

Due dates: March 21, April 11 and April 27, 2017

Project description

In 2016, New Orleans was ranked #1 in population growth of traditional cities. On the other hand, the percentage of IT professionals within Louisiana's workforce is only 1.05%, which is 36% lower than the national average, and 93%, 121%, 195% lower than Alabama, Florida and Texas respectively. Louisiana deserves to claim more professional jobs by creating and enhancing programs that train the workforce for coding jobs, professional IT jobs. This is not only because coding jobs represent a large and growing part of the job market (there were nearly 7 million job openings in the U.S. last year for roles requiring coding skills), this is also because the jobs requiring coding skills are projected to grow 12% faster than the job market overall in the next 10 years; IT jobs are expected to grow 25% faster than the overall market¹. Jobs valuing coding skills have average salaries at or above the average for all jobs. The wages associated with each role type correspond with the intensity of programming skill requirements. Nationwide, the highest earning roles are in IT, with an average advertised salary of \$90,000².

While creating one's own business is a solution, finding a job is a more common way to start a college graduate's careers. To be qualified for a job, one often has to retool him/herself by learning and training. In this project, your task is to build an information system that can help optimize training programs, job assistance services, career planning, and company recruitment. The central efforts for this task is to design a database that support these services. If your design is good and if Louisiana wants to keep the crown of workforce training in nation, the State has to use your program.

Workers are classified as staff (salary worker), wage-workers, and volunteers. Companies and organizations hold jobs. Workers acquire knowledge and skills through two types of learning activities, academic courses and commercial trainings. The elements of these two types of activities are courses. The mission of learning is to fill up knowledge/skill gaps between the workers' possession and the job's requirements. In industry, practitioners refer this kind of learning activities as *training*, so do we. Generally, academic courses help students learn knowledge; training courses help trainees gain skills. Each course covers a set of knowledge/skills. For example, CSCI 4125 covers: data modeling (medium level), relational models (medium level), SQL (advanced level), JDBC programming (beginner level), RDBMS design (medium level), and database concurrency control (advanced level).

A person can work on one or more jobs. A job is a slot that pays the worker. Every good-pay job requires many skills and sometimes a certificate. Jobs can be classified into categories. Job categories form hierarchies. There are standards of job category descriptions such as those provided by the Department of Labor at the federal level or in many states.

Some important entities along with their important attributes are listed below. The attributes surrounded by a pair of braces ({ }) may have multiple values, the attribute that carries a plus (+) is a composite field which you should flatten in relational modeling. Note, the purpose of the following information is to describe the systems, **not** to give you an actual schema of the data model. For example, "company" in "Job(job_code, type, pay_rate, pay_type, company)" means the identifier of a company, rather than the company object. A focus of your design is the database schema configuration. In your database, you should consider the nature of the data and the queries to process. You will need additional relationships such as Has_Skill and Work_at as well as entities defined by yourselves.

Person(per_id, name, address⁺, zip_code, email, gender, {phone})
 where per_id is a the primary key.

Job_category(cate_code, title, description, pay_range_high, pay_range_low, {core_skill}, parent_cate)
 where cate_code is a unique identifier, and core_skill refers to one or multiple skills represented by k_code. Parent_cate is a cate_code representing a higher level category that the current one belong to.

Job(job_code, emp_mode, {required_skill}, {preferred_skill}, pay_rate, pay_type, company, *more ...*)

¹ Projections data are from the Bureau of Labor Statistics' Employment Projections program. <http://www.bls.gov/emp/>

² "BEYOND POINT AND CLICK", a report by Burning Glass Technologies Corp., June 2016, www.burning-glass.com

where job_code is a unique identifier, emp_mode can be “full-time”, “part-time”, and pay_type can be wage or salary, pay_rate is the hourly rate for wage or annual pay for salary. Note, the relationship between jobs and cate_code is commonly considered many-to-one. However, some job can belong to multiple high-level categories.

Company(comp_id, address*, zip_code, primary_sector, {speciality}, website)

Where primary_sector refers to the business sector the company belongs to, such as tourism, oil and mining, entertainment, service, education, technology. Specialities can be construction, logistics service, computer hardware, software, electronics, basic materials, IT consulting, transportation.

Course (c_code, title, level, description, status, retail_price)

where c_code is a unique identifier, and status is either active or expired.

Prereq (c_code, required_code)

Section (c_code, sec_no, complete_date, year, offered_by, format, price)

Where the triple (c_code, sec_no, year) form a unique identifier for every section, format can be (classroom, online-sync, online-selfpaced, or correspondence), offered_by refers to a university or a training company.

Knowledge_Skill(ks_code, title, description, level)

where ks_code is a unique identifier (you may use the skill code defined by the Department of Labor), and level can be “beginner”, “medium” or “advanced”.

In a person’s career, working on multiple jobs is common. Sometimes, a person’s working history can make him/her invaluable at certain situation. Recruiters often seek applicants’ experience. For this reason, the database should not only represent the job a person is currently holding, it should also track every worker’s job history.

The database will support at least two services, the *career planning service*, the *job hunting assistance*, and the *training service*. The *career planning service* needs information about workers’ education, training and skills in order to help find workers suitable for a job. The *training service* can recommend the courses for a person who pursues a job by recognizing the missing skills of the person for the job. The features of *job hunting assistance* are up to your experience and imagination.

Data Resources

In this semester, you are required to use the real data in the job market through Web sites such as:

- NOLA.com <http://jobs.nola.com/Jobs/technology>. You can browse by category
- Monster <https://www.monster.com/jobs/q-it-jobs-l-new-orleans,-la.aspx>
- indeed <https://www.indeed.com/l-New-Orleans,-LA-jobs.html>
- and WorkNOLA.com <https://worknola.com/job-search>
<https://worknola.com/365-connect/web-developer-back-end-net>
<https://worknola.com/firstline-schools/2016-2017-systems-development-manager>

Visit these Web sites and get an idea about what matter for real jobs in the market. For example, on the left side of the Web page of WorkNOLA, you see panels of Job Function, Industry, Job Type, Employer, Required skills, and Preferred skills. indeed.com additionally shows salary estimate, location, experience level.

The job ads in some site such as NOLA.com appear in a very loosely formatted manner. As a consequence, the same attribute may be referred in different terminologies. For instance, “Required skills” is sometimes called “Qualifications” or “Our Expectation”, “Qualifications & characteristics” means “Ideal Candidates”. You have your freedom to include or drop out the less essential attributes such as “Qualifications & characteristics”.

I will provide you with Java program that can fetch job data from specified Web sites including the above four sites. You should be able to use it for daily/weekly data feed.

Team organization

You are required to form a team made of two people. Each team member should work on one or more different service. Every team member should perform every aspect in the development process of database applications such as E-R modeling, SQL, and Java. Do not divide your tasks horizontally such as one write SQL, the other writes Java. Such kind of task division will give the Java guy serious disadvantages in earning a fair grade.

Queries

The following queries are given to illustrate the capabilities of your system. As we understand the job market Web sites better and agree on the features of your job hunting assistance, I will adjust this list. Your database should support these queries:

1. List a company's workers by names.
2. List a company's staff by salary in descending order.
3. List companies' labor cost (total salaries and wage rates by 1920 hours) in descending order.
4. Find all the jobs a person is currently holding and worked in the past.
5. List a person's knowledge/skills in a readable format.
6. List the skill gap of a worker between his/her job(s) and his/her skills.
7. List the required knowledge/skills of a job/ a job category in a readable format. (two queries)
8. List a person's missing knowledge/skills for a specific job in a readable format.
9. List the courses (course id and title) that each alone teaches all the missing knowledge/skills for a person to pursue a specific job.
10. Suppose the skill gap of a worker and the requirement of a desired job can be covered by one course. Find the "quickest" solution for this worker. Show the course, section information and the completion date.
11. Find the cheapest course to make up one's skill gap by showing the course to take and the cost (of the section price).
12. If query #9 returns nothing, then find the **course sets** that their combination covers all the missing knowledge/skills for a person to pursue a specific job. The considered course sets will not include more than three courses. If multiple course sets are found, list the course sets (with their course IDs) in the order of the ascending order of the course sets' total costs.
13. List all the job categories that a person is qualified for.
14. Find the job with the highest pay rate for a person according to his/her skill qualification.
15. List all the names along with the emails of the persons who are qualified for a job.
16. When a company cannot find any qualified person for a job, a secondary solution is to find a person who is almost qualified to the job. Make a "missing-one" list that lists people who miss only one skill for a specified job.
17. List the skillID and the number of people in the missing-one list for a given job code in the ascending order of the people counts.
18. Suppose there is a new job that has nobody qualified. List the persons who miss the least number of skills and report the "least number".
19. For a specified job category and a given small number k , make a "missing- k " list that lists the people's IDs and the number of missing skills for the people who miss only up to k skills in the ascending order of missing skills.
20. Given a job category code and its corresponding missing- k list specified in Question 19. Find every skill that is needed by at least one person in the given missing- k list. List each skillID and the number of people who need it in the descending order of the people counts.
21. In a local or national crisis, we need to find all the people who once held a job of the special job category identifier.
22. Find all the unemployed people who once held a job of the given job identifier.
23. Find out the biggest employer in terms of number of employees or the total amount of salaries and wages paid to employees.
24. Find out the job distribution among business sectors; find out the biggest sector in terms of number of employees or the total amount of salaries and wages paid to employees.
25. Find out the ratio between the people whose earnings increase and those whose earning decrease; find the average rate of earning improvement for the workers in a specific business sector.

26. Find the leaf-node job categories that have the most openings **due to lack of qualified workers**. If there are many opening jobs of a job category but at the same time there are many qualified jobless people. Then training cannot help fill up this type of job. What we want to find is such a job category that has the largest difference between vacancies (the unfilled jobs of this category) and the number of jobless people who are qualified for the jobs of this category.
27. Find the courses that can help most jobless people find a job by training them toward the jobs of this category that have the most openings due to lack of qualified workers.
28. List all the courses, directly or indirectly required, that a person has to take in order to be qualified for a job of the given category, according to his/her skills possessed and courses taken. (required for graduate students only)

Development Tasks

- (1) Make an E-R diagram showing your data model. In the diagram, show the important attributes of each entity. For every relationship, show the cardinality and the participation status, as well as any important attributes associated with the relationship. Be sure to indicate the primary key of each strong entity. Note, this is a data-model E-R diagram, not a detailed relation schema diagram for implementation. I will accept hand-draw diagrams. Feel free to discuss your E-R diagram in class or during my office hours.
- (2) Reduce the data-model E-R diagram to the database relations schema. Draw a relation-schema diagram for your implementation.
- (3) Enumerate the concerned functional dependencies (to be learned in Chapter 8) in your system.
- (4) Revise the database schema design by producing a lossless-join 3NF schema that preserves functional dependencies (to be learned in Chapter 8).
- (5) Populate the tables with adequate data such that all the questions can be demonstrated. Be sure to write your INSERT statements in SQL scripts. You are also required to write a SQL script that cleans up everything in the database; this will be a good exercise to realize the dependencies enforced by foreign key references. In your development process, you must re-populate your database after cleaning up it for multiple times.
- (6) Write the SQL statements that carry out the 26 queries listed above, or 28 for the graduate students. Write a user interface in Java to drive each of these queries with JDBC. A graphical interface (GUI) will be appreciated but not necessary.
- (7) Design and implement **some** of the Java classes such as Course, Job_Category, Job, and Person with the JDBC technique. These Java classes should support creation (such as creating a course, a job category, or a job) and deletion (such as removing a job or setting a course *inactive*.) *Note, task (6) carries relative light weight of this project; do not spend too much time on it.*
- (8) Design and implement the database application that manages **business processes**. Write application programs to support business **processes** such as the following:
 - a) A company accepts a new employee;
 - b) A person's job hunting;
 - c) A company finds the right people for a job with training;
 - d) Evaluate the opportunities in all the business sectors for the career planning service;

Due Dates

Due date of Phase 1 (10%) – March 21: The deliverables of task (1), hand in a hardcopy of your E-R diagram.

Due date of Phase 2 (20%) – April 11: The deliverables of tasks (2), (5) and (6).

Due date of Phase 3 (70%) – April 27: The entire project report including everything such as the revised diagram of task (1) and revised SQL statements of task (6).

Every group must sign up for an hour-long project inspection with me in the week of **May 1**. At least two test cases should be prepared and documented for queries 8 through 28, as well as for services (a) through (d) specified in

Development Task (8). Each test case must include the prepared input data/setting and a specification of the expected output/results. Early submission and demonstration of project earn bonus points.

Graduate Students Are Required to Complete Additional Work (CSCI 5125) [Undergraduate students can earn extra points by working out the following requirements.]

(1) Some jobs require one or more certificates. One can obtain a certificate by passing a set of exams provided by some companies. Most certificates can expire. Many certificates are bound to a tool. A set of certificates are potential for CSCI 4125 students including Oracle Database Developer I and Oracle Database Developer II. These two certificates are bound to tool Oracle 11g. To obtain a certificate, one has to pass a number of courses. The courses are among the courses we considered previously. They could be offered either by colleges or training companies. A conceptual entity of certificate is given below.

Certificate(*cer_code*, title, description, *expire_date*, *issued_by*, *t_code*)

where *cer_code* is a unique identifier, *t_code* shows the bounded tools of this certificate if any, and *issued_by* refers to a company.

Graduate students' implementation of the project should manage certificates. The queries regarding jobs and skills should handle considerations of certifications.

(2) Implementation of Query 28 is mandatory for graduate students, optional for undergraduate students for bonus points.

(3) Design and implement experiments on concurrency control using this database system. Each experiment can be demonstrated in SQLPLUS or in JDBC implementations. Every experiment must be documented with problem specification, applied solution, and presented with three test cases (with input data/setting and the expected output/results). Every graduate student must have at least four different experiments prepared and documented.