Group 2 Final Project Report

Team Number: Group 2
Team Name: Second to None

Overview of Business Case

Our business is a hypothetical national company, “Healthy Aging LLC”, that refers patients being discharged home from skilled nursing or acute care facilities to agencies providing home health services. We need a better process to refer patients being discharged to home health providers, in their area, who accept their insurance. This is an important aspect of continuity of care, to ensure they get the services they need to continue to recuperate once they are discharged home. Patient discharge operations need access to data to refer those patients to a home health agency (or agencies) that can provide the services they need and accept their insurance.

Team Description with Roles

Team Members:

1. Sun Afolabi - Chairperson

Within our group we had a variety of results from our team role survey. Darcie got the result of executive as the primary result and then chairperson second. Sun received chairperson as his primary leader role. Matt received a score of expert with analyst as a second ranking score. Ashley received explorer but executive was a close second. Roseanna received the primary score of the driver. Shawn got the top score for Analyst. We do feel these roles seemed to fit who we were within our groups and we did seem to work well together. We used communication methods such as WhatsApp, and conference calls to keep in touch and stay up to date with assignments and work. When working on specific assignments, we were able to break up the work in a way that allowed us to use each other’s strengths.

Dataset Description and Motivation for Choice

We will create fictional patient data. Patient data will include basic demographics, insurance information, and information about home health services needed.

Data about home health agencies will come from the Medicare data website, https://data.medicare.gov/Home-Health-Compare/Home-Health-Care-Agencies/6jpm-

sxkc/data.Home health agency data, which is available online, includes information about location, services provided, and quality.

We ended up using data for the following tables: Patient, Patient_Service, Provider, Provider_Service, Provider_Insurance, Provider_Patient_Service, and Service. From the original ERD model we chose to not complete the Insurance table due to the fact that we did not have queries that needed that data. We only listed insurance categories (Medicare, Medicaid, or Private) within Patient and Provider, so we did not need Insurance name, ID, phone, zip, or credit rating for our queries. This decision was based on length of time to collect the data needed as well as the fact that this data would not be needed within the queries that we determined were needed for our business model.

Three significant changes were made to the final data model as compared to our original ERD. They were the way in which we handled insurance, locations, and the provider-patient relationship. Changes to insurance were driven by data considerations as mentioned above. Changes to locations was based on the feedback that you provided on our ERD; we agreed with your assessment that our original configuration was overly complicated and could have led to redundancies in storing location data.

Changes to the provider-patient relationship were a bit more extensive, and upon further reflection we did not, in our initial development phase, give careful enough consideration to this relationship. Basically our company is in the business of matching up two many-to-many relationships: patient-service (a patient can request many services and a service can be used by many patients) and provider-service (a provider can offer many services, and a service can be offered by many providers). We originally conceived of this as a simple task of matching patient and provider. But, as it turned out it was a little more complex, because we weren't really matching patients and providers, but patient-service-requests and provider-service-offerings. To make this data model work in reality, we dropped the patient_provider table and created two intersection tables for the many-to-many relationships (patient_service and provider_service); we then added another intersection table (provider_patient_service) that holds primary key values from the other two primary intersection tables.

The solution works, and brings together all the data that we need. However, admittedly, it is not elegant. It necessitates some ungainly JOINS. And so, while this works, with more time we could perhaps generate a better solution. That being said, it was a fun exercise thinking through the relationship and crafting a practical resolution.

Data Model

Healthy Aging, LLC, is a nationwide home health referral agency that needs a database to track the patients that use their services, the providers that contract with them, and the services the providers offer. They will also track location and insurance companies in order to match providers with patients. Ultimately, the aim is to match each patient to a provider in the patient's location, accepting the patient's insurance, and offering the service(s) the patient needs.

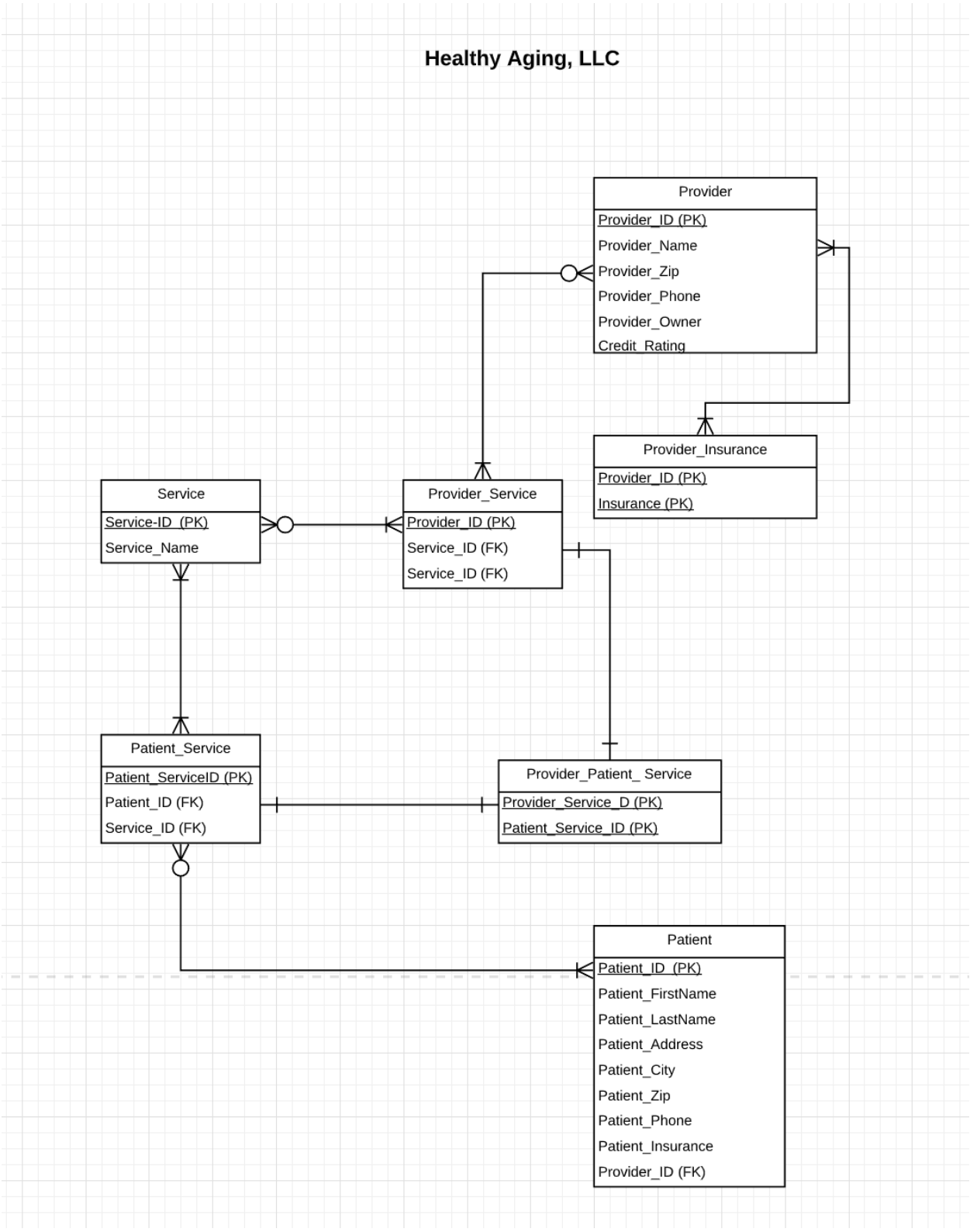
Each patient must be prescribed at least one of the services following home health services: nursing care, physical therapy, occupational therapy, speech therapy, social services, or home health. Healthy Aging tracks each patient's information including their name, address, location, insurance, and phone number, as well as the service(s) they are prescribed to receive. A patient shall only have one form of insurance listed in the database.

Providers must offer at least one of the services listed above, but most providers offer more than one service. In addition to the service(s) offered, Healthy Aging keeps track of each provider's name, phone number, address, insurance accepted, and quality of patient care rating. A provider must accept at least one insurance company but may accept many.

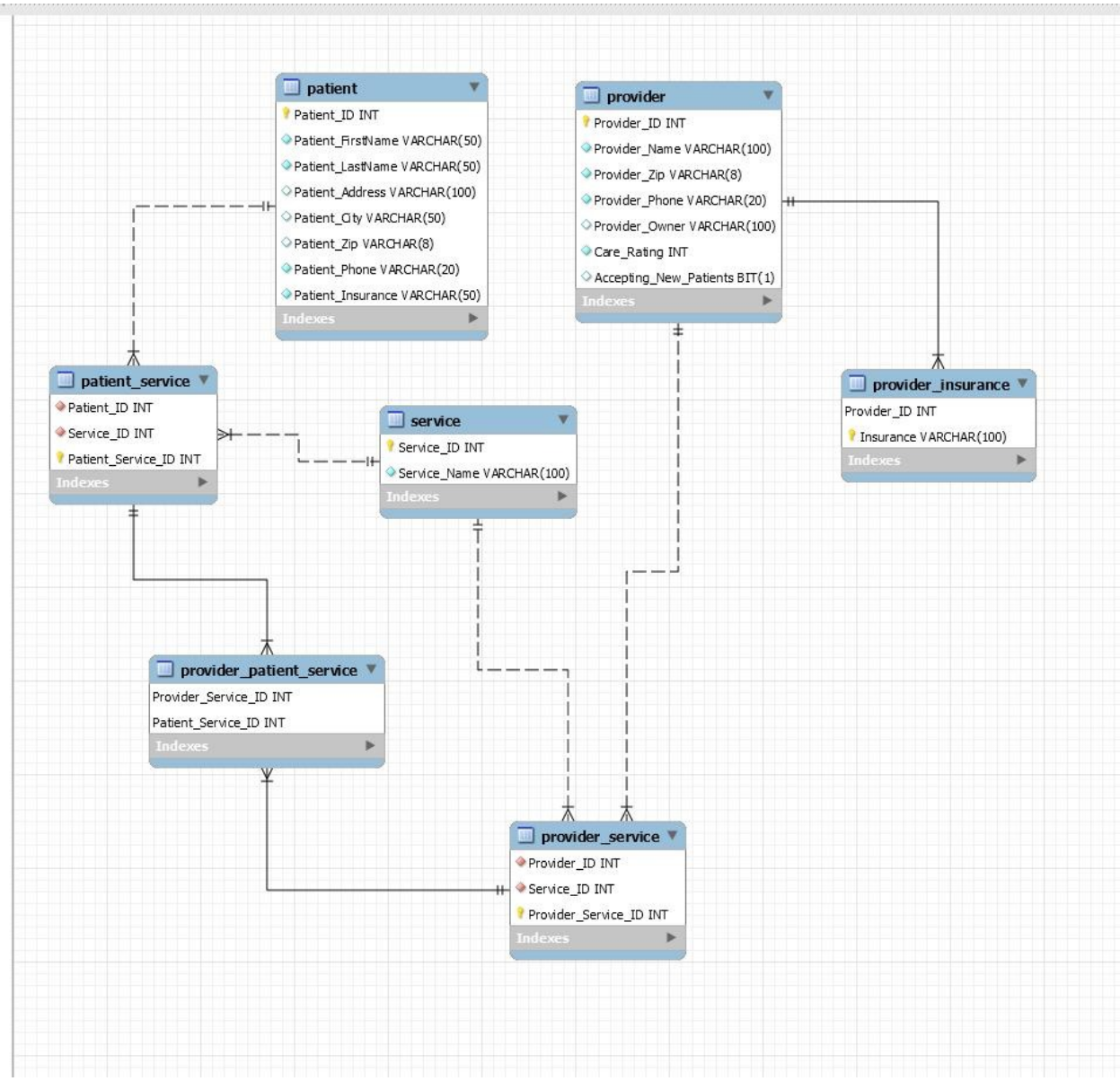
Each patient must be assigned to at least one provider and a provider can be assigned to multiple patients, but a provider does not need to be associated with any current patients to be in the system. For a patient to be assigned to a provider, they must be within the same general location and the provider must accept the patient's insurance.

For simplification, we have chosen to use three categories for insurance (Medicare, Medicaid, and Private) instead of attempting to list all insurance companies and build out which companies each provider accepts, since most providers accept a large number of private insurance companies and in-network participation changes frequently.

Healthy Aging, LLC



Screenshot of Data Model from MySQL



Normalization Process

The first step in getting our data to 1NF was to break up any fields that were storing more than one piece of data. A good example of this was patient address fields; in the ERD a single address field was intended to hold all address data (street, city, state, and zip). Instead, we broke that data into separate columns. Another example was Patient Name, where we created separate first and last name columns.

Achieving 2NF was fairly easy. In our original model we had a few tables with composite keys, but those tables were simple, storing only intersecting data (mostly keys), so we weren't faced with any partial dependencies.

What's more, we also weren't looking at many transitive dependencies, either, given that our tables were simplified, so 3NF was also easy to attain. The bigger problem in solidifying full 3NF was ensuring that we had no insert-update-deletion anomalies. We removed some entities – like location – and changed some others – like insurance – which left us with a threat of deletion anomalies. We have not necessarily solved this completely (storing insurance only in Provider_Insurance might leave us open to losing insurance data if all records of one insurance type are deleted); in a real world scenario, further iterations of our business model would necessitate that we look at this with more detail.

Sample Queries and Motivation for Choice

After reviewing our home health data we had to decide what scenarios we can create to aid our referral business and patient needs. One query we chose was to generate a list of patients seeking services in Dayton, the query identified the patient name, the services they're seeking, their city, and is sorted alphabetically by last name. We chose to use this scenario for Healthy Aging since we are a referral company, it would be helpful to identify the needs of patients on a regional basis. Utilizing this data would allow us analyze how our services are being used, which services are used least and most in each city, and overall improve our network of providers to better serve communities in need.

Another query we used was to generate a list of providers offering nursing services with a quality rating above 4 stars. The list is sorted by highest rating and identifies the provider name, zip code, the service, and the rating. This scenario is ideal for Healthy Aging because like any other service, patients seeking healthcare services are also interested in receiving quality care and it is beneficial to them to know how these companies rate on a scale. We can use this data to appeal to our patients by enabling them to make informed decisions on the quality of their healthcare needs and to ensure we are referring our patients to the best possible providers.

It is also vital for Healthy Aging to be able to pinpoint patients related to specific providers. The query we used to exemplify this need was to generate a list of patients under the care of the provider Nexus Home Health, Inc. The query identifies each patient's ID, name, and phone number. This scenario would occur when there is a question about which patients are registered to the provider through our referral agency. The information would be relevant to Healthy Aging in case we would like to follow up with the patients in regards to their experience with their current provider. This would help to continue a good relationship with patients and make sure they are receiving high quality care.

Problems Encountered and Solutions

We struggled to come up with a realistic scenario using existing data sources, to minimize the amount of fictionalized data we would have to create. In the end, we re-purposed addresses from an existing business database and assigned those addresses to individuals and other businesses.

We would have struggled with SQL, but Matt uses SQL at work and was able to make sure our group stayed on track with our data tables and queries.

One of the major changes we made was simplifying insurance companies to insurance categories. Although we found a CMS data source listing all national insurance companies and plans, it would have taken an extraordinary amount of time to clean up the data and build the connection to patients and providers for potentially thousands of insurance companies. In the real world, insurance networks change so often that it would have to be verified anyway, so using categories did seem like a realistic solution. Another option would have been to change the scope of the company to a single state or locality, and limit providers and insurance companies to those zip codes. However, we already had the patient data created, and we chose to keep our company national.

Like probably everyone else, during the last half of the project, time management became difficult as our work schedules were upended by COVID-19. Several of us work in healthcare, and others' work was also impacted heavily. We did our best to split up the tasks and use Google Docs to coordinate, so that each person could participate when they could. We also made special effort to be flexible in the timing of our group meetings via WebEx and Zoom, and everyone was able to attend every meeting despite scheduling challenges.

Summary

In the end, our group utilized our individual strengths and skills to create a database that successfully matches our (fictional) patients with providers who could provide them the home-based health services they would need as they transition home from the hospital or skilled nursing facility. This is an important aspect of continuity of care and a major part of preparing to discharge patients. Although along the way we had to adjust how we were to tackle this problem, we did accomplish the initial goal we set.

As we developed this project, we had to adapt to the provider data we had. We originally planned to include Durable Medical Equipment and Medical Supplies as services needed (since in the real world, those are often important things to set up to provide on-going care and recovery at home). However, those services were not included in the data from CMS, and it was rather

daunting to create that data for over 9000 home health providers. Instead, we chose to exclude those services. In the real world, those would be separate companies anyway, so it seemed like a realistic decision.

As discussed above, we also had to simplify insurance information. Again, this was based on finding a realistic solution to a problem, that wouldn't impact our ability to meet the requirements of the assignment.

By completing this project, we learned from each other, worked together to tackle the problems even when we each, individually, might have chosen a different solution to one or more of them, and learned to trust ourselves as we applied our new (to most of us) ERD and SQL skills and knowledge. In the end, that accomplishment is the one that will live on after this semester ends.