

PowerBI – Teknisk analyse	Ansvarlig	AXH
	Oprettet	16-03-2017
	Side	1 af 5

## PowerBI - Teknisk analyse

PowerBI er delt mellem PowerBI Desktop, og PowerBI (service). Sidstnævnte tilbyder et subset af funktionaliteten i førstnævnte, primært omkring opsætning af grafer. Interfacemæssigt ligner de hinanden meget, selvom den ene er en desktop app ala Excel, og den anden er en SPA.

Det typiske flow er at rapporter og dashboard laves i PowerBI desktop, og publiceres til PowerBI servicen, hvor andre kan se dem (evt. via et anonymt tilgængeligt link med en nøgle i url'en der er view-only)

To modes:

- Import mode
- DirectQuery mode

**Import mode** loader alle data ind i PowerBis datastore, og efterbehandler dem her. Den understøtter således alle datakilder PowerBI understøtter (Google Analytics, Github m.m.m). Ulempen er, at data publiceres som et snapshot til tabeller der ligger i kopi i PowerBI servicen - dvs. nogen skal manuelt "trykke på knappen" i PowerBI desktop med jævne mellemrum, for at få opfrisket data. Der er dog angiveligt muligheder for at automatisere dette, vistnok via et Excel-ark på Onedrive (har ikke kigget på det, lyder meget brittle). REST api'et for PowerBI tabellerne er disabled for data i import mode, med mindre de er skabt via REST api'et (altså IKKE via import fra eksempelvis Excel), men vil man oprette tabellerne med api'et er bør data kunne pumpes ind i Import-mode tabeller på PowerBI service ad denne vej (utestet).

**DirectQuery** mode lader PowerBI servicen hente data direkte ved kilden - der spørges ved brug efter credentials når man publicerer fra PowerBI Desktop til PowerBI servicen første gang. Det har den fordel at data opdateres automatisk, til gengæld er man begrænset på en række punkter (fx kan man ikke definere lokale tabeller, DAX columns er yderst begrænsede). Dataanalyse er dog i vidt omfang muligt via custom measures. Datakilder skal eksPLICIT understøtte DirectQuery for at det virker - det gør SQL server, mens alt andet der er afprøvet (Google Analytics, Excel, Github) ikke gør. Vigtigt er også, at man ikke kan mikse - vil man bruge DirectQuery skal alle datakilder understøtte DirectQuery (det er så vidt vides ikke engang muligt at have mere end en database i spil i den samme rapport). I praksis betyder det nok at man må skubbe data i en form for SQL server "data warehouse" for at kunne tilgå dem der.

DirectQuery virker umiddelbart mod Azure SQL - vil man bruge en on-prem SQL server skal der installeres en connector service (utestet).

### Kolonner og filtre:

PowerBI tager afsæt i Excel, og de fleste ting udtrykkes således som kolonner og datafiltre. Fra andre systemer ville man måske forvente at en given graf kunne acceptere en formel, der beregnede værdien af Y-aksen ud fra kolonnerne. I PowerBI definerer man i stedet en custom column eller et custom measure (kan ikke redegøre præcist for forskellen) på den relevante tabel, laver eventuelle cross-lookups her så værdien beregnes, og kan herefter anvende den nye kolonneværdi i grafen.

### Brugerfiltrering:

Brugerinteraktion med fx hvilket datorange der analyseres/afrapporteres styres via en såkaldt "Slicer". Sliceren kræver, at alle tabeller der skal filtreres på én gang har en relation til slicing kolonnen. Relationen

kan enten stamme fra SQL, eller være sat op i PowerBi (hvor relationer kan føjes på i den dertil indrettede designer).

Hertil kan sættes mere eller mindre globale filtre op på hhv. Rapport, Side (en rapport har en eller flere sider) eller Visual (Card, line graph, bar chart etc.)

### Eksempel på measures:

Givet tabellerne

```
CREATE TABLE [dbo].[Dates](
    [Id] [int] IDENTITY(1,1) NOT NULL,
    [Date] [date] NOT NULL,
    CONSTRAINT [PK_Dates] PRIMARY KEY CLUSTERED
(
    [Id] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
)
```

```
CREATE TABLE [dbo].[AnonymizedLogins](
    [Id] [int] IDENTITY(1,1) NOT NULL,
    [DateId] [int] NOT NULL,
    [AnonymizedUserId] [nvarchar](250) NOT NULL,
    [System] [nvarchar](50) NOT NULL,
    [Aktiv] [bit] NULL,
    [CvrValideret] [bit] NULL,
    CONSTRAINT [PK_AnonymizedLogins] PRIMARY KEY CLUSTERED
(
    [Id] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
)
```

GO

```
ALTER TABLE [dbo].[AnonymizedLogins] WITH CHECK ADD CONSTRAINT
[FK_AnonymizedLogins_Dates] FOREIGN KEY([DateId])
REFERENCES [dbo].[Dates] ([Id])
GO
```

```
ALTER TABLE [dbo].[AnonymizedLogins] CHECK CONSTRAINT [FK_AnonymizedLogins_Dates]
GO
```

```
CREATE TABLE [dbo].[GroupCounts](
    [Id] [int] IDENTITY(1,1) NOT NULL,
    [DateId] [int] NOT NULL,
    [TotalAktive] [int] NULL,
    [TotalCvrValiderede] [int] NULL,
    CONSTRAINT [PK_GroupCounts] PRIMARY KEY CLUSTERED
(
    [Id] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
)
```

GO

```
ALTER TABLE [dbo].[GroupCounts] WITH CHECK ADD CONSTRAINT [FK_GroupCounts_Dates]
FOREIGN KEY([DateId])
REFERENCES [dbo].[Dates] ([Id])
GO
```

```
ALTER TABLE [dbo].[GroupCounts] CHECK CONSTRAINT [FK_GroupCounts_Dates]
GO
```

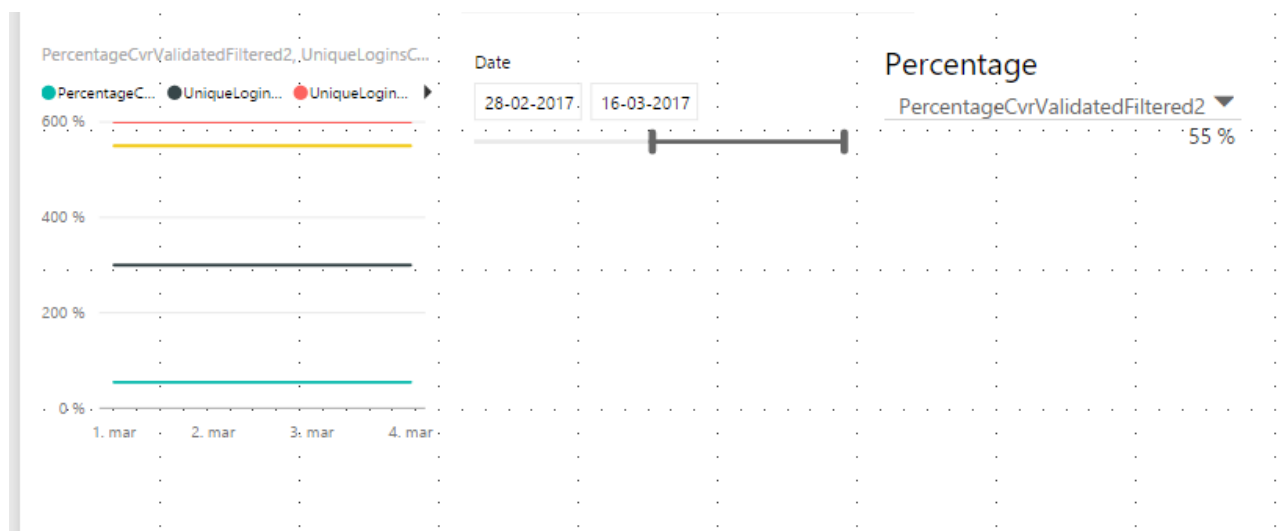
Kan man lave følgende measures i DAX på AnonymizedLogins for at se gennemsnit antal CVR-validerede der er logget på i en given periode:

```
UniqueLoginsCvrValidatedFiltered2 =
CALCULATE(DISTINCTCOUNT(AnonymizedLogins[AnonymizedUserId]);AnonymizedLogins[CvrValideret];ALLSELECTED(AnonymizedLogins))
```

```
AverageTotalCvrValideredeFiltered2 =
CALCULATE(AVERAGE(GroupCounts[TotalCvrValiderede]);ALLSELECTED(GroupCounts))
```

```
PercentageCvrValidatedFiltered2 =
DIVIDE([UniqueLoginsCvrValidatedFiltered2];[AverageTotalCvrValideredeFiltered2])
```

Baseret på en slicer på Date-tabellens Date felt. Herefter kan procenten trækkes ind på et card - vælg measuren og formatter (via ribbon) som procent. Kigger man data på et line chart er der en værdi per kolonne, men det er den samme værdi for hele datarandet. Intuitivt ville man måske forvente at ovenstående i stedet var et simpelt udtryk på selve visual'en, men sådan gør man ikke i PowerBi - på visual'en kan man kun vælge hvilke kolonner / measures der skal vises, hvordan de skal formatteres, og filtre på hvilke rækker der skal vises.





**Nyttige artikler:**

<https://www.simple-talk.com/sql/bi/direct-query-power-bi/>

<https://www.simple-talk.com/sql/reporting-services/working-with-sql-server-data-in-power-bi-desktop/>