

**Temario**

<b>DEPARTAMENTO</b>	<u>CEM</u>	<b>MATERIA</b>	<u>Big Data</u>
<b>PROFESOR</b>	<u>Daniel Vallejo</u>	<b>PERIODO</b>	<u>Primavera, 2018.</u>

**Objetivo general**

Entender la estructura y organización de los sistemas “Big Data” en general, entendiendo los conceptos de diseño e implementación. Se revisan los conceptos teóricos generales aplicados a los sistemas “Big Data” existentes actualmente en el mercado.

**Resultados de Aprendizaje**

Identificar los componentes fundamentales de un sistema “Big Data”. Entender los diversos aspectos que se toman en cuenta para el diseño de un sistema “Big Data” para un sistema computacional particular. Ser capaz de poder diseñar e implementar un sistema “Big Data” mínimo para un hardware básico.

**Requisitos para cursar la materia**

Estadística, Bases de Datos, Programación.

## **Temario detallado**

### R Programming

#### I Project 1: Weighted Dice

- 1 The Very Basics
- 2 Packages and Help Pages

#### II Project 2: Playing Cards

- 1 R Objects
- 2 R Notation
- 3 Modifying Values
- 4 Environments

#### III Project 3: Slot Machines

- 1 Programs
- 2 S3
- 3 Loops
- 4 Speed

## **Data Science**

### I Explore

- 1 Data Visualization with ggplot2
- 2 Workflow: Basics
- 3 Data Transformation with dplyr
- 4 Workflow: Scripts

### II Wrangle

- 1 Tibbles with tibble
- 2 Data Import with readr
- 3 Tidy Data with tidyr
- 4 Relational Data with dplyr

### III Program

- 1 Pipes with magrittr
- 2 Functions
- 3 Vectors
- 4 Iteration with purr

## IV Model

- 1 Model Basics with modelr
- 2 Model Building
- 3 Many Models with purr and broom

## V Communicate

- 1 Markdown
- 2 Graphics for Communication with ggplot2
- 3 R Markdown Formats

## **Hadoop**

### I Hadoop Fundamentals

- 1 Meet Hadoop
- 2 MapReduce
- 3 The Hadoop Distributed Filesystem
- 4 YARN

### II MapReduce

- 1 Developing a MapReduce Application
- 2 How MapReduce Works
- 3 MapReduce Types and Formats
- 4 MapReduce Features

### III Hadoop Operations

- 1 Setting Up a Hadoop Cluster
- 2 Administering Hadoop

### IV Related Projects

- 1 Pig
- 2 Hive
- 3 Spark
- 4 HBase

### V Case Studies

- 1 Composable Data at Cerner
- 2 Biological Data Science: Saving Lives with Software

### 3 Cascading

#### **Bibliografía**

Garrett Golemund, "Hands-On Programming with R", O'Reilly, 2014.

Hadley Wickham and Garrett Golemund, "R for Data Science", O'Reilly, 2016.

White, Tom, "Hadoop: The Definitive Guide", 4th Edition, Sixth Edition, O'Reilly, 2015.

Marz y Warren, "Big Data", First Edition, Manning, 2015.

Provost y Fawcett, "Data Science for Business", O'Reilly, 2013.

#### **Forma de Evaluar**

- Exámenes (4)            80%  
(Tres Exámenes Parciales y un Examen Final, cada uno un 20%)
- Programas                20%

**Nota Importante:** Para que el estudiante pueda aprobar el curso tiene que tener al menos 75/100 de promedio en los exámenes. Debido a las características del curso, ningún trabajo o examen se podrá entregar/hacer extemporáneo. Esto es, no habrá posibilidad de reposición de algo.

#### **Fechas Exámenes Parciales**

Las fechas tentativas de los exámenes parciales son:

Semana 3 de clases

Semana 6 de clases

Semana 9 de clases

Semana 12 de clases

#### **Fecha del Examen Final**

Lunes 7 de Mayo del 2018

El examen final puede ser programado por control escolar (CACE) durante un periodo de 10 días hábiles después del último día de clases luego es responsabilidad del estudiante permanecer en la universidad o cerca del campus hasta el último día de exámenes finales.

### **Con respecto a la asistencia a clases**

La asistencia a clases es obligatoria. La asistencia a clases corresponde al 10% de la calificación final del curso. El tener más del 50% de inasistencias implicará no poder presentar el examen parcial correspondiente.

### **Información Adicional**

Daniel Vallejo

Oficina; IA-231

Teléfono: 2-29-2639

E-mail: [daniel.vallejo@udlap.mx](mailto:daniel.vallejo@udlap.mx)

Página Web:

<http://comunidad.udlap.mx/portafolios/daniel.vallejo/Paginas/Default.aspx>

### **Horas de Oficina**

Lunes, Miércoles y Viernes: de 10:00 a 11:00 PM.

Otra hora previa cita.

(Mandar un correo electrónico para concertar una cita)

### **Acerca del correo electrónico y página web**

Las formas más comunes de comunicación entre el grupo y el profesor serán a través de la página del curso y el correo electrónico (e-mail). Ustedes deben de checar su cuenta de correo de la UDLAP (OWA) por lo menos una vez al día para verificar si no les he mandado ningún aviso y de manera similar deben de visitar la página del curso por lo menos una vez a la semana con la finalidad de

chechar si hay nuevo material del curso. Cualquier correo que me manden deberá tener la siguiente información: nombre y apellido del remitente, curso y sección en la que está inscrito y el asunto a tratar. De no recibir tal información sólo contestare solicitando me re-envíen el correo con la información antes mencionada. También les recuerdo que solo me pueden mandar correos de su dirección de la UDLAP (OWA).

## **Plagio**

Según la idea de plagio que se tiene en general es algo así como “tratar de pasar las ideas de otras personas como propias, aún bajo el supuesto del consentimiento de esas otra persona”. El plagio se considera algo inapropiado pues mina la confianza entre los que nos dedicamos a cuestiones científicas y académicas. Ejemplos típicos de plagio son: copiar en un examen, copiar tareas, presentar un programa en el cual no tuve ninguna participación, ya sea porque lo hicieron los otros integrantes de mi equipo o porque simplemente lo baje de la red y lo presento como si yo lo hubiera hecho. Mi posición particular ante esta práctica es de “cero tolerancia”, esto es, no la toleraré en absoluto. En caso de que desde mi punto de vista alguien incurra en esta práctica, analizaré el hecho de manera objetiva, y si encuentro elementos suficientes la consecuencia será que el estudiante reprobará el curso. El alumno por supuesto tendrá el derecho de apelar ante quien considere conveniente.

## **Reglamento de Estudiantes**

Para cualquier asunto no tratado en este documento pero que ustedes consideren que pueda no estar permitido ustedes tienen la responsabilidad de checar el Reglamento General de Estudiantes y los lineamientos que de éste se deriven. El Reglamento se encuentra en la siguiente liga:

[http://comunidad.udlap.mx/departamentos/normatividad/Documents/Estudiantes\\_General%20Reglamento.pdf](http://comunidad.udlap.mx/departamentos/normatividad/Documents/Estudiantes_General%20Reglamento.pdf)