

PRASHANTH MUDDAM

Email: prashanthmuddam2@gmail.com| Mobile: +1 (203) 901-6723 | LinkedIn| Github

SUMMARY

Experienced Data Engineer with 4+ years of experience specializing in data analysis, visualization, and pipeline optimization. Improved data accuracy by 20% through comprehensive exploratory data analysis and SQL-based data cleansing. Capable of creating actionable insights using Tableau and streamlining ETL processes. Proficient in Python, SQL, and AWS technologies, with demonstrated success in enhancing data processing efficiency and accelerating insights generation. Mastered Artificial Intelligence(AI), machine learning, and deep learning algorithms—such as Random Forest, XGBoost, Support Vector Machines (SVM), k-Means, and Neural Networks—to drive advanced analytics initiatives, streamlining data processing by 40% while ensuring compliance with industry regulations through robust pipeline maintenance.

EDUCATION

Master of Business Analytics
Sacred Heart University, Fairfield, CT

Aug 2023 – Dec 2024

CERTIFICATIONS

Earned Microsoft Certified: Azure Data Engineer Associate.
Completed Tableau Certification – Edureka.

TECHNICAL SKILLS

Programming Languages: Python, R, Java, SQL, PL/SQL, Scala, Shell Scripting, Bash.
Big Data Tools: Hadoop, HDFS, Sqoop, Hive, Pig, Pyspark, Kafka, Spark SQL, MapReduce, Flink.
Databases: MySQL, Oracle, MS SQL, DB2, NoSQL, SQL Server, MongoDB, DynamoDB.
Frameworks/Libraries: ArcGis, Matplotlib, SciPy, Scikit-Learn, PyTorch, Django, FastAPI, Restful API, SQLAlchemy.
Cloud Platforms & Services: AWS, GCP, Microsoft Azure.
AWS Stack: Aws Athena, S3, Lambda, SNS, EC2, Sagemaker, Kinesis, RDS, FireHose, Databricks, IAM, Route53, Quicksight, Cloudwatch, SQS, KMS, ECS, Stepfunctions, Elastic search.
Data Visualization Tools: Tableau, Power BI, Excel, Erwin.
ETL & Data Integration: SQL Server Integration Services (SSIS), Talend, Alteryx, Airflow.
CI/CD & Automation: Jenkins, Gitlab, Ansible, Terraform.
Project Management & Methodologies: Jira, Agile, Scrum.
Version Control: Git, GitHub.
Additional skills: Data Wrangling, NLP, Model Evaluation, Hyper-parameter Tuning, Data Cleaning.
Soft Skills: Problem-Solving, Communication, Team Collaboration, Analytical Thinking, Attention to Detail.

PROFESSIONAL EXPERIENCE

United Health Group
Data Engineer

Dec 2023 - Present
Charlotte, NC

Technologies: SQL, Python, Pandas, Tableau, NumPy, Matplotlib, MySQL, SQL Server, Power BI, Apache Cassandra, AWS(Glue, S3, Lambda, Redshift, Codepipeline), Apache Airflow, Teradata, DBT,Kubernetes, MongoDB, Jira.

- Developed real-time data ingestion pipelines with Python, Apache Kafka (MSK), and Kinesis Firehose to accelerate the processing of electronic health records (EHR), insurance claims, and IoT-generated patient data.
- Engineered and maintained scalable ETL/ELT pipelines using AWS Glue, AWS Lambda, and EMR, accelerating data processing speed by 40% and ensuring seamless data ingestion into Amazon Redshift and S3, supporting critical healthcare data workflows.
- Utilized DBT and AWS Data Pipeline to orchestrate data workflows and perform reliable transformations across structured and semi-structured datasets.
- Leveraged Amazon S3, Redshift, RDS (PostgreSQL), DynamoDB, and Amazon Keyspaces (Cassandra) to optimize data storage and retrieval, improving performance by 50% and reducing operational costs by 15% while ensuring HIPAA compliance.
- Implemented AWS S3 as a centralized data lake, integrating Databricks Delta Lake to support ACID transactions, versioned data, and faster analytical queries.
- Enhanced machine learning model performance by leveraging healthcare data APIs such as FHIR and HL7 to enrich patient datasets, leading to a 5% improvement in patient readmission prediction accuracy.
- Deployed Machine learning models in production environments using Docker, Kubernetes(EKS), and AWS SageMaker, reducing deployment inconsistencies by 35%.
- Implemented CI/CD pipelines with AWS CodePipeline and CodeBuild, automating model deployment, testing, and monitoring, reducing deployment time by 30%.

- Monitored production pipelines with AWS CloudWatch and AWS X-Ray, enabling proactive debugging and ensuring reliability and scalability of workflows.
- Automated encryption key rotation and access auditing using AWS CloudTrail and AWS KMS, ensuring compliance with HIPAA regulations and enhancing data security by reducing the risk of unauthorized access to protected health information (PHI)
- Automated data reporting and visualization using Pandas, NumPy, Matplotlib, and Seaborn, reducing manual reporting by 40% and enabling actionable insights.
- Created interactive dashboards in Amazon QuickSight and Tableau, boosting stakeholder engagement and supporting data-driven decision-making by 25%.

Accenture

Data Engineer

January 2021 – June 2023

Accenture, Hyderabad, India

Technologies: Python, AWS(Glue, Lambda, EMR, DyanamoDB, Redshift, Codepipeline, CloudFormation), PostgreSQL, Hadoop, Hive, Apache Spark, Kafka, Talend, Nagios, QlikView.

- Engineered and boosted data pipelines using Python within AWS Glue, improving data management efficiency by 45% while ensuring secure data access via IAM roles and policies.
- Structured and deployed serverless data processing workflows using AWS Lambda, reducing infrastructure overhead by 30% and enhancing response times for event-driven financial transactions.
- Architected real-time data streaming solutions using AWS Kinesis, improving real-time analytics processing speeds by 50%, enabling faster fraud detection and risk assessment.
- Programmed data ingestion, transformation, and storage using Apache Spark and AWS Lake Formation, ensuring 99.9% data accuracy and compliance with financial regulatory standards (SOX, GDPR, PCI DSS).
- Utilized Python-based Spark applications to establish distributed processing environments, efficiently loading terabytes of transactional data into Amazon Redshift tables, reducing query execution time by 40%.
- Developed and executed data warehousing solutions in Amazon Redshift, optimizing ETL pipelines with SSIS, which reduced manual reporting efforts by 40% and improved data accuracy through SSRS report automation.
- Enhanced Hive queries on Amazon EMR using Hadoop, YARN, Python, and PySpark, reducing data processing time by 40% and accelerating credit risk modeling workflows.
- Administered and maintained PostgreSQL and DynamoDB databases, ensuring 99.99% uptime and data integrity for highly sensitive financial applications.
- Operated Hadoop, CDH, Hive, and Spark to process and analyze petabyte-scale financial datasets, improving reporting efficiency by 50% and delivering actionable insights for trading and risk management.
- Applied CI/CD pipelines with AWS CodePipeline and automated infrastructure provisioning using CloudFormation, accelerating deployment speed by 60% while ensuring robust system monitoring via Nagios.
- Designed and streamlined Power BI reports by implementing DAX calculations and data modeling, improving data query efficiency by 40%.