



1. INFORMACIÓN DEL CURSO¹

Denominación: Gestión de la Información en Bases de Datos Corporativas (<i>Data Warehouse</i>)	Tipo: curso-taller	Nivel: Pregrado
Área de formación: Optativa Abierta	Modalidad: <input type="checkbox"/> Escolarizada <input type="checkbox"/> No escolarizada <input checked="" type="checkbox"/> Mixta	Prerrequisitos: IC548 Desarrollo de Bases de Datos para Web
Horas totales: 80 horas Horas teoría: 40 Horas práctica: 40	Créditos: 8	Clave: 129437
Elaboró: Centro Universitario de los Valles (Mtro. Jorge Ernesto Castillo Rizo) Revisó: Academia de Ciencias de la Computación		Fecha de elaboración: Julio del 2019 Fecha de actualización: Febrero del 2021

Relación con el perfil de egreso

La presente unidad de aprendizaje aporta al perfil de egreso del licenciado en tecnologías de la información el desarrollo de *habilidades para aplicar tecnologías en la gestión de la información, así como el análisis de datos para la toma de decisiones y resolución de problemas en las organizaciones.*

Aporta al perfil de egreso la capacidad *para la identificación de soluciones en la integración y tratamiento de datos, la gestión y el análisis de la información como propuestas de solución en las organizaciones con grandes volúmenes de datos.*

Proporciona al estudiante un conjunto de habilidades y herramientas de software para la construcción de soluciones organizacionales, la integración y análisis de datos.

Relación con el plan de estudios

Dentro del plan de estudios de la carrera, la materia de Gestión de la Información en Bases de Datos Corporativas (*Data Warehouse*) se clasifica en el área de formación **Optativa abierta**, se imparte durante el octavo semestre y tienen como prerrequisito el haber cursado la materia de *Bases de Datos Distribuidas*.

La materia se considera fundamental para la Orientación "Pre-especializante en Gestión y desarrollo de bases de datos" y Orientación "Pre-especializante en Inteligencia de Negocios (BI)" dentro del área de formación especializante selectiva.

Se sugiere previamente haber cursado materias de Estructuras de Datos, Bases de datos relacionados, Bases de datos distribuidas.

Campo de aplicación profesional de los conocimientos que promueve el desarrollo de la unidad de Aprendizaje

El estudiante será capaz de aplicar los conocimientos y habilidades adquiridas en el diseño y la construcción de soluciones de análisis de datos para la integración, interpretación y toma de decisiones en las organizaciones con grandes volúmenes de información.

¹ Este formato se trabajó con base en los términos de referencia del artículo 21 del Reglamento General de Planes de Estudio de la Universidad de Guadalajara.



2. DESCRIPCIÓN

Objetivo general del curso

El estudiante será capaz de desarrollar un Almacén de Datos de acumulación simple, desde su modelado, hasta los procesos de ETL e implementación de un Data Warehouse a través de la metodología Kimball y el uso de herramientas de software disponibles en el mercado para llevar a cabo el procesamiento analítico en línea (OLAP) o la minería de datos (data mining) que facilitan la toma de decisiones estratégicas en las organizaciones.

Objetivos parciales o específicos

Al término de la unidad 1, el estudiante deberá identificar los elementos que conforman un almacén de datos, sus características y generalidades para interpretar el modelo de un almacén de datos mediante los recursos didácticos proporcionados por el asesor.

Al término de la unidad 1, el estudiante deberá de distinguir entre bases de datos operacionales (OLTP) y bases de datos para la toma de decisiones (OLAP) así como describir la arquitectura de un DW para su modelado haciendo uso de los recursos didácticos proporcionados por el asesor.

Al término de la unidad 2, el estudiante será capaz de transformar un modelo de BD relacional en un modelo dimensional, así como identificar las tablas de dimensiones (atributos) y hechos (medidas) para generar almacenes de datos utilizando el modelo multidimensional de la teoría de respuesta al ítem jerárquico.

Al término de la unidad 2, el estudiante será capaz de distinguir los diferentes esquemas de los almacenes de datos para el diseño y construcción eficiente del repositorio a través de la metodología y el modelado multidimensional.

Al término de la unidad 3, el estudiante deberá identificar e implementar las áreas de trabajo requeridas para el desarrollo del proceso de ETL, así como elaborar los flujos de datos y control para la extracción y poblamiento de datos a través de un gestor de bases de datos y los servicios de integración datos.

Al término de la unidad 3, el estudiante deberá implementar el proceso de extracción, limpieza, transformación y carga de datos (ETL) con el fin de componer el repositorio de datos final (Data Mart o Data Warehouse), así como definir un cubo de datos OLAP mediante las tecnologías y herramientas para la construcción de un almacén de datos.

Al término de la unidad 4, el estudiante será capaz de identificar e implementar el proceso analítico en línea sobre los cubos OLAP para el análisis y visualización de los datos utilizando las reglas de Codd y las herramientas para elaborar OLAP.

Al término de la unidad 4, el estudiante será capaz de reconocer técnicas y operaciones de minería de datos para el reconocimiento de patrones, así como la explotación de datos con el fin de extraer o descubrir conocimiento mediante las herramientas de minería de datos.



Contenido temático

Unidad 1. Introducción a los Almacenes de Datos (Data Warehouse)

- 6.1. Conceptos básicos de almacenes de Datos (Data Warehouse).
- 6.2. Características de un Data Warehouse.
 - 35.2.1. Los datos están orientados a temas
 - 35.2.2. Los datos están integrados
 - 35.2.3. Los datos son no volátiles
 - 35.2.4. Los datos varían con el tiempo
- 6.3. Beneficios de los almacenes de datos.
- 6.4. Dificultades de los almacenes de datos.
- 6.5. Comparación entre OLTP y Bases de datos para la toma de decisiones (OLAP).
- 6.6. Arquitectura de un Data Warehouse.
 - 35.6.1. Fuentes de datos
 - 35.6.2. Integración de datos
 - 35.6.3. Repositorio de datos
 - 35.6.4. Análisis
- 6.7. Tipos de Data Warehouse.
- 6.8. Casos de éxito en el uso de bases de datos para la toma de decisiones.

Unidad 2. Análisis y Diseño de Almacenes de Datos (Modelo Dimensional)

- 6.1. Diseño de la base de datos para un almacén de datos.
- 6.2. Conceptos de Modelación Dimensional.
 - 35.2.1. Tabla de hechos
 - 35.2.2. Dimensiones del modelo de datos
 - 3.2.2.1. Elementos de la dimensión
 - 3.2.2.2. Atributos de dimensión
 - 3.2.2.3. Tabla de dimensiones
 - 35.2.3. Dimensión de tiempo
 - 35.2.4. Granularidad
 - 35.2.5. Jerarquías
- 6.3. Diseño del esquema de un Almacén de Datos.
 - 35.3.1. Esquema de estrella
 - 35.3.2. Esquema de copo de nieve
 - 35.3.3. Esquema de copo de estrella
- 6.4. Metodología de diseño de base de datos para almacenes de datos.
- 6.5. Estrategias de almacenamiento para el modelo dimensional : Relacional vs Multidimensional
- 6.6. Modelos dimensionales aplicados a casos de estudio.
 - 35.6.1. Data warehouse de General Hardware Company
 - 35.6.2. Data warehouse de Good Reading Bookstore
 - 35.6.3. Data warehouse de Lucky Rent-A-Car
- 6.7. Ejemplo de un cubo de datos multidimensional



Unidad 3. Construcción de un Almacén de Datos (Extracción Transformación y Carga ETL)

- 6.1. Áreas en la construcción de un almacén de datos
 - 45.1.1. Fuentes de datos
 - 45.1.2. Área de Staging
 - 45.1.3. Área de Data Warehouse
- 6.2. Implementación de flujos en un almacén de datos
 - 45.2.1. Flujos de datos
 - 45.2.2. Flujos de control
- 6.3. Proceso General de Construcción de un data warehouse (ETL)
 - 45.3.1. Extracción de datos
 - 45.3.2. Limpieza de datos
 - 45.3.3. Transformación de datos
 - 45.3.4. Carga de datos
- 6.4. Diseño del diccionario de datos para apoyar el proceso de ETL
- 6.5. Cubos OLAP (OnLine Analytical Processing)
- 6.6. Tecnologías y herramientas para la construcción de almacenes de datos.

Unidad 4. Funcionalidades de un Almacén de Datos

- 6.1. Procesamiento analítico en línea (OLAP)
 - 55.1.1. Conceptos de Análisis Multidimensional.
 - 55.1.2. Estructuras de Datos Multidimensionales
 - 55.1.3. Dispersidad
 - 55.1.4. Reglas de Codd para las herramientas OLAP
 - 55.1.5. Servidores OLAP
 - 5.1.5.1. OLAP Multidimensional (MOLAP)
 - 5.1.5.2. OLAP Relacional (ROLAP)
 - 55.1.6. Entornos de Consultas Gestionadas (MQE)
 - 55.1.7. 4.6 Construcción del cubo de datos por MOLAP.
 - 55.1.8. 4.7 Visualización de datos.
 - 55.1.9. 4.8 Tecnologías y herramientas para OLAP
- 6.2. Minería o explotación de datos (data mining)
 - 55.2.1. Técnicas u operaciones de minería de datos
 - 5.2.1.1. Modelado predictivo
 - 5.2.1.2. Segmentación de Base de Datos
 - 5.2.1.3. Análisis de Vínculos
 - 5.2.1.4. Detección de Desviaciones
 - 55.2.2. El proceso de la minería de datos
 - 55.2.3. Herramientas de Minería de Datos

Estructura conceptual del curso

Data Warehouse

Gestión de la INFORMACIÓN

en bases de datos corporativas

01 MODELADO DE ALMACÉN DE DATOS
Diseño de BD para el Almacén de datos

- Modelado dimensional
- Esquema tipo estrella, copo nieve y constelación
- Tablas de dimensiones, hechos

02 PROCESO DE ETL
Extracción, transformación y carga de datos

- Integración de datos
- Área de staging / Data Warehouse
- Flujos de datos / control

03 ALMACÉN DE DATOS
Consolidación de repositorio de datos

- Repositorio
- Orientado al negocio, integrada, variable en el tiempo, no volátil
- Data set

04 ANÁLISIS Y REPORTES
Generación de conocimiento, presentación de la información

- Reportes
- Procesamiento analítico en línea (OLAP)
- Minería de datos
- Dashboard

Mtro. Jorge Ernesto Castillo Rizo

<https://view.genial.ly/5f362ee0367d4e0d78fc65af/vertical-infographic-timeline-almacenes-de-datos>



Modalidad de evaluación

Instrumento de evaluación	Factor de ponderación
Convencional: Examen teórico -práctico	20%
Desempeño: Prácticas áulicas	30%
Desempeño: Prácticas extra áulicas	20%
Desempeño: Autoevaluación estudiante	5%
Desempeño: Autoevaluación asesor	5%
Desempeño: Proyecto integrador	20%
Total	100%

Elementos del desarrollo de la unidad de aprendizaje (asignatura)

Conocimientos	El estudiante obtendrá los conocimientos sobre: Conceptos y términos clave de los Almacenes de Datos Características de las bases de datos multidimensionales Arquitectura de un almacén de datos Conceptos de modelado multidimensional Diseño de esquemas de un almacén de datos Metodologías de diseño Flujos en un almacén de datos Proceso de ETL Proceso analítico en línea (OLAP) Minería de datos
Habilidades y Destrezas	El estudiante, desarrollará las habilidades y destrezas en: Distinguir bases de datos OLTP y OLAP Identificar tablas de hechos, dimensiones, granularidad y jerarquías Modelar una base de datos multidimensional Aplicar metodología de diseño de almacén de datos Manejo de herramientas tecnológicas para la construcción de un almacén de datos Implementar flujos en un almacén de datos Realizar procesos de ETL (Extracción, Limpieza-Transformación y Carga de datos) Construir cubos OLAP Generar reportes de datos Realizar proceso de minería de datos
Valores y Actitudes	Respeto Tolerancia Compromiso Trabajo colaborativo Proactivo Interés Autoaprendizaje Honestidad, puntualidad, responsabilidad, ética, veracidad, servicio, respeto, equidad, disposición.



3. BIBLIOGRAFÍA BÁSICA

Autor(es)	Título	Editorial	Año	URL o biblioteca digital donde está disponible (en su caso)
Ralph Kimball & Margy Ross	The Data Warehouse Toolkit	Wiley	2013, 3th Edition	http://www.essai.rnu.tn/Ebook/Informatique/The%20Data%20Warehouse%20Toolkit,%203rd%20Edition.pdf
Ralph Kimball, Margy Ross, Warren Thornthwaite & Etal	The Data Warehouse Lifecycle Toolkit	Wiley	2011, 2th Edition	https://es.slideshare.net/nirajbharambe/the-data-warehouse-lifecycle-toolkit
Inmon, William H.	Building the data warehouse	Wiley	2002, 3th Edition	http://fit.hcmute.edu.vn/Resources/Docs/SubDomain/fit/ThayTuan/DataWH/Bulding%20the%20Data%20Warehouse%204%20Edition.pdf
*Curto Díaz, Josep	¿Cómo crear un data warehouse?	Barcelona UOC	2015, 1ra Edición	B.C. CU Valles

4. BIBLIOGRAFÍA COMPLEMENTARIA

Autor(es)	Título	Editorial	Año	URL o biblioteca digital donde está disponible (en su caso)
Williamson, Ben	Big data en educación : el futuro digital del aprendizaje, la política y la práctica	Morata	2018	B.C. CU Valles
López Murphy, Juan José	La ingeniería del Big Data : cómo trabajar con datos	UOC	2017	B.C. CU Valles
Paul Beynon-Davies	Sistemas de Bases de Datos	Reverte	2014	B.C. CU Valles
*Ramez Elmasri, Shamkant B. Navathe	Fundamentos de Sistemas de Bases de Datos	Pearson Addison Wesley	2007	B.C. CU Valles
*Thomas M. Connolly /Carolyn E. Begg	Sistemas de Bases de Datos	Pearson Addison Wesley	2005, 4ta Edición	B.C. CU Valles
*Mark L. Gillenson,	Administración de Bases de Batos	Limusa Willey	2006, 1ra Edición	B.C. CU Valles
*Michel V. Mannino	Administración de Bases de Batos	Mc GrawHill	2007, 3ra Edición	B.C. CU Valles



Enlaces web:

Libros:

<http://www.essai.rnu.tn/Ebook/Informatique/The%20Data%20Warehouse%20Toolkit,%203rd%20Edition.pdf>
<https://es.slideshare.net/nirajbharambe/the-data-warehouse-lifecycle-toolkit>
<http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/books/data-warehouse-dw-lifecycle-toolkit/>
<http://fit.hcmute.edu.vn/Resources/Docs/SubDomain/fit/ThayTuan/DataWH/Bulding%20the%20Data%20Warehouse%204%20Edition.pdf>

General:

https://mva.microsoft.com/en-us/training-courses/implementing-a-data-warehouse-with-sql-server-jump-start-8257?l=KtvLHdKy_7204984382
[https://technet.microsoft.com/es-es/library/hh231699\(v=sql.110\).aspx](https://technet.microsoft.com/es-es/library/hh231699(v=sql.110).aspx)
http://docs.oracle.com/cd/B10500_01/server.920/a96520/toc.htm
<http://wiki.pentaho.com/display/EAl/es/Manual+del+Usuario+de+Spoon>
https://www.ibm.com/support/knowledgecenter/es/SSGU8G_11.50.0/com.ibm.datawhsenode.doc/datawhse.htm

Unidad 1:

<https://es.slideshare.net/rigardo871/almacen-de-datos-12145837>
<https://es.slideshare.net/rimenri/data-warehouse-introduccion>
<https://www.youtube.com/watch?v=jFsRdTcljeU>

Unidad 2:

<https://es.slideshare.net/ecastrom/modelado-de-data-warehouse>
<http://inf.udec.cl/~revista/ediciones/edicion4/modmulti.PDF>
<http://carlosproal.com/dw/dw04.html>
<https://www.youtube.com/watch?v=x2keCD2lCuk&list=PLG1t8jaLbxA8cOqm9zBxbXeGjlgbrnl7a&index=8>

Unidad 3:

<http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/288859/procesos-etl-extracci-n-transformaci-n-carga>
<http://carlosproal.com/dw/dw05.html>
<https://www.youtube.com/watch?v=DlnmRyACbd4&list=PLG1t8jaLbxA8cOqm9zBxbXeGjlgbrnl7a&index=10>
<https://www.youtube.com/watch?v=9AkVz2x0az4>

Unidad 4:

<https://es.slideshare.net/ashir233/olap-56094588>
<https://es.slideshare.net/WalidElbadawy/olap-on-line-analytical-processing>
https://www.youtube.com/watch?v=MD-qd4Kdu_E&list=PLG1t8jaLbxA8cOqm9zBxbXeGjlgbrnl7a&index=15
<https://www.youtube.com/watch?v=d9cEN5pDMC0&index=16&list=PLG1t8jaLbxA8cOqm9zBxbXeGjlgbrnl7a>
<https://es.slideshare.net/keopx/minera-de-datos-38280245>



Perfil del profesor:

Formación académica: ingeniero en sistemas de información, ingeniero en computación, licenciado en tecnologías de la información o ciencias a fines, preferentemente contar una maestría en tecnologías de la información o gestión de la información. El profesor deberá de contar con experiencia en el manejo de bases de datos, sistemas de información, gestión de la información, almacenes de datos, modelado multidimensional, procesos de ETL (extracción, transformación y carga), sistemas OLAP, inteligencia de negocios, minería de datos, data lakes, big data. Debe de dominar el uso de herramientas de software para el tratamiento y análisis de los datos.

Creativo en el proceso de aprendizaje, Diseñar estrategias de aprendizaje y evaluación orientadas al desarrollo de competencias, responsable, íntegro, organizado y coherente.