

Alex Tjin . Jarvis Consulting

As a dedicated Data Engineer, I bring over four years of comprehensive IT development experience to the table. My expertise primarily revolves around data ingestion and integration, where I excel in leveraging Python, SQL, Spark and Azure to seamlessly handle data workflows. I possess extensive knowledge of diverse databases, including Redshift, Oracle, and Postgres, allowing me to efficiently manage and manipulate data at scale. Furthermore, I have honed my skills in constructing robust Data Pipelines utilizing cutting-edge ETL tools like Informatica and Airflow. These tools enable me to orchestrate and automate complex data processes, ensuring streamlined data integration and efficient information flow. Also, I have acquired valuable exposure and hands-on experience with the three major cloud service providers AWS, Azure, and GCP, which enable me to design and implement scalable and performant data solutions that leverage the power and versatility of these platforms.

Skills

Proficient: SQL, Python, Spark, Airflow, Databricks, Azure Data Factory, AWS, Azure Cloud Services, Data Modelling, Git, Docker

Competent: Javascript, Typescript, Node, Angular, REST API, Bash

Familiar: VBA, Data Visualisation, Regression Testing, Hadoop, GCP

Jarvis Projects

Project source code: https://github.com/Jarvis-Consulting-Group/jarvis_data_eng-alextjin

Cluster Monitor [GitHub]: Reduced the workload of the Linux Cluster Administration (LCA) team by developing several scripts that track instance performance. I used Docker in implementation and testing to ensure consistent execution. The 'psql_docker.sh' script facilitates container creation and activation. 'host_info.sh' and 'host_usage.sh' extract platform details, with the latter scheduled using crontab for continuous data collection, minimizing manual intervention. All data is stored in a PSQL instance for future analytics. Git was utilized as the version control tool, with all project scripts accessible on GitHub for collaboration and transparency.

Highlighted Projects

XML to CSV conversion: I built a reusable scripts to handle XML to CSV conversion. lxml is adopted for parsing the data. The script can support parsing the data from different root and layer. All the parent and child data would be captured.

IBOR (Investment Book of Record) Data Package: I was responsible to build data pipelines for the middle office book keeping. We retrieved data from our vendor via APIs request. I utilised AWS s3, Airflow and Spark for data ingestion, transformation and loading phases. Adopted Great Expectation for data quality checking.

RISE: I played a significant role in an on-prem-to-cloud project that involved integrating new data sources. I actively contributed by establishing connections with new data sources and constructing the project-specific database. This involved creating over 650 tables to facilitate efficient data storage and retrieval. Additionally, I designed and developed more than 20 ETL workflows using Informatica, ensuring smooth data transformations and transfers. To optimize operations, I leveraged Airflow and worked on three DAGs (Directed Acyclic Graphs) for workflow orchestration. Furthermore, I implemented bash scripts to automate tasks and monitor the performance of these workflows, enabling streamlined operations and efficient performance monitoring.

Library System Management App: I built a library system management app designed to manage all the functions of a library. Users can do all the CRUD operations for books. The frontend is designed with Angular and the backend is held by Express Web Framework. I also adopted Angular Material, MySQL to enhance the web functionality and data storage. Unit test has been conducted with Jest.

Professional Experiences

Full Stack Developer, RBC (2023-present): Played a pivotal role contributing to both Full Stack Development and Data Engineering initiatives. Being full responsible for comprehensive Data Pipeline development, from Data Modeling to report generation in Power BI, utilizing Python, Spark, and Airflow. Also, contributed in the migration from on-premises to cloud with Azure cloud services like ADF and Databricks.

Full Stack Developer, Jarvis (2023-present): Engaged in various projects and bolstered my skills and proficiency in data engineering, encompassing the area of SQL, Python, Spark and Linux.

Data Engineer, IBM (2021-2023): Leveraged tools like Informatica, Apache Airflow, and AWS services to architect scalable and efficient data integration workflows, encompassed ingesting data from various sources (e.g csv, json, xml), employing Python in data cleansing and transformation, automated routine tasks in Bash. Also, maintained data warehouse integrity and swiftly rectify defects. Built report with IBM Cognos.

Education

City University of Hong Kong (2016-2020), Bachelor of Business Administration, Department of Management - Dean's List 2019 - 2020 - Joseph Lau Student Exchange Awards - Hong Kong Hua-Yan Buddhist Association Exchange Scholarships - GPA: 3.54/4.3

Miscellaneous

- Microsoft Certified: Azure Data Fundamentals
- IBM Agile Explorer
- Basketball player
- Hodophile