

Resume Parsing Database Schema

candidate

Stores main profile-level information for a person/candidate.

| Column | Type | Description |
|------------------|---------------|--|
| candidate_id | UUID (PK) | Primary key, default gen_random_uuid(). |
| full_name | TEXT | Candidate's full name. |
| primary_email | CITEXT | Primary email (case-insensitive). |
| alt_emails | CITEXT[] | Alternate emails. |
| primary_phone | TEXT | Primary phone number. |
| location_city | TEXT | City. |
| location_state | TEXT | State. |
| location_country | TEXT | Country. |
| linkedin_url | TEXT | LinkedIn URL. |
| github_url | TEXT | GitHub URL. |
| portfolio_url | TEXT | Portfolio/personal site URL. |
| resume_text_hash | TEXT (UNIQUE) | Hash of resume text for deduplication. |
| source_file_id | TEXT | File/source identifier (FK logical link to resume_source). |
| parsed_at | TIMESTAMPTZ | Parsing completion timestamp (optional). |
| created_at | TIMESTAMPTZ | Row creation timestamp (default now()). |

resume_source

Stores metadata about uploaded resume files.

| Column | Type | Description |
|----------------|-------------|-----------------------------------|
| source_file_id | TEXT (PK) | Identifier for the uploaded file. |
| filename | TEXT | Original filename. |
| bytes | BIGINT | File size in bytes. |
| mime_type | TEXT | MIME type. |
| uploaded_at | TIMESTAMPTZ | Upload timestamp (default now()). |

experience

One row per employment role for a candidate.

| Column | Type | Description |
|-----------------|------------------------------------|--|
| exp_id | UUID (PK) | Primary key. |
| candidate_id | UUID (FK → candidate.candidate_id) | Candidate reference (ON DELETE CASCADE). |
| company_raw | TEXT | Company name as written in resume. |
| company_norm | TEXT | Normalized company name. |
| title_raw | TEXT | Title as written in resume. |
| title_norm | TEXT | Normalized title. |
| seniority | TEXT | Seniority level (e.g., Junior, Senior). |
| employment_type | TEXT | Full-time, Internship, Contract, etc. |
| location | TEXT | Location text. |
| start_date | DATE | Start date (nullable). |

| | | |
|-------------|-------------|--|
| end_date | DATE | End date (nullable). |
| is_current | BOOLEAN | If currently employed in this role. |
| exp_summary | TEXT | Optional summary paragraph for the role. |
| created_at | TIMESTAMPTZ | Row creation timestamp. |

experience_bullet

Bullet-level detail under an experience; each bullet is a searchable unit.

| Column | Type | Description |
|------------------|-------------------------------|--|
| bullet_id | UUID (PK) | Primary key. |
| exp_id | UUID (FK → experience.exp_id) | Parent experience (ON DELETE CASCADE). |
| bullet_text | TEXT | Bullet content. |
| skills_mentioned | TEXT[] | List of skill tokens found in the bullet (optional). |
| created_at | TIMESTAMPTZ | Row creation timestamp. |

education

Education history entries for a candidate.

| Column | Type | Description |
|--------------|------------------------------------|--|
| edu_id | UUID (PK) | Primary key. |
| candidate_id | UUID (FK → candidate.candidate_id) | Candidate reference (ON DELETE CASCADE). |
| degree_raw | TEXT | Degree name as written. |
| degree_level | TEXT | Normalized degree level (e.g., BE-BTech, MS, PhD). |

| | | |
|------------------|-------------|-------------------------------------|
| field_raw | TEXT | Field of study as written. |
| field_norm | TEXT | Normalized field of study. |
| institution_raw | TEXT | Institution as written. |
| institution_norm | TEXT | Normalized institution name. |
| start_date | DATE | Start date (nullable). |
| end_date | DATE | End date (nullable). |
| gpa | TEXT | GPA text (e.g., 8.7/10). |
| grading_scale | TEXT | Scale descriptor (e.g., /10, /4.0). |
| created_at | TIMESTAMPTZ | Row creation timestamp. |

skill

Master list of normalized skills.

| Column | Type | Description |
|-----------|---------------|--|
| skill_id | UUID (PK) | Primary key. |
| name_norm | TEXT (UNIQUE) | Normalized skill name (canonical). |
| category | TEXT | High-level category (language framework tooling cloud domain other). |

candidate_skill

Mapping between a candidate and skills; includes evidence links.

| Column | Type | Description |
|--------------|------------------------------------|--|
| candidate_id | UUID (FK → candidate.candidate_id) | Candidate reference (ON DELETE CASCADE). |

| | | |
|-------------------|----------------------------|--|
| skill_id | UUID (FK → skill.skill_id) | Skill reference (ON DELETE CASCADE). |
| proficiency_est | INT | Estimated proficiency level (1–5; optional). |
| evidence_count | INT | Number of distinct mentions/evidence points. |
| evidence_exp_ids | UUID[] | Experience IDs where this skill appears. |
| evidence_proj_ids | UUID[] | Project IDs where this skill appears. |
| PRIMARY KEY | (candidate_id, skill_id) | Composite PK. |

project

Project entries for a candidate (parent rows).

| Column | Type | Description |
|-----------------|------------------------------------|---|
| project_id | UUID (PK) | Primary key. |
| candidate_id | UUID (FK → candidate.candidate_id) | Candidate reference (ON DELETE CASCADE). |
| name | TEXT | Project name. |
| role | TEXT | Role on the project. |
| start_date | DATE | Start date (nullable). |
| end_date | DATE | End date (nullable). |
| repo_url | TEXT | Repository or project URL. |
| tech_stack | TEXT[] | Free-text stack (normalize later via skills). |
| project_summary | TEXT | Optional summary/overview. |
| created_at | TIMESTAMPTZ | Row creation timestamp. |

project_bullet

Bullet-level detail under a project; mirrors experience_bullet.

| Column | Type | Description |
|------------------|--------------------------------|--|
| proj_bullet_id | UUID (PK) | Primary key. |
| project_id | UUID (FK → project.project_id) | Parent project (ON DELETE CASCADE). |
| bullet_text | TEXT | Bullet content. |
| skills_mentioned | TEXT[] | List of skill tokens found in the bullet (optional). |
| created_at | TIMESTAMPTZ | Row creation timestamp. |

certification

Professional certifications and credentials.

| Column | Type | Description |
|---------------|------------------------------------|--|
| cert_id | UUID (PK) | Primary key. |
| candidate_id | UUID (FK → candidate.candidate_id) | Candidate reference (ON DELETE CASCADE). |
| name | TEXT | Certification name. |
| detailed_text | TEXT | Raw line/description from resume. |
| issue_date | DATE | Issue date (nullable). |
| credential_id | TEXT | Credential identifier (optional). |
| url | TEXT | Verification or info URL (optional). |
| created_at | TIMESTAMPTZ | Row creation timestamp. |

publication

Publications; keep it lean with optional fields.

| Column | Type | Description |
|--------------|------------------------------------|--|
| pub_id | UUID (PK) | Primary key. |
| candidate_id | UUID (FK → candidate.candidate_id) | Candidate reference (ON DELETE CASCADE). |
| title | TEXT | Publication title (required). |
| year | INT | Year (optional). |
| url | TEXT | URL (optional, e.g., arXiv/Scholar). |
| created_at | TIMESTAMPTZ | Row creation timestamp. |

achievement

Awards or notable achievements.

| Column | Type | Description |
|--------------|------------------------------------|--|
| ach_id | UUID (PK) | Primary key. |
| candidate_id | UUID (FK → candidate.candidate_id) | Candidate reference (ON DELETE CASCADE). |
| title | TEXT | Achievement title. |
| description | TEXT | Description/details (optional). |
| date | DATE | Date (optional). |
| created_at | TIMESTAMPTZ | Row creation timestamp. |

candidate_language

Spoken languages and optional proficiency level.

| Column | Type | Description |
|--------|------|-------------|
|--------|------|-------------|

| | | |
|--------------|---------------------------------------|--|
| candidate_id | UUID (FK → candidate.candidate_id) | Candidate reference (ON DELETE CASCADE). |
| language | TEXT | Language name. |
| level | TEXT | Proficiency (e.g., C1, Native, Professional). |
| PRIMARY KEY | (candidate_id, language) | Composite PK. |

raw_section_text

Raw text per resume section for vectorization (semantic search).

| Column | Type | Description |
|----------------|---------------------------------------|--|
| section_id | UUID (PK) | Primary key. |
| candidate_id | UUID (FK → candidate.candidate_id) | Candidate reference (ON DELETE CASCADE). |
| section_type | TEXT (CHECK) | One of: experience project education skills summary certification publication achievement language other. |
| text | TEXT | Verbatim text content for the section. |
| source_file_id | TEXT | Source file identifier (optional). |
| created_at | TIMESTAMPTZ | Row creation timestamp. |

Indexes

candidate

idx_candidate_email on candidate(primary_email)

idx_candidate_location on candidate(location_country, location_state, location_city)

experience

idx_experience_candidate on experience(candidate_id)

idx_experience_dates on experience(start_date, end_date, is_current)

idx_experience_company_title on experience(company_norm, title_norm)

experience_bullet

idx_experience_bullet_exp on experience_bullet(exp_id)

education

idx_education_candidate on education(candidate_id)

idx_education_level on education(degree_level)

project

idx_project_candidate on project(candidate_id)

project_bullet

idx_project_bullet_project on project_bullet(project_id)

raw_section_text

idx_raw_section_candidate on raw_section_text(candidate_id, section_type)