Schema Design

Need Analysis

- User System:
  - Sign up: Name, billing address, email
  - One user <- many locations
- Location:
  - Able to modify locations
  - Attributes:
    - Full address, including unit number
    - Start date
    - Square feet
    - Num bedrooms
    - Number of occupants of the unit
    - List of registered devices
- Devices
  - Type: AC, fridge, ...
  - Model: Samsung AC R100, ...
  - Ownership: User, Locations. (Can have more than one same model in a location by a user)
  - Can have an ID
- Devices Events(Signal / Information)
  - Switching On / Off (Time)
  - Setting Changed (Time, setting[possibly]). e.g. AC temperature, light bulb brightness, fridge door opened / closed, ...
  - Energy consumption
    - info / 5min: energy use
    - switched off: energy use since last update
- Information format
  - Device ID
  - time stamp
  - label (e.g. energy_consumption, temperature_lowered, door_opened, ...)
  - number (corresponding to label)
- Energy Price
  - Vary according to:
    - Hour
    - Zip Code

Tables Design

1. User & Location Table

Assumption:

1. *Each user can only have one billing address. So billing_address could be an attribute in user table*
2. *Each home location can only belong to one account, so customer_id & start_date should be an attribute in address table*
3. *We assume the location format as the following: (Line 1)street_number street_name (Line 2)unit_number (zip code info)city, state zipcode.*
4. *We also asuume the input addresses are all legal(they really exist, and in the right format)*

customer: (**customer_id**, first_name, last_name, email, billing_street_num, billing_street_name, billing_unit_number, billing_city, billing_state, billing_zipcode, cpassword)
location: (**location_id**, customer_id, location_street_num, location_street_name, location_unit_number, location_city, location_state, location_zipcode, square_feet, num_bedrooms, num_occupants)

2. Device & Event

Assumptions:

1. *All the device models are in the list of device_model table, whenever a new device is promoted, we can modify the database to put it into the table*
2. *User can only register devices of which the models are in the device_model table*
3. *There are only limited number of event_label, and every the event_label revceived by the system should be legal*
4. *As mentioned in the problem desciption, we don't have to model how the system prestored all the event_labels. So we assume that the events are automatically stored into the database. In the project, this process might be simulated by manaully insert data into the model_event table*

device_model(**model_id**, model_type, model_name), *This is for prestoring devices for user to register*
device_registered(**device_id**, model_id, location_id), *This is for devices registered by user*
device_event(**device_id**, event_label, event_datetime, event_number), *event_number corresponds to event_label*

3. Energy Price

energy_price(**zipcode**, **hour_of_day**, price), *Energy prices vary on hourly and locational basis*

Database Creation

In this part, we choose MySQL to implement the schema

1. customer

```
CREATE TABLE customer(
    customer_id INT AUTO_INCREMENT,
    first_name VARCHAR(63) NOT NULL,
    last_name VARCHAR(63) NOT NULL,
```

```
    email VARCHAR(255) NOT NULL,
    billing_street_num INT NOT NULL,
    billing_street_name VARCHAR(127) NOT NULL,
    billing_unit_number VARCHAR(127) NOT NULL,
    billing_city VARCHAR(127) NOT NULL,
    billing_state VARCHAR(16) NOT NULL,
    billing_zipcode VARCHAR(5) NOT NULL,
    cpassword VARCHAR(127) NOT NULL,
    PRIMARY KEY (customer_id)
);
```

2. location

```
CREATE TABLE location(
    location_id INT AUTO_INCREMENT,
    customer_id INT NOT NULL,
    location_street_num INT NOT NULL,
    location_street_name VARCHAR(127) NOT NULL,
    location_unit_number VARCHAR(127) NOT NULL,
    location_city VARCHAR(127) NOT NULL,
    location_state VARCHAR(127) NOT NULL,
    location_zipcode VARCHAR(5) NOT NULL,
    square_feet FLOAT NOT NULL,
    num_bedrooms INT NOT NULL,
    num_occupants INT NOT NULL,
    PRIMARY KEY (location_id),
    FOREIGN KEY (customer_id) REFERENCES customer(customer_id)
);
```

3. device_model

```
CREATE TABLE device_model(
    model_id INT AUTO_INCREMENT,
    model_type VARCHAR(127) NOT NULL,
    model_name VARCHAR(127) NOT NULL,
    PRIMARY KEY (model_id)
);
```

4. device_registered

```
CREATE TABLE device_registered(
    device_id INT AUTO_INCREMENT,
    model_id INT NOT NULL,
    location_id INT NOT NULL,
    PRIMARY KEY (device_id),
    FOREIGN KEY (model_id) REFERENCES device_model(device_id),
    FOREIGN KEY (location_id) REFERENCES location(location_id)
);
```

5. device_event

```
CREATE TABLE device_event(
    device_id INT NOT NULL,
    event_datetime DATETIME DEFAULT CURRENT_TIMESTAMP,
    event_label VARCHAR(63) NOT NULL,
    event_number FLOAT,
    PRIMARY KEY (device_id, event_datetime)
);
```

6. energy_price(**zipcode**, **hour_of_day**, price)

```
CREATE TABLE energy_price(
    zipcode VARCHAR(5),
    hour_of_day INT NOT NULL,
    price FLOAT NOT NULL,
    PRIMARY KEY (zipcode, hour_of_day)
);
```

SQL

1. List all enrolled devices with their total energy consumption in the last 24 hours, for a specific customer identified by customer ID.

```
SELECT dr.device_id, SUM(de.event_number) AS
total_energy_consumption
FROM device_event de
JOIN device_registered dr ON de.device_id = dr.device_id
JOIN location l ON l.location_id = dr.location_id
WHERE de.event_label = 'energy use'
GROUP BY de.device_id and l.customer_id = "specific customer_id"
HAVING MAX(de.event_datetime) >= NOW() - INTERVAL 24 HOUR;
```

2. Calculate the average monthly energy consumption per device type, for the month of August 2022, considering only devices that have been on (i.e., reported data) at least once during that month.

```
SELECT d.device_label, AVG(total_energy_consumption) AS
average_energy_consumption
FROM ( # first compute all the device monthly energy consumption
    SELECT dr.device_id, SUM(de.event_number) AS
total_energy_consumption
    FROM device_event de
    JOIN device_registered dr ON de.device_id = dr.device_id
    WHERE de.event_label = 'energy use'
    AND de.timestamp >= '2022-08-01' AND de.timestamp < '2022-
09-01'
    GROUP BY dr.device_id
    HAVING SUM(de.event_number) IS NOT NULL # only consider
devices that have on at least once
) AS subquery
```

```
JOIN device_registered dr ON subquery.device_id = dr.device_id
JOIN device_model dm ON dr.model_id = dm.model_id
GROUP BY d.device_type;
```

3. Identify cases where a refrigerator door was left open for more than 30 minutes. Output the date and time, the service location, the device ID, and the refrigerator model.

```
SELECT dr.device_od, dr.model_id,
FROM device_event de
JOIN device_registered dr ON de.device_id = dr.device_id
JOIN device_model dm ON dr.model_id = dm.model_id
WHERE m.event_label = 'door opened'
AND dm.model_type = "refrigerator"
AND (
  TIMEDIFF (
  (
    SELECT MIN(timestamp)
    FROM model_event AS m2
    WHERE m2.device_id = m.device_id
    AND m2.timestamp > m.timestamp
    AND m2.event_label = 'door closed'
  )
  , m.event_datetime
  ) > '00:30:00'
  OR
  (
    TIMEDIFF(NOW(), m.event_datetime) > '00:30:00'
    AND NOT EXISTS(
      SELECT 1
      FROM model_event AS m3
      WHERE m3.device_id = m.device_id
      AND m2.timestamp > m.timestamp
    )
  )
)
```

4. Calculate the total energy cost for each service location during August 2022, considering the hourly changing energy prices based on zip code.

```
# calculate the cost for every hour
SELECT SUM(ep.price * de.event_number/12) as monthlyCostSum
FROM device_event de
JOIN device_registered dr ON de.device_id = dr.device_id
JOIN location l ON dr.location_id = l.location_id
JOIN energy_price ep ON ep.zipcode = l.location_zipcode AND
ep.hour_of_day = HOUR(de.event_datetime)
GROUP BY dr.location_id
WHERE de.event_label = 'energy use' AND de.event_datetime
BETWEEN "2022-08-01" AND "2022-08-31"
```

5. For each service location, compute its total energy consumption during August 2022, as a percentage of the average total energy consumption during the same time of other service locations that have a similar square footage (meaning, at most 5% higher or lower square footage). Thus, you would output 150% if a service location with 1000 sqft had 50% higher energy consumption than the average of other service locations that have between 950 and 1050 sqft.

6. Identify service location(s) that had the highest percentage increase in energy consumption between August and September of 2022.