Paul Polsinelli

IFT 410 - 22485

4/4/2021

# Lab 10: BD Storage with Databricks

### **Part II: Creating a Database using Databricks**

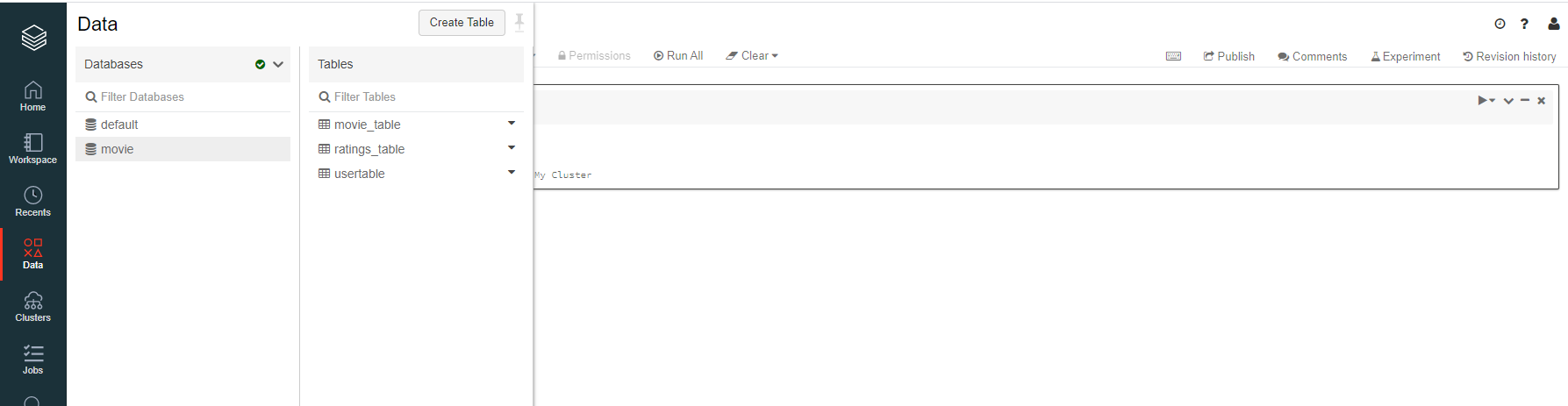
Create database Movie;

Create table Movie.UserTable (UserID INT, UserName STRING, Password STRING, Start\_Date DATE);

Create table Movie.Movie\_Table (MovieID INT, Title String, Genre STRING, Year DOUBLE, CastCrew STRING, Director STRING);

Create table Movie.Ratings\_Table (UserID INT, MovieID INT, Rating DOUBLE, TimeStamp TIMESTAMP);





SET time\_zone='+00:00';

insert into Movie.Ratings\_table values (1,1,1,'2021-01-01 01:00:01');

insert into Movie.Ratings\_table values (2,1,2,'2020-01-01 02:00:01');

insert into Movie.Ratings\_table values (3,1,3,'2019-01-01 03:00:01');

insert into Movie.Ratings\_table values (4,2,4,'2018-01-01 04:00:01');

insert into Movie.Ratings\_table values (5,2,5,'2017-01-01 05:00:01');

insert into Movie.Ratings\_table values (6,2,1,'2016-01-01 06:00:01');

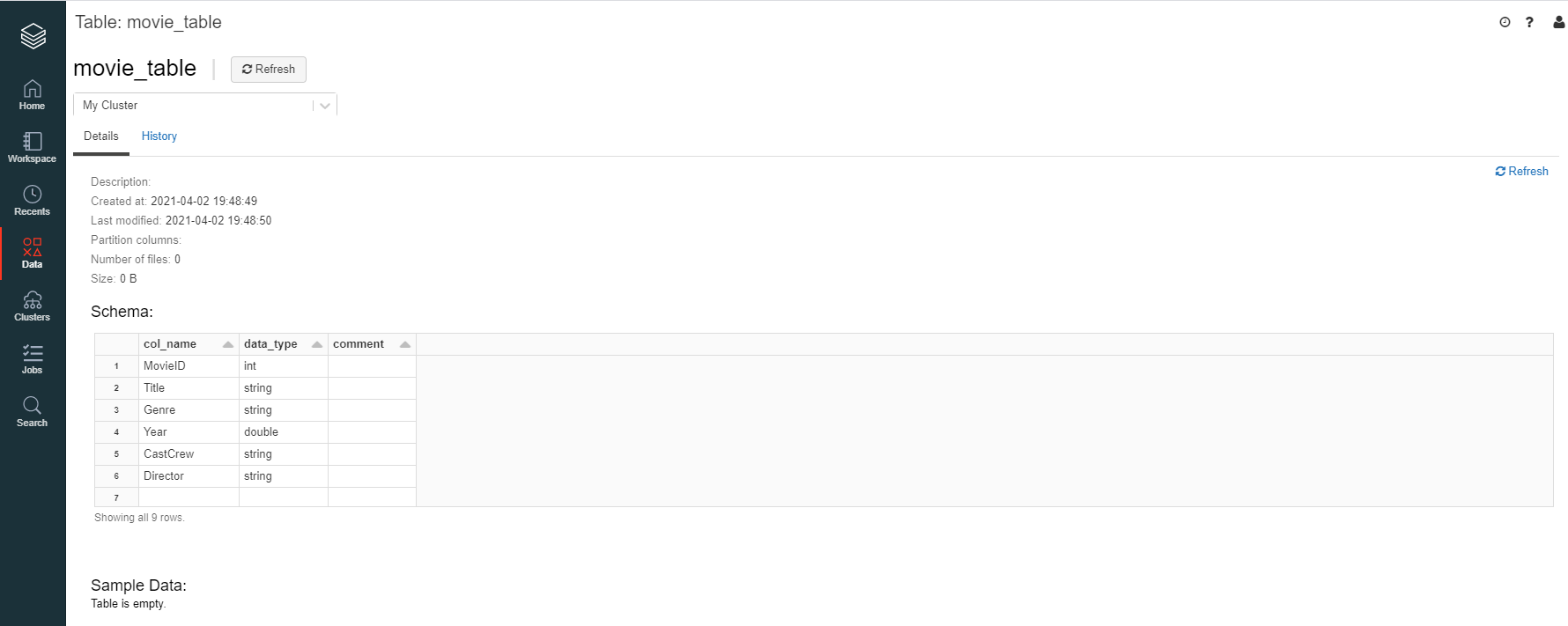
insert into Movie.Ratings\_table values (7,3,2,'2015-01-01 07:00:01');

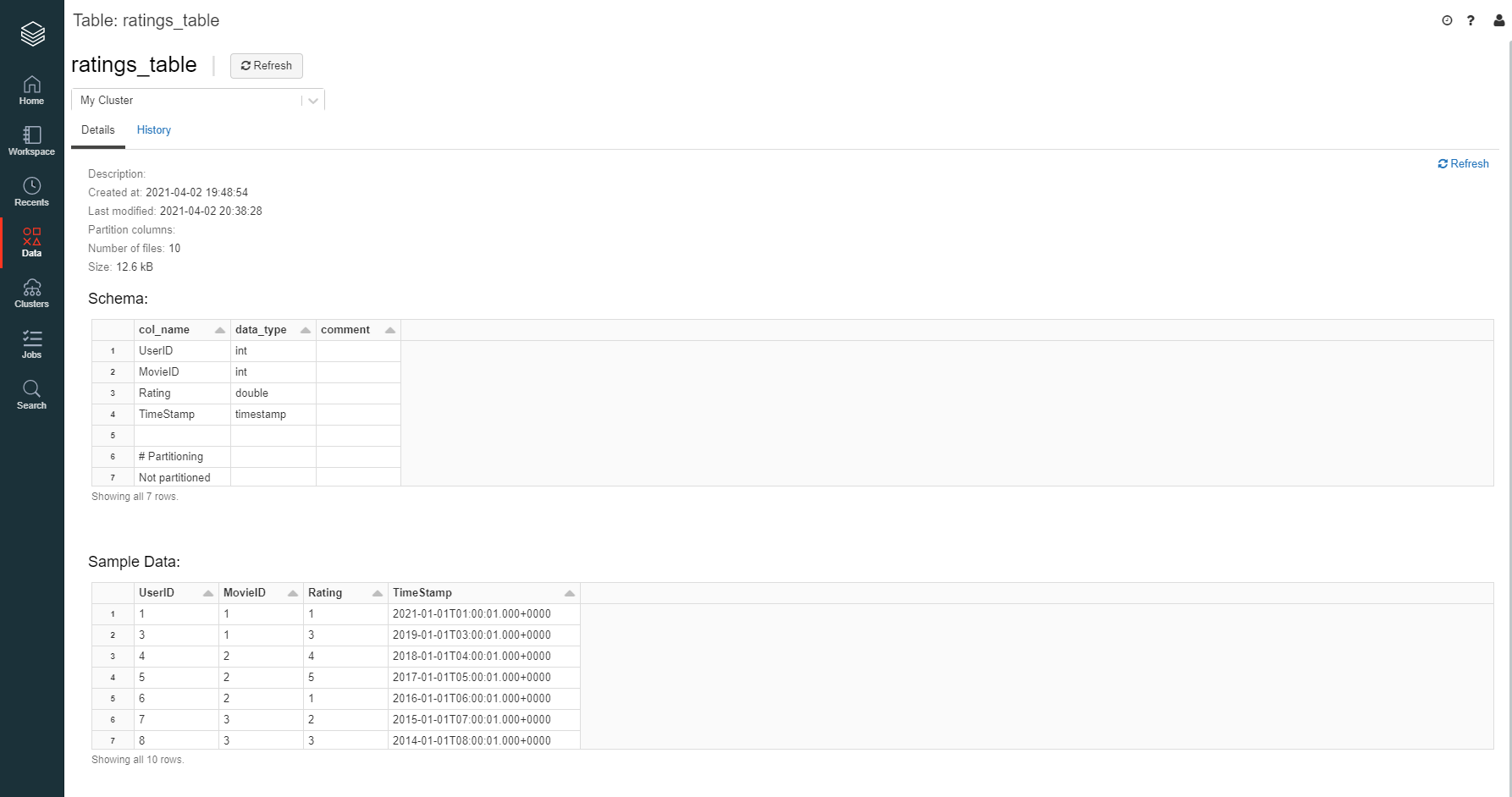
insert into Movie.Ratings\_table values (8,3,3,'2014-01-01 08:00:01');

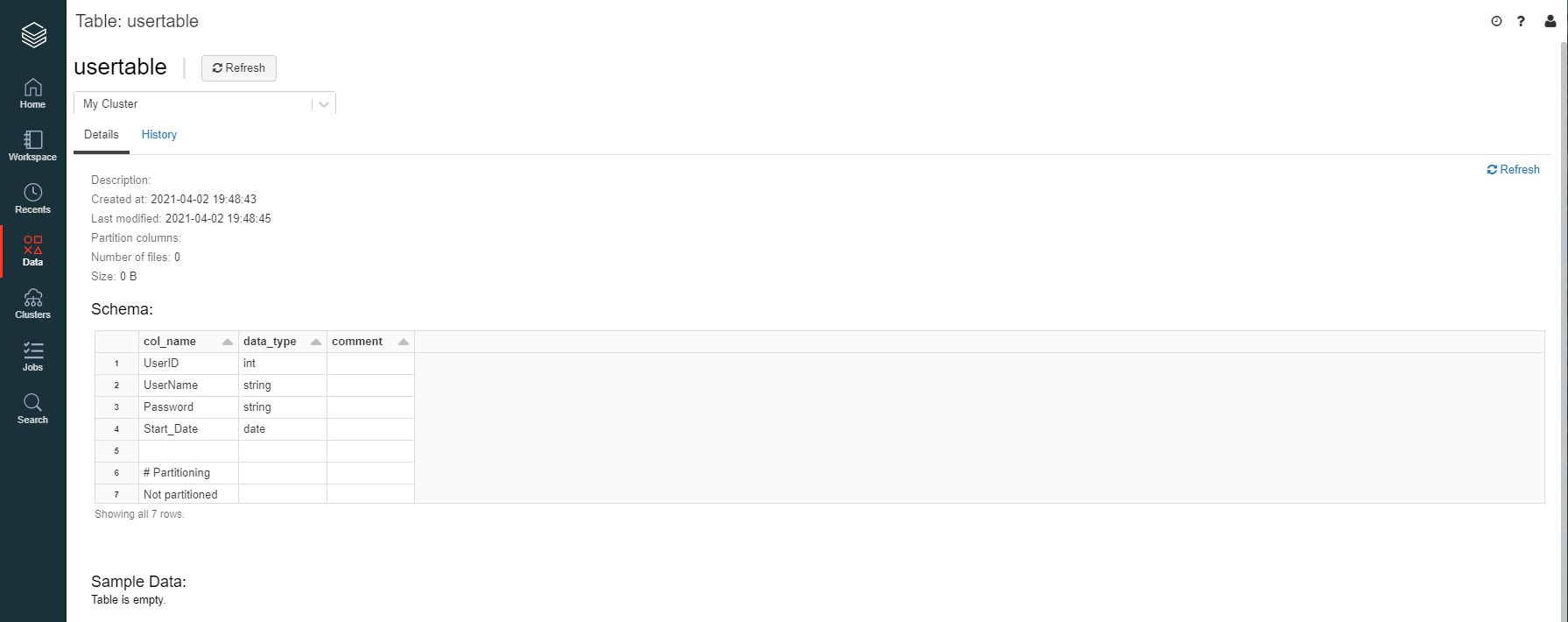
insert into Movie.Ratings\_table values (9,3,4,'2013-01-01 09:00:01');

insert into Movie.Ratings\_table values (10,3,5,'2012-01-01 09:00:01');





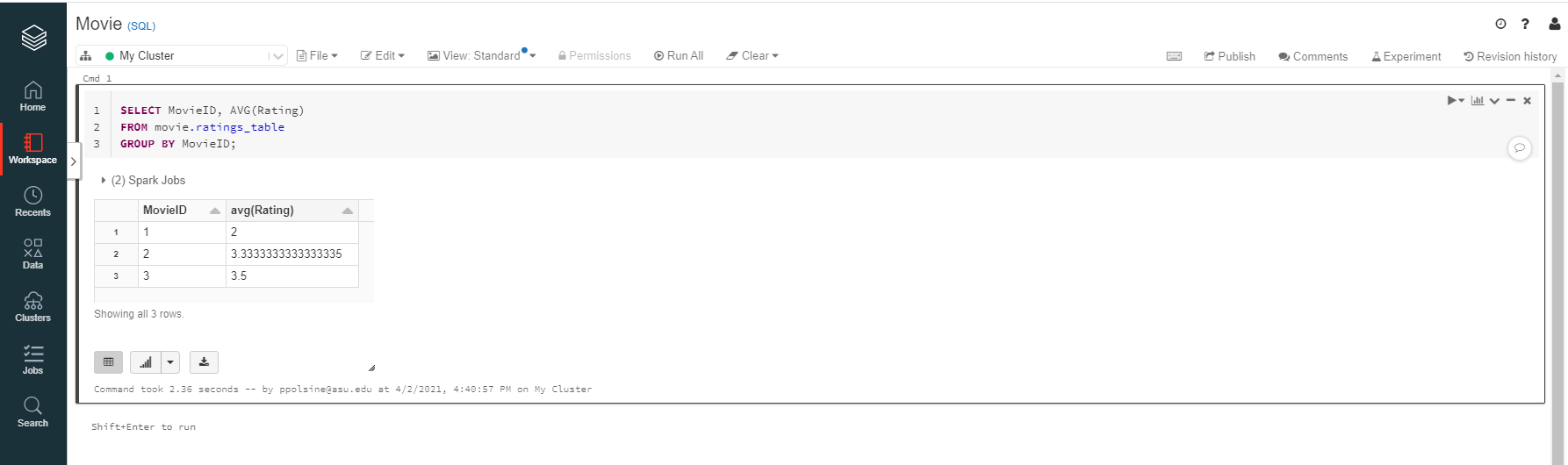




SELECT MovieID, AVG(Rating)

FROM movie.ratings\_table

GROUP BY MovieID;



### **Part III: Querying Data Stored in a CSV File**

### This query shows all the movieID’s that userID 4 gave a rating of 3

val sqlDF = spark.sql("SELECT \_c1 FROM csv.`/FileStore/tables/ratings.csv` where \_c0=4 AND \_c2=3")

sqlDF.show()

