

Anil Babu Ankisettipalli, Hansen Chen, Pranav Wankawala

SAP HANA® Advanced Data Modeling

 **Rheinwerk®**
Publishing

Bonn • Boston

Contents

1	SAP HANA Data Models	21
1.1	SAP HANA Database Architecture Overview	21
1.2	SAP HANA Modeling Paradigms	22
1.2.1	Client and Data Connection	23
1.2.2	Modeled Views	24
1.2.3	Stored Procedures	24
1.2.4	C++ (Application Function Libraries)	25
1.2.5	L Language	25
1.2.6	R Language	25
1.3	Information Views	26
1.3.1	Attribute Views	26
1.3.2	Analytic Views	37
1.3.3	Calculation Views	52
1.4	Analytic Privileges	67
1.4.1	Classical XML-Based Analytic Privilege	67
1.4.2	SQL-Based Analytic Privilege	73
1.5	Stored Procedures	75
1.5.1	SQLScript Procedures	75
1.5.2	L Procedures	81
1.5.3	R Procedures	82
1.6	Application Function Library	86
1.6.1	Business Function Library	86
1.6.2	Predictive Analysis Library	88
1.7	Summary	90
2	Modeling Complex Logic	93
2.1	Achieving Recursive Logic with Hierarchies	93
2.1.1	Creating Hierarchies with Tables	94
2.1.2	Creating a Hierarchy in an Attribute or Calculation View	97
2.1.3	Hierarchy View Attributes	100
2.2	Transposing Columns and Rows	110
2.2.1	Column-to-Row Transposition	110
2.2.2	Row-to-Column Transposition	116
2.2.3	Reversing a Matrix	118

2.2.4	Merging Data from Multiple Records	121
2.2.5	Splitting Strings	122
2.3	Using cube() with Hierarchies	123
2.4	Calculating Running Total	127
2.5	Calculating Cumulative Sum	131
2.6	Filtering Data Based on Ranking	134
2.6.1	Using a Subquery	134
2.6.2	Using Window Functions	135
2.6.3	Manipulating Concatenated Virtual Columns	136
2.6.4	Using a Rank Node in a Calculation View	137
2.7	Controlling Join Paths via Filters	138
2.8	Full Outer Join in a Calculation View	143
2.9	Making Dynamic Queries in a Stored Procedure	148
2.9.1	Changing Tables Dynamically	148
2.9.2	Changing Filters Dynamically	150
2.9.3	Changing Output Columns Dynamically	152
2.10	Showing History Records Side By Side	153
2.11	Sample Data	158
2.11.1	Using RAND()	158
2.11.2	Using \$rowid\$	158
2.11.3	Using Identity Columns	159
2.11.4	Using LIMIT/OFFSET	160
2.11.5	Using the TABLESAMPLE SYSTEM	160
2.12	Using a Vertical Union to Join Tables	161
2.13	Sorting Records	163
2.13.1	Sorting IP Addresses	163
2.13.2	Sorting with Exceptions	166
2.13.3	Sorting with User-Defined Rules	167
2.14	Finding Missing Values	168
2.14.1	Using the NOT IN Clause	169
2.14.2	Using a Self-Join	170
2.14.3	Using a Vertical Union	171
2.14.4	Using Window Functions	171
2.15	Using Window Functions for Complex Grouping	172
2.16	Joining Based on a Date Sequence	178
2.17	Using a Nested Calculation View	185
2.18	Summary	191

3 Scaling for Large Datasets 193

3.1	Partitioning	193
3.1.1	Round-Robin Partitioning	194
3.1.2	Range Partitioning	194
3.1.3	Hash Partitioning	195
3.1.4	Two-Level Partitioning	197
3.2	Using Input Parameters to Enforce Pruning	198
3.3	Creating an Index	201
3.4	Analyzing Query Performance with Tools	205
3.4.1	Explain Plan	205
3.4.2	Visualize Plan	207
3.4.3	Performance Trace	208
3.5	Enforcing Execution Paths	214
3.6	Using a Union with Constant Values Instead of a Join	218
3.7	Manipulating Joins in an Analytic View	224
3.7.1	Breaking a Union of Dimension Tables	225
3.7.2	Making Nonequi Joins	230
3.7.3	Modifying Tables	232
3.8	Time Traveling	236
3.8.1	History Tables	237
3.8.2	Simulated History Tables	240
3.9	Storing Temporary Data	243
3.10	Calculating Count Distinct	247
3.11	Using Cached Views	250
3.11.1	Defining a Result Cache	251
3.11.2	Defining a View Cache	254
3.12	Summary	258

4 Basic Predictive Modeling 259

4.1	Predictive Analytics Lifecycle in SAP HANA	259
4.1.1	Commonly Used Models	261
4.1.2	Predictive Algorithms in SAP HANA	264
4.1.3	Application Function Library	266
4.1.4	Business Example	269
4.2	Data Exploration	270
4.2.1	Understanding Sales Data	270
4.2.2	Correlation and Autocorrelation	279
4.2.3	Deterministic Variables	285

4.3	Data Preparation	291
4.3.1	Predictive Data Types	291
4.3.2	Cleaning and Preparing Data	292
4.4	Modeling	297
4.4.1	Predictive Modeling Tasks	297
4.4.2	Setting Control Parameters	299
4.4.3	Creating and Maintaining Models	299
4.4.4	Validating Models	299
4.4.5	Scoring Models	300
4.4.6	Business Example	300
4.5	Creating Models Using SAP Applications on SAP HANA	308
4.5.1	Application Function Modeler	308
4.5.2	SAP Predictive Analytics	311
4.6	Summary	318
5	Advanced Predictive Modeling	319
5.1	R Script Modeling and Design	319
5.1.1	SAP HANA and R Integration	319
5.1.2	Data Considerations	321
5.1.3	Data Types and RLANG Script	324
5.2	PAL Model Consumption	326
5.3	Real-Time Model Consumption vs. Batch Predictive Modeling	329
5.3.1	Real-Time Model Execution	329
5.3.2	Batch Predictive Model Execution	330
5.4	Impact of Data Partitions in Predictive Modeling	332
5.5	Using Multiple R Servers and Data Partitions	333
5.5.1	Predictive Analysis Library	333
5.5.2	Using Multiple R Servers	335
5.6	Modeling Using R and PAL Simultaneously	337
5.7	Summary	340
6	Simulations and Optimizations	341
6.1	Case Study	341
6.2	Monte Carlo Simulation of Value-at-Risk	342
6.2.1	Random Variable Generation	343
6.2.2	Simulation Model and Process	350
6.2.3	Avoiding Imperative Logic	358

6.3	Portfolio Optimization	363
6.3.1	Variance-Covariance Matrix	365
6.3.2	Modeling for Optimization Constraints and Equations	370
6.3.3	Executing Optimization Models	379
6.4	Summary	380
	The Authors	381
	Index	383