

Database Systems (22C:144)

Part 1: Project Description

Group 3 Uber

Changze Han, Cory Skeers, Kaiqiang Zhang, Josh Kamp, Cameron Chen

Web interface:

This document is an initial description of our project. We will introduce the scope of our project in five sections which are general description, requirements, assumptions, scope and design approach.

We will keep updating this document along with implementing the project.

Individual contributions to the project description:

Changze Han :

- Cover sheet
- Description of the project
- Updated Project design approach

Cory Skeers:

- Requirements

Kaiqiang Zhang:

- Some details about project scope and assumptions

Josh Kamp:

- Project design approach

Cameron Chen:

- Project scope and assumptions

Description of the project

In this project, our group is going to design a database for Uber which is a perfect example of a gig economy company. We will utilize what we've learned in class to build a relational database to keep track of a series of information in order to support the business. Generally, we want our database to be able to perform the following tasks:

1. Enable the company/administrator to access/view all the trip information (including country, city name, date, driver, customer, mileage, price, ...)
2. Enable the company/administrator to access/view the dynamics data of each country's and each city's revenue, driver amount, customer amount and trip amount.
3. Enable drivers to access/view their personal information and driving history.
4. Enable customers to access/view their personal information and trip history.

We will go into more detail about our design plan in the requirements and scope sections. We will also talk about some specific design approaches and a couple of assumptions we made.

Requirements of the project

What is the business purpose or need?

The overall business purpose is three-fold:

1. Contract employees (drivers) need to be able to access and update their information for accurate accounting and bookkeeping
2. Clients (passengers) need to be able to access and update their information to ensure accurate billing
3. The company/administrator needs to be able to access trip frequencies, driver and client histories, peak cities/times, and other factors in order to make informed business decisions such as driver compensation changes, driver quantity updates, sales and marketing initiatives, etc.

What is the problem you are trying to solve?

As a gig economy business model, Uber works by capitalizing on relatively small margins, and is highly dependent on consistent, predictable use by both their drivers and clients. Too few drivers can leave clients unhappy due to long wait times, whereas too many drivers can satisfy clients at the expense of driver satisfaction/profitability. Similarly, rates need to remain competitive to draw in clients, quality drivers must be maintained with competitive compensation, and the business as a whole needs to be maintainable on that margin.

The existence of a well developed relational database will help drivers understand where and when the most profitable trips are, managers be able to better understand their employees and customers, and users have a better overall experience.

What questions will you be answering?

Outside of updating their personal information, the primary questions drivers and clients will have is with regards to their trip totals—drivers their total/average mileage, total/average earnings, or changes in the average number of trips per time period so that they can determine their net compensation per hour, etc.; passengers their total/average trips taken and total/average expenditure, so that they can manage their budget.

Examples may include such driver questions as: “In which week did I drive the most miles, and did it have a higher earnings-to-mile ratio than an average week’s earnings-to-mile ratio?”

The administrator/company will want to answer such questions as: What cities are the most profitable, which drivers are the least/most active; which passengers take the most trips and is it more profitable to have passengers take fewer longer trips or many short trips; is there a passenger demographic we serve more/less frequently than others; is there an increase in trips following an advertisement or promotion. These questions revolve around gathering data to make decisions on passenger fees, driver compensation rates, ideal driver/passenger ratios, and marketing/promotional efforts.

Example company questions may include:

- “In which city is there the greatest difference between its most profitable day of the week and its least profitable day of the week, and how does this city’s driver-to-client ratio compare to the city’s driver-to-client ratio which has the least difference between its most profitable day of the week and its least profitable day of the week?”
- “On average, how many days of the week are the most profitable drivers active, and does it vary by region?”
- “Of the clients who have not made a transaction in the last month, what percentage had been regular customers prior to this period (at least one transaction per two weeks)? What percentage had been regular customers for longer than one month, three months, or six months?”

Assumptions of the project

We will assume that each customer used their own account exclusively, only pays for rides for themselves, and can only be at one place at a time. We will assume that drivers also only conduct business using one phone, one account, and one car. Additionally, we will assume

all customer will pay by card. If we find data that violates these assumptions then we will reevaluate the data we are collecting and revise our assumptions.

Scope of the project

Throughout our project we will be capturing data about trips. The information that we will be focussing on is customer data, driver data, and the trip details such as time, location, duration, and trip type. Things that will be out of our scope is larger world statistics such as the price of gas in the areas that the trips are taking place, the cost of owning a car, and the distance from home that the driver's average trips are occurring.

The customer data will include account ID, name, payment method, phone number, home address(so that customers do not need to type their home address when they want back home), etc. Diver data includes account ID, name, phone number, car information including car plate, brand, type, color, and maximum people it could take.

Project Design Approach

Many types of data that we will be collecting have already been mentioned, so I will only repeat a few here for examples for each relation. We will be dividing the data into at least 9 relations, including but not limited to:

- Customer relation: contains information specific to each customer, such as their name, address, payment methods, phone number, and their trips.
- Driver relation: contains information such as the things Cory put so well, "total/average mileage, total/average earnings, or changes in the average number of trips per time period so that they can determine their net compensation per hour, etc."
- Trip relation: holds the information about each individual trip, the pickup location, and destination, customer, driver, time, date, fate, etc.
- Region relation: will most likely look at the level of cities and hold information such as drivers operating in the area and customers being picked up or dropped off in the area. Using this information, we can derive things such as the density of drivers and customers in that region as well as the average wait time for customers in that region. We can then use those to figure out the ideal driver ratio in that area.
- Booking relation
- Payment relation: each passenger owns a payment relation which records attributes like date, fare, refund...
- Rating relation: records passengers' and drivers' each star rate matching with the date.
- Accidents relation: records information of each reported accident with attributes like driver name, passenger name, date, accident code(different code represents different kinds of accident)

- Report relation: records passengers' and drivers' reports with attributes like name, date, booking number, and report code(different code represents different kind of report)

Once we start constructing the database, we may come up with a few other necessary relations, but these will be the fundamental ones on which we build the database.