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

Executive Summary

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.

Relevant 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.
- Developed new ETL jobs to aggregate data from Pulsepoint's RTB exchange.
- Ingested third party data to make it available internally.
- Guided the team through splitting our ETL monolithic repository.
- Organized the migration of the ETL pipeline from Python 2 to Python 3.
- Optimized Hadoop jobs.
- Implemented data duplication between two Hadoop clusters.
- Participated in on-call rotation.

OrcaTec, LLC - Developer

Atlanta, GA (Telecommute) - 2012-2014
- Developed advanced search tool using Python, TurboGears, and jQuery.
- Created new document production framework from scratch.
- 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.
- Wrote Python framework to manage long running background jobs.
- Reduced multi-hour SQLAlchemy bulk database jobs to minutes.
- Found major security hole (remote code execution) and closed it.
- 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.
- 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/

Relevant Technical Skills

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

Programming and Scripting Languages

	Time Used	Last Used	**Proficiency**
Bash	10 years	2014	Good
C/C++	12 years	2009	Good
Java	2 years	2021	Fair
Javascript	3 years	2021	Good
Perl	6 years	2012	Fair
PHP	2 years	2012	Fair
Python	15 years	2023	Excellent

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

Software Configuration Management Tools

	Time Used	Last Used	**Proficiency**
Git	11 years	2023	Good
Mercurial	4 years	2014	Fair
Subversion	2 years	2010	Fair

Education

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

Project History

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 substitue for Sqoop.
- Updated our ETL pipelines to use FreeBCP in place of Sqoop.

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.

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