Samhitha Tatipalli
Big Data Engineer

thatipallisamhitha@gmail.com
8897558686

Experience Summary

Senior software engineer with 5.6 years of experience in Python, Pyspark and Cloud technologies.

- Currently working as a Data Engineer for client, Armorblox.
- Implemented solutions and pipelines for integrations. Work includes – Data extraction, Data cleansing and integrating with respective tools.
- Analytical and problem-solving skills, applied to Big Data domain Hadoop, AWS, Hive, EMR, Databricks.
- Implemented different Python modules like Pandas, boto3, sqlalchmy, and Selenium in various tasks.
- Implemented armorblox-sdk module for incident data.
- Proficient in writing unit testing code using unittest and integrating the testcode with the build process.
- Good knowledge of ETL, PostgreSQL databases with Python.
- Worked on Automated Ingestion Integration Framework using Python.
- Worked on application migration from Windows platform to Linux platform.
- Experience with Version Control with Bitbucket and GIT.
- Experience in working with continuous deployment using Jenkins.

EDUCATION QUALIFICATIONS

- □ B. TECH (Electronics and Communication.) from Padmasri Dr.B.V.Raju Institute of technology, Hyderabad affiliated to JNTU university in 2017.

TECHNICAL EXPERIENCE:

Languages:	Python, Pyspark, SQL
RDBMS:	PostgreSQL, Oracle
OPERATING SYSTEM:	Windows flavors, Linux flavors
TOOLS:	PyCharm, Jupyter Notebooks, Bitbucket, Jenkins
TECHNOLOGIES	Azure Sentinel, Cortex, Rapid7, Splunk, ServiceNow, Google Analytics.
CLOUD FRAMEWORK/SERVICES	AWS-Lambda, S3, Redshift, Azure- Azure Sentinel. GCP- GCS, Google cloud function.

Certifications:
- Certified Python Developer - vSkills
- Certified IBM Machine learning with Python - Coursera.
- Certified IBM Data analysis with Python – Coursera.

PROFESSIONAL EXPERIENCE: MAJOR ASSIGNMENTS:

Jan 2022 – till date	
Project:	Armorblox Connector
Client	Armorblox
Environment:	Windows: Python, Pandas, Boto3, Github, AWS – Lambda, S3. Azure- Azure Sentinel, GCP- Google cloud function, GCS.
Role:	Data Engineer Lead

Brief description of the project:

This project helps to protect organizations against data loss and email attacks.

Roles and Responsibilities:
- Involved in integrating Armorblox product with different tools as per the client requirements.
- Developed armorblox-sdk package for data regarding email attacks.
- Involved in figuring out the solutions for integrating tools like cortex, rapid7 and ServiceNow.
- Involved in data formatting and parsing the response.
- Developed pipeline for the integration process.
- Implemented lambda function for integrating the tools and stored data into S3 bucket.
- Implemented Google cloud function and stored the data into Google cloud storage using python.
- Involved in resolving complex issues in the project.
- Leading the team.

Sep'2020 – Dec'2021	
Project:	Consumer EDGE, POD and OP
Client:	Johnson & Johnson
Environment:	Linux-Windows: Python, Selenium Module, Pandas, Boto3 and Bitbucket, AWS Redshift
Role:	Data Engineer

Description of the Project:

In this project development for automation of end-to-end processing of Client POS data. Work related to downloading POS data from the client portal, data cleaning and validation and storing in the required format.

Roles and Responsibilities:

- Data extraction, Cleaning and Validation of POS from the Client Portal
- Develop headless selenium automation solution to download the POS data files
- Maintain environment specific configuration
- Integrate automation to automation framework.
- Involved in Build and Build Automation Tools (Jira, Jenkins) with CI/CD process.

Jun '2019– Aug'2020	
Project:	Bolt & Allergy Cast
Client	Johnson & Johnson
Technologies:	Python, GCP, Google Analytics, BigQuery, NLP
Role:	Data Engineer

Brief description of the project:

Bol t& Allergy Cast includes all products data related to skin in US and Canada regions. We extract data using APIs or from GA360 and develop applications using Python.

Roles and Responsibilities:
- Involved in data Extraction using REST APIs and data formatting, data cleansing for the applications.
- Was involved in testing and maintenance of the project.
- Developing scripts for data extraction from Google Analytics.
- Involved in data parsing and formatting after the extraction of data from Google Analytics.
- Worked on loading data into Big Query tables using python and monitoring the tableau reports.
- Involved in enhancing scripts for sentiment analysis using NLP techniques

Jun '2018– May'2019	
Project:	Linear Forecasting
Client	NBCU
Technologies:	Python, Pyspark, Amazon EC2 M4, 2xlarge instance, SQL.
Role:	Data Engineer

Brief description of the project:
This project was prediction of SC3 impressions (minutes) of programs. SC3 is strictly commercial minutes rating (average) of a live show +3 days from live.

Roles and Responsibilities:
- Understand the requirement and business functionality of the application.
- Developed data cleaning pipeline in Python.
- Data pre-processing techniques using Python.
- Engaged in data mining and preparing data for modeling purposes
- Involved in forecasting of data and feature engineering
- Extracted data from AWS (S3).
- Involved in writing SQL queries to extract data.