



Universidad Tecnológica Nacional
Facultad Regional Buenos Aires

PROGRAMA ANALÍTICO DE ASIGNATURA

DEPARTAMENTO: Ingeniería en Sistemas de Información

CARRERA: Ingeniería en Sistemas de Información

NOMBRE DE LA ACTIVIDAD CURRICULAR: Bases de Datos noSQL

Año Académico: 2023

Área: Desarrollo de Software

Bloque: Electivas

Nivel: 5º

Tipo: Electiva

Modalidad: Cuatrimestral

Cargas horarias totales:

<i>Horas reloj</i>	<i>Horas cátedra</i>	<i>Horas cátedra semanales</i>
72	96	6

COMPOSICIÓN DEL EQUIPO DE CÁTEDRA:

Profesor Adjunto: Juan Zaffaroni

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



Universidad Tecnológica Nacional
Facultad Regional Buenos Aires

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 Key/Value, 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.



*Universidad Tecnológica Nacional
Facultad Regional Buenos Aires*

CONTENIDOS

Contenidos analíticos

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 implementan “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



Universidad Tecnológica Nacional
Facultad Regional Buenos Aires

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.



Universidad Tecnológica Nacional
Facultad Regional Buenos Aires

BIBLIOGRAFÍA OBLIGATORIA

- Ajit Singh y Anant, Kumar (2020). Graph Database Modeling with neo4j. Ed O'Reilly
- Carpenter, J. y Hewitt, E. (2020). Cassandra: The Definitive Guide: Distributed Data at Web Scale. Ed O'Reilly
- Pengcheng, Huang y Zuofei, Wang (2018). Redis 4.x Cookbook: Over 80 hand-picked recipes for effective Redis development and administration. Ed Packt
- Pramod Sadalage, Martin Fowler (2012). No SQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Ed Addison Wesley.
- Robinson, I. Webber, J. Eifrem, E. (2015). Graph Databases: New Opportunities for Connected Data 2nd Edición. Ed O'Reilly.
- Shannon Bradshaw, Eoin Brazil (2019). MongoDB: The Definitive Guide: Powerful and Scalable Data Storage. Ed O'Reilly.
- White, T. (2015). Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale. Ed O'Reilly.

BIBLIOGRAFÍA COMPLEMENTARIA

- Eric Redmond, Jim Wilson (2012). Seven databases in seven weeks: a guide to Modern Databases and the NoSQL Movement. Ed Pragmatic Programmers.
- O'Reilly Radar Team (2012). Planning for Big data: A CIO's Handbook to the Changing Data Landscape. Ed O'Reilly.

CORRELATIVAS

Para cursar y rendir

- Cursadas:
 - Redes de Datos
 - Administración de Sistemas de Información
 - Simulación
 - Ingeniería y Calidad de Software
- Aprobadas:
 - Diseño de Sistemas de Información
 - Sistemas Operativos
 - Bases de Datos