

# Ashutosh Samal

7004827527 | ashutosh06.samal@gmail.com | [Linkedin](#) | [leetcode](#)

---

A dedicated individual Motivated by data and results,with four years of experience in the data domain. Ready to join a team of critical thinkers to build any data stack from scratch

## EDUCATION

### Indian Institute of Technology Patna

*Executive Master of Technology in Data Science and Engineering*

Online

2024 Jan – 2026 Jan

### National Institute of Technology Karnataka

*Bachelor of Technology in Electrical and Electronics*

Karnataka, India

2017 Aug – 2021 April

## Relevant Courses

Computer Science,Distributed Sytem, Data Structures and Algorithms, Object-Oriented Programming (OOP), Design and Analysis of Algorithms, Data Science, Machine Learning,statistics

## EXPERIENCE

### Walmart Global Tech,Bangalore : Data Engineer III 2021 July - Current

#### ODS (Operational Data Store) | *Scala,Spark Streaming,GCP*

- Migrated high priority batch jobs to structured streaming reducing ETA for crucial data by 80 percent.
- Enhanced monitoring, alerting and validation framework for streaming pipelines, which helped reduce pipeline breakdowns from 5 incidents to 0 monthly
- Architected a high-performance ETL streaming solution leveraging Spark to manage upwards of 10 terabytes of critical operational data; improved overall system reliability while enabling faster access to actionable insights for stakeholders.

#### Time series forecasting: | *Python,EDA,Regression,statistics*

- Designed and deployed a regression model to forecast the hourly fulfillment capacity of Walmart's distribution center based on available resources and other relevant parameters.
- Conducted Exploratory Data Analysis (EDA),which informed the development of the regression model for forecasting hourly fulfillment capacity.
- Implemented model and data drift checks to monitor performance and determine when model retraining is necessary to maintain accuracy over time.

#### IMS(Inventory management systems) | *Scala,Spark,Hive,GCP,Hudi*

- Set up multiple batch pipelines which handles 100GB's of data every 15 Mins
- Built compression utility for raw data enabling org wide savings of 70% in storage costs
- Spearheaded the integration of Kafka Connect across the team, facilitating seamless streaming of over 10TB of data daily while introducing self-healing mechanisms that reduced system downtimes by an average of 30 minutes per incident.

#### Tech Modernisation: | *Spark,Hive,GCP,scala,Python,Scripting*

- Helped my team to Migrate 100+ Workflows running on cloudera's Hadoop cluster to google data proc cluster.
- Modified the existing data lake to reduce dependency on BigQuery by leveraging Apache Hudi, resulting in a reduction of cloud costs from \$200K/month to \$120K/month.

### Walmart Global Tech, Bangalore: Data Science Intern 2021 Jan - April

- Designed a super-resolution model for text detection and extraction in collaboration with a data science team.
- Implemented a GAN network after researching and analyzing multiple papers to optimize model performance.
- Developed a custom loss function to compare generated and true images, capturing structural and perceptual differences more effectively. This improvement led to better training efficiency and enhanced model accuracy.

## TECHNICAL SKILLS

**Languages:** Python, Scala, SQL

**Big Data skills:** Spark, Spark Structured streaming, Hadoop, Hive, Data modeling, Airflow, Kafka, PySpark, Kafka Connect

**Cloud Services :** Google Cloud Platform (GCP), GCS, Dataproc, BigQuery, Spark Serverless

**Data science skills:** Deep learning,statistics,Machine learning,NLP,keras,Data analysis

**Database :**Azure Sql, Hana, Cassandra