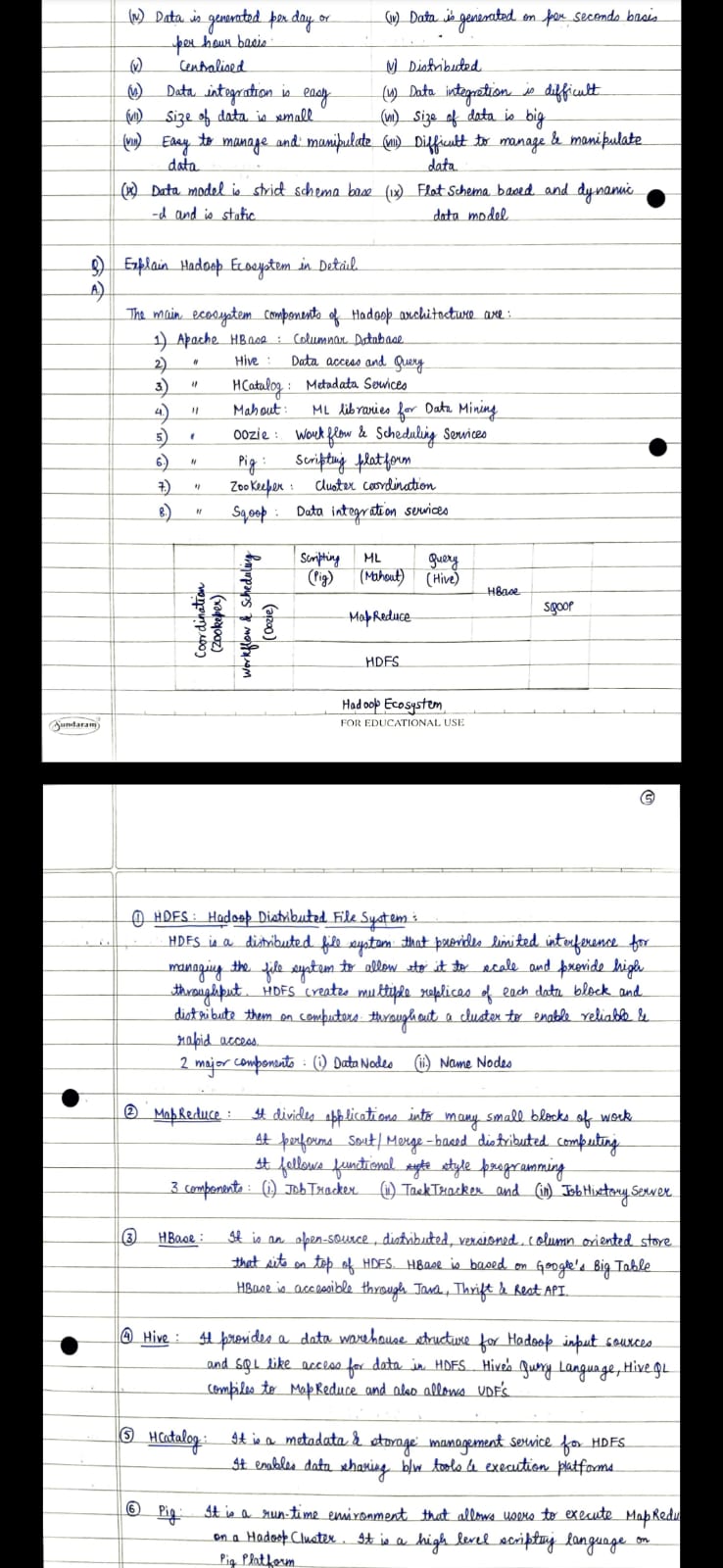
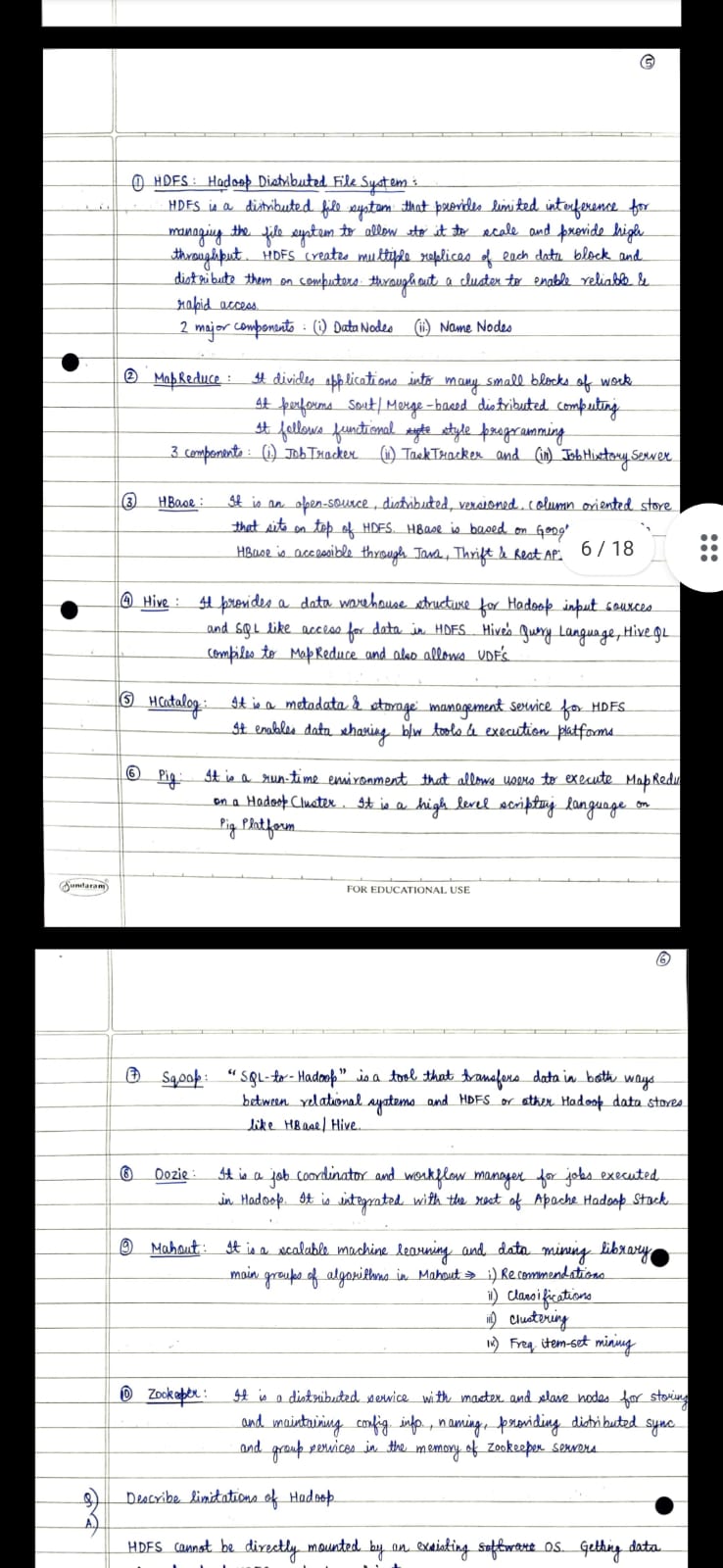
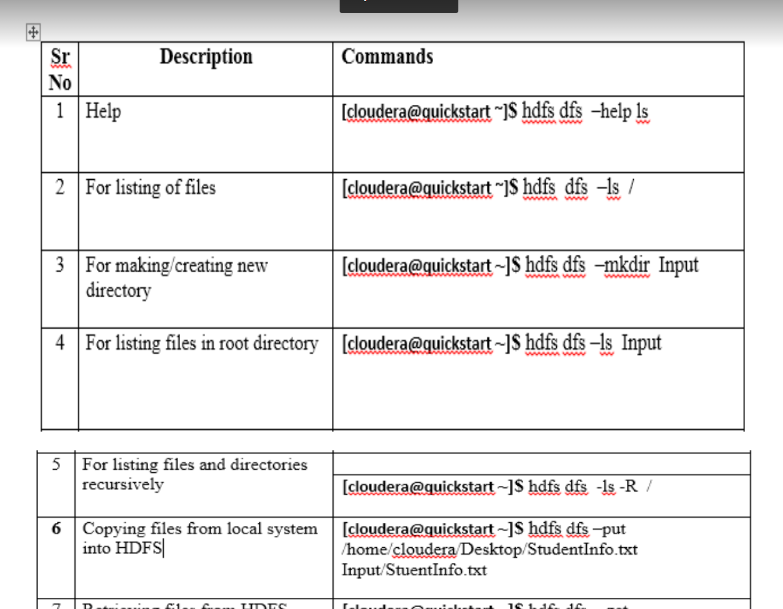
1.Draw and explain Hadoop Ecosystem? write 5 HDFS commands







2.What is HBASE Shell? Create table using HBASE command and

write 5 commands to perform operation on this table.

ANS- HBase is the Hadoop database, which provides random, realtime read/write access to very large

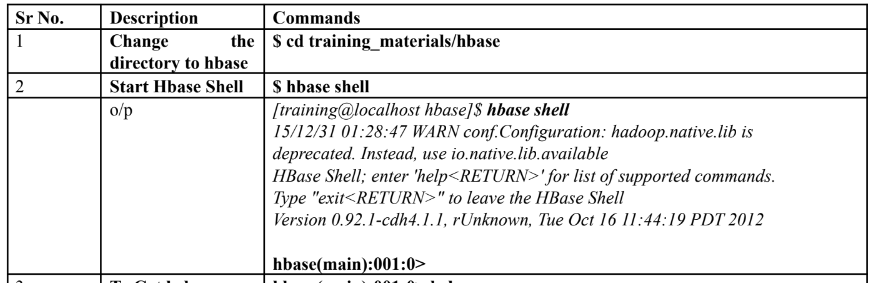
data. See the references on HBase for more information.

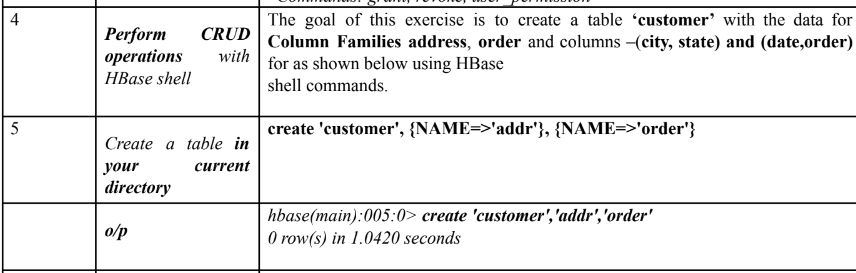
The HBase Shell is a ruby script that helps in interacting with the HBase system using a

command line interface. This shell supports creating, deleting and altering tables and also

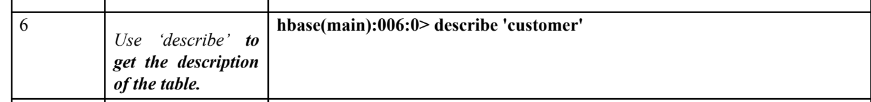
performing other operations like inserting, listing, deleting data and to interact with HBase.

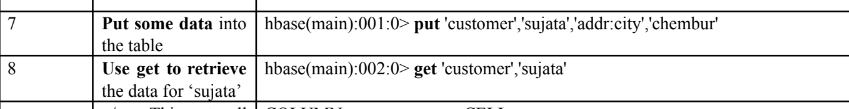
Commands to create a table-

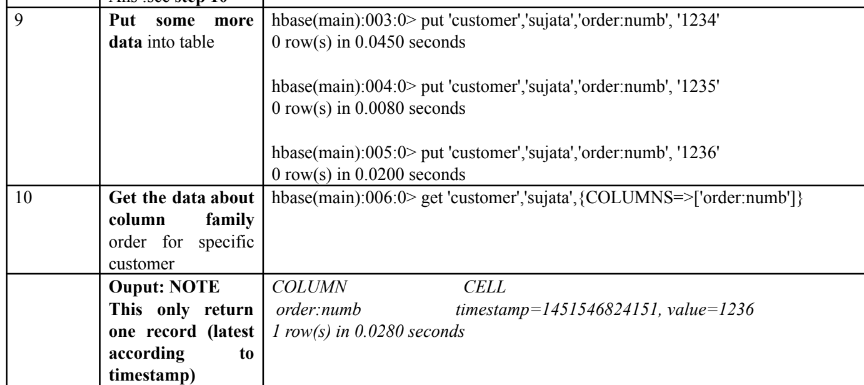




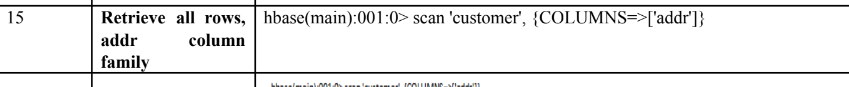
5 commands to perform operation in table-











3.Data Visualization using R

**Data visualization** is the technique used to deliver insights in data using visual cues such as graphs, charts, maps, and many others. This is useful as it helps in intuitive and easy understanding of the large quantities of data and thereby make better decisions regarding it.

**Data Visualization in R Programming Language**

The popular data visualization tools that are available are Tableau, Plotly, R, Google Charts, Infogram, and Kibana. The various data visualization platforms have different capabilities, functionality, and use cases. They also require a different skill set. This article discusses the use of R for data visualization.

R is a language that is designed for statistical computing, graphical data analysis, and scientific research. It is usually preferred for data visualization as it offers flexibility and minimum required coding through its packages.

## Types of Data Visualizations

Some of the various types of visualizations offered by R are:

### **Bar Plot**

There are two types of bar plots- horizontal and vertical which represent data points as horizontal or vertical bars of certain lengths proportional to the value of the data item

Bar plots are used for the following scenarios:

* To perform a comparative study between the various data categories in the data set.

To analyze the change of a variable over time in months or years

### **Histogram**

A histogram is like a bar chart as it uses bars of varying height to represent data distribution. However, in a histogram values are grouped into consecutive intervals called bins.

**Histograms are used in the following scenarios:**

* To verify an equal and symmetric distribution of the data.

To identify deviations from expected values

### **Box Plot**

The statistical summary of the given data is presented graphically using a boxplot. A boxplot depicts information like the minimum and maximum data point, the median value, first and third quartile, and interquartile range.

**Box Plots are used for:**

* To give a comprehensive statistical description of the data through a visual cue.
* To identify the outlier points that do not lie in the inter-quartile range of data.

### **Scatter Plot**

A scatter plot is composed of many points on a Cartesian plane. Each point denotes the value taken by two parameters and helps us easily identify the relationship between them.

**Scatter Plots are used in the following scenarios:**

* To show whether an association exists between bivariate data.
* To measure the strength and direction of such a relationship.

### **Heat Map**

Heatmap is defined as a graphical representation of data using colors to visualize the value of the matrix. heatmap() function is used to plot heatmap.

**Syntax:** heatmap(data)

**Parameters:**data: It represent matrix data, such as values of rows and columns

**Return:** This function draws a heatmap.