DATA MODEL

This is a comprehensive data dictionary explaining each database schema:

1. **Consumer Complaints Database Schema:**
   - ComplaintID (Unique identifier for each complaint):
     An alphanumeric code that uniquely identifies each complaint.
   - Date (Date of the complaint):
     The date when the complaint was registered.
   - ServiceType (Type of service: Telecom or Pay TV):
     Indicates whether the complaint pertains to telecommunications services or pay TV services.
   - ServiceProvider (Name of the service provider):
     The name of the company or organization providing the service.
   - ComplaintType (Type of complaint: Billing, Service Quality, Customer Service, etc.):
     Describes the nature or category of the complaint.
   - CustomerName (Name of the customer):
     The full name of the customer lodging the complaint.
   - CustomerContact (Contact information of the customer):
     Contact details of the customer, which may include a phone number or email address.
   - Description (Detailed description of the complaint):
     A narrative or detailed explanation of the customer's complaint.
   - ResolutionStatus (Status of complaint resolution: Open, In Progress, Resolved, etc.):
     Indicates the current status of resolving the complaint.

2. **Service Provider Performance Database Schema:**
   - ProviderID (Unique identifier for each service provider):
     A unique identifier assigned to each service provider.
   - ProviderName (Name of the service provider):
     The name of the company or organization providing services.
   - NetworkReliability (Metrics related to network reliability):
     Metrics indicating the reliability of the service provider's network.

- CallDropRate (Percentage of dropped calls):
  The percentage of calls that are dropped or disconnected.

- BroadbandSpeed (Average broadband speed):
  The average speed of broadband or internet service provided.

- CustomerSatisfaction (Customer satisfaction ratings):
  Ratings or scores representing customer satisfaction with the service.

3. **Regulatory Compliance Database Schema:**
   - ProviderID (Unique identifier for each service provider):
     A unique identifier assigned to each service provider.

   - ProviderName (Name of the service provider):
     The name of the company or organization providing services.

   - ComplianceStatus (Status of regulatory compliance):
     Indicates whether the service provider is compliant with regulatory requirements.

   - RegulatoryRequirements (Details of regulatory requirements):
     Detailed information about the specific regulatory requirements that apply.

   - AuditDate (Date of compliance audit):
     The date when compliance with regulations was audited.

4. **Billing and Pricing Database Schema:**
   - ProviderID (Unique identifier for each service provider):
     A unique identifier assigned to each service provider.

   - ProviderName (Name of the service provider):
     The name of the company or organization providing services.

   - PricingPlan (Details of pricing plans):
     Information about the different pricing plans offered by the service provider.

   - TariffChanges (Details of tariff changes):
     Information about changes in pricing or tariffs.

   - BillingDisputes (Details of billing disputes):
     Information about disputes related to billing or charges.

5. **Customer Feedback Database Schema:**

- FeedbackID (Unique identifier for each feedback):
  An alphanumeric code that uniquely identifies each feedback or survey response.

- Date (Date of feedback):
  The date when the feedback or survey response was submitted.

- ServiceType (Type of service: Telecom or Pay TV):
  Indicates whether the feedback pertains to telecommunications services or pay TV services.

- ServiceProvider (Name of the service provider):
  The name of the company or organization providing the service.

- CustomerName (Name of the customer):
  The full name of the customer providing the feedback.

- CustomerContact (Contact information of the customer):
  Contact details of the customer, which may include a phone number or email address.

- FeedbackText (Customer feedback or survey responses):
  The textual feedback or responses provided by the customer.

- Rating (Customer rating):
  A numerical rating or score given by the customer.

6. Service Outages and Disruptions Database Schema:

- OutageID (Unique identifier for each outage):
  An alphanumeric code that uniquely identifies each outage or disruption.

- Date (Date of the outage):
  The date when the outage or disruption occurred.

- ServiceType (Type of service: Telecom or Pay TV):
  Indicates whether the outage pertains to telecommunications services or pay TV services.

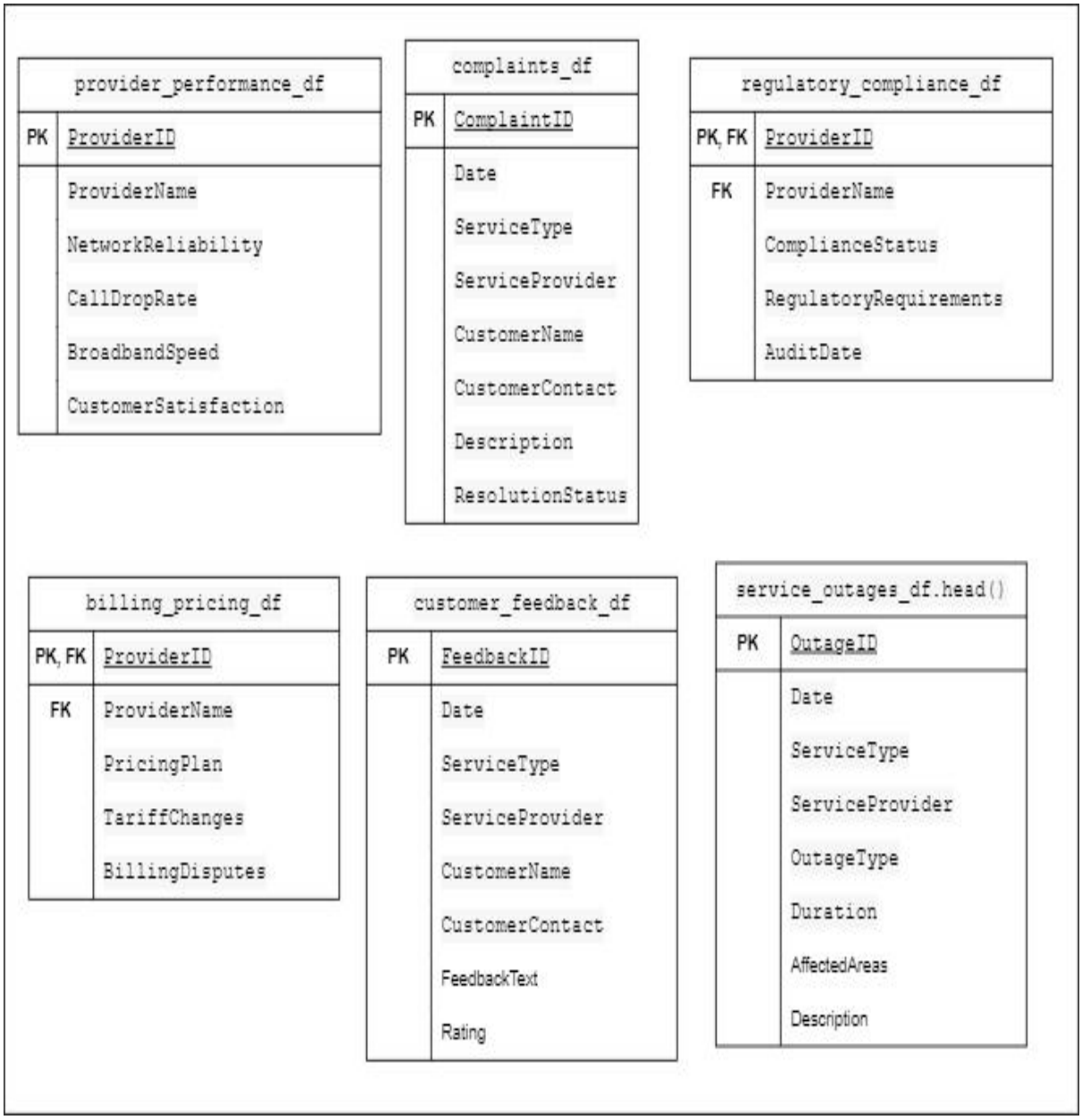
- ServiceProvider (Name of the service provider):
  The name of the company or organization providing the service.

- OutageType (Type of outage: Network, Service, Downtime, etc.):
  Describes the nature or category of the outage.

- Duration (Duration of the outage):
  Indicates how long the outage or disruption lasted.

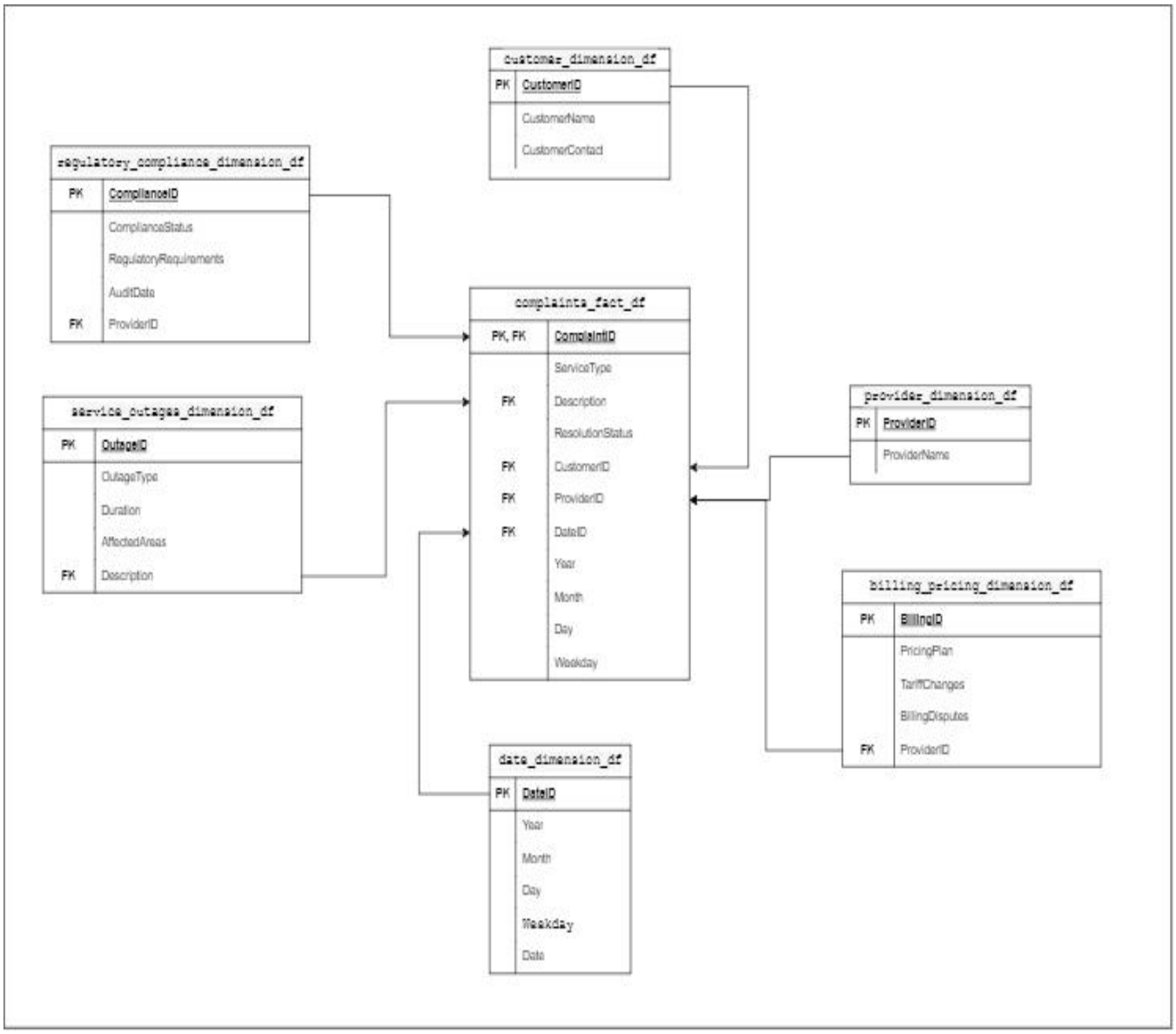
- AffectedAreas (Geographical areas affected):
  Specifies the geographical regions impacted by the outage.

- Description (Description of the outage):
  A narrative or detailed description of the outage or disruption.



Dimension Model and Fact Table

The creation of dimension models and fact tables from the datasets involves a process of organizing and structuring the data for effective analysis and reporting. Here's a comprehensive explanation of the logic behind creating these dimension models and the fact table:



1. **Dimension Model 1 - Customer Dimension:**
   - CustomerID:
     This is a unique identifier for each customer. It is created to establish a link between the fact table and the customer-related information in the complaints dataset.

   - CustomerName:
     This field stores the name of the customer. It provides a way to identify customers by

name.

- CustomerContact:
  This field contains the contact information of the customer, such as a phone number or email address. It helps in communication and further analysis.

To create this dimension, we extracted distinct customer information from the **complaints_dataset**, which includes **CustomerName** and **CustomerContact**. This gives us a list of unique customers.

2. **Dimension Model 2 - Provider Dimension:**
   - ProviderID:
     This unique identifier is assigned to each service provider. It serves as a reference point to link the fact table with information related to service providers in various datasets.

   - ProviderName:
     This field stores the name of the service provider. It allows for easy identification of service providers in the analysis.

We created this dimension by extracting distinct service provider information from various datasets. In this case, we used the **ProviderName** from the **provider_performance_dataset**. This gives us a list of unique service providers.

3. **Dimension Model 3 - Date Dimension:**
   - DateID:
     This unique identifier is created for each date in the dataset. It serves as a key for linking the fact table to date-related information. We generated a list of unique dates within a specified range (from '2022-01-01' to '2023-08-30') to create this dimension. We assigned a unique identifier (DateID) to each date.

   - Date:
     This field stores the date in YYYY-MM-DD format, providing a standardized date representation. We included the actual date in YYYY-MM-DD format.
   - Year:
     The year part of the date is extracted and stored separately. This allows for aggregating data at the year level.

   - Month:
     Similarly, the month part of the date is extracted and stored for aggregating data at the monthly level.

   - Day:
     This field represents the day of the month, allowing for analysis at the daily level.

   - Weekday:
     The day of the week is recorded, which helps in analyzing patterns related to weekdays.

4. **Fact Table - Complaints Fact Table:**
    - ComplaintID (Foreign key to the Complaints dataset):
      This field serves as a foreign key that links the fact table to the original complaints dataset. It allows us to associate each record in the fact table with a specific complaint in the dataset.

    - CustomerID (Foreign key to the Customer Dimension):
      We used the CustomerName and CustomerContact fields from the complaints_dataset to look up the corresponding CustomerID in the Customer Dimension. This links each complaint to a specific customer.

    - ProviderID (Foreign key to the Provider Dimension):
      Similarly, we used the ProviderName from the complaints_dataset to look up the corresponding ProviderID in the Provider Dimension. This links each complaint to a specific service provider.

    - Description:
      We included the Description field directly from the complaints_dataset to provide a detailed description of each complaint.

    - ServiceType (Type of service: Telecom or Pay TV):
      We extracted the ServiceType attribute from the complaints_dataset to categorize each complaint by the type of service it pertains to.

    - ComplaintType (Type of complaint):
      This field represents the category or type of complaint and is extracted directly from the complaints_dataset.

    - ResolutionStatus (Status of complaint resolution):
      We included the ResolutionStatus field from the complaints_dataset to indicate the current status of resolving each complaint.

    - Year (Year part of the date):
      We extracted the year from the date of each complaint and included it in the fact table for time-based analysis.

    - Month (Month part of the date):
      Similarly, we extracted the month from the date to provide a month-level granularity for analysis.

    - Day (Day of the month) and Weekday (Day of the week):
      These attributes were included to allow for further time-based analysis and reporting.

The logic behind creating these dimension models and the fact table is to provide a structured and organized foundation for analyzing and reporting on customer complaints data. By using unique

identifiers and foreign keys, you establish relationships between the fact table and dimension models, enabling efficient data analysis and exploration across different dimensions, including customers, service providers, and dates.