

Siddhant Pratap Singh

Uttar Pradesh, India

8840097852 | siddhant19pratap@gmail.com |
linkedin.com/in/siddhant-pratap-singh-9bb942280 | github.com/siddhant19pratap

Profile

Data Engineer with hands-on experience in AWS-based ETL pipelines, Python, SQL, and PySpark. Skilled in validation, and automation with a focus on delivering scalable, cloud-native solutions.

Education

•B.Tech - Information Technology

JSS Academy of Technical Education, Noida. 8.44 CGPA 2024

• Class XII

Woodbine Gardenia School. Percentage: 90.8 2019

Experience

•Motherson Technology Services Limited

July 2023 – Present
Noida

Engineer/ Data Engineer (1 year 1 month)

- Designed and maintained scalable ETL pipelines using AWS Glue, Athena, S3, Lambda, and CloudWatch, improving pipeline efficiency by 20% and ensuring timely data delivery across the analytics ecosystem.
- Streamlined master data ingestion and validation across dev and prod, achieving 100% accuracy during month-end cycles for 3+ critical business reports.
- Eliminated Excel-based dependencies from ETL workflows, reducing processing time by 30% and cutting manual error rates by 40%.
- Solely developed a PySpark-based pipeline to parse highly nested XML structures,semi-structured data into queryable formats using AWS Glue and Athena, now deployed in production and supporting automated reporting.
- Led monthly validation of high-visibility reports and collaborated with analysts and Power BI teams to define KPIs, improving report accuracy and trust.
- Collaborated with clients to gather change requests, clarify reporting requirements, and assist in deployment and post-production support alongside senior team members.

•Ambak.Com

March 24- May 24
Gurugram

Software Development Intern (3 Months)

- Developed REST APIs for data ingestion and implemented optimizations that improved data retrieval speed by 10%, while ensuring scalable and secure system design.

Projects

•ETL Pipeline Project for Global Automobile Manufacturer

Designed and managed scalable data pipelines to generate business insights from automotive sales data for enterprise clients.

- Tools & technologies used: AWS Glue, AWS Athena, AWS S3, AWS CloudWatch, AWS Lambda, PySpark, Python, SQL
- Developed and optimized ETL pipelines using PySpark on AWS Glue to process high-volume sales data, implemented monitoring with CloudWatch and Lambda, and delivered KPI-based insights for Audi, Skoda, and Volkswagen.

•Youtube Data Analysis using AWS

YouTube Data Analysis is an end-to-end data engineering project built on AWS to process and analyze trending video..

- Tools & technologies used: AWS Glue, AWS Athena, AWS SNS, AWS Catalog, AWS S3, Pyspark
- Processed 1,000+ trending video records using an AWS-based ETL pipeline to analyze viewer behavior and content categories for performance reporting.

•Crypto Mart

CryptoMart is a cryptocurrency tracker app developed for monitoring real-time cryptocurrency data

- Tools & technologies used: React JS, Material UI, and Chart JS, VS Code
- Built CryptoMart to track live crypto data via Coin Gecko API, visualizing price trends and market volume for 10+ coins, enhancing real-time decision-making.

Technical Skills and Interests

Languages: C/C++, Python, Java

Frontend: HTML/CSS, Javascript, React JS

Developer Tools: Git/Github, Google Collab, Power BI, VS Code, Jupyter

Skills: Data Engineering, Data Analytics, PySpark, Pandas, Numpy, Problem Solving, Debugging, Risk and Controls Awareness

Cloud/Databases: Amazon Web Services (AWS), MySQL, Data Lakes, Data Warehouses

AWS Services: AWS Glue, AWS Athena, AWS S3, AWS CloudWatch, AWS Lambda

Soft Skills: Leadership, Teamwork and Collaboration, Time Management, Problem Solving, Critical Thinking

Coursework: DBMS, DSA, OOPS, OS

Certifications: Informatica Data Engineering Certified, AWS Data Engineer Professional Certificate