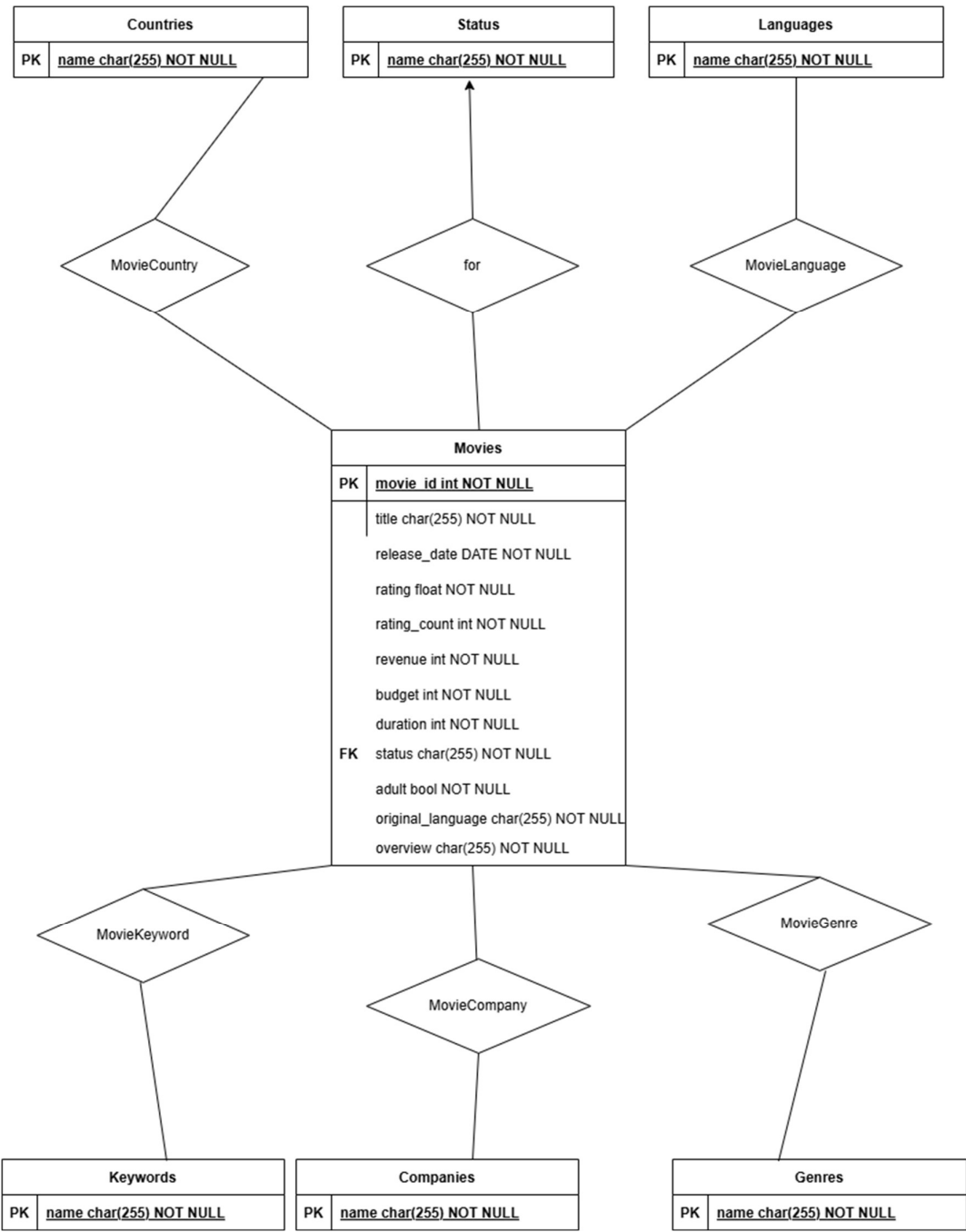


System Documentation

מבנה סכמת מסד הנתונים:



טבלאות -

Movies

טבלה מרכזית המאחסנת מידע ראשי על סרטים הכולל:

- פרטים בסיסיים (כותרת, תאריך הוצאה, משך)
- מדדי ביצועים (דירוג, מספר דירוגים)
- נתונים פיננסיים (הכנסות, תקציב)
- מידע על התוכן (תקציר, תוכן למבוגרים)
- מידע על שפה וסטטוס

Status: מאחסנת ערכי סטטוס אפשריים של סרטים.

Genres: מאחסנת ז'אנרים זמינים.

Companies: מאחסנת חברות הפקה.

Countries: מאחסנת מדינות הפקה.

Keywords: מאחסנת מילות מפתח.

Languages: מאחסנת שפות זמינות.

MovieGenre: מקשרת בין סרטים לז'אנרים.

MovieCompany: מקשרת בין סרטים לחברות הפקה.

MovieCountry: מקשרת בין סרטים למדינות הפקה.

MovieKeyword: מקשרת בין סרטים למילות מפתח.

MovieLanguage: מקשרת בין סרטים לשפות.

בחרנו לעצב את הטבלאות בצורה זו מהסיבות הבאות:

נורמליזציה -

הפרדת נתונים כמו ז'אנרים, חברות, מדינות, מילות מפתח ושפות לטבלאות נפרדות.

יתרונות:

- מניעת כפילות של מחרוזות ארוכות.
- חיסכון משמעותי במקום אחסון.
- עדכון מידע במקום אחד בלבד.

שימוש במפתחות זרים -

- אכיפת קשרים בין טבלאות.
- הבטחת מחיקה אוטומטית של קישורים לא רלוונטיים.
- הבטחת עקביות הנתונים.

סדר טבלאות -

יצירת טבלאות בסדר המתאים למפתחות זרים.

- מניעת שגיאות בעת טעינת נתונים.

אופטימיזציות:

אינדקסי FULLTEXT על עמודות overview או title מאפשרים -

- חיפוש מהיר של מילים.
- תמיכה בחיפוש מילים חלקיות.
- אפשרות לחיפוש ביטויים מדויקים.
- שיפור משמעותי בביצועי חיפוש טקסט חופשי.

אינדקסים נוספים -

idx_movies_release_date:

- מיועד לשאלות הממיינות לפי תאריך הוצאה.
- שיפור ביצועי שאלות של סרטים חדשים (למשל, דף סרטים חדשים).

idx_rating:

- מאיץ שאלות המסננות לפי דירוג.
- יעיל למציאת סרטים מדורגים גבוה (למשל, דף סרטים מומלצים).

idx_revenue_budget:

- אינדקס משולב על הכנסות ותקציב.
- מאיץ חישובי רווחיות.
- יעיל לשאלות אנליטיות.
- יעיל במיוחד עבור שאלות המצריכות מיון לפי הכנסה ואז מיון לפי תקציב.

idx_movie_genre:

- משפר ביצועי חיפוש לפי ז'אנר.
- מאיץ שאלות סטטיסטיות על ז'אנרים.

idx_movie_language:

- מייעל חיפוש לפי שפה.
- משפר ביצועי ניתוח שפות.

פירוט השאלות ומטרותיהן –

שאלתה 1 – מציגה סרטים שהתקציר שלהם מכיל את המילה 'New York'.

מטרה:

- מציאת סרטים שהעלילה שלהם התרחשה בניו יורק.
- מאפשר למשתמשים לחפש סרטים על פי מילות מפתח בתקציר.
- שימושי במיוחד למציאת סרטים לפי נושאים, מיקומים או אלמנטים ספציפיים בעלילה.
- תומך בחיפוש חופשי ואינטואיטיבי.

תמיכת מסד הנתונים:

- אינדקס FULLTEXT על עמודת overview מאפשר חיפוש יעיל.
- שמירת התקציר ב TEXT - מאפשר אחסון טקסט ארוך.

```
SELECT title, overview
FROM Movies
WHERE MATCH(overview) AGAINST("New York");
```

שאלתה 2 – מציגה סרטים שהשם שלהם מכיל את המילה שסיפק המשתמש.

מטרה:

- מאפשר למשתמשים למצוא סרטים ספציפיים לפי שמם.
- תומך בחיפוש חלקי של כותרות.
- מסייע בביווט מהיר למידע על סרט מבוקש.
- ישנו שימוש בשאלתה זו גם בדף הבית.

תמיכת מסד הנתונים:

- אינדקס FULLTEXT על עמודת title .

```
SELECT title
```

```
FROM Movies
WHERE MATCH(title) AGAINST(%s);
```

שאלתה 3 – מציגה חברות ממוינות לפי מספר הסרטים שהן שיחררו אחרי 2020.

מטרה:

- מספק תובנות על פעילות חברות ההפקה.
- מאפשר זיהוי החברות הפעילות ביותר.
- עוזר בניתוח מגמות בתעשייה.
- מופיע גם בדף סטטיסטיקות עבור סרטים משנת 2023.

תמיכת מסד הנתונים:

- טבלת MovieCompany מאפשרת קשר many-to-many .
- אינדקס על release_date לסינון יעיל.
- מפתחות זרים לשמירת שלמות הנתונים.

```
SELECT mc.name AS company, COUNT(*) AS movie_count
FROM MovieCompany mc, Movies m
WHERE m.movie_id = mc.movie_id
AND m.status = "Released"
AND m.release_date >= '2020-01-01'
GROUP BY mc.name
ORDER BY movie_count DESC;
```

שאלתה 4 – מציגה את הז'אנרים הפופולריים ביותר בכל שפה (אם יש כמה ז'אנרים פופולריים אז תציג את כולם).

מטרה:

- מזהה העדפות ז'אנר בשפות שונות.
- מסייע בהבנת טעמי קהל לפי אזורים.
- תומך בהחלטות לגבי הפקות עתידיות.
- מופיע גם בדף סטטיסטיקות.

תמיכת מסד הנתונים:

- טבלאות נפרדות לז'אנרים ושפות.
- אינדקסים על ז'אנר ושפה.

```

SELECT ml.name AS language, g.name AS genre, COUNT(*) AS genre_count
FROM Movies m
JOIN MovieGenre mg ON m.movie_id = mg.movie_id
JOIN Genres g ON mg.name = g.name
JOIN MovieLanguage ml ON m.movie_id = ml.movie_id
GROUP BY ml.name, g.name
HAVING COUNT(*) = (
    SELECT MAX(genre_count)
    FROM (
        SELECT ml1.name AS language, g1.name AS genre, COUNT(*) AS genre_count
        FROM Movies m1
        JOIN MovieGenre mg1 ON m1.movie_id = mg1.movie_id
        JOIN Genres g1 ON mg1.name = g1.name
        JOIN MovieLanguage ml1 ON m1.movie_id = ml1.movie_id
        GROUP BY ml1.name, g1.name
    ) AS genre_counts
    WHERE genre_counts.language = ml.name
)
ORDER BY genre_count DESC, language;

```

שאלתה 5 – מציגה סרטים עם דירוג ורווח מעל הממוצע.

מטרה:

- מזהה סרטים מצליחים מבחינת דירוג ורווחיות.
- מספק תובנות על הקשר בין איכות להצלחה כלכלית.
- מופיע גם בדף סטטיסטיקות.

תמיכת מסד הנתונים:

- אינדקסים על revenue/budget ו rating.

```

SELECT m.title, m.rating, (m.revenue - m.budget) AS profit
FROM Movies m
WHERE m.revenue > 0
AND m.rating > (
    SELECT AVG(m1.rating)
    FROM Movies m1
    WHERE m1.revenue > 0)
AND (m.revenue - m.budget) > (
    SELECT AVG(m2.revenue - m2.budget)
    FROM Movies m2
    WHERE m2.revenue > 0
)
ORDER BY m.rating DESC, profit DESC;

```

מבנה הקוד:

create_db_script.py - מכיל את הקוד האחראי ליצירת מסד הנתונים:

- הגדרת כל הטבלאות הנדרשות.
- יצירת האינדקסים.
- הגדרת מפתחות זרים וראשיים.
- יצירת הקשרים בין הטבלאות.

api_data_retrieve.py - מטפל בהכנסת נתונים:

- מכיל את הלוגיקה להכנסת מידע למסד הנתונים.
- אחראי על שליפת המידע מה-CSV.
- מבצע עיבוד ראשוני של המידע לפני הכנסתו.
- מוודא שהנתונים תקינים ומתאימים למבנה הטבלאות.

queries_db_script.py - מכיל פונקציות עבור השאילתות במסד הנתונים:

- הגדרת כל השאילתות הנדרשות.
- שימוש בפונקציות נפרדות לכל שאילתה מורכבת.
- מימוש הלוגיקה של כל שאילתה.

queries_execution.py - מכיל את הפונקציה הראשית וקורא לשאילתות מהקובץ
queries_db_script.py

- הרצת השאילתות.
- טיפול בשגיאות.
- הדפסת התוצאות.

שימוש ב-MYSQL Connector:

- התחברות למסד הנתונים.
- סגירה נכונה של חיבורים וקורסורים.