

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

Resume: Michael J. Pedersen

Summary

With a 28 year career, I have filled a variety of technical roles. As a manager, I have spent 5 years of my career guiding teams through challenging conditions, ranging from technical challenges through COVID-19 related challenges. I have kept my team together, and we have thrived on the challenges of taking care of the data for a \$200 million company. This includes managing my team through the COVID-19 pandemic, without losing a single member of my team to another company.

I am a [Google Cloud Certified Professional Data Engineer](#). I have spent 9 years working with streaming challenges, ETL challenges, and data latency challenges as we keep our overall latency to 5 hours or less. We have done this using [Hadoop](#), [Python](#), [Kafka](#), and [Alluxio](#). We have worked with different data storage formats, different compute engines for accessing the data, and different ways to coordinate our ETL pipeline to avoid having jobs crash into each other.

Furthermore, as a DevOps Engineer, I have 23 years of experience managing Linux, UNIX, [Windows](#), and OSX/macOS systems. This means that I look at the whole picture, not just System Administration or Software Development. Shepherding a system through the creation and deployment process, and seeing the customer's happiness at having things work the way they need it to, is a particular joy of mine. Making people's lives better is the point of technology, after all.

Finally, as a Software Engineer, I have spent 6 years of my career focused on delivering high quality software to my company's customers, with their focus and needs being on sorting through large numbers of documents in a timely fashion. This has meant understanding ingestion, storage, and display of arbitrary data. It has included custom data visualizations. This was primarily done with [Python](#) and [Ubuntu Linux](#), but has also included work with [Perl](#) and [PHP](#).

I am comfortable in a wide range of working conditions. 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). Programming languages have included Python, PHP, Perl, and Java.

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

Relevant Technical Skills

- **Programming Skills:** [Docker](#), [Jenkins](#), [Jira](#), [IntelliJ IDEA](#), Object-Oriented Design, Object-Oriented Programming, Refactoring
- **Database Skills:** [PostgreSQL](#) Database Administration, Relational Schema Design, Structured Query Language (SQL)
- **Big Data:** [Google Big Query](#), [HDFS](#), [Hive](#), [YARN](#), [Alluxio](#), [Impala](#), [Trino](#), [Kafka](#), [Kubernetes](#), [Dagster](#)
- **Programming and Scripting Languages:** [Bash](#), C/C++, [Java](#), Javascript, [Perl](#), [PHP](#), [Python](#)
- **Software Configuration Management Tools:** [Git](#), [GitHub](#), [GitHub Actions](#), [Mercurial](#), [Subversion](#)
- **Database Servers:** [MySQL](#), [PostgreSQL](#), [Microsoft SQL Server](#)
- **Operating Systems Administered:** Linux ([Debian](#), [RedHat](#), [Suse](#), [Ubuntu](#)), Microsoft Windows (10/2008/7/Vista/2003/XP/NT/98/95), UNIX ([Solaris](#), [AIX](#), [HP-UX](#))
- **Markup Languages:** CSS, HTML, Markdown, XML
- **Applications:** [Ipswitch What's Up](#), [Nagios](#), [OpenStack](#), [Slack](#), [VirtualBox](#), [VMware](#), [Zenoss](#)
- **Networking and Security:** [Checkpoint VPN](#), Cisco, Firewall Design, TCP/IP

Job History

EvolutionIQ - Senior Software Engineer

New York City, NY - Feb 2024 - Current

EvolutionIQ leverages cutting-edge AI and machine learning to revolutionize the insurance industry, optimizing claims management and accelerating individuals' return to the workforce. In January 2025, EvolutionIQ was acquired by CCC Intelligent Solutions, further expanding its impact and innovation in the sector.

As a Senior Software Engineer, I played a key role in delivering projects within the Platform Pipeline Infrastructure team.

- 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.

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

- 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

Pulsepoint is an internet healthcare marketing company with a focus on activating health care providers. Pulsepoint was acquired by WebMD in June 2021.

My role evolved over time from dealing with individual data jobs to overseeing the entire ETL pipeline to leading the entire department.

Director of Infrastructure for Data, May 2018 - Nov 2023

- Architected data streaming that manages 40T of data/day.
- Lead maintainer for ETL pipelines, encompassing over 250 transformations.
- Established new data centers in Europe and in Virginia.
- Migrated data center, moving processing of data pipelines to new data center.
- Split data management team into data platform and data product development.
- Guided the team through splitting our ETL pipelines into multiple repositories.
- Organized the migration of ETL pipelines from Python 2 to Python 3.
- Instituted and formalized processes and procedures for the team.
- Planned capacity to ensure we could handle incoming data throughout the year.
- Replaced [Vertica](#) with [Trino](#).
- Reported on system wide data latency using [ElasticSearch](#), [Kibana](#), and [Grafana](#).
- Conducted interviews for my team and for teams that work closely with my team.
- Automated distribution of incident reports to all affected parties.
- Changed hardware profiles for [Hadoop](#) to remove storage and compute colocation.
- Acted as scrum master for the team.
- Onboarded new team members, helping them to fully integrate into the team.
- Held weekly 1 on 1 meetings with team members.
- Developed new stories (including estimates) for our [Jira](#) board.
- Prioritized tickets for our [Jira](#) board.
- Passed annual HIPAA training for data protection.
- 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.
- Developed roadmap for the Data Platform team.

Data Engineer, Mar 2015 - May 2018

- Participated in on-call rotation.
- 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 cronjobs 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.
- Created ELT jobs to ingest third party data to make it available internally.
- Installed and configured multiple [Hadoop](#) clusters.
- Developed new ETL jobs to aggregate data from Pulsepoint's RTB exchange.

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

- Optimized [Hadoop](#) jobs.
- Maintained [Vertica](#) cluster, including troubleshooting.
- Tested [Cassandra](#) as a potential reporting database.
- Converted [Sqoop](#) jobs to use [FreeBCP](#) instead.
- Collaborated with other teams to help them use the systems to find the data they need.
- Optimized the performance and reliability of [Hadoop](#), ensuring high availability.
- Worked with other teams to define and then implement needed features for our internal ETL pipeline framework.
- Added over a hundred automated tests to our ETL pipeline.
- Performed root cause analysis on ETL and cluster level failures.
- Managed data backfill issues whenever they arose.

Weight Watchers - Systems Engineering Lead

New York City, NY - Nov 2014 - Feb 2015

Weight Watchers is a Fortune 500 company focused on helping customers manage their weight and reduce health problems caused by it.

My role was focused on providing internal support within the company to enable other groups to support the customer base.

- 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

OrcaTec is in the litigation support industry (they help their clients reduce the costs of being sued). OrcaTec is primarily a software-as-a-service company, allowing OrcaTec to host customer data. While working here, my focus has been on improving the GUI. This has involved refactoring code heavily, adding new features, and adding new tests to cover existing and new code.

The team structure at OrcaTec is geographically very diverse. In addition to my own telecommuting, I have teammates in many states. We all work remotely, and we all work together to make the product the best that it can be.

- Mentored other developers in the use of [TurboGears](#), [SQLAlchemy](#), [Python](#), and JavaScript.
- Organized weekly meetings for members of the frontend (OTGUI) team, providing a chance to discuss (in depth) the issues the team was facing.
- 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.
- Wrote [Python](#) framework to manage long running background jobs.
- Reduced multi-hour [SQLAlchemy](#) bulk database jobs to minutes.

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

- Spearheaded conversion from [YUI 2](#) to [jQuery](#) and [jQueryUI](#).
- Documented internal server API, wrote a [Python](#) class to standardize it's use.
- Added tag cloud (using [awesomecloud plugin for jQuery](#)).
- Added support for allowing customers to login using [OpenID](#).
- Developed advanced search tool using [Python](#), [TurboGears](#), and [jQuery](#).
- Created new document production framework from scratch.
- Installed and configured [WSO2 Identity Server](#) for our [OpenID](#) implementation
- Created a tool to allow copying settings between instances.
- Added user preferences to the frontend.
- Resolved intermittent issue with drag/drop events that had been unsolvable by the existing team.
- Implemented login idle timeout functionality.
- Refactored [Python](#) and JavaScript code on a regular basis to reduce code repetition and increase legibility.

Choopa.com - Developer

Sayreville, NJ - Jan 2012 - May 2012

As a developer at Constant.com (renamed from Choopa.com in Jan, 2012), I worked with a variety of technologies, with the heaviest focus being on [OpenStack](#) and [Nagios](#). I helped bring two products to production level availability for their customers (specifically: the Dedicated Cloud Server and Backup systems).

- 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.
- 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.

6th Avenue Electronics - Systems Administrator, DevOps Engineer

Springfield, NJ - Aug 2005 - Apr 2008, Feb 2011 - Dec 2011

In 2007, 6th Avenue began switching from their then-current POS system (named [Tyler](#)) to [SAP](#). At the end of 2010, [SAP](#) was declared unworkable, and the effort was begun to switch back to [Tyler](#).

The environment at 6th Avenue covered a wide range of platforms spread out over 120 servers (both physical and virtual). We had [VMware](#) ESX, [Windows](#) Server 2003, [Windows](#) Server 2008, [CentOS](#) Linux, [Suse](#) Linux, and [Debian](#) GNU/Linux. In 2011, I was brought back to transition the point of sale system and become the IT Manager. At the time the point of sale transition was completed, we had a team of 6 people managing the servers and about 300 desktops.

- Successfully lead migration from [SAP](#) to [Tyler](#) Point of Sale system.
- Developed [Python](#) validation scripts for data going from [SAP](#) into [Tyler](#).
- Automated configuration options within [Tyler](#) that could not be done via import.
- Developed [Python](#) program to copy sales data from [Tyler](#) POS to [PostgreSQL](#).
- Installed and configured [Zenoss](#) for full systems monitoring.
- Implemented [VMware](#) Virtual Infrastructure 3.
- Maintained [Tyler](#) POS/ERP system on [HP-UX](#) (and, later, Linux).
- Maintained Active Directory, including implementation of group policy.

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

- Wrote scripts to satisfy company needs using [AutoIt3](#) and [Python](#).
- Wrote automated installer for the [Tyler](#) client program to incorporate the program plus the mandatory pieces that we needed.
- Developed workaround to resolve issue in point of sales system causing store wide sales terminal lockups.
- Maintained heterogeneous environment (>60 Linux, >40 [Windows](#) servers).
- Implemented ticket tracking system for help desk issues.
- Deployed [Windows Software Update Server](#) for Microsoft product updates.
- Updated customer facing web site to reflect changes to NJ sales tax rates.
- Exported data from [Tyler](#) point of sale system for import into [SAP](#) system.
- Created an internal wiki for use by the IT department, including populating with over 30 pages of documentation at time of deployment.
- Wrote several scripts to extract data from [Tyler](#) POS system before [PostgreSQL](#) database was available.
- Maintained [CommVault](#) backup system and disaster recovery site.
- Developed intranet pages (using AJAX) to allow customer service representatives to find old invoices in the database copy of [Tyler](#)'s data.
- Created intranet pages (using AJAX) to assist in the selling of complex systems.
- Retrieved bulk information from [Tyler](#) point of sale system for audits.
- Performed field certification of [MaxDB](#) system for [CommVault](#), providing reliable and supported backups for [SAP](#) databases.
- Configured all servers for newly implemented [SAP](#) system.
- Spearheaded server room cleanup: Shut down over 30 servers, removed over a mile of wire.
- Rack mounted, installed, and prepared newly arrived servers for use in projects.

Datapipe, Inc. - UNIX Developer

Jersey City, NJ - May 2008 - Jan 2011

Datapipe manages thousands of customers servers. Many of these servers are connected to various shared storage systems, including [3Par](#), [Isilon](#), and backup servers. Datapipe required an ability to do reporting on what data was being stored on these systems for each client, and then report that data back to billing. In addition, Datapipe required monitoring of the backup systems to ensure timely and complete backups of client data. My duties primarily focused on making these systems work well.

My team structure is worth describing briefly as well: My immediate manager worked out of Austin, TX. One coworker worked in the same building as myself, and I had two "extended" teammates who worked in Jersey City, NJ (I worked in Somerset, NJ). The extended team included the [Windows](#) developers, while I was on the UNIX development team.

- Created reporting system called StorageWeb (using [TurboGears](#)), enabling new revenue stream.
- Developed [Python](#) app named unixops, allows server access via one time [SSH](#) keys.
- Optimized [PostgreSQL](#) on [FreeBSD](#). Bulk inserts reduced from hours to 20 minutes.
- Debugged [Python](#), [FreeBSD](#), [Apache](#), and [modwsgi](#) working together.
- Developed multi-threaded back end daemon (in [Python](#)) which connected to the various storage systems and gathered the data about the stored data for reporting before pushing aggregate data to the billing system.
- Developed web interface that would allow users to drill down and see how storage was being used (by client, by server, by data center, and/or by storage type).
- Wrote tool to gather performance data from [3Par](#) InServ nodes and display it via the

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

client portal.

- Updated and maintained the existing backup monitoring tool which reported backup failures to our main ticketing system.
- Repackaged [Bacula](#) (internal name: SureRestore) for all supported platforms.
- Evaluated potential replacements for [Subversion](#), including [Git](#) and [Mercurial](#).

Diversified Systems - Systems Administrator / Developer

Hackettstown, NJ - Sep 2002 - Jul 2005

Diversified Systems is a small company that focuses on low voltage wiring and subcontracting. While there, I wore many hats, and did work on every system. The total number of servers for this company was less than 10, and the entire IT department consisted of myself.

- Developed GUI to [new software system](#) using [PHP](#), [Apache](#), and [Mozilla](#).
- Automated sending faxes to techs, saving five hours/day (in a 10 person office).
- Deployed [Unattended](#), an automated [Windows](#) installation system.
- Implemented [HylaFax](#) fax server for incoming and outgoing faxes, allowing electronic receipt of over 200 pages of faxes per day from field technicians.
- Worked with upper management to completely redesign entire business processes and systems company-wide (accounting, customer service, builder coordination, sales and warehouse management).
- Implemented employee remote-access system, using VPN (virtual private network).
- Implemented employee remote-access system, using VPN (virtual private network).
- Deployed SNMP and monitored daily resource utilization.
- Converted structured portion of data from older system into a database, and provided training to employees on proper usage.
- Created [PHP](#) scripts to provide clean access to unstructured data from older system, and showed employees how to access and use.
- Managed upgrade of entire office to [Windows](#) 2000. This involved testing programs for compatibility, replacing obsolete programs, and determining proper installation procedures for undocumented installations.
- Resolved issues with the various systems which Diversified Systems installs in customer homes, including alarm systems, stereos, central vacuums, and structured wiring on a daily basis.
- Coordinated the activities of field technicians with customers to provide the maximum service level to the customers.
- Authored new scripts using [Perl](#) and [PHP](#).

Ciber, Inc. / Decision Consultants - Member of Technical Staff

Greenwood Village, Co - Mar 1999 - Sep 2002

Decision Consultants (DCI) was acquired by Ciber, Inc., in 2002. While working for DCI, I was contracted out to Coors, IBM, and a .com named "X-Care" (no longer in business). The points below come from all of those places.

- Ran, and later automated, nightly code compilations for patient records program. This effort saved approximately 1000 developer hours per week. Before this effort, corrupted nightly compilations stopped the whole team until resolved (usually an entire day would be lost). After this effort, no corrupted compilations occurred for over six months.
- Revamped and improved scripts used by developers for retrieving the nightly code compilations and to perform their own personal compiles, providing new features as

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

requested.

- Wrote several smaller scripts related to the compilation process, to help developers understand what would be required for their work to be completed.
- Researched/corrected error in [Perl](#), permitting resolution of time-zone conversion issues, enabling global use of data from medical care providers.
- Created [Perl](#) scripts to migrate health care provider data between systems.
- Integrated [AIX/Solaris](#) servers into [Windows](#) NT network, allowing developers on [Windows](#) to access [AIX/Solaris](#) files/printers.
- Developed a set of [Solaris](#) packages allowing deployment of new servers within 2 hours of receipt.
- Developed [Ghost](#)-like utility overnight, meeting next-morning deadline for usable computer loads.
- Configured several Sun Ultra servers to work as part of a network. Original condition was such that they were on a network, but not working together. This involved re-mapping user id numbers, and configuring NFS mounts such that the machines worked together.
- Configured and administered a [CVS](#) repository.
- Created a set of [Solaris](#) packages to allow the deployment of new [Solaris](#) servers within 2 hours of receiving them.
- Coded many scripts to perform several daily tasks.
- Downloaded and tested new tools to be used for the compilation process, to make sure they still produced correct results.
- Installed [Perl](#) modules and programs as requested by developers.
- Documented all new scripts and processes, and informed developers when new documents were available.
- Participated in configuration of training room using [Red Hat Linux](#) with [Kickstart](#).
- Assisted customers in resolution of issues with [Windows](#) 95, [Windows](#) NT, Microsoft Office, and other software packages in use throughout Coors.
- Instructed junior developers in the inner workings of C++.

Robert Half International - Technical Support

Boulder, Co - Jan 1999 - Feb 1999

Robert Half International's client, StorageTek, provided large enterprises with long term backup solutions (typically involving dozens of tape drives, thousands of tape cartridges, and robotic tape libraries to manage all of it).

- Assisted customers of StorageTek in resolution of problems with both hardware and software products.

Sykes Enterprises - Systems Technologist

Denver, Co - Aug 1998 - Dec 1998

Working for Sykes Enterprises, I was contracted out to Sun's internal Resolution Center. I worked with Sun employees around the world to resolve their issues with the workstations and servers they relied on daily.

- Performed remote operating system installations and upgrades.
- Resolved customer issues with [Solaris](#) 2.5.1 and [Solaris](#) 2.6.
- Wrote a [Tcl/Tk](#) script to speed up the process of logging into customers' machines for use in the Resolution Center.
- Wrote a Korn shell script to check a list of users and make sure that all users on the list were valid Sun employees.

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

- Added new users throughout Sun's internal network

Fabian Corporation - System Administrator

Stroudsburg, Pa - Feb 1998 - May 1998

Fabian Corporation was a small virtual hosting provider for web sites during the fledgling web days, even before the dot-com era. A typical customer made a static web site and uploaded it via ftp to show to any web site visitors.

- Added new domains to Linux servers for web site hosting.
- Added new user logins to servers.
- Set up sendmail to forward email addresses from hosted domains to local users and remote users.
- Performed system upgrades (both hardware and software).
- Performed system backups.
- Dealt with security issues through upgrades, and removal of suspicious software.
- Installed firewall.
- Upgraded and recompiled kernel as needed.

MaxTech Corporation - Developer / System Administrator

Rockaway, NJ - Mar 1995 - Dec 1997

I was hired at MaxTech as a customer service representative. Over the time I worked there, I earned the opportunity to participate in system administration and the development of a new call tracking system to be used by the customer service team.

- Assisted customers daily with issues installing, configuring, and using their MaxTech modems.
- Discovered bug in the newly released modem drivers for [Windows](#) 95 and the MaxTech 28.8kbps modems.
- Created a new [Windows](#) based call tracking system to replace the old [MS-DOS](#) based call tracking system. Used [Delphi](#) and [InterBase](#) as the development environment and database.
- Rebuilt [Novell NetWare](#) server that had experienced hard drive crash. Did so while the server was in Atlanta, GA and I was in Rockaway, NJ.
- Fixed issues with the [Lotus cc:Mail](#) SMTP gateway.
- Helped test the new MaxTech website.

Personal and Side Projects - Developer, Systems Administrator

1995-Current

When I'm not working on projects for my employer, I'm working on projects for myself, or side projects for people who get in touch with me to make something for them.

- Studying Android application development at Udacity.com.
- Starting in 2009, I began participating in the [TurboGears](#) project, working primarily on the documentation. In 2011 and 2012, I was the lead project maintainer, and we put out three releases in 2011 alone. As of now, I still manage the server and DNS for turbogears.org, with work on documentation, bug fixing, new features, and mailing list management, as time permits.
- Created [Java](#) plugin for [Openfire XMPP](#) server, allowing [vBulletin](#) forums to have a working XMPP server for their communities.
- Created Linux-based network featuring NIS, NFS, DHCP, Linux firewall (using

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

- iptables), [Samba](#), [SSH](#), [Subversion](#) server, and [Mercurial](#).
- Customized installation of [Request Tracker](#) for San Diego firm.
- Contributed patch to [Mercurial](#) to assist with repository conversions. Specifically, it allows branches to be renamed (useful for repositories that used named branches in [Subversion](#) to change their main trunk location).
- Implemented initial [Pluggable Authentication Module](#) support for [HylaFax](#), which was accepted into [HylaFax](#) 4.2.0.
- Contributed code to [MythTV](#) project, allowing users to save recordings using custom cut lists, which allowed for easy removal of commercials.
- Contributed documentation to [WebGUI](#) project, showing how to design a custom theme for [WebGUI](#).
- Contributed documentation to [libpqxx](#) project, showing how to compile [libpqxx](#) using [MinGW/MSYS](#) on [Windows](#).
- Helped clients, family, and friends resolve various computer and home networking issues.

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

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

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

Project History

Migrate To New Data Center

| | |
|-----------------|---|
| Period | 2022-2023 |
| Company | Pulsepoint |
| Tools | Alluxio , Hadoop , Kafka , Python |
| Platform | CentOS , Kubernetes |

Pulsepoint is in the process of migrating between data centers. A significant portion of the existing hardware has gone past its end of life, so we chose to build a new data center, with new hardware. At the same time, we used the latest versions of all relevant software that we could ([Hadoop](#), [Kubernetes](#), etc).

This provided us with an opportunity to fix some design flaws in the original big data clusters, and we used this chance to make things better for us overall.

The work remaining at this point comes down to verifying that the new versions of the ETL jobs function as expected, producing valid output. The process is expected to complete in 2025.

- Created new clusters, with new versions of relevant software, in the new data center.
- Updated ETL jobs as needed so that they would run exclusively in the new data center.
- Configured those ETL jobs to output copies of their data to the original data center.
- Removed those ETL jobs from the original data center, configuring the original to use the output from the new data center.

Migrate From Python 2 to Python 3

| | |
|-----------------|---|
| Period | 2022-2023 |
| Company | Pulsepoint |
| Tools | Python |
| Platform | CentOS , Kubernetes |

Pulsepoint built the entire ETL pipeline using [Python](#) 2. On January 1, 2020, Python 2 reached its end of life. In order for the ETL pipeline to continue to grow, we needed to migrate to Python 3.

The path we chose was to extract the code that was common to the pipeline, and turn that code into a library. We then began the normal route of making backwards incompatible changes. Because of the scope of this work (nearly 200K lines in Python files), and the work being done during a data center migration, the project is still ongoing. However, over 50K lines have been successfully completed so far.

- Established a library cutoff version, after which the library would no longer support Python 2.
- Began regular release cycles for the library
- Ensured that developers outside of the library maintenance team could use the library to easily migrate ETL jobs.

Dataflow Explorer

| | |
|---------------|------|
| Period | 2015 |
|---------------|------|

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

| | |
|-----------------|---|
| Company | Pulsepoint |
| Tools | Python , Graphviz Dot , Luigi |
| Platform | Mesos , CentOS , NGINX |

At Pulsepoint, we have a large number of data aggregation jobs that are coordinated with each other via Spotify's [Luigi](#) tool. [Luigi](#) has the user create a [Python](#) codebase that resolves which order to do jobs similar to how [GNU Make](#) actually works. A negative side effect of this is difficulty for humans to understand the order of jobs that will be run when the number gets to any significant size.

The Dataflow Explorer would walk the [Python](#) code that represented all of the jobs, and extract the attributes that would allow construction of a dependency tree. It would then pass that tree to the [Graphviz DOT](#) tool, which would run dot to produce an SVG file showing the graph of all the jobs. Finally, it would publish that output onto [Mesos](#) using [NGINX](#), allowing people to browse, zoom, and search the resulting graph.

- Wrote code to walk a [Python](#) code base and extract specific attributes
- Produced syntactically valid [Dot](#) files.
- Automatically published updated versions of the graph for myself and others to use.

Cassandra for User Reporting

| | |
|-----------------|------------------------------|
| Period | 2015 |
| Company | Pulsepoint |
| Tools | Cassandra |
| Platform | CentOS Linux |

Pulsepoint has a fairly significant [Microsoft SQL Server](#) installation, and we were asked if we could use [Cassandra](#) as a replacement for it. We set up a small cluster, and began trying to run various reports against it.

The actual performance was impressive, but we ran into a significant roadblock: Cassandra is, in significant ways, a disk based key/value store. In order to use this as a reporting database, and avoid triggering table scans for the user reporting, we would have had to load up many copies of the same data into different tables with different primary keys.

In the end, this was deemed non-feasible for the number of combinations we would have had to provide, along with the amount of maintenance as new reports could be brought online.

- Deployed a [Cassandra](#) cluster.
- Produced data sets into that cluster.
- Confirmed queries ran, and ran well.
- Ultimately recommended against because of the issues with table scans and primary keys.

California Hadoop Cluster

| | |
|-----------------|------------------------------|
| Period | 2015 |
| Company | Pulsepoint |
| Tools | Hadoop |
| Platform | CentOS Linux |

Pulsepoint needed to establish a disaster recovery site, and had chosen an existing data center to do so. In the process, establishing a [Hadoop](#) cluster was required for business continuity. My task was to get everything configured to the point that the same data jobs running in the primary cluster ran in the backup cluster and provided equivalent data, even though everything was running independently.

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

- Installed [Cloudera Distribution of Hadoop](#) across the cluster.
- Ensured that [HDFS](#), [Hive](#) and [Impala](#) were functioning properly.
- Ensured that the same data jobs running in the primary cluster were running in the secondary cluster.
- Ensured that equivalent output was happening in both data centers.

Sqoop to FreeBCP([FreeTDS](#)) Conversion

| | |
|-----------------|---|
| Period | 2016 |
| Company | Pulsepoint |
| Tools | Sqoop , FreeTDS |
| Platform | Hadoop , Microsoft SQL Server |

[Apache Sqoop](#) has long been deprecated, with its eventual complete retirement in June 2021. As part of Pulsepoint's platform, we needed a replacement for [Sqoop](#) before it was fully retired. We settled on FreeBCP, which is part of the [FreeTDS](#) project. Using this tool, we were able to migrate our processes for transferring data from [Hadoop](#) to [MS SQL Server](#).

- Developed migration strategy to transition from [Sqoop](#) to FreeBCP.
- Tested FreeBCP as a substitute for [Sqoop](#).
- Updated our ETL pipelines to use FreeBCP in place of [Sqoop](#).

Vertica Decommissioning

| | |
|-----------------|---|
| Period | 2018 |
| Company | Pulsepoint |
| Tools | Vertica , Trino |
| Platform | CentOS Linux |

Pulsepoint had used [Vertica](#), but we were outgrowing it in 2017. In 2018, when we came up for the most recent support renewal, we had fully outgrown it and needed to replace it with something else. After trying out several other options (including [Clickhouse](#), [Trino](#), [MariaDB][MARIADB], and others), we settled on Trino as the option that provided us with the best capabilities while being nearest to the performance that [Vertica](#) provided.

- Performance tested existing [Vertica](#) queries.
- Stood up several competitors and compared their performance using the same queries.
- Compared maintenance of these environments to Vertica.
- Finally chose [Trino](#), implemented it, and fully decommissioned [Vertica](#).

Data Management Team Split

| | |
|-----------------|---|
| Period | 2021 |
| Company | Pulsepoint |
| Tools | Git |
| Platform | Jira , GitHub |

As part of the growth of Pulsepoint, the Data Management team reached a point wherein the team was no longer able to do everything that was required: New data products were needed, and the data platform itself needed both maintenance and new features as well. I made the decision to split the team in two, creating a Data Platform team and a Data Product Development team. Each team would be focused on exactly one role, instead of trying to split the focus between two distinct functions.

- Divided the team into two distinct functional teams.
- Divided the code between the two teams to reflect their individual functions.

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

- Divided the [Jira](#) board between the two teams.
- Established new teams on [GitHub](#), with each team getting only the portion of the code belonging to them.

Data Management Code Split

| | |
|-----------------|------------------------|
| Period | 2021 |
| Company | Pulsepoint |
| Tools | Git |
| Platform | GitHub |

Pulsepoint needed to split the Data Management team into a Data Platform team and a Data Product Development team. This also meant splitting the code, since the entirety of the ETL pipeline was in one monolithic repository. The team had to develop a means of crossing repository boundaries to establish the pipeline steps (e.g.: Job A in repository 1 is dependent on Job B in repository 2). We also had to come to agreements on how to determine which team got which pieces of code.

- Developed cross-repository dependency system for ETL jobs.
- Agreed on terms to decide which team got which piece of code from the original repository.
- Created new repositories to get that code.
- Created new teams on [GitHub](#) to assign ownership over the newly divided code.

Advanced Search Tool

| | |
|-----------------|--|
| Period | 2014 |
| Company | OrcaTec, LLC |
| Tools | Python , jQuery , jQueryUI |
| Platform | Server: TurboGears , Browser (Cross Browser) |

At OrcaTec, the primary tool we provided to our customers was the ability to search collections of documents quickly. In addition to having simple search tools, we also had a helper tool in the “Advanced Search”.

This tool allowed the user to search based on a dozen different fields, but was still limited and fragile. It was unable to help the user build queries which combined different fields in a single clause. In addition, it had issues with encoding <> in email addresses, and did not support drag and drop on all of our supported browsers.

When this project was completed, this tool had transformed noticeably. It now is its own miniature investigative tool, allowing customers to easily search through collections of documents. One customer reported narrowing their searches from 80,000 possible documents down to under 2,000 within an hour through use of this tool. Due to extensive test coverage when the code was published, even the problems that were found were quickly fixable. All of this was accomplished while reducing the total code for it by 50%.

- Debugged issues with drag/drop on mobile browsers.
- Designed new interface for maximum flexibility, and to allow easy refinement of queries as they are being built.
- Incorporated user feedback to improve that design.

[Paster](#) to [Apache/mod_wsgi](#) Conversion

| | |
|----------------|--------------|
| Period | 2013 |
| Company | OrcaTec, LLC |

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

Tools

[Python](#), [Apache](#), [mod_wsgi](#), [Paster](#)

Platform

[Ubuntu](#) Linux

[Paster](#) is meant to be used in a development environment, allowing the developer to use a (single threaded) lightweight, easily managed webserver while writing code before it goes to production. At OrcaTec, we were using [Paster](#) both in development and in production. Due to the demands being placed on [Paster](#) (in many instances, loading up documents that were over 100M), the entire system could appear (to one user) to freeze up due to it responding to a request from another user.

After analysis, we were able to determine that [Paster](#) was no longer suitable for our needs. Since [Apache](#), with [mod_wsgi](#), provides an at least adequate performance web server (in comparison to others like Nginx), and the [Apache](#) configuration was already known to the team, we chose to switch from [Paster](#) to [Apache](#). This allowed us to have [Apache](#) itself serve up static files (like images, css files, and javascript files), leaving the dynamic pages to the [Python](#) code.

- Debugged threading/locking/memory usage issues with [Paster](#).
- Recompiled and repackaged [Python 2.6.8](#), [Apache](#), and [mod_wsgi](#) for use with [Ubuntu 10.04](#).
- Developed automatic [Apache](#) configuration for use within our local stack.

StorageWeb

Period

2010

Company

Datapipe

Tools

[FreeBSD](#), [Python](#), [Apache](#), [PostgreSQL](#), [TurboGears](#)

Platform

[FreeBSD](#), Web Browser

Datapipe manages thousands of servers. Many of these servers are connected to various shared storage systems, including [3Par](#), [Isilon](#), and backup servers. Datapipe required an ability to do reporting on what data was being stored on these systems for each client, and then report that data back to billing. StorageWeb was written to fill that need.

- Debugged issues with [Python](#), [FreeBSD](#), [Apache](#), and [modwsgi](#). Turned out to require specific compilation options to get these all working correctly.
- Developed web interface that would allow users to drill down and see how the storage was being used (by client, by server, by data center, by storage type).
- Developed multi-threaded backend daemon which connected to the various storage systems and gathered the data about the stored data for reporting.
- Developed backend daemon that pushed aggregate data to the billing system, allowing billing to finally happen for all clients.

UNIXOps

Period

2010

Company

Datapipe

Tools

[FreeBSD](#), [Python](#), [Apache](#), [PHP](#)

Platform

[FreeBSD](#), Web Browser

Datapipe provides managed hosting for its clients. This means that customers contact Datapipe to report issues on servers, and Datapipe administrators log in to customer machines as root to fix the problems. UNIXOps provides a secure method to allow the administrators a one time SSH key to login to the customer equipment, along with providing detailed logging of everything the administrator does for later review.

- Installed and configured client-side SSL certificate validation for Apache, requiring

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

that machines connecting to UNIXOps provide a valid SSL certificate before being granted any access.

- Developed the code that would follow the workflow of Datapipe: Administrator requests access, UNIXOps configures the access on the client machine, administrator uses that access, and the access is revoked when used or 15 minutes have passed without it being used.

SAP to Tyler Conversion

| | |
|-----------------|--|
| Period | 2011 |
| Company | 6th Avenue Electronics |
| Tools | AutoIt3 , CentOS Linux, Python |
| Platform | Server: CentOS Linux, Client: Windows |

6th Avenue Electronics found that [SAP](#) was not a workable solution for them. The decision was made to switch back to the [Tyler](#) POS system, clearing out old mistakes and improving maintainability. I managed the technical aspects of the migration, while my immediate managers handled the business aspects.

Due to the costs associated with [SAP](#), we had just over three months, in total, to complete the transition. We were successful.

- Wrote several one-off scripts to check data that was sent in various Excel spreadsheets. Validate that all entries in column A of File 1/Sheet 1 are in Column C of File 2/Sheet 1.
- Used [AutoIt3](#) to automate the update of several items that could only be keyed into the client. No import existed at all. This reduced work from several hours down to an hour (including the initial script creation).
- Developed an automated installer that was used to handle installing all components (receipt printer, fonts, initial configuration) on every machine in the company.
- Worked with [Tyler](#) Retail Systems to configure the server properly.
- Developed snapshot backup strategy that reduces downtime for [Tyler](#) to mere minutes per night.

PyTyler - Tyler POS to PostgreSQL Migration Tool

| | |
|-----------------|--|
| Period | 2007, 2011 |
| Company | 6th Avenue Electronics |
| Tools | Python , PostgreSQL , Tyler POS System |
| Platform | HP-UX , Debian GNU/Linux |

[Tyler](#) is a point of sale system used by many smaller retail establishments. [Tyler](#) stores data in a set of proprietary ISAM files. These files do not have a modern access tool available (such as Crystal Reports) to perform reporting.

The users needed an easy way to report on the data, and this meant a tool was needed to copy the data from the on-disk files into a formal SQL server of some variety. In less than a month, I wrote a tool in [Python](#) to read the [Tyler](#) data files and load the information into a [PostgreSQL](#) database on a nightly basis.

This tool copied the entire database, comprising approximately 36,000,000 records, 140 tables, and 22 gigabytes of disk space. The program worked by reading the structure definition from the configuration files and recreating the structure in [PostgreSQL](#). PyTyler would then read each table, row by row, parse the data in the row, and load it into [PostgreSQL](#) server.

This allowed the users to use standard ODBC drivers to access and report on the data.

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

- Developed a tool to read configuration of ISAM files, and generate SQL “create table” statements mirroring the structure of the file.
- Created a specialized reader class which could read the data stored in the ISAM table.
- Developed small web server application to provide status page for administrators while migration tool runs
- Reduced total run time from 13 hours to 5 hours by converting the entire application into a multi-threaded application.
- Verified that data is being copied into the system correctly.
- [Tyler](#) POS system was in production until the closure of the company in December, 2011, so data copy ran every night to bring in updated data from previous day’s activities.

VMware Implementation

| | |
|-----------------|--|
| Period | 2005-2007 |
| Company | 6th Avenue Electronics |
| Tools | VMware Virtual Infrastructure 3, VMware Virtual Center |
| Platform | Linux (Various distributions), Windows Server 2003 |

6th Avenue Electronics, like many companies, had a growing need for individual servers for various internal services. They chose to implement [VMware](#) to reduce hardware costs, downtime, and environmental costs.

- Installed and configured iSCSI based SAN disks.
- Installed and configured all aspects of [VMware](#) Virtual Center and [VMware](#) Virtual Infrastructure 3.
- Developed (and tested) virtual machine templates to allow rapid deployment of new virtual servers using various operating systems ([Windows](#) XP, [Windows](#) 2003, [Debian](#) GNU/Linux, [RedHat](#) Linux).
- Monitored daily usage of [VMware](#) hosts.

SBN Implementation

| | |
|-----------------|--|
| Period | 2004-2005 |
| Company | Diversified Systems |
| Tools | SBN , Sybase 11.0 , PHP |
| Platform | Microsoft Windows 2000, Debian GNU/Linux |

[SBN](#), published by IBSoft, is an ERP system for the alarm industry. Diversified Systems is a subcontractor working in the low voltage electrical industry, including alarm systems, stereo systems, central intercom systems, structured wiring, and central vacuum systems. I implemented all aspects of [SBN](#) at Diversified Systems.

The provided client interface was unsuited for the intended use. This resulted in much in-house development to augment the [SBN](#) client with a web-based interface.

- Configured all aspects of [SBN](#) from base installation to full production mode, with active communication with users at each step.
- Implemented over 50 custom screens and reports using [PHP](#) on an [Apache](#) web server. This included easier access to customer searches, more usable technician schedules, and easier input for large quantities of data.
- Developed automated system for the sending of faxes to field technicians, saving over 5 work hours per day.
- Implemented an automated backup system for the database.

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

- Administered [Sybase](#) instance on day to day basis, resolving issues with full log files, etc.

SQL-Ledger Implementation

| | |
|-----------------|---|
| Period | 2005 |
| Company | Diversified Systems |
| Tools | Perl , Apache |
| Platform | Apache , Debian GNU/Linux |

The [SBN](#) accounting system was inadequate for the needs of Diversified Systems. This lead to the selection and installation of an external accounting package.

- Authored script to automatically migrate necessary data (customers, bills to be collected, etc.) from [SBN](#) to [SQL-Ledger](#).
- Installed and configured [SQL-Ledger](#).

KP-CIS

| | |
|-----------------|--|
| Period | 2001-2002 |
| Company | Ciber, Inc., contracted to IBM |
| Tools | Perl , Cygwin , GNU Make |
| Platform | Server: AIX , Client: Windows NT |

IBM was under contract to develop a complete clinical information system for Kaiser Permanente clinics. I participated as a member of the environment team, focusing on improving the build processes.

- Resolved issues with corrupted builds occurring weekly, resulting in savings of over 1000 work hours every week.
- Developed and improved approximately 450 compilation scripts and Makefiles on [AIX](#) and [Windows](#) NT/2000, fixing dependency issues and allowing reliable use of nightly code compilations.
- Evaluated, tested, integrated, and deployed new compilation tools.