Databricks Utilities

Description

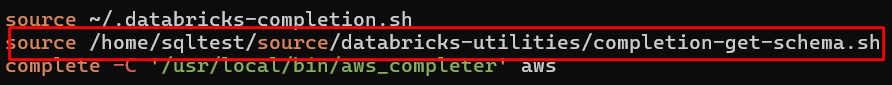
These are small utility tools I use with Databricks.

|  |  |
| --- | --- |
| completion-get-schema.sh | Bash completion script for get-schema.py |
| delete-jobs.sh | Delete all runs of a job |
| get-schema.py | Get the json representation of a data file schema |

Procedures

Completion

Source this in your bash\_profile to get bash completion to work.



Delete Jobs

Edit the script and replace the JOB\_ID by copy/pasting out of the workflow area of the webiste. You will see it in your job at the top right with the name. Login with SSO and then you can run the script to delete jobs.

databricks auth login

./delete-jobs.sh

Get Schema

Point this at a data file and it will send the file’s details to the clipboard or stdout. This will detect file encoding, format, header and delimiter (if CSV), and multiline (if JSON). The resulting JSON can be pasted into the settings files in the pipeline.

./get-schema.py --type json --file powerPlay.json 2>/dev/null

./get-schema.py --type json --file powerPlay.json --output stdout 2>/dev/null

