

## RUI(Rachel) DENG

2901 S Sepulveda Blvd Apt 338, Los Angeles, CA, 434-987-0466, [racheldeng1995@gmail.com](mailto:racheldeng1995@gmail.com)

### SUMMARY

- Expert in **Data Engineering, ETL, Data Warehouse, Data Modelling, Machine Learning, Statistical Modeling, and Data Visualization** with over 2 years' experience.
- Expert in SQL with **Queries, DDL, DML, DQL, DCL, TCL, Stored Procedures, and UserDefined Functions.**
- Expert in **Python** (3.x) and its libraries including **NumPy, Pandas, Scipy, Scikit-learn, Keras, NLTK, etc.**
- Experienced in creating **Data Pipelines, ETL tools with Airflow.**
- Experienced in utilizing cloud services **Amazon Web Services (AWS)** including **Aurora, Athena, EC2, EMR, and S3.**
- Experienced in utilizing **Spark, Presto** with Big Data.
- Excellent in Dashboard Report analysis and implementation by **Tableau and Domo.**
- Experienced in **CI/CD pipeline** to develop Python Script to convert file format.
- Proficient in Machine Learning Algorithms including **Regression, Classification, NLP, deep learning.**
- Experienced in **Time Series Analysis** using **AR, MA, ARIMA, GARCH, and ARCH Models.**

### PROFESSIONAL EXPERIENCE

#### **Sony Interactive Entertainment (PlayStation)**

##### **Data Engineer**

##### **Responsibilities:**

**Los Angeles, CA**

**05/2019 - Present**

- Created, refactored and daily maintained, optimized data pipelines on Airflow.
- Structured data pipelines for separating data source platforms, such as Mysql database, Oracle database, as a result extracting data from various data sources, providing promotion-related data, QA-related data and partner-related data for Operational Reporting tool and other 14 downstream consumers.
- Upgraded pipelines function from Athena-oriented to Spark-oriented, as a result processing data pipelines in parallel and scale as needed serving critical business functions such as provisioning content pipeline data to 15+ consumer groups.
- Updated and published new features or entities in SIE-wide data model and data lake, because of this, 15 downstream consumers are able to access organized and clear data structure and content.
- Cooperated with the Data Ocean team, participated in workshop to discuss big data architecture and partitioning best practices as well as several ways to improve self-service and service provisioning across SIE users. As a result the DO team have landed several changes to their infrastructure in order to support arrays, maps and structs as well as partition overwrites which they previously did not support.
- Updated our EMR cluster(supported Presto)to ensure our EMR infrastructure is compatible with data ocean infrastructure. Refactored the structure of entities that need to be delivered to data ocean, and upgraded pipelines as necessary. As a result, successfully publish and deliver data to the data ocean.
- Supported data migration team, utilized machine learning(natural language processing) to analyze and detect data quality, as a result cleansed hundreds of thousands of product descriptions (removing compatibility notices/modular texts/copyrights), ensuring customers would see the correct content without any duplicate product information on console.
- Provided source data and created dashboards on Tableau and Domo based on business requirements such as partner data, promotion data and FQA data, which has more than 20 users are analyzing data from the dashboards based on this data, as a result showing the data in a timely manner for SIE-wide users to apply enterprise-wide analytics and reporting.

##### **Technology Stack:**

Airflow, AWS Athena, AWS EMR, AWS EC2, AWS S3, Spark, Presto, MySQL, Python 3.x, Machine Learning Algorithms(Natural Language Processing), Tableau, DOMO

#### **Itlize Global LLC**

##### **Data Scientist**

##### **Responsibilities:**

**Piscataway, NJ**

**02/2019 - 05/2019**

- Used SSIS to create ETL packages to Validate, Extract, Transform, and Load data into Data Warehouse and Data Mart.
- Maintained and developed complex SQL queries, stored procedures, views, functions, and reports that meet customer requirements using Microsoft SQL Server 2008.
- Created Views and Table-valued Functions, Common Table Expression (CTE), joins, and complex subqueries to provide the reporting solutions.
- Optimized the performance of queries with modification in T-SQL queries, removed the unnecessary columns and redundant data, normalized tables, established joins, and created index.
- Created SSIS packages using Derived Columns, Condition Split, Aggregate, Execute SQL Task, Data Flow Task, and Execute Package Task.
- Migrated data from SAS environment to SQL Server 2008 via SQL Integration Services (SSIS).
- Developed worksheets and interactive data visualization dashboards with heat map graphs, drill down, and drill through reports.

**Technology Stack:** MS SQL Server 2014, Tableau 9.0, Tableau Server, and MS Excel.

### EDUCATION

Master of Statistics – University of Virginia, Charlottesville.

Bachelor of Economics and Bachelor of Business – Nankai University, Tianjin, China