Data Normalization (1)

IS240 – DBMS

Lecture # 4 - 2010-02-01

M. E. Kabay, PhD, CISSP-ISSMP Assoc. Prof. Information Assurance School of Business & Management, Norwich University <u>mailto:mkabay@norwich.edu</u> V: 802.479.7937

Copyright © 2010 Jerry Post. All rights reserved

Why Normalization?



2

- > Need standardized data definition
 - □ Advantages of DBMS require careful design
 - Define data correctly and the rest is much easier
 - It especially makes it easier to expand database later
 - Method applies to most models and most DBMS
- Similar to Entity-Relationship
- Similar to Objects (without inheritance and methods)
- > Goal: Define tables carefully
 - Save space
 - □ Minimize redundancy
 - Protect data

3

1

Copyright © 2010 Jerry Post. All rights reserved

Keys



- Primary key
 - Every table (object) must have a primary key
 Uniquely identifies a row (one-to-one)
- Concatenated (or composite) key
 Multiple columns needed for primary key
 Identify repeating relationships (1 : M or M : N)
- Key columns are <u>underlined</u>
- > First step
 - Collect user documents
 - □ Identify possible keys
 - ✓ For each attribute_1 can there be many attribute_2? If so, make attribute_2 a key

Topics

- > Why Normalization?
- Definitions
- Client Billing Example
- > First Normal Form
- Second Normal Form
- > Third Normal Form
- > Checking Your Work (Quality Control)

This material was created by Prof Jerry Post and has been reformatted by Prof M. E. Kabay

NORWICH

NORWICH

Class: Employee



- > Relational database: A collection of *tables*.
- Table: A collection of columns (attributes) describing an entity. Individual objects are stored as rows of data in the table.

Copyright @ 2010 Jerry Post. All rights reserved

- Property (attribute): a characteristic or descriptor of a class or entity.
- > Every table has a primary key.
 - The smallest set of columns that uniquely identifies any row
 Primary keys can span more than one column (concatenated keys)
 - CustomerID, Product#, . . .) called a *surrogate key*.

Primary key Properties



Copyright © 2010 Jerry Post. All rights reser



CustomerID	Phone	LastName	FirstName	Address	City	State	Zipcode
1	502-666-7777	Johnson	Martha	125 Main Street	Alvaton	KY	42122
2	502-888-6464	Smith	Jack	873 Elm Street	Bowling Green	KY	42101
3	502-777-7575	Washington	Elroy	95 Easy Street	Smith's Grove	KY	42171
4	502-333-9494	Adams	Samuel	746 Brown Drive	Alvaton	KY	42122
5	502-474-4746	Rabitz	Victor	645 White Avenue	Bowling Green	KY	42102
6	616-373-4746	Steinmetz	Susan	15 Speedway Drive	Portland	TN	37148
7	615-888-4474	Lasater	Les	67 S. Ray Drive	Portland	TN	37148
8	615-452-1162	Jones	Charlie	867 Lakeside Drive	Castalian Springs	TN	37031
9	502-222-4351	Chavez	Juan	673 Industry Blvd.	Caneyville	KY	42721
10	502-444-2512	Roio	Maria	88 Main Street	Cave City	KY	42127

6

Copyright © 2010 Jerry Post. All rights reserved.

ights reserved.



Identifying Key Columns



NORWICH

OrderID Item Quantity 8367* ·229 2 8367 253 4 8367 876 1 8368 555 4 8368 229 1

Orders

Date

5-5-04

5-6-04

OrderItems

OrderID

8367

8368

7

9

part of the key. We can uniquely identify each order simply by its OrderID. Each order has many items. Each item can appear on many orders. So OrderID and Item are both part of the key. We need both the OderID & the Item to uniquely identify a

Copyright © 2010 Jerry Post. All rights reserved

row in the order.



8

Surrogate Keys

Example: Customer

numbers change.

with duplicate values)

Access: AutoNumber

SQL Server: Identity

mechanisms.

> Real world keys sometimes cause problems in a database.

□ E.g., using lastname_firstname as a key does not work!

□ Avoid phone numbers: people may not notify you when

authorized to ask for verification, so you could end up

Oracle: Sequences (but require additional programming)

Drawback: Numbers are not related to any business data, so

the application needs to hide them and provide other look up

Copyright @ 2010 Jerry Post. All rights reserved

Assumptions about uniqueness are often wrong

Avoid SSN (privacy and most businesses are not

Often best to let the DBMS generate unique values

NORWICH

Copyright © 2010 Jerry Post. All rights reserved







