



4.10.3 Database Normalisation

Normalisation

A relational database can be **normalised** to **reduce duplication of data** in the tables and **improve the consistency** of the data.

We will look at 3 normalisation forms: 1st Normal Form, 2nd Normal Form and 3rd Normal Form.

1st Normal Form

- Has a primary key.
- All data is atomic.

2nd Normal Form

- Is in 1st Normal Form.
- Has no partial-key dependencies.

3rd Normal Form

- Is in 2nd Normal Form.
- Has no non-key dependencies.





1st Normal Form

A **flat file** database is a database where all the data is stored in one table.

For a table to be in 1st normal form:

- It must have a primary key
- All the data must be atomic.

A database that is not normalised

CustomerID	Forename	Surname	Address	DateOfDownload	MovieID	MovieTitle	Genre	Format	FileType
1	John	Hall	1 High Street	19/04/2025	1	The Hangover	Comedy	LowRes	MPEG-2
				23/04/2025	2	The Thing	Horror	LowRes	MPEG-2
2	Elle	Thompson	17 Acacia Avenue	19/04/2025	3	Alien	Horror	HiRes	MPEG-4
				23/04/2025	4	Robocop	Sci-fi	HiRes	MPEG-4
				25/04/2025	2	The Thing	Horror	LowRes	MPEG-2
3	Stephen	Fellowes	26 Maple Drive	19/04/2025	5	Minions	Children	HiRes	MPEG-4
				23/04/2025	4	Robocop	Sci-fi	HiRes	MPEG-4

The table is not in 1st normal form because some cells contain multiple values. These must be separated into individual records so that all data is atomic.





A database in NF1

CustomerID	Forename	Surname	Address	DateOfDownload	MovieID	MovieTitle	Genre	Format	FileType
1	John	Hall	1 High Street	19/04/2025	1	The Hangover	Comedy	LowRes	MPEG-2
1	John	Hall	1 High Street	23/04/2025	2	The Thing	Horror	LowRes	MPEG-2
2	Elle	Thompson	17 Acacia Avenue	19/04/2025	3	Alien	Horror	HiRes	MPEG-4
2	Elle	Thompson	17 Acacia Avenue	23/04/2025	4	Robocop	Sci-fi	HiRes	MPEG-4
2	Elle	Thompson	17 Acacia Avenue	25/04/2025	2	The Thing	Horror	LowRes	MPEG-2
3	Stephen	Fellowes	26 Maple Drive	19/04/2025	5	Minions	Children	HiRes	MPEG-4
3	Stephen	Fellowes	26 Maple Drive	23/04/2025	4	Robocop	Sci-fi	HiRes	MPEG-4

CustomerID is no longer unique, so CustomerID and MovieID are used together as a composite key. The combination is unique for each record.

You may notice that we have more duplicate data but that will be eliminated in the next steps.





2nd Normal Form

For a table to be in 2nd normal form:

- It must be in NF1
- It must have no **partial-key** dependencies.

A partial key dependency is where a field depends on only part of the primary key but not the whole primary key. The table has partial key dependencies. One of which is the Forename field which depends on the CustomerID part of the primary key but does not depend on the MovieID part of the primary key.

In this case we need to separate the data in three tables. One table for the fields that depend on the CustomerID, another table for the fields that depend on the MovieID and another table for the fields that depend on both CustomerID and MovieID.

The Database in NF2

CustomerID	Forename	Surname	Address
1	John	Hall	1 High Street
2	Elle	Thompson	17 Acacia Avenue
3	Stephen	Fellowes	26 Maple Drive

MovieID	MovieTitle	Genre	Format	FileType
1	The Hangover	Comedy	LowRes	MPEG-2
2	The Thing	Horror	LowRes	MPEG-2
3	Alien	Horror	HiRes	MPEG-4
4	Robocop	Sci-fi	HiRes	MPEG-4
5	Minions	Children	HiRes	MPEG-4

CustomerID	MovieID	DateOfDownload
1	1	19/04/2025
1	2	23/04/2025
2	3	19/04/2025
2	4	23/04/2025
2	2	25/04/2025
3	5	19/04/2025
3	4	23/04/2025





All the fields in the Customer table depend on the primary key CustomerID.

All the fields in the Movie table depend on the primary key MovieID.

All the fields in the Download table depend on both parts of the composite key; CustomerID and MovieID.

3rd Normal Form

For a table to be in 3rd normal form:

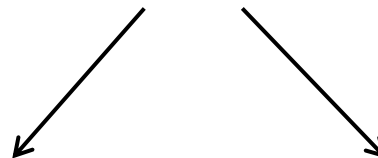
- It must be in NF2
- It must have no **non-key** dependencies.

In the Movie table the FileType field depends on the Format field. A LowRes format always has a filetype MPEG-2 and a HiRes format always has a filetype MPEG-4. This is a non-key dependency.

An easy way to spot non-key dependencies is to look for columns where the values always match up.

Format and FileType are used as fields in a new table with Format as the primary key. The Format field is left in the Movie table to act as a foreign key to link the tables.

MovieID	MovieTitle	Genre	Format	FileType
1	The Hangover	Comedy	LowRes	MPEG-2
2	The Thing	Horror	LowRes	MPEG-2
3	Alien	Horror	HiRes	MPEG-4
4	Robocop	Sci-fi	HiRes	MPEG-4
5	Minions	Children	HiRes	MPEG-4



MovieID	MovieTitle	Genre	Format
1	The Hangover	Comedy	LowRes
2	The Thing	Horror	LowRes
3	Alien	Horror	HiRes
4	Robocop	Sci-fi	HiRes
5	Minions	Children	HiRes

Format	FileType
LowRes	MPEG-2
HiRes	MPEG-4



*The Normalised Database in NF3*

CustomerID	Forename	Surname	Address
1	John	Hall	1 High Street
2	Elle	Thompson	17 Acacia Avenue
3	Stephen	Fellowes	26 Maple Drive

MovieID	MovieTitle	Genre	Format
1	The Hangover	Comedy	LowRes
2	The Thing	Horror	LowRes
3	Alien	Horror	HiRes
4	Robocop	Sci-fi	HiRes
5	Minions	Children	HiRes

Format	FileType
LowRes	MPEG-2
HiRes	MPEG-4

CustomerID	MovieID	DateOfDownload
1	1	19/04/2025
1	2	23/04/2025
2	3	19/04/2025
2	4	23/04/2025
2	2	25/04/2025
3	5	19/04/2025
3	4	23/04/2025

