Dinesh Reddy Kankanala

Phone: 980-303-1749 | Email: dineshreddyk569@gmail.com

PROFESSIONAL SUMMARY

- Over 3+ years of experience dealing with structured and unstructured data, proficiently managing tasks like Data Mining, Data Acquisition, Data Validation, Predictive Modeling, and Data Visualization. Proficient in Python ETL Development, crafting SQL queries, including multi-table joins, subqueries, and window functions, to extract, transform, and analyze complex datasets for actionable insights.
- Crafted SQL queries for Data Modeling and database design, including multi-table joins, subqueries, and window functions, to extract, transform, and analyze complex datasets for actionable insights.
- Employed Python scripts for meticulous data cleaning, incorporating regular expressions, custom functions, and statistical techniques to ensure data accuracy and maintain high data quality standards.
- Expertise in developing and managing end-to-end ETL pipelines, optimizing database performance, and automating reporting processes, resulting in a 20% increase in data-driven decision-making and a substantial reduction in manual reporting efforts.
- Proficiently engineered Python scripts for data processing and ETL (Extract, Transform, and Load) operations, optimizing data workflows for both efficiency and accuracy.
- Strong expertise in Model Evaluation and Statistics, enabling sophisticated data-driven insights and informed decision-making with a focus on data warehousing best practices.
- Proficient in advanced ML/Statistics Algorithms, encompassing Decision Trees, Random Forest, Predictive Modeling, Regression, Classification, Clustering and Time Series Analysis contributing to robust data modeling solutions.
- Leveraged the capabilities of Python packages such as NumPy, Pandas, Seaborn, and Matplotlib to execute intricate data manipulations and in-depth analysis, contributing significantly to data-driven insights and optimizing data warehousing strategies.
- Developed Python scripts to automate data cleaning and reporting tasks, streamlining data management processes.
- Crafted impactful visualizations using Power BI, Tableau, and Excel, enabling stakeholders to make informed decisions based on compelling data representations, while ensuring optimal data warehousing integration.
- Proficient in advanced ML/Statistics Algorithms, encompassing Decision Trees, Random Forest, Predictive Modeling, Regression, Classification, Clustering and Time Series Analysis.
- Demonstrated proficiency in cloud technologies, including Azure Data Factory, Azure Databricks, and AWS CloudFormation, with a focus on streamlining data workflows, optimizing query performance, and implementing real-time data processing solutions.
- Leveraged cloud technologies like Azure Data Factory and Azure Databricks to design and automate data pipelines.
- Utilized Databricks for comprehensive data pipeline development and automation, encompassing cluster management, library implementation, and PySpark development for data transformations and integration. Demonstrated expertise in Azure Databricks configuration and administration, ensuring optimal resource utilization and alignment with ETL/ELT processes.

CORE QUALIFICATIONS

Programming Languages	Python, Scala, SQL, Go, PowerShell & T-SQL
Hadoop Components /Big Data	Apache Spark, Airflow, HDFS, MapReduce, Hive, HCatalog, HBase, Sqoop, Impala, Zookeeper, Kafka & Yarn.
Cloud Platform	AWS (Lambda, S3, EC2, EMR, RDS), Microsoft Azure (Azure Databricks, Azure Data Factory, Azure Data Explorer, Azure HDInsight, ADLS).
Reporting and ETL Tools	AWS GLUE, Tableau & Power BI, Microsoft Excel.
Databases	Oracle, SQL Server, MS Access & NoSQL Database (HBase, MongoDB, DynamoDB).
Big Data Technologies	Hadoop, HDFS, Hive, Oozie, Sqoop, Spark, Machine Learning, Pandas, NumPy, Seaborn, Impala, Zookeeper, Airflow, Informatica, Snowflake, Data Bricks, Kafka .
Data Analysis Libraries	Pandas, NumPy, SciPy, Scikit-learn & Matplotlib
Containerization	Docker & Kubernetes
CI/CD Tools	Jenkins, Ansible, GitLab & Bamboo.
Monitoring Tools	Cloudera Manager
Software Methodologies	Agile, Scrum & Waterfall.
Development Tools	Eclipse, PyCharm, IntelliJ, SSMS & Microsoft Office Suite (Word, Excel, PowerPoint).
Version Control	GitHub

EDUCATION

University of North Carolina at Charlotte May 2023
Master's: Information Technology

WORK EXPERIENCE

Data Analyst   **Jan 2023 to Present**
NOV   **Charlotte, NC**

- Employed SQL and Python to execute comprehensive data analysis, applying techniques such as data visualization, data mining, and data warehousing, which resulted in extracting valuable insights from diverse datasets, contributing to a 20% increase in data-driven decision-making. Built and utilized data warehouses for efficient data storage and retrieval.
- Conducted Exploratory Data Analysis (EDA) using Matplotlib and Seaborn, diligently maintained and monitored adherence program reporting, designed, experimented with, and tested hypotheses, and expertly applied advanced statistical and predictive modeling techniques. These efforts led to real-time decision-making improvements by 15%.
- Developed solution-oriented views and dashboards within Power BI, integrating various chart types, including Pie Charts, Bar Charts, Tree Maps, Circle Views, Line Charts, Area Charts, and Scatter Plots, resulting in a remarkable 30% enhancement in data accessibility and understanding among stakeholders. These visualizations relied heavily on data models defined within the data warehousing system.
- Assisted in the automation of reporting functionality by utilizing Power BI tools and integrating MySQL and Data warehouse data sources, achieving a substantial 40% reduction in manual reporting efforts.
- Employed clustering techniques like K-Means and Hierarchical Clustering to uncover hidden patterns in data, which resulting in a 20% enhancement in customer segmentation strategies, driving targeted marketing campaigns.
- Crafted intricate SQL queries in MySQL, applying query optimization techniques such as indexing, query caching, and EXPLAIN analysis, leading to a notable 20% enhancement in query performance and overall database efficiency. Contributed to efficient data access and improved data warehousing processes.
- Produced comprehensive reports using Power BI, ensuring alignment with business needs, and efficiently deployed them on SharePoint Server. Implemented drill-through and drill-down functionalities in Power BI reports, facilitating a 25% improvement in data exploration and insights generation.
- Collaborated with various data providers to source data and developed Extraction, Transformation, and Loading (ETL) modules, achieving a notable 35% reduction in data loading time and enhancing data source analysis.
- Proficiently oversaw Redshift data warehouses, encompassing query optimization and performance tuning efforts, which resulted in a significant 20% enhancement in query response time and overall data warehouse efficiency.
- Tracked storage expenses and report on budget adherence and analyze cost per unit of storage and recommend cost-saving measures.

Environment: SQL, Python, Matplotlib, Seaborn, Power BI, Redshift, MySQL, SharePoint Server, Data warehouse tools, ETL module tools.

Software Engineer (Data)   **Oct 2020 to Dec 2021**
Tata Consultancy Services   **Pune, IN**

- Collaborated closely with a cross-functional team to design and develop secure, high-performance APIs and data pipelines. Successfully translated user needs into robust data solutions, optimizing ETL processes and data quality within the complex regulations and requirements.
- Leveraging Spark Streaming and Kafka, achieved real-time ingestion of high-volume data from diverse sources, enabling immediate identification and mitigation of potential fraudulent transactions and market fluctuations. This resulted in a 20% reduction in risk detection latency, allowing to react to threats faster and minimize potential losses.
- Extracted and refined technical requirements and implemented agile methodologies to translate user needs into robust pipeline designs and solutions.
- Developed and managed Azure Data Factory, for comprehensive ETL orchestration, incorporating Blob storage for efficient data persistence and backup on the Azure platform.
- Designed and automated ETL pipelines using Databricks for streamlined data processing, ensuring effective workflow management. Utilized Python-based Spark applications to establish distributed environments for loading high-volume files using PySpark into Azure SQL DB tables.
- Implemented streaming pipelines via Azure Event Hubs and Stream Analytics, facilitating data-driven workflow analysis.
- Developed data workflows utilizing Databricks, Scala, and Spark, capturing data from Delta tables in Delta Lakes, contributing to robust ETL processes.
- Developed Spark Streaming scripts for real-time processing, enhancing data accuracy by consuming topics from Kafka. Leveraged Azure Data Factory with Blob storage for storage and backup, employing Python scripting and tools like Airflow for batch data and Kafka for streaming data, building robust ETL pipelines.
- Operated ETL processes using Azure Databricks, employing Kafka for connecting to relational databases.
- Automated data ingestion, transformation, and storage using Apache Spark and Delta Lake, ensuring data quality and integrity.
- Implemented a distributed stream processing platform for low-latency data integration inside and outside Azure, facilitating real-time ETL capabilities.
- Deployed Sqoop for data extraction from Teradata to HDFS and subsequent analysis of patterns back to Teradata.
- Took charge of identifying and implementing SQL Server enhancements to optimize query performance and ensure data integrity.
- Enhanced Hive queries using best practices, Hadoop, YARN, Python, and PySpark to optimize query performance.

Environment: Python, SQL, Oracle DB, Teradata, Azure Data Factory, Azure SQL DB, Spark, Databricks, SSIS, SQL Server, Kafka, Informatica, Apache Spark, Delta Lake, Azure Event Hubs, Stream Analytics, Azure Blob Storage, PowerShell, Apache Airflow, Hadoop, YARN, PySpark, Hive, Teradata, Sqoop, HDFS, Spark, Agile.

Data Analyst Intern   **Jan 2020 – Sep 2020**
Netcidr LTD   **Hyderabad, IN**

- Manipulated and analyzed extensive financial datasets using advanced Excel and SQL, uncovering key patterns that drove strategic investment decisions.
- Conducted thorough variance analysis on monthly and annual financial reports, identifying critical discrepancies that saved the company over $200,000 in misallocated funds.
- Collaborated with cross-functional teams to develop a data governance framework, ensuring data integrity and compliance with regulatory standards.
- Utilized Power BI for robust financial analysis, visualizing data trends and delivering actionable insights to C-level executives.
- Enabled data-driven decision-making through the provision of comprehensive financial dashboards and reports, contributing to a 15% increase in annual growth rate.
- Designed and maintained sophisticated financial models in R, forecasting revenue streams accurately and informing a refined budget allocation strategy.
- Provided insights into customer behavior from complex financial datasets, aiding in a customer retention strategy that reduced churn by 8%.
- Implemented automated data reconciliation processes between reporting platforms, boosting financial data accuracy by 40%.
- Executed data cleaning and wrangling to enhance data quality and usability, improving financial forecast accuracy by 15% and demonstrating data management proficiency.
- Communicated findings through reports, presentations, and data storytelling, influencing executive decision-making and showcasing exceptional communication skills.
- Optimized financial analytics tool usage, SAS and SPSS, for complex data analysis, achieving a 30% reduction in data processing time.
- Established an automated alert system with Python for real-time anomaly detection in transaction data, cutting fraudulent activity by 20%.

Environment: SQL Python, R, Microsoft Excel, Power BI, Shell Scripting, Linear Modeling, Logistic modeling, Decision Tree, Random Forest, Neural Network, NLP, Predictive, Descriptive, Regression modeling Databricks, Classification model, Clustering Model, K-NN, SVM, PCA, Model Evaluation Skills, Statistics, Time series analysis .

ACADEMIC PROJECTS

Web Application Deployment on AWS using Amazon EC2, Amazon RDS, and Amazon S3
- Implemented application architecture on AWS, including instance management, load balancing, database configuration, and storage setup for optimal performance and scalability.
- Utilized Amazon EC2 instances to host web applications, ensuring reliable and efficient deployment.
- Configured Elastic Load Balancer (ELB) to distribute traffic across EC2 instances, enhancing application availability and scalability. Additionally, utilized Amazon RDS for database hosting and Amazon S3 for hosting static content, ensuring efficient data storage and retrieval.

Car Accidents Visualization:
- Dashboard showcasing US crash data (2016-2021) with insights on timing, conditions, and severity prediction. Cleaned dataset, dropping 3,414,349 null values.

Tree types found in Roosevelt National Forest:
- Model for analyzing tree types based on characteristics, evaluating machine learning algorithms and achieving 54% accuracy with linear regression.

CERTIFICATIONS

- **Microsoft Certified: Azure Developer Associate**
- **Certified in AWS Academy Data Analytics by AWS Academy Graduate**