

Simona Zhang

simonazhangzy@gmail.com | [LinkedIn](#) | [Github](#) | +1 617-314-5190

SUMMARY

An experienced **Data Analyst** and **ETL Developer** with over 3-year experience. A certified **Spark Developer** by Databricks with extensive experience in Spark application development, big data transformation, AWS Cloud services. Skilled in data quality validation, data ingestion, ETL pipeline development, and visualization using SQL, Python, Scala, Spark and Tableau.

SKILLS

- **Languages:** Python, Scala, Java, SQL
- **Tools:** AWS(S3, EMR, Athena, Redshift), Spark, Airflow, JIRA, PostgreSQL, MySQL, Tableau, Excel, Git, Linux
- **Professionals:** Google Data Analytics Certificate, AWS Cloud Practitioner, Data Analyst with Datacamp, Spark Developer by Databricks

WORK EXPERIENCE

Data Analyst/ETL Developer at Multiplan company | Boston, MA March 2022 - Present

- Automated validation of raw data in various formats from different insurance carriers by writing a **Spark** job in Scala;
- Maintaining the **ETL** pipelines with Spark API, transforming and loading 10TB+ data to HDFS of Hive and AWS Redshift;
- Fixing data issues by solving bugs in ETL code, ensuring the delivery of correct data to support the work of business analyst;
- Performing exploratory data analysis by writing complex **SQL** queries or ad-hoc querying to supporting business need;
- Employed **Python**, Flask to develop automation tools to boost the efficiency of daily work.

Research Assistant at Old Dominion University | Norfolk, VA Sep 2020 - Dec 2021

- Trained and deployed end-to-end image&CNN-based self-driving algorithms on cars in 3D simulators and physical robots.
- Designed simulated transportation system to collect various driving behavior data of autonomous cars to Mysql database.
- Transformed the deep JSON structure data of transportation systems to non-relational data format for advanced modeling.
- Defined and analyzed indicators to quantify the safety hazards of attacked autonomous cars(best presentation reward).

Game Data Analyst at Tap4Fun | Chengdu, China May 2017 – Dec 2018

- Collected users' feedback using customized **ETL** pipelines from questionnaires and APIs of game communities;
- Worked with **MySQL database** to query large amount of data and analyze patterns to stimulate user's payments;
- Designed **A/B tests** to evaluate the performance of different operation activities, interface layouts.
- Designed and optimized **models** to do player segmentation, life-time value forecasting, and bug detection.

RELATED PROJECTS

Dynamic dashboard development for real-time data monitoring | Python, SQL, Flask, JavaScript May 2023 – July 2023

- Developed data extraction, transformation pipelines using SQL and Python to ensure accurate and up-to-date data;
- Implemented interactive dashboard utilizing Python Flask framework, JavaScript, enabling real-time data analysis.

Scalable Data Validation with Spark| Scala, SQL, bash Jan 2023 – May 2023

- Created a structured data validation logic in Scala, enabling the classification of data as either passing or failing validation;
- Orchestrated the integration of multiple data formats from Amazon S3, enabling validation operations across various file types.
- Generated readable reports for each team member's reference, providing concise insights into the validation results.

Data Warehouse migration to AWS Jan 2023 – May 2023

- Developed code patches with Scala to enrich data and fix data issues caused by migration, facilitating a seamless transition;
- Conducted extensive testing and validation to verify the accuracy and completeness of the migrated data.

EDUCATION

Old Dominion University Norfolk, VA, USA
Master in Modeling and Simulation Engineering Sep 2020 - Dec 2021

Northeastern University Boston, MA, USA
Master in Data Analytics Jan 2019 - Sep 2020

University of Electronic Science and Technology of China Chengdu, China
Bachelor in Finance Sep 2013 - June 2017