Managing the Two Sides of Decision Modeling: Science and Art

Published in www.modernanalyst.com April 28, 2014

By Barbara von Halle

Barbara von Halle is Managing Partner of Knowledge Partners International, LLC (KPI). She is co-inventor of the Decision Model and co-author of The Decision Model: A Business Logic Framework Linking Business and Technology published by Auerbach Publications/Taylor and Francis LLC 2009.
Larry and Barb can be found at www.TheDecisionModel.com.

A previous issue of this column [1] introduced three forms of normalization behind The Decision Model. It also pointed out that these normalization forms are similar in concept to the same normal forms defined for the relational model. That column explained how the purposes of these three normal forms are similar (but different) for both data and for business logic. It gave examples for The Decision Model.

The fact that there are normal forms for The Decision Model is pioneering. It means that The Decision Model is not simply a new representation but is a discipline grounded in a stable, scientific foundation. If history repeats itself, this foundation should lead to The Decision Model's endurance.

Managing the Two Sides of Decision Modeling: Science and Art

As reassuring as it is that the scientific basis sets The Decision Model apart from other approaches, the science is only half of the story. In the real world, good decision modeling is always a balance between science and art.

The science is systematic decomposition of a structure (of data or logic) into a set of smaller structures based on the definitions of successive normal forms. The art, on the other hand, is a general decomposition into a set of smaller structures based on factors not related to detecting and correcting normalization errors.

The Two Faces of Normalization

Unfortunately, the word normalization is often used in two different ways: one for decomposition resulting from correcting normalization violations and the other for decomposition resulting from other criteria. In the latter usage, the science of normalization is lost. In fact, many people consider the words normalization and decomposition to be synonymous. But, although both result in a set of smaller structures from a larger one, they are not the same. So, in a practical sense, what is the difference and why does it matter?

True Normalization (in this column)

True normalization is a science leading to an exact division. Specifically, if there are violations of normalization within a data table or a Rule Family table, the table contains unnecessary redundancy. Unnecessary redundancy is bad. It not only results in extra overhead in maintaining and executing that redundancy, but it is also error prone due to incorrectly maintaining that redundancy. To avoid such unnecessary redundancy and the inevitable problems it causes, the table with normalization violations must be reduced to a proper normal form. So, in both data and decision models, true normalization leads to a representation of data or logic in its most minimal representation (i.e., least redundant).

General Decomposition

On the other hand, general decomposition based on factors other than normalization, is an art. As an art, it does not lead to an exact division because it purposely supports degrees of freedom. When creating decision models (and data models), both the science and the art are important but for different reasons.

The remainder of this column first presents a data example of normalization and general decomposition. It then presents an example of normalization and general decomposition for The Decision Model.

Data Example

Data Normalization

Let's begin with a simple example for relational data modeling.

© 2014 Knowledge Partners International, LLC All Rights Reserved
www.kpiusa.com ▪ information@kpiusa.com

Consider the simple relational table in Figure 1. It contains data for Person based on the following data requirements:

- A Person has a unique Person ID.
- A Person has only one First Name.
- A Person has only one Last Name.
- A person has only one Annual Salary Amount.

Person Information

Person ID (pk)	Person First Name	Person Last Name	Person Annual Salary Amount
1	George	Smith	$100k
54	Jane	Murray	$250k
13	John	Davis	$65k

Figure 1: Single Relational Table for Person Information

First normal form in the relational model, for this discussion, is a set of data attributes associated with a primary key (i.e., unique identifier) and for which there are no repeating or multivalued data attributes.

The relational table in **Figure 1** has a primary key of Person ID and three non-key data attributes: Person First Name, Person Last Name, and Person Annual Salary Amount. A value for Person ID identifies a specific row in the table whose columns contain the corresponding values for each of the three non-key attributes. As this set of data attributes is associated with a primary key and there are no repeating or multivalued data attributes in this relational table, it is in (at least) first normal form.

Second normal form in the relational model prescribes that the full primary key is needed to identify all non-key data attributes. In other words, there is no non-key data attribute that is functionally dependent on only part of the primary key. Since this primary key has only one part to it, this relational table is (at least) in second normal form.

Third normal form in the relational model prescribes that there be no non-key data attributes that are functionally dependent on other non-key attributes. In **Figure 1**, there are no relationships among the non-key

© 2014 Knowledge Partners International, LLC All Rights Reserved
www.kpiusa.com ▪ information@kpiusa.com

attributes in the relational table. For example, Person First Name does not dictate the value of their Last Name; their Last Name does not dictate the value of their Annual Salary, and so forth. So, essentially, the values for these non-key attributes are independent of each other. Therefore, this table is (at least) in third normal form [2]. There is no unnecessary redundancy in this table.

Data Decomposition

Decomposition simply means a separation into constituent parts, but does not prescribe a specific reason for the separation. There can be many reasons to separate resulting in different decompositions.

For example, one way to decompose the relational table in Figure 1 is to separate it into two relational tables. One relational table would hold the data about a Person's name. The other relational table would hold the data about the Person's income. These are shown in Figure 2.A reason for this decomposition may be that access to Person's income may be more restricted than access to Person's name.

There is no unnecessary redundancy in these tables, according to the relational model. You might think that the primary key is unnecessarily redundant because it appears twice (once in each table) because there are two tables. This redundancy does not qualify as unnecessary redundancy because it is the way to "relate" data from multiple tables when the data is exists in relational table format. So, in this case, the primary key in each table is also a foreign key relating to another table. Foreign keys are not unnecessary, they are core to relational representation.

Person Name Information

Person ID (pk)	Person First Name	Person Last Name
1	George	Smith
54	Jane	Murray
13	John	Davis

Person Salary Information

© 2014 Knowledge Partners International, LLC All Rights Reserved
www.kpiusa.com ▪ information@kpiusa.com

Person Id (pk)	Person Annual Salary Amount
1	$100k
54	$250k
13	$65

Figure 2: Multiple Relational Tables for Person Information

Some people may refer to the tables in Figure 2 as being ***more*** normalized (or sometimes, ***over normalized***) than that in Figure 1. However, this is not a true statement. Both relational tables in Figure 2 are in third normal form (actually higher normal form) because they both conform to the definition of third normal form.

What these people really mean is that the tables in Figure 2 are separated into more parts than the one in Figure 1, but the reason for the separation has nothing to do with resolving a normalization error.

Just to prove that there are many ways to decompose a normalized relational table into a set of smaller normalized tables, see Figure 3. It contains another set of relational tables representing Person Information. In this case, each relational table contains the primary key and only one non-key attributes. All of these tables are in third normal form (actually higher normal form).As long as each piece of data is co-located in a table with the proper primary key according to normalization principles, the resulting table is normalized.

Person First Name Information

Person ID (pk)	Person First Name
1	George
54	Jane
13	John

Person Last Name Information

© 2014 Knowledge Partners International, LLC All Rights Reserved
www.kpiusa.com ▪ information@kpiusa.com

Person ID (pk)	Person Last Name
1	Smith
54	Murray
13	Davis

Person Salary Information

Person ID (pk)	Person Annual Salary Amount
1	$100k
54	$250k
13	$65k

Figure 3: Another Set of Relational Tables for Person Information

Most often in practice, all pieces of data are co-located in one relational table with the corresponding primary key. This is common because it represents one logical data entity (in our case, Person) as one relational table holding all related data elements. As the examples above illustrate, having one relational table for one logical entity results in the fewest tables. However, there is nothing in normalization theory that prevents decomposing such a table into smaller ones, each one in third (or higher) normal form. While such decomposition introduces increased complexity due to having more tables when one suffices, it typically has little or no effect on the stability of the data [3].

Reasons for decomposing data into multiple tables may be differences in security/authorization, general governance of updates, and even geographical distribution. Yet, an advantage in the relational model is that separated tables with the same primary key can be combined together through creation of relational views. So, relational theory provides the best of both worlds. Relational views can virtually recombine separated tables into one or virtually divide one table in multiple ones, in most cases.

The Decision Model Example

© 2014 Knowledge Partners International, LLC All Rights Reserved
www.kpiusa.com ▪ information@kpiusa.com

The Decision Model Normalization

Let's now move onto a simple example for decision modeling according to The Decision Model.

Consider the simple Rule Family table in Figure 4. It contains logic for determining if a new driver is in compliance with the restrictions of a "new driver" license. Assume these restrictions are:

- New driver must drive a car that has specific new driver decals on both bumpers.
- New driver must have seat belt on.
- New driver must not be using a hand-held device.

	Conditions								Conclusion	
Row ID	RP	Vehicle Front Decal		Vehicle Rear Decal		Driver Seat Belt		Driver Hand Held Device		New Driver Compliance	
1	1	Is	Present	Is	Present	Is	Worn	Is	Not in use	Is	Compliant
2	2	Is	Absent							Is	Noncompliant
3	3			Is	Absent					Is	Noncompliant
4	4					Is	Not worn			Is	Noncompliant
5	5							Is	In use	Is	Noncompliant

Figure 4: Rule Family for New Driver Compliance

First normal form in The Decision Model [4] means that each row in a Rule Family cannot be decomposed into more than one row reaching the conclusion. What this means in practice are two considerations: a Rule Family cannot have more than one conclusion column (otherwise it can be decomposed for each conclusion column) and the condition columns are not ORed together (otherwise they can be decomposed into separate rows).

The Rule Family in **Figure 4** has only one conclusion column and all of its condition columns are ANDed to reach the conclusion so it is (at least) in first normal form.

© 2014 Knowledge Partners International, LLC All Rights Reserved
www.kpiusa.com ▪ information@kpiusa.com

Second normal form in The Decision Model prescribes that there be no populated condition cells in a Rule Family table that are irrelevant to reaching the conclusion value. In Figure 4, only the relevant condition cells are populated in each row, so the Rule Family is (at least) in second normal form.

Third normal for in The Decision Model prescribes that no populated condition cells in a Rule Family row lead to the values in another populated condition cell in that row. In Figure 4, there are no relationships among populated condition columns. For example, the Vehicle Front Decal condition cell does not dictate the value of the Vehicle Rear Decal cell; the Vehicle Rear Decal value does not dictate the Driver Seat Belt cell, and so on. So the values for these condition cells are independent of each other, which means the Rule Family is in (at least) third normal form.

True to The Decision Model principles, the logic in Figure 4 is complete in that it covers all possible combinations of input values.

The Decision Model Decomposition

Recall that decomposition simply means a separation into constituent parts, but does not prescribe a specific reason for the separation. This means there can be many reasons and, therefore, many ways to separate the Rule Family in Figure 4 into more than one Rule Family.

For example, one way is to separate logic about Vehicle Decal Compliance from logic about Driver Seat Belt Compliance and to separate logic for both of those from that for Driver Device Compliance, as shown in the Rule Family tables in Figure 5.A reasons for this separation may be the need for different views for Vehicle Decal Compliance - perhaps the logic about vehicle decals varies by state. If so, there would be a view for Vehicle Decal Compliance Rule Family for each state's logic.

		Conditions						**Conclusion**	
Row ID	RP	Vehicle Decal Compliance		Driver Seat Belt Compliance		Driver Device Compliance		New Driver Compliance	
1	1	Is	Compliant	Is	Compliant	Is	Compliant	Is	Compliant
2	2	Is	Noncompliant					Is	Noncompliant
3	3			Is	Noncompliant			Is	Noncompliant

© 2014 Knowledge Partners International, LLC All Rights Reserved
www.kpiusa.com ▪ information@kpiusa.com

4	4					Is	Noncompliant	Is	Noncompliant

		Conditions				Conclusion	
Row ID	RP	Vehicle Front Decal		Vehicle Rear Decal		Vehicle Decal Compliance	
1	1	Is	Present	Is	Present	Is	Compliant
2	2	Is	Absent			Is	Noncompliant
3	3			Is	Absent	Is	Noncompliant

		Conditions		Conclusion	
Row ID	RP	Driver Seat Belt		Driver Seat Belt Compliance	
1	1	Is	Worn	Is	Compliant
2	2	Is	Not Worn	Is	Noncompliant

		Conditions		Conclusion	
Row ID	RP	Driver Hand Held Device		Driver Device Compliance	
1	1	Is	Not in Use	Is	Compliant
2	2	Is	In Use	Is	Noncompliant

Figure 5: Multiple Rule Family Tables for New Driver Compliance

Some people may refer to these Rule Family tables as being ***more*** normalized than those in Figure 4, but again that is not true. By now, you know that the Rule Family tables in Figure 5 are more decomposed than

© 2014 Knowledge Partners International, LLC All Rights Reserved
www.kpiusa.com ▪ information@kpiusa.com

those in Figure 4. Yet, both sets of Rule Family tables are in third normal form (or higher) because the decomposition has nothing to do with resolving a normalization error.

Again, to prove that there are many ways to **decompose** a normalized Rule Family into a set of smaller ones see Figure 6. A reason for this decomposition may be that all of the logic for Driver Compliance is governed by one business area. If so, representing it all in one Rule Family enables easier governance.

		Conditions				**Conclusion**	
Row ID	RP	Vehicle Decal Compliance		Driver Compliance		New Driver compliance	
1	1	Is	Compliant	Is	Compliant	Is	Compliant
2	2	Is	Noncompliant				Noncompliant
3	3			Is	Noncompliant	Is	Noncompliant

		Conditions				**Conclusion**	
Row ID	RP	Vehicle Front Decal		Vehicle Rear Decal		Vehicle Decal Compliance	
1	1	Is	Present	Is	Present	Is	Compliant
2	2	Is	Absent			Is	Noncompliant
3	3			Is	Absent	Is	Noncompliant

		Conditions				**Conclusion**	
Row ID	RP	Driver Seat Belt		Driver Hand Held Device		Driver Compliance	
1	1	Is	Worn	Is	Not in Use	Is	Compliant
2	2	Is	Not Worn	Is	Noncompliant	Is	Noncompliant

Figure 6: Another Set of Rule Families for New Driver Compliance

© 2014 Knowledge Partners International, LLC All Rights Reserved

In The Decision Model, there are perhaps more degrees of freedom for decomposing Rule Families than there are for decomposing data. That's because, while the conditions are the primary key (identifier for a Rule Family Table), it is the conclusion fact type around which conditions are grouped into one Rule Family Table. And, a decision modeler has the luxury of "manufacturing" a new conclusion fact type (i.e., an interim conclusion fact type) for any combination of condition columns. The data modeler does not necessarily have the freedom to create newly invented foreign keys in data structures.

In Figure 5, these manufactured conclusion types are: Vehicle Decal Compliance, Driver Seat Belt Compliance, and Driver Device Compliance. These don't exist in the Rule Family of Figure 4. A decision modeler can simply make them up.

Because of this freedom to make up new conclusions having meaning, a decision modeler becomes a true artist in deciding the optimum way to decompose a decision model.

Value of Normalization and Decomposition

The examples in this column started with ones that were already normalized based on the definition of three normal forms. This may not always be the case. However, in practice, you may not detect normalization errors (in data or in The Decision Model) until you populate the tables. That's because it is more difficult to understand the functional dependencies based on structure alone. Most often, normalization errors are detected when you notice redundancy in fully populated tables that seems unnecessary.

The value of normalization is that it reduces a set of data elements or business logic statements into smaller structures by removing all unnecessary redundancies. Therefore, normalization is mandatory and delivers the highest integrity in logic and data structures.

The value of general composition is not related to fixing normalization errors. Rather, it is related to other factors and allows for many variations.

Finding the Balance

In general, there are three approaches to creating decision models: top down, bottom up, and a combination of both.

The bottom up approach means starting first with Rule Family table headings, populate them, and see how they connect together to form a decision model. In the New Driver Compliance example, this may mean starting with a Rule Family for each bullet of the logic description. This results in Rule Families for New Driver Compliance (for the overall), Vehicle Decal Compliance, Driver Seat Belt Compliance, and Driver Hand Held

© 2014 Knowledge Partners International, LLC All Rights Reserved
www.kpiusa.com ▪ information@kpiusa.com

Device Compliance. From here, determine if there is a condition in one of these Rule Families which is a conclusion in another, creating and populating all supporting Rule Families. Finally, be sure all Rule Families are in third normal form.

The top down approach means guessing at a decomposed decision model structure first, populating it, and validating that every Rule Family table is in third normal form.

Top down is most often the best approach because you immediately begin to investigate options for decomposition even before populating Rule Family tables.

In both cases, since decomposition is an art, there is no single step-by-step procedure (as there is with normalization). However, here are some general thoughts which are summarized by Figure 7:

1. Start with the top level conclusion.
    i. For our example, this is New Driver Compliance.
2. Consider decomposition below it (logic branches) for each bullet or paragraph or section of business input or for each business concept (i.e., entity or object or subject) in the input documentation.
    i. For our example, the first composition is by business concept and these are Vehicle Compliance and Driver Compliance.
3. For each decomposition in the first level, consider further decomposition based on the highest level conclusion fact type.
    i. In the example, the highest level of conclusion fact type under Vehicle Compliance is Vehicle Decal Compliance. The highest level of conclusion fact types under Driver is Driver Seat Belt Compliance and Driver Hand Held Compliance.
4. Keep decomposing until all input is raw data.
    i. For the example, there is no need to further decompose Vehicle Decal Compliance into Front and Rear Compliance as these are raw data.
5. Re-evaluate your decomposition based on subsets of logic that are universal (hence, shared across decision models) or require customized logic for different business purposes (hence, have different views for the same conclusion fact type)
    i. For our example, Driver Seat Belt Compliance is universal across all states, but Vehicle Decal Compliance and Driver Hand Held Device Compliance have logic that varies by state.
    ii. Upon re-evaluation, there is still no need to further decompose Vehicle Decal Compliance into Front and Rear Compliance since these are raw data and are neither universal nor need customized view (customization will happen at Vehicle Decal Compliance)

© 2014 Knowledge Partners International, LLC All Rights Reserved
www.kpiusa.com ▪ information@kpiusa.com

6. Be sure all resulting Rule Families are in third normal form.

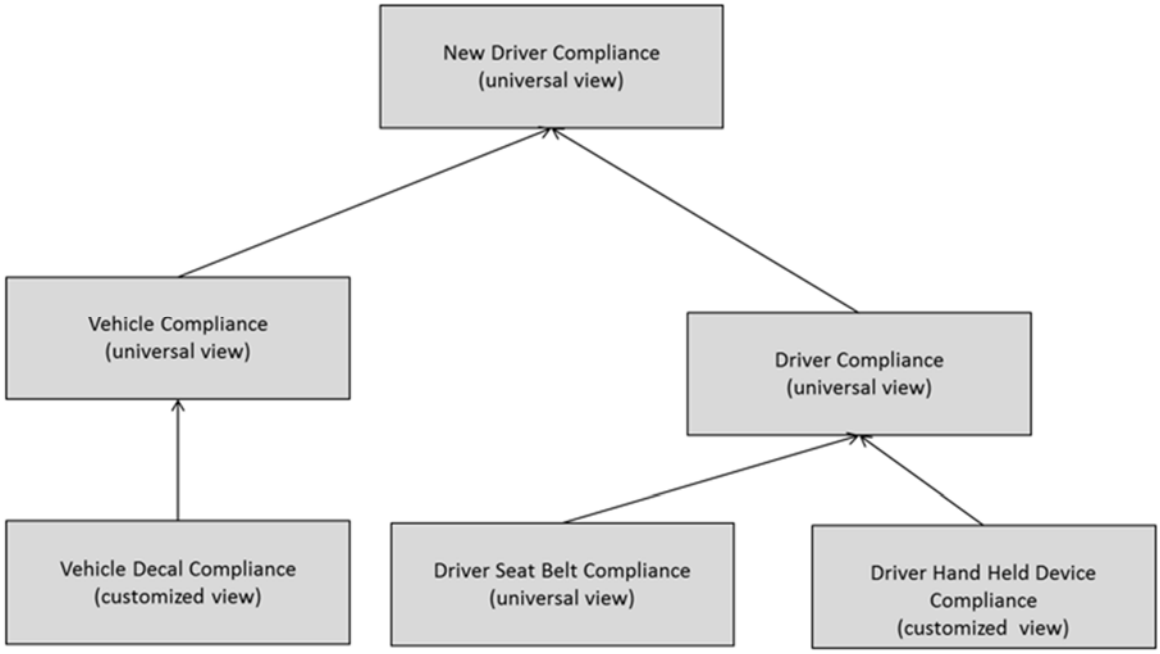

Figure 7: Sample Diagram to Understand Decomposition using OMG Decision Model and Notation (DMN)

In Figure 7, the New Driver Compliance Rule Family is a universal view [5] because its logic applies to all states, specifically the logic always results in a Compliant conclusion only if Vehicle Compliance is Compliant and Driver Compliance is Compliant, otherwise it always results in a Noncompliant conclusion. The Vehicle Compliance Rule Family is also a universal view because its logic applies to all states, specifically the logic always results in a Compliant conclusion only if Vehicle Decal Compliance is Compliant, otherwise it always results in a Noncompliant conclusion. The Driver Compliance Rule Family is a universal view because its logic applies to all states, specifically the logic always results in a Compliant conclusion only if both Driver Seat Belt Compliance is Compliant and Driver Hand Held Device is Compliant, otherwise it always results in a Noncompliant conclusion.

However, the Vehicle Decal Compliance Rule Family and Driver Hand Held Device Compliance Rule Family are customized views, perhaps one for each state, because each state mandates its own compliance when it comes to vehicle decals and hand held devices. Should there come a time when there is a federal law governing hand held devices, for example, the Rule Family for Driver Hand Held Device Compliance would become a universal view.

© 2014 Knowledge Partners International, LLC All Rights Reserved
www.kpiusa.com ▪ information@kpiusa.com

Wrap Up

There are many degrees of freedom within decision modeling, so there are many decision model structures for representing a given set of business logic requirements. Experiment with different decompositions based on governance, opportunity for reuse, and ease of business understanding.

What is the magical balance between the science and the art behind decision modeling? Every Rule Family should be at least in third normal form. Beyond that, it is up to the decision modeler to determine the optimum levels of decomposition to maximize usability.

---

[1] See http://www.modernanalyst.com/Resources/Articles/tabid/115/articleType/ArticleView/articleId/2649/New-Opportunities-for-Business-Analysts-Decision-Modeling-and-Normalization.aspx

[2] For examples of data normalization errors, see Chapter 7 of The Decision Model: A Business Logic Framework Linking Business and Technology, von Halle & Goldberg, © 2009 Auerbach Publications/Taylor & Francis, LLC.

[3] Stability here refers to the ability to service many requirements over time and still be correct and consistent.

[4] von Halle and Goldberg, page 286

[5] In this case, “universal view” refers to a view for one entire country and “customized view” refers to a view for specific states or provinces within that country. The scope of any decision model determines the meaning and boundary of “universal view” as used in this column.

© 2014 Knowledge Partners International, LLC All Rights Reserved
www.kpiusa.com ▪ information@kpiusa.com