

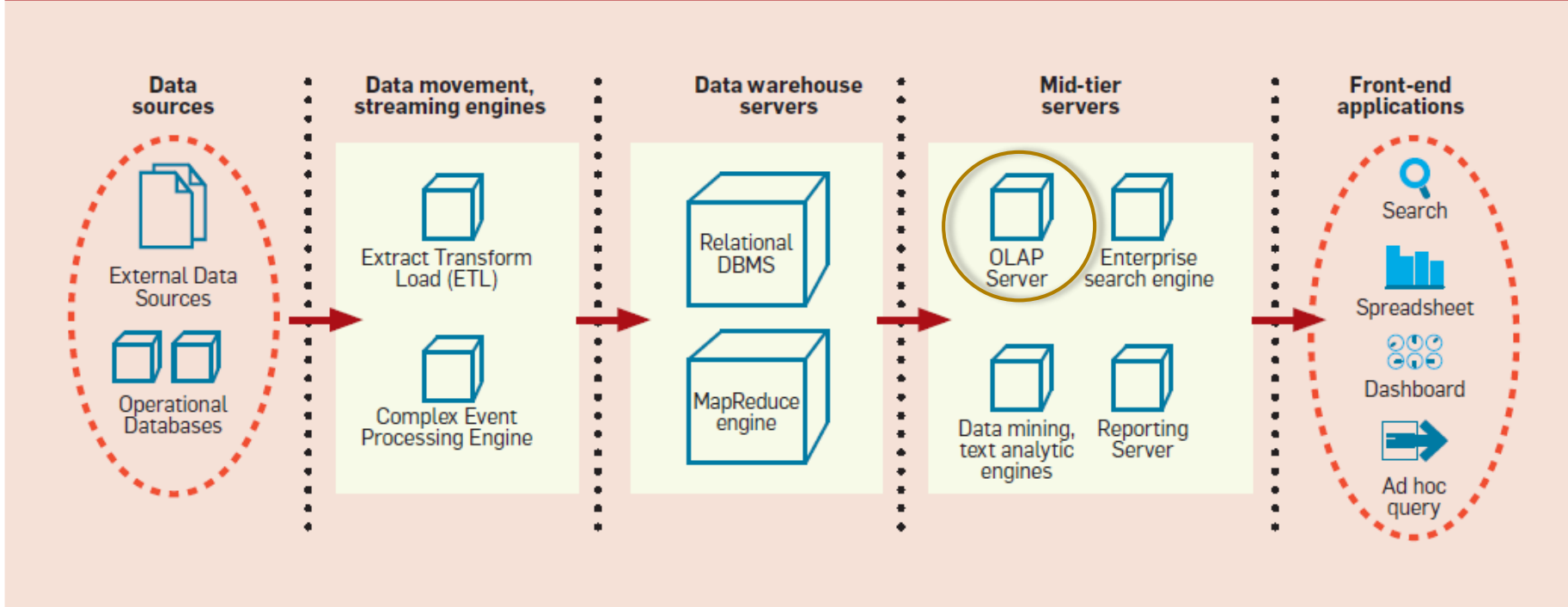
# BUSINESS INTELLIGENCE

## SSAS - SQL Server Analysis Services

# BI Architecture

2

Figure 1. Typical business intelligence architecture.



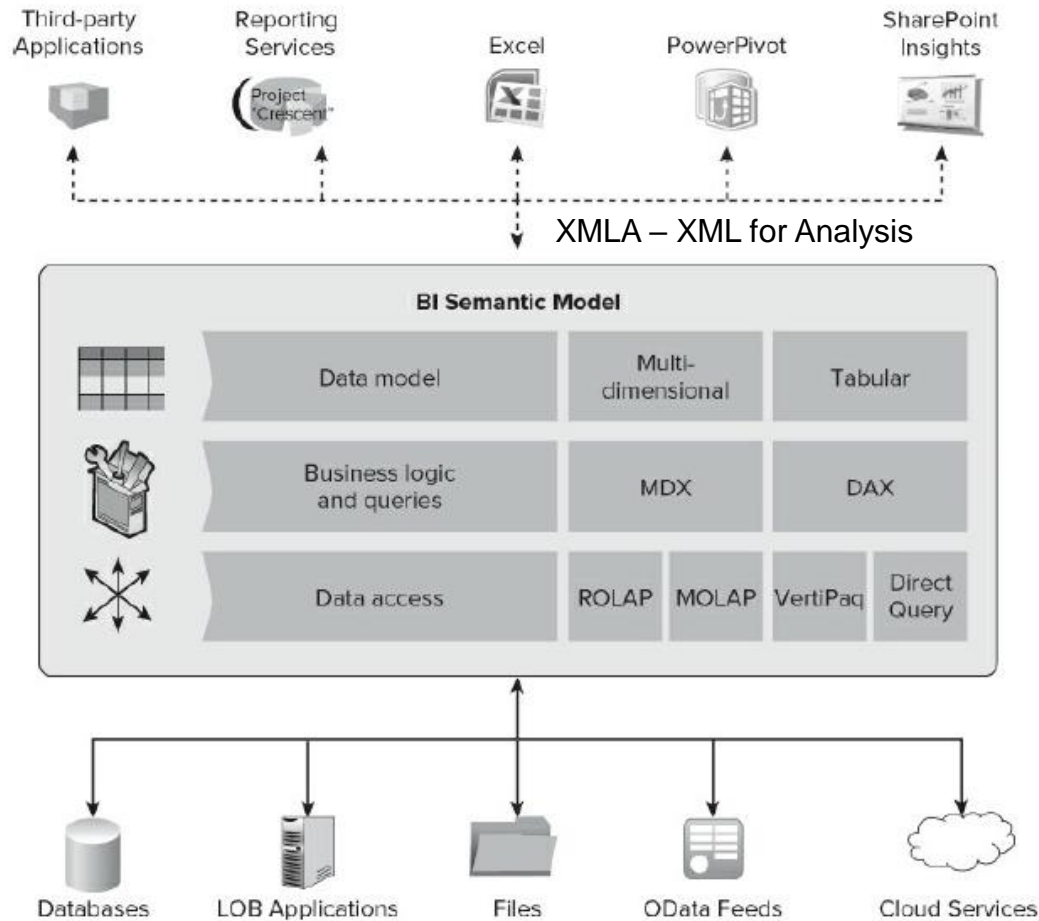
# SSAS: SQL Server Analysis Services

3

- It is both an OLAP Server and a Data Mining Server
  - ▣ Distinct from the RDBMS engine
  - ▣ Can access ODBC, OLE DB, CSV, XML data sources
  
- Most OLAP concepts are covered
  - ▣ Dimensions, hierarchies, measures, attributes, calculated metrics, key performance indexes, actions (URL links, drill-through, report launch), ...
  - ▣ Query language (MDX) for querying data cubes
  
- Docs and samples
  - ▣ Documentation
    - <http://msdn.microsoft.com/en-us/library/bb522607.aspx>
  - ▣ Tutorial from Books on Line
    - <http://msdn.microsoft.com/en-us/library/ms170208.aspx>

# SSAS architettura

4



# SSAS projects

5

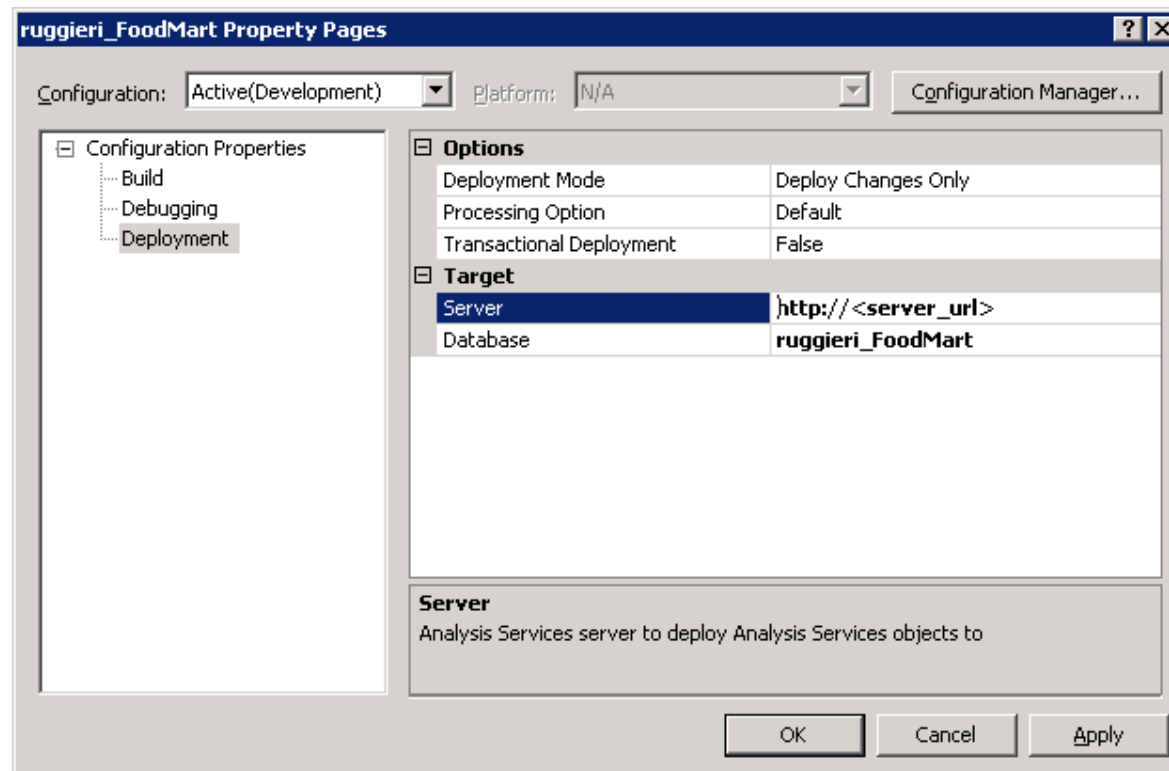
- Developing environment is SSDT
  - ▣ Project type: Analysis services multidimensional and data mining
- IMPORTANT!
  - ▣ Name of SSAS projects must be prefixed by your account

**<account>\_<name>**

# SSAS Server for deployment

6

- Right click on project name
  - ▣ Properties → Deployment



# SSAS project folders

7

- Data source
  - ▣ Sets the data sources
    - Use your login and password to SQL Server
    - Use OLE DB for SQL Server if client/server are different versions (eg., you are using a version < SQL 2016)
  - ▣ Impersonification credentials
    - Specify 'Use the service account'
- Data source view (DSV)
  - ▣ A view of data sources
    - Disconnected access to data sources
    - Names of attributes/tables can be changed (*without affecting the source!*)
    - Calculated attributes and tables (*without affecting the source!*)
    - External keys (*without affecting the source!*)

# SSAS project folders

8

- Dimensions
  - ▣ Type: standard / time
    - Time is useful to derive hierarchies directly from a 'datetime' attribute
  - ▣ Create new wizard
    - Select existing table
    - Key column: primary key (surrogate key)
      - Name column: descriptive key
    - Attributes
      - Select none at the wizard stage
  - ▣ Organize attributes into hierarchies
    - in the dimension structure pane



# Useful attribute properties

9

- AttributeHierarchyVisible
  - ▣ Flat hierarchies with only the attribute is visible
    - This is by default
- OrderBy
  - ▣ Default ordering method in visualization
- DiscretizationMethod
  - ▣ Discretization of continuous attributes into bins
- Type
  - ▣ Leave 'regular'
- Usage
  - ▣ Modes: key, regular and parent

# Build-deploy-processing

10

- Build
  - ▣ Syntactic check of correctness of the SSAS project
  - ▣ Run by the SSDT client environment
- Deploy
  - ▣ The project is copied on the deployment SSAS server
  - ▣ Data cubes are not re-processed
    - Nevertheless, features that do not depend on data re-processing are updated, eg., formatting of numbers, calculated metrics
- Processing
  - ▣ The deployment server re-computes the data cubes by accessing the data sources

# Build-deploy-processing

11

- Issue with current SQL server installation:
  - ▣ A delay of about 30 seconds is experienced at each deploy and at each process operations
  - ▣ Hope it will be solved with a future service pack.
    - Be patient!



# Data exploration

12

- Panel: browsing
  - ▣ Pivot table + filters
- Since SQL Server 2012
  - ▣ Data exploration in Excel

# Calculated members

13

- Calculated metrics:
  - Net sales
  - Margin
  - Sum year to date
  - Average sale amount per customer
  - Average sale amount per sale
  - Rank of products wrt sales
  - ...
- Calculated members:
  - Top 5 selling products
  - ...
- They do not exist on the data cube
  - They are calculated at run-time

# MDX

MultiDimensional eXpressions

Language for querying OLAP cubes  
and for defining calculated  
members

Standard de-facto

# Drill-through actions (rightclick -> 'show details' in Excel)

14

The image displays two overlapping screenshots of Microsoft Visual Studio. The top-left screenshot shows the 'Action Organizer' for a 'Drillthrough Action'. The 'Action Target' is set to 'Sales Fact 1997 1998'. The 'Drillthrough Columns' section is configured with the following table:

Dimensions	Return Columns
MEASURES	Store Sales, Store Cost, Unit Sales, Sales
Customer	Customer, Country, State Province, City

The bottom-right screenshot shows the 'Data Sample Viewer (first 1000 records)' window, which displays a table of data. The table has columns for Sales Fact 19, Sales Fact 19, Sales Fact 19, Sales Fact 19, Customer, and Customer. The data is as follows:

[Sales Fact 19]	[Sales Fact 19]	[Sales Fact 19]	[Sales Fact 19]	[\$Customer].[	[\$Cus
15.16	5.6092	4	1	Trujillo Ira	USA
6.48	2.5272	3	1	Whitten Fred	USA
2.28	1.026	2	1	Norris Timoth	USA
10.36	4.144	4	1	Cummings Ri	USA
2.2	0.814	2	1	Hahn Glenn	USA
2.73	1.3104	3	1	Koman Denis	USA
4.56	1.6416	4	1	Pacheco Kare	USA
2.73	0.8736	3	1	Bunczewski R	USA
2.28	0.912	2	1	Harrison Jose	USA
6.48	2.6568	3	1	Beninati Barry	USA
4.32	1.728	2	1	Martucci Mar	USA
3.3	1.353	3	1	Kartz Kay	USA
2.73	0.9555	3	1	Bliss Roseann	USA
7.77	3.4188	3	1	Haynes Phil	USA
7.44	3.2736	4	1	Trujillo Ruth	USA

# Data cube storage model

15



## **ROLAP** (Relational OLAP)

- ❑ relational engine enhanced with CUBE BY and analytic SQL
  - materialized views + bitmap/columnstore indexes + star-join optimization
  - performance
  - scalability

## **MOLAP** (Multidimensional OLAP)

- ❑ multidimensional array store on disk/memory in binary format
  - very efficient for a small number of hierarchies
  - do not scale well on space data

## **HOLAP** (Hybrid OLAP)

- ❑ trade-off between the previous two solutions
  - most accessed cuboids on MOLAP, others on ROLAP

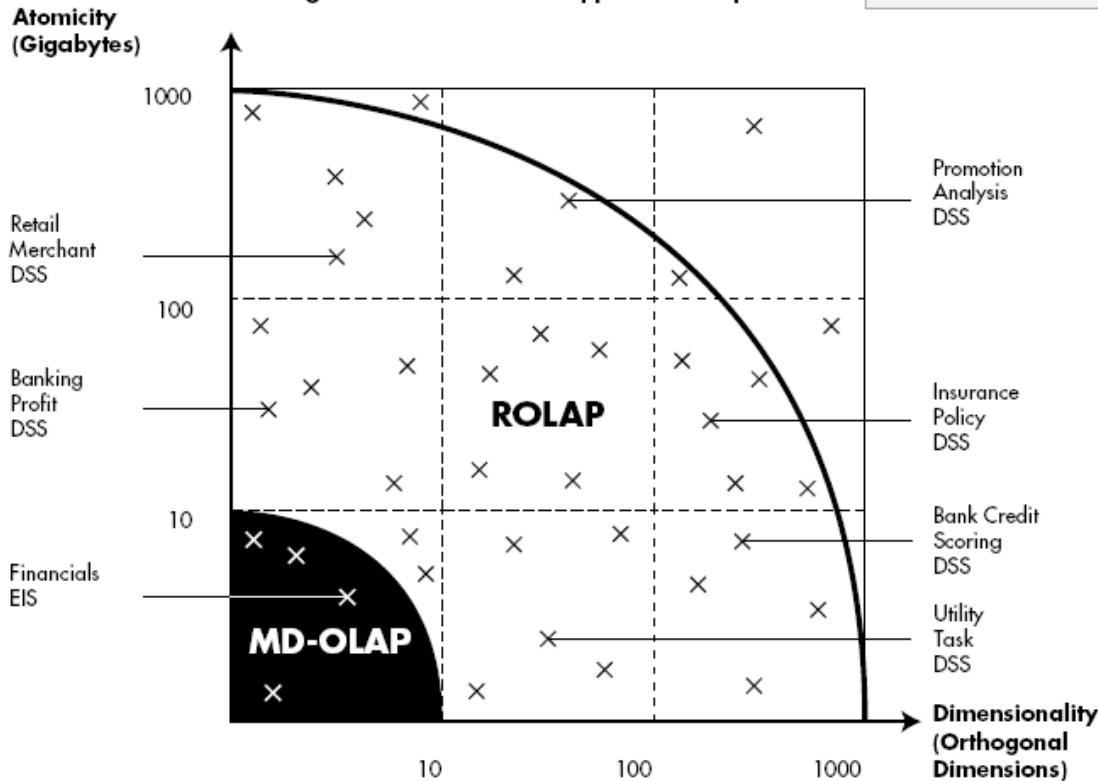
# The ROLAP case

16

## Data storage modes

OLAP Server	MOLAP	ROLAP	HOLAP	Offline
Essbase	Yes	Yes	Yes	
icCube	Yes	No	No	Offline Cubes
Microsoft Analysis Services	Yes	Yes	Yes	Local cubes, PowerPivot for Excel
MicroStrategy Intelligence Server	Yes	Yes	Yes	MicroStrategy Office, Dynamic Dashboards
Mondrian OLAP server	No	Yes	No	
Oracle Database OLAP Option	Yes	Yes	Yes	
Palo	Yes	No	No	
SAS OLAP Server	Yes	Yes	Yes	
TM1	Yes	No	No	
SAP NetWeaver BW	Yes	Yes	No	

Figure 20: ROLAP Meets Application Requirements

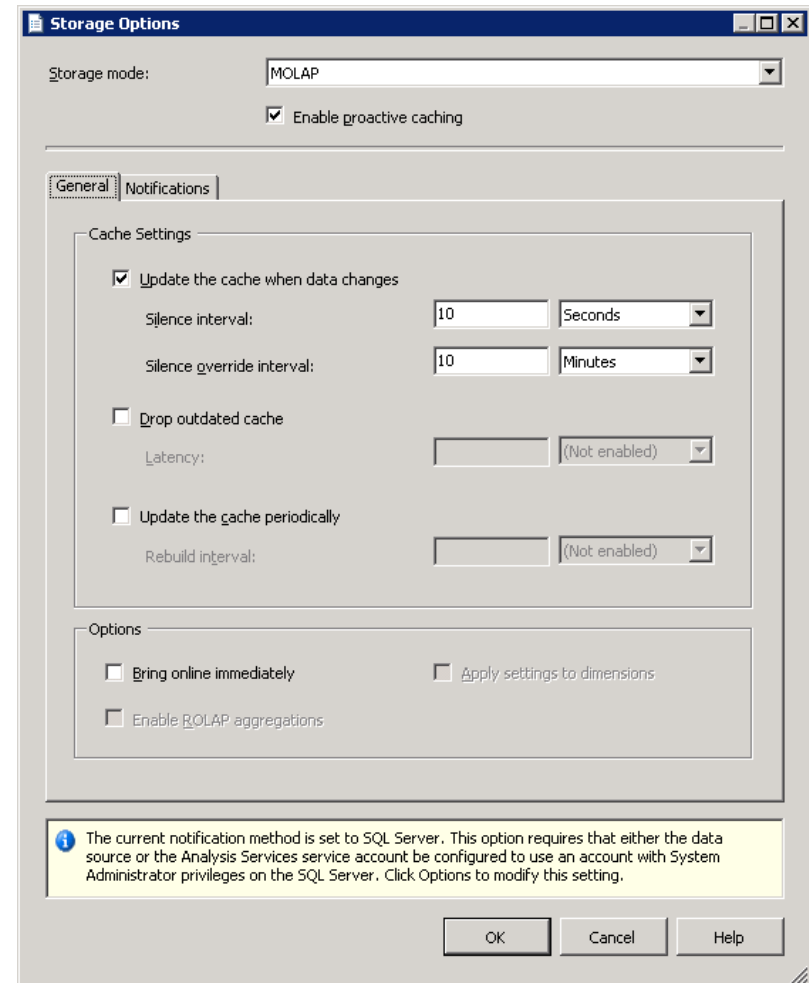
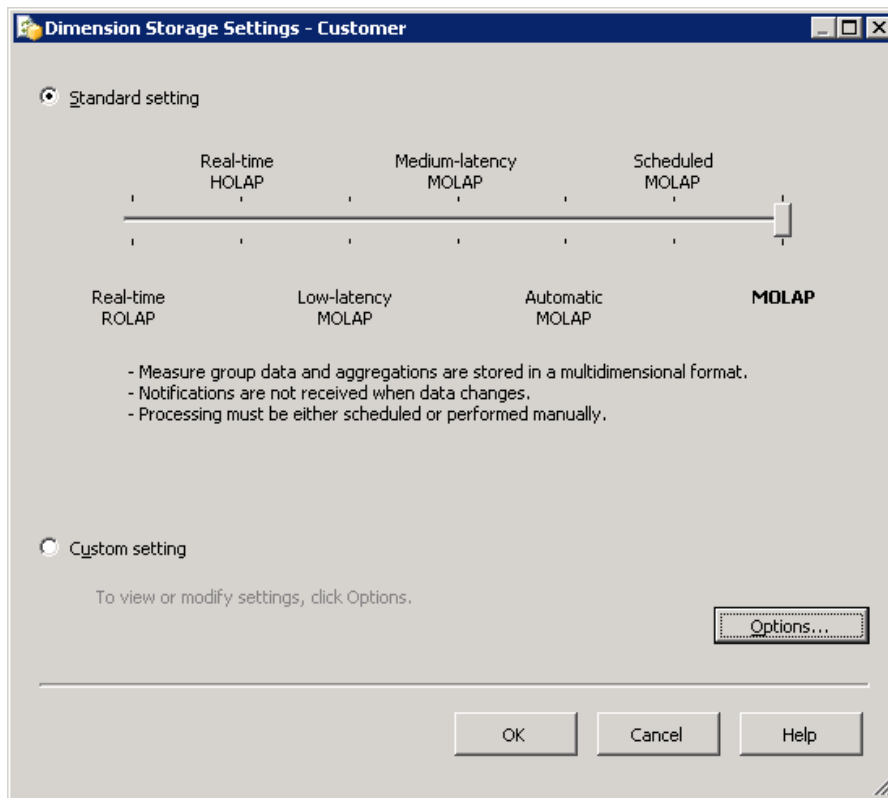




# Storage model

17

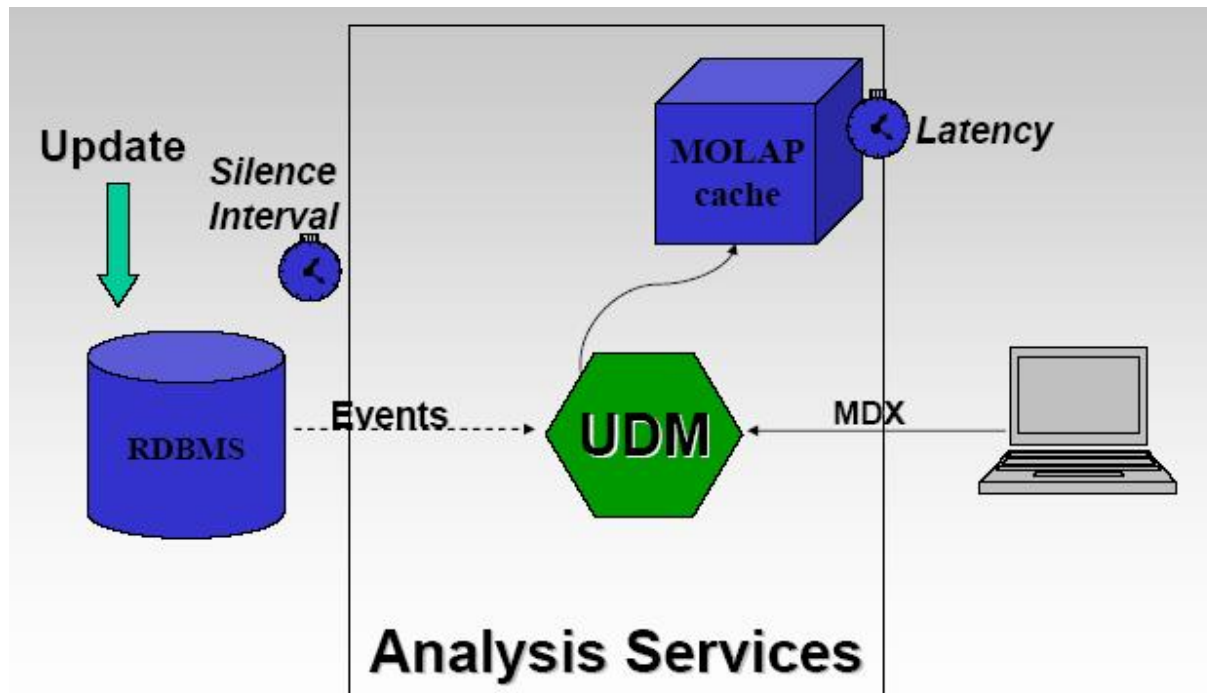
- Can be set for whole cube or for single dimensions



# Proactive caching

18

- Proactive caching
  - ▣ Latence time for refresh
  - ▣ Silence time (after refresh)



# Other features of SSAS

19

- KPI – Key performance index
  - ▣ Metrics with target values shown
- Perspectives
  - ▣ Subsets of objects, e.g., sub-cubes for product manager, store manager ..
- Roles
  - ▣ Access rights management
- ...
- Self-service Business Intelligence
  - ▣ PowerPivot for Excel/SharePoint
  - ▣ Tabular data model (evolution of PowerPivot)
    - Different instance of SSAS, different SSDT project type, different query language (DAX – Data Analysis eXpressions)

# BUSINESS INTELLIGENCE LABORATORY

## SSAS Practice

# Data analysts: final user

21

- Explore a report produced from a multidimensional view, using:
  - a reporting tool
    - Browser, Excel, Microstrategy, ...
  - only data exploration primitives:
    - Drill down and roll-up over pre-defined hierarchies
    - Existing calculated measures
    - Slide and dice
    - Filter and sort

# Q0 on foodmart

22

- What is the distribution of sales
  - by quarter?
  - and by customer city?
- ▣ in absolute value
- ▣ in percentage wrt the total
- ▣ in percentage wrt the country of residence of customers

# Q1 on foodmart

23

- What are the 5 best product categories
  - as per total sales?
  - as per number of items sold?
  - as per number of distinct customers?
- in each quarter of 1998 and gender?
- in the CA state?

# Q1 on foodmart

24

- What are the 5 best product categories
  - as per total sales?
  - as per number of items sold?
  - as per number of distinct customers?
- in each quarter of 1998 and gender?
- in the CA state?



# Data analysts: OLAP designer

25

- Design data cubes and reports
  - ▣ by defining
    - existing hierarchies from the DW
    - existing metrics from the DW
    - calculated members
    - and reprocessing data cubes
  - ▣ using a tool for OLAP design
    - with read-only rights on the DW

# Q2 on foodmart

26

- Which stores are the most profitable
  - ▣ mean profit wrt customers
    - = (total sales – total cost ) / number of customers
  - ▣ mean profit wrt baskets
    - = (total sales – total cost ) / number of baskets
- evaluated
  - ▣ in each quarter of 1998 and gender?
  - ▣ in the CA state?
  - ▣ *in each month wrt previous month?* MDX

# Q2 on foodmart

27

- Which stores are the most profitable
  - ▣ mean profit wrt customers
    - = (total sales – total cost ) / number of customers
  - ▣ mean profit wrt baskets
    - = (total sales – total cost ) / number of baskets
- evaluated
  - ▣ in each quarter of 1998 and gender?
  - ▣ in the CA state?
  - ▣ *in each month wrt previous month?* MDX

# Q3 on foodmart

28

- Which are the 5 product categories with the best margin
  - for each age-range of customers?
- evaluated
  - in each quarter of 1998 and gender?
  - in the CA state?
  - *in each month wrt previous month?* MDX

# Data analysts: DW designer

29

- Design and maintain the DW to satisfy new requirements
  - ▣ by re-designing conceptual and logical shemata
  - ▣ adding new dimensions and attributes
  - ▣ adding new data marts
- and managing the population of data
  - ▣ using ETL tools

# Q4 on foodmart

30

- What quantity (in Kg) has been sold
  - ▣ in each quarter of 1998 and gender?
  - ▣ in the CA state?
  - ▣ *in each month wrt previous month?* MDX

# Q5 on foodmart

31

In September 1998, store 7 changed its type

- ▣ from 'Supermarket' to 'Deluxe Supermarket'
  - ▣ with new store\_id = 25, but it is the same store!
- 
- Which stores are the most profitable
    - ▣ in each quarter of 1998?

# BUSINESS INTELLIGENCE LABORATORY

## MultiDimensional eXpressions (MDX)



# MDX Queries

```
[WITH < formula > [, < formula > ...]]  
SELECT [ < axis > , [ < axis > ...]]  
FROM [ < cube > ]  
[WHERE <set>]
```

33

```
select [Measures].[Store Sales] on columns,  
       [Customer].[Geography].[Country] on rows  
from sales  
where [Time].[WeekDays].[The Day].&[Monday]
```

Axis expr (cuboid)

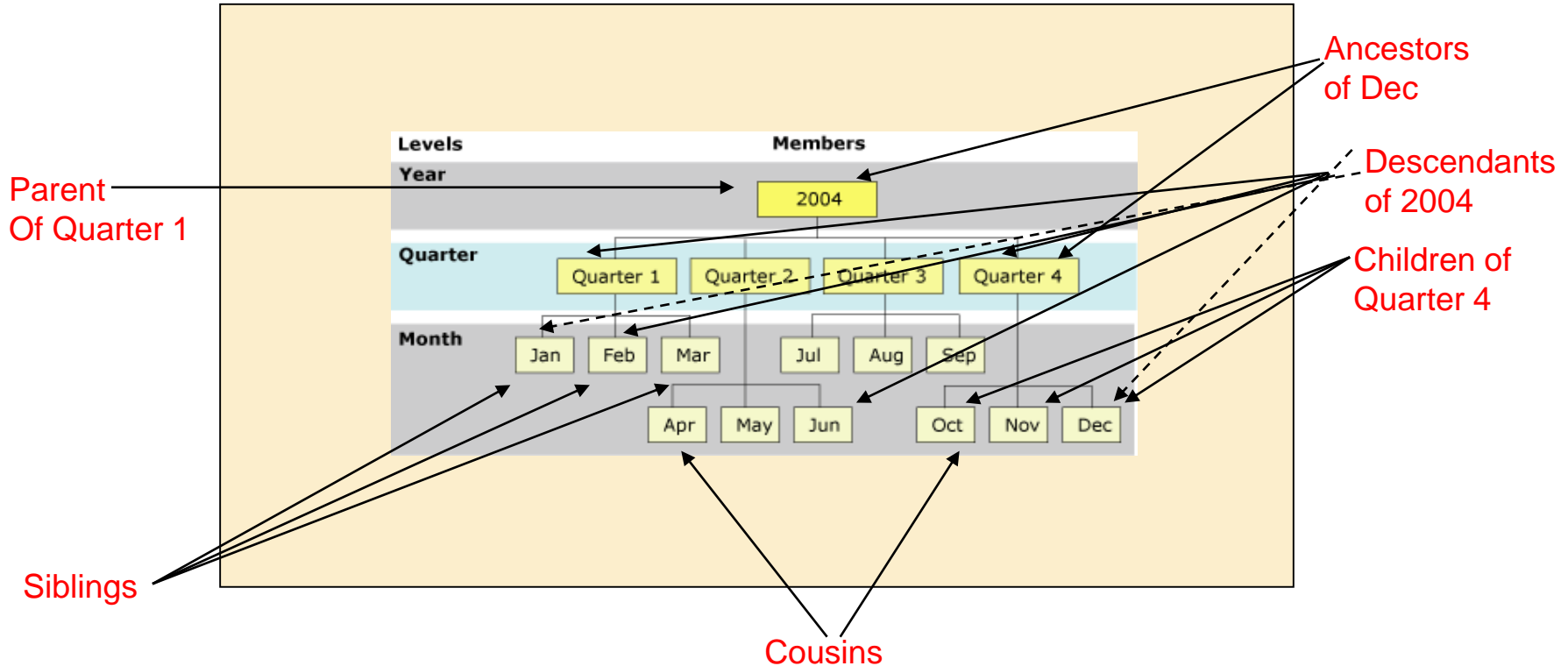
Cube

Set expr (slice)

# Members of hierarchies & nav. functions

34

## Hierarchy



**Syntax:** [DimensionName].[HierarchyName].[LevelName].[MemberName]

**Example:** [Store].[Time].[Quarter].&[Quarter 1]

**Syntax:** [DimensionName].[HierarchyName].[Path from root]

**Example:** [Store].[Time].[All].[2004].[Quarter 1]

# Tuples, Sets, Axis

```
[WITH < formula > [, < formula > ...]]  
SELECT [ < axis > , [ < axis > ...]]  
FROM [ < cube > ]  
[WHERE <set>]
```

35

## □ Axis

- axis ::= [NONEMPTY] Set ON (alias | AXIS(number) | number)
  - aliases COLUMNS, ROWS, PAGES, SECTIONS, and CHAPTERS

## □ Sets

- Set ::= tuple | {tuple, ..., tuple} | set + set | set – set |  
set\_function(parameters)
- Denotes a set of members/tuples

## □ Tuples

- Tuple ::= Member | (Member, ..., Member)
- Denotes a data cube cell by its coordinates
- No two members over the same hierarchy
  - Two member over the same dimension is OK

## □ Cube

- subselects are admitted

# Calculations

```
[WITH < formula > [, < formula > ...]]  
SELECT [ < axis > , [ < axis > ...]]  
FROM [ < cube > ]  
[WHERE <set>]
```

36

- Calculated member
  - ▣ Formula ::= MEMBER alias\_name AS mdx\_expr
- Named set formula
  - ▣ Formula ::= SET alias\_name AS set

## □ Syntax of MDX expressions

$mdx\_expr ::= Numeric \mid ( Tuple, Numeric )$

$Numeric ::= [Measures].[measure\ name]$

$\mid numeric\_function(Parameters)$

$\mid Numeric + Numeric \mid Numeric - Numeric$

$\mid Numeric / Numeric \mid Numeric * Numeric$

### ▣ Meaning:

**Numeric**: the expression **Numeric** is evaluated on the current cell

**( Tuple, Numeric )**: the expression **Numeric** is evaluated on the cell **Tuple**

# Calculations in SSDT + Excel

37

**Calculation Items Pane**

- Sales
  - Freight
  - Order Quantity
  - Profitability
  - Sales Amount
  - Sales Contribution Percentage
  - Sales Quota
  - Sales Year on Year Growth
  - Tax Amount
  - Unit Price
  - Unit Price Discount Percent
- Employee
- Product
  - Overstocked Items
  - Top 10 Products By Profitability
  - Top 10 Products By Sales
  - Class
  - Size Range
  - Size Unit of Measure
  - Standard Cost

**Hierarchy Pane**

- (All)
- Beverages
- Condiments
- Confections
- Dairy Products
- Grains/Cereals
- Meat/Poultry
- Produce
- Seafood

**PivotTable**

Product SKU	Sales Contribution Percentage	Sales 3 Month Average	Sales Year on Year Growth
HL Mtn Frame - Silver, 42	2.0%	\$19,890	-11%
AWC logo cap	0.9%	\$8,970	3%
Bike wash - dissolver	0.3%	\$3,100	7%
Front Derailleur	0.5%	\$4,980	5%
Chain	0.4%	\$4,160	-6%
Classic Vest, L	0.3%	\$3,170	-16%
HL Mountain Pedal	0.9%	\$9,010	20%
Headlights - dual-beam	0.4%	\$4,070	5%
HL Mtn Frame - Black, 38	0.8%	\$7,680	45%
Front Brakes	0.7%	\$7,680	-4%

**Formula 1 (Sales 3 Month Average):**  
`SUM([Calendar By Month].CurrentMember.Lag(3):[Calendar By Month].CurrentMember).[Measures].[Sales Amount]/3`

**Formula 2 (Top 10 Products By Profitability):**  
`TopCount([Product].[Product SKU].Members , 10, [Measures].[Profitability] )`

# Exercise on the FoodMart cube

38

- Re-do the explorative data analysis exercise (queries Q0-Q3) using MDX instead of Excel/BIDS
- Extra queries
  - Q1 extra: top 5 categories wrt sales since 1 Jan 1998 in CA in March 1998
  - extra: how many cities per sales region had more than 4.000 dollars of total sales in March 1998