**Mock Test 1 - Big Data Question no. 3**

**3.**Write a PySpark code to read a JSON file named "students.json" containing student records with the following schema: "name" (string), "age" (integer), "grade" (string). Filter the DataFrame to include only students whose age is greater than 18.

**Sol.**

from pyspark.sql import SparkSession

# Create SparkSession

spark = SparkSession.builder \

.appName("StudentRecords") \

.getOrCreate()

# Read JSON file into DataFrame

students\_df = spark.read.json("students.json")

# Filter DataFrame based on age greater than 18

filtered\_df = students\_df.filter(students\_df.age > 18)

# Show the filtered DataFrame

filtered\_df.show()

# Stop the SparkSession

spark.stop()