The Askew Wall SQL and The Relational Model (background to *The Third Manifesto*) Hugh Darwen HD@TheThirdManifesto.com

for Warwick University CS319

Terminological Equivalences (?)

Posh term	Cuddly term
Relation	Table
(n-)tuple	Row
Attribute	Column
Domain	(data) type
Domain	(object) class

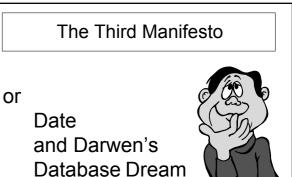
The table represents a relation! Its predicate: *Cuddly term* is the SQL counterpart of Relationland's *Posh term*.

The Perversity of SQL

SELECT CityName FROM City C1 WHERE 4 > (SELECT COUNT(*) FROM City C2 WHERE C1.Population < C2. Population)

The Unperversified Version

SELECT CityName FROM City C1 WHERE (SELECT COUNT(*) FROM City C2 WHERE C2. Population > C1. Population) < 4



References

Relational Database Writings 1985-1989 by C.J.Date with a special contribution "Adventures in Relationland" by H.D. (as Andrew Warden)

Relational Database Writings 1989-1991 by C.J.Date with Hugh Darwen

Relational Database Writings 1991-1994 by C.J.Date

Databases, Types, and The Relational Model : *The Third Manifesto*

by C.J. Date and Hugh Darwen (3rd edition, 2006)

SQL: A Comparative Survey by Hugh Darwen, 2012 (free download from Bookboon.com)

A Brief History of Data 1960: Punched cards and magnetic tapes 1965: Disks and 'direct access' 1970: E.F. Codd's great vision: "A Relational Model of Data for Large Shared Data Banks" 1970: C.J. Date starts to spread the word 1975: Relational Prototypes in IBM: PRTV (ISBL), QBE, System R 1979: First SQL product: Oracle. IBM shipped in 1980. 1987: SQL an international standard 1990: OODB – didn't come to much in the end 2000: XML? (shudder!)

A Brief History of Me

1967 : IBM Service Bureau, Birmingham
1969 : "Terminal Business System" – putting users in direct contact with their databases.
1972 : Attended Date's course on database (a personal watershed)
1978 : "Business System 12" - a relational dbms for the Bureau Service
1985: Death of Bureau Service (and of BS12)
1987 : Joined IBM Warwick dev. lab. Attended a Codd & Date database conference in December
1988 : "Adventures in Relationland" by Andrew Warden.

2004 : Retired from IBM

Blank Slide

(material from lecture not suitable for handouts)

8

10

Blank Slide

(material from lecture not suitable for handouts)

Blank Slide

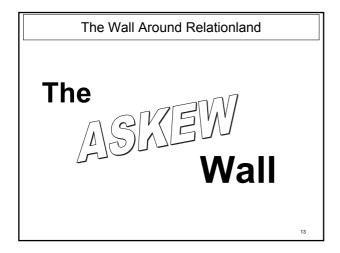
(material from lecture not suitable for handouts)

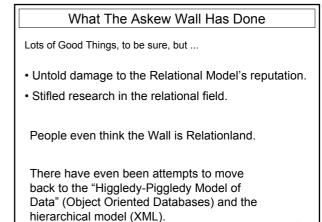
Blank Slide

(material from lecture not suitable for handouts)

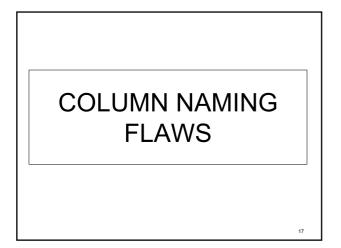
11

Bank Slide (material from lecture not suitable for handouts)





The Good Things The Askew Wall Has Done	The "Fatal Flaws" of SQL
odd's vision has come true in the following respects: • TABLE as the only available structure (though, alas, table ≠ relation)	Anonymous columns (partly addressed in 1992)
 Value at row/column intersection the ONLY method of representing information. e.g., no pointers, no ordering of rows. 	 FROM clause restricted to named tables (fixed in 1992) Duplicate column names
 Orthogonality of tables with respect to data types (domains) over which their columns are defined. 	 Order of columns is significant Duplicate rows
The catalogue is made of tables, too.	• NULL
 Query language ALMOST closed over tables and does embrace relational algebra/calculus principles (as well as regrettably departing from them). 	 Failure to support degenerate cases (e.g. columnless tables)
 Constraints expressed declaratively, in the schema, and enforced by the dbms. 	 Failure to support "=" properly and lots more, and probably more to come
• No "record-level" (or other) subversion.	



	Α	Thema	atic Query	Example		
Ref:	"The Nar	ning of Co	olumns", char	oter 17 in RDB	N 1985-89	
Given :	Nam	e	Subject		Mark	
ExamMarks	Ann	е	Relationa	I DB	92	
	Bori	s	Object DI	В	68	
	Cinc	dy	Object DI	В	56	
	Dav	е	Relationa	l DB	84	
To derive:						
[Name	Subjec	:t	Mark	Avg	
	Anne	Relat	ional DB	92	88	
	Boris	Objec	t DB	68	62	
	Cindy	Objec	t DB	56	62	
	Dave	Relat	ional DB	84	88	18

Anonymous Columns	
Ref: "The Naming of Columns", chapter 17 in RDBW 1985-	89
Example 3:	
Show the average exam mark obtained by all students in each subject.	
SELECT Subject, AVG(Mark) FROM ExamMarks GROUP BY Subject	
The "second" column of this table has no name!	
This is a CORRECTABLE flaw (well, NEARLY correctable). It was corrected (kind of) in 1992.	
It is NOT BYPASSABLE	
	19

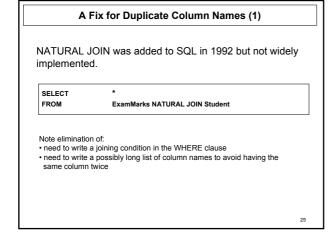
FR	COM Clause Restricted to Named Tables
Example 4:	
	student in each subject the mark obtained and the average dents in that subject.
SELECT FROM	Name, E.Subject, E.Mark, S.??? unnamed column ExamMarks E, (SELECT Subject, AVG(Mark) FROM ExamMarks GROUP BY Subject) S E.Subject = S.Subject
	table flaw (and was corrected in 1992). It is not generally ough sometimes you can create a named view for the
	articular query CAN be done without nesting (exercise for e solution cannot be generalized.
	20

	The FROM clause fix	
Example 4 (fixed):		
obtained a	ach student in each subject the mark nd the average mark obtained by all that subject.	
SELECT	Name, E.Subject,E.Mark,S.Avg	
FROM	ExamMarks AS E,	
	(SELECT Subject, AVG(Mark) AS Avg	
	FROM ExamMarks	
	GROUP BY Subject) AS S	
WHERE	E.Subject = S.Subject	
	nly an optional conformance feature in I think it is <i>very</i> important. Without it SQL is ncomplete.	
	But it's clunky!	21

WITH AvgMar	ks AS
(SELECT	Subject, AVG(Mark) AS Avg
FROM	ExamMarks
GROUP E	3Y Subject)
SELECT Name	e, E. Subject,E.Mark,A.Avg
FROM Exan	Marks AS E, AvgMarks AS A
WHERE E.Su	bject = A. Subject

	Duplicate column names (1)
F	Ref: "In Praise of Marriage", chapter 18 in RDBW 1985-89
mple 5:	
SELECT	*
FROM	ExamMarks E, Student S
WHERE	E.Name = S.Name
	ooking join, but there are two columns called Name. <u>duplicate column</u> s, as it happens.
ese are also cometimes	<u>duplicate column</u> s, as it happens.
ese are also cometimes	duplicate columns, as it happens. such joins generate two columns both called,
ese are also cometimes	duplicate columns, as it happens. such joins generate two columns both called,

[
Duplicate Column Names (2)	
Thanks to AS you can now even do this :	
SELECT Col1 AS X, Col2 AS X FROM T	
Enjoy!	
	24



Why NATURAL JOIN is "Natural"

Relational JOIN is the relational counterpart of logical AND.

Assume predicates: ExamMarks "Name scored Mark in Subject". Student "Name joined in Year, was born on DoB [etc.]"

Then the predicate for

ExamMarks JOIN Student

is "Name scored Mark in Subject and Name joined in Year, was born on DoB [etc.]"

The two Names are the same variable!

26

DUPLICATE ROWS



Ref: "The Duplicity of Duplicate Rows", chapter 5 in RDBW 89-91 "The Keys of the Kingdom", chapter 19 in RDBW 85-89, and:	
"If something is true, saying it twice doesn't make it more true (E.F. Codd)	"
This is a bypassable flaw:	
 Declare at least one candidate key for every base table. and ask for support for system-generated keys. 	
 Always write DISTINCT after the word SELECT and complain to supplier if this makes duplicate-free queries go slower. 	
 Never write the word ALL after UNION and demand decent optimization here, too. but, alas, it is not a correctable flaw. 	
Usability problems should be recognized and solved, but	

NOT by departing from fundamental principles.

Duplicate Rows

Are Duplicate Rows Really Harmful?

Well, they needlessly complicate the language, making it more difficult than it ought to be to define, teach, and learn.

And they allow the implementation to place some of the performance burden needlessly on its users (when is it safe to write UNION ALL or to omit DISTINCT?)

But look at the trap somebody fell into here:

1

29

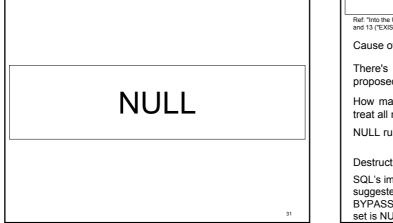
SELECT COUNT(*) AS HowMany, AVG (Salary) AS HowOld FROM (SELECT Emp#, Salary FROM Emp NATURAL JOIN

(SELECT Emp# FROM WorksIn

WHERE Dept# IN ('D3', 'D7')) AS dummy) AS dummy

"How many people work in departments 3 and 7, and what is their average salary?"

In Tutorial D: SUMMARIZE (((WorksIN WHERE Dept# = 'D3' OR Dept# = 'D7') { Emp# } JOIN Emp) { Emp#, Age }) ADD (COUNT AS HowMany, AVG(Age) AS HowOld) 30



NULL

Ref: "Into the Unknown", chapter 23 in RDBW 85-89. See also chapters 8 ("NOT" is not 'Not!") and 13 ("EXISTS is not 'Exists'!" and the whole of part IV(chapters 17-21) in RDBW 89-91

Cause of much debate and anguish.

There's even a split in the relational camp (E.F. Codd proposed "A-marks", "I-marks" and a 4-valued logic).

How many different things can NULL mean? Is it valid to treat all nulls alike?

NULL ruins everything -

- UNION of sets, cardinality of sets.

Destruction of functional dependency theory

SQL's implementation of NULL is even worse than the best suggested by theoreticians. And it's not completely BYPASSABLE, because SQL thinks that the sum of the empty set is NULL! Nor is it CORRECTABLE.

Is NULL a Value?

CREATE TABLE NT (N INTEGER) ; INSERT INTO NT VALUES NULL, NULL ; -- inserts two rows!

Now, what is the cardinality of:

(a) SELECT * FROM NT WHERE N = N (b) SELECT * FROM NT WHERE N <> N (c) SELECT DISTINCT * FROM NT

Which answers are consistent with NULL being a value?

What difference would it have made if it had been a value?

33

3-Valued Logic: The Real Culprit

Relational theory is founded on classical, 2-valued logic.

A relation *r* is interpreted as a representation of the extension of some predicate *P*.

Let t be a tuple with the same heading as r.

If tuple *t* is a member of *r*, then the proposition P(t) is taken to be TRUE; otherwise (*t* is not a member of *r*), P(t) is taken to be FALSE.

There is no middle ground. The Law of The Excluded Middle applies.

There is no way of representing that the truth of P(t) is unknown, or inapplicable, or otherwise concealed from us.

SQL's WHERE clause arbitrarily splits at the TRUE/UNKNOWN divide.

Surprises Caused by SQL's NULL

- SELECT * FROM T WHERE X = Y OR NOT (X = Y) is not equivalent to SELECT * FROM T
- SELECT SUM(X) + SUM(Y) FROM T is not equivalent to SELECT SUM(X + Y) FROM T
- IF X = Y THEN 'Yes'; ELSE 'No' is not equivalent to IF NOT (X = Y) THEN 'No'; ELSE 'Yes'

Why NULL Hurts Even More Than It Once Did

Suppose "x = x" returns "unknown"

Can we safely conclude "*x* IS NULL" ?

Suppose x "is not the null value"?

Can we conclude "x IS NOT NULL"?

Not in modern SQL!

How x= x Unknown Yet x NOT NULL

For example:

- 1. *x* is ROW (1, null) or even ROW(null, null) ROW(...) is a row "constructor".
- x is POINT (1,null)
 POINT(a,b) is a "constructor" for values in the user-defined data type POINT.

37

39

3. x is ROW (POINT(1,1), POINT(null,3))

Consequences?

TABLE_DEE AND TABLE_DUM

38

TABLE_DEE and TABLE_DUM

Ref: "TABLE_DEE and TABLE_DUM", chapter 22 in RDBW 85-89, and "The Nullologist in Relationland, or Nothing *Really* Matters", chapter 13 in RDBW 89-91

Two very important relations that SQL is unaware of.

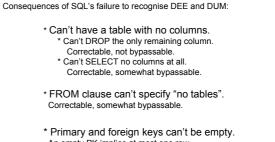
Consider the question, "Do we have any students?"

In Tutorial D: Student { }

SQL counterpart would be SELECT DISTINCT FROM Student

The result is a relation of degree 0 (no attributes) and either one tuple (TABLE_DEE) or none (TABLE_DUM).

Interesting property of TABLE_DEE: for every relation r, r JOIN TABLE_DEE = r

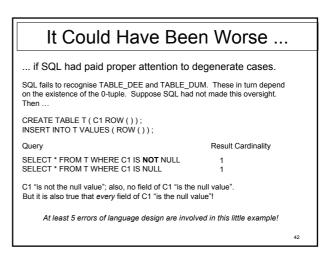


Failure to Recognise DEE and DUM

^c Primary and foreign keys can't be empty. An empty PK implies at most one row. Correctable, not bypassable.

and the above set of nullological observations is still growing

Вура	isses for Absence of DEE and DUM	
mple 6:		
d any student	obtain more than 75 marks in Relational DB?"	
SELECT	DISTINCT 'Yes!'	
	ExamMarks	
FROM	Examiniation	
WHERE	Mark > 75 AND Subject = 'Relational DB'	
	Mark > 75 AND Subject = 'Relational DB'	



MISCELLANEOUS FURTHER FLAWS

43

"=" Is Not "equals"

Modern SQL supports user-defined "equals" functions, for user-defined data types.

We would like to require these to honour the rule that if a=b then for all f, f(a) = f(b)

Unfortunately SQL itself already fails to honour it: 'A' = 'A ', but Length('A') < Length('A ')

Unpleasant consequences for GROUP BY, NATURAL JOIN, DISTINCT, foreign keys, etc.

44

Indeterminacy in SQL

What's the result of: SELECT DISTINCT X FROM (VALUES 'x', 'x ') AS T(X) ?

or:

SELECT MAX(X) AS MaxX, MIN(X) AS MinX FROM (VALUES 'x', 'x ') AS T(X) ?

Because of these difficulties, all expressions involving DISTINCT, MAX and MIN on character strings are classed as "possibly non-deterministic" and therefore may not be used in constraints.

This flaw is rather irrelevant in practice because support for subqueries in constraints is optional and most implementations don't allow those anyway.

Tables and Multisets of RowsSQL's table types are not "data types".So a table can't be the value of a local variable, a
parameter, an operator invocation, or a "field".But users wanted support for "nested tables", like
relation-valued attributes in Tutorial D.So SQL:2003 introduced MULTISET types.

A MULTISET type whose element type is a ROW type is a table type "in disguise"!

SQL's TABLE Operator

TABLE (exp)

If *exp* is a query (table expression), result is the multiset consisting of the rows of *exp*.

If *exp* is a multiset expression, result is the table consisting of the rows of *exp*.

Enjoy! (but beware: TABLE(TABLE(*t*)) is not necessarily equal to *t*) In SQL: SELECT Name, Total_Pay FROM (SELECT (E.*)-Salary + Bonus AS Total_Pay FROM Emp E) AS dummy WHERE Total_Pay > 1000 Required but useless name! In Tutorial D: EXTEND Emp ADD (Salary + Bonus AS Total_Pay) WHERE Total_Pay > 1000 {Name, Total_Pay}

Clunkiness of SELECT-FROM-WHERE

A Murky Story

On the subject of: SELECT(E.*, Salary + Bonus AS Total_Pay

Must qualify * here!

49

In 2005 a UK proposal to correct the silly mistake and allow SELECT *, Salary + Bonus AS Total_Pay was vigorously opposed by the USA, led by Oracle and IBM, and consequently defeated. Why?

Because "* leads to maintenance nightmares, and [we are] not aware of any customer request or requirement for the feature" and "its use should be discouraged".

Why "Maintenance Nightmare"

Because of yet another violation by SQL of the relational model, which stipulates that there is no significance to any *order* in which the attributes of a relation might appear.

But for getting data out of and into a table, SQL uses

SELECT/FETCH INTO :v1, :v2, ... and INSERT INTO target source

The mapping from source columns to target columns is by column number.

Defining the order means defining it for every query operator, including FROM and UNION, which thus fail to be commutative (as they should be, as relational counterparts of AND and OR). Are they associative?

The correct approach is to map columns to variables by name, not by order.

50

"Nobody Ever Asked For It"

A couple of other nice ideas from ISBL that "nobody has ever asked for" :

SELECT all columns except specified one. In Tutorial D:

r { ALL BUT a, b, ... }

Very often you know the ones you don't want and there aren't so many of them either. (In 2005 a UK proposal to add * EXCEPT (...) to SQL was also rejected on the grounds that it would encourage use of the hated *.)

Rename selected columns and keep the others. In Tutorial D:

r RENAME (a AS x, b AS y, ...)

Handy for getting the right attributes to match up for JOIN, UNION, etc.

The Folly of "Structured Queries"			
In old SQL, the WHERE clause could not results of aggregation, so they had to invesame meaning as WHERE):			
SELECT D#, AVG(Salary) AS Avg_Sa FROM Emp GROUP BY D# HAVING AVG(Salary) >999	I		
But would we ever have had HAVING if in SELECT * FROM (SELECT D#, AVG(Sal) AS Avg_S FROM Emp GROUP BY D#) AS dummy WHERE Avg_Sal > 999			

The Folly of Coercion

Consider the so-called scalar subquery, expressed by placing a query inside parentheses:

SELECT D#, (SELECT E#

FROM Emp E

WHERE E.D#=D.D#) AS Emps

53

FROM Dept D

Scalar subquery or nested table ? I.e., is Emps an employee number or a set of employee numbers? (loosely speaking)

(Answer: an employee number)

Why The Flaws Are "Fatal" The Shackle of Compatibility (existing syntax cannot be deleted) The Growth of Redundancy (plugging holes gives new solutions where existing solutions already available) Desired extensions can be difficult or impossible to specify

(e.g., because of nulls)

Also shackled by existing style

Errors Here to Stay

- * Duplicate Rows
- * NULL and 3VL
- * Columns being ordered
- * SELECT-FROM-WHERE
- * Table types not data types

(and probably many others)

The Growth of Redundancy

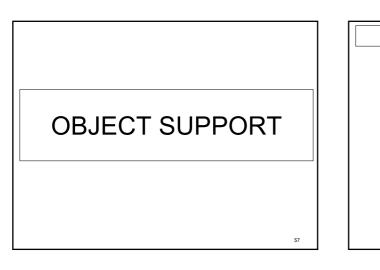
Since SQL:1992, the following features (e.g.) have been redundant:

- subqueries
- correlation names
- · doing joins in longhand
- the HAVING clause
- the GROUP BY clause

56

58

60



Object Oriented Databases

A couple of good ideas:

- Database variable types available for local variables too (but a relational database *must* be restricted to *relation* variables)
- · User-defined types (classes)

A questionable idea:

 Type inheritance via extension rather than by specialisation (e.g., 3D_POINT subtype of 2D_POINT ???)

Bad ideas (for relational database purposes):

- "Persistence is orthogonal to type" (see first good idea)
- Object identifiers (because they are *pointers*)
 Class extents (when used for the purpose of relations)
 Operator ("method") definitions bundled with type definitions
 "Selfish methods" (I.e., the "distinguished parameter")

Rapprochement

A bringing together of objects and relations. Widely sought, because:

- * Some Objectlanders wanted to be able to do what Relationlanders do with tables - specially ad hoc gueries and declarative constraints.
- * Some Relationlanders wanted to do some more complicated things that require user-defined data types of arbitrary complexity.

55

A Multimedium Relation							
BirdName	Info	Pic	Video	Song	Migr		
Robin		Ş	, CO,		Q		
Thrush		Ş			Ø		
Sparrow		Ş			Ø		
Pic is a pi	: ormation an icture of Bi a video of E	rdName,	and	ie, and			

Song is BirdName's song, and

Migr is BirdName's migration route.

SQL for The Bird Table

Correct (and supported):

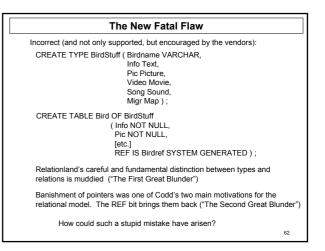
CREATE TYPE Text (T CLOB); CREATE TYPE Picture (P ROW (R INT, G INT, B INT) ARRAY); CREATE TYPE Sound (...) CREATE TYPE Movie (Vid PICTURE ARRAY, Aud SOUND); CREATE TYPE Map (...) ;

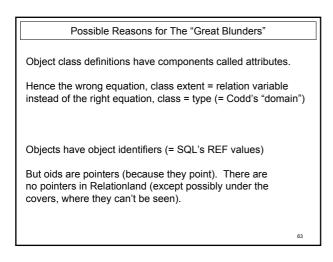
CREATE TABLE Bird (Birdname VARCHAR(30) NOT NULL, Info Text NOT NULL, Pic Picture NOT NULL, Video Movie NOT NULL, Song Sound NOT NULL, Migr Map NOT NULL,

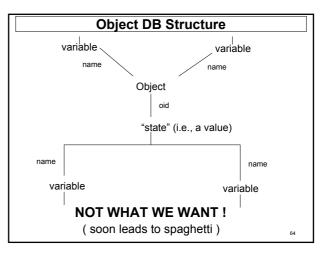
PRIMARY KEY Birdname) ;

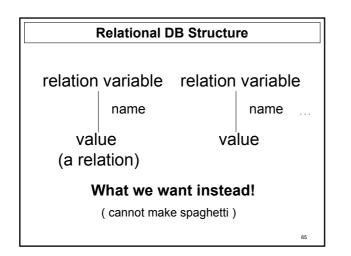
61

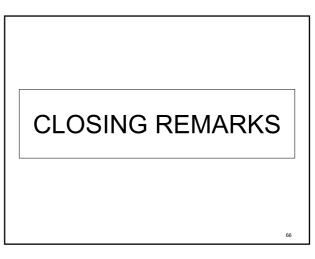
All fine and reasonable so far. $\operatorname{BUT}\ldots$











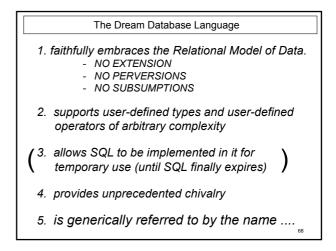
The Relationlander's Promise

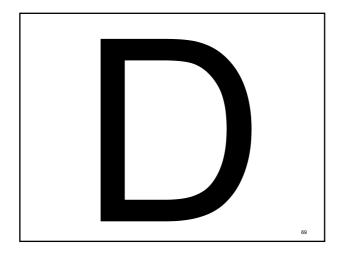
I solemnly promise ...

... *never* to use the word "relational" when I mean SQL, ...

67

... cross my heart and hope to die.





Projects based on TTM Rel: an implementation of Tutorial D by Dave Voorhis, University of Derby D4: a commercial implementation by Alphora, Utah, USA front-end to SQL (so includes nulls, alas) Duro: a C API for relational algebra and transactions, by René Hartmann Further info at www.thethirdmanifesto.com

Some Guiding Principles

Some important principles that we have become particularly conscious of, for various reasons.

Some have always been with us.

Some arise from a retrospective look at our manifesto.

Some may even be said to have informed our manifesto.

Logical Differences

Principle #1 (our motto)

"All logical differences are big differences" (Wittgenstein)

So all logical mistakes are big ones!

And we think all non-logical differences are small ones. In the database context, at least.

Values and Variables

Principle #2

"We retain and embrace a clear distinction between values and variables"

(Object Orientation seems to have blurred this distinction.)

Data Types and Relations Principle #3

Data types and the relational model are orthogonal to each other.

Corollary :

The relational model has no jurisdiction concerning which data types a relational system should support.

(except we must have relation types, and BOOLEAN!)

74

Types are Not Tables Principle #4

Types are to tables as nouns are to sentences!

So we cannot accept the equation object class = relation that some ORDBMSs (and SQL!) attempted to embrace.

"object class = domain" works fine. 75

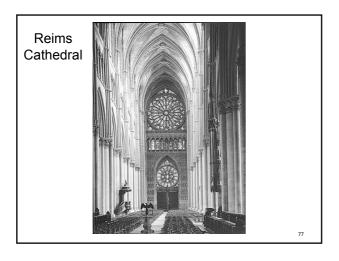
Conceptual Integrity Principle #5

"Conceptual integrity is *the* most important property of a software product" (Fred Brooks, 1975)

Of course, you must *have* concepts before you can be true to any. These had better be:

a.few

b.agreeable to those invited to share them





"This above all: to thine own self be true, And it must follow, as the night the day, Thou canst not then be false to any user."

(from Polonius's advice to D, by WS with HD)

в

Blank Slide

(material from lecture not suitable for handouts)

