



PROMiDAT

IBEROAMERICANO

Programa Iberoamericano de
Formación en Minería de Datos

Manipulación y Preparación de Datos



(506) 2268.8823 - (506) 8708.9091



info@promidat.com



facebook.com/oldemarrodriguez



www.promidat.com

Tutor: El curso será impartido por el ing. Diego Jiménez.

Duración: Cuatro semanas.

Descripción:



En este curso se presentarán los fundamentos del lenguaje SQL. El énfasis principal del curso será examinar diversos componentes del lenguaje, como lo son declaraciones, expresiones, tipos de datos, entre otros. Se le dará especial importancia al uso del lenguaje como herramienta de análisis exploratorio de datos, como punto de partida para el desarrollo de aplicaciones de minería de datos. Para esto se utilizarán diversos motores de bases de datos como *SQL Server* (utilizando T-SQL) y *MySQL*, así como paquetes en R para manipulación de datos.

Objetivos:

En este curso el estudiante será capaz de:

1. Entender la estructura básica del lenguaje SQL como herramienta para consultar bases de datos.
2. Utilizar el lenguaje como mecanismo de extracción de datos e información a partir de repositorios con grandes volúmenes de datos.
3. Hacer uso correcto del lenguaje para construir consultas complejas que permitan obtener información de distintas tablas simultáneamente.
4. Entender el lenguaje SQL desde el punto de vista de teoría de conjuntos y lógica de predicados, permitiendo realizar operaciones usuales como lo son uniones, intersecciones, diferencias, entre otros.
5. Utilizar *SQL Server* y *MySQL* como motores de bases de datos basados en SQL.
6. Importar información de un Administrador de base de datos SQL a R.
7. Exportar resultados obtenidos en R a un Administrador de base de datos SQL.

Metodología:

Basado en la teoría y en la aplicación directa de los conceptos aprendidos. Para esto se dispondrán de las siguientes herramientas.

- Una vídeo conferencia semanal, las cuales quedarán grabadas en Zoom, para que los alumnos la puedan acceder en cualquier momento.
- Trabajos prácticos semanales.
- Foros para plantear dudas al tutor y compañeros.
- Aula virtual en Moodle.

Luego de este curso el estudiante será capaz de:

Desarrollar proyectos de Minería de Datos que involucren alta manipulación de datos utilizando el lenguaje SQL.

Contenido:

1- **SQL:** fundamentos de bases de datos

- a. ¿Qué es SQL?
- b. Bases de datos, tablas, columnas y filas.
- c. Diagramas Entidad-Relación.
- d. Sentencias SQL.
- e. MySQL.
- f. Sintaxis SQL (MySQL).

2- **Consultas SQL:** Exploración de tablas de datos.

- a. Explorar tablas completas o subconjuntos (sentencia **SELECT**).
- b. Derivación/Creación de nuevas columnas.
- c. Ordenamiento de datos (**ORDER BY**).
- d. Concatenar conjuntos de datos (**UNION**).
- e. Cómo unir una o varias tablas (**JOIN**).
- f. Agrupación de individuos (**GROUP BY**).
- g. Funciones sobre grupos (**COUNT, MAX, MIN**, etc.).

3- **R y Bases de Datos:** acceder a bases de datos directamente desde R

- a. Conexión a Bases de datos (DBI y RODBC)
- b. Base de datos SQLite
- c. Creación de estructuras (data frames) en R.
- d. Almacenar datos en la Base de Datos.
- e. Paquete dbplyr (Uso de sintaxis dplyr a una base de datos).

4- Manipulación de datos en R: uso de paquetes dplyr y tidyr.

- a. ¿Qué es Data Wrangling?
- b. tidyverse
- c. Uso de dplyr y tidyr para limpieza, procesamiento y manipulación de datos en R
 - a. Filtros
 - b. Agregaciones
 - c. Transformaciones
- d. Uso de pipes (%>%) en R.

Bibliografía:

1. Alexander M. Decker J. and Wehbe B. "Business Intelligence Tools for Excel Analysts". Wiley, 2014.
2. Beaulieu A. "Learning SQL", O'Reilly, 2009.
3. Ben-Gan I. "SQL Server 2012 T-SQL Fundamentals". O'Reilly, 2012.
4. Date C.J. "SQL and Relational Theory: How to Write Accurate SQL Code". O'Reilly, 2012.
5. DuBois P. "MySQL Developer's Library". Addison-Wesley, 2013.
6. Harrington J.L. "SQL Clearly Explained". Morgan Kaufmann, 2010.
7. Linoff G. "Data Analysis Using SQL and Excel". Wiley, 2008.
8. Mistry R, Misner S. "Introduction Microsoft SQL Server 2008 R2". Microsoft Press, 2010.
9. Seyed M.M, "Saied" T. and Hugh E.W. "Learning MySQL", O'Reilly, 2007.
10. Williams, G. "Data Mining with Rattle and R", Springer, 2011.

11. Winston W. "Microsoft Excel 2013 Data Analysis and Business Modeling", O'Reilly, 2014.
12. Zapawa T. "Excel® Advanced Report Development". Wiley, 2005.