Data Warehouse Matrix Estimation Spreadsheet Instructions

Adam Gimore (Dimodelo Solutions)   7/1/16

Contents

Step 1 Define Variables

The first step is to provide the estimates (in days) for developing Facts and Dimensions of different complexity. These estimates are used calculate the number of days it will take to develop each Dimension and Fact on the Data Warehouse Matrix sheet.

There are two estimates per complexity band. One estimate represents the number of days of development when developing the Data Warehouse with the aid of Dimodelo Architect. The second estimate represents the number of days when developing without the aid of Dimodelo Architect.

We have already filled in the estimates based on our experience of projects using Dimodelo Architect, and projects using a more conservation approach and standard tools like Visual Studio and SQL Server Management studio.

If you choose, you can modify the estimates (in blue) for the 'Without Dimodelo' version of the project.

Estimates

Use the table below to assign a number of days to each of the complexity options available in the Data Warehouse matrix. Use a With and Without Dimodelo Architect value. Our experience is that Dimodelo Architect saves around 65% of the build time.

Complexity Code	Complexity Name	With Dimodelo Estimate (Days)	Without Dimodelo Estimate (Days)	Description
S	Simple	1	3	A Simple Dimension or Fact is source from a single existing source, with mostly a one to one relationship between source columns and destination columns. The Source is readily available and accessible.
M	Medium	1.75	5	A Medium Dimension or Fact is source from a single existing source, with mostly a one to one relationship between source columns and destination columns. However the source many need to be created, purely for reporting purposes, or there may be many attributes or measures.
C	Complex	2.8	8	Complex will involves at least one issues like: the Source is undecided, data quality issues, multiple sources, many attributes, many measures, many dimension associations, or there are complex transformations and/or derived staging tables required to derive measures or attributes.
HC	Highly Complex	4.2	12	Highly Complex will involve a combination of a number of issues like: the Source is undecided, data quality issues, multiple sources, many attributes, many measures, many dimension associations, or there are complex transformations and/or derived staging tables required to derive measures or attributes.

There are variables for the percentage breakdown of the estimate over various development tasks. See below:

Task Breakdown

Use the table below to breakdown the work required for each Fact and Dimension into Analysis, Modelling, Source-Target mapping and Build/Unit Testing tasks. The braekdown is used in the Data Warehouse Matrix spreadsheet to break down all work over these task categories.

Task	Breakdown
Analysis	15%
Modeling	15%
Source Target Mapping	15%
Build & Unit Test, Integrate	55%

Plus a variable to adjust the daily rate. The daily rate is used on the Project Work Breakdown sheet to calculate labour cost.
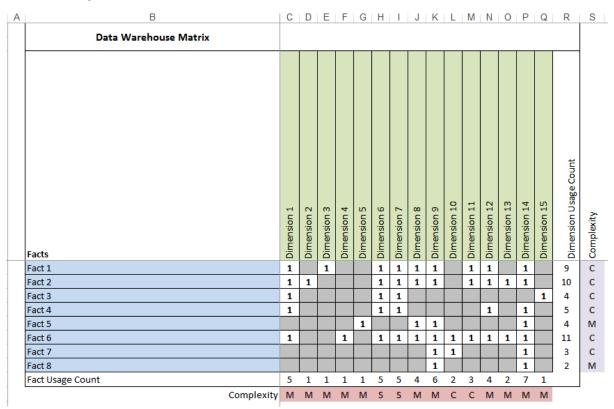
Variables

Daily Rate $ 1,500

Step 2 Define the Data Warehouse Matrix

The Data Warehouse matrix is used to define the scope of the Data Warehouse including the Facts and Dimensions and their relationships. It is a good planning, high level design and communication tool.

To use the Data Warehouse matrix you first need to add the names of your Facts to rows (blue cells) and the names of your Dimensions to the columns (green cells). You need to give each Fact and Dimension a complexity. Available complexities are Simple (S), Medium (M), Complex (C) and Highly Complex (HC). Use the same code as the code on the Variables sheet. Use the purple cells to define the complexity of each Fact and the red cells to define the complexity of each Dimension.

To delete a fact or Dimension simply delete the row or column.

To add an additional Fact or Dimension insert a row or column within the matrix, and then copy the adjacent cells form an existing fact and dimension to the new column/row. Make sure you copy all cells, including the estimates cells.

Data Warehouse Matrix

Facts	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6	Dim 7	Dim 8	Dim 9	Dim 10	Dim 11	Dim 12	Dim 13	Dim 14	Dim 15	Dimension Usage Count	Complexity
Fact 1	1		1			1	1	1	1		1	1		1		9	C
Fact 2	1	1				1	1	1	1		1	1	1	1		10	C
Fact 3	1					1	1								1	4	C
Fact 4	1					1	1					1		1		5	C
Fact 5					1			1	1					1		4	M
Fact 6	1			1		1	1	1	1	1	1	1	1	1		11	C
Fact 7								1	1					1		3	C
Fact 8								1						1		2	M
Fact Usage Count	5	1	1	1	1	5	5	4	6	2	3	4	2	7	1		
Complexity	M	M	M	M	M	S	S	M	M	C	C	M	M	M	M		

Step 3 - Define the relationships

To define the relationships between the Facts and Dimensions, you put a value in the grey column that is the intersect of the Fact and Dimension. The cells have automatic formatting enabled. By putting a 1 in the cell at the intersection of a fact row and dimension column you are indicating that the Fact is associated to that dimension (and therefore the facts can be analysed by attributes of the dimension). When you put a one in a grey cell it will change to a white background.

The Dimension Usage Count and Fact Usage count is a good indicator of a Dimensions importance and the complexity of a Fact.

Step 4 – Estimate Data Warehouse development duration

The spreadsheet automatically calculates the estimates (in days) for the Data Warehouse development for both the Fact and the Dimensions. Both the "With" and "Without" Dimodelo Architect Estimates are calculated. These estimates are used by the project Work Breakdown sheet to calculate cost.

Fact Estimates	Analysis	Modeling	Source Target Mapping	Build & Unit Test	With Dimodelo Estimate	Without Dimodelo Estimate
	0.3	0.3	0.5	1.8	3.0	8.0
	0.3	0.3	0.5	1.8	3.0	8.0
	0.3	0.3	0.5	1.8	3.0	8.0
	0.3	0.3	0.5	1.8	3.0	8.0
	0.2	0.2	0.2	0.9	1.5	5.0
	0.3	0.3	0.5	1.8	3.0	8.0
	0.3	0.3	0.5	1.8	3.0	8.0
	0.2	0.2	0.2	0.9	1.5	5.0
	2.1	2.1	3.2	13	21.0	58.0

Figure 1 Fact Estimates

Dimension Estimates																
Analysis	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.2	0.2	0.3	0.3	0.2	0.2	0.2	0.2	2.4
Modeling	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.2	0.2	0.3	0.3	0.2	0.2	0.2	0.2	2.4
Source Target Mapping	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.2	0.2	0.5	0.5	0.2	0.2	0.2	0.2	3.5
Build & Unit Test	0.9	0.9	0.9	0.9	0.9	0.3	0.3	0.9	0.9	1.8	1.8	0.9	0.9	0.9	0.9	14.1
With Dimodelo Estimate	1.5	1.5	1.5	1.5	1.5	0.5	0.5	1.5	1.5	3.0	3.0	1.5	1.5	1.5	1.5	23.5
Without Dimodelo Estimate	5.0	5.0	5.0	5.0	5.0	3.0	3.0	5.0	5.0	8.0	8.0	5.0	5.0	5.0	5.0	77.0

Figure 2 Dimension Estimates

A total days summary is provided.

```
With Dimodelo -  Total Days    47
                 vs
Without Dimodelo - Total Days  135
```

Step 5 - Project Costing

Every project is different, but we have endeavoured to develop a project template that is a good representation of the phase 1 of the Data Warehouse component of a BI project. Note it doesn't include any reporting activity.

DIMODELO
solutions
DATA WAREHOUSE AUTOMATION

Typical Work Breakdown

	Days		$	Notes
Analysis				
Business Analysis	10	$	15,000	Investigate requirements, business processes, sources.
High Level Design	5	$	7,500	Develop a high level design for the DW.
PreRequisites				
Establish Development Environment	3	$	4,500	Establish Connectivity, Logins, Source Control,
Establish Test Environment	2	$	3,000	Development copies of databases etc
Build and Unit Test				
Data Warehouse				
Dimensions	77	$	115,500	**Without Dimodelo Architect**
Facts	58	$	87,000	**Without Dimodelo Architect**
Cube/Tablular Model	8	$	12,000	Including any calculated Measures
Manual Data Entry	4	$	6,000	Master Data Services. Setup MDS
				This represents time required to respond to issues found in
System/UAT Testing	5	$	7,500	Testing
				Several Deployments over course of project, including TEST
Deployment	5	$	7,500	and PROD.
Security	5	$	7,500	Implement data security.
Project Management	15	$	22,500	Time for reporting status, communicating project strategy.
Documentation & Handover	5	$	7,500	
Without Dimodelo Architect Total	202	$	303,000	
Savings With Dimodelo Architect				
Dimension Savings	-50.15	-$	75,225	
Fact Savings	-37.7	-$	56,550	
With Dimodelo Architect Total	114.15	$	171,225	
Dimodelo Architect Savings		$	131,775	

contact@dimodelo.com

You may want to make changes to the project plan and estimates, but don't modify Data Warehouse Fact and Dimensions estimates. Their value comes from the Data Warehouse Matrix spread sheet.

You will need to adjust the Daily Rate (on the variable sheet) to reflect the Daily rate attributable to development effort on the project. Daily rate isn't just "annual salary"/"number of work days" in a year (approx 200). This usually includes an overhead for providing the workplace and all other costs associated with employment.

The costing provides a cost estimate for the project without the aid of Dimodelo Architect and a cost estimate with the aid of Dimodelo Architect.

One of the great advantages of Dimodelo Architect is it makes it very easy to change your design and iterate quickly. Going forward there are additional major cost savings using Dimodelo Architect over and above the savings made on initial development.

About Dimodelo Solutions

Dimodelo Solutions are the creators of Dimodelo Architect, a Data Warehouse Automation tool targeting the Microsoft SQL Server platform.

Dimodelo Architect has been proven to reduce the time it takes to build a Data Warehouse by 65%.

To contact us, or arrange a free demo please email contact@dimodelo.com

You can visit our website and download a free Trial at www.dimodelo.com