Database Management Project: Student Scholarship Finder Application

Group 6: Xiuyi Feng, Massiel Sanchez

Project Statement

Scholarships and financial aid are important resources for many college students, particularly those from low-income families. Despite the numerous scholarships and financial aid options from federal, government, and other organizations, many students are unaware of their eligibility or even the existence of these aids. To address this, we propose the Smart Scholarship and Financial Aid System, which aims to streamline the scholarship and financial aid discovery and application process for students in the City University of New York (CUNY) system.

Using student data, the system will filter and suggest scholarships that align with each student's eligibility. Detailed information about each scholarship including its requirements, along with a direct application link will be provided to the student.

Attributes

	Domain-Name	DB-Name	Domain(Example)	TYPE	Atomic?	Repeating_ Group?	Table/Entity	PK?	FK?
Sname	Student Name	Sname	Anthony Gonzalez	varchar(120)	Yes	Yes	Student	Yes	TBD
Sgender	Student Gender	Sgender	Male	varchar(32)	Yes	Yes	Student	No	TBD
Srace	Student Race/Ethnicity	Srace	White-Hispanic	varchar(120)	Yes	Yes	Student	No	TBD
Cname	College Name	Cname	Hunter College	varchar(120)	Yes	Yes	Student	No	TBD
SID	Student ID	SID	12345	varchar(32)	Yes	Yes	Student	Yes	TBD
SGPA	Student GPA	SGPA	3.89	varchar(10)	Yes	Yes	Student	No	TBD
Smajor	Student Major	Smajor	Computer Science	varchar(120)	Yes	Yes	Student	No	TBD
Aincome	Student Income	Aincome	30,000	int	Yes	Yes	Student	No	TBD
Dstatus	Student Dependency Statu	Dstatus	Dependent	varchar(64)	Yes	Yes	Student	No	TBD
Rstatus	Student Residency Status	Rstatus	In-State Student	varchar(120)	Yes	Yes	Student	No	TBD
Ssname	Scholarship Name	Ssname	Latinos in STEM Scholarship Name	varchar(120)	Yes	Yes	Scholarship	No	TBD
SsID	Scholarship ID	SsID	305	varchar(20)	Yes	Yes	Scholarship	Yes	TBD
requiredGPA	Scholarship GPA Requireme	requiredGPA	3.5+	varchar(20)	Yes	Yes	Scholarship	No	TBD
Srequirement	Scholarship Requirements	Srequirement	Reccomendation letter,Resume, Tra	varchar(120)	No	Yes	Scholarship	No	TBD
Adeadline	Scholarship Deadline	Adeadline	8/28/2023	datetime	Yes	Yes	Scholarship	No	TBD
Aamount	Scholarship Award Amount	Aamount	$3,000	int	Yes	Yes	Scholarship	No	TBD
Alink	Scholarship Link	Alink	www.examplewebsite123456.com	varchar(32)	Yes	Yes	Scholarship	No	TBD

The dataset includes comprehensive information about students including their names, gender, race, major, student ID, and more. Additionally, the dataset includes information about various scholarships for which the students may qualify for, including the scholarship names, requirements, deadlines, and other relevant details.

Normalization

To achieve first normal form (**1NF**) we must

- Remove Non-Atomic Fields
- Remove Repeating Groups
- Find PK

In our attribute catalog the only non-atomic field was '*Srequirement*'. To remove the non-atomic attribute we converted the multi-domain attribute into multiple atomic attributes, each referring to a single domain.

1. Remove non-atomic fields:

Attribute	Domain-Name	DB-Name	Domain(Example)	TYPE	Atomic?	Repeating_Group?	Table/Entity	PK?	FK?
Srequirement	Scholarship Requirements	Srequirement	Essay, Reccomendation letter, Resume, Transcript	varchar(120)	No	Yes	Scholarship	No	TBD

New Attributes	Domain-Name	DB-Name	Domain(Example)	TYPE	Atomic?	Repeating_Group?	Table/Entity	PK?	FK?
scholarship_essay	Scholarship Required Essay	scholarship_essay	Essay	varchar(24)	Yes	Yes	Scholarship	No	TBD
scholarship_resume	Scholarship Required Resume	scholarship_resume	Resume	varchar(24)	Yes	Yes	Scholarship	No	TBD
scholarship_letter	Scholarship Required Letter	scholarship_letter	Recommendation Lett	varchar(24)	Yes	Yes	Scholarship	No	TBD
scholarship_transcript	Scholarship Required Transcript	scholarship_transcript	Transcript	varchar(24)	Yes	Yes	Scholarship	No	TBD

Normalization cont.

We identified three repeating groups in our new list of attributes: Students, Scholarships and Scholarship Requirements. To remove the repeating groups we grouped the related fields into separate tables.

2. Remove Repeating Groups:	
New Atomic Attributes:	
Sname	Group: Student
Sgender	
Srace	
Cname	
SID	
SGPA	
Smajor	
Aincome	
Dstatus	
Rstatus	
Ssname	Group: Scholarship
SsID	
requiredGPA	
scholarship_essay	Group: Scholarship Requirements
scholarship_resume	
scholarship_letter	
scholarship_transcript	
Adeadline	
Aamount	
Alink	

Normalization cont.

Student Table									
Sname	Sgender	Srace	Cname	SID	SGPA	Smajor	Aincome	Dstatus	Rstatus
Scholarship Table									
Ssname	SsID	requiredGPA	Adeadline	Aamount	Alink				
Scholarship Requirement Table:									
SSID	Scholarship Requirement								

separate tables for related entities

Normalization cont.

Finally, we identified the primary key of each table. The Student table uses the student ID, the Scholarship table uses the Scholarship ID and lastly, a composite key consisting of the scholarship Id and specific scholarship requirement is the key of the Requirement table.

3. Find PK									
Student Table									
Sname	Sgender	Srace	Cname	SID (Pk)	SGPA	Smajor	Aincome	Dstatus	Rstatus
Scholarship Table									
Ssname	SsID (Pk)	requiredGPA	Adeadline	Aamount	Alink				
Scholarship Requirement Table:									
(PK)ssID + requirenment	SsID (FK)	scholarship_essay							
(PK)ssID + requirenment	SsID (FK)	scholarship_resume							
(PK)ssID + requirenment	SsID (FK)	scholarship_letter							
(PK)ssID + requirenment	SsID (FK)	scholarship_transcript							

Normalization cont.

To achieve second normal form (2NF), partial dependencies must be removed. After performing 1NF, we did not identify any partial dependencies. All attributes within their respective tables rely entirely on the complete primary key.

Further, to achieve third normal form (3NF), we have to eliminate transitive dependencies. Our tables had none.

Final Tables:

Student Table

Sname	Sgender	Srace	Cname	SID (Pk)	SGPA	Smajor	Aincome	Dstatus	Rstatus
Christopher Diaz	Male	Hispanic	Hunter College	23799	3.75	Math	40,000	Dependent	In-State

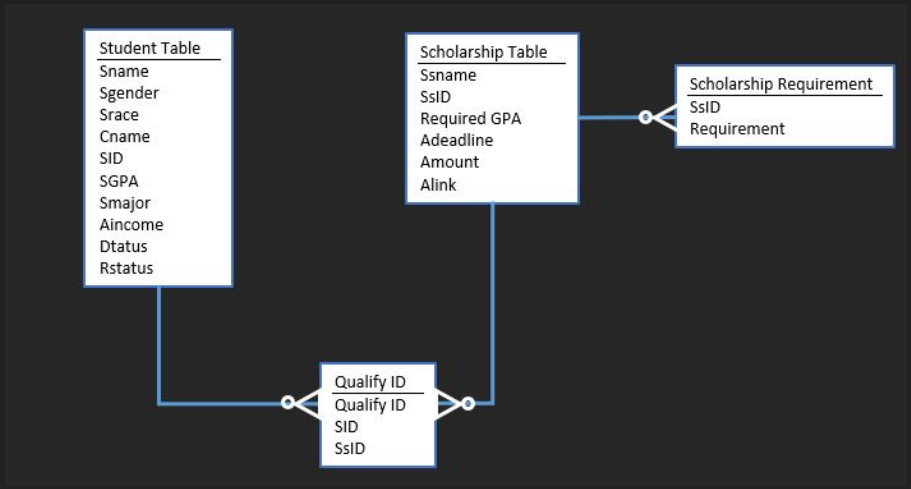
Scholarship Table

Ssname	SsID (Pk)	requiredGPA	Adeadline	Aamount	Alink
Honors Scholarship	101	3.7+	8/10/2023	10,000	www.example.com

Scholarship Requirement Table:

	SSID	Scholarship Requirement					
(PK)ssID + requirenment	101	Essay		QUALIFY ID(PK)	SID (FK)	SsID (FK)	
(PK)ssID + requirenment	102	Resume		1	23799	101	
(PK)ssID + requirenment	103	Letter					

ERD - Crowfoot Notation



Cardinality Statements

1 Student can qualify for zero or more Scholarships. Each qualification corresponds to a single student. Zero or one to many

1 Scholarship can be awarded to zero more students. One scholarship can be qualified for by many students, but each qualification is specific to a single scholarship. Zero-or one to many

1 Scholarship can have zero or more requirements

Pre-normalization Table Definition

```
MariaDB [project]> CREATE TABLE pre_normalized(
    ->     Sname varchar(120),
    ->     Sgender varchar(32),
    ->     Srace varchar(120),
    ->     Cname varchar(120),
    ->     SID varchar(32),
    ->     SGPA float,
    ->     Smajor varchar(120),
    ->     Sincome int,
    ->     Dstatus varchar(120),
    ->     Rstatus varchar(120),
    ->     Ssname varchar(300),
    ->     SsID varchar(32),
    ->     requiredGPA float,
    ->     Srequirement varchar(300),
    ->     Sdeadline date,
    ->     Aamount int,
    ->     Alink varchar(350)
    -> );
Query OK, 0 rows affected (0.020 sec)
```

**This SQL DDL statement defines the pre_normalized table*

Loading Raw Data From CSV file

Sname	Sgender	Srace	Cname	SID	SGPA	Smajor	Aincome	Dstatus	Rstatus	Ssname	SsID	requiredG	Srequirem	Sdeadline	Aamount	Alink
Anthony Gonzale	Male	White-His	Hunter Co	12345	3.89	Computer	30000	Dependen	In State	Latinos in	305	3.5+	Essay;Rec	#######	3000	www.examplewebsite89.com
Giselle Knowles	Female	Black	Brooklyn C	12355	4	Music	150000	Independe	Out of Sta	Formation	306	3.5+	Resume; F	#######	25000	www.examplewebsite89.com
Kelly Wu	Female	Asian	Hunter Co	12356	3.4	Psycholog	40000	Dependen	In State	Roosevelt	303	3.0+	Reccomer	#######	2500	www.examplewebsite89.com
Jeffery Matos	Male	Black-Hisp	City Colleg	12357	3.1	Math	37187	Dependen	In State	Latinos in	305	3.5+	Essay;Rec	#######	3000	www.examplewebsite89.com
Tanya Mcquoid	Female	White	City Colleg	12358	3.45	Economics	113559	Dependen	In State	Roosevelt	303	3.0+	Reccomer	#######	2500	www.examplewebsite89.com
Aureliano Buendi	Male	White-His	City Colleg	12359	2.9	English	36294	Dependen	In State	English Ma	301	3.0+	Resume;E	#######	10000	www.examplewebsite89.com
Vanessa Saint	Female	Black	Hunter Co	12360	3.83	Media Stu	108573	Independe	In State	Athena Hc	302	3.7+	Transcript	#######	5000	www.examplewebsite89.com
Natalie Louis	Female	Black-Hisp	Baruch	12361	3.4	Business	54406	Dependen	In State	Roosevelt	303	3.0+	Reccomer	#######	2500	www.examplewebsite89.com
Christopher Molt	Male	White	City Colleg	12362	2.5	Business	104707	Independe	Out of Sta	Frank D. S	300	3.5+	Resume;T	9/2/2023	15000	www.examplewebsite89.com
Adriana La Cerva	Female	White	Baruch	12363	3.3	Psycholog	67241	Dependen	Out of Sta	Frank D. S	300	3.5+	Resume;T	9/2/2023	15000	www.examplewebsite89.com
Brett Brown	Male	Black	Baruch	12364	3.9	Art	70113	Dependen	In State	Athena Hc	302	3.7+	Transcript	#######	5000	www.examplewebsite89.com
Sasha Carter	Female	Black	Brooklyn C	12365	4	Music	85381	Dependen	In State	Formation	306	3.5+	Resume;R	#######	25000	www.examplewebsite89.com
Colin Beckett	Male	White	Brooklyn C	12366	3.94	English	19341	Dependen	In State	English Ma	301	3.0+	Resume;E	#######	10000	www.examplewebsite89.com
Kevin Chen	Male	Asian	Hunter Co	12367	3.87	Computer	46010	Independe	In State	Athena Hc	302	3.7+	Transcript	#######	5000	www.examplewebsite89.com
David Lynch	Male	White	City Colleg	12368	3.01	Art	48619	Dependen	Out of Sta	Roosevelt	303	3.0+	Reccomer	#######	2500	www.examplewebsite89.com
Jennifer Melfi	Female	White	City Colleg	12369	3.55	Psycholog	29388	Dependen	In State	Roosevelt	303	3.0+	Reccomer	#######	2500	www.examplewebsite89.com
Carlos Diaz	Male	Black-Hisp	Hunter Co	12370	3.78	Psycholog	115976	Dependen	In State	Athena Hc	302	3.7+	Transcript	#######	5000	www.examplewebsite89.com
Amaarae Genfi	Female	Black	City Colleg	12371	3.32	Music	89957	Dependen	In State	Formation	306	3.5+	Resume;R	#######	25000	www.examplewebsite89.com
Kimberly Jones	Female	Black	Brooklyn C	12372	3.6	Computer	110805	Dependen	In State	Roosevelt	303	3.0+	Reccomer	#######	2500	www.examplewebsite89.com
Azema Khan	Female	Asian	Baruch	12373	3.8	Business	115333	Dependen	In State	Athena Hc	302	3.7+	Transcript	#######	5000	www.examplewebsite89.com

dummy data in csv file

```
MariaDB [project]> LOAD DATA LOCAL INFILE "C:\\Users\\Massiel\\Desktop\\raw_data.csv"
    -> INTO TABLE pre_normalized
    -> FIELDS TERMINATED BY ','
    -> LINES TERMINATED BY '\n'
    -> IGNORE 1 ROWS
    -> (Sname, Sgender, Srace, Cname, SID, SGPA, Smajor, Sincome, Dstatus, Rstatus, Ssname, SsID, requiredGPA, Srequirement, @Sdeadline, Aamount, Alink)
    -> SET Sdeadline = STR_TO_DATE(@Sdeadline, '%m/%d/%Y');
Query OK, 20 rows affected, 20 warnings (0.012 sec)
Records: 20  Deleted: 0  Skipped: 0  Warnings: 20
```

Loading data into ***pre_normalized*** table. Had to change the format of the date data in column '***Sdeadline***'.

DDL Statements of Normalized Tables

```
MariaDB [project]> create table Students(
    ->     Sname varchar(120),
    ->     Sgender varchar(32),
    ->     Srace varchar(120),
    ->     Cname varchar(120),
    ->     SID varchar(32),
    ->     SGPA float,
    ->     Smajor varchar(120),
    ->     Sincome int,
    ->     Dstatus varchar(120),
    ->     Rstatus varchar(120)
    -> );
Query OK, 0 rows affected (0.020 sec)
```

This SQL DDL statement defines the **Students table*

```
MariaDB [project]> create table Scholarships(
    -> Ssname varchar(300),
    -> SsID varchar(32),
    -> requiredGPA float,
    -> Sdeadline date,
    -> Aamount int,
    -> Alink varchar(350)
    -> );
Query OK, 0 rows affected (0.025 sec)
```

This SQL DDL statement defines the **Scholarships table*

DDL Statements of Normalized Tables

```
MariaDB [project]> create table scholarship_req(
    -> SsID varchar(32),
    -> Requirement varchar(250)
    -> );
Query OK, 0 rows affected (0.024 sec)
```

This SQL DDL statement defines the **scholarship_req table*

```
MariaDB [project]> create table qualify_id(
    -> qualify_id int auto_increment Primary Key,
    -> SID varchar(32),
    -> SsID varchar(32)
    -> );
Query OK, 0 rows affected (0.019 sec)
```

This SQL DDL statement defines the **Qualify_ID table*

Loading Data into Normalized Tables

Loading data from pre_normalized table into the students table:

```
MariaDB [project]> insert into Students
    -> (Sname, Sgender, Srace, Cname, SID, SGPA, Smajor, Sincome, Dstatus, Rstatus)
    -> Select Sname, Sgender, Srace, Cname, SID, SGPA, Smajor, Sincome, Dstatus, Rstatus
    -> From pre_normalized;
Query OK, 20 rows affected (0.011 sec)
Records: 20  Duplicates: 0  Warnings: 0
```

Loading Data into Normalized Tables

Loading data from pre_normalized table into Scholarships Table:

```
MariaDB [project]> insert into scholarships
    -> (Ssname, SsID, requiredGPA, Sdeadline, Aamount, Alink)
    -> Select distinct Ssname, SsID, requiredGPA, Sdeadline, Aamount, Alink
    -> from pre_normalized;
Query OK, 6 rows affected (0.010 sec)
Records: 6  Duplicates: 0  Warnings: 0
```

Loading Data into Normalized Tables

Loading data into the scholarship_req table:

```
MariaDB [project]> INSERT INTO scholarship_req (SsID, Requirement)
    -> SELECT DISTINCT SsId, TRIM(SUBSTRING_INDEX(SUBSTRING_INDEX(Srequirement,';',numbers.n),';', -1))
    -> AS Requirement
    -> FROM pre_normalized
    -> JOIN (
    ->     SELECT 1 AS n UNION ALL SELECT 2 UNION ALL SELECT 3 UNION ALL SELECT 4
    -> ) AS numbers ON CHAR_LENGTH(Srequirement) - CHAR_LENGTH(REPLACE(Srequirement, ';', '')) >= numbers.n - 1;
Query OK, 12 rows affected (0.016 sec)
Records: 12  Duplicates: 0  Warnings: 0
```

SsID	Srequirement
301	Resume;Essay
300	Resume;Transcript
300	Resume;Transcript
301	Resume;Essay

SsID	Requirement
301	Resume
301	Essay
300	Resume
300	Transcript

Loading Data into Normalized Tables

Loading data into the *qualify_id* table:

```
MariaDB [project]> insert into qualify_id
    -> (SID, SsID)
    -> Select SID, SsID
    -> From pre_normalized;
Query OK, 20 rows affected (0.006 sec)
Records: 20  Duplicates: 0  Warnings: 0
```

Possible Queries

- Find Students with certain GPAs or Majors
  - select * from students where 3.5 < SGPA;
  - select * from students where Smajor = ‘Math’;
- Return Scholarship IDs that a Student is eligible for using qualify_id table
  - *select q.SsID, stu.Sname, stu.SID from qualify_id q join Students stu on q.SID = stu.SID;*
- Return requirements of Scholarship student is eligible for given Scholarship ID
  - *select * from scholarship_req where SsID = 'xxx’;*
- Return requirements of scholarship student qualifies for
  - *select * from scholarship_req req natural join (select q.SsID, stu.Sname from qualify_id q join Students stu on q.SID = stu.SID) qualify;*
- *Return scholarships with required GPA 3.5+*
  - *select * from scholarships where 3.5 <= requiredGPA;*

More queries

- Return Students with GPA greater or = 3.5
  - select * from students where 3.5 <= SGPA;
- Add students that qualify for certain scholarship into qualify_ID table

```
MariaDB [project]> insert into qualify_ID (SID, SsID)
    -> select stu.SID, 306
    -> from students stu
    -> where 3.5 <= stu.SGPA
    -> AND NOT EXISTS
    -> (select 1 from qualify_ID q where q.SID = stu.SID and q.SsID = 306);
Query OK, 9 rows affected (0.008 sec)
Records: 9  Duplicates: 0  Warnings: 0
```

Acknowledgements

- ChatGPT and other online resources were used to resolve errors