

Erika Jacobs

ErikaAshley3@gmail.com • (480) 529-1864 • www.erika-jacobs.com

Work Experience

Nationwide Insurance Company, September 2020-Present

Scottsdale, AZ

Data Engineer

- Design, develop, and maintain scalable ETL pipeline for telematics data quality initiatives using Spark SQL and S3 on TBs of data. Extract car signals from thousands of JSON/CSV files to create telematics trip stakeholder output.
- Conduct exploratory data quality analysis for big data from multiple IoT vendors, and identify data quality issues in existing architecture. Work with internal customers to investigate data quality issues and provide solutions.
- Identified substantial error in production machine learning model's scoring code that went undetected for more than a year. Improved model runtime by 30%, and improved efficiency of code by developing loops and UDFs.
- Automate and support manual scoring process for driven trips as part of Nationwide's Smartride program using Spark SQL.

Nationwide Insurance Company, January 2017-September 2020

Scottsdale, AZ

Business Analyst

- Identified opportunity to develop ETL scripts to extract data and remove macros from hundreds of Excel documents. Cleaned data, changed variable data types, and built data mart in SQL Server. Opened use cases for integration organization-wide.
- Managed monthly ETL pipelines for data production, and managed prioritization amongst responsibilities.
- Automated consolidation and merging of datasets, and verified accuracy of processed data sources for reporting.
- Conducted ad-hoc quantitative analyses, built exhibits, and summarized multiple metrics from large and complex datasets using SQL queries.

Nationwide Insurance Company, December 2015-January 2017

Scottsdale, AZ

Senior Analyst

- Cleaned and manipulated datasets for analysis and verified accuracy of all data.
- Built SQL tables and views for analysis. Wrote queries to consolidate and combine complex datasets.
- Generated dashboards and visualizations, with output including tables, metrics, charts and graphs.
- Maintained positive working relationships with leadership clients across different departments.

Project Experience

[NASA Data Warehouse](#)

Python (Boto3, Psycopg2, Requests, Pandas), AWS (Redshift, S3)

- Reads unstructured weather data from the NASA space weather API. Cleans and structures data into multiple data frames. Creates security role and opens port to connect to Redshift. Stages data in S3 to load into Redshift.

[COVID-19 Dashboard](#)

Python (PySpark, Boto3, S3FS, Dash), AWS (S3), DataBricks

- Designed data pipeline that ingested COVID-19 data from GitHub, transformed data, and aggregated results using Spark SQL. Exported and staged data in S3. Developed and deployed dashboard of case trends per country and state on Heroku using HTML and CSS.

Skills

- **Languages:** Python, SQL, Scala, R, HTML, CSS
- **Databases:** SQL Server, MySQL, PostgreSQL, Redshift, Athena, RDS, DynamoDB, MongoDB
- **Platforms:** AWS (S3, EC2, IAM, CLI, EMR, Glue), Linux, Bash
- **Frameworks:** DataBricks, Spark, Kafka, Airflow, Hadoop (HDFS, MapReduce, Hive)
- **Tools:** GitHub, R Studio, Access, Excel, Tableau, Visual Studio (SSIS)
- **Application:** data modeling, cloud architecture, data mining, statistics, machine learning, agile/scrum

Education

Penn State University

University Park, PA

Masters in Applied Statistics - 3.76 GPA

Arizona State University

Tempe, AZ

Bachelor of Arts in Psychology, Theatre

Activities and Awards

- Mathematics Professor through the Maricopa Community Colleges 2019-2020
- Toastmasters Competent Communicator and Competent Leader Awards