Import:

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

Rename:

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

Merge:

A screenshot of a computer program

Description automatically generated

|  |  |
| --- | --- |
| 1. | **Compute the number of movies each actor was in. The output should have two columns: actor, count. The output should be ordered by the count in descending order** |
|  | A white background with black dots  Description automatically generated |
| 2. | **Compute the highest-rated movie per year and include all the actors played in that movie. The output should have only one movie per year, and it should contain four columns: year, movie title, rating, a semicolon-separated list of actor names. This question requires a join between *movies.tsv* and *movie-ratings.tsv* files. There are two approaches to this problem. The first is to figure out the highest-rated movies per year and then join with a list of actors. The second one is to perform the join first and then figure out the highest-rated movies per year and a list of actors. The result of each approach is different from the other one. Why do you think that is?** |
|  | **First Approach**      **Second approach**      The reason why the result is different between the two approaches is because in the first approach, you find the highest-rated movies per year first and then join with the list of actors, which might include multiple actors. In the second approach, you join the two datasets first, which results in a single row for each movie-actor combination, and then you find the highest-rated movie per year from the joined data. Compared to the second method where we combined the two tables first, this method automatically filter out all the movies that are only exclusive to only one table, so that when we calculate the maximum rating, it results in a complete table when we print it out. |
| 3. | **Determine which pair of actors worked together most. Working together is defined as appearing in the same movie. The output should have three columns: actor1, actor2, and count. The output should be sorted by the count in descending order. The solution to this question requires doing self-join.** |

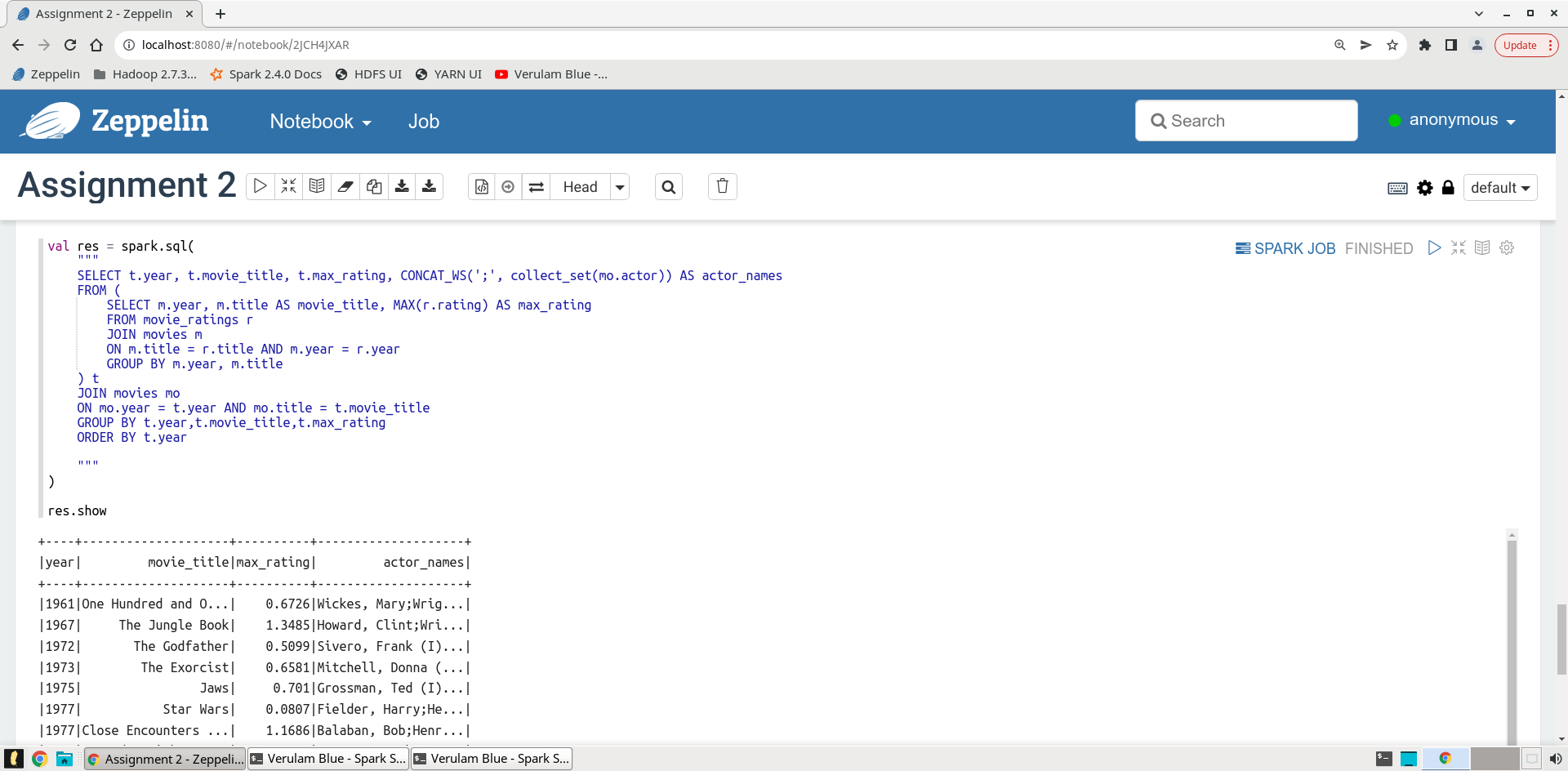
**Nomor 1**

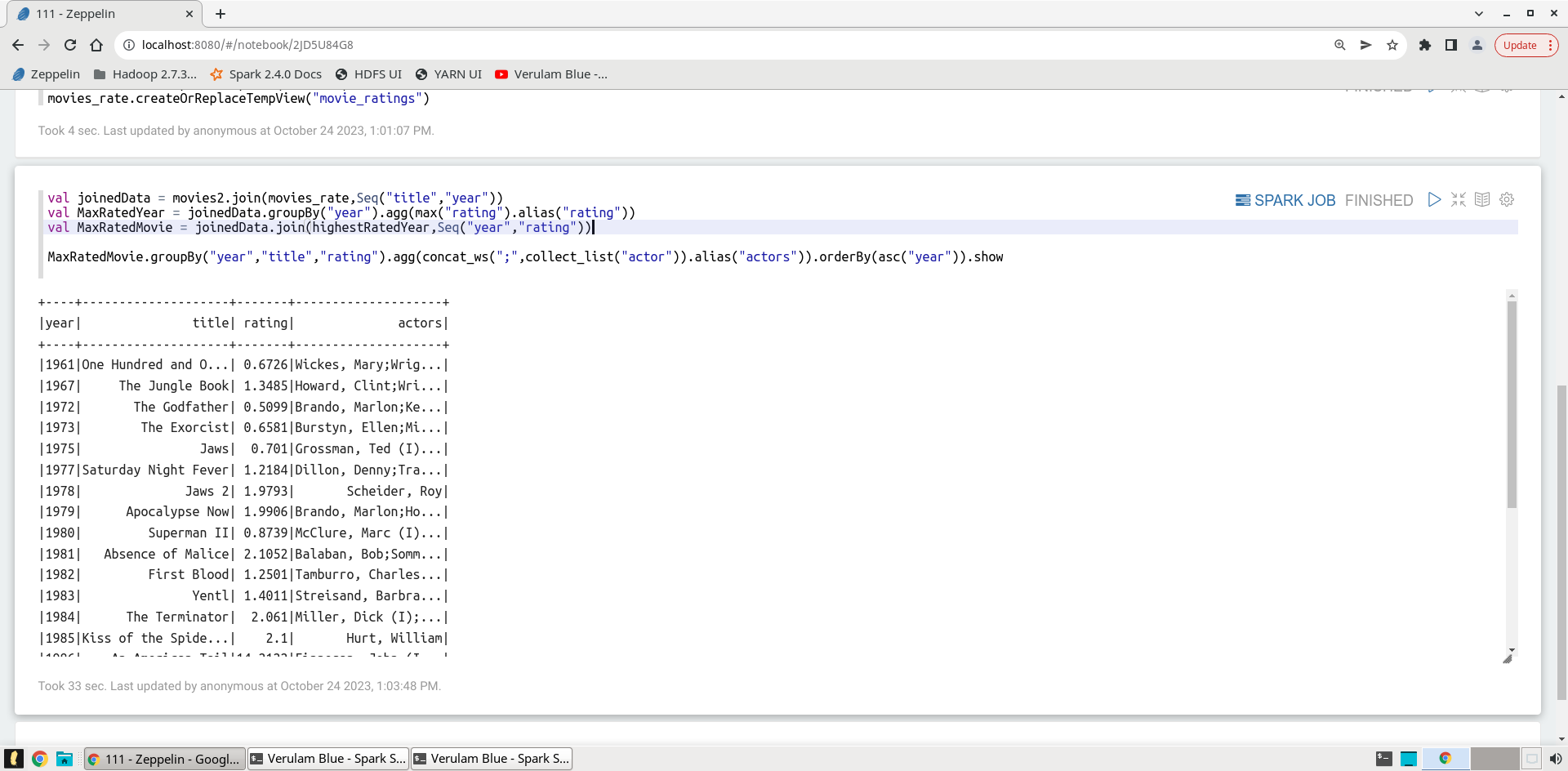
Sebuah gambar berisi teks, software, Laman internet, nomor

Deskripsi dibuat secara otomatis

Sebuah gambar berisi teks, software, Ikon komputer, Software multimedia

Deskripsi dibuat secara otomatis

**Nomor 2 Cara 1**

**Nomor 2 Cara 2**

Perbedaan **pendekatan 1** (figure max rating first join last) dan **pendekatan 2** (Join first figure max rating last) :

Perbedaan antara kedua pendekatan ini terletak pada urutan operasinya.

- Pendekatan 1 mengidentifikasi film dengan rating tertinggi per tahun sebelum bergabung dengan informasi aktornya, sedangkan Pendekatan 2 melakukan penggabungan terlebih dahulu dan kemudian menentukan film dengan rating tertinggi per tahun.

- Pendekatan 1 hanya mempertimbangkan film dengan rating tertinggi setiap tahunnya, terlepas dari apakah film tersebut memiliki aktor terkait dalam kumpulan datanya, sedangkan Pendekatan 2 yang melakukan join terlebih dahulu hanya mempertimbangkan film dengan aktor saat menentukan film dengan rating tertinggi setiap tahunnya

- Kesimpulannya, perbedaan hasil antara kedua pendekatan ini karena pendekatan 1 hanya berfokus pada rating film dan kemudian mencoba mencari aktor terkait, sedangkan pendekatan 2 menggabungkan informasi film dan rating dengan informasi aktor sebelum mengidentifikasi film dengan rating tertinggi per tahun

**Sebuah gambar berisi teks, software, Ikon komputer, Laman internet

Deskripsi dibuat secara otomatisNomor 3**