2-28-2024 Townhall Q&A

1. **When will the HIVE tables be visible for R coding?**
   A- This will be available soon as we have completed the migration from Hive to Glue. We are still working through some of the latency issues.

2. **Can you clarify if you want one decommissioning survey per organization or by contract? If we have more than one contract that uses SAS Viya, do we submit multiple surveys?**
   A- For organizations that have multiple contracts/organizations in HARP, we are asking for one survey per contract/organization that has SAS Viya users. If that is the case for you, then we would ask that you submit multiple surveys.

3. **What are the next steps after the survey for SAS Viya? I completed the survey but did not know all the information. At some point hope we get to work with someone who can provide detailed info from within SAS Viya (i.e., list off all jobs running in SAS Viya by org and last run date, etc...).**
   A- We will be sharing next steps after we have a chance to analyze the survey. We are planning to work with the individual organizations if needed.

4. **What is the timeline to transition CASLIBs schemas such as Internet Quality Improvement and Evaluation System (iQIES)-CASPER into Databricks? Most of our work starts with pulling data from there and we cannot transition any of our code until we have access to these schemas in Databricks.**
   A- There is no timeline now. We are actively working with the IT team on getting the data that is currently in CASLIBS, to be made available, within the Centralized Data Repository (CDR) outside of those CASLIBS.

5. **Does this apply to files on the workbench? I get access denied when trying to delete comma separated values (CSVs) from the workbench.**
   A- If you are writing these CSV files using a Databricks notebook to your workbench Simple Storage Service (S3) bucket and you are trying to use SAS Viya to delete those CSV files, we would recommend not using this route. If you create something with Databricks and put it under your workbench and want to delete it, we recommend using Databricks to simply drop the CSV files, or you can utilize the new unified file management (UFM). UFM is the replacement for FileCloud which has specific access directly to your workbench or S3 bucket, and you could manage files that route as well like you would with FileCloud.

6. **Is Databricks already available to start our transition?**
   A- Yes. CCSQ Databricks was rolled out in 2023. You just must request the Databricks role within HARP for the specific organization and your account will automatically be created and added to the specific group. Then, you can log in and begin using it.

7. **Can someone who did not write out the table change the owner?**
   A- Only if a user wrote the table on the SAS side as that uses a service principal account

for all users within the group. Person A could write the table in SAS and Person B could run the alter owner table macro in SAS. On Databricks, only the user that wrote the table can alter the table.

8. **How could we check who owns the table?**
A- Users can check this through the Catalog Explorer in Databricks.

9. **Why is SAS being decommissioned?**
A- There are several factors in the decommissioning of the SAS version that we have in the CDR. The major factor is that it is coming to the end of its life on Tuesday, December 31, 2024. There is a potential for extended support until December 2025.

10. **With SAS Viya being decommissioned, by any chance do you provide desktop SAS for SAS users, if needed?**
A- There are currently no plans for providing a desktop version of SAS.

11. **Contractors use output data to do more reporting or analysis locally. We need to know where the output in Databricks is stored and how to download it? Can it be in database or SAS format (with data type embedded, not just CSV). Teach us how to work as a team - in coding, store codes, share codes, work on codes together. CMS may need to audit our work.**
A- We suggest reviewing the current training offerings we have available in our Training Hub under the Databricks tab for best practices in starting off using Databricks.

12. **Can the deadline be postponed for another day or two?**
A- The survey has closed as of 11:59 p.m. on Monday, March 5, 2024. If you have not submitted on behalf of your organization, please reach out to #help-service-center-sos and submit an inquiry to the D&A team.

13. **We have multiple contracts. When can we use tables created in another schema? Otherwise, we will have to create duplicated common tables in multiple schemas or, is it prohibited because of data government policy?**
A- If there are two separate contracts, they are specific to their own contract, even if you are working on both. If you need certain type of access across contracts within your database that can be requested and would need approval of CMS to grant that so you can read from other databases.

14. **I usually get a cascade of error messages when I ask for permissions. What would cause this?**
A- It depends on what the error messages are and what you are attempting. Contact the Help Desk for assistance.

15. **Will the notebooks that Derek used in the demonstration be made available to the user community?**
A- This has been made available within the workspace under cdr-pda-qdas > public_scripts > Table_Owner_issues_Demo.

16. **I have found the Catalog Explorer functionality to be useful, but it has a limit to the number of tables it will display, which hinders its effectiveness. Is the display limited to 10k tables?**
A- No. Currently, there are tables within the Catalog Explorer with databases with fewer tables that are slow. We are currently working to pinpoint the issue and offer a solution or workaround.

17. **Will we have access to the SQL functionality in Databricks? Currently those functions are not active.**
A- Yes, currently SQL can be written in the cell of any notebook. However, the dedicated SQL editor will be enabled in a future release.

18. **Can we have more time to answer the polls in general?**
A- We appreciate this feedback/question and will take this into consideration future Townhalls and Data Camps.

19. **Can you speak more about the Hive to Glue migration? We are aware that there was an impact on certain tables (rearranged columns, missing data) etc. How were these issues resolved, and should we be concerned that there may be additional impacts that we may just not have noticed?**
A- We made a copy of all metadata in Hive, and we put it in Glue. As a part of that migration process, it was identified that any tables that were created from flat files like CSV or text, anything other than Parquet, they had the column orders rearranged to alphabetical order. This does not matter in Parquet, but it does matter for those flat files. We copied what it was in Hive and rearranged it. The data itself was not impacted at all and it was a purely metadata issue. Once we identified it we ran a script in all tables that were identified to be sourced from flat files and it resolved the issue. If you are still having an issue with any column order issue, just let us know and we will run it again on your database or tables.

20. **Is there any way I can find the size of the whole database/schema either in SAS Viya or in Databricks? Do we have permission for that level of information?**
A- If you are referring to file size as in gigabytes or megabytes or are table counts, we can accommodate either. If you contact us, we can look it up for you.

21. **I am trying to use Databricks to delete and then get the error in Databricks. What would cause this?**
A- This would happen if another user created that table within Databricks or the table was created using SAS.

22. **Why not upgrade to the new version or the new workbench? The Visual Studio product is excellent. (Is our workbench available in Databricks? Currently, we use SAS Viya to view/organize our data and files.)**
A- To clarify, what users call the workbench that is available in SAS is a folder in, Amazon Web Service (AWS) S3 and it is a place where you can store and access all your files. Right now, it is accessible through SAS and Databricks. We can provide some code that you can run to access it through Databricks. With the decommissioning of SAS, you

are not going to lose your workbench because it is just a folder that you can connect SAS, Databricks, any tool, UFM is connected to it.

23. **If SAS Viya is going away, as a QIES Business Intelligence Center (QBIC) State Agency General User (SAGU) user, I found out I could not get access to Databricks at this time. When will this be available to us so we can transition our code to Databricks?**
A- That is correct QBIC users do not have access to Databricks currently. CMS is working with iQIES leadership to provide additional information on next steps in the coming weeks.

24. **Is our workbench available in Databricks? Currently, we use SAS Viya to view/organize our data and files.**
A- Yes, it is accessible through SAS and Databricks.

25. **Does the response for CAS also apply to the End Stage Renal Disease Quality Reporting System (EQRS) Redshift schemas, which being that it is in work and will be available later? We cannot begin moving forward with transition until we can access those schemas.**
A- Yes. We must open that connection up on the Databricks side. It does require a few additional steps within that process to open that connection and make any direct connections to Red Shift available, within the Databricks CDR platform itself. If you have concerns about it you can reach out and we could provide more information and details, as well as potential timelines.

26. **If I wanted to transition my SAS code to Databricks using R language, it will not be fully functional for another three-four months in the 10 months' time we have to transition before SAS Viya is no longer supported on December 31, 2024?**
A- R language is already available. The connection to the CDR will become available sooner than three to four months. The move from Hive to Glue was done to open that connection up. In addition to addressing the latency issues, there are other things going on in parallel right now to set up that ability for R to connect to the CDR like Python and Spark SQL can.

27. **We have millions of lines of code to support our CMS clients and decommissioning SAS would be a monumental change and effort to both CMS and contractors.**
A- We fully understand and appreciate that this will be a heavy lift for many members of our community. The results of the survey will be used to identify the next steps so we can support the transition away from SAS going forward.

28. **Is it recommended to continue to use the S3 workbench for data archival purposes or is there an alternate archival tool that is recommended to use with Databricks?**
A- This depends on the type of data. If you want to simply store data within a database table, you can use your org-specific database within the CDR. Or you can use your S3 workbench to store other types of files and things as Databricks can connect directly to their S3 workbench and read/write from and to it just like SAS Viya can. You can also

use the new UFM tool that is replacing FileCloud for internal team storage and to have network shares to connect directly to their S3 workbench for uploading files.

29. **If Viya is being decommissioned, is there a way for organizations that have and use base SAS in-house to access the data from the CDR to pull into our systems and run our SAS code there?**
A- No, the data cannot leave the CMS Federal Information Security Modernization Act (FISMA) boundary.

30. **Other CMS teams provide virtual machines with SAS Enterprise Guide to access their databases (e.g., Integrated Data Repository (IDR)), has that option been considered?**
A- Currently, we are not considering this option.

31. **When will this be available to us so we can transition our code to Databricks? While SAS Viya is still supported and with the decommission of FileCloud, will the code path used in SAS Viya to access our S3 bucket data change?**
A- The code path will not change for Databricks and accessing any workbench/S3 bucket directory. You can also use the UFM tool just as you did with FileCloud to access your workbench/S3 bucket directory for uploading files.

32. **Do we have a definitive date for the refresh of CDR data? It has been delayed by a few weeks already. We are asking specifically about Patient Assessment Data (Assessment_iQIES) and Provider Data (Provider_iQIES).**
A- The iQIES data refresh was complete as of Monday, March 4, 2024.

33. **The range of statistical models supported by SparkR, Sparklyr, and PySpark is limited. Consequently, most of our existing modeling code is incompatible with Spark. To accommodate the needs of multiple users (for instance, 35 users), we will need clusters that support the regular use of R and Python for our computations. Are there any plans for this?**
A- The CCSQ D&A team plans to evaluate with CMS leadership the need for clusters to support the volume necessary for the community to perform their everyday work without disruption.

34. **What about what happened to the Parquet files? We will need training in where to store large datasets and what to do when the workbench is no longer there and now that files can disappear from issues with Glue agent.**
A- Parquet files still exist as the underlying file type on all database tables. The workbench is only an S3 bucket mounted onto the SAS servers and is accessible through a Databricks Notebook. Your organization databases still exist and will remain and be available to store data tables within them.

35. **Can we get clarification on what functionality will persist past December 2024 to help address the concerns around decommission or is it really we are stating outright the cut off is now December 2024.**
A- The end of full support by the vendor of CCSQ's version of SAS Viya scheduled for

December 2024. Ony an extended support may be available after that date. What that support would look like, we currently do not have any details.

36. **Why is the SQL functionality in databricks greyed-out? When would you use SQL vs data ingestion vs notebook? but why is there a SQL functionality integrated into databricks? Is there an advantage to using that?**
A - You can use SQL in a notebook, starting a notebook cell with %sql or selecting SQL from the language drop down. The SQL warehouse which is a functionality that we are currently working on making available to end users. More information on that to come soon.