Ashutosh Samal

7004827527 | ashutosh06.samal@gmail.com | Linkedin | leetcode

INTRO

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

EDUCATION

National Institute of technology Karnataka — Karnataka, India
Bachelor of technology in Electrical and electronics — *Aug. 2017 – April 2021*

Indian Institute of technology Patna — Online
Executive Master of technology in Data science and engineering — *Jan. 2024 – April 2026*

EXPERIENCE

Data Engineer 3, Walmart Global Tech,Bangalore 2021 - Current
Building robust ETL pipelines for moving, exposing and visualising big data. Understanding data needs from customers and providing optimised solutions.

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

- Migrated high priority batch jobs to structured streaming reducing ETA for crucial data by 80 percent.
- Enabled monitoring, alerting and validation framework which helped reduce pipeline breakdowns from 10 incidents to 0 monthly
- Designed and implemented ETL data pipeline handling around 1 TB operational data in Spark

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

- Implemented batch pipelines which handles 100GB's of data every 15 Mins
- Designed a config based solution, enabling new pipelines to be made within 1 day
- Split the pipeline to hot (10% highly used data) and cold data bringing data latency from 8 hours to 2 minutes for hot data (99.5% reduction)
- Lead and did sizing of kafka connect for my whole team to process huge volume of data

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

- Proactively led Development Cycle Optimisation processes
- Helped my team to Migrate 100+ Workflows running on cloudera's Hadoop cluster to google data proc cluster.
- Worked on modifying existing datalake to use HUDI file format to reduce the dependency on big query and reduce cloud cost

Data science intern , Walmart global tech, Bangalore 2021 Jan - April
Worked With data science team in the Field of computer vision to plan and design Super resolution model for text detection and extraction

- Read and analyzed multiple research papers and models to optimize our model

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: Pytroch,Sklearn,Deep learning,statistics,Machine learning ,Pandas,NumPy
Database :Azure Sql, Hana, Cassendra

PERSONAL PROJECTS

Mask Detection and alerting system | *Python,Pytorch, openCv,MongoDB*

- Designed two Deep learning models to detect the persons face and then to recognize if a person is wearing a mask or not
- Trained a face recognition model on a dataset so that we can detect the person's ID who is not wearing a mask.
- Maintained a database with person's ID and other contact details to send automated SMS if someone is not wearing a mask.

Data engineering on YELP dataset | *Spark,Hive,GCP,scala*

- Designed end to end data pipeline in GCP for yelp dataset which have five kinds of data which can be found here
- Worked on modeling final data warehouse and implemented SCD2 and other technique in the final table.