

# Temas Selectos de Base de Datos

---

**Materia: Temas Selectos de Bases de datos**

## **OBJETIVO DEL CURSO**

El alumno explicará las tendencias de las bases de datos y los sistemas de información en la industria

## **PROGRAMA DEL CURSO**

### **1 Calidad en Bases de Datos**

**Objetivo:** El alumno distinguirá los diferentes indicadores de calidad en los datos dentro de los sistemas de información.

#### 1.1 Introducción

1.1.1 Causas de problemas de calidad de bases de datos

1.1.2 Calidad de Información

Dimensiones de Calidad 1.1.3

Calidad de Datos

Dimensiones de Calidad

#### 1.2 Marcos de trabajo para la mejora de Calidad

1.2.1 Objetivos

1.2.2 Medición

1.2.3 Evaluación

1.2.4 Limpieza

#### 1.3 Pasos esenciales en un Proyecto empresarial de calidad de datos

1.3.1 Necesidades del negocio

1.3.2 Análisis de la información

1.3.3 Evaluación de calidad de datos

1.3.4 Evaluación del impacto en el negocio

1.3.5 Identificación de causas de mala calidad

1.3.6 Plan de mejora

1.3.7 Prevención de errores

1.3.8 Corrección de errores

1.3.9 Implementar controles

1.3.10 Comunicar acciones

#### 1.4 Talend Open Studio for Data Quality

##### 1.4.1 Prácticas

### **2 Algoritmos de de-duplicación de datos**

**Objetivo:** El alumno distinguirá, comparará y programará algunos algoritmos para la de-duplicación de registros

#### 2.1 Introducción

#### 2.2 El proceso de correspondencia de datos

2.2.1 Ejemplo de correspondencia de datos

2.2.2 Pre procesamiento de datos-Codificación

2.2.2.1 Eliminación de caracteres y tokens

2.2.2.2 Estandarización y tokenización

2.2.2.3 Segmentación en campos de salida

2.2.2.4 Verificación

2.2.3 Indexado

2.2.3.1 Definición de llaves para bloqueaje

2.2.3.2 Soundex

# Syllabus de Temas Selectos de Base de Datos

---

- 2.2.3.3 Phonex
- 2.2.3.4 Phonix
- 2.2.3.5 Nysiis
- 2.2.3.6 Double metaphone
- 2.2.3.7 Fuzzy Soundex
- 2.2.4 Comparación de pares de registros
  - 2.2.4.1 Justificación
  - 2.2.4.2 Comparación exacta, por truncamiento y por codificación
  - 2.2.4.3 Comparacion de cadenas
  - 2.2.4.4 Otras técnicas de comparación de cadenas
  - 2.2.4.5 Comparación numérica
  - 2.2.4.6 Comparación de fechas, edades y tiempo
  - 2.2.4.7 Comparación de distancias geográficas
  - 2.2.4.8 GEOGRAPHICAL DISTANCE COMPARISON
  - 2.2.4.9 Comparación de registros
- 2.2.5 Clasificación
  - 2.2.5.1 Clasificación basada en umbrales
  - 2.2.5.2 CLASIFICACIÓN PROBABILISTICA.
  - 2.2.5.3 Clasificación basada en costos
  - 2.2.5.4 Clasificación basada en reglas
  - 2.2.5.5 Clasificación basada en métodos supervisados
  - 2.2.5.6 Clasificación colectiva
- 2.2.6 Evaluación de calidad y complejidad en la correspondencia de datos
  - 2.2.6.1 Medición de calidad de la correspondencia de datos
  - 2.2.6.2 Medición de la complejidad de la correspondencia de datos
- 2.2.7 Uso e Implementación de algunos algoritmos en FEBRL

## **3 Introducción a Manejo de Datos Maestros**

- 3.1 Conceptos básicos
- 3.2 Arquitectura
- 3.3 Estilos de Implementación de MDM
- 3.4 Talend MDM
- 3.5 Prácticas

## **4 Introducción a Bases de datos en memoria**

- 4.1 Conceptos básicos
- 4.2 Arquitectura
- 4.3 Instalación