

Spark SQL Exercise

Exercise

- Download from the repository (github.com/forons/BigDataExamples) the following files in the 'files' folder:
 - `course_name_credits.csv`
 - `student_course_grade.csv`
 - `student_name.tsv`
- Upload the data to Databricks (Data -> Add Data -> Drag and Drop files)
- **Find out the average grade for each student, displaying the name**
- **Find out the maximum grade for each course, displaying its name**
- **Decrease the grade by 4 and add a column to the dataset with the word **PASS** if the grade is ≥ 18 , and **FAIL** if the grade is < 18**
- **Save the output of this 3 tasks in a new file: one csv, one json, one parquet**