MAINFRAME MEDICAL DATABAS

Tyler Zamski, Caitlin Mooney, Andres Gonzalez, Michael Hernandez, Professor Alan Eliscu

Abstract

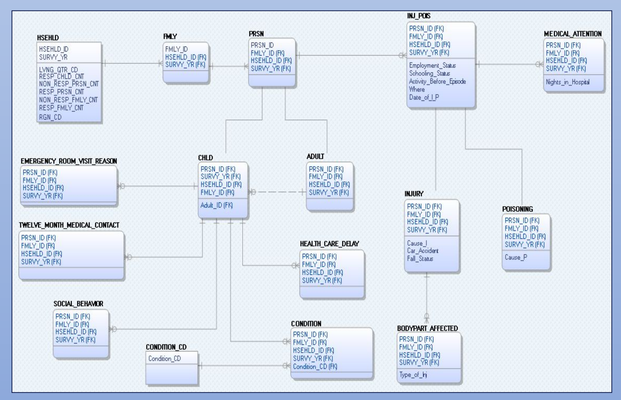
A DB2 database is being developed to test the viability of storing massive amounts of medical data within a mainframe environment. This study was done by first collecting a test dataset of a few gigabytes that contained relevant healthcare information. The data was uploaded to a z/OS flat file, where it was then transformed by an assembly code into a “cleaned” format that would work with the database specifications. Simultaneously, a database model was created in ERwin data modeler, which was used to generate DDL code that could be used by DB2’s load utility to create database tables. After populating these database tables with the cleaned data, it can now be queried using DML SQL commands.

Introduction

Mainframes have mainly been used in industry to allow data transactions to occur with incredible security and speed. However, no one has implemented medical data within a mainframe’s environment. Our project focuses on automating the process of One of the project goals is to automate the uploading of readable medical data to the mainframe. Using ERwin data modeler software, a Vista TN3270 terminal provided by IBM, in addition to Assembly Language code, we mapped the process to bring raw data from the National Health Interview Surveys, into the mainframe environment to be queried.

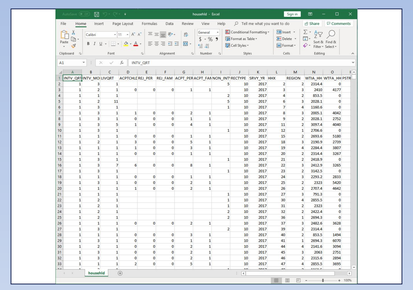
Database Architecture

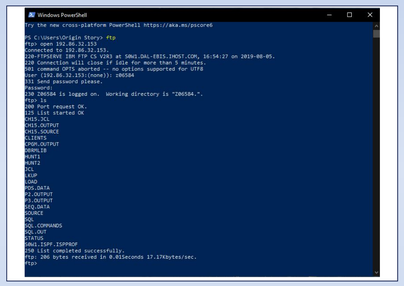
This database model was created using the ERwin Data Modeler. The hierarchy was established to allow navigation between the datasets. The Erwin software creates tables with DDL reports, which are then given addresses to be referenced similar to a glossary. This process then allows our assembly code to find the value for any data entry, and translates our raw data into readable content for analysis.



Raw Data

The data used was collected from CDC.gov. It is a collection of 6 datasets from the 2017 NHIS. One dataset includes over 600 columns and 70,000 records. The datasets are also in .csv file format.



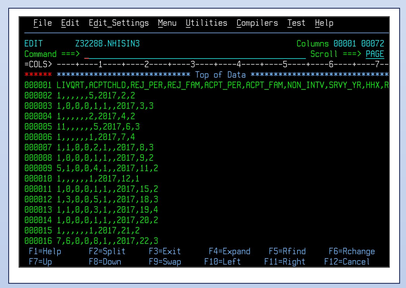


FTP

To get our data onto the mainframe we used File Transfer Protocol. FTP is universally used between operating systems and allows us to ensure the integrity of the file in the mainframe environment.

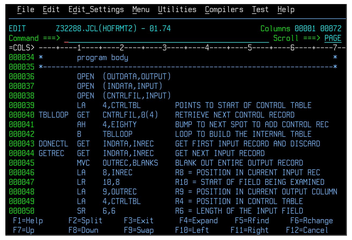
z/OS Flat File

When the file arrives on the mainframe it is misaligned and cannot be read by the database. Datasets had to be manually formatted until we chose to write code to automate this process.



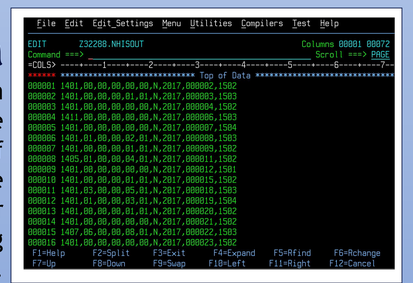
Assembly Code

The decision to use assembly language was made after testing python and c++. This code manually goes through data entries to reference our DDL reports and tables.



Formatted Data

After developing a custom assembly program we were able to automate the formatting of the data set live on the mainframe. This furthers our goal for eventually automating the entire process.



Results

The database “ACCOUNTS.BERGEN” was created within the Dallas mainframe that is publicly available for the “Master the Mainframe” Challenges. The lookup code table “LKUP” was loaded and formatted within the mainframe environment without the need to be modified locally. The process for uploading, formatting, and indexing is almost completely automated.

Conclusion

This entire exercise proves that database manipulation is possible for previously formatted medical datasets. With a small team, good code, and some time, these datasets can be formatted and used in a way that is beneficial to already existing medical standards. This process can be easily replicated in a professional medical environment for both clerical and research purposes.

References:

GOV, CDC. “NHIS - 2017 Data Release.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 2017, www.cdc.gov/nchs/nhis/nhis_2017_data_release.htm.

Ehrman, John R. Assembler Language Programming for IBM z System™ Servers. 1st ed., IBM Corporation 2015, 2015.

https://www.ibm.com/it-infrastructure/z/education/master-the-mainframe

https://erwin.com/products/erwin-data-modeler/

The content of this poster was developed and supported under a grant from the U.S. Department of Education. However, the content do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the U.S. Federal Government.