

Práctica 1

Objetivo: Analizar las consultas enfocándose en la capacidad de la realización de consultas, almacenaje y memoria principal.

Para las siguientes tablas

- Movie_Ratings: Es una tabla de Películas que fueron evaluadas por usuarios, dándoles un puntaje del 1 al 5.
- MovieLens Datasets: Es un datasets que contiene 943 usuarios que califican a 1682 películas.

Archivo u.data contiene los siguientes atributos:

user id | item id | rating | timestamp

a) Analizar el contenido del datasets y consultas en memoria.

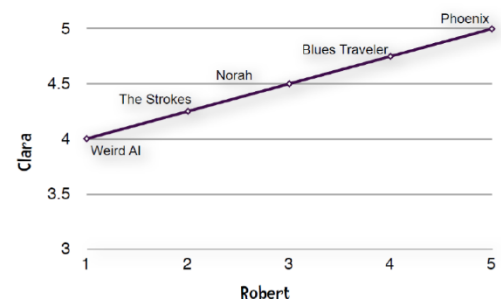
- Movie_Ratings:
 1. ¿Cuántas películas fueron calificadas por Thomas?
 2. ¿Cuántas películas fueron calificadas por Matt?
 3. ¿Qué película fue la más calificada por los usuarios?, ¿cuántos usuarios la calificaron? Disregue por número de usuarios que la calificaron con un determinado puntaje.
 4. ¿Qué película fue la menos calificaron?, ¿cuántos usuarios la calificaron? Disregue por número de usuarios que la calificaron con un determinado puntaje.
- MovieLens:
 5. ¿Qué película fue la más calificada por los usuarios?, ¿cuántos usuarios la calificaron? Disregue por número de usuarios que la calificaron con un determinado puntaje.
 6. ¿Qué película fue la menos calificaron?, ¿cuántos usuarios la calificaron? Disregue por número de usuarios que la calificaron con un determinado puntaje.

b) Analizar lectura/escritura en el disco y almacenamiento en memoria

Coeficiente de Correlación de Pearson

	Blues Traveler	Norah Jones	Phoenix	The Strokes	Weird Al
Clara	4.75	4.5	5	4.25	4
Robert	4	3	5	2	1

Usuarios tienen diferentes escalas



Coeficiente de Correlación de Pearson

- Aproximación de Pearson:

$$r = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} \sqrt{\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}}}$$

Aproximación de Pearson

$$r = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} \sqrt{\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}}}$$

	Blues Traveler	Norah Jones	Phoenix	The Strokes	Weird Al
Clara	4.75	4.5	5	4.25	4
Robert	4	3	5	2	1

$$\sum_{i=1}^n x_i y_i = (4.75 \times 4) + (4.5 \times 3) + (5 \times 5) + (4.25 \times 2) + (4 \times 1) = 70$$

$$\frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} \begin{cases} \nearrow \sum_{i=1}^n x_i = 4.75 + 4.5 + 5 + 4.25 + 4 = 22.5 \\ \searrow \sum_{i=1}^n y_i = 4 + 3 + 5 + 2 + 1 = 15 \end{cases}$$

$$= \frac{22.5 \times 15}{5} = 67.5$$

$$\sum_{i=1}^n x_i^2 = (4.75)^2 + (4.5)^2 + (5)^2 + (4.25)^2 + (4)^2 = 101.875$$

$$\frac{(\sum_{i=1}^n x_i)^2}{n} = \frac{22.5^2}{5} = 101.25$$

$$\sqrt{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} = \sqrt{101.875 - 101.25} = \sqrt{.625} = .79057$$

$$\sqrt{\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}} = \sqrt{55 - 45} = 3.162277$$

$$r = \frac{2.5}{.79057(3.162277)} = \frac{2.5}{2.5} = 1.00$$

- Calcular la Aproximación de Pearson para:

Movie_Ratings

Para los usuarios que obtuvieron como respuesta en la pregunta 3 y 4.

- MovieLens

Para los usuarios que obtuvieron como respuesta en la pregunta 5 y 6.

Para el datasets MovieLens

9. ¿Cuánto tiempo se demoró en realizar las consultas?
10. ¿Cuánto ocupó su RAM? Comparar antes de realizar las consultas y luego de realizarlas
11. Cómo mejoró los tiempos en almacenamiento.