



**Shubham Gupta**

**API / Cloud Engineer**

|  |  |
| --- | --- |
| **Experience Summary** | |
| * Data Engineer expertise in SQL, Spark, Python & AWS Certified Developer * Engage in the entire software development life cycle (SDLC), including requirement gathering, analysis, development, testing & deployment. * Utilized Apache Airflow to create robust and scalable workflows, ensuring seamless orchestration of data processes * Evaluate and devise enhancement strategies to boost key performance indicators (KPIs) around 40% across all business units * Created Lambda functions for event-driven processing, such as responding to new file uploads in S3 locations * Applied AWS Glue services to build and manage robust 10+ Glue ETL jobs, seamlessly integrating them with Apache Airflow * Leveraged AWS Glue's automatic schema discovery and evolution capabilities for handling complex data structures and changes * Employed Redshift as a data warehouse solution, empowering Tableau dashboards to present valuable insights to business users * Conceived Lambda functions for event-driven processing, checking file upload at S3 in last 6 hours and get notified through AWS SNS * Developed efficient Spark and Python code for data extraction, transformation, and loading, leveraging the power of Apache Airflow pipelines running on EMR clusters * Developed python http trigger, queue trigger and time trigger function * Gathering data from user or through Gmail API or twitter API and upload to blob queue to future process | |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Work Experience** | | | |
| **Organization Cognizant Technology Solutions** | | | |
| **Project Name** | KDH & CDLS | **Team Size** | 8 |
| **Start Date** | 23 DEC 2020 | **End Date** | 10th NOV 2023 |
| **Project Description** | Working in Kite pharmaceutical's Commercial Data Lake Setup & Kite Data Hub applications to maintain overall patient journey data with projected forecast and overall performance | | |
| **Role & Contribution** | * Engage in the entire software development life cycle (SDLC), including requirement gathering, analysis, development, testing, deployment, and documentation * Automating email notification using Lambda, SNS * Simplifying existing data pipeline using amazon internal ETL tool * Developed efficient Spark and Python code for data extraction, transformation, and loading, leveraging the power of Apache Airflow pipelines running on EMR clusters * Building data quality check for existing data pipeline * Identified performance bottlenecks, tune Query & Performance, Triage and resolve them. | | |
| **Technology & Tools** | AWS(Glue,lambda,EMR,S3,MWAA,Redshift), Spark, SQL, Python, Git, Jira, Agile  Project -2 | | |
| **Project Name** | | Ai sense- Azure | **Team Size** | 4 |
| **Start Date** | | 10th SEPT 2022 | **End Date** | 10th JAN 2023 |
| **Project Description** | | Project is based on Azure library that do Sentiment analysis and generate metadata files like Main words, transcribe, translate, Sentiment using azure text analytics library for all sort of files gathering data through API or directly. | | |
| **Role & Contribution** | | * Developed Dev, Test and prod env in Azure function app for project * Developed python http trigger, queue trigger and time trigger function * Gathering data from user or through Gmail API or twitter API and upload to blob queue to future process * Enhanced Security and implemented IP based security and key level security of http API | | |
| **Technology & Tools** | | Python, Azure function app, azure database, azure blob storage, VS code,  GitHub, docker | | |

|  |  |
| --- | --- |
| **Achievements / Awards** | |
| **Awards** | * Received “Team player” Award in the Year 2024 for the successful volunteering for Data Booth IgnAIte Event in Pune |