**SREENATH REDDY VEMIREDDY**

Senior Big Data Engineer

**Phone:** +1 (469) 793-3883

**Email:** [vemisreenathreddy@gmail.com](mailto:vemisreenathreddy@gmail.com)**LinkedIn:** <https://www.linkedin.com/in/sreenath-09bb91183>**Portfolio:** [vemireddysreenath.github.io](https://vemireddysreenath.github.io/)

**Professional Summary:**

Highly accomplished **Senior Big Data and Azure Data Engineer with** over **10 years of experience** in building, automating, and optimizing large-scale data engineering solutions. Adept at leveraging open-source Big Data technologies and cloud-native platforms like **Azure and AWS** to solve complex business challenges across **finance, insurance, and healthcare domains**. Proven track record of managing **end-to-end data migrations**, regulatory compliance pipelines, analytics platforms from scratch, **cloud integration** and **regulatory reporting.**

* **Extensive Data Pipeline Development:**  
  Built scalable, high-throughput data pipelines using **Hadoop, Spark**, **PySpark**, and **Hive** to support critical reporting and analytics use cases. Designed data models to handle structured and semi-structured data at scale.

Hands-on experience with **Cloudera** and **Hortonworks** distributions of Hadoop for enterprise-grade data processing, deployment, and cluster management.

* **Advanced Data Lake & Cloud Integration:**  
  Successfully migrated terabytes of data from on-prem and hybrid sources (e.g., Oracle, Vertica, S3) into **Azure Data Lake Storage and SQL databases**, enhancing data availability and performance.
* **Metadata-Driven Automation & Frameworks:**  
  Designed and implemented custom **metadata-driven Python frameworks** for schema validation, lineage tracking, alerting, and audit logging—reducing manual effort and operational overhead.
* **Regulatory Reporting Expertise:**  
  Delivered data solutions for **MAS 637 and PILLAR3** reporting using PySpark, ADA Framework, and Airflow, ensuring strict compliance with regulatory bodies like the Monetary Authority of Singapore.
* **ETL Workflow Orchestration:**  
  Expertise in designing complex DAGs using **Apache Airflow** and Zena for ETL orchestration. Automated failure alerts, error handling, retries, and monitoring across the data lifecycle.
* **Cloud Security and Governance:**  
  Integrated with **Azure Key Vault** for secure access to credentials and secrets. Defined access controls and lineage using **Collibra**, ensuring data security and governance.
* **Cross-functional Collaboration:**  
  Worked with **compliance teams, analysts, QA engineers, and business stakeholders** to deliver reliable and actionable data products.  
  Contributed to key projects with clients such as **DBS Bank**, **Country Financial**, **PayPal**, **ICICI Bank**, **AbbVie**, **Apple**, and **Nokia**.
* **Performance Optimization & Troubleshooting:**  
  Continuously monitored and optimized **Spark jobs**, **SQL queries**, and **ADF pipelines** for better performance. Diagnosed bottlenecks and applied best practices for code and system-level improvements.
* **Agile & DevOps Methodologies:**  
  Familiar with **Agile** delivery, version control using **Git**, CI/CD pipelines using Azure **DevOps**, and automated testing in Big Data environments.
* **Mentoring and Leadership:**  
  Guided junior developers and data engineers by providing architectural guidance, code reviews, and training. Led small teams in project planning, task allocation, and delivery assurance.
* **Domain Versatility:**  
  Applied big data solutions across multiple industries:
  + **Finance:** Risk analysis, customer profiling, compliance reporting
  + **Healthcare:** Patient data unification and insights
  + **Insurance:** Policy and claims data transformation
  + **Banking:** Regulatory audit pipelines and trend analysis
  + **Geospatial:** Map data ingestion and visualization pipeline support

### **Technical Skills**

* **Big Data Technologies:** Hadoop, Apache Spark, PySpark, Hive, HDFS, Impala, Pig, Sqoop, Cloudera, Hortonworks
* **Cloud Platforms:** Microsoft Azure (Azure Data Factory, Azure Blob Storage, Azure SQL Database, Azure Key Vault), AWS (S3)
* **Data Analytics & BI Tools:** Adobe Analytics, Power BI, Apache Superset, SAS, Collibra, Watch UI
* **Programming & Scripting Languages:** Python, Scala, Java, Shell Script, SQL
* **Orchestration & Workflow Tools:** Apache Airflow, Zena, Autosys (JIL scripting)
* **Development Tools & IDEs:** Git, IntelliJ, Eclipse, PyCharm, Jupyter Notebook, Azure DevOps
* **Databases:** MySQL, Oracle, SQL Server, Vertica, SAP HANA, MongoDB
* **Frameworks (Internal / Custom):** ADA Framework, Metadata-Driven Python Frameworks
* **Other Skills:** Data Modeling, ETL Design, Data Migration, Regulatory Reporting, Metadata Management, Data Governance, Performance Optimization, Agile Delivery

**Professional Experience**

**Hadoop Lead | FirstMeridian Business Services Ltd (Client: Country Financial), India**  
*Jan 2024 – Mar 2025*

**Project:** Comm-Agg End-to-End Data Migration Solution  
**Description:** This project involved creating a robust and efficient system for migrating data from Guidewire S3 to Azure SQL using Blob Storage. The solution applies Slowly Changing Dimension (SCD2) methodology to ensure accurate historical data tracking. Additionally, the process includes developing automated, metadata-driven data migration pipelines for incremental loads, schema management, and the generation of comparison reports to assure data quality.

* Developed and optimized data migration pipelines using Azure Data Factory (ADF), ensuring seamless integration between Guidewire S3 and Azure SQL.
* Implemented SCD2 methodology to manage historical changes in data, allowing for precise tracking of data over time.
* Designed and created automated incremental data loads, reducing manual efforts and improving data accuracy.
* Applied metadata-driven automation for schema management, streamlining the maintenance of database schemas.
* Built and executed Python scripts for error handling, triggering email notifications, and generating comparison reports for data validation between ODS and DL Hive tables.
* Orchestrated data transfer from Azure SQL to Hadoop, improving the organization’s ability to analyze and store large-scale datasets.
* Led data processing activities, including copying data to landing zones, core zones, and active views in Hadoop.
* Automated remote execution of pipelines and stored procedures using Zena, improving system automation and reducing operational overhead.

**Technologies Used:** Azure Data Factory (ADF), Blob Storage, Azure SQL, Python, Hadoop, Hive, Impala, Zena, Shell Scripting

**Consultant (Big Data Engineer) | U3 InfoTech Pte Ltd (Client: DBS Bank), Singapore**  
*Jun 2022 – Jan 2024*

**Project:** MAS 637 and PILLAR3 Reporting  
**Description:** The MAS 637 project involved generating reports for the Monetary Authority of Singapore (MAS) to ensure regulatory compliance. This project analyzed pool statuses across various user types and tracked the changes over a 12-month period. PILLAR3 reporting focused on risk management, using predictive analytics to evaluate user behavior and default risks based on historical data.

* Developed and optimized PySpark scripts within the ADA (Advanced DBS Analytics) framework to analyze user pool statuses across different time periods.
* Automated trend analysis and default risk evaluations for users, improving regulatory compliance and decision-making.
* Provided insights into financial transactions and default status for users by integrating various data sources within the bank’s system.
* Delivered high-quality, compliant reports to the Monetary Authority of Singapore, ensuring that data and analysis met regulatory standards.
* Applied predictive analytics techniques to evaluate potential default risks, improving the bank's risk management strategies.
* Enhanced the reporting process by implementing Presto for faster data retrieval and analysis.
* Created and managed metadata structures within Collibra to support proper data governance, lineage tracking, and regulatory compliance.

**Technologies Used:** ADA (In-house framework), PySpark, Spark SQL, Presto, Hive, Hadoop, Airflow, Jupyter, Collibra

**Project:** SAS Exit  
**Description:** The SAS Exit project was designed to migrate legacy SAS scripts into DBS Bank’s ADA platform. The migration involved converting complex SAS reports into PySpark-based solutions, improving the performance and scalability of the reporting system.

* Led the migration of complex SAS reporting scripts to PySpark within the ADA framework, significantly improving the performance of data processing.
* Designed a robust architecture for the ADA platform that met the requirements of the SAS reports and ensured seamless integration with existing systems.
* Developed and maintained metadata structures in Collibra, improving data governance and ensuring compliance.
* Automated the execution of PySpark jobs using Apache Airflow, ensuring timely report generation and reducing manual intervention.
* Collaborated with key stakeholders to understand business requirements and ensure the migration of reports aligned with their expectations.

**Technologies Used:** SAS, ADA (In-house framework), PySpark, Airflow, Collibra, Python, Spark SQL, Presto, Hive, Hadoop

**Consultant – Big Data Migration | Ernst & Young (Clients: PayPal & ICICI), Chennai, India**  
*Jul 2020 – Jun 2022*

**Project:** CRM Data Migration  
**Description:** This project involved migrating CRM data from multiple on-premises and cloud environments to an Azure-based platform. The goal was to ensure smooth data transfer while generating insightful reports for the business.

* Led the migration of CRM data to Azure SQL using Azure Data Factory (ADF), ensuring reliable and efficient data processing.
* Developed scalable ADF pipelines to handle data ingestion from multiple sources, including Vertica and Oracle databases.
* Managed a team of junior engineers, providing guidance on best practices for cloud migration and data governance.
* Automated the scheduling of data transfer tasks using Azure Logic Apps, improving overall efficiency and reducing delays in data processing.

**Technologies Used:** Azure Data Factory (ADF), Blob Storage, SQL Database, Oracle, Key-Vault, Vertica, Azure Logic Apps

**Project:** H2H (HANA to Hadoop) | PayPal  
**Description:** The H2H project involved migrating SAP HANA reports to a Hadoop-based environment, ensuring improved performance and scalability for reporting and analytics.

* Led the design and migration of SAP HANA reports to the Hadoop ecosystem, focusing on Spark and Hive for data processing.
* Developed Spark and Hive-based reports to improve data accessibility and reporting performance.
* Utilized Hadoop, HDFS, and MongoDB to store and manage large volumes of data, improving data storage and processing efficiency.
* Collaborated with business teams to deliver reports through various channels, including email, web UI, and dashboards.
* Optimized the data migration process to ensure accurate and timely delivery of reports to end-users.

**Technologies Used:** PySpark, HDFS, Hive, MongoDB, GIT, Shell, Custom PayPal Frameworks

**Software Engineer – Data Analytics | ExperisIT Pvt Ltd (Client: AbbVie), Chennai, India**  
*May 2019 – Jul 2020*

**Project:** IAP (Integrated Analytics Platform)  
**Description:** The IAP project was designed to centralize patient data from various applications into Hadoop for analysis. The goal was to process and analyze patient activity data, allowing healthcare providers to make informed decisions.

* Ingested data from various source applications into Hadoop, ensuring seamless integration and data processing.
* Developed Spark applications to analyze patient activity data, providing a comprehensive view of patient behavior and outcomes.
* Built efficient Hive and Impala queries to transform and analyze the data, enabling the creation of actionable insights for healthcare providers.
* Automated batch data processing using Autosys, ensuring timely updates and reducing manual interventions.
* Managed the end-to-end pipeline for data ingestion, processing, and reporting.

**Technologies Used:** Apache Spark, Hadoop, Hive, Scala, Impala, Shell Scripting, Autosys

**Engineer – Geospatial Data Solutions | RMSI Pvt Ltd (Client: Apple), Hyderabad, India**  
*Mar 2017 – Apr 2019*

**Project:** SDC (Spatial Data Collaborator) | Apple  
**Description:** This project involved working with geospatial data sources to support the development of Apple’s map features. We processed and integrated data from multiple sources to create accurate and up-to-date maps.

* Ingested geospatial data from multiple sources (USGS, TIGER) into Hadoop for processing and integration into Apple’s mapping systems.
* Developed Hadoop-based pipelines to process large geospatial datasets, enabling more accurate and comprehensive map features.
* Worked with Apple’s mapping team to ensure the geospatial data was accurately merged, improving map development.
* Automated the data ingestion process from SFTP servers, reducing manual data processing time.

**Technologies Used:** Hadoop, Hive, HDFS, Python, Shell Scripting, MapReduce, USGS, TIGER Data Sources

**Trainee Engineer – Data Engineering | TriGeo Technologies Pvt Ltd, Hyderabad, India**  
*Jul 2015 – Mar 2017*

**Project:** Mapping Data Integration | Nokia  
**Description:** This project focused on integrating Nokia's geospatial data into Hadoop for large-scale analytics, enabling the creation of advanced mapping and navigation solutions.

* Assisted in the integration of Nokia's geospatial data into Hadoop for large-scale analysis.
* Worked on developing data ingestion pipelines to process geospatial data from various sources, improving data quality and consistency.
* Contributed to developing tools to automate the ingestion process, reducing manual effort and improving operational efficiency.

**Technologies Used:** Hadoop, Hive, HDFS, Python, Shell, Spark

**Education**:

**Bachelor's degree in engineering** *June 2011 – May 2015*  
Jawaharlal Nehru Technological University – Hyderabad