On the analysis of HL7 messages to track patient location and demographic trends

James Tsai (jjtsai@mail.smu.edu), Student and Wid Sogata (wsogata@mail.smu.edu), Student

***Abstract*—Currently there is no accurate methodology to track general demographic information of patients accessing the services of the medical center. For instance, the distances patients are traveling in order to access the services is not generally known. The most reliable source of truth can be derived from the HL7 messages since all patients that arrive at the medical must first be electronically registered and admitted. The goal is to parse this information from the HL7 messages and derive useful information and population trends.**

I. Introduction

THE HL7 messaging standard is used for transfer of clinical and administrative data between software applications used by various healthcare providers. Medical centers utilize this standard in order to exchange information between disparate systems that otherwise would not communicate with each other. For the purposes of this project, we are interested in the registration or admission of a patient. The two message types representing this event are the A01 message and the A04 message. Since the integration team archives all HL7 messages in a file-based format, it contains all the registration and admission of every patient that arrives at the medical center for any service. It is a matter of parsing and aggregating patient information contained in these message types in order to construct a baseline where we can perform further analysis.

II. Conceptual Design

In Figure 1, the data flow and components for our proposal is shown. The box labeled "Project Scope" contains the components that will be created.

To achieve our goal, we propose a 3-tier architecture. Since we are dealing with patient information, we first implement a Python script to parse only the A01 and A04 HL7 messages and grab only the zip code, age (calculated, if date-of-birth is present) and gender from the PID segment, without propagating any of the patient identifying information such as Name, Address, or Date-Of-Birth. We will also grab the date of the event from the MSH segment. We then calculate the distance of the patient based on the latitude/longitude of their zip code and the latitude/longitude of the medical center's zip code. The information we derived will then be inserted into a MySQL database on a daily basis for the previous days of HL7 messages. The trigger for this will be a daily cron job using a scheduler such as the crontab function in the UNIX operating system. The final component is to display the visualization of the data through the application Tableau.

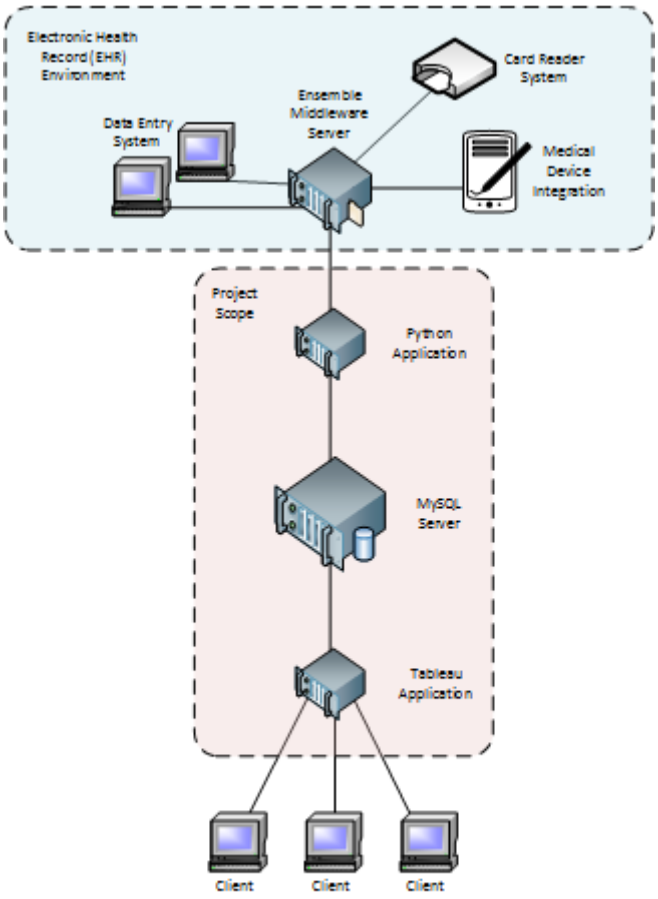

Figure 1. Conceptual Design

III. System Components

Main system components for this project:
Python ver. 2.7.10 – Python Software Foundation
My SQL ver. 5.6.27 – Oracle Corporation
Tableau Desktop ver. 9.0 – Tableau Software

IV. Technical Work

In this project we collect sample of HL7 messages data in the form of several flat files. The data from these files have exactly the same structure and format as the ones generated by Ensemble server.

The following is actual steps in completing the project:

- Install Python. Parse and extract message data using regular expression from flat files into Python objects. HL7 Messages have a limited number of levels. The top level is the message which is delimited by a carriage return. Each message contains a number of segments. Each of the segments are comprised of fields. The content of a field is either a primitive data type (such as a string) or a composite data type comprised of one or more components. Components are in turn comprised of sub-components. Segments/fields that are included in this proposal for analysis are:
  - MSH-7 (Date/Time)
  - MSH-9.2 (Message Type)
  - PID-7.1 (Date of Birth)
  - PID-8 (Gender)
  - PID-11.5 (Zip Code)
- Install MySQL database and Workbench. Design and create a database schema with necessary objects and relationships to contain HL7 data from Python script feed.
- Install Tableau Desktop software. Configure and connect Tableau software to the database and create necessary HL7 data objects for analysis and visualization.

V. Sample Data

HL7 message sample data:

MSH|^~\&|TEST|TEST|IEPRD|TEST|20151002104146|7196|
ADT^A01|185992044|T|2.3|||AL||||||
EVN|A01|20151002104146||ADT_EVENT|7196^CSHS^ED^
REGISTRARAAA^^^^^CSMC^^^^^MAH|20151002104141|
PID|1||E2528374^^^^ EPI
~200710904^^^GSMRN^GSMRN||CUPCAKE^REDVELVE
T^^^^^D~CUPCAKE^REDVELVET^^^^^L|SMITH^^^|1955
0909|F||White|25 WESTST^^PASADENA^CA^91108^USA^
P^^LOS ANGELES|LOS ANGELES|(333)4445555^P^PH^^^
333^4445555||EN^ENGLISH|MARRIED|NR|7734563|000-00
-0001|||NON-HISPANIC|^^NYC^^^||||||N||
PV1|1|ED|ER^ IE3
7^01^CAAC^D^^MAH^^^^|ER|||||EMD|||||||49103634192|SE
LF||||||||||||||||||AD|^^^CAAC^^^^^^^||20151002104000|||||||||PV
2|||||||||||||||||||N|||||||||||||Car||||||||||
AL1|1||^^|||||||||||||||
DG1|1||^ouch|||GT1|1|1229779|CUPCAKE^REDVELVET^^||
25 WEST
ST^^PASADENA^CA^91108^USA^^^LOSANGELES|(333)
444-5555^^^^^333^4445555||19550909|F|P/F|SLF|000-00-
0001|||||^^^^^USA|||Full||||||||||||||||||||||||||
UB2|||||||ZCL|||||||||||||N|||

VI. Schedule

First Prototype, October 30, 2015
Final Prototype, December 1, 2015
Submission, December 7, 2015

VII. Analysis

The system prototype will use one year's worth of HL7 data, representing one year's worth of patient activity at the medical center. The amount of information collected in one year will be substantial, and specific trends and descriptive statistics will most likely be significant. The number of A01 and A04 messages that will be analyzed across one year is approximately 2 million based on average A01 and A04 volume of 40,000 per week.

Statistical analysis to be performed:

- Patient demographics (gender, age, zip code)
- Patient volume across various months and seasons
- Patient distance from medical center

VIII. Conclusion

Once this system is in operation, insight into patient demographics trends can potentially improve hospital services. One possible example of this is the ability for the medical center to make wiser choices when making decisions on expanding the medical center services, and focusing on the population trends and choosing locations with the highest patient densities.

References

[1] Health Level Seven Implementation Support Guide for HL7 Standard Version 2.3, Health Level Seven, 1998, pp 4.9–4.10.
[2] Timothy Boronczyk, *Jump Start MySQL*, Sitepoint, 2015, ch. 1-6.
[3] Mark Lutz, *Learning Python 5th Edition*, O'Reilly Media, 2013.
[4] Albert Lukaszewski, PhD, MySQL for Python, Packt Publishing, ch.1-4.