**D205 Data Acquisition: Performance Assessment**

Andrew Fagundes

Department of Information Technology, Western Governors University

Dr. William Sewell

February 26, 2024

Table of Contents

[Part A: Research Question 3](#_Toc159941007)

Part B: Entity Relationship Diagram……………………...................................................3

Part C: SQL Query Answering Research Question……………………………………….5

Part D: How often should CSV file be used? How is it relevant in the business?...............6

Part E: Panopto Video Link……………………………………………………………….6

Part F: References…………………………………………………………………………6

**Part** **A:** **Research** **Question**

For part A, the following research question was created: “How many of each internet services are used in the Providence County of RI?” **A1:** This research question involved the use of the tables *customer, services* and *location.* The *services* table had the different Internet services. The *location* table had the county information. The *customer* table was used to join the *services* table together through its primary key, customer\_ID, along with the foreign key from *services* table, customer\_ID. The *customer* table was used to link all of the information as it is the main table with five foreign keys that are used to link the other tables: *services, payment, location, job,* and *contract.* All the columns used had text data types.

**Part B: Entity Relationship Diagram**

For part B, pgAdmin was used to create the new *Services* table from the add-on CSV file. To create the table, the following SQL code was used:

CREATE TABLE services

(

customer\_id text NOT NULL,

internet\_service text,

phone text,

multiple text,

online\_security text,

online\_backup text,

device\_protection text,

tech\_support text

PRIMARY KEY (customer\_id),

CONSTRAINT customer\_id\_pkey FOREIGN KEY (customer\_id)

REFERENCES public.customer (customer\_id) MATCH SIMPLE

ON UPDATE NO ACTION

ON DELETE NO ACTION

NOT VALID

);

The code above illustrates how the *services* table connects to the *customer* table through the primary key, customer\_id. A one to many relationship was established between the *services* table and *customer* table since *services* uses a foreign key, customer\_id. Once the table was created in pgAdmin, the data from the CSV file had to be inserted. The code to do this is the following:

INSERT INTO services (customer\_id, internet\_service, phone, multiple, online\_security, online\_backup, device\_protection, tech\_support)

VALUES

(‘A00088’, ‘Fiber Optic’, ‘Yes’, ‘Yes’, ‘No’, ‘No’, ‘No’, ‘No’),

(‘A04204’, ‘DSL’, ‘Yes’, ‘Yes’, ‘Yes’, ‘Yes’, ‘Yes’, ‘Yes’),

(‘A04378’, ‘None’, ‘Yes’, ‘Yes’, ‘No’, ‘No’, ‘No’, ‘Yes’),

(‘A04830’, ‘DSL’, ‘Yes’, ‘Yes’, ‘Yes’, ‘No’, ‘Yes’, ‘No’),

…;

The SQL code above is used for loading the CSV file but due to its size, the import/export function of pgAdmin was used. After loading the data from the CSV file, an ERD diagram was created using the ERD tool in pgAdmin.A graph with many boxes

Description automatically generated with medium confidence

**Part C: SQL Query Answering Research Question**

To answer the research question in part A, the following SQL code was used:

SELECT services.internet\_service, COUNT (\*)

FROM customer, services, location

WHERE customer.customer\_id = services.customer\_id

AND location.location\_id = customer.location\_id

AND location.county = ‘Providence’

GROUP BY services.internet\_service;

The results of the query were the following (data file attached in submission):

A screenshot of a computer

Description automatically generated

**Part D: How often should CSV file be used? How is it relevant in the business?**

There are two approaches to processing data: online transaction processing (OLTP) and online analytical processing (OLAP). Since we are analyzing data that is consolidated and historical, OLAP would be the best fit. The amount of time the add-on file should be updated is weekly or monthly, depending on how of the analysis for business decisions are made. This would make sense since it seems like the primary use for this database would be for marketing efforts in acquiring customers.

**Part E: Panopto Video**

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=7ccca823-83b0-456a-8db4-b123013b0458>

**Part F: References**

No sources, web or otherwise, were used in the production of this project and its report. All course material came from Datacamp.com.