REPORTING

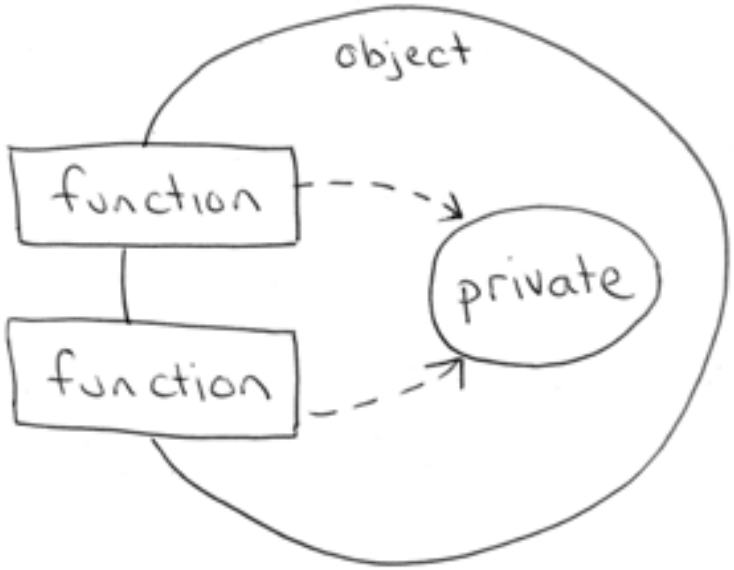
Unit 3 – Data exploitation. Query languages and visualization
S3 –2 – MDX

1

Business intelligence

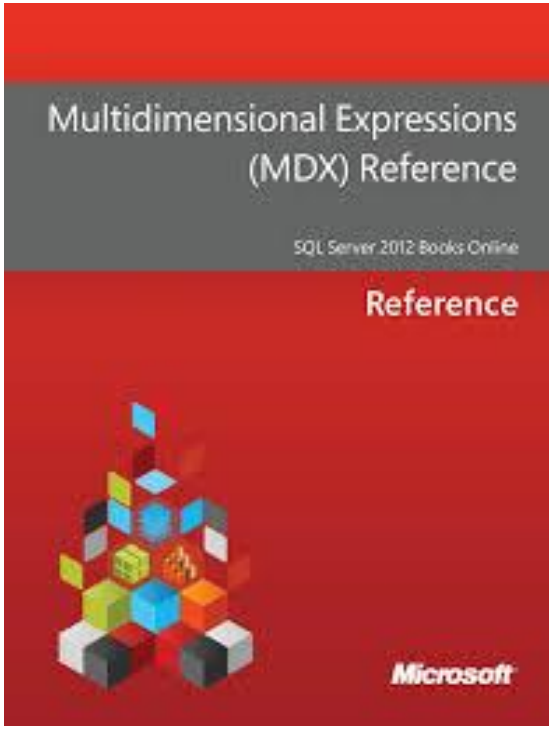
- **CUBES**
- **LANGUAGE**
- **M**ulti**D**imensional e**X**pression.
- Microsoft in 1997

- Do you remember studying Object-Oriented Programming?



Outline

1. Introduction
2. Tuples, Sets & Cells
3. MDX Spells
4. MDX Query Syntax



MDX: An Introduction

What is MDX? query language for OLAP cubes and multidimensional databases, designed to retrieve and analyze data stored in cubes.
Primarily used in MOLAP systems like Microsoft SQL Server Analysis Services (SSAS), Oracle Essbase, and IBM Cognos.

Key Features of MDX:

- Retrieval of data from multidimensional structures.
- Definition of calculated members and custom measures.
- Aggregation and slicing of data for analysis.

Why Use it?

- Advanced capabilities for querying hierarchical and multidimensional datasets, making it ideal for complex analytics in BI systems.

Tuples, Sets & Cells

- **Fact Tables:** Contain measures, which are numerical values or metrics used for analysis.
- **Dimension tables:** Contain Members. Members are values in a dimension that describe or categorize the data in the fact table. These values can be qualitative or quantitative.
- E.g.: 2 dimension cube
  - **1 measure**: discharged patients.
  - Time Dimension with 4 members: Jan to April.
  - Hospital Dimension with 4 members: H1,H2,H3,H4.

Discharged	H1	H2	H3	H4
January	20	44	81	44
February	15	32	78	32
March	23	65	88	65
April	19	67	67	67

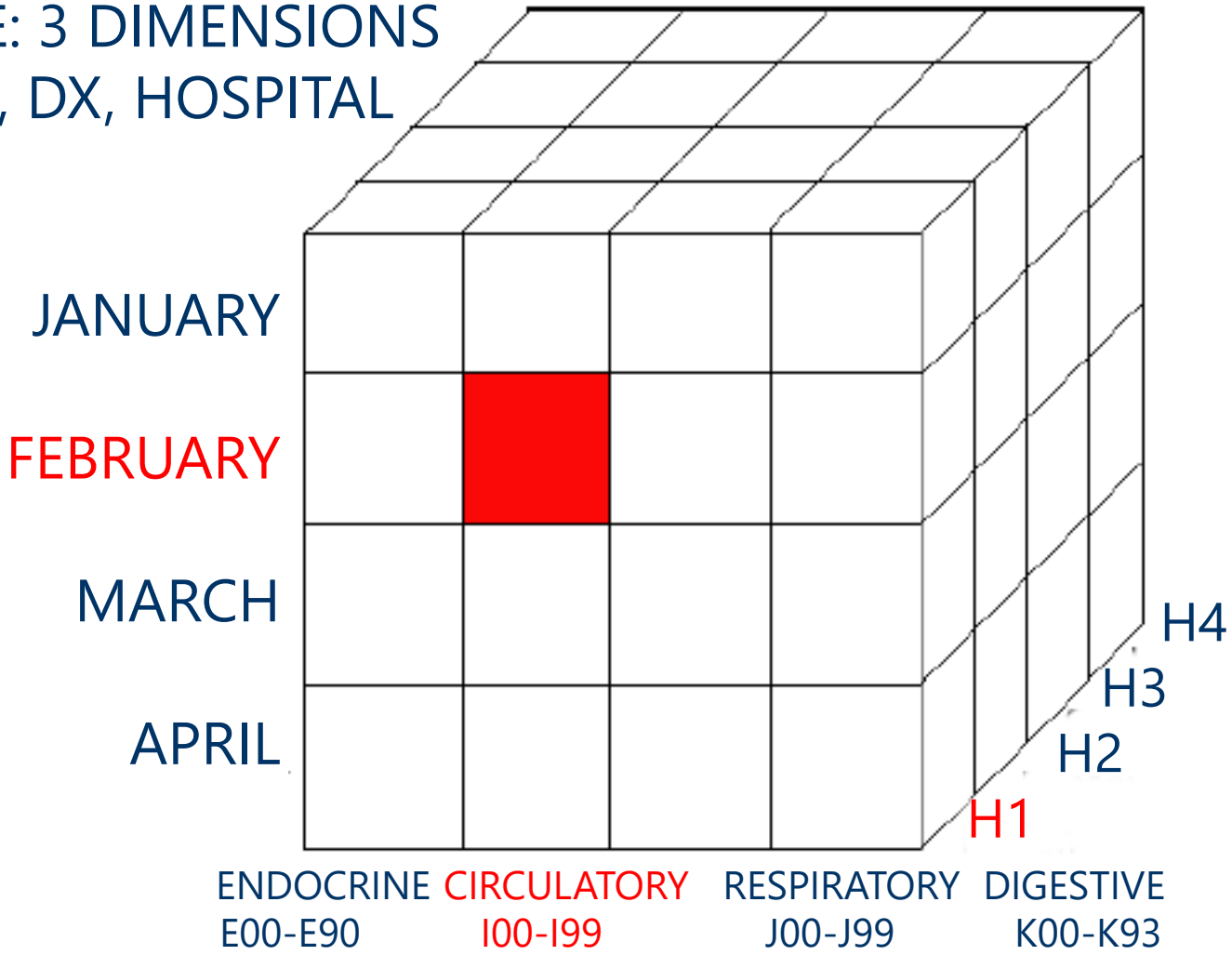
Tuples, Sets & Cells

- **Fact Tables:** Contain measures, which are numerical values or metrics used for analysis.
- **Dimension tables:** Contain Members. Members are values in a dimension that describe or categorize the data in the fact table. These values can be qualitative or quantitative.
- E.g.: 2 dimension cube
  - **2 measure**: no. discharged patients, total cost (M€).
  - Time Dimension with 4 members: Jan to April.
  - Hospital Dimension with 4 members: H1,H2,H3, H4.

Discharged	H1	H2	H3	H4
January	20 1.5M€	44 4.1M€	81 10.5M€	44 4.1M€
February	15 1.1M€	32 3.9M€	78 10.4M€	32 3.9M€
March	23 1.6M€	65 5.4M€	88 10.7M€	65 5.4M€
April	19 1.5M€	67 5.6M€	67 9.5M€	67 5.6M€

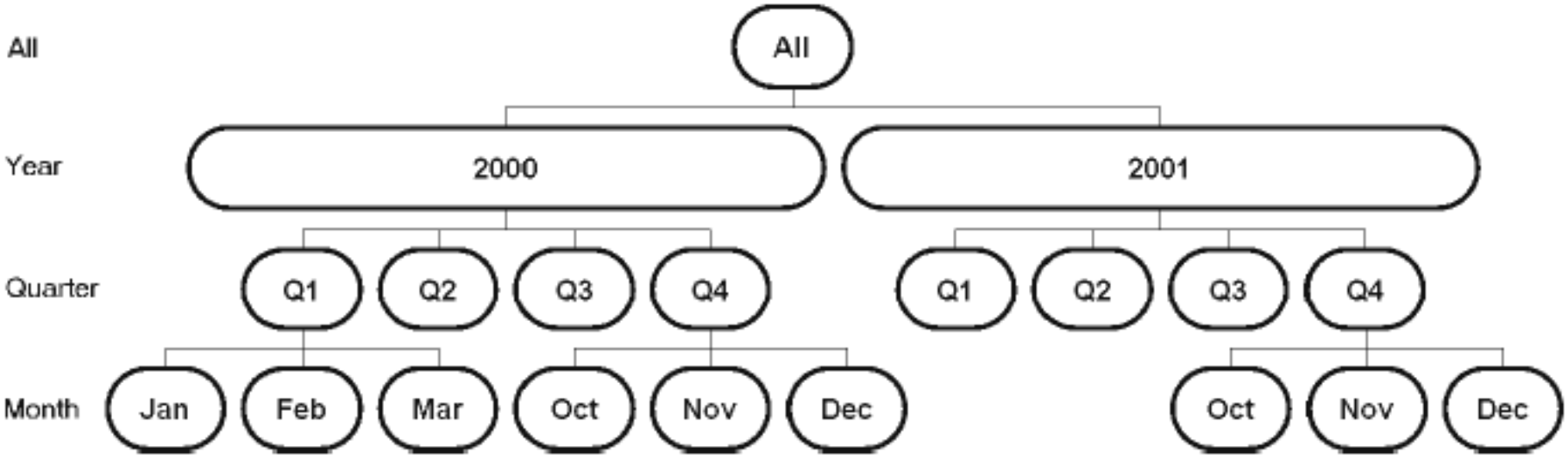
Tuples, Sets & Cells

CUBE: 3 DIMENSIONS
TIME, DX, HOSPITAL



Tuples, Sets & Cells

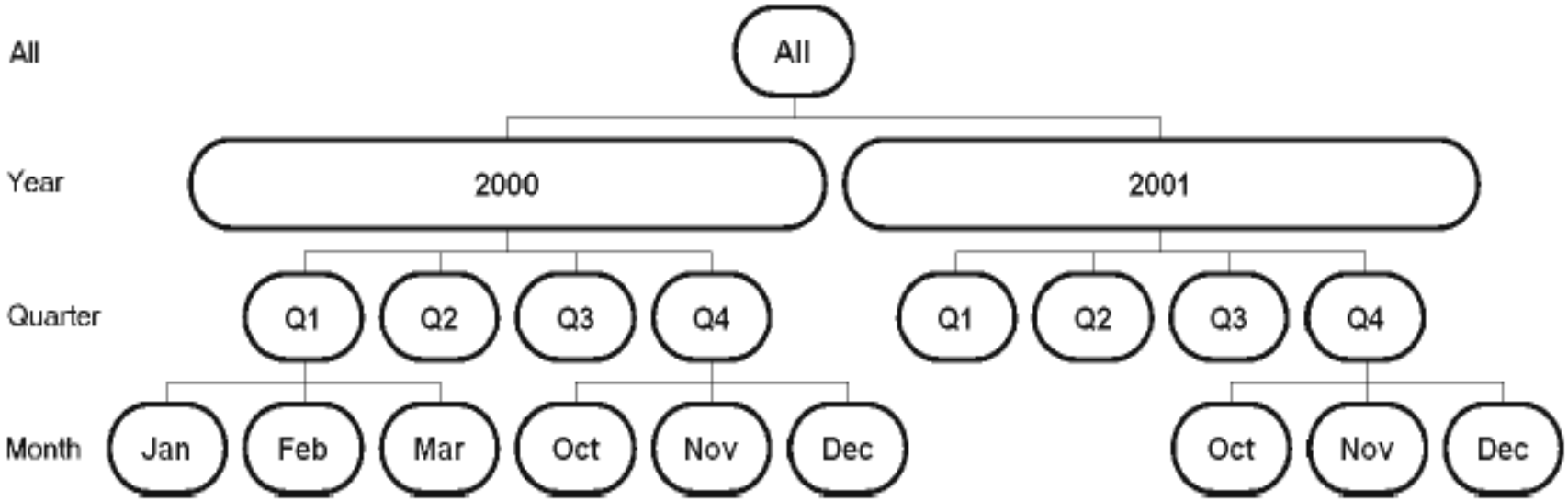
- Dimension has **hierarchies.**
- Hierarchy has **levels**: All, Year, Quarter, Month.
- A hierarchy organizes members of a dimension into multiple levels for drill-down or roll-up operations.
- All → Highest aggregation level.



Tuples, Sets & Cells

- Naming Conventions

[Time].[All].[2000].[Q4].[Oct] = [Time].[2000].[Q4].[Oct]



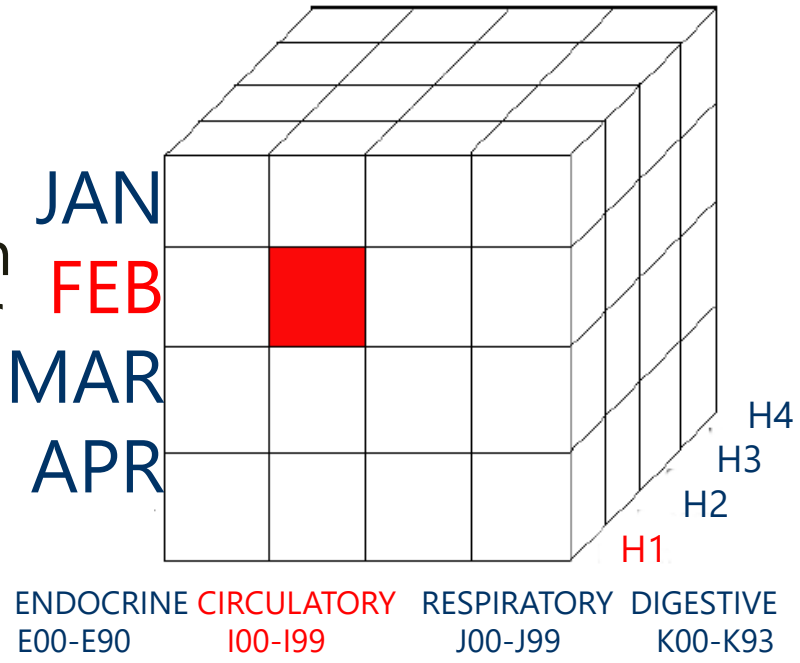
Tuples, Sets & Cells

Naming Conventions: **Tuple**
Tuple in pseudo-MDX:
([Time].[Feb],[Dx].[Circ],[Hosp].[H1])

Def1: "***Tuple*** is the intersection of **one member from each dimension**"

Def2: "A tuple is the intersection of **one (<u>and only one</u>)** member taken from one or several of the dimensions in the cube."

(tuple=single cell in the cube ??)



Tuples, Sets & Cells

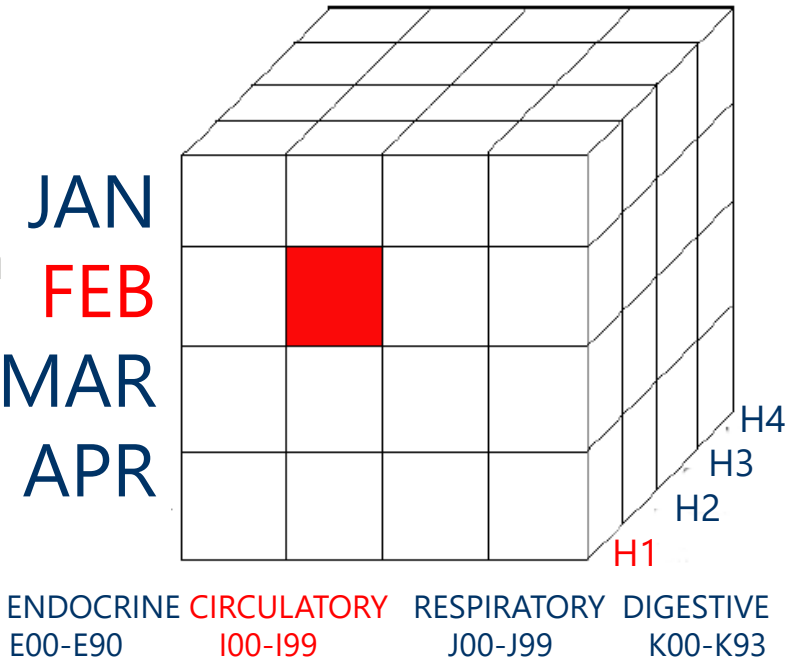
- Naming Conventions: **Tuple**

Tuple in pseudo-MDX: (x,y,z)=(y,z,x)
([Time].[Feb],[Dx].[Circ],[Hosp].[H1])

Def1: "***Tuple*** is the intersection choosing **one member of each dimension**"

Def2: "A tuple is the intersection of **one (<u>and only one</u>)** member taken from one or several of the dimensions in the cube."

(tuple=single cell in the cube ??)



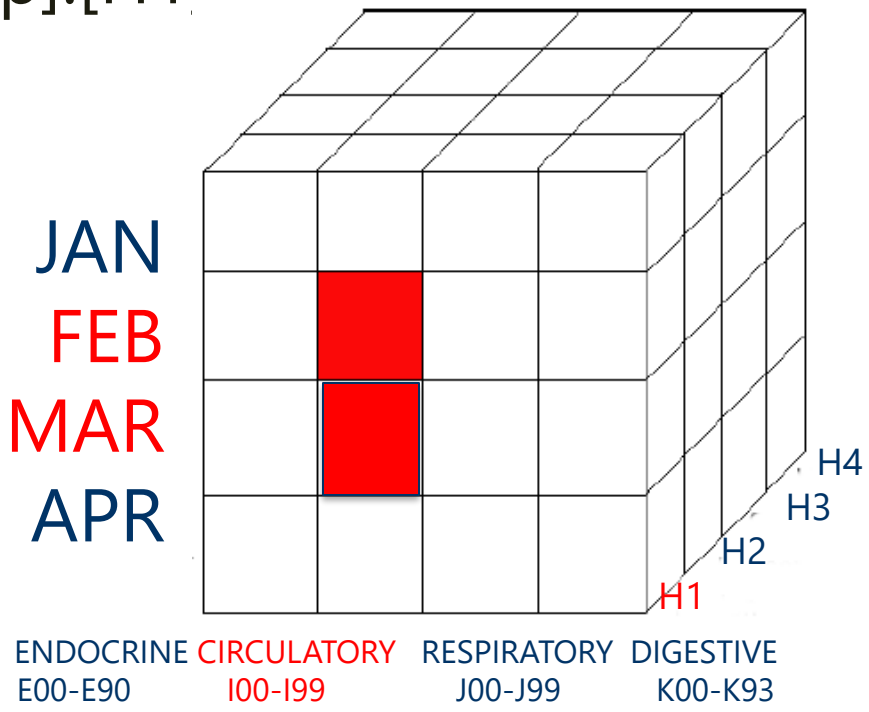
Tuples, Sets & Cells

- Naming Conventions: **Set**

Set in pseudo-MDX: {(x1,y1,z1),..,(xn,yn,zn)}
{([Time].[Feb],[Dx].[Circ],[Hosp].[H1]),
([Time].[Mar],[Dx].[Circ],[Hosp].[H1])}

"***Set*** is a set of tuples with the same dimensionality"
(set of cells in the cube)

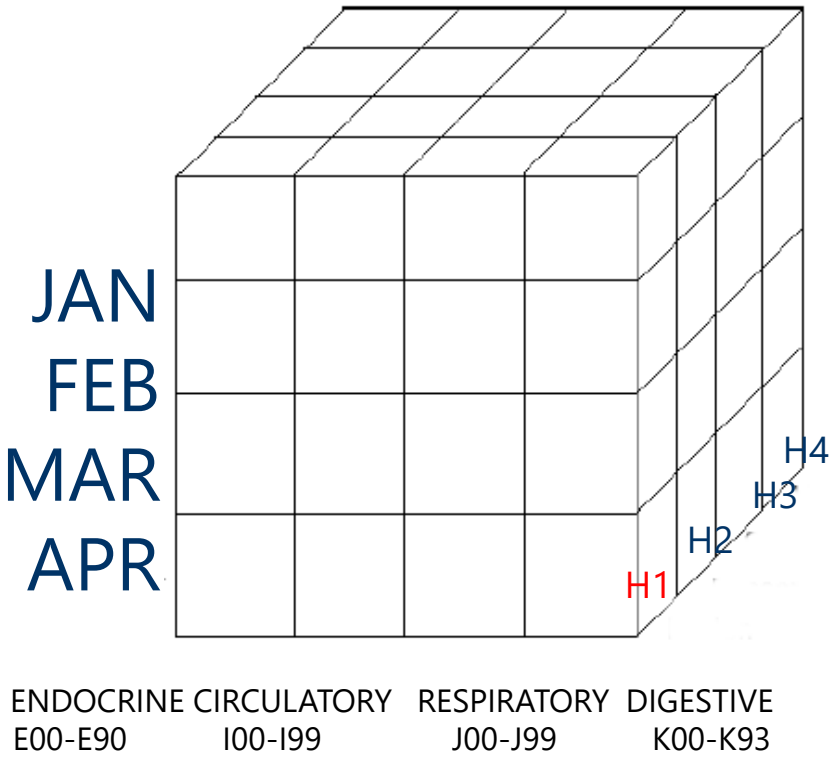
A set can be aggregated in MDX:
AVG(SET)→FLOAT



QUESTION 1: Tuples, Sets & Cells

- Question: Tuple or Set?

[DX].[Circulatory],[Hosp].[H1]

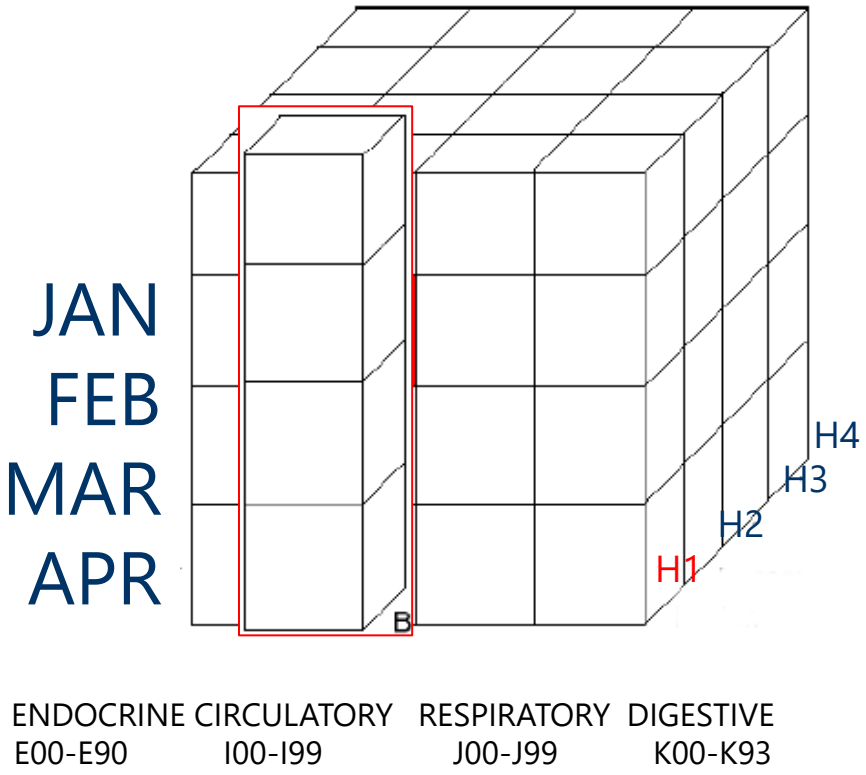


QUESTION 1:Tuples, Sets & Cells

- Question: Tuple or Set?

[DX].[Circulatory],[Hosp].[H1]

Is a TUPLE!
(but MANY CELLS!)



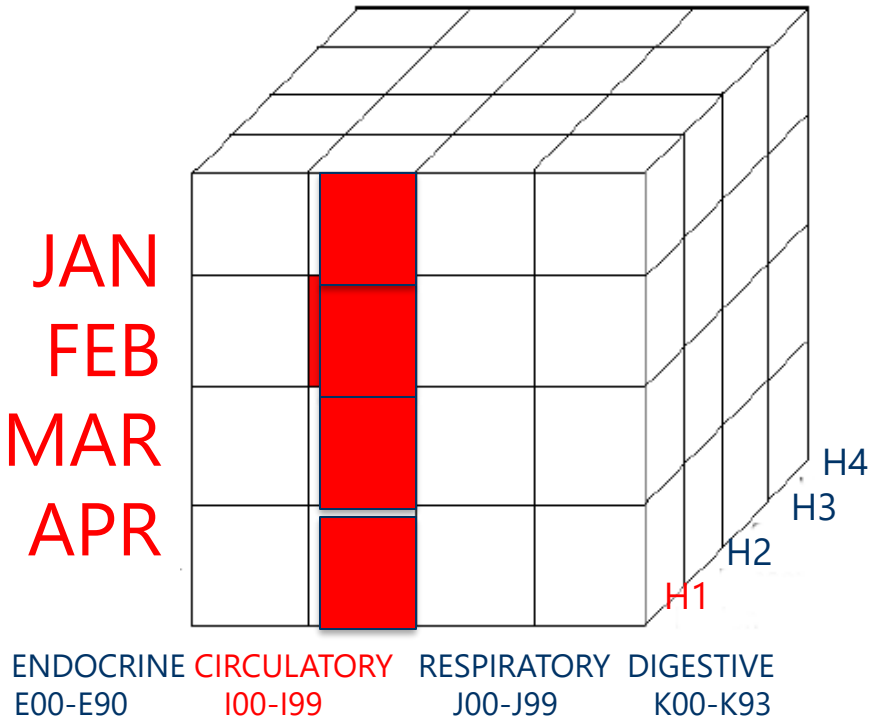
QUESTION 2: Tuples, Sets & Cells

- Question: Difference between?

a) ([Dx].[Circ],[Hosp].[H1])

b)

{([Dx].[Circ],[Hosp].[H1],[Time].[Jan]),
([Dx].[Circ],[Hosp].[H1],[Time].[Feb]),
([Dx].[Circ],[Hosp].[H1],[Time].[Mar]),
([Dx].[Circ],[Hosp].[H1],[Time].[Apr])
}



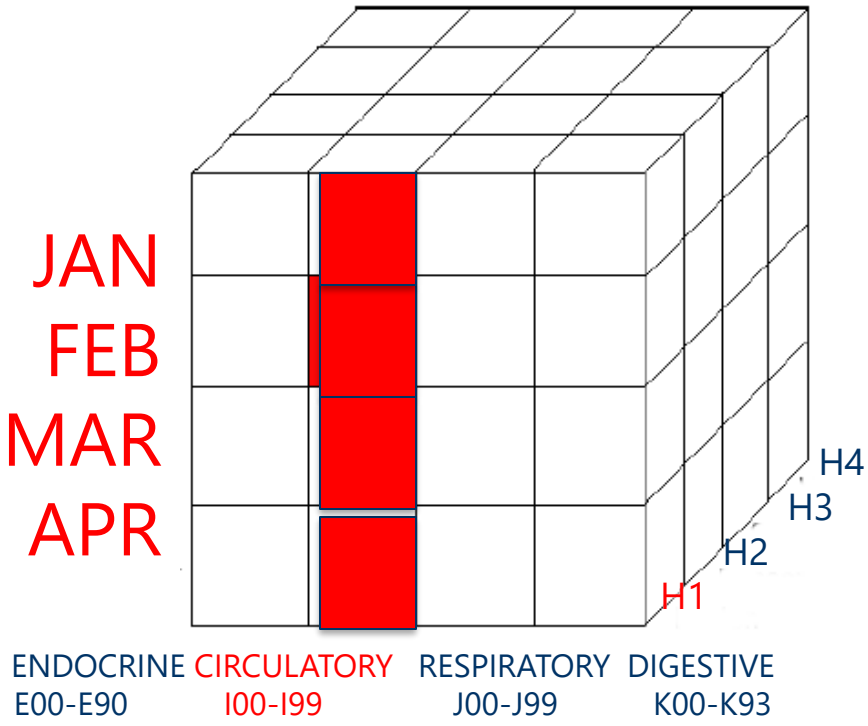
QUESTION 2: Tuples, Sets & Cells

- Question: Difference between?

a) ([Dx].[Circ],[Hosp].[H1]) **IS A TUPLE** (SEE DEFINITION 1)

b)

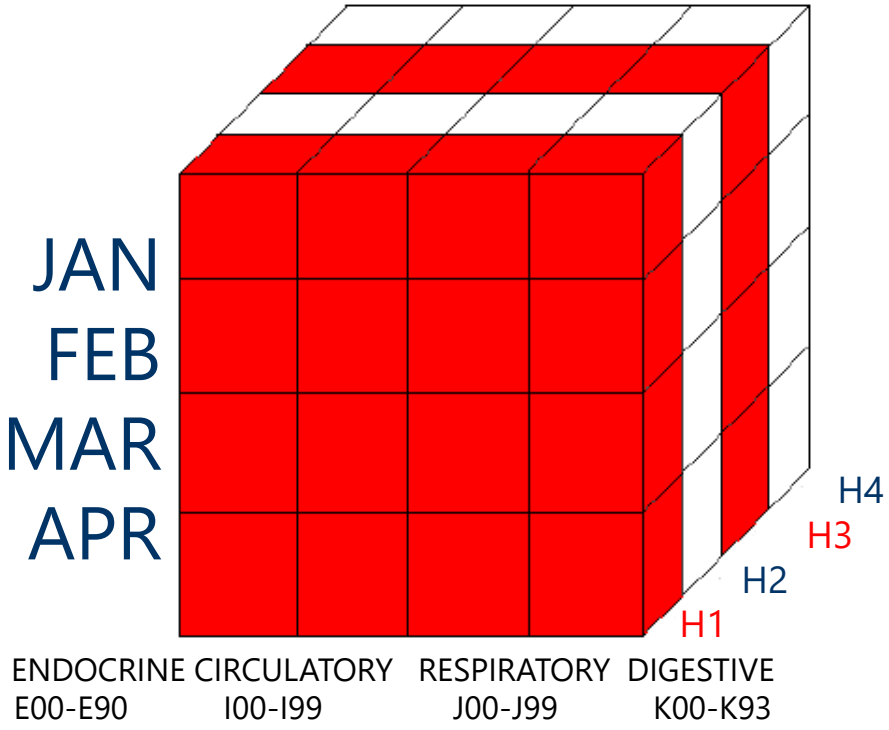
{([Dx].[Circ],[Hosp].[H1],[Time].[Jan]),
([Dx].[Circ],[Hosp].[H1],[Time].[Feb]),
([Dx].[Circ],[Hosp].[H1],[Time].[Mar]),
([Dx].[Circ],[Hosp].[H1],[Time].[Apr])
} **IS A SET**



QUESTION 3: Tuples, Sets & Cells

- Question: Tuple or Set?

[Hosp].[H1], [Hosp].[H3]

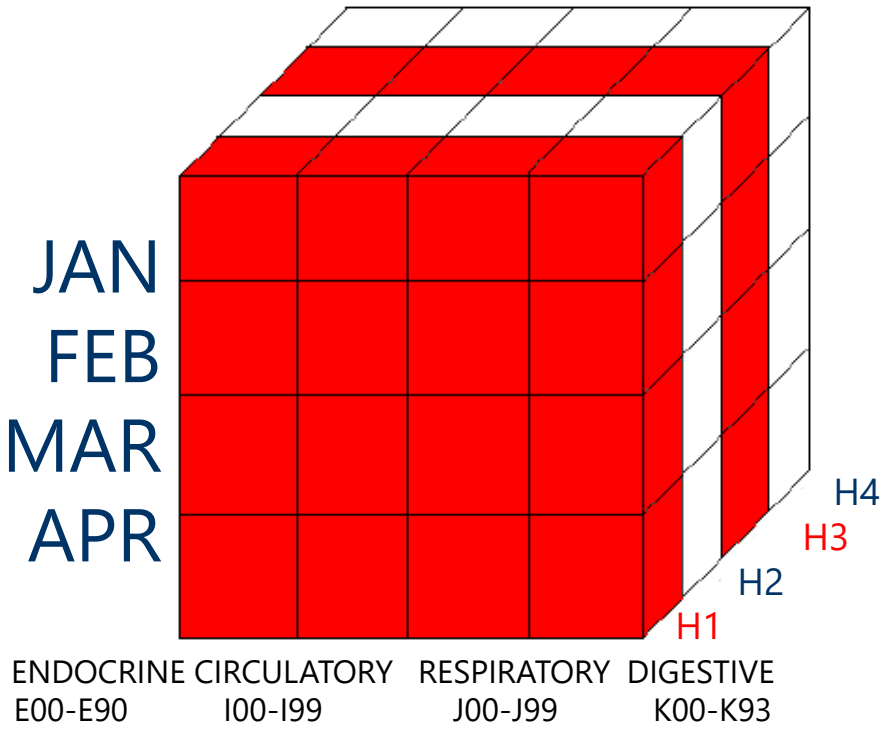


QUESTION 3: Tuples, Sets & Cells

- Question: Tuple or Set?

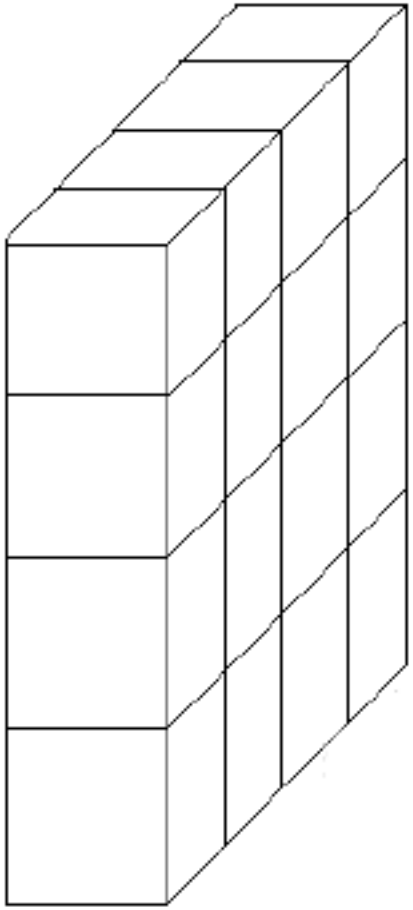
[Hosp].[H1], [Hosp].[H3]

is a Set
(see Def2!)

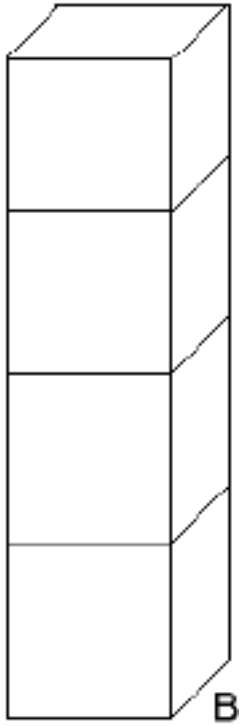


RECAP: Tuples, Sets & Cells

[DX].[Circulatory]



[DX].[Circulatory],[Hosp].[H1]

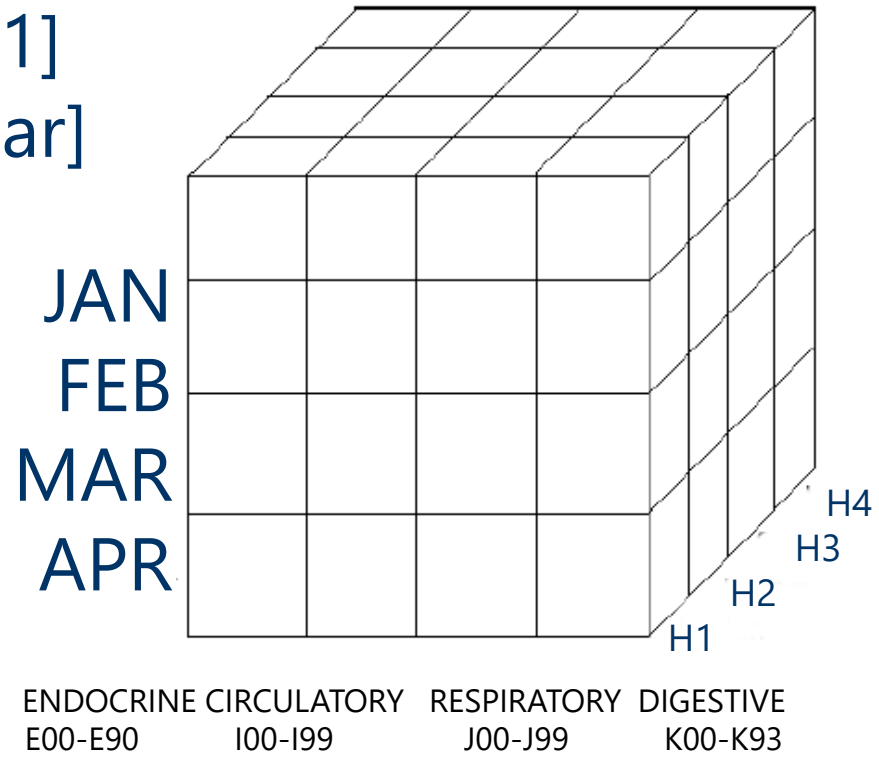


ALL THESE ARE TUPLES SINCE THEY HAVE THE "CAPACITY TO POINT TO A SINGLE CELL"
(actually they don't)

QUESTION 4: Tuples, Sets & Cells

- Question: Do these tuples point to a single cell?

[DX].[Circulatory],[Hosp].[H1], [Time].[Mar]
[DX].[Circulatory],[Hosp].[H1]
[DX].[Circulatory],[Time].[Mar]
[Hosp].[H1]



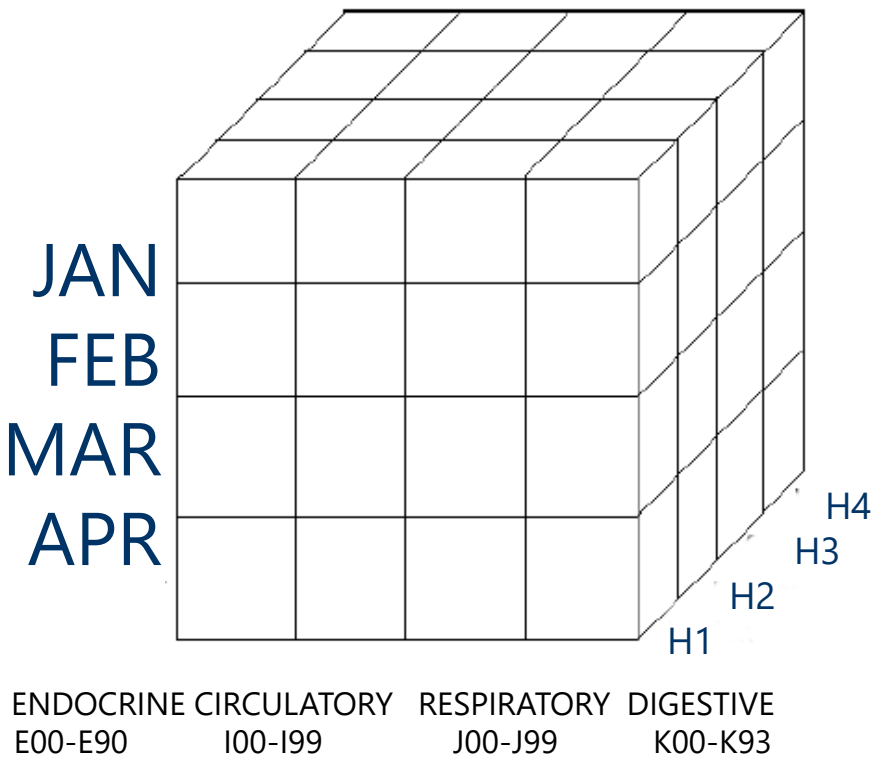
QUESTION 4: Tuples, Sets & Cells

- Question: Do these tuples point to a single cell?

[DX].[Circulatory],[Hosp].[H1], [Time].[Mar]
[DX].[Circulatory],[Hosp].[H1]
[DX].[Circulatory],[Time].[Mar]
[Hosp].[H1]

(or an aggregated cell)…
YES

If we consider that all dimensions have a '*DEFAULT MEMBER*'
In MDX if you don't specify a member of a dimension the default member is implied



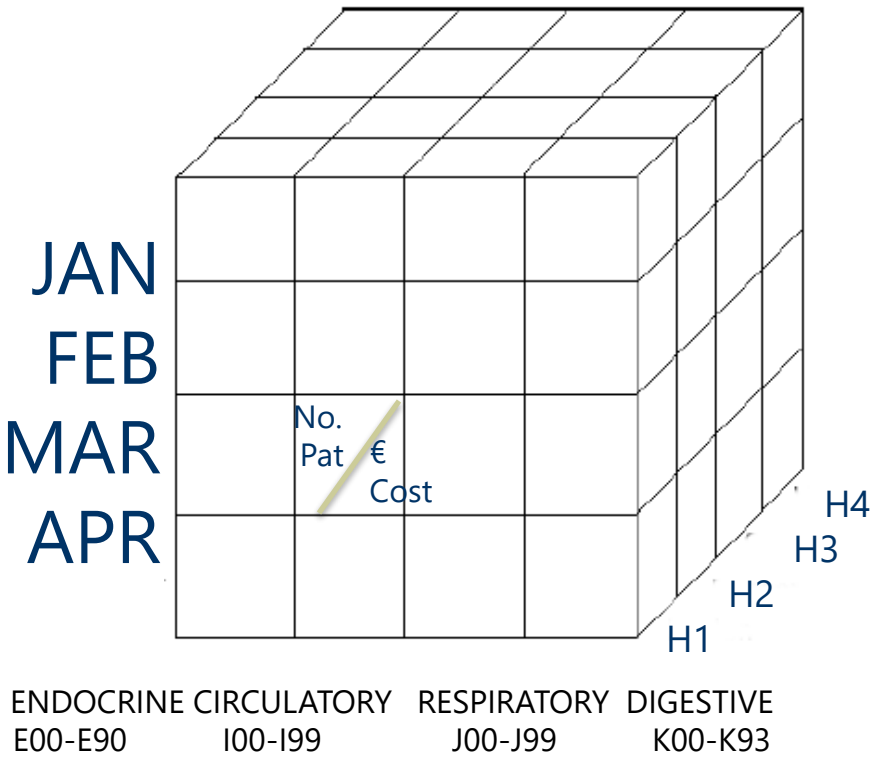
WHAT IF…

- Measures like dimensions

Suppose a cube **with 2 measures**:
No.Patients and €Cost.

[Hos].[H1],[TIME].[Mar],
[Dx].[Car],**[Measures].[NoPat]**

Measure behaves like member of a dimension

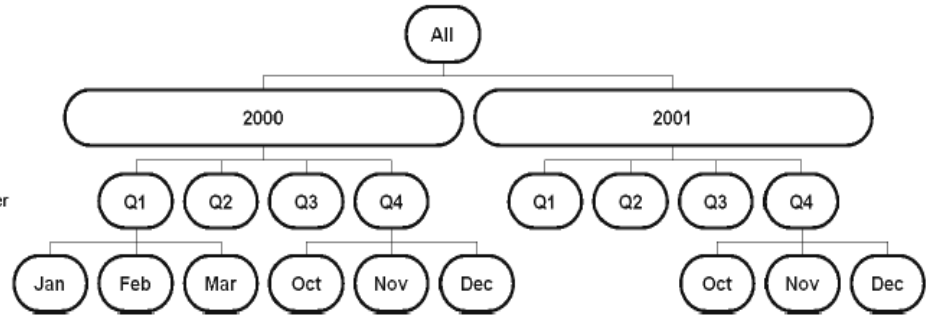


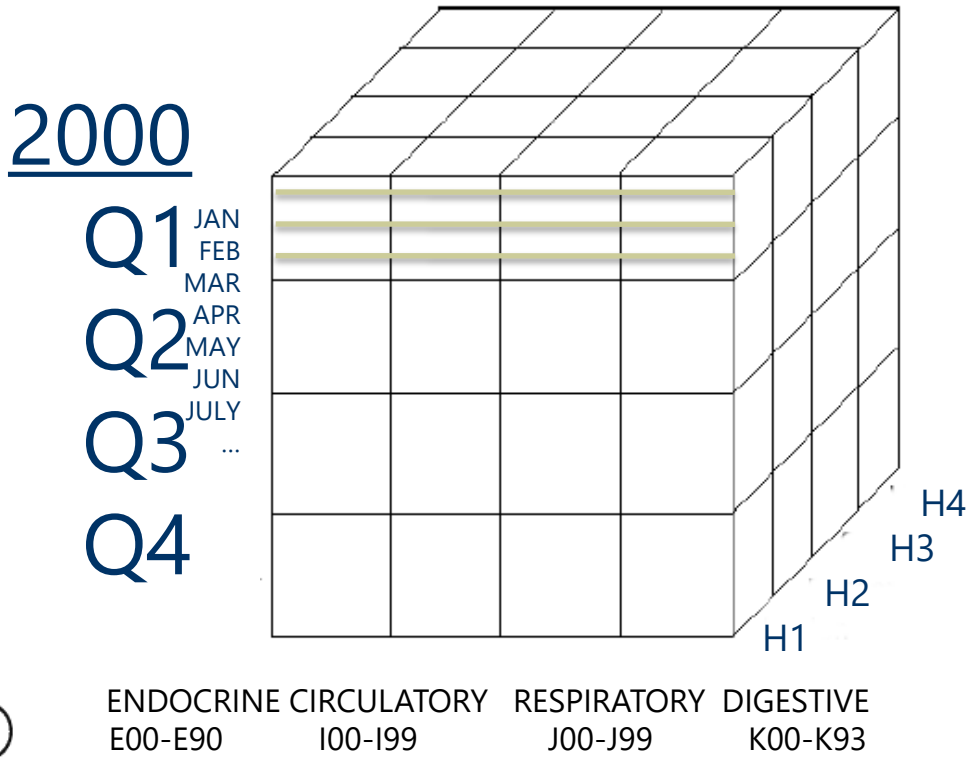
QUESTION 5: Tuples, Sets & Cells

- Measures & Hierarchies

[Hos].[H1],**[TIME].[Q1]**,[Dx].[Cir]

SET?
TUPLE?
CELL?



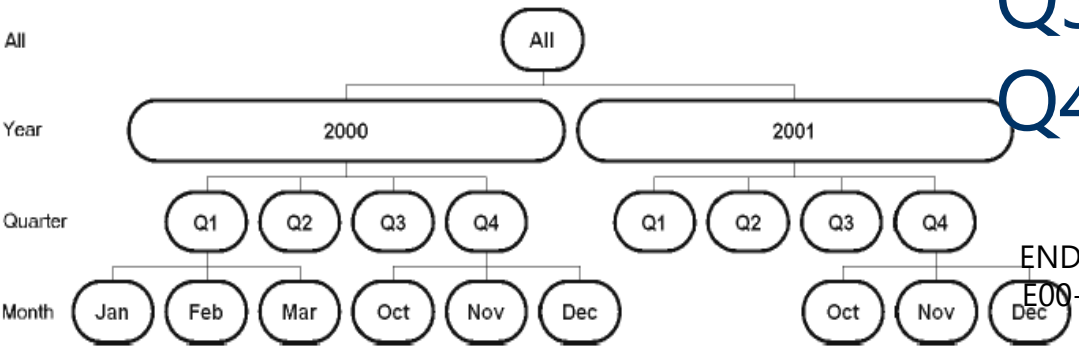


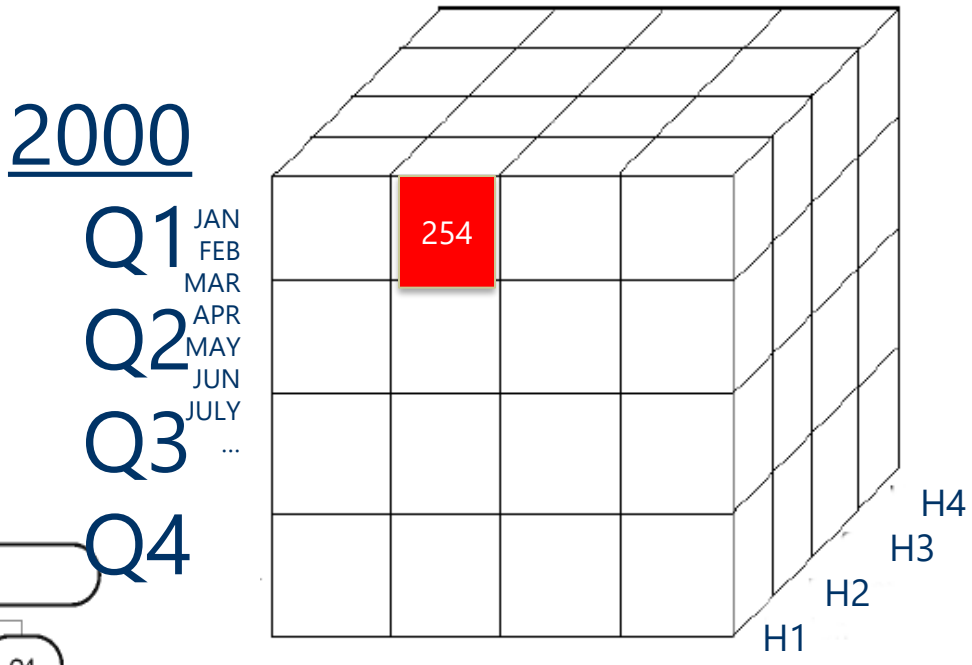
QUESTION 5: Tuples, Sets & Cells

- Measures & Hierarchies

[Hos].[H1],**[TIME].[Q1]**,[Dx].[Cir]

TUPLE!
CELL! (AGGREGATION)





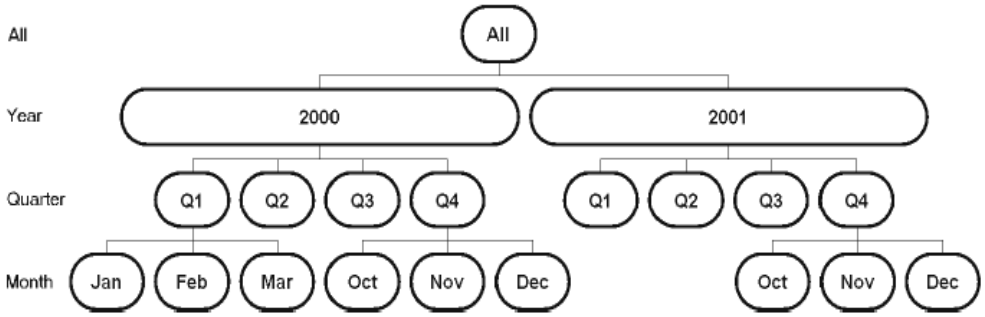
MDX

- Designed for multidimensional data retrieval and analysis.
- Supports hierarchical navigation through dimensions.
- Enables slicing, dicing, drilling down/up, and pivoting data.
- SQL returns a subset of a 2D table. MDX returns a multidimensional subset from a cube.
- Case-insensitive but conventionally capitalizes reserved keywords.
- Square brackets [ ] are used to reference members and dimensions.
- Curly braces { } are used for defining sets.
- A Typical SQL query has SELECT, FROM and WHERE.

MDX Basics

MDX ≠ SQL ? SQL: Country="Spain" MDX: Location.[Spain]

	ALL (TIME)
COST	45,300,000 €



- **SELECT**
  {[TIME].[*ALL*]} **ON *COLUMNS*** *(COL dimension)*
  {[Measure].[Cost]} **ON *ROWS*** *(ROW dimension)*
  FROM [MyCube]
  (shows costs of [HOSP].[H1], default member of HOSP)
  (also for [Dx].[Circulatory])

SELECT-> Axis or layour of the results
FROM->Refers to the Cube we query

MDX Basics

	COL 1	COL 2	COL 3
ROW A			
ROW B			
ROW C			

SELECT
{column headers} ON COLUMNS → SET
{row headers} ON ROWS → SET
FROM [cube] →name

SELECT defines the resulting set with the subset of multidimensional data from the cube.

We define the number of axes and the members from each dimension to include in each axis

MDX Basics

```
SELECT
    {[Measure].[Patient]} ON COLUMNS
    {[Hospital].[Hosp1],
     [Hospital]. [Hosp2],
     [Hospital]. [Hosp3],
     [Hospital]. [Hosp4]}    ON ROWS
    FROM [MyCube]
```

(shows a default member of TIME)

	PATIENT
HOSP 1	23
HOSP 2	65
HOSP 3	88
HOSP 4	65

FROM-> What is the source of the multidimensional data? A cube (restricted to 1 cube).

With LookupCube() we can bypass that restriction.

MDX Basics

```
SELECT
    {[Measure].[Patient]} ON COLUMNS
    {[Hospital].Children}    ON ROWS
    FROM [MyCube]
```

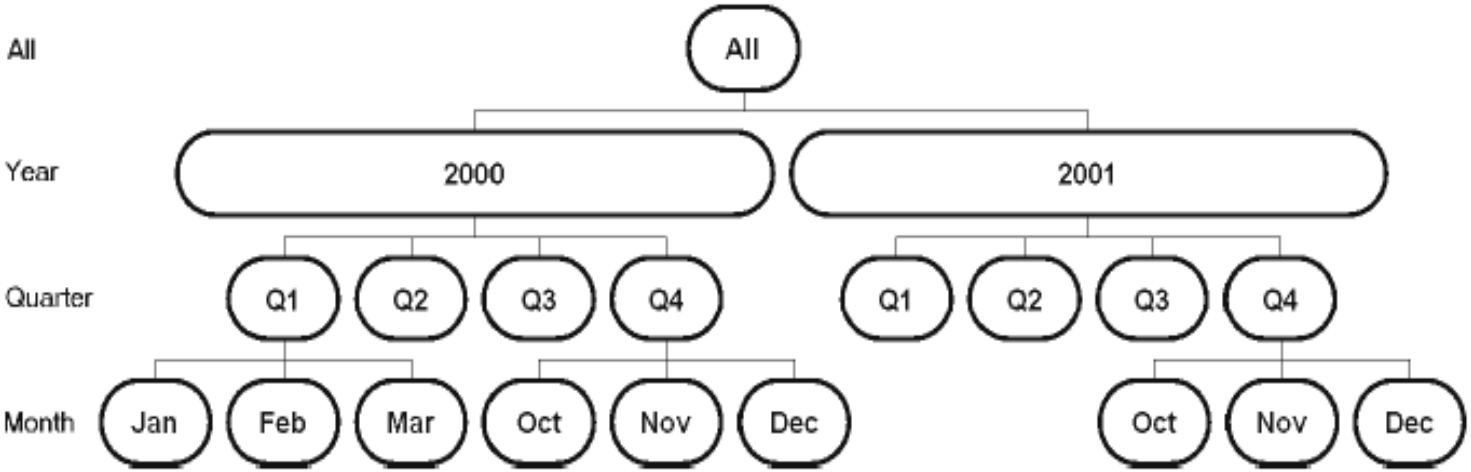
	PATIENT
HOSP 1	23
HOSP 2	65
HOSP 3	88
HOSP 4	65

MDX Basics

- QUESTION:

Cost in H1,H2 during 2000 (by Q), for circulatory diseases.

Y2000	HOSP1	HOSP2
Q1	2M€	0.3M€
Q2	3.2M€	0.7M€
Q3	1.5M€	0.6M€
Q4	0.4M€	0.5M€



<u>Hint:</u>

cost/circulatory are default members

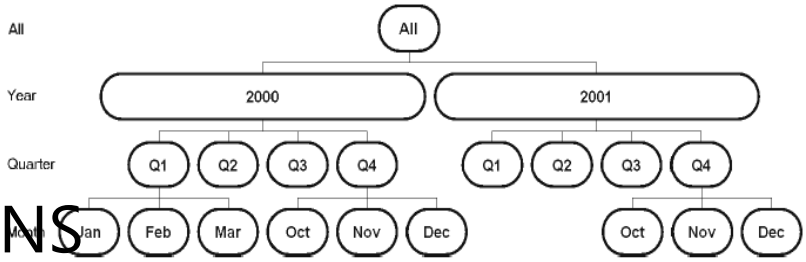
MDX Basics

QUESTION:

Cost in H1,H2 in 2000 (by Q), for circulatory diseases.

```
SELECT
    {[Hospital].[Hosp1],
     [Hospital].[Hosp2]} ON COLUMNS
    {[Time].[All].[2000].Children}    ON ROW
    FROM [MyCube]
```

Y2000	HOSP1	HOSP2
Q1	2M€	0.3M€
Q2	3.2M€	0.7M€
Q3	1.5M€	0.6M€
Q4	0.4M€	0.5M€



MDX Basics

- QUESTION:

Patients no. in H1,H2 during 2000 (by Q), for circulatory diseases.

Y2000	HOSP1	HOSP2
Q1	121 pat	78 pat
Q2	165 pat	61 pat
Q3	115 pat	41 pat
Q4	120 pat	76 pat

Hint:

Number of patients is NOT a default member

MDX Basics

QUESTION:

Patients in H1,H2

during 2000 (by Q), for

circulatory diseases.

```
SELECT
      {[Hospital].[Hosp1],
       [Hospital].[Hosp2]} ON COLUMNS
      {[Time].[All].[2000].Children}    ON ROWS
FROM [MyCube]
WHERE ([Measures].[NoPat])
```

Hint:

Number of patients is NOT a default member

Y2000	HOSP1	HOSP2
Q1	121 pat	78 pat
Q2	165 pat	61 pat
Q3	115 pat	41 pat
Q4	120 pat	76 pat

MDX Basics

WHERE clause

- Not restricted to measures.
- Not restricted to 1 dimension.
- It is a *SLICER/DICER*.

MDX Basics

WHERE clause

Not restricted to measures.

```
SELECT
    {[Hospital].[Hosp1],
     [Hospital].[Hosp2]} ON COLUMNS
    {[Time].[All].[2000].Children}    ON ROWS
    FROM [MyCube]
    WHERE ([Dx].[Respiratory])
```

Y2000	HOSP1	HOSP2
Q1	1M€	0.4M€
Q2	1.2M€	0.1M€
Q3	0.5M€	0.5M€
Q4	0.4M€	0.3M€

WHERE-> optional. Defines the slicer dimension, that we use to filter the multidimensional data.

MDX Basics

WHERE clause

Not restricted to 1 dimension.

```
SELECT
    {[Hospital].[Hosp1],
     [Hospital].[Hosp2]} ON COLUMNS
    {[Time].[All].[2000].Children}    ON ROWS
    FROM [MyCube]
    WHERE ([Dx].[Respiratory],[Measures].[NoPat])
```

Y2000	HOSP1	HOSP2
Q1	61 pat	28 pat
Q2	75 pat	41 pat
Q3	105 pat	11 pat
Q4	112 pat	56 pat

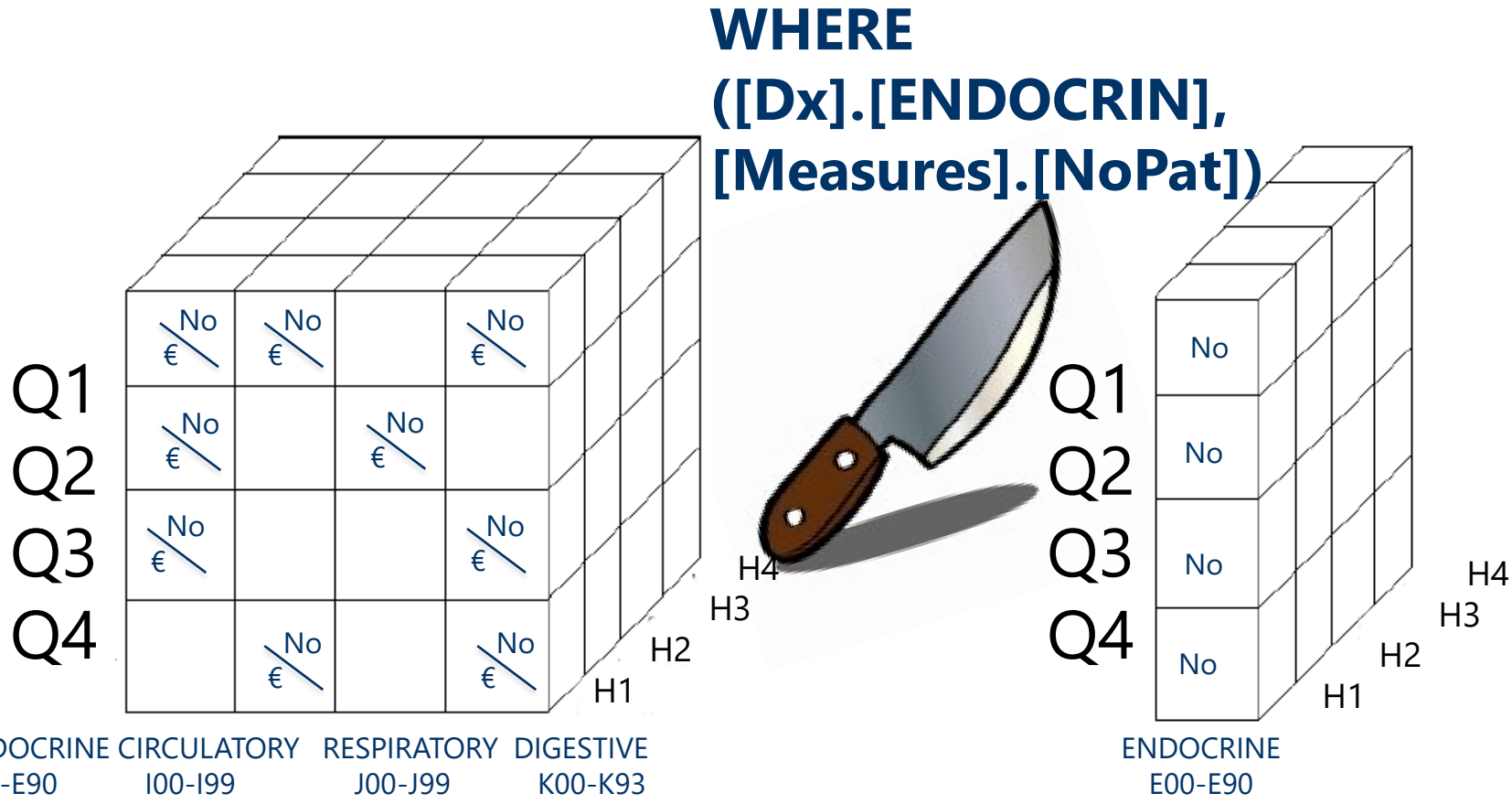
MDX Basics

ORDERED BY Number of PATIENTS?

```
SELECT
    {[Hospital].[Hosp1],
     [Hospital].[Hosp2]} ON COLUMNS
    ORDER({[Time].[All].[2000].Children},
([Dx][Respiratory],[Measures].[NoPat]),BDESC)    ON
ROWS
    FROM [MyCube]
    WHERE ([Dx].[Respiratory],[Measures].[NoPat])
```

MDX Basics

- WHERE clause
  - It is a *SLICER/DICER*.



3. MDX syntax

() [] {} .

MDX syntax

- Brackets []
  - Dimensions: [Time]
  - Members: [2000]
- Dots .
  - Separators: [Time].[2000].[Q3]
- Parentheses ()
  - Tuples: ([DX].[Circulatory],[Hosp].[H1])

MDX syntax

- Braces {}
  - Sets: {[Hosp].[H1], [Hosp].[H3]}
    {[Dx].Children}
    { ([Dx].[Circ],[Hosp].[H1],[Time].[Jan]),
      ([Dx].[Circ],[Hosp].[H1],[Time].[Feb]),
      ([Dx].[Circ],[Hosp].[H1],[Time].[Mar]),
      ([Dx].[Circ],[Hosp].[H1],[Time].[Apr]) }

MDX syntax

```
SELECT
    { SET } ON COLUMNS
    { SET } ON ROWS
    FROM [cube]
    WHERE (TUPLE)
```

MDX syntax

- QUESTION: Correct? Why?

```
SELECT
([Measures].[NoPatients]) ON COLUMNS,
{[Time].[2000].Children} ON ROWS
FROM [MyCube]
```

MDX syntax

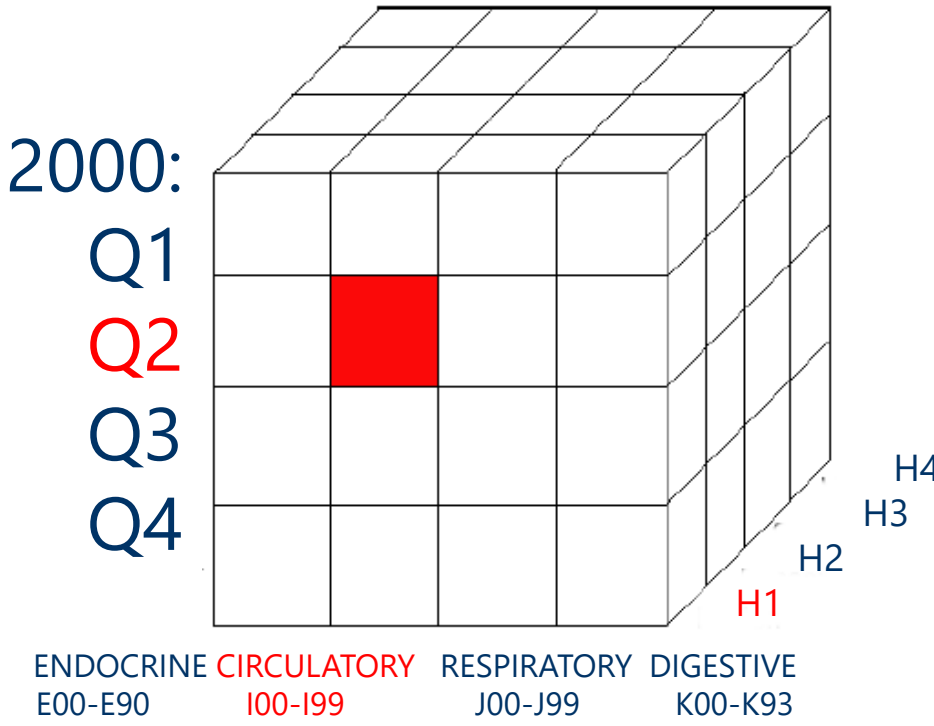
- QUESTION: Correct? Why?

```
SELECT
{[Dx].Children} ON COLUMNS,
{[Time]. [2000].[Q1].[May].Children} ON ROWS
FROM [MyCube]
WHERE {[Measure].[cost],[Hosp].[H2]}
```

MDX syntax

- Name of a CELL.
  - In a cube, each cell has a name.

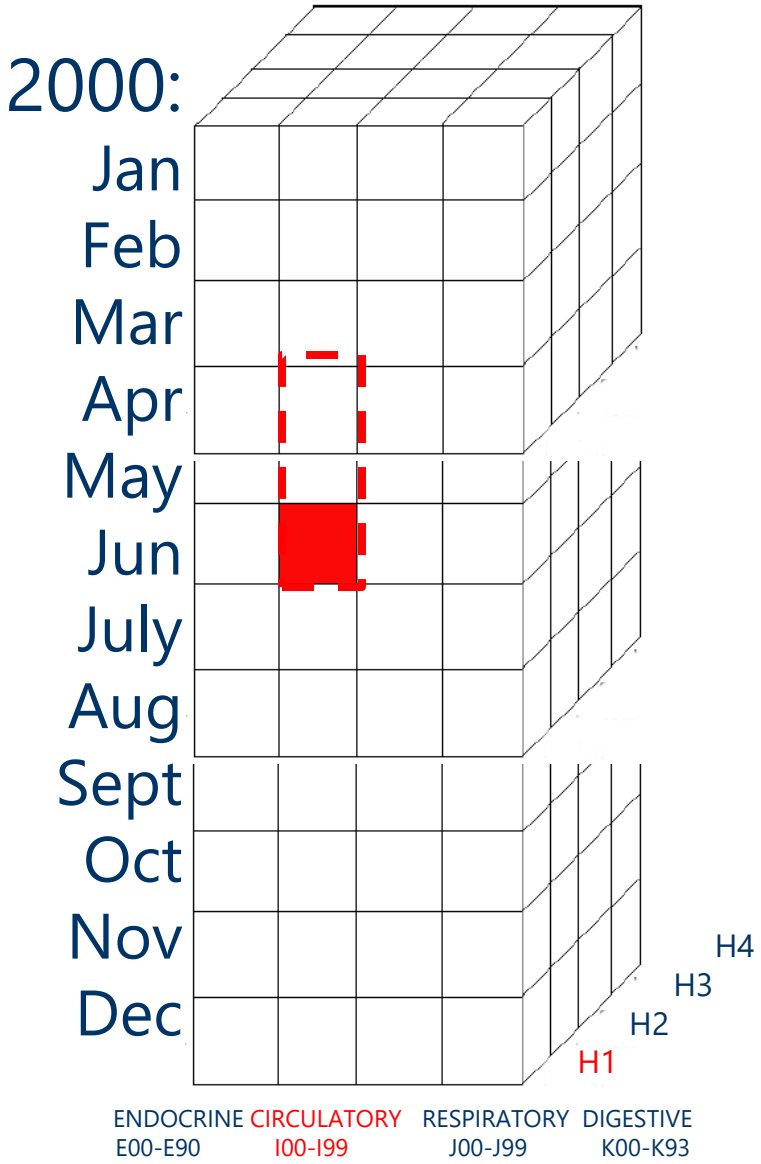
The name of this cell is:
([Time].[2000].[Q2],
[Dx].[Circulatory],
[Hospital].[H1])



MDX syntax

- Name of a CELL.

The name of this cell is:
([Time].[2000].[Q2].[Jun],
[Dx].[Circulatory],
[Hospital].[H1])

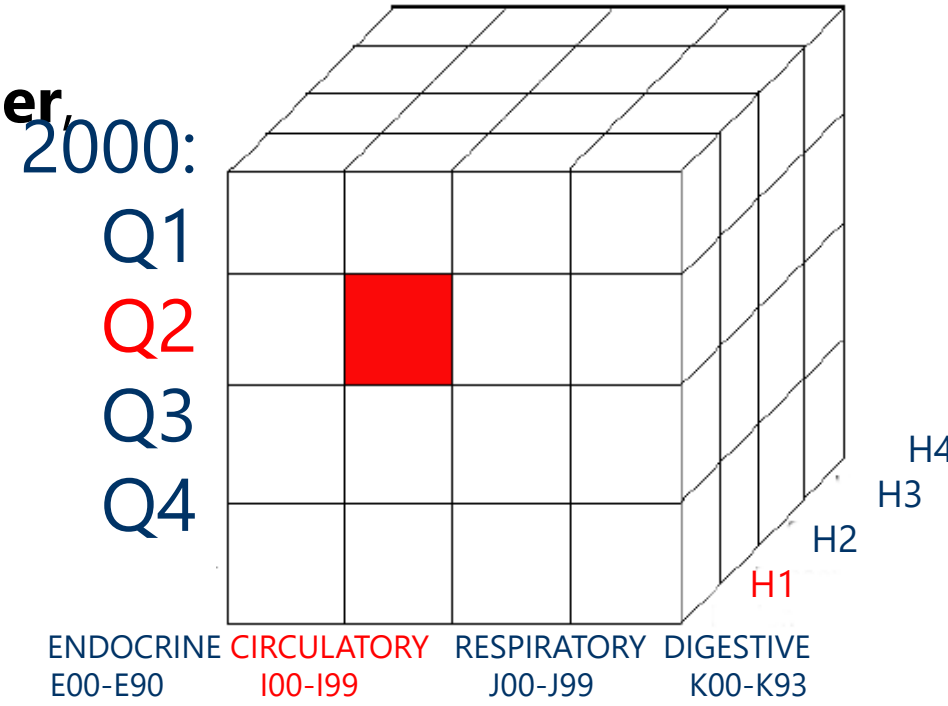


MDX syntax

- Relative Cell Referencing:
  - CurrentMember, PrevMember, NextMember.

The name of this cell is:

([Time].[2000].**[Q3].PrevMember**,
[Dx].[Circulatory],
[Hospital].[H1])



MDX syntax

- Calculated Members: +-*/ %

"Attention improvement on circulatory patients of the 1st quarter of years 1999 and 2000".

Calculus:
([Hosp].[H1],[Dx].[Circ],[Time].[2000].[Q1],[Measures].[NoPatient])
-
([Hosp].[H1],[Dx].[Circ],[Time].[1999].[Q1],[Measures].[NoPatient])

Calculated elements bring flexibility to MDX to generate complex queries. MDX supports multiple logical and arithmetic clauses.

MDX syntax

- Calculated Members: +-*/ %

"Growth cost throughout year 2000 in H1 for Circulatory patients".

```
WITH MEMBER [Measures].[Increment] AS
([Measures].[Inputs] - [Measures].[Outputs]) SELECT
{[Measures].[Inputs], [Measures].[Outputs],
[Measures].[Increment]} ON COLUMNS,
[Time].[Year].Members ON ROWS FROM [MyCube]
```

MDX syntax

- Calculated Members: +-*/ %

"**Growth** cost **throughout** year 2000 in H1 for Circulatory patients".

  - Growth cost: cost(t)- cost(t-1) (increment/derivate)
  - Obviate: H1 for Circulatory patients

MDX Syntax

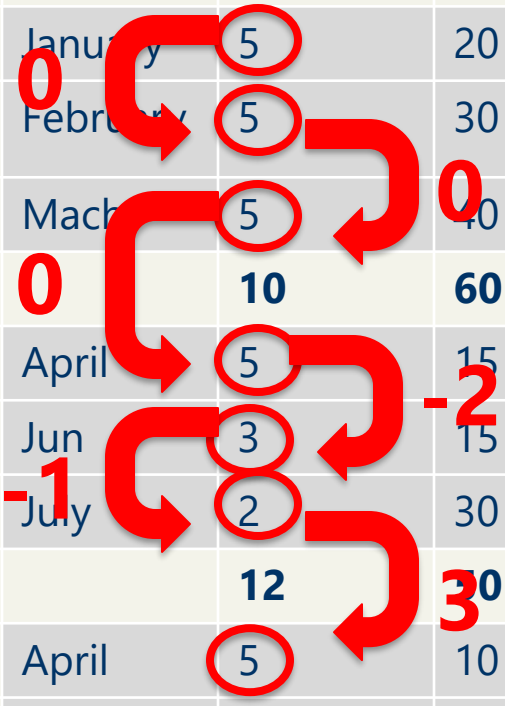
"Growth cost throughout year 2000".

Year	Quarter	Month	Cost M€	No Patient
2000			57	280
	Q1		**15**	**90**
		January	5	20
		February	5	30
		Mach	5	40
	Q2		**10**	**60**
		April	5	15
		Jun	3	15
		July	2	30
	Q3		**12**	**50**
		April	5	10
		Jun	5	10
		July	2	30
	Q4		**20**	**80**
		April	5	15
		Jun	5	15
		July	10	50

MDX Syntax

"Growth cost throughout year 2000".

Year	Quarter	Month	Cost M€	No Patient
2000			57	280
	Q1		**15**	**90**
		January	5	20
		February	5	30
		March	5	40
	Q2		**10**	**60**
		April	5	15
		Jun	3	15
		July	2	30
	Q3		**12**	**80**
		April	5	10
		Jun	5	10
		July	2	30
	Q4		**20**	**80**
		April	5	15
		Jun	5	15
		July	10	50



MDX Syntax

"Growth cost throughout year 2000".

Year	Quarter	Month	Cost M€	No Patient
2000			57	280
	Q1		**15**	**90**
		January	5	20
		February	5	30
		Mach	5	40
	Q2		**10**	**60**
		April	5	15
		Jun	3	15
	Q4		**20**	**80**
		April	5	15
		Jun	5	15
		July	10	50

```
[Time].CurrentMember,[Measures].[Cost]
    -
    [Time].CurrentMember.PrevMember,[Measures].[Cost]
```

MDX syntax

- Calculated Members: +-*/ %

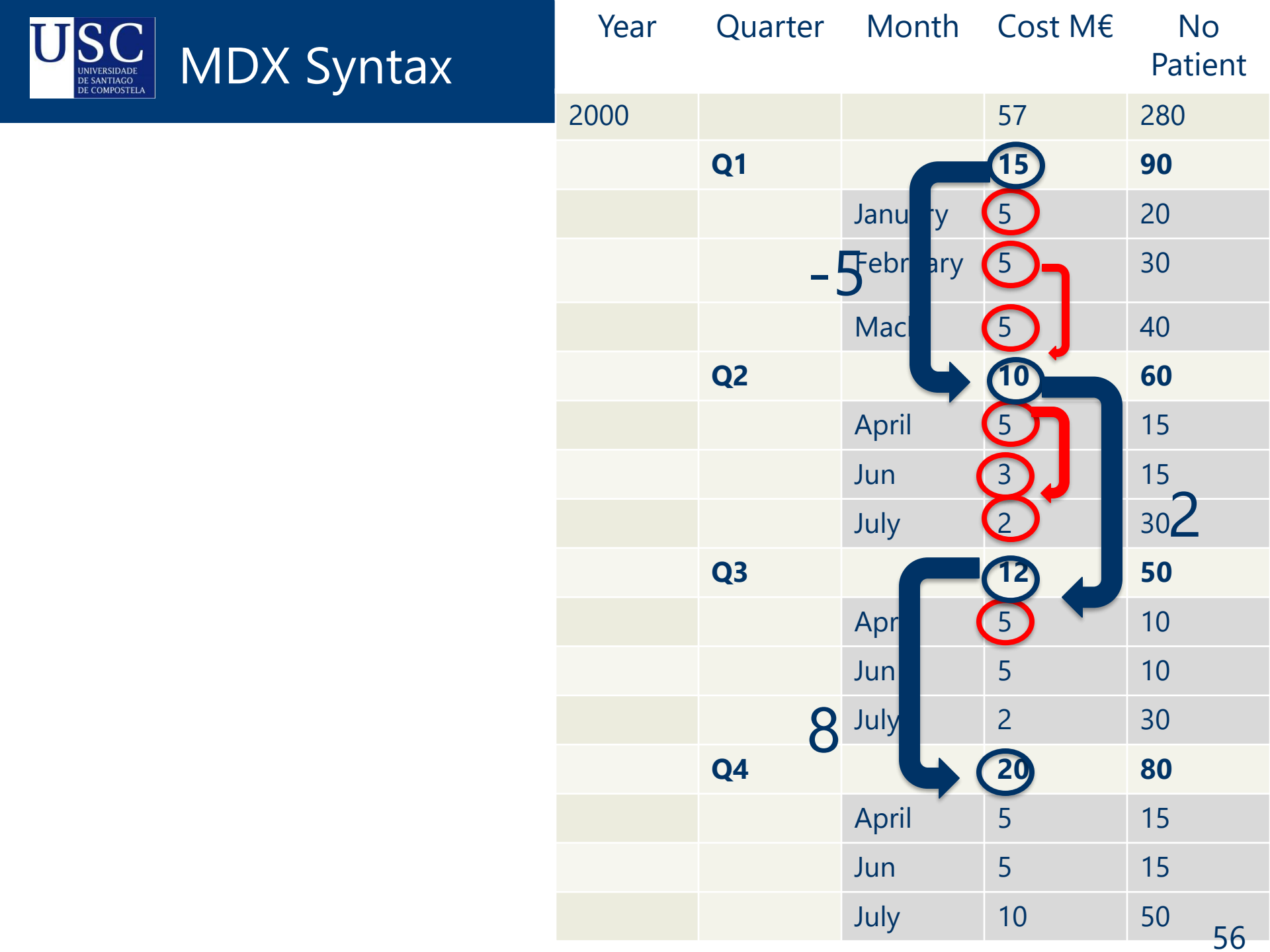
"**Growth** cost **throughout** year 2000 in H1 for Circulatory patients".

- ([Time].**CurrentMember**,[Measures].[Cost]
  - [Time].**CurrentMember**.**PrevMember**,[Measures].[Cost])

YES… WHAT IF WE FOCUS ON **THE QUARTER GROWTH** ?????

MDX Syntax

Year	Quarter	Month	Cost M€	No Patient
2000			57	280
	Q1		**15**	**90**
		January	5	20
		February	5	30
		March	5	40
	Q2		**10**	**60**
		April	5	15
		Jun	3	15
		July	2	30
	Q3		**12**	**50**
		April	5	10
		Jun	5	10
		July	2	30
	Q4		**20**	**80**
		April	5	15
		Jun	5	15
		July	10	50



MDX syntax

- Calculated Members: +-*/ % AVG SUM …

"**Growth** cost **throughout** year 2000 in H1 for Circulatory patients".

- ([Time].**CurrentMember**,[Measures].[Cost]
  - [Time].**CurrentMember**.**PrevMember**,[Measures].[Cost])

YES… WHAT IF WE FOCUS ON **THE SEMESTER GROWTH** !!!

… SAME EXPRESSION. THAT'S THE COOL THING. DEPENDS ON THE TIME DIMENSION, WHICH DEFINES THE *CURRENT MEMBER* PROPERTY

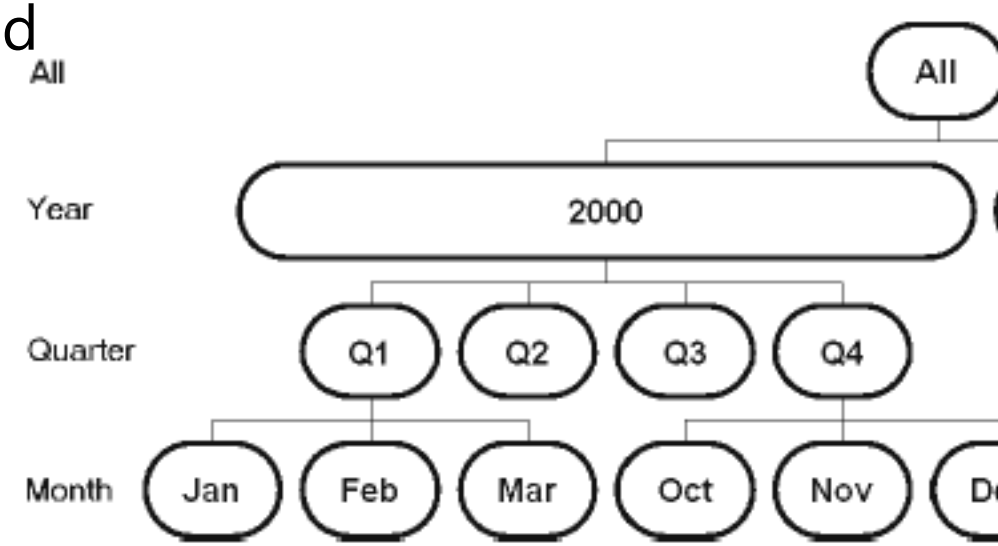
MDX syntax

- OTHER FUNCTIONS:
  - Sum (X)→Number : Sums all members of X
  - X.Lag(N) : N positions back from X.
  - X.Lead(M) : M position forward from X.
  - YTD(X)→Set : YearToDate: Members of the Year until member X.
    e.g. YTD(March) → {Jan, Feb, March}

  ClosingPeriod, OpeningPeriod, ParallelPeriod,

MDX syntax

- Hierarchy navigation:
  - *Member*.Children
  - *Member*.Parent
  - *Member*.FirstChild / LastChild
  - Descendants(X,n)
  - Ancestors(X,n)
  - Siblings / Cousins
  - ~~Aunt~~ / ~~Uncle~~



Security

As with SQL, MDX can be vulnerable to MDX Injection
Do we want to prevent unauthorized access to the cube?

Do we want to prevent unauthorized modifications?
If we can inject in the first dimension of SELECT we can write a full custom query… very dangerous.

If the injection is in WHERE, blind injection can be used.

We have to sanitize the inputs, implement least privilege, testing for injection vulnerabilities, audit, etc.

MDX

- Bibliography and Resources:

- Mark Whitehorn et al. Fast Track to MDX (2nd Ed). Springer. 2004.
- Microsoft, "Key Concepts in MDX (Analysis Services)", https://docs.microsoft.com/en-us/analysis-services/multidimensional-models/mdx/key-concepts-in-mdx-analysis-services?view=asallproducts-allversions
- InterSystems, "Introduction to MDX Queries", https://docs.intersystems.com/irislatest/csp/docbook/DocBook.UI.Page.cls?KEY=D2GMDX_CH_MDX_INTRO