

OLAP Systems and Multidimensional Queries II

Krzysztof Dembczyński

Intelligent Decision Support Systems Laboratory (IDSS)
Poznań University of Technology, Poland



Software Development Technologies
Master studies, second semester
Academic year 2014/15 (winter course)

Review of the previous lectures

- Mining of massive datasets
- Evolution of database systems: operational vs. analytical systems.
- Dimensional modeling.
- Operational vs. analytical systems.
- Extraction, transformation and load of data.

Review of the previous lectures

- Mining of massive datasets
- Evolution of database systems: operational vs. analytical systems.
- Dimensional modeling.
- Operational vs. analytical systems.
- Extraction, transformation and load of data.
- OLAP Systems and multidimensional queries I:

Review of the previous lectures

- Mining of massive datasets
- Evolution of database systems: operational vs. analytical systems.
- Dimensional modeling.
- Operational vs. analytical systems.
- Extraction, transformation and load of data.
- OLAP Systems and multidimensional queries I:
 - ▶ ROLAP → SQL

Outline

- 1 Motivation
- 2 MOLAP
- 3 MDX
- 4 Summary

Outline

① Motivation

② MOLAP

③ MDX

④ Summary

Multidimensional reports

- OLAP servers provide an effective solution for accessing and processing large volumes of high dimensional data.
- OLAP systems provide tools for multidimensional reporting.

	All	F	M
All Products	60,398	30,017	30,381
Accessory	23,358	11,570	11,788
Mountain	16,898	8,399	8,499
Road	15,552	7,756	7,796
Touring	4,590	2,292	2,298

Operators in multidimensional data model

- Roll up – summarize data along a dimension hierarchy.
- Drill down – go from higher level summary to lower level summary or detailed data.
- Slice and dice – corresponds to selection and projection.
- Pivot – reorient cube.
- Raking, Time functions, etc..

Toronto				
New York	818	7462	43	591
Chicago	1087	968	38	872
Vancouver	854	882	89	623
Q1	605	825	14	400
Q2	680	952	31	512
Q3	812	1023	30	501
Q4	927	1038	38	580
	TV	Computer	Phone	Security

Querying OLAP systems

- We can use relational databases and SQL for OLAP applications:

```
SELECT Name, AVG(Grade)
FROM Students_grades G, Student S
WHERE G.Student = S.ID
GROUP BY Name;
```

Querying OLAP systems

- We can use relational databases and SQL for OLAP applications:

```
SELECT Name, AVG(Grade)
FROM Students_grades G, Student S
WHERE G.Student = S.ID
GROUP BY Name;
```

- But the query result is a relational table, not a multidimensional report.

Querying OLAP systems

- We can use relational databases and SQL for OLAP applications:

```
SELECT Name, AVG(Grade)
FROM Students_grades G, Student S
WHERE G.Student = S.ID
GROUP BY Name;
```

- But the query result is a relational table, not a multidimensional report.
- Relational model does not exactly match the multidimensional view on data.

Querying OLAP systems

- We can use relational databases and SQL for OLAP applications:

```
SELECT Name, AVG(Grade)
FROM Students_grades G, Student S
WHERE G.Student = S.ID
GROUP BY Name;
```

- But the query result is a relational table, not a multidimensional report.
- Relational model does not exactly match the multidimensional view on data.
- We need systems and query languages for multidimensional reporting.

Querying OLAP systems

- We can use relational databases and SQL for OLAP applications:

```
SELECT Name, AVG(Grade)
FROM Students_grades G, Student S
WHERE G.Student = S.ID
GROUP BY Name;
```

- But the query result is a relational table, not a multidimensional report.
- Relational model does not exactly match the multidimensional view on data.
- We need systems and query languages for multidimensional reporting.
- **MOLAP** (Multidimensional OLAP) and **MDX** (Multidimensional Expressions).

Outline

- 1 Motivation
- 2 MOLAP
- 3 MDX
- 4 Summary

MOLAP

- **MOLAP Servers** use array-based multidimensional storage engines.
- Optimization techniques:
 - ▶ Two-level storage representation: dense cubes are identified and stored as array structures, sparse cubes employ compression techniques,
 - ▶ Materialized cubes.

MOLAP

Example

- Logical model consists of four dimensions: customer, product, location, and day
- In case of 100 000 customers, 10 000 products, 1 000 locations and 1 000 days, the data cube will contain 1 000 000 000 000 000 cells!
- Huge number of cells is empty: a customer is not able to buy all products in all locations

MOLAP

- Data can be naturally stored with aggregates.

	Products	Mountain	Road	Touring	All
Time	2010	5076	4005	3560	12641
	2011	6503	4503	3445	14451
	All	11579	8508	7005	27092

MOLAP

- Advantages of MOLAP Servers:
 - ▶ Multidimensional views are directly mapped to data cube array structures – efficient access to data,
 - ▶ Can easily store subaggregates.
- Disadvantages of MOLAP Servers:
 - ▶ Scalability problem in the case of larger number of dimensions,
 - ▶ Not tailored for sparse data,
 - ▶ Young technology,
 - ▶ There are no existing standards.

HOLAP

- **HOLAP Servers** are a hybrid approach that combines ROLAP and MOLAP technology.
- HOLAP benefits from the greater scalability of ROLAP and the faster computation of MOLAP.

Outline

- ① Motivation
- ② MOLAP
- ③ MDX**
- ④ Summary

Multidimensional expressions

- For OLAP queries, MDX is an alternative to SQL.

Students' grades	ID	Academic year	Student	Course	Professor	Grade
	01	D1	S4	C1	P1	3.9
	02	D1	S4	C1	P2	4.0
	03	D1	S5	C2	P1	4.4
	04	D1	S5	C2	P2	4.4
	01	D2	S4	C1	P1	3.5
	02	D2	S4	C1	P2	4.0
	03	D2	S5	C2	P1	4.1
	04	D2	S5	C2	P2	4.0
	05	D3	S6	C1	P1	3.6
	06	D3	S6	C1	P2	3.9
	07	D3	S7	C2	P3	4.8

Avg(Grade)
by Academic year
and Professor

Academic year/ Professor	P1	P2	P3
D1	4.1	4.2	
D2	3.8	4.0	
D3	3.6	3.9	4.8

Multidimensional expressions

- For OLAP queries, MDX is an alternative to SQL.

MDX

```
SELECT {[Time].[1997],[Time].[1998]} ON COLUMNS,  
{[Measures].[Sales],[Measures].[Cost]} ON ROWS  
FROM Warehouse  
WHERE ([Store].[All Stores].[USA])
```

Multidimensional expressions (MDX)

- An MDX query must contain the following information:
 - ▶ The number of **axes** on which the result is presented.
 - ▶ The **set** of **tuples** to include on each axis of the MDX query.
 - ▶ The name of the **cube** that sets the context of the MDX query.
 - ▶ The set of members or tuples to include on the **slicer** axis.

Multidimensional expressions (MDX)

- Main concepts of MDX:
 - ▶ Dimension,
 - ▶ Hierarchy,
 - ▶ Level,
 - ▶ Member,
 - ▶ Member property,
 - ▶ Measure,
 - ▶ Tuple,
 - ▶ Set,
 - ▶ Axis.

SQL vs. MDX

- Single member
 - ▶ SQL: where City = 'Redmond'
 - ▶ MDX: [City].[Redmond]
- Multiple members (a set)
 - ▶ SQL: where City IN ('Redmond', 'Seattle')
 - ▶ MDX: { ([City].[Redmond]), ([City].[Seattle]) }

SQL vs. MDX

- SQL:

```
SELECT Sum(Sales), City FROM Sales_Table  
WHERE City IN ('Redmond', 'Seattle')  
GROUP BY City
```

- MDX:

SQL vs. MDX

- SQL:

```
SELECT Sum(Sales), City FROM Sales_Table  
WHERE City IN ('Redmond', 'Seattle')  
GROUP BY City
```

- MDX:

```
SELECT Measures.Sales ON 0,  
NON EMPTY {[City].[Redmond]}, {[City].[Seattle]} ON 1  
FROM Sales_Cube
```

SQL vs. MDX

- SQL:

```
SELECT Sum(Sales) FROM Sales_Table  
WHERE City IN ('Redmond', 'Seattle')
```

- MDX:

SQL vs. MDX

- SQL:

```
SELECT Sum(Sales) FROM Sales_Table  
WHERE City IN ('Redmond', 'Seattle')
```

- MDX:

```
SELECT Measures.Sales ON 0  
FROM Sales_Cube  
WHERE {[City].[Redmond]}, {[City].[Seattle]}
```

Multidimensional expressions (MDX)

MDX

```
SELECT  
{[CARS].[All CARS].[Chevy], [CARS].[All CARS].[Ford]} ON ROWS,  
{[DATE].[All DATE].[March], [DATE].[All DATE].[April]} ON COLUMNS  
FROM MDDBCARS;
```

Multidimensional expressions (MDX)

MDX

```
SELECT  
{[CARS].[All CARS].[Chevy], [CARS].[All CARS].[Ford]} ON ROWS,  
{[DATE].[All DATE].[March], [DATE].[All DATE].[April]} ON COLUMNS  
FROM MDDBCARS;
```

	March	April
Chevy	\$155 000.00	\$ 75 000.00
Ford	\$ 55 000.00	\$175 000.00

Multidimensional expressions (MDX)

MDX

```
SELECT
{[CARS].[ALL CARS].[CHEVY], [CARS].[ALL CARS].[FORD]} ON ROWS,
{[DATE].[ALL DATE].[MARCH], [DATE].[ALL DATE].[APRIL]} ON COLUMNS
FROM MDDBCARS
WHERE ([MEASURES].[SALES_N])
```


Multidimensional expressions (MDX)

MDX

```
SELECT
{[CARS].[ALL CARS].[CHEVY], [CARS].[ALL CARS].[FORD]} ON ROWS,
{[DATE].[ALL DATE].[MARCH], [DATE].[ALL DATE].[APRIL]} ON COLUMNS
FROM MDDBCARS
WHERE ([MEASURES].[SALES_N])
```

Sales_N	March	April
Chevy	1 000	700
Ford	600	1 500

Multidimensional expressions (MDX)

MDX

```
SELECT  
{[CARS].[ALL CARS].[CHEVY], [CARS].[ALL CARS].[FORD]} ON COLUMNS,  
{[DATE].[ALL DATE].[JANUARY]:[DATE].[ALL DATE].[APRIL]} ON ROWS  
FROM MDDBCARS
```

Multidimensional expressions (MDX)

MDX

```
SELECT
{[CARS].[ALL CARS].[CHEVY], [CARS].[ALL CARS].[FORD]} ON COLUMNS,
{[DATE].[ALL DATE].[JANUARY]:[DATE].[ALL DATE].[APRIL]} ON ROWS
FROM MDDBCARS
```

	Chevy	Ford
January	\$66 000.00	\$ 79 000.00
February	\$55 000.00	\$ 72 000.00
March	\$155 000.00	\$ 55 000.00
April	\$ 75 000.00	\$175 000.00

Multidimensional expressions (MDX)

MDX

```
SELECT  
{[CARS].[ALL CARS].[CHEVY], [CARS].[ALL CARS].[FORD]} ON COLUMNS,  
{[DATE].[ALL DATE].[YEAR].MEMBERS} ON ROWS  
FROM MDDBCARS
```

Multidimensional expressions (MDX)

MDX

```
SELECT  
{[CARS].[ALL CARS].[CHEVY], [CARS].[ALL CARS].[FORD]} ON COLUMNS,  
{[DATE].[ALL DATE].[YEAR].MEMBERS} ON ROWS  
FROM MDDBCARS
```

	Chevy	Ford
1998	\$566 000.00	\$ 479 000.00
1999	\$545 000.00	\$ 672 000.00
2000	\$745 000.00	\$ 527 000.00
2001	\$345 000.00	\$ 622 000.00

Multidimensional expressions (MDX)

MDX

```
SELECT  
{ [CARS] . [ALL CARS] . [FORD] . CHILDREN } ON COLUMNS,  
{ [DATE] . [ALL DATE] . [YEAR] . MEMBERS } ON ROWS  
FROM MDDBCARS
```

Multidimensional expressions (MDX)

MDX

```
SELECT  
{ [CARS] . [ALL CARS] . [FORD] . CHILDREN } ON COLUMNS,  
{ [DATE] . [ALL DATE] . [YEAR] . MEMBERS } ON ROWS  
FROM MDDBCARS
```

	Ford Mustang	Ford Taurus	...
1998	\$56 000.00	\$ 79 000.00	
1999	\$54 000.00	\$ 72 000.00	
2000	\$72 000.00	\$ 52 000.00	
2001	\$34 000.00	\$ 22 000.00	

Multidimensional expressions (MDX)

MDX

```
SELECT {([CARS].[ALL CARS].[CHEVY], [MEASURES].[SALES_SUM]),  
([CARS].[ALL CARS].[CHEVY], [MEASURES].[SALES_N]),  
([CARS].[ALL CARS].[FORD], [MEASURES].[SALES_SUM]),  
([CARS].[ALL CARS].[FORD], [MEASURES].[SALES_N])  
} ON COLUMNS,  
{[DATE].[ALL DATE].[YEAR].MEMBERS} ON ROWS  
FROM MDDBCARS
```


Multidimensional expressions (MDX)

MDX

```
SELECT {([CARS].[ALL CARS].[CHEVY], [MEASURES].[SALES_SUM]),  
([CARS].[ALL CARS].[CHEVY], [MEASURES].[SALES_N]),  
([CARS].[ALL CARS].[FORD], [MEASURES].[SALES_SUM]),  
([CARS].[ALL CARS].[FORD], [MEASURES].[SALES_N])  
} ON COLUMNS,  
{[DATE].[ALL DATE].[YEAR].MEMBERS} ON ROWS  
FROM MDDBCARS
```

	Chevy		Ford	
	Sales_Sum	Sales_N	Sales_Sum	Sales_N
1998	\$566 000.00	450	\$ 479 000.00	450
1999	\$545 000.00	475	\$ 672 000.00	670
2000	\$745 000.00	750	\$ 527 000.00	490
2001	\$345 000.00	325	\$ 622 000.00	640

Multidimensional expressions (MDX)

MDX

```
SELECT
{CROSSJOIN({[CARS].[ALL CARS].[CHEVY], [CARS].[ALL CARS].[FORD]},
{[MEASURES].[SALES_SUM], [MEASURES].[SALES_N]})
} ON COLUMNS,
{[DATE].[ALL DATE].[YEAR].MEMBERS} ON ROWS
FROM MDDBCARS
```

Multidimensional expressions (MDX)

MDX

```
SELECT
{CROSSJOIN({[CARS].[ALL CARS].[CHEVY], [CARS].[ALL CARS].[FORD]},
{[MEASURES].[SALES_SUM], [MEASURES].[SALES_N]})
} ON COLUMNS,
{[DATE].[ALL DATE].[YEAR].MEMBERS} ON ROWS
FROM MDDBCARS
```

	Chevy		Ford	
	Sales_Sum	Sales_N	Sales_Sum	Sales_N
1998	\$566 000.00	450	\$ 479 000.00	450
1999	\$545 000.00	475	\$ 672 000.00	670
2000	\$745 000.00	750	\$ 527 000.00	490
2001	\$345 000.00	325	\$ 622 000.00	640

Multidimensional expressions (MDX)

MDX

```
SELECT NON EMPTY [Store Type].[Store Type].MEMBERS ON COLUMNS,  
FILTER([Store].[Store City].MEMBERS, (Measures.[Profit], [Time].[1997])  
> 250000) ON ROWS FROM [Sales] WHERE (Measures.[Profit],  
[Time].[Year].[1997])
```

Multidimensional expressions (MDX)

MDX

```
SELECT NON EMPTY [Store Type].[Store Type].MEMBERS ON COLUMNS,  
FILTER([Store].[Store City].MEMBERS, (Measures.[Profit], [Time].[1997])  
> 250000) ON ROWS FROM [Sales] WHERE (Measures.[Profit],  
[Time].[Year].[1997])
```

Profit	Normal	24 Hours
Toronto	\$66 000.00	\$196 000.00
Vancouver	\$111 000.00	\$156 000.00
New York	\$59 000.00	\$196 000.00
Chicago	\$75 000.00	\$211 000.00

Multidimensional expressions (MDX)

MDX

```
SELECT Measures.MEMBERS ON COLUMNS,  
ORDER({[Store].[Store City].MEMBERS}, Measures.[Sales Count], DESC) ON  
ROWS  
FROM [Sales]
```

Multidimensional expressions (MDX)

MDX

```
SELECT Measures.MEMBERS ON COLUMNS,  
ORDER({[Store].[Store City].MEMBERS}, Measures.[Sales Count], DESC) ON  
ROWS  
FROM [Sales]
```

	Profit	Sales Count	...
New York	\$747 000.00	2 196 000	
Chicago	\$785 000.00	1 956 000	
Toronto	\$666 000.00	1 916 000	
Vancouver	\$711 000.00	1 596 000	

Multidimensional expressions (MDX)

MDX

```
WITH  
MEMBER [Time].[Year Difference] AS [Time].[2nd half] - [Time].[1st half]  
SELECT { [Account].[Income], [Account].[Expenses] } ON COLUMNS,  
{ [Time].[1st half], [Time].[2nd half], [Time].[Year Difference] } ON  
ROWS  
FROM [Financials]
```


Multidimensional expressions (MDX)

MDX

```
WITH  
MEMBER [Time].[Year Difference] AS [Time].[2nd half] - [Time].[1st half]  
SELECT { [Account].[Income], [Account].[Expenses] } ON COLUMNS,  
{ [Time].[1st half], [Time].[2nd half], [Time].[Year Difference] } ON  
ROWS  
FROM [Financials]
```

	Income	Expenses
1st Half	5 000	4 200
2st Half	8 000	7 000
Year Difference	3 000	2 800

Multidimensional expressions (MDX)

MDX

```
WITH  
MEMBER [Account].[Net Income] AS  
([Account].[Income] - [Account].[Expenses]) / [Account].[Income]  
SELECT { [Account].[Income], [Account].[Expenses], [Account].[Net  
Income] } ON COLUMNS,  
{ [Time].[1st half], [Time].[2nd half] } ON ROWS  
FROM [Financials]
```

Multidimensional expressions (MDX)

MDX

```
WITH  
MEMBER [Account].[Net Income] AS  
  ([Account].[Income] - [Account].[Expenses]) / [Account].[Income]  
SELECT { [Account].[Income], [Account].[Expenses], [Account].[Net  
Income] } ON COLUMNS,  
{ [Time].[1st half], [Time].[2nd half] } ON ROWS  
FROM [Financials]
```

	Income	Expenses	Net Income
1st Half	5 000	4 200	0.16
2st Half	8 000	7 000	0.125

Multidimensional expressions (MDX)

MDX

```
WITH
MEMBER [Time].[Year Difference] AS
' [Time].[2nd half] - [Time].[1st half],
SOLVE_ORDER = 1
MEMBER [Account].[Net Income] AS
'([Account].[Income] - [Account].[Expenses]) / [Account].[Income]',
SOLVE_ORDER = 2
SELECT
{ [Account].[Income], [Account].[Expenses], [Account].[Net Income] } ON
COLUMNS,
{ [Time].[1st half], [Time].[2nd half], [Time].[Year Difference] } ON
ROWS
FROM [Financials]
```

Multidimensional expressions (MDX)

	Income	Expenses	Net Income
1st Half	5 000	4 200	0.16
2st Half	8 000	7 000	0.125
Year Difference	3 000	2 800	0.066

Multidimensional expressions (MDX)

MDX

```
WITH
MEMBER [Time].[Year Difference] AS
' [Time].[2nd half] - [Time].[1st half],
SOLVE_ORDER = 2
MEMBER [Account].[Net Income] AS
'([Account].[Income] - [Account].[Expenses]) / [Account].[Income]',
SOLVE_ORDER = 1
SELECT
{ [Account].[Income], [Account].[Expenses], [Account].[Net Income] } ON
COLUMNS,
{ [Time].[1st half], [Time].[2nd half], [Time].[Year Difference] } ON
ROWS
FROM [Financials]
```

Multidimensional expressions (MDX)

	Income	Expenses	Net Income
1st Half	5 000	4 200	0.16
2st Half	8 000	7 000	0.125
Year Difference	3 000	2 800	-0.035

Multidimensional expressions (MDX)

MDX

```
WITH SET [Quarter1] AS  
'GENERATE([Time].[Year].MEMBERS, {[Time].CURRENTMEMBER.FIRSTCHILD})'  
SELECT [Quarter1] ON COLUMNS, [Store].[Store Name].MEMBERS ON ROWS  
FROM [Sales]  
WHERE (Measures.[Profit])
```


Multidimensional expressions (MDX)

MDX

```
WITH SET [Quarter1] AS  
'GENERATE([Time].[Year].MEMBERS, {[Time].CURRENTMEMBER.FIRSTCHILD})'  
SELECT [Quarter1] ON COLUMNS, [Store].[Store Name].MEMBERS ON ROWS  
FROM [Sales]  
WHERE (Measures.[Profit])
```

	1998Q1	1999Q1	...
Saturn	\$147 000.00	\$196 000.00	
Media Markt	\$185 000.00	\$156 000.00	
Avans	\$166 000.00	\$116 000.00	

Outline

- ① Motivation
- ② MOLAP
- ③ MDX
- ④ Summary

Summary

- Three types of OLAP servers: ROLAP, MOLAP, and HOLAP.
- Several approaches for querying data warehouses.
- ROLAP servers: **SQL** and its OLAP extensions.
- MOLAP servers: **MDX**.

Bibliography

- Mark Whitehorn, Robert Zare, and Mosha Pasumansky. *Fast Track to MDX*. Springer, 2002