# **Solution Design and Testing for Test Scores Analysis**

This document explains the technology, programming language being used, design and any assumptions made along with instructions for executing the code and evidence of testing for the Test Scores Analysis exercise.

**Technology & Programming language:**

The solution has been implemented on Azure Databricks environment with the sample file uploaded onto Azure Blob Storage and coded in Databricks notebook with PySpark as the programming language for applying transformations and achieve the desired result.

**Design:**

The databricks notebook takes File Path and File Name as inputs using widgets, reads the contents of the file as a string which would then be split into list of elements. At this stage, a dataframe ***test\_scores\_df*** has been created from this list, which would be used to derive the highest score and full name of the respective people. The formatted string for presenting highest score and the full names of top scorers are being captured in different dataframes which would be combined at the final stage to print/output the desired result in requested format.

**Assumptions:**

Please find below the list of assumptions made while implementing this solution:

1. A databricks workspace with a mount point is available for the application user.
2. If exists, header row would be: **First Name,Second Name,Score**
3. Column delimiter: “,” (comma)
4. Comma (“,”) is not allowed in any of the data elements.
5. Data in the file is not malformed.

**Instructions to run the application:**

Please find below the instructions and the respective sequence to run the application:

1. Import the notebook source code file ***Test\_Scores\_Analysis.py*** into a databricks workspace.
2. Upload the test file (Eg: **TestData.csv**) onto any of the accessible paths on mount point for the respective workspace.
3. Start the databricks cluster.
4. Click on **Run all** on the notebook imported.
5. Pass the respective values for prompts: FilePath and FileName

Eg: FileName: TestData.csv

FilePath: /dbfs/mnt/sink\_lake/raw-src/test/csv/

**Testing:**

The implemented solution has been tested and is generating the results as expected for the test file **TestData.csv**. Please find below screen of output of execution of the notebook.

A screenshot of a computer

Description automatically generated