**Krishna Deepika**

[**deepikachavali.777@gmail.com**](mailto:deepikachavali.777@gmail.com) **| 862-930-6997**

**Summary**

* **Certified Azure and AWS Data Engineer** with 4+ years of experience designing and delivering **cloud-native data pipelines** and distributed data systems across AWS, Azure, and GCP.
* Proven expertise in **ETL/ELT workflows**, real-time streaming with Kafka/MSK and Spark Structured Streaming, and **scalable pipeline orchestration** with Airflow and Argo.
* Skilled in **Generative AI (Gemini, Agentic) integration**, embedding LLM-powered automation into pipelines for anomaly detection, intelligent reporting, and advanced analytics.
* Strong foundation in **data warehousing and modeling** (Snowflake, Redshift, BigQuery, Synapse), schema design, lineage tracking, and **query optimization** for performance at scale.
* Experienced in **containerization (Docker, Kubernetes)** and **CI/CD (GitHub Actions, Jenkins, Terraform)** for secure, automated deployments in production environments.
* Adept at **cross-functional collaboration**, translating business requirements into **production-grade data products** that drive measurable business value.
* Hands-on background in **data governance, metadata management, and BI enablement** using PostgreSQL, Cassandra, Power BI, and Tableau.

**TECHNICAL SKILLS**

| Programming & Scripting: | Python, SQL, Java, Scala, R, BashBig |
| --- | --- |
| Data & Streaming: | Apache Spark, Kafka, Flink, Kinesis |
| Cloud & ETL Tools: | AWS (S3, Lambda, Glue, Redshift, SageMaker), Azure (ADF, Blob, Synapse), Airflow, dbt |
| Data Warehousing & Modeling: | Snowflake, Redshift, BigQuery, Synapse, Star/Snowflake Schema Design, Partitioning, Clustering |
| Data Pipelines & Orchestration: | Batch & Real-time Pipelines, Apache Airflow, Argo, dbt |
| DevOps, CI/CD & Infra-as-Code: | GitHub Actions, Jenkins, Terraform, Docker, Kubernetes, Git |
| MLOps & AI Infrastructure: | SageMaker Pipelines, MLflow, Model Monitoring, AI/LLM Infrastructure, LLMOps, RAG Pipelines |
| Vector & NoSQL Databases: | Pinecone, FAISS, Cassandra, PostgreSQLData |
| Visualization & BI: | Power BI, Tableau, Grafana |
| Collaboration & Productivity: | JIRA, Slack, Confluence, Trello, Notion |
| Business & Communication: | Stakeholder Engagement, Requirement Translation, Data Governance, Documentation |

**EDUCATION**

* Master of Science, Data Science and Business Analytics. University of North Carolina at Charlotte.
* Artificial Intelligence. Winter Program**.** Asia University, Taiwan.
* Bachelor of Technology**.** SRM University, India.

# **CERTIFICATION**

* AWS Certified: Data Engineer Associate. May 2025
* Microsoft Certified: Azure Data Engineer Associate | Nov 2024
* Coursera: Machine Learning in Production | Feb 2022
* Microsoft Verzeo : Machine Learning/AI Intern
* Coursera: Python (Crash course/ Intermediate/ Advance)

# **PROFESSIONAL EXPERIENCE**

# **Acer America** **July 2024 – Present**

# **Role: Data Engineer**

# **Responsibilities:**

* Designed and deployed cloud-native data pipelines using Apache Spark (RDD/DataFrame) and AWS Glue, reducing daily processing latency by 62% across 5M+ device telemetry records.
* Built real-time ingestion systems using Amazon MSK (Kafka) and Kinesis Data Streams, enabling sub-minute analytics for performance alerts and device monitoring.
* Orchestrated 85+ production pipelines using Apache Airflow with SLA monitoring, auto-retry logic, and integrated Slack/JIRA alerts to ensure high reliability.
* Modeled and maintained analytical datasets in Amazon Redshift, S3, and PostgreSQL, incorporating schema versioning, lineage tracking, and access controls for ML and BI use cases.
* Developed automated data quality checks with Great Expectations and Python, reducing anomalies by 40% and improving reporting trustworthiness.
* Collaborated with data scientists to deploy predictive failure detection models using Amazon SageMaker, achieving a 25% improvement in proactive maintenance accuracy.
* Led phased migration from on-prem ETL to S3, Lambda, and Step Functions, boosting scalability and cutting operational costs by 45%.
* Implemented Terraform-based IaC and built GitHub Actions CI/CD pipelines for automated testing, deployment, and rollback of production workflows.

**Acer America Jan 2024 – Jun 2024**

**Roles: Data Engineer**

# **Responsibilities:**

* Built and deployed scalable ETL pipelines using Python and Apache Airflow, optimizing data refresh cycles and reducing batch latency by 28% across critical sales and inventory systems.
* Developed high-performance SQL queries and Redshift materialized views for Power BI dashboards, reducing report load times by 35% and enhancing executive decision-making.
* Engineered cloud-based data ingestion layers using Amazon S3, AWS Lambda, and Amazon Event Bridge, supporting a 20% increase in data volume with stable throughput.
* Created robust data validation frameworks using Pandas, custom rule engines, and anomaly detection scripts, leading to a 40% drop in data inconsistencies and higher stakeholder trust.
* Contributed to the migration of legacy ETL systems to Amazon Redshift and AWS Glue, resulting in a 50% reduction in maintenance overhead and significantly faster query performance for analytics teams.

**University of North Carolina at Charlotte Aug 2023 – Dec 2023**

**Roles: Graduate Research Assistant**

# **Responsibilities:**

* Supported academic teams in building automated ETL pipelines for research and departmental reporting using Python, SQL, and Airflow.
* Developed Snowflake data models and orchestrated cloud-based ingestion workflows using dbt and AWS S3, simulating production-grade architecture.
* Built a real-time analytics pipeline using Kafka + Spark Structured Streaming, delivering mock e-commerce insights to PostgreSQL and Power BI dashboards for visualization.
* Simulated an IoT data monitoring system using Flink, Kinesis, Cassandra, and Grafana, enabling real-time telemetry tracking and performance alerting.
* Managed infrastructure-as-code for deployment using Terraform, and integrated CI/CD pipelines via GitHub Actions for automated testing and release control.

**Cooper Standard March 2020 - June 2022**

**Role: Data Analyst**

# **Responsibilities:**

* Performed data analysis and created insightful reports for the client, helping improve operational efficiency by 15% through data-driven decision-making.
* Utilized SQL and Python to extract, clean, and analyze large datasets (over 50,000 records), ensuring data integrity and accuracy.
* Developed interactive dashboards using Power BI, enabling real-time tracking of key performance indicators (KPIs) for stakeholders.
* Conducted exploratory data analysis on customer data, identifying trends and insights that led to a 10% increase in customer satisfaction.
* Collaborated with cross-functional teams to define and implement data strategies, improving reporting timelines by 20%.
* Automated manual reporting processes using Python scripts, reducing reporting time by 30%.
* Identified data anomalies and inconsistencies, investigating and resolving issues to maintain high data quality standards.
* Contributed to the preparation of monthly and quarterly performance reports, providing insights into trends and future projections.

**Mphasis July 2019 - Feb 2020**

**Role: Data Analyst**

# **Responsibilities:**

* Assisted in gathering, cleaning, and transforming large datasets for analysis using Excel, SQL, and Python, improving data quality by 30%.
* Conducted exploratory data analysis (EDA) to uncover insights and trends, providing actionable recommendations that led to a 15% increase in team efficiency.
* Developed automated reporting dashboards using Power BI, reducing manual report generation time by 40%.
* Collaborated with cross-functional teams to understand data requirements and translated them into clear analytical solutions, enhancing communication between departments.
* Utilized SQL to query relational databases and performed complex joins and aggregations to support business decisions.