

Overview

Highly experienced data engineer with over 7 years of experience in programming languages such as Python, big data technologies like Hadoop, Apache Spark, and Hive, and database technologies like SQL and database design. Proficient in data orchestration tools like Apache Airflow and cloud platforms like GCP. Expertise in designing, developing, and optimizing data pipelines on GCP and on-prem, as well as creating automation tools. Strong background in consulting on big data solutions and training customers on various data engineering and data science skills. Proven track record of leading data engineering projects and improving performance through analysis and optimization of queries, architectures, and algorithms. Committed to staying current with industry trends and technologies to provide value to any organization in the field of data engineering.

Key Skills

Programming: Python | Java

Big Data Technologies: Hadoop | Apache Spark | Hive | Oozie

Database: SQL | Database Design

Data Orchestration: Apache Airflow

Cloud: GCP | Azure

Infrastructure as Code: Terraform

Data Management and Integration: data build tool (dbt) | Informatica | Alteryx

Containerization and Version Control: Docker | Git

Data Science: Applied Machine Learning | Deep Learning | Natural Language Processing | Computer Vision

Others: APIs | Applied Mathematics & Statistics

Professional Experience

Lead Data Engineer | March 2021 – Present

Ford Motor Company | Dearborn, MI, United States of America

- Led the development of data engineering pipelines on GCP as one of the early adopters of the technology at Ford Motor Company
- Led the development of data engineering pipelines on-prem for many teams spanning across different domains such as Strategy and Enterprise Analytics, Manufacturing Analytics, and Quality
- Successfully designed and developed multiple data products
- Improved the performance of existing pipelines through analysis and optimization of queries, architectures, and algorithms
- Participated as an Apache Spark instructor in the 2021 DataTech Bootcamp
- Built Ford-specific utility tools to assist Data Engineers in automating their tasks
- Provided technical assistance to team members on code design and optimization.

Data Scientist/Engineer | December 2017 – Feb 2021

Cognitus | Paris, France

- Assisted companies, international clients, teams, and governments in adopting big data solutions through consulting on the design and deployment of big data ecosystems for storage, streaming, processing, and visualization using Hadoop ecosystem on-prem and cloud products
- Improved the speed and efficiency of legacy solutions by analyzing client requirements, designing, and building big data architectures and data processing pipelines to process large amounts of raw data in real-time and batch modes using SQL, Apache Spark, and Hive
- Designed architecture and led the development of a distributed, scalable, and fully extensible smart surveillance solution by coordinating with the development team and stakeholders to drive execution using Python, REST, Postgres, RabbitMQ, and Docker
- Spearheaded the development of 4 solutions in the computer vision and natural language processing domains using Python and Docker
- Trained 30+ customers on enterprise data science/engineering including courses on machine learning (scikit-learn and SparkML), deep learning (Keras), distributed data processing (pyspark), and programming (Java and Python) to enable companies to build their own data science teams.

Enterprise Solutions Architect | July 2020 – March 2021

Cogent Softech | Dearborn Heights, MI, United States of America | Remote Job

- Improved delivery speed and reduced errors by closing the communication gap between customers and the development team through thorough analysis of client requirements and clear documentation
- Researched and planned the development of 5 new solutions on fast-paced timelines by designing scalable and fault-tolerant cloud and on-premises architectures
- Ensured developers had the necessary skills by advising and planning learning paths and collaborating remotely using project management tools (Trello and JIRA) and communication channels (Slack and email) to guide project plans and updates.

Doctoral Student | February 2015 - November 2017

VEDECOM | Versailles, France

- Researched transportation multimodality for smart cities and organized and integrated data by implementing 3 software solutions using Semantic Web technologies in Python and Java
- Enhanced the ability to combine heterogeneous transportation data sources by proposing 3 integration solutions on the schema level, instance level, and service level
- Improved multimodal trip planning by extending the Connection Scan Algorithm to support non-scheduled services
- Published 8 articles in international conferences and journals through utilizing excellent communication and writing skills.

Research & Development Intern | February 2014 - August 2014

3iConsulting | Palaiseau, France

- Conducted research on the integration of supply chain management systems and business process management solutions
- Improved scalability and efficiency of legacy solutions by implementing research on supply chain and business process management solutions to adopt a service-oriented architecture using Java and REST.

Education

Doctorate in Computer Science, University of Paris Saclay, Versailles, France

Master's in Computer Science, Lebanese University - Faculty of Science, Beirut, Lebanon

Bachelor's in Computer Science, Lebanese University - Faculty of Science, Beirut, Lebanon

Certifications

[Data Engineering, Big Data, and Machine Learning on GCP Specialization](#), Coursera

[IBM Data Science Specialization](#), Coursera

[Data Scientist with Python Track](#), DataCamp

[Deep Learning Specialization](#), deeplearning.ai

[Natural Language Processing Specialization](#), deeplearning.ai

[Machine Learning](#), Stanford University on Coursera

[Associate Android Developer](#), Google Developer Certification