Apoorva Chandanshive

+91-9284028847 | apoorvachandanshive.sae.comp@gmail.com | in LinkedIn | ⎇ GitHub

Summary

Data Engineer with 1.5 years of experience designing **real-time data streaming pipelines** and **cloud-based data lakes** using AWS, Kafka, and PySpark. Proven success in reducing manual reporting by **70%** and improving data accuracy by **30%** for influencer marketing analytics. Skilled in automation, ETL orchestration, and scalable backend integration.

Education

Sinhgad Academy of Engineering   Pune, Maharashtra
*Bachelor of Engineering in Computer Engineering - **CGPA - 8.40***   *2020 – 2023*

Experience

Vhub.ai | Data Engineer   Dec 2024 – Present

- Engineered end-to-end YouTube real-time streaming pipeline using **AWS MSK (Kafka)**, **Kafka Connect**, and Python to ingest influencer metrics (views, likes, comments, subscribers). Designed **S3 sink connectors** with partitioning and **BigQuery** sync via **AWS Glue**, achieving 99.9% data completeness and 70% reduction in manual reporting for brands like Dot & Key.
- Designed S3 sink connectors with partitioning and implemented AWS Glue ETL jobs using PySpark for data cleansing, transformation, and aggregation before syncing to BigQuery, improving data accuracy by 30%.
- Developed **Python scripts** (Requests, Pandas, BigQuery) to track **Instagram ROI**, improving **data accuracy by 30%** and **reporting efficiency by 25%**.
- Automated data ingestion processes for partner analytics, tracking performance for brands like Dot & Key, ensuring 100% data completeness and improving efficiency.
- Built a scalable **sales dashboard backend** using Flask, BigQuery, and Firestore, reducing API latency by **40%** and supporting **500+ concurrent users**; implemented **JWT authentication** for secure data access.

Vodafone Intelligent Solutions (VOIS) | Graduate Engineer Trainee   Aug 2023 – Mar 2024

- Collaborated within the Development team to create **Python programs** for efficiently loading CSV file data into Oracle tables, reducing **data processing** errors by 20%.
- Played a key role in Conducting 10+ **Root Cause Analyses (RCAs)**, to determine system failure causes and implement preventive measures, following **Agile methodology** for timely project delivery.

Projects

AIS Datalake Project
- Built automated data pipeline for daily TV data ingestion (channels, schedules, user activity, billing) from S3/SFTP using AWS, Python, and Spark, implementing Bronze/Silver/Gold data lake architecture for analytics in Redshift
- **Technologies:** AWS S3, AWS Glue, PySpark, Amazon Redshift, Python, ETL Pipelines

Cloud Infrastructure Optimization ⎇
- Optimized cloud infrastructure by implementing Auto Scaling Groups and Elastic Load Balancer across multiple Availability Zones, reducing costs by 30% and improving system uptime to 99.9%
- **Technologies:** AWS EC2, Auto Scaling Groups, Elastic Load Balancer, CloudWatch, Availability Zones

TECHNICAL SKILLS

Programming & Scripting: Python, Unix Shell Scripting
Data Engineering & ETL: Databricks, Apache Airflow, PySpark, AWS Glue, ETL Pipelines, Data Lake Architecture(Bronze/Silver/Gold)
Databases & Warehousing: MySQL, BigQuery, Firestore, Amazon Redshift
Cloud & Infrastructure: Amazon Web Services (S3, Auto Scaling, Load Balancer, Availability Zones)
Analytics & Frameworks: Pandas, NumPy, Matplotlib, Flask