Chapter 4 - HiveQL

Solutions

1. How many rows are in the Movies table? How many rows are in the UserRatings?

```
SELECT count(1) FROM movies;
SELECT count(1) FROM userratings;
```

2. Using the Movies table, find the name of all movies released in 1990.

SELECT movie_name FROM movies WHERE release_date LIKE '%1990';

3. List the movieid of the most rated films in the UserRatings table.

```
SELECT count(1) AS ratings, movieid
FROM userratings
GROUP BY movieid
ORDER BY ratings DESC limit 10;
```

4. Use a join to list the names of the movies you found in step 3.

```
SELECT count(1) AS ratings, movie_name
FROM movies JOIN userratings
ON (movies.movieid = userratings.movieid)
GROUP BY movie_name ORDER BY ratings DESC LIMIT 10;
```

5. Do any movies have no ratings?

They all have a rating:

```
SELECT  movie_name
FROM movies LEFT OUTER JOIN userratings
ON (movies.movieid = userratings.movieid)
WHERE userratings.movieid IS NULL;
```

6. What is the highest rated sci_fi movie?

```
SELECT movie_name, avg(rating) AS avgrating
FROM movies JOIN userratings
ON (movies.movieid = userratings.movieid)
WHERE sci_fi = 1
GROUP BY movie_name
ORDER BY avgrating DESC LIMIT 1;
```

7. What is the highest rated sci_fi movie that has at least 10 user ratings?

```
SELECT movie_name, avgrating FROM
      (SELECT movie_name, avg(rating) AS avgrating, count(1) AS
      numratings
      FROM movies JOIN userratings
      ON (movies.movieid = userratings.movieid)
      WHERE sci_fi = 1
      GROUP BY movie_name) subq
WHERE numratings > 10
ORDER BY avgrating DESC LIMIT 1;
```