STRUCTURED SPREADSHEET MODELLING AND IMPLEMENTATION WITH MULTIPLE DIMENSIONS

PART 2: IMPLEMENTATION

PAUL MIREAULT
HONORARY PROFESSOR
HEC MONTRÉAL
paul.mireault@ssmi.international

EuSpRIG, 2018-07-05, LONDON
# Table of Contents

- Examples and their problems
- Review of the SSMI Conceptual model
- Implementation principles from Data Warehouse Design
- Structured implementation
  - Model Management Variables
  - Primary and Foreign Keys
  - Formulas with multi-dimensional variables
- Maintenance
**Zero or One Dimension**

*The most common spreadsheet*

- Zero dimension: no aggregation needed (or possible!)

- One dimension:
  - Variables in rows.
  - One dimension with instances in columns, like Months.
  - Aggregate calculation performed on row.
**TWO DIMENSIONS**

Developers start using “tricks”

- Use columns for one dimension.
- Use a set of worksheets for the other dimension.

![Excel spreadsheet example](image)
### Three Dimensions

**Implementation Trick:** Repeated sheets for one dimension, columns for the another dimension and repeated blocks of variables for the third dimension.

Revenue is a variable with 72 instances: 3 Products x 8 Quarters x 3 Regions.
Aggregate Calculations

Aggregate Variables

- From Revenue(Product, Region, Quarter) we can calculate 7 aggregate variables:
  - Revenue PR (Product, Region)
  - Revenue PQ (Product, Quarter)
  - Revenue QR (Quarter, Region)
  - Revenue P (Product)
  - Revenue Q (Quarter)
  - Revenue R (Region)
  - Total Revenue
**Calculating Aggregates**

**Different formulas for the same concept**

- **Revenue(Product, Region):** =SUM(F6:M6)
- **Revenue(Product, Quarter):** =F6+F11+F16
- **Revenue(Quarter, Region):** =Apple!F6+Banana!F6+Cherry!F6 or =SUM(Apple:Cherry!F6)
- **Revenue(Product):** =SUM(F6:M6,F11:M11,F16:M16) or =SUM(F6:M6)+SUM(F11:M11)+SUM(F16:M16)
- **Revenue(Region):** =SUM(Apple:Cherry!F6:M6) or =SUM(Apple!F6:M6)+SUM(Banana!F6:M6)+SUM(Cherry!F6:M6)
**CALCULATING AGGREGATES**

**DIFFERENT FORMULAS FOR THE SAME CONCEPT**

\[
\text{Revenue}_{\text{Product}} = \sum_{\text{Region Quarter}} \text{Revenue}_{\text{Product}, \text{Region}, \text{Quarter}}
\]

- \(= \text{SUM}(F6:M6, F11:M11, F16:M16)\)
- \(= \text{SUM}(F6:M6) + \text{SUM}(F11:M11) + \text{SUM}(F16:M16)\)

\[
\text{Revenue}_{\text{Region}} = \sum_{\text{Product Quarter}} \text{Revenue}_{\text{Product}, \text{Region}, \text{Quarter}}
\]

- \(= \text{SUM}([\text{Apple}, \text{Cherry}]!F6:M6)\)
- \(= \text{SUM}([\text{Apple}]!F6:M6) + \text{SUM}([\text{Banana}]!F6:M6) + \text{SUM}([\text{Cherry}]!F6:M6)\)

\[
\text{Revenue}_{\text{Quarter}} = \sum_{\text{Product Region}} \text{Revenue}_{\text{Product}, \text{Region}, \text{Quarter}}
\]

- \(= \text{SUM}([\text{Apple}, \text{Cherry}]!F6) + \text{SUM}([\text{Apple}, \text{Cherry}]!F11) + \text{SUM}([\text{Apple}, \text{Cherry}]!F16)\)
- \(= [\text{Apple}]!F6 + [\text{Apple}]!F11 + [\text{Apple}]!F16 + [\text{Banana}]!F6 + [\text{Banana}]!F11 + [\text{Banana}]!F16 + [\text{Cherry}]!F6 + [\text{Cherry}]!F11 + [\text{Cherry}]!F16\)

\[
\text{Revenue}_{\text{Product, Region}} = \sum_{\text{Quarter}} \text{Revenue}_{\text{Product, Region, Quarter}}
\]

- \(= \text{SUM}(F6:M6)\)

\[
\text{Revenue}_{\text{Product, Quarter}} = \sum_{\text{Region}} \text{Revenue}_{\text{Product, Region, Quarter}}
\]

- \(= F6 + F11 + F16\)

\[
\text{Revenue}_{\text{Region, Quarter}} = \sum_{\text{Product}} \text{Revenue}_{\text{Product, Region, Quarter}}
\]

- \(= [\text{Apple}]!F6 + [\text{Banana}]!F6 + [\text{Cherry}]!F6\)
- \(= \text{SUM}([\text{Apple}, \text{Cherry}]!F6)\)

\[
\text{Total Revenue} = \sum_{\text{Product Region Quarter}} \text{Revenue}_{\text{Product, Region, Quarter}}
\]

- \(= \text{SUM}([\text{Apple}, \text{Cherry}]!F6:M6, [\text{Apple}, \text{Cherry}]!F11:M11, [\text{Apple}, \text{Cherry}]!F16:M16)\)
- \(= \text{SUM}([\text{Apple}]!F6:M6, [\text{Apple}]!F11:M11, [\text{Apple}]!F16:M16, [\text{Banana}]!F6:M6, [\text{Banana}]!F11:M11, [\text{Banana}]!F16:M16, [\text{Cherry}]!F6:M6, [\text{Cherry}]!F11:M11, [\text{Cherry}]!F16:M16)\)
FUTURE MODIFICATIONS
PROBLEMS ARE LURKING IN THE WINGS

• What if we add a Product? a Region? a Quarter?

• The manipulations are more complex, and increase the probability of making an error.

• It is very hard to present a different view: Sheets for each Region, with Products and Quarters. Sheets for each Quarter, with Products and Regions.

• The problem is that the developer mixes the Interface (the output layout) with the Application (the calculations).
**FOUR DIMENSIONS**

- Very rarely seen in the wild. Developers avoid them.
- Well.. the truth is not many know how handle them.
THE SSMI METHODOLOGY

In Information Systems, we use a **Conceptual Model**.

- It describes what the user **needs**, without references to the technology used to implement it.

- The Formula List and the Formula Diagram are the conceptual model.

### Formula List

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>Average price of widgets</td>
<td>Input</td>
<td>= Revenue - Total Cost</td>
</tr>
<tr>
<td>Profit</td>
<td>Profit of each region</td>
<td>Output, repeating</td>
<td>= Revenue - Total Cost</td>
</tr>
<tr>
<td>DemParA</td>
<td>First Demand function parameter</td>
<td>Parameter</td>
<td>247,1,000</td>
</tr>
<tr>
<td>DemParB</td>
<td>Second Demand function parameter</td>
<td>Parameter</td>
<td>1,1009</td>
</tr>
<tr>
<td>Fixed Cost</td>
<td>Fixed cost of manufacturing the widgets</td>
<td>Parameter</td>
<td>3,500,000</td>
</tr>
<tr>
<td>Mfg Cost</td>
<td>Cost of manufacturing one widget</td>
<td>Parameter</td>
<td>1219</td>
</tr>
<tr>
<td>Distribution</td>
<td>Proportion of the Demand sold in each region</td>
<td>Parameter, repeating</td>
<td>48%, 23%, 23%</td>
</tr>
<tr>
<td>Delivery Cost</td>
<td>Cost of delivery of widgets in each region</td>
<td>Parameter, repeating</td>
<td>568, 830, 690</td>
</tr>
<tr>
<td>Total Demand</td>
<td>Demand of widgets, formula given by the market research specialist</td>
<td>Calculated</td>
<td>= DemParA * Demand*Volume</td>
</tr>
<tr>
<td>Regional Demand</td>
<td>Portion of the Demand sold in each region</td>
<td>Calculated, repeating</td>
<td>= Total Demand * Distribution</td>
</tr>
<tr>
<td>Total Cost</td>
<td>Total Cost of selling widgets in each region</td>
<td>Calculated, repeating</td>
<td>= Regional Fixed Cost + Variable Cost</td>
</tr>
<tr>
<td>Regional Fixed Cost</td>
<td>Portion of the Fixed cost allocated in each region</td>
<td>Calculated, repeating</td>
<td>= Fixed Cost * Distribution</td>
</tr>
<tr>
<td>Variable Cost</td>
<td>Variable Cost of the widgets sold in each region</td>
<td>Calculated, repeating</td>
<td>= Regional Demand * Unit Cost</td>
</tr>
<tr>
<td>Unit Cost</td>
<td>Unit cost of one widget in each region</td>
<td>Calculated, repeating</td>
<td>= Mfg Cost + Delivery Cost</td>
</tr>
<tr>
<td>Revenue</td>
<td>Revenue of each region</td>
<td>Calculated, repeating</td>
<td>= RegionalDemand * Price</td>
</tr>
<tr>
<td>Total profit</td>
<td>Total profit of all regions</td>
<td>Output</td>
<td>= DEM(PROFITS)</td>
</tr>
</tbody>
</table>

### Formula Diagram
CONCEPTUAL MODEL EXAMPLE

Acme TechnoWidget Company

• Four Dimensions:
  • Product: Standard, Deluxe
  • Region: N, SE, SW, E and W
  • Month: January to December
  • Sector: Government, Military, Private, Education
Expected Advantages

• Complex problems need the developer’s full attention: using Excel in the modelling stage is a source of interruptions of the creative process.

• The dimension set of a variable is clearly defined.

• Use of a diagram leads to significantly less omission errors (O’Donnel, 2001).

• Use of a diagram does not significantly take more time (O’Donnel, 2001).
DATA WAREHOUSE DESIGN AND IMPLEMENTATION CONCEPTS

- Constellation Schema

- A Fact Table has a composite PK that refers to a Dimension Table.
Tables with Non-Unique Foreign Keys
**Renaming Foreign Keys**

- **Sector-Product**
  - SP
  - $S$ in SP
  - $P$ in SP
  - Product Distribution per Sector
  - Annual Sector-Product Unit Sales
  - Annual Sector-Product Sales
  - Amount
  - Price

- **Month-Sector-Product**
  - MSP
  - $M$ in MSP
  - $MS$ in MSP
  - $MP$ in MSP
  - $SP$ in MSP
  - MSP Sales Amount
  - MSP Unit Sales

- **Product**
  - Product Code
  - Product
  - Base Price Multiplier
  - Unit Production Cost

- **Region**
  - Region Code
  - Region
  - Unit Delivery Cost

- **Month-Sector**
  - MS
  - Monthly Sales Distribution per Sector

- **Month-Product**
  - MP
  - MP Sales Amount
  - MP Unit Sales

- **Sector**
  - Sector Code
    - Sector
    - Rebate Percentage
    - Sector Price Factor
    - DemParA
    - DemParB
    - Sector Base Price
    - Sector Annual Demand Units

- **Sector-Region**
  - SR
  - Region Sales Distribution per Sector

- **Month-Sector-Product-Region**
  - MSPR
    - $M$ in MSPR
    - $SR$ in MSPR
    - $PR$ in MSPR
    - MSP in MSPR
    - $MSP$ in MSPR
    - MSPR Variable Cost
    - MSPR Unit Sales

- **Month**
  - Month Code
    - Month
    - Monthly Cost
    - Monthly Variable Cost
    - Monthly Profit
    - Monthly Unit Sales
    - Monthly Sales Amount

- **Month-Product-Region**
  - MPR
    - MPR
    - MPR Unit Sales
# Tables with Unique Foreign Keys

## Sector-Product
- **SP**
  - Sector Code
  - Sector
  - Rebate Percentage
  - Sector Price Factor
  - DemParA
  - DemParB
  - Sector Base Price
  - Sector Annual Demand Units
- **P in SP**
  - Product Distribution per Sector
  - Annual Sector-Product Unit Sales
  - Annual Sector-Product Sales
  - Amount
  - Price

## Region
- **Region Code**
  - Region
- **Unit Delivery Cost**

## Product
- **Product Code**
  - Product
- **Base Price Multiplier**
  - Unit Production Cost

## Month-Sector
- **MS**
  - Monthly Sales Distribution per Sector

## Month-Product
- **MP**
  - MP Sales Amount
  - MP Unit Sales

## Sector-Region
- **SR**
  - Region Sales Distribution per Sector

## Month-Sector-Product
- **MSP**
  - MSP Sales Amount
  - MSP Unit Sales

## Month-Sector-Product-Region
- **MSPR**
  - MSPR Variable Cost
  - MSPR Unit Sales

## Month
- **Month Code**
  - Month
- **Monthly Cost**
- **Monthly Variable Cost**
- **Monthly Profit**
- **Monthly Unit Sales**
- **Monthly Sales Amount**

## Month-Product-Region
- **MPR**
  - MPR Unit Sales
Determining the Worksheets
Using the Dimension Sets and the Variable Types

• If a dimension set contains at least one calculated variable, create a worksheet with the dimension set name.

• If a dimension set contains at least one data variable, create a worksheet with the dimension set name followed by “Data”.
MODEL MANAGEMENT VARIABLES

• In the implementation, we need variables that are not part of the Formula Diagram.

• We need to manage the Primary Keys of each dimension set, as well as the Foreign Keys we will use in reference formulas and in aggregation formulas.

• Some variables are useful for the developer, like the last column of each dimension set.

• Other variables can perform error-checking and warn the user of an anomaly.
In different dimension sets, the Excel columns have no relation.
MODEL MANAGEMENT VARIABLES
LAST COLUMN OF EACH DIMENSION SET

- Calculate the last column used in each dimension set.

- Simple dimension sets:
  - Number of elements + 1 (for the labels in Column A) + initialization and documentation columns.

- Compound dimension sets:
  - Product of the size of its simple dimensions + 1 + other columns, or
  - Calculate size from imported data.
A variable in the same dimension set is referenced by name.

A variable in a smaller dimension set is referenced with an INDEX-MATCH formula.

INDEX(Variable, MATCH(Foreign Key, Primary Key, Exact match code))
**Using Foreign Keys**

**Aggregating a Variable from a Bigger Dimension Set**

We create an area with the same size as the bigger dimension set.

The variable in the smaller dimension set is calculated with a xIF type formula.

SUMIF(Foreign Key, Primary Key, Variable)
The developer may decide that it is easier for the user to enter some data in a format that is different than the linearized form of the Data worksheets.

He prepares and Interface sheet with the data entry in the desired format.

The data variables, in the Data worksheets, reference the corresponding Interface worksheet cells

- Direct cell references, or
- LOOKUP or INDEX-MATCH references.
**Interface and Interface Preparation**

**Input Data Example**

![Excel Sheets Example](image-url)
Presenting an output variable from the zero-dimensional dimension set is straightforward: use the variable’s name.

Presenting an output variable from a one-dimensional dimension set is also straightforward: use the variable’s name if the presentation is aligned with the variable’s definition, or use INDEX-MATCH with a properly constructed primary key.
Interface and Interface Preparation

Output Data, multi-dimensional

• Presenting an output variable from a multi-dimensional dimension set is a bit more complex.

• We first create the layout we want: every value needs to be characterized by its dimensions.
## Interface and Interface Preparation

### Layout of Multi-Dimensional Variables

<table>
<thead>
<tr>
<th>Interface</th>
<th>Product S</th>
<th>Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interface</th>
<th>Product D</th>
<th>Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unit Sales by Month, Product and Region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unit Sales by Month and Product</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unit Sales by Month</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sales Amount by Month and Product</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Profit by Month</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
We then create a Model Management worksheet to prepare the primary keys that will be used.
### Interface and Interface Preparation

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>P</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Interface Preparation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>INDEX-MATCH preparation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Product Index</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Product Code</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Region Index</td>
<td>Region Code</td>
<td>Month Index</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>N</td>
<td>SE</td>
<td>SW</td>
<td>E</td>
<td>W</td>
<td>N</td>
<td>SE</td>
<td>SW</td>
<td>E</td>
<td>W</td>
<td>N</td>
<td>SE</td>
<td>SW</td>
</tr>
<tr>
<td>7</td>
<td>Region Code</td>
<td>Month Code</td>
<td>1</td>
<td>Jan</td>
<td>Feb</td>
<td>Mar</td>
<td>Apr</td>
<td>May</td>
<td>Jun</td>
<td>Jul</td>
<td>Aug</td>
<td>Sep</td>
<td>Oct</td>
<td>Nov</td>
<td>Dec</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table above shows the interface and interface preparation for different product codes and region codes, along with the corresponding month codes and index values.
We finally complete the Interface sheet with INDEX-MATCH references of the desired variable and direct references to the constructed primary keys.
## Interface and Interface Preparation

### Layout of Multi-Dimensional Variables

![Excel spreadsheet screenshot](image.png)

- **=INDEX(MPR_Unit_Sales;MATCH('Interface Preparation'!H10;MPR;0))**

<table>
<thead>
<tr>
<th>Interface</th>
<th>Enter Base Price: $140,000</th>
<th>Total Profit $302,799</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product S</td>
<td>Regions N</td>
<td>161</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>SW</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>93</td>
</tr>
<tr>
<td>Product D</td>
<td>Regions N</td>
<td>164</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>SW</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>89</td>
</tr>
<tr>
<td>Unit Sales by Month and Product</td>
<td>Product S</td>
<td>641</td>
</tr>
<tr>
<td></td>
<td>Product D</td>
<td>581</td>
</tr>
<tr>
<td>Unit Sales by Month</td>
<td>1221</td>
<td>1314</td>
</tr>
<tr>
<td>Sales Amount by Month and Product</td>
<td>Product S</td>
<td>$54,660</td>
</tr>
<tr>
<td></td>
<td>Product D</td>
<td>$91,844</td>
</tr>
<tr>
<td>Profit by Month</td>
<td>$42,043</td>
<td>$38,584</td>
</tr>
</tbody>
</table>
IMPACT OF A NEW SECTOR

Determining the changes in the worksheets

- We add a new sector.
- Every worksheet with a dimension set containing Sector will need more columns.
- Every aggregation formula using a variable from a dimension set containing Sector will need more columns.
LIVE EXCEL DEMONSTRATION

- Month-Product-Sector-Region
  - RN1 to WD1
  - Monthly Unit Sales and Monthly Variable Cost in Month
    - RN1 to WD1
  - Monthly Sales Amount in Month
    - CT1 to DR1
Questions?