Skills Network Lab formerly called Data Science Workbench, it's a virtual lab environment, where you can use tools like Apache Zeppelin, Jupyter NoteBook etc.

**Jupyter Notebook** is an interactive document that allows you to execute code in smaller chunks called cells.

**Apache Zeppelin** is also similar to Jupyter Notebook, it allows interactive data analytics. its supports interpreters such as - Apache Spark, Python, JDBC, Markdown and Shell. Apache Zeppelin provides built-in Apache Spark integration.

apache zeppelin can be used with apache spark to convert a .csv file into a Spark data frame. the steps to do this are -

1. Converting .csv to rdd.
2. Data cleaning using map and filter.
   1. creates RDD of tuples from .csv
   2. create the schema represented to define name and type of column.
   3. apply the schema to the rdd.
3. Create the data frame using the toDF() function. A **DataFrame** is a distributed collection of data organized into named columns.
4. Registers the DataFrame as a temporary table.
5. Retrieve data using SQL commands. use %sql before writing the SQL query.

zeppelin imports SparkContext and SqlContext by default. and you can download the file using the link of the file from any dataset. the zeppelin notebooks are .json files.

**RStudio** in Building Analytics tools, RStudio IDE allows you to analyze data, take advantage of many statistical packages, create beautiful visualizations and Web applications.

getwd - returns current working directory.

**IBM Watson Studio,** formerly known as Data Science Experience or DSX, in an enterprise-ready environment for data scientists and developers,