

Michael J. Pedersen
42 Henry St
Oxford, NJ 07863
datacyclist@gmail.com
(908) 283-0318

Executive Summary

Manager with 5 years of experience 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. Most importantly, all of this has been done while keeping an eye on our budget, making sure to keep our costs down to keep the company profitable right through COVID-19.

Job History

Pulsepoint - Data Engineer (2015-2018), Director of Infrastructure for Data (2018-2023)

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

- 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 flows to new data center.
- Split data management team into data platform and data product development.
- Guided the team through splitting our ETL monolithic repository.
- Organized the migration of the ETL pipeline 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](#).
- Acted as scrum master for the team.
- 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.
- Onboarded new team members, helping them to fully integrate into the team.
- Held weekly 1 on 1 meetings with team members.
- Participated in on-call rotation.
- Developed new stories (including estimates) for our Jira board.
- Prioritized tickets for our Jira board.
- Passed annual HIPAA training for data protection.

OrcaTec, LLC - Developer

Atlanta, GA (Telecommute) - 2012-2014

- 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.

For more history going back to 1995, please visit my website at <https://www.icelus.org/>
Michael J. Pedersen datacyclist@gmail.com 908-283-0318

Relevant Technical Skills

Programming Skills

	Time Used	Last Used	Proficiency
Object-Oriented Design	10 years	2023	Very Good
Object-Oriented Programming	10 years	2023	Very Good
Refactoring	6 years	2023	Very Good

Database Skills

	Time Used	Last Used	Proficiency
PostgreSQL Database Administration	1 year	2011	Fair
Relational Schema Design	14 years	2023	Very Good
Structured Query Language (SQL)	14 years	2023	Very Good

Big Data

	Time Used	Last Used	Proficiency
HDFS	9 years	2023	Very Good
Hive	9 years	2023	Good
YARN	9 years	2023	Good
Alluxio	3 years	2023	Good
Impala	9 years	2023	Fair
Trino	3 years	2023	Good
Kafka	9 years	2023	Very Good
Kubernetes	4 years	2023	Good

Education

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

Project History

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.
- 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.