|  |
| --- |
| **OMOP Data Set Processing**  **Steps Taken** |

David A Stumpf, MD, PhD[[1]](#footnote-1)

Downloaded data set: <https://www.ohdsi.org/data-standardization/> SynPUF% and unzipped to C:\Data\Consulting\MATTER\Neuropath\Horizon 2020\OHDSI\OMOP Test Data. This was a set of csv files. The data is simulated and has limitations. The provider NPI, patients are not real; there are missing elements such as costs, address, etc.

Downloaded SQL Query to create tables. <https://raw.githubusercontent.com/OHDSI/CommonDataModel/v5.2.2/Sql%20Server/OMOP%20CDM%20ddl%20-%20SQL%20Server.sql>. Index and constraints still need to be done. Files at <https://github.com/OHDSI/CommonDataModel/tree/v5.2.2/Sql%20Server>.

This was trimmed of the top rows and the comment rows and then saved. This was used to create tables with python code. However, there was a missing comma in one row which needed to be inserted. After that, the table were created.

The csv file names are the same as the SQL server names. However, there were csv files that were not in the SQL query to create their table.

Table uploads:

|  |  |
| --- | --- |
| Successful load | Failed for lack of table |
| care\_site  cdm\_source  concept  concept\_ancestor  concept\_class  concept\_relationship  concept\_synonym  condition\_era  condition\_occurrence  cost  death  device\_exposure  domain  drug\_era  drug\_exposure  drug\_strength  location  measurement  observation  observation\_period  payer\_plan\_period  person  procedure\_occurrence  provider  relationship  visit\_occurrence | achilles\_analysis.csv  achilles\_heel\_results.csv  achilles\_results.csv  achilles\_results\_derived.csv  achilles\_results\_dist.csv  concept\_hierarchy.csv |

At this point, the SQL server tables are:

 database mdf=15.3 mB; log= 38.4 mB. After shrinking: 15.3 and 8.3 mB.

MANUAL UPLOAD

Put cvs files in the import directory. Then run these queries in the Neo4j browser:

LOAD CSV WITH HEADERS FROM 'file:///omop/Person.csv' as line FIELDTERMINATOR '|' create (p:Person{person\_id:toInteger(line.person\_id),year\_of\_birth:toInteger(line.year\_of\_birth),month\_of\_birth:toInteger(line.month\_of\_birth),race:toString(line.race),ethnicity:toString(line.ethnicity),state:toString(line.state)})

LOAD CSV WITH HEADERS FROM 'file:///omop/Provider.csv' as line FIELDTERMINATOR '|' create (p:Provider{provider\_id:toInteger(line.provider\_id),NPI:toInteger(line.NPI),care\_site\_id:toInteger(line.care\_site\_id)})

LOAD CSV WITH HEADERS FROM 'file:///omop/CareSite.csv' as line FIELDTERMINATOR '|' create (p:CareSite{care\_site\_id:toInteger(line.care\_site\_id),care\_site\_type:toString(line.care\_site\_type)})

create index on :Person(person\_id)

create index on :Provider(provider\_id)

create index on :CareSite(care\_site\_id)

LOAD CSV WITH HEADERS FROM 'file:///omop/PersonProvider.csv' as line FIELDTERMINATOR '|' match (p:Provider{provider\_id:toInteger(line.provider\_id)}) match (people:Person{person\_id:toInteger(line.person\_id)}) merge (people)-[r:PersonProvider{visit\_ct:toInteger(line.visit\_ct)}]->(p)

1. © David A Stumpf, MD, PhD 2020 [↑](#footnote-ref-1)