MAESTRÍA EN ESTADÍSTICA APLICADA/ MANEJO DE DATOS

**Taller Unidad 3: Recopilación, integración y manipulación y almacenamiento de datos.**

**Nombres de participantes Grupo4:**

| **Mariuxi Quimis** | **Ricardo Mora** |  |
| --- | --- | --- |
| **Maricela Carpio** | **Carola Flores** |  |

**Instrucciones del taller:**

* El taller es una actividad que se realiza en un grupal, luego de los aportes y desarrollo, cada integrante debe subir la resolución a la plataforma Canvas.
* Se desarrolla en Rstudio.
* En la resolución de cada tema, se debe presentar el código y los resultados generados. En caso de no estar alguno de los elementos se penalizará con el 50% de la puntuación del tema.

**Temas del Taller:**

**Desde MySQL**

1. Explore la Base de Datos Demo de MySQL Sakila, sobre la renta de DVD (2005), los datos se organizan desde las perspectivas de Customers, Movie, Rent.

```{r}

#install.packages("RODBC") #instalación de paquete RODBC

library(RODBC) #referencia a la librería RODBC

con<-odbcConnect("MySQLconexion2", uid = "root", pwd="12345678") #Establecer la conexión con el DBMS

```

```{r warning=FALSE}

library(sqldf)

df\_customer <-sqlQuery(con,"Select \* from customer") #Ejecutar una sentencia SQL

df\_film <-sqlQuery(con,"Select \* from film")

df\_rental <-sqlQuery(con,"Select \* from rental")

df\_payment <-sqlQuery(con,"Select \* from payment")

df\_inventory <-sqlQuery(con,"Select \* from inventory")

```

Diagrama

Descripción generada automáticamente con confianza baja

Fuente de imagen: MySQL

Recomendación: Comprenda la estructura de las tablas Customer (Cliente), rental (alquiler), payment (pago por el alquiler) , inventory (inventario o disponibilidad) y film (DVD) sus conexiones y datos.

**Desde Rstudio:**

1. Establezca una conexión a la base de datos Sakira en MySQL, luego con las funciones de Rstudio cargue los datos de las tablas customer, rental y payment, los datos son almacenados en dataframe. Muestre para cada dataframe su estructura de campo.

Utilizando la función de sqldf:

1. Determine el listado de clientes con estado inactivo (igual a cero) que están vinculados al local 2 (store\_id). Muestre las primeras 10 filas.

df\_customer<-sqlQuery(con,"Select \* from customer")

df\_t3<-sqldf("SELECT \* FROM df\_customer AS c WHERE c.active= 0 AND c.store\_id=2 LIMIT 10")

df\_t3

1. Determine todos los films cuyo título inicie con “A” y con descripción que contenga la palabra “DRAMA”. Muestre las primeras 10 filas.

df\_t4<-sqldf("SELECT \* FROM df\_film AS f WHERE f.title LIKE 'A%' AND f.description LIKE '%DRAMA%' LIMIT 10")

df\_t4

1. Determine todo el historial de rentas y pagos del cliente NATALIE MEYER (nombre apellido). El reporte debe tener los siguientes campos: CustomerID, first\_name, last\_name, rental\_id, rental\_date, payment\_id, payment\_date, amount. Muestre las primeras 10 filas.

df\_t5 <- sqldf ("SELECT A.customer\_id,

A.first\_name,

A.last\_name,

B.rental\_id,

B.rental\_date,

C.payment\_id,

C.payment\_date,

C.amount

FROM df\_customer A, df\_rental B, df\_payment C

WHERE A.customer\_id = B.customer\_id

AND B.rental\_id = c.rental\_id

AND A.first\_name = 'NATALIE' AND A.last\_name = 'MEYER' "

)

head(df\_t5,10)

1. Considerando el siguiente listado de títulos de filmes ('TRADING PINOCCHIO', 'VANILLA DAY', 'WORKER TARZAN', 'ACADEMY DINOSAUR'), determine todos los alquileres vinculados a algunos de estos filmes. El reporte debe tener los campos de rent y film (title, description, release\_year). Ordene por filmes de forma alfabética. Muestre las primeras 10 filas.
2. Determine todos los alquileres realizados en el mes de junio del 2005 con valores de pago mayor a USD 5. El reporte debe presentar, el nombre del cliente y título del film. Muestre las primeras 10 filas.

*Sugerencia: en el DML puede utilizar las funciones de year (fecha) y month(fecha)*

*df\_t7 <-sqlQuery(con,"SELECT \**

*FROM df\_rental A, df\_payment B*

*WHERE A.customer\_id = B.customer\_id*

*AND create\_date BETWEEN '2005-07-01' AND '2005-07-31'*

*AND amount > 5 ")*

head(df\_t7,10)

1. Determine todos los alquileres realizados en el segundo trimestre del 2005, que fueron solicitados por clientes vinculados al local 2 (store\_id). Muestre las primeras 10 filas.