VEMIREDDY SREENATH REDDY

Sr Big Data Developer   Passport No: N8548163

● **6+ years of** IT experience in **Big Data with Hadoop**, **Spark** and its components like Spark core, Spark SQL, Hive, HDFS and **Azure** (Data Factory, Blob Storage, SQL Database)**.**

● Excellent enthusiastic to work with new technologies, analytical skills, communication, interpersonal and strong ability to perform as part of a team and individual.

Work History

2022-06 – Current

Consultant
U3 Info Tech Pvt Ltd, **Singapore**
- Working for a Banking client for risk model monitoring.

2020-07 – 2022-04

Consultant
Ernst & Young, Chennai, Tamil Nādu, India
- Worked with multiple clients for data migration and reporting projects by using Cloud and On-Prem environments.

2019-05 - 2020-07

Software Engineer
Experis IT Pvt Ltd, Chennai, Tamil Nādu, India
- As per the client requirements, I have designed and developed the frameworks for generate the analytical reports for daily, weekly, monthly, yearly.

2017-03 - 2019-01

Engineer
RMSI Pvt Ltd, Hyderabad, India
- *Worked as a Data Engineer and collaborated with various teams & managed to understand the requirements and gave my best work.*

Education

2011-06 - 2015-05

Bachelor of Technologies: Mechanical Engineering
Jawaharlal Nehru Technological University - Hyderabad

Contact

Address
08-41, The Inflora, Flora Drive, Singapore, 506849

Phone
+65 93763634
+91 8801469535(WhatsApp)

E-mail
vemisreenath@gmail.com

LinkedIn
www.linkedin.com/in/vemireddy-sreenath-reddy-09bb9

Skills

Hadoop, Spark, PySpark

Cloudera, Hortonworks

Python, Scala, Core Java, Hue, Superset, Hive, Impala, SQL, Shell, Pig, Sqoop, Airflow

AZURE (Data Factory, Blob Storage, SQL Database)

GitHub

PyCharm, IntelliJ, Eclipse, Jupyter Notebook

Projects

Project Title: MAS 637 Reporting   **2023-03 - Current**

MAS Reporting Description: MAS 637 reporting project is involved analyzing the status of different users, including AUTO, RML, CC, CL, and WM. This analysis focused on the pool IDs associated with each user and determined the status for the previous 12 months as well as the current month. The objective was to identify any significant changes or trends in the status of these pools and share the findings with the Singapore government's Monetary Authority of Singapore (MAS). By fulfilling this responsibility, I contributed to ensuring regulatory compliance and transparency in the financial sector.

Client: DBS

Technologies: ADA (Inhouse framework), Python, Spark, spark SQL, Presto, Hive, Hadoop, Jupyter, Airflow, Collibra.

Responsibilities:

- PySpark Code Developer: Developed PySpark scripts in ADA framework for analyzing pool statuses and identifying trends over 12 months.
- Data Analyst: Conducted comprehensive data analysis using PySpark to identify significant changes and patterns in pool statuses.
- Compliance Reporter: Generated accurate and compliant reports based on identified trends, ensuring adherence to regulatory guidelines.

Project Title: SAS Exit   **2022-06 - Current**

SAS Exit Description: SAS Exit is a project where we have migrated all the existing SAS scripts into DBS inhouse platform called ADA (Advanced DBS Analytics platform). The Project is all about calculating various reports, especially impact caused by the default accounts from various products which back offers to it retail or non-retail customers, such as Credit cards, Mortgage loans, Cash line, Auto loans etc...

Client: DBS

Technologies: SAS, ADA (Inhouse framework), Python, Spark, spark SQL, Presto, Hive, Hadoop, Airflow, Collibra.

Responsibilities:

- Requirement Analyst: Understood SAS scripts and prepared requirement documents to capture project specifications and objectives.
- Architect: Developed an architecture in ADA framework that effectively met the project requirements, ensuring seamless integration and efficient data processing.
- Metadata Designer: Designed and created metadata structures in Collibra to facilitate data governance and management.
- PySpark Conversion Specialist: Converted various SAS scripts into PySpark code, leveraging the capabilities of the ADA framework for enhanced data analysis.
- Airflow Scheduler: Created Directed Acyclic Graphs (DAGs) in Airflow to schedule and automate the execution of PySpark scripts, ensuring timely data processing.

Project Title: CRM (Customer Relationship Management) Data Migration   **2022-01 – 2022-04**

Description: Migrating the data from multiple sources to Azure cloud environment and generating the reports on SQL Database.

Client: ICICI

Technologies: Azure Data Factory, Blob Storage, SQL Database, Oracle, Key-Vault and Vertica.

Responsibilities:

- Created the ADF Pipeline to manage multiple sources and destinations.
- Migrating the data from Vertica On-premises and other environments to Cloud.
- Handling the team.

Project Title: H2H (Hana to Hadoop)   **2020-07 – 2021-12**

Description: H2H is a migration project from SAP-HANA to Hadoop environment. Need to migrate all possible hana reports into Hadoop with better approaches and as client's requests, hand overed the reports through mails, watch (UI web), different file types and hive tables.

Client: PayPal

Technologies: PySpark, HDFS, Hive, Python, GIT, Shell and Custom PayPal Frameworks.

Responsibilities:

- Prepared the business requirement, Design docs and Application docs.
- Converted the SAP-HANA reports into Hadoop system design and analysis.
- Involved in the Data loading and pre-processing the data using hive and spark.
- Contributed the data move to the mongo and watch reporting UI.
- Developed most of the reports using spark for better performance and approaches.

Project Title: **IAP (Integrated Analytics Platform)**   **2019-05 - 2020-07**

Description: Integrated Analytics Platform (IAP) is an analytical project to understand the patient's thoughts in the organization. As an initial task, we have taken the feeds from different applications of data from source systems to understand the data and ingested the same into the Hadoop environment. This project holds the 360-degree view of the patient activities, and we are processing the events of patients into hive tables and analyzing the same.

Client: AbbVie

Technologies: Apache Spark, Hadoop, Hive, Scala, Impala, Shell, Autosys

Roles: Big Data Engineer

Responsibilities:

- Involved in Workflow and pipeline design.
- Ingestion of all source application data into IAP applications.
- Managed the different kinds of data as a data engineer.
- Involved in Spark application development.
- Written the hive and impala queries for transforming and analyzing the data.
- Developed Shell Scripts for handling spark jobs.
- Scheduled the applications using Autosys, so created JIL files for that.
- Managed all kind of report requirements using PySpark.

Project Title: SDC (Spatial Data Collaborator)   **2017-03 -2019-01**

Description: This project is handles with multiple GIS data source like USGS, Tiger Data (US Govt) etc. As an initial task we need to collect the data from multiple sources to Hadoop environment. Finally merge the data to map development phase.

Client: Apple

Technologies: HDFS, Pig, Hive, HBase, Java, GIT, Shell and Custom apple Software's and Frameworks.

Responsibilities:

- Developed the scripts to read the data from multiple sources and write into Hadoop environment.
- Involved in data movement from SFTP servers to Hadoop env.
- Involved in ingestion data and merge data into map development phase.

Technologies: HDFS, Pig, Hive, HBase, Java, GIT, Shell and Custom apple Software's and Frameworks.

Responsibilities:

- Developed the scripts to read the data from multiple sources and write into Hadoop environment.
- Involved in data movement from SFTP servers to Hadoop env.
- Involved in ingestion data and merge data into map development phase.