Entity:

* what we store data about
* Person/thing –row. Each row is a tuple. Also called record.
* Entity type is for which the data is being stores. Eg person, or thing.

Attributes:

* What the data is. The attributes of the entity.
* name, address, age- each column, also called field.

View mechanism:

* Change the way the database looks.
* For different users different kinds of views can be made. (to show only certain amount of information).

SQL

* Data definition language –DDL- defines the data.- creating table , defining structure
* Data manipulation language- DML- manipulates the language- adding and manipulating values. Update, delete, search

Data integrity:

Entity integrity: that every entity should be unique in the table. For that, ID can be used to show every entity is unique, even if some attribute values match.

Referential integrity: if there is another table with some primary key (ID) then the primary key value (ID) should be a valid ID.

Domain integrity: what values should be stored for every column. Ex: phone no should be numbers and 10 digit numbers. Rules for each column values.

Keys-

* used to define uniqueness, less error.
* Protect integrity.

Foreign key- It is a key in child table which refers to primary key in parent table. The customer\_id in items table is a primary key in customer table but it is a foreign key in items table. The FOREIGN KEY constraint prevents invalid data from being inserted into the foreign key column, because it has to be one of the values contained in the parent table.

Super key- Set of any number of columns which make every row unique in the table.

Candidate key- least no of columns making every entry unique. Can contain null values

Primary key- parent table. It is one single column, which makes every entry unique. unique, never change, never NULL.

**Alternate Key:** The candidate key other than the primary key is called an alternate key.

Natural vs surrogate keys: surrogate are the ones which are not present in the db and are added by admin to make the entries unique. Natural are already present in the db.

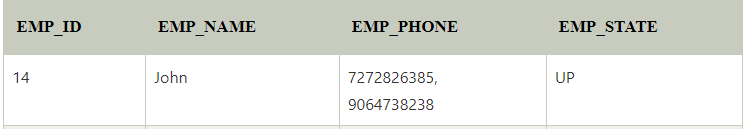
| **Super Key** | **Candidate Key** |
| --- | --- |
| Super Key is an attribute (or set of attributes) that is used to uniquely identifies all attributes in a relation. | Candidate Key is a subset of a super key. |
| All super keys can’t be candidate keys. | But all candidate keys are super keys. |
| Various super keys together makes the criteria to select the candidate keys. | Various candidate keys together makes the criteria to select the primary keys. |
| In a relation, number of super keys is more than number of candidate keys. | While in a relation, number of candidate keys are less than number of super keys. |
| Super key attributes can contain NULL values. | Candidate key attributes can also contain NULL values. |

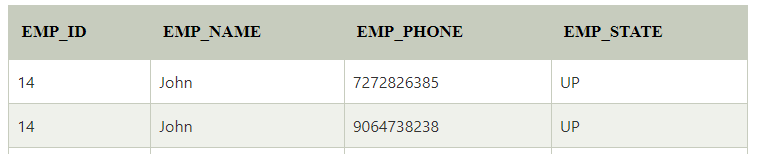
**Data modification anomalies can be categorized into three types:**

* **Insertion Anomaly:** Insertion Anomaly refers to when one cannot insert a new tuple into a relationship due to lack of data.
* **Deletion Anomaly:** The delete anomaly refers to the situation where the deletion of data results in the unintended loss of some other important data.
* **Updatation Anomaly:** The update anomaly is when an update of a single data value requires multiple rows of data to be updated.

First Normal Form (1NF)

* It states that an attribute of a table cannot hold multiple values or multiple data type values. It must hold only single-valued attribute.
* Table should have a primary key

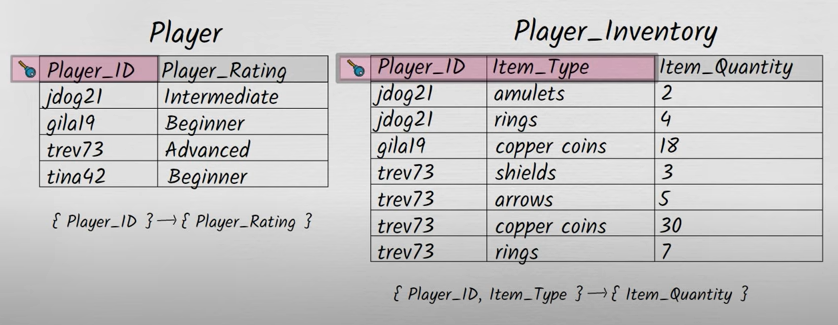




Second Normal Form (2NF)

* In the second normal form, all non-key attributes must be dependent on the entire primary key.
* By making 2 tables now each table has separate primary key and all other attributes are directly connected to their entire primary key value.





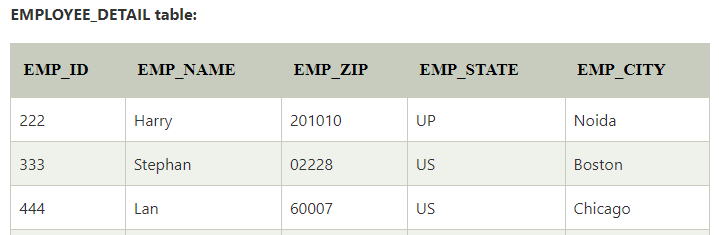
Third Normal Form(3NF)/Boyce Codd normal form

* There should not be any transitive dependency for non-key attributes, then the relation must be in third normal form. Every non key attribute should depend on the primary key, whole key directly not indirectly.
* Example:
* Super key in the table above: {EMP\_ID}, {EMP\_ID, EMP\_NAME}, {EMP\_ID, EMP\_NAME, EMP\_ZIP}....so on .

**Candidate key:** {EMP\_ID}

**Non-prime attributes:** In the given table, all attributes except EMP\_ID are non-prime.

Here, EMP\_STATE & EMP\_CITY dependent on EMP\_ZIP and EMP\_ZIP dependent on EMP\_ID. The non-prime attributes (EMP\_STATE, EMP\_CITY) transitively dependent on super key (EMP\_ID). It violates the rule of third normal form.



To this :

