Scalable Data Engineering, Exercise 1

Henry Haustein
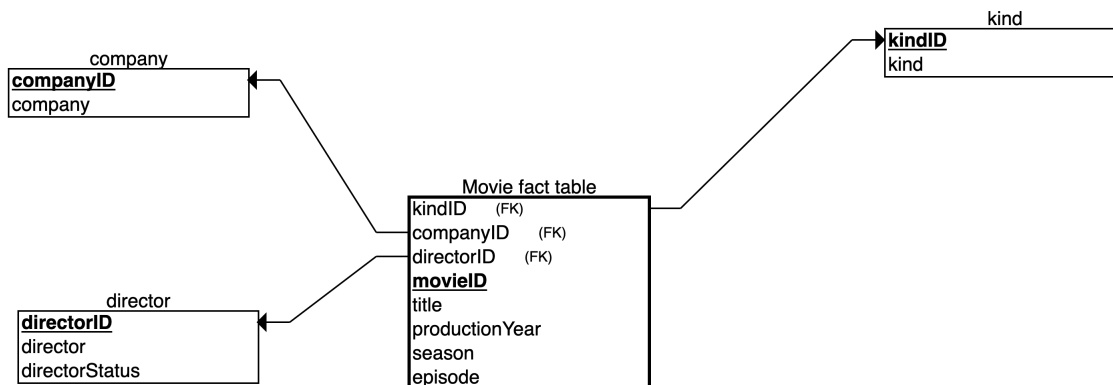
Task 1

(a) False

(b) False

(c) False (?) it just provided a link to the data sources

(d) True

(e) False. The data cube is the base just for MOLAP. ROLAP uses a relational database.

(f) False. Depends on your goals/data.

(g) True

(h) False. It can have multiple fact tables.

(i) False. You have to transform it back to a data cube for better readability.
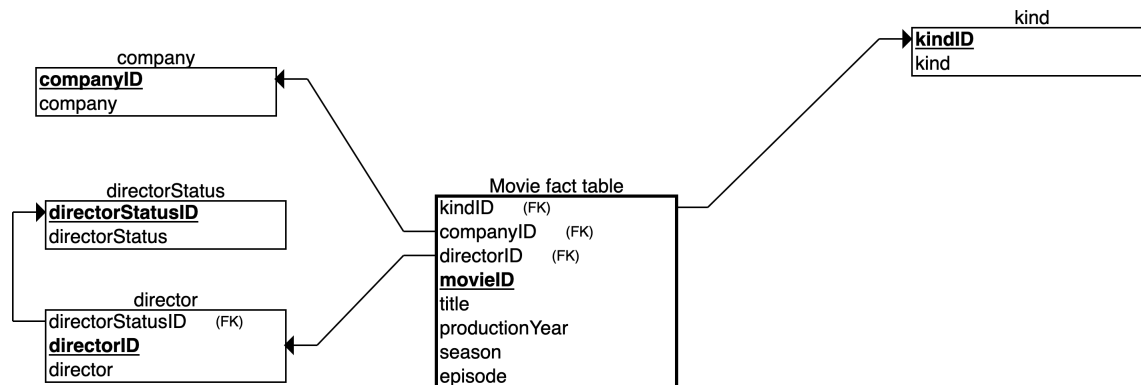
(j) True. If properly normalised.

Task 2

	one fact table	multiple fact tables
partially normalised	star schema	galaxy schema
normalised	snowflake schema	snowstorm schema

Task 3

Task 4

kind
kindID
kind

company
companyID
company

directorStatus
directorStatusID
directorStatus

director
directorStatusID   (FK)
directorID
director

Movie fact table
kindID   (FK)
companyID   (FK)
directorID   (FK)
movieID
title
productionYear
season
episode

Task 5

When storing all information in a single large table, each tuple (= row) consists of 5 strings and 1 integer, so in total $5 \cdot 128 \cdot 8$ Bit $+ 3 \cdot 64$ Bit $= 5.312$ Bit. With 1000 rows in this table, we need in total $5.312.000$ Bit $= 664.000$ Byte. *In the exercise they say the season and episode columns are empty, but we need to save them anyway.*

For the star schema we have 4 tables:

- fact table: $1000 \cdot (7 \cdot 64$ Bit $+ 128 \cdot 8$ Bit$) = 1.472.000$ Bit
- company table: $1000 \cdot (64$ Bit $+ 128 \cdot 8$ Bit$) = 1.088.000$ Bit
- director table: $1000 \cdot (64$ Bit $+ 2 \cdot 128 \cdot 8$ Bit$) = 2.112.000$ Bit
- kind table: $7 \cdot (64$ Bit $+ 128 \cdot 8$ Bit$) = 7.616$ Bit

In total this are $4.679.616$ Bit $= 584.952$ Byte.

For the snowflake schema we have 5 tables:

- fact table: $1000 \cdot (7 \cdot 64$ Bit $+ 128 \cdot 8$ Bit$) = 1.472.000$ Bit
- company table: $1000 \cdot (64$ Bit $+ 128 \cdot 8$ Bit$) = 1.088.000$ Bit
- director table: $1000 \cdot (2 \cdot 64$ Bit $+ 128 \cdot 8$ Bit$) = 1.152.000$ Bit
- director status table: $3 \cdot (64$ Bit $+ 128 \cdot 8$ Bit$) = 3.264$ Bit
- kind table: $7 \cdot (64$ Bit $+ 128 \cdot 8$ Bit$) = 7.616$ Bit

In total this are $3.722.880$ Bit $= 465.360$ Byte.