

Name: Renuka Huddar

NUID: 002750433

Data Management & Database Design

Notes on Jobs Database

- ❖ To create a database for people looking for jobs in finance using fields like location, salary, job type, etc. Using languages like SQL and python (data visualization)
- ❖ To extract data, create a Twitter bot for scrapping the web.
- ❖ The Bot retrieves job postings and inserts them into the jobs database.
- ❖ It also extracts keywords from Twitter and then adds them to the database, which makes it easy for users to look for jobs using keywords.
- ❖ JSON files remove duplicate data and store useful information in the database.
- ❖ Used CSV files to store scraped data.
- ❖ This database also shows the eligibility of job seekers and connects them with an appropriate employer.
- ❖ The jobs database also recommends job-based on the users' interest and postings from different job websites like indeed.com, Glassdoor, etc.
- ❖ This database has roles like Analyst roles, Artificial Intelligence Roles, Biotech roles, finance, education, etc.
- ❖ This also shows company ratings by glassdoor in a CSV file. The file has a review link and employee title.
- ❖ Includes companies like Morgan Stanley, Wells Fargo, Bank of England, Prodigy Finance, etc.
- ❖ Primary Key is used for Job ID in indeed table and jobs urls.
- ❖ In this database, normalization is done to reduce the redundancy of the data and implement real-life use cases.
- ❖ There is one source of the information on which web scrapping is done and web API is extracted.
- ❖ There are two tables, one is Indeed and the other is Twitter. These two tables create the database which has fields like location, salary, job role, etc.

- ❖ In SQL, views are created to avoid the complexity and security of the data.
- ❖ In SQL, joins are of 4 types: INNER JOIN, OUTER JOIN, RIGHT JOIN, and FULL JOIN. They are used to connect various tables.