Dinesh Reddy Kankanala

LinkedIn | (E) dineshreddyk569@gmail.com | (M) +1 (980) 303-1749

SUMMARY & OBJECTIVE

Over 3+ years of experience dealing with both structured and unstructured data. Proficient in data mining, acquisition, validation, modeling, and visualization, I uncover hidden patterns and translate them into clear, compelling visuals for both technical and non-technical audiences. My expertise extends to building robust data pipelines with Apache Spark, Hadoop, and cloud platforms, ensuring clean, reliable data for analysis. I prioritize data quality and collaborate with stakeholders to deliver data infrastructure that empowers decision-making.

EDUCATION

University of North Carolina At Charlotte, *Jan 2022 - May 2023* Charlotte, NC

Masters: Information Technology

GPA: 3.80/4.00

Jawaharlal Nehru Technological University, *Aug 2016 – June 2020* Hyderabad, IN

Bachelors: Electronics & Communication Engineering

TECHNICAL SKILLS

Programming Languages: Python, SQL, PowerShell & Scala.

Big Data Tools: Oozie, Sqoop, Spark, Pandas, NumPy, Seaborn, Airflow, Snowflake, Data Bricks.

Hadoop Components: HDFS, MapReduce, Hive, HCatalog, HBase, Sqoop, Impala, Zookeeper, Kafka & Yarn.

Cloud Services: AWS (Lambda, S3, EC2, EMR, RDS), Microsoft Azure (Azure Databricks, Azure Data Factory, Azure Data Explorer, Azure HDInsight, ADLS).

Reporting Tools: AWS GLUE, Tableau & Power BI, Microsoft Excel.

Databases: Oracle, SQL Server, MS Access & NoSQL Database (HBase, MongoDB, DynamoDB).

Operating Systems: Windows and macOS.

PROFESSIONAL EXPERIENCE

National Oilwell Varco (via DeltaSoft)**,** *Data Engineer, Jan 2023 - Present* 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.
- Designed a high-level ETL architecture to efficiently transfer data from source servers to the enterprise data warehouse, optimizing data flow and ensuring data integrity.
- Utilize advanced time series analysis methods to forecast future trends and anomalies within temporal data, leading to reduced forecasting errors and improved inventory management.
- Actively contributed to enhancing performance by designing partitions in cubes using SQL Server Analysis Services (SSAS), resulting in a remarkable improvement in query response time.
- Designed data warehousing solutions in Oracle and automate ETL processes using Oracle Data Integrator (ODI), resulting in a significant reduction in manual data processing efforts and improved data storage and retrieval.
- Built interactive Power BI dashboards with calculated fields and parameters, increasing data accessibility by 30% through publishing live reports and workbooks.
- 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.
- Implemented regular data integrity checks to identify and rectify any corruption or loss and tracked read/write speeds and I/O latency to identify performance bottlenecks.
- Monitored backup schedules and ensured they are executed as planned and Reported on the effectiveness of tiering strategies and suggest adjustments.
- Leveraged AWS Lambda and EC2 Auto Scaling to enable scalable data processing, resulting in a 30% reduction in operational costs while ensuring seamless performance adjustments based on workload demands.
- Generated reports on successful backups and conducted periodic recovery tests and generated reports on storage & monitoring.
- Actively participated in agile workflows, including sprint planning and retrospectives, gaining insights into efficient project management methodologies, which contributed to a 15% improvement in project delivery timelines during the internship period.
- Expertly integrated code repositories with Git and GitHub, ensuring robust version control and facilitating collaborative development and code review processes.

Environment: SQL, Python, Oracle, SQL Server Analysis Services, Oracle Data Integrator, Power BI, Git, GitHub, AWS Lambda, EC2.

Tata Consultancy Services, *Data Engineer, Oct 2020 – Dec 2021* Pune, IN

- Collaborated closely with a cross-functional team to design and develop secure, high-performance APIs. Extracted and refined technical requirements and implemented agile methodologies to translate user needs into robust pipeline designs and solutions.
- 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.
- Developed and managed Azure Data Factory, for comprehensive ETL orchestration, incorporating Blob storage for efficient data persistence and backup on the Azure platform.
- Built re-usable Snap Logic Pipelines to ingest data from Oracle DB and SFTP file systems. Operated ETL processes using Azure Databricks, employing Kafka for connecting to relational databases.
- Developed Spark Streaming scripts for real-time processing, enhancing data accuracy by consuming topics from Kafka and store the stream data to HDFS using Python and NoSQL databases such as HBase and Cassandra.
- Migrated Existing MapReduce programs to Spark Models using Python. Implemented Spark using Scala and Spark SQL for faster testing and processing of data.
- Designed and automated ETL pipelines using Databricks for streamlined data processing, ensuring effective workflow management. Established distributed environments for loading high-volume files using PySpark into Azure SQL DB tables.
- Developed data workflows utilizing Databricks, Scala, and Spark, capturing data from Delta tables in Delta Lakes, contributing to robust ETL processes.
- Utilized Snap Logic best practices to tweak the performance of Pipelines. Experienced in creating complex SSIS packages using proper control and data flow elements.
- Automated data ingestion, transformation, and storage using Apache Spark and Delta Lake, ensuring data quality and integrity.
- 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: Spark Streaming, Kafka; Azure Data Factory, Blob Storage (Azure); Snap Logic; Azure Databricks, HDFS, Python, HBase, Cassandra; PySpark; Databricks, Scala, Spark SQL, Delta Lake; SSIS; Apache Spark; SQL Server; Hadoop, YARN.

Synergist Ltd, *Data Analyst Intern, Jan 2020 – Sept 2020* Hyderabad, IN

- Leveraged SQL and Python to extract, manipulate, and analyze complex datasets, uncovering valuable insights and trends. Assisted in data cleaning, transformation, and preparation to ensure high data quality for analysis.
- Designed and implemented complex SQL queries, including subqueries, window functions, and recursive CTEs, to extract and manipulate data for in-depth analysis.
- Utilize essential Python packages, including NumPy, Pandas, Seaborn, and Matplotlib, to conduct comprehensive data analysis and visualization, enabling data-driven insights that contributed to a 25% increase in data accuracy.
- Built automated ETL (Extract, Transform, and Load) processes using SQL Server Integration Services (SSIS), reducing manual data processing efforts by 40%.
- Collaborated with cross-functional teams to develop a data governance framework, ensuring data integrity and compliance with regulatory standards.
- Optimized financial analytics tool usage, SAS and SPSS, for complex data analysis, achieving a 30% reduction in data processing time.
- Spearheaded data visualization efforts by creating highly informative and visually captivating reports and dashboards using Tableau and Excel, resulting in a 20% improvement in data accessibility and understanding among stakeholders.
- Applied advanced data analysis techniques such as data mining and data warehousing to uncover meaningful insights from various datasets, resulting in a 30% improvement in identifying critical business trends and opportunities.

Environment: SQL, Python (NumPy, Pandas, Seaborn, Matplotlib), SQL Server Integration Services (SSIS), SAS, SPSS, Tableau, Excel.

ACADEMIC PROJECTS

Speech To Text Recognition UNC Charlotte, NC

- Developed an interactive Speech-to-Text application leveraging Streamlit,and established libraries like Speech Recognition, PyAudio and pocketsphinx for robust speech recognition capabilities.
- Designed an intuitive interface with interactive widgets for effortless conversion of spoken content into text.

Data Visualization UNC Charlotte, NC

- Studied the trend and relation between GDP, Military Expenditure and Population of various countries worldwide.
- Developed a Dashboard using D3 & React JS with advanced visualization techniques to study the past 30 years data.
- Added various interactively interlinked elements to understand and discover the various insights in the data.

CERTIFICATIONS

- AWS Academy Data Analytics Certification
- Microsoft Certified: Azure Data Engineer Associate