The Teradata Travel and Hospitality Industry Data Model

Overview and Application

By Steve Hoberman, Steve Hoberman & Associates, LLC

SPONSORED BY

TERADATA®

Table of Contents

Executive Summary

In the Teradata white paper titled, *Leveraging the Industry Data Model*, I provided an overview to the Enterprise Data Model and the Teradata industry data models (iDM). In this white paper, I will provide detail about the Teradata® Travel and Hospitality Data Model (THDM). Specifically, this paper provides an overview of the Teradata THDM and a scenario illustrating how the THDM can be leveraged. The goal of this paper is to increase your awareness of how the THDM helps organizations like yours attain the big picture more quickly and accurately than building an EDM from scratch, thus permitting your organization to answer complex strategic and tactical business questions faster and more accurately.

"An industry data model is a prebuilt data model that captures how an organization in a particular industry works or should work."

TERADATA

Introduction

I anxiously entered Detroit International airport knowing that I booked a flight with an extremely tight connection with the goal of getting me in to Nova Scotia that same night for a class I was teaching the following day. A quick glance at the departures monitors confirmed my fear—my flight was cancelled! In despair, I went to the ticket counter of the major airline carrier with whom I was flying to explain the situation. Before I could finish my first sentence, the ticket agent smiled and said they automatically booked me on a flight leaving in about 15 minutes, and if I could get through security in time, my ticket would be waiting at the gate. I raced through security, caught my flight, and made it to Nova Scotia that night.

I was very impressed with this level of service. In the grand scheme of things, rebooking a customer is a minor transaction. However, knowing their frequent flyers, proactively identifying which customers are going to be missing connections, and then identifying alternate travel routes to get those customers to their destinations is no easy task. This requires knowledge of the *big picture*, that is, knowing how all of the concepts within the organization fit together, such as passenger, flight, and airport.

Knowing the big picture is critical to travel and hospitality organizations that need to maintain (and hopefully increase) profitability within the constraints of a tight economy. There is a lot of money at stake in getting the right information to the right people at the right time. This maxim can be achieved by having a single well-understood big picture of the organization. A single representation of passenger, for example, enables graceful growth of operational information and the building blocks for powerful business intelligence (BI) applications.

A well understood big picture of the organization needs to be captured and communicated in the form of a model. A model is a set of symbols and text used to make a complex landscape easier to grasp. Our content-rich world can overwhelm our senses and make it very challenging to focus only on the relevant information needed to make intelligent decisions. A complex *geographic* landscape is made understandable via a model called a map. A complex *information* landscape is made understandable via a data model. A data model uses symbols and text to help business leaders, developers, and analysts better understand a set of data elements and their corresponding business rules. In addition, every model has a defined scope. A map might be limited to New York City or represent the big picture in the form of a globe. Likewise, a data model can represent a specific functional area, such as supply chain, or it can represent the big picture, in the form of an enterprise data model (EDM).

TERADATA®

An EDM is a subject-oriented and integrated data model describing all of the data produced and consumed across an entire organization. Subject oriented means that the concepts on a data model fit together as the CEO sees the company, as opposed to how individual functional or department heads view the company. One person can play many roles including possibly being a passenger and an employee. If Bob is a passenger and an employee, his name and other pertinent information is represented once, instead of repeating Bob's information for each role he plays. Integration goes hand in hand with subject orientation and implies a single version of the truth along with a mapping back to the chaotic real world. For example, if a person's last name lives in ten applications within an organization, the integrated EDM would show Person Last Name only once, and in addition, capture the mapping back to these ten applications, such as the person's last name as a passenger and an employee.

There are resource and skill challenges with creating and maintaining an EDM, and, therefore, instead of reinventing the wheel, organizations are increasingly purchasing starter EDMs in the form of industry data models. An industry data model is a prebuilt data model that captures how an organization in a particular industry works or should work. Teradata Corporation offers ten industry data models (iDMs):

- Teradata Communication Data Model

- Teradata Financial Services Data Model

- Teradata Healthcare Data Model

- Teradata Life Sciences Data Model

- Teradata Manufacturing Data Model

- Teradata Media and Entertainment Data Model

- Teradata Retail Data Model

- Teradata Transportation and Logistics Data Model

- Teradata Travel and Hospitality Data Model

- Teradata Utilities Data Model

Teradata Travel and Hospitality Data Model Overview

The Teradata Travel and Hospitality Data Model is the big picture for a general travel and hospitality organization, containing more than 50 broad subject areas, such as Travel Reservation and Purchase, Customer Account Loyalty, Promotion and Travel Performance. I have studied industry models that were extremely generic, and, therefore, only contained a handful of generic entities, such as Party. These generic models appear elegant yet require extremely complex mappings to the real source system to produce any value. The THDM does contain a handful of these generic concepts (e.g., Party and Travel Transaction), yet these generic concepts are used to link more granular and concrete parts of the business together (e.g., a reservation and a purchase are both travel transactions) and even to link different iDMs together (e.g., an organization that operates both manufacturing and retail lines of business can recognize that Bob is a party playing the role of an employee on the manufacturing side and a party playing the role of consumer on the retail side). Due to the details provided in the THDM, the source system mapping becomes more manageable. The current version of the THDM is extremely robust, containing more than 1,100 entities and 4,000 attributes, but these numbers—and model features—are continuously updated through new releases.

The THDM provides a comprehensive view of the travel and hospitality business. This model addresses specific lines of business including airlines, cruise lines, passenger rail and bus services, rental car agencies, hotels, casinos and Internet Travel Booking services in addition to providing a holistic view of reservations, customers, and travel partners across the travel industry.

Teradata Professional Services consultants work directly with clients in the field and provide feedback for enhancements to future releases of the THDM. Recent releases have included a number of new features and functionality to include enhanced customer loyalty program tracking (including partner programs), maintenance and repair operations and (within Hospitality and Gaming) Patron Retention and Slot Optimization. Each iteration of the THDM results in THDM customers benefiting from these enhancements resulting from many THDM

TERADATA.

implementations, and changes in the industry, such as the introduction of RFID technology.

The THDM exists in an Erwin® Data Modeler file. ERwin Data Modeler is one of the more popular data modeling tools that supports reports for viewing and printing the models and their meta data. In addition, The THDM documentation includes both hard copy and PDF files of three books. These include a reference guide common across all iDMs, a reference guide specific to the THDM, and an appendix, which includes an extremely comprehensive set of business questions for the travel and hospitality business, a glossary, and all of the subject area, entity, and attribute definitions.

The THDM has a number of very important characteristics:

Operational
The THDM captures how a travel and hospitality company works instead of how a travel and hospitality company typically does reporting. In other words, the vast majority of the structures in the THDM capture the data elements and business rules that govern the day-to-day operation of the business. For example, the THDM captures the rule that a Casino can contain many game stations. In addition, there are sections of the THDM that have been added to facilitate business monitoring, such as the tracking of those VIP game players whose actual winnings exceeded their expected winnings. Using this same Casino and game station example, there is a subject area for capturing trends and analytical models, which can lead to better understanding the profitability at both the casino and game station levels.

Logical
An industry data model is a business data and rules representation for a specific set of business requirements. If a requirement is to capture promotion information, the industry data model would contain the data elements and business rules pertinent to the promotion. It is completely independent of both application and technology, built using the process of normalization. Normalization ensures all data elements are correctly assigned to entities based on their dependency on a primary key.

Extensible
The THDM contains the common information that companies share within an industry, and, therefore, it is meant to be a jump-start toward creating a complete enterprise data warehouse for a company. Most companies use the THDM as a starter model, and add new structures, remove existing structures, and enrich the provided definitions to make them more meaningful to their organization. Customizations are expected and encouraged.

Abstract
The THDM contains some amount of abstraction to maintain the third normal form (3NF) structure. Abstraction means combining like things together under generic terms, such as Location (Web Site, Store, and Kiosk) and Party (Individual, Organization, and Casino Player), to facilitate integration and to gracefully handle future requirements.

Global
The structures and terms on the THDM are designed for international use, and not just U.S.-based. For example, the term 'postal code' is chosen over 'zip code,' and 'territory' is used instead of 'state.' This facilitates communication on global projects and mappings back to global source systems, such as ERP systems including SAP R/3.

Standard
The THDM follows the International Standards Organization (ISO) 11179 metadata standard. A class word is the last part of a data element name that represents the high-level category in which the data element belongs. Examples of class words are name, code, identifier, date, quantity, and amount. For example, the class word for Person Last Name is 'Name'.

Digestible
The THDM is sectioned into subject areas. Subjects are neatly captured in separate views, and the use of color distinguishing each subject area makes it easier to digest the larger models. In addition, there are certain subject areas that are common across the industry data models, such as Party, Geography, and Promotion. These subject areas have a common core in each iDM, and then are extended where appropriate within each of the models.

TERADATA®

Travel and Hospitality Data Model Scenario

Port Jumper (PJ) is a medium-sized cruise company that departs exclusively from ports in the Northeast U.S. PJ has been continuously losing market share, and the CFO of PJ is at a loss to explain the specific reasons behind the declining market share other than to relate it to a flat economy. Without understanding the cause, it is difficult to come up with a turnaround plan. For example, should the focus be on reducing costs, introducing new products, or retaining profitable customers?

PJ has grown rapidly to meet the anticipated large market of retired baby boomers. Lots of siloed operational and reporting systems make it nearly impossible to answer any business questions that cross departments or business functions, including those of importance to the CFO. The CIO needs to ensure applications are built to answer the CFO's questions by building an enterprise data model.

The Approach

Jamie Jones is a highly-skilled data analyst in PJ's enterprise data management team, responsible for building PJ's EDM. She built four data models: white board conceptual data model (CDM), enterprise CDM, enterprise

industry data model (iDM), and an enterprise physical data model (PDM). The white board CDM was built without any reference to the THDM. The enterprise CDM was built leveraging the Teradata CDM that accompanies the Teradata THDM. The enterprise iDM was built using the THDM in four different roles, which are discussed later in this paper. The enterprise PDM was built based on the PJ enterprise iDM as well as the PDM portion of the Teradata THDM. Figure 1 summarizes each of these models, and the following sections will provide the details along with examples.

White Board Conceptual Data Model

Jamie organized a series of meetings with business analysts, functional analysts, and department managers with a goal of creating a single, high-level view of the organization. She met with groups of one to five individuals and built their view of the organization using whiteboards and flipcharts. For those individuals who preferred not to see data models, Jamie worked with them to jointly create a listing of key concepts and their definitions. As you might expect, the finished model had severe integration issues. Sets of entities were not related to each other, and there were many cases where the same concept had two or more definitions, and similar concepts had completely different names and rules. This is actually a very good thing because it documents the integration issues, and

Model	Purpose	Built all at once or incrementally
White board CDM	Captures the current understanding of the business on one piece of paper.	All at once
Enterprise CDM	Captures a proposed integrated view of the business on one piece of paper	Incremental
Enterprise iDM	Captures a cross-functional, objective, and detailed view of business data.	Incremental
Enterprise PDM	Captures a detailed view of the business data taking into account the constraints of the database management system and user queries.	Incremental

Figure 1. Types of EDMs.

TERADATA

A reservation is a formal agreement between PJ and an outside party to provide a cruise experience at a given date and price. [from the booking department manager]

A reservation is any interaction between PJ and a person or organization outside of PJ that might have an interest in doing business with us. [from the marketing department manager]

A reservation is any paid amount received from a person or organization for an anticipated cruise vacation. [from an accounting department representative]

Figure 2. Three different definitions of Reservation.

Figure 3. Subset of Teradata CDM.

acknowledging the problem is a prerequisite to solving the problem. Jamie called this initial model the white board conceptual data model (CDM) because most of it was created in partnership with the business groups standing at white boards and flipcharts. It represented each business area in its own terms.

The concept of Reservation will be used to illustrate the four different types of models in this section.

Reservation is just one of the 1,100 entities on the THDM (albeit an important entity), and it exists in the Travel Reservation Subject Area. There were three different definitions of Reservation identified in the white board CDM, as shown in Figure 2.

This white board CDM had more than 200 entities. It was built all at once using a top down approach. A top down approach is one where the model is built purely from the business perspective and not from an existing systems perspective.

Enterprise Conceptual Data Model

The THDM comes with a conceptual data model (CDM) that contains more than 50 key concepts and their relationships for the travel and hospitality industry. It was built by including at least one major entity from each subject area and then generalizing the rules among these
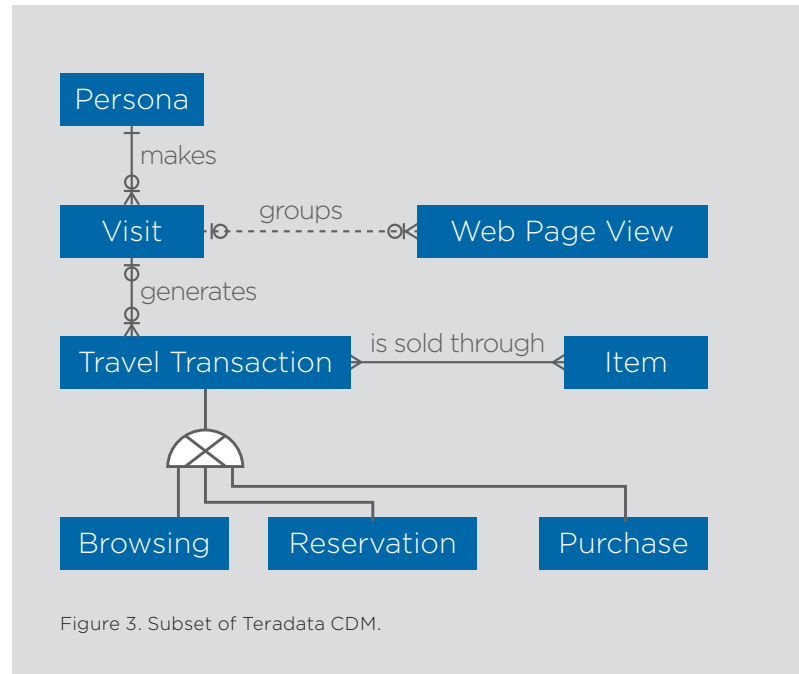
remaining entities. For example, more than 10 entities, including Reservation, Reservation Status, Lodging Reservation, and Dining Reservation, are represented by just the single Reservation entity on the Teradata CDM. Figure 3 contains a subset of the Teradata CDM.

The semi-circle with an 'X' in the middle is a subtype symbol. It identifies a "parent" entity (in this case Travel Transaction) called a supertype, as well as those entities sharing common data elements and relationships (in this case Browsing, Reservation, and Purchase) called the subtypes. This model subset contains the following rules:

- Each Persona can make many Visits.
- Each Visit must be made by a single Persona.
- Each Visit may group many Web Page Views.
- Each Web Page View may be tied to a single Visit.
- Each Visit may generate many Travel Transactions.
- Each Travel Transaction may be generated by a single Visit.
- Each Travel Transaction may lead to the sale of many Items.
- Each Item is sold through many Travel Transactions.
- Each TTravel Transaction can be either for Browsing, Reservation, or Purchase.

TERADATA.

- Browsing is a Travel Transaction.
- Reservation is a Travel Transaction.
- Purchase is a Travel Transaction.

The Teradata CDM allows an organization to achieve a high-level big picture of the organization without getting overwhelmed by jumping straight into a complex logical design. Jamie took a first pass at fitting the white board CDM into the Teradata CDM.

After spending time speculating how the pieces might fit together, she organized a second series of meetings. These meetings took place in groups of 10 to 15 individuals, and Jamie purposely invited people with very different views about the same concepts. She showed them the CDM retrofitted with each of their views and encouraged open communication so that when the meeting was over, there was either agreement on the model or issues that needed to be reconciled.

Figure 4 contains the portion of the Enterprise CDM after the terminology and definitions surrounding the term
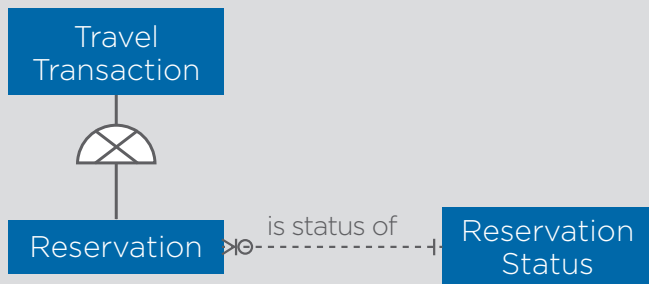

Figure 4. Subset of Enterprise CDM.

'Reservation' were resolved.

This model subset contains these rules:

- Each Travel Transaction can be a Reservation.
- Reservation is a Travel Transaction.
- Each Reservation Status can be a status to many Reservations.
- Each Reservation may have a single Reservation Status.

Travel, Transaction, and Reservation were already available in the Teradata CDM, each containing robust definitions. Although Reservation Status did not appear on the Teradata CDM, it does appear as a logical entity in the Travel Reservation subject area. Jamie decided it would help communicate how pending reservations and booked reservations relate if she added this entity into the CDM.

The definition for TRAVEL TRANSACTION in the Teradata CDM is:
An entity that represents a travel transaction which can contain the Browsing, Reservation, and Purchase of Travel Products. Our current convention is to create a new transaction for each Travel Trip.

The definition for RESERVATION in the Teradata CDM is:
The sub-type of the Travel Transaction entity that represents the domain of requests for a specific itinerary associated with a Travel Transaction. Examples: Hotel Reservation, Airline Reservation, and Rental Car Reservation.

The definition for RESERVATION STATUS in the Teradata CDM is:
The domain of status types for a specific itinerary item associated with booked Reservations. Status conditions include: Open, Closed, Cancelled.

The booking department manager's definition for reservation (See Figure 2) most closely matched the definition for Reservation, the marketing department manager's definition most closely matched the more generic concept of Travel Transaction, and she was ok with using this term with the caveat that the Teradata definition will need to be expanded to include more details, and the accounting department representative's definition most closely matched one of the states a reservation goes through and, therefore, including Reservation Status on the model satisfied him. In each of these cases, the definitions were expanded to include examples specific to PJ.

Jamie's CDM based on the Teradata CDM was regarded as a large success within business and IT circles. Jamie credits the success to first trying to understand the organization and then leveraging the Teradata model. Even business folks with very strong viewpoints found it easier to adopt the Teradata HDM terminology rather than get into a win/lose debate with colleagues from different departments. And now, on to the industry data model.

TERADATA

Enterprise Industry Data Model

As you might expect, the enterprise industry data model (iDM) required more effort than the prior two models. It had more detail and required the most discussions to resolve the integration issues.

Note that some of the integration issues remained unresolved yet well documented. Version one of the Enterprise iDM for PJ contained more than 900 entities and 3,000 data elements. It was built using a hybrid approach. A hybrid approach means it was built from both a top down and bottom up perspective. Top down is driven from the business requirements, which takes the form of the Enterprise CDM, and bottom up means start with the existing system's environment.

THDM Roles

An industry data model can play up to four different roles within an organization: blueprint, template, encyclopedia, and invisible. These are described below in order of decreasing reliance on the THDM (e.g., blueprint requires the most reliance on the THDM and invisible the least). The degree of reliance is determined by available modeling resources and knowledge of a particular business process.

Blueprint (The industry data model is *the* model)

The THDM contains the concept of an analytical model within the Model Score and Forecast subject area. The definition of analytical model is: "Describes a process used to predict, cluster, or classify information. Typically used in data mining and knowledge discovery. Examples: Booking Forecast, MRO Forecast, Customer Scoring, and Segmentation, a model that describes the propensity of a customer to engage in a particular activity." Analytical model is a concept that the organization has not even considered relevant, yet after understanding its potential value of predicting future market share and profitability, they decided to add it to their EDM. This involved adding more than a dozen new entities to their EDM exactly as they appeared in the THDM, including the actual analytical model entity.

Template (The industry data model is an integration point)

The THDM concept of Reservation becomes an important integration point for the company. Each of the Reservation data elements from the source systems was mapped into THDM data elements. A sample mapping appears in Figure 5.

Note, this mapping is overly simplified, as usually there can be complex transformation rules, as well as other types of meta data that need to be reconciled, such as format, granularity, and nullability. This mapping does illustrate the usefulness of subtyping, as the Reservation Booking Date is really a Transaction_Date (recall that Reservation is a subtype of Travel Transaction).

Many integration battles are quickly defused using the THDM, because instead of win/lose definition debates among business areas, it becomes a mapping exercise where both parties agree on a single, external, unbiased view.

Encyclopedia (The industry data model is referenced where needed)

There is a need within the organization to understand products better and relate these features to the actual travel transactions. The THDM provides a comprehensive data model for Item (product) and also includes, in detail, how Item relates to Travel Transaction. Jamie researched

Source			THDM	
Source system	Table or file	Data element	Entity	Data element
XYZ	BOOKING	Effective_Date		Transaction_Date
ABC	RESERVATION	Begin_Date	TRAVEL TRANSACTION	Transaction_Date
X3000	APPOINTMENT	Start_Date		Transaction_Date

Figure 5. Data element sample mapping.

TERADATA.

this area in the THDM and was able to understand the data and rules behind the model so she could add these concepts to the existing EDM. In some cases, terms, rules, and definitions from the THDM needed to be changed to fit the existing EDM. David Schoeff, Teradata Principal Consultant, compares this approach with how someone would use an encyclopedia. According to David, "There can be a substantial amount of modeling needed to build an organization's EDM, and the iDMs can serve as a valuable reference to save some modeling time and reduce risk by ensuring all concepts are present in the model."

Invisible (The industry data model is not consulted)
The THDM is not used at all. The Promotion area is extremely well modeled within PJ and has been rigorously maintained for the past five years. For this area, the THDM was not consulted at all and instead PJ's existing promotion structures were connected to the relevant THDM structures.

Enterprise Physical Data Model

The Enterprise PDM was based on the PDM portion of the THDM, and optimized incrementally on a project-by-project basis. An in-depth business questions analysis was performed, and sets of business questions were bundled into project deliverables. Jamie found it challenging to extract questions from the business folks. Luckily, the THDM came with hundreds of business questions, and she used this list as a brainstorming technique with the business to agree on a set of common questions. In fact, there are more than a dozen business questions about Web bookings that became the scope for an entire reporting solution. An example of one of these questions is: "Calculate the percentage of bookings on the Web versus other booking channels."

Smooth Sailing Ahead

All of PJ's future operational and business intelligence applications relied on the EDM as a starting point for design. There was continuous reconciliation with the enterprise model. So each application starts with the EDM, and then contributes new ideas back to the EDM. This keeps the EDM up-to-date and continuously valuable. Knowing the big picture saves design time and allows for each new application to fit together cleanly with existing applications. The THDM proves to be indispensable in creating this big picture.

Conclusion

The Teradata Travel and Hospitality Data Model saves organizations substantial amounts of time and money by providing a detailed and well-proven data model as a foundation for creating their own enterprise data model. This model keeps up with the industry, as there is a new release of the THDM every year. In addition, the THDM can be easily extended as the business grows, and it provides the organization with a common understanding of business terms.

About the Author

Steve Hoberman is a world-recognized innovator and thought leader in the field of data modeling. He has worked as a business intelligence and data management practitioner and trainer since 1990, and is a popular presenter at industry conferences, both nationally and internationally. Steve is a columnist and frequent contributor to industry publications, as well as the author of *Data Modeler's Workbench* and *Data Modeling Made Simple*. He is the founder of the Design Challenges group and inventor of the Data Model Scorecard.™ He can be reached at **me@stevehoberman.com**.