

# Satyam Kumar

+91 7042746946 | satyamkumar2107@gmail.com | [LinkedIn](#)

## Technical Skills

---

**Programming Languages :** Python, SQL

**Big Data Technologies :** Spark , Pyspark , Spark SQL , Kubernetes , Kafka , Hive

**Cloud platforms :** Amazon S3, AWS EC2 , AWS Glue, AWS EMR , AWS Redshift, Databricks, ADLS , ADF

**Orchestration & DevOps Tools :** Airflow, Kubernetes, Git, Docker, Linux, Data Modelling

## Experience

---

### Deloitte - Data Engineer

Feb 2023 - Sep 2025

- Built reusable data blocks for stakeholder use cases such as product eligibility, behaviour insights, UX optimization, and digital signature adoption, enhancing data delivery speed and flexibility for analytics and campaign targeting.
- Built an end-to-end ingestion and transformation pipeline to move and prepare data from DynamoDB to S3, helping replace manual paper forms with digital signatures.
- Enabled Event-driven data processing with AWS Glue and S3, speeding up data availability for analytics by 40%. Optimized client data models by reducing schema complexity by ~50%, which improved query performance and reduced storage costs.
- Implemented an automated alerting mechanism using Power Automate to notify team members via Microsoft Teams upon AWS Glue job failures, reducing reliance on email monitoring and enabling faster incident response.
- Re-engineered and optimized legacy ETL jobs from Informatica to PySpark, improving performance by ~40% and ensuring compatibility with AWS-native architecture.
- Built scalable ingestion pipelines to standardize raw vendor and drug order data in a staging layer on S3, ensuring schema consistency and data quality.
- Designed and implemented Airflow DAGs to fully automate pipeline orchestration, reducing manual intervention by 80 -90% and ensuring timely data refreshes.
- Owned and executed UAT, identifying and resolving critical issues ahead of production cut-over.
- Independently developed a low-code internal tool using Power Apps + Power Automate to streamline assessments, improving efficiency across business units.

### Concert AI - Data Engineer

Sep 2025 - Present

- Converted a Python notebook into a fully optimized AWS Glue + PySpark pipeline by rewriting the entire logic in Spark , improving throughput , reducing ML model runtime by ~65%, and delivering 3× faster processing on multi-GB datasets.
- Automated creation of schemas by cloning 200+ tables, reducing environment setup time from 1–2 hours to <10 minutes.
- Designed pipelines to union data across three large schemas (~300M rows) and built dedicated breast-cancer schemas, improving data completeness to 99%+ and query performance by ~40%
- Built QC validation schemas and optimized Redshift SQL operations, reducing production defects by ~50% and improving job stability by 90%
- Automated Precision360 Excel export (multi-sheet), reducing manual work from 45–60 minutes → 2–3 minutes, improving analyst productivity by ~95%.

## Education

---

### Jain University

Master of Computer Applications ( MCA )

Bangalore, Karnataka

March 2023 - May 2025

### CMR University

Bachelor of Computer Applications ( BCA )

Bangalore, Karnataka

July 2019 - May 2022

## Certifications and Awards

---

- **AWS Certified Cloud Practitioner**
- **AWS Certified Solutions Architect – Associate**
- **Databricks Data Engineer Certificate**
- Nominated and Awarded **[Spot Award]** for developing "Internal ILA assessment" tool for Deloitte.
- Awarded **[Applause Award]** for outstanding impact as a Data Engineer for driving development resulting in significantly enhanced business data insights, decision-making and cost saving.