Relational Data Model

Tanzim Hossain
Lecturer
Department of Software Engineering
Daffodil International University

# Building Blocks

Tables

- A two-dimensional arrangement of data.
- A table consists of a <u>heading</u> defining the table name and column names and a <u>body</u> containing rows of data.

StdSSN	StdFirstName	StdLastName	StdCity	StdState	StdZip	StdMajor	StdClass	StdGPA
123-45-6789	HOMER	WELLS	SEATTLE	WA	98121-1111	IS	FR	3.00
124-56-7890	BOB	NORBERT	BOTHELL	WA	98011-2121	FIN	JR	2.70
234-56-7890	CANDY	KENDALL	TACOMA	WA	99042-3321	ACCT	JR	3.50

Relationships

- Connection between rows in two tables.
- Relationships are shown by column values in one table that match column values in another table.
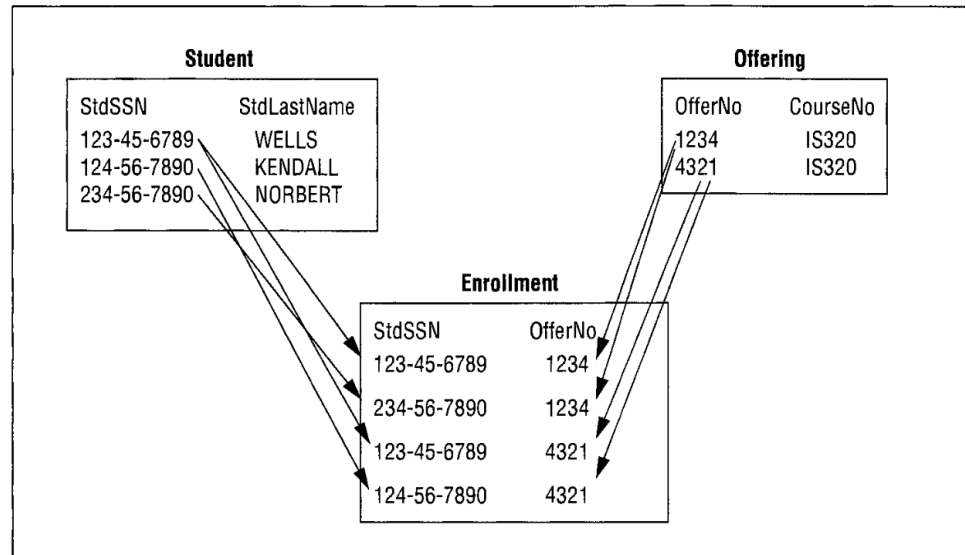
Alternative Terminologies

- The <u>table-oriented</u> terminology appeals to end users; the <u>set-oriented</u> terminology appeals to academic researchers; and the <u>record-oriented</u> terminology appeals to information systems professionals.

Table-Oriented	Set-Oriented	Record-Oriented
Table	Relation	Record type, file
Row	Tuple	Record
Column	Attribute	Field

Integrity Rules

Entity Integrity

- Entity integrity means that each table must have a column or combination of columns with unique values.

- Unique means that no two rows of a table have the same value.

- For example, StdSSN in Student is unique and the combination of StdSSN and OfferNo is unique in Enrollment.

- Entity integrity ensures that entities (people, things, and events) are uniquely identified in a database.

Referential Integrity

- Referential integrity means that the column values in one table must match column values in a related table.

- For example, the value of StdSSN in each row of the Enrollment table must match the value of StdSSN in some row of the Student table.

- Referential integrity ensures that a database contains valid connections.

Definitions

Superkey & Candidate Key

- A column or combination of columns containing unique values for each row.

- The combination of every column in a table is always a superkey because rows in a table must be unique.

- Candidate key is a minimal superkey.

- A superkey is minimal if removing any column makes it no longer unique.

Null Value, Primary & Foreign Key

- Null value: a special value that represents the absence of an actual value.

- A null value can mean that the actual value is unknown or does not apply to the given row.

- Primary key: A specially designated candidate key.

- The primary key for a table cannot contain null values.

- A primary key consisting of more than one column is known as a composite or a combined primary key.

Null Value, Primary & Foreign Key

- Foreign key: A column or combination of columns in which the values must match those of a candidate key.

- A foreign key must have the same data type as its associated candidate key.

- In the CREATE TABLE statement of SQL, a foreign key must be associated with a primary key rather than merely a candidate key. Why?

Integrity Rules

- <u>Entity integrity rule:</u> No two rows of a table can contain the same value for the primary key.

- In addition, no row can contain a null value for any column of a primary key.

- <u>Referential integrity rule:</u> Only two kinds of values can be stored in a foreign key:
  1. A value matching a candidate key value in some row of the table containing the associated candidate key, or
 2. a null value.

Self-Reference

- A relationship in which a foreign key refers to Referential Integrity for Self-Referencing (Unary) Relationships the same table.

- Self-referencing relation ships represent associations among members of the same set.

FacSSN	FacFirstName	FacLastName	FacCity	FacState	FacDept	FacRank	FacSalary	FacSupervisor	FacHireDate	FacZipCode
098-76-5432	LEONARD	VINCE	SEATTLE	WA	MS	ASST	$35,000	654-32-1098	01-Apr-95	98111-9921
543-21-0987	VICTORIA	EMMANUEL	BOTHELL	WA	MS	PROF	$120,000		01-Apr-96	98011-2242
654-32-1098	LEONARD	FIBON	SEATTLE	WA	MS	ASSC	$70,000	543-21-0987	01-Apr-95	98121-0094
765-43-2109	NICKI	MACON	BELLEVUE	WA	FIN	PROF	$65,000		01-Apr-97	98015-9945
876-54-3210	CRISTOPHER	COLAN	SEATTLE	WA	MS	ASST	$40,000	654-32-1098	01-Apr-99	98114-1332
987-65-4321	JULIA	MILLS	SEATTLE	WA	FIN	ASSC	$75,000	765-43-2109	01-Apr-00	98114-9954

Graphical Representation of Referential Integrity

Referential Integrity
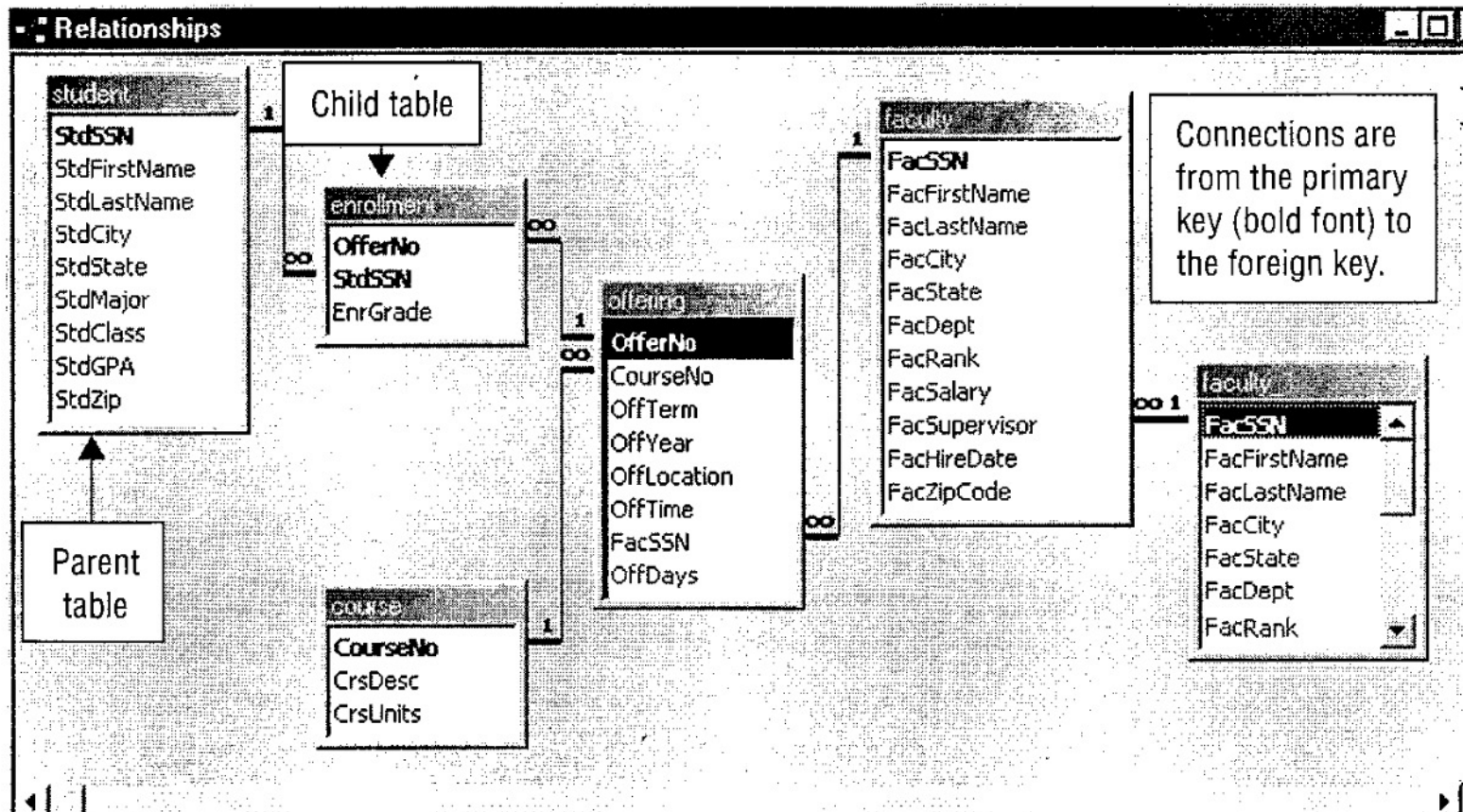
Parent Table / Child Table

- In a relationship, the primary key table is known as the <u>parent</u> or "1" table (for example, Student) and the foreign key table (for example, Enrollment) is known as the <u>child</u> or "M" (many) table.

1-M Relations

- The relationship from Student to Enrollment is called "1-M" (one to many) because a student can be related to many enrollments but an enrollment can be related to only one student.

- Similarly, the relationship from the Offering table to the Enrollment table means that an offering can be related to many enrollments but an enrollment can be related to only one offering.

M-N Relations

- M-N (many to many) relationships are not directly represented in the Relational Model.

- An M-N relationship means that rows from each table can be related to many rows of the other table.

- For example, a student enrolls in many course offerings and a course offering contains many students.

M-N Relations

- In the Relational Model, a pair of 1-M relationships and a linking or associative table represents an M-N relationship.

- The linking table Enrollment and its relationships with Offering and Student represent an M-N relationship between the Student and Offering tables.

Delete / Update Referenced Row

Delete/Update Referenced Row

- For each referential integrity constraint, you should carefully consider actions on referenced rows in parent tables of 1-M relationships.

- For example, the first row of the Course table with CourseNo "IS320" is referenced by the first row of the Offering table.

- It is natural to consider what happens to related Offering rows when the referenced Course row is deleted or the CourseNo is updated.

Delete/Update Referenced Row

- More generally, these concerns can be stated as:
  - <u>Deleting a referenced row:</u> What happens to related rows (that is, rows in the child table with the identical foreign key value) when the referenced row in the parent table is deleted?
  - <u>Updating the primary key of a referenced row:</u> What happens to related rows when the primary key of the referenced row in the parent table is updated?

Restrict & Cascade

- Restrict: Do not allow the action on the referenced row. For example, do not permit a Student row to be deleted if there are any related Enrollment rows. Similarly, do not allow Student.StdSSN to be changed if there are related Enrollment rows.

- Cascade: Perform the same action (cascade the action) to related rows. For example, if a Student is deleted, then delete the related Enrollment rows. Likewise, if Student.StdSSN is changed in some row, update StdSSN in the related Enrollment rows.

Nullify

- Nullify: Set the foreign key of related rows to null.

- For example, if a Faculty row is deleted, then set FacSSN to NULL in related Offering rows.

- Likewise, if Faculty. FacSSNis updated, then set FacSSN to NULL in related Offering rows.

- The nullify action is not permitted if the foreign key does not allow null values.

Default

- Default: Set the foreign key of related rows to its default value.

- For example, if a Faculty row is deleted, then set FacSSN to a default faculty in related Offering rows.

- The default faculty might have an interpretation such as "to be announced."

- The default action is an alternative to the nullify action as the default action avoids null values.

Relational Algebra

Relational Algebra

- You can think of relational algebra similarly to the algebra of numbers except that the objects are different: algebra applies to numbers and relational algebra applies to tables.

- In algebra, each operator transforms one or more numbers into another number.

- Similarly, each operator of relational algebra transforms a table (or two tables) into a new table.

Restrict / Project

- Restrict produces a subset of the rows, while project produces a subset of columns.
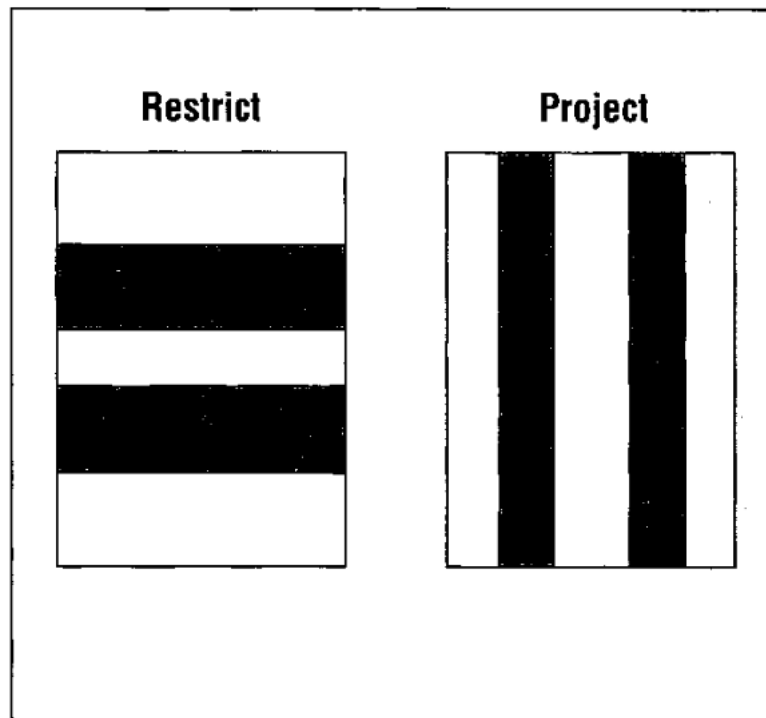
Restrict / Project

- Restrict uses a condition or logical expression to indicate what rows should be retained in the output.

- Project uses a list of column names to indicate what columns to retain in the output.

- Sometimes after a subset of columns is retrieved, there are duplicate rows. When this occurs, the project operator removes the duplicate rows.

Restrict / Project

- Note that if the primary key or a candidate key is included in the list of columns, the resulting table has no duplicates.

- This side effect is due to the mathematical nature of relational algebra. In relational algebra, tables are considered sets.

- Because sets do not have duplicates, duplicate removal is a possible side effect of the project operator.

Extended Cross Product

- The extended cross product (product for short) operator shows everything possible from two tables.

- The product of two tables is a new table consisting of all possible combinations of rows from the two input tables.

- The number of rows in the resulting table is the product of the number of rows of the two input tables.

- In contrast, the number of result columns is the sum of the columns of the two input tables.

Sample Student & Enrollment Tables

StdSSN	StdLastName	StdMajor	StdClass
123-45-6789	WELLS	IS	FR
124-56-7890	NORBERT	FIN	JR
234-56-7890	KENDALL	ACCT	JR

OfferNo	StdSSN	EnrGrade
1234	123-45-6789	3.3
1234	234-56-7890	3.5
4321	124-56-7890	3.2

Product of Student & Enrollment Tables

Student.StdSSN	StdLastName	StdMajor	StdClass	OfferNo	Enrollment.StdSSN	EnrGrade
123-45-6789	WELLS	IS	FR	1234	123-45-6789	3.3
123-45-6789	WELLS	IS	FR	1234	234-56-7890	3.5
123-45-6789	WELLS	IS	FR	4321	124-56-7890	3.2
124-56-7890	NORBERT	FIN	JR	1234	123-45-6789	3.3
124-56-7890	NORBERT	FIN	JR	1234	234-56-7890	3.5
124-56-7890	NORBERT	FIN	JR	4321	124-56-7890	3.2
234-56-7890	KENDALL	ACCT	JR	1234	123-45-6789	3.3
234-56-7890	KENDALL	ACCT	JR	1234	234-56-7890	3.5
234-56-7890	KENDALL	ACCT	JR	4321	124-56-7890	3.2

Extended Cross Product

- The extended cross product operator can combine any two tables.

- Other table combining operators have conditions about the tables to combine.

- Because of its unrestricted nature, the extended cross product operator can produce tables with excessive data.

- The extended cross product operator is important because it is a building block for the join operator.
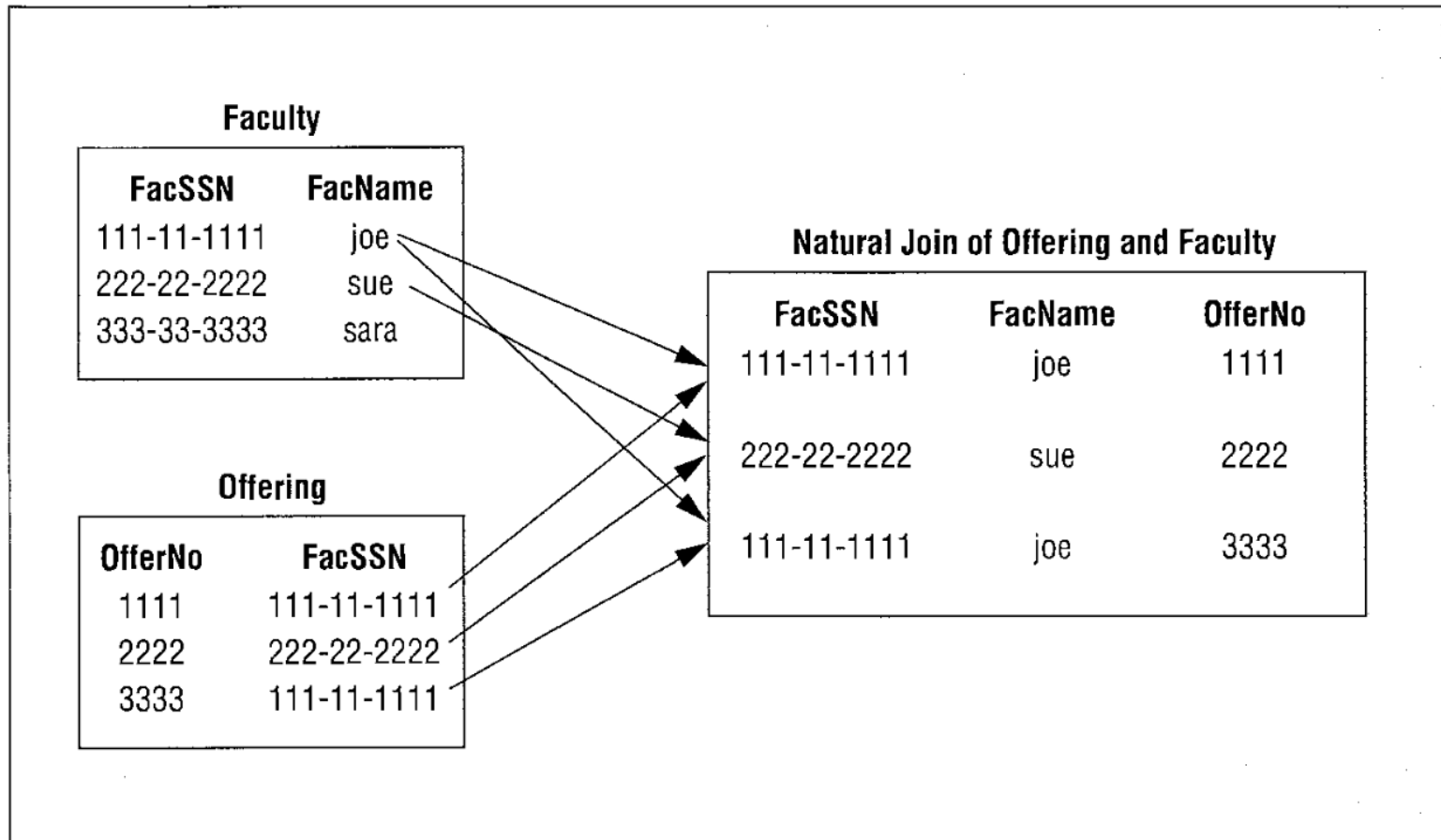
Join Operators

- An operator that produces a table containing rows that match on a condition involving a column from each input table

- The join operator builds a new table by combining rows from two tables that match on a join condition.

- Typically, the join condition specifies that two rows have an identical value in one or more columns.

- When the join condition involves equality, the join is known as an <u>equi-join.</u>

Natural Join

- A commonly used join operator where the matching condition is equality (equi-join), one of the matching columns is discarded in the result table, and the join columns have the same unqualified name.

- An <u>unqualified name</u> is the column name without the table name.

- The full name of a column includes the table name.
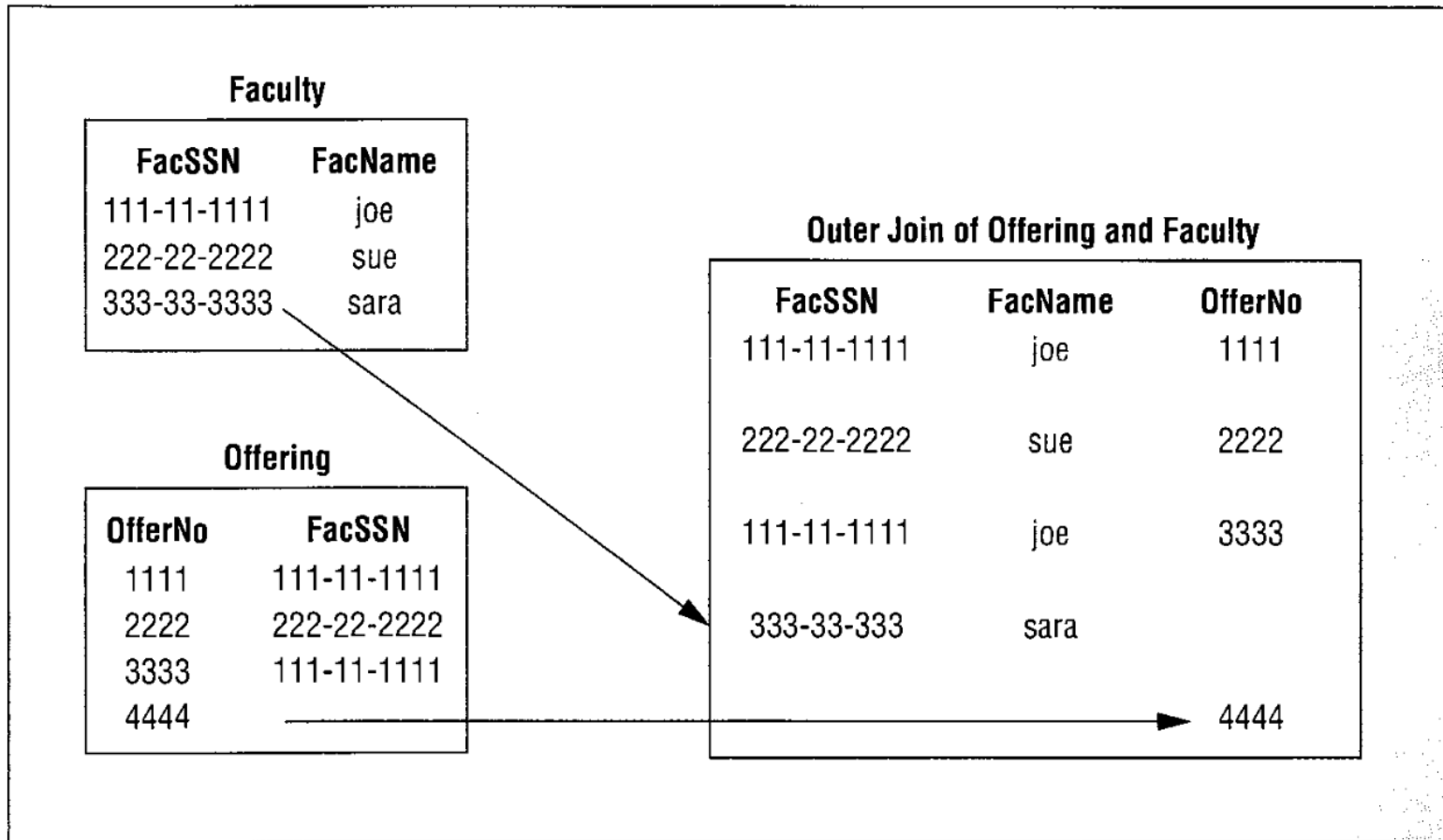
Natural Join

Outer Join

- The result of a join operation includes the rows matching on the join condition.

- Sometimes it is useful to include both matching and nonmatching rows.

- For example, you may want to know offerings that have an assigned instructor as well as offerings without an assigned instructor.

- In these situations, the outer join operator is useful.

Outer Join

- The outer join operator provides the ability to preserve nonmatching rows in the result as well as to include the matching rows.

Outer Join

Faculty

FacSSN	FacName
111-11-1111	joe
222-22-2222	sue
333-33-3333	sara

Offering

OfferNo	FacSSN
1111	111-11-1111
2222	222-22-2222
3333	111-11-1111
4444	

Outer Join of Offering and Faculty

FacSSN	FacName	OfferNo
111-11-1111	joe	1111
222-22-2222	sue	2222
111-11-1111	joe	3333
333-33-333	sara	
		4444

Outer Join

- Each table has one row that does not match any row in the other table.

- The third row of Faculty and the fourth row of Offering do not have matching rows in the other table.

- For nonmatching rows, null values are used to complete the column values in the other table.

Full Vs. One-sided Outer Join

- The outer join operator has two variations.
  - The full outer join preserves nonmatching rows from both input tables. It is an operator that produces the matching rows (the join part) as well as the nonmatching rows from both input tables.
  - Because it is sometimes useful to preserve the non- matching rows from just one input table, the one-sided outer join operator has been devised. an operator that produces the matching rows (the join part) as well as the nonmatching rows from the designated input table.

Full Outer Join

FacSSN	FacLastName	FacDept	FacRank
098-76-5432	VINCE	MS	ASST
543-21-0987	EMMANUEL	MS	PROF
876-54-3210	COLAN	MS	ASST

StdSSN	StdLastName	StdMajor	StdClass
123-45-6789	WELLS	IS	FR
124-56-7890	NORBERT	FIN	JR
876-54-3210	COLAN	MS	SR

StdSSN	StdLastName	StdMajor	StdClass	FacSSN	FacLastName	FacDept	FacRank
123-45-6789	WELLS	IS	FR				
124-56-7890	NORBERT	FIN	JR				
876-54-3210	COLAN	MS	SR	876-54-3210	COLAN	MS	ASST
				098-76-5432	VINCE	MS	ASST
				543-21-0987	EMMANUEL	MS	PROF

One-sided Outer Join

FacSSN	FacLastName	FacDept	FacRank
098-76-5432	VINCE	MS	ASST
543-21-0987	EMMANUEL	MS	PROF
876-54-3210	COLAN	MS	ASST

OfferNo	CourseNo	OffTerm	FacSSN
1111	IS320	SUMMER	
1234	IS320	FALL	098-76-5432
2222	IS460	SUMMER	
3333	IS320	SPRING	098-76-5432
4444	IS320	SPRING	543-21-0987

OfferNo	CourseNo	OffTerm	Offering.FacSSN	Faculty.FacSSN	FacLastName	FacDept	FacRank
1111	IS320	SUMMER					
1234	IS320	FALL	098-76-5432	098-76-5432	VINCE	MS	ASST
2222	IS460	SUMMER					
3333	IS320	SPRING	098-76-5432	098-76-5432	VINCE	MS	ASST
4444	IS320	SPRING	543-21-0987	543-21-0987	EMMANUEL	MS	PROF

Union/Intersection/Difference
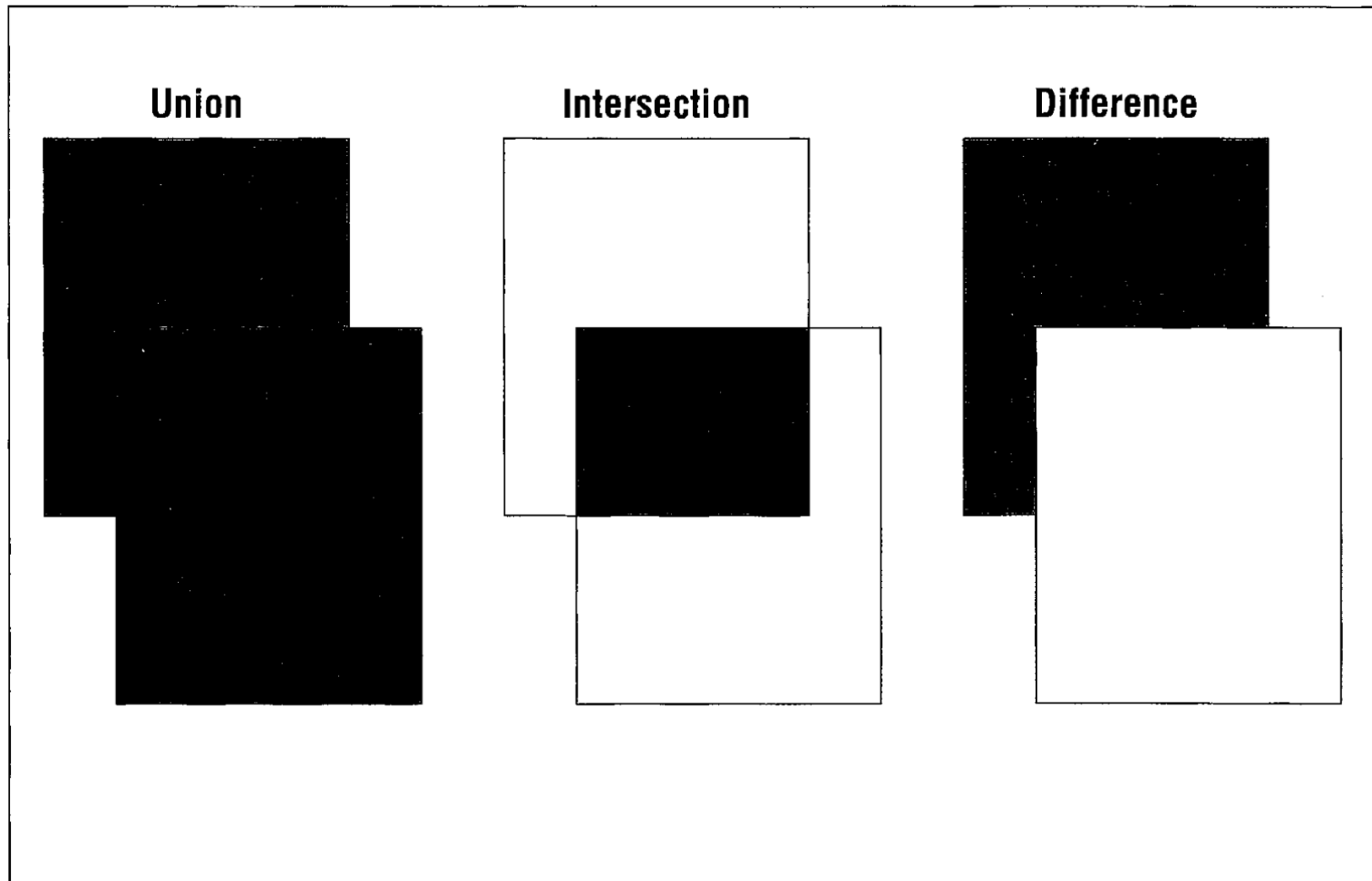
Union/Intersection/Difference

- A union operation retrieves all the rows in either table. For example, a union operator applied to two student tables at different universities can find all student rows.

- An intersection operation retrieves just the common rows. For example, an intersection operation can determine the students attending both universities.

- A difference operation retrieves the rows in the first table but not in the second table. For example, a difference operation can determine the students attending only one university.

Union Compatibility

- A requirement on the input tables for the traditional set operators.

- Each table must have the same number of columns and each corresponding column must have a compatible data type.

Summarize Operator

- An operator that produces a table with rows that summarize the rows of the input table.

- Aggregate functions are used to summarize the rows of the input table.
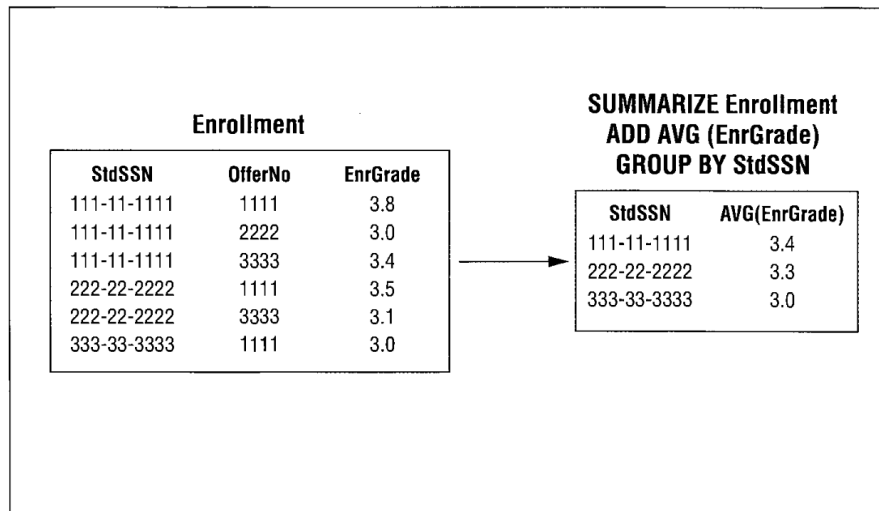
Summarize Operator

- An operator that produces a table in which the values of a column from one input table are associated with all the values from a column of a second input table.

OfferNo	StdSSN
1234	123-45-6789
1234	234-56-7890
4235	123-45-6789
4235	234-56-7890
4235	124-56-7890
6321	124-56-7890

StdSSN
123-45-6789
124-56-7890
234-56-7890

OfferNo
4235