Databases - Normalization

Keys

- take example of an **Employee** table:
  Employee (
  Employee ID,
  FullName,
  SSN,
  DeptID
)

  1. Candidate Key: are individual columns in a table that qualifies for uniqueness of all the rows. Here in Employee table EmployeeID & SSN are Candidate keys.

  2. Primary Key: is the columns you choose to maintain uniqueness in a table. Here in Employee table you can choose either EmployeeID or SSN columns

  3. Alternate Key: Candidate column other the Primary column, like if EmployeeID is PK then SSN would be the Alternate key.

  4. Super Key: If you add any other column/attribute to a Primary Key then it become a super key, like EmployeeID + FullName is a Super Key.

  5. Composite Key: If a table does not have a single column that qualifies for a Candidate key, then you have to select 2 or more columns to make a row unique. Like if there is no EmployeeID or SSN columns, then you can make FullName + DateOfBirth as Composite primary Key. But still there can be a narrow chance of duplicate row.

Database Tables and Normalization

- Edgar F. Codd, the inventor of the relational model, introduced the concept of normalization and what we now know as the First Normal Form (1NF) in 1970.

- Codd went on to define the Second Normal Form (2NF) and Third Normal Form (3NF) in 1971
- Codd and Raymond F. Boyce defined the Boyce-Codd Normal Form (BCNF) in 1974.

- Informally, a relational database table is often described as "normalized" if it is in the Third Normal Form.

Database Tables and Normalization

- Normalization is the process where a database is designed in a way that removes redundancies, and increases the clarity in organizing data in a database

- Normalization usually involves dividing large tables into smaller (and less redundant) tables and defining relationships between them.

- The objective is to isolate data so that additions, deletions, and modifications of a field can be made in just one table

Normalization

There is a sequence to normal forms:
1NF is considered the weakest,
2NF is stronger than 1NF,
3NF is stronger than 2NF, and
BCNF is considered the strongest

Also,
any relation that is in BCNF, is in 3NF;
any relation in 3NF is in 2NF; and
any relation in 2NF is in 1NF.

Normalization

We consider a relation in BCNF to be fully normalized.

The benefit of higher normal forms is that update semantics for the affected data are simplified.

This means that applications required to maintain the database are simpler.

A design that has a lower normal form than another design has more redundancy. Uncontrolled redundancy can lead to data integrity problems.

Normal form	Brief definition
First normal form (1NF)	*An entity is in the first normal form if it contains no repeating groups.*
Second normal form (2NF)	*An entity is in the second normal form if all of its attributes depend on the whole primary key.*
Third normal form (3NF)	*An entity is in the third normal form if it contains no transitive dependency*

Objectives of Normalization Beyond 1NF

The objectives of normalization beyond 1NF (First Normal Form) were stated as follows by Codd:

- To free the collection of relations from undesirable insertion, update and deletion anolomies
- To reduce the need for restructuring of collections as new data types are introduced
- To make the relational model more informative to users

Insertion Anomaly

An **Insert Anomaly** occurs when certain attributes cannot be inserted into the database without the presence of other attributes. For example we can't add a new course unless we have at least one student enrolled on the course.

StudentNum	**CourseNum**	**Student Name**	**Address**	**Course**
S21	9201	Jones	Edinburgh	Accounts
S21	9267	Jones	Edinburgh	Business
S24	9267	Smith	Glasgow	Physics
S30	9201	Richards	Manchester	Computing
S30	9322	Richards	Manchester	Maths

Composite Keys

Update Anomaly

- An **Update Anomaly** exists when one or more instances of duplicated data is updated, but not all.
- For example, consider Jones moving address - you need to update all instances of Jones's address.

StudentNum	**CourseNum**	**Student Name**	**Address**	**Course**
S21	9201	Jones	Edinburgh	Accounts
S21	9267	Jones	Edinburgh	Business
S24	9267	Smith	Glasgow	physics
S30	9201	Richards	Manchester	Computing
S30	9322	Richards	Manchester	Maths

Deletion Anomaly

- A **Delete Anomaly** exists when certain attributes are lost because of the deletion of other attributes.
- For example, consider what happens if Student S30 is the last student to leave the course - All information about the course is lost.

StudentNum	**CourseNum**	**Student Name**	**Address**	**Course**
S21	9201	Jones	Edinburgh	Accounts
S21	9267	Jones	Edinburgh	Business
S24	9267	Smith	Glasgow	physics
S30	9201	Richards	Manchester	Computing
S30	9322	Richards	Manchester	Maths

Second Normal Form: 2NF

Definition:

An entity is in the second normal form if all of its attributes depend on the whole primary key (Composite Key)

- **In relational terms, every column in a table must be *functionally dependent* on the whole primary key of that table.**
- **Functional dependency indicates that a link exists between the values in two different columns.**

Converting From 1NF to 2NF

- **To convert first-normal-form tables to second-normal-form tables, remove columns that are not dependent on the Composite key**

Third Normal Form: 3NF

Definition:

An entity is in the third normal form if it contains no transitive dependencies.

- **A transitive dependency is one which is dependant on a non-key attribute.**

Converting From 2NF to 3NF

- **To convert to third normal form, remove attributes that depend on non primary key attributes.**

Normalization Example

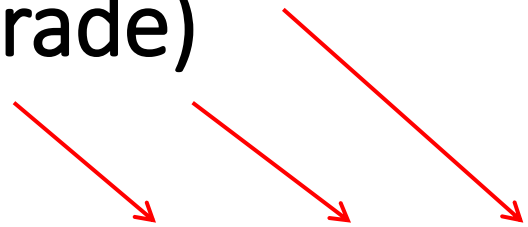
Take the following table.

StudentID is the primary key.

Student ID	StudentName	Address	Mentor Name	Mentor Office	Subject	Subject Cost	Grade
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309	English	€50	B
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309	Maths	€50	A
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309	Info Tech	€100	B+

Is it 1NF?

No. There are repeating groups (subject, subjectcost, grade)



Student ID	**StudentName**	**Address**	**Mentor Name**	**Mentor Office**	**Subject**	**Subject Cost**	**Grade**
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309	English	€50	B
					Maths	€50	A
					Info Tech	€100	B+

How can you make it 1NF?

To Convert to 1NF Remove Repeating Groups

Student ID	StudentName	Address	Mentor Name	Mentor Office	Subject	Subject Cost	Grade
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309	English	€50	B
					Maths	€50	A
					Info Tech	€100	B+

Becomes

Repeating Groups have been removed

Student ID	Student Name	Address	Mentor Name	Mentor Office
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Student Table

Student ID	Subject	Subject Cost	Grade
A00002516	English	€50	B
A00002516	Maths	€50	A
A00002516	Info Tech	€100	B+

Subject Table

To Convert to 1NF Remove Repeating Groups

Student ID	StudentName	Address	MentorName	MentorOffice	Subject	SubjectCost	Grade
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309	English	€50	B
					Maths	€50	A
					Info Tech	€100	B+

Becomes

Student ID	Student Name	Address	Mentor Name	Mentor Office
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Student ID	Subject	Subject Cost	Grade
A00002516	English	€50	B
A00002516	Maths	€50	A
A00002516	Info Tech	€100	B+

The primary key of the Subject table is Student ID and Subject

To Convert to 1NF Remove Repeating Groups

Student ID	StudentName	Address	MentorName	MentorOffice	Subject	SubjectCost	Grade
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309	English	€50	B
					Maths	€50	A
					Info Tech	€100	B+

Becomes

Student ID	Student Name	Address	Mentor Name	Mentor Office
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Student ID	Subject	Subject Cost	Grade
A00002516	English	€50	B
A00002516	Maths	€50	A
A00002516	Info Tech	€100	B+

Is it 2NF?

To Convert to 2NF Remove Attributes Which Are Not Dependent on Composite Key

Student ID	Student Name	Address	Mentor Name	Mentor Office
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Student ID	Subject	Subject Cost	Grade
A00002516	English	€50	B
A00002516	Maths	€50	A
A00002516	Info Tech	€100	B+

Which Attribute is not Dependent on the whole Primary key/ Composite key?

To Convert to 2NF Remove Attributes Which Are Not Dependent on Composite Key

Student ID	Student Name	Address	Mentor Name	Mentor Office
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Student ID	Subject	Subject Cost	Grade
A00002516	English	€50	B
A00002516	Maths	€50	A
A00002516	Info Tech	€100	B+

Subject Cost depends on Subject but not on Student ID so it needs to be removed

To Convert to 2NF Remove Attributes Which Are Not Dependent on Composite Key

Student ID	Student Name	Address	Mentor Name	Mentor Office
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Student ID	Subject	Subject Cost	Grade
A00002516	English	€50	B
A00002516	Maths	€50	A
A00002516	Info Tech	€100	B+

Becomes

Student ID	Student Name	Address	Mentor Name	Mentor Office
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Student Table

Student ID	Subject	Grade
A00002516	English	B
A00002516	Maths	A
A00002516	Info Tech	B+

Grades Table

Subject	Subject Cost
English	€50
Maths	€50
Info Tech	€100

Subject table

To Convert to 2NF Remove Attributes Which Are Not Dependent on Composite Key

Student ID	Student Name	Address	Mentor Name	Mentor Office
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Student ID	Subject	Subject Cost	Grade
A00002516	English	€50	B
A00002516	Maths	€50	A
A00002516	Info Tech	€100	B+

Subject Cost has been moved to a subject table.

Becomes

Student ID	Student Name	Address	Mentor Name	Mentor Office
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Subject	Subject Cost
English	€50
Maths	€50
Info Tech	€100

Student ID	Subject	Grade
A00002516	English	B
A00002516	Maths	A
A00002516	Info Tech	B+

To Convert to 2NF Remove Attributes Which Are Not Dependent on Whole Primary Key

Student ID	Student Name	Address	Mentor Name	Mentor Office
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Student ID	Subject	Subject Cost	Grade
A00002516	English	€50	B
A00002516	Maths	€50	A
A00002516	Info Tech	€100	B+

Becomes

Student ID	Student Name	Address	Mentor Name	Mentor Office
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Subject	Subject Cost
English	€50
Maths	€50
Info Tech	€100

Student ID	Subject	Grade
A00002516	English	B
A00002516	Maths	A
A00002516	Info Tech	B+

Is it 3NF?

To Convert to 3NF Remove Transitive Dependencies

Student ID	Student Name	Address	Mentor Name	Mentor Office
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Subject	Subject Cost
English	€50
Maths	€50
Info Tech	€100

Student ID	Subject	Grade
A00002516	English	B
A00002516	Maths	A
A00002516	Info Tech	B+

Which Attribute is not Dependent on the Primary key?

To Convert to 3NF Remove Transitive Dependencies

Student ID	Student Name	Address	Mentor Name	Mentor Office
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Subject	Subject Cost
English	€50
Maths	€50
Info Tech	€100

Student ID	Subject	Grade
A00002516	English	B
A00002516	Maths	A
A00002516	Info Tech	B+

Mentor Office is dependant on Mentor Name (not StudentID)

To Convert to 3NF Remove Transitive Dependencies

Student ID	**Student Name**	**Address**	**Mentor Name**	**Mentor Office**
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan	X309

Student ID	**Subject**	**Grade**
A00002516	English	B
A00002516	Maths	A
A00002516	Info Tech	B+

Subject	**Subject Cost**
English	€50
Maths	€50
Info Tech	€100

Becomes

Student ID	**Student Name**	**Address**	**Mentor Name**
A00002516	Tom Byrnes	10 Glen Croi	Jane Ryan

Mentor Name	**Mentor Office**
Jane Ryan	X309

Student ID	**Subject**	**Grade**
A00002516	English	B
A00002516	Maths	A
A00002516	Info Tech	B+

Subject	**Subject Cost**
English	€50
Maths	€50
Info Tech	€100

Another Normalization Example

Table: Project

Project Number	Project Name	Employee No	Employee Name	Job Description	Hourly Rate	Hours Charged
15	Evergreen	103	June	Elec Eng	80	40
		101	John	DB Design	95	40
		105	Alice	DB Design	95	25
18	AmberWave	114	Darlene	Analyst	95	40
		118	James	Support	20	40
		104	Anne	Analyst	95	30
22	Rolling Tide	105	Alice	DB Design	95	15
		104	Anne	Analyst	95	10
25	Star Flight	107	Maria	Programmer	35	40

Is it 1NF?

No. There are repeating groups

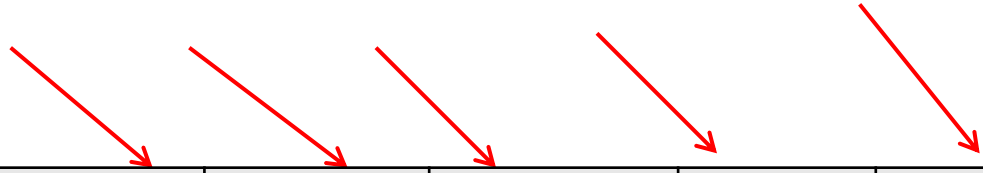


Table: Project

Project Number	Project Name	Employee No	Employee Name	Job Description	Hourly Rate	Hours Charged
15	Evergreen	103	June	Elec Eng	80	40
		101	John	DB Design	95	40
		105	Alice	DB Design	95	25
18	AmberWave	114	Darlene	Analyst	95	40
		118	James	Support	20	40
		104	Anne	Analyst	95	30
22	Rolling Tide	105	Alice	DB Design	95	15
		104	Anne	Analyst	95	10
25	Star Flight	107	Maria	Programmer	35	40

How can you make it 1NF?

To Convert to 1NF Remove Repeating Groups

Table: Project

Project Number	Project Name	Employee No	Employee Name	Job Description	Hourly Rate	Hours Charged
15	Evergreen	103	June	Elec Eng	80	40
		101	John	DB Design	95	40
		105	Alice	DB Design	95	25
18	AmberWave	114	Darlene	Analyst	95	40
		118	James	Support	20	40
		104	Anne	Analyst	95	30
22	Rolling Tide	105	Alice	DB Design	95	15
		104	Anne	Analyst	95	10
25	Star Flight	107	Maria	Programmer	35	40

Becomes

Repeating Groups have been removed

Table: Project

Project Number	Project Name
15	Evergreen
18	AmberWave
22	Rolling Tide
25	Star Flight

Table: Assignment

Project Number	Employee No	Employee Name	Job Description	Hourly Rate	Hours Charged
15	103	June	Elec Eng	80	40
15	101	John	DB Design	95	40
15	105	Alice	DB Design	95	25
18	114	Darlene	Analyst	95	40
18	118	James	Support	20	40
18	104	Anne	Analyst	95	30
22	105	Alice	DB Design	95	15
22	104	Anne	Analyst	95	10
25	107	Maria	Programmer	35	40

To Convert to 1NF Remove Repeating Groups

Table: Project

Project Number	Project Name
15	Evergreen
18	AmberWave
22	Rolling Tide
25	Star Flight

Table: Assignment

Project Number	Employee No	Employee Name	Job Description	Hourly Rate	Hours Charged
15	103	June	Elec Eng	80	40
15	101	John	DB Design	95	40
15	105	Alice	DB Design	95	25
18	114	Darlene	Analyst	95	40
18	118	James	Support	20	40
18	104	Anne	Analyst	95	30
22	105	Alice	DB Design	95	15
22	104	Anne	Analyst	95	10
25	107	Maria	Programmer	35	40

Is it 2NF?

To Convert to 2NF Remove Attributes Which Are Not Dependent on Whole Primary Key

Table: Project

Project Number	Project Name
15	Evergreen
18	AmberWave
22	Rolling Tide
25	Star Flight

Table: Assignment

Project Number	Employee No	Employee Name	Job Description	Hourly Rate	Hours Charged
15	103	June	Elec Eng	80	40
15	101	John	DB Design	95	40
15	105	Alice	DB Design	95	25
18	114	Darlene	Analyst	95	40
18	118	James	Support	20	40
18	104	Anne	Analyst	95	30
22	105	Alice	DB Design	95	15
22	104	Anne	Analyst	95	10
25	107	Maria	Programmer	35	40

Which Attributes are not Dependent on the whole Primary key / Composite Key?

To Convert to 2NF Remove Attributes Which Are Not Dependent on Whole Primary Key

Table: Project

Project Number	Project Name
15	Evergreen
18	AmberWave
22	Rolling Tide
25	Star Flight

Table: Assignment

Project Number	Employee No	Employee Name	Job Description	Hourly Rate	Hours Charged
15	103	June	Elec Eng	80	40
15	101	John	DB Design	95	40
15	105	Alice	DB Design	95	25
18	114	Darlene	Analyst	95	40
18	118	James	Support	20	40
18	104	Anne	Analyst	95	30
22	105	Alice	DB Design	95	15
22	104	Anne	Analyst	95	10
25	107	Maria	Programmer	35	40

Employee Name, Job Description, Hourly Rate are not dependant on Project Number

To Convert to 2NF Remove Attributes Which Are Not Dependent on Whole Primary Key

Table: Project

Project Number	Project Name
15	Evergreen
18	AmberWave
22	Rolling Tide
25	Star Flight

Table: Assignment

Project Number	Employee No	Hours Charged
15	103	40
15	101	40
15	105	25
18	114	40
18	118	40
18	104	30
22	105	15
22	104	10
25	107	40

Table: Employee

Employee No	Employee Name	Job Description	Hourly Rate
103	June	Elec Eng	80
101	John	DB Design	95
105	Alice	DB Design	95
114	Darlene	Analyst	95
118	James	Support	20
104	Anne	Analyst	95
107	Maria	Programmer	35

Is it 3NF?

To Convert to 3NF Remove Transitive Dependencies

Table: Project

Project Number	Project Name
15	Evergreen
18	AmberWave
22	Rolling Tide
25	Star Flight

Table: Assignment

Project Number	Employee No	Hours Charged
15	103	40
15	101	40
15	105	25
18	114	40
18	118	40
18	104	30
22	105	15
22	104	10
25	107	40

Table: Employee

Employee No	Employee Name	Job Description	Hourly Rate
103	June	Elec Eng	80
101	John	DB Design	95
105	Alice	DB Design	95
114	Darlene	Analyst	95
118	James	Support	20
104	Anne	Analyst	95
107	Maria	Programmer	35

Which Attribute is not Dependent on the Primary key?

To Convert to 3NF Remove Transitive Dependencies

Table: Project

Project Number	Project Name
15	Evergreen
18	AmberWave
22	Rolling Tide
25	Star Flight

Table: Employee

Employee No	Employee Name	Job Description	Hourly Rate
103	June	Elec Eng	80
101	John	DB Design	95
105	Alice	DB Design	95
114	Darlene	Analyst	95
118	James	Support	20
104	Anne	Analyst	95
107	Maria	Programmer	35

Table: Assignment

Project Number	Employee No	Hours Charged
15	103	40
15	101	40
15	105	25
18	114	40
18	118	40
18	104	30
22	105	15
22	104	10
25	107	40

Hourly Rate is dependant on Job Description not on EmployeeNo

To Convert to 3NF Remove Transitive Dependencies

Table: Project

Project Number	Project Name
15	Evergreen
18	AmberWave
22	Rolling Tide
25	Star Flight

Table: Employee

Employee No	Employee Name	Job Description
103	June	Elec Eng
101	John	DB Design
105	Alice	DB Design
114	Darlene	Analyst
118	James	Support
104	Anne	Analyst
107	Maria	Programmer

Table: Assignment

Project Number	Employee No	Hours Charged
15	103	40
15	101	40
15	105	25
18	114	40
18	118	40
18	104	30
22	105	15
22	104	10
25	107	40

Table: Job Grade

Job Description	Hourly Rate
Elec Eng	80
DB Design	95
DB Design	95
Analyst	95
Support	20
Analyst	95
Programmer	35

Links to Watch

- https://www.youtube.com/watch?v=xoTyrdT9SZI - normalization
- https://www.youtube.com/watch?v=mUtAPbb1ECM - 1NF
- https://www.youtube.com/watch?v=R7UblSu4744&list=PLLGlmW7jT-nTr1ory9o2MgsOmmx2w8FB3&index=3 -2NF
- https://www.youtube.com/watch?v=aAx_JoEDXQA&list=PLLGlmW7jT-nTr1ory9o2MgsOmmx2w8FB3&index=4 – 3NF
- https://www.youtube.com/watch?v=NNjUhvvwOrk&list=PLLGlmW7jT-nTr1ory9o2MgsOmmx2w8FB3&index=5 -BCNF