4.6 Aggregation (Extra)

- To specify mathematical aggregate functions on collections of values from the database.
- Common functions applied to collections of numeric values include SUM, AVERAGE, MAXIMUM, and MINIMUM. The COUNT function is used for counting tuples or values.
- Another common type of request involves grouping the tuples in a relation by the value of some of their attributes and then applying an aggregate function independently to each group.
- It is important to note that, in general, duplicates are not eliminated when an aggregate function is applied; this way, the normal interpretation of functions such as SUM and AVERAGE is computed.
- However, NULL valuesThere is no single agreed-upon notation for specifying aggregate functions. In some cases a “script A” ($\mathcal{A}$) is used.
- are not considered in the aggregation.
- We can define an AGGREGATE FUNCTION operation, using the symbol I (pronounced script F), to specify these types of requests as follows:

<grouping attributes> $\mathfrak{I}$ <function list> (R)

4.6 Aggregation (Extra)

Employee

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5

Examples

- Retrieve each department number, the number of employees in the department, and their average salary, while renaming the resulting

$$\rho R_{(Dno,\ No\_of\_employees,\ Average\_sal)}\ (_{Dno}\ \Im\ COUNT\ Ssn,\ AVERAGE\ Salary\ (EMPLOYEE))$$

R

Dno	No_of_employees	Average_sal
5	4	33250
4	3	31000
1	1	55000

- We specified a list of attribute names—between parentheses in the RENAME operation—for the resulting relation R. If no renaming is applied, then the attributes of the resulting relation that correspond to the function list will each be the concatenation of the function name with the attribute name in the form <function>_<attribute>.

Dno	Count_ssn	Average_salary
5	4	33250
4	3	31000
1	1	55000

$$_{Dno}\ \Im\ COUNT\ Ssn,\ AVERAGE\ Salary(EMPLOYEE)$$

  - *Note that this is an arbitrary notation, consistent with what SQL would do.*

- If no grouping attributes are specified, the functions are applied to all the tuples in the relation, so the resulting relation has a single tuple only.

Count_ssn	Average_salary
8	35125

$$\Im\ COUNT\ Ssn,\ AVERAGE\ Salary(EMPLOYEE)$$