

Indicaciones para la ejecución del programa Tarea_1.py y test_Tarea_1.py

Información Preliminar

Para la creación de este programa se asume que si existen estudiantes listados en el archivo 'estudiante.csv' pero no aparecen en el archivo 'nota.csv' se debe a que los estudiantes no tienen cursos matriculados durante el periodo de referencia y por ende no se toman en cuenta para la obtención de los mejores N promedios. Así mismo si un estudiante ha llevado un mismo curso múltiples veces se tomarán en cuenta todas las notas obtenidas en este curso para el calculo de su promedio ponderado.

Ejecución del Programa

- Se ejecuta en la consola el comando `spark-submit Tarea_1.py estudiante.csv curso.csv nota.csv`

A continuación, se detallan las ejecuciones del programa:

- Se muestra la tabla estudiante:

Numero de Carnet	Nombre Completo	Carrera
1	Person1LastName1	Administracion de...
2	Person2LastName2	Biotecnologia
3	Person3LastName3	Electromecanica
4	Person4LastName4	Mecanica
5	Person5LastName5	Administracion de...
6	Person6LastName6	Biotecnologia
7	Person7LastName7	Electrica
8	Person8LastName8	Electrica
9	Person9LastName9	Mecanica
10	Person10LastName10	Electronica
11	Person11LastName11	Administracion de...
12	Person12LastName12	Electronica
13	Person13LastName13	Biotecnologia
14	Person14LastName14	Mecanica
15	Person15LastName15	Fisica
16	Person16LastName16	Quimica
17	Person17LastName17	Quimica
18	Person18LastName18	Biotecnologia
19	Person19LastName19	Quimica
20	Person20LastName20	Electronica

- Se muestra la tabla Curso

24/04/10 04:10:00 INFO DAGScheduler: Job 1 finished

Codigo de Curso	Credito	Carrera_c
261	3	Administracion de...
262	5	Administracion de...
263	4	Administracion de...
264	2	Administracion de...
265	5	Administracion de...
266	2	Administracion de...
267	3	Administracion de...
268	5	Administracion de...
269	3	Administracion de...
260	4	Administracion de...
131	4	Biotecnologia
132	4	Biotecnologia
133	2	Biotecnologia
134	5	Biotecnologia
135	5	Biotecnologia
136	5	Biotecnologia
137	5	Biotecnologia
138	5	Biotecnologia
139	5	Biotecnologia
130	4	Biotecnologia

only showing top 20 rows

- Se muestra la tabla Nota

24/04/10 04:16:09 INFO DAGScheduler: Job 2 finished

Numero de Carnet	Codigo de Curso	Nota
1	261	82.2
1	262	64.2
1	263	72.0
1	264	43.5
1	265	44.0
1	267	37.5
1	268	83.8
1	269	63.1
5	261	84.3
5	262	92.9
5	263	81.6
5	264	70.6
5	265	33.8
5	266	81.1
5	268	62.1
5	269	76.3
5	260	100.0
11	261	69.0
11	262	66.3
11	264	91.4

- Se ejecuta la función 'unir_datos' y se muestran los resultados

24/04/10 04:18:10 INFO CodeGenerator: Code generated in 32.133072 ms

Codigo de Curso	Numero de Carnet	Nota	Nombre Completo	Carrera	Credito
261	1	82.2	Person1LastName1	Administracion de...	3
262	1	64.2	Person1LastName1	Administracion de...	5
263	1	72.0	Person1LastName1	Administracion de...	4
264	1	43.5	Person1LastName1	Administracion de...	2
265	1	44.0	Person1LastName1	Administracion de...	5
267	1	37.5	Person1LastName1	Administracion de...	3
268	1	83.8	Person1LastName1	Administracion de...	5
269	1	63.1	Person1LastName1	Administracion de...	3
261	5	84.3	Person5LastName5	Administracion de...	3
262	5	92.9	Person5LastName5	Administracion de...	5
263	5	81.6	Person5LastName5	Administracion de...	4
264	5	70.6	Person5LastName5	Administracion de...	2
265	5	33.8	Person5LastName5	Administracion de...	5
266	5	81.1	Person5LastName5	Administracion de...	2
268	5	62.1	Person5LastName5	Administracion de...	5
269	5	76.3	Person5LastName5	Administracion de...	3
260	5	100.0	Person5LastName5	Administracion de...	4
261	11	69.0	Person11LastName11	Administracion de...	3
262	11	66.3	Person11LastName11	Administracion de...	5
264	11	91.4	Person11LastName11	Administracion de...	2

only showing top 20 rows

- Se ejecuta la función 'agregaciones_parciales' y se muestran los resultados

Nombre Completo	Carrera	promedio_ponderado
Person156LastName156	Electronica	68.47
Person608LastName608	Electronica	70.87
Person572LastName572	Electromecanica	56.86
Person717LastName717	Electromecanica	69.74
Person429LastName429	Quimica	61.0
Person124LastName124	Administracion de...	73.17
Person418LastName418	Administracion de...	67.39
Person207LastName207	Electromecanica	57.87
Person691LastName691	Mecanica	67.06
Person782LastName782	Mecanica	87.27
Person389LastName389	Electrica	79.59
Person724LastName724	Electronica	71.16
Person632LastName632	Quimica	64.2
Person821LastName821	Electromecanica	74.71
Person29LastName29	Electrica	70.4
Person982LastName982	Quimica	80.95
Person985LastName985	Quimica	60.76
Person884LastName884	Electrica	65.56
Person261LastName261	Quimica	71.84
Person411LastName411	Administracion de...	56.3

- Se ejecuta la función 'resultados_finales' y se muestran los resultados. En este caso se utilizo el parámetro N=3

24/04/18 04:18:14 INFO DAGScheduler: Job 14 finished: showString at

Nombre Completo	Carrera	Mejores_promedios
Person471LastName471	Administracion de...	82.4
Person136LastName136	Administracion de...	80.49
Person892LastName892	Administracion de...	80.15
Person954LastName954	Biotecnologia	87.39
Person143LastName143	Biotecnologia	82.94
Person115LastName115	Biotecnologia	81.3
Person478LastName478	Electrica	85.25
Person739LastName739	Electrica	84.37
Person731LastName731	Electrica	81.15
Person750LastName750	Electromecanica	86.18
Person940LastName940	Electromecanica	85.03
Person886LastName886	Electromecanica	84.31
Person22LastName22	Electronica	81.86
Person947LastName947	Electronica	80.67
Person176LastName176	Electronica	80.52
Person782LastName782	Mecanica	87.27
Person824LastName824	Mecanica	87.25
Person921LastName921	Mecanica	82.36
Person772LastName772	Quimica	83.09
Person969LastName969	Quimica	82.87

A continuación, se muestran el comando de ejecución del archivo conteniendo las pruebas unitarias:

- En la consola se ejecuta el comando `pytest test_Tarea_1.py`
- La imagen de abajo comprueba que las tres pruebas dentro del archivo pasan la prueba.

```
C:\Fernando Gutierrez\Programa Ciencia de los Datos\Big Data\bigdataclass> docker run -i -t bigdata /bin/bash
bash-5.0# cd testsample
bash-5.0# pytest test_Tarea_1.py
===== test session starts =====
platform linux -- Python 3.7.10, pytest-7.4.4, pluggy-1.2.0
rootdir: /src/testsample
collected 3 items

test_Tarea_1.py ... [100%]

===== warnings summary =====
../../../../usr/lib/python3.7/site-packages/pyspark/context.py:317
/usr/lib/python3.7/site-packages/pyspark/context.py:317: FutureWarning: Python 3.7 support is deprecated in Spark 3.4.
  warnings.warn("Python 3.7 support is deprecated in Spark 3.4.", FutureWarning)

-- Docs: https://docs.pytest.org/en/stable/how-to/capture-warnings.html
===== 3 passed, 1 warning in 38.40s =====
bash-5.0#
```