

JANKI RANA

AWS Data Engineer

📞 +1 (289) 941-1912 | ✉️ janki.mahida30@gmail.com | 🌐 Portfolio

PROFESSIONAL SUMMARY

Data Engineer with **5+ years** of experience designing and building scalable data pipelines and distributed data systems using **AWS** and **Apache Spark**. Expert in **PySpark**, **Airflow**, and **cloud-based data lake architectures**. Proven track record of processing large-scale datasets (500GB+ daily), optimizing data workflows, and enabling data-driven decision-making. **AWS Certified Solutions Architect** with strong expertise in **ETL development**, **data modeling**, and **performance tuning**.

CERTIFICATIONS

- **AWS Certified Solutions Architect – Associate**
Validation No: b2910348269449e4ac43933bcb1d0ae2 | [Verify Certificate]
- **AWS Solution Architect** – Udemy (Stephen Marrak)
Certificate No: UC-90270908-6e68-4e51-9fed-fa36883c2706

SKILLS

- **Languages:** Python, PySpark, SQL, Shell Scripting
- **Data Engineering:** ETL Pipelines, Data Lakes, Data Warehousing, Data Modeling (Star Schema), Data Quality
- **Big Data Technologies:** Apache Spark, Hadoop (HDFS, Hive), Kafka
- **Cloud Platforms:** AWS (Glue, S3, Athena, Redshift, Lambda, EKS), Azure
- **Workflow Orchestration:** Apache Airflow
- **DevOps/IaC:** Terraform, Docker, Kubernetes, Jenkins, GitHub Actions
- **Databases:** PostgreSQL, SQL Server, MongoDB, Cassandra, DynamoDB
- **BI Analytics:** Tableau, Power BI, Excel

PROFESSIONAL EXPERIENCE

Senior Data Engineer

Wipro Limited

Oct 2023 – Present

Canada (Remote)

- Designed and developed scalable **ETL pipelines** using **AWS Glue** and **PySpark**, processing **500GB+ of data daily**
- Built and optimized **data lake architecture** on **AWS S3** with **Athena**, improving query performance by **40%**
- Developed batch and **near real-time** data pipelines integrating **Kafka**, **AWS Lambda**, and **Redshift**
- Implemented efficient **data partitioning** and indexing strategies, reducing processing time significantly
- Migrated legacy ETL workflows from **IBM DataStage** to **AWS Glue**, improving throughput by **35%**
- Automated infrastructure provisioning using **Terraform**, reducing deployment time by **50%**
- Optimized **Spark jobs** by tuning memory usage, **parallelism**, and **data skew** handling
- Ensured **data quality** and validation using **Python-based frameworks**
- Collaborated with data analysts and business teams to design **scalable data models**

Data Engineer

Infosys Limited

Jan 2021 – Dec 2021

India (Remote)

- Developed **ETL workflows** using **Python** and **IBM DataStage** for enterprise data integration
- Integrated large-scale datasets into **Hadoop ecosystem (HDFS, Hive, Spark)**
- Performed **query optimization** and **performance tuning** on big data pipelines
- Built **reusable data transformation scripts** to improve efficiency
- Ensured high **data accuracy** through validation and cleansing techniques

Junior Data Engineer

Dec 2018 – Dec 2020

Tata Consultancy Services (TCS)

India (Remote)

- Developed and maintained **ETL pipelines** for **enterprise data warehouse** systems
- Assisted in **data transformation**, mapping, and integration processes
- Performed **debugging** and **troubleshooting** of ETL workflows
- Supported deployment and **production releases** of data pipelines
- Improved pipeline performance through **code optimization**

EDUCATION

Seneca College, Newnham Campus

Jan 2023 – Aug 2023

Graduate Certification in Cloud Architecture & Administration

GPA: 3.8 / 4.0

Conestoga College, Waterloo Campus

Jan 2022 – Aug 2022

Graduate Certification in Big Data Solutions Architecture

GPA: 3.65 / 4.0

Gujarat Technological University, Gujarat, India

Jul 2014 – Aug 2018

Bachelor of Engineering (Computer / Information Technology)

CGPA: 9.22 / 10.0

PROJECTS

Real-Time Data Pipeline (Kafka + AWS) | *Kafka, AWS Glue, S3, Redshift*

- Built end-to-end streaming pipeline: Kafka → AWS Glue → S3 → Redshift
- Enabled near real-time analytics with scalable architecture
- Implemented partitioning and performance optimization techniques

AWS Data Lake Architecture | *S3, Glue Data Catalog, Athena*

- Designed scalable data lake solution using S3, Glue Data Catalog, and Athena
- Implemented schema evolution and efficient query optimization
- Reduced reporting latency and improved data accessibility