

# Data Vault Course

## Data Vault 2.0 Boot Camp and Certification

---

### Course Description

This class covers what you need to know as a practitioner in the world of Data Warehousing and Business Intelligence. This is a foundation course. This is a 3-day (in person) course that covers end-to-end best practices.

### Intended Audience

IT Professionals desiring to become DV2.0 certified. This class is a prerequisite for anyone wishing to achieve DV2.0 Certified Practitioner status. CDVP2 is now the only sanctioned and accepted form of Data Vault certification.

### Topics Covered

- Architecture – Including NoSQL, Big Data, Hybrid Systems, and Relational stores
- Methodology – Including CMMI, Six Sigma, Optimization, Automation, and Generation
- Implementation – Including Performance and Tuning, Set Logic, ELT vs ETL, Parallelism
- Modeling – Including replacing of surrogates with Hash Keys, data layout, data co-location

This class takes you through the why/what/how of Data Vault 2.0. It includes the coverage of the business justifications, then follows with the technical descriptions of the architecture, implementation, methodology, and modeling. Included in the topics are reaching agility, practicing Six Sigma, measuring and optimizing at CMMI level 5, the KPA's and KPI's of Data Warehousing, and more.

We also discuss the use of Hadoop and NoSQL platforms alongside the relational world. The objective is to enrich your understanding of how and when to apply Big Data Solutions. The course finishes with descriptions on ETL and ELT design time paradigms, including templates, best practices and working SQL.

- How and when to apply KPA's and KPI's for measurement and optimization in business intelligence programs
- What the impacts are of CMMI Level 5 optimization on data warehousing methodologies
- The best practices for automation and generation of ETL / ELT that is highly scalable
- How Data should be laid out in MPP formats
- What Co-Location is, and data re-distribution in MPP environments
- How to deal with joins to unstructured and semi-structured data sets
- What the difference is between Schema on Read and Schema on Write
- How to seamlessly integrate BIG DATA solutions to existing relational database systems
- How to model your data warehouse using Data Vault 2.0 Modeling Techniques

These topics, and much much more are covered in the class. The end objective is to enhance your skill set so that you are a qualified practitioner/expert in Enterprise Business Intelligence projects. Whether it's data acquisition, data provisioning, change management, or project management you now have the skills to deal with whatever is thrown your way.

We will be discussing DV2 on Hive / Hadoop, the benefits, pros, and cons, some suggestions on how to build it and leverage it properly. We will be talking about Satellites on HDFS, Hubs & Links on Hive. We'll discuss data modeling implications, and using SERDe definitions at query time.

#### Overview Agenda:

##### Day 1:

- What is Data Vault 2.0?
- DV1.0 Compared to DV2.0
- Issues Faced Today
- Data Vault 2.0 In Business
- Managed Self-Service BI
- Agile Delivery and DV2.0 Methodology
- Agile Requirements Gathering
- Technical Numbering
- Roles and Releases

##### Day 2:

- Data Vault 2.0 Architecture Components
- Business Processes to Business Keys
- Hashing and Sequences
- Common Terminology
- Multi-Tenancy (Introduction)
- Core DV Structures
- Dependent Child (Data Modeling)
- Driving Key
- Link to Link Denormalization

##### Day 3:

- Link Unit of Work
- Applying Links
- Non-Historized Link
- Satellites in Depth
- Record Source Tracking
- Applying Set Logic
- ELT / ETL Performance Tuning
- Loading Architecture
- Column Comparison and Hash Differences
- Zero Key Concept
- Staging Loads
- DV2.0 Loading Templates / Standards

- End-Dating Satellite Templates
- Real-Time Loading
- Point-in-Time and Bridge Tables
- Dealing with Corrupted Data

The practitioner certification test is ***included*** as a part of the course. The test is offered privately, usually one (1) week after the class completes in a proctored setting. You will need to contact your instructor for further details and to set up a time and place to take the exam. We highly recommend studying, as the test is ***not easy***. It is, however, closed book, closed note.