

## PROFESSIONAL SUMMARY

---

- **Senior Data Engineer** with 7+ years of experience developing scalable data pipelines and Lakehouse architectures using **Databricks**, **Spark**, and **Snowflake**.
- Hands-on with **Azure services** such as **ADF** and **Delta Lake** for building reliable, automated data workflows.
- Mentored and guided cross-functional data engineering teams, establishing best practices in **ETL design**, **version control**, and **performance optimization** to deliver scalable, high-quality data solutions.
- **Databricks-certified engineer** with hands-on experience in building **Delta Lake pipelines**, automating workflows, and optimizing **Spark** for large-scale analytics.
- Passionate about building **cloud-native, data-driven systems** that enhance data reliability, accelerate insight generation, and empower teams to make informed, impactful business decisions.
- Experienced in designing end-to-end data solutions that **improve data quality**, **reduce pipeline latency**, and enable **real-time business intelligence**.

## TECHNICAL SKILLS

---

- **Languages:** Python, PySpark, Scala, Java, SQL, Shell Scripting
- **Data Ingestion and Processing:** Spark, Kafka, Snowpipe, Airflow, ADF
- **Cloud and Storage:** Snowflake, Databricks, HDFS
- **Database Technologies:** PostgreSQL, Hive, SQL Server, HBase
- **DevOps and Version Control:** Git, CI/CD
- **Performance and Governance:** Data Quality Validation, Cost Optimization, Metadata Management
- **Operating Systems:** Linux, Windows

## EXPERIENCE

---

- **Senior Data Engineer** Bangalore, India  
*May 2024 – Present*  
*IBM*
  - **Migration to Real-Time Snowflake:** Architected the migration from **Databricks batch pipelines** to a real-time **Snowflake ingestion** framework, reducing data latency from 4 hours to under 10 minutes.
  - **Continuous Ingestion:** Designed and implemented **Snowpipe + Streams + Tasks** for continuous ingestion from S3, improving data freshness and scalability.
  - **Cost Optimization:** Optimized transformation logic and SQL queries, achieving **30% cost reduction** and **40% faster report refresh times**.
  - **DBT Model Development:** Built modular **DBT models** for curation and validation, providing clean, auditable datasets for analytics and compliance.
  - **Automated Deployments:** Automated deployments by integrating **DBT** with **Airflow** and **CI/CD**, enabling version-controlled data releases.
  - **Collaboration:** Partnered with **data science** and **product teams** to align datasets with key business metrics, accelerating insight delivery.
  - **Quality and Alerting:** Implemented **data quality** and **alerting frameworks** to proactively detect anomalies in production data streams.
  - **Mentoring:** Mentored and supported **data engineering teams**, sharing best practices in **Databricks**, **Spark** optimization, and workflow automation to enhance productivity and code quality.

- **Modak Analytics (Client: GSK)**

Hyderabad, India

Dec 2017 – Mar 2024

*Software Development Engineer*

- **Metadata Ingestion Framework:** Built a **metadata ingestion framework** to capture operational metadata from **GDrive** and **S3** into **PostgreSQL**, reducing manual effort by 60%.
- **Kafka-Based Pipelines:** Developed **Kafka-based pipelines** processing 500K files daily and curating semi-structured data into **Hive** datasets for global analytics.
- **Scalable ETL Jobs:** Created scalable **Spark/Scala ETL jobs** handling terabyte-scale datasets with **35%** faster processing.
- **Migration to Spark-Scala:** Led a team in converting existing **Java-based ETL processes** into **Spark-Scala frameworks**, improving data processing efficiency and reducing runtime by 25%.
- **Team Development:** Conducted technical interviews and onboarding for data engineering candidates, building a high-performing team aligned with project standards and best practices.
- **Databricks Pipelines:** Collaborated with **Azure** teams to operationalize **Databricks pipelines** using **ADF** for reliable orchestration.
- **Monitoring and Alerts:** Implemented **error monitoring** and **email alert routines**, reducing production incidents by 40%.
- **Requirement Translation:** Worked closely with **business stakeholders** to translate requirements into maintainable and scalable ETL.

## CERTIFICATIONS

---

- Databricks Certified Data Engineer Associate
- Databricks Data Lakehouse Fundamentals
- Databricks Azure Data Platform Architect
- Coursera – Data Analysis, Management and Visualization

## EDUCATION

---

- **National Institute of Technology**

Calicut, India

*B.Tech – Computer Science and Engineering*

2013-2017