

Numeric Data Platform

Fleksibel indsamling og deling af data

Centrale og decentrale BI-løsninger

Business Intelligence løsninger kan typisk opdeles i dels centraliserede data warehouse systemer med tilhørende kuber og rapporteringsværktøjer, dels mindre decentrale løsninger, hvor rapportering og analyse sker direkte på kildesystemerne.

Vi kender fordele og ulemper ved begge approaches:



De data warehouse baserede løsninger understøtter rapportering på tværs af virksomhedens forskellige datakilder og muliggør avanceret behandling og berigelse af data, til understøttelse af enhver tænkelig dataanalyse. Til gengæld er sådanne løsninger tidskrævende og dyre at etablere, og ikke mindst tunge at justere i takt med ændrede forretningsbehov.



De decentrale løsninger med rapporteringsværktøjer, der går direkte på kildedata, er velegnede til at løse specifikke rapporteringsbehov ud fra de aktuelle forretningsønsker. Til gengæld er det vanskeligt at sammenstille data på tværs af virksomheden og muligheden for berigelse af data er meget begrænset. Hertil kommer at honorering af sikkerhedskrav til dataadgang ofte er bekostelige eller utilstrækkelige.

Trends

Blandt vores kunder oplever vi i disse år en glidning væk fra de store centrale data warehouse løsninger over mod de mere fleksible decentrale løsninger, ikke mindst på baggrund af den store markedsfokus på PowerBI og tilsvarende løsninger, der lokker med flot grafik og kort implementeringstid. Omvendt ved enhver erfaren BI-udvikler, at der kræves en solid bagvedliggende infrastruktur med ETL og datamodellering, hvis virksomhedens fulde analyse og rapporteringsbehov i praksis skal kunne understøttes.

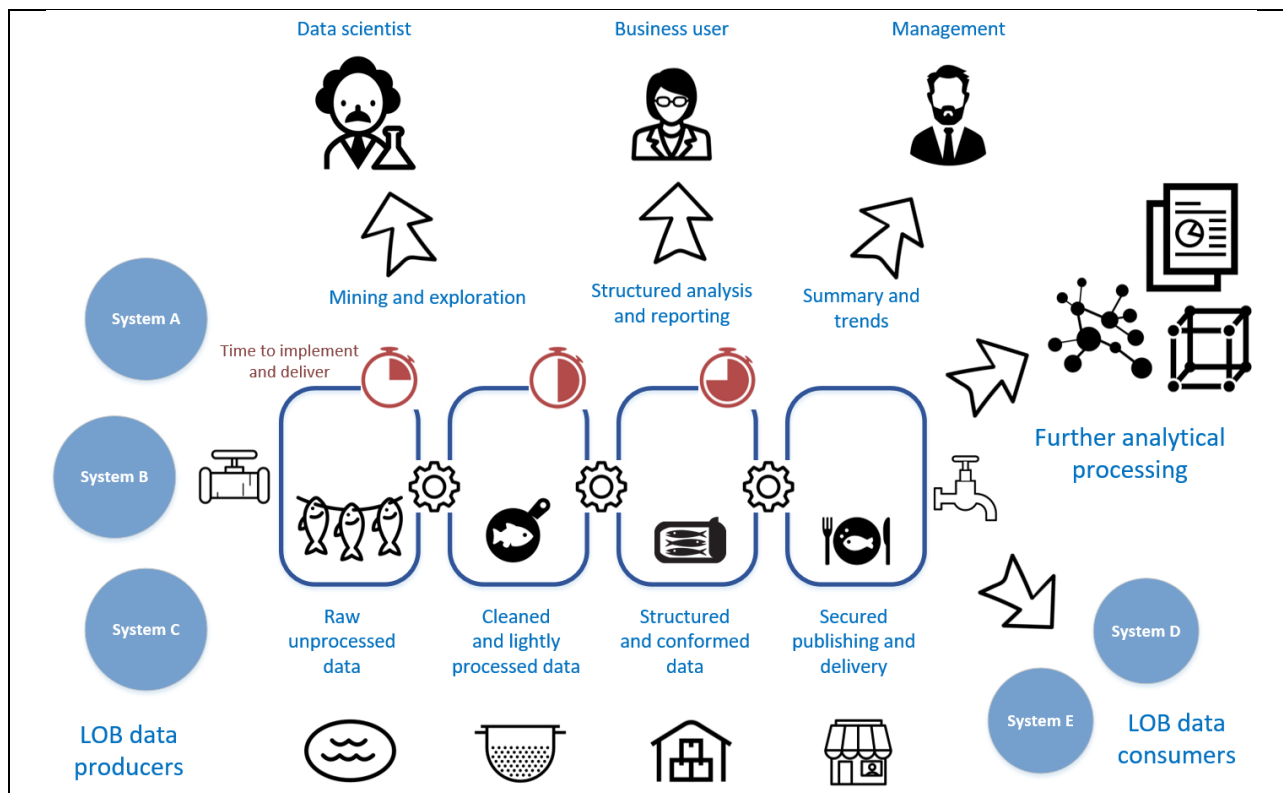
Ændrede behov

Generelt ønskes en løsningsplatform, der på en fleksibel og hurtig måde samler virksomhedens data, muliggør den nødvendige berigelse og udstiller data i den form, der bedst understøtter brugernes differentierede behov. Modsat ønsker man at undgå den statiske datamodel og omfangsrige ETL proces, der fordyrer BI projekterne og mindsker fleksibiliteten og dermed løsningens evne til at ændres i takt med forretningskravene.



Numeric data platform

På baggrund af disse ændrede behov, har Numeric i samarbejde med en række kunder udviklet en dataplatform, der kombinerer mange af styrkerne fra det traditionelle strukturerede data warehouse, med mere moderne teknikker såsom data lakes, automatisk data discovery og near realtime delivery. Resultatet er en dataplatform, som understøtter en høj grad af fleksibilitet og tilpasningsdygtighed til løbende ændring af forretningsbehov.



Fleksibilitet

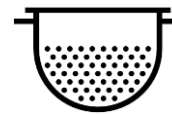
Fleksibiliteten opnår vi bl.a. ved at optimere de processer, der indgår i etableringen af BI løsningen og den efterfølgende drift og vedligeholdelse. Ved hjælp af konfigurationsstyret opsætning og data discovery kan den centrale datamodel, tilsammen med automatisk løbende dataindlæsning fra kildesystemer, etableres på få timer.



Slutbrugerne kan, med ganske kort etableringstid, få adgang til virksomhedens vigtigste data fra én central placering, baseret på kildedata opdateret med den ønskede frekvens, med mulighed for near-realtime delivery. En sådan indsamling af rådata direkte fra kildesystemerne er især relevant for data scientists, som typisk ønsker direkte og fuld adgang til virksomhedens rådata, men som vil slippe for selv at foretage den praktiske indsamling af data.

Rensning og sortering af data

Med udgangspunkt i de indsamlede rådata kan der foretages yderligere bearbejdning af data. Typisk vil der etableres en struktur til den grundlæggende rensning og sortering af data, hvorved de umiddelbare datakvalitetsproblemer kan håndteres og data kan grupperes i mere håndterbare størrelser. Denne efterbehandling foretages i direkte forlængelse af den løbende dataindlæsning, og medfører ikke forsinkelser i importen. På dette stadie er det desuden muligt at foretage simpel transformation og sammenstilling af data, hvilket ofte vil være et ønske fra slutbrugerne.



De let rensede og sorterede data vil typisk være udgangspunktet for de brugere, som ønsker at arbejde meget direkte med data, men som samtidig vil undgå selv at skulle håndtere datakvalitetsproblemer. Typiske konsumenter vil være data scientists og BI-løsninger baseret på værktøjer såsom PowerBI.

Hvis løsningen anvendes i en data hub arkitektur, vil de rensede og sorterede data typisk også være udgangspunktet for de konsumerende systemers hentning af data.

Strukturerede modeller

Hvis der er behov for yderligere bearbejdelse af data, så understøtter Numeric data platform samtidig muligheden for traditionelle ETL-processer til etablering af tværgående datastrukturer, hvor data fra forskellige kildesystemer konsolideres i fælles entiteter og beriges på baggrund af forretningsregler. Herudfra kan eks. dannes konforme dimensioner og facts. Transformationen af data sker på baggrund af de indsamlede rådata, og kan defineres præcist efter de gældende krav. Platformen giver fuld adgang til customisering af løsningens komponenter samt muligheden for at indsætte egenudviklede komponenter.



Med Numeric data platform kan der således etableres et egentligt data warehouse, som kan understøtte virksomhedens behov for centraliseret dataindsamling og understøttelse af tværgående processer. Data warehouse løsningen vil herved kunne danne udgangspunkt både for den centrale rapportering og analyse, og samtidig understøtte de decentrale BI-løsninger, som måtte eksistere lokalt i virksomheden.

Dataadgang

Et grundlæggende princip for platformen er, at data kan tilgås på forskellige niveauer. Data scientist kan anvende rådata indhentet direkte fra kildesystemerne. Forretningsbrugere der ønsker at anvende mere bearbejdede data, kan i stedet tilgå de strukturerede modeller. Data kan ligeledes danne udgangspunkt for yderligere modellering i en analytisk platform baseret på kuber eller lign.

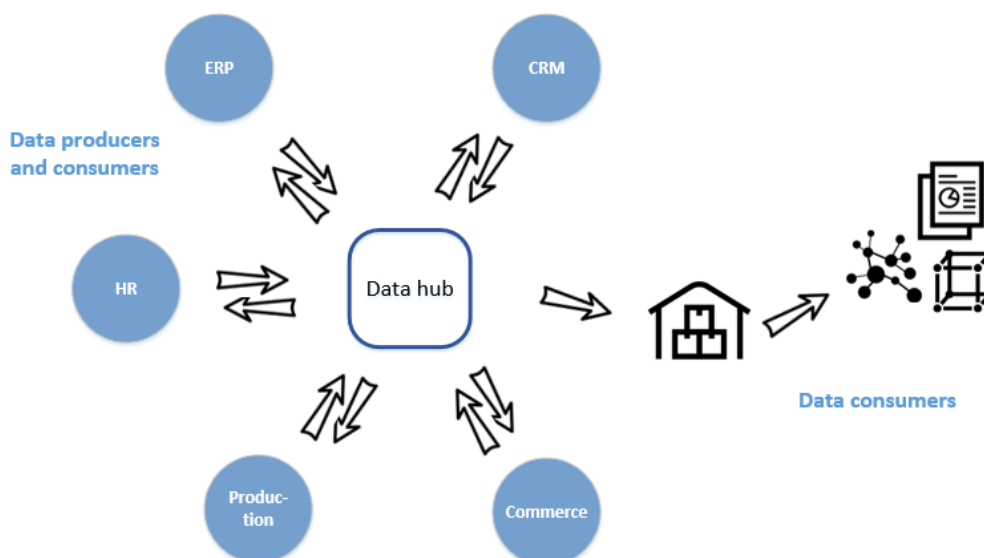


Uanset hvilke data der tilgås, og hvem de tilgås af, er al dataadgang omfattet af stringent adgangskontrol og logning. Adgangskontrollen kan styres på forskellige niveauer, herunder gruppe, individ og system, samt på entitetsniveau og rækkebaseret. Herved undgås at styring af dataadgange skal vedligeholdes i et flertal af systemer med forskellig funktionalitet og ansvarsdeling. Desuden understøttes de rapporteringsbehov, som bl.a. følger af EU's persondataforordning (GDPR).

Data hub arkitektur

Numeric Data Platform er grundlæggende baseret på en data hub arkitektur, hvor platformen skal ses som den centrale komponent i et samlet systemlandskab med data producenter og konsumenter. Løsningens øvrige funktionalitet er teknisk set "bare" en udbygning af data hub strukturen.

Behovet for en data hub skal særligt ses som ønsket om at strukturere dataudvekslingen mellem virksomhedens forskellige it-systemer. Typisk er dataudvekslingen baseret på point-to-point integrationer, der tilsammen skaber en spaghetti-arkitektur på tværs af systemlandskabet, som uvægerligt medfører kompliceret og bekostelig governance. Med en data hub arkitektur etableres løbende indsamling af relevante data fra virksomhedens it-systemer til en central database, hvorfra data løbende kan indlæses til øvrige systemer. Herved opnås fordele i form af en mere centraliseret styring at den samlede dataudveksling, en forenkling af driftsprocesser samt bedre overblik og kontrol med dataadgang.



Numeric Data Platform understøtter direkte etableringen af en data hub arkitektur i form af etablering af en central datamodel baseret på kildesystemernes dataentiteter, hurtig etablering af løbende dataindsamling, optimeret deltabaseret dataimport, effektiv adgangsstyring og solide driftsrutiner.

Drift og vedligeholdelse

Dataplattformens driftsprocesser er optimeret til sikring af et meget lavt niveau af driftsopgaver. Dataindlæsning fra kildesystemer understøtter både fuld og delta-baseret load, der sikrer en minimal belastning af de berørte kildesystemer. Alle dataprocesser logges løbende og samler derigennem detaljeret information om afviklingstider og eventuelle fejl. Logning kan aflæses grafisk via den indbyggede portal og giver både overordnede og detaljerede driftsinformation.

Dataplattformen løsningskomponenter afvikles som udgangspunkt på samme server og er udelukkende baseret på Microsoft SQL Server. Herved kan driftsopgaverne koncentreres til kun én server og ét softwareprodukt.

Målgruppe

Numeric Data Platform er relevant for virksomheder, som vil understøtte organisationens behov for løbende dataindsamling til brug for rapportering og analyse, og som samtidig lægger vægt på en høj grad af fleksibilitet i forhold til brugernes ændrede ønsker.

Numeric Data Platform er ligeledes relevant for virksomheder, som ønsker at etablere en centraliseret dataudveksling mellem organisationens it-systemer frem for den gængse point-to-point arkitektur.

Endelig er Numeric Data Platform relevante for virksomheder, som ønsker at etablere et egentlig data warehouse, som understøtter solide tværgående datamodeller til brug for organisationens centrale dataopsamling og rapportering.

Kom videre

Har du lyst til at høre mere om hvordan Numeric Data Platform kan understøtte din virksomheds databehov, så kontakt lead architect Henrik Nordtorp (M 5116 6068 / E henrik.nordtorp@numeric.dk) eller managing partner Kaare Thyregod (M 4142 1568 / E kaare.thyregod@numeric.dk).