

# William Spoth

Website: [willspoth.com](http://willspoth.com) | Email: [wmspoth@gmail.com](mailto:wmspoth@gmail.com) | GitHub: [willspoth](https://github.com/willspoth)

## TECHNICAL SKILLS

---

**Languages:** Scala/Java, Python, SQL, C

**Technologies:** PostgreSQL, AWS, NiFi, Spark, Docker, Kubernetes

**Frameworks:** Serverless, DBT, GraphQL, Spring, Flask

**Tools:** Maven, Git, Jupyter Notebook, Microcontrollers

## EDUCATION

---

### University at Buffalo

**Doctor of Philosophy, Computer Science and Engineering**

Master of Science in Computer Science; GPA: 3.95/4.0

Bachelor of Science in Computer Science; GPA: 3.64/4.0

Bachelor of Arts in Psychology; GPA: 3.64/4.0

Buffalo, NY

February 2022

February 2018

May 2016

May 2016

## EXPERIENCE

---

### ParkHub | Data Engineer

February 2021 - Current

- Migrated critical analytics backend from custom software to DBT
- Improved OLTP and OLAP database performance, stability, and usability
- Implemented query monitoring, rewriting, database layout, partitioning, indexes, and major version upgrades
- Create and improve partner integrations, this includes utilizing state-of-the-art ETL tools such as Apache NiFi and AWS Glue
- Maintain and develop features for our real-time and historical analytic reporting software offering

### AWS Redshift | Software Development Engineer Intern

August 2020 - November 2020

- Prototype a massive system port to merge execution engines for Redshift and Spectrum
- Enable Redshift queries directly over S3 data without performing ingestion
- Prototype Apache Arrow scanner implementation

### CUBRC | Software Engineer Intern

May 2019 - August 2019

- Implemented non performance degrading support for historical queries in PostgreSQL
- Created automatic system deployment scripts using Kubernetes and Docker
- Updated existing systems to Spring Boot microservices
- Managed data pipelines and refactored RPC calls to external systems

### University at Buffalo | Database Teaching Assistant

Spring 2018

- Facilitated learning and implementation of operator pipelines, optimizer, and indexes
- Troubleshoot logical and performance issues using java profiler
- Created auto-grading system using RaspberryPi's and Docker

### University at Buffalo | Research Assistant

Fall 2017 - Current

- Incorporate research projects into existing production systems
- Co-review research papers
- Assist in preliminary grant research and writing

## FOCUS AREA

---

### Databases

- Created working database from scratch in Java including optimizer, indexes, and file formatter
- Created various functionality in PostgreSQL, Oracle, and SQLite using PL/SQL
- Implemented and refactored significant systems to include functionality such as Grafana and Apache Spark as core components

### Machine Learning

- Extensive use of TensorFlow, Keras, SparkML, Scikit-learn, and Smile libraries to create data models and visualizations
- Implemented a novel clustering technique to sort JSON datasets using Apache Spark and compared it to multiple Scikit-learn clustering algorithms
- Created simple multi-layered neural network example in TensorFlow and demonstrated uses and core concepts to non-domain audience

### Data Science

- Presented differentially private type-system with an intuitive Jupyter Notebook Demo
- Experience creating Bayesian models and visualizing data with predictive distributions
- Experience creating unique data mining tools using Map-Reduce and production database systems like PostgreSQL and Oracle

## INVITED TALKS

## COMPANY

Adaptive Schemas and Elastic Query for No-SQL Data  
Model based compression for JSON Collections  
Entity extraction for heterogenous JSON collections  
Why data staging is the hardest part of TensorFlow  
GRANTS

Oracle (2019)  
Snowflake (2019)  
Datometry (2019)  
Stark & Wayne (2018)

---

### Physical Layout Optimization & Query Transformation for messy JSON collection

\$95,635 from Oracle University Relations (#Award liu2641)  
Oliver Kennedy, William Spoth

## PUBLICATIONS

## CONFERENCE

### Loki: Streamlining Integration and Enrichment

William Spoth, Poonam Kumari, Oliver Kennedy, Fatemeh Nargesian

HILDA 2020

### Your notebook is not crumby enough, REPLace it

Michael Brachmann, William Spoth, Oliver Kennedy, Boris Glavic, Heiko Mueller, Sonia Castelo, Carlos Bautista, Juliana Freire

CIDR 2020

### Data Debugging and Exploration with Vizier

Mike Brachmann, Carlos Bautista, Sonia Castelo, Su Feng, Juliana Freire, Boris Glavic, Oliver Kennedy, Heiko Mueller, Remi Rampin, William Spoth, Ying Yang

SIGMOD 2019

### SchemaDrill: Interactive Semi-Structured Schema Design

William Spoth, Ting Xie, Oliver Kennedy, Ying Yang, Beda Hammerschmidt, Zhen Hua Liu, Dieter Gawlick

HILDA 2018

### Adaptive Schema Databases

William Spoth, Bahareh Sadat Arab, Eric S. Chan, Dieter Gawlick, Adel Ghoneimy, Boris Glavic, Beda Hammerschmidt, Oliver Kennedy, Seokki Lee, Zhen Hua Liu, Xing Niu, Ying Yang

CIDR 2017

## PROJECTS

[Mimir](#): Probabilistic data explorer

- Migrated backend machine learning models and core query executor from SQLite to Apache Spark
- Created core probabilistic schema resolver
- Implemented functional dependency based entity resolver

[json-schema-scala](#): JSON schema validator with precision metrics

- Parses JSON schemas into easily traversable Scala objects
- Validates JSON records against schema
- Reports helpful schema metrics like tightness of attributes, object decision trees, and schema diff
- Utilizes FastParse, performant parser library

[Json Explorer](#): JSON entity extractor

- Extracts ER-style entities and relationships from mixed JSON collections
- Discovers structure misuse through structure entropy analysis
- Utilizes Apache Spark for parallelization
- Exports schemas using json-schema specification

**FBtition**: Publicly verifiable petition signature protocol

- Block-chain based authentication
- PGP and web-of-trust validation
- Automatic removal of bot votes

## ACHIEVEMENTS & ACTIVITIES

ACM Member

ACM-SIGMOD Member

Best Graduate Teaching Assistant Award 2018

BSA Eagle Scout Excellence Scholarship