

**UNIVERSIDAD NACIONAL DE ASUNCIÓN**  
**FACULTAD POLITÉCNICA**  
**PROGRAMA DE ESTUDIO**  
**INGENIERIA EN INFORMÁTICA**

Consejo Directivo - Resolución 13/05/18-00 Acta 876/11/03/2013 Anexo 07

**I. - IDENTIFICACIÓN**

1. Materia : Electiva 7 – Data Warehouse II Especialización en ETL y capas de Presentación
2. Semestre : Décimo
3. Horas semanales : 7 horas
  - 3.1. Clases teóricas: 7 horas
  - 3.2. Clases prácticas: 0 horas
4. Total real de horas disponibles: 112 horas
  - 4.1. Clases teóricas: 98 horas
  - 4.2. Clases prácticas: 14 horas

**II. - OBJETIVO**

El principal objetivo de esta asignatura es capacitar al estudiante para reconocer el potencial y la problemática del análisis de sistemas de información para la toma de decisiones y especializarse en las distintas etapas del ETL (*Extract, Transform and Load*), control de calidad de datos y análisis y modelado de datos. Así como conocer las distintas capas de presentación de datos. Conocer las tendencias del Data warehouse.

**III. - PRE-REQUISITO**

El alumno debe haber aprobado el 80% de los créditos de la carrera y 300 horas de pasantía para cursar la Electiva 7 de acuerdo al Plan 2008 de la Carrera de Ingeniería en Informática.

**IV. - CONTENIDO**

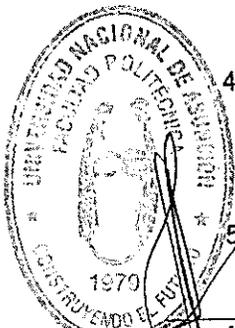
***Workflow del ETL***

***Tema 1: Extrayendo.***

1. Mapeo Lógico del Dato, diseñando el mapeo lógico de dato y construyendo el mapeo.
2. El desafío de extraer datos de diferentes fuentes.
3. Trabajando con Flat Files, XML Source, Weblog Source, y ERP system.
4. Extrayendo datos cambiados, extrayendo cambios, tipos de extracción, detectando eliminación o sobre escritura de los hechos en la fuente.

***Tema 2: Limpiando e integrando***

1. Definiendo de la calidad de Datos
  - 1.1 Diseñando Objetivos, entendiendo como se constituyen las keys.
  - 1.2 Balanceando prioridad de conflictos y definiendo políticas.
2. Limpieza de Datos entregable definición de la política del dato entregable.
  - 2.1 Error de la tabla de hechos y error en la dimensión.
  - 2.2 Auditoría de las dimensión y ETL.
3. Proceso y sus mediciones.
  - 3.1 Detección de anomalías.
  - 3.2 Tipos de ejecución.
  - 3.3 Datos y valores reglas de ejecución.
  - 3.4 Estructura de la ejecución.
  - 3.5 Revisión del Flujo del proceso.
4. Conformando los datos entregables.
  - 4.1 Conocer la cantidad de filas.
  - 4.2 Columnas nulas.
  - 4.3 Columnas numéricas y rango de fechas.
  - 4.4 Restricciones del tamaño de las columnas.
5. Entregables Conformadas.
  - 5.1 Diseñando dimensiones conformadas y hechos conformados.



### Tema 3: Entregando Tablas de Hechos

1. Estructura básica y clave subrogada.
2. Granos fundamentales: Análisis y desarrollo de distintos ETL.
  - 2.1 Tablas de hechos de transacciones.
  - 2.2 Tablas de hechos de *snapshot* periódicos.
  - 2.3 Tablas de hechos acumulativos.
3. Preparación para carga de tablas de hechos.
  - 3.1 Manejo de Índices.
  - 3.2 Particiones.
  - 3.3 Log de *rollback*.
  - 3.4 Carga de Datos.
  - 3.5 Carga incremental.
  - 3.6 Inserción de Datos.
  - 3.7 Actualización y corrección de datos.
  - 3.8 Hechos negativos.
  - 3.9 Borrado de Hechos.
  - 3.10 Recolección del beneficio en distintas monedas.
  - 3.11 Hechos que arriban tarde.
4. Agregaciones
  - 4.1 Distintos diseños y requerimientos de agregaciones.
  - 4.2 Administración de requerimientos
  - 4.3 Inclusión de vistas materializadas.
5. Entregado de dimensiones a cubos OLAP (*OnLine Analytical Processing*).
  - 5.1 Fuentes de los Cubo.
  - 5.2 Procesamiento de dimensiones.
  - 5.3 Cambios en las dimensiones.
  - 5.4 Proceso de hechos.
  - 5.5 Interacción entre el proceso OLAP y el ETL.

### Tema 5: Implementación y Operaciones

1. El mercado actual de ETL.
2. Lenguajes de *script* actuales.
3. El tiempo es la esencia: ordenando datos antes de leerlo, extrayendo un conjunto de datos, y creando extracciones agregadas.
4. Usando *bulk loader utility* para acelerar los *insert*.
5. Manejo de las característica de las base de datos para mejorar el desempeño.
  - 5.1 El orden de los objetos.
  - 5.2 El efecto de las agregaciones.
  - 5.3 El impacto en el desempeño usando funciones escalar.
  - 5.4 Evitando *Triggers*.
  - 5.5 Superando el Cuello de botella ODBC (*Open Data Base Connectivity*).
  - 5.6 Beneficio de procesamiento paralelo.
  - 5.7 Aumentando la partición de ETL.
6. Reduciendo Entradas/Salidas.
  - 6.1 Eliminando lecturas/escrituras.
  - 6.2 Filtrado de datos.
  - 6.3 Particionado y paralelizado.
  - 6.4 Bulkloading/Eliminando Logging.
  - 6.5 Borrado de restricciones e índices de Base de Datos.
  - 6.6 Eliminación del ruido del tráfico.

### Tema 6: Herramientas de BI y Capa de Presentación del Data Warehouse

1. Distintas herramientas de BI (*Business Intelligence*) orientados a distintos niveles de usuario.
2. ScoreCard.
3. Dashboard.
4. Key Performance Indicator - KPI.
5. OLAP.
6. Aplicaciones analíticas.



7. Dataware.
8. Reportes Ad-Hoc.

**Tema 7: Distintos Casos de Estudios**

1. Order Transaction.
2. Snapshot Acumulativo para Órdenes.
3. Customer Relationship Management - CRM.
4. Accounting.
5. Administración de Recursos Humanos.
6. Servicios Financieros, etc.

**V. - ESTRATEGIAS METODOLÓGICAS.**

1. Exposición oral.
2. Laboratorio.
3. Formación de grupos y trabajo en equipo para resolver problemas durante las clases.
4. Trabajos prácticos.

**VI. - MEDIOS AUXILIARES**

1. Pizarra.
2. Proyector Multimedia.
3. Sala de máquinas para clases prácticas.
4. Bibliografía de apoyo.

**VII. - EVALUACIÓN**

La evaluación se realizará de acuerdo a las reglamentaciones de la Facultad y la Carrera.

**VIII. - BIBLIOGRAFÍA**

**Bibliografía Básica**

1. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning by Ralph Kimball and Joe Caserta, 2004.

**Bibliografía Complementaria**

1. The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses, Ralph Kimball. 2008.

