Zachary (Zack) Novak

Address: 4301 Kenmore Avenue Parma OH 44134
(440) 840-8527 || zack@znovak.io

TECHNICAL SKILLS:

- **Development**: Java, Scala, Python, Atlassian Tools (Jira, Confluence, Crucible), Bash, Shell, Apache (Airflow, Kafka, Spark), PuTTY, SVN, Git, Numpy, scikit, Pandas, JSP, Docker, AJAX, SQL, WinMerge, VirtualBox, SFTP, GPG, Venv, Git/BitBucket, JDBC, Gradle, Maven, Jenkins, Alteryx, JetBrains (PyCharm, IntelliJ), Eclipse
- **Cloud**: Amazon AWS technologies (S3, SES, IAM, EC2, Lambda), Databricks
- **Database**: MySQL, Postgres, MSSQL, SSRS, Amazon AWS (Redshift, RDS), InfluxDB. Neo4J, SQL Workbench, Hive
- **Front-End Development**: HTML, CSS, Jquery, JavaScript, Bootstrap, Spring, Flask, Thymeleaf
- **Analytics**: Elasticsearch, Tableau, Excel, R, Python

RELEVANT WORK EXPERIENCE:

CardinalCommerce | **Mentor, Ohio**   January 2019 – Present
Data Engineer

- Lead and developed automated parsing system for of 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 currently maintain 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 by all facets of the company, 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.
- Developed and currently maintain a healthcheck system that spot checks other services within Cardinal's Data Engineering team to alert of services being down and follows automated restart procedures if applicable.
- Optimized Airflow batch processing jobs by investigating known and suspected bottlenecks to ultimately lead to a 300% speed increase in jobs.
- Created and pushed for usage of standardized Airflow deployment system that includes testing changes in local Airflow VM instance and automated deployment using SVN, boto3, Jenkins, and bash shell scripting.
- Updated existing data processing jobs to use latest versions of libraries, and latest schemas while also developing an internal job QA that runs to detect schema changes of the data being read (expected schema vs actual schema).
- 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.
- Lead and pushed for use of documentation and knowledge sharing within team for greater visibility when troubleshooting and preventing knowledge siloes.

- Closely worked with Cardinal's data scientists and analytics team to assist them in projects by preparing custom data sets for use with analytic and machine learning projects.
- Upgraded legacy scripts and projects to use data collected from external APIs where applicable for more accurate / reliable data rather than static files created manually by internal employees.
- Investigated Cardinal's overall data pipeline and created documentation for others to reference on where a dataset is sourced / derived from, when it can be expected to be populated, and made recommendations to my manager on where lag can be cut down to increase processing speed.
- Contributed to Cardinal's Data Portal showing a dataset's information including input, output, owner, stakeholders, fields, normalized fields, field description, and schema using neo4j.
- Investigated database indexes within MSSQL used for our external reporting system and deleted unnecessary indexes improving overall query performance and freeing up system memory.
- Lead backfill projects on Elasticsearch indexes and created a standardized procedure with how to create a new index, how to update an existing indexes' template, and how to perform a backfill on multiple ES indexes.
- Provided technical support to business users on how to search for transactions within Elasticsearch (ES), how to create dashboards / visualizations, and how to overall increase their experience with ES.

CardinalCommerce | **Mentor, Ohio**   May 2016 – Aug 2016, May 2017 – Aug 2017, May 2018 – Aug 2018
Business Intelligence Engineering Intern / Big Data Intern
- Worked within Cardinal's Big Data group to assist in implementing new methods of data aggregation into their current production environment as well as preparing scheduled jobs
- Migrated singular scheduled Spark jobs into grouped jobs under separate AWS clusters using Apache Airflow
- Maintained and updated internal Big Data website with monitoring dashboards, reports, and triggers to push notifications to appropriate client managers, security, and/or production specialists
- Worked on projects involving design, development, and delivery of data analysis and reporting for internal and external consumers
- Performed functions dealing with report writing, capacity planning, ETL jobs, client data delivery, and testing within reporting database schema

OTHER WORK EXPERIENCE:
Diebold Nixdorf | **North Canton, Ohio**   Sept 2017 – May 2018
Jira Specialist / Project Management Intern

EDUCATION:
The University of Akron, College of Business Administration   Akron, Ohio
Bachelor of Business Administration in Information Systems Management   Graduated December 2018
GPA: 3.65 / 4.00   Dean's List: 4 semesters
Outstanding 2018 – 2019 Management Information Systems Graduate