**Documentation**

**Schema Design Decisions:**

1. Companies Table:

- Primary key: id

- UNIQUE constraint on name to prevent duplicate companies

- NOT NULL constraints on critical fields (name, created\_date, is\_customer)

2. Contacts Table:

- Primary key: id

- Foreign key to companies table (company\_id)

- UNIQUE constraint on email to prevent duplicate contacts

- NOT NULL constraints on critical fields (email, first\_name, last\_name, status, created\_date, last\_modified)

3. Opportunities Table:

- Primary key: id

- Foreign keys to companies and contacts tables

- NOT NULL constraints on critical fields (name, company\_id, amount, stage, created\_date, is\_closed)

- CHECK constraint on probability to ensure it's between 0 and 100

4. Activities Table:

- Primary key: id

- Foreign keys to contacts and opportunities tables

- NOT NULL constraints on critical fields (contact\_id, type, subject, timestamp)

- CHECK constraint on duration\_minutes to ensure it's positive

5. Indexes:

- Created on frequently queried columns and foreign keys to improve query performance

6. Data Types:

- Used appropriate data types for each column (INTEGER, TEXT, REAL, DATE, DATETIME, BOOLEAN)

7. Normalization:

- Tables are in 3NF (Third Normal Form) to minimize data redundancy and dependency

**Documentation of cleaning rules and assumptions**

* Standardization of company name
* Check for duplicates
* Normalization of phone numbers
* \*Fill missing values in companies DataFrame.If null category goes unknown , if int goes 0-> to confirm with stakeholder
* If important fields are null, remove row

**Sample queries demonstrating data access**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated