

Salim El Awad

elawadsalim@gmail.com

(832)-547-1787

EXPERIENCE

Lead Data Engineer

Oct 2022 - Present

Enterprise Products Partners – Houston, TX

- Lead role in developing quant machine learning framework used by the data scientists to explore, train, and deploy their trading models.
- Design proper data pipelines and leverage correct data standards and technologies that reduce complexity, minimize failure points, and ensure scalability.
- Maintained and monitored data quality and integrity of the data pipelines.
- Automated data processing tasks to increase efficiency and accuracy of data.

Data Engineer

Feb 2019 – Oct 2022

Enterprise Products Partners (Contract) – Houston, TX

- Develop and maintain applications to average and interpolate real-time temperature data from sensors in the braised aluminums heat exchangers. This data is then fed into multiple ETL pipelines used for data modeling, ad-hoc analysis, data historians, and business intelligence dashboards.
- Work with the Data Science team to productionize data wrangling code in Python and operationalize Keras multilayer perceptron neural networks.
- Implement data flows and data transformations connecting OSI Pi, MapR, InfluxDB, and PowerBI using Python, Scala, Spark, Drill, StreamSets, Kafka, and SQL.
- Schedule all ETL jobs and near-time applications using a containerized Airflow solution. Define directed acyclical graphs (DAGs) for most of our workflows.
- Develop data-intensive applications with API's and streaming data pipelines.
- Assist data analysts and data scientists with query optimization, performance tuning, and data processing.
- Document and maintain source-to-target mappings and data lineage.

Software Data Engineer

Apr 2017 – Jan 2019

Sanchez Oil and Gas – Houston, TX

- Set up Airflow using PostgreSQL and Celery as our workflow scheduler. Defined directed acyclical graphs (DAGs) for many of our workflows.
- Built automation tool to ingest flowback emails in spreadsheet format which were downloaded, parsed, and written to their corresponding SQL tables.
- Created and dockerized a software solution to ingest data from SCADA systems. This data was sent to a distributed JSON database (MapR-DB) and produced to a real-time topic that was consumed by other applications for real-time analysis.
- Developed a real-time operation alert system by consuming data from multiple topics and writing to tables in a time-series database (InfluxDB). This data was read by microservices that would generate alerts if predefined conditions were met.
- Built batch ETL services that extracted public well data and Sanchez field data, transforming and storing the data for easy access by the Data Science and Operation teams within the company.

Teaching Assistant

Jan 2016 – Jan 2017

Computer Science Department - Houston, TX

- Taught Introduction to Computer Science and Computer Science and Programming.
- Used problem solving and critical thinking skills to create assignments, grade homework and exams.
- Strengthened communication and listening skills through teaching 25 students in weekly labs and holding office hours to assist students.

Software Engineer Intern

May 2016 – Aug 2016

NetIQ – Houston, TX

- Generated reports with JasperSoft for the Access Review Database using scrum methodology.
- Collaborated with Software, Data, and DevOps Engineers on developing automated tools that reduced the time spent updating reports manually.

Research Assistant**Jun 2016 – Sep 2016***Computer Science Department – Houston, TX*

- Researched classification of mixed-language usage by Hispanic populations in social media.
- Collected large amount of mixed-language data from various online sources using APIs and web-crawlers.
- Analyzed, tokenized, and classified large amounts of collected data.

EDUCATION**Major - Computer Science, BS****May 2015 – May 2017***University of Houston*

Graduated with summa cum laude (3.95/4 GPA)

AWARDS

University of Houston Computer Science Department Valedictorian

May 2017

University of Houston Dean's List (4)

Fall 2015 – Spring 2017**SKILLS****Programing Languages:** Python, Scala, Spark (Python and Scala), R**Operating Systems:** Linux (CentOS, Ubuntu, Raspbian), Unix (MacOS), Windows**ETL Experience:** Kafka, Airflow, Docker, Kubernetes, Apache Drill, GitLab CI/CD, Apache Sqoop, Streamsets**Databases:** InfluxDB, PostgreSQL, MongoDB, MapR-DB (similar to Hbase), Microsoft SQL Server, Sqlite3**Languages:** Bilingual in English and Spanish