

---

# Modeling the Agile Data Warehouse with Data Vault

Hans Hultgren

---

## Contents

<b>FORWARD</b> .....	<b>4</b>
<b>ABOUT THE AUTHOR</b> .....	<b>7</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>8</b>
<b>CHAPTER 1 DATA VAULT DEFINED</b> .....	<b>19</b>
1.1 DATA VAULT IS A DATA MODELING APPROACH.....	20
1.2 THE UNDERLYING PREMISE OF DATA VAULT.....	21
1.3 SINGLE GRAIN OF KEY .....	22
1.4 CONCEPT CONSTELLATIONS.....	24
1.5 WHERE DATA VAULT ENSEMBLE FITS IN .....	25
1.6 DATA VAULT CORE CONSTRUCTS.....	25
1.7 DATA VAULT METHODOLOGY.....	27
1.8 WHY DATA VAULT .....	28
1.9 WHEN TO USE DATA VAULT MODELING .....	29
<b>CHAPTER 2 ENSEMBLE MODELING</b> .....	<b>33</b>
2.1 ENSEMBLE MODELING .....	34
2.2 ENSEMBLE DEFINED .....	34
2.3 DATA VAULT ENSEMBLE.....	35
2.4 APPLYING THE DATA VAULT ENSEMBLE.....	36
2.5 ENTITY VERSUS DATA VAULT ENSEMBLE.....	37
<b>CHAPTER 3 DATA MODELING</b> .....	<b>39</b>
3.1 MODELING AND DESIGN.....	40
3.2 DATA MODELING APPROACHES.....	40
3.3 THE NORMALIZED DATABASE & 3 <sup>RD</sup> NORMAL FORM .....	41
3.4 DIMENSIONAL MODELING AND THE STAR SCHEMA.....	45
3.5 COMMON FACTORS AND DATA VAULT MODELING.....	47
3.6 INTEGRATION & THE ENTERPRISE DATA WAREHOUSE.....	49
3.7 APPLYING MODELING TECHNIQUES TO EDW .....	49
3.8 OPTIMAL MODELING FOR DIFFERENT LAYERS .....	50
3.9 ADDRESSING THE ISSUES WE EXPERIENCE TODAY.....	52
3.10 MODELING APPROACH SUMMARY.....	54
<b>CHAPTER 4 MODELING WITH DATA VAULT</b> .....	<b>57</b>
4.1 DATA VAULT MODELING .....	58
4.2 THE BUSINESS KEY .....	60
4.3 THE MODELING PROCESS.....	62
4.4 A DATA VAULT MODEL .....	65
<b>CHAPTER 5 COLORS OF DATA VAULT</b> .....	<b>67</b>
5.1 DATA MODELING COLOR CATEGORY ANALYSIS.....	68
5.2 COLORS OF 3 <sup>RD</sup> NORMAL FORM .....	72
5.3 COLORS OF STAR SCHEMA .....	73
5.4 COLORS ANALYSIS FOR DATA VAULT .....	74
5.5 COMPARING COLORS FOR THE EDW .....	76
5.6 SUMMARY OF THE COLORS.....	80

<b>CHAPTER 6</b>	<b>CORE CONSTRUCTS: HUB</b> .....	<b>83</b>
6.1	THE HUB .....	84
6.2	APPLYING HUBS IN DATA VAULT MODELING.....	84
6.3	IDENTIFYING AND MODELING HUBS.....	86
6.4	HUB ATTRIBUTES .....	87
<b>CHAPTER 7</b>	<b>CORE CONSTRUCTS: LINK</b> .....	<b>93</b>
7.1	THE LINK .....	94
7.2	APPLYING LINKS IN DATA VAULT MODELING.....	95
7.3	RELATIONSHIP INTEGRITY / CONTINUITY.....	95
7.4	LINKS AND CARDINALITY.....	96
7.5	COMMON APPLICATIONS OF A LINK .....	99
7.6	RELATIVE CARDINALITY OF N-WAY RELATIONSHIPS.....	102
7.7	IDENTIFYING AND MODELING LINKS.....	105
7.8	MODELING RECURSIVE RELATIONSHIPS.....	106
7.9	LINK ATTRIBUTES .....	107
<b>CHAPTER 8</b>	<b>CORE CONSTRUCTS: SATELLITE</b> .....	<b>113</b>
8.1	THE SATELLITE .....	114
8.2	APPLYING SATELLITES IN DATA VAULT MODELING.....	115
8.3	DESIGNING AND MODELING SATELLITES.....	116
8.4	SATELLITE ATTRIBUTES .....	117
<b>CHAPTER 9</b>	<b>THINKING DIFFERENTLY WITH DATA VAULT</b> .....	<b>123</b>
9.1	THINKING DIFFERENTLY WITH DATA VAULT .....	124
9.2	CHANGING MODELING PARADIGMS.....	125
9.3	IMPLICATIONS OF THE NEW PARADIGMS .....	126
9.4	SATELLITE KEYS .....	127
9.5	LINKS AND CARDINALITY.....	128
9.6	FOREIGN KEYS IN A SATELLITE.....	132
9.7	SUMMARY OF THINKING DIFFERENTLY.....	133
<b>CHAPTER 10</b>	<b>MODELING A CORE BUSINESS CONCEPT</b> .....	<b>135</b>
10.1	THE CORE BUSINESS CONCEPT.....	136
10.2	ABOUT THE ENTERPRISE VIEW .....	137
10.3	WHERE TO FIND THE ENTERPRISE VIEW.....	139
10.4	MODELING THE CORE BUSINESS CONCEPT .....	141
10.5	CORE CONCEPT SUPER-TYPES AND SUB-TYPES .....	143
<b>CHAPTER 11</b>	<b>ENTERPRISE DATA WAREHOUSING</b> .....	<b>147</b>
11.1	INTRODUCTION TO ENTERPRISE DATA WAREHOUSING.....	148
11.2	DEFINITION AND CHARACTERISTICS.....	149
<b>CHAPTER 12</b>	<b>DWBI</b> .....	<b>157</b>
12.1	DWBI .....	158
12.2	BI IN THE ORGANIZATION .....	159
12.3	DWBI VISUALIZATION.....	164
12.4	SUMMARY .....	165
<b>CHAPTER 13</b>	<b>DATA INTEGRATION</b> .....	<b>167</b>
13.1	DATA INTEGRATION.....	168

13.2	THE SEMANTIC GAP .....	169
13.3	INTEGRATION IS NOT BASED ON DATA .....	170
13.4	ETLAND DATA VIRTUALIZATION.....	171
13.5	ETL MECHANICS AND THE PIPELINE.....	172
<b>CHAPTER 14</b>	<b>MOVING FROM LOGICAL TO PHYSICAL DESIGN .....</b>	<b>175</b>
14.1	MOVING FROM LOGICAL TO PHYSICAL .....	176
14.2	THE REALITY OF SOURCES.....	176
14.3	ABSTRACTIONS AND GENERIC FORMS .....	176
14.4	SUMMARY OF GENERIC CONSTRUCTS.....	185
<b>CHAPTER 15</b>	<b>BIG DATA.....</b>	<b>189</b>
15.1	BIG DATA .....	190
15.2	HUGE DATA VOLUMES .....	191
15.3	N-STRUCTURED & VERY COMPLEX DATA .....	191
15.4	STREAMING AND SHAPE-SHIFTING DATA.....	192
15.5	BIG DATA AND DATA VAULT .....	192
<b>CHAPTER 16</b>	<b>MASTER DATA MANAGEMENT .....</b>	<b>197</b>
16.1	MASTER DATA MANAGEMENT .....	198
16.2	MDM AND DATA WAREHOUSING.....	199
16.3	THE DATA VAULT.....	200
16.4	MDM INITIATIVES .....	201
16.5	MDM AND THE DATA VAULT.....	203
<b>CHAPTER 17</b>	<b>DATA WAREHOUSE ARCHITECTURE.....</b>	<b>207</b>
17.1	DATA WAREHOUSE ARCHITECTURE.....	208
17.2	ARCHITECTURAL STATES OF DATA .....	209
17.3	ABOUT THE LAYERS .....	211
17.4	ARCHITECTURAL LAYERS AND SEMANTICS.....	212
17.5	DATA VAULT ARCHITECTURE .....	216
17.6	THE RAW VERSUS BDW LAYERS .....	217
17.7	BELLY UP TO THE BAR .....	218
17.8	FROM UPSTREAM TO DOWNSTREAM .....	221
17.9	ARCHITECTING THE FLOW OF DATA.....	221
17.10	PERSISTING DATA IN THE ARCHITECTURE.....	224
17.11	REFERENCE MODELS AND TAXONOMIES.....	227
17.12	VIEWS, DATA VIRTUALIZATION, AND MORE .....	229
17.13	ARCHITECTURE SUMMARY .....	231
<b>CHAPTER 18</b>	<b>KEY ALIGNMENT .....</b>	<b>233</b>
18.1	WHY KEY ALIGNMENT?.....	234
18.2	LEVEL OF EXISTING ALIGNMENT .....	237
18.3	KEY ALIGNMENT IN THE DATA WAREHOUSE.....	237
18.4	ALIGNING THE KEYS IN THE EDW ARCHITECTURE.....	239
<b>CHAPTER 19</b>	<b>CODES &amp; REFERENCE TABLES .....</b>	<b>243</b>
19.1	CODES AND REFERENCE TABLES.....	244
19.2	CODES AND REFERENCE TABLES DEFINED .....	246
19.3	USING REFERENCE TABLES.....	247

19.4	REFERENCE TABLES AND DATA WAREHOUSING .....	248
19.5	REFERENCE TABLES IN THE DATA VAULT EDW .....	249
<b>CHAPTER 20 INFORMATION MODELS, TAXONOMIES &amp; INDUSTRY REFERENCE MODELS .....</b>		<b>253</b>
20.1	CHAPTER FORWARD.....	254
20.2	SCOPE OF THE VARIOUS MODELS.....	255
20.3	TAXONOMIES AND THE EDW BUCKETS OF DATA .....	256
20.4	DOMAIN VALUES AND REFERENCE MODELS .....	260
20.5	TAXONOMY MODELS IN THE EDW .....	263
20.6	INDUSTRY REFERENCE MODELS .....	264
20.7	INFORMATION MODELS.....	268
20.8	MORE MODELS .....	269
<b>CHAPTER 21 SATELLITE DESIGN TOPICS .....</b>		<b>273</b>
21.1	SATELLITE KEYS AND GRAIN .....	274
21.2	SATELLITE GRAIN MODELING ISSUES.....	275
21.3	MULTI-VALUED SATELLITE.....	276
21.4	TOGGLING BETWEEN 3NF AND DATA VAULT.....	280
21.5	CONCEPT CONSTELLATIONS .....	281
21.6	SATELLITE DESIGN AND THE NUMBER OF SATELLITES .....	283
21.7	SATELLITE DESIGN VARIATIONS .....	286
<b>CHAPTER 22 LINK DESIGN TOPICS .....</b>		<b>289</b>
22.1	LINK DESIGN TOPICS OVERVIEW .....	290
22.2	BUSINESS DRIVEN VERSUS SOURCE DRIVEN .....	291
22.3	AN EVENT ALWAYS REQUIRES A HUB .....	292
22.4	RELATIONSHIP CONSISTENCY & UNIT OF WORK .....	296
22.5	WHAT DRIVES THE UNIT OF WORK .....	297
22.6	RELATIVE CARDINALITY.....	298
22.7	LINK ANALYSIS AND DESIGN PROCESS .....	300
<b>CHAPTER 23 HUB DESIGN TOPICS .....</b>		<b>305</b>
23.1	ALTERNATE IDS FOR BUSINESS KEYS .....	306
23.2	DUPLICATE BUSINESS KEYS .....	310
23.3	USING DATE/TIME AS PART OF HUB KEY.....	320
23.4	THE IMPORTANCE OF HUB BUSINESS KEYS .....	321
<b>CHAPTER 24 CORE BUSINESS CONCEPT DESIGN TOPICS.....</b>		<b>323</b>
24.1	CORE BUSINESS CONCEPT.....	324
24.2	THE CENTRAL, ENTERPRISE-WIDE VIEW.....	324
24.3	THE CORE CONCEPT LEVEL .....	326
24.4	ATTRIBUTE SWARMING .....	327
24.5	THE BUZZ AROUND CORE CONCEPTS.....	330
<b>CHAPTER 25 ATTRIBUTE DESIGN .....</b>		<b>335</b>
25.1	CORE DATA WAREHOUSE ATTRIBUTES.....	336
25.2	COMPOUND ATTRIBUTES.....	337
25.3	CONTEXT ATTRIBUTES BECOMING ENTITIES .....	339
25.4	ATTRIBUTE CLUSTERS .....	343

25.5	SPLIT APART AND DRAW TOGETHER.....	345
<b>CHAPTER 26</b>	<b>ARCHITECTURE DESIGN CONSIDERATIONS .....</b>	<b>349</b>
26.1	ARCHITECTURAL LAYERS ANALYSIS.....	350
26.2	LAYERS BETWEEN SOURCES AND MARTS .....	351
26.3	FLOW OF DATA AND THE EDW PIPE.....	354
<b>CHAPTER 27</b>	<b>LOADING THE EDW .....</b>	<b>357</b>
27.1	LOADING THE EDW .....	358
27.2	TECHNICAL BENEFITS OF UNIFIED DECOMPOSITION.....	360
27.3	THE TECHNICAL BENEFITS OF DATA VAULT .....	360
27.4	LOADING PROCESS FUNDAMENTALS .....	361
27.5	LOADING AUTOMATION .....	363
27.6	LOADING THE EDW SUMMARY .....	364
<b>CHAPTER 28</b>	<b>SOURCING THE EDW .....</b>	<b>367</b>
28.1	SOURCING THE EDW .....	368
28.2	NO BUSINESS ACCESS TO DATA WAREHOUSE LAYER .....	368
28.3	THE ROLE OF DATA MARTS .....	369
28.4	SETTING YOUR DATA MARTS FREE.....	370
28.5	BUILDING DIMENSIONS FROM THE DATA VAULT .....	371
28.6	BUILDING FACTS FROM THE DATA VAULT.....	374
28.7	AUTOMATION .....	377
28.8	SOURCING THE EDW SUMMARY .....	378
<b>CHAPTER 29</b>	<b>HELPFUL ANCILLARY TABLES .....</b>	<b>381</b>
29.1	TRANSFORMATION ENABLING STRUCTURES.....	382
29.2	BRIDGE TABLES.....	383
29.3	POINT IN TIME (PIT) TABLE.....	385
<b>CHAPTER 30</b>	<b>ADVANCED CONCEPTS.....</b>	<b>391</b>
30.1	ALL DATA IS UNSTRUCTURED.....	392
30.2	ENTERPRISE DATA INTEGRATION IS IMPOSSIBLE.....	393
30.3	THE TRANSACTIONAL LINK .....	394
30.4	SATELLITES AND NAME/VALUE PAIR .....	401
30.5	DATA DRIVEN VERSUS MODEL DRIVEN .....	402
30.6	SATELLITE DESIGN AND NVP SUMMARY .....	406
30.7	ANCHOR MODELING .....	406
<b>CHAPTER 31</b>	<b>COMPARISON EXAMPLE.....</b>	<b>411</b>
31.1	COMPARING MODELING APPROACHES.....	412
<b>TABLE OF FIGURES</b>	<b>.....</b>	<b>428</b>