

## Experience

---

### Data Engineer, Team Lead                      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 Spark applications to make the bootstrapping of Infor products faster and repeatable. The applications cover data transformation, schema and constraints validation, outlier detection and ML. The time-to-delivery dropped from 15-20 business days to 2-4, and the forecast tuning from 5-10 business days to several hours. Keywords: Scala, Spark, Delta Lake, EMR/Dataproc, S3/GS, Deequ, TransmogrifAI, 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.
- Develop, schedule and monitor data pipelines using Airflow on AWS MWAA.

### Data Science Intern                      Infor                      Feb 2018 – June 2018

- Built a web scraper to collect data from a retailer's website. Keywords: Python, Scrapy framework.
- Retrained Inception-v3 on the collected dataset to generate a multi-label predictive model.
- Built a web-based recommendation system that allows users to upload an image and get the most similar ones stored in a NoSQL database using the multi-label model. Keywords: Python, Angular, Flask, MongoDB.

### Software Engineering Intern                      EURANOVA                      July 2017 – August 2017

- Developed a web-based application for video stream analysis to help marketing agencies gain insight about brands' presence in local channels. Keywords: Python, Flask, Angular, PostgreSQL.
- Containerized the application and deployed it via Gitlab CI/CD pipeline. Keywords: Docker.

### Research Intern                      Microsoft                      June 2016 – July 2016

- Comparative research of NoSQL document databases. Keywords: Microsoft Azure, MongoDB, CouchDB, DocumentDB.
- Performed bulk imports and exports on each database and generated performance reports to serve as a deciding factor to migrate the existing MongoDB database to DocumentDB service on Azure.

## Education

---

### National Engineering School of Tunis (ENIT)                      2015-2018

- National Engineering Degree in Computer Science (**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 (Core, SQL, MLlib), Delta Lake, Livy, Deequ, Hue, Hive, Hadoop.
- Cloud Services: Amazon Web Services, Google Cloud Provider.
- Version Control Systems: Git, Mercurial.

## Courses & Certifications

---

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