SVSS SOBHIT VARADA

+1 (945) 527-5343 | svsssobhitvarada@gmail.com | Plano, TX, USA | linkedin.com/in/sobhit-varada-016899165/

PROFESSIONAL SUMMARY

Detail-oriented Data Engineer Intern with nearly 2 years of experience in data analysis and visualization, leveraging tools like Databricks and Tableau to drive impactful insights. I've successfully designed data-driven reports that improved reporting efficiency by 30% and optimized ETL workflows with Snowflake integration, ensuring robust data management. My work analyzing over 10 million records has enhanced decision-making by 20%, demonstrating my ability to translate complex data into actionable strategies.

SKILLS

Skills: Python, R, SQL, Apache Spark, Hadoop, MySQL, Postgres, Snowflake, Redshift, Power BI, Tableau, Excel/Numbers/Sheets, Microsoft Azure, AWS, Git, Data Analysis
Languages: Hindi, Tamil, Telugu

EDUCATION

University of Texas - Dallas **August 2023 - May 2025**
Master's, Information Systems *GPA: 3.58*

srm institute of science and technology **August 2018 - May 2022**
Bachelor's, Mechanical Engineering

PROFESSIONAL EXPERIENCE

innover solutions **Frisco, TX, USA**
Data Engineer Intern *May 2024 - Present*

- Enhanced reporting efficiency by 30% by designing dashboards with Python, SQL, and Tableau, using Databricks for ETL and integrating Snowflake as the database on Azure cloud technologies.
- Streamlined ETL workflow monitoring by deploying pipelines on Azure with Apache Spark and Databricks, using Airflow for orchestration and aligning with Agile sprints.
- Improved data integration by connecting Databricks, Snowflake, and GCP BigQuery using SQL and backend data structures for scalable analytics.
- Increased automation by scripting data pipeline management in Python on Azure and Linux/UNIX, participating in continuous integration with Jenkins.
- Supported scalable data processing by using Apache Spark for ETL and gaining exposure to Hadoop, Hive, and Pig on UNIX/Linux systems.

Cognizant **Hyderabad, AP, India**
Business Analyst *June 2021 - July 2023*

- Improved business decision-making by 20% by analyzing 10M+ records using SQL and Python in Agile sprints, building data workflows on Linux, UNIX, Azure, and GCP Dataproc.
- Reduced reporting time by 40% by building dashboards in Power BI and Tableau, integrating backend data from Hadoop, Hive, BigQuery, Airflow, Spark, and Pig.
- Ensured 99% data accuracy by validating data with Python, SQL, and HBase, and automating checks with Jenkins for continuous integration on Azure and GCP.
- Optimized operations by 15% by designing scalable data solutions with Java, C++, Apache Hadoop, Spark, and strong data structures knowledge.
- Supported secure data integration for mobile app and web services using HTML, Node.js, Linux/Unix, Hive, and BigQuery.

PROJECTS & OUTSIDE EXPERIENCE

Movie Recommendation System Analysis | Tableau, MySQL, R, Excel

- Ensured 98% data integrity in movie rating analysis by developing robust ETL pipelines with Python, SQL, and Unix/Linux commands, applying database management principles and Agile methodologies.
- Achieved 85% recommendation accuracy by applying data structures and algorithms in R and SQL to analyze collaborative filtering, integrating backend data visualizations with Tableau and HTML dashboards.

- Enhanced user insight and reporting capabilities by engineering web services and cloud technologies with Tableau and GCP BigQuery, orchestrating data workflows and optimizing queries for dynamic analytics.
- Streamlined workflow execution and monitoring by utilizing Apache Airflow to schedule data pipelines across Hadoop, Spark, and Dataproc environments on AWS and GCP, ensuring SLA adherence and leveraging Jenkins for continuous integration.
- Improved scalability and performance of analytics platform by optimizing query execution with Hive, Pig, and HBase on cloud-native platforms, applying JVM tuning and backend enhancements within an Agile team.

Data Analysis and Visualization|COVID-19 Impact Analysis|Tableau, Excel, R

- Enhanced data quality and reliability by utilizing Python, R, and SQL on UNIX/Linux systems to preprocess and structure a dataset of 50,000+ records, leveraging advanced data structures.
- Delivered actionable insights by developing Tableau dashboards integrated with backend data processed using Spark and Hive.
- Improved workflow automation by implementing Airflow on GCP Dataproc to automate ETL jobs across Hadoop and cloud technologies.
- Supported scalable data infrastructure by leveraging Apache Pig and HBase for processing and optimizing database performance with Jenkins.
- Increased query efficiency by optimizing BigQuery and Hive jobs using backend systems, JVM tuning, and cloud technologies in AWS and Azure.

Flight Booking Database | SQL, Database Design

- Increased database efficiency by 30% for airline booking by optimizing SQL queries and database design using Triggers, Views, and Indexes on Oracle backend.
- Streamlined job scheduling and monitoring by automating data orchestration with Airflow and integrating with cloud technologies such as Dataproc and GCP.
- Boosted query performance for analytics by implementing SQL best practices and leveraging Hive, Pig, and BigQuery on Hadoop clusters.
- Ensured secure and reliable deployments by supporting database operations on Linux and UNIX, and contributing to Jenkins-based continuous integration.
- Enabled scalable distributed data management by applying data structures, JVM-based technologies, and experience with Apache Hadoop, Spark, and HBase in backend design.

Analyzing Sales Data| Python

- Improved sales forecasting accuracy by 25% through analysis of large-scale sales datasets using Python and SQL, employing data cleansing, linear regression, and random forest regression, and leveraging Dataproc on GCP for scalable cloud processing.
- Enabled actionable business insights by developing data visualization pipelines with Matplotlib and Seaborn on Linux/UNIX systems, integrating outputs into backend web services for enhanced reporting.
- Streamlined data workflows and improved job monitoring efficiency by orchestrating ETL jobs with Airflow in an Agile environment, ensuring SLA adherence and iterative process enhancements.
- Enhanced database management and query optimization by designing and optimizing SQL queries for Hive, Pig, and BigQuery, improving data retrieval performance and integration with Hadoop clusters.
- Strengthened data security and continuous integration practices by managing credentials, implementing Jenkins for CI, and supporting deployments on Apache Spark and Hadoop clusters across AWS, GCP, and Azure.