

# DATA VAULT CDVDM

Certified Data Vault Data Modeler Course

---

Sydney Australia

December 3-5 2012

In cooperation with



---

GENESEE ACADEMY, LLC

Course Description and Outline

# DATA VAULT CDVDM

---

## Certified Data Vault Data Modeler Course

### Forward

Thank you for your interest in the Data Vault certification course. Your successful completion of this course will result in a designation of Certified Data Vault Data Modeler – CDVDM.

Data Vault modeling is rapidly becoming the standard modeling approach for the data warehouse. Over 600 major international companies are applying data vault modeling techniques today.

The data vault principles are specifically well suited for data warehousing programs and can provide the organization with some very compelling benefits. These include auditability, agility, adaptability, alignment with the business, and support for operational data warehousing initiatives.

To gain these benefits the organization will need to commit to both EDW program level factors as well as specific data vault modeling patterns, rules and methods. This course will guide you through these modeling patterns, rules and methods to prepare you for active data vault modeling in your organization.

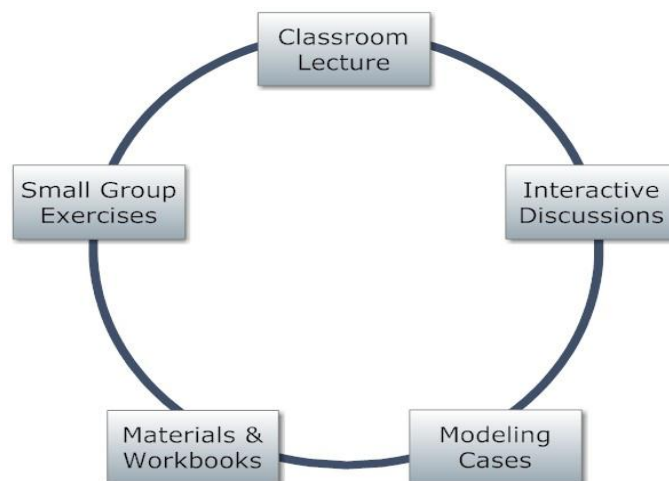
We look forward to seeing you in the classroom.

## The CDVDM Course

Welcome to the Certified Data Vault Data Modeler - CDVDM seminar! This course will guide you through the Data Vault modeling approach from modeling constructs and patterns to applying data vault principles in your DWBI program. This course covers also loading paradigms, architectures, and how to develop an effective overall data vault data warehouse program. Since data vault may be new for many of you, this course also includes a summary of the benefits of using data vault techniques.

## CDVDM Course Structure

The CDVDM course is a three (3) day intensive classroom based course. The course consists of five core components plus the certification exam. The classroom time is 40% Classroom Lecture, 30% Small Group Exercises and 30% Interactive Discussions.



The course materials consist of a 120 page course book that follows the classroom lectures and several modeling case books for the small group sessions. Class interactive discussions are based on these cases plus several workbooks.

## Target Audience

The target audience for this course includes data warehousing and business intelligence professionals, data modelers, data architects, model managers, data warehouse DBAs, and ETL professionals. Because data vault modeling concepts are closely aligned with the business aspects of DWBI programs our target audience also includes program managers, business analysts, information modelers, information architects, BICC professionals, and data scientists.

## Course Outline

### Day One

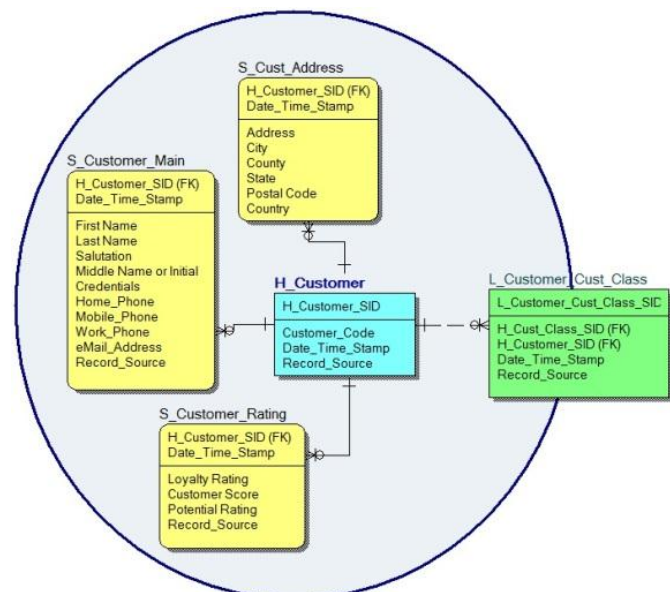
- 0830 Welcome & Introductions
- 0900 Data Vault Overview
- 0945 Core Constructs: Hub
- 1000 Core Constructs: Link
- 1015 Core Constructs: Satellite
- 1030 Break
- 1045 Data Vault Modeling
- 1115 First Small Group Case Exercise
- 1200 Class Discussion
- 1230 Lunch
- 0130 Defining the Business Key
- 0145 Link Design with the Coffee Example
- 0215 Second Small Group Case Exercise
- 0300 Break
- 0315 Class Discussion
- 0345 Data Modeling for the Data Warehouse
- 0400 The 3NF Data Warehouse
- 0415 The Dimensional Data Warehouse
- 0430 New EDW Requirements
- 0500 Wrap up for the Day

### Day Two

- 0830 Q&A from Day One
- 0845 Why Data Vault for DW
- 0915 Satellite Design
- 0945 Third Small Group Case Exercise
- 1030 Break
- 1045 Class Discussion
- 1100 Continue Third Case
- 1145 Class Discussion & Groups Present
- 1230 Lunch
- 0130 Unit of Work with Discussion
- 0200 Bridge Tables
- 0215 PIT and Grouping Links
- 0230 Class Workbook
- 0300 Break
- 0315 Reference Tables
- 0345 Raw & BDV Layers
- 0400 Aligning Business Keys
- 0415 Data Vault DW Architecture
- 0445 Loading Patterns
- 0500 Wrap up for the Day

### Day Three

- 0830 Q&A from Day Two
- 0900 Working with Data Vault
- 0930 Data Vault & Agility
- 1000 Data Vault & Big Data
- 1015 Business Key Topics
- 1030 Break
- 1045 Class Workbook
- 1115 Class Discussion
- 1130 Top 10 List
- 1145 Recap for Exam
- 1230 Lunch
- 0130 Recap & Discussions
- 0200 CDVDM EXAM
- 0500 Drinks



## Requirements and Pre-Requisites

This course does not require a computer or other device. Students should be prepared to take notes. All materials, work books, case books and exercises will be provided.

There are no pre-requisites for this course. It is helpful if you bring any specific topics that you are interested in concerning data vault. It can also be useful if you have the time for web research on the topic prior to class.

## Registration and Fees

Registration fee is 2950 AUD for this course.

This fee includes all materials and meals during the training.

Please contact [Jackie Harper](#) at MIP +61 2 92 600 700

Or click here to [Register](#).

## Contact Genesee Academy

For any other questions, please contact Genesee Academy

[Kristin@GeneseeAcademy.com](mailto:Kristin@GeneseeAcademy.com)

+1 303 526 0340

Thank you again for your interest!

We look forward to seeing you in the classroom.

*Please see next page for a quick Data Vault primer*

## The Data Vault Fundamentals

The data vault consists of three core components, the **Hub**, **Link** and **Satellite**. Above all other DV Program rules and factors, the commitment to the consistency and integrity of these constructs is paramount to a successful DV Program.

The **Hub** represents a Core Business Concept such as Customer, Vendor, Sale or Product. The Hub table is formed around the Business Key of this concept and is established the first time a new instance of that business key is introduced to the EDW. It may require a multiple part key to assure an enterprise wide unique key however the cardinality of the Hub must be 1:1 with a single instance of the business concept. The Hub contains no descriptive information and contains no FKs. The Hub consists of the business key only, with a warehouse machine sequence id, a load date/time stamp and a record source.



Fig. 1 Hub

A **Link** represents a natural business relationships between business keys and is established the first time this new unique association is presented to the EDW. It can represent an association between several Hubs and sometimes other Links. It does maintain a 1:1 relationship with the unique and specific business defined association between that set of keys. Just like the Hub, it contains no descriptive information. The Link consists of the sequence ids from the Hubs and Links that it is relating only, with a warehouse machine sequence id, a load date/time stamp and a record source.



Fig. 2 Link

Notice the similarity between the Hub and the Link. Both represent the first time a core business concept (Hub) or natural business relationship (Link) is introduced to the DW.

The **Satellite** contains the descriptive information (context) for a business key. There can be several Satellites used to describe a single business key (or association of keys) however a Satellite can only describe one key (Hub or a Link). There is a good amount of flexibility afforded the modelers in how they design and build Satellites. Common approaches include using the subject area, rate of change, source system, or type of data to split out context and design the Satellites. The Satellite is keyed by the sequence id from the Hub or Link to which it is attached plus the date/time stamp to form a two part key.

Note that the Satellite then is the only construct that manages time slice data (data warehouse historical tracking of values over time).

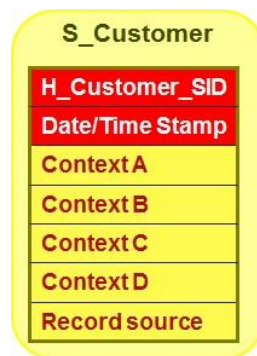


Fig. 3 Satellite

A Satellite does not have a Sequence ID of its own and in fact cannot have a different key than the Hub or Link sequence to which it is attached. Further, a Satellite does not have any foreign key constraints (no snow-flaking, branching or bridging).

These three constructs are the building blocks for the DV EDW. Together they can be used to represent all integrated data from the organization. The Hubs are the business keys, the Links represent all relationships and the Satellites provide all the context and changes over time.

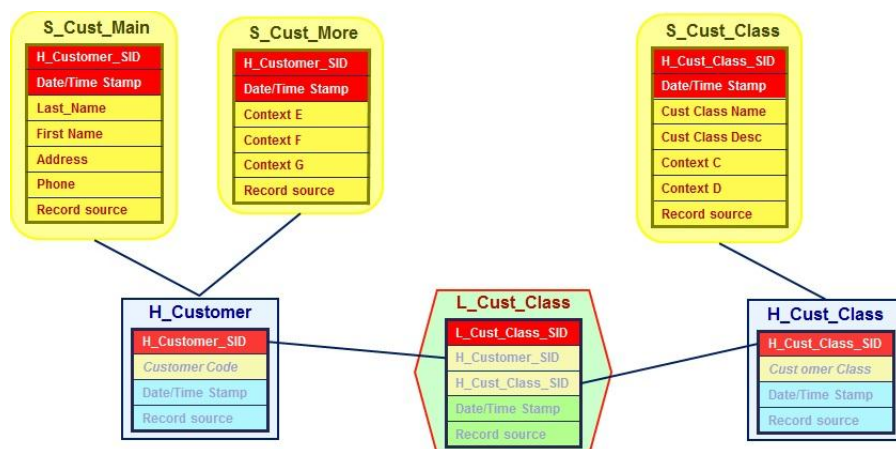


Fig. 4 Data Vault Model