datamængder.

Distribueret behandling

Udførelsen af en proces på tværs af flere computere forbundet med et computernetværk.

Økonometri

Anvendelsen af statistiske og matematiske teorier i økonomi til test af hypoteser og forudsigelse af fremtidige tendenser. Econometrics bruger økonomiske modeller, tester dem gennem statistiske forsøg og sammenligner derefter resultaterne med virkelige eksempler. Det kan opdeles i to hovedkategorier: teoretisk og anvendt.

Uddrag, transformer og indlæs (ETL)

En proces, der bruges i datalager til at forberede data til brug i rapportering eller analyse. Forkortelsen for ekstrakt, transformation og belastning. Det refererer til processen med at 'udtrække' rådata', 'transformere' ved at rense / berige dataene til 'egnet til brug' og 'indlæse' i det passende lager til fremtidig brug.

Undersøgende analyse

Finde mønstre inden for data uden standardprocedurer eller metoder. Det er et middel til at opdage dataene og finde datasættets vigtigste egenskaber.

Hadoop

Apache Hadoop er en af de mest anvendte software-rammer i big data. Det er en samling af programmer, der muliggør opbevaring, hentning og analyse af meget store datasæt ved hjælp af distribueret hardware (så dataene kan spredes på mange mindre lagerenheder snarere end en meget stor).

Tingenes internet

Almindelige enheder, der er tilsluttet internettet til enhver tid hvor som helst via sensorer, der indsamler, analyserer og transmitterer data for at øge deres anvendelighed.

Juridisk dataoverensstemmelse

I den juridiske terminologi henviser ordet "Overholdelse" til overholdelse af landets lov. Forretningsmæssigt betyder overholdelse i tilfælde af enhver organisation nøje overholdelse af de love, regler, retningslinjer og specifikationer, der er relevante for en forretningsenheds livscyklus. Juridisk dataoverholdelse bruges ofte i sammenhæng med skybaserede løsninger, hvor dataene gemmes i et andet land eller kontinent. Datalagring på en server eller et datacenter i et fremmed land skal overholde nationens datasikkerhedslove.

Reaktionstid

Enhver forsinkelse i et svar eller levering af data fra et punkt til et andet.

Ældre system



Forældede computersystemer, programmeringssprog eller applikationssoftware, der bruges i stedet for tilgængelige opgraderede versioner. Ældre systemer er også forbundet med terminologi eller processer, der ikke længere er anvendelige i nuværende sammenhænge eller indhold, hvilket skaber forvirring.

Belastningsafbalancering

Processen med at distribuere arbejdsbyrde over et computernetværk eller computerklynge for at optimere ydeevnen.

Placeringsanalyse

Placeringsanalyse bringer kortlægning og kortdrevet analyse til virksomhedens forretningssystemer og datalager. Det giver dig mulighed for at knytte geospatiale oplysninger til datasæt.

Placeringsdata

Data, der beskriver en geografisk placering.

Logfil

En logfil defineres som en fil, der opretholder et register over begivenheder, processer, meddelelser og kommunikation mellem forskellige kommunikationssoftwareapplikationer og operativsystemet. Logfiler findes i den eksekverbare software, operativsystemer og programmer, hvor alle meddelelser og procesdetaljer registreres.

Maskingenererede data

Alle data, der oprettes automatisk fra en computerproces, applikation eller anden ikke-menneskelig kilde.

Maskinelæring

Brug af algoritmer til at tillade en computer at analysere data med det formål at "lære", hvilken handling der skal foretages, når et bestemt mønster eller en bestemt begivenhed opstår.

MapReduce

Softwareproceduren til opdeling af en analyse i stykker, der kan distribueres på forskellige computere forskellige steder. Den distribuerer først analysen (kort) og samler derefter resultaterne tilbage i en rapport (reducer). Flere virksomheder, herunder Google og Apache (som en del af Hadoop-rammen), leverer MapReduce-værktøjer.

Mashup

Metoden til at flette forskellige datasæt i en enkelt applikation for at forbedre output. For eksempel at kombinere jobfortegnelser med demografiske data.

Metadata

De data, der tjener til at give kontekst eller yderligere oplysninger om andre data. For eksempel udgør oplysninger om titlen, emnet, forfatteren, skrifttypen, forbedringerne og størrelsen på datafilen til et dokument metadata om det pågældende dokument. Det kan også beskrive de betingelser, hvorunder de data, der er gemt i en database, blev erhvervet, dets nøjagtighed, dato, klokkeslæt, en metode til kompilering og behandling osv.

Naturlig sprogbehandling

Softwarealgoritmer designet til at give computere mere nøjagtig forståelse af dagligdags menneskelig tale, så vi kan interagere mere naturligt og effektivt med dem.

NoSQL



Databasehåndteringssystemer, der ikke (eller ikke kun) bruger relationelle tabeller, der normalt bruges i traditionelle databasesystemer. Det refererer til datalagrings- og hentningssystemer, der er designet til håndtering af store datamængder, men uden tabelkategorisering (eller skemaer).

Objektdatabaser

De gemmer data i form af objekter, som de bruges af objektorienteret programmering. De adskiller sig fra relations- eller grafdatabaser, og de fleste af dem tilbyder et forespørgselsprog, der gør det muligt at finde objekt med en deklarativ programmeringsmetode.

Operative databaser

Operative databaser udfører regelmæssig drift af en organisation og er generelt meget vigtige for en virksomhed. De bruger generelt online transaktionsbehandling, der giver dem mulighed for at indtaste, indsamle og hente specifikke oplysninger om virksomheden

Parse

Opdelingen af data, såsom en streng, i mindre dele til analyse.

Forudsigende analyse

Brug af statistiske funktioner på et eller flere datasæt til at forudsige tendenser eller fremtidige begivenheder.

Offentlige data

Offentlige oplysninger eller datasæt, der blev oprettet med offentlig finansiering.

Forespørgsel

En anmodning om data eller information fra en databasetabel eller kombination af tabeller. Disse data kan genereres som resultater, der returneres af Structured Query Language (SQL) eller som piktogrammer, grafer eller komplekse resultater, f.eks. Trendanalyser fra data-mining-værktøjer. SQL er det mest kendte og udbredte forespørgselsprog.

R

Et populært open source-softwaremiljø, der bruges til analyse. R giver en bred vifte af statistiske (lineære og ikke-lineære modeller, klassiske statistiske tests, tidsserie-analyse, klassificering, klyngedannelse, ...) og grafiske teknikker og er meget udvidelig.

Realtidsdata

Data, der oprettes, behandles, lagres, analyseres og visualiseres inden for millisekunder

Anbefalingsmotor

En algoritme, der analyserer en kundes køb og handlinger på et e-handelswebsted og derefter bruger disse data til at anbefale supplerende produkter.

Referenceoplysninger

Data, der beskriver et objekt og dets egenskaber. Objektet kan være fysisk eller virtuelt.

Grundårsagsanalyse

Processen med at bestemme hovedårsagen til en begivenhed eller et problem.

Skalerbarhed

Systemets eller processens evne til at opretholde acceptable ydeevneniveauer, når arbejdsbyrden eller omfanget øges.

Skema

Strukturen, der definerer organisering af data i et databasesystem.

This programme has been funded with support from the European Commission. The author is solely responsible for this publication (communication) and the Commission accepts no responsibility for any use that may be made of the information contained therein.



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

Søg data

Samlede data om søgeudtryk brugt over tid.

Semistrukturerede data

Data, der ikke er struktureret af en formel datamodel, men som giver andre måder at beskrive dataene og hierarkierne på.

Sentiment analyse

Anvendelsen af statistiske funktioner på kommentarer, som folk laver på nettet og gennem sociale netværk for at bestemme, hvordan de har et produkt eller en virksomhed.

Server

En fysisk eller virtuel computer, der betjener anmodninger om en softwareapplikation og leverer disse anmodninger over et netværk.

Software-As-A-Service (SAAS)

Den voksende tendens hos softwareproducenter til at levere deres programmer over skyen - hvilket betyder, at brugerne betaler for den tid, de bruger på at bruge den (eller mængden af data, de får adgang til) snarere end at købe software direkte.

Rumlig analyse

Det refererer til analyse af geodata såsom geografiske data eller topologiske data for at identificere og forstå mønstre og regelmæssigheder inden for data distribueret i geografisk rum.

Structured v Unstructured Data: Strukturerede data er alt, hvad der kan placeres i en tabel og organiseres på en sådan måde, at de vedrører andre data i den samme tabel. Ustrukturerede data er alt, hvad der ikke kan - for eksempel e-mail-meddelelser, indlæg på sociale medier og optaget menneskelig tale.

Opbevaring

Enhvert middel til vedvarende lagring af data.

Structured Query Language (SQL)

Et programmeringssprog designet specielt til at administrere og hente data fra et relationsdatabasesystem.

Tag

Et tag er et stykke information, der beskriver de data eller indhold, det tildeles. Mærker er ikke-hierarkiske nøgleord, der bruges til internetbogmærker, digitale billeder, videoer, filer og så videre.

Taxonomi

Taxonomi henviser til klassificering af data i henhold til et forudbestemt system med det resulterende katalog. Det giver en konceptuel ramme for nem adgang og hentning.

Tekstanalyse

Anvendelsen af statistiske, sproglige og maskinindlærings teknikker på tekstbaserede kilder for at få mening eller indsigt.

Transaktionsdata

Data, der ændres uforudsigeligt. Eksempler inkluderer betalingspligtige og tilgodehavende data eller data om produktforsendelser.



Bind

Mængden af data, der spænder fra megabyte til brontobytes.

Visualisering

En visuel abstraktion af data designet med det formål at udlede mening eller kommunikere information mere effektivt.

XML-databaser

XML-databaser gør det muligt at gemme data i XML-format. XML-databaser er ofte knyttet til dokumentorienterede databaser.

Dataene, der er gemt i en XML-database, kan forespørges, eksporteres og serieiseres i ethvert format, der er nødvendigt.

