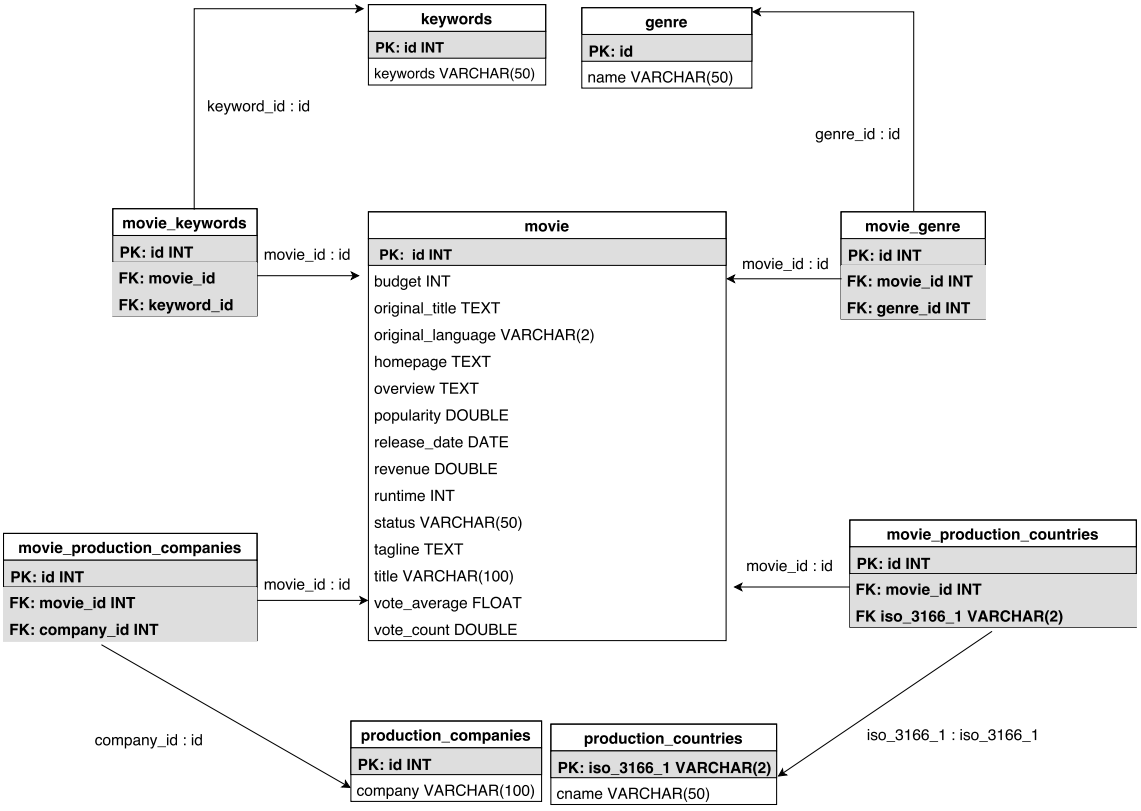
1 Data



Above are all the schemas and their relationships with other schemas. PK stands for primary key and FK stands for foriegn key. All the entries in the movies csv that where not atomic, I broke them down into two seperate tables. One those two tables I used to show the relationship to the **movie** table and the other one just kept specific information such as **keywords**, **genre**, and **production_companies**. All tables are in BCNF because (1) all atributes are atomic, (2) all non-key attributes depend only on the candidate key, (3) and in all $X \rightarrow Y$ FD's, X is a candidate key. For example, all atributes in the **movie** table only depend on the candidate key **id**.

2 Queries

1. ```
   SELECT avg(budget) FROM movie;
   ```

29045039.8753

2. ```
   SELECT title, company
   FROM
       (
           SELECT title, company, iso_3166_1
           FROM movie
           INNER JOIN movie_production_countries
           ON movie.id=movie_production_countries.movie_id
           INNER JOIN movie_production_companies
           ON movie_production_companies.movie_id=movie_production_countries.
               movie_id
           INNER JOIN production_companies company
           ON movie_production_companies.company_id = company.id
       ) AS full
   WHERE full.iso_3166_1='US';
   ```

Four Rooms	Miramax Films
Four Rooms	A Band Apart
Star Wars	Lucasfilm
Star Wars	Twentieth Century Fox Film Corporation
Finding Nemo	Pixar Animation Studios

3. ```
   SELECT title, revenue
   FROM movie
   ORDER BY revenue DESC LIMIT 5;
   ```

Avatar	2787965087
Titanic	1845034188
The Avengers	1519557910
Jurassic World	1513528810
Furious 7	1506249360
   ```

4.
```
SELECT mys.title, genre.name
FROM
  (
      SELECT title, g2.name, g.movie_id
      FROM movie
        INNER JOIN movie_genre g ON movie.id = g.movie_id
        INNER JOIN genre g2 ON g.genre_id = g2.id
      WHERE g2.name="Science Fiction"
  ) AS sci
  INNER JOIN
    (
      SELECT title, g2.name, g.movie_id
      FROM movie
        INNER JOIN movie_genre g ON movie.id = g.movie_id
        INNER JOIN genre g2 ON g.genre_id = g2.id
      WHERE g2.name = "Mystery"
    ) AS mys
  ON sci.movie_id=mys.movie_id
  INNER JOIN movie_genre ON movie_genre.movie_id=mys.movie_id
  INNER JOIN genre ON genre.id=movie_genre.genre_id;
```

Tomorrowland	Adventure
Tomorrowland	Family
Tomorrowland	Mystery
Tomorrowland	Science Fiction
Inception	Action

5.
```
SELECT title, popularity
FROM movie
WHERE popularity > (SELECT avg(popularity) FROM movie);
```

Four Rooms	22.87623
Star Wars	126.393695
Finding Nemo	85.688789
Forrest Gump	138.133331
American Beauty	80.878605