Normalization

CS 341 Database Systems

Database Tables and Normalization

Table is basic building block in database design

Normalization is process for assigning attributes to entities such that

- Reduces data redundancies
- Helps eliminate data anomalies
- Produces controlled redundancies to link tables

Normalization stages

- 1NF - First normal form
- 2NF - Second normal form
- 3NF - Third normal form
- 4NF - Fourth normal form
- BCNF – Boyce Codd normal form
- 5NF – Fifth normal form

Informal Normalization Guidelines

- Attribute types from multiple entity types should not be combined in a single relation
- Avoid excessive amount of *NULL* values in a relation

Normalization Forms

Normal Form	Characteristic
First normal form (1NF)	Table format, no repeating groups, and PK identified
Second normal form (2NF)	1NF and no partial dependencies
Third normal form (3NF)	2NF and no transitive dependencies
Boyce-Codd normal form (BCNF)	3NF and every determinant is a candidate key (special case of 3NF)
Fourth normal form (4NF)	BCNF and no independent multivalued dependencies
Fifth normal form (5NF or PJNF)	4NF and cannot have lossless decomposition into smaller tables

Table 6.1 A Sample Report Layout

Project Number	Project Name	Employee Number	Employee Name	Job Class	Charge/Hour	Hours Billed	Total Charge
15	Evergreen	103	June E. Arbough	Elec. Engineer	$ 84.50	23.8	$ 2,011.10
		101	John G. News	Database Designer	$105.00	19.4	$ 2,037.00
		105	Alice K. Johnson *	Database Designer	$105.00	35.7	$ 3,748.50
		106	William Smithfield	Programmer	$ 35.75	12.6	$ 450.45
		102	David H. Senior	Systems Analyst	$ 96.75	23.8	$ 2,302.65
				Subtotal			**$10,549.70**
18	Amber Wave	114	Annelise Jones	Applications Designer	$ 48.10	24.6	$ 1,183.26
		118	James J. Frommer	General Support	$ 18.36	45.3	$ 831.71
		104	Anne K. Ramoras *	Systems Analyst	$ 96.75	32.4	$ 3,134.70
		112	Darlene M. Smithson	DSS Analyst	$ 45.95	44.0	$ 2,021.80
				Subtotal			**$ 7,171.47**
22	Rolling Tide	105	Alice K. Johnson	Database Designer	$105.00	64.7	$ 6,793.50
		104	Anne K. Ramoras	Systems Analyst	$ 96.75	48.4	$ 4,682.70
		113	Delbert K. Joenbrood *	Applications Designer	$ 48.10	23.6	$ 1,135.16
		111	Geoff B. Wabash	Clerical Support	$ 26.87	22.0	$ 591.14
		106	William Smithfield	Programmer	$ 35.75	12.8	$ 457.60
				Subtotal			**$13,660.10**
25	Starflight	107	Maria D. Alonzo	Programmer	$ 35.75	24.6	$ 879.45
		115	Travis B. Bawangi	Systems Analyst	$ 96.75	45.8	$ 4,431.15
		101	John G. News *	Database Designer	$105.00	56.3	$ 5,911.50
		114	Annelise Jones	Applications Designer	$ 48.10	33.1	$ 1,592.11
		108	Ralph B. Washington	Systems Analyst	$ 96.75	23.6	$ 2,283.30
		118	James J. Frommer	General Support	$ 18.36	30.5	$ 559.98
		112	Darlene M. Smithson	DSS Analyst	$ 45.95	41.4	$ 1,902.33
				Subtotal			**$17,559.82**
				Total			**$48,941.09**

A closer look

Table 6.1 A Sample Report Layout

Project Number	Project Name	Employee Number	Employee Name	Job Class	Charge/Hour	Hours Billed	Total Charge
15	Evergreen	103	June E. Arbough	Elec. Engineer	$ 84.50	23.8	$ 2,011.10
		101	John G. News	Database Designer	$105.00	19.4	$ 2,037.00
		105	Alice K. Johnson *	Database Designer	$105.00	35.7	$ 3,748.50
		106	William Smithfield	Programmer	$ 35.75	12.6	$ 450.45
		102	David H. Senior	Systems Analyst	$ 96.75	23.8	$ 2,302.65
				Subtotal			**$10,549.70**
18	Amber Wave	114	Annelise Jones	Applications Designer	$ 48.10	24.6	$ 1,183.26
		118	James J. Frommer	General Support	$ 18.36	45.3	$ 831.71
		104	Anne K. Ramoras *	Systems Analyst	$ 96.75	32.4	$ 3,134.70
		112	Darlene M. Smithson	DSS Analyst	$ 45.95	44.0	$ 2,021.80
				Subtotal			**$ 7,171.47**
22	Rolling Tide	105	Alice K. Johnson	Database Designer	$105.00	64.7	$ 6,793.50
		104	Anne K. Ramoras	Systems Analyst	$ 96.75	48.4	$ 4,682.70
		113	Delbert K. Joenbrood *	Applications Designer	$ 48.10	23.6	$ 1,135.16

Need for Normalization

- The project number (PROJ_NUM) is intended to be primary key or at least part of a PK, but it contains NULLs
- Table entries invite data inconsistencies
- Table displays data anomalies

Data Anomalies

Update

Insertion

Deletion

Update Anomalies

Modifying the JOB_CLASS for employee number 104 requires (potentially) many alterations, one for each EMP_NUM = 104.

Table 6.1 A Sample Report Layout

Project Number	Project Name	Employee Number	Employee Name	Job Class	Charge/Hour	Hours Billed	Total Charge
15	Evergreen	103	June E. Arbough	Elec. Engineer	$ 84.50	23.8	$ 2,011.10
		101	John G. News	Database Designer	$105.00	19.4	$ 2,037.00
		105	Alice K. Johnson *	Database Designer	$105.00	35.7	$ 3,748.50
		106	William Smithfield	Programmer	$ 35.75	12.6	$ 450.45
		102	David H. Senior	Systems Analyst	$ 96.75	23.8	$ 2,302.65
				Subtotal			**$10,549.70**
18	Amber Wave	114	Annelise Jones	Applications Designer	$ 48.10	24.6	$ 1,183.26
		118	James J. Frommer	General Support	$ 18.36	45.3	$ 831.71
		104	Anne K. Ramoras *	Systems Analyst	$ 96.75	32.4	$ 3,134.70
		112	Darlene M. Smithson	DSS Analyst	$ 45.95	44.0	$ 2,021.80
				Subtotal			**$ 7,171.47**
22	Rolling Tide	105	Alice K. Johnson	Database Designer	$105.00	64.7	$ 6,793.50
		104	Anne K. Ramoras	Systems Analyst	$ 96.75	48.4	$ 4,682.70
		113	Delbert K. Joenbrood *	Applications Designer	$ 48.10	23.6	$ 1,135.16

Insertion Anomalies

Just to complete a row definition, a new employee must be assigned to a project. If the employee is not yet assigned, a phantom project must be created to complete the employee data entry.

Table 6.1 A Sample Report Layout

Project Number	Project Name	Employee Number	Employee Name	Job Class	Charge/Hour	Hours Billed	Total Charge
15	Evergreen	103	June E. Arbough	Elec. Engineer	$ 84.50	23.8	$ 2,011.10
		101	John G. News	Database Designer	$105.00	19.4	$ 2,037.00
		105	Alice K. Johnson *	Database Designer	$105.00	35.7	$ 3,748.50
		106	William Smithfield	Programmer	$ 35.75	12.6	$ 450.45
		102	David H. Senior	Systems Analyst	$ 96.75	23.8	$ 2,302.65
				Subtotal			**$10,549.70**
18	Amber Wave	114	Annelise Jones	Applications Designer	$ 48.10	24.6	$ 1,183.26
		118	James J. Frommer	General Support	$ 18.36	45.3	$ 831.71
		104	Anne K. Ramoras *	Systems Analyst	$ 96.75	32.4	$ 3,134.70
		112	Darlene M. Smithson	DSS Analyst	$ 45.95	44.0	$ 2,021.80
				Subtotal			**$ 7,171.47**
22	Rolling Tide	105	Alice K. Johnson	Database Designer	$105.00	64.7	$ 6,793.50

Deletion Anomalies

Suppose that only one employee is associated with a given project. If that employee leaves the company and the employee data are deleted, the project information will also be deleted.

Table 6.1 A Sample Report Layout

Project Number	Project Name	Employee Number	Employee Name	Job Class	Charge/Hour	Hours Billed	Total Charge
15	Evergreen	103	June E. Arbough	Elec. Engineer	$ 84.50	23.8	$ 2,011.10
		101	John G. News	Database Designer	$105.00	19.4	$ 2,037.00
		105	Alice K. Johnson *	Database Designer	$105.00	35.7	$ 3,748.50
		106	William Smithfield	Programmer	$ 35.75	12.6	$ 450.45
		102	David H. Senior	Systems Analyst	$ 96.75	23.8	$ 2,302.65
				Subtotal			**$10,549.70**
18	Amber Wave	114	Annelise Jones	Applications Designer	$ 48.10	24.6	$ 1,183.26
		118	James J. Frommer	General Support	$ 18.36	45.3	$ 831.71
		104	Anne K. Ramoras *	Systems Analyst	$ 96.75	32.4	$ 3,134.70
		112	Darlene M. Smithson	DSS Analyst	$ 45.95	44.0	$ 2,021.80
				Subtotal			**$ 7,171.47**
22	Rolling Tide	105	Alice K. Johnson	Database Designer	$105.00	64.7	$ 6,793.50

Normalization Process

- The objective of normalization is to ensure that each table conforms to the concept of well-formed relations, that is, tables that have the following characteristics:
  - Each table represents a single subject.
  - No data item will be unnecessarily stored in more than one table (in short, tables have minimum *controlled* redundancy).
  - All nonprime attributes in a table are dependent on the primary key—the entire primary key and nothing but the primary key.
  - Each table is void of insertion, update, or deletion anomalies ensuring the integrity and consistency of the data.

Unnormalized Form (UNF)

- A table that contains one or more repeating groups.

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	84.50	23.8
		101	John G. News	Database Designer	105.00	19.4
		105	Alice K. Johnson *	Database Designer	105.00	35.7
		106	William Smithfield	Programmer	35.75	12.6
		102	David H. Senior	Systems Analyst	96.75	23.8
18	Amber Wave	114	Annelise Jones	Applications Designer	48.10	24.6
		118	James J. Frommer	General Support	18.36	45.3
		104	Anne K. Ramoras *	Systems Analyst	96.75	32.4
		112	Darlene M. Smithson	DSS Analyst	45.95	44.0
22	Rolling Tide	105	Alice K. Johnson	Database Designer	105.00	64.7
		104	Anne K. Ramoras	Systems Analyst	96.75	48.4
		113	Delbert K. Joenbrood *	Applications Designer	48.10	23.6
		111	Geoff B. Wabash	Clerical Support	26.87	22.0
		106	William Smithfield	Programmer	35.75	12.8
25	Starflight	107	Maria D. Alonzo	Programmer	35.75	24.6
		115	Travis B. Bawangi	Systems Analyst	96.75	45.8
		101	John G. News *	Database Designer	105.00	56.3
		114	Annelise Jones	Applications Designer	48.10	33.1
		108	Ralph B. Washington	Systems Analyst	96.75	23.6
		118	James J. Frommer	General Support	18.36	30.5
		112	Darlene M. Smithson	DSS Analyst	45.95	41.4

Figure 6.1 Base Data for a Construction Company Report

Table name: RPT_FORMAT **Database name: Ch06_ConstructCo**

PROJ_NUM	PROJECT_NAME	EMP_NUMBER	EMP_NAME	JOB_CLASS	CHARGE_HOUR	HOURS_BILLED
15	Evergreen	103,101,105, 106, 102	June E. Arbough, John G. News, Alice K. Johnson *, William Smithfield, David H. Senior	Elec. Engineer, Database Designer, Database Designer, Programmer, Systems Analyst	85.5, 105., 105., 35.75, 98.75	23.8, 19.4, 35.7, 12.6, 23.8
18	Amber Wave	114, 118, 104, 112	Annelise Jones, James J. Frommer, Anne K. Ramoras *, Darlene M. Smithson	Applications Designer, General Support, Systems Analyst, DSS Analyst	48.1, 18.36, 96.75, 45.95	25.6, 45.3, 32.4, 45.
22	Rolling Tide	105, 104, 113, 111, 106	Alice K. Johnson, Anne K. Ramoras, Delbert K. Joenbrood *, Geoff B. Wabash, William Smithfield	DB Designer, Systems Analyst, Applications Designer, Clerical Support, Programmer	105., 96.75, 48.1, 26.87, 35.75	65.7, 48.4, 23.6, 22., 12.8
25	Starflight	107, 115, 101, 114, 108, 118, 112	Maria D. Alonzo, Travis B. Bawangi, John G. News *, Annelise Jones, Ralph B. Washington, James J. Frommer, Darlene M. Smithson	Programmer, Systems Analyst, Database Design, Applications Designer, Systems Analyst, General Support, DSS Analyst	35.75, 96.75, 105., 48.1, 96.75, 18.36, 45.95	25.6, 45.8, 56.3, 33.1, 23.6, 30.5, 41.4

Functional Dependencies

- A functional dependency X → Y, between two sets of attribute types X and Y implies that a value of *X uniquely determines a value of Y*
  - there is a functional dependency from X to Y or Y is functionally dependent on X
- X → Y holds if whenever two tuples have the same value for X, they must have the same value for Y
- Example: Consider r(A,B) with the following instance of r.

A	B
1	4
1	5
3	7

On this instance,
$A \rightarrow B$ does **NOT** hold,
but $B \rightarrow A$ does hold.

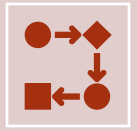
Partial Vs Transitive Dependency

- **Partial**
  - Dependency on part of composite primary key
- **Transitive**
  - One nonprime attribute depends on another nonprime attribute
  - A transitive dependency exists when there are functional dependencies such that X → Y, Y → Z, and X is the primary key.
    In that case, the dependency X → Z is a transitive dependency because *X determines the value of Z via Y.*

First Normal Form (1 NF)

- The first normal form (1 NF) states that every attribute type of a relation must be **atomic** and **single valued**
  - No composite or multivalued attribute types (domain constraint!)
- SUPPLIER(SUPNR, NAME(FIRST NAME, LAST NAME), SUPSTATUS)
- SUPPLIER(SUPNR, FIRST NAME, LAST NAME, SUPSTATUS)

Conversion to 1NF



Step 1: Eliminate the Repeating Groups

Step 2: Identify the Primary Key

Step 3: Identify All Dependencies

Unnormalized Form (UNF)

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	84.50	23.8
		101	John G. News	Database Designer	105.00	19.4
		105	Alice K. Johnson *	Database Designer	105.00	35.7
		106	William Smithfield	Programmer	35.75	12.6
		102	David H. Senior	Systems Analyst	96.75	23.8
18	Amber Wave	114	Annelise Jones	Applications Designer	48.10	24.6
		118	James J. Frommer	General Support	18.36	45.3
		104	Anne K. Ramoras *	Systems Analyst	96.75	32.4
		112	Darlene M. Smithson	DSS Analyst	45.95	44.0
22	Rolling Tide	105	Alice K. Johnson	Database Designer	105.00	64.7
		104	Anne K. Ramoras	Systems Analyst	96.75	48.4
		113	Delbert K. Joenbrood *	Applications Designer	48.10	23.6
		111	Geoff B. Wabash	Clerical Support	26.87	22.0
		106	William Smithfield	Programmer	35.75	12.8
25	Starflight	107	Maria D. Alonzo	Programmer	35.75	24.6
		115	Travis B. Bawangi	Systems Analyst	96.75	45.8
		101	John G. News *	Database Designer	105.00	56.3
		114	Annelise Jones	Applications Designer	48.10	33.1
		108	Ralph B. Washington	Systems Analyst	96.75	23.6
		118	James J. Frommer	General Support	18.36	30.5
		112	Darlene M. Smithson	DSS Analyst	45.95	41.4

UNF

PROJ_NUM	PROJECT_NAME	EMP_NUMBER	EMP_NAME	JOB_CLASS	CHARGE_HOUR	HOURS_BILLED
15	Evergreen	103,101,105, 106, 102	June E. Arbough, John G. News, Alice K. Johnson *, William Smithfield, David H. Senior	Elec. Engineer, Database Designer, Database Designer, Programmer, Systems Analyst	85.5, 105., 105., 35.75, 98.75	23.8, 19.4, 35.7, 12.6, 23.8
18	Amber Wave	114, 118, 104, 112	Annelise Jones, James J. Frommer, Anne K. Ramoras *, Darlene M. Smithson	Applications Designer, General Support, Systems Analyst, DSS Analyst	48.1, 18.36, 96.75, 45.95	25.6, 45.3, 32.4, 45.
22	Rolling Tide	105, 104, 113, 111, 106	Alice K. Johnson, Anne K. Ramoras, Delbert K. Joenbrood *, Geoff B. Wabash, William Smithfield	DB Designer, Systems Analyst, Applications Designer, Clerical Support, Programmer	105., 96.75, 48.1, 26.87, 35.75	65.7, 48.4, 23.6, 22., 12.8
25	Starflight	107, 115, 101, 114, 108, 118, 112	Maria D. Alonzo, Travis B. Bawangi, John G. News *, Annelise Jones, Ralph B. Washington, James J. Frommer, Darlene M. Smithson	Programmer, Systems Analyst, Database Design, Applications Designer, Systems Analyst, General Support, DSS Analyst	35.75, 96.75, 105., 48.1, 96.75, 18.36, 45.95	25.6, 45.8, 56.3, 33.1, 23.6, 30.5, 41.4

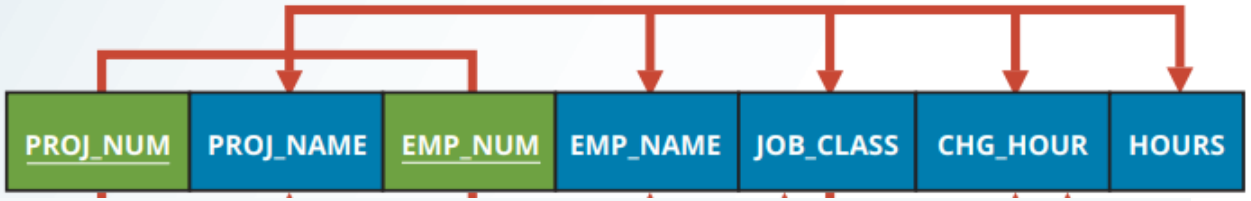
Identify Primary Key

Table name: DATA_ORG_1NF **Database name: Ch06_ConstructCo**

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	84.50	23.8
15	Evergreen	101	John G. News	Database Designer	105.00	19.4
15	Evergreen	105	Alice K. Johnson *	Database Designer	105.00	35.7
15	Evergreen	106	William Smithfield	Programmer	35.75	12.6
15	Evergreen	102	David H. Senior	Systems Analyst	96.75	23.8
18	Amber Wave	114	Annelise Jones	Applications Designer	48.10	24.6
18	Amber Wave	118	James J. Frommer	General Support	18.36	45.3
18	Amber Wave	104	Anne K. Ramoras *	Systems Analyst	96.75	32.4
18	Amber Wave	112	Darlene M. Smithson	DSS Analyst	45.95	44.0
22	Rolling Tide	105	Alice K. Johnson	Database Designer	105.00	64.7
22	Rolling Tide	104	Anne K. Ramoras	Systems Analyst	96.75	48.4
22	Rolling Tide	113	Delbert K. Joenbrood *	Applications Designer	48.10	23.6
22	Rolling Tide	111	Geoff B. Wabash	Clerical Support	26.87	22.0
22	Rolling Tide	106	William Smithfield	Programmer	35.75	12.8
25	Starflight	107	Maria D. Alonzo	Programmer	35.75	24.6
25	Starflight	115	Travis B. Bawangi	Systems Analyst	96.75	45.8
25	Starflight	101	John G. News *	Database Designer	105.00	56.3
25	Starflight	114	Annelise Jones	Applications Designer	48.10	33.1
25	Starflight	108	Ralph B. Washington	Systems Analyst	96.75	23.6
25	Starflight	118	James J. Frommer	General Support	18.36	30.5
25	Starflight	112	Darlene M. Smithson	DSS Analyst	45.95	41.4

First Normal form (1NF)

Figure 6.3 First Normal Form (1NF) Dependency Diagram



1NF (PROJ_NUM, EMP_NUM, PROJ_NAME, EMP_NAME, JOB_CLASS, CHG_HOURS, HOURS)

Note

The term **first normal form (1NF)** describes the tabular format that conforms to the definition of a relational table in which:

- All of the key attributes are defined.
- There are no repeating groups in the table. In other words, each row/column intersection contains one and only one value, not a set of values.
- All attributes are dependent on the primary key.

Converted to 1NF - Proceed to identify dependencies

Figure 6.2 A Table in First Normal Form

Table name: DATA_ORG_1NF **Database name: Ch06_ConstructCo**

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	84.50	23.8
15	Evergreen	101	John G. News	Database Designer	105.00	19.4
15	Evergreen	105	Alice K. Johnson *	Database Designer	105.00	35.7
15	Evergreen	106	William Smithfield	Programmer	35.75	12.6
15	Evergreen	102	David H. Senior	Systems Analyst	96.75	23.8
18	Amber Wave	114	Annelise Jones	Applications Designer	48.10	24.6
18	Amber Wave	118	James J. Frommer	General Support	18.36	45.3
18	Amber Wave	104	Anne K. Ramoras *	Systems Analyst	96.75	32.4
18	Amber Wave	112	Darlene M. Smithson	DSS Analyst	45.95	44.0
22	Rolling Tide	105	Alice K. Johnson	Database Designer	105.00	64.7
22	Rolling Tide	104	Anne K. Ramoras	Systems Analyst	96.75	48.4
22	Rolling Tide	113	Delbert K. Joenbrood *	Applications Designer	48.10	23.6
22	Rolling Tide	111	Geoff B. Wabash	Clerical Support	26.87	22.0
22	Rolling Tide	106	William Smithfield	Programmer	35.75	12.8
25	Starflight	107	Maria D. Alonzo	Programmer	35.75	24.6
25	Starflight	115	Travis B. Bawangi	Systems Analyst	96.75	45.8
25	Starflight	101	John G. News *	Database Designer	105.00	56.3
25	Starflight	114	Annelise Jones	Applications Designer	48.10	33.1
25	Starflight	108	Ralph B. Washington	Systems Analyst	96.75	23.6
25	Starflight	118	James J. Frommer	General Support	18.36	30.5
25	Starflight	112	Darlene M. Smithson	DSS Analyst	45.95	41.4

Partial
PROJ_NUM

PARTIAL DEPENDENCIES:
(PROJ_NUM → PROJ_NAME)

Figure 6.2 A Table in First Normal Form

Table name: DATA_ORG_1NF **Database name: Ch06_ConstructCo**

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	84.50	23.8
15	Evergreen	101	John G. News	Database Designer	105.00	19.4
15	Evergreen	105	Alice K. Johnson *	Database Designer	105.00	35.7
15	Evergreen	106	William Smithfield	Programmer	35.75	12.6
15	Evergreen	102	David H. Senior	Systems Analyst	96.75	23.8
18	Amber Wave	114	Annelise Jones	Applications Designer	48.10	24.6
18	Amber Wave	118	James J. Frommer	General Support	18.36	45.3
18	Amber Wave	104	Anne K. Ramoras *	Systems Analyst	96.75	32.4
18	Amber Wave	112	Darlene M. Smithson	DSS Analyst	45.95	44.0
22	Rolling Tide	105	Alice K. Johnson	Database Designer	105.00	64.7
22	Rolling Tide	104	Anne K. Ramoras	Systems Analyst	96.75	48.4
22	Rolling Tide	113	Delbert K. Joenbrood *	Applications Designer	48.10	23.6
22	Rolling Tide	111	Geoff B. Wabash	Clerical Support	26.87	22.0
22	Rolling Tide	106	William Smithfield	Programmer	35.75	12.8
25	Starflight	107	Maria D. Alonzo	Programmer	35.75	24.6
25	Starflight	115	Travis B. Bawangi	Systems Analyst	96.75	45.8
25	Starflight	101	John G. News *	Database Designer	105.00	56.3
25	Starflight	114	Annelise Jones	Applications Designer	48.10	33.1
25	Starflight	108	Ralph B. Washington	Systems Analyst	96.75	23.6
25	Starflight	118	James J. Frommer	General Support	18.36	30.5
25	Starflight	112	Darlene M. Smithson	DSS Analyst	45.95	41.4

Partial
EMP_NUM

PARTIAL DEPENDENCIES:
(PROJ_NUM ⇒ PROJ_NAME)
(EMP_NUM ⇒ EMP_NAME, JOB_CLASS, CHG_HOUR)

Figure 6.2 A Table in First Normal Form

Table name: DATA_ORG_1NF **Database name: Ch06_ConstructCo**

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	84.50	23.8
15	Evergreen	101	John G. News	Database Designer	105.00	19.4
15	Evergreen	105	Alice K. Johnson *	Database Designer	105.00	35.7
15	Evergreen	106	William Smithfield	Programmer	35.75	12.6
15	Evergreen	102	David H. Senior	Systems Analyst	96.75	23.8
18	Amber Wave	114	Annelise Jones	Applications Designer	48.10	24.6
18	Amber Wave	118	James J. Frommer	General Support	18.36	45.3
18	Amber Wave	104	Anne K. Ramoras *	Systems Analyst	96.75	32.4
18	Amber Wave	112	Darlene M. Smithson	DSS Analyst	45.95	44.0
22	Rolling Tide	105	Alice K. Johnson	Database Designer	105.00	64.7
22	Rolling Tide	104	Anne K. Ramoras	Systems Analyst	96.75	48.4
22	Rolling Tide	113	Delbert K. Joenbrood *	Applications Designer	48.10	23.6
22	Rolling Tide	111	Geoff B. Wabash	Clerical Support	26.87	22.0
22	Rolling Tide	106	William Smithfield	Programmer	35.75	12.8
25	Starflight	107	Maria D. Alonzo	Programmer	35.75	24.6
25	Starflight	115	Travis B. Bawangi	Systems Analyst	96.75	45.8
25	Starflight	101	John G. News *	Database Designer	105.00	56.3
25	Starflight	114	Annelise Jones	Applications Designer	48.10	33.1
25	Starflight	108	Ralph B. Washington	Systems Analyst	96.75	23.6
25	Starflight	118	James J. Frommer	General Support	18.36	30.5
25	Starflight	112	Darlene M. Smithson	DSS Analyst	45.95	41.4

Transitive JOB_CLASS

TRANSITIVE DEPENDENCY:
(JOB_CLASS → CHG_HOUR)

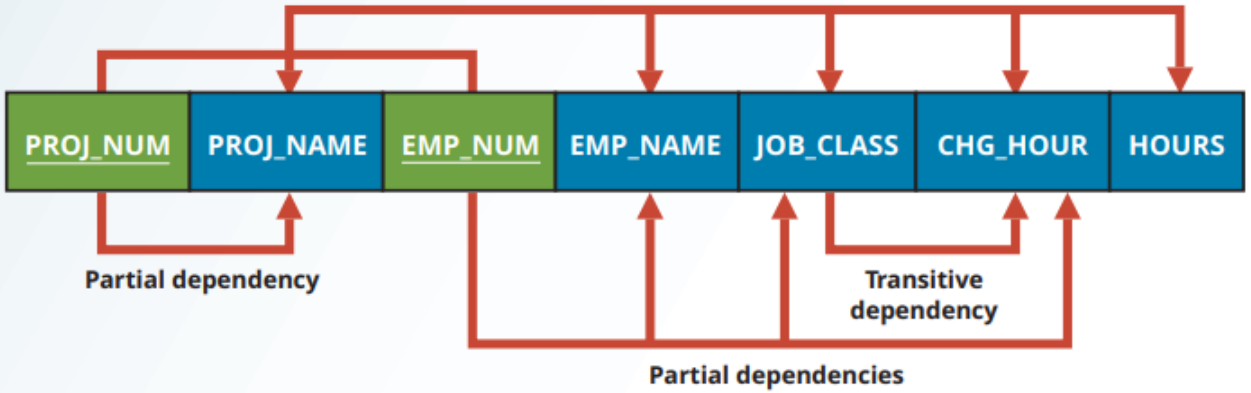
Figure 6.2 A Table in First Normal Form

Table name: DATA_ORG_1NF **Database name: Ch06_ConstructCo**

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	84.50	23.8
15	Evergreen	101	John G. News	Database Designer	105.00	19.4
15	Evergreen	105	Alice K. Johnson *	Database Designer	105.00	35.7
15	Evergreen	106	William Smithfield	Programmer	35.75	12.6
15	Evergreen	102	David H. Senior	Systems Analyst	96.75	23.8
18	Amber Wave	114	Annelise Jones	Applications Designer	48.10	24.6
18	Amber Wave	118	James J. Frommer	General Support	18.36	45.3
18	Amber Wave	104	Anne K. Ramoras *	Systems Analyst	96.75	32.4
18	Amber Wave	112	Darlene M. Smithson	DSS Analyst	45.95	44.0
22	Rolling Tide	105	Alice K. Johnson	Database Designer	105.00	64.7
22	Rolling Tide	104	Anne K. Ramoras	Systems Analyst	96.75	48.4
22	Rolling Tide	113	Delbert K. Joenbrood *	Applications Designer	48.10	23.6
22	Rolling Tide	111	Geoff B. Wabash	Clerical Support	26.87	22.0
22	Rolling Tide	106	William Smithfield	Programmer	35.75	12.8
25	Starflight	107	Maria D. Alonzo	Programmer	35.75	24.6
25	Starflight	115	Travis B. Bawangi	Systems Analyst	96.75	45.8
25	Starflight	101	John G. News *	Database Designer	105.00	56.3
25	Starflight	114	Annelise Jones	Applications Designer	48.10	33.1
25	Starflight	108	Ralph B. Washington	Systems Analyst	96.75	23.6
25	Starflight	118	James J. Frommer	General Support	18.36	30.5
25	Starflight	112	Darlene M. Smithson	DSS Analyst	45.95	41.4

First Normal form (1NF) dependency diagram

Figure 6.3 First Normal Form (1NF) Dependency Diagram



1NF (PROJ_NUM, EMP_NUM, PROJ_NAME, EMP_NAME, JOB_CLASS, CHG_HOURS, HOURS)

PARTIAL DEPENDENCIES:
(PROJ_NUM → PROJ_NAME)
(EMP_NUM → EMP_NAME, JOB_CLASS, CHG_HOUR)

TRANSITIVE DEPENDENCY:
(JOB_CLASS → CHG_HOUR)

Note

The term **first normal form (1NF)** describes the tabular format that conforms to the definition of a relational table in which:

- All of the key attributes are defined.
- There are no repeating groups in the table. In other words, each row/column intersection contains one and only one value, not a set of values.
- All attributes are dependent on the primary key.

Second Normal Form (2 NF)

- A relation R is in the second normal form (2 NF) if it satisfies 1 NF and every non-prime attribute type A in R is **fully functional dependent** on the key of R OR there are **no partial dependencies**
- If the relation is not in second normal form, we must:
  - Decompose it and set up a new relation for each partial key together with its dependent attribute types
  - Keep a relation with the original primary key and any attribute types that are fully functional dependent on it

Conversion to 2NF

1NF (PROJ_NUM, EMP_NUM, PROJ_NAME, EMP_NAME, JOB_CLASS, CHG_HOURS, HOURS)

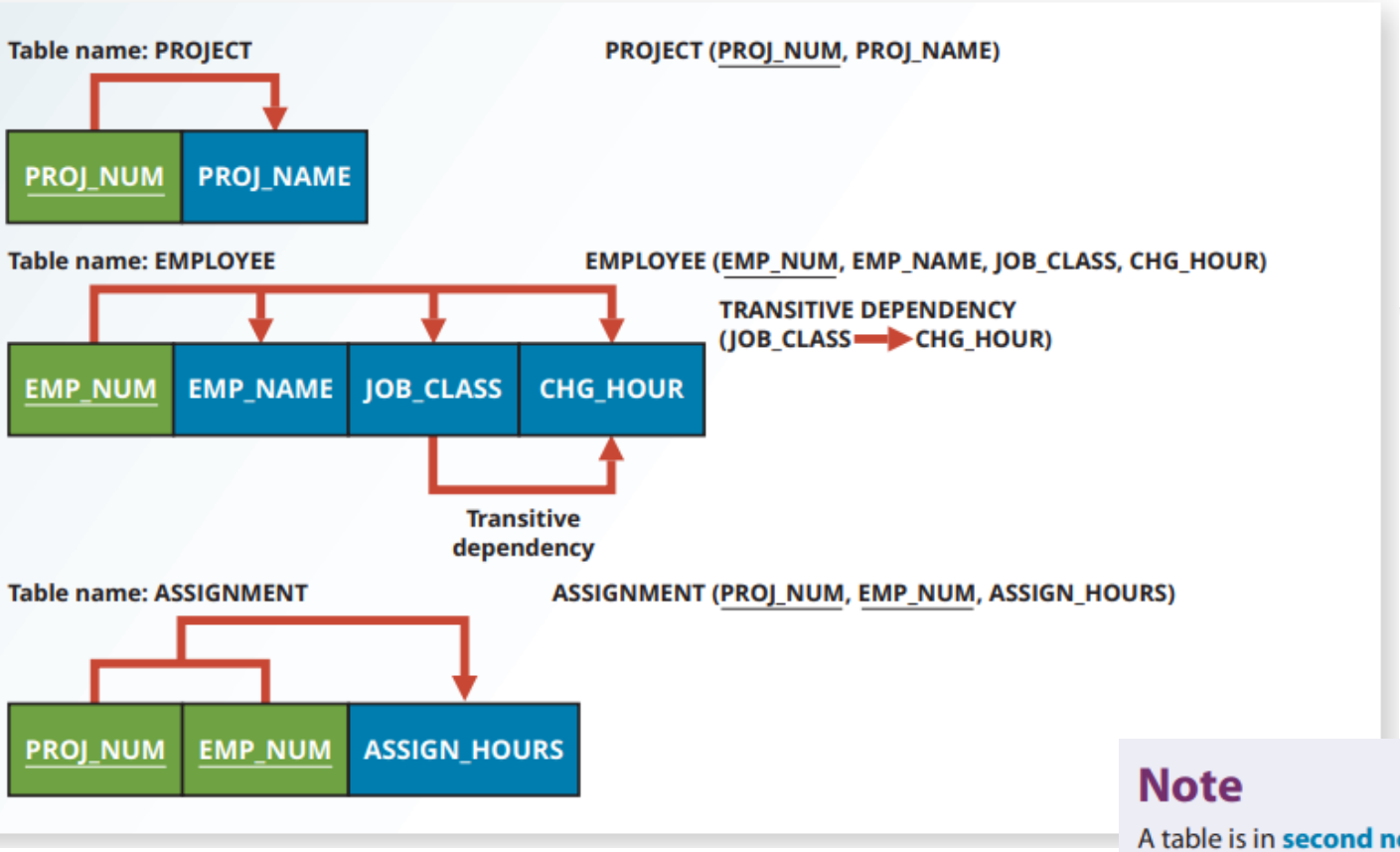
PARTIAL DEPENDENCIES:
(PROJ_NUM → PROJ_NAME)
(EMP_NUM → EMP_NAME, JOB_CLASS, CHG_HOUR)

- Start with 1NF format:
- Write each key component on separate line
- Each component will become key in a new table
- Write dependent attributes after each key

Example:
Proj_Num
Emp_Num
Proj_Num, Emp_Num

PROJECT (PROJ_NUM, PROJ_NAME)
EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS, CHG_HOURS)
ASSIGNMENT (PROJ_NUM, EMP_NUM, ASSIGN_HOURS)

Figure 6.4 Second Normal Form (2NF) Conversion Results



Converted to 2NF

Note

A table is in **second normal form (2NF)** when:

- It is in 1NF

and

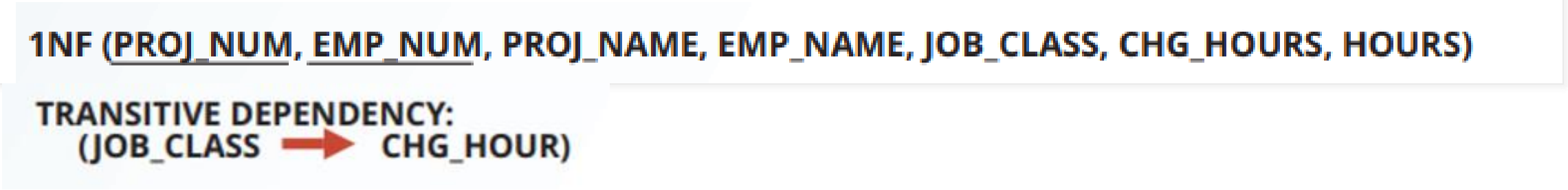
- It includes no partial dependencies; that is, no attribute is dependent on only a portion of the primary key.

It is still possible for a table in 2NF to exhibit transitive dependency. That is, the primary key may rely on one or more nonprime attributes to functionally determine other nonprime attributes, as indicated by a functional dependence among the nonprime attributes.

Third Normal Form (3 NF)

- A functional dependency $X \rightarrow Z$ in a relation R is a transitive dependency if there is a set of attribute types Y that is neither a candidate key nor a subset of any key of R, and both $X \rightarrow Y$ and $Y \rightarrow Z$ hold
- *A relation is in the third normal form (3 NF) if it satisfies 2 NF and no non-prime attribute type of R is transitively dependent on the primary key*
- If the relation is not in third normal form, we need to decompose the relation R and set up a relation that includes the non-key attribute types that functionally determine the other non-key attribute types

Conversion to 3NF



- Create separate table(s) to eliminate transitive functional dependencies

PROJECT (PROJ_NUM, PROJ_NAME)
ASSIGNMENT (PROJ_NUM, EMP_NUM, ASSIGN_HOURS)
EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS)
JOB (JOB_CLASS, CHG_HOUR)

- In 3NF
- Contains no transitive dependencies

Figure 6.5 Third Normal Form (3NF) Conversion Results

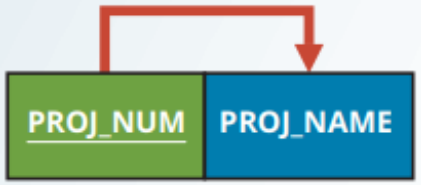


Table name: PROJECT

PROJECT (PROJ_NUM, PROJ_NAME)

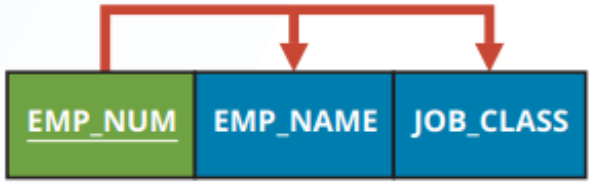


Table name: EMPLOYEE

EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS)

Table name: JOB

JOB (JOB_CLASS, CHG_HOUR)

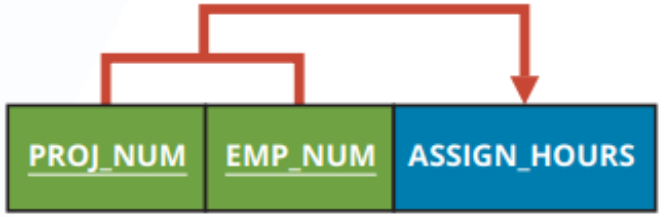


Table name: ASSIGNMENT

ASSIGNMENT (PROJ_NUM, EMP_NUM, ASSIGN_HOURS)

Converted to 3NF

Note

A table is in **third normal form (3NF)** when:

- It is in 2NF

and

- It contains no transitive dependencies.

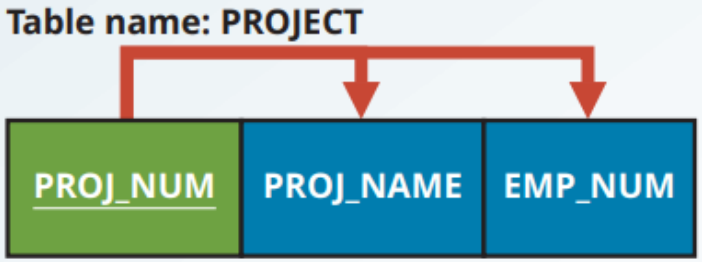
Improving the Design

- *Minimize Data Entry Errors*
  - Job_class to Job_code with Job_class/description as attribute
- *Evaluate Naming Conventions*
  - CHG_HOUR to JOB_CHG_HOUR to indicate belonging to Job table
- *Refine Attribute Atomicity*
  - EMP_NAME to EMP_LNAME, EMP_FNAME, EMP_INITIAL

Improving the Design

- *Identify New Attributes*
  - Consider the real-world scenario and add required attributes e.g EMP_HIREDATE
- *Identify New Relationships*
  - Initial report required modelling project manager so add new relation
- *Refine Primary Keys as Required for Data Granularity*
  - Granularity refers to the level of detail represented by the values stored in a table's row. Data stored at its lowest level of granularity is said to be atomic data. Re-evaluate granularity of ASSIGN_HOURS

Figure 6.6 The Completed Database



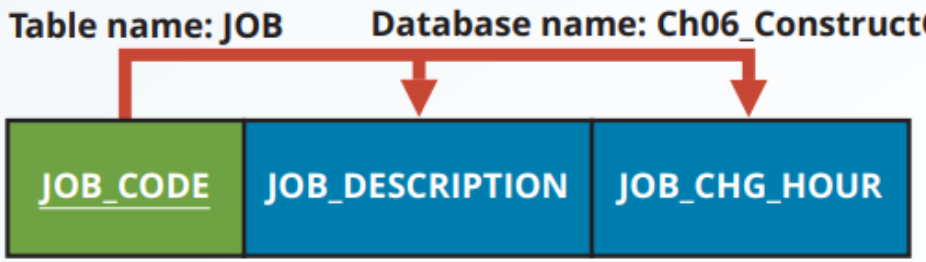


Table name: PROJECT

PROJ_NUM	PROJ_NAME	EMP_NUM
15	Evergreen	105
18	Amber Wave	104
22	Rolling Tide	113
25	Starflight	101

Table name: JOB

JOB_CODE	JOB_DESCRIPTION	JOB_CHG_HOUR
500	Programmer	35.75
501	Systems Analyst	96.75
502	Database Designer	105.00
503	Electrical Engineer	84.50
504	Mechanical Engineer	67.90
505	Civil Engineer	55.78
506	Clerical Support	26.87
507	DSS Analyst	45.95
508	Applications Designer	48.10
509	Bio Technician	34.55
510	General Support	18.36

Table name: ASSIGNMENT

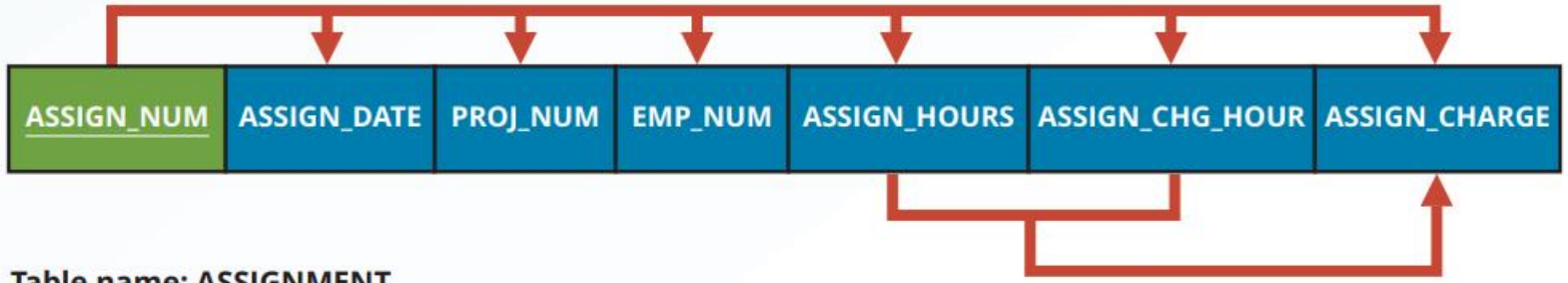


Table name: ASSIGNMENT

ASSIGN_NUM	ASSIGN_DATE	PROJ_NUM	EMP_NUM	ASSIGN_HOURS	ASSIGN_CHG_HOUR	ASSIGN_CHARGE
1001	04-Mar-22	15	103	2.6	84.50	219.70
1002	04-Mar-22	18	118	1.4	18.36	25.70
1003	05-Mar-22	15	101	3.6	105.00	378.00
1004	05-Mar-22	22	113	2.5	48.10	120.25
1005	05-Mar-22	15	103	1.9	84.50	160.55
1006	05-Mar-22	25	115	4.2	96.75	406.35
1007	05-Mar-22	22	105	5.2	105.00	546.00
1008	05-Mar-22	25	101	1.7	105.00	178.50
1009	05-Mar-22	15	105	2.0	105.00	210.00
1010	06-Mar-22	15	102	3.8	96.75	367.65
1011	06-Mar-22	22	104	2.6	96.75	251.55
1012	06-Mar-22	15	101	2.3	105.00	241.50
1013	06-Mar-22	25	114	1.8	48.10	86.58
1014	06-Mar-22	22	111	4.0	26.87	107.48
1015	06-Mar-22	25	114	3.4	48.10	163.54
1016	06-Mar-22	18	112	1.2	45.95	55.14
1017	06-Mar-22	18	118	2.0	18.36	36.72
1018	06-Mar-22	18	104	2.6	96.75	251.55
1019	06-Mar-22	15	103	3.0	84.50	253.50
1020	07-Mar-22	22	105	2.7	105.00	283.50
1021	08-Mar-22	25	108	4.2	96.75	406.35
1022	07-Mar-22	25	114	5.8	48.10	278.98
1023	07-Mar-22	22	106	2.4	35.75	85.80

Figure 6.6 The Completed Database *(Continued)*

Table name: EMPLOYEE **Database name: Ch06_ConstructCo**

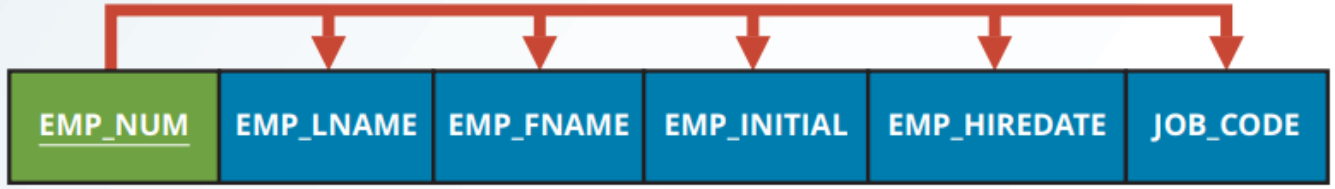


Table name: EMPLOYEE

EMP_NUM	EMP_LNAME	EMP_FNAME	EMP_INITIAL	EMP_HIREDATE	JOB_CODE
101	News	John	G	08-Nov-00	502
102	Senior	David	H	12-Jul-89	501
103	Arbough	June	E	01-Dec-97	503
104	Ramoras	Anne	K	15-Nov-88	501
105	Johnson	Alice	K	01-Feb-94	502
106	Smithfield	William		22-Jun-05	500
107	Alonzo	Maria	D	10-Oct-94	500
108	Washington	Ralph	B	22-Aug-89	501
109	Smith	Larry	W	18-Jul-99	501
110	Olenko	Gerald	A	11-Dec-96	505
111	Wabash	Geoff	B	04-Apr-89	506
112	Smithson	Darlene	M	23-Oct-95	507
113	Joenbrood	Delbert	K	15-Nov-94	508
114	Jones	Annelise		20-Aug-91	508
115	Bawangi	Travis	B	25-Jan-90	501
116	Pratt	Gerald	L	05-Mar-95	510
117	Williamson	Angie	H	19-Jun-94	509
118	Frommer	James	J	04-Jan-06	510

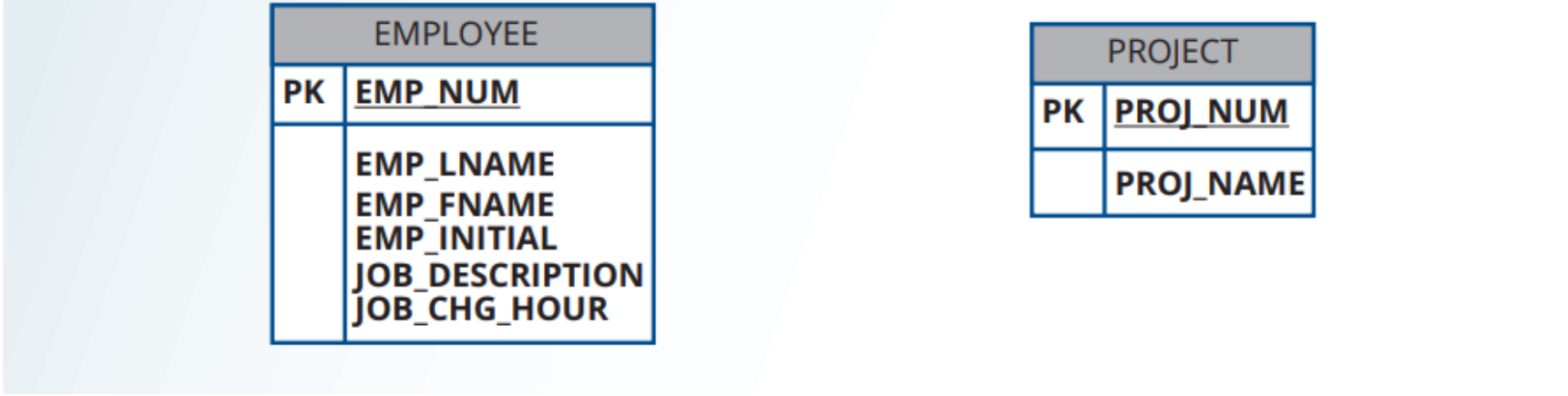
Normalization and Database Design

- Normalization should be part of the design process
- E-R Diagram provides macro view
- Normalization provides micro view of entities
  - Focuses on characteristics of specific entities
  - May yield additional entities
- Difficult to separate normalization from E-R diagramming
- Business rules must be determined

Session 02

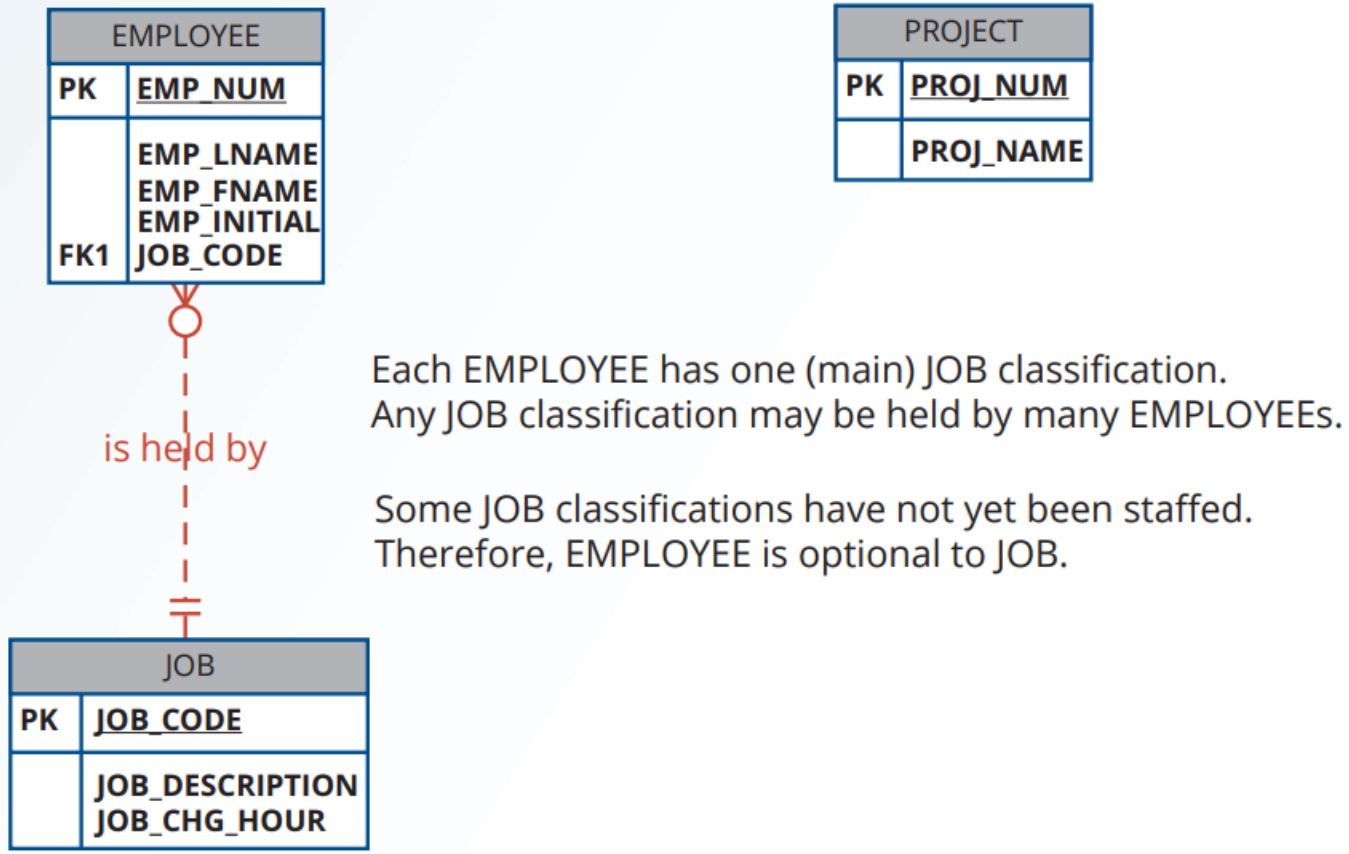
Initial ERD for Contracting Company

Figure 6.14 Initial Contracting Company ERD



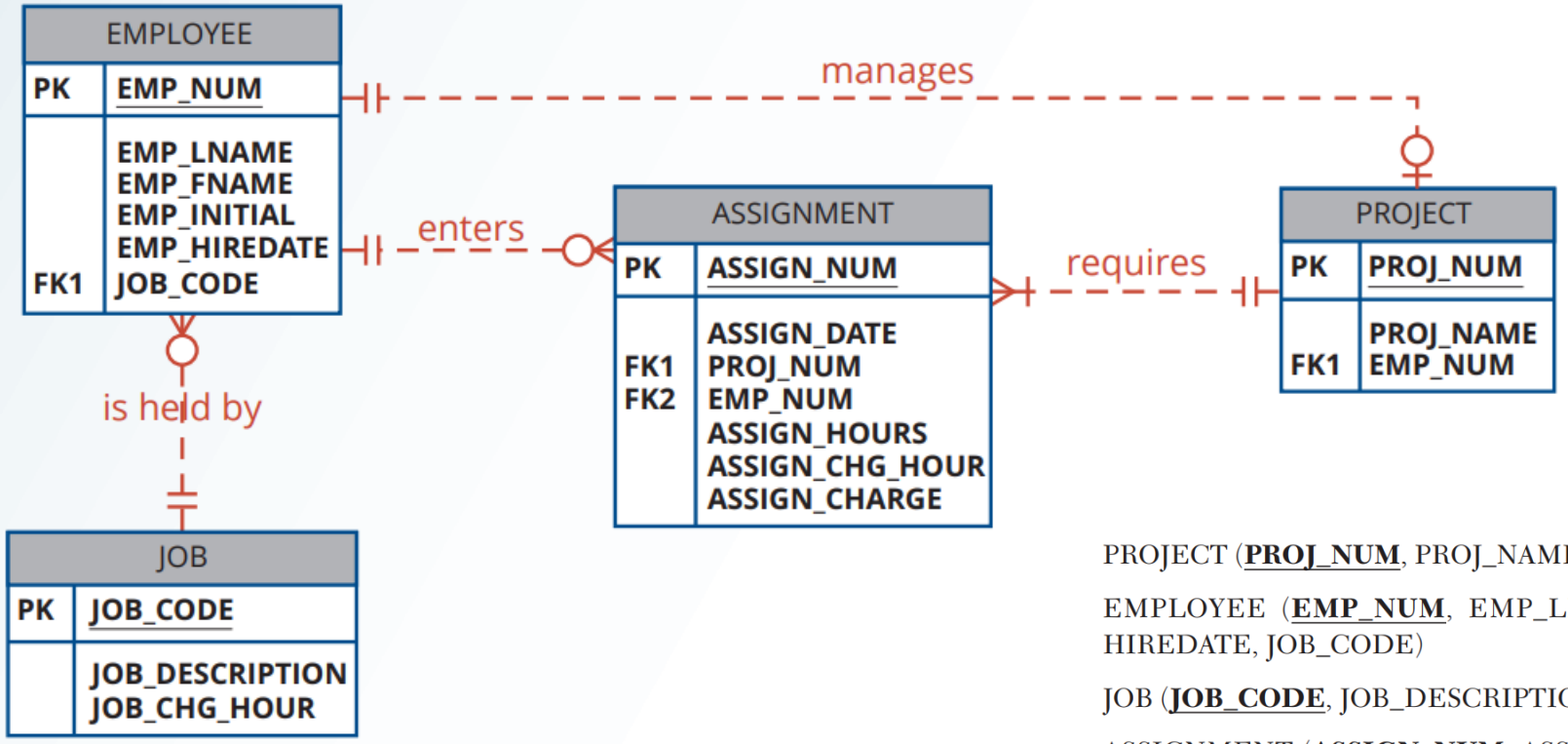
Modified ERD for Contracting Company

Figure 6.15 Modified Contracting Company ERD



Final ERD for Contracting Company

Figure 6.17 Final Contracting Company ERD



PROJECT (**PROJ_NUM**, PROJ_NAME, EMP_NUM)

EMPLOYEE (**EMP_NUM**, EMP_LNAME, EMP_FNAME, EMP_INITIAL, EMP_HIREDATE, JOB_CODE)

JOB (**JOB_CODE**, JOB_DESCRIPTION, JOB_CHG_HOUR)

ASSIGNMENT (**ASSIGN_NUM**, ASSIGN_DATE, PROJ_NUM, EMP_NUM, ASSIGN_HOURS, ASSIGN_CHG_HOUR, ASSIGN_CHARGE)

Example – Normalize to 3 NF

ORDER

Customer No: 001964
Name: Amna Ali
Address: House 101
Block 7 Gulshan-e-Iqbal
Karachi

Order Number: 00012345
Order Date: 25-Sep-2023

Product Number	Product Description	Unit Price	Order Quantity	Line Total
ER101	Eraser	20.00	5	100.00
PN502	Pencil	10.00	10	100.00
PN404	Pen	100.00	1	100.00

Order Total: 300.00

ORDER (**orderNo**, orderDate, custNo, custName, custAdd,
(*prodNo, prodDesc, unitPrice, ordQty, lineTotal****)********, orderTotal

UNF

orderNo	orderDate	custNo	custName	custAdd	()*	orderTotal
00012345	25-Sep-2023	001964	Amna Ali	House 101 Block 7 Gulshan-e-Iqbal Karachi	[(ER101, Eraser, 20.00, 5, 100.00), (PN502, Pencil, 10.00,10, 100.00), (PN404, Pen, 100.00, 1, 100.00)]	300.00

Converting to 1 NF

orderNo	orderDate	custNo	custName	custAdd	()*	orderTotal
00012345	25-Sep-2023	001964	Amna Ali	House 101 Block 7 Gulshan-e-Iqbal Karachi	ER101, Eraser, 20.00, 5, 100.00	300.00
00012345	25-Sep-2023	001964	Amna Ali	House 101 Block 7 Gulshan-e-Iqbal Karachi	PN502, Pencil, 10.00,10, 100.00	300.00
00012345	25-Sep-2023	001964	Amna Ali	House 101 Block 7 Gulshan-e-Iqbal Karachi	PN404, Pen, 100.00, 1, 100.00	300.00

Converting to 1 NF

orderNo	orderDate	custNo	custName	custAdd	prodNo	prodDesc	unitPrice	ordQty	lineTotal	orderTotal
00012345	25-Sep-2023	001964	Amna Ali	House 101 Block 7 Gulshan-e-Iqbal Karachi	ER101	Eraser	20.00	5	100.00	300.00
00012345	25-Sep-2023	001964	Amna Ali	House 101 Block 7 Gulshan-e-Iqbal Karachi	PN502	Pencil	10.00	10	100.00	300.00
00012345	25-Sep-2023	001964	Amna Ali	House 101 Block 7 Gulshan-e-Iqbal Karachi	PN404	Pen	100.00	1	100.00	300.00

Identifying Dependencies

- 1NF

ORDER (**orderNo**, orderDate, custNo, custName, custAdd,
prodNo, prodDesc, unitPrice, ordQty, lineTotal, orderTotal)

Partial dependencies

orderNo → orderDate, custNo, CustName, custAdd,orderTotal

prodNo → prodDesc, unitPrice

Transitive dependencies

custNo → custName, custAdd

Converting to 2NF

Partial dependencies

orderNo → orderDate, custNo, CustName, custAdd, orderTotal

prodNo → prodDesc, unitPrice

ORDER (**orderNo**, orderDate, custNo, custName, custAdd, orderTotal)

PRODUCT (**prodNo**, prodDesc, unitPrice)

ORDER_DETAILS (**orderNo**, **prodNo**, ordQty, lineTotal)

Converting to 3NF

Transitive dependencies

custNo → custName, custAdd

ORDER (**orderNo**, orderDate, custNo, orderTotal)

PRODUCT (**prodNo**, prodDesc, unitPrice)

ORDER_DETAILS (**orderNo**, **prodNo**, ordQty, lineTotal)

CUSTOMER (**custNo**, custName, custAdd)

Normalization

UNF

ORDER (**orderNo**, orderDate, custNo, custName, custAdd,
*(prodNo, prodDesc, unitPrice, ordQty, lineTotal)**, orderTotal

1NF

ORDER (**orderNo**, orderDate, custNo, custName, custAdd,
prodNo, prodDesc, unitPrice, ordQty, lineTotal, orderTotal)

2NF

ORDER (**orderNo**, orderDate, custNo, custName, custAdd, orderTotal)
PRODUCT (**prodNo**, prodDesc, unitPrice,)
ORDER_DETAILS (**orderNo**, **prodNo**, ordQty, lineTotal)

3NF

ORDER (**orderNo**, orderDate, custNo, orderTotal)
PRODUCT (**prodNo**, prodDesc, unitPrice,)
ORDER_DETAILS (**orderNo**, **prodNo**, ordQty, lineTotal)
CUSTOMER (**custNo**, custName, custAdd)