

# Normalización de Base de Datos

Alfonso Tobar Arancibia

Data Scientist

03-11-2020



[github.com/Rcubes/clases-ml/Slides](https://github.com/Rcubes/clases-ml/Slides)

# Normalización

El objetivo principal de la Normalización de la Base de Datos es disminuir la Redundancia de manera de ahorrar espacio de almacenamiento como de evitar problemas en el mantenimiento de la base de datos.

id_estudiante	Nombre	Generación	Modalidad	Fecha Inscripción Academia
1	Francisco	17	Online	10-07-2020
1	Francisco	17	Online	10-07-2020
1	Francisco	17	Online	10-07-2020
1	Francisco	17	Online	10-07-2020
2	Maite	17	Online	10-07-2020
2	Maite	17	Online	10-07-2020
2	Maite	17	Online	10-07-2020
2	Maite	17	Online	10-07-2020

# Ejemplo

id_estudiante	Nombre	Generación	Módulo
1	Francisco	17	Python, FDS, ML, SQL
2	Maite	17	Python, FDS, ML, SQL

El primer problema que se puede ver a partir de eso es que tenemos varios campos ingresados en una sola columna, lo cual es óptimo.

¿Cómo determino qué Alumno han tomado el módulo de ML?

```
SELECT NOMBRE  
FROM ESTUDIANTES  
WHERE MODULO = 'ML'
```

# Solución

## Primera forma Normalizada

id_estudiante	Nombre	Generación	Módulo
1	Francisco	17	Python
1	Francisco	17	FDS
1	Francisco	17	ML
1	Francisco	17	SQL
2	Maite	17	Python
2	Maite	17	FDS
2	Maite	17	ML
2	Maite	17	SQL

### Requisitos:

- Cada elemento debe tener valores atómicos (indivisibles).
- Cada columna debe tener elementos del mismo tipo de dato.
- Cada columna debe tener nombres únicos.
- El orden que se ingresen los datos no importa.

## Otro problema...

id_estudiante	Modulo	Nota	Profesor
1	1	8	Gianni
1	3	9	Alfonso
2	2	10	Gianni
2	3	7	Alfonso

- Se tiene un caso de Clave Primaria compuesta.
- Se tiene redundancia de datos en la Columna Profesor, ya que la columna Profesor tiene una dependencia parcial.

# Solución

## Segunda forma Normalizada

id_estudiante	Modulo	Nota
1	1	8
1	3	9
2	2	10
2	3	7

Modulos	Nombre	Nombre Profe
1	Python	Gianni
2	FDS	Gianni
3	ML	Alfonso
4	SQL	Alfonso
5	Big Data	?
6	Proyecto	?

### Requisitos:

- Todos los Requisitos de la Primera forma Normalizada.
- Que no haya dependencia parcial.

## Último problema...

id_estudiante	Modulo	Nota	Evaluación	N Hitos
1	1	8	Prueba	0
1	3	9	Informe	4
2	2	10	Informe	4
2	3	7	Informe	4

- También existe un tipo de redundancia que en este caso se manifiesta en el Número de Hitos.
- También hay dependencia transitiva.

# Solución

## Tercera forma Normalizada

id_estudiante	Modulo	Nota
1	1	8
1	3	9
2	2	10
2	3	7

Modulos	Nombre	Nombre Profe	Evaluación	N Hitos
1	Python	Gianni	Prueba	0
2	FDS	Glanni	Informe	4
3	ML	Alfonso	Informe	4
4	SQL	Alfonso	Prueba	0
5	Big Data	?	?	?
6	Proyecto	?	?	?

### Requisitos:

- Todos los Requisitos de la Primera y Segunda forma Normalizada.
- Que no haya dependencia Transitiva.



# Esquema Final

