

Divya Singh

Data Engineer

8318879987 | divyasingh71039@gmail.com | [linkedin.com/in/divya-singh-88644a226](https://www.linkedin.com/in/divya-singh-88644a226) |

SUMMARY

Results-driven Data Engineer with 2+ years of experience in designing and implementing big data solutions across Azure, AWS, and Databricks. Expertise in ETL pipelines, data warehousing, and cloud migrations, with a strong command of PySpark, SQL, and distributed computing. Proven ability to optimize data workflows (10x performance gains) and lead cross-cloud data projects while ensuring security and scalability.

TECHNICAL SKILLS

Languages: Python, SQL, PySpark, Scala (basic)

Big Data Tools: Apache Spark, Hadoop, Hive, Kafka, Airflow, Delta Lake

Cloud Platforms: Azure (Databricks, Synapse, Data Factory, Data Lake), AWS (S3, Glue, Lambda, Redshift)

Data Engineering: ETL/ELT, Data Warehousing, Streaming (Structured Streaming), Data Modeling)

DevOps/CI-CD: Git, GitHub, AWS CodeCommit, Databricks Workflows

EXPERIENCE

Associate Data Engineer

Feb 2023 – Current

Celebal Technologies.

Jaipur, India

- **Project:** Cross-Cloud Data Pipeline Implementation for Adhesive Brand
- Developed and managed a cross-cloud data integration pipeline involving Azure and AWS services.
- Implemented data ingestion from Azure Synapse to AWS S3 using Azure Data Factory with purging logic.
- Designed workflows for handling daily, weekly, and monthly incremental files.
- Utilized AWS Glue for ETL processing across stage and core schemas, including archival of processed files.
- Automated data movement from stage to core schema using AWS Lambda and Step Functions.
- Enabled alerting and monitoring through AWS SNS.
- Managed version control and collaboration using AWS CodeCommit.
- **Technologies Used:** Azure Synapse, Azure Data Factory, Azure Data Lake Storage, AWS S3, AWS Glue, AWS Lambda, Step Functions, SNS, Amazon Redshift, CodeCommit.
- **Project:** UC Migration for a Leading Car Manufacturing Company
- Migrated various types of tables (managed and external) and views from a source system to a Hive catalog in Databricks.
- Successfully migrated Delta Live table pipelines and jobs, ensuring minimal disruption to business operations.
- Migrated notebooks of jobs and DLTs, as well as compute resources, ensuring a seamless transition and maintaining data integrity.
- Utilized a UC-enabled Databricks workspace and external location permissions to facilitate the migration of external tables, ensuring secure and efficient data transfer.
- Coordinated with cross-functional teams to ensure a smooth migration process, minimizing downtime and ensuring business continuity.
- Developed and implemented a comprehensive testing strategy to validate data integrity and pipeline functionality post-migration.
- **Technologies and Tools Used:** Databricks, Hive, Delta Live, UC-enabled workspace, PySpark.
- **Project:** Hive and Spark Migration to Databricks
- Contributed to the project's progress through active participation in the Code Review Process.
- Improved code readability by implementing code changes and removing repetition, resulting in a 20% increase in efficiency.
- Optimized Spark jobs, achieving 10x better performance for critical jobs and reducing job completion time from 4+ hours to 0.5 hours.
- Reduced pre-production unit testing time by 50% for enhanced reliability of Hive, Hudi, and Avro table DDL migration.
- Migrated around 3,500 Hive jobs and around 650 Spark jobs to Databricks.
- Migrated over 500 Airflow DAGs to Databricks.
- Implemented table ACLs on external Hive Metastore to provide and restrict relevant access to users.
- **Technologies and Tools Used:** Databricks, Airflow, GitHub, AWS S3, PySpark.

EDUCATION

Abdul Kalam Technical University

Bachelor's of Electronic and Communication Engineering

Lucknow, India

2018 - 2022

CERTIFICATIONS

- **Databricks Certified Data Engineer Professional - Databricks**
- **AZ900 - Azure Fundamentals**
- **DP900 - Azure Data Fundamentals**
- **Databricks Lakehouse Fundamentals - Databricks**
- **Databricks Generative AI - Databricks**