MAESTRÍA EN ESTADÍSTICA APLICADA/ MANEJO DE DATOS

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

**Nombres de participantes Grupo3:**

| **EMILY ALCÍVAR TOALA** | **WLADIMIR MONTIEL RIVERA** | **GEORGE ACOSTA CHONG** |
| --- | --- | --- |
| **CRISTIAN TOMALÁ MAZZINI** |  |  |

**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

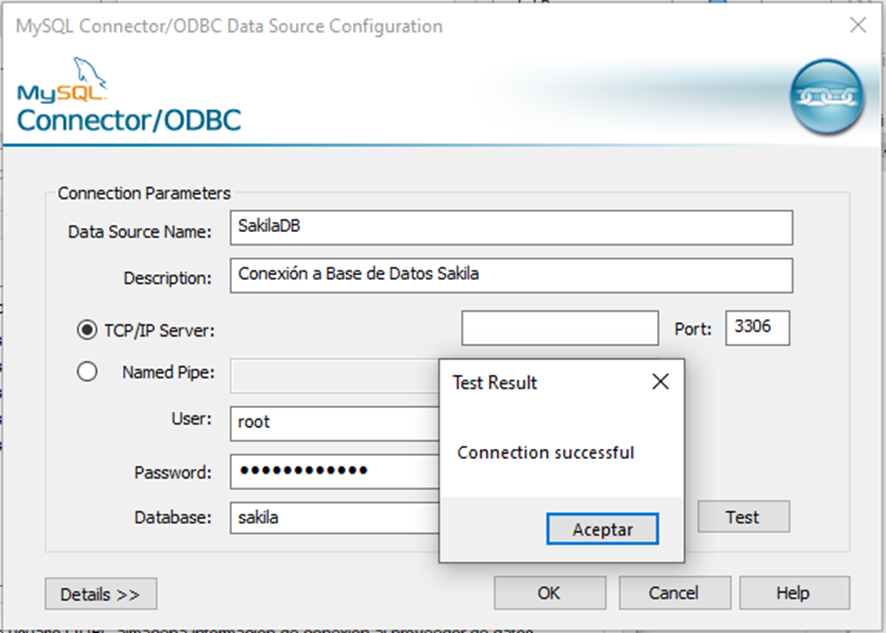
Descripción generada automáticamente con confianza baja

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.

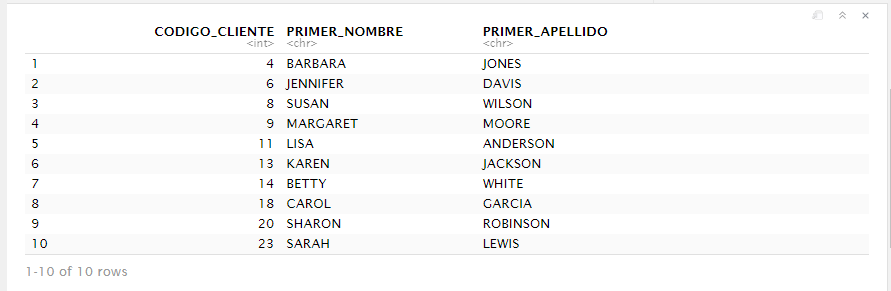


| #install.packages("RODBC")  library(RODBC)  con<-odbcConnect("SakilaDB", uid = "root", pwd="2808\_Ct\_1982")  df\_customer<-sqlQuery(con,"Select \* from customer")  df\_rental<-sqlQuery(con,"Select \* from rental")  df\_payment<-sqlQuery(con,"Select \* from payment") |
| --- |

Utilizando la función de sqldf:

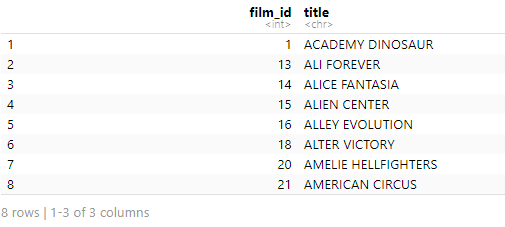
1. Determine el listado de clientes con estado inactivo (igual a cero) que están vinculados al local 2 (store\_id). Muestre las primeras 10 filas.

| library(sqldf)  df\_clientes<-sqlQuery(con,"SELECT \* FROM CUSTOMER") df\_locales<-sqlQuery(con,"SELECT \* FROM STORE")  df\_res3<-sqldf("SELECT CLI.customer\_id CODIGO\_CLIENTE, CLI.first\_name PRIMER\_NOMBRE, CLI.last\_name PRIMER\_APELLIDO FROM df\_clientes CLI, df\_locales LOC WHERE CLI.store\_id = LOC.store\_id AND CLI.ACTIVE = 1 AND LOC.store\_id = 2", connection=NULL)  head(df\_res3, 10) |
| --- |



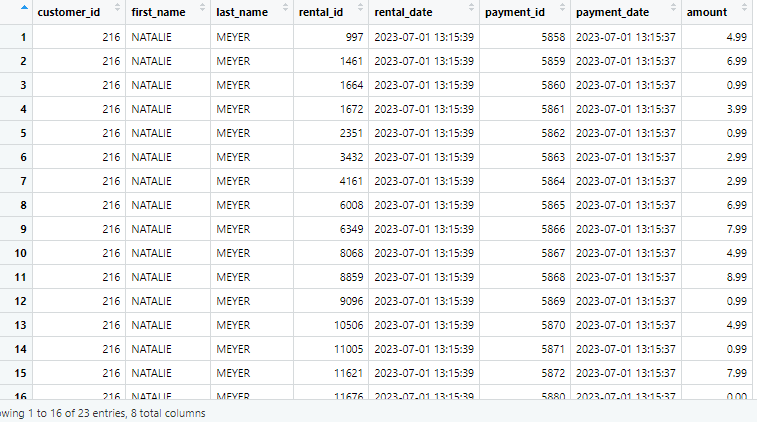
1. Determine todos los films cuyo título inicie con “A” y con descripción que contenga la palabra “DRAMA”. Muestre las primeras 10 filas.

| library(sqldf)  df\_film<-sqlQuery(con, "Select\* from film")  df\_drama<-sqldf("SELECT f.film\_id, f.title, f.description FROM df\_film f  WHERE f.title LIKE 'A%' AND f.description LIKE '%DRAMA%' ")  head(df\_drama,10) |
| --- |



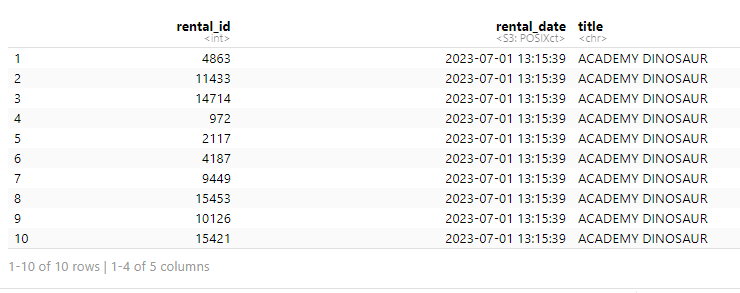
1. Determine toto el historial de rentas y pagos del cliente NATALIE MEYER (nombre apellido). El reporte debe tener los siguientes campos: CustomerID, first\_name, last\_name, rental\_id, rental\_date, payment\_id, payment\_date, amount. Muestre las primeras 10 filas.

| df\_customer<-sqlQuery(con,"Select \* from customer")  df\_rental<-sqlQuery(con,"Select \* from rental")  df\_payment<-sqlQuery(con,"Select \* from payment")  preg5<-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' ") |
| --- |



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.

| preg6<-sqldf("SELECT C.rental\_id, C.rental\_date, B.title, B.description, B.release\_year  FROM df\_rental C, df\_inventory A, df\_film B  WHERE C.inventory\_id=A.inventory\_id  AND A.film\_id=B.film\_id  AND B.title IN ('TRADING PINOCCHIO', 'VANILLA DAY', 'WORKER TARZAN', 'ACADEMY DINOSAUR')  ORDER BY title ")  head(preg6,10) |
| --- |



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)*

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.