

SUDARSHAN J N RAO

DATA ENGINEER

+1 (667) 290-5370 | sudarshanjn141@gmail.com | Baltimore, MD



SUMMARY

- Led full-cycle Data Engineering projects with around 4 years of experience, adeptly navigating **SDLC, Agile, and Waterfall methodologies** to foster seamless collaboration and ensure the timely delivery of resilient solutions.
- Proficiently developed **ETL pipelines using Python, and SQL**, integrating with diverse databases including **MySQL, Redis, MongoDB, DynamoDB, PostgreSQL, and Teradata** for streamlined data processing.
- Applied advanced statistical and machine learning algorithms (**NumPy, Pandas, SciPy, Scikit-learn, TensorFlow**) to build predictive models, enhancing accuracy and insights for strategic decision-making.
- Expertly crafted visually compelling dashboards using **Tableau, Power BI, and Advanced Excel**, incorporating **Pivot Tables, VLOOKUP, and VBA Macros** for insightful data storytelling.
- Implemented cloud-based solutions on **Microsoft Azure (ADF, SSMS, Blob Storage), AWS (EC2, RDS, S3)**, optimizing storage, retrieval, and processing of data.
- Applied big data technologies, including **Apache Hadoop, HDFS, Spark, Hive, Pig, and Zookeeper**, for scalable and efficient processing of large-scale datasets.
- Applied statistical methodologies (**ANOVA, Hypothesis Testing, Regression Analysis**) and proficiency in Linear Algebra for advanced analytics, extracting actionable insights from complex datasets.
- Led end-to-end A/B testing, clustering, classification, regression, and forecasting projects, incorporating association rules, enhancing business outcomes through data-driven optimizations.

SKILLS

Methodologies:	SDLC, Agile, Waterfall
Programming Language:	Python, R, SQL
Packages:	NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, TensorFlow, Seaborn, Plotty, TensorFlow, NLTK
Visualization Tools:	Tableau, Power BI, Advanced Excel (Pivot Tables, VLOOKUP, VBA Macros)
IDEs:	Visual Studio Code, PyCharm, Jupyter Notebook, IntelliJ IDEA, Eclipse
Database:	MySQL, Redis, MongoDB, DynamoDB, PostgreSQL, Teradata
Cloud Technologies:	AWS (EC2, RDS, S3), Microsoft Azure (ADF, SSMS, Blob Storage)
Other Technical Skills:	Google Analytics, DAX, Power Query, Alteryx, SAS, JIRA, SAP, SSIS, SSRS, Machine Learning Algorithms, Probability distributions, Confidence Intervals, ANOVA, Hypothesis Testing, Regression Analysis, Linear Algebra, Advance Analytics, Data Mining, Data Visualization, Data warehousing, Data transformation, Data, Storytelling, Association r rules, Clustering, Classification, Regression, A/B Testing, Forecasting & Modelling, Data Cleaning, Data Wrangling.
Big Data Tools:	Apache Hadoop, HDFS, Spark, Hive, Pig, Zookeeper
Version Control Tools:	Git, GitHub
Operating Systems:	Windows, Linux, Mac iOS

EXPERIENCE

Data Engineer | Goldman Sachs, MD

Oct 2022 - Present

- Interacted with business and customers to understand the architecture of the applications.
- Designed and created Data Marts, Databases, Indexes, Views, Aggregations, Stored Procedures, Partitions and Data Integrity.
- Performed research to identify source and nature of data required for ETL solutions.
- Employed Parquet for efficient data serialization, enhancing data transfer speed and reducing storage overhead.
- Integrated data sources from Kafka (Producer and Consumer API) for data stream-processing in Spark using AWS Network.

- Implemented and managed complex data workflows using Apache Airflow to orchestrate ETL processes and dependencies.
- Executed SQL queries using PySpark SQL for data manipulation, leveraging the expressive power of SQL for analytics tasks.
- Extracted, Transformed and Loaded (ELT/ETL) using Azure Data Factory, building Data flows, pipelines and transformed activity, building SSIS packages and executed SSIS in Azure Data Factory using SSIS integration runtime. Replaced existing ETL/SSIS solution with Spark job using Azure Databricks.
- Conducted thorough statistical significance analysis using tools such as Python's SciPy library to determine the reliability and significance of observed changes in performance metrics.
- Parsed JSON and log data and designed the data flows using Apache NiFi - Processors, Funnel.
- Implemented the project development using Agile processes in Kanban boards and 2-week sprints.

Data Engineer | NextGen Healthcare, India

Sept 2019 - Jan 2022

- Designed, developed, and maintained end-to-end data pipelines for processing and transferring healthcare data efficiently.
- Utilized ETL processes to extract data from various sources, transform it according to business rules, and load it into data warehouses.
- Implemented security measures and ensured compliance with healthcare data regulations, particularly HIPAA.
- Created and optimized data models to support analytical and reporting needs, ensuring data integrity and performance.
- Managed relational databases (SQL Server, Oracle) for efficient storage and retrieval of healthcare data.
- Implemented solutions using Apache Spark and Hadoop for processing and analyzing large healthcare datasets.
- Instituted robust data quality checks and governance processes to ensure accurate and reliable healthcare data.
- Deployed automation and orchestration tools such as Apache Airflow for scheduling and overseeing data workflows.
- Used Python libraries like NumPy and SciPy for data modeling, enhancing the efficiency of healthcare data analysis.
- Utilized Amazon S3 for scalable and reliable storage of healthcare datasets, ensuring accessibility and durability.
- Deployed and configured Tableau Server to provide a centralized platform for sharing and collaborating on healthcare data visualizations.
- Implemented PySpark for distributed data processing, optimizing performance in handling large healthcare datasets.
- Built various graphs for business decision making using Python matplotlib library.
- Created Informatica mappings using various Transformations like Joiner, Aggregate, Expression, Filter and Update Strategy.

EDUCATION

Master of Science in Information Systems – University of Maryland Baltimore County, Baltimore, MD

Bachelor of Technology in Information Science – RNSIT, Bangalore, India

CERTIFICATION

AWS Certified Solutions Architects