

- Databricks provide different workspaces for Data Engineers, ML Engineers & Data Analysts.
- Use GitHub repo [Databricks-Certified-Data-Engineer-Associate](#) to import code notebooks into your personal GitHub repo.
- For multi node cluster, access sharing is allowed in Python & SQL only.
- Databricks runtime version should be latest above 11.3 LTS only for practicing the codes provided in GitHub Repo.
- Configure for Driver and Worker nodes.
- DBU/h – Databricks Units Per Hour, cost of Databricks is performed based on DBUs consumed. Depending on nodes provisioned in cluster, DBUs will be measured. For example, single node cluster will consume less DBUs than multi node cluster.
- For practicing codes for certification, Single node cluster is sufficient.
- You can terminate/resume clusters once they are created.
- You can edit cluster configuration once they are created, but cluster may restart once you edit and save the configuration.
- Event logs are captured for all cluster activities.
- Driver logs will get log generated within cluster notebooks and libraries.
- Make sure to terminate cluster once you have stopped using it, otherwise it will add to your cost.