

Joe Celko's Data, Measurements and Standards in SQL

Joe Celko



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK* OXFORD -PARIS* SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY*TOKYO

Morgan Kaufmann Publishers is an imprint of Elsevier

**MORGAN
KAUFMANN**

Contents

Introduction.....	xiii
PART 1 HISTORY, STANDARDS, AND DESIGNING DATA	1
CHAPTER 1 Scales and Measurements.....	3
1.1. Measurement Theory.....	4
1.1.1. Range, Granularity and Your Instruments.....	6
1.1.2. Range.....	8
1.1.3. Granularity, Accuracy and Precision.....	8
1.2. Defining a Measurement.....	10
1.3. Tolerance.....	11
1.3.1. Scale Conversion Errors.....	12
1.4. Validation.....	12
1.5. Verification.....	13
1.5.1. Erroneous Values.....	13
1.5.2. Phony Values.....	14
1.5.3. Degree of Trust Versus Risk of Error.....	14
CHAPTER 2 Validation.....	17
2.1. Look-Up Tables.....	18
2.1.1. Auxiliary Tables for Noncomputed Data.....	20
2.2. Check Digits.....	21
2.2.1. Error Detection Versus Error Correction.....	22
2.2.2. Check Digit Algorithms.....	22
2.3. Declarations, not Functions, not Procedures.....	30
2.4. Patterns and Regular Expressions.....	34
2.4.1. Tricks with Patterns.....	35
2.4.2. Results with NULL Values and Empty Strings.....	37
2.4.3. Like is not Equality.....	37
2.4.4. Avoiding the Like Predicate with a Join.....	38
2.4.5. CASE Expressions and Like Predicates.....	40
2.4.6. Similar to Predicates.....	41
2.4.7. Tricks With Strings.....	42
2.4.8. Regular Expression Web Sites.....	45
2.5. Nondatabase Validation.....	45
CHAPTER 3 Data Encoding Schemes.....	47
3.1. Bad Encoding Schemes.....	48
3.2. Encoding Scheme Types.....	51
3.2.1. Enumeration Encoding.....	51

3.2.2. Measurement Encoding.....	52
3.2.3. Abbreviation Encoding.....	53
3.2.4. Algorithmic Encoding.....	53
3.2.5. Hierarchical Encoding Schemes.....	55
3.2.6. Vector Encoding.....	56
3.2.7. Concatenation Encoding.....	57
3.3. Atomic Versus Scalar.....	58
3.4. Transition States.....	59
3.4.1. State Transitions.....	59
3.4.2. State Transition DDL.....	60
3.4.3. State Transition Tables.....	61
3.4.4. Automatic State Transition Tables.....	63
3.5. General Guidelines for Designing Encoding Schemes	63
3.5.1. Use Existing Encoding Standards.....	63
3.5.2. Allow for Expansion.....	64
3.5.3. Use Explicit Missing Values to Avoid NULLs	64
3.5.4. Translate Codes for the End User.....	65
3.5.5. One True Look-up Table.....	66
3.6. Keep the Codes in the Database.....	69
3.7. Multiple Character Sets.....	70
CHAPrEFM Scales.....	71
4.1. Bit Flags are not Scales.....	72
4.1.1. BITs.....	73
4.1.2. BITs and Booleans.....	74
4.1.3. BITVectors.....	75
4.1.4. Replacing BITs.....	76
4.2. Dimensionless Measurements.....	78
4.3. Types of Scales.....	78
4.3.1. Nominal Scales.....	79
4.3.2. Categorical Scales.....	79
4.3.3. Absolute Scales.....	81
4.3.4. Ordinal Scales.....	82
4.3.5. Rank Scales.....	83
4.3.6. Interval Scales.....	84
4.3.7. Ratio Scales.....	85
4.4. Using Scales.....	85
4.5. Scale Conversion.....	86
4.6. Derived Units.....	88
4.7. Punctuation and Standard Units.....	90
4.8. General Guidelines for Using Scales in a Database.....	92

CHAPTER 5	Data with Ignorance.....	95
5.1.	Get it Right.....	95
5.2.	Replace Bad Values with a General Dummy Value.....	96
5.3.	Replace Bad Values with a Statistical Dummy Value.....	96
5.4.	Replace Bad Values to Complete a Statistical Distribution ...	97
5.5.	Replace Bad Values with Statistical Profiling.....	98
CHAPTER 6	Keys.....	99
6.1.	Uniqueness.....	99
6.2.	A Key Cannot Be NULL.....	99
6.3.	Invariant Or Universally Controlled Values.....	100
6.4.	Surrogate And Physical Locators Keys.....	101
6.4.1.	Physical Locators for Performance.....	102
6.4.2.	Physical Locators for Lack of a Proper Key.....	103
6.4.3.	Trusted Sources.....	104
PART2	A SAMPLING OF STANDARDS.....	105
CHAPTER 7	Dates.....	107
7.1.	ISO-8601 standard.....	107
7.1.1.	Year Field.....	108
7.1.2.	Month Field.....	109
7.1.3.	WeekField.....	109
7.1.4.	Day Field.....	110
7.1.5.	Time Field.....	111
7.1.6.	Time Zones and Daylight Saving Time.....	112
7.2.	Putting it all Together.....	113
7.3.	Durations and Intervals.....	113
7.3.1.	Time Intervals.....	114
CHAPTERS	Sex Codes.....	115
8.1.	Sex Codes.....	115
8.2.	Other Sources.....	117
CHAPTER 9	Ethnicity and Race Codes.....	119
9.1.	Race Versus Ethnicity.....	119
9.2.	U.K. Ethnic Groups.....	120
CHAPTER 10	ISO-3166 and Other Country Codes.....	123
10.1.	ISO 3166-1.....	123
10.2.	ISO 3166-2.....	123
10.3.	ISO 3166-3.....	124

CHAPTER 11	Language Codes.....	125
CHAPTER 12	Currency Codes.....	127
	12.1. Noncurrency Units.....	128
CHAPTER 13	National Identification Numbers.....	129
	13.1. Social Security Numbers.....	130
	13.1.1. SSN Area Numbers.....	130
	13.1.2. SSN Group Numbers.....	134
	13.1.3. SSN Serial Numbers.....	135
	13.1.4. SSN Validation.....	135
	13.2. Social Insurance Number.....	136
	13.2.1. SIN First Digit.....	136
	13.2.2. SIN Validation.....	137
	13.3. Swedish Personal Identity Number.....	137
	13.3.1. Personnummer Format.....	137
	13.4. Eu Biometric Passports.....	138
	13 4.1. Fingerprint Classification Systems.....	139
CHAPTER 14	Occupations.....	141
	14 1. National Occupational Classification (NOC).....	142
CHAPTER 15	Colors.....	145
	15.1. International Color Consortium.....	146
CHAPTER 16	Telephone Numbers.....	147
	16.1. The International Telephone Number Components... ..	147
	16.2. Subscriber Local Exchange or Prefix.....	149
CHAPTER 17	E-Mail Addresses.....	151
CHAPTER 18	Universal Postal Union.....	153
	18.1. ZIP Code.....	154
	18.1.1. Basic ZIP Code.....	154
	18.1.2. ZIP + 4.....	154
	18.1.3. Postal Zone Charts.....	155
	18.1.4. ZIP Code Validation.....	156
	18.2. Canadian Postal Codes.....	156
	18.3. Postcodes in the United Kingdom.....	158
	18.3.1. Postcode Formats.....	158
	13.3.2. Greater London Postcodes.....	159
	18.4. Case Expression for Many International Postal Codes ..	160
CHAPTER 19	Hierarchical Triangular Mesh.....<.....	163

CHAPTER 20 Shoe Sizes.....	167
CHAPTER 21 International Clothing Sizes.....	169
CHAPTER 22 ICD Codes.....	173
22.1 Local Versions.....	173
22.2. Mental and Behavioral Disorders.....	174
CHAPTER 23 Vehicle Identification Number (VIN).....	175
23.1. Vin Format.....	175
CHAPTER 24 Freight Containers.....	179
24.1. Freight Container Codes.....	179
24.1.1. Map Letters to Numeric Values.....	180
24.1.2. Assign Weights to Each Position.....	180
24.1.3. Compute the Check Digit from the Weighted Sum.....	180
24.2. Size and Type Codes.....	180
24.3. Related ISO Standards.....	180
CHAPTER 25 Credit Card Numbers.....	183
25.1. Card Issuers.....	183
25.2. Account Number.....	184
25.3. Other Numbers.....	184
25.4. Personal Identification Numbers.....	185
25.5. PCI DSS and Related Standards.....	185
25.6. Tools and More Information.....	186
CHAPTER 26 SWIFT And Related Banking Standards.....	187
26.1. BIC Codes.....	188
26.2. International Bank Account Number (IBAN).....	189
26.2.1. IBAN Check Digits.....	190
CHAPTER 27 Data Universal Numbering System.....	193
CHAPTER 28 Global Trade Item Number.....	195
28.1. GTIN Family.....	196
28.1.1. Prefixes for UPC.....	196
28.2. ISBN.....	197
CHAPTER 29 Digital Object Identifier (DOI).....	199
29.1. DOI Syntax.....	199
29.2. ISBN-A.....!	201
CHAPTER 30 Audiovisual Media.....	203
30.1. Format.....	204
30.1.1. Root.....	204

	30.1.2. Episode or Part.....	204
	30.1.3. Version.....	204
	30.2. Cataloging AV Materials.....	205
CHAPTER 31	Isin And Related Securities Identifiers.....	207
	31.1. Cusip.....	208
	31.2. Sedol and Other Countries.....	209
	31.3. Classification of Financial Instruments.....	209
CHAPTER 32	Temperature Scales.....	211
	32.1. Celsius Scale.....	212
	32.2. Fahrenheit Scale.....	213
	32.3. Kelvin Scale.....	213
	32.4. Other Temperature Scales.....	214
	32.4.1. Rankine Scale.....	214
	32.4.2. Delisle Scale.....	214
	32.4.3. Newton Scale.....	215
	32.4.4. Reaumur Scale.....	215
	32.4.5. Romer Scale.....	215
CHAPTER 33	National Animal Identification System (NAIS).....	217
	33.1. Premises Identification Number (PIN).....	217
	33.2. Animal Identification Number (AIN and GIN).....	218
	33.3. AnimalTracking and other Considerations.....	221
CHAPTER 34	Iso 216 Paper Sizes ("A," "B," And "C" Series).....	223
	34.1. "B" Series.....	224
	34.2. "C" Series.....	225
	34.3. Tolerances.....	225
	34.4. Non-ISO Paper Sizes.....	225
CHAPTER 35	Compass Points.....	227
	35.1. Traditional Compass Points.....	227
	35.2. Other Compass Point Systems.....	228
CHAPTER 36	Unicode.....	231
	36.1. Types of Written Languages.....	232
	36.2. Practical and Political Problems.....	232
	36.3. Normalization.....	233
CHAPTER 37	Driver's Licenses.....	235
	37.1. ID-1 Cards.....	236
	37.2. U.S. Driver's Licenses.....	236
	37.3. Enhanced Driver's License (EDL).....	237

CHAPTER 38	Currency Units and Near Money.....	239
38.1.	Stock Exchanges and Nondecimal Units.....	240
38.2.	Decimalization in the United Kingdom.....	240
38.3.	Physical Currency Choices.....	241
38.4.	Coupons.....	243
38.4.1.	Types of Coupons.....	244
38.4.2.	Trade Groups and References.....	245
CHAPTER 39	Recipes and Food Preparation.....	247
39.1.	Weight Versus Volume.....	247
39.2.	Scaling a Recipe.....	248
CHAPTER 40	Portable Document Format (PDF).....	249
40.1.	ISO 32000.....	249
40.2.	Things that Paper Cannot do.....	251
CHAPTER 41	Temporal Data.....	253
41.1.	Notes On Calendar Standards.....	253
41.1.1.	The Julian and Gregorian Calendars.....	254
41.1.2.	Computerizing Calendars.....	256
41.1.3.	Leap Seconds.....	257
41.1.4.	Calendar Tables.....	258
41.2.	SQL Temporal Data Types.....	261
41.2.1.	Tips for Handling Dates, Timestamps and Times.....	262
41.2.2.	Date Format Standards.....	262
41.2.3.	Handling Timestamps.....	263
41.2.4.	Handling Times.....	265
41.3.	Expressions with Date Arithmetic.....	267
41.4.	The Nature of Temporal Data Models.....	268
41.5.	References on Temporal Data and SQL.....	269
CHAPTER 42	Additive Congruential Generators.....	271
CHAPTER 43	Traditional and Metric Typographic Units.....	275
43.1.	DIN 16507-2 Font Sizes.....	277
	Index.....	281