Contact Data Standardization Database Design (contact_db)

1. Purpose and Overview

The contact_db database is designed to store raw contact data (e.g., from an exported list) and, critically, to **normalize and enrich the associated company data**.

The primary challenge this design addresses is the inconsistencies in company names within the raw contact list (e.g., "Google Inc.", "Goggle", "Google, LLC"). The database structure ensures that every raw company name is mapped to a single, standardized, and manually managed "clean" company profile, which can then be enriched with internal data (size, revenue, notes).

2. Entity Relationship Diagram (Conceptual)

The design implements a **Many-to-One** relationship:

1. **Many** raw company names (from the contacts table, via the company_name_mapping table) link to **One** clean company profile (companies table).
2. **Many** contacts link to **One** raw company name.
3. **Many** contacts link indirectly to **One** clean company profile.

The **company_name_mapping** table serves as the crucial **junction** for standardization, allowing multiple raw name variations to point to a single company_id.

3. Schema Details

Table: contacts (Source/Raw Contact Data)

This table stores the original, imported data, including the raw, potentially unstandardized company name.

Field Name	Data Type	Description
id	integer	**Primary Key** (Contact ID).
first_name	varchar(50)	Contact's first name.
last_name	varchar(50)	Contact's last name.
url	varchar(255)	URL associated with the

		contact (e.g., LinkedIn profile).
email_address	varchar(100)	Contact's email address.
company	varchar(100)	**Raw company name** from the source data (e.g., "Google Inc."). This is the field that needs cleaning.
position	varchar(100)	Contact's job title/position.
connected_on	date	Date the contact was made.

Table: company_name_mapping (Standardization Link)

This is the core mapping table, linking every unique raw company name to a single clean company profile ID. This table is populated by the **Raw Name Mapping Tool**.

Field Name	Data Type	Description
raw_name	varchar(100)	**Primary Key / Foreign Key (to contacts.company)**. The unique, unstandardized company name string.
company_id	integer	**Foreign Key (to companies.company_id)**. The ID of the clean, standardized company profile this raw name belongs to.

Table: companies (Clean, Managed Profiles)

This table holds the standardized company profiles and all enriched, manually managed data points. This data is managed via the **Company Management Tool**.

Field Name	Data Type	Description

company_id	integer	**Primary Key** (Clean Company ID). Auto-generated unique identifier.
company_name_clean	varchar(100)	**The standardized company name** (e.g., "Google").
target_interest	boolean	Flag indicating if this is a company of sales/marketing interest (TRUE/FALSE).
size_employees	integer	Estimated number of employees.
annual_revenue	numeric	The numeric part of the annual revenue (e.g., 5.2).
revenue_scale	varchar(10)	The scale of the revenue (e.g., 'M' for Million, 'B' for Billion).
headquarters	varchar(100)	Location of the company's headquarters.
notes	text	Free-form field for research, strategy, and qualitative notes.