

DAVID PAULINA

Lead Data Engineer | Hadoop PySpark Developer | AI/ML Specialist

Washington, DC | (202) 415-7906 | griot1academy@gmail.com

LinkedIn: [linkedin.com/in/david-paulina-a8b571a6](https://www.linkedin.com/in/david-paulina-a8b571a6) | GitHub: github.com/DavidPaulina

Professional Summary

Results-driven Senior Data Engineer with over 12 years of extensive experience designing, developing, and deploying robust big data processing systems, cloud architectures, and scalable enterprise solutions. Specialized in delivering impactful projects across diverse sectors including Financial Services, Educational Technology, and Banking. Proven expertise in leveraging PySpark, Hadoop, AWS cloud services, and Snowflake to build secure, high-performance data platforms. Skilled in integrating advanced AI and Machine Learning models to enhance analytics capabilities, support strategic decision-making, and optimize operational efficiency. Committed to implementing effective data management strategies and governance frameworks, ensuring compliance, data integrity, and fostering continual team innovation.

Technical Skills

- Big Data Technologies: Apache Spark, PySpark, Hadoop, Hive, Airflow, dbt
- Programming Languages: Python (Advanced), SQL (Advanced), JavaScript, Node.js, Bash
- Cloud Platforms: AWS (Lambda, S3, API Gateway, EC2, CloudWatch, EMR)
- Databases & Warehousing: Snowflake
- DevOps: Docker, CI/CD, GitHub Actions
- Data Analysis Tools: Pandas, NumPy, Jupyter Notebooks, Cloudera
- Development Tools: Git/GitHub, VS Code, Linux (Ubuntu, RHEL)

Professional Experience

Lead Data Engineer | Citecht, Washington, DC

June 2023 – Present

- Architected scalable real-time analytics platforms using PySpark and Hadoop
- Managed AWS infrastructure ensuring high availability and robust security
- Automated ETL workflows, significantly improving data ingestion efficiency
- Developed comprehensive data governance frameworks
- Integrated AI/ML solutions for enhanced analytics and decision-making
- Collaborated effectively with business stakeholders to meet strategic objectives
- Mentored junior engineers, promoting best practices and knowledge transfer
- Optimized performance through targeted pipeline tuning
- Documented data processes clearly to ensure operational consistency

- Implemented and maintained CI/CD pipelines to streamline software deployment

Senior Data Engineer | GRIOT Academy, Washington, DC

April 2021 – June 2023

- Developed ETL pipelines handling large educational datasets
- Automated processes reducing data workflow processing time by 40%
- Designed real-time analytics dashboards enhancing decision-making
- Built robust AWS infrastructure supporting scalable analytics
- Established data governance and quality assurance protocols
- Mentored junior team members and promoted professional development
- Optimized data processing for efficiency and scalability
- Created thorough technical documentation
- Worked closely with teams to ensure alignment with business goals
- Led continuous improvement initiatives and enhancements

ETL/ Cloud Data Engineer | Trust Bank, Washington, DC

December 2018 – April 2021

- Engineered secure and efficient financial data pipelines
- Successfully transitioned legacy systems to AWS cloud environments
- Established comprehensive data validation frameworks
- Ensured compliance with financial data regulations
- Automated monitoring and proactive alerting systems
- Optimized and continuously improved system performance
- Provided detailed documentation and training
- Coordinated cross-functional initiatives and communications
- Delivered ongoing technical support and enhancements
- Collaborated with compliance and regulatory stakeholders

Data Engineer | Citigroup, New York, NY

February 2017 – January 2018

- Developed modular ETL frameworks for financial data management
- Achieved performance improvement (35%) through optimization
- Implemented robust data cleansing and validation procedures
- Participated actively in agile team environments
- Enhanced data governance and compliance practices
- Supported real-time analytics and data ingestion
- Documented methods and delivered technical training
- Monitored systems and proactively addressed performance issues
- Facilitated internal team knowledge sharing
- Provided expert troubleshooting and swift issue resolution

Technology Director | GRIOT Academy, Washington, DC

January 2006 – April 2021

- Directed technology strategy and educational tool development
- Developed and managed web-based educational platforms
- Successfully managed virtual learning transitions
- Maintained infrastructure supporting extensive educational programs
- Established cybersecurity policies and compliance standards
- Led, supervised, and trained technology staff
- Conducted technology training sessions for faculty and students
- Managed budgets and resource allocations effectively
- Ensured compliance with educational technology standards
- Collaborated with academic stakeholders to enhance technology integration

Education

Master of Science in Artificial Intelligence

University of Colorado Boulder, Boulder, CO (Expected August 2025)

Certifications & Professional Development

- AWS Certified Machine Learning Specialty, Amazon Web Services
- Harvard CS50x Computer Science Foundations (In Progress)
- University of Colorado Boulder AI/ML MasterTrack Program (Applying 2025)
- AWS Skill Builder: Serverless, Lambda, IAM

Professional Interests

- Big Data Analytics & Distributed Computing
- Machine Learning & Artificial Intelligence
- Cloud-Native Architectures
- Educational Technology & STEM Outreach