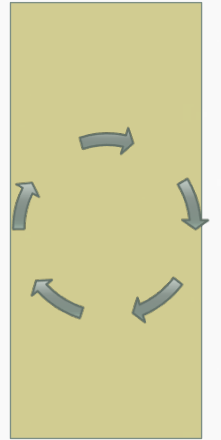


WHAT GOES AROUND COMES AROUND

MEHVISH SALEEM



AUTHORS



Joseph Hellerstein



Michael Stonebraker

CONTENT

- Introduction
- Hierarchical IMS Era
- Network CODASYL Era
- Relational Era
- Comparison
- Conclusion

INTRODUCTION

BACKGROUND

- Provides a summary of 35 years of data model proposals
- Discusses data models divided into 9 eras
- Talks about the pros, cons and lessons learned from each era
- Example used:

Supplier (sno, sname, scity, sstate)

Part (pno, pname, psize, pcolor)

Supply (sno, pno , qty, price)

IMS (INFORMATION MANAGEMENT SYSTEM)

IMS ERA

- Late 1960's and 1970's
- Hierarchical data model: a collection of instances of record types
- Uses a “record-at-a-time” language called DL/I for data manipulation
- Stored sequentially or through hashing

Supplier (sno,
sname, scity,
sstate)



Part (pno, pname,
psize, pcolor, qty,
price)

Part (pno,
pname, psize,
pcolor)



Supplier (sno,
sname, scity,
sstate, qty, price)

TWO HIERARCHICAL ORGANIZATIONS

PROS AND CONS

- Simple
- Some support for logical data independence
- Data Redundancy
- Child cannot exist without parent
- Lack of physical data independence

CODASYL (COMMITTEE ON DATA SYSTEMS LANGUAGES)

CODASYL ERA

- Reports released in 1970s
- A directed graph model
- Database consists of a collection of record and set instances of the set type
- Record-at-a-time data manipulation language

Supplier (sno,
sname, scity,
sstate)

Part (pno,
pname, psize,
pcolor)

Supplies

Supplied_by

Supply(qty, price)

A CODASYL DIRECTED GRAPH

PROS AND CONS

- Flexible
- No redundancy
- Can deal with corner cases
- Complex model
- Long long load times due to the graph structure
- Poorer physical and logical independence

RELATIONAL

HOW DOES IT WORK?

- Proposed in 1970 by Ted Codd
- Data structured as relations – sets of tuples
- Uses a set-at-a-time query language

sno	sname	scity	sstate
16	GS	Boston	MA
24	SS	Detroit	Mi

sno	pno	qty	price
16	27	100	\$20.00
24	42	1000	\$0.08

pno	pname	psize	pcolor
27	Saw	7	silver
42	bolts	12	gray

RELATIONAL DATABASE

THE GREAT DEBATE

- A debate between Tedd Codd and Charlie Bachman (inventor of the network model) lasted through most of the 1970s
- SIGFIDET '74
- Conclusion left on the commercial marketplace

COMPARISON

- Arguments against CODASYL:
 - Too complex
 - No acceptable data independence
 - Not flexible enough
- Arguments against Relational:
 - Difficult to understand the new languages
 - Might not be possible to implement the model efficiently

SO WHO WON?

- VAX (32-bit minicomputers) supported relational and mainframes supported non-relational
- IBM announced dual support for IMS and DB/2 (an early relational database)
- SQL established as the standard relational database language

WHAT DID WE LEARN?

- The importance of data independence
- Tree structured models are restrictive
- Optimization of record-at-a-time queries is hard
- Directed graphs are more flexible than hierarchies but more complex
- Set-at-a-time languages offer more data independence
- Technical debates are usually settled by market giants

THANK YOU!