- Databricks Notebooks supports Python, SQL, R & Scala languages.
- You need to attach a cluster during creation of Databricks Notebook.
- You can change the language in notebook anytime.
- You can collaborate between team members by sharing the notebooks.
- Name of the notebook can be changed anytime by clicking on the notebook name.
- You can view the status of attached cluster in the notebook. Green circle next to cluster name indicates that cluster is up and running.
- You can run the notebook cell by clicking on Shift + Enter.
- You can change the language of the cell by clicking on the language in the notebook. It will add %lang in the notebook which are called as Magic commands. Magic commands are built-in commands that provide same output regardless of notebook language.
- %md (Markdown magic command) allows us to have a cell with formatted text which can be used to add comments in the notebook with plenty of formatting options like font, hyperlink formatting etc...,
- %run (Run magic command) allows us to run another notebook within a notebook by providing location of the notebook. You can get access to all variables in current notebook which have been provided in referenced notebook. It will help to reuse our code.
- %fs (File System Magic command) allows us to run file system commands within Databricks workspace.
- dbutils command is another way to deal with file system operations known as Databricks Utilities.
- dbutils.help() function will provide necessary help for dbutils command.
- dbutils.fs allow us to interact with Databricks file system or DBFS.
- Use Tab button in keyboard to help with auto-completion of code.
- dbutils.fs.ls(location) will list all files and folders within location provided.
- dbutils is more powerful than fs magic command since you work with Python directly without changing the cell language.
- display() function is used to print the output of Databricks cell in more structured format and you can download output as csv format. It is limited to preview only 1000 records.
- No need to use display function when you are working with SQL language as output is provided in tabular format by default with SQL language.
- Download the notebook using export option in File command provided in notebook.
- DBC cloud file (DBC Archive format) is a zip file containing collection of directories and notebooks. DBC files can be uploaded in any Databricks workspace to move or share notebooks.
- You can undo the changes in notebooks by clicking on version history of notebook and restore to that version if necessary.