Basic Relational Data Model
Keys & Constraints

Hafiz Tayyeb Javed

Assistant Professor (CS)

Department of Computer Science

FAST-NU, CFD Campus

Objectives

- Terminology
- Characteristics of Relations
- Relational Data Model Notations
- Key constraints
- Others Constraints

Terminology

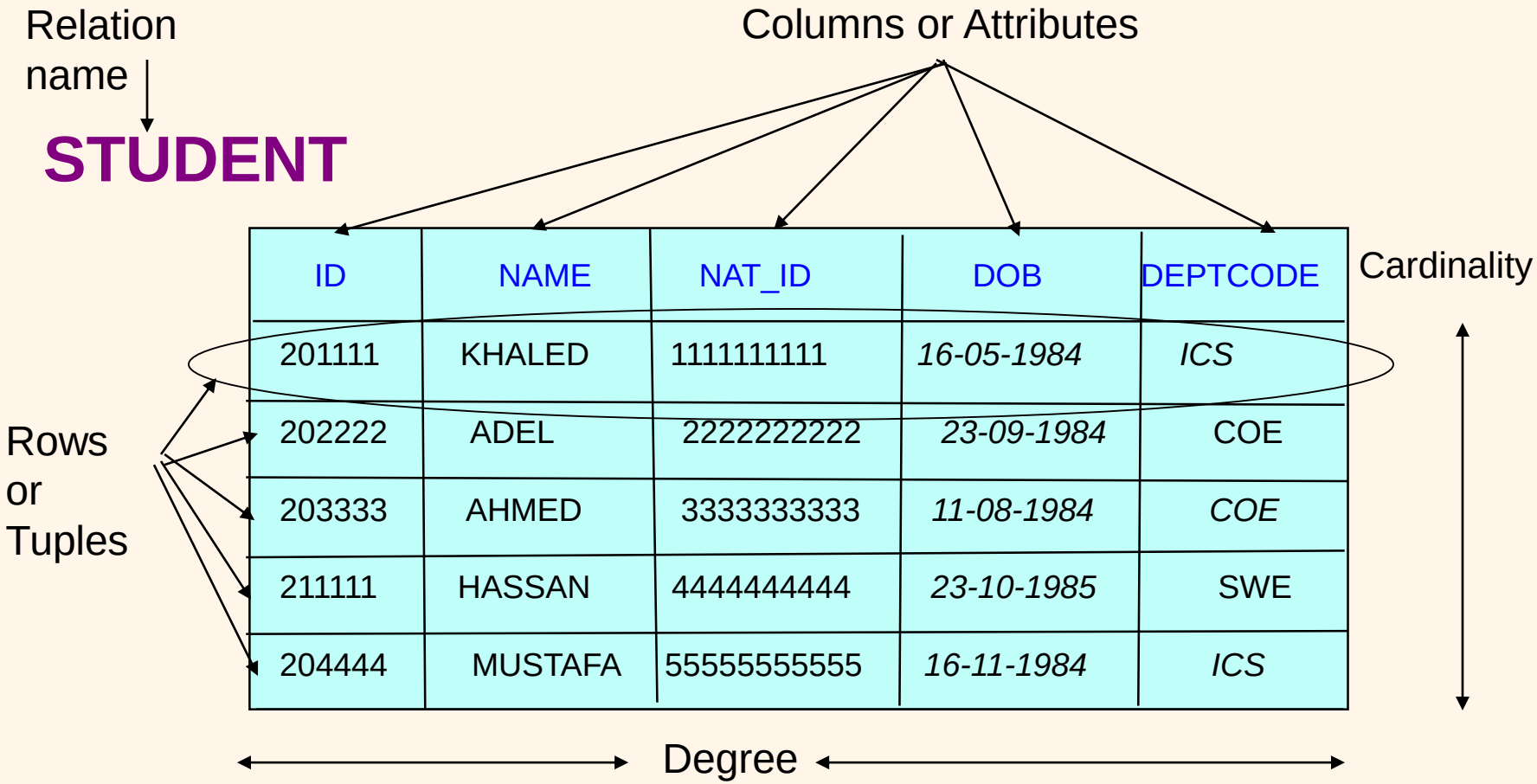
- A relation is a table with columns and rows.
- Attribute is a named column of a relation.
- Domain is a set of allowable values for one or more attributes.
- Tuple is a row of a relation.
- Degree is a number of attributes in a relation.
- Cardinality is a number of tuples in a relation.
- Relationship is a link or dependency between relations.
- Relational Database is a collection of relations.

Terminology

A relation is a table with columns and rows.

A table can be written as

STUDENT(ID, Name, Nat_ID, DOB, DeptCode)



Characteristics of Relations

- Each relation in the same relational database schema has a distinct name
- Each value in a tuple is atomic
- Each attribute in a relation has a distinct name.
- Values of an attribute are all from the same domain.
- Each tuple is distinct.
- Order of attributes has no significance.
- Order of tuples has no significance, theoretically.

Relational Data Model Notations

The letters Q, R, S denote the abstract relation names.

$R(A_1, A_2, A_3, ..., A_n)$ denotes a relation schema R of degree n.

Example: STUDENT(RollNo, Name, Phone, Address, Mobile, DOB)

COURSE(<u>CourseID</u>, Title, CreditHours)

Both $t[A_i]$ and $t.A_i$ refers to the value v_i in t for attribute A_i

Example: in second tuple, both t[name] and t.name refers to “Adel”

Dom(DEPTCODE) = {‘COE’,’ICS’,’SWE’}

Relation's Definition

Relation: A *relation* is a subset of Cartesian product of all domains

R(r) $\subseteq$ Dom(A1) x dom(A2) x ..... x Dom(An)

Where A1, A2, ..., An represent attributes of a relation r.

Evaluation:

Given

A1 = {1, 2}, A2 = {a}, A3 = {x, y}

Then relation

X = {(1, a, y), (2, a, x), (2, a, y)}; 3 tuples of a relation

is a subset of

Y = Dom(A1) x dom(A2) x Dom(A3)

= {(1, a, x), (1, a, y), (2, a, x), (2, a, y)}

Note: Degree of a relation is the number of attributes, in above case degree of relation is 3.

Alternate Definition of Relation

For each tuple ti is a mapping from R to D

Where D = Dom(A1) $\cup$ Dom(A2)$\cup$ ... $\cup$ Dom(An)

And t[Ai] $\in$ Dom(Ai)

Then a tuple can be considered as attribute and value.

For example: The following representation shows a tuple definition

t = <(id, 21587), (Name, Muhammad,), (Hphone, Null), (Wphone, 866-1141)>

Data Models

- Object-Based Data Models
  - Entity-Relationship
  - Semantic
  - Functional
  - Object-Oriented
- Record-Based Data Models
  - Relational Data Model
  - Network Data Model
  - Hierarchical Data Model
- Physical Data-Models

Keys

A **key** is a combination of one or more columns that is used to identify rows in a relation

A **composite key** is a key that consists of two or more columns

Super Key

A super key is a combination of columns that uniquely identifies any row within a relational database management system (RDBMS) table.

Roll#	FirstName	LastName	Address	City	NIC#	Deptno
...						
12i-2565	...					CS
	Ali					CS
	Ali					EE
...						

{Roll#}, {Roll#, NIC#}, {NIC#}
{Roll#, FirstName, LastName, Address, City, NIC#, Deptno}

Example-1

Roll Number	First Name	Last Name
CSU0001	Shabbir	Bhimani
CSU0002	Jain	Rao
CSU0003	Nitin	Uday

Now we have the following as super keys

Roll Number	First Name	
Roll Number	**First Name**	**Last Name**

Example-2

Following table consists of four columns
EmployeeID, Name, Job, DeptID

Examples of superkeys in this table would be {employeeID, Name}, {employeeID, Name, job}, and {employeeID, Name, job, departmentID}

In a real database we don't need values for all of those columns to identify a row

We only need, per our example, the set {EmployeeID}.

This is a minimal superkey

So, employeeID is a candidate key.

EmployeeID can also uniquely identify the tuples.

FirstName and FatherName (for NADRA)

Candidate Keys

A **candidate key** is a key that determines all of the other columns in a relation
Candidate key columns help in searching fewer duplicated or unique records.
Examples
In PRODUCT relation
Prod# is a candidate key
Prod_Name is also a candidate key
In ORDER_PROD
(OrderNumber, Prod#) is a candidate key
If BirthDate (or DrivingLicense#) is known then Name, address, telno etc. can be found from fewer searched records.

Identifying Candidate Keys

For Bank: AccountHolder (or Customer)
ACC#, Fname, Lname, DOB, CNIC#, Addr, City, TelNo, Mobile#, DriveLic#

For PTCL: Customer (single telno holders)
Consumer#, Fname, Lname, DOB, CNIC#, Addr, City, TelNo, Mobile#

For NADRA: Citizen (CNIC#, Fname, Lname, FatherName, DOB, OldCNIC#, PAddr, PCity, TAddr, TCity, TelNo, Mobile#)

Identifying right record using Candidate Key – an Example

Let's analyze the following set of columns

{FirstName, LastName, DOB, TelNo, NatCardID}

First Name	Last Name	DOB	TelNo	NatCardID
Fiaz	Ali			
Fyaz	Ali			110116987231

Candidate Keys cont….

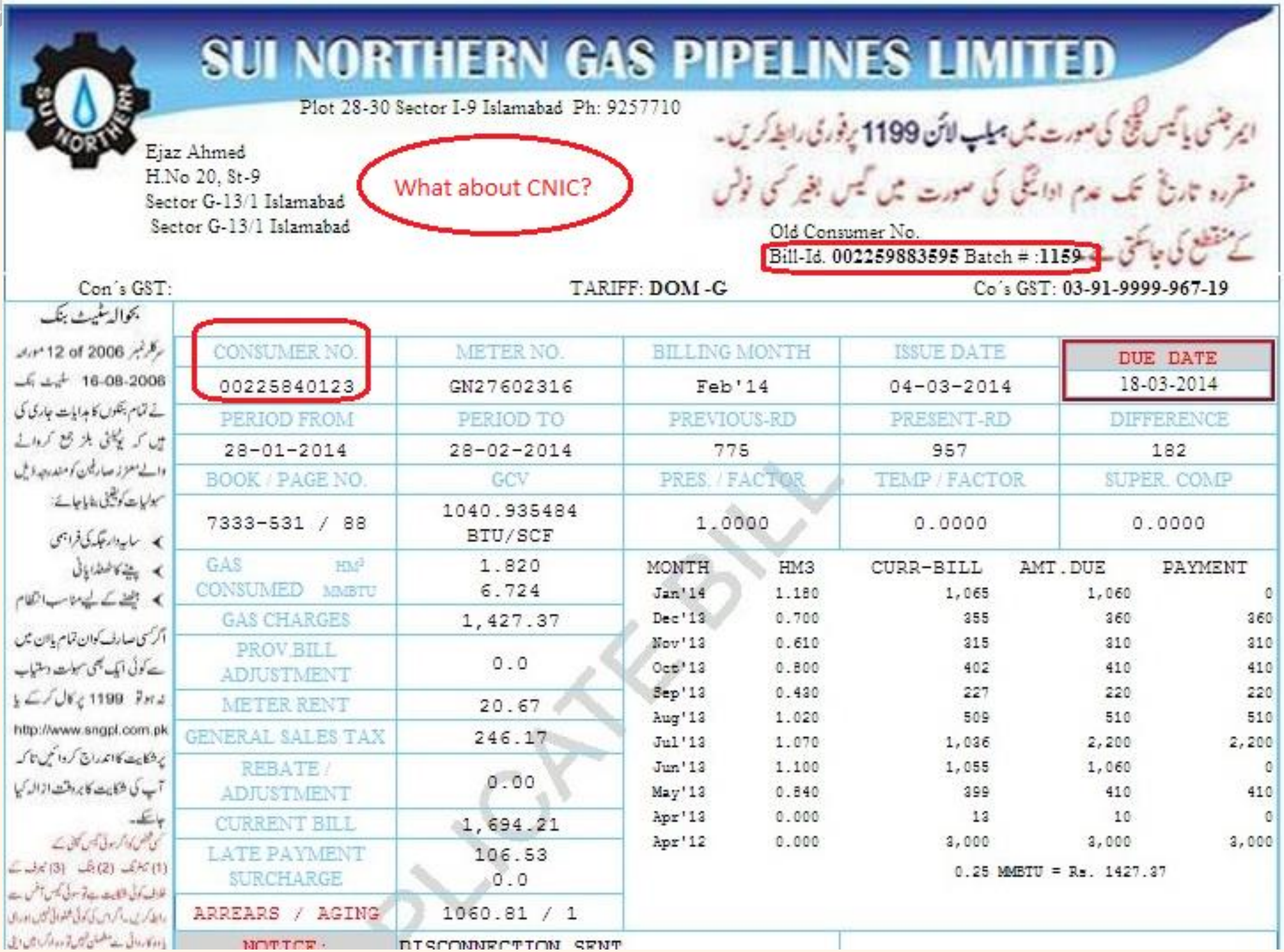
Example-1: Branch (branch-name, assets, branch-city)
Candidate Key: {branch-name}

Example-2: Customer (cname, natid, address, city, telno)
cname, address, city can be duplicated individually and cannot determine a record.

The following combinations distinguish customer records or tuples.
{cname, telno}
{natid}
{natid, cname}
As {natid} ⊆ {natid, cname}, then {natid, cname} is not candidate key and {natid} is a candidate key

Example-3: Employee(empno, name, birth_date, address, city, telno, citizenship_id)
empno, telno, citizenship_id are possible candidate keys ???

Exercise:
Course (cid, cname, deptno)
Semester (sid, syear, startdate)
ClassAllocation (sid, cid, sec#, building#, room#)
Identify candidate keys in each of the above relations



SUI NORTHERN GAS PIPELINES LIMITED

Plot 28-30 Sector I-9 Islamabad Ph: 9257710

ایمرجنسی یا گیس لیکیج کی صورت میں ہیلپ لائن 1199 پر فوری رابطہ کریں۔
مقررہ تاریخ تک عدم ادائیگی کی صورت میں گیس بغیر کسی نوٹس کے منقطع کی جا سکتی ہے

Ejaz Ahmed
H.No 20, St-9
Sector G-13/1 Islamabad
Sector G-13/1 Islamabad

What about CNIC?

Old Consumer No.
Bill-Id. 002259883595 Batch # :1159

Con's GST: TARIFF: **DOM -G** Co's GST: 03-91-9999-967-19

http://www.sngpl.com.pk

CONSUMER NO.	METER NO.	BILLING MONTH	ISSUE DATE	DUE DATE
00225840123	GN27602316	Feb'14	04-03-2014	18-03-2014
PERIOD FROM	PERIOD TO	PREVIOUS-RD	PRESENT-RD	DIFFERENCE
28-01-2014	28-02-2014	775	957	182
BOOK / PAGE NO.	GCV	PRES. / FACTOR	TEMP / FACTOR	SUPER. COMP
7333-531 / 88	1040.935484 BTU/SCF	1.0000	0.0000	0.0000

GAS CONSUMED HM³ / MMBTU	1.820 / 6.724
GAS CHARGES	1,427.37
PROV.BILL ADJUSTMENT	0.0
METER RENT	20.67
GENERAL SALES TAX	246.17
REBATE / ADJUSTMENT	0.00
CURRENT BILL	1,694.21
LATE PAYMENT SURCHARGE	106.53 / 0.0
ARREARS / AGING	1060.81 / 1
NOTICE:	DISCONNECTION SENT

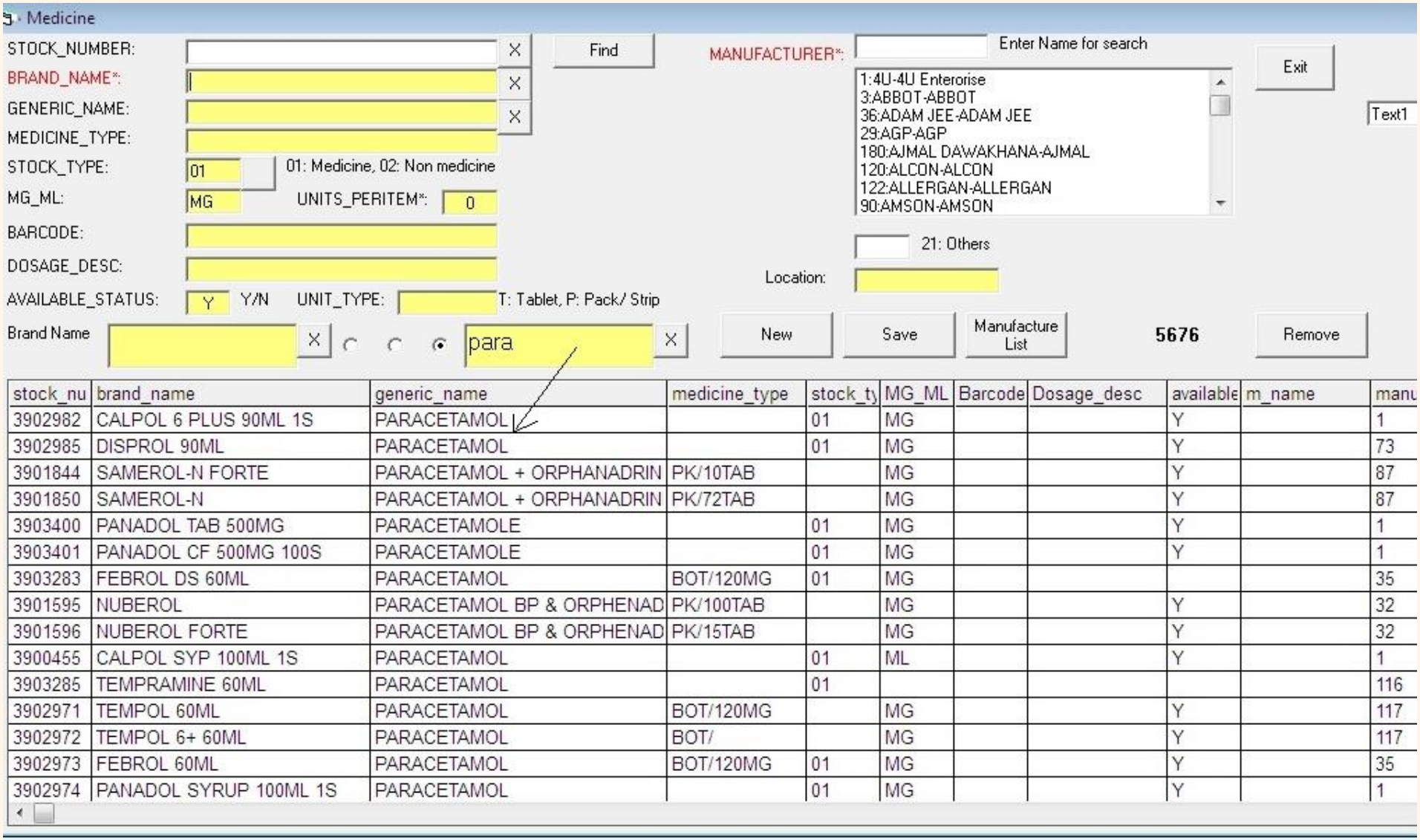
MONTH	HM3	CURR-BILL	AMT.DUE	PAYMENT
Jan'14	1.180	1,065	1,060	0
Dec'13	0.700	355	360	360
Nov'13	0.610	315	310	310
Oct'13	0.800	402	410	410
Sep'13	0.430	227	220	220
Aug'13	1.020	509	510	510
Jul'13	1.070	1,036	2,200	2,200
Jun'13	1.100	1,055	1,060	0
May'13	0.840	399	410	410
Apr'13	0.000	13	10	0
Apr'12	0.000	3,000	3,000	3,000

0.25 MMBTU = Rs. 1427.37

DUPLICATE BILL

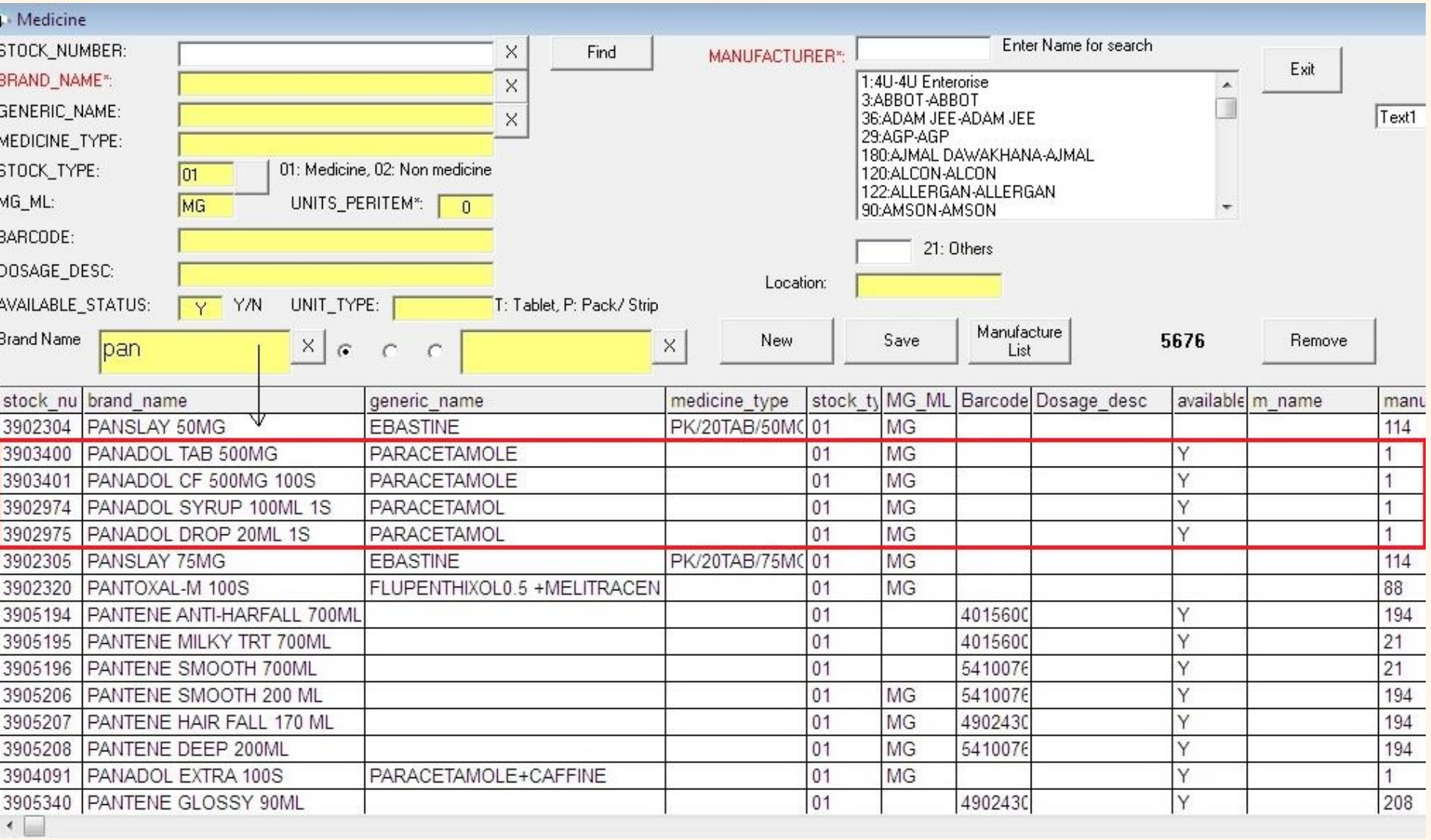
Question: With one CNIC, there may be many consumer nos (houses/ shops) , explain?

Candidate Key during search



stock_nu	brand_name	generic_name	medicine_type	stock_ty	MG_ML	Barcode	Dosage_desc	available	m_name	manu
3902982	CALPOL 6 PLUS 90ML 1S	PARACETAMOL		01	MG			Y		1
3902985	DISPROL 90ML	PARACETAMOL		01	MG			Y		73
3901844	SAMEROL-N FORTE	PARACETAMOL + ORPHANADRIN	PK/10TAB		MG			Y		87
3901850	SAMEROL-N	PARACETAMOL + ORPHANADRIN	PK/72TAB		MG			Y		87
3903400	PANADOL TAB 500MG	PARACETAMOLE		01	MG			Y		1
3903401	PANADOL CF 500MG 100S	PARACETAMOLE		01	MG			Y		1
3903283	FEBROL DS 60ML	PARACETAMOL	BOT/120MG	01	MG					35
3901595	NUBEROL	PARACETAMOL BP & ORPHENAD	PK/100TAB		MG			Y		32
3901596	NUBEROL FORTE	PARACETAMOL BP & ORPHENAD	PK/15TAB		MG			Y		32
3900455	CALPOL SYP 100ML 1S	PARACETAMOL		01	ML			Y		1
3903285	TEMPRAMINE 60ML	PARACETAMOL		01						116
3902971	TEMPOL 60ML	PARACETAMOL	BOT/120MG		MG			Y		117
3902972	TEMPOL 6+ 60ML	PARACETAMOL	BOT/		MG			Y		117
3902973	FEBROL 60ML	PARACETAMOL	BOT/120MG	01	MG			Y		35
3902974	PANADOL SYRUP 100ML 1S	PARACETAMOL		01	MG			Y		1

Candidate Key during search ...



stock_nu	brand_name	generic_name	medicine_type	stock_ty	MG_ML	Barcode	Dosage_desc	available	m_name	manu
3902304	PANSLAY 50MG	EBASTINE	PK/20TAB/50MC	01	MG					114
3903400	PANADOL TAB 500MG	PARACETAMOLE		01	MG			Y		1
3903401	PANADOL CF 500MG 100S	PARACETAMOLE		01	MG			Y		1
3902974	PANADOL SYRUP 100ML 1S	PARACETAMOL		01	MG			Y		1
3902975	PANADOL DROP 20ML 1S	PARACETAMOL		01	MG			Y		1
3902305	PANSLAY 75MG	EBASTINE	PK/20TAB/75MC	01	MG					114
3902320	PANTOXAL-M 100S	FLUPENTHIXOL0.5 +MELITRACEN		01	MG					88
3905194	PANTENE ANTI-HARFALL 700ML			01		401560(		Y		194
3905195	PANTENE MILKY TRT 700ML			01		401560(		Y		21
3905196	PANTENE SMOOTH 700ML			01		5410076		Y		21
3905206	PANTENE SMOOTH 200 ML			01	MG	5410076		Y		194
3905207	PANTENE HAIR FALL 170 ML			01	MG	490243(		Y		194
3905208	PANTENE DEEP 200ML			01	MG	5410076		Y		194
3904091	PANADOL EXTRA 100S	PARACETAMOLE+CAFFINE		01	MG			Y		1
3905340	PANTENE GLOSSY 90ML			01		490243(		Y		208

Dependents

For example: **Dependent** means family members (wife, son, daughter etc.), who depends on father. Father is an employee in some organization

Primary Keys

A **primary key** is a candidate key selected as the primary means of identifying rows in a relation:

- There is one and only one primary key per relation
- The primary key is NOT NULL, UNIQUE
- The ideal primary key is short, numeric(alpha), indexed, fixed length and never changes
- The primary key may be a composite key

Primary Keys cont....

A primary key is a minimal identifier that is used to identify tuples uniquely. This means that no subset of the primary key is sufficient to provide unique identification of tuples. NULL value is not allowed in primary key attribute.

Example:

- STU_ID

(Practically PK must have UNIQUE, NOT NULL, INDEXED on an attribute)

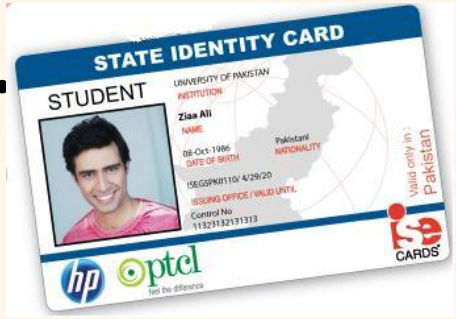
STUDENT

STU_ID	NAME	CID_ID	DOB	DEPTCODE

Although, candidate keys are STU_ID, CIT_ID. Organization ITSELF issues its own identifier PK. PK is a property of an organization. Driving license# is a property of Ministry of Motors/ transportation, it must not be used PK in university or other organizations. It is a candidate key for other organization for search purposes.

Note: It is not recommended to create PKs for columns like Quantity, Salary, Price and column that has datatype DATE

Primary Keys cont…



Example-1:
STUDENT(StuID, FirstName, FamilyName, DOB, …)

Example-2: Building (B#, BName, Location, Region),
B# is a primary key. Although, BName is unique, not null but it is not short.

Example-3: Customer (cname, citizenid, address, city, telno)
This relation indicates the information about personal details. There is a chance that cname is duplicated, some may have citizenid and telno as null. This forces us to introduce new a attribute such as cust# that would be a primary key.
Customer (cust#, cname, citizenid, address, city, telno)
Example-4:
BankBranch(Branch-Name, City, TotalAsset)

What will be a PK that you will suggest?
What is a candidate key?
Which attribute is unique
There may be many branches in one city then finalize this relation with possible constraints

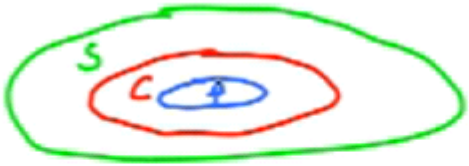
Primary Keys cont…

PK for this table

Do not make PK

Entity Set: VENDOR

VENDOR ID	V NAME	V STREET	V CITY	V STATE	V ZIP
V110	Jersey Vegetable Co.	2 Main St.	Patterson	NJ	07055
V25	General Provisions	125 Common St.	Boise	ID	44830
V250	Spices Unlimited	25 Salty Lane	East Hampton	NY	10027
V75	Pasta Supply, Inc.	34 Henry St.	Philadelphia	PA	09098



Primary Key is a Candidate Key

Candidate is a Super Key

Primary Key Instances - Samples

512235665 610112654	Social Security Number (By Birth Asian, Europian, HomeLand)
2003389	Student ID (With Registration year)
DMM051 JUB004	Product ID (with Manufacturing Loc, branches)
312668852369	Bank Account# (followed by Branch Code)
100211 300411	SKU# (By Category) or a barcode
Serial#	Serial# starts with 1 and incremented by 1
005384 007004	Company has many branches with code 005, 007 etc., invoices issued from branches are recognized with branch codes

Use of PK or Indexing

```
SQL> set timing on
SQL> select PATIENT_ID, max(PATIENT_VISITNO)
  2  from PATIENT_VISIT
  3  where PATIENT_VISITNO = (select max(PATIENT_VISITNO)
  4     from PATIENT_VISIT
  5     where DEPEND_SNO='0')
  6  group by PATIENT_ID
  7  having count(*)>=1;

PATIENT_ID MAX(PATIENT_VISITNO)
---------- --------------------
9999993                    1316

Elapsed: 00:00:00.12
```

```
Table: PATIENT_VISIT
Index: Yes, PK
Test Query execution time:less than 1 second

Table: PATIENT_VISIT1
Index: NO
Test Query execution time:12 seconds

SQL> select count(*) Total_Records
  2  from PATIENT_VISIT;

  Total_Records
---------------
         625721

Elapsed: 00:00:00.04
```

```
SQL> select PATIENT_ID, max(PATIENT_VISITNO)
  2  from PATIENT_VISIT1
  3  where PATIENT_VISITNO = (select max(PATIENT_VISITNO)
  4     from PATIENT_VISIT1
  5     where DEPEND_SNO='0')
  6  group by PATIENT_ID
  7  having count(*)>=1;

PATIENT_ID MAX(PATIENT_VISITNO)
---------- --------------------
9999993                    1316

Elapsed: 00:00:11.54
SQL>
```


```
PRIMARY KEY ( PATIENT_ID, DEPEND_SNO, PATIENT_VISITNO )
```

Surrogate Keys

A **surrogate key** as an artificial column added to a relation to serve as a primary key:

- DBMS supplied
- Short, numeric and never changes – an ideal primary key!
- Has artificial values that are meaningless to users
- Normally hidden in forms and reports

Surrogate Keys cont....

RENTAL_PROPERTY without surrogate key:

RENTAL_PROPERTY (Street, City,
State/Province, Zip/PostalCode, Country, Rental_Rate)

RENTAL_PROPERTY with surrogate key:
RENTAL_PROPERTY (PropertyID, Street, City,State/Province,
Zip/PostalCode, Country, Rental_Rate)

Other examples include Invoice#, ComplaintRef#
ATMTransaction (Card#, Serial#, Amount, DrawDate)

Surrogate Keys cont....

Other examples include

Invoice#

ComplaintRef#

ATMTransaction (Card#, Serial#, Amount, TransDate, TransType)

InsurancePaid(Policy#, PaidYear, PaidDate, VehicleReg#, Amount)

ATM Transaction Sample

بنك الرياض
riyad bank

LOCATION: KING FAHED UNIVERSITY 954
TERM. ID: 0009549L03203596
ATM ID : 03596
A/C NO : ********99940
AID : A0000002281010

DATE	TIME	TRANS NO
25/09/11	09:33	571188

CASH WITHDRAWAL SAR 500.00
TOTAL AMOUNT DEBITED TO ACCOUNT IS:

SAR 500.00

10% OFF OUR 4 PRODUCTS. CREDIT CARDS, PERSONAL LOAN, MORTGAGE AND AUTO LEASING

YOUR NEW BALANCE IS:
SAR 20,263.74

NCB الأهلي

DATE : 15/09/11
TIME : 11:05
ATM ID : EASTRG3003101313
CARD NUM : 588849********8032
APPL NAME: SPAN
APPL AID : A0000002281010

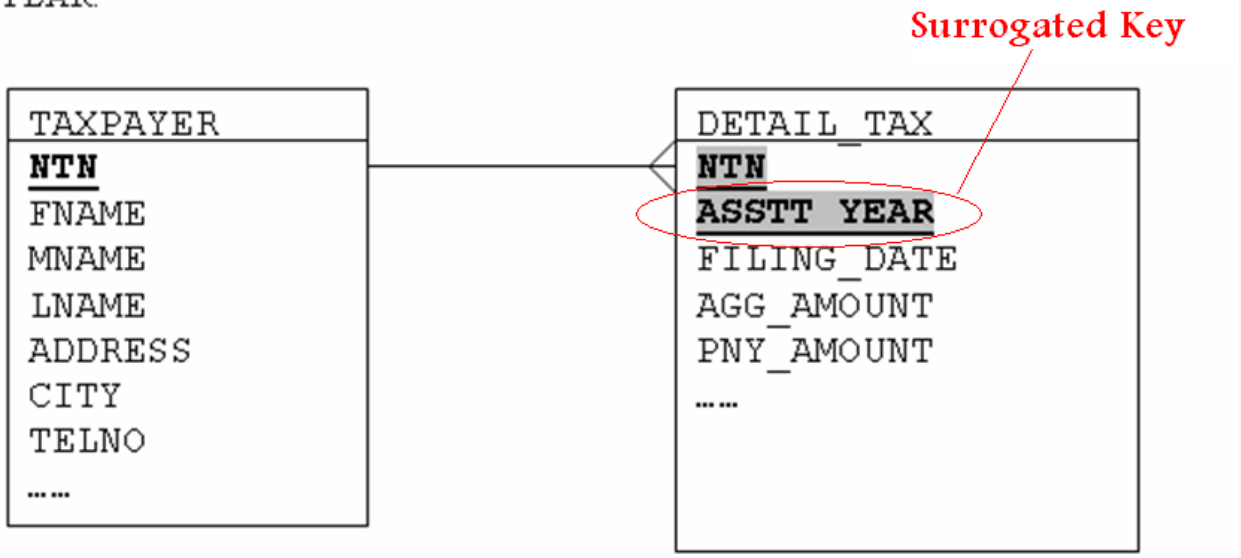
AL-TASHEILAT FUEL STATION
KOBAR
EASTERN REGION-DAMMAM

STAN: 003775
APPROVAL CODE: 739098
RRN 9393 CASH WITHDRAWAL
SAUDI RIYALS 5,000.00
YOUR BALANCE IS 7,141.64 SAR

THANK YOU FOR USING SPAN

Surrogate Keys cont....

A tax payer with primary key NTN pays tax every year and tax is assessed in next or later year called ASSTT_YEAR. An assessment or taxpayer status is identified every time with NTN and ASSTT_YEAR.



NTN and ASSTT_YEAR are both primary keys in DETAIL_TAX table, known as composite key of this table. NTN of DETAIL_TAX table is a FK references to NTN of TAXPAYER table.

Exercise: What is different between PK and unique key?

Exercise: How to avoid that a record can be duplicated with all attributes other than PK attribute?

Exercise: Give three more examples from real life where composite keys can be applied

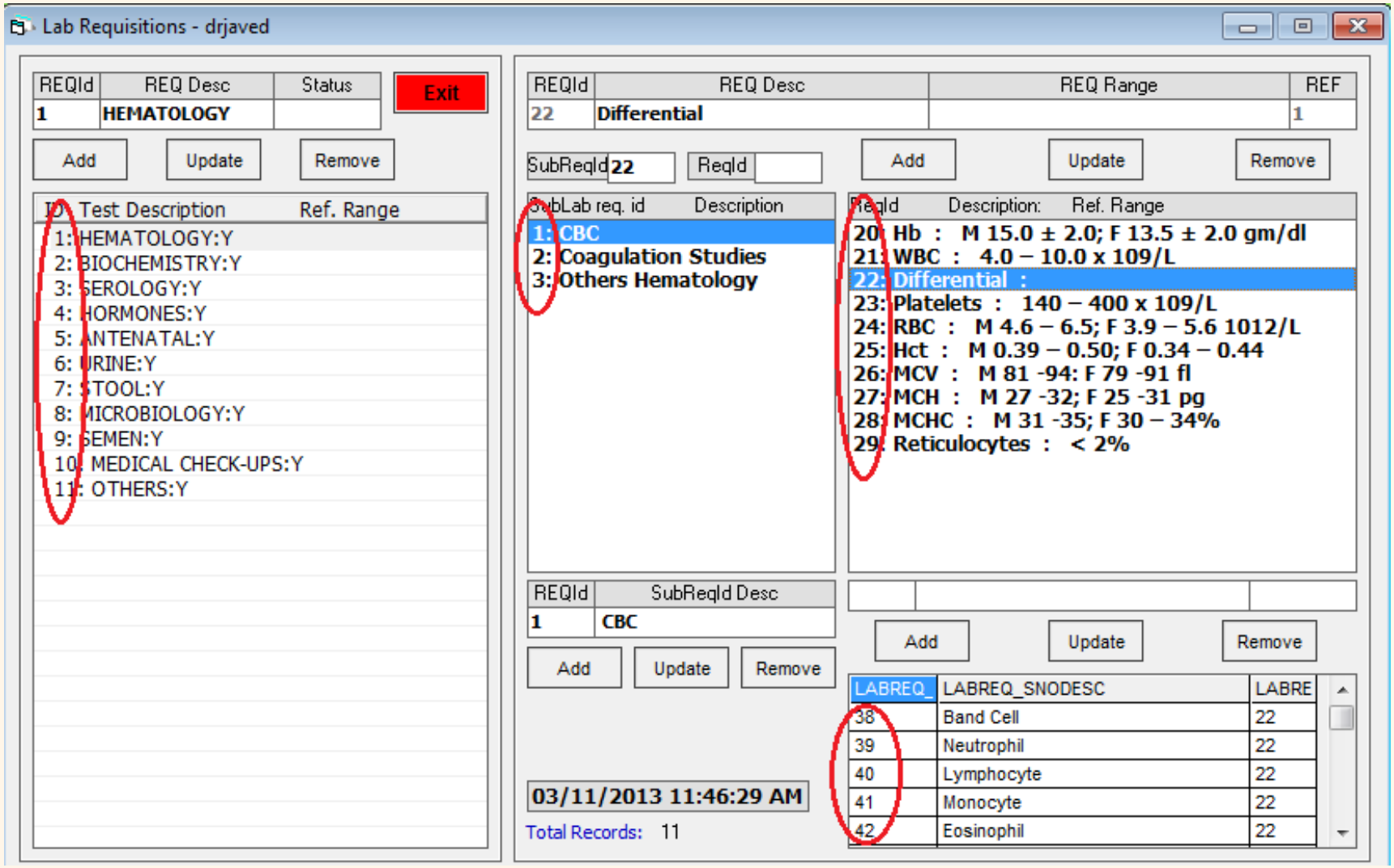
Surrogate Keys cont....

PatientID	Visit#	Visit Date
558741	1	...
558741	2	
558741	3	
...		
589210	1	
589210	2	
599211	1	
599211	2	
...		

Better Solution
using composite keys

PatientID	Visit#	Visit Date
558741	1	...
558741	2	
558741	3	
589210	4	
589210	5	
599211	6	
599211	7	
...	...	
	830000	and more

Example – End User View



Foreign Keys (FK)

A **foreign key** is an attribute that refers to a primary key of same or different relation to form a link (constraint) between the relations:

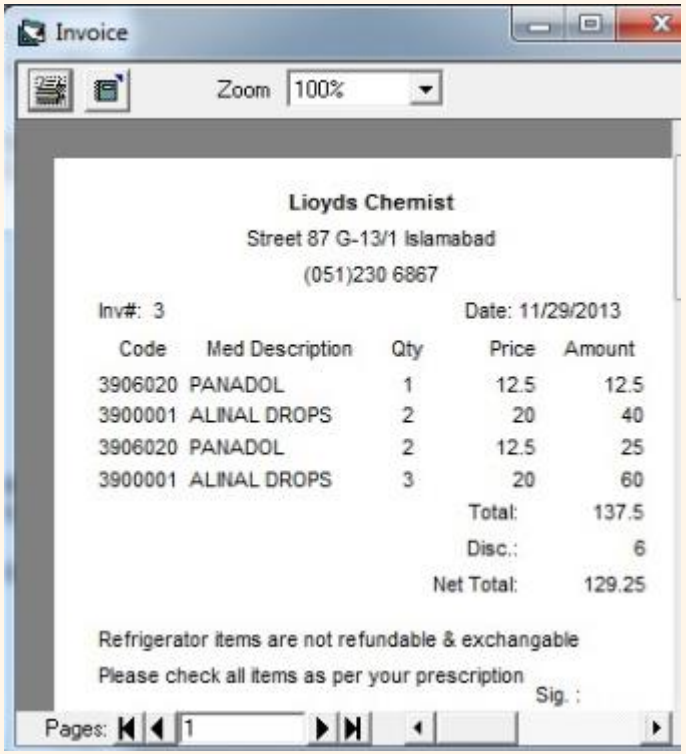
- A foreign key can be a single column or a composite key
- The term refers to the fact that key values are *foreign* to the relation in which they appear as foreign key values
- Ideally Data type, Length/ Size of FK and referring PK must be same

Foreign Keys

NOTE: The primary keys of the relations are underlined and any foreign keys are in *italics* in the relations below:
Name of FK column may be different from the name of referencing PK

Example-1:
DEPARTMENT (DeptID, DepartmentName, BudgetCode, ManagerName)
EMPLOYEE (EmployeeNumber, EmployeeName, *DeptID*)

Customer-Orders



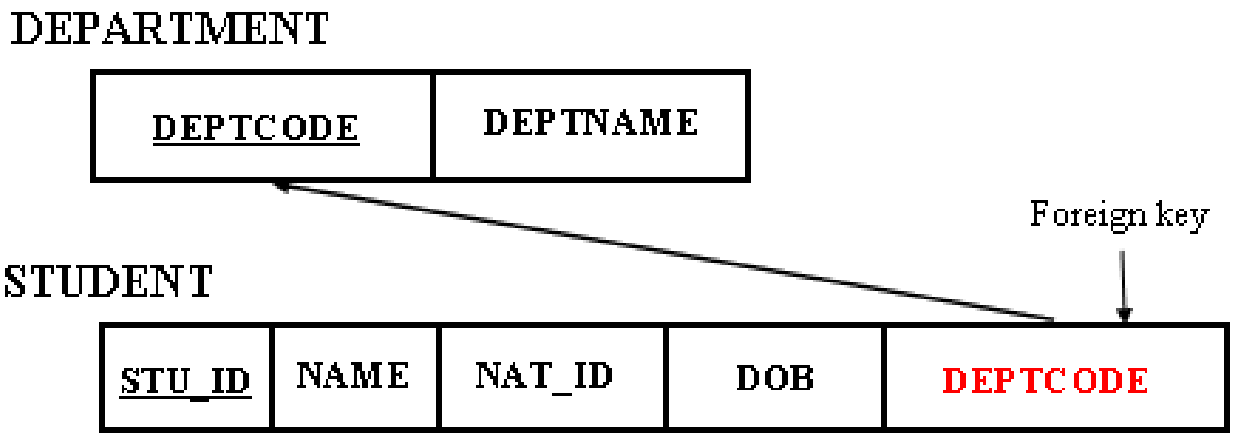
Example-2:

CUSTOMER(CustId, Name, Address, Telno, Email)
ORDERS(Invoice#, InvType, InvDate, *CustomerId*)

Can we call Invoice# as a Surrogated Key?
Does Invoice# always contain serial numbers?

Foreign Key with a Referential Integrity Constraint

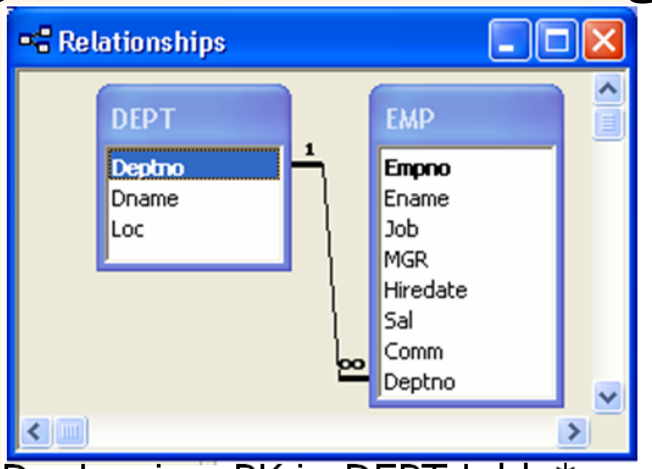
- An attribute or a set of attributes within one relation that matches the candidate key (or PK) of some (possibly the same) relation.



How to use FK in SQL:

```
select d.deptname, s.name, s.DOB
from Department d, Student s
where s.deptcode = e.deptcode;
```

Foreign Key with a Referential Integrity Constraint



Deptno is a PK in DEPT table*

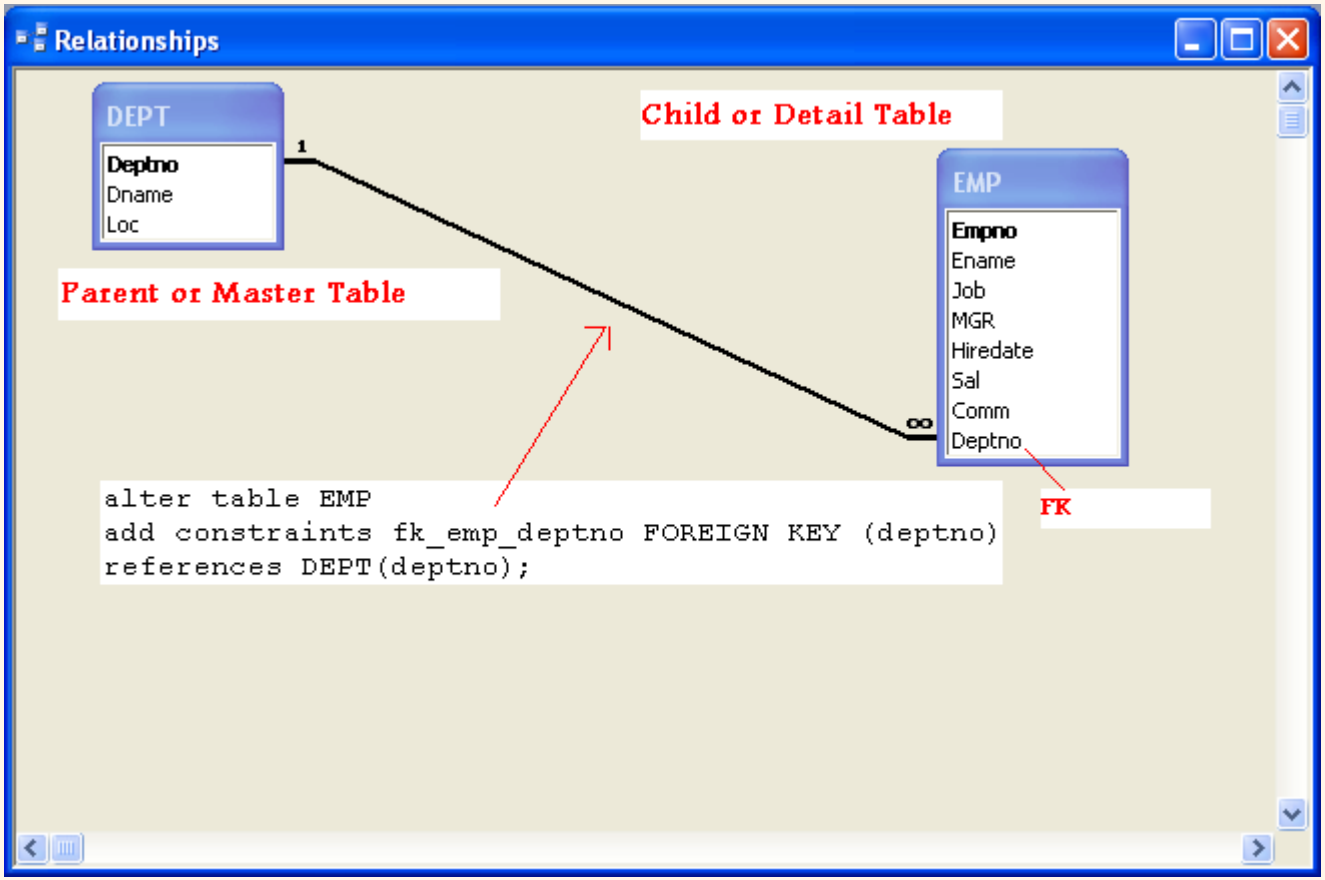
Empno is a PK in EMP table

Deptno is a FK in EMP table*

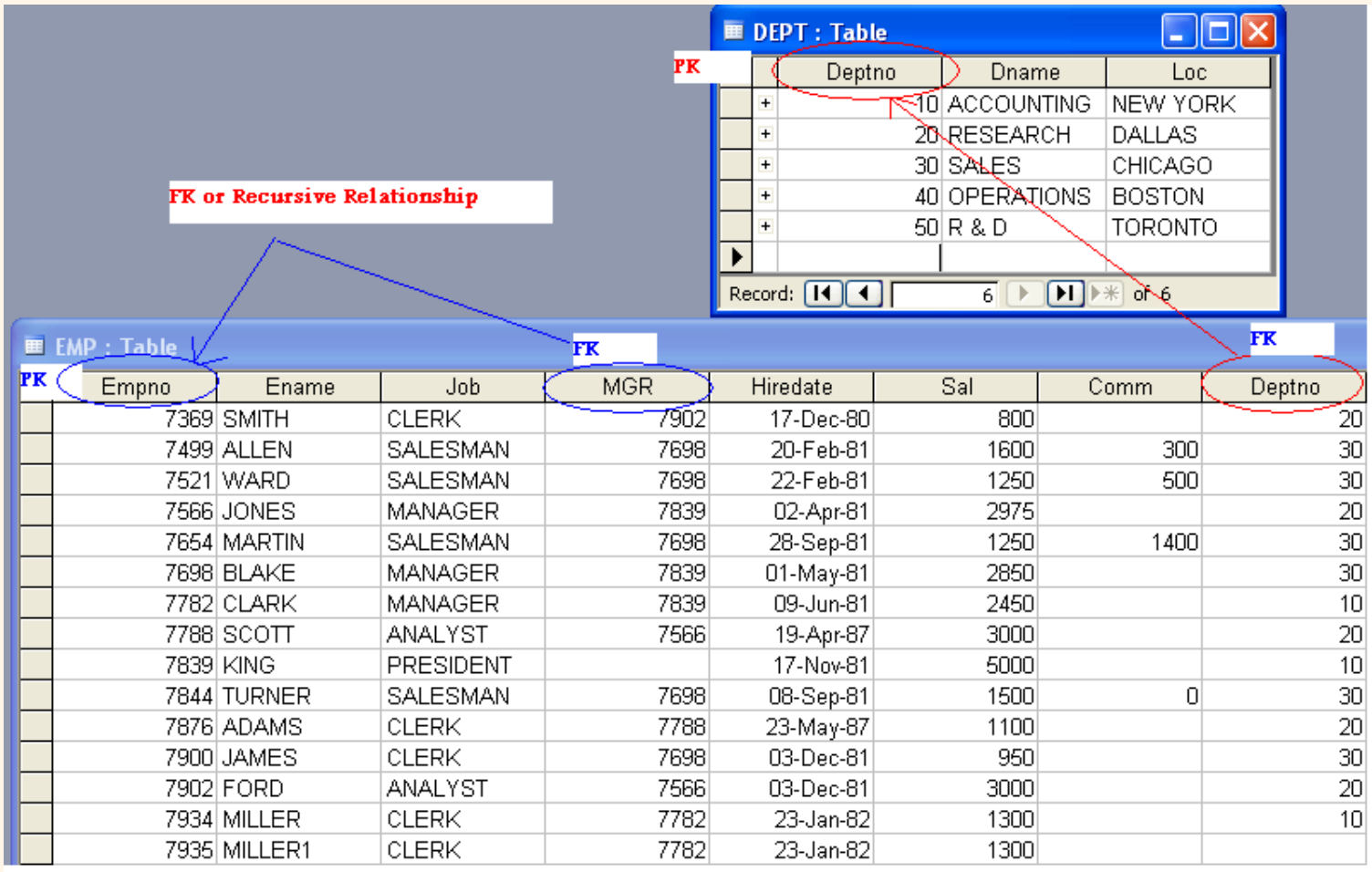
(it references to an attribute Deptno of DEPT table)

NOTE: An attribute within one relation that matches the key (Primary/same) in the same relation, is called recursive relation

Creating Foreign Key in MS Access & Oracle

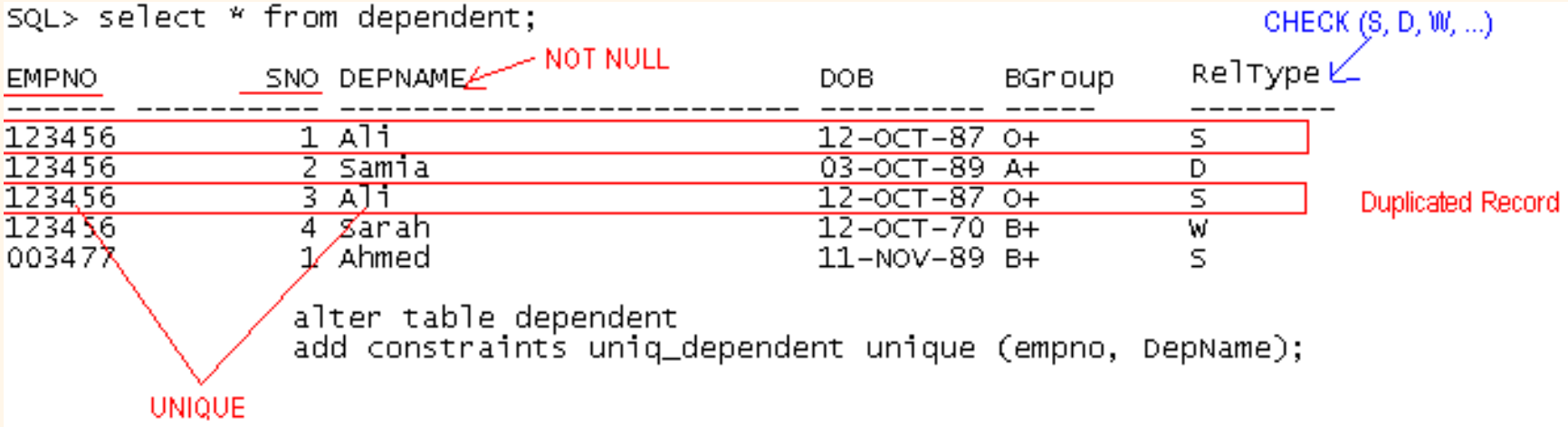


FK and Recursive Relationship



Other Constraints Examples

Unique constraint can be applied on single or multiple columns



CHECK (Max_Participants>=Enrolled_Participants)

CHECK (Deptno>0)

CHECK (RelType IN ('S', 'D', 'W'))

Default 'Y'

Database Triggers