

4-11-2024 Townhall Q&A

1. **Q- Is Python training available to all?**

A- We currently have Python training available within the [Training Hub](#). You can review some example Python programs and notebooks from Session 5 of the Fall 2023 [Data Camp](#) by going in the Databricks Notebook, Go to workspace -cdr-pda-qdas folder.

2. **Q- How do we get access to Confluence?**

A- Instructions on getting Confluence access are posted here: <https://qnetconfluence.cms.gov/display/HS/QualityNet+Atlassian>

3. **Q- Are there any updates for access to the HIVE tables using R?**

A- That feature is being prioritized with CMS for the upcoming Program Increment. We will update users once we have more information.

4. **Q- The SAS Viya user interface (UI) lets us see all our internally created quarterly folders. Those are saved in our project dedicated Amazon Web Services (AWS) Simple Storage Service (S3) bucket/workbench. This helps us easily navigate through programs, logs, results, and datasets. Databricks allows us to save data to the S3 workbench but does not have the UI to navigate the folders/data. Is there a resolution to this?**

A- Databricks has a future feature called Volumes that will give us the ability to do that, and we are working proactively to enable that. In the meantime, that workspace is pointed to an S3 bucket that should be the same one that you can see in Managed File Transfer (MFT), and you should be able to do that there.

5. **Q- Is the Portal able to take the new format of data use agreements (DUA) now?**

A- Yes, it will with the new release of Portal. [Instructions](#) on how to format DUAs are available on Portal.

6. **Q- Is the new Portal a *replacement* of the current Data & Analytics (D&A) Confluence? Or is it an augmentation of?**

A- The D&A Portal will eventually replace the current Confluence space. Confluence is currently still available and there will not be any changes to Confluence until two weeks after the release of Portal. We invite you to visit the new [Portal](#) site to find training materials and content.

7. **Q- What is the link to the D&A Portal?**

A- <https://ccsqdataanalytics.cms.gov/>

8. **Q- Just to confirm, the Databricks environment is available now and we can work with trainers to transition our SAS code to SQL?**

A- Yes, Databricks is available for CDR users, except those who are members of the Internet Quality Improvement & Evaluation System (iQIES) organizations.

9. **Q- Is the Support page of the D&A Portal a replacement for ServiceNow?**
A- They can both still be used interchangeably. If you have ServiceNow access, you can still go in and submit a ticket as you have done in the past. The Support page on Portal now goes over different ways to be able to submit a help desk ticket that gets routed to support.
10. **Q- Is the timeline for SAS transition still the same? What support if any will be available in 2025?**
A- The timeline is still the same. All we know at this time is that Standard support will not be available past February 2025.
11. **Q- Is there a timeline for ability of utilization of job clusters in workflows?**
A- We do not have an active date for this currently. However, it is on the roadmap.
12. **Q- Are there resources for the differences between base R and Python and the Apache Spark variations (SparkR/Sparklyr/PySpark) used in Databricks.**
A- This will be covered in our upcoming [Spring 2024 Data Camp](#).
13. **Q- How do we connect MFT to HIVE to load local tables?**
A- Every organization has their own workbench, which is the same workbench that you access with SAS and should be enabled in MFT. You should be able to use MFT to load and unload data into that workbench and then from the S3 path, load, or unload data into Databricks for use there.
14. **Q- Will Medicare Administrative Contractors (MACs) receive information on training sessions?**
A- Yes. The MACs will receive training. Be on the lookout for communications on future training.
15. **Q- Is it possible to create a shared folder in the Databricks workspace that multiple users can access and control the rights, but would not be shared with all users? When I create a shared folder, it must be shared with all users.**
A- All users can provide specific sharing/access to other users on folders and files within their Workspace. They can manage it by right-clicking on the folder/file and selecting Share (Permissions). However, if they want to share something from their group's workspace folder with a user in their group, that should already be available. For example, we have a cdr-pda-qdas folder that all users within the group can access and see and access.
16. **Q- Is the Databricks coding assistant leveraging a large language model (LLM)? If so, which LLM is the Databricks coding assistant using?**
A- It is currently a Databricks feature. Databricks just released an open source LLM.
17. **Q- For individuals new to their roles, where is the best place to start to learn about data and access options?**
A- Those users can start by reviewing the information in the new [D&A Portal](#).

18. Q- Will there be integration of R Studio functionality within the Databricks environment?

A- There should already be the ability to integrate a type of R Studio within Databricks: <https://docs.databricks.com/en/spark/rstudio.html>

19. Q- How do you use MFT to load, are there instructions?

A- You can find instructions on the [MFT Confluence](#) page.

20. Q- Will any SAS functionality be lost during the period that SAS is still available but no longer supported (i.e.,1/1/2025 to 12/31/2025)?

A- We are not 100% sure, but we are expecting no functionality will be lost. However, SAS has not been able to guarantee us that.

21. Q- Is there a possibility of being able to see all the files from the workbench on Databricks side soon or ever? I know that workbench file access is available through MFT, and we can also access and process them through our code on Databricks side but cannot see them all on side menu like in SAS Viya.

A- That requires Databricks Volumes which we are looking at enabling. There are a couple of different technical limitations that we have with our current architecture. However, we are looking to add this in the future.

22. Q- What are the current access control mechanisms for tables at the Hive table level for users? Is there any documentation available to understand these mechanisms?

A- Everybody is given a database for themselves, which can be used in tables created. You will have carte blanche access to all those tables you create. Other tables are managed either through the DUA process or specifically if it is not a dataset that is governed by DUA. There is another process for that, and we can gradually manage table access to any different organization at this point.

23. Q- Is there an IMPL environment for Databricks? Will the log in process be the same?

A- There is. However, we typically do not give people access to that unless it is a Bring Your Own Analytics (BYOA) access. Our IMPL environment is primarily for integrating other systems rather than user direct access. For testing purposes, we recommend that people use the PROD Databricks.

24. Q- Earlier in the meeting, someone mentioned they have provided a structure for groups to put tables into. Is there a way to structure this (i.e., create subfolders within it to separate the tables into different folders)?

A- Users can create whatever folders they need inside their workbench. We are currently working on a non-production lane for the future, and we will update you on specific when they come.

25. Q- Are there any plans to allow us as users of the Databricks compute cluster, to install any library we need on the cluster, or all the installations need to go through a request to you?

A- You should have the ability to load the libraries that you need. Users can add their desired libraries to their cluster after they start it up.

26. **Q- If a state agency is currently using SAS to pull CMS data, will they be able to use Databricks for CMS data after the transition?**

A- The data is still stored in the same place. If the state agency has access to the AWS S3 location that has the data, they should still be able to query it with SAS.