

# Zachary (Zack) Novak

---

## SUMMARY

My name is Zack Novak, and I am a data engineer. I have a rapid ability to learn and apply new technologies to solve complex problems while also leveraging a high ability to discern when battle tested technologies solve the problem as effectively, quicker, and more reliably.

## SKILLS

- **Development:** Python, Shell, Spark, Git, SQL, Postman
- **Cloud:** AWS, Databricks
- **DevOps:** Terraform, Helm, Kubernetes, Docker
- **Orchestration:** Airflow, Prefect, Argo Workflows
- **Database:** MySQL, Postgres
- **Analytics:** Elasticsearch, Tableau, Excel, Python

## CERTIFICATIONS:

- Certified Associate Developer for Apache Spark 2.4 with Scala 2.11

## EXPERIENCE

### Second Front Systems, Remote – Data Engineer

Apr. 2021 – Present

- Key technologies: Python, Shell, Git, Docker, Kubernetes, Argo Workflows, GitHub Actions, AWS (RDS, ECR, EKS, EC2, VPC, IAM, S3, Lambda, Batch), Databricks, Postgres, Postman
- Refactored existing data pipelines created before joining to be more fault tolerant, and cut out middlemen data providers for increased data quality and quantity while reducing data costs.
- Implemented internal data team automation processes according to test driven development, Don't Repeat Yourself, and Keep It Stupid Simple principles by using GitHub Templates, private PyPI repositories, and base Docker images.
- Iteratively researched and chose software libraries and SAAS solutions while recognizing constraints from Department of Defense and internal Second Front requirements.
- Wrote IaC (Terraform and Helm Charts) for data applications.
- Lead workshops to introduce the data and development team to data team practices, tools, packages, processes, etc.
- Developed several python packages for interacting with AWS, validating and normalizing data, alerting using Slack and AWS SNS/SES/PinPoint, and more.
- Created Postgres Database maintenance strategy using DDL, ORM, Terraform, restoring from snapshot, etc.

## **Progressive Insurance, Mayfield Vlg, OH – Data Engineer**

Mar. 2020 – Apr. 2021

- Key technologies: Python, Prefect, Git, Terraform, Docker, Shell, AWS (EC2, VPC, S3, Redshift, DynamoDB, IAM), Flask
- Contributed to development of model serving applications directly plugged into Progressive's quoting flow for commercial and personal vehicles.
- Created an infrastructure as code solution using Terraform to deploy a Prefect orchestration execution module to business groups throughout Progressive.
- Developed and maintained internal Flask website that automates AI/ML team's work intake, reducing communication burdens on team, improving customer experience, and shortening new work intake process by ~900%.
- Shortened code elevation process from ~ 4 weeks to ~ 5 days through shell automation, unit / integration tests, and approved security infrastructure templates.
- Introduced new technologies to Progressive by leading security reviews, evaluating software against currently used technologies, and highlighting why the new software would be beneficial.

## **Visa Inc., Mentor, OH – Data Engineer**

Jan. 2019 – Mar. 2020

- Key technologies: Spark, Scala, Python, Airflow, SVN, Shell, Kafka, Elasticsearch, Tableau, MySQL, AWS (EC2, S3, Redshift), Databricks, Neo4J, Spring, Thymeleaf,
- Assisted in development of a data dictionary using neo4j that enabled business users to self-service requests as to where data fields were coming from, standard values to expect, and the field's lineage throughout our data system.
- Lead creation of automated parsing system for all internal system logs including > 2 year backfill process. Based on this data, I developed tools and dashboards that analyzed the numerous services within the logs to provide actionable insights into user behavior, identified business opportunities, and drove several business decisions regarding log analysis and service design.
- Developed and maintained an anomaly detection system that works on a customizable schedule for all of our cloud data sources that checks for abnormalities in input data using historical data and automatically alerts appropriate internal stakeholders of abnormalities.
- Due to widespread adoption of Databricks / Spark within CardinalCommerce, I created an "Introduction to Spark" user guide, a "Spark Optimization Guide" that helps to improve query / notebook speed using both beginner and advanced techniques and lead several lunch and learn sessions geared towards working with Spark.
- Optimized Airflow batch processing jobs by investigating known and suspected bottlenecks to ultimately lead to a 300% speed increase in jobs.
- Worked with program / product managers as a technical lead to design, test, implement, and maintain data solutions projects with 3rd party clients while complying with Visa and federal data compliance regulations.
- Automated report / file delivery to 3rd party clients using SFTP, pgp/gpg key encryption, boto3, and bash scripting.

## **EDUCATION**

### **The University of Akron, Akron, OH – Bachelor of Business Administration in Information Systems Management**

Aug. 2015 – Dec. 2018, Akron, OH

Outstanding 2018 – 2019 Management Information Systems graduate.