Session 7: EXPLORING APACHE PIG

Case Study Movie Rating

Task 1:

What are the movie titles that the user has rated?

Solution:

```
MovieTask1.pig ✖
movies = LOAD '/user/acadgild/hadoop/movies.csv' USING PigStorage(',') AS( movieid:int, title:CharArray, genres:CharArray);
ratings = LOAD '/user/acadgild/hadoop/ratings.csv' USING PigStorage(',') AS( userid:int, movid:int,
rating:CharArray,timestamp:int);
joinData = JOIN movies BY movieid,ratings BY movid;
movieTitle = FOREACh joinData GENERATE title;
result = DISTINCT movieTitle;
DUMP result;
```

Output:

```
The Edit View Search Terminal Help
(Films to Keep You Awake: The Baby's Room (Películas para no dormir: La habitación del niño) (2006))
(Amazon Jack 2: The Movie Star (a.k.a. Hugo the Movie Star) (Jungledyret 2 - den store filmhelt) (1996))
(Dragon Ball Z: Super Android 13! (Doragon bôru Z 7: Kyokugen batoru!! San dai sûpâ saiyajin) (1992))
(History of Postwar Japan as Told by a Bar Hostess (Nippon Sengoshi - Madamu onboro no Seikatsu) (1970))
(My Sex Life... or How I Got Into an Argument (Comment je me suis disputé... (ma vie sexuelle)) (1996))
(Texas Chainsaw Massacre: The Next Generation (a.k.a. The Return of the Texas Chainsaw Massacre) (1994))
(All the Boys Are Called Patrick (Charlotte et Veronique ou tous les garcons s'appellent Patrick) (1959))
(Christiane F. (a.k.a. We Children from Bahnhof Zoo) (Christiane F. - Wir Kinder vom Bahnhof Zoo) (1981))
(Gurren Lagann: The Lights in the Sky are Stars (Gekijô ban Tengen toppa guren ragan: Ragan hen) (2009))
(Madhouse (There Was a Little Girl) (And When She Was Bad) (Flesh and the Beast) (Scared to Death) (1981))
(Dragon Ball GT: A Hero's Legacy (Doragon bôru GT: Gokû gaiden! Yûki no akashi wa sû-shin-chû) (1997))
(Everything Happens to Me (a.k.a. Why Did You Pick on Me?) (Chissà perché... capitano tutte a me) (1980))
(Investigation of a Citizen Above Suspicion (Indagine su un cittadino al di sopra di ogni sospetto) (1970))
(Dragon Ball Z: Broly Second Coming (Doragon bôru Z 10: Kiken na futari! Sûpâ senshi wa nemurenai) (1994))
(Puppetry of the Penis: Live at the Forum (Puppetry of the Penis: The Ancient Art of Genital Origami) (2001))
(Life As a Fatal Sexually Transmitted Disease (Zycie jako smiertelna choroba przenoszona droga plciowa) (2000))
(Riuscirà l'avvocato Franco Benenato a sconfiggere il suo acerrimo nemico il pretore Ciccio De Ingras? (1971))
(Samurai III: Duel on Ganryu Island (a.k.a. Bushido) (Miyamoto Musashi kanketsuhen: kettô Ganryûjima) (1956))
(Victory in the Ukraine and the Expulsion of the Germans from the Boundaries of the Ukrainian Soviet Earth (1945))
(Dragon Ball Z: Wrath of the Dragon (Doragon bôru Z 13: Ryûken bakuhatsu!! Gokû ga yaraneba dare ga yaru) (1995))
(Manufacturing Reality: Slavoj Zizek and the Reality of the Virtual (Slavoj Zizek: The Reality of the Virtual) (2004))
(Dragon Ball Z: Broly - The Legendary Super Saiyan (Doragon bôru Z 8: Moetsukiro!! Nessen retsusen-chô gekisen) (1993))
(Fullmetal Alchemist the Movie: Conqueror of Shamballa (Gekijô-ban hagane no renkinjutsushi: Shanbara wo yuku mono) (2005))
(To Each His Own Cinema (Chacun son cinéma ou Ce petit coup au coeur quand la lumière s'éteint et que le film commence) (2007
))
(Dragon Ball Z: The History of Trunks (Doragon bôru Z: Zetsubô e no hankô!! Nokosareta chô senshi - Gohan to Torankusu) (1993
))
(More About the Children of Noisy Village (a.k.a. More About the Children of Bullerby Village) (Mer om oss barn i Bullerbyn)
(1987))
(Revolutionary Girl Utena: Adolescence of Utena (a.k.a. Revolutionary Girl Utena the Movie) (Shoujo kakumei Utena: Adolescenc
e mokushiroku) (1999))
(Dragon Ball Z: Bardock - The Father of Goku (Doragon bôru Z: Tatta hitori no saishuu kessen - Furiiza ni itonda Z senshi Kak
arotto no chichi) (1990))
(Dragon Ball Z the Movie: The World's Strongest (a.k.a. Dragon Ball Z: The Strongest Guy in The World) (Doragon bôru Z: Kono
yo de ichiban tsuyoi yatsu) (1990))
2018-12-19 14:08:01,387 [main] INFO  org.apache.pig.Main - Pig script completed in 15 minutes, 26 seconds and 767 millisecond
s (926767 ms)
You have new mail in /var/spool/mail/acadgild
```

Task 2:

How many times a movie has been rated by the user?

Solution:

```
MovieTask2.pig ✖
movies = LOAD '/user/acadgild/hadoop/movies.csv' USING PigStorage(',') AS( movieid:int, title:CharArray, genres:CharArray);
ratings = LOAD '/user/acadgild/hadoop/ratings.csv' USING PigStorage(',') AS( userid:int, movid:int,
rating:CharArray,timestamp:int);
group_data = GROUP ratings BY movid;
count_data = FOREACH group_data GENERATE group,COUNT(ratings.userid) as CountOfUser;
joinData = JOIN movies BY movieid,count_data BY group;
result = FOREACh joinData GENERATE title,CountOfUser;
DUMP result;
```

Output:

```
(Starquest II (1996),1)
(Rivers of Sand (1974),1)
(Altar of Fire (1976),1)
(Šíleně smutná princezna (1968),1)
(The Wonders of Aladdin (1961),1)
(Phobos. Fear Kills (2010),1)
(The Final Storm (2010),1)
(In a Heartbeat (2017),17)
("Blood,1)
(Jungle Woman (1944),1)
(To Be Fat Like Me (2007),1)
(Cadet Kelly (2002),7)
(The Scheming Gambler's Paradise (1905),1)
(The Man with the Rubber Head (1901),1)
(The Living Playing Cards (1905),1)
(The Hilarious Posters (1906),1)
(The Devilish Tenant (1909),1)
(The Untameable Whiskers (1904),1)
(The Imperceptable Transmutations (1904),1)
(Pooh's Heffalump Halloween Movie (2005),1)
(The One-Man Band (1900),1)
(The Fat and Lean Wrestling Match (1900),2)
(Deep Hearts (1981),1)
(Mom (2017),2)
(The Morning After (2015),1)
(St. Michael Had a Rooster (1972),1)
(House of Horrors (1946),1)
(Shadow of the Blair Witch (2000),1)
(The Burkittsville 7 (2000),1)
(Caged Heat 3000 (1995),1)
(Robin Hood (1991),1)
(Subdue,1)
(Century of Birthing (2011),1)
(Betrayal (2003),1)
(Satan Triumphant (1917),1)
2018-12-19 14:30:17,432 [main] INFO  org.apache.pig.Main - Pig script completed in 6 minutes, 41 seconds and 406 milliseconds
 (401406 ms)
You have new mail in /var/spool/mail/acadgild
```