

PROFESSIONAL EXPERIENCE:

The Vanguard Group

May 2019 - Present

Data Engineer

Master the World of ETF's (MWE) is a Project designed to build a data repository for everything about ETF's. Vanguard's global cash flow report is generated every month using this repository.

Responsibilities:

- Automated ETL process to fill MWE data lake from data vendors like FactSet and Bloomberg
- Reduced the expense of vendor data about 40% by implementing the raw-transform model
- Design of Configuration based PySpark Application which is cost effective and low maintenance
- Implemented vendor's third-party application on AWS EC2 instance after passing security requirements
- Active collaboration with business team for requirements gathering and business need
- Worked with various file formats like Parquet, CSV, JSON and compression formats like gzip
- Design and development of Tableau Dashboards with cash flow comparisons and summary
- Create technical documentation on confluence page for reference and reporting

Environment: Python, Pyspark, AWS EMR, AWS S3, Hive, SPARK SQL, Maven, Hue, Presto, Tableau, SQL, Linux, Control-m, Spring Tool Suite, Scrum Methodology

Fidelity Investments

Aug 2018 - Apr 2019

Data Engineer

Fidelity Investments uses its Data Analytics Platform to act as a Data Hub to all the fidelity labs startup projects

Responsibilities:

- Design, Development and Testing of Apache Spark based Applications used for Data ETL
- Salesforce, Google Analytics Integration using REST API's and the end points available in them
- Handling the installation, configuration, and integrating of any Big Data tools and frameworks required to provide requested capabilities
- Perform various Data oriented operations like data extraction, data validation, data cleansing, data modelling and data loading
- Internal Data Virtualization using Openlink virtuoso.
- Design and develop Tableau reports depending on requirement to perform visualizations, trends, summary and detailed reports
- Work in accordance with company's compliance policies to maintain Data Security and Privacy

Environment: HDFS, Spark, Hive, AWS EMR, AWS S3, Oozie, SPARK SQL, Maven, Scala, SQL, Linux, IntelliJ

Infosys, Hyderabad India

Jan 2016 - Nov 2016

Systems Engineer

Responsibilities:

- **Systems Engineer trainee:**
 - Trained in OOPS concepts, PL/SQL with python as major programming language
 - Finished course with top 2 percentile score.
- **Systems Engineer:**
 - Developed a POC Web application Using Python on Django Web Framework pattern
 - Application was intended to store the patient data and medical history for a hospital
 - Utilized PyUnit for testing the functionality of the application.

TECHNICAL SKILLS

- **Languages and Frameworks:** Python, Scala, Java.
- **AWS Services:** EMR, EC2, S3, IAM, Cloud Formation, RDS, Cost Explorer
- **Big Data Ecosystem:** Hue, Spark, Hive, HDFS
- **Databases:** PostgreSQL, Oracle, My SQL, Maria DB
- **Virtualization:** Openlink Virtuoso
- **CI/CD Tools:** SourceTree, Bitbucket, Jira, Confluence, jFrog Artifactory, Jenkins, Control-m
- **Visualization:** Tableau
- **File Formats:** Json, Parquet, Csv, Xml, Yaml
- **IDE's:** Spring Tool Suite, IntelliJ, Eclipse

EDUCATION:

Master of Computer Science, New Jersey Institute of Technology

2018

Bachelor of Engineering, CBIT, Hyderabad, India

2016