Content-Container Data Access Strategies

Content for Functionality - Containers for Performance

D L Childs
<iis@umich.edu>

1. INTRODUCTION

A database management system, DBMS, manages data for application processing and for storage preservation.

Since data represented for processing is not always ideal for preservation, and since data represented for preservation is not always ideal for processing, accessing application-friendly data from storage-friendly data *poses a genuine challenge.*[4]

Large increases in volumes of available data and diversity of applications have developers questioning use of traditional data access strategies for development of future DBMSs.[6]

2. DATA ACCESS

Computers do not access data. Computers do not process nor store data. Computers process **data representations**, and store **representations of data representations**.

Data access is actually **data representations** accessed from **representations of data representations**.

3. CONTENT & CONTAINERS

The difficulty with traditional data access strategies is a lack of precision for manipulating data representations. This can be resolved with the introduction of two new terms.

Data Content: a data representation for processing.

Data Container: a data representation for preservation.

Using these terms for actual data representations allows management of **content** independently of **containers.**

Contents are equivalent if they represent the same data. (In the same sense that **SIX, VI, 0110, 6** are equivalent.)

4. TABLES are CONTENTS

The Relational Data Model, RDM[3], introduced a formal definition of content, **Tables**. Tables are **not** files, but RDM implementations, unfortunately, bind Tables to files.

This binding of **contents** to **containers** violates the idea of **data independence**. It is also the main barrier to any possible implementation of an optimal data access strategy.

5. DATA & Set Operations

The RDM introduced advantages of using set operations to manage application data. Though Codd envisioned the advantages of also using set operations to manage storage data, the foundations required to represent *structured sets* were not available at the time, but they are available now[1].

Set-Processing Research & Development Series [CCDAS: 16/03/28]

6. XSP: Structured Sets

Structured sets, or extended sets, are just classical sets with and added membership condition, as formally defined under the axioms supporting extended set processing, XSP[2].

This additional parameter can be used to formally define physical representations of abstract logical relationships. All data representations have a formal set definition under XSP.

7. XSP: Access Strategies

Since all data representations are XSP sets, **content** and **containers** can be managed using set operations. Thus any conceivable data access strategy can be implemented with set operations. Now the only *genuine challenge* to data access development is the imagination of developers.

8. SQL Rejuvenation

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

9. CONCLUSION

Today's DBMSs use a single base **content-container** access for all applications. This *one size fits all* imposes a significant difficulty for development of future DBMSs.[5]

Systems with multiple applications sharing large volumes of available data, can use content-container access strategies to provide each application with just the right content, in just the right format, in just the right time.

10. REFERENCES

[1] Blass, Childs: *Axioms & Models for an Extended Set Theory*, http://www.math.lsa.umich.edu/%7Eablass/XST_Axioms.pdf

[2] Childs: *XSP: Extended Set Processing* http://xsp.xegesis.org/ExcSum.pdf

[3] 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

[4] Larsen: *The Business Value of Intelligent Data Access* - For a wide range of reasons, designing for and maintaining optimal data access poses a genuine challenge to even the most sophisticated enterprises.- 2009 http://www.ca.com/files/whitepapers/data-access-white-paper-final-030909_201548.pdf

[5] Stonebraker: *"One Size Fits All": An Idea Whose Time Has Come and Gone* - Paper argues that traditional DBMS architectures using a single storage organization to support a variety of applications is no longer applicable. ICDE, 2005

[6] Stonebraker: *The End of an Architectural Era* - The DBMS vendors (and the research community) should start with a clean sheet of paper and design systems for tomorrow's requirements, not continue to push code lines and architectures designed for yesterday's needs. - 2007 http://www.ahzf.de/itstuff/papers/vldb07hstore.pdf

[7] **XSP Technology**: (webpage), http://xsp.xegesis.org