sqlbi

sqlbi.com

# BISM Introduction

presented by
Marco Russo
marco@sqlbi.com



## Who am I

- BI Experts and Consultants
- Problem Solving
- Complex Project Assistance
- DataWarehouse Assesments and Development
- Courses, Trainings and Workshops
- Microsoft Business Intelligence Partners
- Book Writers





## Latest conferences

- PASS Europe 2009 – Neuss – Germany
- PASS 2009 – Seattle – USA
- SQL Conference 2010 – Milan – Italy
- Teched 2010 – New Orleans – USA
- 24 Hours of PASS 2010 – Online
- PASS 2010 – Seattle – USA



## Agenda

- What is BISM?
- What is going to change?
- The road to BISM: PowerPivot
- BISM vs UDM
  - Patterns design comparison
  - Dimensional vs Relational Modeling
- What to do today?

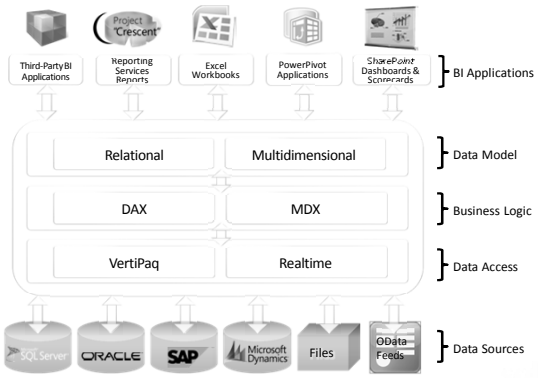


# One bad news

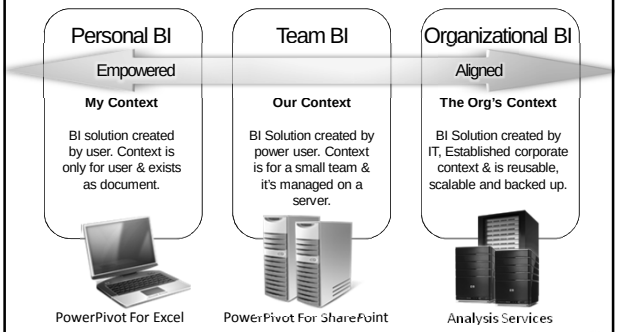
## BISM – BI Semantic Model

- One semantic model for Business Intelligence
  - For reporting, analytics, scorecards, dashboards
  - Unification of OLAP (UDM) and report (SMDL) models
- Can be queried through:
  - MDX
  - DAX
  - SQL (?)
- Different engine than SSAS UDM
  - Vertipaq – in-memory columnar database
  - Realtime – query passthrough to SQL Server

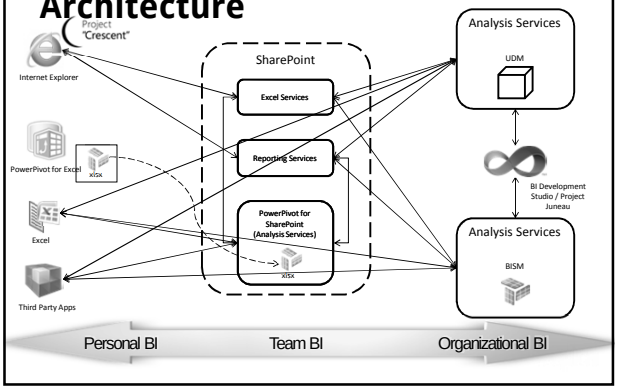
## BI Semantic Model




## Seamless Transition Across BI Spectrum




## PowerPivot + Analysis Services Architecture




## Instance level choice

- A SSAS instance can be:
  - UDM (the classical one)
  - PowerPivot for SharePoint (since SQL2008 R2)
  - BISM (only Denali)
- You can install Denali with different roles in side-by-side instances on the same server

## Road to BISM: PowerPivot

- PowerPivot for Excel
  - Free AddIn for Microsoft Excel 2010
  - Available since June 2010
  - Different releases for 32/64 bits
  - Contains the Vertipaq Engine
    - SSAS running in process with Excel
- PowerPivot for SharePoint
  - Vertipaq running server side
  - Integrated with SharePoint

## Road to BISM: Denali

- Denali
  - New release of SQL Server
  - SSAS / Vertipaq running server side
  - Works standalone (PowerPivot for SP cannot)
  - Vertipaq / SSAS selected at instance level
  - New release of PowerPivot
    - Parent-child hierarchies, perspectives, KPIs
    - New DAX functions (like Ranking)
- PowerPivot 1.0 today:
  - Is the road to learn new Denali features

## Vertipaq: change your mind!

- Vertipaq in Denali
  - Not a simple evolutionary step
  - It is a complete new environment
  - New calculation paradigms
  - New data modeling techniques
  - New formula authoring experience
- Want to learn BISM?
  - Forget MDX
  - Forget SQL
  - Then, Learn DAX. ☺

## The role of DAX in BISM

- DAX is
  - A formula language (like in PowerPivot 1.0)
  - A query language (new to BISM)
    - Syntax based on CALCULATETABLE (existing in 1.0)
    - But now you can define result shape (ADDCOLUMNS)
    - And control the order of results (EVALUATE / ORDER BY)
  - The fastest way to query BISM
    - MDX available, but might add some overhead

## What is changing?

| DENALI | SQL 2008 R2 |
|---|---|
| BISM | UDM |
| DAX | MDX |
| RELATIONAL | MULTIDIMENSIONAL |
| IN MEMORY | ON DISK |
| FAST VERTIPAQ ENGINE | AGGREGATIONS |
| LEAF LEVEL COMPUTATION | CALCULATED MEMBERS |
| FILTER / ROW CONTEXT | MDX FUNCTIONS, TUPLES ETC |
| SELF SERVICE + CORPORATE | CORPORATE ONLY |
| EASY DATA INTEGRATION | BI PROFESSIONALS NEEDED |

## Relational vs Multidimensional

- UDM
  - Facts, Dimensions, Bridges, Snowflakes
  - Relationships: Regular, Referenced, M2M
  - Hierarchies, Attribute Relationships
- BISM
  - Tables
  - Relationships (1:N, One Column only)

## Complexity in BISM

- BISM: simple data model
- Complexity is moved into:
  - Calculated columns
  - Measures
  - DAX, DAX, DAX, DAX, DAX and… **DAX**!
- Every column is a dimensional attribute
- Every column is a measure

A SMALL EXAMPLE OF HOW TO AGGREGATE DATA IN DAX

# Semi Additive Measures

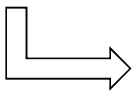
sqlbi.com

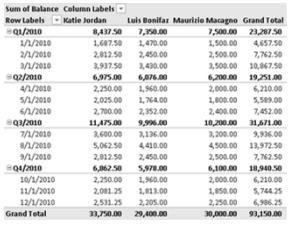
## Semi Additive Measures

- Additive Measure
  - SUM over all dimensions
- Non Additive Measure
  - Different function over all dimensions
  - Example: average of the sale price
- Semi Additive Measure
  - SUM over some dimensions
  - Different function over other dimensions
  - Time is the standard exception for aggregations
  - Examples
    - Warehouse stocking
    - Current account balance

## Current Account Balance

| Name | Occupation | Country | Date | Quarter | Balance |
|---|---|---|---|---|---|
| Katie Jordan | Farmer | USA | 1/1/2010 | Q1/2010 | 1,687.50 |
| Luis Bonifaz | IT Consultant | Argentina | 1/1/2010 | Q1/2010 | 1,470.00 |
| Maurizio Macagno | IT Consultant | Italy | 1/1/2010 | Q1/2010 | 1,500.00 |
| Katie Jordan | Farmer | USA | 2/1/2010 | Q1/2010 | 2,812.50 |
| Luis Bonifaz | IT Consultant | Argentina | 2/1/2010 | Q1/2010 | 2,450.00 |
| Maurizio Macagno | IT Consultant | Italy | 2/1/2010 | Q1/2010 | 2,500.00 |
| Katie Jordan | Farmer | USA | 3/1/2010 | Q1/2010 | 3,937.50 |
| Luis Bonifaz | IT Consultant | Argentina | 3/1/2010 | Q1/2010 | 3,430.00 |





| Sum of Balance | Column Labels | | | |
|---|---|---|---|---|
| Row Labels | Katie Jordan | Luis Bonifaz | Maurizio Macagno | Grand Total |
| Q1/2010 | 8,437.50 | 7,350.00 | 7,500.00 | 23,287.50 |
| 1/1/2010 | 1,687.50 | 1,470.00 | 1,500.00 | 4,657.50 |
| 2/1/2010 | 2,812.50 | 2,450.00 | 2,500.00 | 7,762.50 |
| 3/1/2010 | 3,937.50 | 3,430.00 | 3,500.00 | 10,867.50 |
| Q2/2010 | 6,975.00 | 6,076.00 | 6,200.00 | 19,251.00 |
| 4/1/2010 | 2,250.00 | 1,960.00 | 2,000.00 | 6,210.00 |
| 5/1/2010 | 2,025.00 | 1,764.00 | 1,800.00 | 5,589.00 |
| 6/1/2010 | 2,700.00 | 2,352.00 | 2,400.00 | 7,452.00 |
| Q3/2010 | 11,475.00 | 9,996.00 | 10,200.00 | 31,671.00 |
| 7/1/2010 | 3,600.00 | 3,136.00 | 3,200.00 | 9,936.00 |
| 8/1/2010 | 5,062.50 | 4,410.00 | 4,500.00 | 13,972.50 |
| 9/1/2010 | 2,812.50 | 2,450.00 | 2,500.00 | 7,762.50 |
| Q4/2010 | 6,862.50 | 5,978.00 | 6,100.00 | 18,940.50 |
| 10/1/2010 | 2,250.00 | 1,960.00 | 2,000.00 | 6,210.00 |
| 11/1/2010 | 2,081.25 | 1,813.00 | 1,850.00 | 5,744.25 |
| 12/1/2010 | 2,531.25 | 2,205.00 | 2,250.00 | 6,986.25 |
| Grand Total | 33,750.00 | 29,400.00 | 30,000.00 | 93,150.00 |

- Month level correct
- Quarter level wrong
- Year level wrong

## Semi Additive Measures in UDM

- Aggregation function:
  - LastChild
  - LastNonEmpty
- Performances not very impressive
- Aggregations always needed
- The aggregation is handled by the data model
- In DAX, there is no default aggregation function. Thus, DAX is needed.

## Semi Additive Measures

- Create a calendar table
- Add a new measure:
  - CALCULATE: to set the filter
  - LASTDATE: to find the last child

```
LastBalance = CALCULATE (
    SUM( Balances[Balance] ),
    LASTDATE(Date[Date])
)
```

## In Memory vs On Disk

- UDM
  - Aggregations
  - Usage Based Optimization
  - Aimed to optimize Storage Engine
- BISM
  - No Storage Engine: everything in memory
  - No aggregations, fast in memory computation
  - Non standard aggregations = Measures

## Calculated Columns / Measures

- Measures
  - Similar to MDX Calculated Members
  - Work on aggregations
  - Might go leaf-level, if needed
- Calculated columns
  - New kind of calculation
  - Computed once in the data model
  - Can access the whole data model
  - Different from calc columns in the UDM DSV

## Calculated Columns / Measures

- GrossMargin = SalesAmount – ProductCost
  - Calculated column
- GrossMargin%=GrossMargin / SalesAmount
  - Cannot be computed row by row
- Measures needed

$$\sum_{k=0}^{n}\left(\frac{GrossMargin}{SalesAmount}\right) \neq \frac{\sum(GrossMargin)}{\sum(SalesAmount)}$$

## In Memory vs On Disk

- Vertipaq and DAX
  - Fast in memory computations
  - Leaf level calculations
  - Calculated columns
- The data model can be different
- Less ETL code
- More DAX code
  - No locking, No I/O, very fast
  - Data is not persisted on disk for other clients

## Self Service vs Corporate BI

- Self Service BI
  - Creative Solutions in the hand of Power Users
  - No need to ask for help to IT
  - Less work for IT
- Example
  - Banding of prices

BANDING IN THE SELF SERVICE BI WORLD

## Banding

sqlbi.com

## Analysis of product sell price

- Price changes over time
  - Discounts
  - Price variations
- Continuous dimension
- High fragmentation
- BANDING
  - From 0 to 100 USD
  - From 101 to 500

| Row Labels | Reseller Order Quantity | Reseller Sales Amount |
|---|---|---|
| 2.29 | 674 | $925,21 |
| 4.99 | 2.571 | $7.476,60 |
| 7.95 | 2.411 | $11.188,37 |
| 8.6442 | 3.289 | $16.779,84 |
| 8.99 | 6.284 | $32.826,92 |
| 9.5 | 1.197 | $6.573,39 |
| 19.99 | 1.130 | $13.514,69 |
| 20.24 | 774 | $9.377,71 |
| 23.5481 | 1.877 | $26.419,61 |
| 24.49 | 3.621 | $52.507,99 |
| 25 | 1.086 | $16.225,22 |
| 27.12 | 448 | $7.280,43 |
| 33.6442 | 6.692 | $131.508,29 |
| 34.2 | 55 | $1.949,40 |
| 34.99 | 6.409 | $127.204,64 |
| 37.99 | 6.055 | $128.847,58 |
| 39.14 | 618 | $14.469,82 |
| 40.49 | 1.317 | $31.995,20 |
| 40.4909 | 531 | $12.900,38 |
| 44.54 | 547 | $14.530,43 |
| 46.09 | 56 | $1.548,62 |
| 46.0673 | 6.587 | $187.952,11 |

## The quick and dirty solution

```
= IF (
    FactResellerSales[DiscountedPrice] <= 5,
    "01 LOW",
    IF (
        FactResellerSales[DiscountedPrice] <=30,
        "02 MEDIUM",
        IF (
            FactResellerSales[DiscountedPrice] <=100,
            "03 MEDIUM",
            IF (
                FactResellerSales[DiscountedPrice] <= 500,
                "04 HIGH",
                "05 VERY HIGH"))))
```

**Even if this works… a better data model would be welcome!**

## Banding: a Data Driven Model

| BandName | FromPrice | ToPrice |
|---|---|---|
| VERY LOW | 0 | 5 |
| LOW | 5 | 30 |
| MEDIUM | 30 | 100 |
| HIGH | 100 | 500 |
| VERY HIGH | 500 | 9999 |

```
SELECT
    P.BandName,
    SUM (S.ExtendedAmount)
FROM dbo.FactResellerSales S
    JOIN PriceBands P
    ON S.UnitPrice BETWEEN P.FromPrice AND
P.ToPrice
GROUP BY
    P.BandName
```

## Data Model with PowerPivot

```
SELECT
    P.BandName,
    SUM (S.ExtendedAmount)
FROM dbo.FactResellerSales S
    JOIN PriceBands P
    ON S.UnitPrice = P.Price
GROUP BY
    P.BandName
```

| BandName | Price |
|---|---|
| VERY LOW | 1 |
| VERY LOW | 2 |
| VERY LOW | 3 |
| VERY LOW | 4 |
| VERY LOW | 5 |
| LOW | 6 |
| LOW | 7 |
| … | … |

## Band Expansion

| BandName | FromPrice | ToPrice |
|---|---|---|
| VERY LOW | 0 | 5 |
| LOW | 5 | 30 |
| MEDIUM | 30 | 100 |
| HIGH | 100 | 500 |
| VERY HIGH | 500 | 9999 |

| BandName | Price |
|---|---|
| VERY LOW | 1 |
| VERY LOW | 2 |
| VERY LOW | 3 |
| VERY LOW | 4 |
| VERY LOW | 5 |
| LOW | 6 |
| LOW | 7 |
| … | … |

- Improve the Data Model
- Cannot do this in PowerPivot
- Write VBA code to transform data
- Needs some VBA knowledge

## Banding With DAX

- Link only the configuration table
- No need to perform expansion
- The complexity moves to the expression

- Again, DAX programming needed

## Banding With CALCULATE

- Works with the filter context
- Leverages
  - CALCULATE
  - VALUES
  - FILTER

```
= CALCULATE(
   VALUES (PriceBands[PriceBand]),
   FILTER (
      PriceBands,
      FactSales[DiscountedPrice] >= PriceBands[MinPrice]
    && FactSales[DiscountedPrice] < PriceBands[MaxPrice]
   )
)
```

SOME INSIGHTS INTO BISM TIME INTELLIGENCE FUNCTIONS

# Time Intelligence

sqlbi.com

## Aggregations Over Time

- Many interesting aggregations
  - YTD: Year To Date
  - QTD: Quarter To Date
  - MTD: Month To Date
- They all need a Calendar Table
- And some understanding of CALCULATE ☺
- Tool dimension is not affordable
  - No DateTool dimension!

## Year To Date

```
CALCULATE(
    SUM ( SalesOrderDetail[LineTotal] ),
    DATESYTD ( OrderDate[Date] )
)
```

- DATESYTD: Returns a set of dates
  - From the start of the year
  - Up to the parameter date
- CALCULATE
  - Creates a filter contexts
  - Perform the SUM operation

## PARALLELPERIOD

```
CALCULATE (
    SUM( SalesOrderDetail[LineTotal] ),
    PARALLELPERIOD
        (OrderDate[Date], -1, YEAR)
)
```

- Returns a set of dates (a table)
  - Shifted in time
  - The whole period is returned, regardless dates in the first parameter

## Moving Annual Total

```
CALCULATE(
    SUM (SalesOrderDetail[LineTotal]),
    DATESBETWEEN (
        OrderDate[Date],
        NEXTDAY(
            SAMEPERIODLASTYEAR(
                LASTDATE( OrderDate[Date] )
            )
        ),
        LASTDATE( OrderDate[Date] )
    )
)
```

## Counting Working Days

- How many working days in a date range?
  - Easily solved with Calendar table
  - Define a new column «WorkingDays»
  - Aggregate with SUM
- Handles any date range
- Works on periods with «holes»
- Difference between BISM and UDM
  - UDM solution requires ETL
  - BISM solution can be self-contained

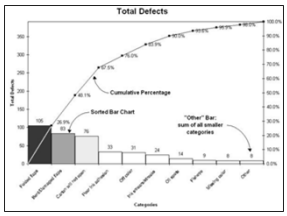
CALCULATED COLUMNS AT WORK WITH ABC PARETO ANALYSIS

# ABC PARETO ANALYSIS

sqlbi.com

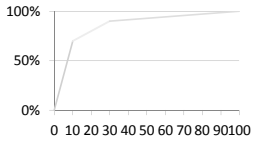
## ABC and Pareto Analysis

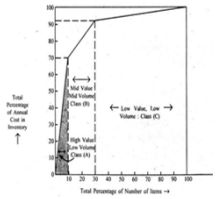
- Pareto principle
  - 80% of effects come from 20% of the causes




## ABC and Pareto Analysis

- ABC Analysis
  - Class A contains items for 70% of total value
  - Class B contains items for 20% of total value
  - Class C contains items for 10% of total value




## ABC Analysis with Excel

The formula is straightforward:

=IF(H6<=0,7;"A";IF(H6<=0,9;"B";"C"))

| Product | TotalSold | Incremental | GrandTotal | Incr.% | Class |
|---|---|---|---|---|---|
| Mountain-200 Black, 38 | 4.400.592,80 | 4.400.592,80 | 109.846.381,40 | 4,01% | A |
| Mountain-200 Black, 42 | 4.009.494,76 | 8.410.087,56 | 109.846.381,40 | 7,66% | A |
| Mountain-200 Silver, 38 | 3.693.678,03 | 12.103.765,59 | 109.846.381,40 | 11,02% | A |
| Mountain-200 Silver, 42 | 3.438.478,86 | 15.542.244,45 | 109.846.381,40 | 14,15% | A |
| Mountain-200 Silver, 46 | 3.434.256,94 | 18.976.501,39 | 109.846.381,40 | 17,28% | A |
| Mountain-200 Black, 46 | 3.309.673,22 | 22.286.174,61 | 109.846.381,40 | 20,29% | A |
| Road-250 Black, 44 | 2.516.857,31 | 24.803.031,92 | 109.846.381,40 | 22,58% | A |
| Road-250 Black, 48 | 2.347.655,95 | 27.150.687,88 | 109.846.381,40 | 24,72% | A |
| Road-250 Black, 52 | 2.012.447,78 | 29.163.135,65 | 109.846.381,40 | 26,55% | A |
| Road-150 Red, 56 | 1.847.818,63 | 31.010.954,28 | 109.846.381,40 | 28,23% | A |
| ... | ... | ... | ... | ... | ... |
| HL Mountain Frame - Black, 42 | 901.590,2... | 76.101.720,96 | 109.846.381,40 | 69,28% | A |
| Road-650 Red, 62 | 879.827,94 | 76.981.548,90 | 109.846.381,40 | 70,08% | B |
| Road-650 Red, 48 | 878.666,66 | 77.860.215,56 | 109.846.381,40 | 70,88% | B |
| Road-650 Black, 58 | 869.632,78 | 78.729.848,34 | 109.846.381,40 | 71,67% | B |
| Touring-2000 Blue, 54 | 772.302,01 | 79.502.150,36 | 109.846.381,40 | 72,38% | B |
| Touring-1000 Blue, 54 | 743.353,03 | 80.245.503,38 | 109.846.381,40 | 73,05% | B |
| Road-550-W Yellow, 44 | 717.825,91 | 80.963.329,29 | 109.846.381,40 | 73,71% | B |
| Road-350-W Yellow, 44 | 694.003,92 | 81.657.333,21 | 109.846.381,40 | 74,34% | B |
| Touring-1000 Yellow, 54 | 667.158,15 | 82.324.491,36 | 109.846.381,40 | 74,95% | B |
| Touring-2000 Blue, 60 | 635.723,72 | 82.960.215,08 | 109.846.381,40 | 75,52% | B |

## ABC Analysis

- Calculate the class

```
SortedWeightProduct =
    Product[CumulatedProduct]
      / SUM( Product[SalesAmountProduct] )

[ABC Class Product] =
    IF( Product[SortedWeightProduct] <= 0.7, "A",
      IF( Product[SortedWeightProduct] <= 0.9, "B", "C" ) )
```

| | Name | ProductModelID | SalesAmountProduct | CumulatedProduct | SortedWeightProduct | ABC Class Product |
|---|---|---|---|---|---|---|
| 748 | HL Mountain Frame -... | 5 | 930,780.68 | 75,200,130.73 | 68.46 % | A |
| 743 | HL Mountain Frame -... | 5 | 901,590.23 | 76,101,720.96 | 69.28 % | A |
| 761 | Road-650 Red, 62 | 30 | 879,827.94 | 76,981,548.90 | 70.08 % | B |
| 763 | Road-650 Red, 48 | 30 | 878,666.66 | 77,860,215.56 | 70.88 % | B |
| 765 | Road-650 Black, 58 | 30 | 869,632.78 | 78,729,848.34 | 71.67 % | B |

Contact | Product | ProductModel | SalesOrderDetail | SalesOrderHeader

A COMPLEX EXAMPLE INVOLVING MANY TO MANY RELATIONSHIPS

# M2M and Audience

sqlbi.com

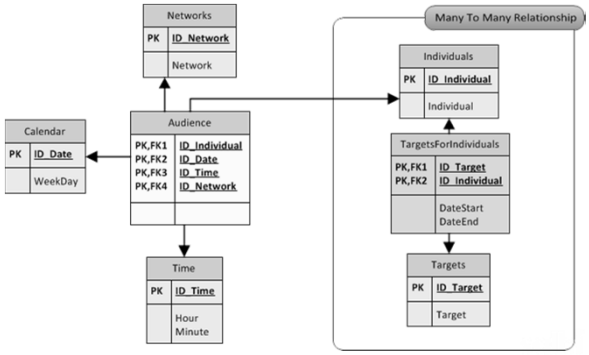
## Audience

- Taken from a real world analysis
- Audience data of TV broadcasts
  - Simple data model
  - Pretty big database
  - Analysis at a very detailed level
- Current solution
  - Ad-hoc software
- Proof of Concept
  - Analysis with PowerPivot and Vertipaq

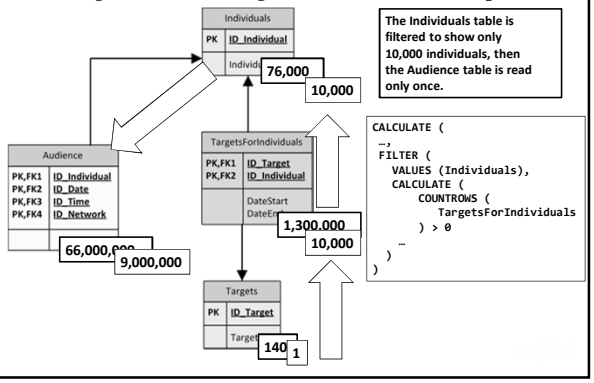
## Audience: background

- Info taken from the remote control
- Once a minute
  - Individual watching
  - Channel watched
- Once a day
  - All the individuals watching
  - Detect the influence of each individual
- Individual
  - Categories
  - Targets

## Audience: The Data Model




## Many-to-many relationship




IS DIMENSIONAL MODELING STILL NECESSARY FOR POWERPIVOT / BISM?

## Dimensional vs Relational Modeling

sqlbi.com

## Dimensional vs Relational: UDM

- UDM
  - Strong separation between facts and dimensions
  - Advanced relationship types
  - Great aggregation level computations
  - Poor leaf level computation performances
- Designed, optimized for dimensional models
- No brain option: go dimensional

## Dimensional vs Relational: BISM

- BISM
  - No concept of fact and dimension: only tables
  - Very simple concept of relationship
  - No aggregations, in memory computation
  - Great leaf level computation performances
- Optimized for simple models
- Two options
  - Dimensional modeling
  - Relational modeling

## Dimensional vs Relational: BISM

- Relational Modeling
  - Does not mean third normal form!
  - Users like facts and dimensions
- Denormalization
  - Needed for the user experience
  - And for performances
- Each relationship has a cost, remember it!

## BISM: Advanced Relationships

- Advanced Relationships in BISM
  - Are not supported
  - But can be implemented through DAX
  - Between (Banding)
  - Many to many (Audience)
  - Parent / Child (Naturalization - some support in BISM)
  - Reference Dimensions (RELATED)
  - LastChild (CALCULATE)
  - Fact Relationships (not needed)
- Thus, nothing is really missing

## BISM: New Power

- Calculated columns
  - Materialize complex calculations
  - ABC Pareto is a good example (see book/blogs)
  - Have access to the whole data model
    - Not like calculated columns in UDM DSV!
- Linked tables
  - Two steps computation
- Easy integration with many sources
- Power to the end user

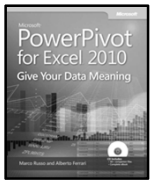
## BISM: DAX

- Brand new language
- Mixes programming and querying
- Much cleaner and easier than MDX
- Designed for efficiency on modern CPU
- Hard to enter the «DAX Mind»
  - For MDX and SQL programmers
  - Don't search same concepts, they are not there!
- Once you master it… awesome!
- Simple, not easy

## What is changing?

| DENALI | SQL 2008 R2 |
|---|---|
| BISM | UDM |
| DAX | MDX |
| RELATIONAL | MULTIDIMENSIONAL |
| IN MEMORY | ON DISK |
| FAST VERTIPAQ ENGINE | AGGREGATIONS |
| LEAF LEVEL COMPUTATION | CALCULATED MEMBERS |
| FILTER / ROW CONTEXT | MDX FUNCTIONS, TUPLES ETC |
| SELF SERVICE + CORPORATE | CORPORATE ONLY |
| EASY DATA INTEGRATION | BI PROFESSIONALS NEEDED |

A NEW BOOK FROM SQLBI




## PowerPivot for Excel 2010: give your data meaning

sqlbi.com

# Questions
and
# Answers

sqlbi.com

## Links

- SQLBI Website
  **www.sqlbi.com**
- PowerPivot Workshop
  **www.powerpivotworkshop.com**
- Marco Russo blog
  **www.sqlblog.com/blogs/marco_russo**
- Alberto Ferrari blog
  **www.sqlblog.com/blogs/alberto_ferrari**

For any question contact us at
**info@sqlbi.com**

# Thank you!

sqlbi.com