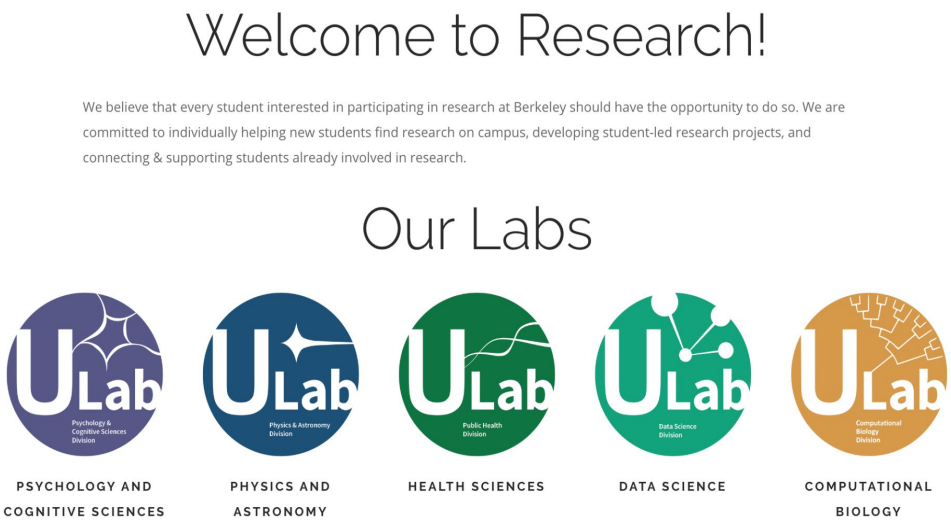
Undergraduate Lab @ Berkeley

Team 6 Database Design Project

Members: Arjun Vats(CEO), Jiho Yeom(CCO), Roy Yoon, Sam Kim, Junbeom Chi, Cosmin Deshmukh, Krish Rambhiya

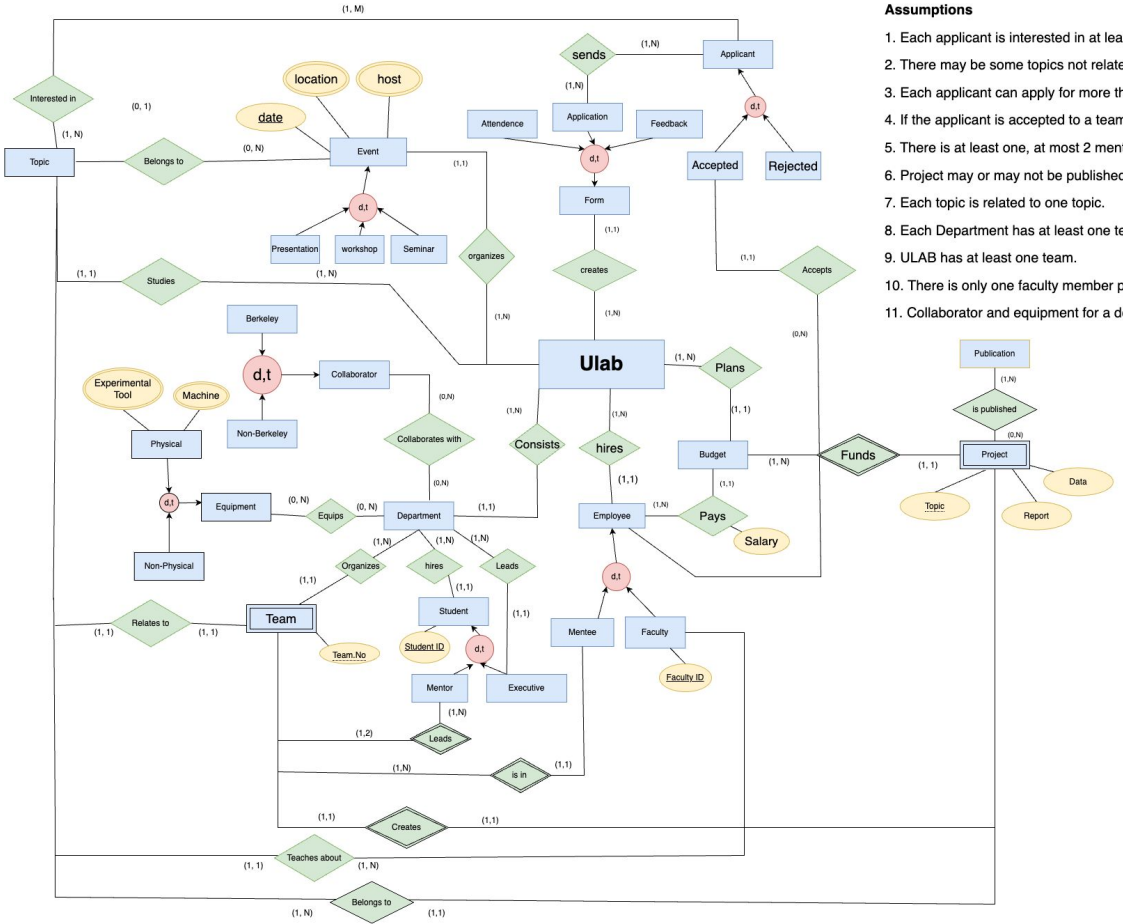
About **ULAB**

1. A student run research organization, endorsed by faculty across the College of Letters and Sciences and College of Engineering at Berkeley.
2. ULAB consists of a core leadership group looking after the operations of its 5 labs which each have an executive body and multiple research groups.
3. Each research group has 1-2 mentors who are experienced undergraduate researchers and many mentees who conduct the research into their chosen topic.



Revised EER Diagram

https://app.diagrams.net/#G1t4klOMgw4VaBzx0Dqsq5IhM_zZkqkJGh



Assumptions

1. Each applicant is interested in at least one topic.
2. There may be some topics not related to events held by ULab.
3. Each applicant can apply for more than one.
4. If the applicant is accepted to a team, the applicant can accept the offer or not.
5. There is at least one, at most 2 mentors per team.
6. Project may or may not be published to publications.
7. Each topic is related to one topic.
8. Each Department has at least one team.
9. ULAB has at least one team.
10. There is only one faculty member per department.
11. Collaborator and equipment for a department may or may not be needed.

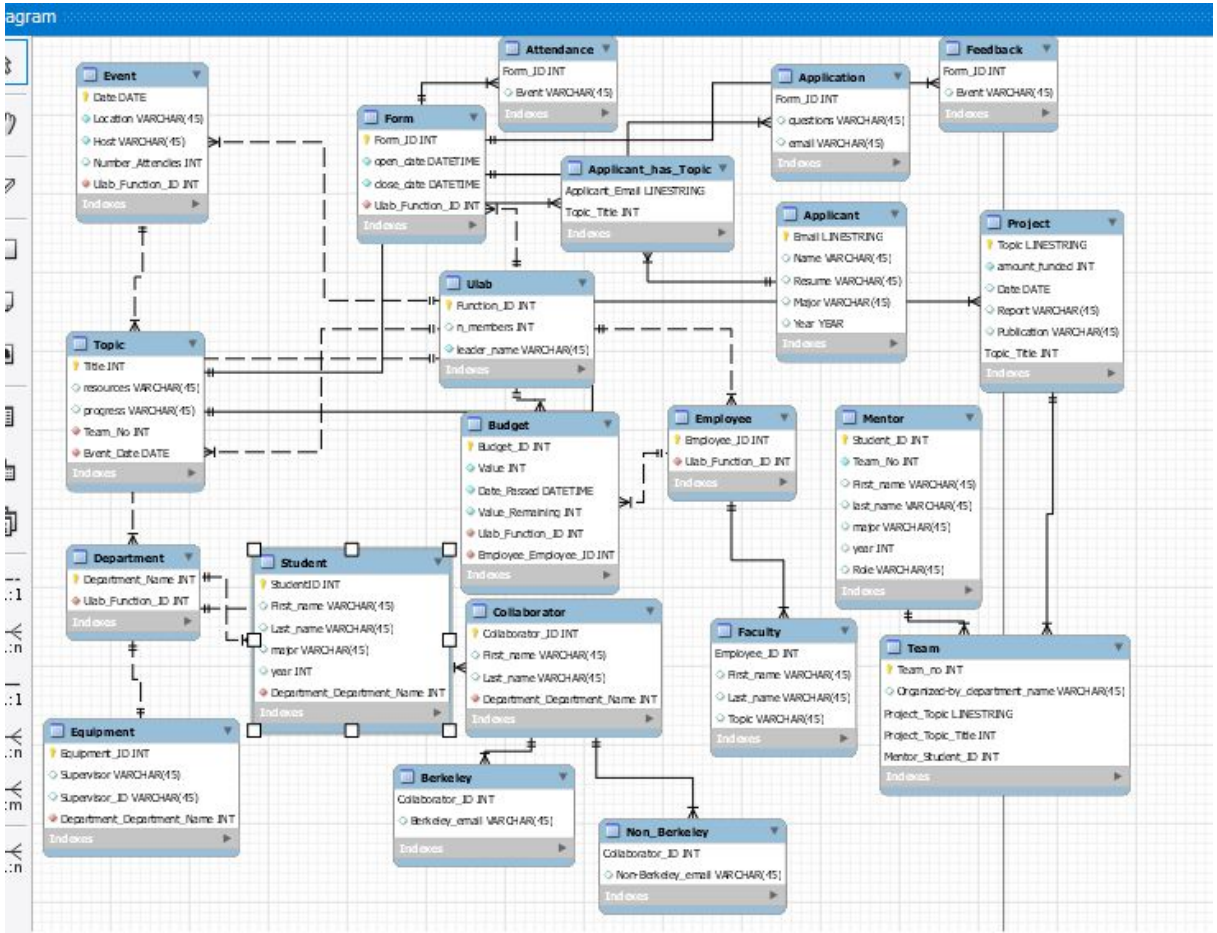
EER Diagram to Relational Schema

1. Ulab(Function_ID, n_members, leader_name)
2. Budget(Budget_ID, Value, Date_Passed, Value_Remaining |Belongs_to_Function_ID[1])
3. Employee(employee_id)
    a Faculty(employee_id**3 , First_name, last_name, Topic)
4. Student(Student_ID, First_name, last_name, major, year)
    a Mentor(Student_ID**8, Team_NO, First_name, last_name, major, year, role, project)
    b Executive(Student_ID**8, First_name, last_name, major, year, role)
    c Mentee(Student_ID**8,Team_NO, First_name, last_name, major, year, role, project)
5. Department(Department_Name | Function_ID**1)
6. Team(Team-no., organized-by-Department_Name**5, Topic)
7. Equipment(Equipment_id, supervisor, supervisor_id)
    a Physical(Physical_id**7, experimental tool, machine)
    b Non-Physical(Non_Physical_id**7, Paper)
8. Collaborator(Collaborator_ID,First_name, last_name)
    a Berkeley(Collaborator_ID**8,Berkeley_email)
    b Non-Berkeley(Collaborator_ID**8, Non-Berkeley_email)
9. Project(Topic, created-by-Team_no..organized-by-Department_Name**5 , amount_funded, date, report, publication)
10. Form(Form_ID, open_date, close_date, n_responses | Function_ID**1)
    a Application(Form_ID**10, questions | Email[15])
    b Attendance(Form_ID**10, event)
    c Feedback(Form_ID**10, event)

Relational Schema Contd.

11. Applicant(Email, Name, Resume, Answers, Major, Year)
    a Accepted(Email**11, Member_ID,)
    b Rejected(Email**11, Rejection_Reason)
12. Event(Date, Location, Host, Number_Attendees | Function_ID**1)
    a Presentation(Date**12, Slide_Deck)
    b Workshop(Date**12, Topic, Materials)
    c Seminar(Date**12, Panelists, Question_Bank)
13. Topic(Title, resources, progress | Team_no.**6, employee_id**3)
14. Publication(Publication_id, format)
15. Interested_In(Email**11,Title**13)
16. Sends(Email**11,Form_ID**10)
17. Belongs_to(Date**12,Title**13)
18. Equips(Department_name**5,Equipment_id**13)
19. Collaborates_with(Department_name**5,Collaborator_ID**8)
20. Published_in(Topic, created-by-Team_no.,organized-by-Department_Name**5 ,Publication_id**14)

MySQL relationship View



Interesting Queries #1

Prediction on how many times the research would be referenced

Getting investors and collaborators is the most important thing to maintain the research. So each department needs to prepare reports about their results so they can present them to their stakeholders anytime. Evidence such as the number of publications by topic and the past collaborators who have contributed can be examples. In order to get what we need, we first retrieve the publications written from each department and aggregate them by topic. There may be more than one department included in the same topic, but that is fine since we are preparing statistics for each department. Then we make a table that includes the year, topic, the count of publications, the count of referenced, and the name of the collaborators. However, this table may not be sufficient to change investors' firm wallets. The created table can be further utilized making a prediction of any kind. One could be the number of referenced that enables the investors to give a glimpse of the potential of our current working topics. If the number of referenced increases, then it means that the topic is trending. We would first use the table to train the LSTM model and make a prediction of the number of referenced and show it in a line graph through plotly library with the year as the x-axis and the number of referenced as the y-axis.

MySql Query for query2

select distinct
ulab2.project.amount_funded,
ulab2.project.Date1,
ulab2.project.project_description,
ulab2.department.Department_Name,
ulab2.budget.Date_Passed,
ulab2.budget.Value from ulab2.project,
ulab2.budget, ulab2.department
where ulab2.project.Department_Id =
ulab2.budget.deparment_id
and ulab2.budget.deparment_id =
ulab2.department.Department_ID
and ulab2.budget.Budget_ID =
ulab2.project.budget_id;

```
select * from ulab2.project, ulab2.budget, ulab2.department
where ulab2.project.Department_Id = ulab2.budget.deparment_id
and ulab2.budget.deparment_id = ulab2.department.Department_ID
and ulab2.budget.Budget_ID = ulab2.project.budget_id;

select distinct ulab2.project.amount_funded,  ulab2.project.Date1, ulab2.project.project_description, ulab2.departmer
ulab2.budget.Date_Passed, ulab2.budget.Value from ulab2.project, ulab2.budget, ulab2.department
where ulab2.project.Department_Id = ulab2.budget.deparment_id
and ulab2.budget.deparment_id = ulab2.department.Department_ID
and ulab2.budget.Budget_ID = ulab2.project.budget_id;

ALTER TABLE ulab2.publication
ADD Refs INT;
```

MySQL query for query 1.

```
select * from ulab2.project,
ulab2.budget, ulab2.department
where
ulab2.project.Department_Id =
ulab2.budget.deparment_id
and ulab2.budget.deparment_id
=
ulab2.department.Department_I
D
and ulab2.budget.Budget_ID =
ulab2.project.budget_id;
```

```
ALTER TABLE
ulab2.publication
ADD Refs INT;

update ulab2.publication
set ulab2.publication.Refs = 10
where
ulab2.publication.Publication_ID
> 0;
```

WorkBench Image

```
select * from ulab2.project, ulab2.budget, ulab2.department
where ulab2.project.Department_Id = ulab2.budget.deparment_id
and ulab2.budget.deparment_id = ulab2.department.Department_ID
and ulab2.budget.Budget_ID = ulab2.project.budget_id;

select distinct ulab2.project.amount_funded,  ulab2.project.Date1, ulab2.project.project_description, ulab2.depa
ulab2.budget.Date_Passed, ulab2.budget.Value from ulab2.project, ulab2.budget, ulab2.department
where ulab2.project.Department_Id = ulab2.budget.deparment_id
and ulab2.budget.deparment_id = ulab2.department.Department_ID
and ulab2.budget.Budget_ID = ulab2.project.budget_id;

ALTER TABLE ulab2.publication
ADD Refs INT;

update ulab2.publication
set ulab2.publication.Refs = 10
where ulab2.publication.Publication_ID > 0;
```

Resulting Table For Query 1

	project_description	pub_num	Refs	First_name	Last_name	project_id
0	Science	12	11	Jane	Doe	1
1	Particle Accelerator	6	10	Jane	Doe	2
2	Classifier	6	15	Jane	Doe	3
3	Photon perceptor	3	82	Jane	Doe	4
4	Classifier	15	56	John	Tamer	7
5	Photon perceptor	3	43	John	Tamer	8
6	Classifier	2	19	Jared	Goff	11
7	Photon perceptor	4	5	Jared	Goff	12

Table generated using MySql query, and exported to jupyter notebook and read as a dataframe using pandas libraries.

Table After Aggregate Functions

29]:

project_description	pub_num	Refs
Classifier	23	90
Particle Accelerator	6	10
Photon perceptor	10	130
Science	12	11

Grouping by project type to show number of references and publications.

Interesting Queries #2

Correlation between funding and the time taken for publication

For a client, it is important to look for any factors that are related to the publication of the project. The purpose of Ulab is to offer opportunities for students to be involved in and conduct the research, and publication of the research project is one factor to prove that the project is conducted successfully. For this query, we decided to look at funding for the project as a factor since funding is obviously a crucial factor for managing a Ulab committee. Therefore, it would be beneficial for a client to be aware whether the funding for the project has an effect on making the project successful. For this query, we will pull out the data table ‘Project’ using SQL function with four attributes: amount funded, date, publication, Deparment_name. We will also pull out the data table ‘Budget’ with two attributes: date passed, and Ulab function. We will join the department name and Ulab function to match the project with the right budget time data. We will calculate the time by subtracting the budget date from project date to calculate the time taken for the project. With the time and budget assigned for each project as attributes, we will then use seaborn library to plot a scatterplot for visualization between these two attributes. We will then add a best fit line to the visualization. Lastly, we will use the numpy library to get correlation coefficients as indicators on whether there is a strong correlation between two variables.

Query 2 Results

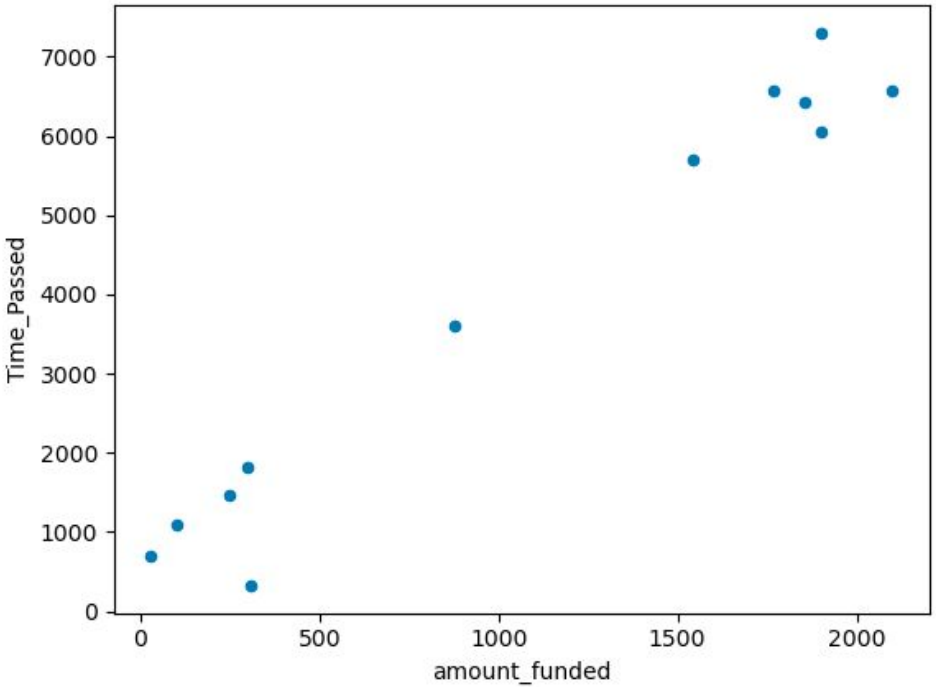
	amount_funded	Date1	project_description	Department_Name	Date_Passed	Value
0	1900	2000-02-02	Science	Deparment1	2020-02-02	1000
1	1856	2003-07-03	Particle Accelerator	Deparment1	2021-02-02	100
2	300	2017-02-02	Classifier	Deparment1	2022-02-02	300
3	876	2009-03-12	Photon perceptor	Deparment1	2019-02-02	1000
4	2100	2000-02-02	Science	Deparment2	2018-02-02	6000
5	1900	2003-07-03	Particle Accelerator	Deparment2	2020-02-02	7000
6	250	2017-02-02	Classifier	Deparment2	2021-02-02	30
7	30	2020-03-12	Photon perceptor	Deparment2	2022-02-02	7000
8	1768	2000-02-02	Science	Deparment3	2018-02-02	1140
9	1543	2003-07-03	Particle Accelerator	Deparment3	2019-02-02	800
10	100	2017-02-02	Classifier	Deparment3	2020-02-02	300
11	308	2020-03-12	Photon perceptor	Deparment3	2021-02-02	100

Table generated using MySql query, and exported to jupyter notebook and read as a dataframe using pandas libraries.

Table after Aggregate Functions

	amount_funded	Date1	project_description	Department_Name	Date_Passed	Value	Time_Passed
0	1900	2000-02-02	Science	Deparment1	2020-02-02	1000	7305
1	1856	2003-07-03	Particle Accelerator	Deparment1	2021-02-02	100	6424
2	300	2017-02-02	Classifier	Deparment1	2022-02-02	300	1826
3	876	2009-03-12	Photon perceptor	Deparment1	2019-02-02	1000	3614
4	2100	2000-02-02	Science	Deparment2	2018-02-02	6000	6575
5	1900	2003-07-03	Particle Accelerator	Deparment2	2020-02-02	7000	6058
6	250	2017-02-02	Classifier	Deparment2	2021-02-02	30	1461
7	30	2020-03-12	Photon perceptor	Deparment2	2022-02-02	7000	692
8	1768	2000-02-02	Science	Deparment3	2018-02-02	1140	6575
9	1543	2003-07-03	Particle Accelerator	Deparment3	2019-02-02	800	5693
10	100	2017-02-02	Classifier	Deparment3	2020-02-02	300	1095
11	308	2020-03-12	Photon perceptor	Deparment3	2021-02-02	100	327

Resulting Graph



It is clear from the resulting graph that the more time that is passed, the more a project will get funded.

Thank you!