



Datos Generales

1. Nombre de la Asignatura	2. Nivel de formación	3. Clave de la Asignatura
Tópicos selectos de grandes bases de datos II	Maestría	
4. Prerrequisitos Ninguno	5. Área de Formación Especializante selectiva	6. Departamento Sistemas de información
7. Modalidad Presencial	8. Tipo de Asignatura: Curso-Taller	
9. Carga Horaria 60 horas		
Teoría: 40 horas	Práctica: 20 horas	Total: 60
		Créditos: 7
10. Trayectoria de la asignatura		

Contenido del Programa

11. Presentación El maestrante en ciencia de datos aprenderá los fundamentos de Big Data y su aplicación en las ciencias económico-administrativas como son: contabilidad, macroeconomía, microeconomía, finanzas corporativas, negocios y gobierno y políticas publicas Con este curso el futuro maestro en ciencia de datos, tendrá una perspectiva amplia de Big Data.
12.- Objetivos del programa Objetivo General Obtener habilidades de Big data, en sus dos ramos más usados: Hadoop y Spark
13.-Contenido Contenido temático Unidad 1 Introducción a Big Data Unidad 2 Bases de datos no relacionales Unidad 3 Hadoop Unidad 4 Spark
Contenido desarrollado Unidad 1 Introducción a Big Data Objetivo particular de la unidad: El alumno conocerá los conceptos básicos de Big data y como se diferencia de small data

Desarrollo

1. Introducción a Big Data

- 1.1. Características de Big Data
- 1.2. V's de Big Data
- 1.3. Impacto de Big Data
- 1.4. Fuentes de Big Data
- 1.5. Plataforma de Big Data
- 1.6. Ecosistemas de Big Data
- 1.7. Ejemplos de Big Data
- 1.8. Aplicaciones de Big Data en finanzas, economía, mercadotecnia y políticas públicas

Unidad 2 Bases de datos no relacionales

Objetivo particular de la unidad: El alumno conocerá los diferentes tipos de bases de datos no relacionales, cuando usar cada una y en que difieren de las bases de datos relacionales

Desarrollo

2. Bases de datos no relacionales

- 2.1.1. En memoria
 - 2.1.1.1. Redis
- 2.1.2. Llave-Valor
 - 2.1.2.1. CouchBase
- 2.1.3. Documentos
- 2.1.4. Grafos

Unidad 3 Hadoop

Objetivo particular de la unidad: El alumno comprenderá los conceptos básicos de Hadoop

- 2.1.5. Introducción a Hadoop
 - 2.1.5.1. Componentes de Hadoop
 - 2.1.5.2. Sistema de archivos distribuidos de Hadoop
 - 2.1.5.3. Clusters de Hadoop
- 2.1.6. MapReduce
- 2.1.7. YARN
- 2.1.8. Cargar datos a Hadoop
 - 2.1.8.1. De una base de datos relacional
 - 2.1.8.1.1. Sqoop
 - 2.1.8.2. Flume
- 2.1.9. Data click
- 2.1.10. Hive
 - 2.1.10.1. Introducción a Hive
 - 2.1.10.2. Hive DDL
 - 2.1.10.3. Hive DML
 - 2.1.10.4. Operadores y funciones de Hive

Unidad 4 Spark

Objetivo particular de la unidad: El alumno conocerá conceptos claves de Spark y como usarlo con Python

Desarrollo

3. Spark

3.1. Introducción a Spark

3.1.1. Spark

3.1.2. RDD

3.1.2.1. Arquitectura

3.1.2.2. Operaciones

3.1.2.3. Caching

3.1.2.4. Niveles de almacenamiento

3.1.3. Colecciones paralelas y datasets externos

3.1.4. Variables compartidas

3.1.5. Pares llave-valor

3.2. PySpark

3.3. SparkML

3.3.1. Tipos de datos

3.3.2. Spark MLlib arboles de decisión y bosques aleatorios

3.3.3. Spark MLlib clustering

3.3.4. GraphX

3.3.4.1. Componentes de GraphX

14. Actividades Prácticas

Ejercicios en línea y ejercicios

15.- Metodología

Lectura previa, estudios de casos, discusión de temas, cursos en línea, participación de los alumnos.

16.- Evaluación

1) Ejercicios en línea	80%
2) Proyecto final	20%
Total	100%

17.- Bibliografía

1.

Otros materiales

- <https://cognitiveclass.ai/learn/spark>
- <https://cognitiveclass.ai/learn/hadoop>
- <https://cognitiveclass.ai/learn/big-data>
- <https://learn.couchbase.com/store>



18.- Perfil del profesor

Profesor con nivel mínimo de maestría, de reconocido prestigio, con amplia experiencia en la docencia, programación y análisis de datos

19.- Nombre de los profesores que imparten la materia

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

21.- Instancias que aprobaron el programa (Junta Académica y/o Coordinación del programa)