Database Project

Stephen Antogiovanni

Domain Description

Title: **Biology Lab Database**

The Biology Lab Database is designed to store and manage information related to an accredited research biology lab. This database will include information about relevant experiments, principal investigators, biological samples, inventory equipment, data analysis software, safety requirements, and other relevant entities. The purpose of the database is to provide a centralized repository of information that can be used to support various biological research endeavors across research institutions, government-funded labs, universities, companies, or any other biological labs.

Entities and Relationships
The Biology Lab Database will include several key entities, each of which will be related to one another in various ways. These entities include:

Principal Investigator: This attribute would include information about the researcher (PI) who leads the research. The attributes will be the PI's ID and name.

Research Assistant: This entity will contain information about the specific personnel that the research lab grants access to. The information would include a research assistant ID, the ID of their principal investigator, and name. The research assistant ID unique identifies these participants of the research lab.

Lab Staff: This entity will contain information specific to the type of personnel that works in each biology lab that differentiates them from one another. This would include their unique lab staff ID, their corresponding PI ID or research assistant ID (null if not a PI or vice versa), contact information (email address, phone number, mailing address), title (either principal investigator, research assistant, graduate research assistant, postdoctoral research assistant, research technician, or research coordinator), and their affiliation (name of university, institution, or company they work at).

Experiments: This entity will contain information about specific experiments, such as the experiment ID to unique identify the experiment. It will contain the PI ID, experiment name, start and end dates, lab room number, and building ID.

Inventory: This entity will contain information about the general supplies and equipment used by researchers in the lab. The attributes will be item ID, item name, vendor name, storage location, lab room number, quantity, and purchase date.

Lab: This entity will contain information about the physical lab, specifically the room number and building ID.

Samples: This entity will contain information about the biological samples used in the experiments, including the name, lab room number, quantity, vendor name, collection date, and storage location. Each sample is differentiated from another sample by its name, lab room number, quantity, and collection date of such sample used in each experiment.

Data Analysis: This entity will contain information about the various data analysis pipelines used by the research lab including software tools, machines, and other related tools used to process the data extracted from the experiments. The attributes are pipeline name, pipeline type, vendor name, storage location, and lab room number.

Safety Protocols: This entity will contain information about lab safety protocols, such as chemical hazards, emergency procedures, and PPE (personal protective equipment requirements). The attributes will be safety protocol name, experiment ID, and description.

Publications: This entity will contain all relevant information on previous lab publications. The attributes will be publication ID, publication title, journal title, PI ID, and publication date. The publication ID uniquely identifies a new publication that emerged from an experiment.

Grants/Funding: This entity will contain information about any grants or funding via external or internal support that the research lab receives. The attributes will include a grant ID, agency name, grant title, grant type (research, training, or equipment grant), grant amount (in dollars), start and end dates of the award period, funding status (active, pending, or completed), experiment ID, and PI ID. The grant ID uniquely identifies a new grant awarded to an experiment.

General Assumptions:

- **Weak Entity Sets:**
  - The following entities are represented as weak entity sets:
    - Inventory
    - Samples
    - Data Analysis
  - These entities are weak because an experiment does not necessarily need to utilize any inventory, such as a microscope or test tube, since some experiments can rely on observations or use of pure data.
  - This is also true for biological samples, such as organisms, chemicals, or solutions. An experiment does not necessarily need to utilize traditional samples, although most do, therefore it is a weak entity.
  - Lastly, this holds true for data analysis since not all experiments need performance-based tools to sequence DNA or related tasks, some can be basic experiments.

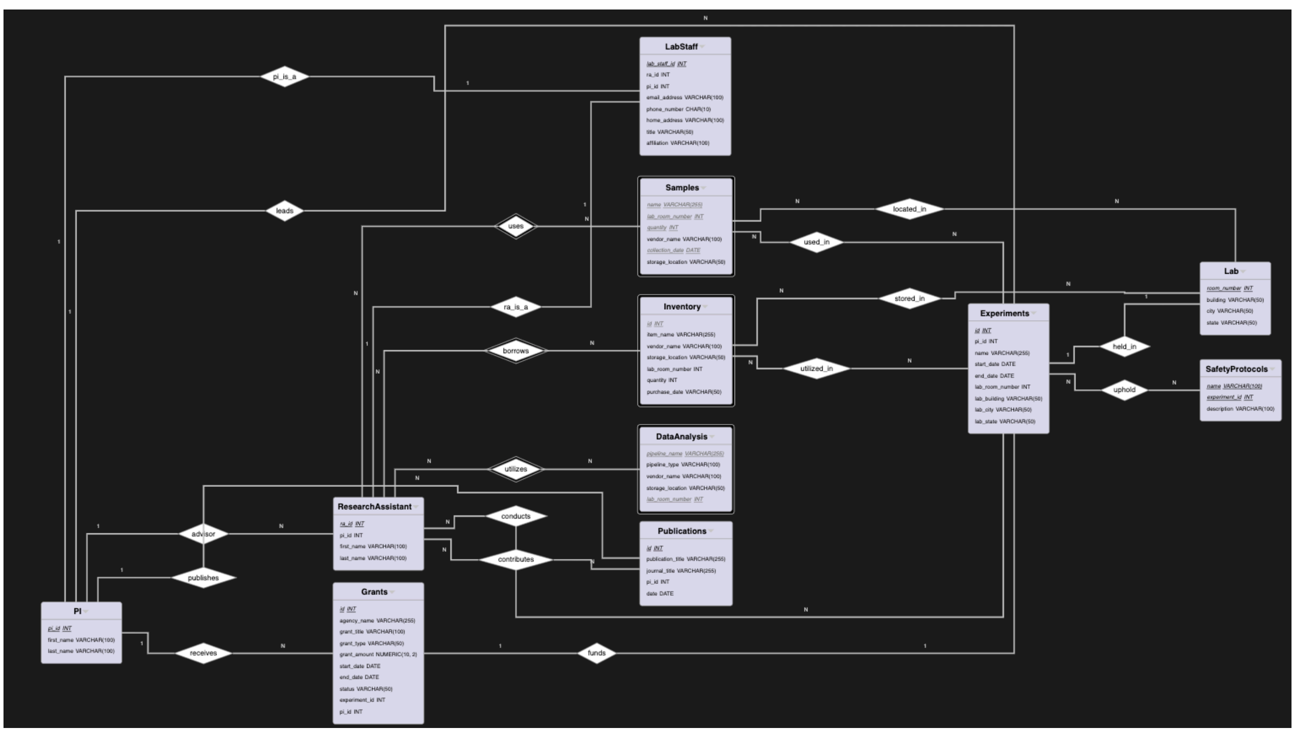
- **Cardinalities:**
    - **One-to-One:**
        - PI → Lab Staff
            - Only one PI person can be the same lab staff member.
        - Research Assistant → Lab Staff
            - Only one research assistant person can be the same lab staff member.
        - Grants → Experiments
            - Only one unique grant can be awarded to one experiment, despite many grants can be awarded to the same experiment, however these are not grants of the same type. Each grant is unique in id, amount, purpose, dates, etc.
        - Experiments → Lab
            - For the simplicity of this lab database, I decided to have each experiment be in its own physical lab room to avoid confusion and integrity constraints when doing my seed data.
    - **One-to-Many:**
        - PI → Research Assistant
            - A PI can have multiple research assistants under its supervision, but not vice versa – a research assistant can belong to multiple PI's (for the simplicity of this lab database, we are having research assistants be involved in only one research group).
        - PI → Experiments
            - A PI can run and lead multiple experiments and for the simplicity of this lab database, experiments cannot be run by more than 1 PI.
        - PI → Grants
            - A PI can receive many grants (not of the same type since each is unique) across their experiments, but grants cannot be distributed to many PI's for the simplicity of this lab database.
        - PI → Publications
            - A PI can be responsible for many publications that may emerge as a result of a successful experiment, but publications cannot be published by many PI's, for the simplicity of this lab database.
    - **Many-to-One:**
        - Inventory → Lab
            - Many inventory items such as graduated cylinders belong to one physical lab space, to avoid cross contamination, loss of equipment over time, and to separate specialized inventory that is specific to certain lab spaces.
        - Samples → Lab

      - Many biological samples such as live specimens and chemicals belong to one physical lab space, since their participation is specific for that certain lab space.
  - **Many-to-Many:**
    - Research Assistant → Experiments
      - Many research assistants can participate in many experiments (assuming these experiments are led by the same PI each time, for the simplicity of this lab database).
    - Research Assistant → Data Analysis
      - Many research assistants can use data analysis tools that belong to their respective experiments that utilize them.
    - Research Assistant → Inventory
      - Many research assistants can use the inventory that it is available to the lab space that their experiment is conducted in.
    - Research Assistant → Samples
      - Many research assistants use the many samples that may be available to their respective experiment.
    - Experiments → Safety Protocols
      - Many experiments can uphold many standard protocols which may be universal across many of the same experiments.
    - Research Assistant → Publications
      - Many research assistants can contribute to multiple publications, that can emerge out of the same PI-research group (since I am restricting each research assistant to belong to one PI for the simplicity of this lab database).
    - Inventory → Experiments
      - Many inventory items can be shared across experiments (only of the same lab space and led by the same PI) for the simplicity of this lab database.
    - Samples → Experiments
      - Many biological samples can be used, studied, and observed across different experiments, but only of the same lab space and led by the same PI for the simplicity of this lab database.

- **Note on *pi_id* Differences in 2 Different Tables:**
  - The *pi_id* in the *research_assistant* table refers to which PI is assigned to the respective research assistant; in contrast to the *pi_id* in the *lab_staff* table potentially being NULL for instantiating a research assistant (it is valid for a normal PI).
  - Also, this means that the *ra_id* in the *lab_staff* table is potentially NULL for instantiating a PI (it is valid for a normal research assistant).

- **Scope Limit:**
  - For the simplicity of this biology lab database, I have decided on the following restrictions:
    - RA's can participate in many experiments, however these experiments must be led by the same PI each time.
    - Each experiment must be in its own physical lab room.
    - A PI can run and lead multiple experiments, but experiments cannot be run by more than 1 PI.
    - A PI can receive many grants (not of the same type since each is unique) across their experiments, but grants cannot be distributed to multiple PI's.
    - A PI can be responsible for many publications that may emerge as a result of a successful experiment, but publications cannot be published by many PI's.
    - Only one unique grant can be awarded to one experiment, despite many grants can be awarded to the same experiment, however these are not grants of the same type.
    - Inventory items and samples belong to their own physical lab room.

Entity-Relationship Diagram:



*ER Diagram was not as blurry in VSCode.

Relational Schema:

*pi (**pi_id**, first_name, last_name)*

*research_assistant (**ra_id**, pi_id, first_name, last_name)*

*lab_staff (**lab_staff_id**, ra_id, pi_id, email_address, title, affiliation)*

*phone_numbers (**phone_number_id**, lab_staff_id, type, phone_number)*

*home_addresses (**id**, street, city, state, country, postal_code)*

*experiments (**id**, pi_id, name, start_date, end_date, lab_room_number, building_id)*

*inventory (**id**, item_name, vendor_name, storage_location, lab_room_number, quantity, purchase_date)*

*lab (**room_number**, **building_id**)*

*lab_building (**id**, building, street, city, state, country, postal_code)*

*samples (**name**, **lab_room_number**, **quantity**, vendor_name, **collection_date**, storage_location)*

*data_analysis (**pipeline_name**, pipeline_type, vendor_name, storage_location, **lab_room_number**)*

*safety_protocols (**name**, **experiment_id**, description)*

*publications (**id**, publication_title, journal_title, pi_id, date)*

*grants (**id**, agency_name, grant_title, grant_type, grant_amount, start_date, end_date, status, experiment_id, pi_id)*

Boyce-Codd Normal Form Decomposition:

A relation is in BCNF if and only if for every non-trivial functional dependency ($\alpha \rightarrow \beta$), α is a superkey or candidate key of the relation. A superkey is a set of attributes that uniquely identifies all the tuples in a relation, while a candidate key is a minimal superkey. In other words, a candidate key is a superkey where no subset of its attributes can uniquely identify all the tuples. Therefore, if a relation is in BCNF, there are no non-trivial functional dependencies between the attributes of a relation other than those implied by the candidate keys. This means that α cannot be a non-prime attribute if β is given as a prime attribute.

*pi (**pi_id**, first_name, last_name)*

Nontrivial Functional Dependencies:

- pi_id → first_name
- pi_id → last_name

Result: This schema is already in 1NF since it does not containing multiple values or composite attributes. It is also already in 2NF since it has a single candidate key (pi_id) and the non-key attributes (first_name, last_name) depend entirely on the primary key. It is also already in 3NF since there are no transitive dependencies present. Lastly, it is already in BCNF since it there are no non-trivial functional dependencies present that depend on a proper subset of a candidate key.

---

*research_assistant (**ra_id**, pi_id, first_name, last_name)*

Nontrivial Functional Dependencies:

- ra_id → pi_id
- ra_id → first_name
- ra_id → last_name

Result: This schema is in BCNF since there are no partial dependencies since the primary key ra_id determines all other attributes. There are also no transitive dependencies present.

---

*lab_staff (**lab_staff_id**, ra_id, pi_id, email_address, title, affiliation)*

Nontrivial Functional Dependencies:

- lab_staff_id → email_address, title, affiliation
- lab_staff_id → email_address
- lab_staff_id → title
- lab_staff_id → affiliation
- lab_staff_id → ra_id
- lab_staff_id → pi_id

Result: The lab_staff_id determines the corresponding email_address, title, and affiliation for a lab staff member. The lab_staff_id uniquely determines the ra_id or pi_id for the lab staff member

(Note: a lab staff member can have at most one associated ra_id or pi_id; if they are a RA, their pi_id will be null. If they are a PI, their ra_id will be null).

There are no partial dependencies since the primary key lab_staff_id determines all other attributes. Since there is a transitive dependency between ra_id and pi_id, the presence of the previous two schemas (pi and research_assistant) ensure that it is in BCNF. However, these schemas could have been optimized into the following:

- *lab_staff (**lab_staff_id,** first_name, last_name, email_address, title, affiliation)*
- *lab_staff_pi (**lab_staff_id,** pi_id)*
- *lab_staff_ra (**lab_staff_id,** ra_id)*

This would result in avoiding NULL values. However, I kept my original schemas because when seeding the data, the *pi_id* in the *research_assistant* table refers to which PI is assigned to such research assistant; in contrast to the *pi_id* in the *lab_staff* table potentially being NULL for instantiating a research assistant.

---

*phone_numbers (**phone_number_id**, lab_staff_id, type, phone_number)*

Nontrivial Functional Dependencies:

- phone_number_id → lab_staff_id, type, phone_number
- phone_number_id → lab_staff_id
- phone_number_id → type
- phone_number_id → phone_number
- lab_staff_id → type
- lab_staff_id → phone_number

Result: This schema is in BCNF since the determinant (phone_number_id) is a candidate key. There are also no non-trivial functional dependencies where the determinant is not a candidate key; therefore there are no non-trivial FD's that violate BCNF.

---

*home_addresses (**id**, street, city, state, country, postal_code)*

Nontrivial Functional Dependencies:

- id → street, city, state, country, postal_code
- id → street
- id → city
- id → state

- id → country
- id → postal_code

Result: Since the determinant (id) determines only a single value, it is in 1NF. There are also no non-trivial functional dependencies where the determinant (id) is not a candidate key. Therefore, this schema is already in BCNF.

---

*experiments (**id**, pi_id, name, start_date, end_date, lab_room_number, building_id)*

Nontrivial Functional Dependencies:

- id → pi_id, name, start_date, end_date, lab_room_number, building_id
- id → pi_id
- id → name
- id → start_date
- id → end_date
- id → start_date
- id → lab_room_number
- id → building_id

Result: Since the determinant (id) determines only a single value, it is in 1NF. There are also no non-trivial functional dependencies where the determinant (id) is not a candidate key. Therefore, this schema is already in BCNF.

---

*inventory (**id**, item_name, vendor_name, storage_location, lab_room_number, quantity, purchase_date)*

Nontrivial Functional Dependencies:

- id → item_name, vendor_name, storage_location, lab_room_number, quantity, purchase_date
- id → item_name
- id → vendor_name
- id → vendor_name
- id → storage_location
- id → lab_room_number
- id → quantity
- id → purchase_date

Result: Since the determinant (id) determines only a single value, it is in 1NF. There are also no non-trivial functional dependencies where the determinant (id) is not a candidate key. Therefore, this schema is already in BCNF.

---

*lab (**room_number**, **building_id**)*

Nontrivial Functional Dependencies:

- room_number → building_id
- building_id → room_number

Result: Since each determinant (room_number, building_id) determines only a single value, it is in 1NF. There are also no non-trivial functional dependencies where the determinant is not a candidate key. Therefore, this schema is already in BCNF.

---

*lab_building (**id**, building, street, city, state, country, postal_code)*

Nontrivial Functional Dependencies:

- id → building, street, city, state, country, postal_code
- id → building
- id → street
- id → city
- id → state
- id → country
- id → postal_code

Result: Since each determinant (id) determines only a single value, it is in 1NF. There are also no non-trivial functional dependencies where the determinant is not a candidate key. Therefore, this schema is already in BCNF.

---

*samples (**name**, **lab_room_number**, **quantity**, vendor_name, **collection_date**, storage_location)*

Nontrivial Functional Dependencies:

- (name, lab_room_number, quantity, collection_date) → *vendor_name, storage_location*
- (name, lab_room_number, quantity, collection_date) → *vendor_name*
- (name, lab_room_number, quantity, collection_date) → *storage_location*

Result: Since each determinant (name, lab_room_number, quantity, collection_date) determines only a single value, it is in 1NF. There are also no non-trivial functional dependencies where the determinant (name, lab_room_number, quantity, collection_date) is not a candidate key. Therefore, this schema is already in BCNF.

---

*data_analysis (**pipeline_name**, pipeline_type, vendor_name, storage_location, **lab_room_number**)*

Nontrivial Functional Dependencies:

- (pipeline_name, lab_room_number) → pipeline_type, vendor_name, storage_location
- (pipeline_name, lab_room_number) → pipeline_type
- (pipeline_name, lab_room_number) → vendor_name
- (pipeline_name, lab_room_number) → storage_location

Result: Since each determinant (pipeline_name, lab_room_number) determines only a single value, it is in 1NF. There are also no non-trivial functional dependencies where the determinant (pipeline_name, lab_room_number) is not a candidate key. Therefore, this schema is already in BCNF.

---

*safety_protocols (**name**, **experiment_id**, description)*

Nontrivial Functional Dependencies:

- (name, experiment_id) → description

Result: Since each determinant (name, experiment_id) determines only a single value, it is in 1NF. There are also no non-trivial functional dependencies where the determinant (name, experiment_id) is not a candidate key. Therefore, this schema is already in BCNF.

---

*publications (**id**, publication_title, journal_title, pi_id, date)*

Nontrivial Functional Dependencies:

- id → publication_title, journal_title, pi_id, date
- id → publication_title
- id → journal_title
- id → pi_id
- id → date

Result: Since each determinant (id) determines only a single value, it is in 1NF. There are also no non-trivial functional dependencies where the determinant (id) is not a candidate key. Therefore, this schema is already in BCNF.

---

*grants (**id**, agency_name, grant_title, grant_type, grant_amount, start_date, end_date, status, experiment_id, pi_id)*

Nontrivial Functional Dependencies:

- id → agency_name, grant_title, grant_type, grant_amount, start_date, end_date, status, experiment_id, pi_id
- id → agency_name
- id → grant_title
- id → grant_type
- id → grant_amount
- id → start_date
- id → end_date
- id → status
- id → experiment_id
- id → pi_id

Result: Since each determinant (id) determines only a single value, it is in 1NF. There are also no non-trivial functional dependencies where the determinant (id) is not a candidate key. Therefore, this schema is already in BCNF.

Transaction and Query Executions:

QUERY 1 → Query involving four or more relations and outer join:

Write a query to get the PI ID, combined PI first and last names, experiment name, and safety protocol description for those PIs that belong to the 'BioScience Research Center', the 'Advanced Technologies Corporation', and the 'Global Research Institute' along with their corresponding experiments and safety protocols, even if certain experiments do not have any associated safety protocols.

```
select p.pi_id, p.first_name, p.last_name, e.name, s.description
from pi as p
join lab_staff as ls on p.pi_id = ls.pi_id
join experiments as e on ls.lab_staff_id = e.pi_id
join lab_building as lb on ls.affiliation = lb.building
```

```
left join safety_protocols as s on e.id = s.experiment_id
where lb.building IN ('BioScience Research Center', 'Advanced Technologies
Corporation', 'Global Research Institute');
```

pi_id	pi_name	name	description
2	Emily Johnson	Examining the Effect of Soil Nutrient Composition on Microbial Diversity in Agricultural Systems	This protocol describes the proper disposal methods for biological waste generated in the laboratory.
4	Maria Rodriguez	Analyzing the Relationship Between Pollinator Behavior and Plant Reproductive Success	This protocol outlines the safety procedures for handling biohazardous materials in the laboratory.
7	Olivia Martinez	Assessing the Impact of Ocean Acidification on Coral Reef Ecosystems	This protocol provides guidelines for the regular maintenance and inspection of laboratory ventilation systems to ensure optimal performance.
8	Daniel Taylor	Examining the Effects of Air Pollution on Lung Function and Respiratory Health in Urban Environments	This protocol explains the proper use and importance of personal protective equipment in the laboratory.
12	William Lopez	Analyzing the Effects of Light Intensity on Algal Growth and Oxygen Production in Aquatic Systems	This protocol provides instructions for the safe handling, storage, and transport of compressed gas cylinders.
9	Sophia Anderson	Studying the Effect of Herbivore Grazing on Plant Defense Mechanisms and Herbivore Population Dynamics	This protocol provides guidelines for safely cleaning up chemical spills to prevent accidents and contamination.
14	Alexander Walker	Exploring the Relationship Between Dietary Habits and Gut Microbiota Composition in Humans	This protocol outlines safety guidelines for working with electrical equipment and avoiding electrical hazards.
20	James Mitchell	Investigating the Effects of Climate Change on Bird Migration and Breeding Patterns	This protocol details the safety measures and precautions to be followed when working with radioactive materials.
2	Emily Johnson	Studying the Role of Nitrogen Fixation in Plant Growth and Soil Fertility	This protocol explains the correct storage and labeling practices for chemicals to ensure safety and prevent accidents.
10	Matthew Thomas	Examining the Impact of Deforestation on Biodiversity and Ecosystem Functioning	This protocol details the proper handling, storage, and disposal of hazardous waste materials generated in the laboratory.
16	Benjamin Adams	Analyzing the Effects of Water Pollution on Fish Behavior and Reproduction	This protocol explains the use and maintenance of emergency eye wash and shower stations for immediate decontamination.
4	Maria Rodriguez	Investigating the Effects of Environmental Toxins on Developmental Disorders in Animal Models	NULL
18	Ethan Young	Analyzing the Effects of Climate Change on Insect Phenology and Population Dynamics	NULL
8	Daniel Taylor	Studying the Role of Microplastics in Aquatic Food Chains and Ecological Impacts	NULL
7	Olivia Martinez	Examining the Impact of Habitat Loss on Butterfly Migration and Population Decline	NULL
12	William Lopez	Analyzing the Effects of Noise Pollution on Animal Communication and Behavior	NULL
2	Emily Johnson	Analyzing the Effects of Climate Change on Plant-Pollinator Interactions and Reproduction	NULL
9	Sophia Anderson	Examining the Effects of Climate Change on Alpine Plant Communities and Biodiversity	NULL
10	Matthew Thomas	Analyzing the Relationship Between Biodiversity and Ecosystem Stability in Forested Areas	NULL
16	Benjamin Adams	Investigating the Effects of Heavy Metal Pollution on Aquatic Invertebrate Populations	NULL
14	Alexander Walker	Examining the Relationship Between Microbial Diversity and Nutrient Cycling in Marine Sediments	NULL
20	James Mitchell	Investigating the Role of Plant Secondary Metabolites in Defense Against Herbivores	NULL

QUERY 2 → Query involving four or more relations and subqueries:

Write a query to get the PI ID, experiment names, and their corresponding distinct inventory item names (hint: use group_concat) for such experiments that are conducted in any lab located in the 'Genetics and Genomics Institute' building. Include any experiments that are associated with a PI who has the title of 'Principal Investigator' and where the lab staff title is either 'Research Coordinator', 'Research Assistant', or 'Research Technician'.

pi_id	name	inventory_items
3	Exploring the Relationship Between Genetic Diversity and Disease Resistance in Plant Populations	Microarray Scanner,Microarray Slides
3	Exploring the Role of Genetic Variation in Disease Susceptibility among Human Populations	Gel Electrophoresis System,Petri Dishes
3	Studying the Effects of Hormone Disruption on Fish Reproduction and Endocrine System	Microarray Scanner,Microarray Slides
3	Studying the Effects of UV Radiation on Skin Pigmentation and DNA Damage in Human Cells	Quartz Cuvettes,UV-Vis Spectrophotometer
15	Exploring the Relationship Between Genetic Variation and Drug Response in Human Populations	Agarose,Gel Electrophoresis System
15	Investigating the Impact of Antibiotic Use on the Development of Antibiotic Resistance in Bacterial Populations	Agar Plates,Hot Plate Stirrer
15	Studying the Role of Genetic Factors in Age-Related Diseases and Longevity	Agarose Gel,Gel Electrophoresis System
19	Examining the Role of Epigenetics in Gene Expression and Developmental Plasticity	Electroporation Cuvettes,Electroporator
19	Exploring the Impact of Antibiotic Use in Livestock on Antibiotic Resistance in Human Pathogens	Microplate Reader,Microplates
19	Investigating the Role of Gut Microbiota in Human Health and Disease	Glass Slides,Incubator

---

QUERY 3 → Query involving aggregate function and grouping:

Write a query to get the PI ID, combined first and last names of the PI who leads each experiment (hint: use || ' ' ||), and the count of experiments for those PIs who have conducted more than 2 experiments. Group the results by PI ID.

```
select e.pi_id, pi.first_name || ' ' || pi.last_name as pi_name, count(*) as
experiment_count
from experiments as e, pi
where e.pi_id = pi.pi_id
group by e.pi_id
having count(*) > 2;
```

pi_id	pi_name	experiment_count
1	John Smith	3
2	Emily Johnson	3
3	David Garcia	4
5	Samantha Davis	3
6	Michael Wilson	4
13	Ava Lee	4
15	Charlotte Hall	3
19	Amelia Rivera	3

QUERY 4 → Query involving subquery with set comparison:

Write a query that involves a subquery with a set comparison to find all experiments where the average inventory count (quantity) is greater than 5. Group by experiment name.

```
select e.name, avg(i.quantity) as avg_inventory_count
from experiments as e
join inventory as i on e.lab_room_number = i.lab_room_number
where e.name in (
    select e.name
    from experiments as e
    join inventory as i on e.lab_room_number = i.lab_room_number
    group by e.name
    having avg(i.quantity) > 5
)
group by e.name;
```

SQL ▾ 1 / 1 1 – 45 of 4

name	avg_inventory_count
Analyzing the Effects of Climate Change on Insect Phenology and Population Dynamics	50.5
Analyzing the Effects of Climate Change on Migratory Bird Stopover Ecology	50.5
Analyzing the Effects of Climate Change on Plant-Pollinator Interactions and Reproduction	25.5
Analyzing the Effects of Light Intensity on Algal Growth and Oxygen Production in Aquatic Systems	20.5
Analyzing the Effects of Noise Pollution on Animal Communication and Behavior	5.5
Analyzing the Effects of Urban Gardens on Pollinator Diversity and Plant-Pollinator Networks	25.5
Analyzing the Impact of Habitat Fragmentation on Wildlife Migration Patterns	300.0
Analyzing the Relationship Between Biodiversity and Ecosystem Stability in Forested Areas	25.5
Analyzing the Relationship Between Pollinator Behavior and Plant Reproductive Success	200.0
Assessing the Impact of Ocean Acidification on Coral Reef Ecosystems	105.0
Examining the Effects of Air Pollution on Lung Function and Respiratory Health in Urban Environments	350.0
Examining the Effects of Climate Change on Alpine Plant Communities and Biodiversity	50.5
Examining the Effects of Pesticide Exposure on Bee Colony Health and Survival	26.5
Examining the Impact of Deforestation on Biodiversity and Ecosystem Functioning	26.5
Examining the Impact of Habitat Loss on Butterfly Migration and Population Decline	25.5
Examining the Impact of Invasive Species on Native Ecosystems and Species Interactions	27.5
Examining the Relationship Between Microbial Diversity and Nutrient Cycling in Marine Sediments	50.5
Examining the Role of Epigenetics in Gene Expression and Developmental Plasticity	25.5
Exploring the Impact of Antibiotic Use in Livestock on Antibiotic Resistance in Human Pathogens	25.5
Exploring the Relationship Between Bacterial Diversity and Soil Health in Agricultural Systems	25.5
Exploring the Relationship Between Dietary Habits and Gut Microbiota Composition in Humans	125.0
Exploring the Relationship Between Genetic Diversity and Disease Resistance in Plant Populations	50.5
Exploring the Relationship Between Genetic Traits and Athletic Performance in Human Populations	26.5
Exploring the Relationship Between Microbial Communities and Nutrient Cycling in Forest Ecosystems	25.5
Exploring the Relationship Between Soil Microbes and Carbon Sequestration in Agroecosystems	50.5
Exploring the Role of Genetic Variation in Disease Susceptibility among Human Populations	51.5
Investigating the Effects of Climate Change on Bird Migration and Breeding Patterns	100.5
Investigating the Effects of Different Fertilizer Formulations on Crop Yield and Soil Health	125.5
Investigating the Effects of Environmental Toxins on Developmental Disorders in Animal Models	50.5
Investigating the Effects of Heavy Metal Pollution on Aquatic Invertebrate Populations	5.5
Investigating the Effects of Urbanization on Wildlife Habitat Use and Connectivity	101.0
Investigating the Impact of Antibiotic Use on the Development of Antibiotic Resistance in Bacterial Populations	11.0
Investigating the Impact of Microplastic Contamination on Freshwater Organisms and Ecosystems	100.5
Investigating the Impact of Plastic Pollution on Marine Organisms and Ecosystem Health	100.5
Investigating the Impact of Temperature on Plant Growth and Photosynthesis Efficiency	102.5
Investigating the Role of Gut Microbiota in Human Health and Disease	100.5
Investigating the Role of Plant Secondary Metabolites in Defense Against Herbivores	25.5
Studying the Effect of Herbivore Grazing on Plant Defense Mechanisms and Herbivore Population Dynamics	30.0
Studying the Effects of Hormone Disruption on Fish Reproduction and Endocrine System	25.5
Studying the Effects of UV Radiation on Skin Pigmentation and DNA Damage in Human Cells	25.5
Studying the Impact of Light Pollution on Nocturnal Animal Behavior and Ecosystem Functioning	50.5
Studying the Role of Genetic Factors in Age-Related Diseases and Longevity	10.5
Studying the Role of Microbes in Soil Nutrient Cycling and Plant Nutrient Uptake	6.5
Studying the Role of Microplastics in Aquatic Food Chains and Ecological Impacts	51.5
Studying the Role of Nitrogen Fixation in Plant Growth and Soil Fertility	50.5

QUERY 5 → Query involving grouping and aggregate function:

Write a query to get the vendor names and the count of samples for each vendor from the samples relation. Group by vendor name and include only those vendors that have a sample count of two or more (hint: use having).

```
select s.vendor_name, count(*) as sample_count
from samples as s
group by s.vendor_name
having count(*) >= 2;
```

vendor_name	sample_count
Cellular Technologies Ltd.	3
Enzyme Technologies Ltd.	2
Genomics Solutions Inc.	3
Microbial Solutions Inc.	4
Microbiology Innovations Inc.	2
Molecular Research Supplies	2
Molecular Technologies Corp.	3

QUERY 6 → Query involving grouping and aggregate function and subquery:

Write a query to find the maximum number of inventory items for all experiments (include experiment ID) where there are research assistants assigned to such experiment and that live in the state 'TX'. Group by experiment ID.

```
select e.id as experiment_id, max(i.item_count) as max_inventory_items
from experiments as e
join lab as l on l.room_number = e.lab_room_number
join research_assistant as ra on ra.pi_id = e.pi_id
join lab_staff as ls on ls.ra_id = ra.ra_id
join home_addresses as ha on ha.id = ls.lab_staff_id
join (
    select lab_room_number, count(*) as item_count
    from inventory
    group by lab_room_number
) as i on i.lab_room_number = e.lab_room_number
where ha.state = 'TX'
group by e.id;
```

experiment_id	max_inventory_items
8	2
10	2
14	2
19	2
28	2
30	2
33	2
44	2

QUERY 7 → Query involving set operations:

Write a query to get distinct vendor names from the samples relations from the 'Advanced Technologies Corporation' and the 'Data Analytics Institute'; use the set operation UNION to create the unified list.

```sql
select distinct vendor_name
from samples as s
join lab as l on l.room_number = s.lab_room_number
join lab_building as lb on lb.id = l.building_id
where lb.building = 'Advanced Technologies Corporation'

UNION

select distinct vendor_name
from samples as s
join lab as l on l.room_number = s.lab_room_number
join lab_building as lb on lb.id = l.building_id
where lb.building = 'Data Analytics Institute';
```

vendor_name
Fungi Research Labs
Genomics Technologies Inc.
Microbial Solutions Inc.
Microbiology Innovations Inc.
Molecular Technologies Corp.
Neuroscience Research Labs
Plant Genetics Solutions
Plant Nursery Supplies
Protein Solutions Inc.
Proteomics Technologies Inc.