



Laboratorio de Seminario de Sistemas 2

CATEGORÍA:	Obligatorio	SEMESTRE:	Primer Semestre 2018
AUXILIAR:	Pablo Echeverria	HORARIO:	19:00 – 20:40
EDIFICIO Y SALÓN:	T-3 304		
DÍAS QUE SE IMPARTE EL LABORATORIO:	Martes		

DESCRIPCIÓN DEL LABORATORIO:

A través de este curso los alumnos aprenderán a programar soluciones para procesar enormes cantidades de datos y crear modelos a partir de los datos, que permita clasificar la información partiendo de un conjunto de reglas inferidas de los datos.

OBJETIVO GENERAL:

Crear proyectos de inteligencia de negocios en las organizaciones utilizando las herramientas de tendencia.

Objetivos Específicos:

- Implementar proyectos de Inteligencia de negocios.
- Identificar y comprender los conceptos de Big Data.
- Comprender y explicar los diversos componentes del marco Spark.
- Transformar datos en información.
- Desarrollar una solución de inteligencia empresarial a través de Pentaho.
- Diseñar modelos predictivos bajo el concepto de minería de datos.
- Extraer de manera automatizada patrones ocultos, tendencias, anomalías, vinculaciones complejas y significados en grandes volúmenes de datos.
- Comprender conceptos básicos de la ciencia de los datos.
- Hacer uso del lenguaje R, desde la instalación del programa hasta la realización de los fundamentos de la Ciencia de los Datos.
- Evolucionar hacia el perfil de Data Scientist.
- Evolucionar hacia el perfil de Arquitecto Big Data.

METODOLOGÍA:

Se impartirán clases magistrales, con la participación de los estudiantes para el manejo de aplicaciones orientadas a cubrir los temas del contenido..

Autoaprendizaje y lectura constante de información relacionada al laboratorio.

Exámenes cortos, tareas, prácticas y proyectos, para el control del progreso.

EVALUACIÓN DEL RENDIMIENTO ACADÉMICO:

Según el Reglamento General de Evaluación y Promoción del Estudiante de la Universidad de San Carlos de Guatemala, la nota mínima para ganar el laboratorio es de **61 puntos**.

PROCEDIMIENTO

Instrumento de evaluación	Ponderación
Tarea 1: ETL con SQL Data Import Export	5 pts
Tarea 2: AdventureWorksDW, IntegrationServices, ReportingServices, AnalysisServices	5 pts
Tarea 3: PowerBI	10 pts
Tarea 4: Pentaho	10 pts
Tarea 5: Apache Spark	10 pts
Tarea 6: RapidMiner	10 pts
Tarea 7: R	10 pts
Proyecto:	40 pts
Nota de promoción	100 puntos

CONTENIDO

Unidad 1: Sistemas Datawarehousing

Tratamiento, conversión y transformación de los datos
Bases de datos multidimensionales
Herramientas "queries and reporting"
Análisis multidimensional (OLAP)

Unidad 2: Proceso ETL

Principales herramientas ETL
Introducción a Pentaho
Configuraciones iniciales
Instalación de Pentaho
PDI - Pentaho Data Integration
Ejemplo práctico de la herramienta Pentaho

Unidad 3: Procesando Big Data con Apache Spark

¿Porque usar Spark?
Arquitectura Spark y sus componentes
Modos de desarrollo en Spark
Framework Spark
Preparación del entorno
Introducción a Spark y a RDD
Transformaciones sobre RDD

Unidad 4: Minería de datos: Uso de RapidMiner

Data Access: Conectando con la información
Exploración de los datos: Descubriendo patrones
Data Blending: Analisis predictivo

Unidad 5: Programando con R

¿Qué es R?
Conociendo a R
Configuración
Aritmética básica y objetos
Operadores lógicos
Scripts en R
Importar datos - read.csv
Funciones básicas
Tipos de objetos
Seleccionando datos
Bucles y programación con R
Gráficos