

# Gurudatta Sai Prasad Akkala

34 Midway Ln Apt 101, Ewing, NJ 08628

+15513287259 [✉ agdsprasad@gmail.com](mailto:agdsprasad@gmail.com) [🌐 LinkedIn Profile](#)

## PROFESSIONAL SUMMARY

Impact-driven Data Engineer with 4+ years of experience developing and optimizing scalable cloud data pipelines on AWS and Azure. Demonstrated success in reducing processing times by up to 65% and cutting manual effort by 70% through automation and real-time data solutions. Skilled in delivering end-to-end data architectures supporting fraud detection, compliance analytics, and business intelligence. Proven ability to lead cross-functional teams and drive complex data initiatives from development to production.

## WORK EXPERIENCE

### Cross River | Sr. Data Engineer | New Jersey

October 2024 – Present

- Improved a real-time fraud detection pipeline in Snowflake by streamlining data ingestion and preprocessing, reducing execution time by 40% and increasing accuracy by 25% in detecting suspicious transactions and account anomalies across banking systems
- Managed Spark-Airflow pipelines to process 1TB+ of transactional data daily, ensuring 95% SLA reliability for real-time AML monitoring of high-risk patterns. Enhanced customer segmentation for KYC initiatives and supported compliance with BSA and FATF standards.
- Enhanced ETL pipelines using Airflow, Python (parallel processing, incremental reads), and Kafka to automate batch and real-time data ingestion into Snowflake, cutting manual effort by 70% and improving reporting latency by 3 hours for critical banking analytics.
- Augmented Snowflake (v6.0) schemas with clustering on timestamp and customer\_id, enabling partition pruning and reducing query latency by 65%. Streamlined Matillion ETL (v2.44) by implementing parallel loading and incremental CDC logic.

### NueHee | QA Tester (Remote)

June 2023 - August 2023

- Led comprehensive manual testing across EHR platforms and patient-facing applications, ensuring a 98% defect-free release rate and driving a 15% improvement in user satisfaction by identifying and resolving critical issues before deployment.
- Conducted comprehensive testing of the telemedicine module, identifying and logging 50+ bugs in JIRA, which led to a 40% reduction in post-deployment issues and ensured a smooth rollout across 500+ clinics.
- Maintained detailed and audit-ready test documentation in TestRail and Confluence, ensuring 100% compliance with FDA regulations and expediting data-driven decision-making during regulatory inspections.
- Improved QA efficiency by centralizing reusable test assets and fostering cross-functional alignment through daily standups and bug triage sessions, cutting test prep time by 25% and boosting bug resolution speed by 35%.

### HealthPlix | Data Engineer | Bengaluru

July 2020 – July 2022

- Built a scalable AWS data lake to process high-volume insurance data, using EMR with Spark, S3 for staging, and Redshift for analytics. Integrated AWS Glue Catalog to ensure data governance and GLBA compliance, streamlining secure insights for downstream teams.
- Fine-tuned Spark workloads on AWS EMR to cut data transformation time by 50% and reduce cloud costs by 70%. Designed Kafka-based CDC pipelines for real-time policy and claims updates, enabling low-latency data capture at scale.
- Executed scalable data pipelines on Apache Spark to analyze insurance data and predict customer preferences using Random Forest and XGBoost. This improved engagement by 15% and reduced churn by 10%, enhancing marketing efforts and customer retention.
- Implemented a HIPAA-compliant AWS framework using Lambda, DynamoDB, and Kerberos to enhance real-time data access control and encryption, significantly reducing malicious access risks by 80% while ensuring regulatory compliance for sensitive insurance data.

### Digit Insurance | Data Analyst | Hyderabad

May 2019 – June 2020

- Worked with cross-functional teams to upgrade the Clinical Trial Management System, ensuring data compliance and streamlining processes to cut rework by 30%, speed up delivery, and enhance real-time data analysis for improved clinical trial decision-making.
- Developed tracking modules for Serious Adverse Events and Hospitalizations, streamlining patient safety monitoring by 40% and supporting real-time, data-driven decisions to enhance healthcare outcomes and response times.
- Refined database models using ER/Win to integrate clinical and laboratory data, improving data consistency by 25%. Enhanced cross-system interoperability, streamlining data access and analysis to support more efficient healthcare operations and decision-making.
- Streamlined clinical data pipelines and engineered complex T-SQL stored procedures to automate data workflows, reducing processing time by 30% and improving report turnaround by 35%, which supported faster, data-driven decisions across clinical operations.

## SKILLS

- Programming Languages:** C, C++, Python, SQL, R, Java, JavaScript, HTML
- Python Libraries:** NumPy, Pandas, NLTK, Matplotlib, Seaborn
- Frameworks:** Flask, Django
- Cloud Platforms:** Amazon Web Services (Redshift, Lambda, Glue, S3, EC2), Google Cloud Platform (GCP), Microsoft Azure
- Big Data Technologies:** PySpark, Apache Spark, Hadoop, Hive, HDFS, HBase, Kafka, Databricks, MapReduce, Scala, Pig
- ETL Tools:** Airflow, Apache NiFi, Sqoop, Flume, Oozie

- **Databases:** SQL Server, PostgreSQL, MySQL, Oracle, MongoDB, DynamoDB, Amazon Redshift, Workbench
- **Visualization Tools:** Tableau, Power BI
- **Security Tools:** Snort, Nmap, Nessus, Burp Suite, Metasploit, Wireshark
- **DevOps & CI/CD:** Jenkins, Docker, Kubernetes, Git, Github, Terraform, Gradle, Maven
- **Applications & Tools:** Amazon Connect, Data Warehousing
- **IDEs & Operating Systems:** Eclipse, IntelliJ, PowerShell, Visual Studio, Windows, Linux, UNIX
- **Methodologies:** Agile/Scrum, Waterfall

## EDUCATION

<b>Master of Science in Management Information Systems (GPA: 3.7/4.0)</b> Stevens Institute of Technology, School of Business, Hoboken, NJ	<b>May 2024</b>
<b>Master of Computer Applications (GPA: 7.5/10.0)</b> Acharya Nagarjuna University, Guntur, India	<b>June 2022</b>
<b>Bachelor of Computer Science (GPA: 7.7/10.0)</b> Acharya Nagarjuna University, Guntur, India	<b>March 2020</b>

## PROJECTS

<b>Network Intrusion Detection System (NIDS)</b>	<b>September 2023 – December 2023</b>
<ul style="list-style-type: none"> <li>• Developed a scalable detection pipeline using PySpark and Hadoop to analyze network traffic in the UNSW-NB15 dataset.</li> <li>• Applied Random Forest and SVM models for anomaly detection of network threats with 92% accuracy.</li> <li>• Collaborated in a team of 3 to operationalize the pipeline, end-to-end solution, improving detection and system reliability by 20%.</li> </ul>	
<b>Vulnerability Assessment &amp; Penetration Testing (VAPT)</b>	<b>January 2023 – April 2023</b>
<ul style="list-style-type: none"> <li>• Developed an automated vulnerability scanning workflow using AWS Lambda, integrating Burp Suite and OWASP ZAP to schedule</li> <li>• Orchestrated periodic security scans, reducing manual effort by 60% and enabling real-time log ingestion for audit analysis.</li> </ul>	
<b>Secure Web Application Development</b>	<b>September 2022 – December 2022</b>
<ul style="list-style-type: none"> <li>• Integrated 2FA and OAuth 2.0 authentication into a data-sensitive e-commerce platform.</li> <li>• conducted code-level audits to detect and remediate SQL injection and XSS vulnerabilities</li> <li>• Reduced security risks by 25% and ensuring data integrity across transactional pipelines.</li> </ul>	

## CERTIFICATIONS

- AWS Fundamentals, Migrating to AWS Cloud, Architecting Solutions on AWS, AWS Cloud Technical Essentials, Building Data Lakes on AWS, Exam Prep: AWS Certified Solutions Architect - Associate | Amazon Web Services
- Python for Data Science, AI & Development | IBM, Crash Course on Python | Google
- Databases and SQL for Data Science with Python | IBM, SQL | Sololearn
- Hands - On Tableau Training for Data Science | Udemy
- Complete Microsoft Power BI Bootcamp | Udemy

## LEADERSHIP & TEAMWORK

- Led a team in the Strategy Simulation: Competitive Dynamics and Wintel V2, achieving a 15% increase in simulated market share. **February 2022**
  - Fine-tuned collaboration and optimized resource allocation to outperform simulation benchmarks.
- Scored 96% in a National-Level Online Quiz on Computer Science. **June 2021**
- Achieved 100% in National-Level Quizze on Object-Oriented Programming using Java. **August 2020**
- Participated in leadership and community service activities through National Cadet Corps (NCC), developing discipline and teamwork skills. **August 2017**