



**UNIVERSIDAD AUTÓNOMA DE  
YUCATÁN  
FACULTAD DE MATEMÁTICAS**



**MISIÓN**

**Formar profesionales altamente capacitados, desarrollar investigación y realizar actividades de extensión en matemáticas y computación, así como en sus diversas aplicaciones.**

**Manejo de Datos no Estructurados**

**Dr. Francisco J. Moo Mena, M. en C. Juan F. Garcilazo Ortiz, Dr. Francisco A. Madera Ramírez**

**ASIGNATURA OPTATIVA DE LICENCIATURA**

**AGOSTO 2015 – NOVIEMBRE 2015**

## **Manejo de Datos no Estructurados**

**Optativa**  
**Horas: 75**  
**Hrs/Sem:4**  
**Créditos:7**  
**Clave:**

### **Objetivo del Curso:**

Aplicar técnicas que permitan, a partir de la recuperación de datos no estructurados, la generación de información relevante en distintas áreas del conocimiento mediante el uso de herramientas de software actuales.

### **Descripción:**

Con el auge actual en la generación de contenido en las grandes redes sociales, y otras aplicaciones modernas de Internet, surge la necesidad de contar con personal calificado en el manejo de grandes volúmenes de datos. A diferencia de las bases datos tradicionales que almacenan información bien estructurada, los datos emanados de Internet, en su mayoría, no presentan una estructura clara como para poder utilizarlos de manera convencional. Es por eso que se requiere, por un lado, de técnicas que permitan extraer dichos datos, analizarlos y presentarlos en forma de información relevante; y por otro lado de herramientas computacionales que permitan procesar de manera eficiente grandes volúmenes de datos. En este curso se utilizarán diversas herramientas no convencionales, para implementar técnicas que permitan el manejo eficiente de esos datos.

### **Contenido:**

#### **1. Introducción (5 sesiones).**

*Objetivo: El participante analizará las principales fuentes actuales para la generación de datos no estructurados.*

- 1.1 Computación en la nube.
- 1.2 Big Data.
- 1.3 Internet de las cosas.
- 1.4 Definición.

#### **2. Bases de datos no estructuradas (15 sesiones).**

*Objetivo: El participante estudiará las características de las bases de datos no estructuradas, así como las técnicas de recuperación de datos en diversos medios.*

- 2.1 Características.
- 2.2 Ejemplos de Base de datos no estructuradas.
- 2.3 Repositorios de datos públicos.
- 2.4 Recuperación de datos en redes sociales.

### **3. Manejo de datos estáticos (15 sesiones).**

*Objetivo: El participante generará información relevante a partir de datos no estructurados disponibles en bases de datos y repositorios públicos actuales.*

- 3.1 Manejo de datos con R.
- 3.2 Manejo de datos con Python.
- 3.3 Hadoop y Map Reduce.
- 3.4 Analítica de datos.

### **4. Manejo de datos dinámicos (15 sesiones)**

*Objetivo: El participante generará información relevante a partir de datos no estructurados obtenidos en tiempo real usando herramientas actuales.*

- 4.1 Características de Storm.
- 4.2 Integración de Storm con otras herramientas.
- 4.3 Gestión de datos en tiempo real.

### **Estrategias de Enseñanza-Aprendizaje:**

Estudios de casos, conferencia, interrogatorio, análisis y síntesis de información, resolución de ejercicios de programación, presentaciones orales y escritas, proyecto integrador.

### **Criterios de Evaluación:**

- Proyecto 40 puntos.
- Tareas 50 puntos.
- Participación 10 puntos.

### **Referencias**

El siguiente listado representa una bibliografía básica, la cual será enriquecida con diversas publicaciones especializadas a lo largo del curso.

1. Amy Newman and Howard Cohen. Cloud Computing: Discover the Skills that Power the Cloud. QuinStreet, Inc., 2012.
2. Aaron Weiss. 2007. Computing in the clouds. *netWorker* 11, 4 (December 2007), 16-25.
3. Brian Hayes. 2008. Cloud computing. *Commun. ACM* 51, 7 (July 2008), 9-11.
4. Fang Hao, T. V. Lakshman, Sarit Mukherjee, and Haoyu Song. 2010. Enhancing dynamic cloud-based services using network virtualization. *SIGCOMM Comput. Commun. Rev.* 40, 1 (January 2010), 67-74.
5. Faouzi Kamoun. 2009. Virtualizing the Datacenter Without Compromising

- Server Performance. Ubiquity 2009, August.
6. Luis Joyanes. Computación en la nube – Estrategias de cloud computing en las empresas. Alfaomega, 2012.
  7. Lizhe Wang, Rajiv Ranjan, JinJun Chen, Boualem Benatallah, Eds. Cloud Computing – Methodology, Systems, and Applications. CRC Press, 2012.
  8. Stuart Douglas, ed. End to End Cloud Computing. Intel Technology Journal, vol. 16, Issue 4, 2012.
  9. Alex Holmes (2012), Hadoop in Practice. Manning Publications.
  10. Foster Provost, and Tom Fawcett (2013), Data Science for Business: What you need to know about data mining and data-analytic thinking. O'Reilly Media.
  11. Jared Lander, (2013), R for Everyone: Advanced Analytics and Graphics. Addison-Wesley Data & Analytics Series
  12. Nina Zumel (2014), Practical Data Science with R. Manning Publications.
  13. Tom White (2012), Hadoop: The Definitive Guide. Yahoo Press.
  14. Viktor Mayer-Schönberger, and Kenneth Cukier (2014), Big Data: A Revolution That Will Transform How We Live, Work, and Think. Eamon Dolan/Mariner Books.
  15. Jonathan Leibiusky, Gabriel Eisbruch, and Dario Simonassi (2012), Getting Started with Storm. O'Reilly Media.
  16. P. Taylor Goetz and Brian O'Neill (2014), Storm Blueprints: Patterns for Distributed Real-time Computation. Packt Publishing.
  17. Quinton Anderson (2013), Storm Real-Time Processing Cookbook. Packt Publishing.
  18. Zacharias Voulgaris (2014), Data Scientist: The Definitive Guide to Becoming a Data Scientist. Technics Publications, LLC; First edition.

### **Perfil Profesiográfico del Profesor:**

Profesor con estudios mínimos de maestría en ciencias de la computación, experiencia docente y de investigación en áreas afines a la asignatura.

**Fecha de Elaboración:** Agosto 2015.

### **Elaboraron:**

Dr Francisco José Moo Mena.  
 M. en C. Juan Francisco Garcilazo Ortiz  
 Dr. Francisco Alejandro Madera Ramírez