Maxime Theokritoff

CSCI - 6622

5b – Database Design

a - Technical specifications

-Individual Queries for the monetary value of a Skill.
Will implement a JOBS table to hold Salary information linked to a SKILLS Table to hold skills names

-Individual Queries for the number of jobs available with a specific skill
Will implement a SKILLS Table with the Skills Name linked to a JOBS Table with the Job status.

-The Database needs to be set up for Business queries
Companies information has been set up

-Skills might be evaluated against geographical area
An CONTACT INFROMATION Table is linked to an OFFICES and a PERSONS Table to be able to query for information on SKILLS and JOBS, linked to PERSONS.

-The skills obtained from a degree could be queried as well
A DEGREES Table is linked to a PERSONS Table linked to JOBS and SKILLS Tables.

B - Assumptions list

1) **We assume we might have missed some of the assumptions that were made.**

2) **We assume this database is an enhanced minimum viable product and that it will be improved further on**. E.g. adding the time dimension, having more intermediate tables to categorizes skills: managing project and project management.

3) We assume there will be a database administrator.

4) When setting the varchar length, we roughly assumed the entry length. We always added a few possible characters to give flexibility and make sure unexpended inputs would work. E.g. leading spaces.

5) For filed asking for a long text like a description we set the value to text. This way we expect to never run out of space.

6) Year values are set as int as we assume, they could be used for computations.

7) We assume that multiple people or offices can have the same contact information. Husband and wife, corporate office and subsidiary redirecting to a call center, …

8) We assume data will be queried from multiple sources. Jobs information will come from Guru, LinkedIn … Companies and times information might come from Bloomberg, Morning Star…

9) We assume people and jobs should have multiple skills

10) We assume that the same degree can generate different skills based on the individual who obtained it and therefore only link degrees to people and not to skills.

11) We assume our JOBS table will have opened positions and currently staffed positions.

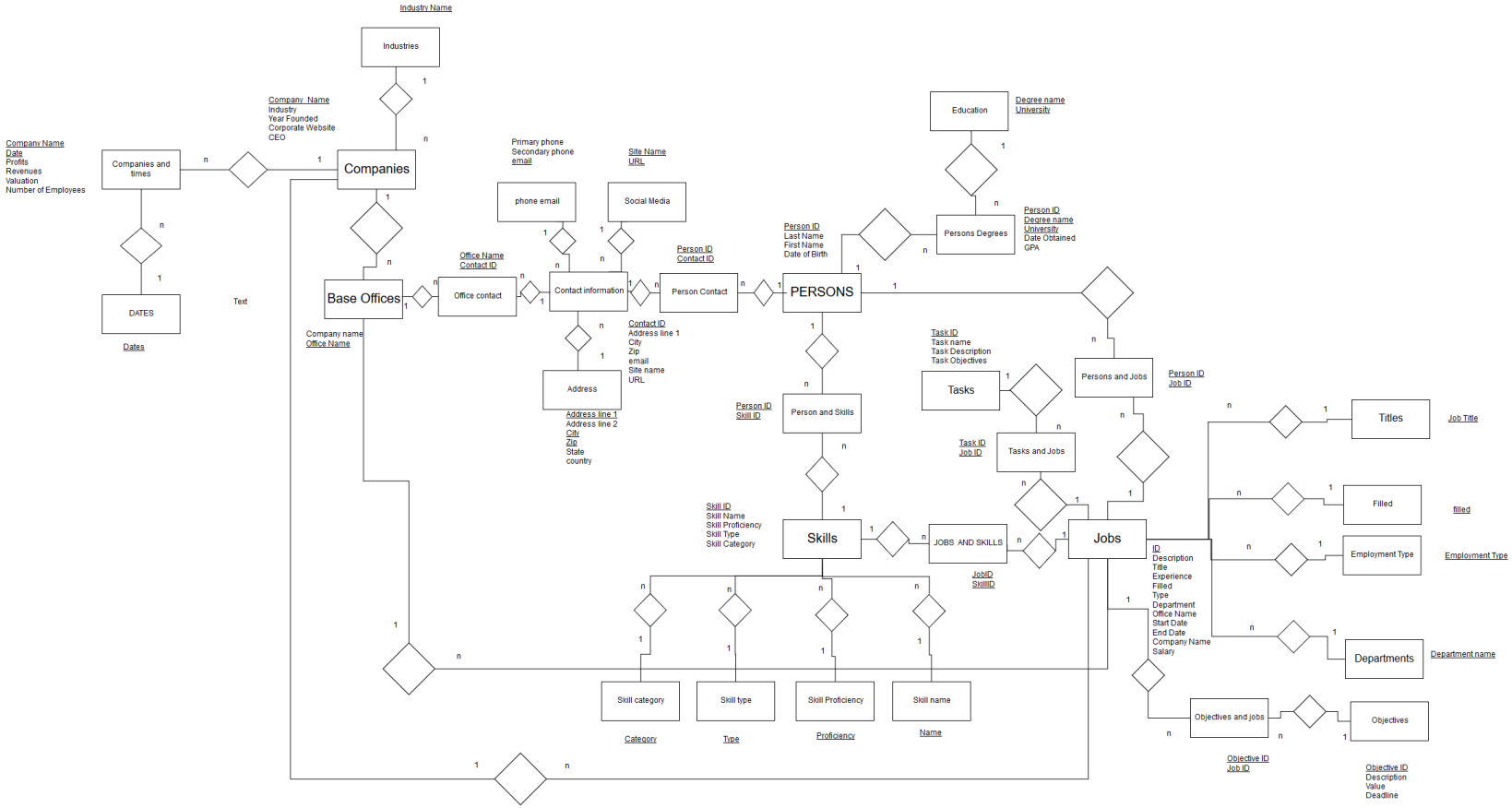
12) We assumed jobs are linked to people, skills, offices and companies.

13) We assume most table and column names give sufficiently good information on their content.

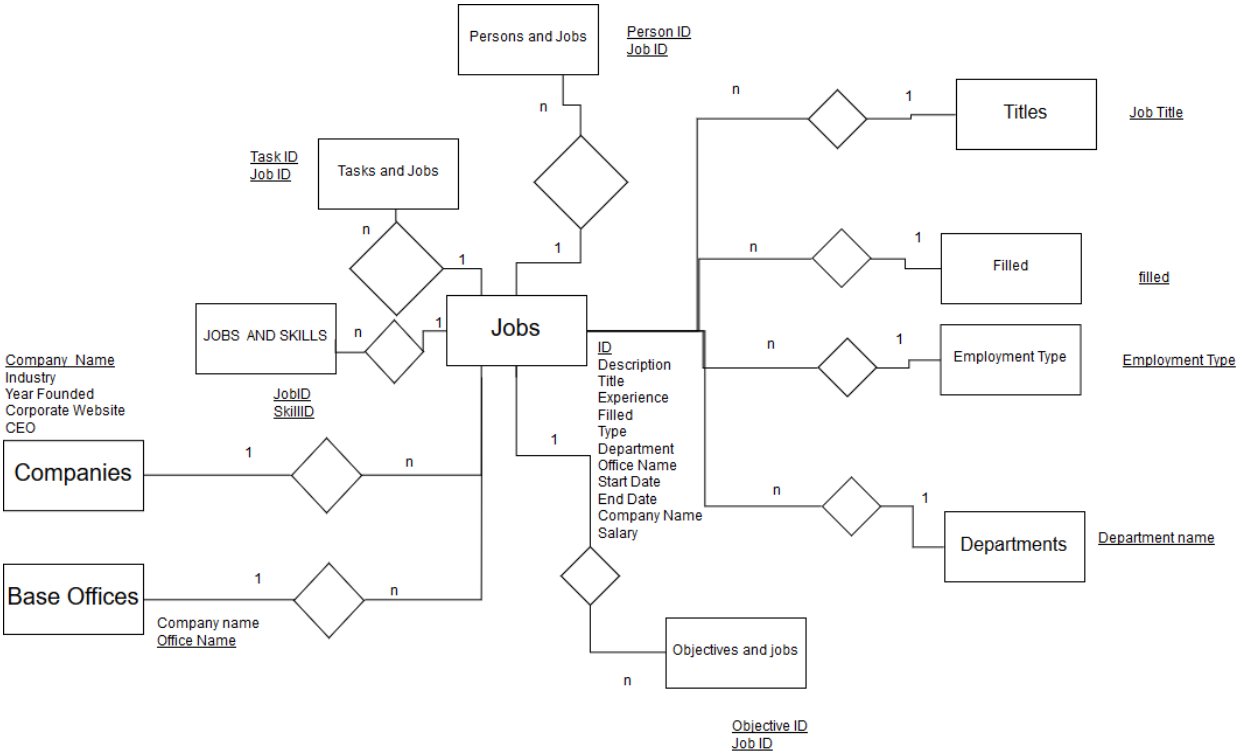
c- Tables descriptions

Our design gravitates around three main tables, JOBS, SKILLS And PERSONS. We will first describe these tables as well as the tables linked exclusively to them. We will then start describing the links between these tables made by cross-reference tables.

To the best of our knowledge, we have achieved the 3rd normal form for the SKILLS database. Details about the 3rd normal form criteria can be found in document 3 – Project Assesment section d.


Industry Name
Industries
Company Name
Date
Profits
Revenues
Valuation
Number of Employees
Companies and times
Company Name
Industry
Year Founded
Corporate Website
CEO
Companies
DATES
Dates
Text
Primary phone
Secondary phone
email
Site Name
URL
phone email
Social Media
Office Name
Contact ID
Person ID
Contact ID
Base Offices
Office contact
Contact information
Person Contact
Company name
Office Name
Contact ID
Address line 1
City
Zip
email
Site name
URL
Address
Address line 1
Address line 2
City
Zip
State
country
Person ID
Last Name
First Name
Date of Birth
PERSONS
Education
Degree name
University
Persons Degrees
Person ID
Degree name
University
Date Obtained
GPA
Person ID
Skill ID
Person and Skills
Task ID
Task name
Task Description
Task Objectives
Tasks
Persons and Jobs
Person ID
Job ID
Task ID
Job ID
Tasks and Jobs
Skill ID
Skill Name
Skill Proficiency
Skill Type
Skill Category
Skills
JOBS AND SKILLS
JobID
SkillID
Jobs
ID
Description
Title
Experience
Filled
Type
Department
Office Name
Start Date
End Date
Company Name
Salary
Titles
Job Title
Filled
filled
Employment Type
Employment Type
Departments
Department name
Skill category
Skill type
Skill Proficiency
Skill name
Category
Type
Proficiency
Name
Objectives and jobs
Objective ID
Job ID
Objectives
Objective ID
Description
Value
Deadline

JOBS



The JOBS table has 12 columns in it. This table is expected to hold thousands if not millions of records.

Job ID is its primary key. It was created as all the other columns could de identical for two different jobs except maybe description. As description is a text field and will hold large amount of text, having it as a primary key is not an option. Job ID is used as the foreign key for all tables linked to jobs except the reference tables.

The other column names are self-explanatory except may be filled and type. Filled will hold information on the recruiting stage the job is in, are applications being review, has the job been submitted, is the positions filled? Type will give information on the type of job, part time, full time…

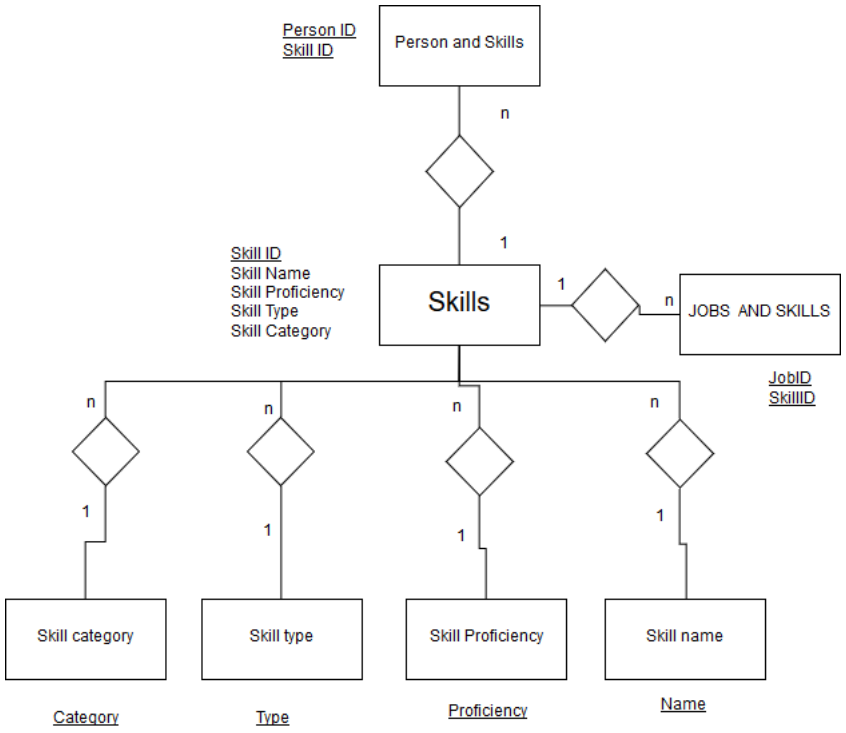
The JOBS table has 4 reference tables TITLES, FILLED, EMPLOYMENT TYPE and DEPARTEMENTS. These tables are used to control the possible entries for the JOBS tables on the related fields which are foreign keys.

Th JOBS table is also linked to the COMPANIES AND BASE OFFICES tables. The JOBS tables will have foreign keys for company name and base office. As a job is linked to a company and generally is linked to an office.

The JOBS table has many cross-reference tables. Two are unique to it, TASKS and OBJECTIVES. We will therefore mention them here and not in the cross-reference table section. Both are cross-reference tables a task can be assign to different jobs, e.g. working in a team and jobs will have multiple tasks.

Objectives follow the same logic, they differ from tasks as they are more specific and have constraints, due date, value…

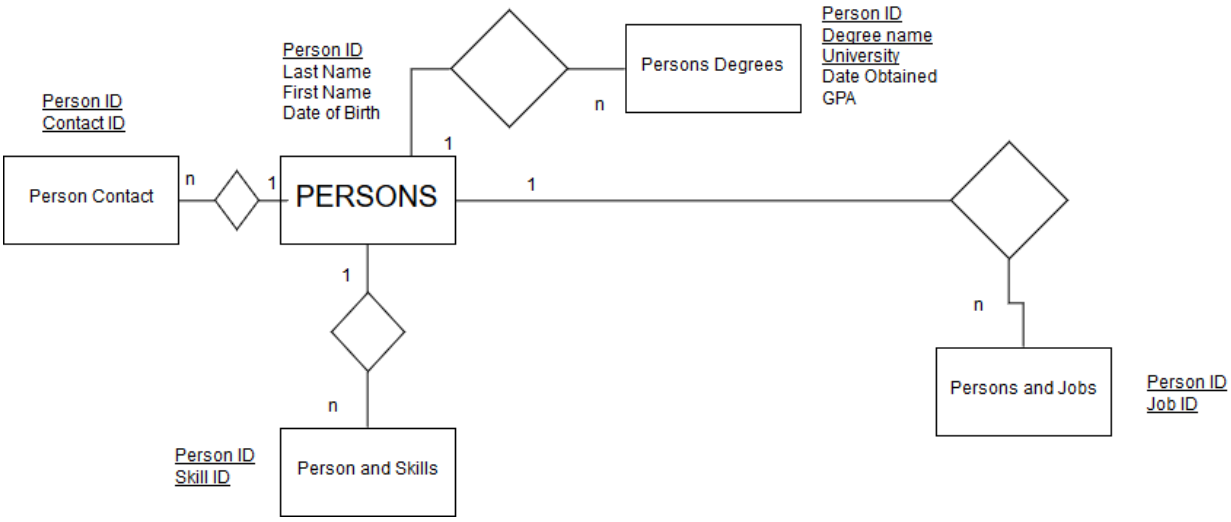
SKILLS



The SKILLS table holds the information on skills, its primary key is skill ID. This had to be done as we expect skill name to have different proficiencies. A skill like coding in C++ can be achieved at different levels, beginner, intermediate… The skill proficiency is grade by numbers from 1 to 5. The database administrator would have to harmonize all skill level to this scale.

The SKILLS table columns are all linked to reference tables to enforce data consistency. A skill type should be hard or soft. Soft skills could potential be identified by algorithms when importing jobs descriptions or people profiles. A skill category would be used to simplify queries. This filed is a way to group skills into families. E.g. Business skills would be one category.

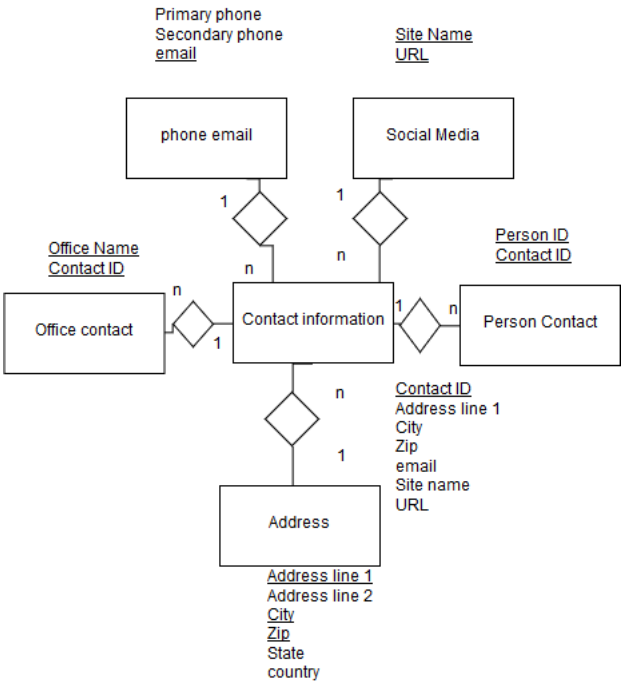
PERSONS



The persons table primary key is a persons ID. This was done as first name, last names and date of birth are not enough to ensure uniqueness. The persons table has no reference table and is only linked to cross-reference tables.

Cross-reference tables

These tables are used to break up the many to many relationships. Most of these tables' relationships are easy to see. A person has many skills and a skill can be acquired by many persons.



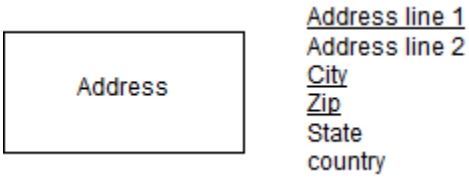
The two tables for which this many to many relationships can be questioned are: OFFICES AND CONTACTS and PERSONS AND CONTACTS.

One office can have multiple contact information, I might want to get the email of the finance department at Disney NYC or their Marketing department email. Different offices could also have the same contact information as they might have centralized their support.

For PERSONS AND CONTACTS, a father and son could share the same address. A Person could also have multiple emails.

The last table to mention here is the COMPANIES AND TIMES cross-reference table. It is noteworthy as it contains more than 2 columns. This design choice was made as profits, revenues, valuation and numbers of employees all depend on a company as well as a specific date. Therefore, we decided to store them in the same table.

Other tables design choices



The ADDRESSES table is a composite key as we are sure we can uniquely identify an address by its address line1, City and Zip code.

The SOCIAL MEDIA table's primary key is a composite key made of URL and Site Name. This choice was made to ensure that we have both a URL and a website name.

List of tables

Main tables

JOBS

SKILLS

PERSONS

List of reference tables

DATES

INDUSTRIES

SKILLS CATEGORIES

SKILLS TYPES

SKILLS PROFICIENCIES

SKILLS NAMES

DEPARTEMENTS

EMPLOYMENT TYPE

FILLED

TITLES

EDUCATION

List of Cross reference tables

OBJECTIVES AND JOBS

JOBS AND SKILLS

PERSONS AND JOBS

PERSONS AND SKILLS

PERSONS AND CONTACTS

PERSONS AND DEGREES

OFFICE AND CONTACTS

COMPANIES AND TIMES

TASKS AND JOBS

List of other/support tables

COMPANIES

BASE OFFICES

PHONE EMAIL

SOCIAL MEDIA

ADDRESS

TASKS

OBJECTIVES

CONTACTS INFORMATION