



Introducción a Mysql

Conociendo la herramienta y los conceptos necesarios para crear nuestras primeras bases de datos relacionales



Juan Manuel Spinelli

Data Engineer en COREBI Analytics & Big Data - Data
Science Enthusiast - Provincia de Córdoba, Argentina
Argentina



COREBI Analytics & Big Data



Acámica



¿Que vamos a aprender en este curso?

Detalles

Durante las próximas 4 clases vamos a aprender:

- Tipos de datos
- Tipos de bases de datos
- Base de datos relacional
- Fases en la creación de una bases de datos relacional
- Uso de Mysql
- Comandos
- Joins
- Procedimientos almacenados
- Triggers



¿Cómo van a ser las clases?

Desarrollo

Las clases se van a llevar adelante de la siguientes forma:

- Parte teórica
- Parte práctica
- Material para practicar entre clases



Clases

Temas



Clase 1

Temas

Teoría:

- Que es un dato
- Tipos de datos
- Que es una base de datos
- Tipos de bases de datos
- Base de datos relacional

Teórico práctico:

- Normalización:
 - Primer fase normal
 - Segunda fase normal
 - Tercera fase normal
 - Clave primaria
 - Clave foránea

Problema a analizar / resolver para la próxima clase

Instalar Mysql / Workbench



Clase 2

Temas

Puesta en común:

- Ejercicio práctico clase anterior

Teórico práctico:

- Diferencias entre DDL, DML y DCL
- Generar un diagrama entidad relación
- Sintaxis
- Crear una base de datos
- Crear las tablas definidas en el problema planteado la semana anterior
- Insertar datos
- Consultar datos
- Actualizar datos
- Eliminar datos

Ejercicios prácticos a resolver para la próxima semana



Clase 3

Temas

Puesta en común:

- Ejercicio práctico clase anterior

Teórico práctico:

- Resumen de operadores
- In, or, and, between, lower, upper, etc.
- Distinct, count, max, min,
- Order by, group by, having
- Joinear tablas (Tipos de Joins)

Ejercicios prácticos a resolver para la próxima semana



Clase 4

Temas

Puesta en común:

- Ejercicio práctico clase anterior

Teórico práctico:

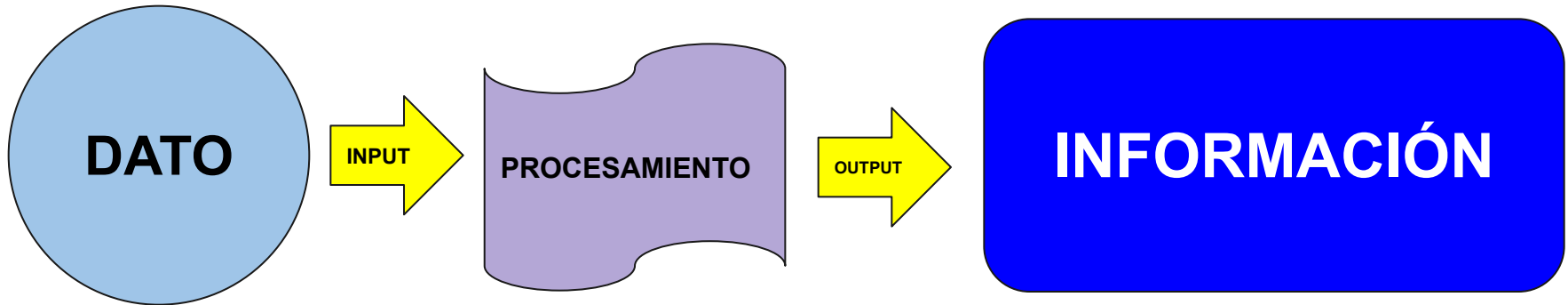
- Introducción a subconsultas
- Introducción a Vistas, procedimientos almacenados y triggers
- Exportar nuestra base de datos



Empezamos

¿Que es un dato?

Datos es un término que se refiere a hechos, eventos, transacciones, etc., que han sido registrados.





¿Que es un dato?

El dato por sí solo no tiene un objetivo, pero ese dato junto a otros y luego de un procesamiento se pueden transformar en información y en ese punto nos permiten responder preguntas o hacer análisis.

Los datos, por ejemplo, pueden ser números, palabras o variables, en cambio, la información son estos datos procesados para que haga sentido en un contexto específico. Se puede decir entonces que los datos se usan para obtener información.



Analicen la imagen:

- Tienen distinto color
- Cada color representa un monto
- Tienen distinto tamaño
- Son muchas
- Están usadas
- Hay repetidas
- Antiguas
- Valor económico
- Valor emocional





DATO \neq
INFORMACIÓN



Algunas características de los datos

Cuando pensamos en un dato tenemos que tener en cuenta algunas cosas como:

- Deben permanecer en el tiempo (Edad **no**, fecha de nacimiento **si**)
- Debe tener un significado
- Debe ser manipulable mediante operadores (Suma, resta, comparaciones, etc)



Tipos de datos

Datos cualitativos: Dan el enfoque o el punto de vista de la muestra. Color, textura, etc.

Datos cuantitativos: Miden o calculan algo para llegar a un punto en su investigación. Tamaño, total, precio, etc.

Datos categóricos: Colocan el objeto por describir en una categoría. Nuevos, usados, etc.

Datos discretos: Son datos numéricos enteros. Tengo 30 monedas, no 29.5.

Datos continuos: Pueden recibir cualquier valor. La medida de una moneda es 7,82 mm



¿Que es una base de datos?

Es una serie de datos organizados y relacionados entre sí, los cuales son recolectados y explotados por los sistemas de información de una empresa o negocio en particular

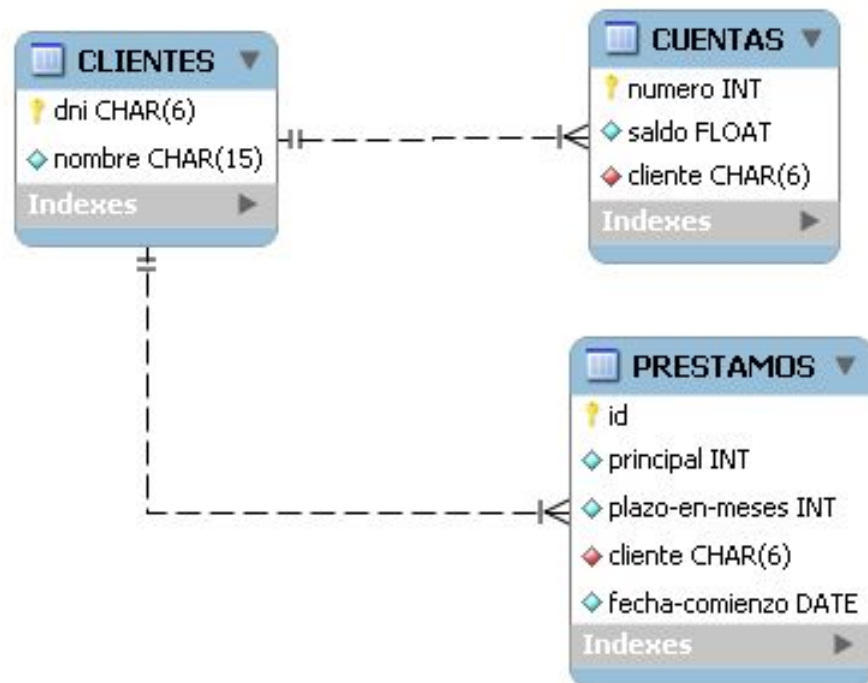
Estas bases nos permiten:


- Control sobre la redundancia de datos
- Consistencia de datos
- Integridad de los datos
- Mejora en la accesibilidad a los datos



Tipos de bases de datos

Bases de dato relacional

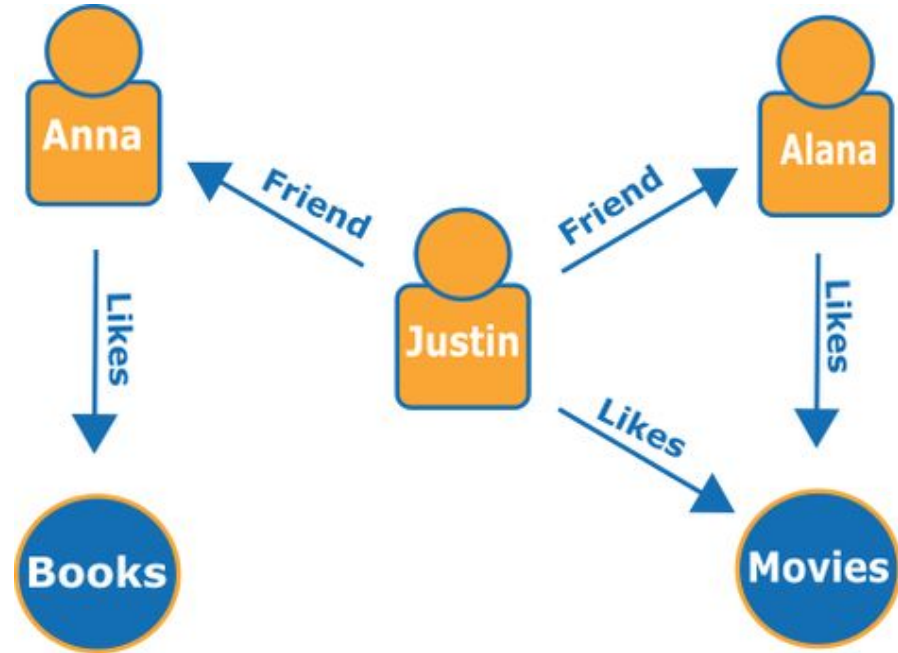




Bases de datos no relacional

```
{ "empinfo" :  
  {  
    "employees" : [  
      {  
        "name" : "Scott Philip",  
        "salary" : f44k,  
        "age" : 27,  
      },  
      {  
        "name" : "Tim Henn",  
        "salary" : f40k,  
        "age" : 27,  
      },  
      {  
        "name" : "Long Yong",  
        "salary" : f40k,  
        "age" : 28,  
      }  
    ]  
  }  
}
```

Bases de datos de grafos





Otros tipos

*Click para leer un
poco más*



Base de datos relacional

Las bases de datos relacionales se basan en la organización de la información en trozos pequeños, que se relacionan entre ellos mediante la relación de identificadores.

En el ámbito informático se habla mucho de ACID :

- Atómicas: Si una parte de una transacción falla, todo se anula.
- Coherentes: Los datos tienen que cumplir con las reglas definidas.
- Aisladas: Cada transacción es independiente.
- Duraderas: Si la transacción se ejecutó Ok, los cambios tienen que ser permanentes.


Atributo

A_1	A_2	...	A_n
Valor			

Esquema

Tupla

Relación



Práctica teórica base de datos relacional y normalización

Normalizar, regularizar, poner en orden.

Vamos a solucionar la redundancia evitando así anomalías en la base de datos.



Tipos de anomalías

- Anomalías de inserción
- Anomalías de modificación
- Anomalías de eliminación



Anomalías de inserción

Nombre	Apellido	Edificio	Departamento
Andrea	Lopez	Torre Uno	A-1
Jonathan	Desia	Torre Dos	B-3



Anomalías de inserción

Nombre	Apellido	Edificio	Departamento
Andrea	Lopez	Torre Uno	A-1
Jonathan	Desia	Torre Dos	B-3

Nombre	Apellido	Edificio	Departamento
Andrea	Lopez	Torre Uno	A-1
Jonathan	Desia	Torre Dos	B-3
???	???	Torre Uno	A-2



Anomalías de modificación

Nombre	Apellido	Edificio	Departamento
Andrea	Lopez	Torre Uno	A-1
Jonathan	Desia	Torre Dos	B-3
Lucas	Garcia	Torre Uno	A-2
Amanda	Perez	Torre Dos	B-1



Anomalías de modificación

Nombre	Apellido	Edificio	Departamento
Andrea	Lopez	Torre Uno	A-1
Jonathan	Desia	Torre Dos	B-3
Lucas	Garcia	Torre Uno	A-2
Amanda	Perez	Torre Dos	B-1

Nombre	Apellido	Edificio	Departamento
Andrea	Lopez	Torre 1	A-1
Jonathan	Desia	Torre Dos	B-3
Lucas	Garcia	Torre Uno	A-2
Amanda	Perez	Torre 2	B-1



Anomalías de eliminación

Nombre	Apellido	Edificio	Departamento
Andrea	Lopez	Torre Uno	A-1
Jonathan	Desia	Torre Dos	B-3
Lucas	Garcia	Torre Uno	A-2



Anomalías de eliminación

Nombre	Apellido	Edificio	Departamento
Andrea	Lopez	Torre Uno	A-1
Jonathan	Desia	Torre Dos	B-3
Lucas	Garcia	Torre Uno	A-2

Nombre	Apellido	Edificio	Departamento
Andrea	Lopez	Torre Uno	A-1
Jonathan	Desia	Torre Dos	B-3
Lucas	Garcia	Torre Uno	A-2



Entonces...

Cuando hablamos de normalización, hablamos de un ciclo repetitivo con la finalidad de lograr la integridad de los datos, evitando la redundancia y corrigiendo las anomalías.



Normalización

Primera Fase Normal

Atributo

Tupla

Esquema

Relación

A_1	A_2	...	A_n
Valor			



Primera fase normal

La forma de los grupos repetitivos:

Una relación está en primera forma normal si y solo si no tiene grupos repetitivos

- No debe haber tuplas repetidas
- No debe importar el orden de las tuplas
- Tiene que existir una clave primaria
- Los atributos tienen que ser atómicos




NOMBRE	APELLIDO	MATERIAS
Andrea	Lopez	Bases de datos, Data Science, Ingenieria de datos
Jonathan	Desia	Arquitectura de datos, Arquitectura de software, Estadistica
Lucas	Garcia	Estadistica, Computer vision, Matematicas

- 
- No hay clave primaria
 - El atributo “Materias” no es atómico

NOMBRE	APELLIDO	MATERIAS
Andrea	Lopez	Bases de datos, Data Science, Ingenieria de datos
Jonathan	Desia	Arquitectura de datos, Arquitectura de software, Estadistica
Lucas	Garcia	Estadistica, Computer vision, Matematicas



PK	NOMBRE	APELLIDO	MATERIAS 1	MATERIAS 2	MATERIAS 3
1	Andrea	Lopez	Bases de datos	Data Science	Ingenieria de datos
2	Jonathan	Desia	Arquitectura de datos	Arquitectura de software	Estadistica
3	Lucas	Garcia	Estadistica	Bases de datos	Data Science

- 
- Si un alumno quiere hacer menos materias desperdiciamos espacio
 - Si un alumno quiere hacer más materias no podemos agregarlo

PK	NOMBRE	APELLIDO	MATERIAS 1	MATERIAS 2	MATERIAS 3	
1	Andrea	Lopez	Bases de datos	Data Science	Ingenieria de datos	
2	Jonathan	Desia	Arquitectura de datos	Arquitectura de software	Estadistica	
3	Lucas	Garcia	Estadistica	Bases de datos	Data Science	
4	Carlos	Gomez	Matematicas			
5	Sabrina	Caro	Arquitectura de datos	Arquitectura de software	Estadistica	Bases de datos



ALUMNOS		
ID_ALUMNO	NOMBRE	APELLIDO
1	Andrea	Lopez
2	Jonathan	Desia
3	Lucas	Garcia
4	Carlos	Gomez
5	Sabrina	Caro

MATERIAS	
ID_MATERIA	DETALLE
1	Bases de datos
2	Arquitectura de datos
3	Estadistica
4	Matematicas
5	Ingenieria de datos
6	Data Science
7	Arquitectura de software

ALUMNOS		
ID_ALUMNO	NOMBRE	APELLIDO
1	Andrea	Lopez
2	Jonathan	Desia
3	Lucas	Garcia
4	Carlos	Gomez
5	Sabrina	Caro

ALUMNO_X_MATERIA	
ID_ALUMNO	ID_MATERIA
1	1
1	6
1	5
2	2
2	7
2	3
3	3
3	1
3	6
4	4
5	2
5	7
5	3
5	1

MATERIAS	
ID_MATERIA	DETALLE
1	Bases de datos
2	Arquitectura de datos
3	Estadística
4	Matemáticas
5	Ingeniería de datos
6	Data Science
7	Arquitectura de software

MATERIAS	
ID_MATERIA	DETALLE
1	Bases de datos
2	Arquitectura de datos
3	Estadística
4	Matemáticas
5	Ingeniería de datos
6	Data Science
7	Arquitectura de software
8	ETL
9	NLP

ALUMNOS		
ID_ALUMNO	NOMBRE	APELLIDO
1	Andrea	Lopez
2	Jonathan	Desia
3	Lucas	Garcia
4	Carlos	Gomez
5	Sabrina	Caro
6	Andres	Robles

ALUMNOS		
ID_ALUMNO	NOMBRE	APELLIDO
1	Andrea	Lopez
2	Jonathan	Desia
3	Lucas	Garcia
4	Carlos	Gomez
5	Sabrina	Caro
6	Andres	Robles

ALUMNO_X_MATERIA	
ID_ALUMNO	ID_MATERIA
1	1
1	6
1	5
2	2
2	7
2	3
3	3
3	1
3	6
4	4
5	2
5	7
5	3
5	1
6	8
6	9

MATERIAS	
ID_MATERIA	DETALLE
1	Bases de datos
2	Arquitectura de datos
3	Estadística
4	Matematicas
5	Ingenieria de datos
6	Data Science
7	Arquitectura de software
8	ETL
9	NLP



Normalización

Segunda Fase Normal



Segunda fase normal

Una relación está en segunda forma normal si y sólo si:

- Está en primera forma normal
- Todos Los atributos dependen de la clave primaria completa
- Ningún atributo depende únicamente de una parte de la llave primaria
- Si la llave primaria no es una llave compuesta ya la relación está en segunda forma normal



Segunda fase normal

Procedimiento:

- Identificar las dependencias parciales
- Mover a una nueva entidad cada una de las dependencias parciales
- En todas las entidades sólo deben quedar dependencias completas



COD_PROD	PRODUCTO	COD_DEPOSITO	CIUDAD	CANT	COSTO	TOTAL
1	Notebook Lenovo Ideapad S145 Ryzen 3 3200U 4Gb 1TB 15.6 W10	1	Córdoba	20	95000	1900000
2	Notebook Lenovo V15 i5 1035G1 4Gb 1TB 15.6 Free	1	Córdoba	32	98000	3136000
3	Notebook Lenovo V330 i5 8250U 4G 1TB S/DVD 15.6	2	Buenos Aires	45	98000	4410000
4	Notebook Hp 250 G7 i5 8265u 1T 4G 15.6 Free	2	Buenos Aires	60	99000	5940000

- 
- Hay dependencias parciales, en la misma tabla se pueden ver datos de productos y depósitos

COD_PROD	PRODUCTO	COD_DEPOSITO	CIUDAD	CANT	COSTO	TOTAL
1	Notebook Lenovo Ideapad S145 Ryzen 3 3200U 4Gb 1TB 15.6 W10	1	Córdoba	20	95000	1900000
2	Notebook Lenovo V15 i5 1035G1 4Gb 1TB 15.6 Free	1	Córdoba	32	98000	3136000
3	Notebook Lenovo V330 i5 8250U 4G 1TB S/DVD 15.6	2	Buenos Aires	45	98000	4410000
4	Notebook Hp 250 G7 i5 8265u 1T 4G 15.6 Free	2	Buenos Aires	60	99000	5940000

- 
- De esta forma puedo separar al producto

COD_PROD	PRODUCTO	COD_DEPOSITO	CIUDAD	CANT	COSTO	TOTAL
1	Notebook Lenovo Ideapad S145 Ryzen 3 3200U 4Gb 1TB 15.6 W10	1	Córdoba	20	95000	1900000
2	Notebook Lenovo V15 i5 1035G1 4Gb 1TB 15.6 Free	1	Córdoba	32	98000	3136000
3	Notebook Lenovo V330 i5 8250U 4G 1TB S/DVD 15.6	2	Buenos Aires	45	98000	4410000
4	Notebook Hp 250 G7 i5 8265u 1T 4G 15.6 Free	2	Buenos Aires	60	99000	5940000

- 
- De esta forma puedo separar al depósito

COD_PROD	PRODUCTO	COD_DEPOSITO	CIUDAD	CANT	COSTO	TOTAL
1	Notebook Lenovo Ideapad S145 Ryzen 3 3200U 4Gb 1TB 15.6 W10	1	Córdoba	20	95000	1900000
2	Notebook Lenovo V15 i5 1035G1 4Gb 1TB 15.6 Free	1	Córdoba	32	98000	3136000
3	Notebook Lenovo V330 i5 8250U 4G 1TB S/DVD 15.6	2	Buenos Aires	45	98000	4410000
4	Notebook Hp 250 G7 i5 8265u 1T 4G 15.6 Free	2	Buenos Aires	60	99000	5940000

COD_PROD	PRODUCTO	COSTO
1	Notebook Lenovo Ideapad S145 Ryzen 3 3200U 4Gb 1TB 15.6 W10	95000
2	Notebook Lenovo V15 i5 1035G1 4Gb 1TB 15.6 Free	98000
3	Notebook Lenovo V330 i5 8250U 4G 1TB S/DVD 15.6	98000
4	Notebook Hp 250 G7 i5 8265u 1T 4G 15.6 Free	99000

COD_PROD	COD_DEPOSITO	CANT	TOTAL
1	1	20	1900000
2	1	32	3136000
3	2	45	4410000
4	2	60	5940000

COD_DEPOSITO	CIUDAD
1	Córdoba
1	Córdoba
2	Buenos Aires
2	Buenos Aires

COD_PROD	PRODUCTO	COSTO
1	Notebook Lenovo Ideapad S145 Ryzen 3 3200U 4Gb 1TB 15.6 W10	95000
2	Notebook Lenovo V15 i5 1035G1 4Gb 1TB 15.6 Free	98000
3	Notebook Lenovo V330 i5 8250U 4G 1TB S/DVD 15.6	98000
4	Notebook Hp 250 G7 i5 8265u 1T 4G 15.6 Free	99000

COD_PROD	COD_DEPOSITO	CANT	TOTAL
1	1	20	1900000
2	1	32	3136000
3	2	45	4410000
4	2	60	5940000

COD_DEPOSITO	CIUDAD
1	Córdoba
1	Córdoba
2	Buenos Aires
2	Buenos Aires

COD_PROD	PRODUCTO	COSTO
1	Notebook Lenovo Ideapad S145 Ryzen 3 3200U 4Gb 1TB 15.6 W10	95000
2	Notebook Lenovo V15 i5 1035G1 4Gb 1TB 15.6 Free	98000
3	Notebook Lenovo V330 i5 8250U 4G 1TB S/DVD 15.6	98000
4	Notebook Hp 250 G7 i5 8265u 1T 4G 15.6 Free	99000

COD_PROD	COD_DEPOSITO	CANT	TOTAL
1	1	20	1900000
2	1	32	3136000
3	2	45	4410000
4	2	60	5940000

COD_DEPOSITO	CIUDAD
1	Córdoba
2	Buenos Aires



Normalización

Tercera Fase Normal



Tercera fase normal


Una relación está en segunda forma normal si y sólo si:

- Está en segunda forma normal
- Ningún atributo "no determinante" depende transitivamente de la clave primaria
- Ningún atributo "no primo" depende funcionalmente de otro atributo "no primo"

Explicación complicada pero vamos al ejemplo así lo entendemos mejor.



Cod_Jugador	Nombre	Apellido	Cod_Manager	Nombre_Manager	Apellido_Manager
1	Andrea	Juarez	102	Juan	Perez
2	Claudio	Lopez	103	Jhon	Doe
3	Ezequiel	Gallo	102	Juan	Perez
4	Milagros	Alam	103	Jhon	Doe



Cod_Jugador	Nombre	Apellido	Cod_Manager	Nombre_Manager	Apellido_Manager
1	Andrea	Juarez	102	Juan	Perez
2	Claudio	Lopez	103	Jhon	Doe
3	Ezequiel	Gallo	102	Juan	Perez
4	Milagros	Alam	103	Jhon	Doe

Está en primera forma normal porque:

- No hay tuplas repetidas
- No importa el orden de las tuplas
- Atributos atómicos
- Existe la clave primaria
- No hay grupos repetidos

Está en segunda forma normal porque:

- No existe una clave primaria compuesta



Cod_Jugador	Nombre	Apellido	Cod_Manager	Nombre_Manager	Apellido_Manager
1	Andrea	Juarez	102	Juan	Perez
2	Claudio	Lopez	103	Jhon	Doe
3	Ezequiel	Gallo	102	Juan	Perez
4	Milagros	Alam	103	Jhon	Doe
A	B	C	D	E	F



Cod_Jugador	Nombre	Apellido	Cod_Manager	Nombre_Manager	Apellido_Manager
1	Andrea	Juarez	102	Juan	Perez
2	Claudio	Lopez	103	Jhon	Doe
3	Ezequiel	Gallo	102	Juan	Perez
4	Milagros	Alam	103	Jhon	Doe
A	B	C	D	E	F

$A \rightarrow BCD$

$D \rightarrow EF$




Cod_Jugador	Nombre	Apellido	Cod_Manager	Nombre_Manager	Apellido_Manager
1	Andrea	Juarez	102	Juan	Perez
2	Claudio	Lopez	103	Jhon	Doe
3	Ezequiel	Gallo	102	Juan	Perez
4	Milagros	Alam	103	Jhon	Doe
A	B	C	D	E	F

A → **BCD**


D → **EF**

D es una dependencia funcional transitiva porque no es una clave primaria




Cod_Jugador	Nombre	Apellido	Cod_Manager
1	Andrea	Juarez	102
2	Claudio	Lopez	103
3	Ezequiel	Gallo	102
4	Milagros	Alam	103

A → **BCD**



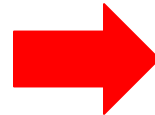
Cod_Manager	Nombre_Manager	Apellido_Manager
102	Juan	Perez
103	Jhon	Doe
102	Juan	Perez
103	Jhon	Doe

D → **EF**




Cod_Jugador	Nombre	Apellido	Cod_Manager
1	Andrea	Juarez	102
2	Claudio	Lopez	103
3	Ezequiel	Gallo	102
4	Milagros	Alam	103

A → **BCD**




Cod_Manager	Nombre_Manager	Apellido_Manager
102	Juan	Perez
103	Jhon	Doe
102	Juan	Perez
103	Jhon	Doe

D → **EF**



Cod_Jugador	Nombre	Apellido	Cod_Manager
1	Andrea	Juarez	102
2	Claudio	Lopez	103
3	Ezequiel	Gallo	102
4	Milagros	Alam	103

A → **BCD**



Cod_Manager	Nombre_Manager	Apellido_Manager
102	Juan	Perez
103	Jhon	Doe

D → **EF**



Clave primaria

La clave primaria identifica un registro único de una tabla.

ID_PAIS	DETALLE
1	ARGENTINA
2	BRASIL
3	URUGUAY
CLAVE PRIMARIA	

Clave foránea

La clave foránea relaciona los datos de un registro de una tabla con los de otra, de este modo indica cómo están relacionadas las tablas. Los datos en los campos de ambas tablas deben coincidir, aunque los nombres de los campos no sean los mismos.

ID_PAIS	DETALLE
1	ARGENTINA
2	BRASIL
3	URUGUAY
CLAVE PRIMARIA	

ID_PROVINCIA	DETALLE	ID_PAIS_FK
1	CORDOBA	1
2	RIO DE JANEIRO	2
3	COLONIA	3
CLAVE PRIMARIA		CLAVE FORÁNEA



**Para la próxima
clase**

nombre	apellido	nacionalidad	equipo	posicion
Lionel Andres	Messi	Argentina	FC Barcelona	Extremo derecho
Cristiano	Ronaldo	Portugal	Juventus	Extremo izquierdo
Neymar da Silva	Santos Junior	Brasil	Paris Saint-Germain	Extremo izquierdo
Jan	Oblak	Eslovenia	Atletico Madrid	Arquero
Eden	Hazard	Belgica	Real Madrid	Extremo izquierdo
Marc-Andre	ter Stegen	Alemania	FC Barcelona	Arquero
Virgil	Van Dijk	Holanda	Liverpool	Defensa central
Sergio Leonel	Aguero	Argentina	Manchester City	Centro delantero



Para la próxima clase

- Releer la diapositiva y el material para que podamos empezar con la parte práctica aplicando los conceptos.
- Analizar el archivo xlsx con la tabla y tratar de normalizar lo que se pueda.
- La próxima clase vamos a hacer la puesta en común del análisis que hayan hecho sobre la tabla.
- Instalar Mysql y Mysql-Workbench

