PYSPARK

ASSIGNMENT 23-12-2023

BASIC PYSPARK COMMANDS

**Initializing SparkSession**

from pyspark.sql import SparkSession

# Initialize SparkSession

spark = SparkSession.builder.appName("SparkBasics").getOrCreate()

**Creating DataFrames**

From a list of tuples

data = [("Alice", 25), ("Bob", 30), ("Charlie", 35)]

columns = ["Name", "Age"]

df = spark.createDataFrame(data, columns)

df.show()

from a dictionary:

data = [{"Name": "Alice", "Age": 25}, {"Name": "Bob", "Age": 30}, {"Name": "Charlie", "Age": 35}]

df = spark.createDataFrame(data)

df.show()

**Reading data**

**From CSV**

df = spark.read.csv("path\_to\_your\_file.csv", header=True, inferSchema=True)

df.show()

From Json

df = spark.read.json("path\_to\_your\_file.json")

df.show()