Introduction to Database Concepts
&
Using SQLite with R

#RLadiesIstanbul

Mehmet Ali Ekmiş

@Obase, Project Engineer

@İstanbul Üniversitesi, Econometrics (BSc)

linkedIn.com/in/maliekmis

twitter.com/maliekmis

1() Introduction

2() The History of Database

3() Databases, Why?

4() Relational Model

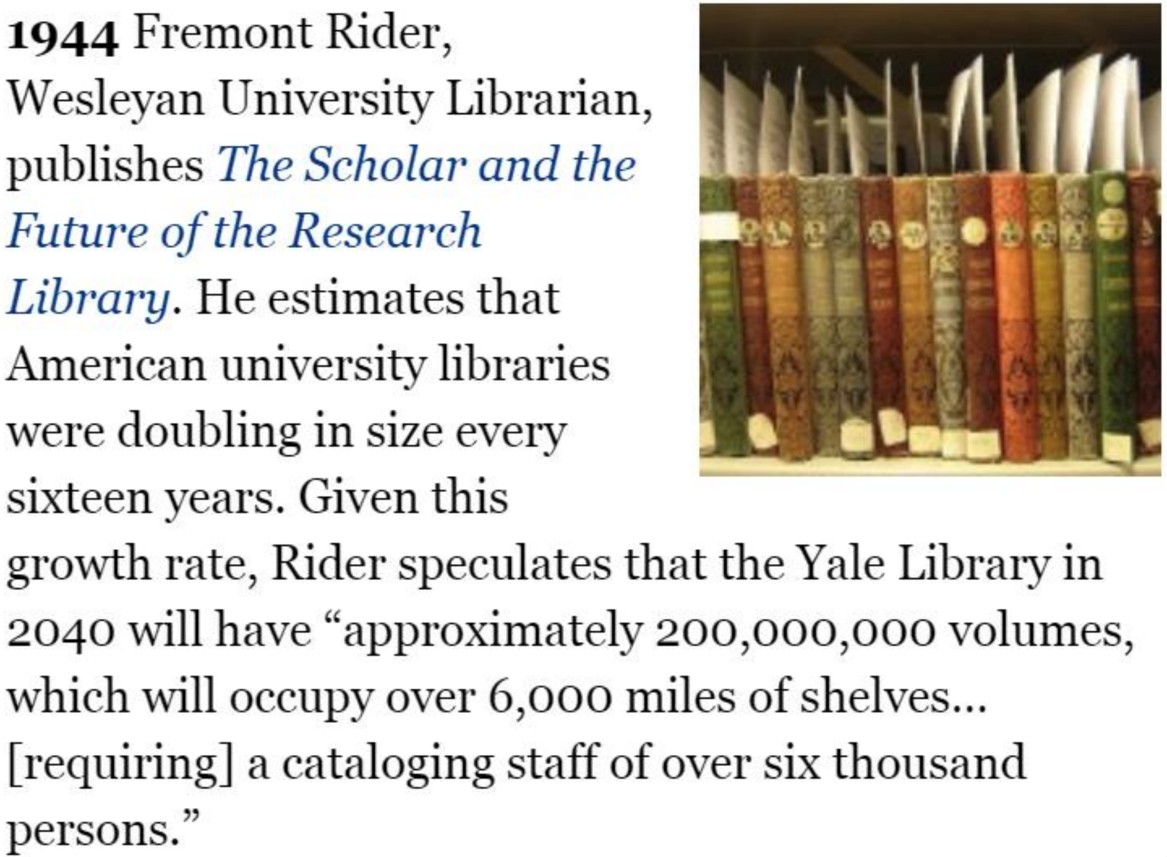
5() Structured Query Language (SQL)

6() Smart, Fast, Reliable: SQLite

7() Demo: SQLite with R

A Data! Data! Data! I can’t make bricks without clay!

- Sir Arthur Conan Doyle

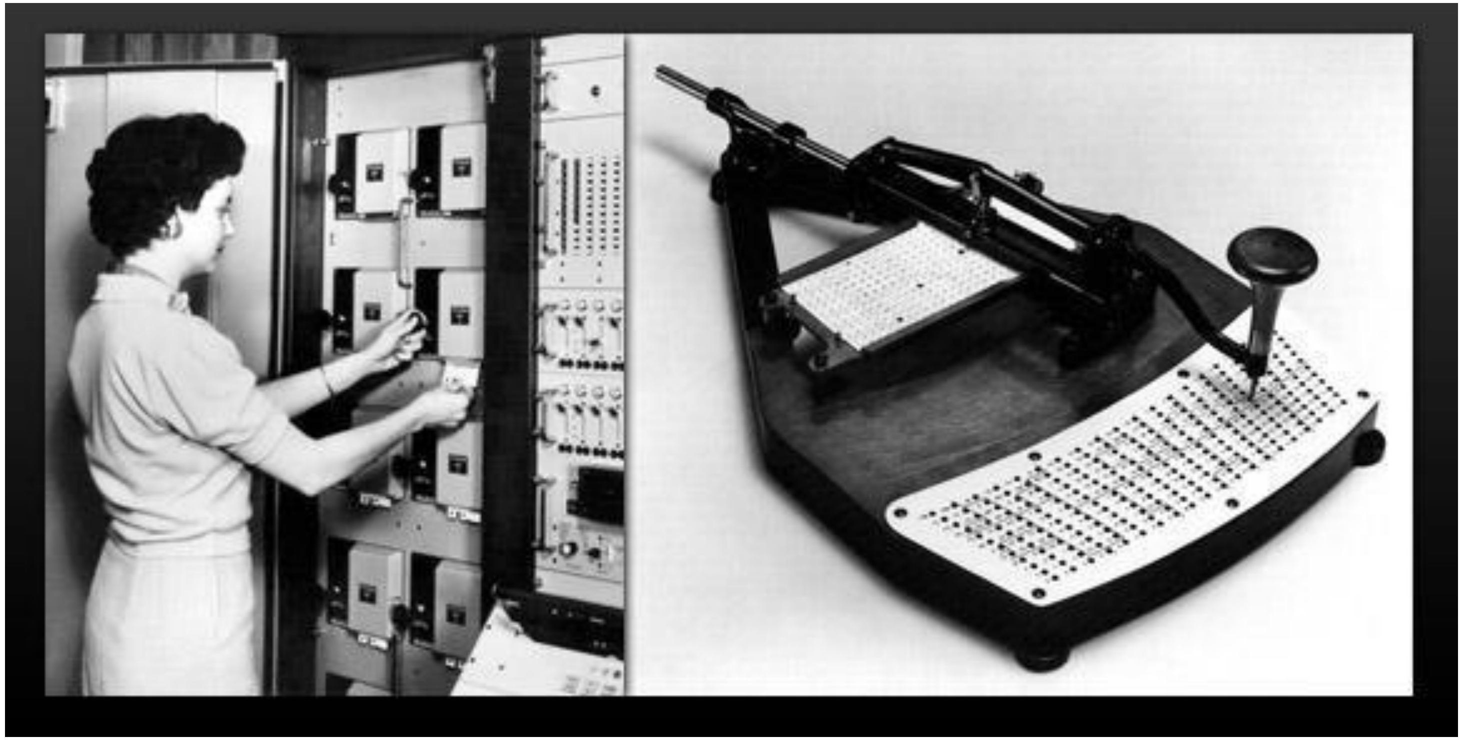


1944 Fremont Rider, Wesleyan University Librarian, publishes *The Scholar and the Future of the Research Library*. He estimates that American university libraries were doubling in size every sixteen years. Given this growth rate, Rider speculates that the Yale Library in 2040 will have "approximately 200,000,000 volumes, which will occupy over 6,000 miles of shelves… [requiring] a cataloging staff of over six thousand persons."

Forbes, December '13

Database; A shared collection of logically related data, designed to meet the information needs of an organization.

The History of Database



Punched card reader (L) and writer, Image from A Brief History of Communication Technology, 1910s

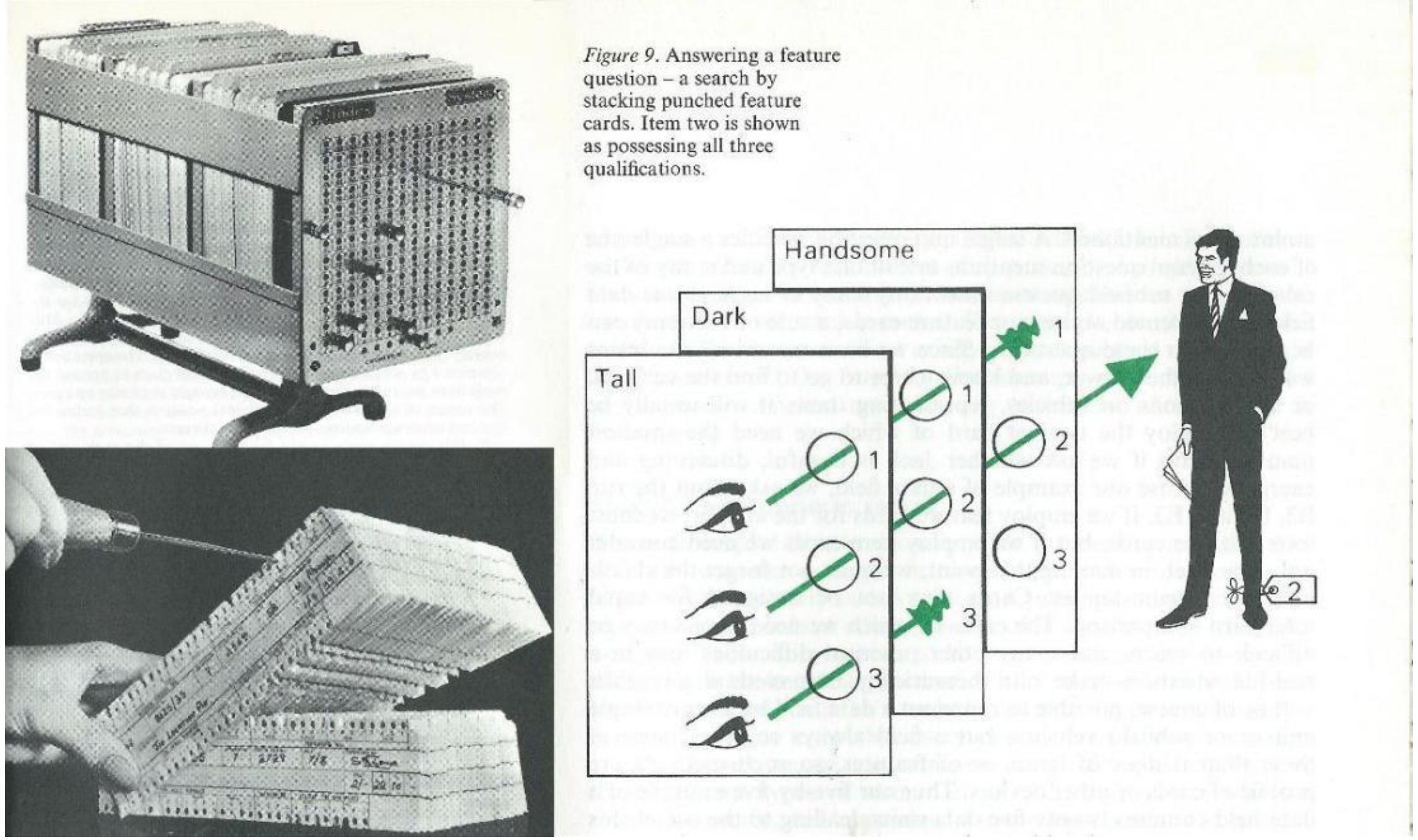
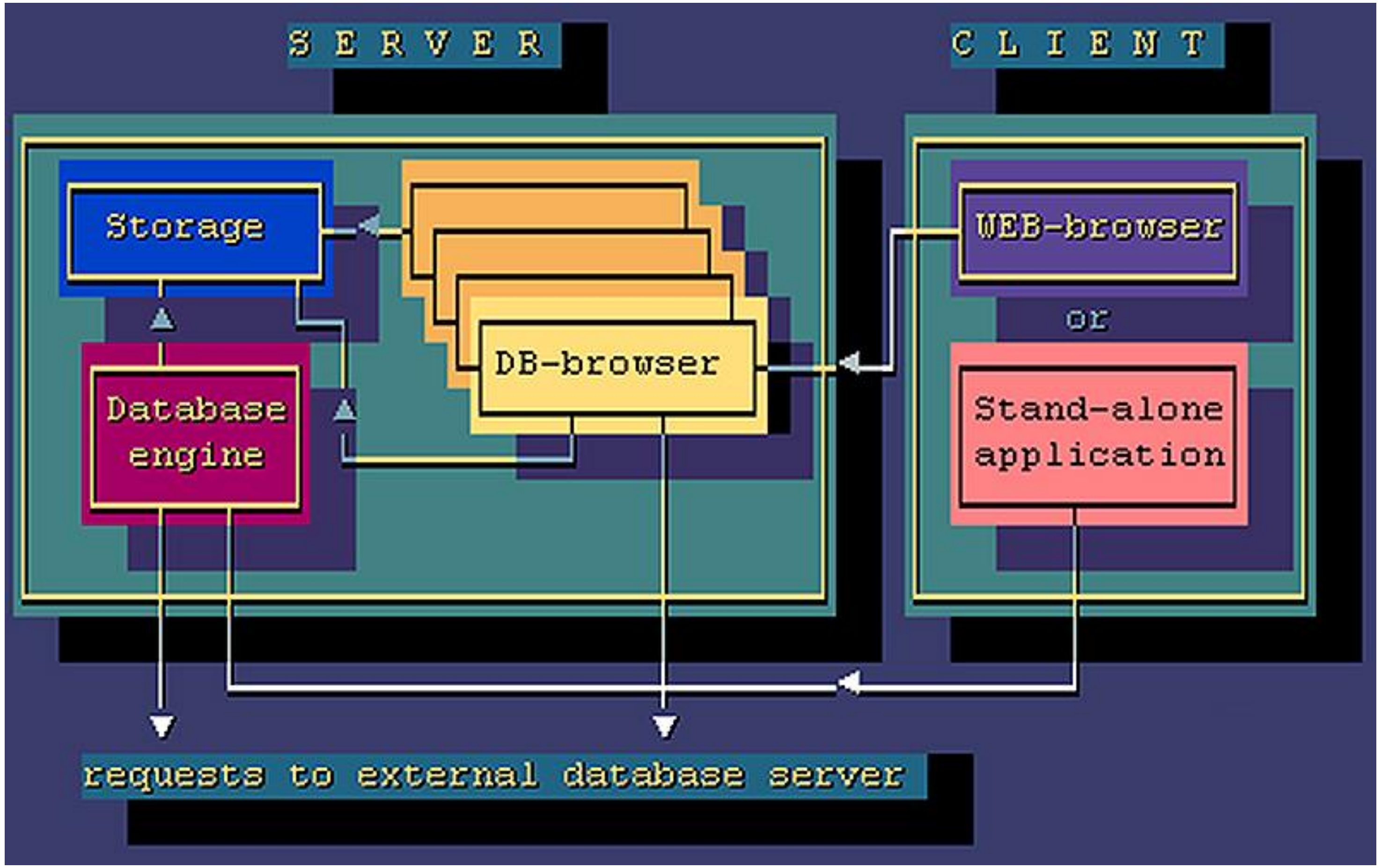
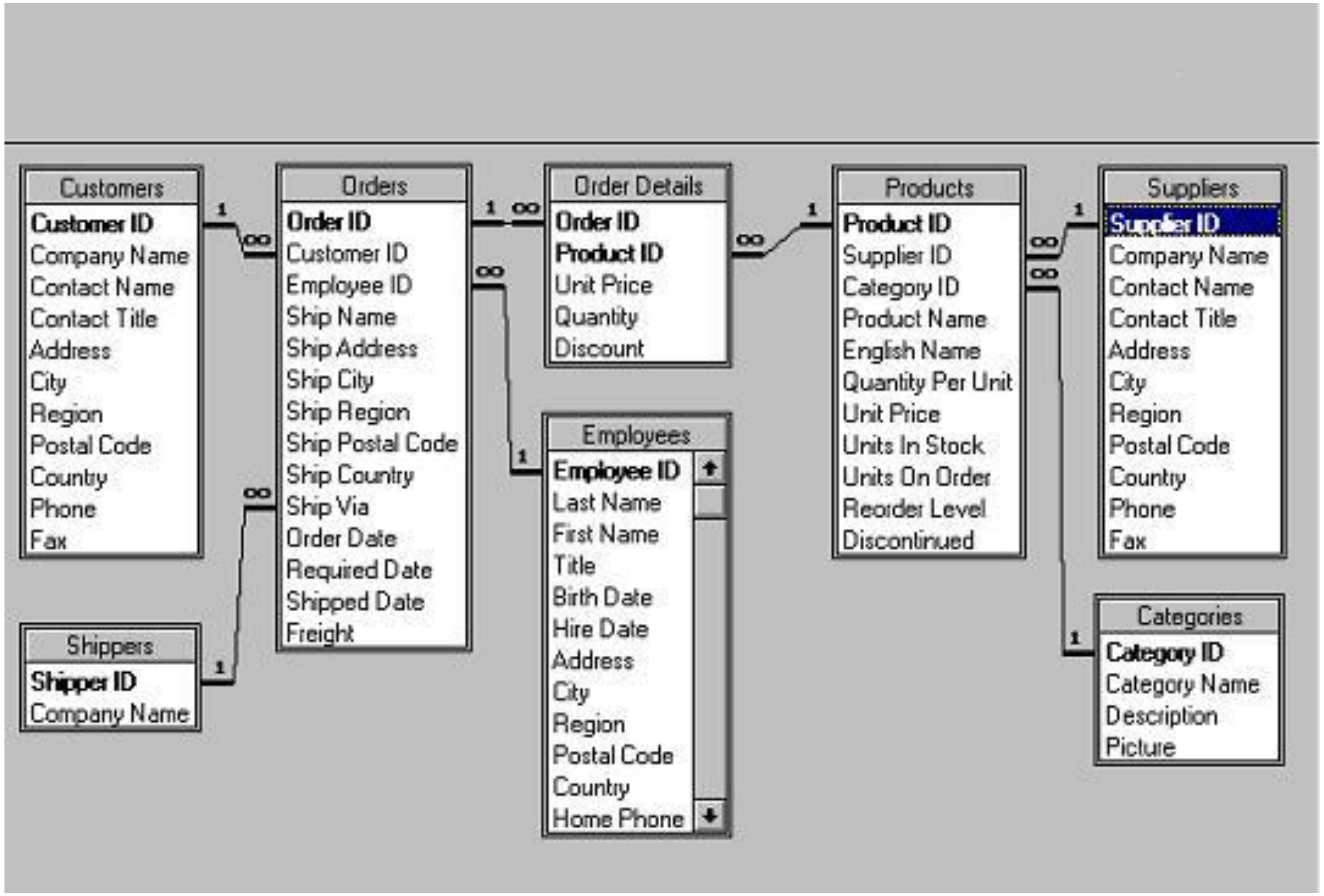


Figure 9. Answering a feature question – a search by stacking punched feature cards. Item two is shown as possessing all three qualifications.

Punch cards and tabulating used by office environment from 1910 to the mid-1960s



Database management system (DBMS), 1960s



Customers
Customer ID
Company Name
Contact Name
Contact Title
Address
City
Region
Postal Code
Country
Phone
Fax

Orders
Order ID
Customer ID
Employee ID
Ship Name
Ship Address
Ship City
Ship Region
Ship Postal Code
Ship Country
Ship Via
Order Date
Required Date
Shipped Date
Freight

Order Details
Order ID
Product ID
Unit Price
Quantity
Discount

Products
Product ID
Supplier ID
Category ID
Product Name
English Name
Quantity Per Unit
Unit Price
Units In Stock
Units On Order
Reorder Level
Discontinued

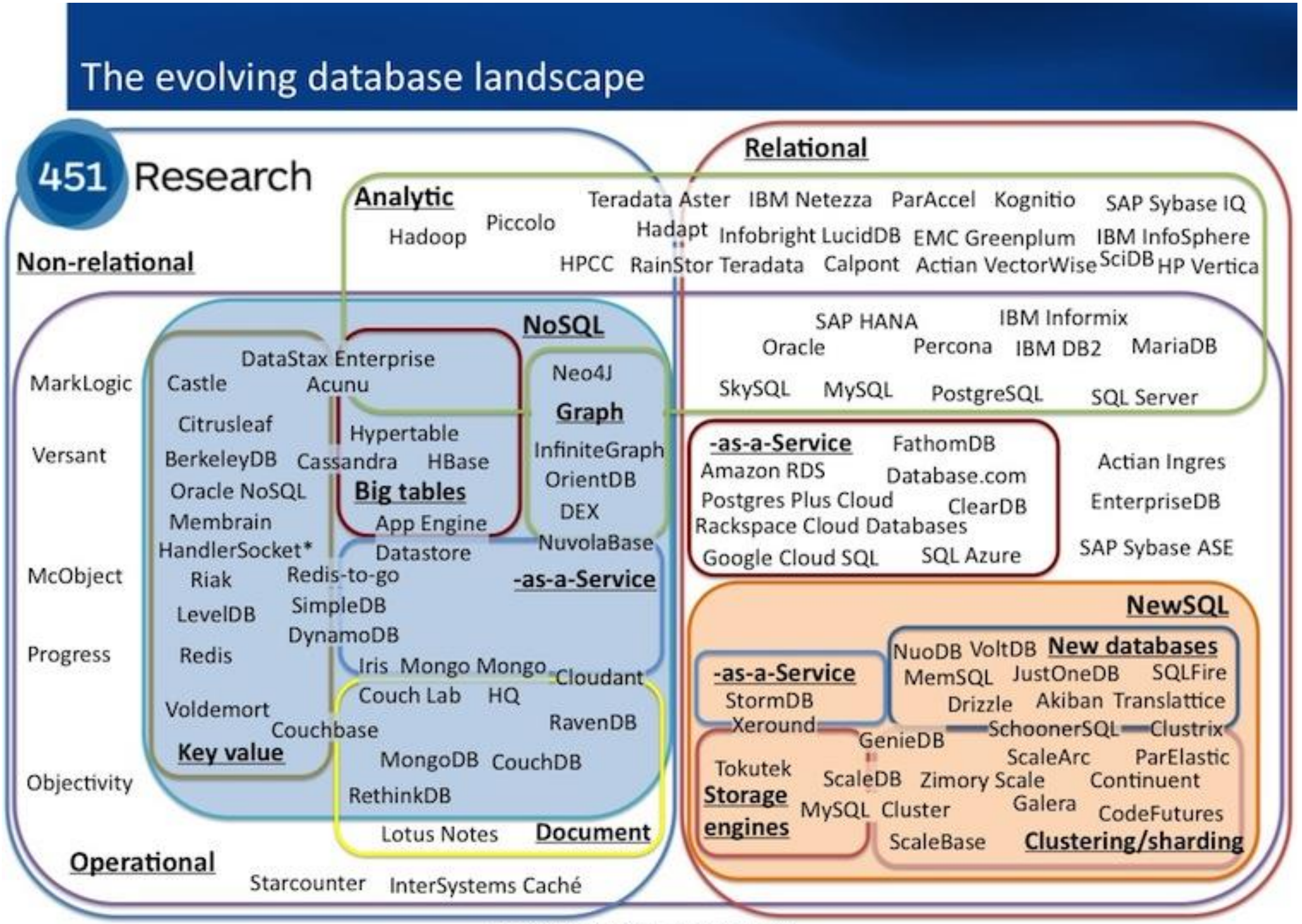
Suppliers
Supplier ID
Company Name
Contact Name
Contact Title
Address
City
Region
Postal Code
Country
Phone
Fax

Shippers
Shipper ID
Company Name

Employees
Employee ID
Last Name
First Name
Title
Birth Date
Hire Date
Address
City
Region
Postal Code
Country
Home Phone

Categories
Category ID
Category Name
Description
Picture

Relational Database Management System (RDBMS), 1970s



© 2012 by The 451 Group. All rights reserved

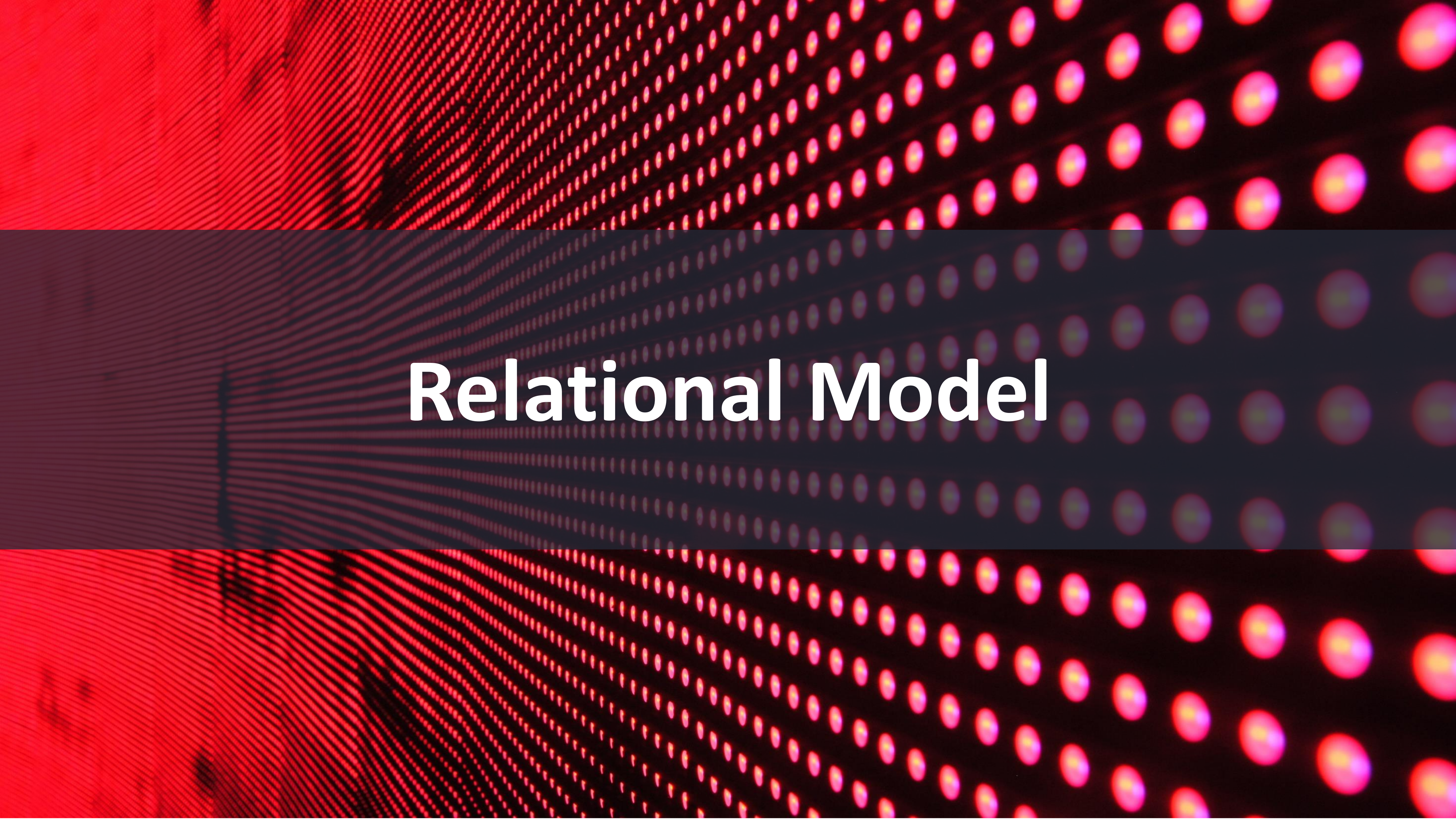
Database Landscape

Databases, Why?

- Can store very large numbers of records efficiently
- Very quick and easy to find information.
- Easy to add new data and to edit or delete old data.
- Data can be searched easily, eg 'find all Ford cars'.
- Data can be imported into other applications.
- Users can access the same database at the same time.
- Security be better than in paper files.

Types / uses of data

- Logistics and payments
- Retail sales, inventories
- Customer service
- Access control
- Navigation / tracking
- Social media events
- Proximity awareness
- Sport performance
- Credit risk
- Geospatial
- Proximity awareness
- and much more...



Relational Model

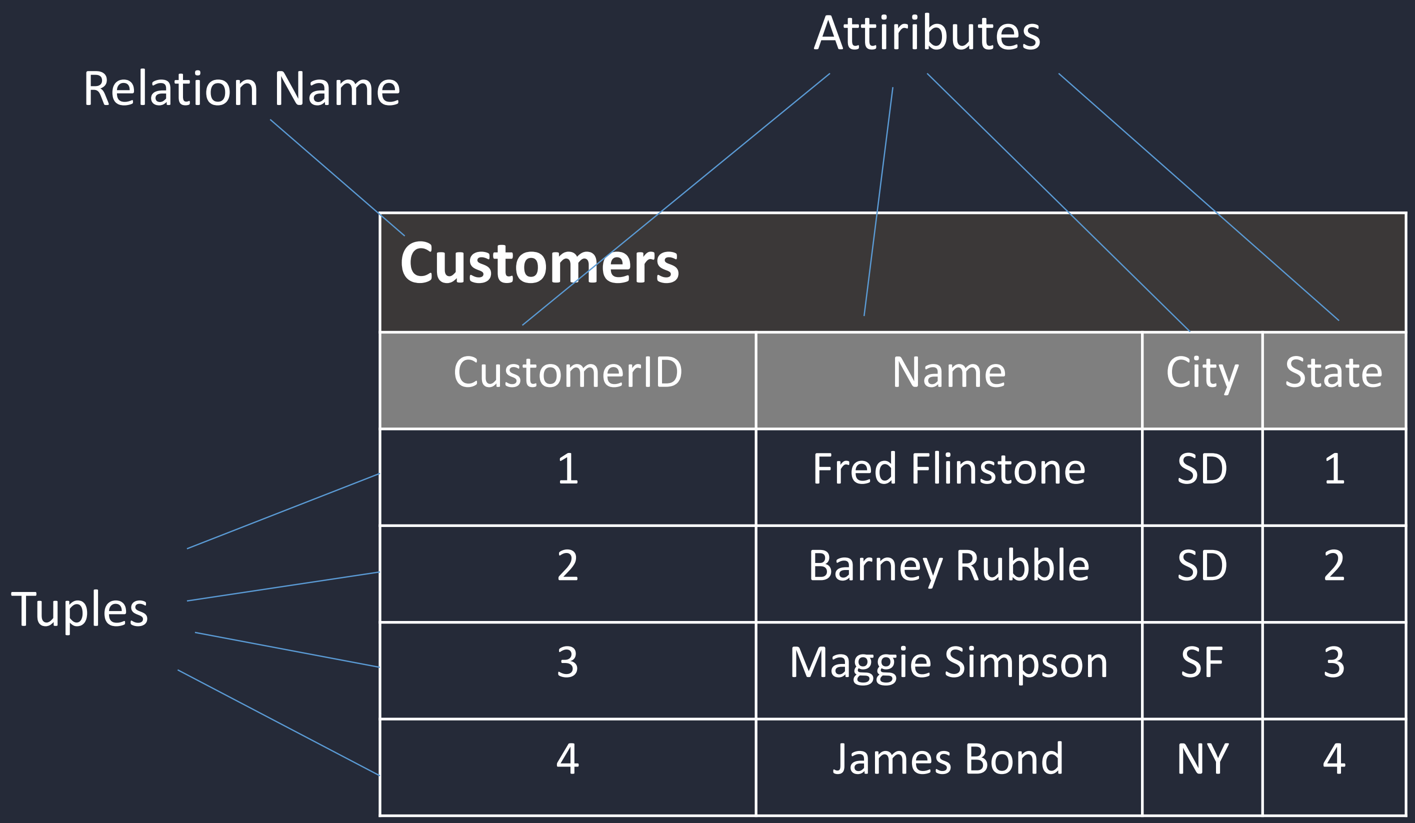
Relational model, is defined as a model that allows you to group its data items into one or more independent tables that can be related to one another by using fields common to each related table.

- A database consists of several **tables** (relations)

Customers			
CustomerID	Name	City	State

Account	
AccountNum	Name

- Columns in the tables are named by **attributes**
- Each attribute has an associated **domain** (e.g. for Customer.State: {CA, NY, WA, …})
- Data in a table consist of a set of **tuples** (rows)



Attiributes

Relation Name

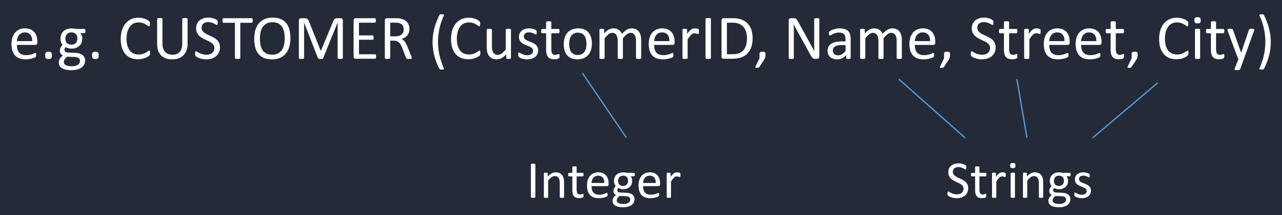
Tuples

Customers

CustomerID	Name	City	State
1	Fred Flinstone	SD	1
2	Barney Rubble	SD	2
3	Maggie Simpson	SF	3
4	James Bond	NY	4

Relational Schema

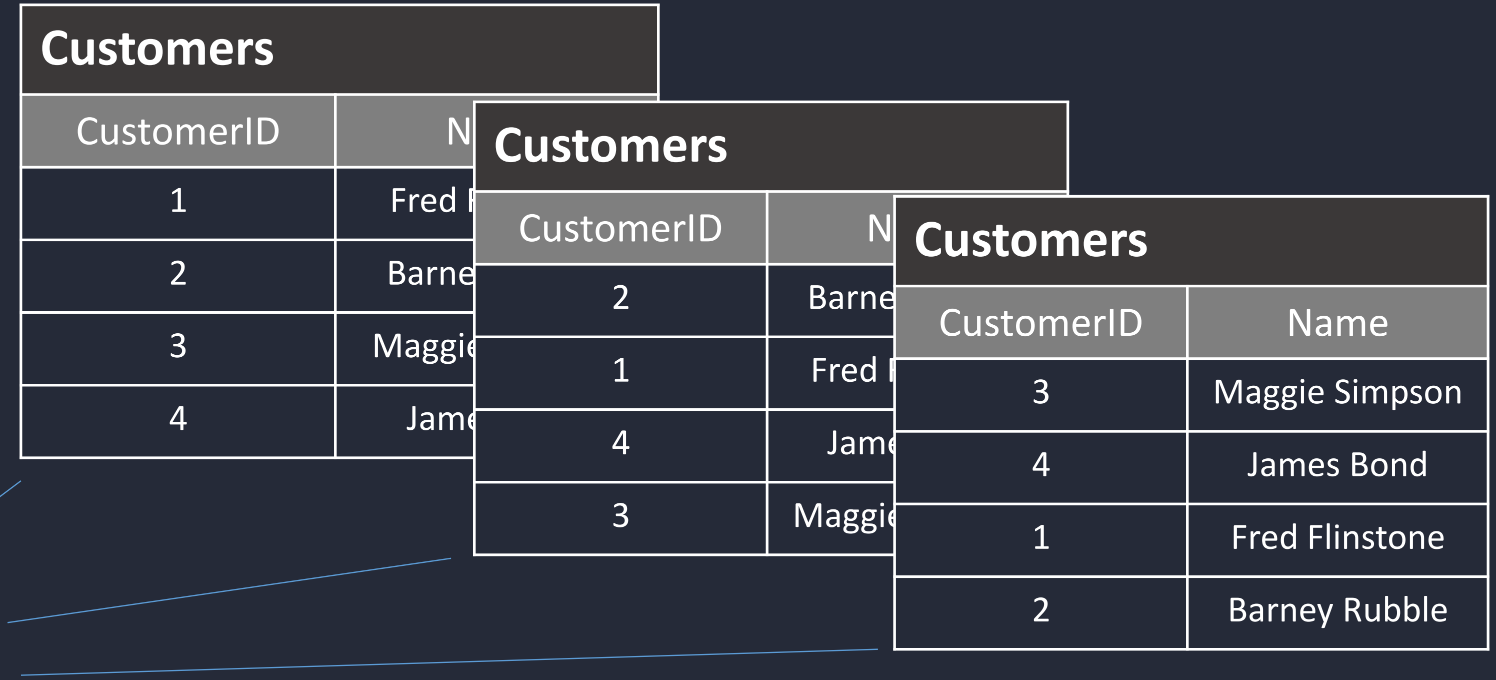
- Type declaration
- Consists of:
  - Relation name
  - Set of attributes
  - Domain of each attribute
  - Integrity constraints



e.g. CUSTOMER (CustomerID, Name, Street, City)

Relations are Unordered

- The tuples are not considered to be ordered, even though they appear to be so when displayed in tabular form



Customers

CustomerID	N
1	Fred F
2	Barne
3	Maggie
4	Jame

Customers

CustomerID	N
2	Barne
1	Fred F
4	Jame
3	Maggie

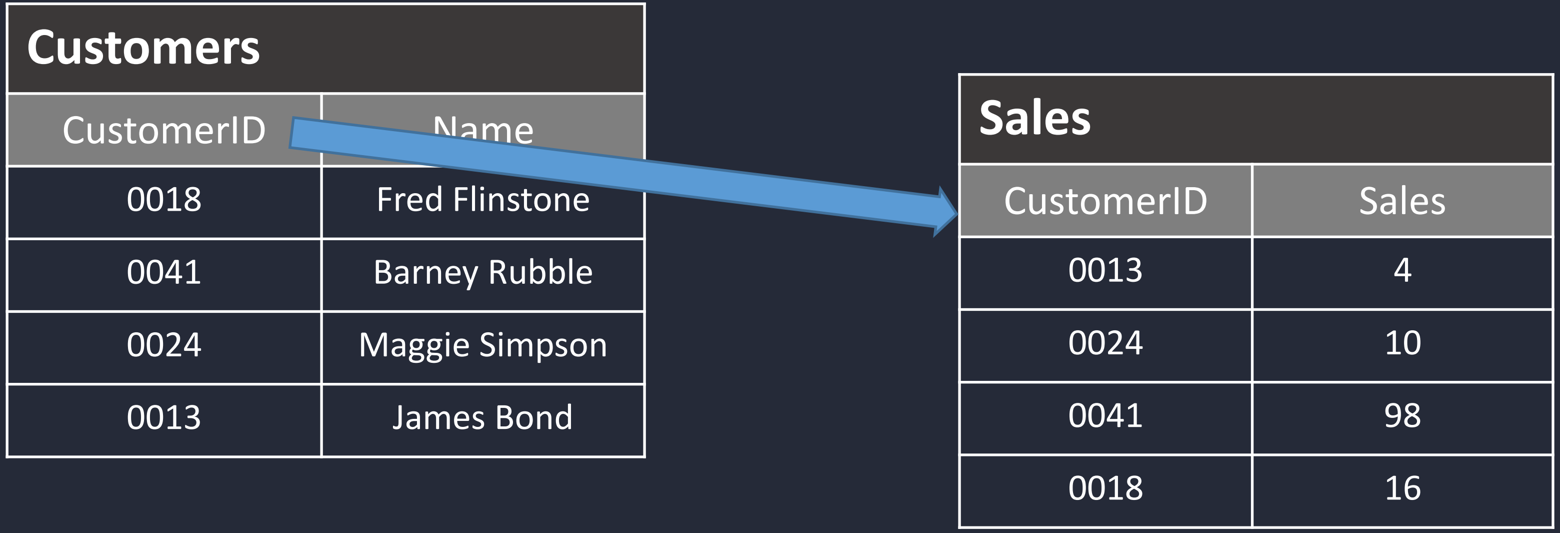
Customers

CustomerID	Name
3	Maggie Simpson
4	James Bond
1	Fred Flinstone
2	Barney Rubble

Visual representations of the same relational instance

Relational Model - Integrity Constraints

- Keys
- Primary Keys
- Entity Integrity
- Referential Integrity



Customers

CustomerID	Name
0018	Fred Flinstone
0041	Barney Rubble
0024	Maggie Simpson
0013	James Bond

Sales

CustomerID	Sales
0013	4
0024	10
0041	98
0018	16

Structured Query Language (SQL)

What is SQL?

- SQL stands for **Structured Query Language**.
- SQL lets you access and manipulate databases.
- SQL is an **ANSI** (American National Standards Institute) **standard**

SQL is a standart, but...

- Although SQL is an ANSI (American National Standards Institute) standard, there are many different versions of the SQL language. (PL/SQL, T-SQL ...)
- However, to be compliant with the ANSI standard, they all support at least the major commands (such as SELECT, UPDATE, DELETE, INSERT, WHERE) in a similar manner.

What SQL can do?

- SQL can execute queries against a database.
- SQL can retrieve data from a database.
- SQL can insert tupples in a database.
- SQL can update tupples in a database.
- SQL can delete tupples from a database.
- SQL can create new databases.
- SQL can create new tables in a database.
- SQL can set permissions on tables, procedures, and views
- and much more...

SQL Parts (DML&DDL)

- SQL can be divided into two parts: **The Data Manipulation Language (DML), Data Definition Language (DDL)**

Category	SQL Command	Description
Data Definition Language (defines the different structures in a database)	**CREATE**	**Creates a new table, a view of a table, or other object in the database.**
	ALTER	**Modifies an existing database object, such as a table.**
	DROP	**Deletes an entire table, a view of a table or other objects in the database.**
Data Manipulation Language (manipulates data in a database)	**SELECT**	**Retrieves certain records from one or more tables.**
	INSERT	**Creates a record.**
	UPDATE	**Modifies records**
	DELETE	**Deletes records**

basic SELECT statement

```
SELECT *|{ [DISTINCT] column|expression [alias],...]
FROM table;
```

- **SELECT** identifies *what* columns
- **FROM** identifies *which* tables

Smart, Fast, Reliable: SQLite

SQLite is a public-domain software package that provides a *relational database management system*, or RDBMS.

SQLite is defined by the following features

- Serverless
- Zero Configuration
- Cross-Platform
- Self-Contained
- Small Runtime Footprint
- Transactional
- Full-Featured
- Highly Reliable

Demo: SQLite with R

Thank you.

in linkedIn.com/in/maliekmis

twitter.com/maliekmis