

SHIVA PANDEY

Data Engineer || +91-8878229150 || pandey.shiva623@gmail.com || Noida (NCR)
<https://www.linkedin.com/in/shiva-pandey-285644310/> || <https://github.com/iamshivapandey>

Professional Summary:

- Data Engineer with almost 2 years of experience in building and optimizing big data pipelines using Python, PySpark, AWS, Databricks, Kafka, and Iceberg. Skilled in databases, ETL processes, data lakes, data warehousing, migration and SQL.

Technical Skills:

- **Languages:** Python (pandas, numpy, boto3, selenium, matplotlib, seaborn etc.), Java, SQL, C.
- **Operating Systems:** Linux, Windows.
- **Version Control & SCM:** Git, Github, Bitbucket.
- **Big Data Tools:** Spark (PySpark), Kafka, Airflow, Hadoop, Hive, HDFS, AWS (Glue, Athena, ECS, Lambda, EC2, S3, EMR, SQS, CloudWatch, Fargate), Databricks, Snowflake, Iceberg, ETL, ELT, Databases (MySQL, PostgreSQL, ClickHouse), Data Migration, Data Lake, Data Pipeline, Data Warehousing, Azure Data Factory (ADF), Azure Functions, Power BI, Tableau.

Work Experience:

- **Demandhelm** 04/2023 to Present
 - **Data Engineer**
 - Built a scalable big data pipeline using AWS Glue, Iceberg, Pyspark, Fargate, Lambda, S3, SQS and Snowflake.
 - Led the database migration from MySQL/PostgreSQL to Apache Iceberg and migrated 350 ETL scripts from Pandas to PySpark, boosting processing efficiency by 80%. Deployed revamped ETLs on AWS ECS (Fargate), reducing costs by 70%.
 - Worked on creating queries for backend according to the requirements from developers.
 - Developed and optimized Amazon Marketing Cloud (AMC) queries and transformation scripts, utilizing Python and AWS Lambda for ETL processes into Iceberg.
 - Automated a few ETL scripts using Airflow on ec2.

Projects:

- **Real Time Weather Data Pipeline** 30 Days
 - **Tech Used:** Kafka, APIs, Python (pandas, requests), Spark, MySQL, SQL, S3, Linux, Git.
 - Built a real-time weather big data pipeline using Python, Apache Kafka, and AWS to process big data. Fetched live data via RapidAPI, streamed through Kafka to AWS S3, processed with PySpark for cleaning and filtering. Stored the cleaned data in MySQL for further analysis.
- **Pneumonia Detection Website** 30 Days
 - **Tech Used:** Python (stream lit), Aws Ec2, TensorFlow, Linux, Git.
 - Developed a machine learning model using TensorFlow to detect infections in human lungs and integrated it into a Python-based Stream lit web application. Deployed the application on AWS EC2 for accessible, real-time diagnosis.
- **Docker Image of Spark Setup** 2 Days
 - **Tech Used:** Docker, Python, Notebook, Spark, Java, Git.
 - Configured Apache Spark with Python in a Docker environment, enabling seamless setup and reuse. Created a Docker file and built an image and pushed to docker hub to simplify deployment, allowing users to run Spark without repeated setup challenges.
- **Crypto Twitter Bot for Daily Updates** 4 Days
 - **Tech Used:** Python (selenium, pillow), Web Scraping, APIs, Git, GitHub Actions, Cron.
 - Built a Twitter bot with Python Selenium to scrape cryptocurrency data and post daily updates via the Twitter API, automated with GitHub Actions.

Certifications:

1. NPTEL Cloud Computing
2. Databricks
3. Amazon Marketing Cloud
4. Hacker Rank SQL

Education:

1. **Cdac-Acts Pune** 03/2023 to 08/2023
PG Diploma in Big Data Analytics || 7.3 CGPA
2. **Madhav Institute of Technology & Science Gwalior** 06/2020 to 06/2022
Masters in Computer Application (MCA) || 8.5 CGPA

Soft Skills:

- Effective Communicator, Collaborative Team Leader, Hard Working, Co-operative.

Date:

Shiva Pandey