

Michael J. Pedersen : (908) 283-0318 : datacyclist@gmail.com : [My Site](#) : [LinkedIn](#)

Executive Summary

DevOps Engineer with 28 years of experience managing Linux, UNIX, [Windows](#), and OSX/macOS systems. Programming languages have included Python, PHP, Perl, and Java. Work environments have been heterogeneous (several flavors of Linux, [Windows](#), and OSX/macOS), small to medium sized (from 10 to 1200 servers, 20 to 300 workstations), and mixed locations (all local to all remote teams).

Relevant Technical Skills

- **Big Data:** [Google Big Query](#), [HDFS](#), [Hive](#), [YARN](#), [Alluxio](#), [Impala](#), [Trino](#), [Kafka](#), [Kubernetes](#), [Dagster](#)
- **Database Servers:** [MySQL](#), [PostgreSQL](#), [Microsoft SQL Server](#)
- **Database Skills:** [PostgreSQL](#) Database Administration, Relational Schema Design, Structured Query Language (SQL)
- **Programming and Scripting Languages:** [Bash](#), C/C++, [Java](#), Javascript, [Perl](#), [PHP](#), [Python](#)
- **Programming Skills:** [Docker](#), [Jenkins](#), [Jira](#), [Intellij IDEA](#), Object-Oriented Design, Object-Oriented Programming, Refactoring
- **Software Configuration Management Tools:** [Git](#), [GitHub](#), [GitHub Actions](#), [Mercurial](#), [Subversion](#)

Relevant Job History

EvolutionIQ - Senior Software Engineer

New York City, NY - Feb 2024 - Current

- Orchestrated [Google Cloud](#) infrastructure with [Terraform](#), managing [GKE](#), [BigQuery](#), and developer virtual machines for scalability and efficiency.
- Redesigned EvolutionIQ's CI/CD pipeline, enhancing efficiency in development workflows.
- Fixed critical bugs in locally developed [GitHub Actions](#), ensuring smooth and reliable environment promotion.
- Built a template repository to guide EvolutionIQ teams in using [GitHub Actions](#) for deployment, as well as setting up data pipelines with [Dagster](#).
- Developed a Python-based CI/CD library to manage complex actions beyond [Bash](#)'s capabilities.
- Implemented [nektos/act](#) to enable developers to test build and deploy actions locally before committing to [GitHub](#).
- Refined [data retention policies](#) to guarantee compliance and ensure critical data availability for legal requirements.
- Implemented SLA monitoring with [Grafana](#), [SendGrid](#), and [PagerDuty](#) to proactively detect and alert customers and engineers about overnight job failures, ensuring swift issue resolution.
- On-call support on a rotating basis.

Pulsepoint - Data Engineer and Director of Infrastructure for Data

New York City, NY & Newark, NJ (Telecommute) - Mar 2015 - Nov 2023

Michael J. Pedersen : (908) 283-0318 : datacyclist@gmail.com : [My Site](#) : [LinkedIn](#)

Director of Infrastructure for Data, May 2018 - Nov 2023

- Architected data streaming that manages 40T of data/day.
- Established new data centers in Europe and in Virginia.
- Migrated data center, moving processing of data pipelines to new data center.
- Replaced [Vertica](#) with [Trino](#).
- Deployed and configured [Alluxio](#) for caching and data orchestration.
- Performance tuned [Kafka](#).
- Tested new tools for suitability, including [MariaDB](#), [Clickhouse](#), and [Kudu](#).
- Switched build server from [TeamCity](#) to [Jenkins](#), recreating all build jobs.

Data Engineer, Mar 2015 - May 2018

- Upgraded [Kafka](#) with zero downtime for producers and consumers.
- Enabled integration with [Active Directory](#) for [Hadoop](#) systems.
- Built tool to graphically show the ETL pipelines.
- Transitioned ETL pipeline from crontabs to [Mesos](#) and then into [Kubernetes](#).
- Troubleshooting of issues with [Hadoop](#), [Kafka](#), [SQL Server](#), and [Kubernetes](#).
- Production maintenance of data pipelines, including after hours support.
- Implemented data duplication between two [Hadoop](#) clusters.
- Upgraded [Hadoop](#) clusters with minimal downtime.
- Optimized the performance and reliability of [Hadoop](#), ensuring high availability.
- Added over a hundred automated tests to our ETL pipeline.
- Performed root cause analysis on ETL and cluster level failures.

Weight Watchers - Systems Engineering Lead

New York City, NY - Nov 2014 - Feb 2015

- Developed lightweight monitoring tool for use within my group.
- Configured [Vormetric](#) products to ensure [HIPAA](#) compliance for customer data.
- Worked to transfer from [Rackspace Cloud](#) to [Openstack](#) based private cloud.

OrcaTec, LLC - Developer

Atlanta, GA (Telecommute) - Jun 2012 - Oct 2014

- Found major security hole (remote code execution) and closed it.
- Debugged and resolved memory issues that were causing systems to shut down.
- Incorporated memcached into our stack to handle sessions and cached data.
- Switched web server from [Paster](#) to [Apache](#) with [mod_wsgi](#).
- Corrected Unicode handling errors in the code.
- Added holds and matters framework, allowing customers to state that documents belong to specific cases and should not be deleted while the cases are ongoing.
- Identified weaknesses in the database model, and added code to prevent those weaknesses from being hit.

Choopa.com - Developer

Sayreville, NJ - Jan 2012 - May 2012

- Developed library to manage [OpenStack](#) nodes, and gather billing information.
- Built [Nagios](#) configuration file generator for in-house web interface for [Nagios](#).
- Configured [Bacula](#) backup system as replacement for custom backup scripts.
- Reconfigured [Nagios](#) monitoring, reducing full check from 8 hours to 2 minutes.

Michael J. Pedersen : [\(908\) 283-0318](tel:9082830318) : datacyclist@gmail.com : [My Site](#) : [LinkedIn](#)

- Refactored in-house [Nagios](#) web interface. This reduced the workload from six files down to one when adding new checks.
- Several smaller bug fixes and features throughout the internal code base.

Education

Bachelor of Science in Computer Science, 2000
East Stroudsburg University, East Stroudsburg, Pennsylvania

Professional Certificates

[Google Cloud Certified Professional Data Engineer](#) (Jan 2024)
Online Course - Google