

# Jack Cheng

314-441-1406 | jqcheng9413@yahoo.com | linkedin.com/in/jiaqicheng9413/ | github.com/jackson9413

## Experience

**Data Warehouse Developer**, BCITO – Wellington, New Zealand Mar 2022 – Present

- Managed data for 30,000+ students, 15,000+ employers, and 300+ teachers in 7+ systems
- Optimized ETL processes with Azure Data Factory, SQL, Python, Bash improving data processing time by 50%
- Developed Power BI dashboards and Python reports leading to a 40% increase in data-driven decision-making
- Designed and maintained star and snowflake data schemas improving query performance and storage by 35%
- Implemented CI/CD pipelines using Azure DevOps, Git, and Github and collaborated in an agile environment reducing deployment and delivery time by 30%

**Data Analytics and Wrangling Analyst**, Equifax – Auckland, New Zealand Sep 2021 – Mar 2022

- Visualized data volume, trends, and validation results and provided insights by designing Tableau dashboards resulting in a 25% improvement in data comprehension and decision-making efficiency
- Developed Python, SQL, and Bash scripts for ETL processes reducing data validation time by 25%
- Managed over 3 million customer data and migrated data and processes to GCP cutting manual efforts by 40%

**Data Engineer**, Health Quality and Safety Commission – Wellington, New Zealand Aug 2020 – Sep 2021

- Developed ETL scripts by using Python, SQL, R, and Bash increasing data processing efficiency by 35%
- Performed statistical and distribution analysis, deployed machine learning and deep learning models, utilized GIS tools to improve data reliability, accuracy, and processing efficiency by 20%, 30%, and 25% respectively
- Visualized data using Excel, Power BI, R Shiny, Python Dash leading to a 20% reduction in data misinterpretation and improving stakeholder engagement and comprehension of data by 30%

## Projects

**Data Engineering Capstone Project** [github.com/jackson9413/data-engineering-capstone-project](https://github.com/jackson9413/data-engineering-capstone-project)

- Developed STAR schema to create different dimension and fact tables based on the data sourced publicly
- Utilized on-premise tools like Python, PySpark, and SQL to develop ETL pipelines
- Used AWS S3 to store all the staging data and built a data warehouse by using AWS Redshift to store staging tables and normalized tables

## Skills

**Languages:** Python, R, SQL, HTML/CSS, JavaScript, PySpark, Bash

**Others:** Power BI, Tableau, Azure, AWS, GCP, Databricks

## Certificates

Microsoft Certified: Power BI Data Analyst Associate; Microsoft Certified: Azure Data Engineer Associate

Microsoft Certified: Azure AI Engineer Associate; Microsoft Certified: Azure Data Scientist Associate

Databricks Certified Data Engineer Associate

## Education

**Georgia Institute of Technology**, MS in Analytics, Online Aug 2023 – Present

**Udacity**, Nanodegree Data Engineering, Online Feb 2022 – May 2022

**University of Canterbury**, MS in Applied Data Science, New Zealand Jun 2019 – Jun 2020

**Jilin University**, BS in Science, China Sep 2013 – Jun 2017