



Datos Generales

1. Nombre de la Asignatura Almacenamiento de Grandes Bases de Datos	2. Nivel de formación Maestría	3. Clave de la Asignatura	
4. Prerrequisitos Procesamiento de Grandes Bases de Datos	5. Área de Formación Formación Básica Particular Obligatoria	6. Departamento Sistemas de Información	
7. Modalidad Presencial	8. Tipo de Asignatura Curso - Taller		
9. Carga Horaria 48 horas			
Teoría: 30	Práctica: 18	Total: 48	Créditos: 7
10. Trayectoria de la asignatura			

Contenido del Programa

11. Presentación
La tecnología ha revolucionado el mundo actual y su ritmo de crecimiento parece no tener límite. Una de las herramientas que cambió drásticamente los sistemas informáticos son las bases de datos y el <i>Big data</i>. En la actualidad estos conceptos se han convertido en herramientas con la cual las empresas hacen negocios, la industria mantiene sus activos y administración de inventarios y las instituciones académicas y científicas acceden a información relevante para la innovación y el avance de la tecnología. Sin embargo, las grandes bases de datos requieren un adecuado diseño inicial y un monitoreo constante para poder obtener el mejor desempeño del motor de bases de datos reflejado en tiempos de respuesta muy cortos. Este curso parte del núcleo académico básico del posgrado de ciencia de datos y pretende proporcionar la bases teóricas y técnicas que permitan procesar información obtenida de grandes bases de datos empleando diferentes estrategias mediante línea de comando empleando lenguajes <i>No SQL</i>.
12.- Objetivos del programa
Adquirir los conocimientos básicos necesarios de <i>software</i> y <i>hardware</i> para el almacenamiento de grandes bases de datos en el contexto de economía-finanzas, negocios-mercadotecnia y políticas públicas.
Objetivo General
Desarrollar las habilidades y destrezas básicas para el diseño, la implementación, la gestión, la

manipulación, la consulta y la administración de grandes bases de datos.

13.-Contenido

Contenido temático

1. Introducción al procesamiento de grandes bases de datos
2. Grandes bases de datos y el paradigma *NoSQL*
3. Gestión eficiente de una gran base de datos
4. Generación de conocimiento para la acción en grandes bases de datos
5. Herramientas para gestionar macrodatos
6. Competencias para la gestión de datos masivos
7. Introducción a las bases de datos *NoSQL*
8. Conceptos básicos de gestión de bases de datos mediante lenguaje NoSQL
9. Operaciones *CRUD*
10. *Queryng* en *MongoDB*
11. *Framework* de agregación
12. Indexación y replicación
13. *Sharding*
14. Optimización
15. *GridFS*
16. Operaciones de administración

Unidad 1 Introducción al procesamiento de Grandes Bases de Datos

Objetivos particulares de la unidad:

Identificar las principales diferencias entre bases de datos convencionales y grandes bases de datos.

Definir que es una gran base de base de datos y sus necesidades técnicas particulares.

Conocer los principales componentes de una gran base de datos

Desarrollo:

Lectura del libro *BIG DATA: UNA BREVE INTRODUCCIÓN*. Autores: Holmes, Dawn E. ISBN: 9788494886058, 9788494886041 Editorial Antoni Bosch editor. Año de Edición: 2018

Unidad 2 Grandes bases de datos y el paradigma *NoSQL*

Objetivos particulares de la unidad:

Identificar las principales diferencias entre bases de datos convencionales y grandes bases de datos.

Definir que es una gran base de base de datos y sus necesidades técnicas particulares.

Conocer los principales componentes de una gran base de datos

Desarrollo:

Lectura de capítulo del libro *BIG DATA: GESTIÓN Y EXPLOTACIÓN DE GRANDES VOLÚMENES DE DATOS*. Autores: García Alsina, Montserrat ISBN: 9788491167129. Editorial UOC, Año de Edición: 2017.

Unidad 3 Gestión eficiente de una gran base de datos

Objetivos particulares de la unidad:

Aprender acerca de modelado de datos por que los modelos de datos son importantes.

Conocer los elementos básicos de modelado de datos.

Identificar que son reglas de negocios y cómo influyen en el diseño de las bases de datos.

Clasificar los modelos de datos por nivel de abstracción.

Desarrollo:

Lectura de capítulo del libro *BIG DATA: GESTIÓN Y EXPLOTACIÓN DE GRANDES VOLÚMENES DE DATOS*. Autores: García Alsina, Montserrat ISBN: 9788491167129. Editorial UOC, Año de Edición: 2017.

Unidad 4 Generación de conocimiento para la acción en grandes bases de datos

Objetivos particulares de la unidad:

Identificar la importancia de la generación de conocimiento para la acción en grandes bases de datos.

Conocer algunos modelos de gestión organizativa en grandes bases de datos.

Reutilizar información para difusión retroalimentación.

Desarrollo:

Lectura de capítulo del libro *BIG DATA: GESTIÓN Y EXPLOTACIÓN DE GRANDES VOLÚMENES DE DATOS*. Autores: García Alsina, Montserrat ISBN: 9788491167129. Editorial UOC, Año de Edición: 2017.

Unidad 5 Herramientas para gestionar macrodatos

Objetivo particular de la unidad:

Definir las principales herramientas para gestionar macrodatos.

Conocer los principales conceptos para gestión de macrodatos.

Identificar las plataformas, tecnologías y aplicaciones más empleadas.

Desarrollo:

Lectura de capítulo del libro *BIG DATA: GESTIÓN Y EXPLOTACIÓN DE GRANDES VOLÚMENES DE DATOS*. Autores: García Alsina, Montserrat ISBN: 9788491167129. Editorial UOC, Año de Edición: 2017.

Unidad 6 Competencias para la gestión de datos masivos

Objetivos particulares de la unidad:

Identificar las competencias generales y particulares necesarias para la gestión de datos masivos.

Conocer las habilidades y conocimientos, así como los perfiles del personal técnico gestor de grandes bases de datos.

Desarrollo:

Lectura de capítulo del libro *BIG DATA: GESTIÓN Y EXPLOTACIÓN DE GRANDES VOLÚMENES DE DATOS*. Autores: García Alsina, Montserrat ISBN: 9788491167129. Editorial UOC, Año de Edición: 2017.

Unidad 7 Introducción a las bases de datos *NoSQL*

Objetivos particulares de la unidad:

Conocer las generalidades de las bases de datos *NoSQL*

Reconocer la importancia del papel de las bases de datos en las aplicaciones *software*

Identificar las limitaciones de las bases de datos relacionales.

Conocer los modelos de bases de datos *NoSQL* orientados hacia agregados

Aplicar el modelo de distribución de datos de las bases de datos *NoSQL*

Definir el teorema *CAP*

Desarrollo:

Lectura de capítulos de los libros *INTRODUCCIÓN A LAS BASES DE DATOS NoSQL USANDO MongoDB*. Autores: Sarasa, Antonio. ISBN: 9788491162506, 9788491162667. Editorial UOC, Año de Edición: 2016 e *INTRODUCCIÓN A LAS BASES DE DATOS NSQL: clave-valor usando Redis*, Autores: Sarasa, Antonio ISBN: 9788491804857 Editorial UOC Año de Edición: 2019.

Unidad 8 Conceptos básicos de lenguaje *NoSQL*

Objetivos particulares de la unidad:

Aprender comandos y funciones básicos de *NoSQL*.

Usar *NoSQL* para administración de datos.

Emplear *NoSQL* para la manipulación de datos.

Consultar información útil en la base de datos mediante comandos con *NoSQL*.

Desarrollo:

Lectura de capítulo del libro *INTRODUCCIÓN A LAS BASES DE DATOS NoSQL USANDO MongoDB*. Autores: Sarasa, Antonio. ISBN: 9788491162506, 9788491162667. Editorial UOC, Año de Edición: 2016.

Unidad 9 Operaciones *CRUD*

Objetivos particulares de la unidad:

Conocer algunas funciones de los operadores *CRUD* para insertar, borrar y actualizar documentos

Modificar *arrays*

Emplear el comando *Upsert*

Actualizar de múltiples documentos y evaluar la eficiencia de las modificaciones

Desarrollo:

Lectura de capítulo del libro *INTRODUCCIÓN A LAS BASES DE DATOS NoSQL USANDO MongoDB*. Autores: Sarasa, Antonio. ISBN: 9788491162506, 9788491162667. Editorial UOC, Año de Edición: 2016.

Unidad 10 *Queryng* en *mongoDB*

Objetivos particulares de la unidad:

Conocer algunos de los comandos útiles para cuestionar y obtener información de los datos.

Introducir al usuario al cuestionamiento de la base de datos en un ambiente *MongoDB*

Emplear el método *find()*, operadores condicionales: *\$lt*, *\$lte*, *\$gt*, *\$gte*, los operadores *\$not*, *\$ and* y *\$or*

Utilizar Expresiones regulares

Realizar consultas sobre *arrays*, sobre documentos embebidos

Consultar empleando el comando *\$where*

Configurar de las consultas y otras características

Desarrollo:

Lectura de capítulo del libro *INTRODUCCIÓN A LAS BASES DE DATOS NoSQL USANDO MongoDB*.

Autores: Sarasa, Antonio. ISBN: 9788491162506, 9788491162667. Editorial UOC, Año de Edición: 2016.

Unidad 11 La *framework* de agregación

Objetivos particulares de la unidad:

Comprender la estrategia de trabajo para obtener resultados a partir de combinaciones de datos.

Diseñar técnicas de trabajo en tuberías.

Conocer cómo se usan los métodos de agregación y operadores de propósito único.

Desarrollo:

Lectura de capítulo del libro *INTRODUCCIÓN A LAS BASES DE DATOS NoSQL USANDO MongoDB*.

Autores: Sarasa, Antonio. ISBN: 9788491162506, 9788491162667. Editorial UOC, Año de Edición: 2016.

Unidad 12 Indexación y replicación

Objetivos particulares de la unidad:

Conocer los métodos para diseño y creación de índices

Crear un índice simple y un índice compuesto

Crear Índices de subdocumentos

Especificar opciones sobre los índices

Eliminar índices de la base de datos

Reeditar índices de una colección

Seleccionar de índices

Emplear el comando *hint* ()

Optimizar del almacenamiento de pequeños objetos

Conocer el empleo de índices geoespaciales

Replicar bases de daos completas o partes de estas

Conocer los procesos de replicación y la configuración de nodos secundarios

Identificar las estrategias de recuperación ante un fallo del primario

Desarrollo:

Lectura de capítulo del libro *INTRODUCCIÓN A LAS BASES DE DATOS NoSQL USANDO MongoDB*.

Autores: Sarasa, Antonio. ISBN: 9788491162506, 9788491162667. Editorial UOC, Año de Edición: 2016.

Unidad 13 Sharding

Objetivos particulares de la unidad:

Conocer estrategias de partición horizontal de datos

Definir el concepto de *sharding*

Método de *sharding*

Configurar de la distribución de datos

Crear un clúster a partir de métodos de *sharding*

Adicionar uno o varios *shards* a un clúster

Eliminar un *shard* de un clúster

Convertir una replica *set* en un *sharded* clúster

Desarrollo:

Lectura de capítulo del libro *INTRODUCCIÓN A LAS BASES DE DATOS NoSQL USANDO MongoDB*.

Autores: Sarasa, Antonio. ISBN: 9788491162506, 9788491162667. Editorial UOC, Año de Edición: 2016.

Unidad 14 Optimización

Objetivos particulares de la unidad:

Conocer cuáles son los métodos más recomendables para optimizar una base de datos en un entorno *mongoDB*

Emplear *mongoDB Profiler*, *Explain()*,

Conocer los valores *queryPlanner* y *executionStats*

Utilizar *Server Info*

Conocer el optimizador de consultas y la herramienta *Index Filter*

Realizar observaciones

Emplear índices que cubren una consulta y evaluar la selectividad de las consultas

Desarrollo:

Lectura de capítulo del libro *INTRODUCCIÓN A LAS BASES DE DATOS NoSQL USANDO MongoDB*.

Autores: Sarasa, Antonio. ISBN: 9788491162506, 9788491162667. Editorial UOC, Año de Edición: 2016.

Unidad 15 Métodos de fragmentación y partición de archivos

Identificar los métodos de fragmentación y partición de archivos más recomendables para la gestión de grandes bases de datos en ambiente *MongoDB*

Utilizar los comandos *GridFS* y *Mongofile*,

Conocer y desplegar las colecciones e índices *GridFS*

Cargar un documento en *GridFS*

Buscar, recuperar y eliminar archivos

Desarrollo:

Lectura de capítulo del libro *INTRODUCCIÓN A LAS BASES DE DATOS NoSQL USANDO MongoDB*.

Autores: Sarasa, Antonio. ISBN: 9788491162506, 9788491162667. Editorial UOC, Año de Edición: 2016.

Unidad 16 Operaciones de administración

Conocer los métodos de operaciones de administración más recomendables para grandes bases de datos en ambiente *MongoDB*

Introducir al usuario a métodos de operaciones de administración de grandes bases de datos

Realizar *backups* del servidor

Restaurar una copia de seguridad

Exportar datos, crear usuarios y otras operaciones de administración

Desarrollo:

Lectura de capítulo del libro *INTRODUCCIÓN A LAS BASES DE DATOS NoSQL USANDO MongoDB*.

Autores: Sarasa, Antonio. ISBN: 9788491162506, 9788491162667. Editorial UOC, Año de Edición: 2016.

14. Actividades Prácticas

El curso esta diseñado con una visión de aprender haciendo, basado en un enfoque de enseñanza

basada en problemas, tomando en cuenta que el grupo está integrado con alumnos con diferentes habilidades y destrezas derivadas de su formación profesional diversa.

Se estimula el proceso de aprendizaje mediante las lecturas del texto de referencia fuera del aula la resolución de tareas en el aula virtual del curso, complementándose con ejercicios prácticos (manos a la obra) con el lenguaje *NoSQL* en un ambiente empleado *MongoDB* durante sesiones presenciales parte del taller.

15.- Metodología

El curso taller tiene un enfoque práctico para facilitar el proceso de enseñanza aprendizaje, complementando las lecturas y actividades del libro de texto, con ejercicios prácticos durante las sesiones practicas semanales.

16.- Evaluación

- 1) Reportes de actividades y ejercicios semanales en aula virtual 60%
- 2) Tareas 20%
- 3) Actividades presenciales en el aula durante talleres 20%

17.- Bibliografía

1. BIG DATA: UNA BREVE INTRODUCCIÓN Autores: Holmes, Dawn E. ISBN: 9788494886058, 9788494886041 Editorial Antoni Bosch editor Año de Edición: 2018
2. BIG DATA: GESTIÓN Y EXPLOTACIÓN DE GRANDES VOLÚMENES DE DATOS. Autores: García Alsina, Montserrat ISBN: 9788491167129. Editorial UOC, Año de Edición: 2017.
3. INTRODUCCIÓN A LAS BASES DE DATOS *NoSQL* USANDO *MongoDB*. Autores: Sarasa, Antonio. ISBN: 9788491162506, 9788491162667. Editorial UOC, Año de Edición: 2016.
4. INTRODUCCIÓN A LAS BASES DE DATOS *NSQL*: clave-valor usando Redis, Autores: Sarasa, Antonio ISBN: 9788491804857 Editorial UOC Año de Edición: 2019.

Otros materiales

XAMPP <https://www.apachefriends.org/es/index.html>

MYSQL Workbench <https://www.mysql.com/products/workbench/>

mongoDB <https://www.mongodb.com/>

18.- Perfil del profesor

El profesor contará con grado de Maestro o Doctor en alguna de las LGAC del Programa, es especial que cuente con conocimientos sólidos en el manejo de grandes volúmenes de datos.

19.- Nombre de los profesores que imparten la materia

Dr. Raúl Cuauhtémoc Baptista Rosas

20.- Lugar y fecha de su aprobación (incluyendo la última actualización)

Zapopan, Jal. 13 de enero de 2020



21.- Instancias que aprobaron el programa (Junta Académica y/o Coordinación del programa)
Junta Académica y Coordinación de la Maestría en Ciencia de los Datos, y el Departamento de
Sistemas de Información.