MAESTRÍA EN ESTADÍSTICA APLICADA/ MANEJO DE DATOS

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

**Nombres de participantes Grupo5:**

| **Wilver Santana Alvarado** | **Marcos Guaman** |  |
| --- | --- | --- |
| Sonia Tigua Moreira | **Kevin Acosta Barreno** |  |

**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.

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.

library(sqldf)

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

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

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

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

df\_category<-sqlQuery(con,"Select \* from category")

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

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\_result3 <-sqldf("SELECT \* FROM df\_customer WHERE active = 0 AND store\_id = 2")

head(df\_result3,10)

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

df\_result4 <-sqldf("SELECT \* FROM df\_film WHERE title LIKE 'A%'AND film\_id IN (SELECT film\_id FROM df\_film\_category WHERE category\_id IN (SELECT category\_id FROM df\_category WHERE name = 'Drama'))", connection=NULL)

head(df\_result4,10)

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.

df\_t5<-sqldf("SELECT

c.customer\_id, c.first\_name, c.last\_name,

r.rental\_id, r.rental\_date, p.payment\_id,p.payment\_date,p.amount

FROM df\_customer AS c

INNER JOIN df\_payment AS p ON c.customer\_id=p.customer\_id

INNER JOIN df\_rental AS r ON p.rental\_id=r.rental\_id

WHERE c.first\_name='NATALIE' AND c.last\_name ='MEYER'")

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.

SELECT f.title,f.description,f.release\_year

from sakila.film AS f INNER JOIN sakila.inventory AS i ON f.film\_id=i.film\_id

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

ORDER BY f.title limit 10

query <- "SELECT r.\*, f.title, f.description, f.release\_year

FROM rental AS r

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

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

WHERE f.title IN (SELECT title FROM titles)

ORDER BY f.title

LIMIT 10"

result

result

<- sqldf(query)

print(result)

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.

query <- "SELECT c.first\_name || ' ' || c.last\_name AS customer\_name, f.title

FROM rental AS r

JOIN payment AS p ON r.rental\_id = p.rental\_id

JOIN customer AS c ON r.customer\_id = c.customer\_id

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

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

WHERE YEAR(r.rental\_date) = 2005 AND MONTH(r.rental\_date) = 6 AND p.amount > 5

LIMIT 10"

result <- sqldf(query)

print(result)

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

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.

query <- "SELECT r.\*, c.customer\_id, c.first\_name, c.last\_name

FROM rental AS r

JOIN customer AS c ON r.customer\_id = c.customer\_id

WHERE QUARTER(r.rental\_date) = 2 AND r.store\_id = 2

LIMIT 10"

result <- sqldf(query)

print(result)