

## Programa de Asignatura

### Historia del programa

| Lugar y fecha de elaboración | Participantes   | Observaciones (Cambios y justificaciones)   |
|------------------------------|---|---|
| Cancún, Q. Roo, 23/03/2017   | Dr. Julio César Ramírez Pacheco<br>Dr. David Israel Flores Granados<br>M. en C. Nancy Aguas García<br>M. en C. Anilú Gómez Pantoja<br>M. en C. Francisco Manzano Pinzón | Creación del programa para incorporarse como asignatura de elección libre en Ingeniería en Datos e Inteligencia Organizacional. |

### Relación con otras asignaturas

| Anteriores  | Posteriores  |
|---|--|
| a) Bases de datos<br>b) Análisis de grandes volúmenes de datos<br>c) Minería de datos | a) Tópicos selectos de ingeniería de datos<br>b) Procesamiento de datos en la nube |
| Tema(s)<br>Todos  | Tema(s)<br>Todos   |

| Nombre de la asignatura                                    | Departamento o Licenciatura                       |
|--|---|
| Herramientas para la gestión de grandes volúmenes de datos | Ingeniería en Datos e Inteligencia Organizacional |

| Ciclo | Clave  | Créditos | Área de formación curricular |
|-------|--------|----------|------------------------------|
| 3 - 4 | ID3468 | 6        | Licenciatura Elección Libre  |

| Tipo de asignatura | Horas de estudio |    |    |    |
|--------------------|------------------|----|----|----|
|                    | HT               | HP | TH | HI |
| Taller             | 16               | 32 | 48 | 48 |

## Objetivo(s) general(es) de la asignatura

---

### Objetivo cognitivo

Revisar características y principales aplicaciones de las herramientas de software que se utilizan para la gestión de grandes volúmenes de datos.

### Objetivo procedimental

Usar herramientas de software para la solución de problemas que involucren la gestión de grandes volúmenes de datos.

### Objetivo actitudinal

Fomentar el trabajo colaborativo y la responsabilidad en la solución de problemas para la adquisición de las habilidades requeridas.

## Unidades y temas

---

### Unidad I. INTRODUCCIÓN

Revisar características y principales aplicaciones de las herramientas de software que se utilizan para la gestión de grandes volúmenes de datos.

#### 1) Principales herramientas

a) Propósito

b) Clasificación

c) Arquitectura

d) Algoritmos

#### 2) Interacción con otras herramientas

#### 3) Casos de estudio

### Unidad II. HERRAMIENTAS PARA EXTRACCIÓN

Usar herramientas de software para la extracción de grandes volúmenes de datos.

- 1) Importancia de la lectura de conjuntos de datos
- 2) Elección de la herramienta: R, Python, para Web Scraper
- 3) Configuración y principales módulos y/o funciones
- 4) Automatización
- 5) Aplicación

### Unidad III. HERRAMIENTAS PARA TRANSFORMACIÓN

Usar herramientas de software para transformación de grandes volúmenes de datos en la integración de nuevos conjuntos.

- 1) Importancia de la transformación de datos
- 2) Elección de la herramienta: R, Python, Perl, XML, ATL
- 3) Configuración y principales módulos y/o funciones
- 4) Reglas de transformación
- 5) Aplicación

### Unidad IV. HERRAMIENTAS PARA CARGA DE DATOS

Usar herramientas de software para carga de grandes volúmenes de datos en la escritura de conjuntos de datos destino.

- 1) Importancia de la carga de datos
- 2) Elección de la herramienta: Bases de datos relacional, no relacional u orientada a grafos.
- 3) Configuración y principales módulos y/o funciones
- 4) Aplicación

## Actividades que promueven el aprendizaje

---

### Docente

Promover el trabajo colaborativo en la definición de propuestas de solución a problemas determinados.

Coordinar la discusión de casos prácticos.

Realizar foros para la discusión de temas o problemas.

### Estudiante

Realizar tareas asignadas

Participar en el trabajo individual y en equipo

Resolver ejercicios y/o problemas

Discutir temas en el aula

Participar en actividades extraescolares

## Actividades de aprendizaje en Internet

---

El estudiante deberá acceder a los enlaces para la búsqueda de información:

<https://www.r-project.org/>

<https://www.python.org/>

<https://www.perl.org/>

<https://www.xml.com/>

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

<https://www.microsoft.com/es-es/sql-server/>

<http://www.oracle.com>

<http://cassandra.apache.org/>

<https://neo4j.com/>

Se promoverá el uso de mecanismos asíncronos (correo electrónico, grupo de noticias, WWW y tecnologías de información) como medio de comunicación.

## Criterios y/o evidencias de evaluación y acreditación

---

### Criterios

Examen

Evidencias individuales

Evidencias equipo

### Porcentajes

30

20

30

|                     |     |
|---------------------|-----|
| Evidencias grupales | 20  |
| Total               | 100 |

## Fuentes de referencia básica

---

### Bibliográficas

Bahga, A., Madiseti, V. (2016) Big Data Science & Analytics: A Hands-On Approach (1a edición) EUA: VPT.

Kimball, R., Caserta, J. (2004) The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data (1a edición) EUA: Wiley.

Laberge, R. (2011) The Data Warehouse Mentor: Practical Data Warehouse and Business Intelligence Insights (1a edición) EUA: McGraw-Hill Education.

Manoochchri, M. (2013) Data Just Right: Introduction to Large-Scale Data & Analytics (Addison-Wesley Data and Analytics) (1a edición) EUA: Addison-Wesley Professional.

Steel, B., Chandler, J, y Reddy, S. (2016) Algorithms for Data Science (1a edición) EUA: Springer.

### Web gráficas

## Fuentes de referencia complementaria

---

### Bibliográficas

Caster, M., Bouman, R. y Dongen, J.(2010) Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration (1a edición) EUA: Wiley.

Chodorow, K. (2013). MongoDB: the definitive guide (1ª Edición). Sebastopol, EUA: O'Reilly Media.

Gillespie, C., Lovelace, R. (2017). Efficient R Programming (1a edición). EUA: O'Reilly.

Karau, H., Konwinski, A., Wendell, P., & Zaharia, M. (2015). Learning spark: lightning-fast big data analysis (1ª Edición). Sebastopol: O'Reilly Media.

Prieto A. (2013). ORACLE 12c/11g/10g Configuración, Gestión y Administración Avanzada de Bases de Datos. (1a edición) EUA: CreateSpace Independent Publishing Platform

Ryza, S., Laserson, U., Owen, S., & Wills, J. (2015). Advanced Analytics with Spark: Patterns for Learning from Data at Scale (1ª Edición). Sebastopol: O'Reilly Media.

Sarka, D., Radivojevic, M., & Durking W. (2017). SQL Server 2016 Developer's Guide. (1ª Edición). Birmingham, UK: Pack Publishing Ltd.

### Web gráficas

## Perfil profesiográfico del docente

---

### Académicos

Contar con licenciatura o posgrado en Ciencias de, la Computación, Sistemas, Datos, Estadística, o áreas afines.

### Docentes

Tener experiencia docente mínima de 3 años a nivel superior en asignaturas afines.

### Profesionales

Tener experiencia preferentemente en el desarrollo de sistemas de información, desarrollo y administración de base de datos y/o DW/BI.