

# Erika Jacobs

## Work Experience

### **Data Engineer - Square**

**May 2021 – Present**

- Optimized critical ETL through applying incremental load. Reduced daily batch runtime by 40%, and reduced overall cost by 16%. Reduced number of lines of code for separate ETL by 66%. Conducted data quality analysis on pipelines and fixed critical errors in data quality such as duplication and incorrect joins.
- Developed API call to insert data into Snowflake. Developed infrastructure to automate Python code using Terraform and Docker.
- Enhanced data model of domain by ensuring accurate atomic grain of each table and removal of redundancy. Removed over 97% of unnecessary rows from large table in data model. Did analysis on appropriate table clustering and made judgements on normalization. Added new table to model with previously unavailable data. Enhanced data model reduced runtime of KPI dashboard by 50%.
- Consulted on data cost reduction task force and reduced data warehousing costs by at least \$3,000 per month.

### **Data Engineer - Nationwide**

**September 2020 - April 2021**

- Designed, developed, and maintained scalable ETL pipeline for telematics data quality initiatives on TBs of data using Spark SQL and S3. Extracted car signals from thousands of JSON/CSV files to create telematics trip stakeholder output.
- Conducted exploratory data quality analysis on big data from multiple IoT vendors, and identified data quality issues in architecture. Worked with internal customers to investigate quality issues and provided 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.
- Automated and supported manual scoring process for driven trips using Spark SQL.

### **Business Analyst - Nationwide**

**January 2017 - September 2020**

- 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. Cleaned and manipulated datasets for analysis and verified accuracy of all data.

### **Senior Analyst - Nationwide**

**December 2015 - January 2017**

- 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.

## Skills

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

## Education

### **Penn State University**

*Masters in Applied Statistics*

**University Park, PA**

### **Arizona State University**

*Bachelor of Arts in Psychology, Theatre*

**Tempe, AZ**

## Activities and Awards

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