

Vandana Arya

Big Data Engineer

Skilled Data Engineer with 3 years of expertise in Big Data for delivering impactful solutions for enhanced data workflows and analytics

✉ vandana.arya7749@gmail.com

☎ 7349080397

📍 Bangalore

🌐 [linkedin.com/in/vandana-arya-30072b1b9](https://www.linkedin.com/in/vandana-arya-30072b1b9)

WORK EXPERIENCE

Associate Engineer Detect Technologies

11/2022 - Present

Achievements/Tasks

- Implemented lakehouse architecture using delta lake to support both transactions and analysis on the same system removing data redundancy
- Enhanced performance of spark jobs by **37%** analyzing execution through spark UI, optimizing the resources and following strategic optimization techniques reducing **29%** processing time
- Closely collaborated with **2** cross functional team, providing data lake solution on HDFS for raw, clean and processed data, with Hive tables facilitating a **36%** efficiency for cleaning and processing pipelines
- Led a POC for resource calculation in a team of **2**, resulting in a **25%** improvement in query performance

Engineer - Cloud and Infra Services Larsen & Toubro Infotech

01/2021 - 11/2022

Bangalore

Achievements/Tasks

- Mentored a team of **4** new joiners, imparting business logic and implementation skills, leading to **27%** improvement in overall team efficiency
- Implemented **3** end-to-end data ingestion and processing pipeline using ADF extracting data from AWS S3, MySQL, http into ADLS, processing with Databricks and populating in Azure SQL Database
- Executed comprehensive data cleaning with Databricks enhancing efficiency for **2** cross-functional teams
- Migrated data from on-prem to Azure cloud using ADF

RPA Engineer Unilever Industries Pvt. Ltd.

01/2020 - 05/2020

Bangalore

(Internship)

- Efficiently automated **4** tasks accelerating processing time by **56%** and conserving **9** man hours

EDUCATION

Bachelor of Engineering Visvesvaraya Technological University

08/2016 - 08/2020

CGPA - 8.5

Electronics and Communication Engineering

- Developed Unique Face Identification system improving accuracy up to **95%** enhancing the face detection based security system

SKILLS

Pyspark

Apache Spark

HDFS

Hadoop

Spark-SQL

Python

SQL

Pytest

Agile

Performance Tuning

Distributed Processing

Azure Databricks

ADLS

Data Structures

Git

ADF

Hive

Azure Unity Catalog

Delta Lake

Jenkins

Lakehouse Architecture

Data Modelling

ACHIEVEMENTS

Alumni Services Coordinator (10/2023)

Conducted interviews with **50** college final year students equipping for real-world job interviews

Orated Career Insights, Fostering Success (10/2023)

Led impactful career awareness session for over **500** final-year students bridging academic-industry gap for successful career transitions with best practices

Outstanding Performer (06/2022)

Recognized by the team lead for achieving top performance in a team of **11** members, demonstrating outstanding proficiency and dedication

Published an IEEE journal on unique face Identification based on ML (09/2020)

Journal presents a face detection system with ML algorithms, enhancing security system accuracy by **95%**

CERTIFICATES

Azure Databricks - Delta Engine and Delta Architecture (01/2024 - Present)

Knowledgeable about Delta Engine and Delta Architecture implementation on Azure Databricks

Distributed Processing with Pyspark (10/2023 - Present)

Certified for data processing with pyspark high level APIs, spark and yarn architecture with other functionality for developing highly performant queries

Distributed Storage and Processing Fundamentals (09/2023 - Present)

Holds a good knowledge for data storage and processing in a distributed environment

Microsoft Certified: Azure Data Fundamentals (DP 900) (2021 - Present)

Foundational knowledge of core data concepts and implementations using Microsoft Azure data services

LANGUAGES

English

Full Professional Proficiency

Hindi

Native or Bilingual Proficiency