can use pyspark, apache spark

also provide cluster instances i.e., for parallel processing

2 ways to use - Community version, paid/cloud version

helps to implement ML flow - CI/CD pipelines

whenever we create multiple instances in Azure/AWS & get data from cloud & process it in databricks

helps in DE, ML & DS problems

supports AWS, Azure & GCP

Databricks from community version if we upgrade we go to paid version

Community version - notebook, table, cluster, ML flow

1st create a cluster & give a name

by default runtime scala & pyspark version would be selected

1 instance with 15 GB memory however it can be upgraded

options in databricks cluster

libraries - can install or upload libraries

we can upload data - using export data - from file/S3/DFS/other sources(redshift, kinesis, cassander, JDBC....) & partners integration

uploaded tips dataset to databricks

create cluster

go to home & create a blank notebook