Progressive Education Society's

**Modern College of Engineering, Pune**

**MCA Department**

**A.Y.2024-25**

**(410904A) Big Data Analytics Laboratory**

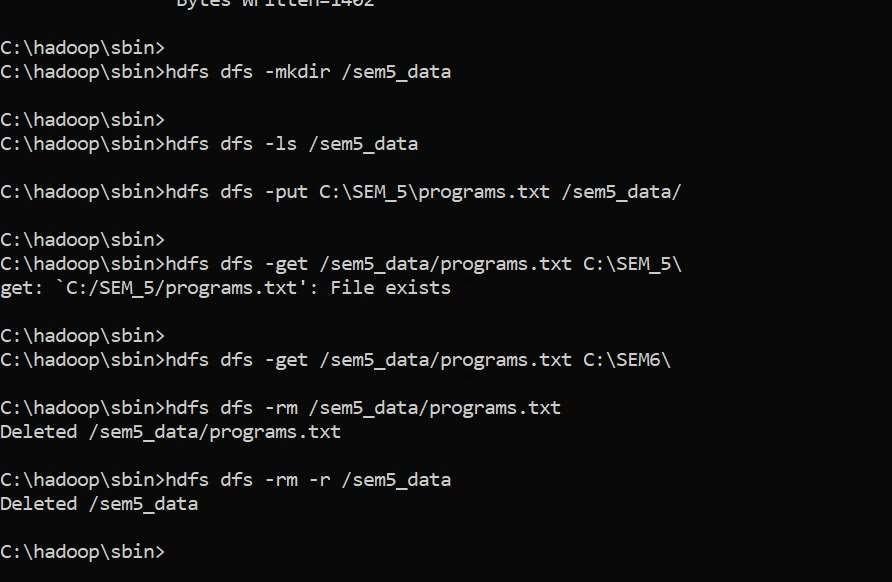
\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

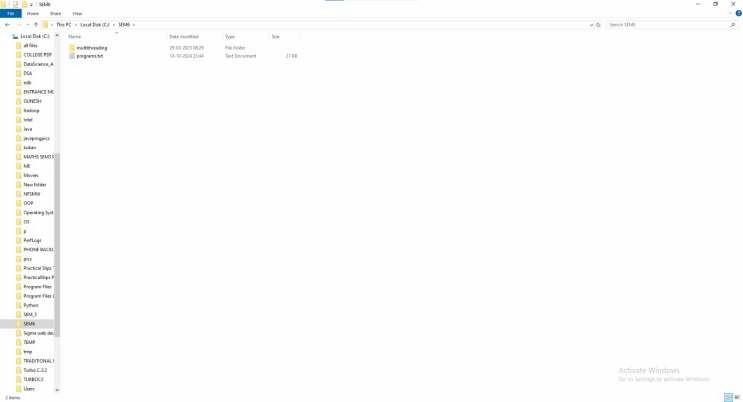
Class: SY-MCA Shift / Div.: S3/A Roll Number.: 52062

Name: Laxman Shinde Assignment No: 5 Date of Implementation: 15/10/2024

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* **Q1) To implement the following file management tasks in Hadoop System (HDFS):**

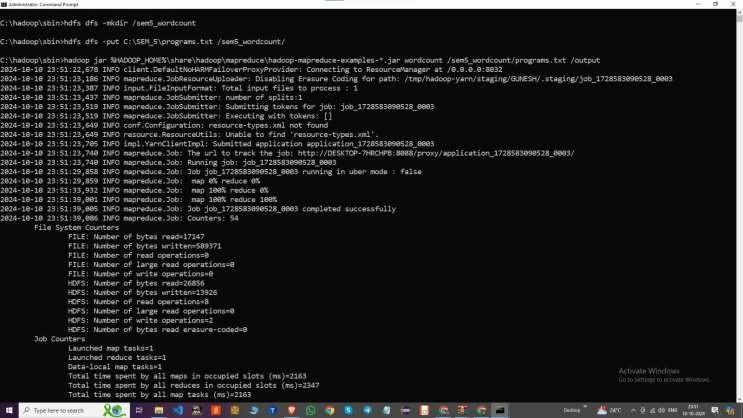
**Adding files and directories, Retrieving files, Deleting files.**

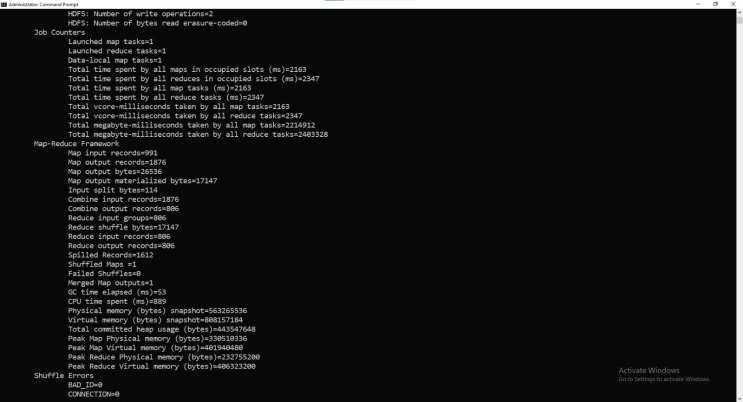


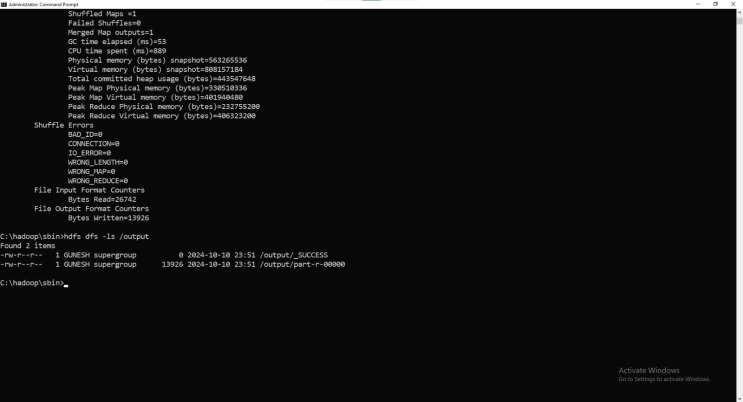


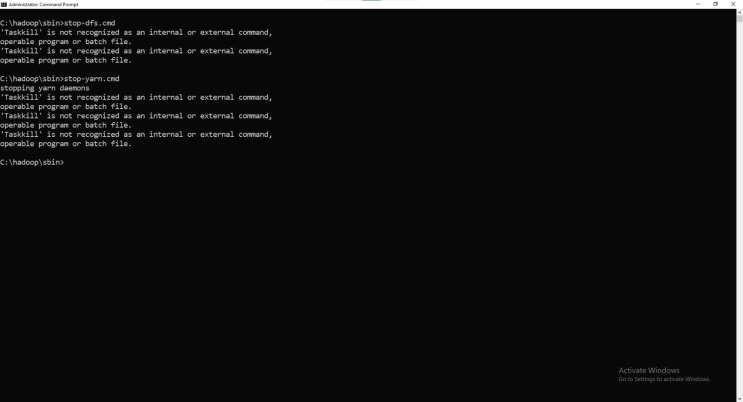
**Q2) To run a basic Word Count MapReduce program to understand MapReduce Paradigm:**

**To count words in a given file, to view the output file, and to calculate execution time.**









**Q3)To study and implement basic functions and commands in R Programming.**

# Basic Arithmetic Operations addition <- 5 + 3 # Addition subtraction <- 9 - 2 # Subtraction multiplication <- 4 \* 7 # Multiplication division <- 8 / 2 # Division exponentiation <- 2^3 # Exponentiation modulo <- 9 %% 2 # Modulo

# Print arithmetic results cat("Addition: ", addition, "\n") cat("Subtraction: ", subtraction, "\n") cat("Multiplication: ", multiplication, "\n") cat("Division: ", division, "\n") cat("Exponentiation: ", exponentiation, "\n") cat("Modulo: ", modulo, "\n\n")

# Variable Assignment x <- 10 y <- 3.5 name <- "R Programming"

cat("x = ", x, ", y = ", y, ", name = ", name, "\n\n")

# Data Types numeric\_var <- 42.5 # Numeric

integer\_var <- as.integer(42) # Integer logical\_var <- TRUE # Logical (boolean) char\_var <- "Hello" # Character

cat("Numeric: ", numeric\_var, "\n") cat("Integer: ", integer\_var, "\n") cat("Logical: ", logical\_var, "\n") cat("Character: ", char\_var, "\n\n")

# Vectors vec <- c(1, 2, 3, 4, 5) cat("Vector: ", vec, "\n")

# Access elements of the vector cat("First element of vector: ", vec[1], "\n") cat("Second to fourth elements: ", vec[2:4], "\n\n")

# Sequences sequence <- 1:10 # Sequence of numbers from 1 to 10 cat("Sequence: ", sequence, "\n\n")

# Functions square <- function(x) { return(x^2) }

result <- square(6) cat("Square of 6: ", result, "\n\n")

# Control Structures # If-Else Statement if (x > 5) { cat("x is greater than 5\n")

} else { cat("x is less than or equal to 5\n")

}

# For Loop cat("For loop from 1 to 5:\n") for (i in 1:5) { cat(i, "\n")

}

# While Loop cat("\nWhile loop:\n") count <- 1 while (count <= 5) { cat(count, "\n") count <- count + 1

}

# Apply Functions numbers <- c(2, 4, 6, 8, 10) squares <- sapply(numbers, square) cat("\nSquares of numbers: ", squares, "\n\n")

# Data Frames data <- data.frame(

Name = c("John", "Sara", "Alex"),

Age = c(25, 30, 28),

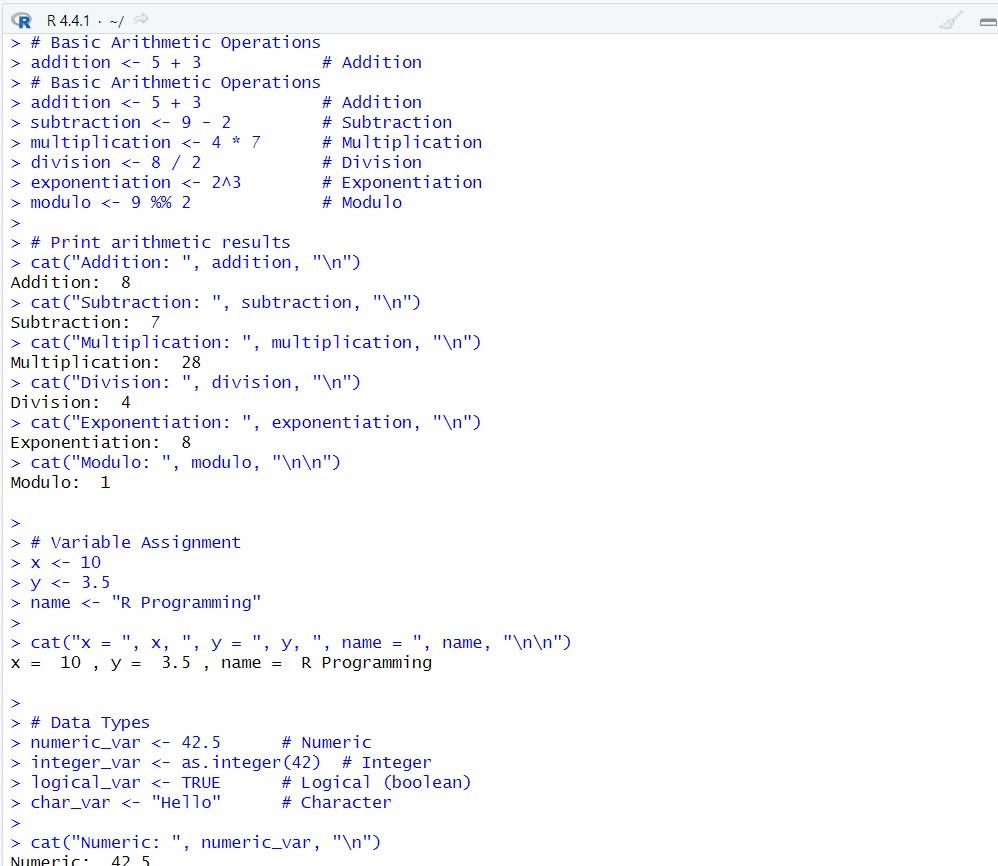
Occupation = c("Doctor", "Engineer", "Artist")

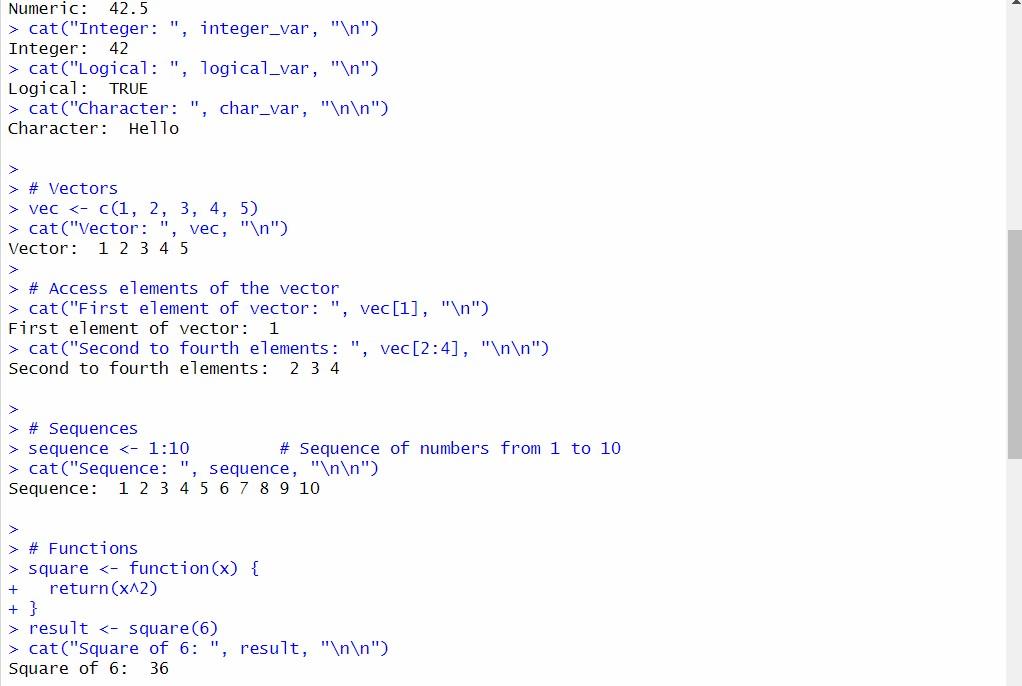
)

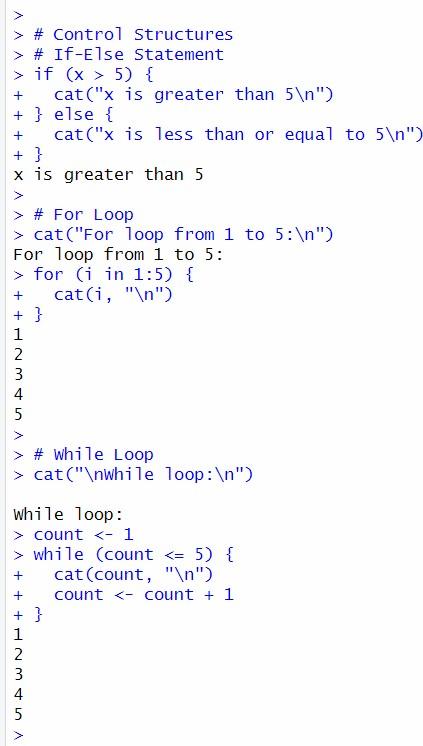
print(data)

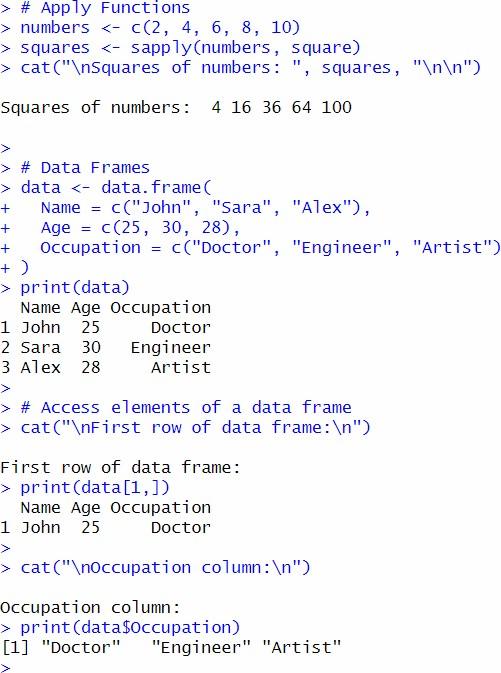
# Access elements of a data frame cat("\nFirst row of data frame:\n") print(data[1,])

cat("\nOccupation column:\n") print(data$Occupation)









**Q4) To build WordCloud, a text mining method using R for easy to understand and visualization than a table data.** install.packages("tm") install.packages("wordcloud") install.packages("RColorBrewer")

library(tm) library(wordcloud) library(RColorBrewer)

text <- "R programming is powerful for data analysis and visualization.

It provides easy-to-use tools for building machine learning models, statistical analysis, and more. Creating WordCloud in R is an exciting way to visualize text data."

corpus <- Corpus(VectorSource(text)) corpus <- tm\_map(corpus, content\_transformer(tolower)) corpus <- tm\_map(corpus, removePunctuation) corpus <- tm\_map(corpus, removeNumbers) corpus <- tm\_map(corpus, removeWords, stopwords("english")) corpus <- tm\_map(corpus, stripWhitespace)

tdm <- TermDocumentMatrix(corpus) matrix <- as.matrix(tdm) word\_freqs <- sort(rowSums(matrix), decreasing = TRUE) df <- data.frame(word = names(word\_freqs), freq = word\_freqs)

colors <- brewer.pal(8, "Dark2") wordcloud(words = df$word, freq = df$freq, min.freq = 1, random.order = FALSE, colors = colors)

