



## Student Symposium Q&A and Helpful Tips

### Why can't I create files in my local directory?

We ran into issues with students creating very large SAS datasets and storing them in their personal folders. The personal folders are sized for small result files of a total size of a few gigabytes. The intent was for the teams to store their large intermediate results in their HDFS team (e.g., /contest/teamxx) and HDFS user folders (e.g. /user/student01).

There are two easy options for storing SAS datasets on HDFS. The easiest approach is to create a folder in HDFS using the `proc hadoop` and `HDFS mkdir` command. Then the folder can be referenced using a libname with SPDE engine. For example, the following code creates a folder called 'results', defines a libname called 'results' and then uses it as the output of a data step:

```
proc hadoop username='student01' password='password';
    hdfs mkdir='contest/teamxx/results';
run;
libname results SPDE 'contest/teamxx/results';
.....
data results.result1;
.....
```

The space taken up by files in your directory should be less 2 gigabytes. Please take action to move your large SAS datasets to HDFS. This can be done either running a data step to copy the dataset over to HDFS or by deleting the local copy and rerunning the code to store the results to and HDFS folder rather than your local folders.

### Why does it take so long to read my data?

Along with the concern about storing large datasets in your local directory are some of the special issues with working with large datasets in HDFS and working with Hadoop. Keep in mind that Hadoop is running in a separate server, and that results from jobs may be copied to SAS if the output is not another dataset in HDFS. If that output is large, it can causes issues. It is important to remember that some actions causes SAS to copy the entire dataset over to SAS and then process. One of those actions is `proc print`, even with a limit on the number of records to print. It will first copy all records in and then show you only the specified number. It is important to first create a small summary and print that. For example you want to print 10 records:

```
proc print data=engngra2.english_all_2grams (obs=100);
run;
```

There are no easy solutions. The best approach is to select a sample based on known values. It may also be possible to run `proc freq` to dataset on HDFS and examine the size of the results using `proc contents`. Keep in mind that working with large datasets requires special care.





# SAS® GLOBAL FORUM 2016

## IMAGINE. CREATE. INNOVATE.

April 18-21 | Las Vegas, NV | [sasglobalforum.com](http://sasglobalforum.com)



### How can I create a persistent data source?

The installation we have setup provides two ways to create datasets. On the Hadoop file system (HDFS) there is a folder for each team member and a folder for each team. You can define these folders as libraries or create subfolders and store datasets there. If you have looked at the tips, you will see how to create any new folders using the `proc hadoop hdfs` command. You can also create and store datasets in your personal folders using SAS® Studio.

If you are experienced with creating SAS datasets, there is a difference when storing datasets on the HDFS file system. You need to change the libname. Take a look at the instructions on how to store a SAS dataset in HDFS using the SPDE engine.

I would also take a look at the instructions on how to reference the existing data sources we copied to HDFS. The FARS datasource needs to be accessed using a libname with SPDE engine (see example). For the other datasources you will need to supply format information using the `proc hdmf`.

### I keep getting an error “No Sufficient Permission” when I try to create a new folder under our team folder. What is wrong?

You will need to supply your user and password in order to run under your authority for `proc hadoop`. For example:

```
proc hadoop username='rayhooker' password='password';  
hdfs mkdir=/user/rayhooker/testagain';  
run;
```

where you put in your user and password. We are actually going to send out a note. You would want to do the same for any libname pointing to an HDFS folder if you want to create new datasets. When accessing the data sources we loaded, it does not matter because they are readable by everyone. If you are creating new datasets under your HDFS team or your HDFS user folder, you will need your user and password on the libname.

For example:

```
libname teamdata spde '/contest/team00/data' user='rayhooker' password='password' hdfs host=default;
```

Also keep in mind that you will need to use a `proc hadoop hdfs mkdir` to create any new folders before storing to them. Such for the data folder for team00 is created as follows:

```
proc hadoop username='rayhooker' password='password';  
hdfs mkdir='/contest/team00/data';  
run;
```

You can create and write to folders in your personal HDFS user folder or to your team folder.



# SAS® GLOBAL FORUM 2016

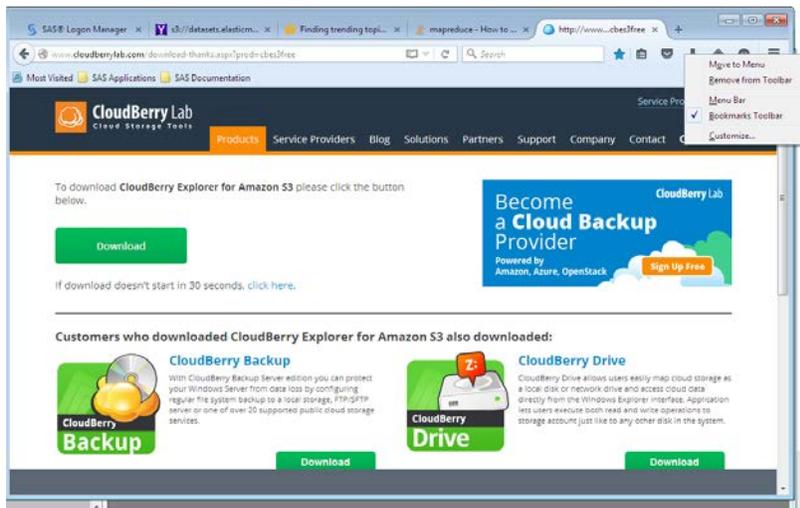
IMAGINE. CREATE. INNOVATE.

April 18-21 | Las Vegas, NV | sasglobalforum.com



## Can I use SAS® Enterprise Miner™ or must I use SAS® Studio?

SAS® Enterprise Miner™ is installed. The client may be accessed from you AWS Workspace from a link on the Firefox Bookmarks Toolbar. If it is not visible, simply right click on menu bar and select to display the Bookmarks Toolbar.



Click on the SAS Applications and select “Enterprise Miner”. It will allow you to launch Java Web Start Launcher to install the client locally.

Please note that if you want to access any of the tables from large data sources that we have loaded, you will need to contact us to register it as a library in SAS Metadata.

# SAS® GLOBAL FORUM 2016

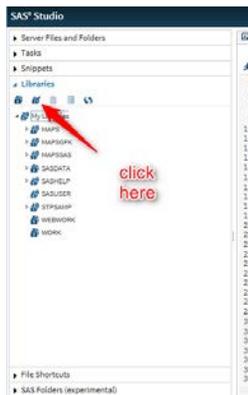
IMAGINE. CREATE. INNOVATE.

April 18-21 | Las Vegas, NV | [sasglobalforum.com](http://sasglobalforum.com)

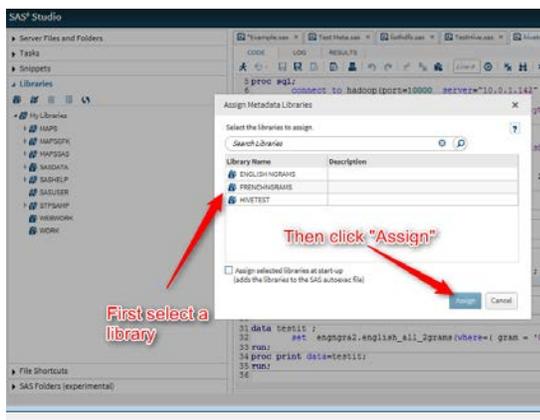


## How do I access SAS Metadata Libraries from SAS® Studio?

SAS provides the ability to register libraries in SAS Metadata, so that they are available to everyone without having to create a libname. In order to access these from SAS® Studio, you need to assign them. First expand the “Libraries” tab and click on the icon to “Assign SAS Metadata Libraries.”



Next select the library to assign.



At that point you can select any of the tables in that libname by simply typing the libname, a period and the table name.

If you have any questions, please contact us at [symposium@sas.com](mailto:symposium@sas.com).