Robert L. Zacchigna

About - Robert-Zacchigna.github.io | Contact - Robert-Zacchigna.github.io | linkedin.com/in/robert-zacchigna

Seasoned engineer with 6+ years of experience designing/implementing machine learning, data pipelines and analytics solutions for financial services and fintech companies. Proficient in technologies like PySpark, Terraform, Git and SQL, with a focus on delivering data-driven insights to support financial monitoring and interpolating trends initiatives.

PROFESSIONAL EXPERIENCE

Capital One **May 2022 – Present**
Senior Assoc Software Engineer *Chicago, IL*

- Designed technical documentation for and wrote internal python ML libraries/utilities to help facilitate data orchestration using an AWS Fargate deployed Airflow platform, resulting in a +35% reduction in duplicated/unneeded code, +20% pipeline development efficiency and a vastly more maintainable platform.
- Rebuilt legacy Python application to utilize event-driven architecture (SNS, SQS, PySpark and Glue), in AWS, resulting in a far more scalable and maintainable application and a +15% decreased resource usage/cost among +200 users.
- Facilitated the solution architecture (IaC - Terraform) of new external partner data pipelines (Glue, PySpark, Fargate, SNS, SQS) using a micro-service architecture, as the designated SME within my team (including creating technical documentation). This architecture proved to be ~15% more resource efficient than other proposed solutions.

Discover Financial Services **June 2019 – May 2022**
Data Engineer *Riverwoods, IL*

- Troubleshooted, architected and implemented improved enterprise financial data pipelines leveraging PySpark, SQL, and Kafka resulting in a +20% performance increase in loading high-velocity financial data into Snowflake.
- Optimized deployed financial data models utilizing PySpark, Jenkins, and in-house modeling tools to on-premises servers, resulting in +15% lower resource cost/usage while adhering to stringent data governance and security standards.
- Collaborated, designed and implemented a containerized (OpenShift) multi-platform development application with cross-functional teams to address current internal development tool short-comings, resulting in +30% increase in development efficiency across +800 internal users.

EDUCATION

Bellevue University, Master of Science – Data Science

- Personal website containing my projects: https://robert-zacchigna.github.io/

Illinois State University, Bachelor of Information Technology – Systems Analyst/Development
Honors: *Cum laude*, National Society of Collegiate Scholars, Dean's List

TECHNICAL SKILLS

- **Data Science:** Modeling, Machine Learning, Tensorflow, FastAPI, PyTorch, Pandas, PySpark, Spark, NumPy
- **Data Engineering:** Analytics, Databricks, ETL, Airflow, Glue, Kafka, Orchestration, Data Warehouse/Lake/Processing
- **DevOps:** Docker, Kuernetes, Git, GitOps, Agile, Jenkins, CI/CD, Unit Testing, MLOps, Artifactory, Linux
- **Cloud:** AWS, Terraform, Fargate, EC2, ECS, EMR, Glue, RDS, S3, EFS, SNS, IAM, Lambda
- **Database:** NoSQL, MySQL, Postgres, SQLite, MongoDB, Aurora, DynamoDB, Snowflake
- **Programming:** Python, Scala/Java, SQL, Bash, Typescript/Javascript, REST API, PowerShell
- **Visualization:** Tableau, Plotly, Power BI, matplotlib, seaborn

CERTIFICATIONS AND PROJECTS

- AWS Certified Machine Learning – Specialty
- AWS Certified Solutions Architect – Associate

- **Python Etrade API Module (PyEtrade)**
  - Significantly contributed to an open source python Etrade API library enhancing code quality, unit-tests and cleanliness.
  - Realigned missing functions and parameters to be in line with API documentation specifications and adding several missing/undocumented API endpoints to the module.
  - Received praise and recognition from the original author for my significant code contributions (+200 Stars).
- **Malaria Cell Image Classification (FastAI and Keras/Tensorflow Comparison)**
  - Cross compared FastAI and Keras neural network models, for determining if an image (~7000 images) of a cell is infected with malaria or not, to see the differences between how the models are developed and performed.
  - Both models scored very well, with a +94% prediction accuracy.
- **Multi-Processing Media File Management CLI (MKVAudioSubsDefaulter)**
  - Created a Python CLI to more easily edit/organize tracks and manage metadata of Matroška (MKV) media files without having to remux (re-encode) the files.
  - Multi-processing allows this to be easily scalable and automatable to a large amount of files, processing over 1500 files in less than 4 mins (+375 files/s).