**ASSIGNMENT DATE-23-12-2023**

**FUNCTIONS OF SPARK:**

DataFrames in PySpark

Creating Dask DataFrames in PythonCreating DataFrames

Looking at Data in DataFrames

Selecting, Renaming, Filtering Data in a Pandas DataFrame

Manipulating, Droping, Sorting, Aggregations, Joining, GroupeBy  DataFrames

**Tips and Tricks to Use Spark Commands:**

Below are the different tips and tricks of Spark commands:

1. Beginners of Spark may use Spark-shell. Since Spark is built on Scala, using Scala Spark shell is an excellent option. However, Python Spark shell is also available, which can be used by those proficient in Python.
2. Spark shell has a lot of options to manage the resources of the cluster. Below Command can help you with that:

spark-shell-help

In Spark, working with long datasets is the usual thing. But things go wrong when bad input is taken. It’s always a good idea to drop bad rows by using the filter **function of Spark**. A good set of input will be a great go.

Spark chooses a good partition on its own for your data. But it’s always a good practice to keep an eye on partitions .

PYSPARK CHEAT SHEET:

