

# BHANU TEJA KURAKULA

United States | [bhanutk2@illinois.edu](mailto:bhanutk2@illinois.edu) | [linkedin](#) | [github](#) | [Portfolio](#) | +1-(217)926-2758  
[AWS Solutions Architect-Associate](#) | [AWS Machine Learning - Specialty](#) | [Databricks Certified Data Engineer](#)

## SUMMARY

Data & AI Professional with 2+ years experience in building scalable data pipelines that can handle high volume quality data. Experience in the full data spectrum—from engineering workflows to delivering predictive insights that matter.

## EDUCATION

### University of Illinois, Urbana-Champaign

Master of Science - Information Management (Data Science & Analytics pathway)

August 2023 - May 2025

(GPA 3.96/4.0)

**Relevant Coursework** : Global Health Informatics, Data Governance, Data Statistics & Informatics, Data Warehouse and BI

### Amrita University

Bachelor of Technology - Computer Engineering

July 2018 - June 2022

(GPA 3.35/4.0)

## SKILLS

<b>Programming Languages</b>	: Python, SQL, Spark, R, JS
<b>Platforms &amp; Frameworks</b>	: Databricks, AWS, Alteryx, SQL Server, Snowflake, Confluent Kafka
<b>Tools</b>	: Tableau, Airflow, Terraform, Git, Postgres, MongoDB, Grafana, PowerBI
<b>Libraries</b>	: Langchain, Numpy, Pandas, Scikit-learn, Matplotlib, Streamlit, Tensorflow, Plotly, Faust
<b>Miscellaneous</b>	: Jira, Github Actions, Excel, React JS, Azure Devops, HTML, REST APIs

## WORK EXPERIENCE

### Data Engineer Intern - AGCO Corporation

June 2024 - Present

- Engineered **10+ complex ETL pipelines handling 150GB+/month sales data on AWS**, standardizing across the EU.
- Implemented **data quality checks**, error handling mechanisms & unit tests ensuring data integrity across all stages of pipeline.
- Translated 20+ Alteryx workflows into PySpark jobs, **saving \$20K in licensing costs**, making it cost effective.
- Implemented **infra provisioning by( IaC )** & other best devops practices reducing deployment time by 30%.
- Optimized object storage management by partitioning data, improving data retrieval efficiency and query performance.
- Developed over 15+ Tableau prep ETL workflows and a scoreboard as part of migration effort.
- Collaborated with cross-functional teams to build data maps and documented user manuals to ensure smooth handover.
- Designed **10+ Azure DevOps** dashboards to monitor the team progress, impact, cost, expenses etc.

**Technologies Used**: Alteryx, Tableau, Prep, SSMS, SQL Server, Oracle, AWS(Glue, S3, Athena, CloudWatch, SNS, RDS)

### Data Engineer(AI & Analytics) - CTS

September 2022 - July 2023

- Enhanced production warehouse reliability by **overseeing 200+ metrics and 100+ jobs** ensuring uninterrupted data flow.
- Improved performance by writing **optimized SQL queries** resulting in 20% reduced query execution time.
- Collaborated with the upstream team to **reduce Tableau scorecard delay by 15-20 minutes** on a year over year basis.
- Presented a ML Prototype on fraudulent data claims to over 10+ stakeholders demonstrating potential integration.

**Technologies Used**: Snowflake, Grafana, Tableau, SQL, Airflow, Python

### Data Engineer Intern - CTS

March 2022 - July 2022

- Engineered an **ETL pipeline** for a dataset with **4M+ records**, leveraging various **GCP services** as part of L&D program.
- Implemented a batch data pipeline using **Delta Live Tables (DLT)** to process a **2GB** dataset, created interactive dashboards using Databricks SQL for data visualization and reporting.

**Technologies Used**: Databricks, Pyspark, SQL, Excel, GCP(GCS, Looker, BigQuery, Functions, Data Fusion, Dataflow)

### Data Science and Analytics Intern - PhysioAI

January 2022 - February 2022

- Enhanced user experience by developing a real-time yoga assistant to accurately detect and classify over **30+ yoga poses**.
- Designed and implemented a **dimensional data model to track user progress across 40+ metrics**.

**Technologies Used**: Postgres, SQL, Python, Tensorflow, Streamlit, Data Modeling, Deep Learning, Machine Learning

## PROJECTS

**Decentralized App for Covid-19 Prognosis using AI** | Dash-Plotly, Tensorflow, Flask, ReactJs, Ganache, Solidity  
Developed a POC to securely store medical records, trained **Convolutional neural networks** models with transfer learning.

**Credit Card Rewards Pipeline** | Terraform, Docker, ECR, Pyspark, Redshift, Glue, S3, Lambda, Looker Studio, Prefect  
Developed a serverless pipeline, transformed data, optimized schema, orchestrated workflow & visualized insights.

**Kafka-Based Real-Time Data Streaming Pipeline** | Confluent Kafka, docker

Developed a Kafka-based pipeline, processed data, ensured privacy compliance, and provided real-time insights.

**Coupon Accepting MLOps Pipeline** | MLFlow, EC2, S3, RDS, Terraform, Docker, Prefect, Github Actions

Built a MLOps pipeline, enabling experiment tracking, workflow orchestration, and automated deployment using **CI/CD**.