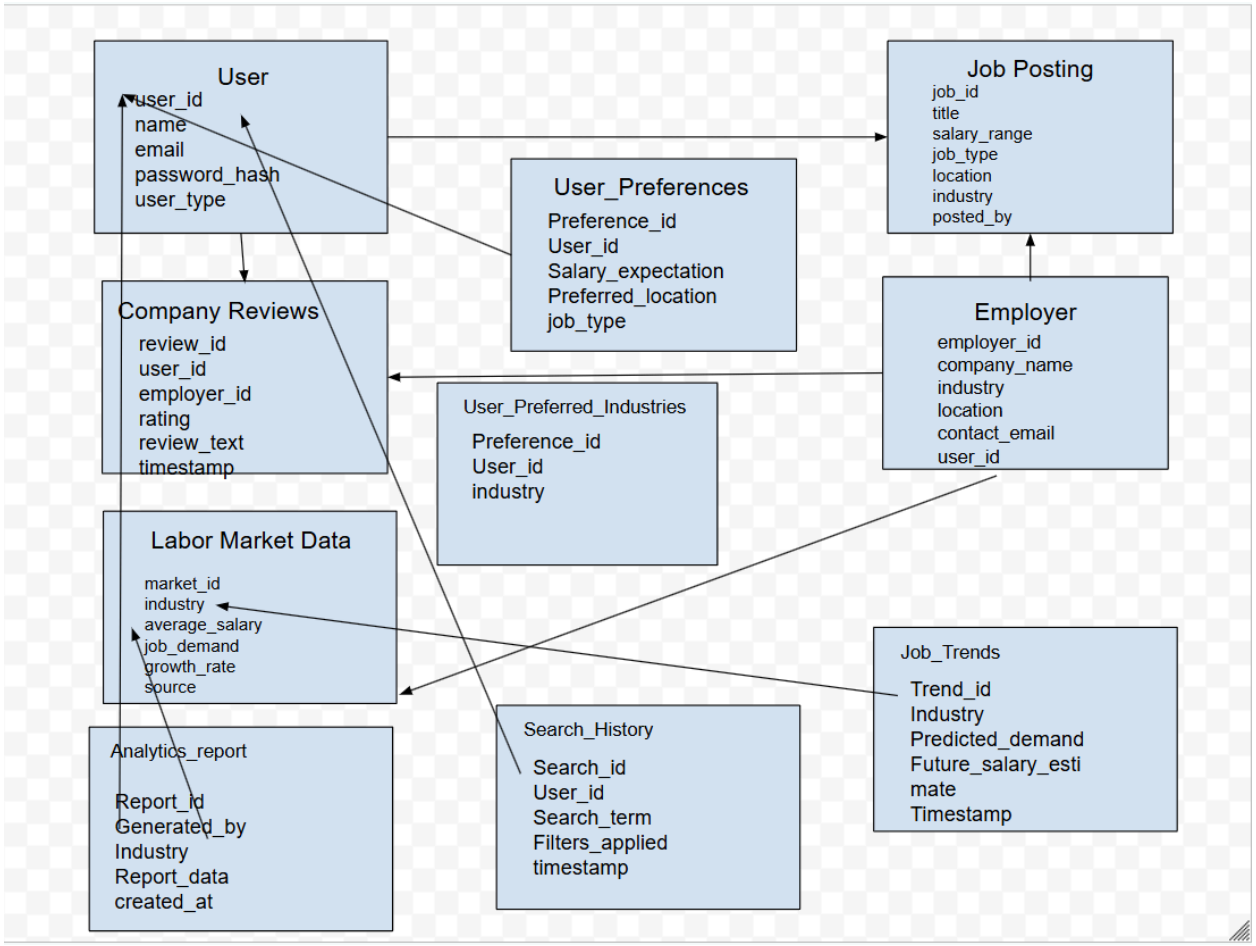
ERD



1. Entities & Attributes

Entities & Attributes

1. User & Account Management

Entities:

1. **User (Account)**
    - user_id (PK)

    - name
    - email
    - password_hash
    - user_type (Enum: Job Seeker, Employer, Admin)
2. **User Preferences**
    - preference_id (PK)
    - user_id (FK to User)
    - preferred_industries (Array)
    - salary_expectation
    - preferred_location
    - job_type (Full-time, Part-time, Remote, etc.)

Relationships:

- **User → User Preferences → One-to-One** (Each user has one set of job preferences)

2. Job Listings & Employers

Entities:

3. **Job Posting**
    - job_id (PK)
    - title
    - salary_range
    - job_type
    - location
    - industry (FK to Labor Market Data)
    - posted_by (FK to Employer)
4. **Employer**
    - employer_id (PK)
    - company_name
    - industry
    - location
    - contact_email
    - user_id (FK to User) → *For employer accounts*
5. **Company Reviews**
    - review_id (PK)
    - user_id (FK to User)
    - employer_id (FK to Employer)
    - rating (1-5)

- review_text
- timestamp

Relationships:

- **Employer → Job Posting → One-to-Many** (Each employer can post multiple jobs)
- **User → Company Reviews → One-to-Many** (A user can write multiple reviews)
- **Job Posting → Industry (Labor Market Data) → Many-to-One** (Jobs belong to an industry)

4. Labor Market Data & Trends

Entities:

8. **Labor Market Data**
   - market_id (PK)
   - industry
   - average_salary
   - job_demand (e.g., High, Medium, Low)
   - growth_rate
   - source (e.g., BLS, Glassdoor)
9. **Job Trends (Predictions)**
   - trend_id (PK)
   - industry (FK to Labor Market Data)
   - predicted_demand (Increase, Stable, Decline)
   - future_salary_estimate
   - timestamp

Relationships:

- **Labor Market Data → Industry Trends → One-to-Many** (Each industry has multiple predictions over time)
- **Job Posting → Labor Market Data → Many-to-One** (Each job belongs to one industry)

5. Search & Analytics

Entities:

10. **Search History**

- search_id (PK)
- user_id (FK to User)
- search_term
- filters_applied
- timestamp

11. **Analytics & Reports**

- report_id (PK)
- generated_by (FK to User)
- industry (FK to Labor Market Data)
- report_data (JSON or separate table)
- created_at

Relationships:

- **User → Search History → One-to-Many** (A user can have multiple searches)
- **User → Reports → One-to-Many** (Users can generate reports)
- **Reports → Labor Market Data → Many-to-One** (Reports are based on labor trends)

2. Assumptions of the UML Diagram

This section describes the assumptions made for each entity in our UML diagram, explaining why they were modeled as entities rather than attributes. Additionally, we provide an overview of the key attributes and relationships, including their cardinality.

User

Assumptions

- The User is an independent entity representing the system's users.
- Each User can write multiple Company Reviews, but each review belongs to only one User.
- The email field must be unique to ensure each user has a distinct identity.

Key Attributes

- user_id: INT [PK] - Unique identifier for each user.
- name: VARCHAR(50) - The user's name.
- email: VARCHAR(50) UNIQUE - A unique email address for user authentication.

- password_hash: VARCHAR(100) - Hashed password for security.
- user_type: ENUM('Job Seeker', 'Employer') - Defines the role of the user.

Relationships

- User and Company Reviews have a **1-N** relationship, meaning a user can write multiple reviews.
- User and Job Applications have a **1-N** relationship, meaning a user can apply to multiple jobs.

Employer

Assumptions

- The Employer represents a company posting jobs.
- An Employer is linked to a User entity to manage access control.
- Each employer can post multiple job listings.

Key Attributes

- employer_id: INT [PK] - Unique identifier for each employer.
- company_name: VARCHAR(100) - Name of the employer.
- industry: VARCHAR(50) - The industry in which the employer operates.
- location: VARCHAR(100) - Location of the employer.
- contact_email: VARCHAR(50) - Contact email for the employer.
- user_id: INT [FK to User.user_id] - Links to the User table.

Relationships

- Employer and Job Posting have a **1-N** relationship, meaning an employer can post multiple jobs.
- Employer and Company Reviews have a **1-N** relationship, meaning employers receive multiple reviews from users.

Job Posting

Assumptions

- A Job Posting represents an employment opportunity.
- Each job posting is created by one employer but can receive multiple applications.

Key Attributes

- job_id: INT [PK] - Unique identifier for each job.
- title: VARCHAR(100) - Job title.
- salary_range: DECIMAL(10,2) - Salary range for the job.
- job_type: ENUM('Full-time', 'Part-time', 'Remote') - Type of job.
- location: VARCHAR(100) - Job location.
- industry: VARCHAR(50) - Industry category.
- posted_by: INT [FK to Employer.employer_id] - Links the job to an employer.

Relationships

- Job Posting and Job Applications have a **1-N** relationship, meaning each job can receive multiple applications.
- Job Posting and Labor Market Data have a **M-N** relationship, meaning jobs belong to an industry labor dataset.

Company Reviews

Assumptions

- A Company Review allows users to leave feedback about employers.
- Each review is associated with exactly one employer and one user.

Key Attributes

- review_id: INT [PK] - Unique identifier for the review.
- user_id: INT [FK to User.user_id] - The user who wrote the review.
- employer_id: INT [FK to Employer.employer_id] - The employer being reviewed.
- rating: INT - Rating score (1-5).
- review_text: TEXT - The review content.
- timestamp: TIMESTAMP - When the review was written.

Relationships

- Company Reviews and Users have a **N-1** relationship.
- Company Reviews and Employers have a **N-1** relationship.

Labor Market Data

Assumptions

- Labor Market Data provides statistical information about industry trends.
- Jobs belong to specific industries for analysis purposes.

Key Attributes

- market_id: INT [PK] - Unique identifier.
- industry: VARCHAR(50) - Industry category.
- average_salary: DECIMAL(10,2) - Average salary in this industry.
- job_demand: ENUM('High', 'Medium', 'Low') - Demand level for jobs.
- growth_rate: DECIMAL(5,2) - Industry growth percentage.
- source: VARCHAR(100) - Source of the data.

Relationships

- Labor Market Data and Job Posting have a **M-N** relationship.

-

3. Database Normalization (BCNF/3NF)

Step 1: First Normal Form (1NF)

- No multi-valued attributes.
- preferred_industries moved into a separate table User Preferred Industries.
- This is done with a foreign key to User so that each preferred industry is stored as a separate row.

Step 2: Second Normal Form (2NF)

- Removed partial dependencies by moving user_name, user_email to the User table.\
- This reduces redundancies. Tables accessing user information now use user_id rather than repeating user_name and user_email.

Step 3: Third Normal Form (3NF)

- Removed transitive dependencies by moving employer_name to the Employer table.

Step 4: Boyce-Codd Normal Form (BCNF)

- Verified that all dependencies are on superkeys.

4. Final Relational Schema

User(user_id: INT [PK], name: VARCHAR(255), email: VARCHAR(255), password_hash: VARCHAR(255), user_type: ENUM)

User_Preferences(preference_id: INT [PK], user_id: INT [FK to User], salary_expectation: DECIMAL, preferred_location: VARCHAR(255), job_type: ENUM)

User_Preferred_Industries(preference_id: INT [PK], user_id: INT [FK to User], industry: VARCHAR(255))

Employer(employer_id: INT [PK], company_name: VARCHAR(255), industry: VARCHAR(255), location: VARCHAR(255), contact_email: VARCHAR(255), user_id: INT [FK to User])

Job_Posting(job_id: INT [PK], title: VARCHAR(255), salary_range: DECIMAL, job_type: ENUM, location: VARCHAR(255), industry: VARCHAR(255), posted_by: INT [FK to Employer])

Company_Reviews(review_id: INT [PK], user_id: INT [FK to User], employer_id: INT [FK to Employer], rating: INT, review_text: TEXT, timestamp: TIMESTAMP)

Labor_Market_Data(market_id: INT [PK], industry: VARCHAR(255), average_salary: DECIMAL, job_demand: ENUM, growth_rate: DECIMAL, source: VARCHAR(255))

Job_Trends(trend_id: INT [PK], industry: VARCHAR(255) [FK to Labor_Market_Data], predicted_demand: ENUM, future_salary_estimate: DECIMAL, timestamp: TIMESTAMP)

Search_History(search_id: INT [PK], user_id: INT [FK to User], search_term: VARCHAR(255), filters_applied: TEXT, timestamp: TIMESTAMP)

Analytics_Reports(report_id: INT [PK], generated_by: INT [FK to User], industry: VARCHAR(255) [FK to Labor_Market_Data], report_data: JSON, created_at: TIMESTAMP)