

Extending the Multidimensional Model for Linking Cubes

Alberto Sabaini¹ **Esteban Zimányi**² **Carlo Combi**¹

Department of Computer Science, University of Verona, Italy
`{alberto.sabaini, carlo.combi}@univr.it`
Department of Computer and Decision Engineering,
Université Libre de Bruxelles, Belgium
`ezimanyi@ulb.ac.be`

Introduction & Objectives

Motivating Scenario

Connecting Multidimensional Cubes

Drill Across Link Operator

Conclusions

Introduction & Objectives

Motivating Scenario

Connecting Multidimensional Cubes

Drill Across Link Operator

Conclusions

Introduction

- Every multidimensional fact is based on a set of dimensions and measures
- Dimensions are organized as hierarchies of levels that allow analysts to aggregate measures at different degrees of detail
- Dimensions specify different ways data can be queried, viewed, aggregated, and sorted

Introduction

- Designing interactions between dimensions and facts is a major challenge [3]
- Another issue is to represent connections between different facts in the same schema
- Shared dimensions are needed for navigating multidimensional cubes
- Non-shared dimensions provide users the ability to view and analyze data that would be otherwise not available

Objectives

- We introduce an explicit link that relates two multidimensional cubes
- Such a link indicates that they represent different aspects of the same reality
- We argue that the standard drill across operator is not suited for analyze such connection
- We provide an extension called drill across link

Introduction & Objectives

Motivating Scenario

Connecting Multidimensional Cubes

Drill Across Link Operator

Conclusions

Pharmacovigilance

Adverse Reactions



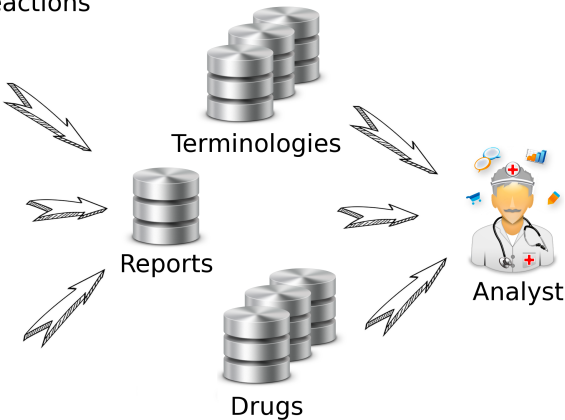
"Still, let's do an x-ray just to be sure."



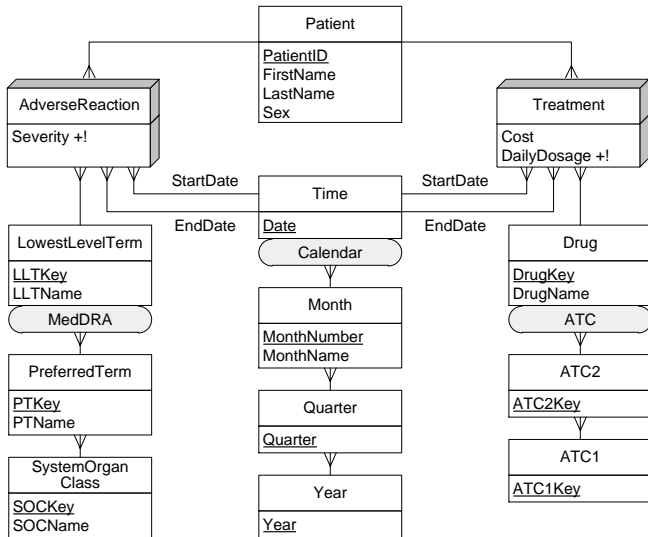
"It's cabin fever, but I've never seen it this acute."



"It's cabin fever, but I've never seen it this acute."



Facts Analysis and Model



Treatment Examples

TKey	Patient	Drug	Cost	DailyDosage
T1	Alice	Tylenol	65	40
T2	Alice	Tylenol	20	20
T3	Alice	Aspirin	60	30
T4	Bob	Aspirin	80	30
T5	Bob	Tylenol	60	30
T6	Charlie	Tylenol	30	40
T7	Charlie	Aspirin	35	50
T8	Charlie	Tylenol	70	20

Adr Examples

AdRKey	Patient	Reaction	Severity
A1	Alice	Hepatitis	6
A2	Alice	Urticaria	1
A3	Bob	Hepatitis	4
A4	Bob	Urticaria	5
A5	Charlie	Urticaria	9
A6	Charlie	Hepatitis	7

Introduction & Objectives

Motivating Scenario

Connecting Multidimensional Cubes

Drill Across Link Operator

Conclusions

Connecting Multidimensional Cubes

- Pharmacovigilance deals in assessing suspected cases of adverse reactions induced by drug administrations
- Usually drill across operator combines cubes by means of their shared dimensions

Consider the following query

What is the maximum daily dosage for drugs suspected to have induced a skin disorder reaction?

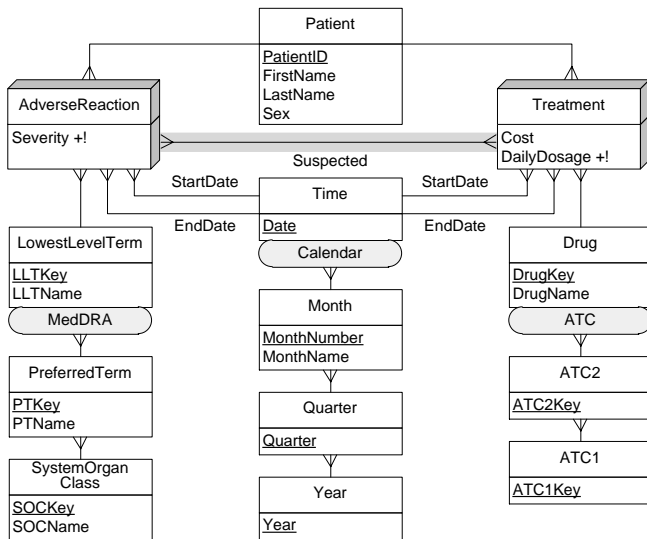
Issues and Restrictions

- By only using shared dimensions, the result might be incorrect
- The standard drill across operator merges facts from the first cube with facts in the second one
- The connection is based on equal values in the shared dimensions.
- However, some of the mentioned facts from the first cube may not be related to the facts in the second one.

Issues and Restrictions

- Shared dimensions represent an implicit connection between cubes
- Instead of relying on common members of shared dimensions, we could rely on explicit links between cubes

Multidimensional Model



Introduction & Objectives

Motivating Scenario

Connecting Multidimensional Cubes

Drill Across Link Operator

Conclusions

Drill Across Link Operator

- The drill across link operator combines the cells from two data cubes that are connected by an explicit link
- The fact schema of the resulting instance corresponds to the union of all dimensions and measures from both source facts
- Only one copy of shared dimensions will be kept

Combining Shared Dimensions

- A shared-dimension could have different coordinates in the two source cubes
- *StartDate* is a role-playing dimension in both Treatment and AdverseReaction cubes
- Imposing these dates to be the same would limit the analysis capabilities
- By renaming these roles, users may avoid such a behavior

Bridge Table

TreatmentKey	AdvReactionKey
T1	A1
T2	A1
T2	A2
T3	A1
T3	A2
T4	A3
T5	A4
T6	A5
T7	A6
T8	A6

Drill Across Link Operator

TKey	AdRKey	Patient	Drug	Cost	DailyDosage	Reaction
T1	A1	Alice	Tylenol	65	40	Hepatitis
T2	A1	Alice	Tylenol	20	20	Hepatitis
T2	A2	Alice	Tylenol	20	20	Urticaria
T3	A1	Alice	Aspirin	60	30	Hepatitis
T3	A2	Alice	Aspirin	60	30	Urticaria
T4	A3	Bob	Aspirin	80	30	Hepatitis
T5	A4	Bob	Tylenol	60	30	Urticaria
T6	A5	Charlie	Tylenol	30	40	Urticaria
T7	A6	Charlie	Aspirin	35	50	Hepatitis
T8	A6	Charlie	Tylenol	70	20	Hepatitis

Measure Aggregation

- Many-to-many relationships may introduce errors when computing aggregates
- The same problem arises in cube-dimension many-to-many relationships
- Some of the measure values may appear several times

Consider the following query

Show the total cost and the maximum daily dosage of treatments per adverse reaction

TKey	AdRKey	Patient	Drug	Cost	DailyDosage	Reaction
T1	A1	Alice	Tylenol	65	40	Hepatitis
T2	A1	Alice	Tylenol	20	20	Hepatitis
T2	A2	Alice	Tylenol	20	20	Urticaria
T3	A1	Alice	Aspirin	60	30	Hepatitis
T3	A2	Alice	Aspirin	60	30	Urticaria
T4	A3	Bob	Aspirin	80	30	Hepatitis
T5	A4	Bob	Tylenol	60	30	Urticaria
T6	A5	Charlie	Tylenol	30	40	Urticaria
T7	A6	Charlie	Aspirin	35	50	Hepatitis
T8	A6	Charlie	Tylenol	70	20	Hepatitis

Multidimensional Normal Forms

- The *double counting* issue can be analyzed through multidimensional normal forms (MNFs) [2]
- Correct measure aggregation is ensured by MNFs requirements
- 1MNF requires each measure to be functionally determined by the set of associated leaf levels
- The resulting cube does not satisfy the 1MNF

Multidimensional Normal Forms

- The double-counting issue may be solved by distributing measure values among multiple instances [4]
- Measures need to be adjusted when the sum aggregation operator is applied

Consider the following query

Show the total cost and the maximum daily dosage of treatments per adverse reaction

Counting Multiple Treatment Instances

TreatmentKey	ReplicationCount
T1	1
T2	2
T3	2
T4	1
T5	1
T6	1
T7	1
T8	1

TKey	AdRKey	Patient	Drug	Cost	DailyDosage	Reaction
T1	A1	Alice	Tylenol	65	40	Hepatitis
T2	A1	Alice	Tylenol	20 10	20	Hepatitis
T2	A2	Alice	Tylenol	20 10	20	Urticaria
T3	A1	Alice	Aspirin	60 30	30	Hepatitis
T3	A2	Alice	Aspirin	60 30	30	Urticaria
T4	A3	Bob	Aspirin	80	30	Hepatitis
T5	A4	Bob	Tylenol	60	30	Urticaria
T6	A5	Charlie	Tylenol	30	40	Urticaria
T7	A6	Charlie	Aspirin	35	50	Hepatitis
T8	A6	Charlie	Tylenol	70	20	Hepatitis

Single Cube Alternative

- We point out that two related facts may be represented as a single cube
- This allows users to by pass the use of the standard drill across operator, intrinsically overcoming its limitations
- All dimensions would in fact be available
- However, due to data replication, issues related to measure aggregation would be introduced even for simple queries

Introduction & Objectives

Motivating Scenario

Connecting Multidimensional Cubes

Drill Across Link Operator

Conclusions

Conclusions

- The inclusion of non-shared dimensions provides the analyst to view data according to different perspectives
- To overcome this limitation, we introduced an explicit link that relates two multidimensional cubes
- We extended the standard drill across operator by proposing an extension called drill across link
- We addressed the double-counting problem that arises when merging two cubes

Thank you!



Ralph Kimball and Margy Ross.

The data warehouse toolkit: The definitive guide to dimensional modeling.

John Wiley & Sons, 2013.



Jens Lechtenbörger and Gottfried Vossen.

Multidimensional normal forms for data warehouse design.

Information Systems, 28(5):415–434, 2003.



William Rowen, Il-Yeol Song, Carl Medsker, and Edward Ewen.

An analysis of many-to-many relationships between fact and dimension tables in dimensional modeling.

In Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW 2001), Interlaken, Switzerland, 2001.



Alejandro Vaisman and Esteban Zimányi.

Data Warehouse Systems: Design and Implementation.

Springer, 2014.