
UNIVERSITÀ DI PISA

LUCA PALLA - 533605

GIULIA CALVO - 544434

# LDS- GROUP 21_DB: SECOND MIDTERM

# ASSIGNMENT 0

*For every year, the users ordered by total number of answers:*

1. We started loading AnswerId and UserID with the operator “Origine OLE DB”
2. Next we searched, using “Ricerca”, GeoID and CountryCode contained respectively in the Table user1 and Geography
3. With the data obtained we used “Aggregazione” to make a Group by with the function count of the total number of answers by country
4. We saved in csv “assignment_0_2019” the file and we used Python to confront the results and be sure that was correct

# ASSIGNMENT 1

*A group (identified by GroupId) is said to have an age mismatch if the difference between the date of birth of the youngest participating student and the oldest is greater than 365 days. List all the groups with an age mismatch:*

1. With the operator “Origine OLE DB” we loaded UserId and OrganizationID
2. Using “Ricerca” we loaded GroupId from Organization Table, DateOfBirthId from user1 table and Date from Date table
3. Using “Ordinamento” we ordered by GroupID and DateofBirthId taking only unique values: the reason of this last choice is for optimization of operational running time
4. Using “Multicast” we split the process calculating on the left the youngest participating student of the group and on the right the oldest one
5. We order by GroupID again because the “Merge Join” needs it.
6. We calculated on the output of the merge two different measures: using “Colonna Derivata” we computed the difference of Date of birth in days for each group between the youngest participating student and the oldest.
7. Using another “Colonna Derivata” named Mismatch if the previous measure was more than 365 we gave 1, else 0.
8. At the end, using “Suddivisione condizionale” on the Mismatch we saved the groups with the value 1 in a Csv file “assignment_1_2019”; as before we checked on Python.

# ASSIGNMENT 2

*For each continent the ratio between correct answers of males and correct answers of females:*

1. With the operator “Origine OLE DB” we loaded UserId and IsCorrect
2. Using “Suddivisione condizionale” we focused only on correct Answer: this was made for the operational optimization of the whole process
3. Using “Ricerca” we loaded GeoId and Gender from user1 table and Continent from Geography.
4. Using again “Suddivisione condizionale” on Gender we split the category: on the left using “Aggregazione” we obtained The Count of Correct answers of Male by Continent; on the right we did the same for the Female
5. To use the Merge, we ordered the previous result and joined them
6. Using “Colonna Deivata” we obtained the Ratio asked by the assignment and wrote the output on a Csv File, checking all on Python.

*All of the choice we performed during this part of the project in visual studio were with the aim of Optimizing with respect of Memory and Velocity of the operators.*