



UNIVERSIDAD
NACIONAL
DE COLOMBIA



Bases de Datos Introducción

Ing. Elizabeth León Guzmán, Ph.D.

MISSION MINTIC 2020

AGENDA



DATOS/INFORMACIÓN

MODELOS DE DATOS

INTRO MODELO RELACIONAL

NoSQL

¿Qué son los datos?

Un dato
es una representación simbólica de un
atributo o variable

Variable es
característica de la **población** que se desea medir

¿Qué son los datos?

Población: Personas que viven en mi barrio



¿Cuáles características tiene la población?

- Edad
- Género
- Nombre
- Dirección
- Teléfono

25
10
12
30
1
40
28
32

datos

Datos

Las variables cuantitativas son las que se expresan mediante un número. Se puede realizar operaciones aritméticas con ellas.

Las variables cualitativas son las que expresan características o cualidades, y no pueden ser medidas con números.



Datos

Ejemplo:

Población: Estudiantes del grupo 1 MINTIC
Variable: Edad

edad (x) = 23



Dato

La variable es cuantitativa,
rango
Real o discreta

Datos

Ejemplos:

Población: Ciudades de Colombia

Variables: Temperatura

Población

Otros ejemplos ?



Datos

Un dato es una representación simbólica de un atributo o variable



<http://www.datanami.com/>

Datos estructurados y
relacionados
“Información”



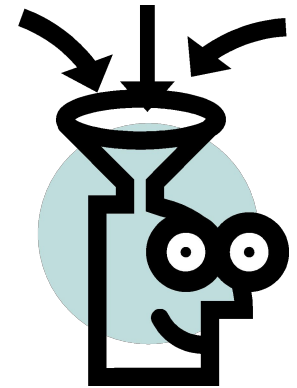
Datos almacenados
en “bases de datos”

Datos

Los datos son la materia prima bruta



INFORMACIÓN



Datos - Información



Datos almacenados

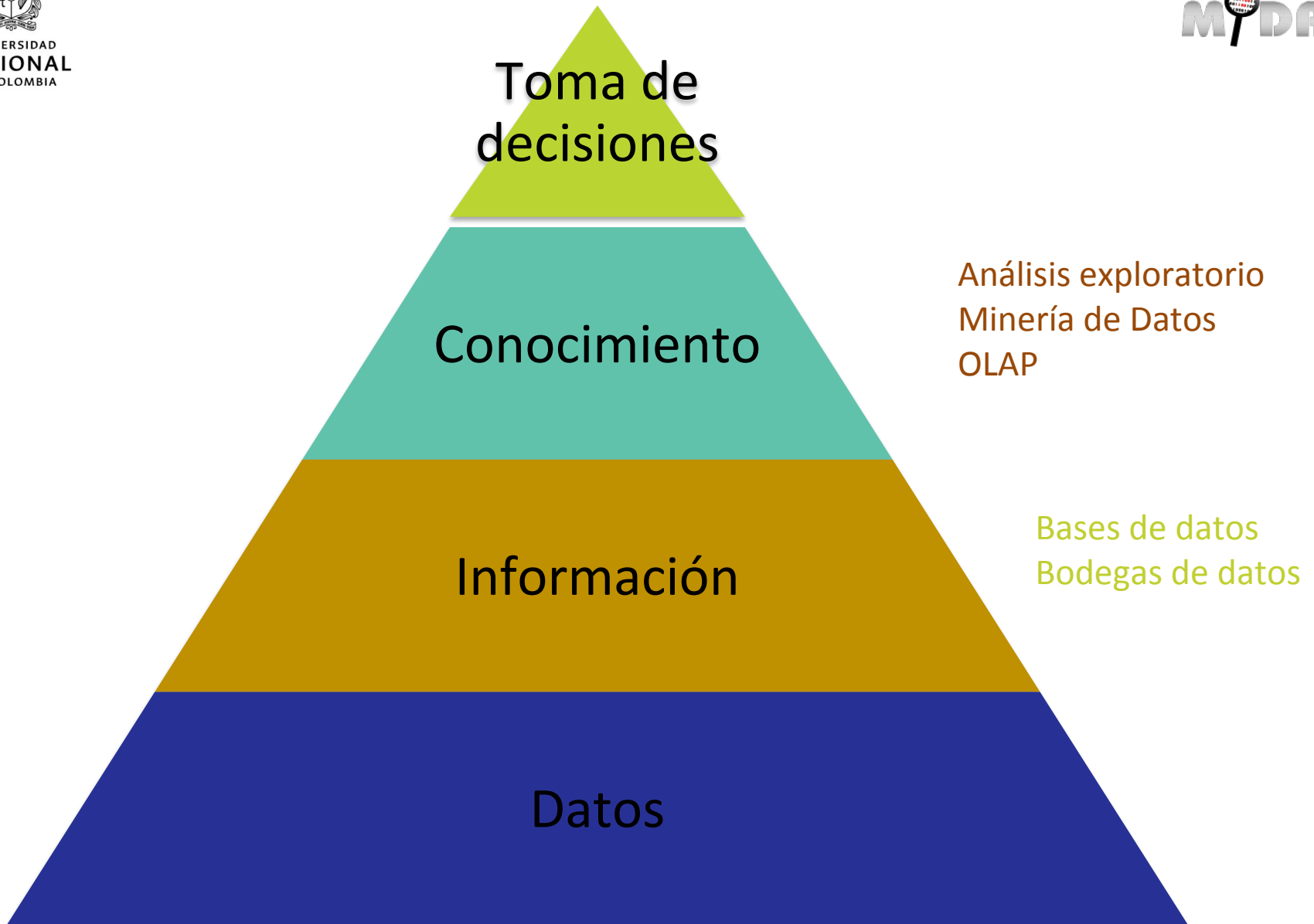
- Bases de datos
- Web
- Archivos (excel, pdf, txt, csv, etc)

Bases de Datos: Definición

Una colección de **datos relacionados**, y una descripción de estos datos, diseñados para cumplir con las necesidades de **información** de una organización

(Connolly & Begg)





Modelos de datos

Estructura para almacenar datos, se definen los tipos y cómo se relacionan

- Modelo relacional
- Modelo jerárquico
- Modelo de red
- Modelo POO
- Modelos NoSQL

Modelo relacional

Es el más usado desde el año 1970

Columna1	Columna2	Columna3	Columna4	Columna5
dato	dato	dato	dato	dato
dato	dato	dato	dato	dato
dato	dato	dato	dato	dato
dato	dato	dato	dato	dato

Modelo relacional

Es el más usado desde el año 1970

Estudiante

Nombre	Apellido	Edad	Genero	teléfono
Jorge	Díaz	20	m	3562819
María	Martínez	23	f	9873209
Rosa	Gómez	19	f	1743829
Pedro	Suarez	21	m	6386472

Modelo Relacional

Estudiante

Código	Nombre	Apellido	Edad	Genero	teléfono
100	Jorge	Díaz	20	m	3562819
101	María	Martínez	23	f	9873209
102	Rosa	Gómez	19	f	1743829
103	Pedro	Suarez	21	m	6386472

Asignatura

Código	Nombre	Créditos
10	Programación	4
20	Bases de Datos	3
30	Matemáticas	4
40	Software	3

Inscripción

Código_estudiante	Cod_asignatura	Semestre
100	10	2020-I
100	20	2020-I
100	10	2019-II
102	10	2020-I
102	20	2020-I
102	30	2020-I
103	30	2019-I
103	40	2019-I

Operaciones CRUD de BD



Creation - Crear- Inserción: Insertar datos en la bd



Read - Lectura: Leer datos (consultas) para obtener información



Update - Actualizaciones: Cambiar valor del dato



Delete - Borrado: Eliminar datos

Lenguaje
SQL : Structured Query Language

Not Only SQL

2009

Dynamo de Amazon
Bigtable de Google



No SQL

¿Por qué ahora?



Conectividad

Computación en la Nube

Variedad de datos



“Big Data”

“Big Data”

“Big Data” son todos los datos



Volumen

Peta bytes
Zetta bytes
Tablas, archivos



Variedad

- Datos Estructurados
- Datos No Estructurados
- Imágenes, texto, multimedia



Velocidad

- Flujos (streams)
- Real time
- Batch



Veracidad

- Autenticidad
- Integridad
- Integración
- Confiabilidad
- Ruido, inconsistencias, vagos, errores, etc.



Valor

Conocimiento
Análisis

5 Vs

Información - Almacenamiento

Bases de Datos Relacionales

Tradicionales, orientadas a lo transaccional



Bodegas de Datos

Almacenar datos históricos, agregados

Datos estructurados

Bases de Datos NoSQL

Orientadas a consulta
Formatos JSON, XML

documentales

grafos

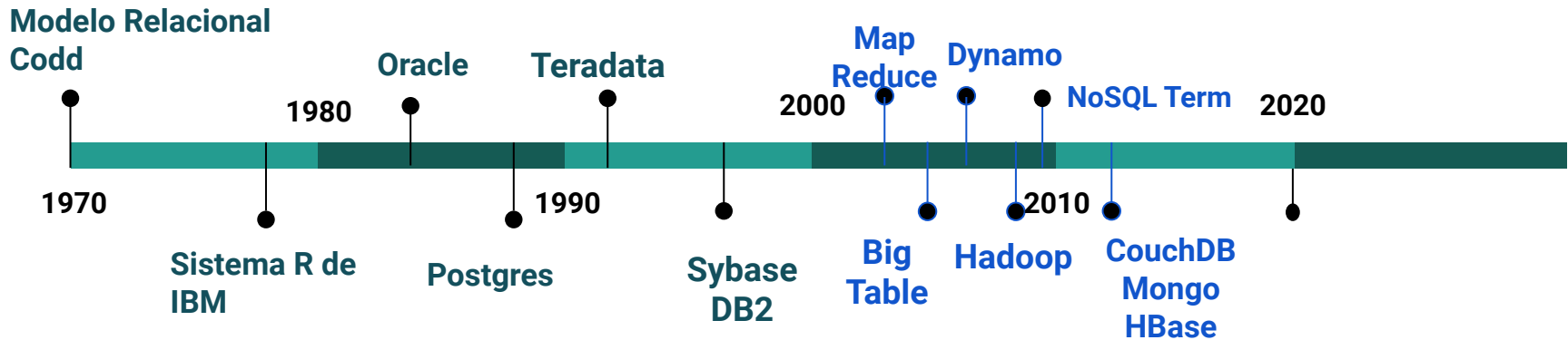


Sistema de Archivos

HDFS

Datos semi o NO estructurados

Breve Historia....



Código abierto

Modelos más flexible para almacenar, organizar y capturar mayor cantidad de datos.

¿Por qué NoSQL?

- Aplicaciones web ejecutan miles de transacciones por segundo:
 - Publicidad (Ad Serving)
 - Servicio de email (Email Services)
 - Carritos de compra (Shopping Carts)
 - Operaciones financieras (Financial Trades)
- Bases de datos tradicionales no capacitadas para el manejo de esas “tasas” de generación de datos.

¿Por qué NoSQL?

- Modelo flexible
- Evita gastos innecesarios en transacciones
- Evita la complejidad de consultas SQL

Permite:

- Cambios en la BD de manera fácil y frecuente.
- Desarrollo rápido
- Manejar grandes volúmenes de datos (ejemplo: Google)
- No definir esquema

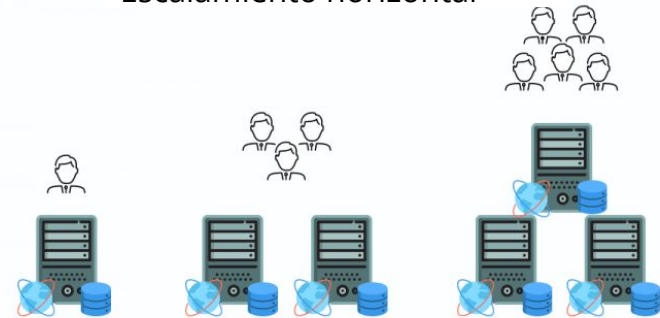
NoSQL

- Evitar la complejidad innecesaria.
- Alto rendimiento. Orientadas a consulta
- Escalabilidad Horizontal a bajo costo

Escalamiento vertical



Escalamiento horizontal



NoSQL



Amazon DynamoDB



Tipos de BD NoSQL

Cuatro grandes modelos



- Clave-Valor (Key-Value)
- Documento
- Familias de Columnas
- Grafos

Clave Valor



- Más popular y sencilla
- Datos son asociados por claves a valores. Los valores no conforman estructuras particulares.
- Estructura de “diccionario”
- Datos no requieren formato.

árbol : Un árbol es una planta, de tallo leñoso, que se ramifica a cierta altura del suelo.

Clave  **Valor** 

La clave debe ser única

Clave Valor



Clave (key)	Valor (Value)
101	La Ciudad y los Perros de Mario Vargas Llosa
102	Cien Años de Soledad de Gabriel García Márquez
201	La Divina Comedia de Dante Alighieri
202	Don Quijote de La Mancha de Miguel de Cervantes

`get ("102") -> "Cien Años de Soledad de Gabriel García Márquez"`

`put ("902", "La Iliada de Homero")`

Clave Valor



Clave (key)	Valor (Value)
101	La Ciudad y los Perros de Mario Vargas Llosa
102	Cien Años de Soledad de Gabriel García Márquez
201	La Divina Comedia de Dante Alighieri
202	Don Quijote de La Mancha de Miguel de Cervantes

No estructura

```
get ("102") -> "Cien Años de Soledad de Gabriel García Márquez"
```

```
put ("902", "La Iliada de Homero")
```

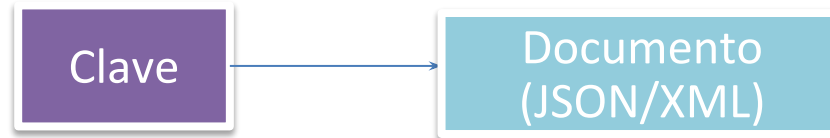
Clave Valor



Compañías que usan bd clave-valor:

- GitHub
- At&T
- BestBuy

Clave - Documento



Los claves referencian a *documentos* (JSON/XML)

Clave	Valor
101	{Titulo: "La Ciudad y los Perros", Autor:" Mario Vargas Llosa"}
102	{Titulo: "Cien Años de Soledad", Año: "1967"}
201	{Titulo: "La Divina Comedia", Año: "1304", Autor: "Dante Alighieri"}
202	{Titulo: "Don Quijote de La Mancha", Autor: "Miguel de Cervantes"}

strings, numeros, booleans, arreglos,
listas, diccionarios o objetos

Clave - Documento

clave	título	año	autor
101	La Ciudad y los Perros		Mario Vargas Llosa
102	Cien Años de Soledad	1967	
201	La Divina Comedia	1304	Dante Alighieri
202	Don Quijote de La Mancha		Miguel de Cervantes Saavedra

documento

documento

documento

documento

```
{
  101:{
    titulo: "La ciudad de los perros"
    autor: "Mario Vargas Llosa"
  }
}
```

```
{
  102:{
    titulo: "Cien Años de Soledad"
    año: 1967
  }
}
```

```
{
  201:{
    titulo: "La Divina Comedia"
    año: 1304
    autor: "Dante
  }
}
```

```
{
  202:{
    titulo: "Don Quijote de la Mancha"
    autor: "Miguel de Cervantes
    Saavedra"
  }
}
```

Clave - Documento

clave	titulo	año	autor
101	La Ciudad y los Perros		Mario Vargas Llosa
102	Cien Años de Soledad	1967	
201	La Divina Comedia	1304	Dante Alighieri
202	Don Quijote de La Mancha		Miguel de Cervantes Saavedra

← documento

← documento

← documento

← documento

```
{
  101:{
    titulo: "La ciudad de los perros"
    autor: "Mario Vargas Llosa"
  }
}
```

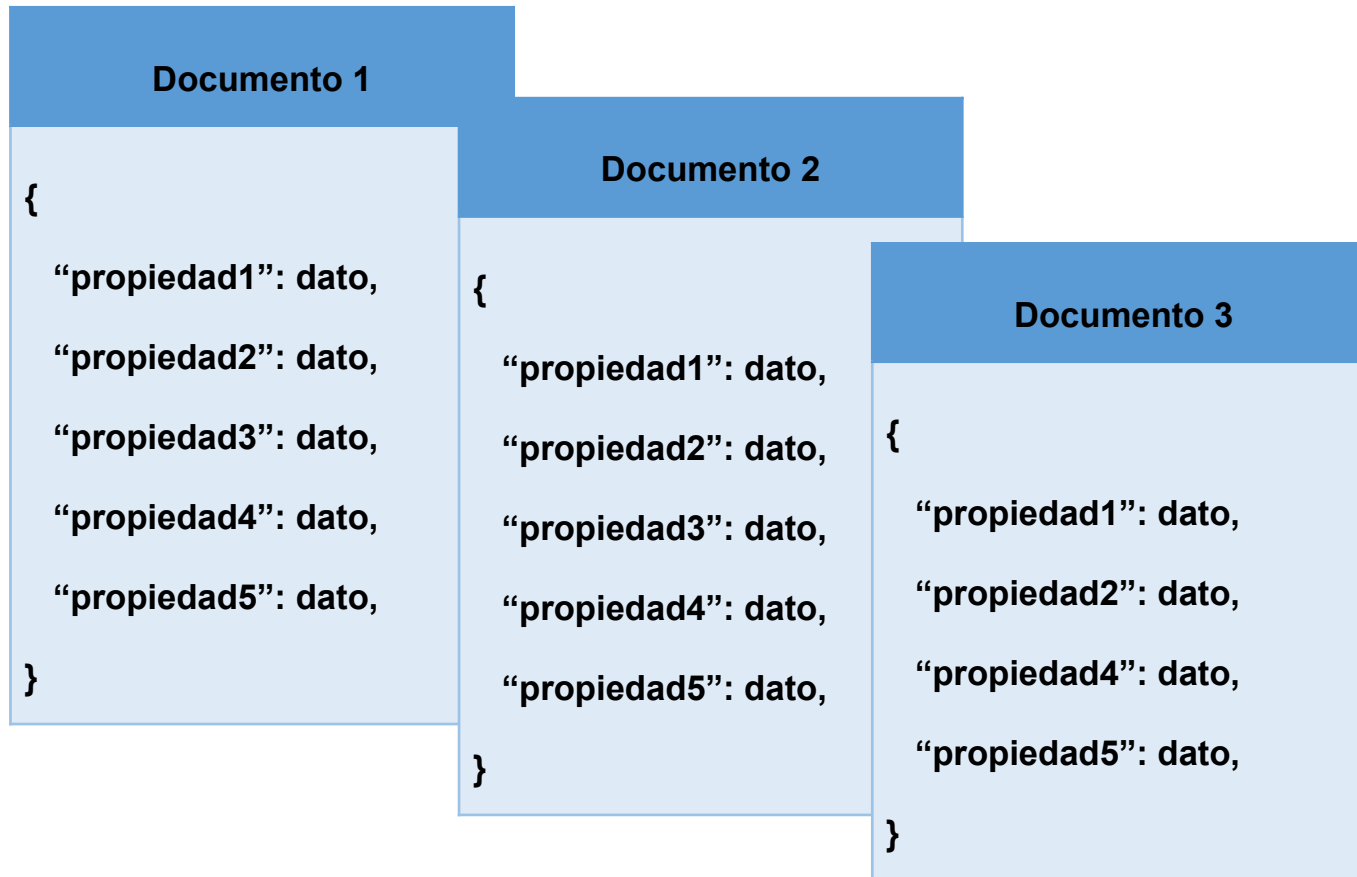
```
{
  102:{
    titulo: "Cien Años de Soledad"
    año: 1967
  }
}
```

```
{
  201:{
    titulo: "La Divina Comedia"
    año: 1304
    autor: "Dante
  }
}
```

```
{
  202:{
    titulo: "Don Quijote de la Mancha"
    autor: "Miguel de Cervantes
    Saavedra"
  }
}
```

Colección de
documentos de
Libros

Clave - Documento



Colección de documentos

Colección de libros en JSON

```
{  "Libros": [
    {
      _id: 101,
      titulo: "La ciudad de los perros",
      autor: "Mario Vargas Llosa",
    },
    {
      _id: 102,
      titulo: "Cien años de Soledad",
      año: 1967,
    },
    {
      _id: 201,
      titulo: "La Divina Comedia",
      año: 1304
      autor: "Dante Alighieri",
    },
    {
      _id: 202,
      titulo: "Don Quijote de la Mancha",
      autor: "Miguel de Cervantes Saavedra",
    }
  ]
}
```

Clave Documento

	Personal		Oficina	
clave	nombre	telefono	dirección	teléfono
001	Juan	3129765437	Av 20 de Julio 3-20	3789029
002	Luis		Cll 22 89-30	1092922
003	Maria	3107653902	Av 20 de Julio 3-20	3789029
004	Ana	3124527654		
005	Natalia	3124527654	Av del Sol 3 67	73647678
006	Andrés	3028680217	Cra 20 9 86	64734635

```

{
  codigo:001,
  personal: {
    nombre: "Juan"
    telefono: 3129765437
  },
  oficina: {
    dirección: "Av 20 de Julio 3-20"
    telefono: 3789029
  }
}

```

Sub-documentos
embebidos

Clave - Documento

```
[
  {
    codigo:001,
    personal: {
      nombre: "Juan"
      telefono: 3129765437
    }
    oficina: {
      dirección: "Av 20 de Julio 3-20 "
      telefono: 3789029
    }
  }
  {
    codigo: 002,
    personal: {
      nombre: "Luis"
    }
    oficina: {
      dirección: "Cl1 22 89-30"
      telefono: 1092922
    }
  }
]
```

```
{
  codigo: 003,
  personal: {
    nombre: "Maria"
    telefono: 3107653902
  }
  oficina: {
    dirección: "Av 20 de Julio 3-20 "
    telefono: 3789029
  }
}
{
  codigo: 004,
  personal: {
    nombre: "Ana"
    telefono: 3124527654
  }
}
...
]
```

Clave - Documento

```
[
  {
    codigo:001,
    personal: {
      nombre: "Juan"
      telefono: 3129765437
    }
    oficina: {
      dirección: "Av 20 de Julio 3-20 "
      telefono: 3789029
    }
  }
  {
    codigo: 002,
    personal: {
      nombre: "Luis"
    }
    oficina: {
      dirección: "Cl1 22 89-30"
      telefono: 1092922
    }
  }
]
```

```
{
  codigo: 003,
  personal: {
    nombre: "Maria"
    telefono: 3107653902
  }
  oficina: {
    dirección: "Av 20 de Julio 3-20 "
    telefono: 3789029
  }
}
{
  codigo: 004,
  personal: {
    nombre: "Ana"
    telefono: 3124527654
  }
}
...
]
```


Clave - Documento

Dos valores

```
{  
  codigo:001,  
  personal: {  
    nombre: "Juan"  
    apellido: "Sandoval"  
    telefono: [3129765437, 3212109287]  
  }  
  oficina: {  
    dirección: "Av 20 de Julio 3-20"  
    telefono: 3789029  
  }  
  codigo: 010,  
  personal: {  
    nombre: "Mariana"  
    apellido: "Diaz"  
    telefono: [3456788921, 3126673843]  
    correo: ["mariana@google.com", "mdiaz@yahoo.com" ]  
  }  
  oficina: {  
    dirección: "Av 3 89-10"  
    telefono: 97897866356  
  }  
}
```

Ejercicio

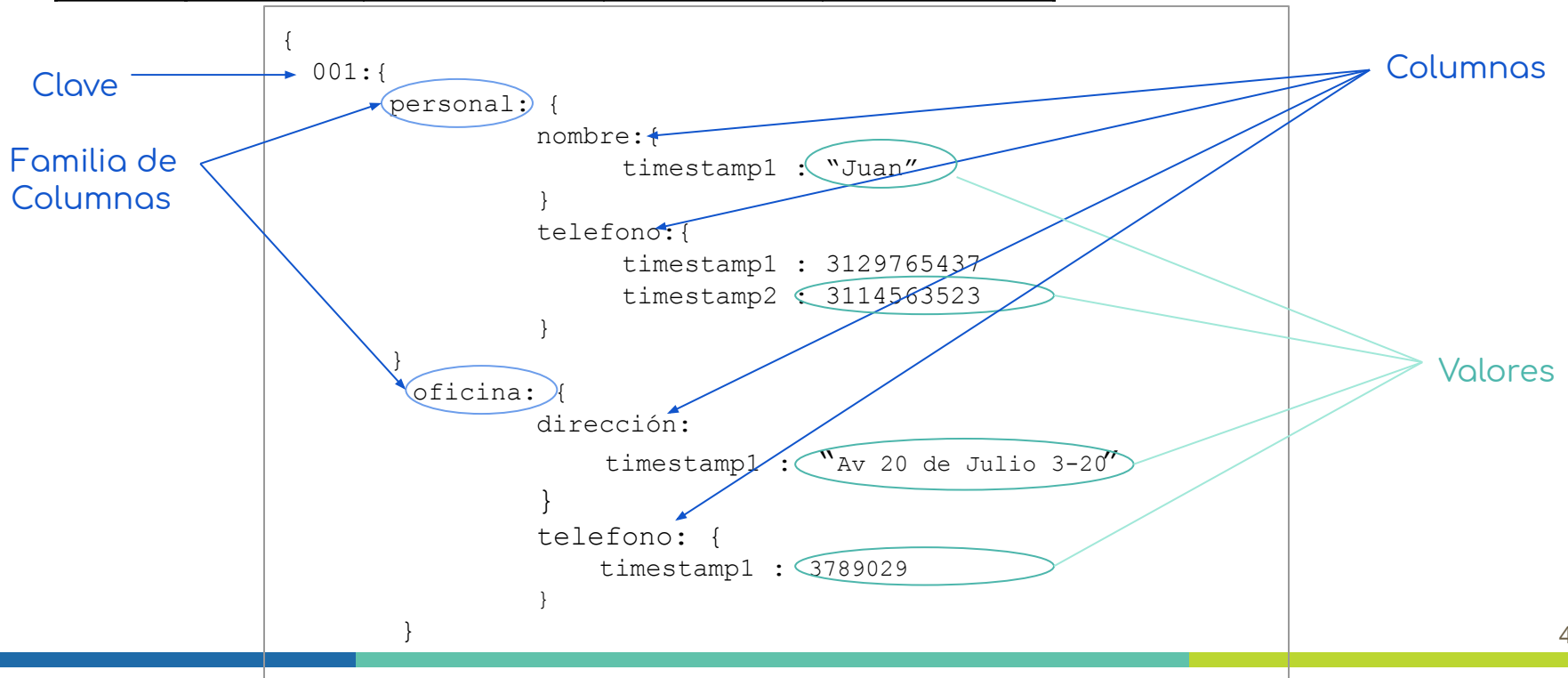
Crear un archivo JSON para almacenar datos de una agenda de personas:

- nombre,
- apellidos,
- dirección (una o más direcciones),
- teléfono (uno o más teléfonos),
- correo electrónico (ninguno, uno o más correos)

Familia de Columnas

	Personal		Oficina	
clave	nombre	telefono	dirección	teléfono
001	Juan	312 3114563523	Julio 3-20	3789029
002	Luis	3242432433	Cir 22 89-30	1092922
003	Maria	3107653902	Av 20 de Julio 3-20	3789029
004	Ana	3124527654	Av 20 de Julio 3-20	3789029
005	Natalia	3124527654	Av del Sol 3 67	73647678
006	Andrés	3028680217	Cra 20 9 86	64734635

Adiciona un dato, por ejemplo un nuevo teléfono. No sobrescribe, genera otra tupla para ese valor con un **timestamp**

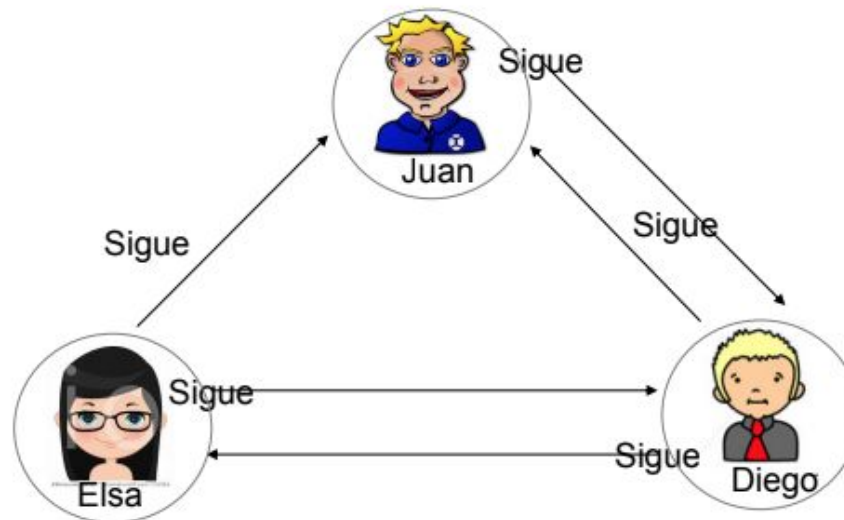


Bases de Datos en grafos

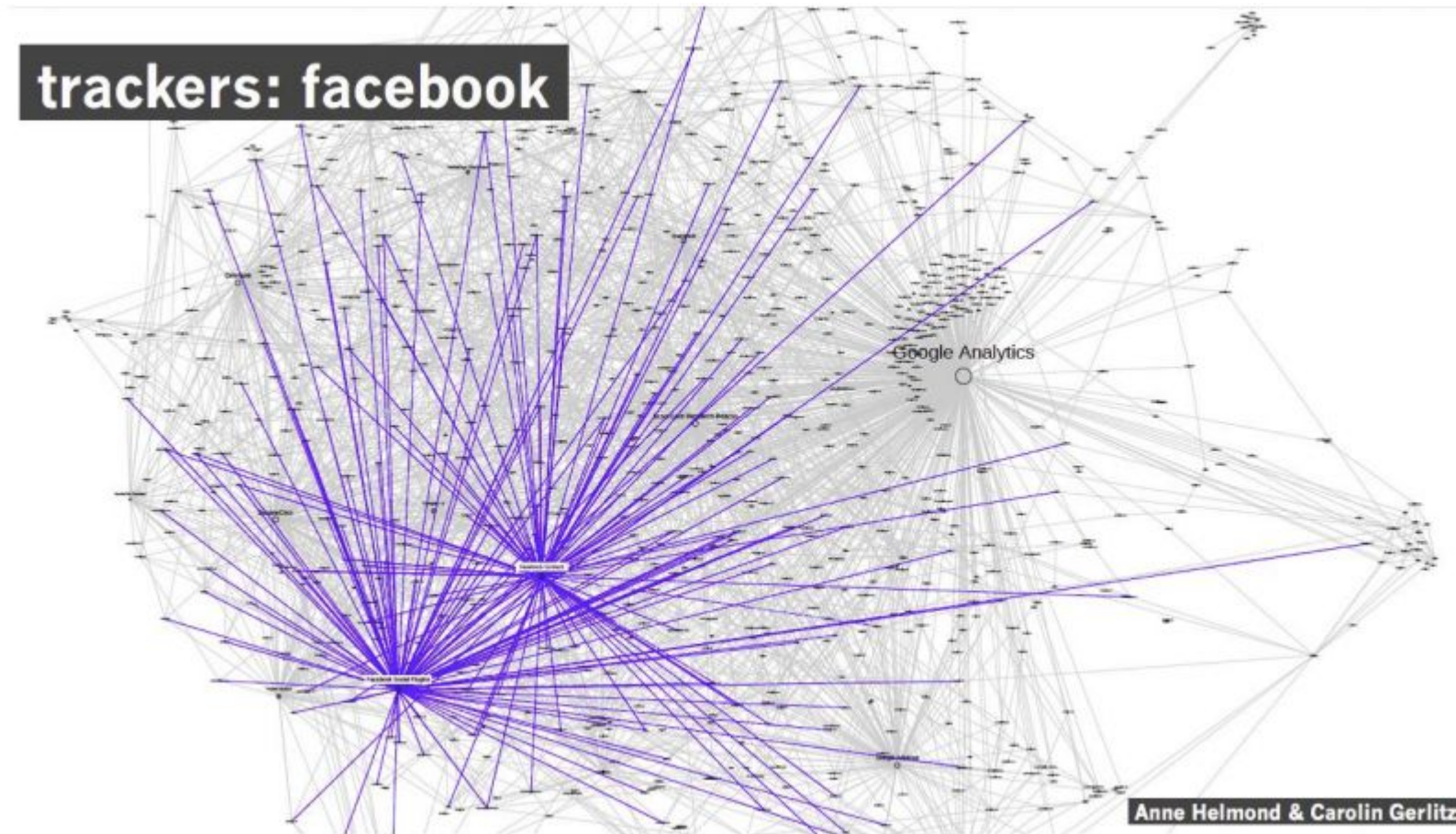
Bases de datos que usan como estructura un grafo para guardar los datos

Estructura de datos, que se compone de:

- Nodos → Entes (personas, lugares, cosas, etc)
- Propiedades → Información de los nodos y enlaces
- Enlace → Relaciones



Grafo de Internet



Modelado de NoSQL

- No se necesita considerar esquema. Los atributos características pueden ser adicionadas como van apareciendo.
- No hay celdas sin usar
- No es necesario tipos de datos. Son implícitos
- La mayoría de consideraciones son tenidas en cuenta en la capa de aplicación

Tarea

Crear archivos JSON desde Python para almacenar datos de ventas. Tener en cuenta información de una venta:

factura, cliente, información del cliente (opcional: nombre, teléfono, correo), vendedor, productos, precio de productos, cantidad de productos, total de la factura.

Referencias

- [1] León, E., Notas Asignatura Bases de Datos Unal.
- [2] Gillenson, M., Fundamentals of Data Management System.



UNIVERSIDAD
NACIONAL
DE COLOMBIA



Gracias

eleonguz@unal.edu.co