

Shubham Shakya

☎ (+91) 9773539888 | ✉ shubhamshakya61@gmail.com | 📺 [Shubham](#) | 🌐 [shubham23471](#)

EXPERIENCE

Fractal Analytics

Jan 2022 – Present

Senior Data Engineer

- **Designed and developed a high-performance, cloud-agnostic DLT (AUTOSAR) log parser**, transforming **terabytes** of hex-encoded logs into a structured, human-readable tabular format. This solution served as a **centralized platform** throughout the organization, allowing seamless access to decoded DLT logs with full-text search capabilities using **ELK**.
- Led a team of **3 Data Engineers** to **design and implement a custom analytics framework**, enabling data engineers and analysts to apply business logic to processed logs. This solution provided tailored insights to **stakeholders, product owners, testers, and developers**, resulting in a **30%** improvement in debugging efficiency and reducing incident resolution time by **40%**.
- These **two solutions** accelerated the **application development lifecycle** by **40%** for **developers** and **testers** while reducing logs decoding **cost** by **70%**.
- Conducted multiple **brown bag sessions** and office hours to support testers and developers in adopting our cloud-based DLT Viewer alternative, driving a **3x increase** in user adoption within **4 months**.
- **Designed and developed a self-service platform** using **Spark Operator on Kubernetes**, enabling users to submit and execute custom Spark jobs seamlessly via **self-service portal**.
- **Led the migration** of Spark workloads from **Databricks to AKS**, managing the full infrastructure setup, including provisioning AKS clusters, configuring node pools, and integrating storage mounts with NFS and Blobfuse.
- Built **CI/CD pipelines** with **Azure DevOps** to automate the image build process, deployment, and monitoring of **Spark applications on AKS**, ensuring seamless operations and scalability.
- This **migration** improved **system performance and cost efficiency**, leading to a **20%** increase in **user adoption** as more teams leveraged the **self-service Spark platform on AKS**.

Tech Stack: *Python, PySpark, PostgreSQL, Azure Service Bus, Kafka, Docker, Kubernetes, ELK, Azure Databricks, Azure Data Factory, ETL, ELT, Azure Functions, Azure DevOps, Azure Delta Lake, Delta Tables.*

Ericsson

Aug 2021 – Dec 2021

Data Engineer

- Responsible for gathering requirements to setup design solution for ingesting data from SFTP location to **Apache Cassandra** using **PySpark**.
- Build end-to-end data ingestion pipelines for both near to real-time processing and batch processes.
- Data pipeline migration from **SPSS** to **Google Data Fusion (CDF)**.

EXL

June 2019 – Aug 2021

Data Engineer

- Collaborated with an onshore team of **Data Scientists** to understand **data requirements** for running **six predictive models**, ensuring efficient data preparation and processing.
- Developed and optimized **PySpark, Python, and Hive scripts**, incorporating **business rules** and setting up the **ETL process** on **on-premises servers**, improving data processing efficiency.
- Managed **monthly production runs**, handled **backfills**, and resolved **production bugs**, ensuring smooth data pipeline operations.
- **Developed a Flask-based backend** to consume **JSON** data from **Apache NiFi**, process it using Python, store the results in Hive, and execute predictive models for advanced analytics.

Paytm

June 2016 – June 2019

Software Engineer

- **Automated Data Reporting:** Developed and executed **SQL queries** and **Python scripts** to generate **daily transactional reports**, reducing manual effort by **40%** and increasing report accuracy.
- **Data-Driven Insights:** Designed and automated **daily and monthly report generation** using **Python**, providing managers and stakeholders with **data-driven insights** on product growth.
- **Data Visualization and Cleaning:** Utilized **Python libraries** to clean, transform, and visualize **500,000+** sales data records, resulting in:
 - * **40% reduction** in manual data processing.
 - * Increased **data accuracy** and reliability.
 - * Enhanced **insights-driven decision-making** for stakeholders.

EDUCATION

G.L Bajaj Institute of Technology and Management

Bachelor of Electronics and Communication Engineering

Greater Noida

Aug 2012 – May 2016

TECHNICAL SKILLS

Languages: PySpark, Python, SQL (Postgres), Kafka, Azure Service Bus, Azure DevOps.

Frameworks: Flask, FastAPI.

Developer Tools: Git, Kubernetes, Docker, Azure Cloud, VS Code, Visual Studio, PyCharm.