

Vibhavari Bellutagi

 vibhavari-bellutagi

 vibhavari.bellutagi@gmail.com

 +33752209047

 Portfolio

Data Engineer

SUMMARY

Seasoned Data Engineer with 8 years of experience in software development, specialising in data engineering for over 4+ years. Skilled in designing and implementing scalable ETL pipelines, optimising data infrastructure, and delivered high-impact projects across diverse industries. Proficient in building data pipelines, cloud-based solutions, and leveraging Data Processing tools/technologies on AWS to drive business efficiency and impact.

SKILLS

Big Data Technologies: Hadoop, Spark, Hive, Presto | **Programming Languages:** Python, Scala (with Spark) | **Data Collection Systems:** Kafka | **Data Warehouse:** Postgres, Redshift | **Cloud:** AWS - EMR, Glue, S3, Lambda, API Gateway, Sagemaker, DynamoDB, Athena | **IaaS:** Terraform | **CI/CD:** Jenkins, GIT | **API Development:** FastAPI (REST APIs)

PROJECTS

Industrialised Claims Recovery Data Pipeline on AWS

Jan '24 - Dec '24

AgileLab

Remote, France

Situation

- Automated claims processing to improve recovery rates.
- Data scientists built NLP models, but they weren't production-ready.
- Needed a scalable, industrialized pipeline on AWS.

Task

- Convert Jupyter notebooks into a production-ready AWS pipeline.
- Automate and scale claim processing.
- Ensure data quality and monitoring.

Action

- Refactored code, added tests, and built a deployment strategy.
- Used SageMaker (NLP processing), S3 (storage), Glue (extraction).
- Orchestrated with Step Functions, Lambda, and tracked claims in DynamoDB.

Results

- Processed ~30k claims/month with 20% growth.
- Achieved ~40% claim recovery, adding ~€200K in financial gains.
- Delivered a scalable, automated, and reliable solution.

Tech Stack: Python, AWS Step Function, Amazon Lambda, AWS Glue, Amazon S3, Data quality, Pytest, Data Monitoring, Git, Jenkins.

Industrialised NLP-Based Policy Document Knowledge Base

Mar '23 - Dec '23

AgileLab

Remote, France

Situation

- The company required a policy document knowledge base for quick retrieval by stakeholders and agencies.
- The NLP model, built for semantic and similarity search, was not production-ready.

Task

- Productionize the NLP knowledge base with automation and scalability.
- Develop an API framework for seamless front-end integration.
- Ensure real-time search, reliability, and minimal manual intervention.

Action

- Extracted policies using AWS Glue and automated workflows with Step Functions.
- Built a FastAPI backend for front-end access.
- Refactored ML code, added unit and integration tests, and automated deployment with Jenkins CI/CD.

Results

- Indexed 30,000+ policy documents for real-time search.
- Eliminated manual document review, improving efficiency.

- Reduced manual deployment effort by 60%, enabling faster updates.

Tech Stack: Python, AWS Step Function, Amazon Lambda, AWS API Gateway, FastAPI, AWS Glue, Amazon S3, Pytest, Git, Jenkins

Optimizing Customer Data Processing with FlinkSQL

Jun '22 - Feb '23

AgileLab

Remote, France

Situation

- The Digital Integration Hub needed real-time customer data updates.
- Spark batch jobs caused delays of minutes to hours.
- FlinkSQL enabled near real-time processing with exactly-once guarantees.

Task

- Convert Spark SQL queries to FlinkSQL-compliant queries.
- Optimize queries for performance and efficiency.
- Develop a custom CDC connector and enable incremental processing.

Action

- Optimized queries with predicate pushdown and salting to reduce skew.
- Resolved join issues and improved performance.
- Optimized state storage using HDFS savepoints and entropy injection on S3.

Results

- Data updates in under a minute, improving user experience.
- Reduced cloud costs by 30% with query optimization and incremental processing.

Tech Stack: FlinkSQL, EMR, S3, Git, Github Actions, Python, Data Processing, Data Engineering

Querying Anonymised Data

Sep '21 - May '21

AgileLab

Remote, France

Situation

- Needed real-time data anonymisation with role-based access control for GDPR compliance.
- Users required a REST API to submit queries securely.

Task

- Implement on-the-fly anonymisation with user-role-based access.
- Ensure fast query execution while maintaining security and compliance.

Action

- Deployed Spark jobs on Livy, enabling REST API-based queries.
- Created metadata-driven anonymisation rules for role-based access control.

Results

- Achieved 2-3 second query execution, improving efficiency.
- Ensured GDPR compliance with secure, role-based data access.

Tech Stack: Apache Spark, Livy, Presto

Optimizing Network Connectivity Analytics with ETL Pipelines

Sep '18 - May '21

Deloitte USI

Bengaluru, India

Situation

- The client needed a Network Connectivity Analytical solution to improve electrical connectivity accuracy and storm readiness.
- Required processing of 30GB+ data daily from remote monitoring devices using data analytics.
- Manual workflow inefficiencies led to duplication and slow troubleshooting.

Task

- Build a scalable ETL pipeline for processing large datasets efficiently.
- Automate workflow orchestration to reduce redundancy.
- Improve audit logging for faster issue detection and resolution.

Action

- Developed an ETL pipeline using Hadoop, Spark, Hive, and Sqoop, handling 30GB daily.
- Created reusable UC4 orchestration workflows, cutting workflow duplication by ~40%.
- Implemented audit data logging, reducing issue detection time by 40% for faster troubleshooting.

Results

- 60% reduction in workflow duplication, increasing operational efficiency.

- Delivered a scalable ETL with audit data logging

Tech Stack: Apache Spark, Hive, SQL, Oracle UC4 Orchestration, Sqoop

Credit Card Transaction Analysis (POC)

Jul '17 - Jun '18

Attra Infotech [Client - Synchrony]

Bengaluru, India

Situation

- The company was new to data engineering and needed a POC to start analyzing its data.
- The goal was to process and analyze ~100K credit card transactions efficiently.
- Required a scalable data processing framework to validate feasibility.

Task

- Develop the company's first data engineering POC using Hive, Sqoop, and HDFS.
- Optimize data ingestion and processing to improve efficiency.

Action

- Designed a Hive-based solution for analyzing ~100K transactions.
- Implemented partitioning, bucketing, and compression to automate external table creation and enable dynamic data loading.

Results

- Successfully delivered the company's first data engineering POC, enabling future scalability.

Tech Stack: Apache Hive, Sqoop, SQL

CERTIFICATIONS

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

Testimonials

[Sara Tamburini, HR Director, AgileLab \(01/2025 \)](#)

During her tenure with us, Vibhavari made significant contributions to the organisation, demonstrating exceptional technical expertise, problem-solving skills, and professionalism. Her primary responsibilities included:

- Designing, developing, and maintaining big data pipelines to process and analyse large volumes of data efficiently.
- Collaborating with cross-functional teams to implement scalable data solutions tailored to client requirements.
- Optimising data models and ensuring data quality and integrity across projects.
- Conducting performance tuning of big data systems to enhance efficiency and reliability.

Vibhavari exhibited advanced skills in technologies such as Apache Spark, AWS, Python and demonstrated strong analytical capabilities that added value to numerous projects. Her proactive approach and ability to work effectively in dynamic environments were commendable. We found Vibhavari to be a highly motivated and diligent professional who consistently went above and beyond expectations. She was a valued member of our team, and her contributions were instrumental in achieving our organisational objectives.

Vincenzo Pafundi, ML Engineer @ Axa Italy ([on LinkedIn](#))

I had the pleasure of working with Vibhavari for nearly two years on Data Engineering and MLOps projects. From the very beginning, she impressed me with her deep technical expertise and her ability to work across multiple areas of a project—ranging from backend development to CI/CD pipelines and code testing. Her strong background as a software engineer was evident in the quality and efficiency of her work.

Beyond her technical skills, Vibhavari is highly autonomous, proactive, and has a remarkable long-term vision that helps shape solid roadmaps and drive high-quality results. She is not only a skilled engineer but also a fantastic team player, always fostering collaboration and sharing knowledge with others. Her strong interpersonal skills make her a great asset to any collaboration.
