

# SAGAR D. SHINGARE

## Senior Data Engineer | Data Architect



+91 9960890489



me@sagarshingare.com



[LinkedIn](#)



Pune, India

### PROFESSIONAL SUMMARY

Senior Data Engineer and Data Architect with overall 14+ years and 8+ years of end-to-end experience delivering **high-scale data platforms, cloud-native ETL/ELT systems, and distributed processing solutions** across Life Sciences, CPG, Retail, Insurance, and Enterprise technology domains. Specialized in **Spark (Scala/Python), Databricks, Hadoop ecosystem, BigQuery, AWS EMR, Delta Lake, and modern Lakehouse architectures**. Adept at building **3 TB+ pipelines**, improving workload efficiency, and engineering reusable frameworks that strengthen data reliability, performance, and analytical readiness. Recognized for strong data modelling expertise, orchestration design, and cross-functional collaboration across complex enterprise environments.

### CORE COMPETENCIES

- **Big Data:** Spark (Core, SQL, Streaming), Hive, HDFS, YARN, MapReduce
- **Cloud Platforms:** AWS (EMR, S3, Lambda, Batch), GCP (BigQuery, Dataflow, Cloud Functions, Composer), Azure (ADF, Databricks)
- **Data Architecture:** Delta Lake, Iceberg, Medallion Architecture, CDC, SCD, Data Quality Frameworks
- **Data Engineering:** ETL/ELT Pipelines, Data Modelling (ERD/Dimensional), Orchestration, Observability
- **Programming:** Scala, Python, SQL, Core Java, Shell Scripting
- **Automation:** Airflow, Composer, CI/CD, Git, Terraform
- **Ecosystem Tools:** Kafka, Snowflake, DBT, Qlik Sense

### PROFESSIONAL EXPERIENCE

#### Data Architect

#### Tiger Analytics

Jan 2022 – Present

- Architected and implemented enterprise-grade data engineering solutions for Life Sciences, Retail, Insurance, Consumer Tech, and AI/ML-driven programs.

#### Key Contributions

- Designed scalable ingestion, ELT/ETL, and transformation pipelines using Spark, Databricks, BigQuery, AWS and GCP-native orchestration tools.
- Developed **standardized and reusable data engineering frameworks**, enabling consistency and faster delivery across large multi-domain initiatives.
- Built production-ready schemas, data models, and medallion-layer architectures supporting analytics, BI, and ML teams.
- Engineered efficient, cloud-optimized data workflows improve reliability, performance, and maintainability across distributed environments.
- Collaborated with architects, data scientists, and product teams to deploy end-to-end data solutions aligned with business goals.

#### Big Data Developer

#### Cognizant

Jan 2020 – Dec 2021

- Delivered high-volume data workflows for CPG forecasting and Life Sciences analytics using Azure Databricks, AWS EMR, and Cloudera.

## Key Contributions

- Built ingestion and analytics pipelines for a global **Sales & Marketing Promotional AI** platform using PySpark and Databricks.
- Designed ERD structures and data models supporting scalable forecasting algorithms.
- Standardized business data from Teradata and enhanced data quality for analytical consistency.
- Integrated **26+ diverse Life Sciences data sources** using Spark (Scala) on Cloudera for Oncology analytics.
- Automated ETL processes using Python and optimized Spark jobs for better runtime efficiency.
- Supported cloud migration initiatives by deploying and tuning pipelines on AWS EMR.

## Technology Specialist

### Infosys

*Apr 2014 – Dec 2019*

- Engineered core components of the Infosys Internal Product Data Lake using Hadoop-based technologies.

## Key Contributions

- Developed ingestion workflows using Spark, Hive, and Sqoop to support enterprise reporting and analytics.
- Performed validation, transformation, and harmonization of ingested data through internal data models.
- Managed Hadoop cluster performance, troubleshooting production issues and ensuring reliable batch execution.
- Migrated data from Oracle to Hadoop using optimized Sqoop and Hive processes.
- Automated deployment and release processes, improving delivery efficiency for data teams.

## Software Developer

### Digital-EDU

*Jan 2011 – Feb 2014*

- Delivered backend systems, mobile applications, and automation solutions for educational institutions.

## Key Contributions

- Enhanced SIMS database reliability and optimized data operations for institution-wide usage.
- Built Linux shell scripts for server configuration, monitoring, and automation.
- Developed mobile applications using Java, integrating NFC and GPS functionalities for attendance and transport tracking.
- Deployed customized solutions across **30+ institutions**, ensuring adoption and workflow alignment.

## EDUCATION

- **Master of Computer Science** – Pune University, 2013
- **Bachelor of Computer Science** – Solapur University, 2009
- **HSC** – Pune Board, 2005
- **SSC** – Pune Board, 2003

## PERSONAL DETAILS

- **Location:** Punawale, Ravet, Pune – 411033
- **Passport:** B6090466