

Samuel Lakes

	Data Engineer
	San Francisco Bay Area
	☐ (513) 519-8753
	shpam@mykolab.com
	https://shpam.github.io/slakes.github.io/
	☐ English
BACKGROUND	
DACI	CGROUND
	ABOUT
	Samuel is originally from Loveland, Ohio. He has earned a degree in Computer Science from California Polytechnic State University. Before starting as a Data Engineer at Intuit, he interned with them for two summers beforehand where he worked on similar problems compared to his full time work. Before that, he was a full stack intern for two years at a startup called Digital Democracy located in San Luis Obispo. While his work focuses mainly in data engineering and cloud computing, his non-work interests range widely, everything from art and music to anthropology, comparative religion, and hiking.
	WORK EXPERIENCE
	Data Engineer, Intuit

Jan, 2019 - Aug, 2020 ☐ 1 year 7 months

As a Data Engineer designed and implemented various data pipelines and python tools to enhance the capabilities of Intuit's data platform while working closely with non-technical teams to bring new insights to data reports.

 Designed, implemented, and tested an automated ETL data pipeline to transfer commissions data between a 3rd party and Intuit's business teams using Hive, Spark, and EMR. Worked within the confines of constantly changing requirements for the on-time transformation of what had previously been an

- antiquated manual process misplacing funds into a well-oiled, mistake free data pipeline and warehousing system with CI/CD; creating new opportunities for revenue and accounting for lost funds.
- Managed, developed, and implemented an auditing framework using AWS S3/EMR and Informatica to comply with CCPA standards to delete user data. Provided confidence to stakeholders for compliance with state and international regulations.
- Designed and implemented a tool to automatically evolve database schemas using the Hadoop framework;
 reducing time to deploy table changes from over one hour to mere minutes without operator error or
 mistakes.
- Collaborated with teams in Finance/HR/Sales to understand their needs and update tables/datamarts to reflect changing requirements, unlocking key insights for their projects and making new data accessible in a timely manner.
- Managed and developed a data parity tool in AWS Athena in order to compare recently ingested tables against data sources, establishing a reporting tool surrounding any ingestion issues which reduced downstream data errors for teams of stakeholders.

Software Engineering Intern, Intuit

May, 2018 - Aug, 2018 □ 2 months

As a Software Engineering Intern assisted with the migration from an Oracle based data engineering platform to an AWS cloud platform.

 Designed and implemented a data parity tool to assist with the migration from Oracle databases to AWS in order to efficiently handle billions of records across many tables. It is used as a fast tool that can ensure data quality and reliability during incidents.

Software Engineering Intern, Intuit

May, 2017 - Nov, 2017 □ 5 months

As a Software Engineering Intern implemented and tested various aspects of Intuit's real-time data pipelines.

- Developed a component of Intuit's money movement system using the Spring framework transferring over \$1 billion every month.
- Implemented a piece of the team's data pipeline using Apache Storm that enriched data in Intuit's financial systems, creating new data for downstream stakeholders.

Software Engineering Intern, Digital Democracy

Oct, 2015 - Jun, 2017 🛛 1 year 7 months

As a Software Engineering Intern worked on various aspects of the front-end and back-end of the Digital Democracy website.

- Architected, managed, and developed a Stripe payment system in Drupal which provides users with a way to upgrade storage space, creating a revenue stream for the startup.
- Architected and managed the development of a semi-autonomous, BeautifulSoup/Selenium based testing system which reported errors with HTML elements of Digital Democracy's website to the QA team, reducing time to discover errors by several hours.
- Developed a data pipeline to handle the cleaning and enrichment of legislative hearing data for the purpose of fact-checking.

SKILLS

Professional Skills

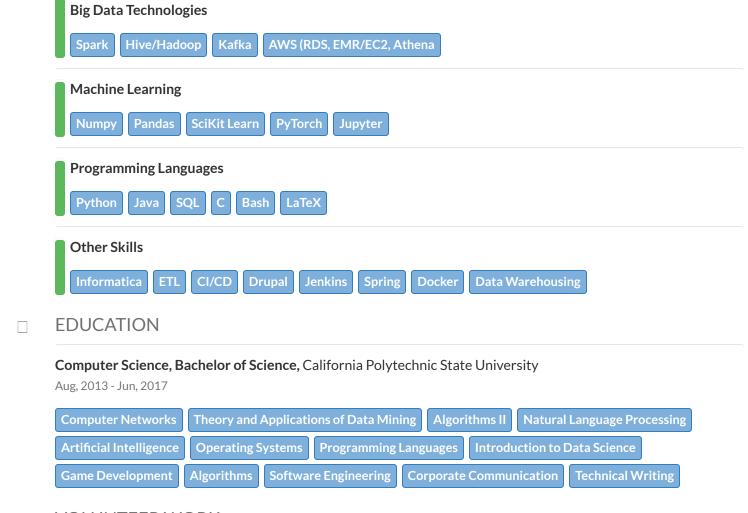
Project Management

Technical Writing

Presentations

Collaboration

Communication



VOLUNTEER WORK

Technical Director of Competitions, Cal Poly Data Science Club

Jan, 2012 - Jan, 2013

Cal Poly's student connection to the growing field of data science through supplementary education, competition, and corporate outreach.

- Planned several internal Data Science competitions to teach machine learning techniques to newer club members.
- Used knowledge of Pandas, Jupyter, and unsupervised learning techniques to help manage teams and give direction and advice in online competitions