

Bases de Datos IA





Agenda

- Definiciones
- Evolución histórica
- Surgimiento de las bases de datos no relacionales
- Necesidades que cubren las bases de datos no relacionales
- Definición de base de datos NoSQL
- Instalación de herramientas



Definiciones

Arquitectura de datos

- Es el diseño estructural de ambientes de información relacionados y compartidos
- Comprende el análisis, organización, disposición y estructuración de la información, como así también la selección y presentación de los datos
- Involucra actividades de modelado o conceptualización de los datos



Definiciones

Base de datos

- Es una colección organizada de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso.
- Generalmente, los datos son almacenados y accedidos de forma electrónica a través de una computadora



Definiciones

DBMS (Database Management System)

- Conjunto integrado de componentes de software que permiten a los usuarios interactuar con una o más bases de datos y proveen acceso a todos los datos contenidos en ella.
- Interactúa con usuarios finales, aplicaciones y la base de datos para capturar y analizar los datos.



Definiciones

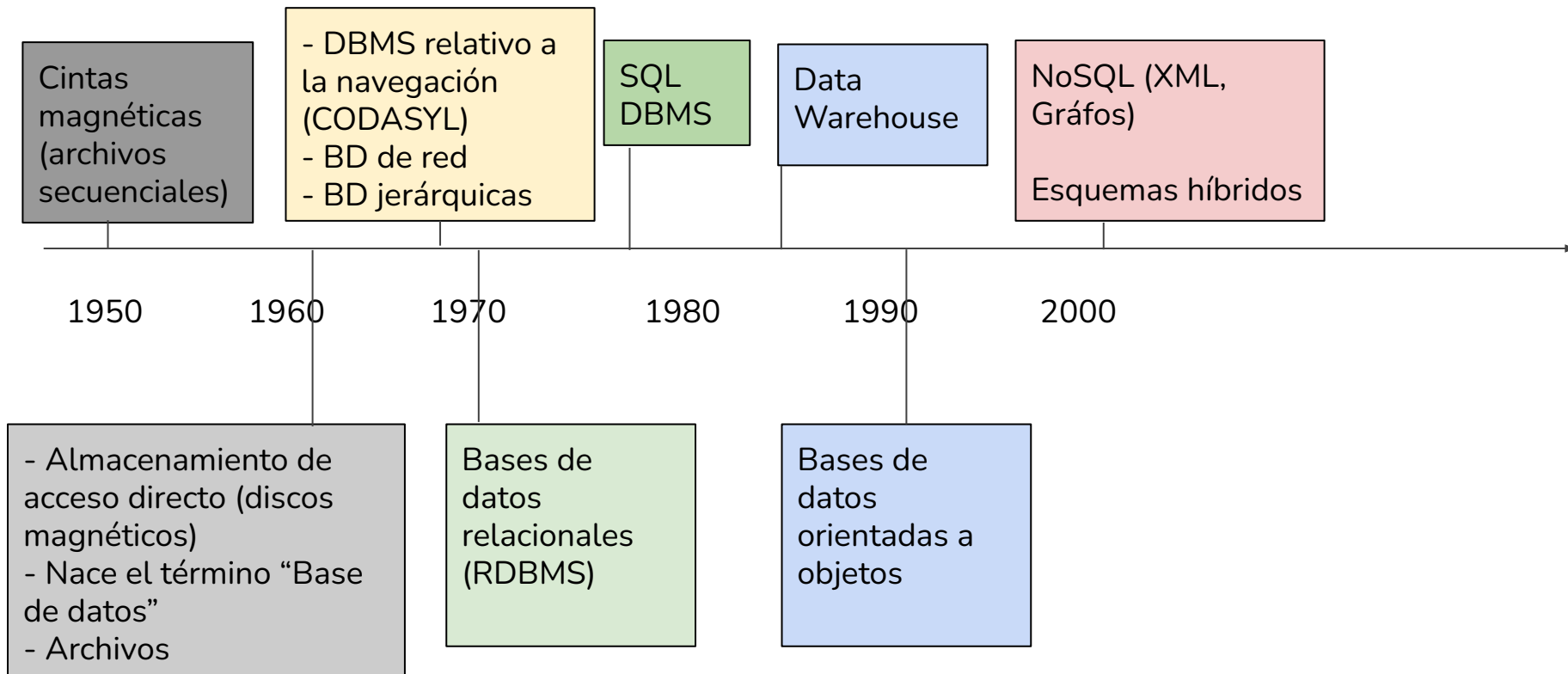
DBMS (Database Management System)

La funcionalidad provista por un DBMS se clasifica en cuatro grupos funcionales:

1. Definición de datos
2. Actualización (altas, bajas, modificaciones)
3. Recuperación
4. Administración



Historia





Clasificación de BD

Basada en el modelo de datos:

- Bases de datos relacionales
- Bases de datos orientadas a objetos
- Bases de datos jerárquicas
- Bases de datos de red
- Bases de datos NoSQL



Clasificación de BD

Basada en la distribución de los datos:

- Bases de datos centralizadas
- Bases de datos distribuidas
 - Sistemas homogéneos
 - Sistemas heterogéneos



Clasificación de BD

Otras clasificaciones:

- Variabilidad de los datos: estáticas o dinámicas
- Contenido: según la información que almacenan
- Número de usuarios: monousuario o multiusuario



Bases de datos relacionales

- Basadas en el modelo relacional propuesto e introducido por E.F. Codd en 1970
- Ampliamente utilizadas durante las últimas 5 décadas
- Centralizadas o distribuidas
- Modelo cliente-servidor
- Lenguaje de consulta estructurado (SQL)
- RDBMS



Bases de datos relacionales

Organización de los datos

- Una o más tablas (relaciones) compuestas por filas y columnas
- Cada tabla/relación representa un tipo de entidad
- Las filas usualmente son llamadas registros o tuplas y las columnas atributos
- Las filas representan instancias del tipo de entidad y las columnas representan los valores de los atributos de dicha instancia
- Cada tabla dispone una clave que permite identificar univocamente cada registro



Bases de datos relacionales

Ejemplo

<i>id-cliente</i>	<i>nombre-cliente</i>	<i>calle-cliente</i>	<i>ciudad-cliente</i>
19.283.746	González	Arenal	La Granja
01.928.374	Gómez	Carretas	Cerceda
67.789.901	López	Mayor	Peguerinos
18.273.609	Abril	Preciados	Valsaín
32.112.312	Santos	Mayor	Peguerinos
33.666.999	Rupérez	Ramblas	León
01.928.374	Gómez	Carretas	Cerceda

(a) La tabla *cliente*

<i>número-cuenta</i>	<i>saldo</i>
C-101	500
C-215	700
C-102	400
C-305	350
C-201	900
C-217	750
C-222	700

(b) La tabla *cuenta*

<i>id-cliente</i>	<i>número-cuenta</i>
19.283.746	C-101
19.283.746	C-201
01.928.374	C-215
67.789.901	C-102
18.273.609	C-305
32.112.312	C-217
33.666.999	C-222
01.928.374	C-201

(b) La tabla *impositor*



Bases de datos no relacionales

- Bases de datos que **no** se basan el modelo relacional para organizar la información
- Modelo de persistencia poliglota
- Utilizadas en aplicaciones de BigData y web de tiempo real
- Existen desde 1970, aunque actualmente se denominan como **NoSQL**. Este concepto se popularizó recientemente (Web 2.0)



Bases de datos no relacionales

Motivaciones:

- Simplicidad del diseño de datos
- Escalabilidad horizontal en cluster de máquinas
- Control preciso sobre la disponibilidad de los datos
- Limitan el problema de adaptación objeto-relación del modelo relacional



NoSQL

- El término *NoSQL* fue introducido por Carlo Strozzi en 1998 para nombrar su base de datos **relacional** de código abierto ([link](#)),
- Johan Oskarsson re-introdujo el término *NoSQL* a principios de 2009 cuando organizó un evento sobre las bases de datos no relacionales distribuidas de código abierto.



¿Por qué NoSQL?

Cinco tendencias que crean desafíos tecnológicos que *NoSQL* permite abordar:

1. Sistemas y clientes online
2. Internet permite conectividad con todo
3. Aplicaciones BigData cada vez más grandes
4. Aplicaciones en la nube
5. Aplicaciones móviles



¿Por qué NoSQL?

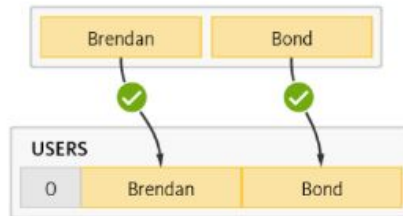
Desarrollo ágil

El modelo relacional no permite adicionar nuevos atributos a demanda de forma simple

Iteration 1 — First, Last

Schema Utilized

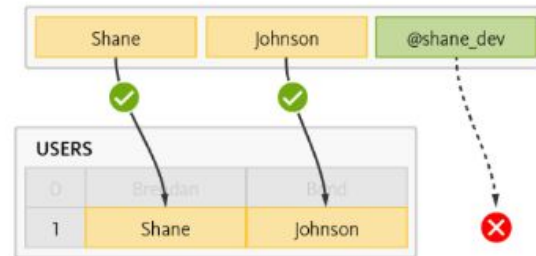
USERS		
ID	First	Last



Iteration 2 — First, Last, *Twitter*

Schema Utilized

USERS		
ID	First	Last

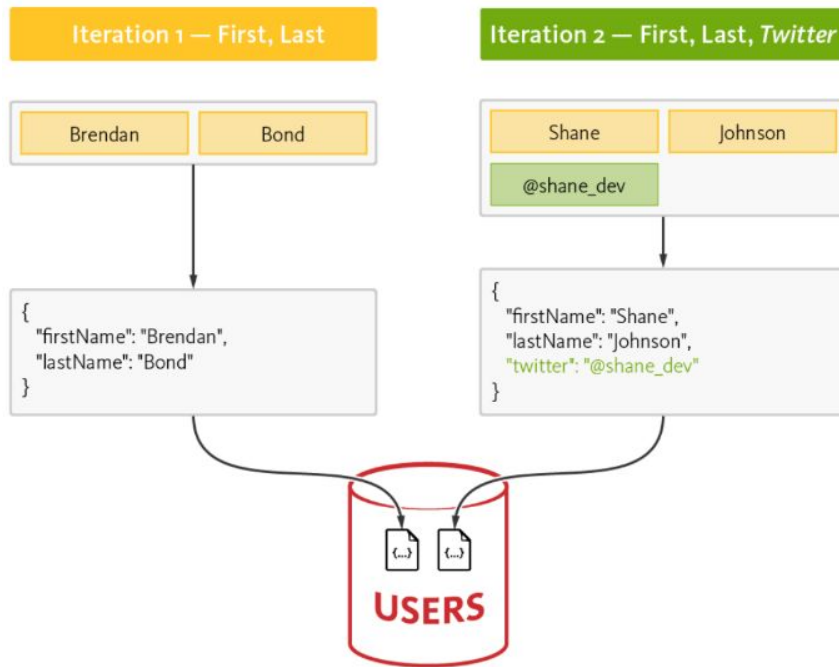




¿Por qué NoSQL?

Desarrollo ágil

El modelo no relacional permite que nuevos atributos sean añadidos a demanda (JSON)

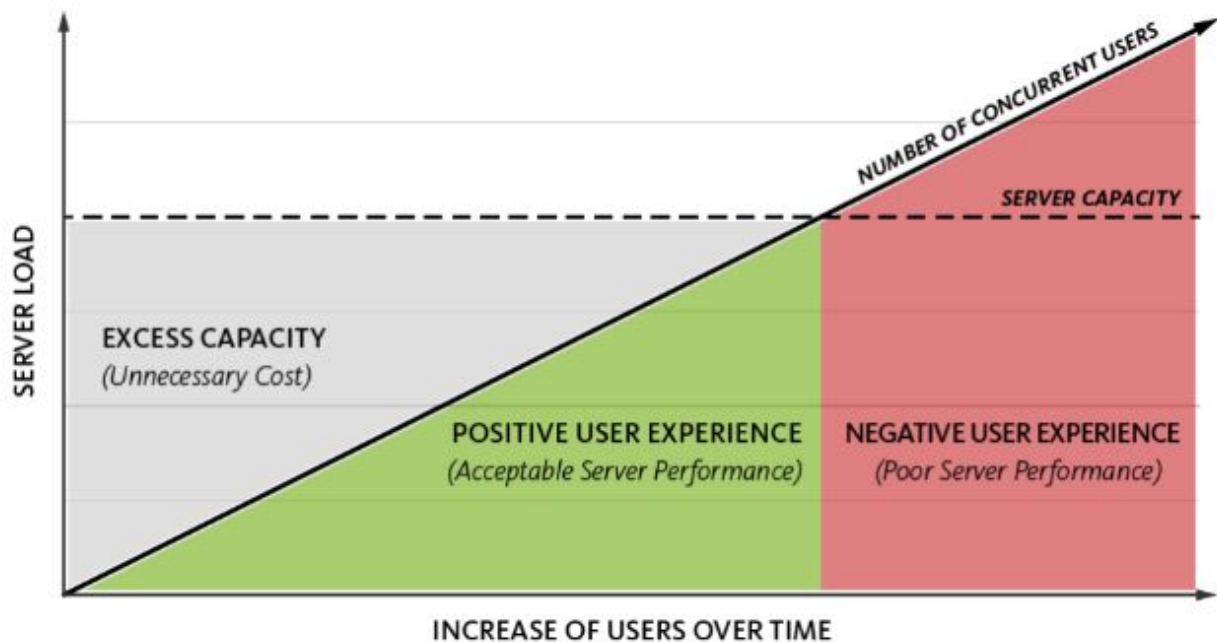




¿Por qué NoSQL?

Elasticidad para el rendimiento a escala

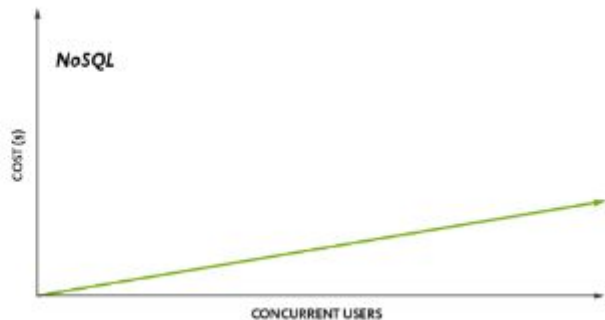
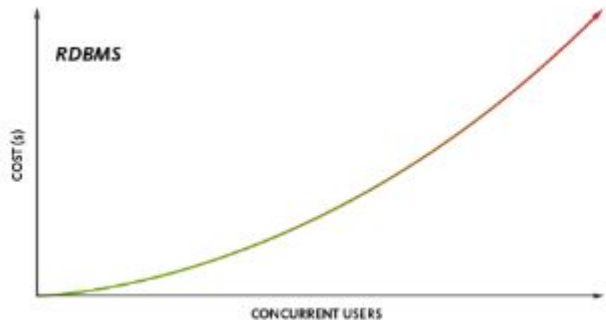
RDBMS - El servidor es demasiado grande o demasiado pequeño, llevando a costos innecesarios o performance pobre





¿Por qué NoSQL?

Costo: añadir capacidad de servicio a demanda así los recursos de hardware son compatibles con la carga de la aplicación





Ranking actual

Rank			DBMS	Database Model	Score		
Jun 2019	May 2019	Jun 2018			Jun 2019	May 2019	Jun 2018
1.	1.	1.	Oracle +	Relational, Multi-model	1299.21	+13.67	-12.04
2.	2.	2.	MySQL +	Relational, Multi-model	1223.63	+4.67	-10.06
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model	1087.76	+15.57	+0.03
4.	4.	4.	PostgreSQL +	Relational, Multi-model	476.62	-2.27	+65.95
5.	5.	5.	MongoDB +	Document	403.90	-4.17	+60.12
6.	6.	6.	IBM Db2 +	Relational, Multi-model	172.20	-2.24	-13.44
7.	7.	↑ 8.	Elasticsearch +	Search engine, Multi-model	148.82	+0.20	+17.78
8.	8.	↓ 7.	Redis +	Key-value, Multi-model	146.13	-2.28	+9.83
9.	9.	9.	Microsoft Access	Relational	141.01	-2.77	+10.02
10.	10.	10.	Cassandra +	Wide column	125.18	-0.54	+5.97

Fuente: <https://db-engines.com/en/ranking>

Ranking actual

356 systems in ranking, June 2020

Rank			DBMS	Database Model	Score		
Jun 2020	May 2020	Jun 2019			Jun 2020	May 2020	Jun 2019
1.	1.	1.	Oracle +	Relational, Multi-model ⓘ	1343.59	-1.85	+44.37
2.	2.	2.	MySQL +	Relational, Multi-model ⓘ	1277.89	-4.75	+54.26
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model ⓘ	1067.31	-10.99	-20.45
4.	4.	4.	PostgreSQL +	Relational, Multi-model ⓘ	522.99	+8.19	+46.36
5.	5.	5.	MongoDB +	Document, Multi-model ⓘ	437.08	-1.92	+33.17
6.	6.	6.	IBM Db2 +	Relational, Multi-model ⓘ	161.81	-0.83	-10.39
7.	7.	7.	Elasticsearch +	Search engine, Multi-model ⓘ	149.69	+0.56	+0.86
8.	8.	8.	Redis +	Key-value, Multi-model ⓘ	145.64	+2.17	-0.48
9.	9.	↑ 11.	SQLite +	Relational	124.82	+1.78	-0.07
10.	↑ 11.	10.	Cassandra +	Wide column	119.01	-0.15	-6.17
11.	↓ 10.	↓ 9.	Microsoft Access	Relational	117.18	-2.72	-23.83
12.	12.	12.	MariaDB +	Relational, Multi-model ⓘ	89.79	-0.30	+4.59

Fuente: <https://db-engines.com/en/ranking>



Ranking actual

13.	13.	13.	Splunk	Search engine	88.08	+0.33	+3.46
14.	14.	14.	Hive	Relational	78.65	-2.89	-0.40
15.	15.	15.	Teradata +	Relational, Multi-model i	73.28	-0.60	-3.36
16.	16.	↑ 20.	Amazon DynamoDB +	Multi-model i	64.87	+0.15	+9.61
17.	17.	↑ 21.	SAP Adaptive Server	Relational	53.09	-0.90	-2.03
18.	18.	↓ 16.	Solr	Search engine	51.26	-1.32	-9.22
19.	↑ 20.	19.	SAP HANA +	Relational, Multi-model i	50.82	+0.29	-5.56
20.	↓ 19.	↓ 18.	FileMaker	Relational	50.16	-0.80	-7.64
21.	↑ 22.	↓ 17.	HBase	Wide column	48.73	-0.99	-9.30
22.	↓ 21.	22.	Neo4j +	Graph	48.27	-1.49	-1.28
23.	23.	↑ 24.	Microsoft Azure SQL Database	Relational, Multi-model i	47.78	+5.03	+18.77
24.	24.	↑ 25.	Microsoft Azure Cosmos DB +	Multi-model i	30.80	+0.13	+2.56

Fuente: <https://db-engines.com/en/ranking>



Instalación de herramientas

- Server PostgreSQL (instalar versión 12):
<https://www.postgresql.org/download/>
- Cliente pgAdmin4:
<https://www.pgadmin.org/download/>



Bibliografía

- Fundamentos De Bases De Datos - Silberschatz
- Introducción a los sistemas de bases de datos - J.C. Date
- NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence - Addison-Wesley
- <https://en.wikipedia.org/wiki/Database>
- https://en.wikipedia.org/wiki/Information_architecture
- <https://www.iainstitute.org/file/whatisiapdf>
- <https://peterjamesthomas.com/wp-content/uploads/2018/02/a-brief-history-of-databases.pdf>
- <http://whatisdbms.com/wp-content/uploads/2017/03/Types-of-DBMS.pdf>
- <https://es.wikipedia.org/wiki/NoSQL>
- <https://www.couchbase.com/resources/why-nosql>
- <http://www.leavcom.com/pdf/NoSQL.pdf>
- <https://db-engines.com/en/ranking>
- https://en.wikipedia.org/wiki/Object-relational_impedance_mismatch