

SUBRAMANIAM SETHURATHINAM

Data Engineer- seeking roles in Data Engineering, Big Data Development and other Data specific domains. Have demonstrated history of working with vast variety of data tools and technologies.

Current Designation: Data Engineer

Total Experience: 3 Year(s) 8 Month(s)

Current Company: Vivriti Capital Pvt Ltd

Notice Period: Currently Serving Notice Period

Current Location: Chennai

Highest Degree: B.Tech/B.E. [Electrical]

Pref. Location: Hyderabad,Bengaluru / Bangalore,Chennai

Functional Area: IT Software - Application Programming / Maintenance

Role: Software Developer

Industry: IT-Software/Software Services

Marital Status: Single/unmarried

Key Skills: **Python,MySQL, Hadoop,Hive, Spark, SCALA, Sqoop, Cloud Computing,AWS, MongoDB, NoSQL**

Verified : Phone Number | Email - id

Last Active: 22-Jan-21

Last Modified: 22-Jan-21

Summary

Result-oriented Professional with over 3.5 years of experience in Software Development,Data Warehousing,MySQL,Python,AWS,Hadoop,Spark,Hive,PostgreSQL,Data Warehousing,Unix,MongoDB

Work Experience

Vivriti Capital Pvt Ltd as Data Engineer

Sep 2019 to Till Date

Building Backend Mechanism for Recommender Tool and Loss Modelling Tool (2019- ongoing)
Productionize the ML model created by data science team and enhance the code for better scaling and performance.

Roles and Responsibilities:

? The algorithms created by the DS team requires input from variety of sources. Our responsibility is to create a seamless data incoming and outgoing mechanism.

? We create data intensive web applications using AWS Elastic Beanstalk and the process will be triggered using SQS message or an api call.

? The calculated results will be stored in the Redshift which can be delivered further.

? Most of the data manipulation processes are done using ORM models in python itself and the Pandas library.

Building Data Pipelines (2019- ongoing)

As a product company, we receive lot of data from different sources. As a data engineer I wrote data pipelines using python and other AWS services.

Roles and Responsibilities:

? Data ingesting mechanism will be determined based on the source. Mostly we use Amazon Lambda or Amazon Elastic Beanstalk for reading data and do basic level of data cleaning and manipulation.

? The process will be usually triggered by SQS messages from the product team

? The cleaned data will be validated basically using PostgreSQL queries and will be pushed into the Amazon Redshift tables.

? Several derived tables need to be created for powering visualization tools such as Tableau. Data stored will also be used by other teams for their business purposes.

? The results will be pushed back to MongoDB collections using pymongo.

Scheduling and Monitoring workflows through Apache Airflow (2019- ongoing)

Scheduling all the batch workflows through Apache Airflow.

Roles and Responsibilities:

- ? Responsible for the data sync between various databases and redshift cluster. This job is done using inbuilt operators present in Airflow.
- ? Monitoring the workflow and fix the issues in case of any failures
- ? Writing custom processors in airflow based on the required use cases.

Aspire Systems India Pvt. Ltd. as Software Analyst May 2017 to Aug 2019

Recommender Tool (2018)

Backend application which uses machine learning and historical data to recommend products to users.

Roles and Responsibilities:

- * Did data analysis, preprocessing steps using pyspark and making spark dataframes out of it.
- * Applied machine learning algorithms specially for recommender systems such as collaborative filtering.
- * Formed user-user similarity matrices and frequency matrices for recommendation.
- * Finally storing the recommendations in HBase.
- * Build the mechanism for the continuously updating recommendation table.

Facebook Data Collector Application (2017-2018)

Python application which collects the data from Facebook API and stores it in Amazon S3 and the process is automated using Linux shell scripts and Cron jobs.

Roles and Responsibilities:

- * Written python program that will get the data from Facebook API for particular page.
- * Created python scripts that will parse the json data and the metrics are extracted.
- * Automated the execution of python script using Cron jobs.
- * Written shell scripts that will merge the data in Amazon S3.
- * Made a external hive table out of the data in S3.

Building NLP based Content Matching Tool (2019)

Our client is a leading service provider in education domain. They have a product that will basically act as a search engine for educational items. Our project is to improve the search results by matching the standards (formulated by government) and resources (available links) using Doc2Vec models.

Roles and Responsibilities:

- * Input data will be pushed to the SFTP server by the product team. We have a script that will read data and convert it into pandas dataframes.
 - * Data cleaning needs to be done as the data contains several anomalies and outliers.
- Doc2Vec model is built using pretrained glove vector. Each standard and resource will be converted into a 50-dimensional vector.
- * Similarity between two vectors will be calculated using cosine similarity (NumPy function).
 - * Top 100 resources for each standard will be extracted from the calculated score.
 - * The final result will be scored in the MongoDB and will be used by the product team to enhance their search results.
 - * A delta mechanism is built so that the scores will be revised every time a resource or standard gets added.

Education

UG: **B.Tech/B.E. (Electrical)** from **GOVERNMENT COLLEGE OF TECHNOLOGY** in **2017**

IT Skills

Skill Name	Version	Last Used	Experience
Python			
Pyspark			
Relational Database - MySQL			
NoSQL Database - MongoDB			
AWS Redshift			
Apache Airflow			
Unix			
Aws services			
JAVA			

Big Data Stack
Apache Nifi
MACHINE LEARNING

Languages Known

Language	Proficiency	Read	Write	Speak
English	Expert			

Affirmative Action

Category: General
Physically Challenged: No

Work Authorization

Job Type: Permanent
Employment Status: Full time