

Experience

Data Engineer Infor July 2018 – Present

- Owner and maintainer of a platform for running Spark applications on managed cloud services.
Keywords: Scala, Play Framework, Docker, ECR, ECS, EMR, Dataproc, ScalaTest, SBT.
- Developed a Spark application for transforming and partitioning input datasets into a target schema to make the bootstrapping of Infor products faster. Validation of the data can also be added to the pipeline in any step depending on the need. This application reduced the integration time by 60%.
Keywords: Scala, Spark, EMR/Dataproc, S3/GS, ScalaTest, ScalaCheck, spark-testing-base.
- Design data models and hierarchy structures to serve as the basis for a proprietary distributed OLTP system called LogicBlox and a BI tool called Birst, both written in Java.
- Develop, schedule and monitor data pipelines using Logicblox's workflow DSL.
- Improve system and batch performance by tuning the application design and optimizing queries.

Data Science Intern Infor Feb 2018 – June 2018

- Built a web scraper to collect data from a retailer's website using Scrapy framework.
- Performed EDA on the collected dataset using D3.js and plotly.
- Migrated the Inception-v3 model from a multi-class to a multi-label classifier.
- Retrained Inception-v3 on the collected dataset to generate a predictive model.
- Built a web-based recommendation system that allows users to upload an image and get the most similar ones stored in a MongoDB database using the multi-label model.

Software Engineering Intern EURANOVA July 2017 – August 2017

- Developed a web-based application using Flask and Angular 4 frameworks for video stream analysis to help marketing agencies gain insight about brands' presence in local channels.
- Dockerized the application and deployed it via Gitlab CI/CD pipeline.

Research Intern Microsoft June 2016 – July 2016

- Comparison of NoSQL document databases: MongoDB, CouchDB & DocumentDB.

Education

National Engineering School of Tunis (ENIT) 2015-2018

- National Engineering Degree in Computer Science and Software Engineering (**Top 1% of the cohort**).

Preparatory Institute for Engineering Studies - El-Manar (IPEIEM) 2013-2015

- Succeeded at the National Exam for entering Engineering Schools (**Top 14% nationwide**).

Skills

- Programming Languages: Scala, Bash, Python, Java.
- Database Management Systems: PostgreSQL, SQLite, MongoDB.
- Big Data Technologies: Spark, Delta Lake, Livy, Deequ, Toree, Hadoop.
- Cloud Services: AWS (EC2, S3, IAM, EMR, ECR, ECS), GCP (Cloud Engine, Cloud Storage, Dataproc, BigQuery).
- Version Control Systems: Git, Mercurial.

Courses & Certifications

- **Udacity** Data Engineering Nanodegree Program
- **Udacity** Intel Edge AI Scholarship Program
- **Coursera Specialization - Google Cloud** Data Engineering with GCP