1.

(1) **title.akas.tsv.gz**

Purpose:

The primary key is titleID and ordering, and the rest of it is some information of titles

Content:

The unique identifier of the title

The unique number of each unique title

Localized title name

Which country the title is used in

What kind of type of this title

More description of the title

Whether the title is an original title or not

(2) **title.basics.tsv.gz**

Purpose:

The primary key is tconst, and the rest of it is some information of titles

Content:

The unique identifier of the title

The type of the title

The more popular name of the title

Original title in original language

Whether the title is adult or not

The release year of the title

If the title is for tv series, then it is its end year; otherwise it shows null

The runtime of the title

An array includes the genres of the title

(3) **title.crew.tsv.gz**

Purpose:

The primary key is tconst, then it includes the information of the writers and the directors of these titles

Content:

The unique identifier of the title

The array of directors

The array of writers

(4) **title.episode.tsv.gz**

Purpose:

The information of the tv episodes. The primary key is tconst

Content:

The unique identifier of the title

The unique identifier of parent tv series of the title

The season number and the episode number

(5) **title.principals.tsv.gz**

Purpose:

The information of the cast, and the primary key is tconst and ordering

Content:

The unique identifier of the title

The unique number of each unique title

The unique identifier of the cast’s name

What kind of the cast’s job

The specific job of the cast

The name of the cast’s character

(6) **title.ratings.tsv.gz**

Purpose:

The evaluation for the titles. The primary key is tconst.

Content:

The average rating

How many people voted

(7) **name.basics.tsv.gz**

Purpose:

Some information for names. The primary key is nconst

Content:

Name

The birth year

The death year or null

The primary profession

An array of titles this person played



2.

|  |
| --- |
| non-adultMovies |
| tconst  genres  ratings  votes |

|  |
| --- |
| name |
| nconst  primaryName  birthYear  deathYear |

|  |
| --- |
| actors |
|  |

|  |
| --- |
| writers |
|  |

|  |
| --- |
| producers |
|  |

|  |
| --- |
| directors |
|  |

0..\*

0..\*

0..\*

0..\*

write

produce

|  |
| --- |
|  |
|  |

direct

0..\*

0..\*

0..\*

0..\*

act

|  |
| --- |
| non-adultMovies |
| PK  tconst  genres  ratings  votes |

3.

|  |
| --- |
| directors |
| FK  PK  movie  PK  director  FK |

|  |
| --- |
| name |
| PK  nconst  primaryName  birthYear  deathYear |

|  |
| --- |
| producers |
| FK  PK  movie  PK  producer  FK |

|  |
| --- |
| actors |
| FK  PK  movie  PK  actor  FK |

|  |
| --- |
| writers |
| FK  PK  movie  FK  PK  writer |

Non\_adultMovie

|  |  |  |  |
| --- | --- | --- | --- |
| tconst | genres | ratings | votes |
|  |  |  |  |

actors

|  |  |
| --- | --- |
| tconst | nconst |
|  |  |

directors

|  |  |
| --- | --- |
| tconst | nconst |
|  |  |

writers

|  |  |
| --- | --- |
| tconst | nconst |
|  |  |

producers

|  |  |
| --- | --- |
| tconst | nconst |
|  |  |

name

|  |  |  |  |
| --- | --- | --- | --- |
| nconst | primaryName | birthYear | deathYear |
|  |  |  |  |

I save the information of movies with the table of non\_adultMovies.The primary key is tconst. Then I save the information of crew with name table. I also connect this 2 tables with actors, directors, writers and producers. There are only 2 attributes of these 4 tables, which are the tconst and nconst are all primary key and foreign key.

Primary keys should be integer type, so the tconst and nconst should be integer. Genres are strings, so I use varchartype. Ratings are float, so I use real type. Votes are integer, so I use integer type. Actors, directors, writers and producers are all foreign key, I use varchar to store the nconst from name. primaryName is string, and birthyear and deathYear are integer.

4.

For name table, I select several columns and insert them into it. For non\_adultMovies table, I select a table from each 2 files and merge them to a new table. For tall he remaining tables, I find the category that match the pattern and select the tconst and nconst that match this category.