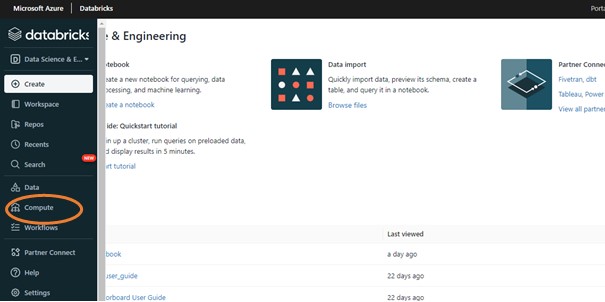
Working with RStudio in CDAP

14/06/2022

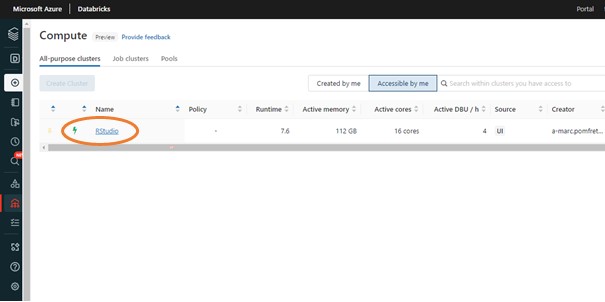
Purpose: This user guide shows how to access RStudio from within the CDAP PROD environment.

## Finding RStudio on CDAP

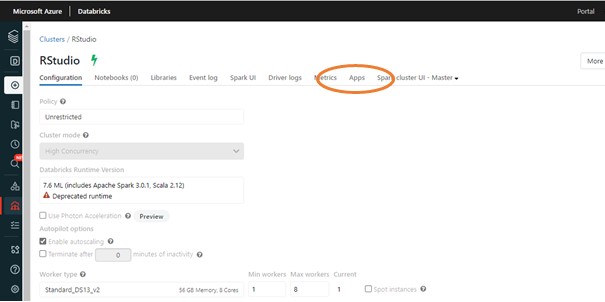
Within the workspace, click on Compute on the menu on the left.



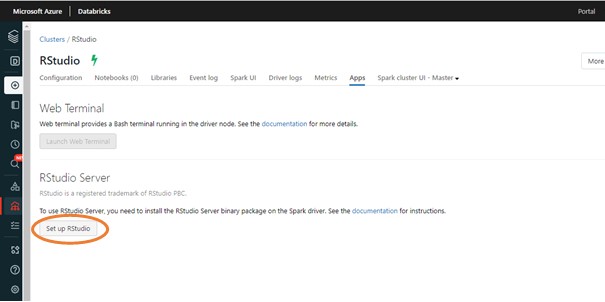
In the table that appears, click on RStudio.



In the horizontal menu, click on Apps.



Click on **Setup RStudio**. This will generate a password which you need to log into RStudio, alongside your Defra or ALB email address which is listed as your username. To copy the password, click on **show** which will reveal the password and copy it.



Click on **Open RStudio**. Use your email address as the username to sign in to RStudio, and paste the password. This will take you to RStudio

If you close the RStudio window but keep other tabs in the browser open, and then reopen RStudio through Databricks, your username and password will be saved and it takes you straight back to RStudio. You can sign out by clicking on the symbol next to your username at the top right of the screen.

To quit the R session you are in, you need to click on **Quit current R session** at the top right hand side of the screen, otherwise everything in your console will still appear once you go back into RStudio.

## Next steps with R and RStudio

If you have used RStudio previously, the environment will look very familiar to you. If you are new to RStudio, we recommend you take a look at courses from the Government Statistical Service. There is a whole host of introductory courses including  
- [Introduction to R course](https://gss.civilservice.gov.uk/training/introduction-to-r/)  
- [Data visualisation in R and Python](https://analysisfunction.civilservice.gov.uk/training/data-visualisation-in-r-and-python/)  
- [Data linkage in R](https://analysisfunction.civilservice.gov.uk/training/data-linkage-in-r/)  
- [R control flow loops and functions](https://analysisfunction.civilservice.gov.uk/training/r-control-flow-loops-and-functions/)  
- [Data linkage in R](https://analysisfunction.civilservice.gov.uk/training/introduction-to-sparklyr/)

There are also more advanced courses:  
- [Introduction to sparklyr](https://analysisfunction.civilservice.gov.uk/training/introduction-to-sparklyr/)  
- [Introduction to Reproducible Analytical Pipelines (RAP)](https://analysisfunction.civilservice.gov.uk/training/introduction-to-reproducible-analytical-pipelines-rap/)  
- [Machine learning in R](https://analysisfunction.civilservice.gov.uk/training/machine-learning-in-r/)

## Getting access to data within CDAP

To access the data held within CDAP, enter setwd("/dbfs"). This will give you access to the Databricks File Store (DBFS), where the data are held.

into the console. Under the files tab (usually on the bottom right), you should now see the folders that contain the data:

You can find data on CDAP by going to mnt > landingr, and explore the data held within the landing zone. **ADD LINK TO DATA GOVERNANCE ETC**

Go back to mnt and you can see the labr folder which contains files from individuals. Set up a new folder for yourself in here, by clicking on New folder, and name the folder according to your email address: