# VINESH DIDDI
# DATA ENGINEER

**Location:** GA | **Phone:** *404* -668-1785 | **Email**: vineshdiddi0314@gmail.com

## SUMMARY


- Over 4+ years of experience in data analytics, demonstrating deep expertise in managing and analyzing diverse datasets.
- Mastery of programming languages such as Python, SQL, and R, providing robust solutions for data analysis and manipulation.
- Extensive experience with databases including MySQL, Oracle, MongoDB, and PostgreSQL, ensuring efficient handling of complex data structures.
- Proficient with AWS services like S3, IAM, EC2, CloudWatch, Lambda, and CloudFront, enabling effective cloud-based data management.
- Advanced skills in data visualization using Tableau, Power BI, and Excel, translating complex datasets into actionable insights for stakeholders.
- Strong background in data manipulation and statistical analysis with NumPy, Pandas, Matplotlib, and Scikit-Learn, facilitating high-level data analysis and predictive modeling.
- In-depth knowledge of big data technologies such as Hadoop, Apache Spark, MapReduce, Kafka, and Airflow, showcasing comprehensive understanding of data processing and workflow automation.
- Expertise in ETL tools like Informatica PowerCenter, Talend, and SSIS, adeptly managing data integration and transformation processes.


## TECHNICAL SKILLS

| | |
|---|---|
| **Programming Languages:** | Python, SQL, R |
| **Databases:** | MySQL, Oracle, MongoDB, PostgreSQL**.** |
| **Cloud Technologies:** | AWS (S3, IAM, EC2, CloudWatch, Lambda, Cloud Front). |
| **Visualization Tools:** | Tableau, Power BI, Excel. |
| **Libraries:** | NumPy, Pandas, Matplotlib, Scikit-Learn. |
| **Big Data Tools:** | Hadoop, Apache Spark, MapReduce, Kafka, Airflow. |
| **ETL Tools:** | Informatica PowerCenter, Talend, Informatica SSIS. |
| **Methodologies:** | Agile, Waterfall. |
| **Analytical Skills:** | Data Mining, Data Cleansing, Statistical Analysis, Data Visualization, Text Mining, ETL, Data Wrangling, Data Warehousing, Critical Thinking, Problem-solving, Communication. |
| **Version Control:** | Git, GitHub**.** |
| **IDE'S:** | PyCharm, Jupyter Notebook. |

## EDUCATION

**Kennesaw State University, Marietta, GA** **Jan 2022 – May 2023**
Master of Science in Computer Science

**St, Peters Engineering College, Telangana, India** **Jun 2015 – Jul 2019**
Bachelor of Technology in Computer Science

## PROFESSIONAL EXPERIENCE

**Data Engineer** **Aug 2023 – Current**
**BCBS, GA**

- Developed Python scripts for automating data ingestion, transformation, and loading processes, significantly reducing manual efforts, alongside conducting regular SQL-based database maintenance tasks to ensure data availability and reliability.
- Created and optimized complex SQL queries that enhanced data retrieval speeds by 30%, enabling more efficient data analysis and supporting timely decision-making across business units.
- Managed large datasets using Hadoop Distributed File System (HDFS) for efficient data handling, and utilized Spark's Data Frame API for structured data processing and SQL queries on distributed datasets.
- Implemented data quality checks using Talend that improved data accuracy by 35%, and designed ETL workflows with Informatica PowerCenter that reduced data redundancies and inconsistencies by 25%.
- Administered Oracle database instances for optimal performance, and managed AWS CloudWatch for monitoring data workflows, along with implementing security best practices using AWS IAM to safeguard data access.
- Orchestrated data storage solutions using Amazon S3 for high durability and availability, and leveraged Tableau for developing complex visualizations to articulate data insights effectively, utilizing features like data blending and advanced calculations.
- Developed ETL processes using SSIS (SQL Server Integration Services), resulting in a 30% reduction in data processing errors, measured by the number of error incidents or data discrepancies.
- Utilized Excel's extensive library of formulae and functions, resulting in a 25% improvement in data processing efficiency, measured by the time taken to complete data manipulation tasks.
- Collaborated within Agile (Scrum) teams to deliver data solutions iteratively, resulting in a 15% increase in project velocity, measured by the number of story points completed per sprint or project timeline adherence.
- Utilized Git for version control and collaborated with cross-functional teams using tools like Pycharm, resulting in a 20% reduction in code conflicts and integration issues, measured by the number of merge conflicts or build failures.

**Data Analyst** **Nov 2019 – Dec 2021**
**Dell Technologies, India**

- Utilized Python for extensive data manipulation and cleansing, applying statistical analysis to derive actionable insights that influenced

strategic decisions, resulting in a 20% increase in decision-making speed across key projects.
- Developed and fine-tuned complex SQL queries, leveraging indexing, window functions, and aggregation pipelines to significantly enhance performance and data retrieval efficiency.
- Utilized Matplotlib for detailed statistical analysis and graphical data visualization, aiding in the interpretation of trends and patterns.
- Designed advanced data models and integrated diverse data sources into Power BI for in-depth analysis and reporting, utilizing DAX and data cleaning techniques in Excel to improve data accuracy and consistency, leading to a 15% increase in reporting efficiency.
- Mastered advanced Excel functions (VLOOKUP, INDEX/MATCH, SUMIF) to build dynamic financial models and waterfall charts, enabling executives to track key performance indicators.
- Administered PostgreSQL databases, optimizing performance through detailed query analysis, indexing strategies, and parameter tuning to improve system responsiveness and efficiency, cutting down data processing times by 40%.
- Constructed data integration workflows using Informatica PowerCenter and SSIS, maintaining robust data flow between disparate systems and platforms.
- Implemented data governance and security measures using AWS services such as IAM, KMS, and CloudTrail, ensuring data privacy and compliance which reduced data breaches by 10% and met 35% of regulatory requirements ahead of deadlines.
- Adhered to Waterfall project management methodologies, ensuring structured and sequential project execution from conception to deployment, which improved project delivery timelines by 20%.

**Data Analyst** **Aug 2018 – Oct 2019**
**Cipla, India**
- Developed Python scripts to automate repetitive data processing tasks, enhancing productivity by reducing processing time by 20% and decreasing manual errors by 30% in data handling workflows.
- Enhanced SQL queries to fetch complex data from different tables in remote databases using joins, database links, and Bulk collects, optimizing data retrieval processes and enhancing efficiency in data analysis.
- Implemented data replication and backup strategies using MySQL's replication and backup tools to ensure data redundancy and disaster recovery.
- Utilized NumPy and Pandas to clean and preprocess data, reducing processing errors by 20% and executing extensive data cleaning, wrangling, and mining on datasets, which streamlined data processing workflows and cut data analysis turnaround time by 25%.
- Implemented AWS services, including Amazon S3, Redshift, and EMR, to enhance data storage, warehousing, and analysis, achieving a 20% reduction in data processing times and a 15% cost saving on data storage.
- Created interactive dashboards in Tableau using calculated fields and custom SQL, increasing user engagement by 20% and providing dynamic visualization experiences that improved decision-making.
- Employed Tableau and custom SQL features to develop dashboards and identify correlations within complex datasets, enhancing data-driven decision-making processes.
- Designed and developed weekly and monthly reports using advanced MS Excel techniques such as charts, graphs, and pivot tables, ensuring comprehensive data analysis and visualization.
- Conducted comprehensive data mining and cleansing, improving the accuracy and reliability of data sources by 25%, which supported critical business initiatives and enhanced the understanding of customer feedback through text mining.

## PROJECTS

### Online Shopping Cart:
- Designed an online shopping cart system using MySQL to manage user accounts, product inventory, shopping carts, and order processing. The system allows users to browse products, add items to their carts, place orders, and track their order history.
- Product Catalog: Displayed various products with details such as title, price, and availability.
- Shopping Cart: Allowed users to add products to their virtual shopping carts.