
EXPERIENCE

Data Engineer

Enterprise Products

Feb 25th 2019 – Present

- 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 17th 2017 – Feb 1st 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.

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), Windows

ETL Experience: Kafka, Airflow, Docker, Kubernetes, Apache Drill, GitLab CI/CD, Sqoop, Streamsets

Databases: InfluxDB, PostgreSQL, MongoDB, MapR-DB, Microsoft SQL Server, GCP, DataProc, BigQuery

Languages: Bilingual in English and Spanish