tusharkumarsahoo08@gmail.com
+91-9438633513
+44-7810389349
linkedin.com/in/tusharkumarsahoo

Tushar Kumar Sahoo

With nearly 5 years of experience in Data Engineering, I've developed expertise in Big Data Architectures, Data Pipelines, Data Warehousing, Data Modelling, and CI/CD Pipelines. I've worked with Apache Spark, Azure Databricks, and Azure Analytics Services within Data Mesh architecture. My dynamic and resourceful approach has significantly reduced costs by optimising and streamlining data processing workflows for a leading oil and gas supermajor. I've successfully led collaborative teams of up to 5 in delivering complex data solutions. I'm eager to leverage my knowledge to drive impactful results and contribute to innovative projects.

Technical Skills

Apache Spark . Databricks . PySpark . Delta Lake . Data Warehousing . Data Modelling . SQL . Python . Azure Data Factory . Azure Data Lake . Microsoft Azure . Azure DevOps . Git . Bazel . Docker . Bash

Soft Skills

Team Leadership . Workload Prioritisation and Delegation . Lateral Thinking . Cross Cultural Competence . Adaptability

Awards

The Chosen One Data Award
Infosys & bp - Dec 2023
- Awarded for making magic happen and delivering exceptional results at UDX.

Client Champion UK Rise Award
Insosys - Aug 2023
- Honoured as Client Champion for delivering breakthrough ideas and solutions.

The Avengers Data Award
Infosys & bp - Dec 2022
- Acknowledged as the Avengers team for outstanding efforts in implementing solutions and resolving issues.

Insta Award 2022
Infosys - Sep 2022
- Commended for persistent problem solving and navigating through complex challenges.

The Eccentric Performer Data Award
Infosys & bp - Dec 2021
- Distinguished as an eccentric performer for pushing boundaries and thought leadership.

Insta Award 2021
Infosys - Jun 2021
- Recognised for developing high quality pipelines and quick learning.

Certifications

- Databricks Certified Data Engineer Associate
- Databricks Certified Lakehouse Fundamentals
- Microsoft Certified Azure Data Fundamentals (DP-900)
- Microsoft Certified Azure Fundamentals (AZ-900)
- Fivetran Certified HVR Associate

Education

Bachelor of Technology in Production Engineering
Veer Surendra Sai University of Technology, Burla, Odisha, India
Jul 2015 - Jun 2019
- Honoured with Certificate of Merit as the second-best student in Production Engineering with a CGPA of 8.75.

Work Experience

Technology Analyst

Infosys Limited I Canary Wharf, London, UK
Jul 2022 - Present
- Lowered Databricks costs by 41% by identifying inefficiencies through billing and usage metrics, implementing Spark optimisations, and refining data processing architectures and warehousing models, which decreased monthly expenses from $1.1M to $679K, resulting in projected annual savings of $4.8M and potential for a further 30% reduction.
- Led a team of 5 in the Unified Data eXperience (UDX) project, managing the coding lifecycle from development to deployment for a 12M LOC monorepo by leveraging Bazel to automate builds and testing, which boosted developer velocity, eased version management, and maintained coding standards, with the work showcased at the DATA AI Summit in June 2024.
- Facilitated the migration of 278 Databricks workspaces by analysing and addressing complexities, managing edge cases specific to the organisation, and refining the migration approach to minimise downtime and mitigate impacts on data and access controls, resulting in enhanced data governance, discoverability, and auditability.
- Engineered and managed batch data pipelines and warehousing solutions that integrated data from Workday and ADP to centralise workforce metrics and enhance data quality, thereby empowering the People and Culture unit to optimise staffing and achieve an 11% decrease in employee turnover.

Senior Systems Engineer

Infosys Limited I Bengaluru, Karnataka, India
Oct 2021 - Jun 2022
- As a pod lead, guided a team of 4 to expedite deliveries and meet high business demands for onboarding use cases within the Data Mesh architecture by gathering requirements, supporting the management team in addressing technical and architectural gaps, clarifying technical aspects for team members, prioritising and delegating tasks, and spearheading high-priority issues.
- Leveraged Databricks AutoLoader to simplify near real-time streaming ingestion of around 100 GB of data per day from SAP ECC, eliminating the need for Azure Event Hubs, reducing cloud costs by approximately 33%, and enabling the Production and Operations business unit to analyse inventory levels and optimise logistics more cost-effectively.
- Designed and implemented a data warehouse model and pipelines to integrate employee travel data from Egencia and SAP Concur, which enabled more effective analysis of travel expenses and policy compliance, improved vendor negotiation strategies for the Procurement department, and reduced travel costs by 14% within 6 months.

Systems Engineer

Infosys Limited I Bengaluru, Karnataka, India
Oct 2019 - Sep 2021
- Crafted a suite of 12 efficient, modular, and metadata-driven ingestion pipelines using Azure Data Factory and Azure Databricks for loading data into Open Lakehouse, implementing a plug-and-play framework that supported various data sources such as Azure Blob Storage, Amazon S3, SharePoint, SFTP, Oracle DB, Azure SQL DB, SAP BW, and Salesforce APIs, thereby providing a scalable solution for diverse and evolving data integration needs.
- Streamlined operations by automating secret renewal and failure notifications with Microsoft Graph, and Databricks APIs, slashing manual tasks by 50% and boosting data pipeline reliability with timely alerts and updates.
- As the youngest member of the team, worked on migrating a legacy fleet card transaction management application to Azure and notably contributed to the real-time data migration of over 300 databases from HP Non-Stop Enscribe and Oracle to Azure SQL Database using Gravic Shadowbase and Fivetran HVR.
- Optimised data replication with Fivetran HVR, minimising compute time by 37% and accelerating the real-time migration of fleet transaction data, which enabled more timely and accurate insights in the new cloud environment.