

9. harjutustund – MDX päringukeel

Kava

- OLAP kuubi täiendamine
 - Hierarhiad dimensioonides
- Mitmedimensionaalne päringukeel MDX
 - Põhimõisted: kuup, dimensioon, hierarhia, tase
 - Hulgad ja korteežid
 - Lihtsamate päringute koostamine

OLAP kuubi täiendamine

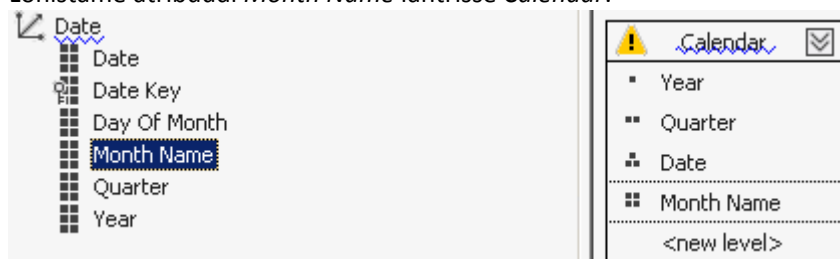
Hierarhiad dimensioonides

Dimensiooni atribuudid on sageli omavahelises hierarhilises seoses. Näiteks aja puhul on teada, et aasta koosneb kvartalitest ning kvartalid kuudest. Seda andmetes esinevat hierarhiat kasutatakse tihti ka andmete analüüsil. Nt. soovib lõppkasutaja üles leida tooted, mille müük ei ole kasumlik. Tootetabelis on tuhandeid tooteid, mistõttu alt-üles liikumine on praktikas ebamugav. Kui toodetele on defineeritud kategooriad ja tooteliinid, saab analüüsi teha ülevalt-alla: alustatakse üldiselt tasemelt, kust valitakse välja kõige probleemsem kategooria. Kategooria siseselt saab üles otsida mitterentaabli tooteliini, mille sees on tavaliselt 5-10 toodet – nende vaatlemine on kasutajale juba jõukohane.

Selleks, et OLAP andmebaasis oleks selline hierarhiline liikumine (*drilldown*) võimalik, tuleb dimensioonil defineerida, millised atribuudid moodustavad hierarhia.

Avame harjutuses 7 loodud ajadimensiooni (Date) ning täiendame loodud hierarhiat, mis kirjeldab üksteises sisalduvaid ajaperioode:

- Lohistame atribuudi *Month Name* lahtrisse *Calendar*.



- Järjestame atribuudid hierarhias järgmiselt:



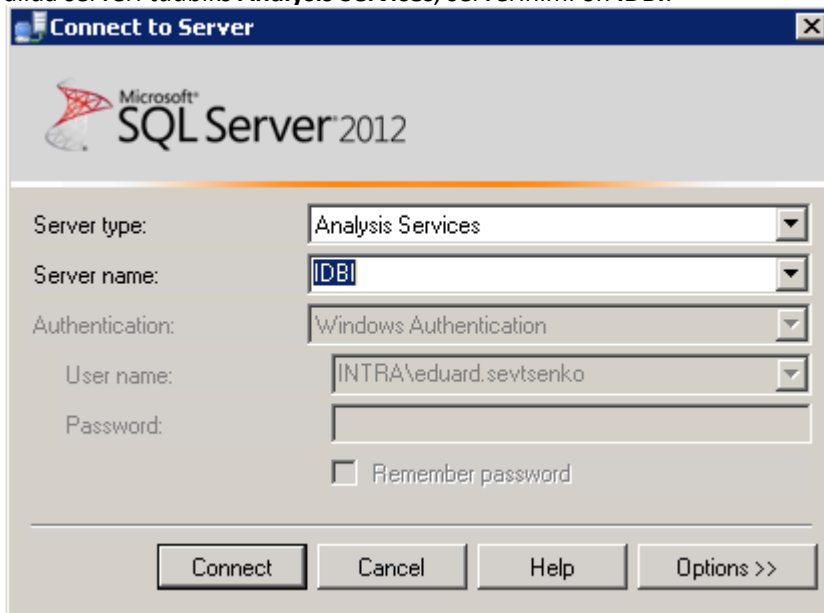
NB! Kõiki võimalikke hierarhiad ei pea tingimata eelnevalt valmis looma, defineerida tasub ainult sellised hierarhiad, millel on mingi äriiline sisu ja mida kasutatakse analüüsis igapäevaselt.

Mitmedimensionaalne päringukeel MDX

Üldist

MDX (Multi-Dimensional eXpressions) on OLAP-andmebaasidele mõeldud päringukeel, mille elemendid meenutavad SQL-keelt. Keel on loodud Microsofti poolt, kuid leiab kasutust ka teistes OLAP-süsteemides (Business Objects, Cognos, Oracle Essbase jt.)

MDX päringute kirjutamiseks kasutame töövahendit SQL Server Management Studio (SSMS). Päringuredaktori käivitamisel tuleb valida serveri tüübiks **Analysis Services**, serverinimi on **IDBI**:



NB! Kui soovite hiljem käivitada SQL päringuid (eelmistele harjutustundide teemade läbivaatamiseks), tuleb ühenduse loomisel määrata serveri tüübiks uuesti **Database Engine**.

Loome SSMSis uue päringu: *File > New > Analysis Services MDX Query* või tööriistaribalt nupp *New Query*.



Järgnevad näited põhinevad OLAP andmebaasil **Adventure Works DW2012_OLAP**, mis asub **IDBI** õppeserveris. OLAP andmebaasi saab määrata/vahetada tööriistaribalt (1). Päringuredaktoris määrame ka, millise kuubi põhjal päring peaks käivituma, näidis-andmebaasis on hetkel loodud kuup **Adventure Works** (2).

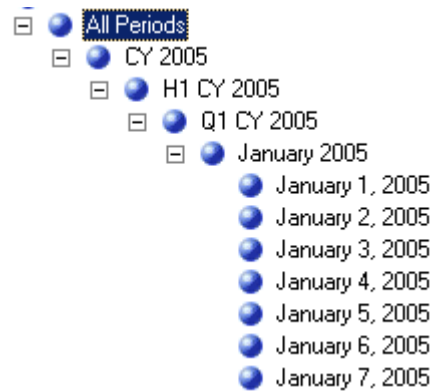


Päringute käivitamine käib sarnaselt SQL redaktoriga - tööriistaribalt **Execute**.

Põhimõisted

Tuletame meelde, et kuup on mitmedimensionaalne andmekogum, mis koosneb **dimensionidest** (*dimension*) ja arvulistest **mõõdikutest** (*measures*).

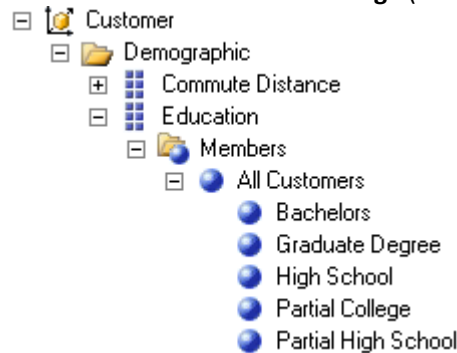
Iga dimension koosneb **hierarhiatest** (*hierarchy*), mis jagunevad omakorda **tasemeteks** (*level*).
Nt. dimensionis *Date* on hierarhia *Calendar*, millel on tasemed *Year*, *Quarter*, *Month*, *Date*.



Joonis 1 - dimensioni *Date* hierarhia *Calendar*

Kui dimensionis ei ole spetsiaalselt hierarhiad defineeritud, moodustavad atribuudid lihtsustatud kahetasemelise hierarhia:

- Ülemine tase grupeerib kokku kõik võimalikud väärtused – All
- Alumisel tasemel on iga erineva väärtuse kohta üks element ehk **liige** (*Member*).



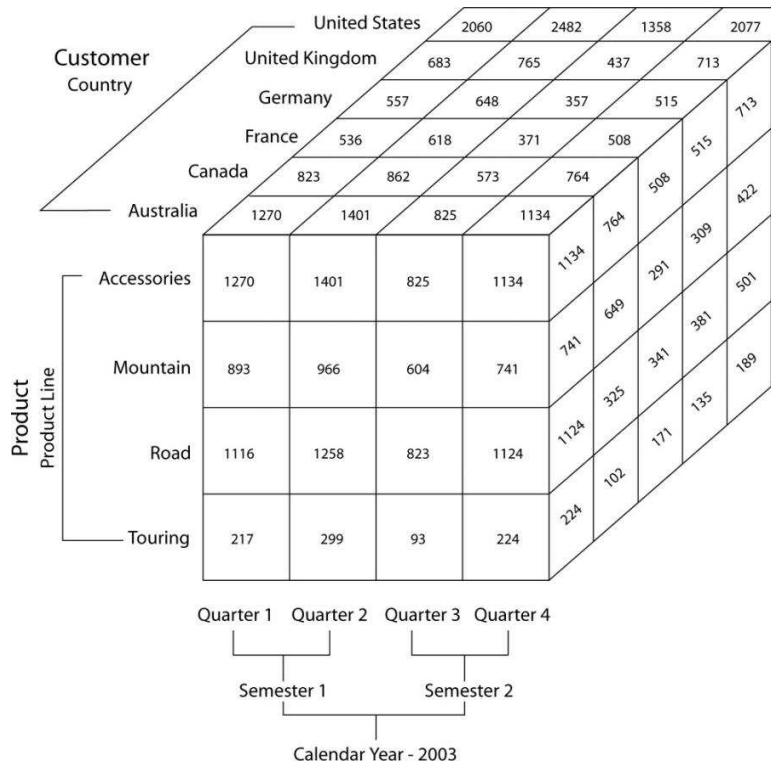
Joonis 2 - atribuudist moodustatud hierarhia

Andmestruktuur kuubis

Olgu antud kuup, mis koosneb kolmest dimensioonist: klient (Customer), kaup (Product) ja kuupäev (Date) ja mõõtudeks oleks müüdü kogused, siis võib sellist andmekogumit ette kujutada kolmemõõtmeliselt (vt. Joonis 3):

- Iga dimensioon on paigutatud ühele teljele.
- Teljele on lisatud jaotised iga dimensiooni liikme (e. väärtuse) kohta.
- Kõigi kolme telje ristumiskohas on nende liikmete kombinatsioonile vastav väärtus.

Jooniselt saame lugeda, et näiteks Austraalias müüdi 2003. aasta 3. kvartalis 604 maastikujalgratast.



Joonis 3 - näitlikustatud kolmest dimensioonist koosnev kuup

Igale dimensiooni liikmele on serveri poolt määratud nimi. Üldisel kujul näeb liikmele viitamine välja järgmine:

[Dimensioon] . [Hierarhia] . [Tase] . [Liige]

Näide: valitud kuupäev kannab nime: [Date] . [Calendar] . [Calendar quarter] . [Q1 CY 2003]

Kui tegemist on lihtatribuudiga, võib taseme ära jätta. Üldkuju on siis [Dimensioon] . [Atribuut] . [Liige]

Näide: liige "Mountain" kannab nime: [Product] . [Product Line] . [Mountain]

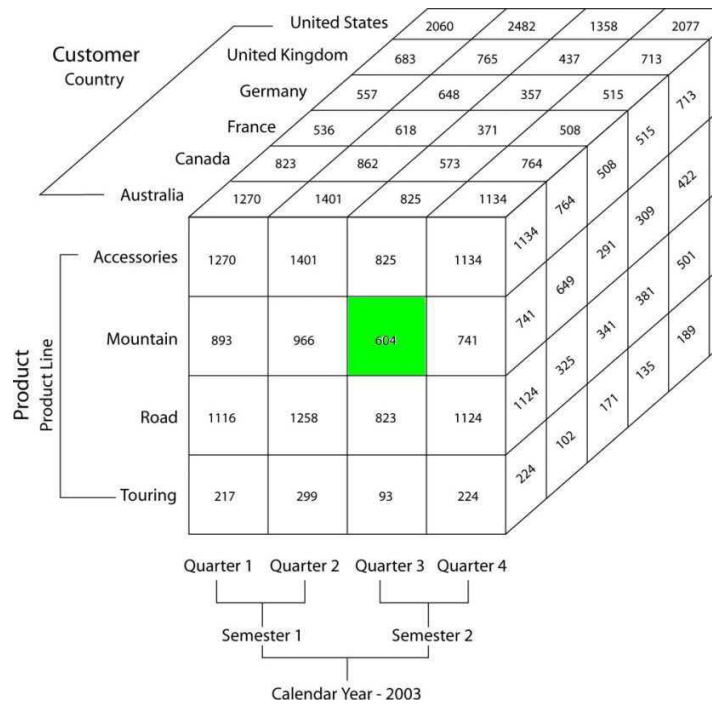
Riigile „Australia“ saab viidata kui: [Customer] . [Country] . [Australia]

Kui dimensiooni piires ei ole atribuudi väärtused unikaalsed ning nime põhjal ei ole võimalik tuvastada täpset liiget (nt. kliendid nimega John Smith), siis on mõistlik kasutada võtmepõhist viitamist. Võtmeatribuut valiti dimensiooni loomisel (vt. HT06 materjal, lk7 - *Key Attribute*). Klientide puhul sobib võtmeks kliendikood või andmeaida tabelisse genereeritud surrogaatvõti.

Näide: [Product] . [Product] . &[447] viitab tootele, mille ProductKey = 447. Koodi kasutamist näitab &-sümbol liikme väärtuse ees.

Kõik graafilised kuubisirviidjad kasutavad võimalusel liikmete viitamisel koodi – see on üheks põhjuseks, miks ka Excelist genereeritud MDX-päring on halvasti loetav.

Element (cell) tähistab kuubis dimensioonide ristumispunkti ning hoiab endas mõõdikuid (numbrilised väärtused). Näiteks soovime leida, mitu maastikujalgratast müüdi Austraalias 2003. aasta 3. kvartalil.



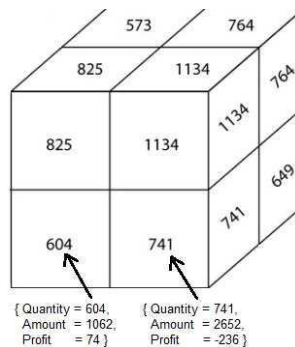
Iga element kuubis on määratud koordinaatide poolt. Elemendi 604 määravad koordinaadid ProductLine=Mountain, Country=Australia, Date=Quarter3. Väärtuse kuubist kätte saamiseks loetleme koordinaate määravad liikmed täisnimedega:

- [Product].[Product Line].[Mountain]
- [Customer].[Country].[Australia]
- [Date].[Calendar].[Calendar Quarter].[Q3 CY 2003]

Tulemuse annab meile järgmine päring:

```
SELECT [Measures].[Internet Order Quantity] ON
COLUMNS FROM [Adventure works]
WHERE ( [Date].[Calendar Quarter of Year].[CY Q3],
[Product].[Product Line].[Mountain],
[Customer].[Country].[Australia])
```

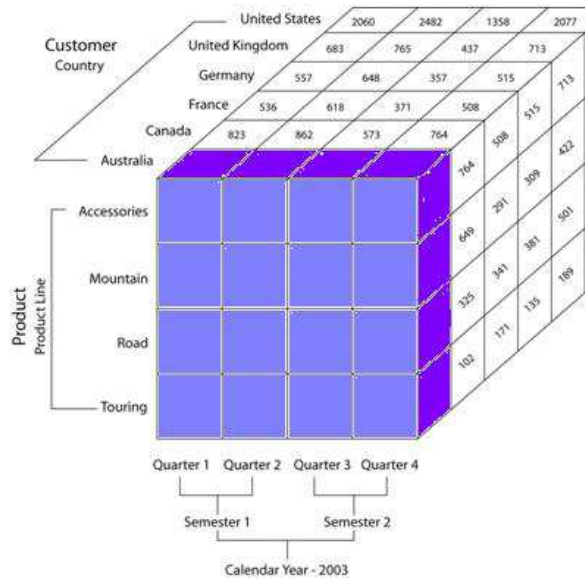
Koordinaatide ristumiskohas on tavaliselt mitu mõõtu (nt. kogus, summa, omahind, kasum jm). Lihtsustatult võib öelda, et iga ristumispunkti sees on väike massiiv, mis sisaldab endas kõiki mõõte:



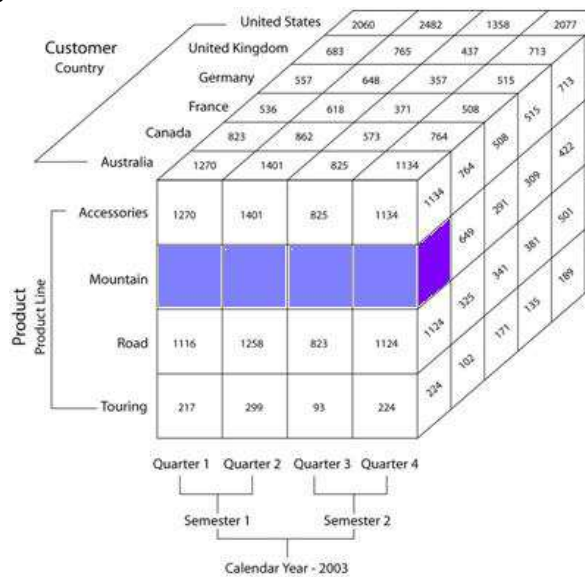
Sisuliselt on mõõtude massiiv ka omaette dimensioon [Measures], mis on olemas igas kuubis. Päringut kirjutades võib mõõtude puhul dimensiooni määramata jätta.

Korteež (*tuple*) identifitseerib kuubist mingi alamosa või lõike. Korteež koosneb ühest või mitmest dimensioonist, millest iga dimensiooni kohta näidatakse ära üks liige.

Näide: ([Customer].[Country].[Australia]) tagastab kuubist osa, kus kliendi asukohaks on Austraalia.



Samas ([Product].[Product Line].[Mountain], [Customer].[Country].[Australia]) lõikab kuubist välja tükikesse, kus osteti maastikujalgrattaid ning ostukohaks oli Austraalia.



Korteeži süntaksi üldkuju on ([Dim1].[Atribuut1].[Liige1], [Dim2].[Atribuut2].[Liige2], ...)
 Liikmete järjekord korteeži määramisel ei ole tähtis, kuid korteež tuleb moodustada erinevate atribuutide või dimensioonidega liikmetest.

Korteeži kasutatakse tavaliselt andmete filtreerimiseks. Järgmises MDX-päringus on vastav filter paigutatud WHERE-klauslisse:

```
SELECT [Measures].[Internet Order Quantity] ON COLUMNS
FROM [Adventure works]
WHERE ([Product].[Product Line].[Mountain], [Customer].[Country].[Australia])
```

Hulk (set) on sama dimensiooni liikmetest koosnev korteež. Kui soovime küsida andmeid kahe riigi kohta: Austraalia ja Kanada, siis selleks defineerime vastava hulga:

```
{ [Customer].[Country].[Australia] , [Customer].[Country].[Canada]}
```

Loetletud kahe riigi summaarse müügisumma leiab päringuga:

```
SELECT [Measures].[Internet Sales Amount] ON COLUMNS
FROM [Adventure Works]
WHERE ( { [Customer].[Country].[Australia] , [Customer].[Country].[Canada]} )
```

Hulga süntaksi üldkuju on { [Dim1].[Atribuut1].[Liige1] , [Dim1].[Atribuut1].[Liige2] , ... }

Liikmete järjekord hulgas määrab elementide järjekorra ka tulemusel (erineb hulga matemaatilisest definitsioonist). Hulga puhul on oluline, et see koosneks sama dimensiooni ja hierarhia/atribuudi liikmetest.

Hulki ja korteeže võib omavahel kombineerida. Nt. leiame Austraaliasse ja Kanadasse müüdud maastikujalgratate käibe:

```
SELECT [Measures].[Internet Sales Amount] ON
COLUMNS FROM [Adventure Works]
WHERE ( { [Customer].[Country].[Australia] , [Customer].[Country].[Canada]} ,
        [Product].[Product Line].[Mountain] )
```

Päringu WHERE-klauselis on korteež, mis koosneb kahest liikmest: asukoht ning toode:

```
( { [Customer].[Country].[Australia] , [Customer].[Country].[Canada]} ,
  [Product].[Product Line].[Mountain] )
```

Asukoht on omakorda hulk, mis koosneb kahest liikmest: Austraalia ning Kanada:

```
{ [Customer].[Country].[Australia] , [Customer].[Country].[Canada]}
```

MDX päringu elemendid

Kahemõõtmelist andmestruktuuri on lihtne visualiseerida tabelina – ridades kliendid ja veergudes aastad. Ka kolmemõõtmelist tulemust on võimalik graafiliselt ekraanil kujutada (ülal kuupide joonised), kuigi need ei ole enam piisavalt ülevaatlikud. Kõrgema dimensionaalsusega andmeid on keeruline juba ette kujutada, ekraanil visualiseerimisest rääkimata. Seetõttu tagastavad enamused MDX-päringuid andmed kahel teljel – read ja veerud.

Telg (axis) määrab päringus ridadele ja veergudele kuvatavad andmed.

- X-telje (tabeli veerud) jaoks on kasutusel COLUMNS
- Y-telje (tabeli read) jaoks ROWS
- Tehniliselt võib päris olla telgi rohkem (nummerdatud 0, 1, 2 jne), kuid enamused klientidele ei oska "kõrgematel" telgedel olevaid andmeid kuvada.

Päringu teljele saab valida üksikud liikmed (nt. [Customer].[Country].[Australia]) või ka hulgad. Kui

soovime saada mingi hierarhia kõikide liikmete hulka (nt. kõik riigid), siis kasutame funktsiooni **.MEMBERS**

Näide: [Customer].[Country].MEMBERS annab hulga kõikidest riikidest.

Järgmine päring leiab kõik riigid ning nende müügi käibe:

```
SELECT Measures.[Internet Sales Amount] ON COLUMNS,
       [Customer].[Country].MEMBERS ON ROWS
FROM [Adventure Works]
```

Ühel teljel on võimalik näidata mitut hierarhiat või atribuuti. Näiteks soovime vaadata riike ning neis leiduvaid linnasid.

Lahehenduseks on korteež:

```
SELECT [Measures].[Internet Order Quantity] ON COLUMNS,
       ([Customer].[Country].[Australia] , [Customer].[City].[Sydney]) ON ROWS
FROM [Adventure Works]
```

Kui korteeži moodustamisel valime liikme asemel hulga, moodustab MDX hulkade vahelise ristkorrutise.

Moodustame korteeži hulkadest: ([Customer].[Country].MEMBERS, [Customer].[City].MEMBERS)

Tulemuseks on päring:

```
SELECT [Measures].[Internet Order Quantity] ON COLUMNS
, ([Customer].[Country].CHILDREN, [Customer].[City].CHILDREN) ON ROWS
FROM [Adventure works]
```

Kui me ei soovi tulemuses näha "Kõik"- liiget, siis kasutame .MEMBERS-funktsiooni asemel .CHILDREN funktsiooni.

MDX päringus tohib üks atribuut või hierarhia paikneda ainult ühel teljel: kas ridadel, veergudel või filtris (WHERE-klausel). Järgmine päring on vigane – [Customer].[Country] asub korraga kahel teljel:

```
SELECT [Internet Order Quantity] ON COLUMNS,
[Customer].[Country].MEMBERS ON ROWS
FROM [Adventure works]
WHERE [Customer].[Country].[Australia]
```

Korrektse päringu saame siis, kui lisame kitsama tingimuse (hetkel Australia) ridadesse – tulemuseks on ju nagunii ainult Austraalia kohta käivad andmed:

```
SELECT [Internet Order Quantity] ON COLUMNS,
[Customer].[Country].[Australia] ON ROWS
FROM [Adventure works]
```

Avaldised

Päringute koostamisel võib tekkida vajadus teha uusi arvutusi kasutades olemasolevaid mõõde. Näiteks soovime iga tooteliini kohta teada kasumit ühiku kohta, mille arvutusvalem on [Internet Gross Profit] / [Internet Order Quantity].

Algselt moodustame päringu, mis leiab arvutuseks vajalikud mõõdud:

```
SELECT
{ [Internet Order Quantity], [Internet Gross Profit]
} ON COLUMNS,
[Product].[Product Line].CHILDREN ON ROWS
FROM [Adventure works]
```

Avaldiste defineerimiseks lisame päringule WITH-klausli, milles loetleme arvutatud mõõdud kujul:

```
MEMBER [mõõdu_nimetus] AS [avaldis]
```

Lisades päringule avaldise definitsiooni ning valides arvutatud mõõdu y- (veergude) teljele, saame:

```
WITH
MEMBER Profit_Per_Unit AS [Internet Gross Profit] / [Internet Order Quantity]
SELECT
{ [Internet Order Quantity], [Internet Gross Profit], [Profit_Per_Unit]
} ON COLUMNS,
[Product].[Product Line].CHILDREN ON ROWS
FROM [Adventure works]
```

Avaldised on mugavad, kui soovime näiteks päringut kasutada mõnes aruandlusrakenduses, kus meil on tarvis kätte saada x-telje (ridade) väärtuste koode (nt. piltide vm lisainfo kuvamiseks või klientrakenduse põhiseks filtreerimiseks).

Selleks lisame avaldise ProdLineKey, mis tuleks rida-haaval arvutada järgmise algoritmi põhjal:

- Küsi .CURRENTMEMBER funktsiooniga viit hetkel aktiivse rea peal kuvatud tooteliini atribuudi liikmele
- Kasuta liikme peal funktsiooni .UNIQUENAME, et kätte saada tema võtmeatribuudi väärtust

Kombineerides eelnevad funktsioonid, saame avaldise:

```
MEMBER ProdLineName AS [Product].[Product Line].CURRENTMEMBER
```


Täiendame eelnevat MDX-päringut:

```
WITH
  MEMBER Profit_Per_Unit AS [Internet Gross Profit] / [Internet Order Quantity]
  MEMBER ProdLineName AS [Product].[Product Line].CURRENTMEMBER.UNIQUENAME
SELECT
  { [ProdLineName]
    , [Internet Order Quantity]
    , [Internet Gross Profit]
    , [Profit_Per_Unit]
  } ON COLUMNS,
  [Product].[Product Line].CHILDREN ON ROWS
FROM [Adventure works]
```

Hulgad avaldistes

Mõnikord tuleb päringus moodustada hulki vastavalt ärireeglitele. Näiteks soovime leida 20 kõige suurema käibega toote kasumlikkust.

Moodustame päringu, mis leiab iga toote kohta kasumi:

```
SELECT
  { [Internet Gross Profit] } ON COLUMNS,
  [Product].[Product].CHILDREN ON ROWS
FROM [Adventure works]
```

Selleks, et leida 20 suurima käibega toodet, tuleks moodustada dünaamiline hulk, mis reastab kõik tooted käibe järgi ning tagastab sellest esimesed 20 elementi. Alustuseks defineerime hulga, mis tagastab kõik tooted:

```
WITH
  SET TopProducts AS [Product].[Product].CHILDREN
SELECT
  { [Internet Gross Profit] } ON COLUMNS,
  TopProducts ON ROWS
FROM [Adventure works]
```

Hulgast esimese N elemendi leidmiseks kasutame funktsiooni HEAD([hulk], [mitu])

```
WITH
  SET TopProducts AS
    HEAD([Product].[Product].CHILDREN, 20)
SELECT
  { [Internet Gross Profit] } ON COLUMNS,
  TopProducts ON ROWS
FROM [Adventure works]
```

Tulemuse kuvatakse esimesed 20 toodet, kuid vaadates andmetele peale, ei ole need kaugeltki kõige suurema käibega. Selleks, et saada korrektset tulemust, tuleks HEAD()-funktsioonile parameetrina ettenud hulk (antud juhul [Product] dimensiooni [Product] atribuudi liikmed) ära sorteerida vastavalt müügikäibele ([Internet Sales Amount]). Kasutame funktsiooni ORDER([hulk], [Mõõt või avaldis]) ja täiendame päringut:

```
WITH
  SET SortedProducts AS
    ORDER([Product].[Product].CHILDREN, [Internet Sales Amount], DESC)
  SET TopProducts AS
    HEAD(SortedProducts, 20)
SELECT
  { [Internet Gross Profit] } ON COLUMNS,
  TopProducts ON ROWS
FROM [Adventure works]
```

Eelneva päringu saab esitada ka kompaktsemalt:

```
WITH
  SET TopProducts AS
    HEAD(ORDER([Product].[Product].CHILDREN, [Internet Sales Amount], DESC), 20)
SELECT
  { [Internet Gross Profit] } ON COLUMNS,
  TopProducts ON ROWS
FROM [Adventure works]
```

Kui päringus defineeritud hulkad on korduvkasutatavad (nt. käibe järgi TOP 20 toodet kasutatakse aruandluses sageli), on mõistlik vastavad hulgad defineerida kuubi tasemel (vt. harjutusest 7 kuubi disainer *Calculations New Named Set*).

Avaldiste ja hulkade kombineerimine

MDX-abil on väga mugav teha arvutusi, mille tulemused sõltuvad "eelmisest" või järgmisest" reast. Näiteks soovime leida 2003 aasta kohta Internetipõhiste müükide kasumi kuude lõikes ning lisada igale kuule võrdluse eelnevaga

Alustame päringust, mis leiab iga kuu kohta kasumi

```
SELECT
    { [Internet Gross Profit]} ON COLUMNS,
    [Date].[Calendar].[Month].MEMBERS ON ROWS
FROM [Adventure works]
WHERE ([Date].[Calendar Year].&[2005])
```

Eelmise kuu kasumi leidmiseks alustame avaldisest, mis leiab meile eelmise kuu nime. Analoogiliselt **.CURRENTMEMBER** funktsioonile on olemas **.PREVMEMBER**, mis annab iga x-teljel (ridades) väljastatud [Date] dimensiooni [Calendar] hierarhia liikme kohta eelmise väärtuse:

```
WITH
    MEMBER LastMonth AS
        [Date].[Calendar].PREVMEMBER.UNIQUENAME
SELECT
    { [Internet Gross Profit], LastMonth } ON COLUMNS,
    [Date].[Calendar].[Month].MEMBERS ON ROWS
FROM [Adventure works]
WHERE ([Date].[Calendar Year].&[2005])
```

Kuna **.PREVMEMBER** tagastab liikme (nt. [Date].[Calendar].[Month].&[2005]&[2]), saame selle kombineerida mõõdikuga, mis annab tulemuseks mõõdiku väärtuse vastava liikme kohta (kasum 2005 veebruaris).

```
WITH
    MEMBER LastMonth AS
        ([Date].[Calendar].PREVMEMBER, [Internet Gross Profit])
SELECT
    { [Internet Gross Profit], LastMonth } ON COLUMNS,
    [Date].[Calendar].[Month].MEMBERS ON ROWS
FROM [Adventure works]
WHERE ([Date].[Calendar Year].&[2005])
```

Protsentuaalse muutuse leidmiseks lisame täiendava avaldise, mis arvutab kasumi kasvu protsentides:

```
WITH
    MEMBER LastMonth AS
        ([Date].[Calendar].PREVMEMBER, [Internet Gross Profit])
    MEMBER Change AS ([Internet Gross Profit] - LastMonth) / LastMonth
        , FORMAT_STRING = "Percent"
SELECT
    { [Internet Gross Profit], LastMonth, Change } ON COLUMNS,
    [Date].[Calendar].[Month].MEMBERS ON ROWS
FROM [Adventure works]
WHERE ([Date].[Calendar Year].&[2005])
```

Kokkuvõte

MDX päringukeel sarnaneb süntaksi poolest SQL-ga, kuid on mõeldud spetsiaalselt OLAP serveritel kasutamiseks. MDX on standardiseeritud, kuid sarnaselt SQL-ga rakendab iga tootja standardit erinevalt. MDX võimaldab pärida suuri andmehulki ning koostada tulemusi, mis SQL-vahenditega osutuvad väga keeruliseks või aeganõudvaks (andmed veergudes jm).

Harjutused

Koostada MDX-päringud, mis leiavad:

- Millises summas, koguses ja mis kasumiga müüdi jalgrattaid 2005. aastal Euroopa riikides?
- Kui palju meeste- ja naistejalgrattaid müüdi 2005 - 2009. aastal? Lisage võrdlus eelmise aastaga.
- Koostage päring, mis näitab 2007. aastal riikide lõikes, kui edukad olid erinevad müügikampaaniad (*Promotion*).

Exceli risttabeli abil saate kontrollida, kas saadud tulemus on korrektne.

Saatke vastused e-mailile (eduard.sevtsenko@ttu.ee)

Materjalid

Kõik hüperlingid on seisuga 15.10.2012

- Key Concepts in MDX - <http://msdn.microsoft.com/en-us/library/ms144884.aspx>
- MDX Query Fundamentals - <http://msdn.microsoft.com/en-us/library/ms145514.aspx>