

Unidad N° 4

Titulo Unidad

Base de Datos Aplicada



Universidad Nacional
de La Matanza

Profesores

- **Jefe de Cátedra:** Jaír Hnatiuk
- **Profesores a cargo de los cursos:**
Ing. Valeria De Rosa

Temario

Modelos de datos heterogéneos: jerárquicos, texto, grafos.
El modelo BASE. Diferencias con ACID.
Datos estructurados y no estructurados.
MongoDB Query Language.
Modeos de Base de datos NoSQL
MongoDB.



Universidad Nacional
de La Matanza

DIIT



Departamento de Ingeniería e
Investigaciones Tecnológicas

Competencias



Universidad Nacional
de La Matanza

DIIT



Departamento de Ingeniería e
Investigaciones Tecnológicas

Competencias

- Identificación, formulación y resolución de problemas de ingeniería en sistemas de información/informática.
- Concepción, diseño y desarrollo de proyectos de ingeniería en sistemas de información / informática.
- Utilización de técnicas y herramientas de aplicación en la ingeniería en sistemas de información / informática.
- Generación de desarrollos tecnológicos y/o innovaciones tecnológicas.
- Desempeño en equipos de trabajo.
- Comunicación efectiva.
- Actuación profesional ética y responsable.
- Evaluación y actuación en relación con el impacto social de su actividad profesional en el contexto global y local.
- Aprendizaje continuo.
- Desarrollo de una actitud profesional emprendedora.

Pre Requisitos

Arquitectura de las bases de datos transaccionales
Modelo ACID
Control de concurrencia. Transacciones



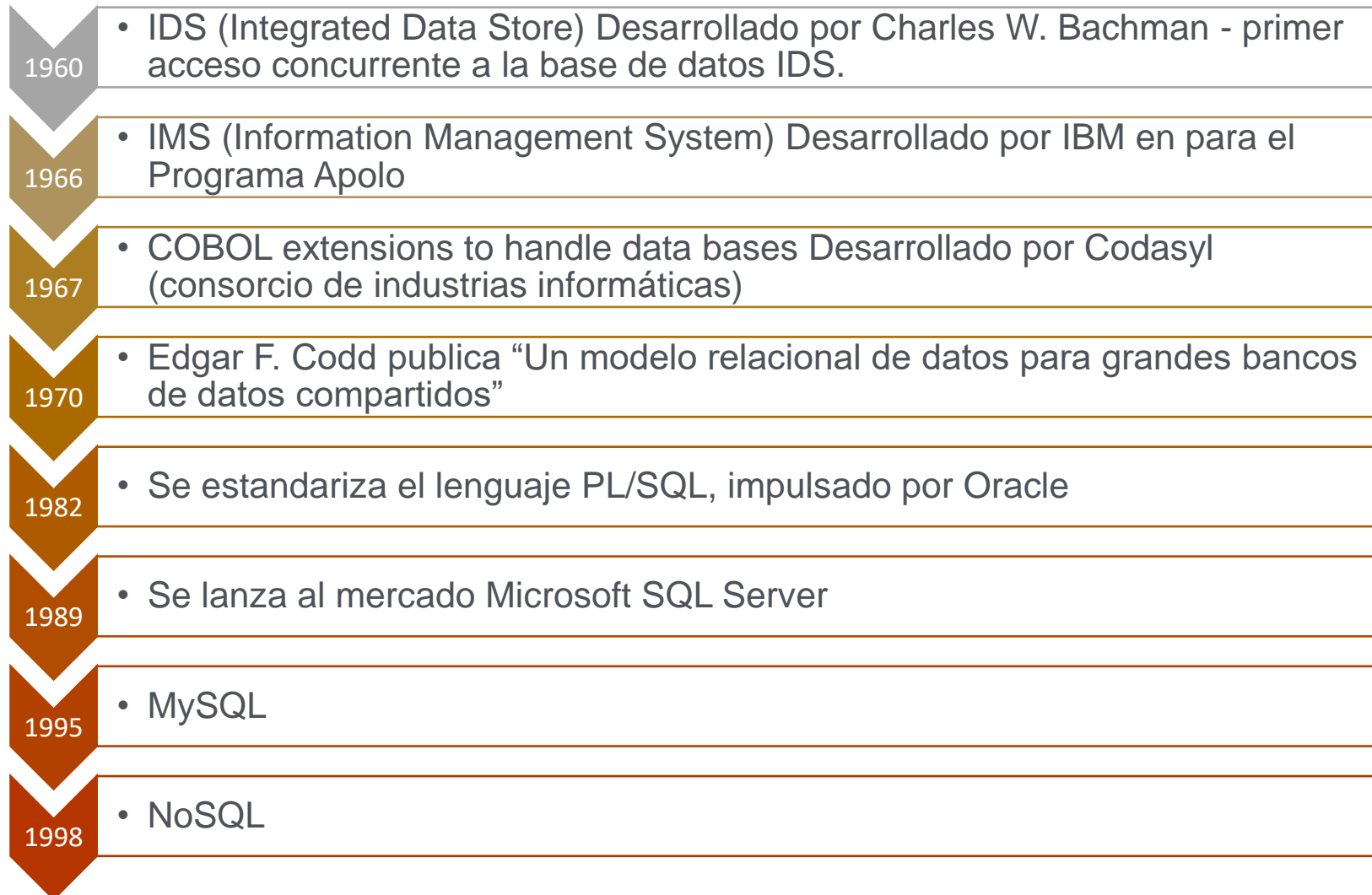
Universidad Nacional
de La Matanza

DIIT



Departamento de Ingeniería e
Investigaciones Tecnológicas

Surgimiento de base de datos NoSQL



Surgimiento de base de datos NoSQL

*Problemáticas a las que se enfrenta los RDBMS
(sistema de gestión de bases de datos relacionales)*

- Costo (Hardware y Gestión)
- Volumen de datos
- Tiempos de respuesta de baja latencia
- Alta disponibilidad
- Problemas de rendimiento

Base de datos NoSQL

El auge de este tipo de base de datos se produjo por la necesidad de cubrir las carencias de las RDBMS

Escalabilidad: servicios en la nube - Entornos distribuidos.

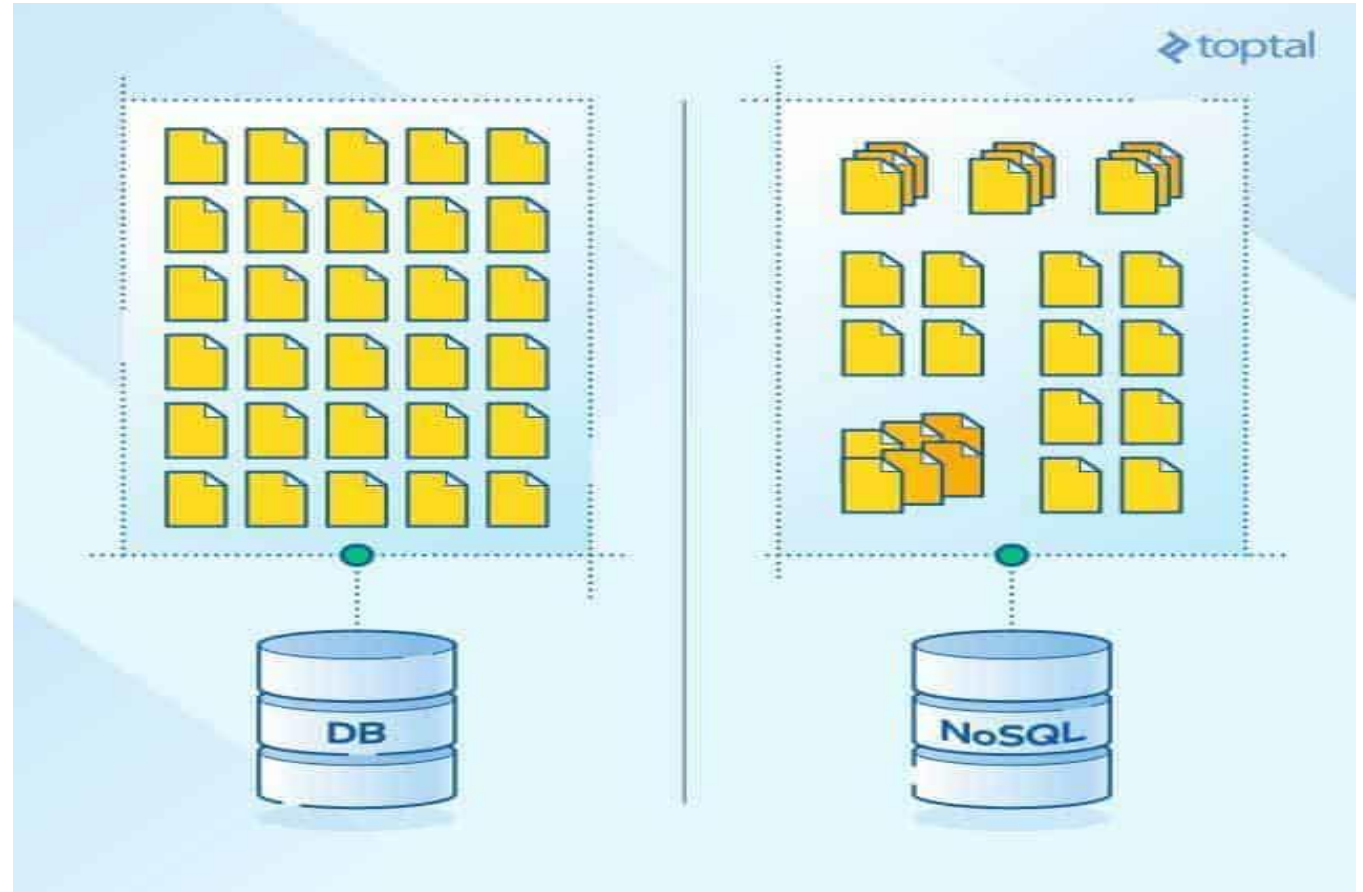
Flexibilidad: modelo de datos flexible - Datos semiestructurados y no estructurados.

Alta disponibilidad: garantiza la continuidad del servicio

Rendimiento: tiempo de respuesta de las consultas muy superior con respecto a las RDBMS tradicionales

Características de bases de datos NoSQL

- Los modelos NoSQL no tienen una estructura definida
- La información no se organiza en registros o campos
- Sistemas distribuidos de datos



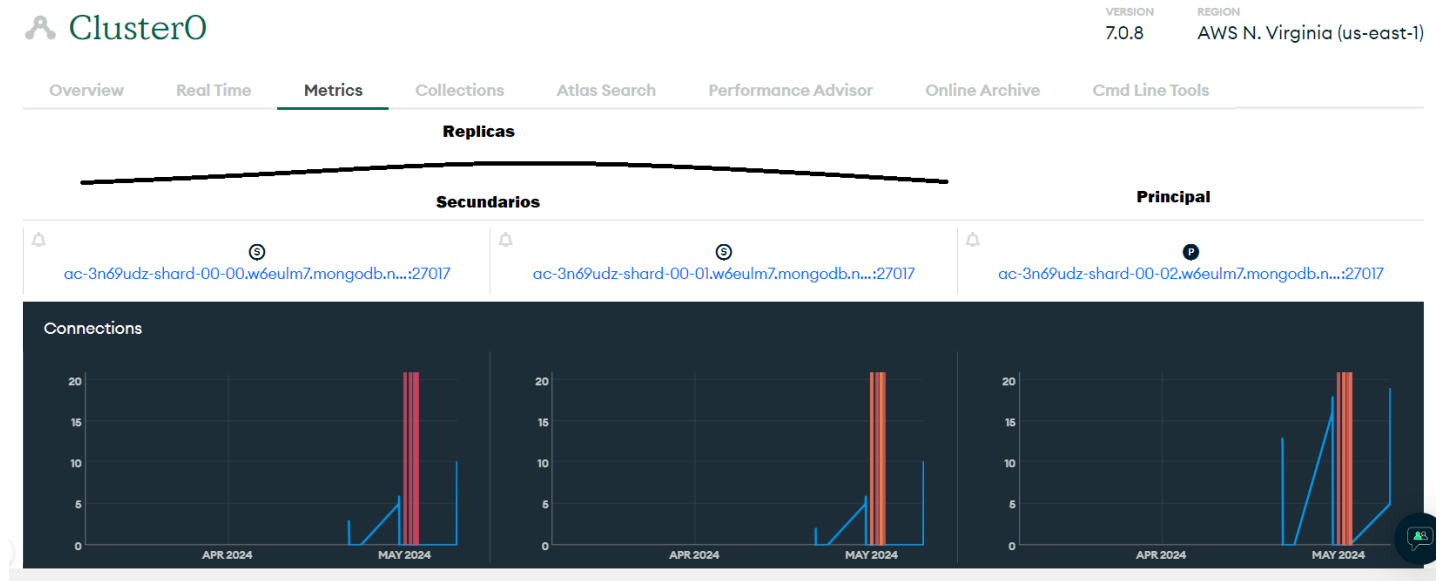
Características de bases de datos NoSQL

Arquitectura

- Clúster en la nube
- Clúster de servidores dedicados

La decisión sobre cual arquitectura implementar se basa en:

- Costo (Capex y Opex)
- Complejidad (infraestructura y el personal técnico calificado)
- Tiempo de implementación (configurar y puesta a punto)



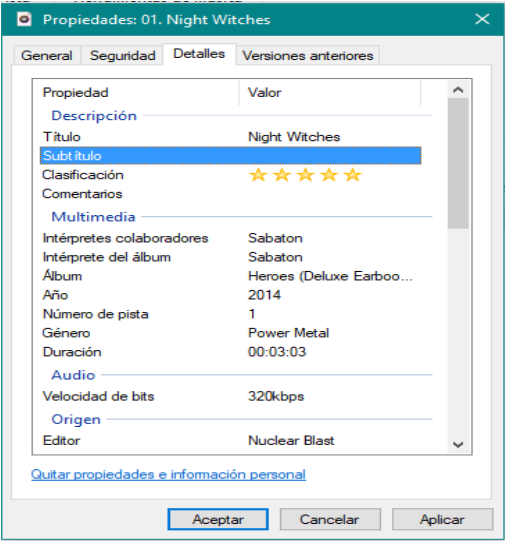
Base de Datos Aplicada

Datos estructurados, semi estructurados y no estructurados

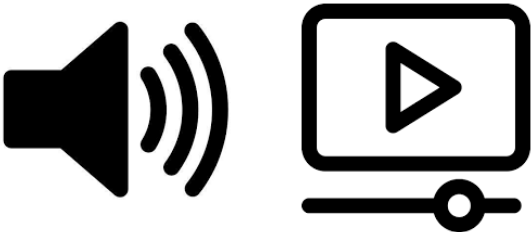
Datos estructurados

TIPO DE DATO	TAMAÑO	WRAPPER
Boolean	1 bit	Boolean
Char	16 bits	Character
Byte	8 bits	Byte
Short	16 bits	Short
Int	32 bits	Integer
Long	64 bits	Long
Float	32 bits	Float
Double	64 bits	Double

Datos semi estructurados



Datos no estructurados

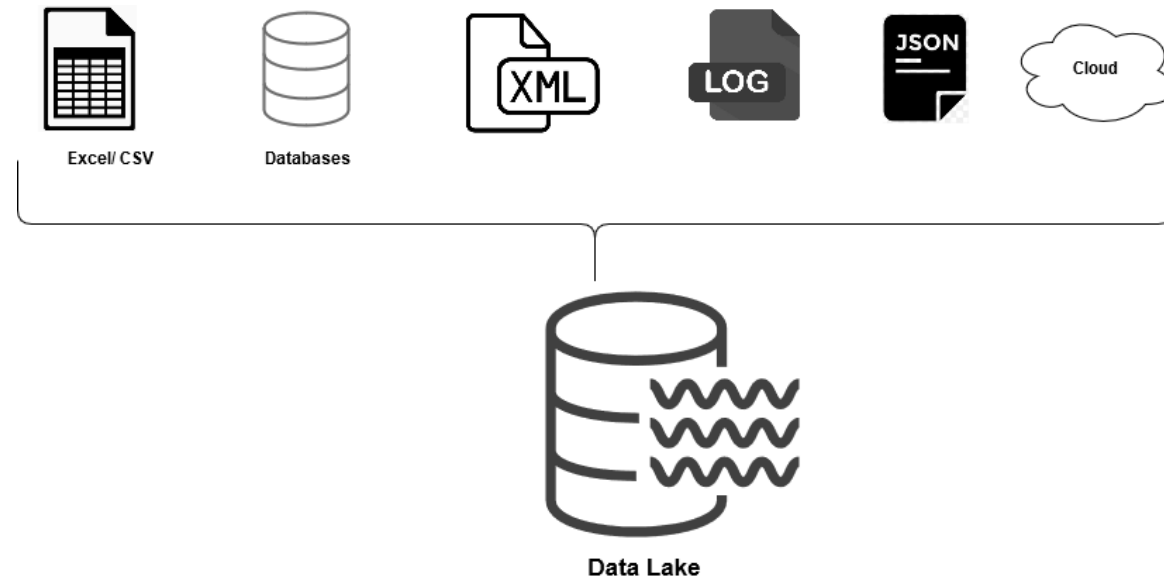


Base de Datos Aplicada

Datos estructurados, semi estructurados y no estructurados

Almacenamiento de datos - Procesamiento de datos

- Reconocimiento Óptico de Caracteres (OCR)
- Procesamiento del Lenguaje Natural (NLP)



Datos estructurados, semi estructurados y no estructurados

Ejemplo

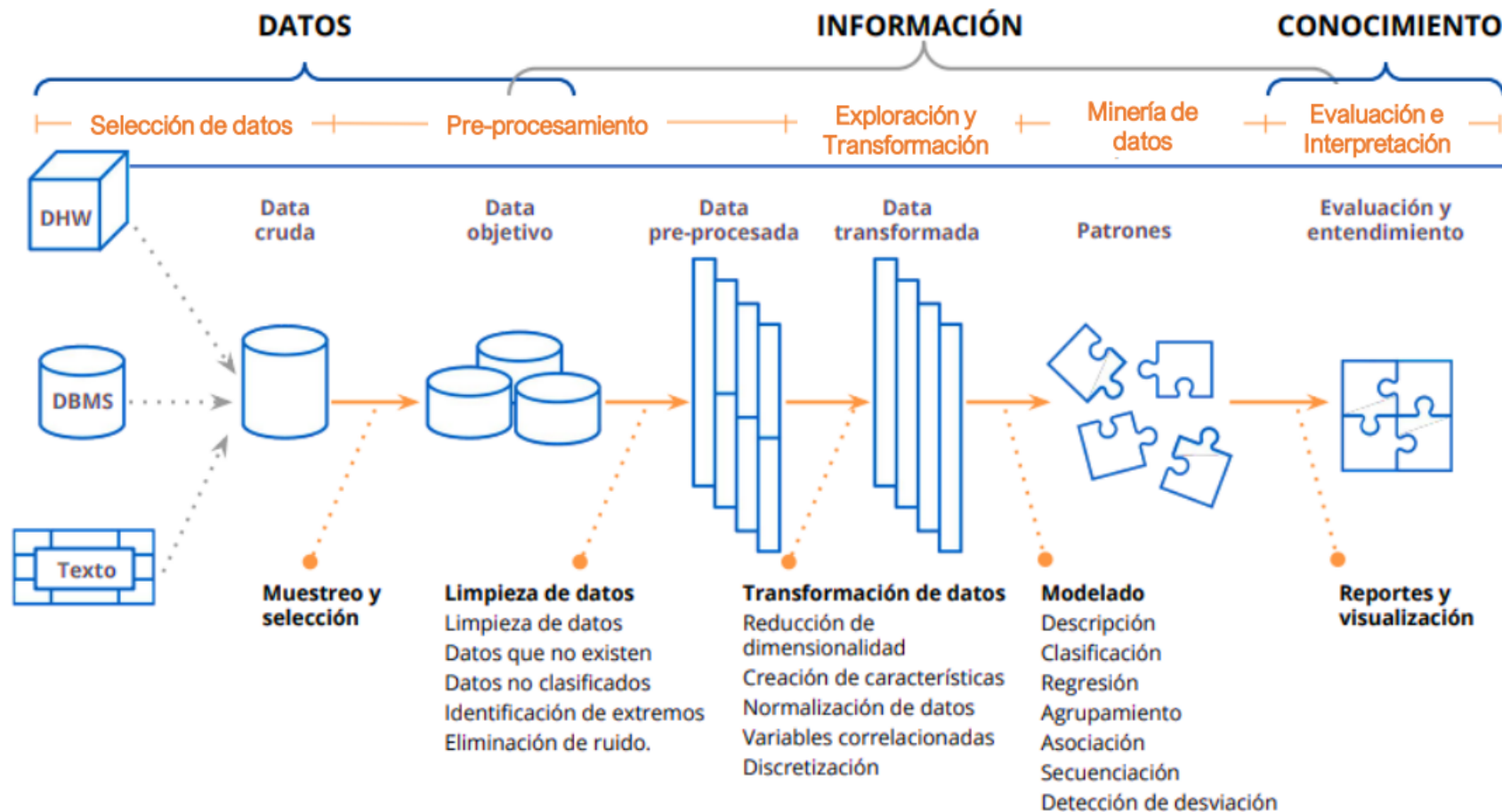
Una empresa de servicios de internet trata de evitar que sus clientes emigren hacia la competencia. Para ello detectan cuales son las actitudes/señales que tiene un cliente antes de cambiarse de empresa.

La empresa cuenta con:

- Información histórica: Facturación, consumos, uso de los datos, plan, costo del plan (datos estructurados)
- Información del cliente: Atención de soporte recibida (llamadas al call center), navegabilidad en las páginas de otra empresa (ticket de consumo 4G), zonas en las que habitualmente se desplaza (antenas propias y de otras empresas), días en los que consume más datos u otros servicios (datos no estructurados)

Datos estructurados, semi estructurados y no estructurados

Proceso de Minería de datos



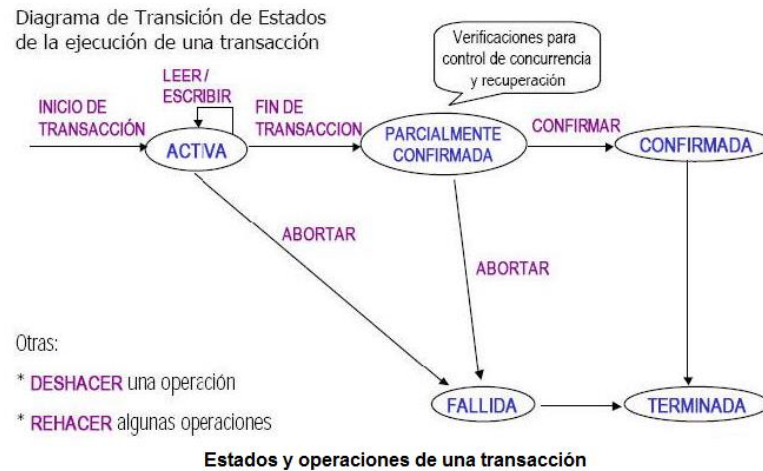
Propiedades de los modelos de BBDD

ACID

- Atomicidad
- Consistencia
- Aislamiento
- Durabilidad

BASE

- Basically Available
- Soft State
- Eventually Consistent



Propiedades de los modelos de BBDD

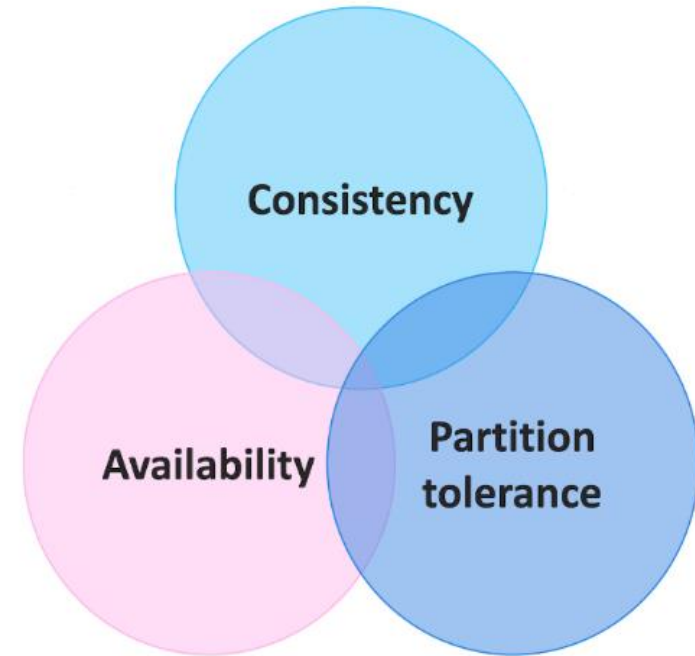
Teorema CAP

Eric Brewer año 2000

C - Consistencia (Consistency): Se refiere a la lectura coherente del valor actual del dato desde cualquier instancia, es decir que los datos se encuentran sincronizados y replicados en todos los nodos a la vez.

A - Disponibilidad (Availability): Se refiere a obtener una respuesta válida y rápida para todas las solicitudes, aunque existan nodos inactivos, es decir el acceso a los datos sin interrupciones.

P - Tolerancia a particiones (Partition tolerance): Se refiere a la capacidad del sistema para permanecer estable y continuar procesando solicitudes a pesar de ocurrir una partición (interrupción) entre la comunicación de los nodos.



Modelos de base de datos NoSQL

- Base de datos de pares clave-valor
- Bases de datos de documentos
- Bases de datos de grafos
- Bases de datos Columnas y familias de columnas

Modelos de base de datos NoSQL

Base de datos de pares clave-valor

Claves: identificadores asociados con valores

Valores: cadena de caracteres, número, imágenes u objetos binarios

Base de datos relacional

ID (Int)	Name (Varchar)	Age(int)
1	Sergio	22
2	Ana	48
3	Pablo	49
4	Juan	12

Base de datos Clave-Valor

Key	Value
1	Sergio, Andres, 22,19/09/1994
2	Ana, 12/08/1969
3	Pablo
4	Juan, 12

Ejemplos de aplicación

Almacenamiento en caché de datos

Administración de sesiones

Administración de perfiles y preferencias de usuario

Recomendación de producto y servicio

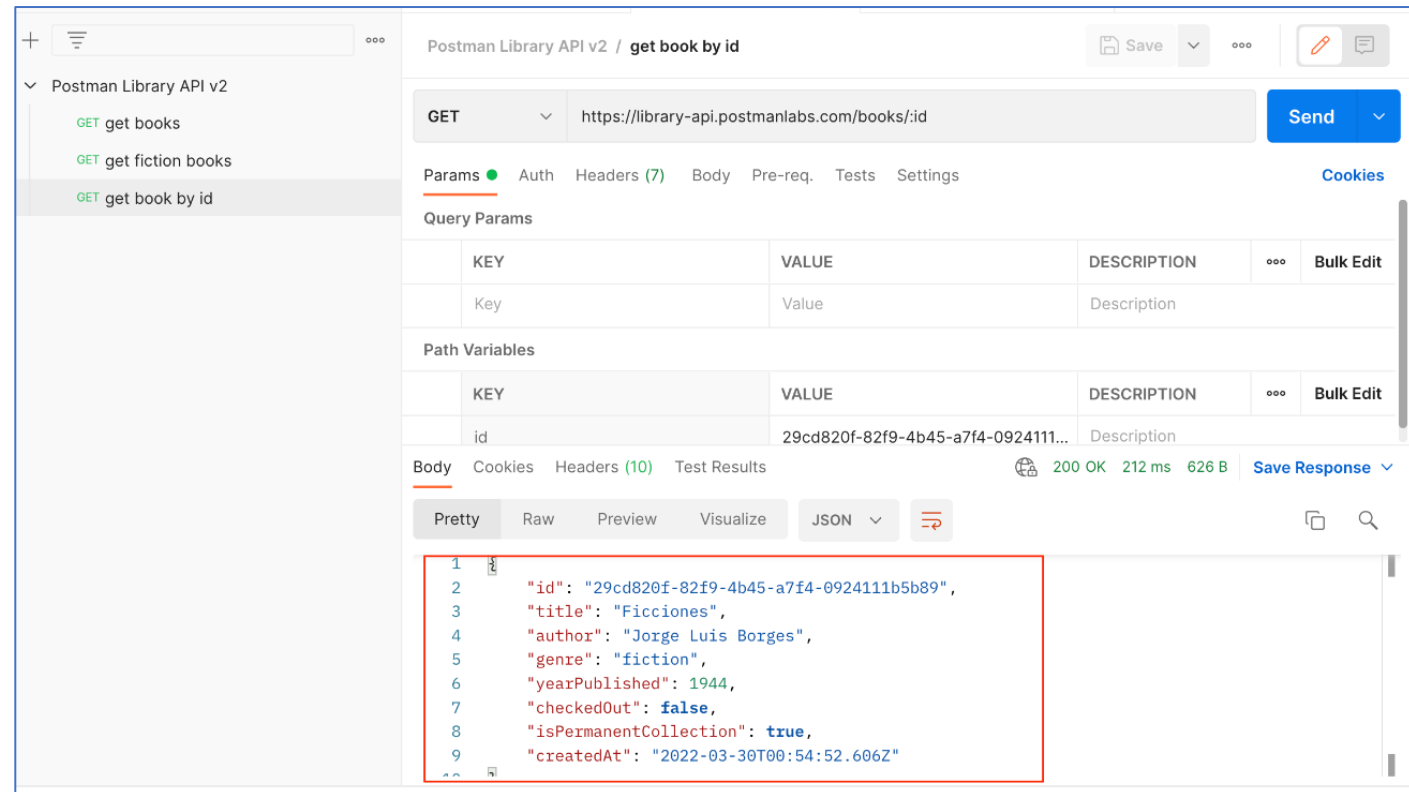
Modelos de base de datos NoSQL

Bases de datos de documentos

Los documentos son entidades semiestructuradas JSON o XML.

Ejemplos de aplicación

- Catálogo de productos
- Administración de contenido
- Administración de inventario



The screenshot shows the Postman interface for an API named 'Postman Library API v2'. The selected endpoint is 'get book by id' with a GET method. The URL is 'https://library-api.postmanlabs.com/books/:id'. The 'Query Params' table is empty. The 'Path Variables' table has one entry: 'id' with the value '29cd820f-82f9-4b45-a7f4-0924111b5b89'. The response status is '200 OK' with a body size of '626 B'. The response body is displayed in 'Pretty' JSON format, showing a book document.

KEY	VALUE	DESCRIPTION	...	Bulk Edit
Key	Value	Description		

KEY	VALUE	DESCRIPTION	...	Bulk Edit
id	29cd820f-82f9-4b45-a7f4-0924111b5b89	Description		

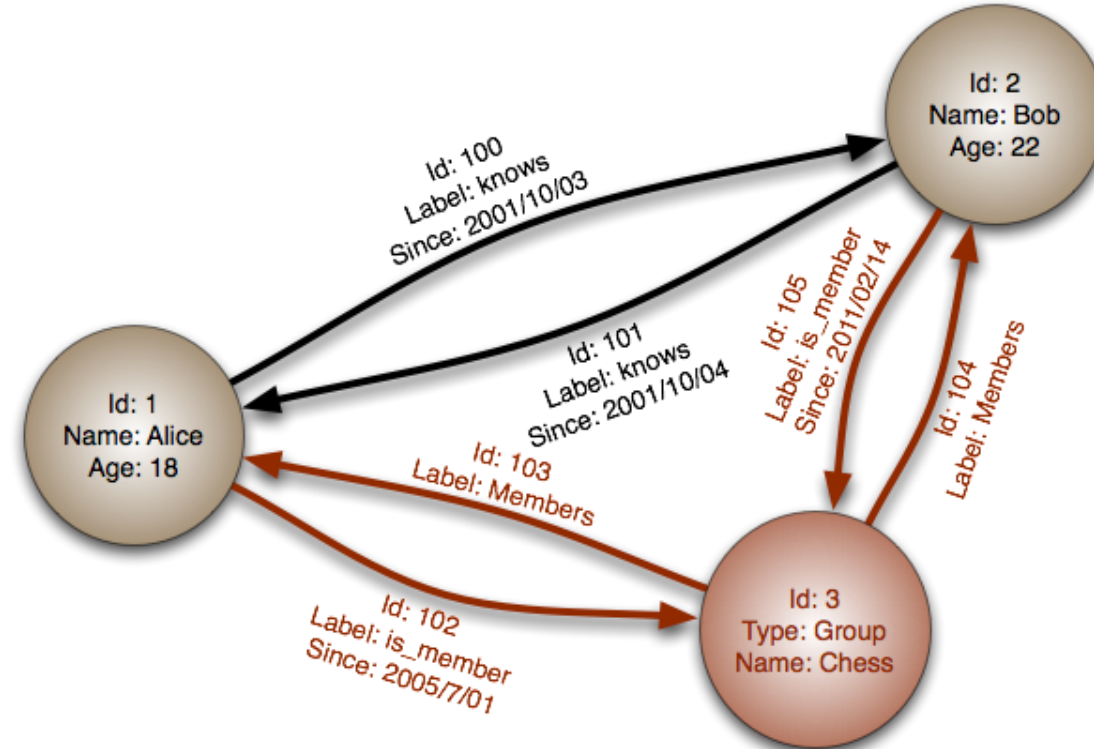
```
1 {
2   "id": "29cd820f-82f9-4b45-a7f4-0924111b5b89",
3   "title": "Ficciones",
4   "author": "Jorge Luis Borges",
5   "genre": "fiction",
6   "yearPublished": 1944,
7   "checkedOut": false,
8   "isPermanentCollection": true,
9   "createdAt": "2022-03-30T00:54:52.606Z"
10 }
```

Modelos de base de datos NoSQL

Bases de datos de grafos

Un nodo es un objeto que tiene un identificador y un conjunto de atributos. (vértice)

Una relación es un enlace entre dos nodos que contienen atributos sobre esa relación (aristas)



Ejemplos de aplicación

- Organigramas
- Gráficos sociales
- Detección de fraudes
- Motores de recomendaciones

Modelos de base de datos NoSQL

Bases de datos Columnas y familias de columnas

Una columna es una unidad básica de almacenamiento en una base de datos de familia de columnas

Ejemplos

- Recomendaciones
- Personalización
- datos del sensor
- Telemetría
- Mensajería
- Análisis de redes sociales
- Análisis web
- Supervisión de la actividad
- El tiempo y otros datos de serie temporal

Base de Datos Relacional

ROWID	Matricula	Modelo	Precio
1	6548 HCF	Fiat Bravo	9861
2	6589 GDB	VW Passat	12500
3	3215 FGD	Ford Fiesta	4589
4	4836 DVN	Audi A6	8956

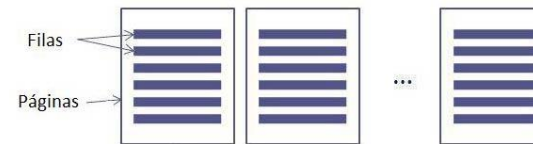
Base de Datos Orientada a Columnas

ROWID	Matricula
1	6548 HCF
2	6589 GDB
3	3215 FGD
4	4836 DVN

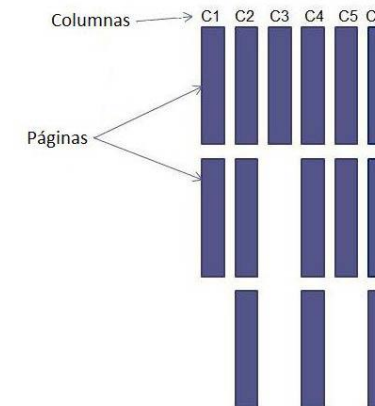
ROWID	Modelo
1	Fiat Bravo
2	VW Passat
3	Ford Fiesta
4	Audi A6

ROWID	Precio
1	9861
2	12500
3	4589
4	8956

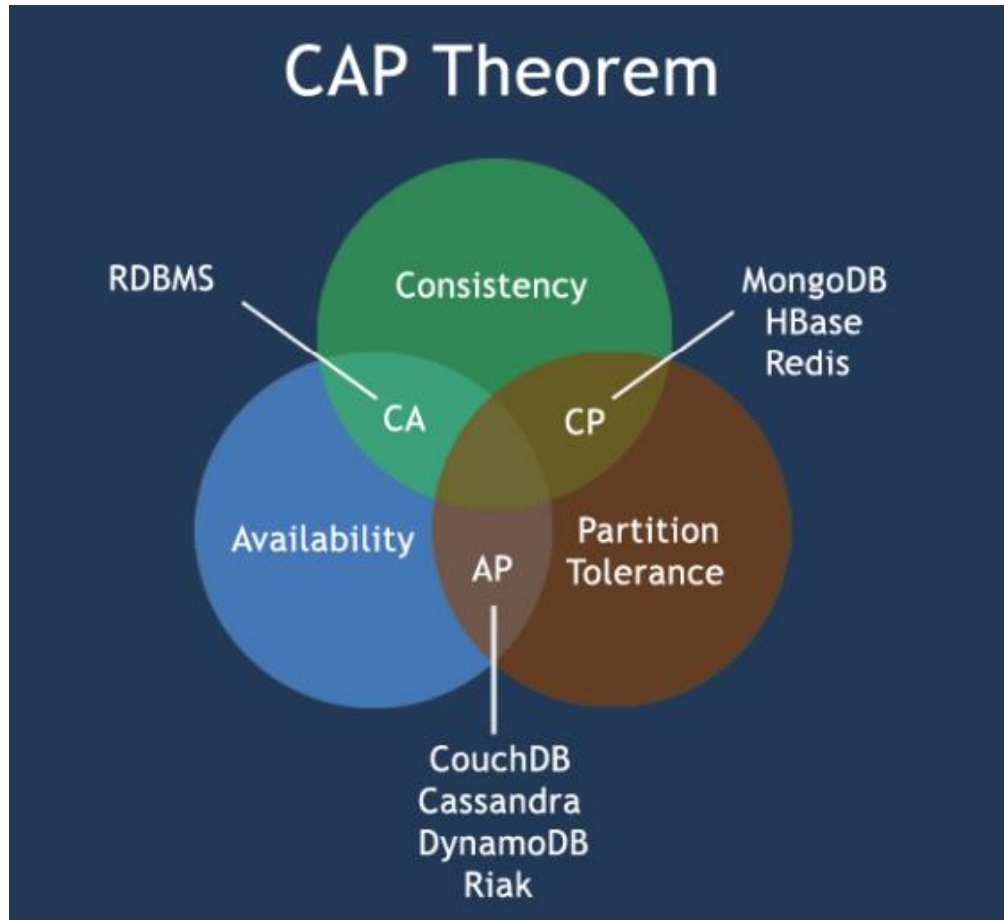
Almacenamiento en filas



Almacenamiento Columnstore



Modelos de base de datos NoSQL



Columnar:

1	Things	A	foo	B	bar	C	baz
2	Things	C	bam	E	coh	People	A Emmanuel
3	Languages	A	C	B	Java	C	Ceylon

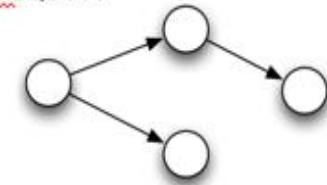
Key-value:

key	value
123	Address@23
126	"Booya"

Documental:

```
{ "user" : {  
  "id": "124",  
  "name": "Emmanuel",  
  "addresses" : [  
    { "city": "Paris", "country": "France" },  
    { "city": "Atlanta", "country": "USA" }  
  ]  
}
```

Graph:



¿Dudas?



Universidad Nacional
de La Matanza

DIIT



Departamento de Ingeniería e
Investigaciones Tecnológicas

Conocimiento Adquirido

Surgimiento de base de datos NoSQL

Características de bases de datos NoSQL

Datos estructurados, semi estructurados y no estructurados.

Propiedades de los modelos de BBDD

ACID

BASE

Diferencias entre ACID y BASE

Teorema CAP

Modelos de base de datos NoSQL

Base de datos de pares clave-valor

Bases de datos de documentos

Bases de datos de grafos

Columnas y familias de columnas



Universidad Nacional
de La Matanza

DIIT



Departamento de Ingeniería e
Investigaciones Tecnológicas

Bibliografía

NoSQL for Mere Mortals - Dan Sullivan

Microsoft

Desafío de conocimientos de la nube - Documentacion

<https://learn.microsoft.com/es-es/azure/architecture/guide/technology-choices/data-store-overview>

IBM - Teorema CAP

<https://www.ibm.com/mx-es/topics/cap-theorem>

ASW-Qué es NoSQL

<https://aws.amazon.com/es/nosql/>



Universidad Nacional
de La Matanza

DIIT



Departamento de Ingeniería e
Investigaciones Tecnológicas