# Jason Shiverick

## Lead Data Scientist and Data Engineer

[jason.shiverick@gmail.com](mailto:jason.shiverick@gmail.com)  
[linkedin](https://www.linkedin.com/in/jason-shiverick-2a9b7b16)  
415-849-5589

### Tech Stack

* Code
* Python, Spark, SQL / NoSQL, Scala, Go
* DevOps
* Git, Docker, Ansible, Airflow, NGINX, Terraform
* AWS
* EC2, S3, EMR, Glue, Athena, RDS, Lambda, Kinesis
* Hadoop
* HDFS, Hive, Impala, Presto
* Stats and ML
* pandas, scipy.stats, numpy, sklearn, lifelines, pymc3, MLlib

### Experience

**Udemy** *Senior Software Engineer, Data Platform* **June 2019 to Present**

* Developed pipeline to process mysql binlogs through kafka into a scala spark streaming application that provides distributed updates and deletes in the apache hudi format on AWS s3.
* Used ansible and terraform to build and maintain production data infrastructure in AWS including multiple AWS EMR clusters that ran up to 880 cores and 7 TB of memory.
* Rotated in on-call schedule to ensure that infrastructure was reliable and ~1K jobs passed on a daily basis.

**Waymo** *Senior Data Scientist* **June 2018 to June 2019**

* Designed and developed the python run feature extractor pipeline. Built on apache beam, flume, borg, python and protobuff, the framework provides users a simple design pattern to extract features *[counts, histograms, rainflow matrix]*, from vehicle and sds log data, then it easily scales over the whole fleet processing 100's of TB in minutes.
* Developed Metropolis Hastings algorithm in Go for sampling from a Weibull posterior with arbitrary priors.
* Developed Monte Carlo simulation in python for accurately forecasting field failures in complex systems.
* Developed Reliability analytics data pipeline and dashboards for report automation.

**Mayfield Robotics** *Data Engineer, Consultant* **March 2018 to June 2018**

* Data Warehouse: Designed and implemented analytics data infrastructure using spark via AWS Glue to process robot logs and disparate data sources into AWS Athena optimized parquet files on S3.

**Tesla** *Senior Data Scientist, Manager* **2013 to 2018**

* Developed a modern approach to advanced warranty simulation in Python that can account for competing failure modes in a repairable system under varying use conditions.
* Developed statistical framework for python: *Weibull analysis, Statistical era finding, Stress-Strength Convolution, Hypothesis testing, Best fit solver, ranking methods, mttf*
* Provided direction on proactive maintenance campaigns and developed prognostics algorithms: *physics of failure, logistic regression.*
* Built and Maintained robust back end infrastructure on top of Docker and Ansible. My design made it easy to provision and manage a Spark cluster and various micro-services between two people while also providing statistical models and TB scale log analytics.
* Established an analytics workflow leveraging git version control, with jira integrations. Designed the ETL workflow using spark, airflow, jupyter and superset.
* Established an extensive code base that provides tools to the organization for extracting, transforming, and analyzing field data at scale.

**Ingersoll Rand** *Reliability Engineer* **2011 to 2013**

**Medtronic INC.** *Product Performance Specialist* **2010 to 2011**

**Boeing Corporation** *Systems Engineer* **2008 to 2009**

### Education

**Graduate Course Work (Reliability Engineering)** **2012 to 2013** University of Maryland (online) College Park, Maryland *ENRE 602: Reliability Analysis* *ENRE 655: Advanced Methods in Reliability Modeling*

**Bachelors of Science in Aerospace Engineering** **2004 to 2008** Iowa State University Ames, Iowa

### Projects

[**Cronicle**](https://github.com/jshiv/cronicle) is a git integrated workflow scheduler that provides a pull model for CI/CD and versioning on job execution.

### Invited Talks

[**PHM Society**](https://www.phmsociety.org/) **2015** automotive panel discussion

[**ARS**](http://www.arsymposium.org/) **2014** Big Data in Reliability: 1st Place