



# **Introduction to Databases**

## **MSc Bootcamp**

**National College of Ireland**  
**Dublin, Ireland**

**Hamilton Niculescu**  
[Hamilton.Niculescu@ncirl.ie](mailto:Hamilton.Niculescu@ncirl.ie)

# Outline

---

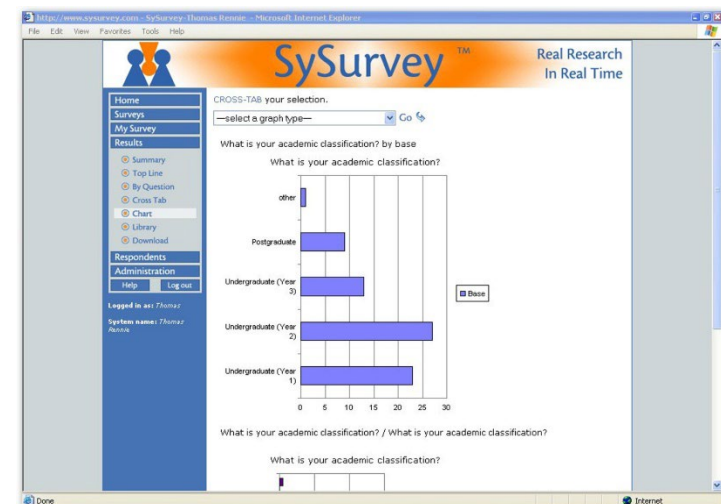
- **File-Based systems and needs for Databases**
- **What is a Database?**
- **Typical functions of a Database Management System (DBMS)**
- **Major components of the DBMS environment**
- **Personnel involved in the DBMS environment**
- **History of the development of DBMSs**
- **Advantages and disadvantages of DBMSs**

# Data vs. Information

- **Data:**
  - Raw facts; building blocks of information
  - Unprocessed information
- **Information:**
  - Data processed to reveal meaning
- Accurate, relevant, and timely information is key to good decision making
- Good decision making is the key to survival in a global environment

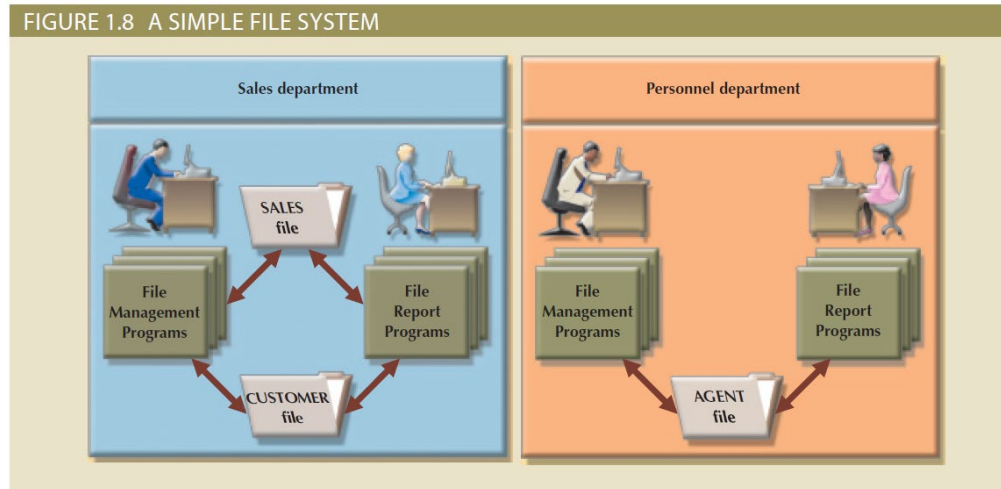
The screenshot shows a web browser window with the URL <http://www.ssysurvey.com>. The page title is "Business Computer Lab Satisfaction Survey". Below the title, there is a blue header bar with the same text. The main content area contains the following text: "This survey is designed to obtain student feedback regarding the services provided by the Business Computer Lab and identify areas in which we need to improve. Please answer each question as accurately as possible." Below this, there is a section titled "Using the Lab" with four questions:

1. What is your academic classification?  
☐ Undergraduate (Year 1) ☐ Undergraduate (Year 2) ☐ Undergraduate (Year 3) ☐ Postgraduate ☐ Other
2. Do you own a computer?  
☐ Yes ☐ No
3. How often do you use the Business Computer Lab?  
☐ Five or more times per week  
☐ Three or four times per week  
☐ Once or twice per week  
☐ Once a month or less
4. What do you primarily use the Business Computer Lab for? (You may check more than one)  
☐ Internet (i.e. Web Browsing, Chat)  
☐ Email Access  
☐ Word Processing (i.e. MS Word)  
☐ Spreadsheets (i.e. MS Excel)  
☐ Presentations (i.e. MS PowerPoint)  
☐ Academic Programs (i.e. MINITAB, SPSS etc)



# A Simple File System

FIGURE 1.8 A SIMPLE FILE SYSTEM



## BASIC FILE TERMINOLOGY

TERM	DEFINITION
Data	Raw facts, such as a telephone number, a birth date, a customer name, and a year-to-date (YTD) sales value. Data have little meaning unless they have been organized in some logical manner.
Field	A character or group of characters (alphabetic or numeric) that has a specific meaning. A field is used to define and store data.
Record	A logically connected set of one or more fields that describes a person, place, or thing. For example, the fields that constitute a record for a customer might consist of the customer's name, address, phone number, date of birth, credit limit, and unpaid balance.
File	A collection of related records. For example, a file might contain data about the students currently enrolled at Gigantic University.

# Sample Customer-Order Data

- Problems using spreadsheet

- **Redundancy**

- Duplication of data or the storing of the same data in more than one place

Diagram illustrating a sample customer order form with annotations for redundancy:

- Heading:** Contains order details (ORDER: 21617, PREMIERE PRODUCTS, DATE: 10/23/2013) and customer information (CUSTOMER: 608, Johnson's Department Store, 372 Oxford, Sheldon FL 33553). This information is repeated in the header and also appears in the order lines table.
- Order lines:** A table with columns: PART NUMBER, PART DESCRIPTION, NUMBER ORDERED, PRICE, and TOTAL. It shows two items: BV06 (Home Gym) and CD52 (Microwave Oven). The same customer information is repeated for each line item.
- Extensions:** A bracket indicates that the price and quantity are multiplied to get the total for each line item.
- Body:** The main table area containing the order lines.
- Footer:** Contains the ORDER TOTAL >> 2189.90.

- Difficulty accessing related data
- Limited security
- Size limitations

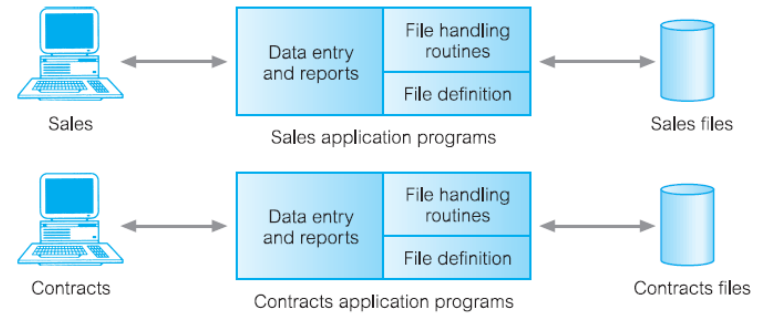
Orders requiring more than one spreadsheet row

Customer Number	Customer Name	Order Number	Order Date	Part Number	Part Description	Number Ordered	Quoted Price	Warehouse	Rep. Number
148	Al's Appliance and Sport	21608	10/20/2013	AT94	Iron	11	\$21.95	3	20
148	Al's Appliance and Sport	21619	10/23/2013	DR93	Gas Range	1	\$495.00	2	20
282	Brookings Direct	21614	10/21/2013	KT03	Dishwasher	2	\$595.00	3	35
356	Ferguson's	21610	10/20/2013	DR93	Gas Range	1	\$495.00	2	65
356	Ferguson's	21610	10/20/2013	DW11	Washer	1	\$399.99	3	65
408	The Everything Shop	21613	10/21/2013	KL62	Dryer	4	\$329.95	1	35
608	Johnson's Department Store	21617	10/23/2013	BV06	Home Gym	2	\$794.95	2	65
608	Johnson's Department Store	21617	10/23/2013	CD52	Microwave Oven	4	\$150.00	1	65
608	Johnson's Department Store	21623	10/23/2013	KV29	Treadmill	2	\$1,290.00	2	65

Sample orders spreadsheet

# File-Based Systems

<b>DreamHome</b> <b>Property for Rent Details</b> <b>Property Number:</b> <u>PG21</u>	
<b>Address</b> <u>18 Dale Rd</u> <b>City</b> <u>Glasgow</u> <b>Postcode</b> <u>G12</u> <b>Type</b> <u>House</u> <b>Rent</b> <u>600</u> <b>No of Rooms</b> <u>5</u>	<b>Allocated to Branch:</b> <u>163 Main St, Glasgow</u> <b>Branch No</b> <u>5003</u>  <b>Staff Responsible</b> <u>Ann Beech</u>
<b>Owner's Details</b>	
<b>Name</b> <u>Carol Farrel</u> <b>Address</b> <u>6 Achray St,</u> <u>Glasgow G32 9DX</u> <b>Tel No.</b> <u>0141-357-7419</u> <b>Owner No.</b> <u>CO87</u>	<b>Business Name</b> _____ <b>Address</b> _____ <b>Tel No.</b> _____ <b>Owner No.</b> _____ <b>Contact Name</b> _____ <b>Business Type</b> _____



## Sales Files

**PropertyForRent** (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)

**PrivateOwner** (ownerNo, fName, lName, address, telNo)

**Client** (clientNo, fName, lName, address, telNo, prefType, maxRent)

## Contracts Files

**Lease** (leaseNo, propertyNo, clientNo, rent, paymentMethod, deposit, paid, rentStart, rentFinish, duration)

**PropertyForRent** (propertyNo, street, city, postcode, rent)

**Client** (clientNo, fName, lName, address, telNo)

## PropertyForRent

propertyNo	street	city	postcode	type	rooms	rent	ownerNo
PA14	16 Holhead	Aberdeen	AB7 5SU	House	6	650	CO46
PL94	6 Argyll St	London	NW2	Flat	4	400	CO87
PG4	6 Lawrence St	Glasgow	G11 9QX	Flat	3	350	CO40
PG36	2 Manor Rd	Glasgow	G32 4QX	Flat	3	375	CO93
PG21	18 Dale Rd	Glasgow	G12	House	5	600	CO87
PG16	5 Novar Dr	Glasgow	G12 9AX	Flat	4	450	CO93


## PrivateOwner

ownerNo	fName	lName	address	telNo
CO46	Joe	Keogh	2 Fergus Dr, Aberdeen AB2 7SX	01224-861212
CO87	Carol	Farrel	6 Achray St, Glasgow G32 9DX	0141-357-7419
CO40	Tina	Murphy	63 Well St, Glasgow G42	0141-943-1728
CO93	Tony	Shaw	12 Park Pl, Glasgow G4 0QR	0141-225-7025

**A client approaches the Sales Department with a view to sale or rent their property**

# File-Based Systems

- Each department accessing their own files through **application programs** written specially for them
- Each set of departmental application programs handles **data entry, file maintenance,** and the generation of a fixed set of **specific reports**
- What is more important, the physical structure and storage of the data files and records are defined in the application code

 Sales by Date

Video Spot	
Sales by Date	
Date	
Tuesday, January 17, 2012	
Payment Type Cash	
SubTotal	\$12.00
Tax1	\$0.72
Tax2	\$0.12
Total	\$12.84
Payment Type Credit Card	
SubTotal	\$16.50
Tax1	\$0.99
Tax2	\$0.17
Total	\$17.66
GRAN TOTAL	
SubTotal	\$28.50
Tax1	\$1.71
Tax2	\$0.29
Total	\$30.50
Movies Rente	9
Movies Return	11

# Limitations of the File-Based approach

---

## 1. Separation and Isolation of Data

- Each program maintains its own set of data
- Users of one program may be unaware of potentially useful data held by other programs
- **For example:** if we want to produce a list of all houses that match the requirements of clients, we first need to create a temporary file of those clients who have 'house' as the preferred type

## 2. Duplication of Data

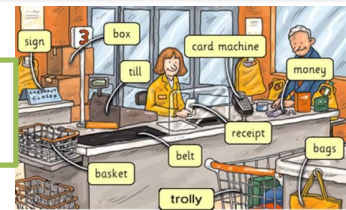
- Same data is held by different programs
- Wasted space and potentially different values and/or different formats for the same item
- **For example:** Payroll and Personnel departments. If a member of staff moves house and the change of address is communicated only to Personnel and not to Payroll, the person's payslip will be sent to the wrong address



# Database

- **A database is a coordinated or organised group of data**
  - **Data is known as facts that can be recorded and have an implicit meaning**
- 
- **The purpose of a database is to keep track of things**
  - **The database is a structure containing categories of information and relationships between these categories**
  - **Categories:** sales reps, customers, orders, parts, etc.
  - **Relationships between categories:** sales rep-to-customer and customer-to-orders
  - **We may say that it is a repository for a collection of computerized data files**

Purchases from the supermarket



Purchases using your credit/ debit card



Booking a holiday at the travel agents or online



Using the local library



# Database approach

- **Arose because:**

- Definition of data was embedded in the application programs, rather than being stored separately and independently
- No control over access and manipulation of data beyond that imposed by application programs

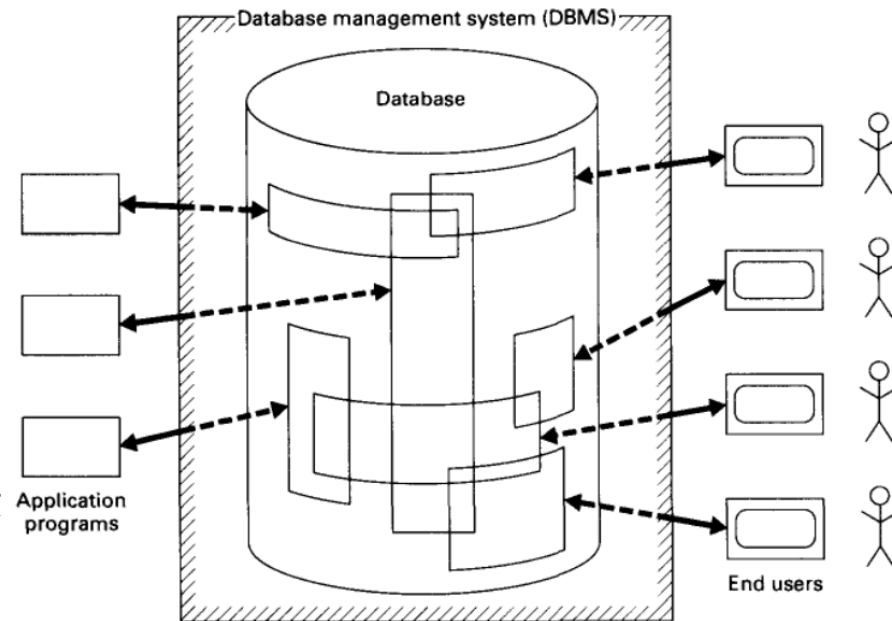


FIG. 1.4 Simplified picture of a database system

- **Result:**

- The Database and the Database Management System (DBMS).

An Introduction to Database Systems, C. J. Date, Addison-Wesley Publisher Company.

# Major Benefits of Database

- **Shared data**
- **Reduced redundancy**
- **Reduced inconsistent data**
- **Transaction support**
- **Support for data integrity**
- **Security enforcement**
- **Support for standards**
- **Conflicting requirements can be met**

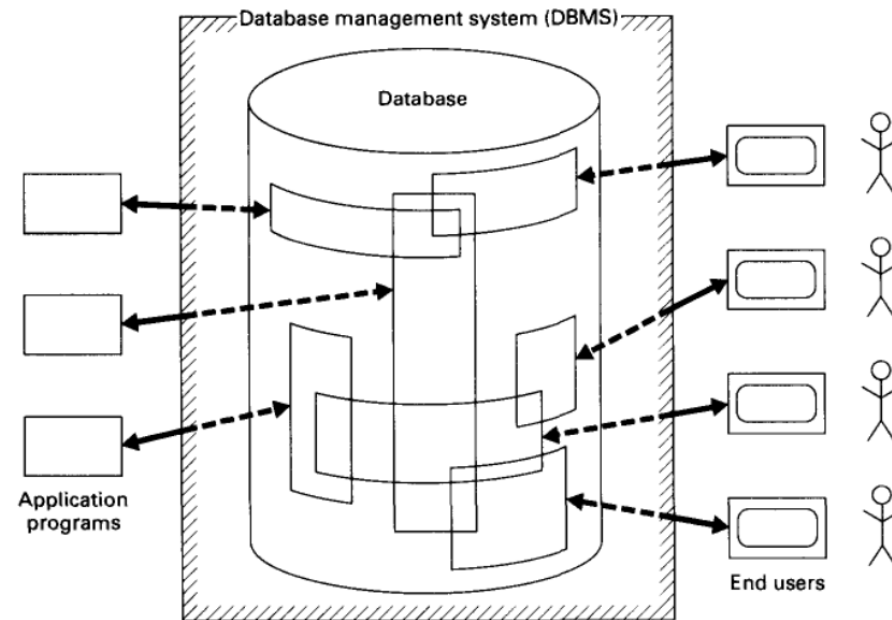


FIG. 1.4 Simplified picture of a database system

An Introduction to Database Systems, C. J. Date, Addison-Wesley Publisher Company.

# Database

---

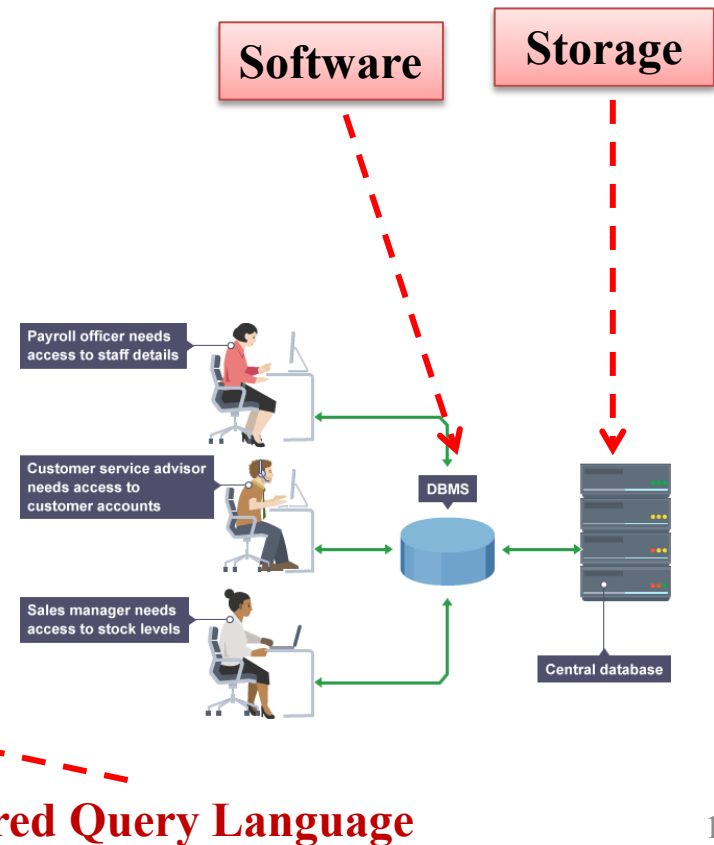
- **Shared collection of logically related data (and a description of this data), designed to meet the information needs of an organisation**
- **System catalogue (data dictionary) provides a description of data to enable the program**

## Data Independence

- **A database separates logical and physical representation of data**
- **Allows changes to application programs without changing the structure of the underlying data and vice versa**

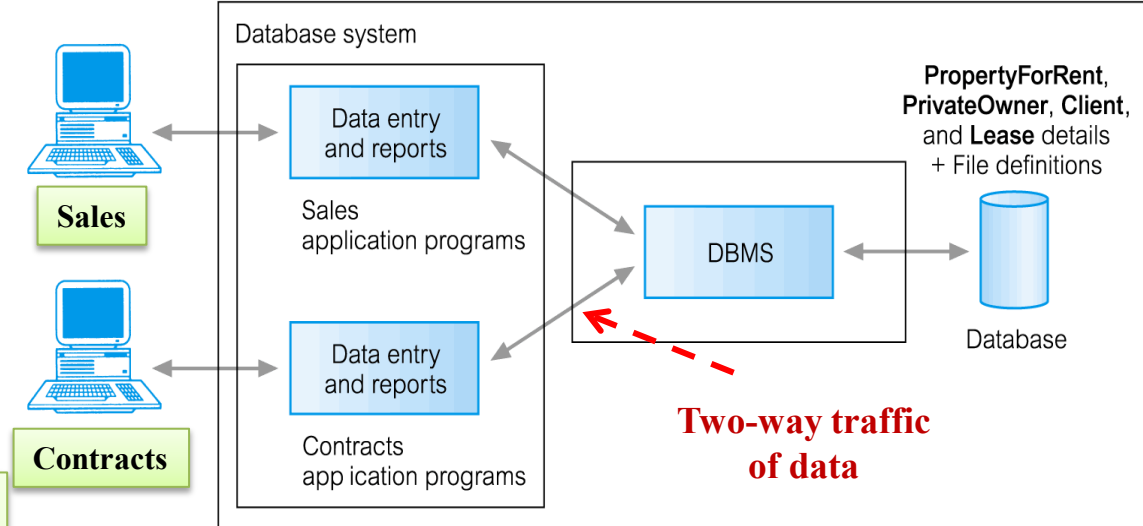
# Database Management System (DBMS)

- The DBMS is the software that interacts with the users' application programs and the database
- A software system that enables users to **define, create, maintain, and control access** to the database
- **Database application program:**  
A computer program that interacts with database by issuing an appropriate request (SQL statement) to the DBMS



# Database Management System (DBMS)

Data independence allows changes to be made to the physical data without affecting the logical level and also allows some changes to be made at the logical level without affecting the views of the users

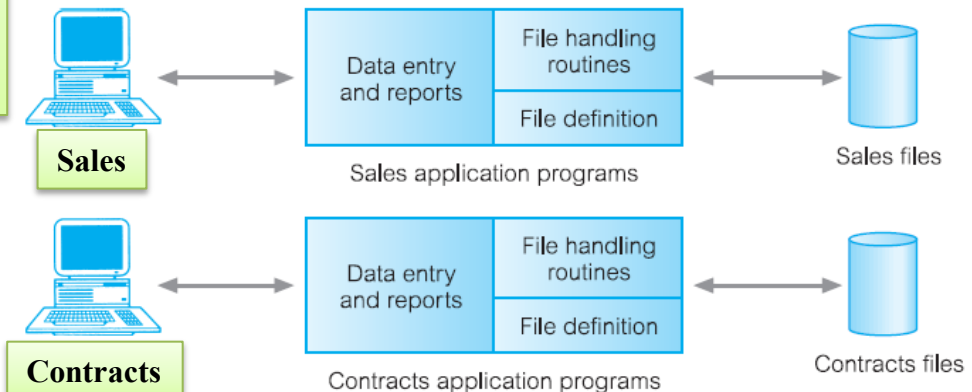


**PropertyForRent** (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)

**PrivateOwner** (ownerNo, fName, lName, address, telNo)

**Client** (clientNo, fName, lName, address, telNo, prefType, maxRent)

**Lease** (leaseNo, propertyNo, clientNo, paymentMethod, deposit, paid, rentStart, rentFinish)



Sales Files

**PropertyForRent** (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)

**PrivateOwner** (ownerNo, fName, lName, address, telNo)

**Client** (clientNo, fName, lName, address, telNo, prefType, maxRent)

Contracts Files

**Lease** (leaseNo, propertyNo, clientNo, rent, paymentMethod, deposit, paid, rentStart, rentFinish, duration)

**PropertyForRent** (propertyNo, street, city, postcode, rent)

**Client** (clientNo, fName, lName, address, telNo)

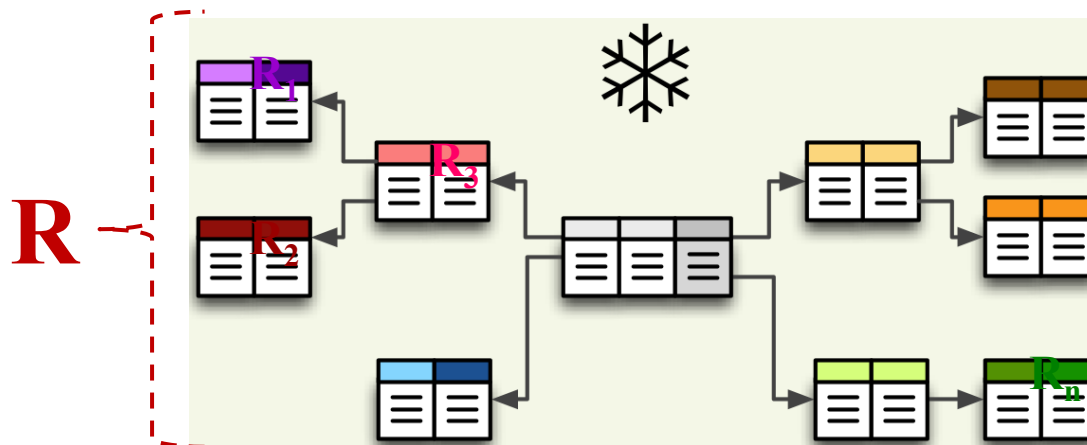
# File Based System (FBS)

# Database Relations

- **Relational Database Schema**

- Set of relation schemas, each with a distinct name.
- If  $R_1, R_2, \dots, R_n$  are a set of relation schemas, then we can write the relational database schema, or simply relational schema,  $R$ , as

$$R = \{R_1, R_2, \dots, R_n\}$$



# Relational Keys

- **Candidate Key**

- A candidate key, **K**, for a relation **R** has two properties:
  1. **Uniqueness** – in each tuple (row) of **R**, the values of **K** uniquely identify that tuple
  2. **Irreducibility** – no proper subset of **K** has the uniqueness property
- There may be several candidate keys for a relation. When a key consists of more than one attribute, we call it **a composite key**

**Read more about it at <http://phlonx.com/blog/fred/index.php/2015/05/21/the-whole-truth-behind-candidate-keys/>**



# Primary Key

- A candidate key is selected to uniquely identify tuples within relation
- Since a relation has no duplicate tuples, it is always possible to identify each row uniquely
  - This means that a relation always has a primary key
  - In the worst case, the entire set of attributes could serve as the primary key

- **Alternate Keys**

- The candidate keys that are not selected to be the primary key are called alternate keys
- For the Branch relation, if we choose **Bno** as the primary key, **Pcode** would then be an alternate key – see next slide

# Alternate Key

Branch

branchNo	street	city	postcode
B005	22 Deer Rd	London	SW1 4EH
B007	16 Argyll St	Aberdeen	AB2 3SU
B003	163 Main St	Glasgow	G11 9QX
B004	32 Manse Rd	Bristol	BS99 1NZ
B002	56 Clover Dr	London	NW10 6EU

**Branch** (Bno, street, city, Pcode)

# Relational Keys

- **Foreign Key**

- Attribute, or set of attributes, within one relation that matches a candidate key of some (possibly same) relation

When an attribute appears in more than one relation, its appearance represents a relationship between tuples of the two relations

- **For example**, the inclusion of **Bno** in both the *Branch* and *Staff* relations is quite deliberate and links each branch to the details of staff working at that branch
- **In the Branch relation, Bno is the primary key**
- **In the Staff relation, Bno will act as a foreign key**

# Relational Keys

Branch				Staff							
Primary Key				Foreign Key							
<u>branchNo</u>	street	city	postcode	<u>staffNo</u>	fName	lName	position	sex	DOB	salary	branchNo
B005	22 Deer Rd	London	SW1 4EH	SL21	John	White	Manager	M	1-Oct-45	30000	B005
B007	16 Argyll St	Aberdeen	AB2 3SU	SG37	Ann	Beech	Assistant	F	10-Nov-60	12000	B003
B003	163 Main St	Glasgow	G11 9QX	SG14	David	Ford	Supervisor	M	24-Mar-58	18000	B003
B004	32 Manse Rd	Bristol	BS99 1NZ	SA9	Mary	Howe	Assistant	F	19-Feb-70	9000	B007
B002	56 Clover Dr	London	NW10 6EU	SG5	Susan	Brand	Manager	F	3-Jun-40	24000	B003
				SL41	Julie	Lee	Assistant	F	13-Jun-65	9000	B005

PropertyForRent									
Primary Key				Foreign Key					
<u>propertyNo</u>	street	city	postcode	type	rooms	rent	ownerNo	staffNo	branchNo
PA14	16 Holhead	Aberdeen	AB7 5SU	House	6	650	CO46	SA9	B007
PL94	6 Argyll St	London	NW2	Flat	4	400	CO87	SL41	B005
PG4	6 Lawrence St	Glasgow	G11 9QX	Flat	3	350	CO40		B003
PG36	2 Manor Rd	Glasgow	G32 4QX	Flat	3	375	CO93	SG37	B003
PG21	18 Dale Rd	Glasgow	G12	House	5	600	CO87	SG37	B003
PG16	5 Novar Dr	Glasgow	G12 9AX	Flat	4	450	CO93	SG14	B003

**Branch** (Bno, street, city, Pcode)

**Staff** (Sno, fName, lName, position, sex, DOB, salary, Bno)

**Property\_For\_Rent** (Pno, street, city, Pcode, type, rooms, rent, Ono, Sno, Bno)

# Properties of Relations

- Relation name is distinct from all other relation names in the relational schema

Customer					
CustomerID (PK)	LastName	FirstName	Address	City	State
C41098X3	Carson	Lewis	121 Center Street	Seattle	WA
CV1099B1	Madison	Sarah	1324 Broadway	Seattle	WA
D345XU24	Brown	Lisa	2201 Second Ave	Seattle	WA

- Each cell of the relation contains exactly one atomic (single) value

Transaction				
TransactionID	TransactionType	TransactionDate	CustomerID (FK)	Amount
10002345	Deposit	2009-2-12 10:25:06	C41098X3	1245.76
10002346	Deposit	2009-2-12 10:27:13	CV1099B1	500.00
10002347	Withdrawal	2009-2-13-14:45:57	C41098X3	200.00

- Each attribute has a distinct name
- Values of an attribute are all from the same domain
- Each tuple (row) is distinct; there are no duplicate tuples
- The order of attributes has no significance
- The order of tuples has no significance, theoretically

**Client**

<u>clientNo</u>	fName	lName	telNo	prefType	maxRent
CR76	John	Kay	0207-774-5632	Flat	425
CR56	Aline	Stewart	0141-848-1825	Flat	350
CR74	Mike	Ritchie	01475-392178	House	750
CR62	Mary	Tregear	01224-196720	Flat	600

**PrivateOwner**

<u>ownerNo</u>	fName	lName	address	telNo
CO46	Joe	Keogh	2 Fergus Dr, Aberdeen AB2 7SX	01224-861212
CO87	Carol	Farrel	6 Achray St, Glasgow G32 9DX	0141-357-7419
CO40	Tina	Murphy	63 Well St, Glasgow G42	0141-943-1728
CO93	Tony	Shaw	12 Park Pl, Glasgow G4 0QR	0141-225-7025

**Viewing**

<u>clientNo</u>	<u>propertyNo</u>	viewDate	comment
CR56	PA14	24-May-04	too small
CR76	PG4	20-Apr-04	too remote
CR56	PG4	26-May-04	
CR62	PA14	14-May-04	no dining room
CR56	PG36	28-Apr-04	

**Registration**

<u>clientNo</u>	<u>branchNo</u>	<u>staffNo</u>	dateJoined
CR76	B005	SL41	2-Jan-04
CR56	B003	SG37	11-Apr-03
CR74	B003	SG37	16-Nov-02
CR62	B007	SA9	7-Mar-03

**Client** (clientNo, fName, lName, telNo, prefType, maxRent)

**PrivateOwner** (Ono, fName, lName, address, telNo)

**Viewing** (clientNo, Pno, viewDate, comment)

**Registration** (clientNo, Bno, Sno, dateJoined)

# Resources / References

---

- **Thomas Connolly, Carolyn Begg 2014, Database Systems: A Practical Approach to Design, Implementation, and Management, 6th Edition Ed., Pearson Education [ISBN: 1292061189] [*available from NCI's library*]**
- **Stephen Morris 2012, Database Principles, 10 Ed., South Western Educational Publishing [ISBN: 1133311970]**
- **Abraham Silberschatz, Henry F. Korth, S. Sudarshan 2010, Database System Concepts, 6th Edition Ed., McGraw Hill Higher Education [ISBN: 0071289593]**
- **Ramez Elmasri and Shamkant Navathe 2013, Fundamentals of Database Systems, 6th Edition Ed., Pearson Education [ISBN: 1292025603]**
- **C.J. Date 2012, Database Design and Relational Theory, O'Reilly Media [ISBN: 1449328016]**
- **Larry Rockoff 2010, The Language of SQL, Course Technology PTR [ISBN: 143545751X]**