Github: https://github.com/GSreeSwetha/FetchAssessment

I have used EC2 instance to create the ppk file

● How would you deploy this application in production?

1. Develop and design the workflow to meet prod requirements.

2. Test the workflow with sample data with production environment credentials to check permissions

3. Implement error handling, more security practices.

4. The data sources, destinations should be accessible and properly configured.

● What other components would you want to add to make this production ready?

We need to add a security layer for the database access, add constraints to the Postgres database.

● How can this application scale with a growing dataset.

Adding more memory, instances and using services like AWS can be helpful built-in scalability features. Tools like Alteryx, Talend are equipped to work with bigdata sets, read the output database and only push the new records.

● How can PII be recovered later on?

For this application the masking is done with SHA-256. The masked data can be verified by hashing the provided input and comparing it to the stored hash. SHA-256 algorithms are difficult to retrieve the original data. If we use MD5 algorithm, they can be decrypted with the key we used to encrypt the information. If we use a macro instead to encrypt the data, for suppose we have used a random number to multiply the records of an integer column, the analyst can decrypt the information using that number but ensuring multiple layers of authorization.

PII data can be stored separately.

● What are the assumptions you made?

The user has required tools installed and have access to the PostgreSQL database and have secured connection.