Ganesh Ankulwar

Senior Technology Lead
5+ Years of Experience

Contact: (+ 91) 8983239677
Email id: ankulwarganesh10@gmail.com
LinkedIn: https://linkedin.com/in/ganesh-ankulwar/
Medium: https://medium.com/@ankulwarganesh10

Technical Skills:

- **Language**: Python, SQL and Java
- **Big Data:** Hadoop, Spark, Pandas, Databricks, Hive, Kafka
- **Web Framework**: Flask
- **Database:** MySQL, Elastic search, Redis, Redshift, Solr
- **DevOps:** Jenkins, Docker, GitLab CI and Kubernetes
- **OS:** Linux, Windows
- **Orchestration Tool:** Airflow
- **Version Control:** Bit bucket, GitLab
- **AWS Services:** S3, EMR, DMS, Kinesis, Lambda, SQS

Certificates and Achievements:

- *AWS Certified Solutions Architect Associate* (2/2020 - 2/2023)
- *Oracle Cloud Infrastructure Foundations Associate* (5/2020 – 11/2021)
- *SAFe 4 Practitioner* (08/2019 – 08/2020)
- *Won Extra Mile Award* (01/2018 – 01/2019)

Work Experience:

Knowledge Lens Pvt. Ltd.

- Senior Technology Lead   *October 2022 - Present*
- Technology Lead   *April 2021 - September 2022*
- Senior Big Data Engineer   *August 2019 - March 2021*
- Big Data Engineer   *August 2017- July 2019*

Projects:

Surveillance Smart:
An Analytical web app for the analysis of Product Complaints and Adverse Events.
- Lead Data Engineering team of 4 members to provide transformed data to the application.
- Design and developed **ETL data pipeline** to ingest data using Databricks.
- **Automated** and Standardized PySpark Databricks notebooks.

- Designed and implemented **architecture** of the Web application to host in Docker Containerized Environment.
- Productionized and improved execution performance of Machine Learning (ML) models like **SpaCy & PyTorch** in distributed framework by 45%.

HLEE (Holistic Lab Experiment & Execution):
A Near Real-Time data ingestion pipeline solution to insert JSON data into tables.
- Upgraded the critical parser component from Java to Python and optimized it.
- Performed POC on AWS DMS and Kinesis for real-time data ingestion to Redshift.
- Developed python script to **migrate historical data** into DB Delta and Redshift.
- Developed crucial **high-performance PySpark** script to ingest data to Data Lake.
- Designed and built a Real-Time data ingestion pipeline solution to ingest data of **100 GB** of data per hour.

FSIP (Field Sales Incentive Program):
A Data warehouse pipeline solution to generate and report incentives of the sales personnel.
- Optimized data pipeline solution by improving the performance by **40%.**
- **Optimized spark jobs** for better cluster utilization and for faster performance
- Developed **ETL PySpark scripts**.
- Developed various **DAG's** in **Airflow** for batch process.
- Designed and Implemented a data pipeline solution for processing **1.2 TB** of data.

Real World Evidence Platform:
A next-generation Analytics Platform for analysis of patient-level medical data.
- Oversaw the complete web services part.
- Built a high-performance analytical layer of the platform using **multithreading and multiprocessing**.
- Developed an export CSV and PDF modules in python which was an essential part of reporting.
- Created import modules for importing the CSV's to the platform.
- Build a script for **snapshotting and restoring** of the Elasticsearch cluster.
- **Designed and Optimized the Elasticsearch** data loading process using PySpark.

Educational Qualification:

- **PG- Diploma in Big Data Analytics** from C-DAC ACTS, Bangalore *(2017)*
- **B.E - Electronics** from K. J. Somaiya C.O.E, Mumbai *(2012-2016)*

Personal Dossier:

Date of Birth: 21-09-1994.
Linguistic Proficiency: English, Marathi, Hindi.
Address: Kharadi, Pune-411014