

# Vishnu Vardhan Mandula

## Data Engineer

[vishnuvardhanmandula@gmail.com](mailto:vishnuvardhanmandula@gmail.com) | 551-328-5322 | [LinkedIn](#)

### PROFESSIONAL SUMMARY

With 4+ years of professional experience in Data Engineer, I am a dedicated and ambitious graduate eager to excel in the information technology field. I specialize in creating, developing and optimizing data plumbing systems and ETL processes. My skills extend to database design, optimization, and SQL proficiency. I am passionate about keeping up with the latest technologies and best practices to foster innovation and enhance project outcomes.

### TECHNICAL SKILLS

**Programming languages:** Python 3 (Pandas, NumPy & matplotlib), Java, SQL, Pyspark, Scala, Bash.

**Big Data:** Hadoop, Hiva, Elasticsearch, Redis, PostgreSQL, MongoDB, MySQL

**Distrusted systems:** Apache Spark, Databricks, Kubernetes, Kafka

**AWS/Azure cloud:** EC2, S3, EMR, Airflow, Lambda, Athena, Glue, IAM, Redshift, DynamoDB, CloudWatch, Sagemaker, Kinesis, Azure SQL Database, Azure Load Balancer, DevOps Tool Integrations.

**Other:** Docker, Git, Kibana, Flask, PyTorch, Salesforce, Tableau, Power BI, Mata Trader 4, Jira, Terraform, Grafana.

### PROFESSIONAL EXPRIENCE

**Senior Data Engineer, Provident Bank**

June 2023 – present | Jersey City,NJ

- Built a **Data Lake** in Amazon S3 using **Python, R, and Pyspark** for the client, imported data from **Snowflake** tables using **CRM Postgres DB, Salesforce, MySQL Server, Amazon RDS**, and Integrated data from multiple sources.
- Implemented automation regression scripts for **ETL** processes across **AWS Redshift, MongoDB, and SQL Server**; contributed to a 15% increase in data reliability, benefiting over 100 business users relying on accurate analytics.
- Transformed vast sets of financial temporal data into actionable insights by leveraging **AWS S3, EMR, Athena, HDFS, Databricks**, and **Apache Spark**, resulting in a 30% increase in data processing efficiency.
- Increased storage capacity by 50% and reduced latency by 250% with **Amazon Aurora Databases and DynamoDB**, implementing **real-time** data processing solutions with **AWS Kinesis** and **AWS Lambda**.
- Optimized the performance of AWS-hosted applications using **CloudWatch** monitoring, which reduced error rates by 10%, and migrated the company's entire workload **to AWS cloud** using **EC2** and **S3** for efficient scaling, which resulted in 40% more efficiency.
- Evaluated performance of business requirements, performed data segmentation, integrated customer data into emails, enforced compliance approvals, and analyzed customer engagement using **Databricks, Snowflake, PySpark, and AWS S3**, enhancing customer insights and compliance adherence.
- Improved data pipeline efficiency by 30% by managing a cloud-based data manipulation pipeline using **SQL, Python, and DBT**. Optimized data transformation and integration across banking applications, leveraging **CloudFormation** and **Jenkins** for streamlined CI/CD deployment.
- Established robust data parsing protocols that transformed raw data into structured formats, supporting the analysis of over 100 datasets monthly and driving strategic initiatives within the organization.
- Created and optimized **SQL queries** for robust data reporting and visualization in **Looker**, enhancing the accuracy and accessibility of financial reports for stakeholders.
- Coordinated **OLTP** and **OLAP** processes by collaborating with cross-functional teams, achieving a 20% increase in data consistency, enhancing strategic planning, and accelerating decision-making capabilities across the organization by 25%
- Streamlined data access protocols within **MongoDB** and **HBase** structures for better performance and scalability, ensuring efficient handling of large datasets and high query throughput.
- Exported the analyzed data to the relational databases using **Amazon EMR** for visualization and generated reports using **Quick sight**. Utilized **Apache Airflow pipelines** to automate **DAGs**, their dependencies, and log files.

- Developed predictive models using regression in collaboration with the healthcare analytics team using **Python, AWS SageMaker, EC2, and S3** which improved analysis of total charges and length of stay for patients with COVID-19 and mental illness.
- Gathered medical records in the filing system which helped decrease outpatient wait time by 13.2%, adhering to **Agile** principles and delivering projects on time.
- Maintained data pipeline up-time of 99.8% while ingesting streaming and transactional data across 8 different primary data sources using **Spark, Redshift, S3, and Python**
- Managed on-premises data infrastructure including **Apache Kafka, Apache Spark, Elasticsearch, Redis, and MongoDB**, using **Docker** containers, which enhanced system reliability and performance.
- Constructed **MapReduce** jobs to validate, clean, and access data and worked with **Sqoop jobs** with incremental load to populate and load into **Hive** External tables, improving data accuracy and accessibility.
- Leveraged statistical modeling techniques including decision trees and generalized linear models (GLM) using SAS and **MATLAB** for predictive modeling in the insurance industry.
- Executed in-depth **data analysis** and transformation processes, achieving compliance with industry regulations while improving the accuracy of predictive models by 25%, and enhancing stakeholder decision-making capabilities.
- Utilized **data warehousing** techniques like **Star Schema, Snowflake Schema**, normalizing, denormalization, transformations, and aggregation to streamline data processes, improving data accessibility and analysis
- Collaborated with ten **cross-functional teams** to identify key data management requirements, resulting in the deployment of a system supporting over 100 data entry points, which improved operational scalability using **Oracle Database** for transactional data management
- Engineered scalable **NoSQL databases** using **MongoDB** and **HBase** to manage large volumes of unstructured data, enhancing data retrieval speeds and storage efficiency.
- Built user-friendly dashboards to analyze over 1M+ data points using **ggplot**, Python matplotlib, **Power BI**, and **Tableau** to analyze important features and model performance.
- Created a centralized repository using **Git** and **GitHub** for medical project data management, ensuring better collaboration and code management within the team.

**Software Engineer- Data Crawling & Analytics, Cliff.AI**

Dec 2019 – May 2020 | Hyd, India

- Led a project to streamline data extraction processes by organizing web scraping scripts using **Python** libraries, **Java, R, and SQL**, ensuring ethical data usage. Reduced manual data collection time by 25 hours monthly, which increased team productivity
- Transformed **raw data** into a usable format using **Pandas** and **NumPy**, improving the quality of datasets for further analysis. Automated Data Extraction, Transformation, and Loading operations using Java-based **ETL** processes, utilizing **J2EE** and **NoSQL** technologies, which reduced manual workload by 29% monthly
- Ingested data from disparate data sources using a combination of **SQL, Google Analytics API, and Salesforce API** using **Python** to create data views in BI tools like **Tableau**.
- Designed and maintained robust data pipelines for annuity computations and risk analysis using **SQL, JSON, and XML** formats.
- Automated data extraction and processing workflows using **VBA**, ensuring seamless integration of diverse data sources into predictive modeling frameworks.
- Led the implementation of **RESTful API** integrations for seamless data exchange between domain systems, leveraging **JSON** and **XML** formats for data transfer.
- Contributed to developing BI tooling solutions, integrating **Power BI** dashboards with existing data infrastructure, and using **Tableau** to create and maintain data visualizations.
- Collaborated with cross-functional teams, including Data scientists & Application developers to guide the development and implementation of Cloud applications, systems, and processes using **DevOps methodologies**.

**EDUCATION****Master's in computer science, Stevens Institute of Technology**

Aug 2022 – Dec 2023 | Hoboken, NJ

**Teaching Assistant** - Instructed students in the use of R, Python, SQL, Excel, and Tableau. Guide students in the analysis and modeling of their Confidential project datasets.

**Bachelor's in computer science, JNTUH**

Hyd, India