

# Ziyu Guo

## Data Engineer

### Summary

### Contact

+1 513-202-6219

ziyu.g@myittechjobs.com

CA

### Education

#### Master of Data Science

University of the Pacific, Stockton,  
CA

#### Bachelor in Chemistry

Lanzhou University, Lanzhou,  
China

### Skills

#### Methodology

SDLC, Agile, Waterfall

#### Programming Language:

R, Python, SQL

#### IDE's:

PyCharm, Jupyter Notebook

#### Big Data Ecosystem:

Hadoop, MapReduce, Hive,  
Apache Spark, Pig

#### ETL Tools:

SSIS

#### Cloud Technologies:

AWS, Azure

#### Packages:

NumPy, Pandas, Matplotlib, SciPy,  
Scikit-learn, Seaborn, TensorFlow

#### Reporting Tools:

Tableau, Power BI, SSRS

#### Database:

MongoDB, MySQL

#### Other Tools:

Git, MS Office

#### Operating Systems:

Windows, Linux

- 3+ years of experience in Data Engineering, Data Analytics, Business Intelligence and Software Development using Open Source, Microsoft (Azure) and Amazon Webservice (AWS) technologies.
- Well-Versed in complete Software Development Life Cycle (SDLC) for the projects using methodologies like Agile and Waterfall.
- Proficient in using Python for Data Engineering and Modeling.
- Ability to Installing, Maintaining and Configuring Hadoop Cluster.
- Proficiency in developing SSIS Packages to extract, transform and load (ETL) data into data warehouse from heterogeneous sources.
- Understanding of cloud platforms like Amazon Web Services, Azure, Databricks (both on Azure as well as AWS integration of Databricks).
- Good knowledge of exploring and processing data science project data using NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow.
- Proficient in development of various reports, dashboards using various Tableau Visualizations.
- Well-Versed in handling, configuration and administration of databases like MySQL databases like MongoDB.
- Proficiency in using Version Control systems like Git.

### Experience

#### MetLife, CA | Sep 2021 – Current

##### Data Engineer

- Implement Agile Methodology for building an internal application.
- Develop fuzzy-text matching program in Python to identify and eliminate redundancy on county instruments.
- Create automate feeds of data using both SSIS and Perl to provide the business with the critical data.
- Implement Generic writable to incorporate multiple data sources into reducer to implement recommendation-based reports using MapReduce programs.
- Design ETL architecture and architecture documents and mapping documents.
- Migrate the Data using Azure database Migration Service (AMS).
- Design, develop, deploy and maintain large scale Tableau dashboards for Product Insight, Devices and Networking, and Cox Premise Equipment.
- Implement End to End solution for hosting the web application on AWS cloud with integration.
- Work on analyzing and examining customer behavioral data using MongoDB.

#### EliteChamp Software Solutions, India (Remote Work) | April 2019 - July 2021

##### Data Engineer

- Used Waterfall life cycle methodology for Iterative development and Rapid delivery of the product.
- Used R machine learning package predicted performance of certain samples.
- Developed various Python scripts to find vulnerabilities with SQL Queries by doing SQL injection, permission checks and performance analysis.
- Used project configuration, parameters, project connections, integration service catalog and environment to perform SSIS configuration.
- Implemented MapReduce programs to perform joins using secondary sorting and distributed cache.
- Used AWS to transform and move large amounts of data into and out of other AWS data stores and databases.
- Orchestrated all Data pipelines using Azure Data Factory and built a custom alerts platform for monitoring.