



ITESO

Universidad Jesuita
de Guadalajara

DEPARTAMENTO DE ELECTRÓNICA, SISTEMAS E INFORMÁTICA

BASES DE DATOS PARA APOYAR LA TOMA DE DECISIONES

MTRO. VÍCTOR ORTEGA

OTOÑO 2012

CONTENIDO

1. INFORMACIÓN DEL PROFESOR.....	3
2. PRESENTACIÓN.....	4
3. PROPÓSITO GENERAL.....	5
4. PROPÓSITOS ESPECÍFICOS.....	5
5. CONTEXTO CURRICULAR.....	6
6. PROYECTO.....	7
7. SITUACIONES DE APRENDIZAJE.....	8
8. EVALUACIÓN.....	13
9. REFERENCIAS.....	13
10. ANEXO 1 →Entregables del proyecto y fases de revisión.....	15
1.0.1 ENTREGABLES DEL PROYECTO.....	15
10.2 FASES DE REVISIÓN DEL PROYECTO.....	17

ASIGNATURA: Bases de datos para apoyar la toma de decisiones	CREDITOS: 4 BCD 4 TIE
CLAVE DE ASIGNATURA Y GRUPO: esi032a	HORARIO: Martes y jueves 11 a 13 horas SALON: W104 – T206
PROGRAMA ACADEMICO: ISC, LTI	IDIOMA: Español
DEPARTAMENTO: ESI	PERIODO ESCOLAR: Otoño 2012

1. INFORMACIÓN DEL PROFESOR

Grado, nombre y categoría: Mtro. Víctor Hugo Ortega Guzmán. Profesor Titular	
Correo Electrónico: vortega@iteso.mx	Página Web: http://cursos.iteso.mx/course/view.php?id=451

2. PRESENTACIÓN

En la actualidad las organizaciones tienen cada vez más datos, tanto estructurados, almacenados en bases de datos, como no estructurados, disponibles en correos electrónicos, videos o mensajes de *facebook*, por mencionar algunos.

Es necesario que un profesional de Sistemas Computacionales o Tecnologías de Información cuente con las competencias necesarias para almacenar y analizar una gran cantidad de datos estructurados, ya sea como miembro de una organización o como consultor externo.

En este curso aprenderás a utilizar técnicas para aprovechar los recursos de información estructurados que las organizaciones han creado en su operación a través del uso de sistemas administrativos durante varios años, lo que requerirá que indagues sobre fuentes de datos internas y externas, construyas cubos de información que facilitan dinámicamente combinar variables de negocios, obtengas información y apliques también algoritmos de minería de datos que te permitan encontrar relaciones ocultas o nuevo conocimiento sobre la organización.

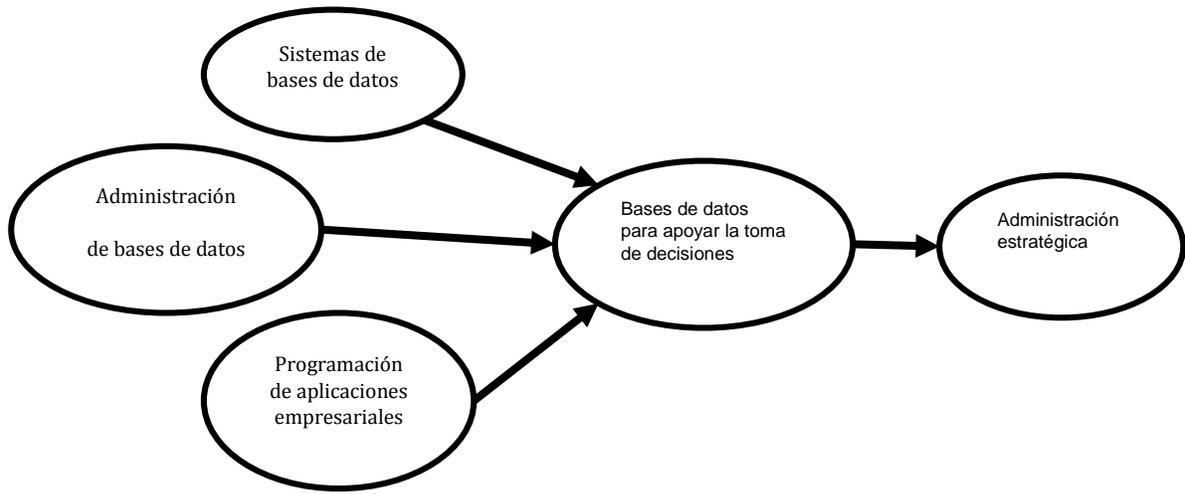
3. PROPÓSITO GENERAL

Aplicar técnicas para almacenar y analizar bases de datos históricas, para contribuir a la toma de decisiones que generen valor agregado en las organizaciones.

4. PROPÓSITOS ESPECÍFICOS

1. Cruzar información de los principales indicadores de las diferentes áreas de negocio a lo largo de varios años de operación a partir de la construcción de cubos de información (*data marts*).
2. Mostrar la situación actual o pasada del negocio respecto a los indicadores y áreas elegidas a partir de la construcción de tableros de control (*scorecards*).
3. Predecir el comportamiento de los indicadores de las áreas de negocio con base en la aplicación de diferentes algoritmos de minería de datos (*data mining*).

5. CONTEXTO CURRICULAR



6. PROYECTO

NOMBRE DEL PROYECTO:

Desarrollo de una solución de Inteligencia de Negocios para la industria automotriz que incluya al menos las áreas de ventas y compras.

PROPÓSITOS DEL PROYECTO:

- Construir un data mart de área de ventas
- Construir un data mart del área de compras
- Construir un tablero de control con los tres principales indicadores para el área de ventas
- Construir un tablero de control con los tres principales indicadores para el área de compras
- Aplicar algoritmos de clasificación, segmentación y agrupamiento a la información área de ventas y compras con la intención de encontrar venta cruzada de productos o segmentos de productos y/o clientes

CRITERIOS DE EVALUACIÓN:

En el [Anexo1](#) se encuentra el [detalle de los entregables](#) y las [fases de revisión](#)

Además de lo anterior:

Trabajo en equipo

Corresponsabilidad: todos los miembros del equipo participan, aportan y dominan el contenido del proyecto.

7. SITUACIONES DE APRENDIZAJE

Este curso toma como situación central los procesos de negocio de ventas y compras de una empresa de autopartes.

Propósito 1: Cruzar información de clientes, productos, proveedores a lo largo de varios años de operación a partir de la construcción de cubos de información (<i>data marts</i>)				
Aprendizajes esperados	Sesión	Actividades BCD	Actividades TIE	Productos e indicadores de evaluación
<ul style="list-style-type: none"> - Establecer semejanzas y diferencias entre el modelo relacional y el modelo dimensional - Catalogar las tablas de las de una BD relacional en transacción, catálogo, histórico o relación - Crear modelos dimensionales de estrella y copo de nieve - Poblar los modelos dimensionales con los datos de las BD relacionales - Crear un cubo - Analizar un cubo - Disposición para explorar nuevas formas de trabajar con bases de datos - Establecer relaciones entre el cubo de información y la generación de valor agregado en la organización 	1	G: A partir del análisis de ejemplos, explicar por qué ciertas empresas ofrecen valor agregado dada la información con la que cuentan y el uso que hacen de ella.		1 Diseños dimensionales: <ul style="list-style-type: none"> - Cuentan solo con las dimensiones necesarias para cumplir con toda la información establecida en el documento de requerimientos. - Cuentan con los datos calculados necesarios para cumplir con dicha información. 2 Cubo de información (con y sin conexión): <ul style="list-style-type: none"> - Atiende toda la información establecida en el documento de requerimientos. - La información refleja correctamente la contenida en la BD relacional. 3 Texto que justifique la manera en que un cubo de información contribuye a la generación valor agregado en una organización: <ul style="list-style-type: none"> - Explica los casos en los que tiene ventajas utilizar el modelo dimensional vs. el modelo relacional. - Comunica con claridad y precisión las ideas - Incluye ventajas y desventajas de crear un cubo de información.
	2	I: Analizar diferentes Cubos OLAP o Data Marts para dar cuenta del tipo de información que ofrecen (práctica 1). Todas las prácticas se llevan a cabo en el laboratorio T206, o si lo prefieres puedes instalar Ms SQL SERVER 2008 R2 Developer en tu computadora portátil.	I: Concluir la práctica 1.	
	3	G: A partir del análisis de ejemplos, describir las características de las bases de datos transaccionales.		
	4	I: Manejar un punto de venta, un CRM y un ERP para reconocer las posibles fuentes de información que los Cubos OLAP o Data Mart pueden utilizar (práctica 2).	I: Concluir la práctica 2.	
	5	I: Analizar las tablas de una base de datos transaccional asociadas a un proceso de negocio para clasificar en transaccional, histórica, de relación o catálogo (práctica 3).	I: Concluir la práctica 3.	

Aprendizajes esperados	Sesión	Actividades BCD	Actividades TIE	Productos e indicadores de evaluación
	6 - 7	G: Realizar modelos dimensionales de estrella a partir de los requerimientos de información presentados en diferentes casos.	<i>I: Concluir los primeros dos casos de modelado</i>	
	8	I: A partir de los requerimientos de información de la práctica 4, realizar un modelo dimensional para construir un Cubo OLAP.	<i>I: Concluir la práctica 4.</i>	
	9	I: A partir del modelo creado en la práctica anterior, construir un Cubo OLAP (práctica 5).	<i>I: Concluir la práctica 5.</i>	
	10-11	E: Analizar las tablas de la base de datos del proyecto asociadas a los procesos de venta y compra para clasificarlas en transaccional, histórica, de relación o catálogo	<i>E: Concluir el análisis de la BD transaccional del proyecto.</i>	
	12-13	E: Con base en los requerimientos de información, hacer el diseño dimensional de cada Cubo de Información del proyecto.	<i>E: Concluir el diseño dimensional de cada cubo.</i>	
	14-15	E: A partir de los diseños aprobados, construir los Cubos de Información (con y sin conexión) requeridos para el proyecto.	<i>E: Concluir la construcción de los Cubos de Información requeridos para el proyecto.</i>	
	16	E: Entregar la Fase1 del proyecto: Los cubos de información funcionando de acuerdo a los requerimientos. I: Responder al examen parcial.	<i>E: Con base en la información que aparece en el ANEXO 1, documentar su proyecto para la primera fase de entrega.</i>	

Propósito 2: Mostrar la situación actual o pasada del negocio respecto a las ventas, compras, mejores productos y utilidades a partir de la construcción de tableros de control (*scorecards*)

Aprendizajes esperados	Sesión	Actividades BCD	Actividades TIE	Productos e indicadores de evaluación
<p>- Establecer semejanzas y diferencias entre el tablero de control y el cubo de información.</p> <p>- Crear un tablero de control</p> <p>-Establecer relación entre el cubo de información y el tablero de control para la generación de valor agregado en la organización</p> <p>-Disposición para explorar nuevas formas de trabajar con bases de datos</p>	17	<p>I: Con base en su historial académico, crear un diseño de tablero control en papel.</p> <p>G: A partir del análisis de ejemplos, recuperar los 13 errores típicos en el diseño de un tablero de control y crear (o mejorar) el diseño original de su tablero.</p> <p>I: Comparar el diseño original contra el nuevo diseño.</p>		<p>1. Tableros de control:</p> <ul style="list-style-type: none"> - Cuentan con los indicadores necesarios para analizar cada área de negocio. - No presenta ninguno de los errores comunes de diseño. <p>2 Texto que justifique la manera en que el tablero de control contribuye a la generación de valor agregado en una organización:</p> <ul style="list-style-type: none"> - Explica los casos en los que es útil emplear un tablero de control. - Comunica con claridad y precisión las ideas.
	18	I: A partir de los requerimientos de información de la práctica 6, construir un tablero de control.	I: Concluir la práctica 6.	
	19-20	<p>E: A partir de los requerimientos de información del proyecto, crear los tableros de control necesarios.</p> <p>E: Entregar la Fase1 del proyecto: Los cubos de información funcionando de acuerdo a los requerimientos.</p>	<p>E: Concluir la creación de los tableros de control requeridos para el proyecto.</p> <p>E: Con base en el ANEXO1, documentar su proyecto para la primera fase de entrega.</p>	

Propósito 3: Predecir el comportamiento de los clientes con base en la aplicación de diferentes algoritmos de minería de datos (data mining)

Aprendizajes esperados	Sesión	Actividades BCD	Actividades TIE	Productos e indicadores de evaluación
<p>- Establecer semejanzas y diferencias entre el cubo de información y los resultados de aplicar algoritmos de minería de datos.</p> <p>- Crear las estructuras necesarias para aplicar cada algoritmo de minería de datos.</p> <p>-Establecer relaciones entre los resultados de aplicar los algoritmos de minería de datos y la generación de valor agregado en la organización.</p> <p>-Disposición para explorar nuevas formas de interpretar los datos</p>	21	G: Inferir la estructura de datos que está implementada en http://es.akinator.com/ y su relación con los algoritmos de clasificación de minería de datos.		<p>1.Gráfica de resultados de la de minería de datos acompañada de su interpretación:</p> <ul style="list-style-type: none"> - Cuenta con las variables necesarias (discretas o continuas) para aplicar cada algoritmo de minería de datos. - Contiene las tablas de entrada –y datos- adecuadas para aplicar cada algoritmo - Puede interpretar el resultado de cada algoritmo en un lenguaje no técnico. - Con base en los datos históricos puede hacer predicciones de comportamientos de clientes - La interpretación los resultados es correcta. <p>2 Texto que justifique la contribución de aplicar cada algoritmo de minería de datos y el valor agregado que aportan en una organización:</p> <ul style="list-style-type: none"> - Explica los casos en los que es útil emplear un algoritmo de minería de datos en lugar de un Cubo de Información - Comunica con claridad y precisión las ideas
	22	I: Aplicar los algoritmos de minería de datos necesarios para resolver cada caso descrito en la práctica (práctica7).	I: Concluir la práctica 7.	
	23-24	G: A partir del análisis de ejemplos, recuperar los pasos necesarios para construir una estructura base para aplicar los algoritmos de clasificación, segmentación o agrupamiento de datos y reconocer en qué caso es mejor aplicar cada uno.	E: Construir las variables y estructuras de entrada para aplicar cada tipo de algoritmo al proyecto.	
	25-26	<p>E: Dadas las hipótesis planteadas en el proyecto, aplicar los algoritmos de minería de datos que pudieran proporcionarles respuestas</p> <p>E: Interpretar las gráficas resultantes de aplicar de cada algoritmo al proyecto.</p>	<p>E: Concluir con la aplicación de los algoritmos de minería de datos en el proyecto.</p> <p>E: Concluir con la interpretación de los resultados de cada algoritmo.</p>	
	27-28	I: Utilizar el lenguaje MDX y DMX para manipular los Cubos de Información y la implementación de la minería de datos (práctica 8).	I: Concluir la práctica 8.	
	29-30	E: Utilizar MDX para automatizar consultas a los Cubos de Información y DMX para hacer una predicción en tiempo real en el proyecto.	E: Hacer la implementación de consultas automáticas a los Cubos de Información y al menos un caso de minería de datos en tiempo real en su proyecto	

Aprendizajes esperados	Sesión	Actividades BCD	Actividades TIE	Productos e indicadores de evaluación
	31	E: Entregar la Fase 2 del proyecto: Cubos de información, tableros de control y algoritmos de minería de datos. I: Responder al examen parcial.	E: Tomando como referencia el ANEXO 1, documentar su proyecto para la segunda Fase de entrega. No olviden incluir la retroalimentación de la Fase 1. E: Preparar la presentación de las conclusiones y hallazgos de su proyecto	
	32	G: Compartir las conclusiones y hallazgos de su proyecto.		

8. EVALUACIÓN

Exámenes parciales complementando teoría con práctica	30
Prácticas de laboratorio	20
Proyecto FASE I (<i>Data Mart y scorecards</i>)	25
Proyecto FASE II (<i>Data Mining</i>)	25

	100

9. REFERENCIAS

Título	Autor	Editorial-Año	Referencia o clave Dewey
Dimensional Modeling: In a Business Intelligence Environment	Chuck Ballard Daniel M. Farrell Amit Gupta Carlos Mazuela Stanislav Vohnik	IBM Redbooks publication Last Update 19 mayo 2008 ISBN 0738496448	http://www.redbooks.ibm.com/abstracts/sg247138.html
Dynamic Warehousing: Data Mining Made Easy	Chuck Ballard John Rollins Jo Ramos Andy Perkins Richard Hale Ansgar Doerneich Edward Cas Milner Janardhan Chodagam	IBM Redbooks publication Publish Date 06 septiembre 2007 ISBN 0738488860	http://www.redbooks.ibm.com/abstracts/sg247418.html?Open&pdfbookmark
The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling	Kimball, Ralph Ross, Margy	Nueva York, EUA: Wiley, 2002, c2002.	005. 74068 KIM
Data mining techniques : for marketing, sales, and customer support	Berry, Michael J.A. Linoff, Gordon	Nueva York, EUA: Wiley, 1997, c1997.	658. 802 VER
Practical DMX Queries for Microsoft SQL Server Analysis Services 2008	Tennick, Art.	Nueva York, EUA: McGraw-Hill, 2011, c2011.	005. 7565 TEN
Practical MDX Queries for Microsoft SQL Server Analysis Services 2008	Tennick, Art.	Nueva York, EUA: McGraw-Hill, 2010, c2010.	005. 7565 TEN
Microsoft SQL Server 2008 R2 Unleashed	Rankins, Ray.	Indianápolis, EUA: Sams, 2011, c2011.	005. 7565 MIC

Página Web / Base de datos	Tecnología	Título
http://www.microsoft.com/sqlserver/2008/en/us/default.aspx	DW, DM, SC	Microsoft SQL Server
http://www.jedox.com/de/home/uebersicht.html	DW	Palo
http://www.cs.waikato.ac.nz/ml/weka/	DM	Weka
http://www.businessobjects.com/product/catalog/xcelcius_engage/trial.asp	SC	Xcelcius
http://www.pentaho.com/	DW, DM, SC	Pentaho

10. ANEXO1 → ENTREGABLES DEL PROYECTO Y FASES DE REVISIÓN

1.0.1 ENTREGABLES DEL PROYECTO

I. DOCUMENTO (DOC O PDF) CON LA SIGUIENTE ESTRUCTURA:

1. Portada
 - a. Nombre de los integrantes
 - b. Carrera y porcentaje de créditos aprobados
 - c. Número de expediente
 - d. Nombre de la institución
 - e. Nombre de la materia
 - f. Nombre del profesor
2. índice
3. Introducción
4. Requerimientos de información por cada cubo
5. Imágenes de los modelos dimensionales utilizados
6. *Stored Procedures* comentados
7. Consultas MDX comentadas
8. Imágenes de los cubos creados
9. Al menos 10 casos de prueba

Hay que hacer una consulta con su cubo capturar la pantalla y hacer la misma consulta sobre la base de datos transaccional (añadir su sentencia SQL) y capturar la pantalla

10. Imágenes del tablero de control creado
11. Hallazgos de los cubos
12. Estructura de la (las) tablas para aplicar los algoritmos de minería de datos
13. Gráficas con hallazgos por algoritmo aplicado
14. Consultas DMX con hallazgos de minería de datos
15. Conclusiones y recomendaciones
16. Fuentes bibliográficas: libros, tutoriales, páginas consultadas de internet, artículos de revistas

II. ARCHIVOS CUB CREADOS DESDE MSEXCEL

Imprescindible probar los cubos al menos en dos computadoras antes de entregarlos

III. ARCHIVO .SQL CON LOS STORED PROCEDURES DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE DATOS

IV. ARCHIVO .SQL CON LAS CONSULTAS MDX QUE RESPONDAN A LOS REQUERIMIENTOS DE INFORMACIÓN

V. PROYECTO DE VISUAL STUDIO QUE CONTENGA LOS CUBOS

VI. RESPALDO DE LA BASE DE DATOS DIMENSIONAL

Una base de datos que contenga los tres cubos

VII. ARCHIVO SWF QUE MUESTRE SU TABLERO DE CONTROL FUNCIONANDO

Puede crearse bajando una versión de prueba de 30 días del software Adobe Captivate

<http://www.adobe.com/products/captivate/> o con el Camstudio que es gratuito <http://camstudio.es/>

VIII. PROYECTO DE VISUAL STUDIO QUE CONTENGA LOS ALGORITMOS DE MINERÍA DE DATOS

IX. RESPALDO DE LA BASE DE DATOS ORIGEN PARA APLICAR LOS ALGORITMOS DE MINERÍA DE DATOS

X. ARCHIVO .SQL CON LAS CONSULTAS DMX CON HALLAZGOS DE MINERÍA DE DATOS

XI. ARCHIVO PPT CON LA PRESENTACIÓN DE SU PROYECTO CON UNA DURACIÓN MÁXIMA DE 10 MINUTOS

10.2 FASES DE REVISIÓN DEL PROYECTO

Fase	Fecha	Comentarios	Semana	Puntos
	6 de septiembre	Entrega de la base de datos fuente y los requerimientos mínimos	4	
	20 de septiembre	Revisión del modelo dimensional por cubo	6	
	27 de septiembre	Revisión de cada cubo	7	
	11 de octubre	Revisión del documento (entregable I)	9	
FASE 1	Jueves 18 de octubre	Entrega para calificación de los entregables I al VII	10	25
	25 de octubre	Revisión del tablero de control	11	
	1 de noviembre	Revisión de las variables y tablas de entrada	12	
	8 de noviembre	Revisión de la interpretación de cada algoritmo	13	
	15 de noviembre	Revisión de las consultas MDX y DMX	14	
	22 de noviembre	Revisión de la presentación (entregable XI)	15	
FASE 2	Jueves 29 de noviembre	Entrega para calificación de los entregables I al IX	16	25
				50