

Lic. En Ciencia de Datos

Materia: Bases de datos estructuradas

OBJETIVOS DE APRENDIZAJE

General:

El(La) alumno(a) aplicará el análisis, diseño, desarrollo e implementación de Sistemas de información transaccionales y analíticos con datos estructurados a través de lecturas, ensayos, ejercicios en clase y prácticas para que al finalizar el curso sea capaz de crear sistemas de información OLTP y OLAP.

Específicos: El(La) alumno(a) comprenderá un ejemplo de negocio, aplicará, analizará y modelará dicho negocio para el diseño y construcción de bases de datos multidimensionales.

El(La) alumno(a) evaluará y utilizará herramientas OLAP para la generación de consultas, informes, tableros de instrumentos y cuadros de mando que correspondan con la estrategia de negocio para el análisis de datos estructurados.

TEMARIO

1. Introducción a los Sistemas de Información
 - 1.1 Concepto Sistema de Información
 - 1.2 Componentes
 - 1.2.1 Usuarios y procesos
 - 1.2.2 Hardware y redes, arquitectura cliente-servidor
 - 1.2.3 Software
 - 1.2.4 Sistemas Operativos
 - 1.2.5 Sistema Manejador de Base de datos
 - 1.2.6 Seguridad, confidencialidad, respaldo y recuperación
 - 1.2.7 Concurrencia
 - 1.2.8 Base de datos
 - 1.2.9 Integridad, consistencia
 - 1.3 Tipos principales de Sistemas de Información
 - 1.3.1 Características de los Sistemas Transaccionales
 - 1.3.2 Características de los Sistemas Analíticos
 - 1.4 Sistemas de Información Transaccionales (OLTP)
 - 1.4.1 Concepto transacción
 - 1.4.2 Propiedades ACID de la transacción
2. Análisis, diseño e implementación de sistemas de Información transaccionales (OLTP)
 - 2.1 Análisis de requerimientos
 - 2.1.1 Diseño conceptual
 - 2.1.2 Modelo Entidad-Relación
 - 2.1.3 Ejercicios de generación de modelo entidad-relación
 - 2.2 Diseño lógico
 - 2.2.1 Modelo Relacional
 - 2.2.2 Normalización
 - 2.2.3 Ejercicios de generación de modelo relacional
 - 2.2.4 Algebra relacional
 - 2.3 Diseño Físico
 - 2.4 SQL
 - 2.4.1 Lenguaje de definición de datos
 - 2.4.2 Lenguaje de control de datos
 - 2.4.3 Lenguaje de consulta de datos
 - 2.4.4 Ejercicios de consultas SQL
 - 2.4.5 Las doce reglas de Codd
 - 2.4.6 El modelo relacional y en memoria como mejor opción para sistemas Transaccionales
3. Análisis, Diseño e implementación de Sistemas de Información Analíticos (OLAP)

- 3.1 Problemática y retos en la inteligencia de negocios
 - 3.1.1 Conceptos básicos
 - 3.1.2 Arquitectura
 - 3.1.3 ¿Qué es un modelo de negocio?
 - 3.1.4 Preguntas para construir un modelo de negocio y notación BPMN
 - 3.1.5 Establecimiento de indicadores de rendimiento clave KPI
- 3.2 Análisis y diseño de la bodega de datos
 - 3.2.1 Modelado multidimensional
 - 3.2.2 Modelo multidimensional
 - 3.2.3 Hechos y dimensiones
 - 3.2.4 Desglose superior e inferior (drill up, drill down)
 - 3.2.5 Modelo copo de nieve (esquema normalizado)
 - 3.2.6 Ejemplo práctico
 - 3.2.7 Modelo estrella (esquema desnormalizado)
- 3.3 Esquema copo de nieve
- 3.4 Creación de bodega de datos con SQL
- 4 Proceso ETL (extracción, limpieza, transformación, integración, actualización)
 - 4.1 Extracción
 - 4.2 Limpieza (depuración, perfilado, corrección, estandarización, correspondencia de datos, consolidación)
 - 4.3 Transformación de acuerdo a las reglas de negocio, estándares, cambio formato, sustitución códigos, valores derivados y valores agregados, definición de nivel de detalle
 - 4.4 Carga (integración, actualización)
 - 4.5 Herramientas ETL
- 5 Componentes de Inteligencia de negocios para analítica descriptiva y predictiva
 - 5.1 Analítica descriptiva
 - 5.2 Consultas complejas en SQL
 - 5.3 Ejemplo práctico de consultas descriptivas
 - 5.4 Analítica predictiva
 - 5.5 Ejemplo práctico de consultas predictivas
 - 5.6 El uso del modelo columnar y en memoria como mejor opción para análisis de datos estructurados
 - 5.7 Introducción a Minería de datos para el análisis de datos
 - 5.8 Herramientas OLAP para diseño
 - 5.9 Herramientas Tableros de instrumentos y Cuadros de mandos (dashboard y Scorecard)
 - 5.10 Herramientas minería de datos

Syllabus Bases de datos estructuradas

DESARROLLO DE CONTENIDOS

A continuación, se describen las actividades que se desarrollaran durante el semestre en la asignatura Bases de datos estructurados.

Sem	Materiales didácticos y situación de aprendizaje	Fechas	Temas e interacciones deseadas
1	Mat: 1-Intro-SisInf Lectura: 1-X-EvolucionSistInfo Ensayo sobre lectura Mat: 2-0-AnalisisyDiseñoOLTP	Ago. 8,10,12	Presentación, evaluación Tema 1: Introducción a los Sistemas de Información 1.1 Concepto Sistema de Información 1.2 Componentes 1.3 Tipos principales de Sistemas de Información 1.4 Sistemas de Información Transaccionales (OLTP) Participación activa sobre ejemplos de la vida real
2	Mat:2-1-DiseñoConceptual-MER Lectura: ArtículoPeterChen Ev. Ejercicios en clase Listo para Practica 1 Modelo-ER	Ago 15,17,19	2.1 Análisis de requerimientos 2.1.1 Diseño conceptual 2.1.2 Modelo Entidad-Relación 2.1.3 Ejercicios de generación de modelo entidad-relación: reclusorios, pacientes, películas, bancos Participación activa sobre el diseño conceptual Especificación de Proyecto OLTP
3	Mat2-2DiseñoLogico.ModeloRel-Norma Lectura: ArtículoEdCodd Listo para Practica 2 Modelo-Relacional	Ago 22,24,26	2.2 Diseño lógico 2.2.1 Modelo Relacional 2.2.2 Normalización 2.2.3 Ejercicios de generación de modelo relacional 2.2.4 Algebra relacional Sistemas: reclusorios, pacientes, películas, bancos Participación de estudiantes sobre dudas del proyecto, ejemplos, propuestas de temas afines al estudiantado. DUDAS PROYECTO OLTP- MER Práctica P1-OLTP-Modelo-ER
4	Mat.2-3-DiseñoFisico-BD Mat.2-4-SQL Demostración conexión Postgres Listo para Practica 3 Diseño-Físico	Ago 29,31, Sep 2	2.3 Diseño Físico 2.4 SQL, 2.4.1 Lenguaje de definición de datos 2.4.2 Lenguaje de control de datos Participación de estudiantes con programas SQL propuestos sobre ejemplos con perspectiva de género DUDAS PROYECTO OLTP- MR DUDAS PROYECTO OLTP- MF Práctica P2-OLTP-ModeloRelacional Práctica P3-OLTP-ModeloFísico
5	Mat.2-4-SQL Lectura: Reglas de Codd Ev. Ensayo Listo para Practica 4 SQL	Sep. 5,7,9	2.4.3 Lenguaje de consulta de datos 2.4.4 Ejercicios de consultas SQL 2.4.5 Las doce reglas de Codd Participación de estudiantes sobre puntos de vista sobre los ejercicios y las reglas de Codd. 2.4.6 El modelo relacional y en memoria como mejor opción para sistemas transaccionales. Participación sobre ensayo, puntos de vista sobre Hanna, motivación. DUDAS PROYECTO OLTP-SQL Práctica P4-OLTP-Consultas
6	12-BD Memoria Lectura: Hanna Fundamentals Ensayo	Sep. 12,14	ENTREGA PROYECTO OLTP 1er. Examen Parcial
7	Lectura: 8-RepasosistemasOLTP,OLAP y BDR Ev. Ensayo Mat.2-Problematíca y Retos en IN-modelo de negocio. Listo para Práctica 1 DWH-modelo negocio en BPMN	Sep. 19,21,23	3 Análisis, Diseño e implementación de Sistemas de Información Analíticos (OLAP) 3.1 Problemática y retos en la inteligencia de negocios 3.1.1 Conceptos básicos 3.1.2 Arquitectura 3.1.3 ¿Qué es un modelo de negocio? 3.1.4 Preguntas para construir un modelo de negocio y notación BPMN usando ejemplos con perspectiva de género 3.1.5 Establecimiento de indicadores de rendimiento clave KPI

Syllabus Bases de datos estructuradas

			Participación sobre construcción de modelos con BPMN Especificación ProyectoOLAP
8	Mat.3-AnálisisyDiseñoOLAP	Sep. 26,28,30	3.2 Análisis y diseño de la bodega de datos 3.2.1 Modelado multidimensional 3.2.2 Modelo multidimensional 3.2.3 Hechos y dimensiones Participación con ejemplos de la vida real que ellos encuentren. <i>DUDAS PROYECTO OLAP- SOBRE MODELADO NEGOCIO</i> Práctica P1-DWH-modelonegocio
9	Mat.3-AnálisisyDiseñoOLAP Act.3-EjerciciosModeladoDWH- Alumnos Explicación de ProyectoOLAP	Oct. 3,5,7	3.2.4 Desglose superior e inferior (drill up, drill down) 3.2.5 Modelo copo de nieve (esquema normalizado) 3.2.6 Ejemplo práctico Participación sobre construcción de modelado DWH <i>DUDAS PROYECTO OLAP SOBRE MODELADO MULTIDIMENSIONAL</i>
10	Mat.3-AnálisisyDiseñoOLAP Act.EjerciciosModeladoDWH- Alumnos Listo para Práctica 2 DWH- DiseñoDWH; Practica3 DWH- CreacionDWH	Oct. 10,12,14	3.2.7 Modelo estrella (esquema desnormalizado) 3.3 Esquema copo de nieve 3.4 Creación de bodega de datos con SQL <i>DUDAS PROYECTO OLAP SOBRE MODELADO MULTIDIMENSIONAL</i> Práctica P2-DWH-DiseñoDWH
11	Mat.4-ETL- CalidadyLimpiezaDatos Listo para Practica 4 DWH- Ev.Extracción-Perfil	Oct. 17,19,21	4 Proceso ETL (extracción, limpieza, transformación, integración, actualización) 4.1 Extracción 4.2 Limpieza (depuración, perfilado, corrección, estandarización, correspondencia de datos, consolidación) Práctica P3-DWH-CreacionDWH Práctica P4-DWH-Extraccion-Perfil
12	Mat.4-ETL- CalidadyLimpiezaDatos Listo para Practica5 DWH-Trans- Limpieza Ev. Practica6DWH-CargaDatos	Oct. 24,26,28	4.3 Transformación de acuerdo a las reglas de negocio, estándares, cambio formato, sustitución códigos, valores derivados y valores agregados, definición de nivel de detalle 4.4 Carga (integración, actualización) 4.5 Herramientas ETL <i>DUDAS PROYECTO OLAP SOBRE IMPLEMENTACIÓN</i> Práctica P5-DWH-Trans-Limpieza
13	Mat.5-ConsultasOLAP Listo para Practica7DWH- ConsultasOLAP	Oct. 31, Nov. 4	5 Componentes de Inteligencia de negocios para analítica descriptiva y predictiva 5.1 Analítica descriptiva 5.2 Consultas complejas en SQL 5.3 Ejemplo práctico de consultas descriptivas Práctica P6-DWH-CargaDatos
14	Act. AlumnosExponenProyecto, Ev. Presentación Act. SegundoExamenParcial	Nov. 7,9,11	Demostración Qlik/RapidMiner <i>Cada equipo expone y entrega proyecto OLAP en 10 mins.</i> <u>Segundo Examen Parcial-AnalíticaDescriptiva</u> Práctica P7-DWH-ConsultasOLAP
15	Mat.6-Analítica Predictiva Mat.7-2-IntroMineríadeDatos Listo para Practica8- ConsultasPredictivas	Nov. 14,16,18	Introducción y Roles en Ciencia de datos 5.4 Analítica predictiva 5.5 Ejemplo práctico de consultas predictivas 5.6 El uso del modelo columnar y en memoria como mejor opción para análisis de datos estructurados 5.7 Introducción a Minería de datos para el análisis de datos 5.8 Herramientas OLAP para diseño 5.9 Herramientas Tableros de instrumentos y Cuadros de mandos (dashboard y Scorecard) 5.10 Herramientas minería de datos
16	Mat.7-2IntroMineríadeDatos Opcional: Lectura y ensayo sobre: 8-Repaso sistemas OLTP,OLAP yBDR ;Act. RapidMiner, Pytho	Nov 23,25	Opcional: Entrega de consultas predictivas con herramienta sobre proyecto OLAP <u>Tercer Examen Parcial</u> Práctica P8-DWH-ConsultasPredictivas
		Nov 28	1er. Examen Final, Retroalimentación
		Dic. 5	2do. Examen Final,Retroalimentación

EVALUACIÓN DEL CURSO

Evaluación de las actividades y el peso relativo de cada grupo de ellas para conformar la calificación final del curso.

Actividad	Porcentaje
Exámenes Parciales	50%
Participaciones	10%
Prácticas OBLIGATORIAS	20%
Proyecto OBLIGATORIO	20%
Total	100%
Examen Final	100%

Por reglamento general de exámenes, se tienen tres oportunidades para acreditar la materia:

- 1.- Presentar TODOS los elementos correspondientes a la evaluación del curso (tabla anterior) en tiempo y forma. Al obtener un promedio mayor o igual a 7.6 se da por acreditado el curso.
- 2.- Presentar el primer examen final y su calificación se pone en actas.
- 3.- Presentar el segundo examen final y su calificación se pone en actas.

Para las calificaciones con enteros de 6 en adelante y decimales .6 sube al siguiente entero. Ej. 5.6 = 6, 6.6=7.

Por reglamento general de exámenes no se puede presentar final para subir de calificación.

No se acredita la materia si se obtiene calificación menor o igual a 5.9 en exámenes finales.

BIBLIOGRAFIA

Básica

1. Coronel, C., & Morris, S. (2016). Database Systems. Mason, OH: Cengage Learning.
2. Date, C. (2007). Introducción a los sistemas de bases de datos. Wilmington (Delaware): Pearson Educación.
3. Inmon, W. (2011). Building the data warehouse. Indianapolis: Wiley.
4. Kimball, R. and Ross, M. (2013). The data warehouse toolkit. Hoboken: Wiley.
5. Lans, R. (2012). Data virtualization for business intelligence systems. [Place of publication not identified]: Morgan Kaufmann.
6. Ricardo, C., & Urban, S. (2017). Databases illuminated. Burlington, MA: Jones & Bartlett Learning.

PAGINA DE LA MATERIA:

<http://cienciadatos.iimas.unam.mx/profesores/pilarang>

<http://profesores.fi-b.unam.mx/pilarang/>

PRÁCTICAS:

Las prácticas se entregan vía correo electrónico dirigido a **basedatosrel@yahoo.com.mx** el día de la fecha de entrega antes de las 2 de la tarde los entregables consisten en lo siguiente:

- 1.- El Asunto del correo electrónico debe ser el **número, tipo de práctica y el equipo**, (ej. Practica 1 DB2, Equipo 3), en el cuerpo del correo deben estar los nombres de los participantes.
- 2.- El Archivo que contenga las imágenes de las pantallas en donde se reflejen la ejecución de los comandos completos que se requieren para realizar todos los pasos de todas las actividades y sus correspondientes resultados.
- 3.- El Archivo script con comandos SQL que se piden en dicha actividad.
- 4.- Los **archivos deben estar adjuntos al correo electrónico, NO deben ser parte del texto del mensaje.**

NOTAS:

Syllabus Bases de datos estructuradas

a) Los correos deben tener en el Asunto el número de práctica y el número de Equipo, en el cuerpo del correo deben estar los nombres completos de todos los integrantes que participaron en la elaboración de la práctica.

b) En caso de que no se envíe a la dirección indicada, fuera de horario establecido o sin los archivos adjuntos correspondientes, se considera la práctica como no entregada. **LO MISMO APLICA PARA EL PROYECTO.**

PARA EL PROYECTO.

Fechas de entrega de prácticas de Temas 1 y 2

Práctica P1-OLTP-Modelo-ER	SEM 3 AGOSTO 26
Práctica P2-OLTP-ModeloRelacional	SEM 4 SEPTIEMBRE 2
Práctica P3-OLTP-ModeloFisico	SEM 4 SEPTIEMBRE 2
Práctica P4-OLTP-Consultas	SEM 5 SEPTIEMBRE 9

Fechas de entrega de prácticas de Temas 3,4 y 5

Práctica P1-DWH-modelonegocio	SEM 8 SEPTIEMBRE 30
Práctica P2-DWH-DiseñoDWH	SEM 10 OCTUBRE 14
Práctica P3-DWH-CreacionDWH	SEM 11 OCTUBRE 21
Práctica P4-DWH-Extraccion-Perfil	SEM 11 OCTUBRE 21
Práctica P5-DWH-Trans-Limpieza	SEM 12 OCTUBRE 28
Práctica P6-DWH-CargaDatos	SEM 13 NOVIEMBRE 4
Práctica P7-DWH-ConsultasOLAP	SEM 14 NOVIEMBRE 11
Práctica P8-DWH-ConsultasPredictivas	SEM 16 NOVIEMBRE 25.

PROYECTO:

El proyecto debe entregarse en papel y en formato electrónico por correo electrónico a la misma dirección que las Prácticas. El documento en pdf debe contener los elementos que se especifican en el formato de entrega que se indica en la hoja de requerimientos, en el mismo orden. Al igual que las prácticas, código en formato txt y pantallas con la ejecución correcta de la creación, inserción de las tablas así como la ejecución de altas, bajas y cambios en los datos que reflejen el aseguramiento de la integridad referencial (si el dato no es correcto, no permite, si lo es si permite) , consultas que se piden, así como la interpretación de los resultados de los análisis.. **ES MUY IMPORTANTE QUE SE PROPORCIONE EL CODIGO PARA BORRAR LAS BASES DE DATOS, ASI COMO LOS CÓDIGOS PYTHON, R,WEKA O RAPIDMINER USADOS.**