### Assignment 1

17

- □ Due Sept 27, 2019 at 10:00pm
- ☐ You are responsible for checking Avenue for news and updates.
- □ Sample script template to be posted

### Keys

18

- A key is a set of attributes that uniquely identifies tuples in a relation.
- More precisely:
  - A set of attributes K is a superkey for a relation r if r cannot contain two distinct tuples t₁ and t₂ such that t₁[K]=t₂ [K];
  - K is a (candidate) key for r if K is a minimal superkey

(that is, there exists no other superkey K' of r that is contained in K as proper subset, i.e,  $\ K' \subset K)$ 

### Example

19

RegNum	Surname	FirstName	BirthDate	DegreeProg
284328	Smith	Luigi	29/04/59	Computing
296328	Smith	John	29/04/59	Computing
587614	Smith	Lucy	01/05/61	Engineering
934856	Black	Lucy	01/05/61	Fine Art
965536	Black	Lucy	05/03/58	Fine Art

- RegNum is a key: i.e., RegNum is a superkey and it contains a sole attribute, so it is minimal.
- □ {Surname, Firstname, BirthDate} is another key

### Beware!

20

RegNum	Surname	FirstName	BirthDate	DegreeProg
296328	Smith	John	29/04/59	Computing
587614	Smith	Lucy	01/05/61	Engineering
934856	Black	Lucy	01/05/61	Fine Art
965536	Black	Lucy	05/03/58	Engineering

☐ There is no pair of tuples with the same values on both Surname and DegreeProg;

i.e., in each program students have different surnames; can we conclude that **Surname** and **DegreeProg** form a key for this relation?

No! There **could be** students with the same surname in the same program

### Existence of Keys

21

- Relations are sets; therefore each relation is composed of <u>distinct</u> tuples.
- It follows that the whole set of attributes for a relation defines a superkey.
- Therefore each relation has a key, which is the set of all its attributes, or a subset thereof.
- The existence of keys guarantees that each piece of data in the database can be accessed,
- Keys are a major feature of the Relational Model and allow us to say that it is "value-based".

### Keys and Null Values

22

- If there are nulls, keys do not work that well:
  - □ They do not guarantee unique identification;
  - They do not help in establishing correspondences between data in different relations

RegNum	Surname	FirstName	BirthDate	DegreeProg
NULL	Smith	John	NULL	Computing
587614	Smith	Lucy	01/05/61	Engineering
934856	Black	Lucy	NULL	NULL
NULL	Black	Lucy	05/03/58	Engineering

- Are the third and fourth tuple the same?
- How do we access the first tuple?

### **Primary Keys**

23

- ☐ The presence of nulls in keys has to be limited.
- Each relation must have a primary key on which nulls are not allowed (in any attribute)
- □ Notation: the attributes of the primary key are <u>underlined</u>
- □ References between relations are realized through primary keys

RegNum	Surname	FirstName	BirthDate	DegreeProg
643976	Smith	John	NULL	Computing
587614	Smith	Lucy	01/05/61	Engineering
934856	Black	Lucy	NULL	NULL
735591	Black	Lucy	05/03/58	Engineering

### Do we Always Have Primary Keys?

24

- □ In most cases, we do have reasonable primary keys (e.g., student number, SIN)
- □ There may be multiple keys, one of which is designated as primary.

### Recap

25

- ☐ A set of fields is a key for a relation if:
  - No two distinct tuples can have same values in all key fields, and
  - 2. This is not true for any subset of the key.
- ☐ If #2 false, then a superkey.
- ☐ If there's >1 key for a relation, one of the keys is chosen to be the primary key.
- E.g., sid is a key for Students. (What about name?) The set {sid, gpa} is a superkey.

### Primary and Candidate Keys

26

Enrolled(sid, cid, grade)

- 1. "For a given student and course, there is a single grade." vs.
- "Students can take only one course, and receive a single grade for that course; further, no two students in a course receive the same grade."
- Be careful to define Integrity Constraints (ICs) correctly at design time.
- ICS are checked when data is updated.

Enrolled(sid, cid, grade)

Enrolled(sid, cid, grade)

• key (cid, grade)

### Foreign Keys

27

- Pieces of data in different relations are correlated by means of values of primary keys.
- Referential integrity constraints are imposed in order to guarantee that the values refer to existing tuples in the referenced relation.
- $\square$  A foreign key requires that the values on a set X of attributes of a relation  $R_1$  must appear as values for the primary key of another relation  $R_2$ .
  - In other words, set of attributes in one relation that is used to `refer' to a tuple in another relation. (Must correspond to primary key of the second relation.) Like a `logical pointer'.

### Referential Integrity

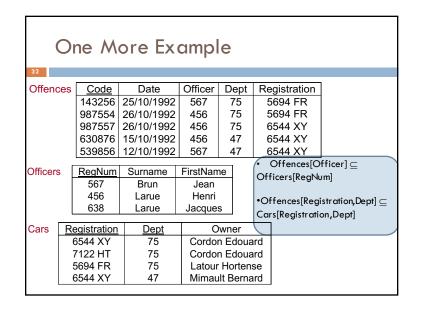
28

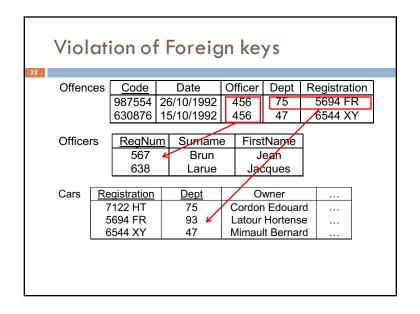
- □ E.g. sid is a foreign key referring to Students:
  - Enrolled(sid: string, cid: string, grade: string)
  - If all foreign key constraints are enforced, <u>referential</u> <u>integrity</u> is achieved, i.e., no dangling references.

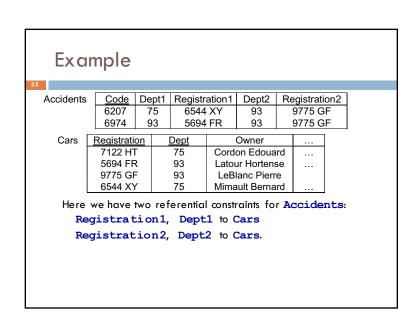
### Referential Integrity (cont'd) □ Only students listed in the Students relation should be allowed to enroll for courses. Enrolled Students sid cid grade sid name login age gpa 53666 Carnatic101 53666 Iones jones@cs 3.4 53666 Reggae203 В 3.2 53688 Smith smith@eecs 18 53650 Topology112 53650 Smith smith@math 3.8 53666 History105

## Consider Students and Enrolled; sid in Enrolled is a foreign key that references Students. What should be done if an Enrolled tuple with a non-existent student id is inserted? Reject it! What should be done if a Students tuple is deleted? Also delete all Enrolled tuples that refer to it. Disallow deletion of a Students tuple that is referred to. Set sid in Enrolled tuples that refer to it to a default sid. Set sid in Enrolled tuples that refer to it to NULL. Similar if primary key of Students tuple is updated.

# Where do ICs Come From? ICs are based upon the semantics of the real-world enterprise that is being described in the database relations. We can check a database instance to see if an IC is violated, but we cannot infer that an IC is true by looking at an instance. An IC is a statement about all possible instances Key and foreign key ICs are the most common; more general ICs supported too.







### Referential Constraints: Comments Referential constraints play an important role in making the relational model value-based. Care is needed in case of referential constraints that involve two or more attributes.