MAESTRÍA EN ESTADÍSTICA APLICADA/ MANEJO DE DATOS

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

**Nombres de participantes Grupo5:**

| **Guillermo Quijije** | **Luis Moncayo** | **Enrique Arias Galan** |
| --- | --- | --- |
| **Marcelo Chávez** | **Iris Castillo Plaza** |  |

**Instrucciones del taller:**

* El taller es una actividad que se realiza en un grupo, 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.

library(RODBC)

library(sqldf)

con<-odbcConnect("sakila", uid = "root")

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

store<-sqlQuery(con,"Select \* from store")

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

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

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

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

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.

str(customer)

str(rental)

str(payment)

str(inventory)

str(film)

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.

r3<-sqldf("SELECT store\_id,active

FROM customer c

WHERE store\_id = 2 AND active = 0

")

head(r3,10)

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

r4<-sqldf("SELECT title

,description

FROM film

WHERE title LIKE 'A%'

AND description LIKE '%drama%'

")

head(r4,10)

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.

r5<-sqldf("SELECT c.customer\_id

,first\_name

,last\_name

,r.rental\_id

,rental\_date

,payment\_id

,payment\_date

,amount

FROM customer c

INNER JOIN payment p ON c.customer\_id = p.customer\_id

INNER JOIN rental r ON r.rental\_id = p.rental\_id

WHERE first\_name = 'NATALIE' AND last\_name = 'MEYER'

"

)

head(r5,10)

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.

acti\_6 <- sqldf("SELECT r.rental\_id, f.title, f.description, f.release\_year

FROM rental r

INNER JOIN inventory i ON r.inventory\_id = i.inventory\_id

INNER JOIN film f ON i.film\_id = f.film\_id

WHERE f.title IN ('TRADING PINOCCHIO', 'VANILLA DAY', 'WORKER TARZAN', 'ACADEMY DINOSAUR')

ORDER BY f.title")

head(acti\_6,10)

1. 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)*

*r7<-sqlQuery(con,"SELECT first\_name*

*,last\_name*

*,title*

*,amount*

*,year(rental\_date) as AÑO*

*,month(rental\_date) as MES*

*FROM rental r*

*INNER JOIN customer c on c.customer\_id =r.customer\_id*

*INNER JOIN payment p ON p.rental\_id = r.rental\_id*

*INNER JOIN inventory i ON i.inventory\_id = r.inventory\_id*

*INNER JOIN film f on f.film\_id = i.film\_id*

*WHERE YEAR(rental\_date) = 2005 AND MONTH(rental\_date) = 6 AND amount >5")*

*head(r7,10)*

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.

r8<-sqlQuery(con,"SELECT rental\_id,rental\_date,year(rental\_date) as AÑO, month(rental\_date) as MONTH, store\_id FROM rental r INNER JOIN customer c on c.customer\_id =r.customer\_id WHERE YEAR(rental\_date) = 2005 AND MONTH(rental\_date) BETWEEN 4 AND 6 AND store\_id = 2")

head(r8,10)