

Maram Deepak Reddy

Data Engineer

deepak.r@mailsworks.com | 660-528-1511 | Greenbelt, MD

SUMMARY

- Data Engineer with 4+ years of experience in Data Extraction, Data Modelling, Statistical Modelling, Data Mining, and Data Visualization.
- Experience in working with the entire process of Software Development Life Cycle (SDLC) and good experience of Agile, and Waterfall Methodologies.
- Expertise in Big Data architecture like Hadoop (AWS, Cloudera) distributed systems, MongoDB, NoSQL.
- Extensive experience in Python including NumPy, SciPy, Scikit-learn, Pandas, Matplotlib and TensorFlow.
- Proficiency in the development and design of various scalable systems using Hadoop technologies in various environments.
- Solid experience working with various native services in AWS Cloud like S3, EMR, Athena, Glue, Redshift etc., for building data pipelines.
- Ability to analyze data using Hadoop Ecosystems including HDFS, MapReduce, Hive and Pig.
- Working knowledge of Amazon's Elastic Cloud Compute (EC2) infrastructure for computational tasks and Simple Storage Service (S3) as a storage mechanism.
- Strong experience and knowledge of real time data analytics using Spark Streaming, Kafka and Flume.
- Good experience in writing Spark applications using Python and Scala.
- Proven experience in designing, developing and deploying SSIS packages in MS SQL server environment using SSIS.
- Well-Versed in performing data analysis on various IDEs like Jupyter Notebook and PyCharm.
- Capable in working with Tableau reports with complex calculations and working on Ad-hoc reporting using Power BI.
- Experience in Azure, Kubernetes, Service, Functions, Data Lake Storage, Data Factory, Synapse Analytics, Monitor, Active Directory.
- Proficient in Big data storage, processing, analysis, and reporting on all major Cloud vendors of Azure.
- Implemented control flow architecture for developing a secure and end-to-end big data application using ADF v2, Azure Data Bricks, Azure Synapse Analytics, Azure Data Lake Storage, Azure SQL DB, and Azure key Vault.
- Extensive experience in developing and deploying data pipelines in cloud such as AWS and Azure.
- Expertise in developing and publishing reports of various types using Power BI, SSRS and Business Objects.
- Good Working knowledge of different data sources as MS SQL Server, PostgreSQL and MySQL.

SKILLS

Methodology: SDLC, Agile, Waterfall

Programming Language: R, Python, SQL

IDE's: PyCharm, Jupyter Notebook, Visual Studio Code

ML Algorithm: Linear Regression, Logistic Regression, Decision Trees, Supervised Learning, Unsupervised Learning, Classification, SVM, Random Forests, Naive Bayes, KNN, K Means, CNN

Big Data Ecosystem: Hadoop, MapReduce, Hive, Apache Spark, Pig

Framework: Kafka, Airflow, Snowflake, Django, Docker

Cloud Technologies: AWS S3, EMR, Glue, Athena, Redshift, IAM, EC2, Lambda, Azure, Virtual Machines, Kubernetes Service, Functions, Data Lake Storage, Data Factory, Synapse Analytics, Monitor, Active Directory

Packages: NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow

Reporting Tools/ Other Software: Tableau, Power BI, SSRS, SSIS

Database: MS SQL Server, PostgreSQL, MongoDB, MySQL

Other Tools: Anaconda, Sitecore, CMS, Jira, Confluence, Jenkins, Git, MS Office, Data Cleaning, Data Wrangling, Critical Thinking, Communication Skills, Presentation Skills, Problem-Solving

Operating Systems: Windows, Linux

EXPERIENCE

State Street, USA | Data Engineer

March 2023 - Current

- Implemented the Agile model throughout the application development lifecycle, enabling iterative progress, continuous feedback, and faster time-to-market; successfully delivered a feature-rich application that meets customer needs and boosted revenue by 15%.
- Performed end-to-end Architecture & implementation assessment of various AWS services like Amazon EMR, Redshift, AWS S3.
- Used Python unit and functional testing modules such as unit test, mock and custom frameworks in-line with Agile Software Development methodologies.
- Devised and developed robust Kafka connectors, seamlessly integrating data from diverse sources into Kafka topics, optimizing data flow by 30%.
- Used Amazon Web Services Elastic Compute Cloud (AWS EC2) to launch cloud instances.
- Implemented a proof of concept deploying this product in AWS S3 bucket and Snowflake.
- Employed data visualization techniques using Matplotlib and Seaborn libraries in Python to effectively communicate complex data insights, resulting in a 20% increase in stakeholder engagement and informed decision-making.

- Integrated AWS Lambda functions with other AWS services such as API Gateway, S3, DynamoDB, SQS, SNS, etc., to build comprehensive serverless solutions.
- Utilized Python packages such as NumPy, Pandas, SciPy, and Matplotlib to analyze data trends, identify key market opportunities, and drive data-informed decision-making; contributed to a 20% increase in revenue through targeted marketing campaigns.
- Involved in designing Hive schemas, using performance tuning techniques like partitioning, bucketing.
- Performed spark for data transformations by writing MapReduce and Pig scripts as per business requirements.
- Developed Spark applications using Spark Core, Streaming, SQL, Data Frames, Datasets & Spark-ML.
- Experience in designing, developing, or deploying applications using serverless architecture, particularly AWS Lambda.
- Developed Spark programs for Batch and Real-Time Processing.
- Used AWS Glue for the data transformation, validate and data cleansing.
- Coordinated with Reporting developers to design and implement visually compelling Tableau reports and dashboards, contributing to a 15% increase in user engagement and data-driven decision making.
- Used Git repository and commands extensively for code check-in.

Groovy Web, India | Data Engineer

Feb 2019 - March 2022

- Utilized MySQL & MongoDB expertise to optimize database performance, design efficient data models, and execute complex queries using microservices improving data retrieval speed by 20%.
- Spearheaded the implementation of the Waterfall Methodology to guide the project through all phases of the SDLC, resulting in the timely production of essential artifacts.
- Utilized MySQL & MongoDB expertise to optimize database performance, design efficient data models, and execute complex queries using microservices improving data retrieval speed by 20%.
- Worked with Azure cloud infrastructure services and was involved in ETL, Data Integration and Migration.
- Worked with Azure Synapse Analytics to Develop end-to-end ETL/ELT applications.
- Designed and implemented database solutions in Azure SQL Data Warehouse and Azure SQL.
- Involved Map Reduce/Spark Python modules for machine learning and predictive analytics in Hadoop on Azure.
- Developed ETL/ELT pipelines using big data and cloud technologies such as Apache Hive, Apache Spark, Azure Data Factory and Azure Databricks.
- Worked with various big data file formats such as Apache Parquet, CSV, AVRO, and JSON while developing Big data Applications using Apache Hive And Apache Spark.
- Capitalized Python for data manipulation, analysis, and automation, utilizing libraries like Pandas, NumPy, and Matplotlib to process and visualize data, resulting in 30%-time savings in data processing tasks.
- Developed Python-based ML models with Azure Data Factory to predict Monitor rules, reducing creation time by 15%.
- Applied statistical techniques to analyze datasets, identifying trends and patterns, and providing actionable insights that contributed to a 10% increase in revenue.

EDUCATION

Master in Applied Computer Science | Northwest Missouri State University, Maryville, Missouri | GPA:3.8

Bachelor in Computer Science and Engineering | JNTUH, India | GPA:3.2