



ASIGNATURA:	IMPLEMENTACION DE BASES DE DATOS NoSQL
DEPARTAMENTO:	ING. EN SIST. DE INFORMACION
AREA:	ELECTIVA
BLOQUE	TECNOLOGÍAS APLICADAS

MODALIDAD:	Cuatrimestral
HORAS SEM.:	6 horas
HORAS/AÑO:	96 horas
HORAS RELOJ	72
NIVEL:	5°
AÑO DE DICTADO:	2018

Objetivos

- Entender la evolución que ha tenido la tecnología en los últimos años y el contexto en el cual surgen estas soluciones de bases de datos no relacionales.
- Comprender el concepto de “Persistencia Políglota”, es decir la aplicación de diferentes soluciones de bases de datos dentro de una misma aplicación para soportar diferentes necesidades funcionales.
- Internalizar la posibilidad de utilizar “Persistencia Políglota” en el diseño de sus aplicaciones.
- Comprender los diferentes modelos de distribución de datos en clusters (Single Server, Sharding, Master-Slave, Peer to Peer) y cuáles de esos modelos son aplicables para cada situación.
- Entender particularidades de la gestión de la consistencia de los datos para las diferentes operaciones sepa qué esquema usar dependiendo el caso.
- Entender qué es Map-Reduce y como este esquema ayuda a resolver problemas de gestión de datos en entornos de alto volumen y distribuidos.
- Comprender conceptualmente la arquitectura de las bases de datos Key-Value, así como sus características, casos en los que hace sentido utilizarlas y casos en los que no. Adicionalmente que pueda tener un primer contacto con bases de datos de este estilo a partir de los ejercicios prácticos basados en Redis.
- Comprender conceptualmente la arquitectura de las bases de datos Documentales, así como sus características, casos en los que hace sentido utilizarlas y casos en los que no. Adicionalmente que pueda tener un primer contacto con bases de datos de este estilo a partir de los ejercicios prácticos basados en MongoDB versión 3.x con motores de procesamiento mmapv1 y wiredTiger.



- Comprender conceptualmente la arquitectura de las bases de datos Column-Family, así como sus características, casos en los que hace sentido utilizarlas y casos en los que no. Adicionalmente que pueda tener un primer contacto con bases de datos de este estilo a partir de los ejercicios prácticos basados en Cassandra.
- Comprender conceptualmente la arquitectura de las bases de datos basadas en Grafos, así como sus características, casos en los que hace sentido utilizarlas y casos en los que no. Adicionalmente que pueda tener un primer contacto con bases de datos de este estilo a partir de los ejercicios prácticos basados en Neo4J.
- Conocer otros Productos para facilitar de almacenamiento, procesamiento y análisis de datos Distribuidos, más allá de NoSQL. En particular el Apache Hadoop.
- Adquirir visión sobre las tendencias en gestión de datos, en particular lo que se conoce como Big Data.

Contenidos Mínimos (Programa Sintético)

- Conceptos asociados a los distintos tipos de bases de datos NoSQL: key-value DB, documents DB, column-family DB, Graph DB.
- Modelos de distribución de datos, consistencia de datos y map reduce.
- Implementación de bases de datos NoSQL, key-value DB (Redis), documents DB (mongoDB), column-family DB (Cassandra), Graph DB (Neo4J).
- Persistencia políglota

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 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.

Bibliografía Obligatoria

. **Bibliografía de apoyo principal:**

No SQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence.

Autores: Pramod Sadalage, Martin Fowler.
Addison Wesley (2012).

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).

Correlativas

Para Cursar:

Cursadas:

- Administración de Recursos
- Redes de Información
- Simulación
- Ingeniería de Software



Aprobadas:

- Diseño de Sistemas
- Sistemas Operativos
- Gestión de Datos

Para rendir:

Aprobadas:

- Administración de Recursos
- Redes de Información
- Simulación
- Ingeniería de Software