

Yang Ye

China, Guangzhou

Tel: +86 18520149491

Linkedin: <https://www.linkedin.com/in/yang-ye/>

Github: <https://github.com/dsuuny?tab=repositories>

Email: D244975154@gmail.com

Data Engineer (M.Sc) | ETL | Data analyze | Hadoop | Machine Learning | Spark | AWS | GCP | Python |

EDUCATION

University of Limerick (First Honor degree)

September 2022 – September 2023

MSc. Software engineering

Limerick, Ireland

North China University of Science and Technology

September 2014 - June 2018

BSc. Information management

Tangshan, China

TECHNICAL SKILLS

Languages: Advanced Python, TypeScript, Java, SQL, NOSQL, HTML5/CSS

Developer Tools: VS Code, MS Visual Studio, IntelliJ Idea, PyCharm, Tableau

Engineering Stack: MySQL, PostgreSQL, Machine Learning, Hive, Spark, HDFS, Data Warehouse, Elasticsearch, MongoDB, Fastapi, React

Others: Agile, Git, MPI, GCP/AWS

WORK EXPERIENCE

SYMBIO (HSBC Contractor)

February 2024 – July 2025

Full Stack Engineer

Guangzhou, China

Data Development Automation Platform:

Technology stack: Python, Google BigQuery, CI/CD, REST APIs, Confluence

- Architected and developed a Python-based automation platform to streamline the entire data development lifecycle.
- Engineered the tool to automatically generate standardized BigQuery JSON code, ensuring consistency and reducing manual errors.
- Implemented an automated pipeline for deploying to the DEV environment, executing tests, and publishing results directly to Confluence pages for improved transparency.
- Integrated with various internal APIs to automate pre-production release workflows, such as ticket creation and approval requests.

Key Impact: Achieved a significant reduction in the data delivery lifecycle, shortening the development phase from **1 week to 1 day** and the production release preparation from **3 days to 1 day**.

ERMS (ETL Requirement Management Service):

Technology stack: FastAPI, React, PostgreSQL, Docker

- Spearheaded the project as **Tech Lead**, guiding a team of 4 engineers through the entire project lifecycle from conception to ongoing development.
- **Led the development team using Agile methodologies** to facilitate iterative progress, manage tasks, and ensure timely delivery.
- Led key technical decisions, including technology stack selection, system architecture, and database schema design.
- Mentored and provided technical training to new team members, fostering their growth and ensuring code quality.

- Implemented core features including version control for requirements, dynamic approval workflows, data lineage visualization, and an AI assistant.

Key Impact: Optimized the requirement lifecycle by streamlining the design process for data modelers and accelerating development for engineers with AI-powered code generation.

CLPS (HSBC Contractor)

Data Engineer

February 2021 – May 2022

Guangzhou, China

Teradata – GCP, AWS & Alicloud Migration:

Technology stack: DataStage + SQL + Python + Control-M

- Designed and developed ETL process from Teradata to GCP, AWS and Alicloud.
- Conducted thorough testing and validation of the ETL process to ensure data integrity during the migration.
- Worked closely with stakeholders to define key performance indicators (KPIs)
- Analyzed existing data structures of Teradata and DataStage.
- Monitored daily batch progress of the migration process and handle batch issue.
- Collaborated with technical teams to implement the optimized solutions, which may have included adjusting data transfer methods, utilizing parallel processing, or optimizing SQL queries.

Xcheck Middleware:

Technology stack: SQL + Python + JinJa

- Developed a verification tool to resolve health check issues arising from discrepancies in source data or logic.
- Leveraged SQL queries to retrieve data from Teradata and BigQuery

HSBC Migration Assis:

Technology stack: Python

- Converted DDL and SQL to be compatible to new environment.
- Assisted users in migrating their SQL queries and scripts from the original Teradata environment to the new platform.
- Developed a multi-threaded tool for efficient table data transfers.

Meiya Pico

ETL Engineer

November 2018 – February 2020

Guangzhou, China

Data standardization annotation system:

Technology stack: SQL + FastAPI + Pandas + scikit-learn

- Designed a text classification model for data annotation.
- Developed a RESTFUL API endpoint using FastAPI to expose the functionality of the text classification model to external systems.

DATA extension:

Technology stack: Java

- Designed and developed ETL extensions tailored to each database's unique data structures, schemas, and data integration needs.
- Developed data cleaning functions for structured and unstructured databases.

E-commerce customer behaviors analysis:**Technology stack:** Python + Pandas + Numpy + POWERBI + SQL

- Understanding Taobao buyers' behavior, including purchase preferences, shopping cart analysis, purchase cycles, and the purchase decision-making process.
- Using Algorithm run analysis on the customer purchase record data from Oracle/mysql/PostgreSQL to predict the potential customer

Merchandise Sales prediction:**Technology stack:** Python + Pandas + Numpy + POWERBI

- Perform exploratory data analysis (EDA) to uncover correlations and dependencies between various factors, such as product categories, pricing, promotions, and sales volumes.
- Analyze the revenue from the sale of goods to gain insights into historical sales patterns, trends, and seasonality.
- Based on the analysis results, sellers optimize pricing, inventory management, promotional activities, and product positioning to maximize customer satisfaction and enhance sales efficiency.

PERSONAL PROJECT EXPERIENCE

Immune cell profiling from single-cell RNA with Neo4J

Github: <https://github.com/MeghanaKshirsagar/ISCBTutorial>

- Constructed Seurat pipeline to perform GEO single data Processing.
- Analyzed the gastric dataset by dimensional reduction method and clustering method.
- Leveraged different annotation tools (SingleR, scMRMA etc.) to label the cell types.
- Developed ETL process from Csv files into Neo4j database and wrote Cypher script to validate key biomarker.

CERTIFICATE

- Oracle Certified Professional
- AWS Certified Developer - Associate
- IELTS 6.0

SKILLS

- Language: Cantonese (Native), Mandarin (Native), English (Fluent)
- Microsoft: Word, Excel, PowerPoint
- Photography and video editing