



ASIGNATURA:	BASE DE DATOS NO SQL	CÓDIGO:	
DEPARTAMENTO:	INGENIERÍA EN SISTEMAS DE INFORMACIÓN	CLASE:	Cuatrimestral
ÁREA:	INGENIERÍA EN SISTEMAS DE INFORMACIÓN	HORAS SEM.:	6 hs.
BLOQUE:	ELECTIVAS	HORAS / AÑO:	Reloj 72hs./ Catedra 96hs

Fundamentación:

En los últimos 30 años el mundo de la informática ha experimentado cambios profundos. Han surgido nuevas arquitecturas de aplicaciones, paradigmas de programación y herramientas de desarrollo de software. Sin embargo, algo ha permanecido constante: el uso de Bases de Datos Relacionales (RDBMS) como soporte básico para el almacenamiento y procesamiento de los datos.

En los últimos 10 años se ha verificado una nueva revolución que, si bien no ha desafiado la preponderancia de las bases de datos relacionales, ha demostrado la necesidad de la utilización de nuevos paradigmas para el almacenamiento de datos. La necesidad de procesar grandes volúmenes de información a partir del uso cotidiano de las redes sociales y dispositivos móviles, requirió nuevas soluciones tecnológicas que se encuadran dentro de lo que podemos denominar bajo la categoría de Bases de Datos NoSQL, nombre genérico en el cual podemos incluir las bases de datos key-value, las documentales, las basadas en columnas y las basadas en grafos.

La adopción de estas tecnologías por las empresas está creciendo, en la medida en que también se incrementa el volumen de datos no estructurados que estas empresas almacenan y gestionan. Como consecuencia de ello, se está generando un mercado laboral para especialistas en estas áreas, que incluyen desde Administradores de Bases de Datos NoSQL hasta Data Scientists.

La fundamentación de la inclusión de esta asignatura en la carrera de Ingeniería en Sistemas bajo la categoría de selectiva, se basa en la de otorgar a aquellos estudiantes que lo deseen, la posibilidad de profundizar en estos conceptos y tener mayores herramientas para insertarse laboralmente en un mercado que consideramos va a ser demandante de profesionales en los próximos años.

Objetivos:

- Reconocer la evolución de la tecnología en los últimos años, las tendencias en gestión de datos (big data) y el contexto en el que surgen las soluciones de bases de datos no relacionales.



- Aplicar diferentes soluciones de bases de datos dentro de una misma aplicación para soportar diferentes necesidades funcionales (“Persistencia Políglota”).
- Utilizar la “Persistencia Políglota” en el diseño de aplicaciones.
- Distinguir modelos de distribución de datos en clusters (Single Server, Sharding, Master-Slave, Peer to Peer) y cuáles son aplicables para cada situación.
- Identificar esquemas de uso y gestión de la consistencia de los datos para el uso en diferentes operaciones.
- Reconocer el esquema Map-Reduce para resolver problemas de gestión de datos en entornos de alto volumen y distribuidos.
- Comprender la arquitectura de las bases de datos KeyValue, Grafos, ColumnFamily y Documentales, sus características y casos para saber cuándo utilizarlo.
- Examinar bases de datos a partir de los ejercicios prácticos basados en Redis, MongoDB versión 3.x con motores de procesamiento mmapv1 y wiredTiger, Cassandra y Neo4J

Programa analítico:

UNIDAD 1: [Surgimiento y conceptualización de Bases de Datos NoSQL]

El valor de las bases de datos RDBMS. Cambios en la evolución tecnológica de los últimos 15 años a partir de los cuales las RDBMS dejaron de ser óptimas. Surgimiento de NoSQL. Necesidades que cubren las bases de datos NoSQL. Definición de base de datos NoSQL. Tipos de bases de datos NoSQL: key-value DB, documents DB, column-family DB, Necesidad de soluciones que implementen “Persistencia Políglota”.

UNIDAD 2: [Modelos de datos basados en Agregados y basados en Relaciones]

Definición de Agregados. Ejemplos. Uso de agregados en bases de datos basadas en key-value, y document. Uso de agregados en bases de datos column-family. Modelado de datos en bases de datos basadas en agregados. Vistas materializadas. Modelos de datos basados en relaciones: bases de datos basadas en grafos. Bases de datos sin esquemas

UNIDAD 3: [Modelos de distribución de datos]

Introducción. Single Server. Almacenamiento de datos en diferentes máquinas: Sharding. Replicación: Master-slave. Replicación: Peer to Peer. Combinación de Sharding y Replicación.

UNIDAD 4: [Consistencia de datos en bases de datos NoSQL]

Consistencia en Update. Consistencia en Read. Consistencia Débil. El teorema CAP. Quorums..

UNIDAD 5: [Map-Reduce]

Introducción. Operación Map. Operación Reduce. Particionamiento y combinación de datos usando Map-Reduce. Ejemplos.

UNIDAD 6: [Bases de Datos Key-Value]



Definición conceptual. Características de almacenamiento: consistencia, transacciones, consultas, estructuras de los datos, escalamiento. Casos en donde aplica usar este tipo de Bases de Datos. Casos en los que no aplica usar este tipo de Bases de Datos. Ejercicios prácticos basados en Redis.

UNIDAD 7: [Bases de Datos basadas en Documentos]

Definición conceptual. Características de almacenamiento: consistencia, transacciones, disponibilidad, consultas, estructuras de los datos, escalamiento. Casos en donde aplica usar este tipo de Bases de Datos. Casos en los que no aplica usar este tipo de Bases de Datos. Ejercicios prácticos basados en mongoDB.

UNIDAD 8: [Bases de Datos basadas en Column-Family]

Definición conceptual. Características de almacenamiento: consistencia, transacciones, disponibilidad, consultas, estructuras de los datos, escalamiento. Casos en donde aplica usar este tipo de Bases de Datos. Casos en los que no aplica usar este tipo de Bases de Datos. Ejercicios prácticos basados en Cassandra.

UNIDAD 9: [Bases de Datos basadas en Grafos]

Definición conceptual. Características de almacenamiento: consistencia, transacciones, disponibilidad, consultas, estructuras de los datos, escalamiento. Casos en donde aplica usar este tipo de Bases de Datos. Casos en los que no aplica usar este tipo de Bases de Datos. Ejercicios prácticos basados en Neo4j.

UNIDAD 10: [Persistencia Políglota]

Uso de diferentes arquitecturas de bases de datos en una misma solución. Ventajas de utilización de servicios en lugar de acceso directo a datos. Elección de la tecnología adecuada.

UNIDAD 11: [Herramientas para almacenamiento, procesamiento y análisis de datos Distribuidos.]

Ecosistema Apache Hadoop. HDFS Hadoop Distributed File System. YARN como gestor de recursos. Motores de Procesamiento Distribuido: MapReduce, Tez, Spark y Spark Streaming, Sqoop como herramienta de Extracción y Carga contra un Motor de BD Externo. Hive, herramienta para realizar consultas SQL sobre una metadata creada sobre archivos de HDFS.

Distribución de carga horaria entre actividades teóricas y prácticas:

Tipo de actividad	Carga horaria total en hs. reloj	Carga horaria total en hs. cátedra
Teórica	49.5	66
Formación Práctica	22.5	30
Formación Experimental	0	0
Resolución de Problemas	0	0



Proyectos de Diseño	0	0
Práctica Supervisada	0	0
Total	72	96

Articulación Horizontal y vertical con otras materias

La asignatura Bases de Datos NoSQL se articula en forma vertical con una asignatura que la antecede en el plan de estudio, específicamente Gestión de datos.

Cada estudiante deberá tener cursada y regularizada cada una de estas asignaturas al momento de comenzar la cursada. El conocimiento de Bases de datos NoSQL, redes y sistemas operativos, son conocimientos previos que se requieren para poder abordar e incorporar eficientemente el contenido impartido en la asignatura Bases de Datos NoSQL.

Cronograma estimado de clases:

Unidad temática	Duración en horas cátedra
1	10
2	10
3	10
4	10
5	10
6	10
7	10
8	10
9	10
10	3
11	3

Bibliografía:

BIBLIOGRAFÍA OBLIGATORIA

- No SQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Autores: Pramod Sadalage, Martin Fowler. Addison Wesley (2012)
- MongoDB: The Definitive Guide: Powerful and Scalable Data Storage. Autores: Shannon Bradshaw , Eoin Brazil O'Reilly (2019).
- Graph Database Modeling with neo4. Autores: Ajit Singh y Anant Kumar O'Reilly (2020).
- Redis 4.x Cookbook: Over 80 hand-picked recipes for effective Redis development and administration. Autores: Pengcheng Huang y Zuofei Wang. Packt (2018).
- Cassandra: The Definitive Guide: Distributed Data at Web Scale. Autores: Jeff Carpenter y Eben Hewitt. O'Reilly (2020).



BIBLIOGRAFÍA COMPLEMENTARIA

- Seven databases in seven weeks: a guide to Modern Databases and the NoSQL Movement.
Autores: Eric Redmond, Jim Wilson.
Pragmatic Programmers (2012).
- Planning for Big data: A CIO's Handbook to the Changing Data Landscape
Autores: O'Reilly Radar Team.
O'Reilly (2012).
- Graph Databases: New Opportunities for Connected Data 2nd Edición
Autores: Ian Robinson, Jim Webber, Emil Eifrem
O'Reilly (2015)
- Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale
Autores: Tom White
O'Reilly (2015)

Correlativas:

PARA CURSAR:

Cursadas: Redes de Información
Administración de Recursos
Simulación
Ingeniería en Software

Aprobadas: Diseño de Sistemas
Sistemas Operativos
Gestión de Datos

PARA RENDIR:

Aprobadas: Redes de Información
Administración de Recursos
Simulación
Ingeniería en Software