

SUMMARY

- Data Engineer & ETL Developer** with 3+ years of experience specializing in designing, implementing, and optimizing robust data infrastructure and pipelines.
- Proficient in programming languages such as R, Python, and SQL, with a focus on leveraging cutting-edge technologies including Hadoop, MapReduce, Hive, and Apache Spark to efficiently process and analyze large datasets.
- Extensive experience in cloud platforms, particularly AWS and Azure, ensuring seamless integration and scalability of data solutions.
- Skilled in utilizing data science libraries such as NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, and Seaborn to extract valuable insights and build machine learning models.
- Expertise in data visualization tools like Tableau, Power BI, and SSRS, enhancing reporting capabilities and providing actionable insights to stakeholders.
- Proficient in database management systems, including MySQL, ensuring optimal data storage, retrieval, and manipulation.
- Proven ability to collaborate and version control using Git, ensuring the integrity and traceability of code. Well-versed in Microsoft Office suite for efficient documentation and reporting.

SKILL

**Methodology:** SDLC, Agile, Waterfall  
**Programming Language:** R, Python, **SQLIDE's:** PyCharm, Jupyter Notebook  
**Big Data Ecosystem:** Hadoop, MapReduce, Hive, Apache Spark, Pig  
**ETL Tools:** SSIS  
**Cloud Technologies:** AWS, Azure  
**Packages:** NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow  
**Reporting Tools:** Tableau, Power BI, SSRS  
**Database:** MongoDB, MySQL  
**Other Tools:** Git, MS Office  
**Operating Systems:** Windows, Linux

EXPERIENCE

Verizon Communications, TX| Feb 2023 – Current | Data Engineer

- Uniting with cross-functional Agile teams to develop and deliver scalable solutions, employing Agile methodologies to enhance adaptability and responsiveness to changing project requirements.
- Utilize programming languages such as R, Python, and SQL to develop and maintain data pipelines, ensuring the seamless extraction, transformation, and loading (ETL) of diverse data sources.
- Create interactive and insightful visualizations using tools like Tableau, Power BI, and SSRS. Enhanced reporting capabilities, providing stakeholders with actionable insights.
- Implementing and maintaining data solutions on cloud platforms including AWS and Azure, ensuring scalability, reliability, and cost-effectiveness. Deployed and managed resources to optimize data workflows in cloud environments.
- Used MS Office suite for efficient documentation and reporting. Prepared detailed technical documentation to facilitate knowledge transfer and maintain data engineering standards.
- Define and track key performance metrics for data processing and ETL workflows. Implemented optimization strategies, resulting in a 15% improvement in data processing efficiency.

VVDN Technologies, India| June 2018 – Dec 2020| Data Engineer & ETL Developer

- Utilized Software Development Life Cycle (SDLC) like Waterfall to configure and develop processes and standard procedures.
- Designed and implemented robust data pipelines to extract, transform, and load (ETL) data from diverse sources using R, Python, and SQL, ensuring the seamless flow of information for analytics and reporting purposes.
- Leveraged data science libraries such as NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, and TensorFlow for advanced analytics, machine learning, and statistical modeling.
- Designed, developed, and maintained Extract, Transform, and Load (ETL) processes using tools such as SSIS, ensuring the timely and accurate flow of data across systems.
- Created visually appealing and informative dashboards using Tableau and Power BI, facilitating data-driven decision-making for stakeholders.
- Utilized MapReduce and Hive for efficient data processing and querying within the Hadoop ecosystem, contributing to a 20% reduction in data processing times.
- Managed and optimized MySQL databases for efficient data storage, retrieval, and manipulation. Ensured data integrity and performance through thorough monitoring and optimization techniques.
- Implemented version control using Git for codebase integrity. Collaborated with cross-functional teams to ensure alignment with project objectives and timelines.

EDUCATION

**Master of Science, Computer Science**  
The University of Texas at Arlington, TX  
**Bachelor of Science and Technology, Computer Engineering**  
Pandit Deendayal Petroleum University, IND

PROJECTS

Smart IoT Infrastructure

- Designed and implemented a Smart IoT Infrastructure project focused on simulating data from IoT devices.
- Leveraged Apache Kafka for real-time data ingestion, Apache Spark for processing and transformation, and Amazon S3 for cloud storage. Developed data simulation scripts, configured Apache Kafka for seamless data flow, and employed Apache Spark for cleaning and normalization.
- Integrated with Amazon S3 for scalable and reliable cloud-based storage. This project demonstrated expertise in IoT data management and data engineering, showcasing proficiency in cutting-edge technologies and cloud solutions.

Machine Learning Project using Deep Learning:

- Completed a Deep Learning project on Handwriting Recognition in Python by creating a custom model based on various research papers.
- Dealt with a large amount of data, with preprocessing taking a significant portion of the project time.

**Stock Data Analysis**

- Executed a Stock Data Analysis project by integrating and analyzing financial API data using Python libraries.
- Extracted, cleaned, and transformed stock price data to derive actionable insights for decision-making in financial markets.
- Leveraged statistical analysis and visualization techniques to identify trends and patterns, enhancing understanding of market behavior and informing investment strategies.

**Twitter Data Analysis**

- Conducted a Twitter Data Analysis project to extract, clean, and analyze data from the social media platform using Python libraries.
- Utilized sentiment analysis to gain insights into public opinion and trends related to a specific topic or event.
- Applied visualization techniques to present findings effectively, enabling stakeholders to make informed decisions based on social media insights.