Analyzing EV Charging Infrastructure: A Data-Driven Approach

Logical Model (Relational Model)

Group 2

Rohan Verma

Priyansh Nileshbhai Vagadiya

617 459 9214

857 339 8070

verma.rohan@northeastern.edu

vagadiya.p@northeastern.edu

Percentage of Effort Contributed by Student1 : 50%

Percentage of Effort Contributed by Student2 : 50%

Signature of Student 1 : Rohan Verma

Signature of Student 2 : Priyansh Nileshbhai Vagadiya

Updated schema for our project :

1. **station :** (id (PK), station_name , station_number
2. **station_address:** (station_id (PK and FK TO ID) , street_address, city, state, zip country, longitude, latitude)
3. **electric_consumption_wrt_station :**(station_id (PK,FK TO ID), electric_consumption_kWh_perday, state)
4. **fuel_type :** (station_id (PK,FK TO ID), fuel_type),
5. **station_compatibility :** (station_id (PK,FK TO ID), evlevel1_charging, evlevel2_charging, dc_fast_count)
6. **acesibility_of_station :** (station_id (PK,FK TO ID), intersection_direction, access_timing, geocode)
7. **pament_methods :** (station_id (PK,FK TO ID), cards , cash , checks)
8. **owner :** (station_id (PK,FK TO ID), owner_type ,federal_agency),
9. **compliance of station :** (station_id (PK,FK TO ID), last_inspection_date, ,next_inspection_date ,safety_issue_reported, ,compliance_status) ,
10. **ev_network :** (station_id (PK,FK TO ID), ev network types).

1. First Normal Form (1NF)

For a relation to be in **First Normal Form (1NF)**, it must meet the following criteria:

- Each column must contain atomic values .
- Each record must have a unique identifier .
- Each attribute must contain only one value per record.

Let's ensure that all attributes are atomic:

- **station** is already in atomic form.
- **station_address** does not violate 1NF as it only contains atomic data.
- **electric_consumption_wrt_stationlocation** has atomic fields such as electric_consumption_kWh_perday.
- **fuel_type** has only one type of fuel per station, which is atomic.
- **compatibility_of_station** and other tables appear to conform to 1NF as they don't contain repeating groups.

2. Second Normal Form (2NF) :

For a relation to be in **Second Normal Form (2NF)**, it must be in 1NF and:

- All non-key attributes must be fully functionally dependent on the primary key.
- This eliminates partial dependencies.

Breakdown of 2NF issues:

1. **station_address**: The primary key is station_id. But the information such as street_address, city, etc., is dependent solely on the station, not on any other part of the key.
    - This table is already in 2NF because each field is dependent on the full key (station_id).
2. **electric_consumption_wrt_stationlocation** : The primary key here is station_id, and the attribute electric_consumption_kWh_perday is fully dependent on station_id. There's no partial dependency, so this table is also in 2NF.
3. **fuel_type**: fuel_type is directly dependent on station_id, so no issues here in terms of 2NF.
4. **station_compatibility**: Same as before, attributes like evlevel1_charging, evlevel2_charging, and dc_fast_count are fully dependent on station_id. This table is in 2NF.
5. **acesibility_of_station**: Attributes like intersection_direction, access_timing, and geocode are dependent on station_id. No partial dependencies are present.
6. **payment_methods**: Same case as the other tables—attributes like cards, cash, and checks are fully dependent on station_id.
7. **owner**: owner_type and federal_agency are fully dependent on station_id.
8. **compliance_of_station**: The columns last_inspection_date, next_inspection_date, safety_issue_reported, and compliance_status are fully dependent on station_id.
9. **ev_network**: ev_network_types are fully dependent on station_id.

So, all tables are already in **2NF**.

3. Third Normal Form (3NF)

For a relation to be in **Third Normal Form (3NF)**, it must be in 2NF, there must be no transitive dependencies.

Breakdown of 3NF issues:

1. **station**: station_name and station_number are directly dependent on station_id, and no transitive dependencies exist here.
2. **station_address**: No non-key attributes are dependent on other non-key attributes. All fields depend only on station_id, so this table is in 3NF.
3. **electric_consumption_wrt_stationlocation**: All attributes depend only on station_id. There are no non-key dependencies that violate 3NF.
4. **fuel_type**: Similarly, fuel_type depends only on station_id, so no transitive dependencies.
5. **compatibility_of_station**: Attributes like evlevel1_charging, evlevel2_charging, and dc_fast_count are all dependent only on station_id.
6. **acesibility_of_station**: Same logic, all attributes depend directly on station_id.
7. **payment_methods**: Each attribute is directly dependent on station_id.
8. **owner**: Same structure, where owner_type and federal_agency depend solely on station_id.
9. **compliance_of_station**: All fields in this table are directly dependent on station_id.
10. **ev_network**: No non-key dependencies that violate 3NF.

Since there are no transitive dependencies in any of the tables, the schema is in **3NF**.

4. Boyce-Codd Normal Form (BCNF)

A relation is in **Boyce-Codd Normal Form (BCNF)** if:

- For every functional dependency ($X \rightarrow Y$), X is a superkey.

In our schema:

- All tables have their primary key as the determinant of the other attributes (i.e., the attributes in each table depend on the primary key station_id).
- There are no cases where a non-superkey determines another attribute. Hence, all tables are in **BCNF**.

5. Fourth Normal Form (4NF)

For a relation to be in **Fourth Normal Form (4NF)**, it must be in Boyce-Codd Normal Form (BCNF) and it must not have any multi-valued dependencies.

Breakdown of 4NF issues:

- **station_address**: No multivalued dependencies.
- **electric_consumption_wrt_stationlocation**: No multivalued dependencies.
- **fuel_type**: Since each station can only have one type of fuel, this is not a multivalued dependency.
- **compatibility_of_station**: This table has attributes like evlevel1_charging, evlevel2_charging, and dc_fast_count, but they depend directly on STATION_ID, and there are no multi-valued attributes.
- **acesibility_of_station**: No multivalued dependencies.
- **payment_methods**: If a station can accept multiple methods (cards, cash, checks), but these are distinct and dependent on station_id, so no multivalued dependency exists.
- **owner**: No multivalued dependencies.
- **compliance_of_station**: No multivalued dependencies.
- **ev_network**: No multivalued dependencies.

Thus, all tables are in **4NF**.