

Apache Spark Assessment

- Task 1
 - Setup/Install local standalone instance of Apache Spark (3.1.2 version) on your laptop/desktop system
- Task 2
 - Write a PySpark/Scala program
 - Read provided CSV file (people.csv) into dataframe
 - Write the data in the above dataframe to local filesystem in delta format
 - Read the above data written in the delta format into dataframe
 - Show record count for above dataframe.

How to submit assessment solution

1. Your GitHub/BitBucket/GitLab profile
 - a. Create a repository for assessment solution
 - b. Create a Readme file under solution repository which includes
 - i. Setup instructions
 - ii. Execution instructions
 - iii. Any other supporting information
 - c. Share repository link
2. Email GitHub/BitBucket/GitLab assessment repo/project link and other relevant details to: vikrams@bizanalytix.com

Useful Resources

- Spark Standalone setup/installation
 - <https://phoenixnap.com/kb/install-spark-on-windows-10>
 - <https://www.geeksforgeeks.org/install-apache-spark-in-a-standalone-mode-on-windows/>
 - <https://sparkbyexamples.com/spark/apache-spark-installation-on-windows/>
- Delta format official documentation
 - <https://docs.delta.io/latest/index.html>
- Apache Spark official documentation
 - <https://spark.apache.org/docs/latest/>