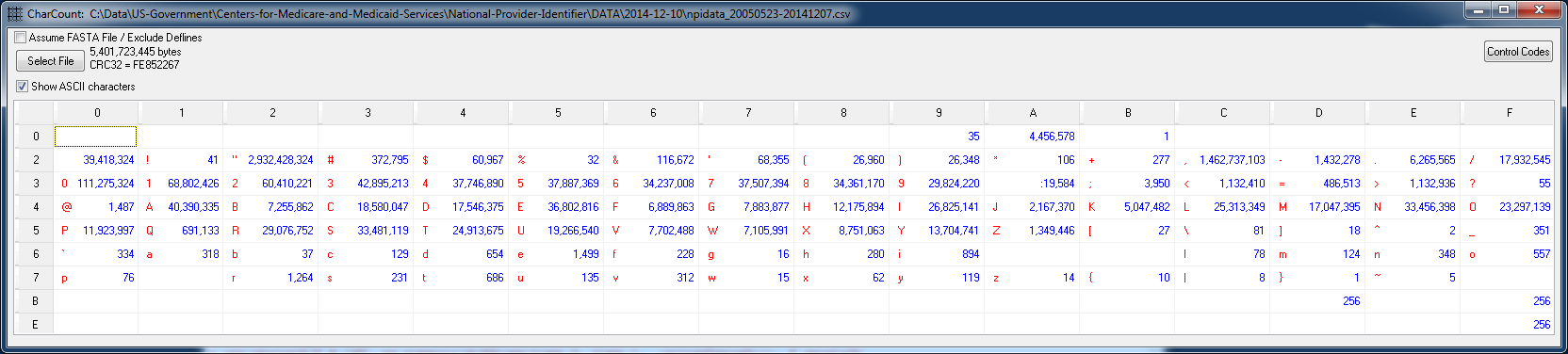
On efg’s prime2 PC:

Run CharCount utility (see <https://github.com/EarlGlynn/CharCount>) on latest npidata file.



35 tabs (hex 09)

Use **CentOS65-BASE** Virtual Linux box for the following:

Update script: **1-Recode.bash**

Review and run the script:

**[239 efg localhost 2014-12-28 20:03:22 /mnt/hgfs/Data/US-Government/Centers-for-Medicare-and-Medicaid-Services/National-Provider-Identifier]**

dos2unix 1-Recode.bash

dos2unix: converting file 1-Recode.bash to UNIX format ...

Modify script to have new directory/file names.

**[241 efg localhost 2014-12-28 20:06:58 /mnt/hgfs/Data/US-Government/Centers-for-Medicare-and-Medicaid-Services/National-Provider-Identifier]**

cat 1-Recode.bash

#! /bin/bash

set -x

# UMKC Center for Health Insights

# Earl F Glynn, 2014-12-01.

cd /mnt/hgfs/Data/US-Government/Centers-for-Medicare-and-Medicaid-Services/National-Provider-Identifier/DATA/2014-12-10

# Show lines with tabs

grep -n $'\x09' npidata\_20050523-20141207.csv > tab-lines.txt

# Changes tabs to spaces

tr $'\x09' ' ' < npidata\_20050523-20141207.csv > npidata\_20050523-20141207-notabs.csv

# Now run 2-CMS-National-Provider-Identifier-Rewrite.R

**[244 efg localhost 2014-12-28 20:08:32 /mnt/hgfs/Data/US-Government/Centers-for-Medicare-and-Medicaid-Services/National-Provider-Identifier]**

./1-Recode.bash

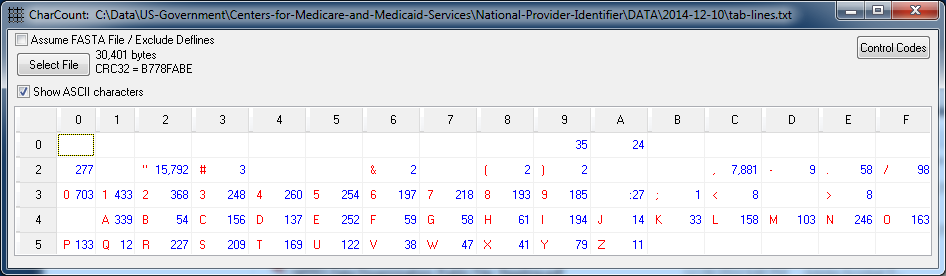
+ cd /mnt/hgfs/Data/US-Government/Centers-for-Medicare-and-Medicaid-Services/National-Provider-Identifier/DATA/2014-12-10

+ grep -n ' ' npidata\_20050523-20141207.csv

+ tr ' ' ' '

[245 efg localhost **2014-12-28 20:13:59** /mnt/hgfs/Data/US-Government/Centers-for-Medicare-and-Medicaid-Services/National-Provider-Identifier]

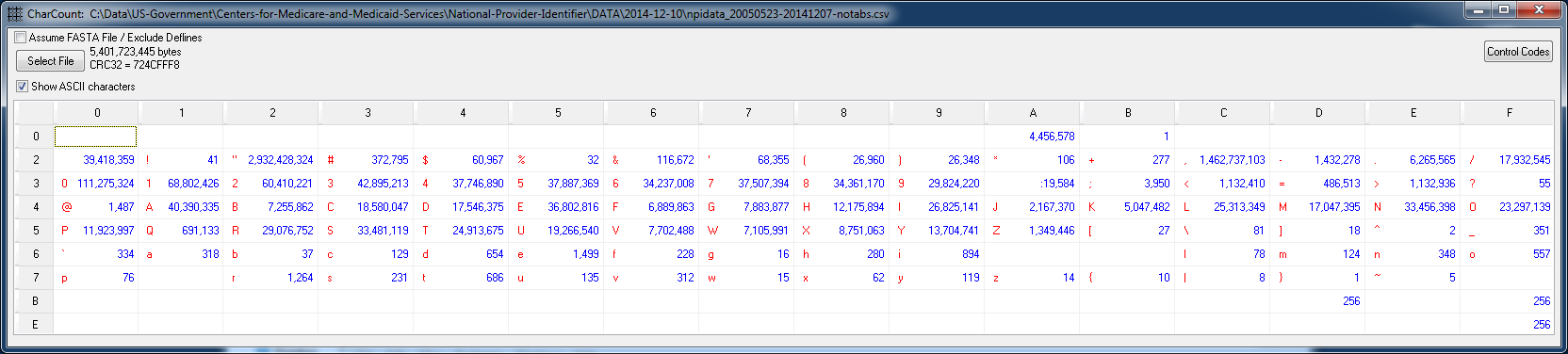
Using CharCount utility, verify all 35 tabs are accounted for in **tabs.line**:



Review **tabs-line.txt** and convince yourself that all tabs can be safely converted to spaces.

Below is proof 35 tabs removed, and 35 spaces inserted.

Verify notabs file has no tabs:



After conversion of all tabs, we can re-write the file with a tab delimiter and remove ~3 billion quotation marks.

The number of tabs above (1,448,512,944) does not make sense.

2,905,858,284 double quotes / 2 = 1,452,929,142 tabs.

National Plan and Provider Enumeration System (NPPES) Downloadable File  
<http://www.cms.gov/Regulations-and-Guidance/HIPAA-Administrative-Simplification/NationalProvIdentStand/DataDissemination.html>

Download National Provider Identifier (NPI) Downloadable File  
<http://nppes.viva-it.com/NPI_Files.html>