

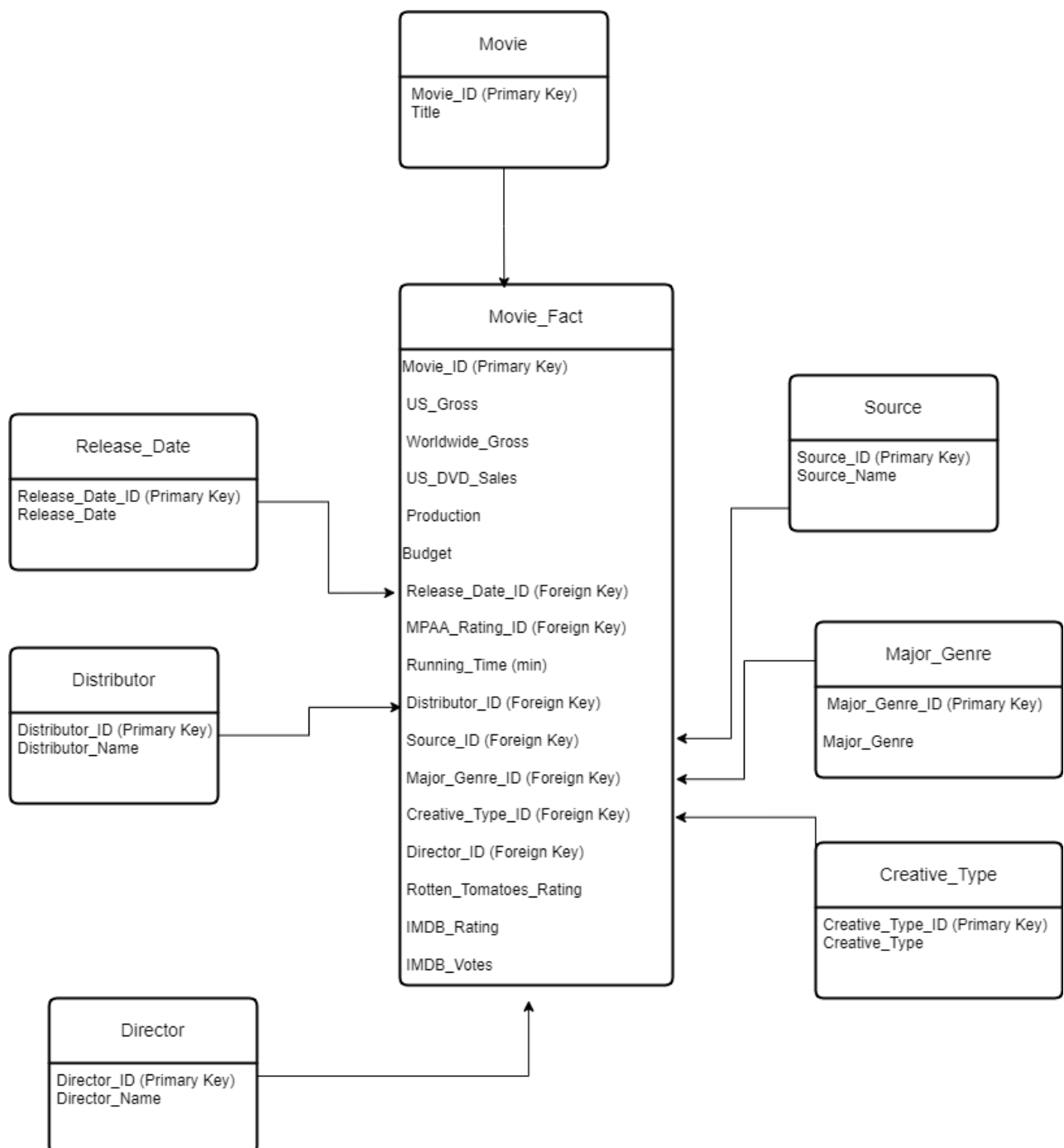
Homework 2 – Business Intelligence

יהונתן כהן
שובל כהן
אורן הייקאני

חלק 2

1. עבור מערך הנתונים נבחר בסכימת הכוכבים מכיוון שהיא מפשטת שאילתות וניתוח על ידי ארגון הנתונים לטבלת Fact מרכזית עבור מדדים הקשורים לסרטים וטבלאות מימדים נפרדות עבור ישויות כמו במאים, תאריכי פרסום וז'אנרים, המאפשרת אחזור נתונים יעיל.

2. תיאור ה-Datawarehouse לאחר שילוב הסכימה:



3. סכימת הכובבים מפשטת את ניתוח הסרטים על ידי מתן טבלאות ממדים נפרדות לז'אנר ולבמאי, המאפשרת שליפה מהירה יותר של מידע רלוונטי ומקלה על קבלת החלטות מושכלת בתעשיית הקולנוע.

חלק 3

1. תהליך ה-ETL עבור מערך הנתונים כולל את הפעולות הבאות בכל שלב:

• Extract

- חילוץ הנתונים מקבצי המקור, כמו קבצי CSV, JSON או מסד נתונים המכיל את מערך הנתונים.

• Transform

- ניקוי והפיכת הנתונים שחולצו כדי שיתאימו לניתוח וטעינה למחסן הנתונים.
- ניקוי נתונים כולל טיפול בערכים חסרים, תיקון פורמטים לא עקביים של נתונים, הסרת כפילויות והבטחת איכות הנתונים.
- הטרנספורמציות עשויות לכלול המרת פורמטים של תאריכים, סטנדרטיזציה של ז'אנרים או דירוגים, גזירת מדדים נוספים או ביצוע נורמליזציה של נתונים.

• Load

- טעינת הנתונים שעברו טרנספורמציה למחסן הנתונים או למערכת היעד, תוך הקפדה על המבנה של סכימת Star.
- הנתונים שעברו טרנספורמציה נטענים לטבלאות המתאימות של מחסן הנתונים, כגון טבלת ה-Fact וטבלאות מימדים (למשל, סרט, תאריך יציאה, במאי וכו').
- ניתן לבצע טעינה באמצעות שאילתות SQL, כלי טעינת נתונים או תוכנת ETL התומכת במסד הנתונים היעד.

תהליך ה-ETL עבור מערך הנתונים כולל חילוץ הנתונים הגולמיים מהמקור, הפיכתם כך שיתאימו למבנה ולתקני האיכות הרצויים, ולאחר מכן טעינתם למחסן הנתונים. זה מכין את הנתונים לשאילתה, ניתוח ודיווח יעילים, ומאפשרים למשתמשים להפיק תובנות חשובות ממערך הנתונים המשולב והמשתנה.