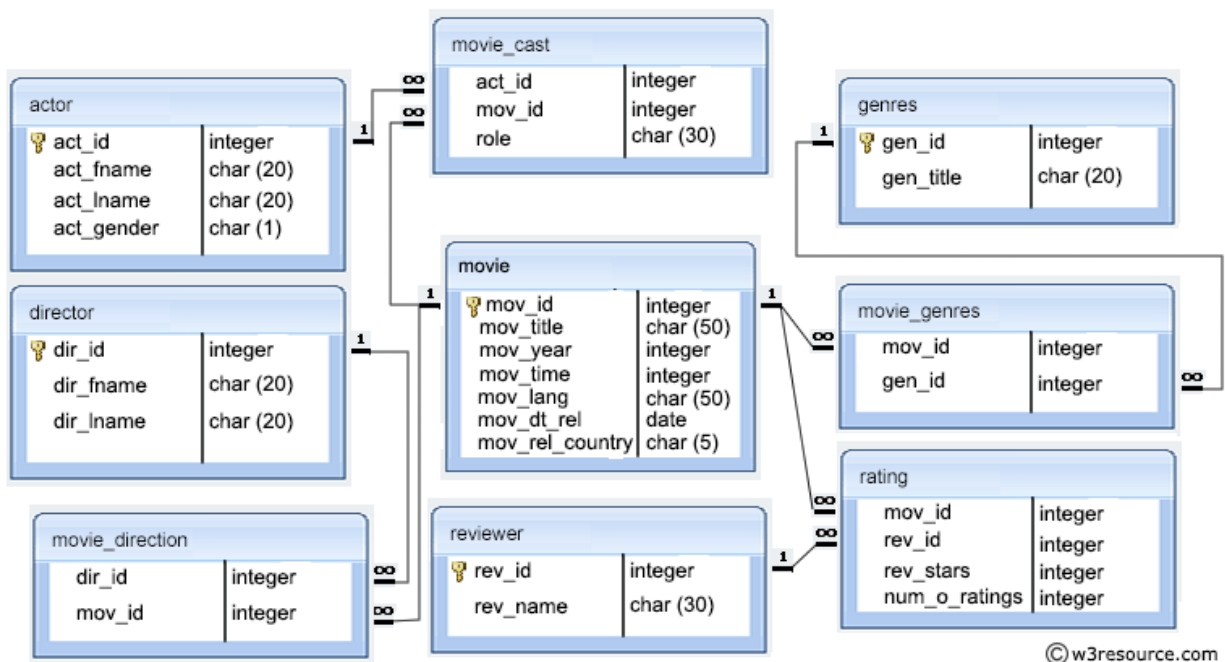


## Interview Challenge (Data Engineer Internship) - 3 days

### เกณฑ์การประเมิน

- การทำงานได้ครบถ้วนตาม requirement
- การทำงานได้โดยไม่เกิดความผิดพลาด
- คุณภาพของ source code (project structure, design pattern, solid principle)
- ความสามารถในการเรียนรู้เครื่องมือและเทคนิคใหม่ในเวลาจำกัด

### โจทย์



### เขียน SQL Script จากฐานข้อมูลภาพยนตร์ตามรูปด้านบน

1. แสดงข้อมูล ผู้กำกับ (director) ที่ได้รับ ความนิยม (rating) สูงที่สุดของแต่ละประเภทหนัง (movie\_genres)
2. แสดงข้อมูล นักแสดงแต่ละคนแสดงหนังแต่ละเรื่องรวมเป็นระยะเวลาเท่าใด โดยคิดเฉพาะหนังที่คะแนนความนิยมไม่ใช่ null
3. แสดงข้อมูล จำนวนครั้งที่ ผู้กำกับ (director) และนักแสดงหญิง (actor) เคยร่วมงานกัน โดยแสดงเฉพาะ 5 ลำดับแรกของนักแสดงหญิงที่ร่วมงานกับผู้กำกับคนนั้นบ่อยที่สุด
4. เปลี่ยนประเภทข้อมูล gen\_title จาก char (20) เป็น integer โดยเรียงตามตัวอักษร เช่น action แปลงเป็น 1, animation แปลงเป็น 2, biography แปลงเป็น 3, ...
5. แก้ไขเพศของนักแสดง ที่ชื่อขึ้นต้นด้วยคำว่า Em, Char หรือ ชื่อลงท้ายด้วยคำว่า dy, sy, lia เป็นเพศหญิง โดยไม่สนใจตัวเล็กตัวใหญ่

### เขียน Python Script

1. สุ่มเลือกผู้ใช้งานในระบบ 20 คน โดยการเรียก API และเก็บข้อมูล ชื่อ นามสกุล และ เพศ  
<https://randomuser.me/api>
2. เรียก API เพื่อทำนายเพศของผู้ใช้ในข้อที่ 1 และเก็บข้อมูล เพศและความน่าจะเป็น (probability)  
[https://api.genderize.io/?name=<first\\_name>](https://api.genderize.io/?name=<first_name>)
3. ตรวจสอบความตรงกันของเพศจากระบบกับเพศจากการทำนาย แสดงผลทาง DataFrame ในรูปแบบดังนี้

first_name	last_name	gender (actual)	gender (predict)	probability	same_gender
Marilyn	Stanley	female	female	1.0	true
Mali	Forsmo	female	male	0.69	false

4. ทุกครั้งที่ run script ต้องทำการสุ่มเลือกผู้ใช้งานใหม่

### ส่งมอบ

- source code และ document ทั้งหมด ผ่าน Git