Denormalization and NoSQL Modelling

Part 1 (Denormalization)

Consider the following normalized relations from a database in a large retain chain:

Store(storeID, region, managerID, squareFeet)

Employee(employeeID, whereWork, employeeName, employeeAddress)

Department(departmentID, managerID, salesGoal)

Schedule(departmentID, employeeID, date)

What opportunities might exist for denormalizing these relations when defining the physical records for this database? Under what circumstances would you consider creating such denormalized records?

Part 2 (NoSQL Modelling)

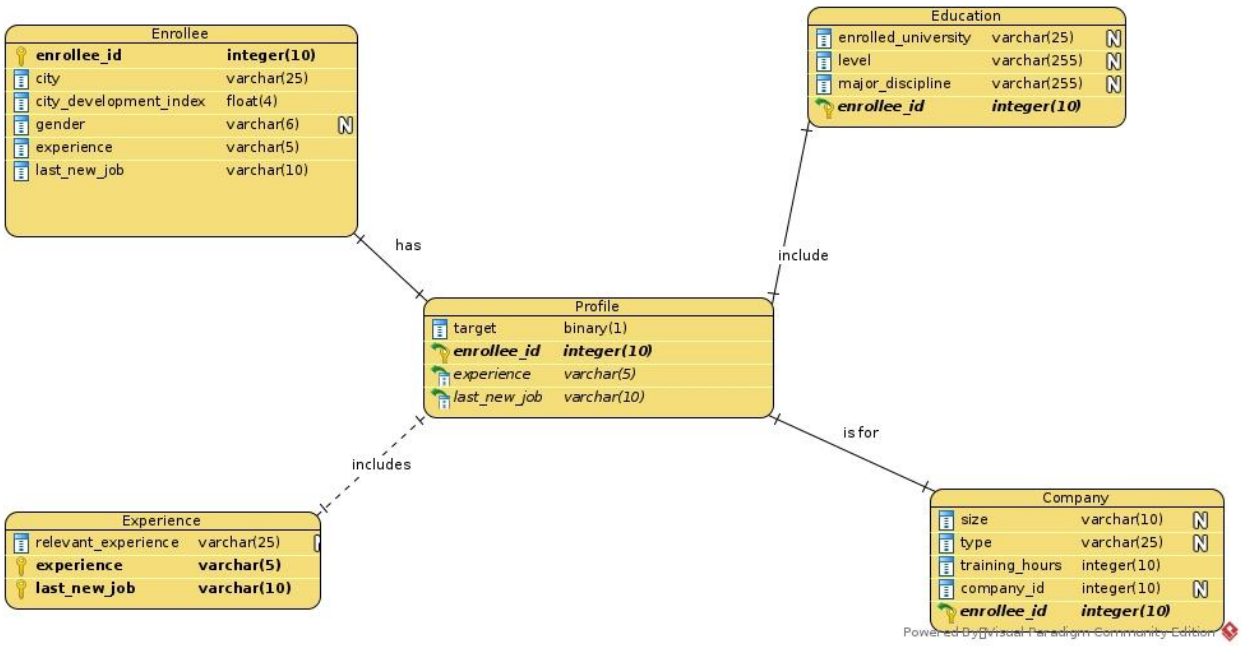
A company which is active in Big Data and Data Science wants to hire data scientists among people who successfully pass some courses which conduct by the company. Many people signup for their training. Company wants to know which of these candidates are really wants to work for the company after training or looking for a new employment because **it helps to reduce the cost and time as well as the quality of training or planning the courses and categorization of candidates**. Information related to demographics, education, experience is in hands from candidate's signup and enrollment.

This dataset designed to understand the factors that lead a person to leave current job for HR researches too.

Data Set

Variable	Definition	Key
enrollee_id	Unique ID for candidate	
city	City code	
city_ development _index	Developement index of the city (scaled)	
gender	Gender of candidate	Male, Female, Other
relevent_experience	Relevant experience of candidate	Has relevent experience, No relevent experience

enrolled_university	Type of University course enrolled if any	no_enrollment, Full time course, Part time course
education_level:	Education level of candidate	Graduate, Masters, PhD, High School, Primary School
major_discipline	Education major discipline of candidate	STEM, Business Degree, Arts, Humanities, No Major, Other,
experience	Candidate total experience in years	
company_size	Number of employees in current employer's company	
company_type	Type of current employer	Pvt Ltd, Funded Startup, Early Stage Startup, Public Sector, Other
last*new*job	Difference in years between previous job and current job	Never, 1, >4, etc
training_hours	training hours completed	
target	Final decision whether the candidate wants to change his/her current job	0 – Not looking for job change, 1 – Looking for a job change



Tasks:

1. Based on the given ERD, state an aggregate boundary
2. Use the visual paradigm to draw aggregate data model for the stated aggregate boundary
3. Use the csv data and the aggregate data model designed in the task 2 to create JSON files. Each JSON file only contains one JSON object.