MWSUG 2024 — Paper AD23

# Q&A with the macro maven: is sql our lingua franca?

Ronald J. Fehd, senior maverick, theoretical programmer,
Fragile-Free Software Institute

## Abstract

**description:** SAS® software provides an implementation of Structured Query Language (sql). Usage of the sql procedure creates an opportunity for the user to learn database terminology and consider how to address their programming problems and reporting as a matter of the design of the data structure.

**purpose:** The purpose of this Q&A dialogue is to review the suitability of sql as a common language for users and programmers.

Its benefits include a knowledge of design of data structure for ease of reporting as well as the ability to research the SAS global symbol table.

The dialogue is an exposition of how we learn both natural and artificial languages, and how we use them to communicate.

**audience:** all levels

**keywords:** data structure: primary key (unique), foreign key (few: category), fact (numeric, summable); procedures: contents, sql; global symbol table

## In this paper:

## Introduction

**audience introduction**

- music skill
- education: degree type, major
- languages: natural: speak second
  artificial: computer language
- SAS years experience
  job category
  industry category

---

**Neuro-Linguistic Programming: NLP**

| NLP deconstructors | psychotherapists |
|---|---|
| Richard Bandler | Milton Erickson |
| Robert Dilts | Fritz Perls |
| John Grinder | Virginia Satir |

The process of observing, analyzing,

distilling, decoding, and installing

the best mental software is called

**modeling**

reprinted from White and Arthur, *Your Seventh Sense, How to Think Like a Comedian*

---

**Dilts: NLP pyramid, 1**


---

**Dilts: NLP pyramid, 2**

| | |
|---|---|
| vision | your resume |
| identity | role |
| values | reusable |
| skills | quality, testing |
| behavior | thorough |
| environment | quiet! |

---

## Learning natural languages

| | | |
|---|---|---|
| **categories of languages** | • pidgin (blank & white) | synthesis of two languages<br>trade: nouns, counting<br>adults learn |
| | • creole (color) | stable, common, daily use<br>children learn |
| | • natural language | culture: writing, literature |
| | • nation | language with an army |
| **steps in learning a second language** | • introduction | hello, goodbye |
| | • counting | time, dates |
| | • directions | right, left,<br>up, down<br>north east south west |
| | • food, drink | water, protein, alcohol |
| | • health | pain, parts of body |

## Learning computer languages

| | | |
|---|---|---|
| **learning any computer language** | • variables | scope: local, global |
| | • conditions | if, case, select, where |
| | • loops | for i = 1 to . . . |
| | • "functions" | f(x), processes, procedures |
| | • syntax | "how to write that idea?" |
| **hierarchy of program types** | • modules | call routines, subroutines |
| | • routines | calls subroutines |
| | • subroutines | do work: create output |

## Learning SAS

**learning SAS, my story**

- data: read column-delimited text
- procedures: contents, format, freq, print, sort, sql, summary, tabulate
- macros: variables and definitions, option sasautos
- programming:
  - operating systems: DOS, MVS
  - %include, autoexec, sasv9.cfg
- sql, database theory,
- scl
- cardinality ratio and type

---

**program: data structure + algorithm**

```
data set
    (label = "...")                ;
     attribute                     ;
     array                         ;
     if 0 then set ...             ;
     drop _*                       ;
do until               (end_of_data);
   set lib.data   end = end_of_data ;
   *...;
   if ... then output;
   end;
stop;
run;
```

---

**macro template**

```
%macro xyz(data = sashelp.class
          ,...
          ,testing = 0
          )/des = '<UNC>\sas-macros\xyz.sas';
%let testing = %eval(not(&testing eq 0)
     or %sysfunc(getoption(mprint)) eq MPRINT
     );
*...;
%if &testing %then %do;
    proc sql; describe table &syslast; quit;
    %end;
%mend xyz;
```

---

## Database theory

**database vocabulary**

- keys:
  - primary — row-id, row-number
  - composite: — two or more
  - foreign — discrete: lookup table dimension table
- facts: — continuous, summable quantity, measurement

---

**database theory: Codd's 12 rules**

0 The foundation rule:
1 The information rule:
2 The guaranteed access rule:
3 Systematic treatment of null values:
4 Dynamic online catalog based on the relational model:
5 The comprehensive data sublanguage rule:
6 The view updating rule:
7 High-level insert, update, and delete:
8 Physical data independence:
9 Logical data independence:
10 Integrity independence:
11 Distribution independence:
12 The nonsubversion rule:

---

**cardinality ratios and types**

example output for sashelp.class, n-obs = 19, n-vars = 5

| mem_ name | var num | card_ type | card_ ratio | n_ levels | name | type | length |
|---|---|---|---|---|---|---|---|
| class | 1 | .unique | 1 | 19 | Name | c | 8 |
| class | 2 | few | 0.1052 | 2 | Sex | c | 1 |
| class | 3 | few | 0.3157 | 6 | Age | n | 8 |
| class | 4 | many | 0.8947 | 17 | Height | n | 8 |
| class | 5 | many | 0.7894 | 15 | Weight | n | 8 |

## QA/QC: Quality assurance or control

**qa/qc personnel**

| date | author | cited | idea |
|---|---|---|---|
| 1890 | Pareto | by Juran | 80/20 rule |
| 1920 | Shewhart | by Deming | PDCA, process control chart |
| 1950s | Deming | Shewhart | PDSA, quality circles |
| 1950s | Juran | Pareto | quality management |
| 1950s | Ishikawa | Deming, Juran | fishbone diagram |
| 1950s | Crosby | | Zero Defects, TQM |

**Phil Crosby: Zero Defects**

- The definition of quality is conformance to requirements (requirements meaning both the product and the customer's requirements)
- The system of quality is prevention
- The performance standard is zero defects (relative to requirements)
- The measurement of quality is the price of nonconformance

---

## Suggested reading

cardinality ratio — Fehd, "Calculating Cardinality Ratio in Two Steps"

predecessors — Fehd, "Advanced Programming Concepts: History of the List Processing and Cardinality Ratio Memes"

languages

computer : `https://en.wikipedia.org/wiki/Computer_language`

creole : `https://en.wikipedia.org/wiki/Creole`

lingua-franca : `https://en.wikipedia.org/wiki/Lingua_franca`

natural : `https://en.wikipedia.org/wiki/Natural_language`

pidgin : `https://en.wikipedia.org/wiki/Pidgin`

database — Agosta, *The Essential Guide to Data Warehousing*
Codd: `https://database.guide/codds-12-rules/`
Kimball and Ross, *The Data Warehouse Toolkit, The Definitive Guide to Dimensional Modeling, Third Edition*

learning sql — Fehd, "How To Use proc SQL select into for List Processing"
Hermansen, "Ten Good Reasons to Learn SAS Software's SQL Procedure"
Lafler, *PROC SQL: Beyond the Basics Using SAS(R), Third Edition*

NLP — White and Arthur, *Your Seventh Sense, How to Think Like a Comedian*
Dilts, *Dilt's Pyramid*

persons

Pareto : 1848–1923; Pareto principle: for many outcomes, roughly 80% of consequences come from 20% of causes. Other names: the 80/20 rule, law of vital few, or principle of factor sparsity; Pareto chart cited

in TQM and Six Sigma.
https://en.wikipedia.org/wiki/Vilfredo_Pareto

Shewhart : 1891–1967; father of statistical quality control; PDCA: Plan, Do, Check, Act; cited by Deming
https://en.wikipedia.org/wiki/Walter_A._Shewhart
process-behavior charts
https://en.wikipedia.org/wiki/Control_chart

Shannon : 1916–2001; father of information theory
https://en.wikipedia.org/wiki/Claude_Shannon

Tufte : 1942–; book, 1983: The Visual Display of Quantitative Information
https://en.wikipedia.org/wiki/Edward_Tufte
https://www.edwardtufte.com/tufte/

Tukey : 1915–2000; book, 1977: Exploratory Data Analysis
https://en.wikipedia.org/wiki/John_Tukey

Deming : 1900–1993; cited Shewhart on statistical process control; 1950s: described quality circles, translated in Japan by Ishikawa; 1983: see "14 points" in book: Out of the Crisis; 1985: Total Quality Management (TQM) based on Deming's ideas; 1993: Plan-Do-Study-Act (PDSA) cycle, successor of Shewhart's PDCA
https://en.wikipedia.org/wiki/W._Edwards_Deming
https://en.wikipedia.org/wiki/Quality_circle

Juran : 1904–2008; 1941: cited Pareto; Juran trilogy: quality planning, quality control, and quality improvement; 1950s: lectured in Japan on quality management https://en.wikipedia.org/wiki/Joseph_M._Juran

Ishikawa : 1915–1989; refined Deming's quality circles, fishbone diagram; 1950s: translated Deming and Juran
https://en.wikipedia.org/wiki/Kaoru_Ishikawa
continual improvement (kaizen)
https://en.wikipedia.org/wiki/Continual_improvement_process
fishbone diagram:
https://en.wikipedia.org/wiki/Ishikawa_diagram

Crosby : 1926–2001; Zero Defects;
quote: "Get things done right the first time"; book: Quality is Free
https://en.wikipedia.org/wiki/Philip_B._Crosby

qa/qc

6σ : Six Sigma https://en.wikipedia.org/wiki/Six_Sigma

Kaizen : https://en.wikipedia.org/wiki/Kaizen

TQM : Total Quality Management
https://en.wikipedia.org/wiki/Total_quality_management

ZD : Zero Defects https://en.wikipedia.org/wiki/Zero_Defects

---

## Conclusion

Gaining an understanding of SQL and database theory has help me solve many of the problems I have been challenged with during my career. Most discussions of programming come back to two issues: (lack of) data structure, and clarity of algorithm.

I hope that this paper has provided an historical overview of issues related to quality of programming using SAS software.

---


**Author Information**

| | |
|---|---|
| Ronald J. Fehd | Ron.Fehd.macro.maven at gmail dot com |
| LinkedIn | https://www.linkedin.com/in/ronald-fehd-5125991/ |
| affiliation | senior maverick, theoretical programmer, Fragile-Free Software Institute |
| also known as | `macro maven` on SAS-L |


---



---

## References


Agosta, Lou (2000). *The Essential Guide to Data Warehousing*. 19 chapters, 499 pp., glossary: 26 pp., references: 4 pp., index: 15 pp. Prentice Hall.

Dilts, Robert (1980). *Dilt's Pyramid*. Neuro-Linguistic Programming (NLP) practitioner checklist. URL: https://www.cstopics.com/en/content/dilt-pyramid-the-neurological-levels.

Fehd, Ronald J. (2010). "How To Use proc SQL select into for List Processing". In: *SouthEast SAS Users Group Conference Proceedings*. Hands On Workshop, 40 pp.; writing constant text, and macro calls, using macro %do loops; URL: http://analytics.ncsu.edu/sesug/2010/HOW06.Fehd.pdf.

— (2016). "Calculating Cardinality Ratio in Two Steps". In: *MidWest SAS Users Group Annual Conference Proceedings*. Tools of Trade, 22 pp.; uses output data set from proc freq nlevels to create a database dimension table similar to both the contents and sql dictionary.columns output data sets; additional information includes cardinality ratio and cardinality-type in (few, many, unique) which is used to separate by or class variables from analysis variables. URL: http://www.mwsug.org/proceedings/2016/TT/MWSUG-2016-TT03.pdf.

— (Sept. 2018). "Advanced Programming Concepts: History of the List Processing and Cardinality Ratio Memes". In: *Western Users of SAS Software Annual Conference Proceedings*. 30 pp.; arrays of macro variables; cardinality ratio; compiled or executed; contents, extended; database vocabulary; data structure and algorithm; dosubl function; list processing; macro arrays; scl functions: open, close, attrn(nobs, nvars), fetchobs, getvarc, getvarn, varname, vartype; associative array. URL: http://www.lexjansen.com/wuss/2018/71_Final_Paper_PDF.pdf.

Hermansen, Sigurd (1997). "Ten Good Reasons to Learn SAS Software's SQL Procedure". In: *SAS Users Group International Annual Conference Proceedings*. 5 pp.; URL: https://support.sas.com/resources/papers/proceedings/proceedings/sugi22/ADVTUTOR/PAPER35.PDF.

Kimball, Ralph and Margy Ross (2013). *The Data Warehouse Toolkit, The Definitive Guide to Dimensional Modeling, Third Edition*. 600 pp. John Wiley & Sons, Inc., New York. URL: https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/books/data-warehouse-dw-toolkit/.

Lafler, Kirk Paul (2019). *PROC SQL: Beyond the Basics Using SAS(R), Third Edition*. 420 pp. SAS Institute. URL: https://sasinstitute.redshelf.com/book/1878360.

White, Karyn Ruth and Jay Arthur (2006). *Your Seventh Sense, How to Think Like a Comedian*. Neuro-Linguistic Programming: (NLP). LifeStar. URL: https://www.yourseventhsense.com/.