

Data

Med denne vejledning får du en introduktion til forskellige dataformer og -typer som analyseteamet arbejder med. Du lærer om de forskellige datakilder og hvordan du kan bruge dem i dit arbejde med data.

Rådata

Data som er uforarbejdede, oprindelige data, der er direkte indsamlet fra en kildefagssystem uden at blevet analyseret, filtreret eller bearbejdet. Disse data opfatter også mange antal enheder, f.eks. et stort antal tabeller.

Bearbejdet data

Data som er rådata, der er blevet omdannet eller analyseret for at gøre det mere brugbart eller forståeligt. Bearbejdning kan omfatte rensning, filtrering, sortering og aggregering.

Bearbejdet data er også organiseret i en struktureret eller semi-struktureret database. F.eks. hvor mange antal rådatatabeller blevet omdannet til én eller få tabeller, som kaldes for "flad-tabeller".

Alle relationelle databaser er en struktureret database, men ikke alle strukturerede database er nødvendigvis relationelle.

En vigtig ting ved en struktureret database er, at det er ofte er en relationel database. Det betyder, at dataene er organiseret i flere tabeller, der har forbindelser til hinanden. Med værktøjer som SQL kan man lave en skema, der definerer regler for, hvordan tabellerne skal relateres. Dette gøres ved at bruge primære nøgler og fremmede nøgler.

SQL

Analyseteamet forventer ikke, at du skal arbejde med SQL, men du kan hente data fra en SQL-database analyseteamet udstiller til rådighed.

Structured Query Language SQL, er et af de mest populære værktøj til at lave forespørgsler på data. SQL gør det nemt og hurtigt at gemme, hente og ændre data. Der findes også SQL-databaser, som er databaser, der bruger SQL som metode til at lave forespørgsler.

Skema i en database er som en blueprint eller en struktur, der definerer regler, hvordan dataene er organiseret. Det beskriver tabellerne i databasen, hvordan de relaterer sig til hinanden, og hvilke typer data hver tabel indeholder.

Primære nøgler er et unikt identifikationsnummer for hver række i en tabel. Det sikre, at hver post kan identificeres entydigt, så der ikke er nogen dubletter.

Fremmede nøgler er et felt i en tabel, der refererer til den primære nøgle i en anden tabel. Det skaber en forbindelse mellem de to tabeller.



Eksempel på en skema, primære nøgler og fremmede nøgler.

Datavarehus

Der er endnu ikke et internt datavarehus til ledelsesinformation. I stedet køber vi en datavarehusløsning fra KMD, som håndterer data fra mange af vores fagsystemer. Denne løsning hedder 'KMD Insight'

Datavarehus er en centraliseret opbevaringsløsning, hvor struktureret relationelle data fra forskellige kilder samles, opbevares og organiseres. På den måde kan brugere få adgang til data fra ét sted med én forbindelse. Brugeren kan bruge disse data til mange forskellige formål, såsom analyser, rapportering og AI-processer.

KMD Insight

Ikke alle har mulighed for at få adgang til KMD Insight og arbejde med data, fordi licensen er meget dyrt.

Dog, hvis din afdeling er seriøs omkring data arbejdet og ønsker at investere tid og licensen, så er det muligt at få adgang til løsningen, hvor analyseteamet vil hjælpe med oplæringen og holde dialog om udviklingen.

KMD Insight er en komplet Business Intelligence løsning fra KMD. Kort sagt betyder det, at løsningen hjælper med at samle data fra Varde Kommunes forskellige fagsystemer (ETL), lagre data; bearbejde data og sørge for at dataene er opdaterede (Datavarehus), og gøre dem nemt tilgængelige (DataMart) til datavisualisering i SAP Web Intelligence (alternativ værktøj til Power BI).

KMD's data-indsamlingsproces henter data fra Varde Kommunes fagsystemer, disse data gennemgår nødvendig validering for at opretholde en fuld historik og sikre korrekt indlæsning til databasen og anvendelse af data.

KMD sørger for at forbedre kvaliteten og berige data med forretningsregler og logik, samtidig med at ensarter dem (skema). De udfører forskellige opgaver som f.eks. datarensning, -validering, -strukturering, -konvetering og tilføjelse af forretningsmæssige regler.

Derefter samles bearbejdet data til én eller få masterdatakilder der indeholder alle relevante oplysninger for ét dataområde. Og gøre dem nemt at få adgang til og anvende data.

DataMart og Semantisk Model

Analyseteamet har lavet databaser, markeret som 'DataMart', hvor vi har bearbejdet data til et specifikt dataområde.

På ledelsesinformation informationsiden kan du finde oplysninger om, hvilke dataområder der er i drift.

DataMart er en del af en større database, som er designet til at fokusere på et specifikt dataområde eller et bestemt formål. Tænk på det som en lille niche butik/kiosk (DataMart) inden for en kæmpe distributionmarked (datavarehus). I stedet for have alle varer (data) i én stor bunke, er der en særlig sted, hvor brugeren kan finde det varer (data) de specifikt har brug for.

Semantisk model handler om at give mening til data. Det er en måde at organisere og beskrive data på, så det bliver lettere at forstå og bruge.

Ved at have en database med semantisk model kan sikre, at alle forstå dataene ens, hvilket er vigtigt, når man arbejder med analyser og data.

Det er muligt at hente data direkte fra et datavarehus, men det kan være svært at få adgang til og forstå dataene. En datamart database med en semantisk model gør det lettere at få adgang til dataene og forstå, hvad det indeholder, selvom brugeren ikke er data teknisk kompetent.

Semistruktureret database

En semi-struktureret database er en type database, der ikke følger en streng og fast struktur som struktureret database, men heller ikke er helt ustruktureret som visse filler eller dokumenter.

Offentlige forvaltninger, som f.eks. KL, Danmark Statistik og FLIS, bruger semi-strukturerede databaser til gøre deres data tilgængelige via API'er. Brugere kan hente disse data i formater som JSON og XML, som er semi-struktureret data.

API

Application Programming Interface API, er en slags snitplade, der gør det muligt for forskellige løsninger at tale sammen. Tænk på det som en menu på en restaurant: du vælger fra menuen, og køkkenet (som du ikke behøver at se) laver maden og leverer den til dig. Du behøver ikke vide, hvordan maden tilberedes - du beder om det, så får du det.

API'er fungerer på samme måde: løsning giver en liste over "funktioner", som en brugere kan bruge, og API'en leverer resultateterne. F.eks. en brugere kan send en forepørgsel om data med bestemte kolonne oplysning i bestemte periode. API'en leverer så dataene med de specifikke parametre, der er defineret i forespørgslen.

Data fra en API er typisk i formateret i JSON og/eller XML.

JSON

JavaScript Object Notation JSON, er et letlæseligt data format, der bruges til at strukturere data, så de nemt kan anvendes mellem løsninger eller gemmes i filer. Det er meget populært, især til webudvikling og API, fordi det er nemt at forstå for både mennesker og computere.

En JSON består af nøgle-værdi par, hvor nøglen er en streng (navn på objektet), og værdien kan være forskellige typer data (f.eks. tekst, tal og/eller lister).

```
{
  "navn": "Anna",
  "alder": 28,
  "adresse": {
    "by": "Varde",
    "postnummer": "6800"
  },
  "interesser": ["Læsning", "Rejser", "Musik"]
}
```

Eksempel på en JSON formatering.

XML

Extensible Markup Language XML, er også et letlæseligt data format, der bruges til at strukturere og opbevare data på en måde, der er både læsbar for mennesker og maskiner.

XML består af tags (mærker) til at afdrænse forskellige dataelementer, og du kan oprette dine egne tags, der passer til de data, du vil bruge. Det er især nyttigt til at udveksle data mellem forskellige løsninger og programmer, fordi det er universelt forståeligt.

```
<?xml version="1.0" encoding="UTF-8"?>
<person>
  <navn>Anna</navn>
  <alder>28</alder>
  <adresse>
    <by>Varde</by>
    <postnummer>6800</postnummer>
  </adresse>
  <interesser>
    <interesse>Læsning</interesse>
    <interesse>Rejser</interesse>
    <interesse>Musik</interesse>
  </interesser>
</person>
```

Eksempel på en XML formatering.

SharePoint List

SharePoint List er en slags tabular tabel, som brugeren kan oprette og bruge i Microsoft SharePoint til at organisere og dele information. Den fungerer som en simpel database, hvor du kan gemme og administrere data i rækker og kolonner, som er en begyndervennlig værktøj til at arbejde med data.

SharePoint List skema giver mere fleksibelt end traditionel databaser, derudover tillader SharePoint List at brugeren kan gemme vedhæftede filer, metadata og bruge af tilpassede kolonner, hvilket giver mere fleksibilitet, og derfor betragtes det som et semi-struktureret database.

ETL

Extract, Transform and Load ETL, er en proces der bruges til at samle, forberede og overføre data fra forskellige kilder til en central database eller et datalager.

ETL bruges primært til at integrere og forberede data, så organisationen kan analysere store mængder information fra forskellige fagsystemer. Det bruges ofte i Business Intelligence (BI) og datavarehuse, hvor man ønsker at få et samlet overblik over organisationens data for at træffe bedre beslutninger.