Vibhavari Bellutagi

Data Platform Engineer

vibhavari-bellutagi | vibhavari.bellutagi@gmail.com | Bengaluru | +91 9742352994

SUMMARY

Data Platform Engineer with 9+ years of experience building Databricks, Spark, Flink, and Kafka-based platforms and scalable data pipelines on AWS. Reduced pipeline runtimes by 80%+, DBU costs by 60–75%, and infrastructure setup from days to minutes using Terraform. Experienced in MCP-enabled Claude automation, governance, observability.

SKILLS

Data Platforms & Processing: Databricks, Apache Spark (PySpark), Apache Flink (FlinkSQL), Kafka (CDC)

AI / LLM Engineering: Claude, Model Context Protocol (MCP), LLM agents, RAG

Cloud & Infrastructure as Code: AWS, Terraform (reusable modules, RBAC/governance as code)

Languages: Python, SQL, Java

Backend & APIs: FastAPI (REST APIs), Node.js

DevOps & Observability: CI/CD (Jenkins, GitHub Actions), pipeline monitoring and alerting (Slack integrations)

PROFESSIONAL EXPERIENCE

Data Platform Engineer

Jul '25 - Aug '26

Mews

France Remote

Built and automated the Databricks platform supporting Mews Business Intelligence, covering infrastructure provisioning, governance, pipeline performance, and AI-assisted operations.

- Re-architected Databricks infrastructure management across three production workspaces using Terraform, Python, & GitHub Actions, placing approximately 80% of SQL Warehouse provisioning & role-based access control policies under version control, **reducing change lead time from weeks to days & manual administration by 50%**
- Optimised critical production data pipelines supporting Mews business intelligence using Databricks & Apache Spark, cutting processing time by over 80% (from 10+ hours to 2 hours) & **database unit compute costs by 60–75%**, reducing infrastructure budget pressure through architectural improvements, query optimisation, & compute tuning
- Provisioned reusable Terraform modules & production-ready Databricks & machine learning operations templates, enabling 5–10 internal teams to independently provision infrastructure, **reducing setup time from days to minutes & cutting deployment delays by 75%**
- Connected Claude to Databricks through Model Context Protocol to automate infrastructure operations & monitor 30+ production pipelines, delivering daily health summaries to Slack, reducing manual monitoring & reporting time by approximately 35%, & cutting monitoring errors by 30%
- Automated workload monitoring & Slack status reporting through Claude across 100+ production workflows, **eliminating approximately 20+ hours of manual status checks** & reducing time to identify failed runs

Data Engineer

Sep '21 - Dec '24

Agile Lab

Aix-en-provence, France

- **Led a collaborative effort** with data engineers and scientists to implement advanced machine learning models in production, resulting in a **50% reduction in deployment time** and substantial enhancements to operational efficiency
- Streamlined ETL pipelines to enhance data quality for ML models, slashing data processing time by **60%** through the utilisation of **Apache Spark** and **AWS Analytics** services.
- Designed and implemented a comprehensive data pipeline using FlinkSQL & Kafka CDC connector to process 100s of GBs of diverse customer data, resulting in a substantial **75-80% decrease in update times for customer profiles.**

- Achieved **2-3 second query execution times** while ensuring GDPR compliance through precise anonymisation rules with Livy, Spark, and Presto, enhancing operational efficiency and data security.
- Spearheaded knowledge transfer sessions and developed a team Wikipedia on Confluence, streamlining onboarding and boosting **new member integration efficiency by 75%**.

Data Engineer Sep '18 - Aug '21

Deloitte USI Bengaluru, India

- Developed and executed a robust ETL pipeline using Hadoop, Spark, Hive, and Sqoop to process 30GBs of data daily, resulting in a **30% enhancement in data-driven decision-making**
- Created reusable orchestration workflows for end-to-end pipelines using the UC4 automation tool, leading to a ~40% reduction in the duplication of workflows.
- Enhanced ETL pipeline auditing efficiency by deploying comprehensive audit data logging, leading to a **40% decrease in issue detection and resolution time**, resulting in expedited troubleshooting and optimised system performance.

Software Engineer Jul '17 - Sep '18

Attra Infotech Bengaluru, India

- Strategised comprehensive solutions for analysing ~100k credit card transaction data during the initial Proof of Concept (POC) using tools like Hive, Sqoop, and HDFS.
- Implemented optimisation techniques such as partitioning, bucketing, & compression to automate Hive external table creation & enable dynamic data loading from HDFS, resulting in a 20% time savings.

EDUCATION

Bachelor of Engineering Jun '11 - Jul '15

Sapthagiri college of Engineering Bengaluru, India

CERTIFICATIONS

- Databricks Data Engineer Associate Certification | Databricks | June-2024
- AWS Data Analytics - Specialty | AWS | Dec-2023