

Ganesh Brahma Kollapudi

Data Engineer

(816) 979-2511 | ganeshbrahma304@gmail.com | Portfolio | LinkedIn | Github

TECHNICAL SKILLS

Languages: Python, PySpark, SQL(MySQL, SQL Server, Oracle, PostgreSQL, Teradata, PL/SQL), NoSQL(Cassandra, HBase), HDFS, Linux, Impala, HiveQL, Oracle, DB2, HTML, CSS, JavaScript
Data Integration: Apache Spark, Ab Initio (exposure/knowledge), Hive, Sqoop, Hadoop, Pig, Kafka, Terraform, Azure Data Factory, Cloudera, Data Mesh, Apache Airflow, Autosys, Atomic, Unity Catalog, SSIS, Teradata, Data warehouse
Cloud Platforms: AWS (S3, EC2, Lambda, RDS, IAM, SES, EventBridge, Cloudwatch, CloudTrail), Azure, GCP,
Developer Tools: Tableau, Power BI, VS Code, Git, CI/CD, DevOps, SDLC, Docker, Kubernetes

EDUCATION

University of Central Missouri

Master's Degree, Computer Science, 3.70 CGPA

Warrensburg, USA

Jan 2024 - Aug 2025

EXPERIENCE

Data Engineering and Analysis Intern

Jun 2025 – Aug 2025

Munich Re

New York, USA

- Built a Deductible Recovery Report uncovering **\$8M** in unpaid deductibles, improving Claims recovery tracking
- Developed **Databricks Delta Live Tables ETL** with **SparkSQL, PySpark/Python**, boosting processing speed **40%** and enforcing data governance via **Unity Catalog**.
- Automated data filtering from multi sheet Excel files in **Azure Data Lake** using **Pandas**.
- Orchestrated **REST API based workflows** for Databricks job dependencies, integrated with **Azure Event Hub** enabling better system reliability and system reliability.

Data Engineer

Nov 2022 - May 2023

Applaudo, Walmart

Freelance

- Developed **batch and real time pipelines** on AWS using Spark and Airflow, ensuring system reliability and reduced runtime by 40%.
- Optimized transformations with **Spark SQL** and AWS tools, improving efficiency **40%** and built **Power BI** dashboards for business analysis.
- Developed automation scripts in **Bash/Shell** to manage pipeline monitoring and reduce manual interventions.

Associate Data Engineer

Oct 2020 - Dec 2023

Cognizant, Molina Healthcare Inc.

Remote

- Boosted AnG pipeline speed **57%** (7h → 3h) by optimizing **Spark/Scala** on **Databricks** with joins, caching, and repartitioning.
- Migrated legacy pipelines from **Cloudera** to **Azure Cloud**, implementing **CI/CD** using **Azure Databricks, Azure Data Factory**, ADLS for scalability and maintainability.
- Developed and maintained optimal **ETL/ELT pipelines** processing **500M+ records/week** into data models using **SSIS, Python, SQL, Airflow** and performed data validation.
- Cut query times **30%** via **PySpark UDFs** and partitioned **Hive(HiveQL)** tables for large scale transformations
- Reduced backlog **30%** (Membermonth) & **20%** (AnG) by implementing data quality triggers, governance workflows, and root cause analysis, improving SLA compliance 20% with **Agile methodology** using JIRA.

Data Analysis Intern

Feb 2020 - Oct 2020

Cognizant

Remote

- Built system from scratch, migrated **MySQL** relational database to **Hadoop** via **Sqoop**, used **HBase** for NoSQL storage, and enabled real time analytics with **Spark Streaming**.
- Executed **ETL** testing on pipeline built via **AWS** and **Terraform** ensuring seamless transition and data accuracy

PROJECTS

AWS MailMatrix | *Python, HTML, CSS, AWS Lambda, RDS, IAM, SES, S3, EventBridge, VPC, SQL*

Jan 2025

AviationProject | *MySQL, Hadoop, Cloudera, Apache Spark, Hive, HBase, Sqoop, Tableau*

Jun 2020

CERTIFICATIONS

Microsoft Certified: Azure Data Engineer Associate • Microsoft Certified: Azure Data Fundamentals • The Data Scientist's Toolbox