

Steps to Create and Populate Master Table in PostgreSQL

Step 1: Plan the Master Table Structure

- Identify key columns required for analysis (e.g., learner_id, opportunity_id, email).
- Define relationships: Primary Keys and Foreign Keys.
- Set appropriate data types and ensure data consistency.

Step 2: Extract Data from Source Tables

- Identify all source CSV files: Learner, Cohort, Cognito, Opportunity, LearnerOpportunity.
- Import data using COPY command into raw PostgreSQL tables.

Step 3: Transform Data for Consistency & Accuracy

- Clean data: lowercase emails, handle NULLs, remove duplicates.
- Example: Use INITCAP() to format stages, COALESCE() to fill NULL login_status.

Step 4: Create and Load Master Table

- Design the master table with relationships between learner, cohort, opportunity, cognito.
- Use JOINS to combine data and INSERT INTO master_table.
- Use DISTINCT ON or ON CONFLICT to handle duplicates.

Step 5: Sample SQL Snippet to Insert Clean Data

```
INSERT INTO master_table (...)  
SELECT DISTINCT ON (l.learner_id, lo.opportunity_id)  
    l.learner_id, l.name, l.email, lo.opportunity_id, INITCAP(o.stage), ...  
FROM learner_raw l  
JOIN learneropportunity_raw lo ON l.learner_id = lo.learner_id  
JOIN opportunity_raw o ON lo.opportunity_id = o.opportunity_id  
LEFT JOIN cohort c ON l.cohort_id = c.cohort_id  
LEFT JOIN cognito cg ON l.cognito_id = cg.cognito_id;
```