Satya S Vepada
Data Engineer
+19725917645 | vepadasatya01@gmail.com |linkedin.com/in/satya12

Summary

With 5 years of experience as a Data Engineer, specialize in data analysis, ETL, and data warehousing. have a proven track record in designing and implementing scalable data solutions using Hadoop and Spark. Proficient in ETL processes, data modeling, and cloud platforms like AWS and Azure. Can create end-to-end data pipelines for diverse needs. Expertise includes big data technologies, data quality, governance, and database optimization. Excel in collaborative environments, translating business requirements into effective data solutions, and staying current with emerging data technologies

---

TECHNICAL SKILLS

Programming: Python, SQL, Scala, Java, Unix/Shell Scripting

Big Data Technologies: Hadoop, Spark, Kafka, Databricks, Cloudera Distribution Hadoop, OOZIE, Airflow

Cloud and Database: Azure Data Factory, HDInsight, VMs, Blob Storage, Amazon EC2, S3, RDS, PostgreSQL, MS SQL Server, Oracle SQL

Big Data Technologies: Hadoop, MapReduce, Hive, Sqoop, Oozie, EMR, Spark, Pig, lambda functions, Zookeeper), Kafka, confluent Kafka,

Frameworks & Tools: NumPy, Scikit-learn, Pandas, Seaborn, Spark, Power BI, Tableau, Git, GitHub, MS Excel, MS Access, Spark, Airflow, ETL, Jira, Informatica power center, Waterfall, Agile/Scrum, Jenkins, Docker, Confluence.

Datastores and Cloud: Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), MySQL, Oracle, MongoDB, Teradata, PostgreSQL, Cassandra, Apache Kafka, Snowflake, Data Bricks, DynamoDB

Cloud Computing: AWS (S3, CloudWatch, Athena, RedShift, EMR, EC2, DynamoDB), Azure (Azure Data Factory, Azure Blob, Azure Databricks), IAM, Secret Manager, S3, Lambda, CloudWatch, Messaging Queue (SNS & SQS), Azure - ADF, Blob Storage, Blob Storage, Amazon EC2, S3, RDS, PostgreSQL, MS SQL Server, Oracle

IDE and Platforms: Pycharm, VS, Eclipse, IntelliJ, Jupyter Notebook, JIRA, Service Now, GitHub, Putty, Linux, Windows, Azure Devops, XZX, Jenkins, CI/CD

WORK EXPERIENCE

Application Programmer - Global business, Bank of America   **10/2023 – Present | Plano, TX**

* Increased successful communication of technical concepts to non-technical stakeholders by 30% through improved requirement analysis and documentation
* Ensuring efficient data integration using Sqoop and handling large structured datasets with Teradata parallel DBMS.
* Utilizing Hive for distributed querying and analysis on HDFS, and scheduling complex data pipelines with Autopsy.
* Creating insightful dashboards and reports with Tableau and ensuring GDPR compliance with an emphasis on PII data.
* Migrating Hive/SQL queries to Spark Data Frame-based transformations, resolving production issues, and collaborating with data science teams for data exploration and experiments using PySpark, HiveQL, and SQL.
* Playing a key role in the migration of Teradata objects to the Snowflake environment and enhancing codebase to meet evolving requirements.

Senior Software Engineer - Data Engineering, Sify - NBME   **07/2022- 08/2023 |Remote, PA**

* Designed and executed a complete data infrastructure migration to AWS cloud using IaC best practices.
* Built and automated data pipelines for data integration, processing, and analytics using tools like Glue, Spark, and Redshift.
* Developed expertise in transforming various data formats (PDF, XML, etc.) into analytical-ready structures using

Python, PySpark, and Spark SQL, achieving a 50% reduction in data processing time compared to previous methods.

* Implemented secure data sharing and access controls for seamless data collaboration.
* Led the creation of CI/CD pipelines for efficient and secure deployments across environments.
* Mentored junior engineers on data engineering best practices and cloud technologies.

Data Engineer | AT&T | USA   SEP 2021 - MAY 2022

* Designed and implemented data pipelines for ingestion, transformation, and loading across diverse data sources (SQL, NoSQL, REST APIs).
* Utilized cloud platforms (AWS, Azure) for data warehousing, big data analytics, and real-time data processing.
* Developed expertise in ETL/ELT processes using industry-standard tools (Apache NiFi, Talend, Airflow, Azure Data Factory).
* Built and automated CI/CD pipelines for efficient deployments.
* Experience with data wrangling, data cleaning, and data quality checks using Python, R, and SQL.
* Increased junior engineer productivity by 95% through mentorship and improved documentation..

Data Engineer | Magna Infotech | India   AUG 2018 - JUL 2019

* Designed and implemented high-performance data pipelines using Spark, Kafka, and Azure Data Factory.
* Migrated data infrastructure to Azure cloud platform, including data lakes, storage, and analytics services.
* Developed expertise in real-time data processing and machine learning pipelines, reducing data processing time by 30% and enabling real-time insights for improved decision-making.
* Automated data governance and data quality checks to ensure data accuracy and compliance.
* Collaborated with stakeholders to define data solutions aligned with business needs.
* Maintained comprehensive data pipeline documentation for knowledge transfer and ongoing optimization.

Quality Analyst | Amazon | India   FEB 2017 - JUL 2018

* Designed and deployed serverless applications and data pipelines on AWS using CloudFormation, Lambda, and DynamoDB.
* Led cloud migration projects to AWS, resulting in cost reduction and improved scalability.
* Implemented CI/CD pipelines for automated deployments and data quality checks, reducing deployment times by 99% and improving data quality by 98%
* Skilled in data analysis using Python, machine learning, and data visualization tools (Tableau).
* Maintained data warehouses and data models for efficient data storage and retrieval.
* Strong understanding of data security and compliance best practices.

EDUCATION

Master's in Industrial Engineering, Central Michigan University   08/2019 - 05/2021| Mt Pleasant, MI

Bachelor's in Industrial Engineering, Gitam University   08/2012 - 05/2016 | Vizag, India