0. all TSQL operates on ghrii2b2.i2b2demodata

1.Move 12 csv files to the ghrii2b2 e drive, csv data directory.

* 1. One for each year
  2. Naming convention is for\_i2b2\_load\_yyyy.csv
  3. We will load 1999 through 2011.

1. Make sure the observation\_fact\_reload exists on i2b2demodata and has been scripted to resemble the observation\_fact
   1. I did this with the SQL manager Create To option and changed the target file name to observation\_fact\_reload.
2. Make sure there are no indexes on the observation\_fact\_reload table.
   1. Delete the indexes if they do exist.
3. Make sure there is no data in the observation\_fact\_reload table.
   1. Truncate table observation\_Fact\_reload

GO

1. Bulk load the observation\_fact\_reload table with the 12 csv files:
   1. Iterate the following code for each year—attached are two iterations for 1999 and 2000.
   2. DBCC SHRINKFILE(i2b2demodata\_log,1509)

GO

BULK INSERT observation\_fact\_reload

from 'S:\for\_i2b2\_load\_1999.csv'

with (

firstrow=2, --Skip the first row because it contains field names

fieldterminator=',', --Data values separated by commas

rowterminator='\n', --Data rows separated by line breaks

maxerrors=0, --Stop loading if any errors occur

keepnulls, --Keep null values (from the CSV) when inserting

errorfile='S:\i2b2\_bulk\_insert\_errors.txt' ) --Dump error messages here

GO

checkpoint

Go

DBCC SHRINKFILE(i2b2demodata\_log,1509)

GO

--A checkpoint forces the transaction log to write all changes to disk so

--the transaction log can be truncated.

--Shrink the database freeing up space that might have gotten chewed up by

--logging during the insert process.

BULK INSERT observation\_fact\_reload

from 'S:\for\_i2b2\_load\_2000.csv'

with (

firstrow=2, --Skip the first row because it contains field names

fieldterminator=',', --Data values separated by commas

rowterminator='\n', --Data rows separated by line breaks

maxerrors=0, --Stop loading if any errors occur

keepnulls, --Keep null values (from the CSV) when inserting

errorfile='S:\i2b2\_bulk\_insert\_errors.txt' ) --Dump error messages here

GO

checkpoint

Go

DBCC SHRINKFILE(i2b2demodata\_log,1509)

GO

ETC….repeat this step for 2001 through 2010 csv files.

For instance:

from 'S:\for\_i2b2\_load\_2001.csv'….

from 'S:\for\_i2b2\_load\_2002.csv' etc.

1. Script indexes from Observation Fact
2. Stop Apache Server.
3. Drop indexes via right click delete.
4. Delete observation\_fact\_backup
   1. Truncate table observation\_fact\_backup
   2. Right click delete
5. rename observation\_fact to observation\_Fact\_backup
6. rename observation\_fact\_reload to observation\_Fact

1. Build 6 indexes on observation\_fact

USE [i2b2demodata]

GO

/\*\*\*\*\*\* Object: Index [OF\_IDX\_ALLObservation\_Fact] Script Date: 10/31/2011 19:24:33 \*\*\*\*\*\*/

CREATE NONCLUSTERED INDEX [OF\_IDX\_ALLObservation\_Fact] ON [dbo].[Observation\_Fact]

(

[Encounter\_Num] ASC,

[Patient\_Num] ASC,

[Concept\_Cd] ASC,

[Start\_Date] ASC,

[Provider\_Id] ASC,

[Modifier\_Cd] ASC,

[ValType\_Cd] ASC,

[TVal\_Char] ASC,

[NVal\_Num] ASC,

[ValueFlag\_Cd] ASC,

[Quantity\_Num] ASC,

[Units\_Cd] ASC,

[End\_Date] ASC,

[Location\_Cd] ASC,

[Confidence\_Num] ASC,

[Sourcesystem\_Cd] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, SORT\_IN\_TEMPDB = OFF, IGNORE\_DUP\_KEY = OFF, DROP\_EXISTING = OFF, ONLINE = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [Index]

GO

USE [i2b2demodata]

GO

/\*\*\*\*\*\* Object: Index [OF\_IDX\_Concept\_PatNum\_Date] Script Date: 10/31/2011 19:24:41 \*\*\*\*\*\*/

CREATE NONCLUSTERED INDEX [OF\_IDX\_Concept\_PatNum\_Date] ON [dbo].[Observation\_Fact]

(

[Concept\_Cd] ASC,

[Patient\_Num] ASC,

[Start\_Date] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, SORT\_IN\_TEMPDB = OFF, IGNORE\_DUP\_KEY = OFF, DROP\_EXISTING = OFF, ONLINE = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [Index]

GO

USE [i2b2demodata]

GO

/\*\*\*\*\*\* Object: Index [OF\_IDX\_Encounter\_Patient] Script Date: 10/31/2011 19:24:56 \*\*\*\*\*\*/

CREATE NONCLUSTERED INDEX [OF\_IDX\_Encounter\_Patient] ON [dbo].[Observation\_Fact]

(

[Encounter\_Num] ASC,

[Patient\_Num] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, SORT\_IN\_TEMPDB = OFF, IGNORE\_DUP\_KEY = OFF, DROP\_EXISTING = OFF, ONLINE = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [Index]

GO

USE [i2b2demodata]

GO

/\*\*\*\*\*\* Object: Index [OF\_IDX\_Modifier] Script Date: 10/31/2011 19:25:08 \*\*\*\*\*\*/

CREATE NONCLUSTERED INDEX [OF\_IDX\_Modifier] ON [dbo].[Observation\_Fact]

(

[Modifier\_Cd] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, SORT\_IN\_TEMPDB = OFF, IGNORE\_DUP\_KEY = OFF, DROP\_EXISTING = OFF, ONLINE = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [Index]

GO

USE [i2b2demodata]

GO

/\*\*\*\*\*\* Object: Index [OF\_IDX\_Start\_Date] Script Date: 10/31/2011 19:25:16 \*\*\*\*\*\*/

CREATE NONCLUSTERED INDEX [OF\_IDX\_Start\_Date] ON [dbo].[Observation\_Fact]

(

[Start\_Date] ASC,

[Patient\_Num] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, SORT\_IN\_TEMPDB = OFF, IGNORE\_DUP\_KEY = OFF, DROP\_EXISTING = OFF, ONLINE = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [Index]

GO

USE [i2b2demodata]

GO

/\*\*\*\*\*\* Object: Index [OF\_IDX\_UPLOADID] Script Date: 10/31/2011 19:25:23 \*\*\*\*\*\*/

CREATE NONCLUSTERED INDEX [OF\_IDX\_UPLOADID] ON [dbo].[Observation\_Fact]

(

[UPLOAD\_ID] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, SORT\_IN\_TEMPDB = OFF, IGNORE\_DUP\_KEY = OFF, DROP\_EXISTING = OFF, ONLINE = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [Index]

GO

1. Start Apache Server
2. If I2B2 is working (someone logs in and confirms that a query returns counts) then you are done. Leave backup file because we have the space.