SQL Rejuvenated

Data, Relationships, RDM, Tables, NoSQL & Set Theory

D L Childs
<iis@umich.edu>

1. INTRODUCTION

SQL[2,5] has always been a Set Query Language. However, it has never been implemented to fully exploit the advantage of its set-theoretic heritage. The fault is not with system developers, it is with the set-theoretic foundations.

2. Relational Data Model

The Relational Data Model, RDM[6], was introduced to provide users formally defined operations for manipulating **relationships** contained in data.

'It is proposed that users interact with a RDM using time-varying relationships (rather than relations)'.[6c]

Since data relationships are represented as **records** and since **n-tuples** from set theory are a conceptually equivalent representation of records, set operations on sets of n-tuples seemed like an ideal formal model for data relationships.

3. Relationships are Sets

Relationships are sets, equivalence classes of **relations**.[6a] Relations are sets, collections of **n-tuples**, but are only well defined when $\mathbf{n = 2}$.[8]

For RDM set operations to be well defined on 'n-ary' relations required n-tuples to be well defined. Since set theory could not satisfy this condition, RDM set operations had to be simulated using **labeled arrays**.[2,5,6b]

4. Arrays are not Sets

In classical set theory, CST, sets are defined as collections of unordered elements. Arrays have both ordered columns and ordered rows. Set operations on arrays are not defined.

It is a significant accomplishment of both the RDM and subsequent RDBMSs that meaningful set operations were defined to manipulate arrays as if they were sets.

5. RDM, SQL & Set Operations

The fundamental SQL retrieval operation is **SELECT**. The fundamental data relationship is a **mapping**. The fundamental data processing operation is a **function**. The fundamental set-theoretic support for all three of these is **Image**.

This set-theoretic connection was well appreciated by the early developers of SQL. The primary operation of SQUARE, the precursor of SQL, was **mapping**[2]. Mappings produce **sets** and are the functional equivalent of CST Image[3].

Set-Processing Research & Development Series [SQLRJ: 16/04/02]

6. Extended Image

The foundation of SQL rests on the CST definition of Image. CST Image only supports binary relations. To adapt the CST Image to support n-ary relations required ingenuity, rules, restrictions, conventions and "work-a-rounds".

If an extended definition of Image had existed, one that supported any possible data representation, then the SQL SELECT operation could be a well defined set operation. There would be no restrictions on the location, diversity, access, nor organization of any stored data.

7. SQL Rejuvenation

Though such a definition for Image did not exist when SQL was being developed, an extended Image[3] definition now exists, as defined under extended set theory, XST[1].

Since every data representation has an identity as a set, XSP[4,10], and since SQL SELECT processes sets, SQL can be extended to process any and all data representations.

8. CONCLUSION

For over thirty years SQL has dominated the DBMS industry. Today the strengths of its set-theoretic capabilities are being overshadowed by the lack of data diversity and unnecessary implementation restrictions.[7]

Adding extended set-theoretic capabilities to existing SQL systems can improve performance by orders of magnitude[9] and allow support for arbitrarily diverse, distributed data.

With extended set-theoretic support, SQL can dominate the industry for another thirty years.

9. REFERENCES

[1] Blass, Childs: *Axioms & Models for an Extended Set Theory*, http://www.math.lsa.umich.edu/%7Eablass/XST_Axioms.pdf
[2] Boyce: *Specifying Queries as Relational Expressions: SQUARE*, IBM 1973, Paper introduced the term **Table**. http://xsp.xegesis.org/SQUARE.pdf
[3] Childs: *Functions as Set Behavior* http://xsp.xegesis.org/Process.pdf
[4] Childs: **XSP**: *Extended Set Processing* http://xsp.xegesis.org/ExcSum.pdf
[5] Chamberlin; Boyce:*SEQUEL*, [renamed SQL] IBM 1974 http://faculty.cs.tamu.edu/yurttas/PL/DBL/docs/sequel-1974.pdf
[6] Codd, E. F.: *A Relational Model of Data for Large Shared Data Banks*, CACM 13, No. 6 (June) 1970 http://www.seas.upenn.edu/~zives/03f/cis550/codd.pdf
a) Relationships are equivalence classes of relations.
b) Arrays are not an essential part of the relational view.
c) http://xsp.xegesis.org/Relations_70.pdf[p. 379-380]
[7] **NoSQL**: (Wikipedia), https://en.wikipedia.org/wiki/NoSQL
[8] Skolem, Thoralf: *Two Remarks on Set Theory,* Mathematica Scandinavica 1957 http://xsp.xegesis.org/T_Skolem.pdf
[9] Stout: *Information Access Accelerator*, IBI 2005 http://xsp.xegesis.org/Iaa_ibi.ppt
[10] **XSP Technology**: (webpage), http://xsp.xegesis.org