

# Aaron R. Smith

529 Victoria Drive, Cedar Park, TX 78613

☎ (512) 758-0313 | ✉ aaronreidsmith@gmail.com | 📷 aaronreidsmith | 🌐 aaronreidsmith

## Skills

---

<b>Languages</b>	Scala, Python, SQL, Java, Bash, R, Raku, Perl, PHP
<b>Big/Streaming Data</b>	Kafka, Spark, Databricks, Kinesis, Hadoop, HBase, Zookeeper, Hive, Pig
<b>Databases</b>	Snowflake, Redshift, MySQL, PostgreSQL, DynamoDB
<b>CI/CD</b>	Azure Pipelines, GitHub Actions, Jenkins, Travis CI, Appveyor, CircleCI
<b>Cloud Platforms</b>	Amazon Web Services, Google Cloud Platform
<b>Infrastructure/Deployment</b>	Terraform, Kubernetes, Docker, Spinnaker

## Work Experience

---

### IronNet Cybersecurity

Apr. 2021 - Present

#### SENIOR DATA ENGINEER

IronNet is the leader in collective defense, allowing companies to share data about potential breaches and keep entire industries secure. Recent projects include:

- Refactoring the legacy analytics infrastructure to be more reliable and scalable (Scala, Spark)
- Migrating the legacy infrastructure to a multi-tenant stream-based architecture, leading to better scalability, quicker onboarding, and faster threat detection (Scala, Kafka)

### Sumo Logic

Mar. 2020 - Apr. 2021

#### DATA ENGINEER II

- Part of team that developed a SQL-based rules engine to process data in real-time, allowing users to write custom queries to alert them about potential security threats (Scala, Kafka, ANTLR)
- Implemented a modified Rete Algorithm in the above rules engine to allow users to chain aggregations together without performance decrease (Scala, Kafka Streams)
- Developed a service to translate SQL-based rules into Sumo Logic's query language as part of the integration of Jask and Sumo Logic (Scala, Akka)

### Talroo

Apr. 2018 - Mar. 2020

#### DATA ENGINEER

Nov. 2018 - Mar. 2020

- Designed and built Talroo's first data warehouse and wrote data pipelines to populate it (Python, Kinesis, Spark, Snowflake)
- Helped create Talroo Insights™, an award-winning business intelligence product, as the sole data engineer on the team in charge of the project (Python, Snowflake, SQL, R)
- Migrated batch ETL processes to streaming (Python, Kinesis, Spark)

#### ASSOCIATE DATA ENGINEER

Apr. 2018 - Nov. 2018

- Part of a small team tasked with re-writing Talroo's platform to increase performance at scale (Java, DynamoDB, Kinesis, PostgreSQL)
- Developed automation scripts to eliminate downtime when altering legacy tables that require table-level locking (Python, MySQL)
- Wrote and maintained integration pipelines between Talroo's data layer and third-party services such as Salesforce (Scala, Spark, MySQL)

## Projects

---

### PMDARIMA

Active contributor to Pmdarima, "a statistical library designed to fill the void in Python's time series analysis capabilities, including the equivalent of R's auto.arima function." Notable contributions include:

- Adding Python 3.7+ support
- Developing and maintaining the project's CI/CD pipelines across GitHub Actions and CircleCI

## Education

---

### Johns Hopkins University

Dec. 2017

M.S. IN BIOINFORMATICS

### Texas A&M University

Dec. 2015

B.S. IN BIOCHEMISTRY AND GENETICS