Ashutosh samal

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

INTRO

A dedicated individual Motivated by data and results,with one 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*

EXPERIENCE

Data engineer — July 2021 – Present
Walmart global tech — *Bangalore, India*
- Planned and Designed a grafana dashboard for my team which plot near real time data of our Pipelines and other systems such as Kafka lag , data lag in our big query tables ,unprocessed data in data lake,cluster metrics etc which helped our team to monitor for any failure.
- Helped my team to Migrate our Workflows running on cloudera's Cloud break cluster to google data proc cluster. Used automic scheduler to schedule the Workflows and modified the workflows to be ephemeral in nature which reduced our overall cost of running the Workflows.
- Provided L1 support to our team during holiday seasons.
- Working on designing ETL pipeline using Scala and spark to ingest data into Big query and Hive from Microsoft SQL server.Also working on data modelling for the tables in hive and big query.

Data science intern — Jan 2021 –june 2021
Walmart global tech — *Banaglore, India*
- Worked With data science team in the Field of computer vision to plan and design Super resolution model for text images
- Used Pytorch and designed a super resolution model which helped in text extraction from blurry images
- Read and analyzed multiple research papers and models to optimize our model

PROJECTS

Mask Detection and alerting system | *Python,Pytorch, openCv,MongoDB* — Aug 2020 – Dec 2020
- 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.

Real time Pipeline for ingesting data to Bigquery | *Spark,Kafka,Python,scala* — Jan 2022 – Present
- Setup a kafka server in GCP to get data from source and put it in a topic for real time processing and a Spark job to dump data from kafka topic to Google GS bucket for further process.
- Will be working on a ETL pipeline to process data and put it in Big query for real time analysis or visualization.

TECHNICAL SKILLS

Languages: Python, Scala, SQL
Big Data tools: Spark,Hadoop,HDFS,Hive,Big query,google Data proc
Data science skill: Pytroch,Deep learning,computer vision
Developer Tools: Git, Google Cloud Platform, PyCharm, IntelliJ,Docker,Automic scheduler
Libraries: pandas, NumPy, Matplotlib