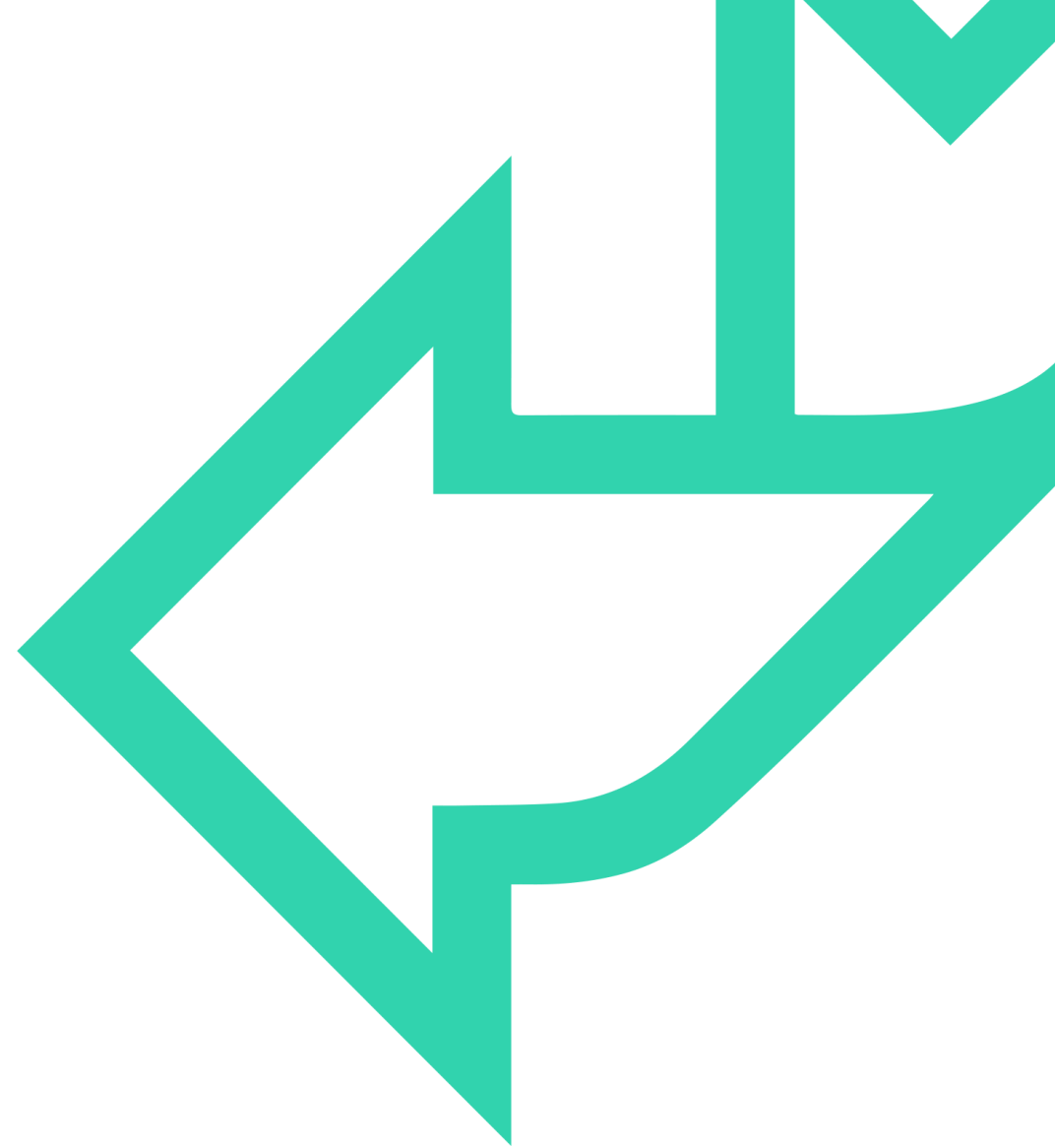




Databases and SQL

Tutor: Rob Menzies
Email: rob.menzies@qa.com





Data Normalisation



THE NEED FOR NORMALISATION

Some Data has a Simple Structure:

- Customer
- Name, Address, Telephone, Balance

Some Data has a Complex Structure:

- But what if Customer has more than one address?
- Home phone, Mobile, business phone....





CUSTOMERS AND ORDERS

QA Games

22 Gamesville Road,
Gamesville,
GM1 8GG

To:

Games-R-Us,
45 Shoppington Ave.
Shoppington
Bucks,
SH8 2AB

ORDER FORM

Order No:100001

Date: 31/07/2011

Customer No: 8573

Quantity	Product ID	Description	Unit Price	Amount
2	849	Alpine Ski Instructor	29.99	£59.98
28	773	SIM Accountant	23.46	£656.88
6	382	Recorder Hero – Legends of Folk	19.99	£119.94
Total Due				£836.80



CUSTOMERS AND ORDERS EXCEL STYLE

How many copies of Alpine Ski Instructor
did *Games-R-Us* order in 2011

	A	B	C	D	E	F	G	H		J	K	L	M	N
1	Ord. No.	Date	Cust. No.	Cust. Name.	Cust. Address	Cust. Town	Cust. County	Cust. PostCode	Prod. ID	Prod. Desc.	Prod. Qty.	Prod. Price	Prod. Amount	Ord. Total
2	100001	31/07/2011	8573	Games-R-Us	45 Shoppingt	Shoppingtc	Bucks.	SH8 2AB	849	Alpine Ski Instr	2	£29.99	£59.98	£836.80
3									773	SIM Accountan	28	£23.46	£656.88	£836.80
4									382	Recorder Hero	6	£19.99	£119.94	£836.80
5	100002	31/07/2011	5644	Gamerz	304 The Mall	Mallville	Herts.	MV9 1AQ	849	Alpine Ski Instr	200	£29.99	£5,998.00	£6,772.50
6									562	Deck Chair Att	50	£15.49	£774.50	£6,772.50
7														

What are total sales of Deck Chair
Attendant Deluxe in the county of
Buckinghamshire?



NORMALISATION

Normalisation method:

- Formal and mechanical approach
- Rigorous, unlike Entity-Relationship method
- Useful for transforming existing files
- Check existing databases, validate extensions
- Lossless Decomposition

Benefits:

- Strategy for key and table selection
- Improved ad-hoc queries and interfaces
- Eliminates insertion anomalies
- Eliminates deletion anomalies
- Reduction in DB modification time
- Correct representation of physical database
- Problems show up early in the design





3 TYPES OF DATA ANOMALY



- **Update Anomalies** happen when the person charged with the task of keeping all the records current and accurate, is asked, for example, to change an employee's title due to a promotion. If the data is stored redundantly (repeated) in the same table, and the person misses any of them, then there will be multiple titles associated with the employee. The end user has no way of knowing which is the correct title.
- **Insertion Anomalies** happen when inserting vital data into the database is not possible because other data is not already there. For example, if a system is designed to require that a customer be on file before a sale can be made to that customer, but you cannot add a customer until they have bought something, then you have an insert anomaly. It is the classic "catch-22" situation.
- **Deletion Anomalies** happen when the deletion of unwanted information causes desired information to be deleted as well. For example, if a single database record contains information about a particular product along with information about a salesperson for the company and the salesperson quits, then information about the product is deleted along with salesperson information.



STEPS TO ACHIEVE 3NF SUMMARY

1NF - No repeating columns

- Add extra columns
- Data should have atomicity

2NF + all nonkey attributes must be dependent on the whole key

- Identify the key
- Identify **Functional Dependencies** (FDs)
- Identify columns that depend on only part of the key
- Create new tables for these columns and partial key

3NF + all nonkey attributes must only be dependent on the key

- Identify **Transitive Dependencies** on nonkey attributes
- Create new tables for those columns



TRANSITIVE DEPENDENCY

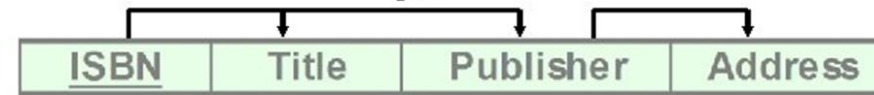
5

Example 1: Determine NF

- ◆ ISBN \rightarrow Title
- ◆ ISBN \rightarrow Publisher
- ◆ Publisher \rightarrow Address

Publisher is a non-key attribute, and it determines Address, another non-key attribute. Therefore, there is a transitive dependency, which means that the relation is NOT in 3 NF.

BOOK



8

ISBN uniquely identifies a specific title and edition of a book, with different ISBNs assigned to distinct formats (like hardcover or paperback), different languages, and different versions of the same title.