Akanksha Tripathi
Data Engineer

Cell: +91-8130520509
akanksha.tripathi1558@gmail.com

PROFESSIONAL SUMMARY

- 6 years of experience in all phases of SDLC including requirement gathering, application design and development.
- Experience in **Talend Data Fabric Tools** with expertise on **Talend** Basic components**, Data Profiling, Data Migration, Data processing, Orchestration, Parallelization, Transformation and ETL** fundamentals.
- Experience in **Google Cloud Plaftorm and GCP components such as Composer, BigQuery, Pub/sub, Cloud Function, DataProc etc.**
- Experience in **Azure Cloud components such as Blob, ADLS, Azure Databricks and Synapse.**
- Experience in using Automation Scheduling tools like **Talend Administration Console (TAC)** and **Control-M.**
- Experience in integrating data from various data source like **MySQL, Hive tables** and **Flat files and Rest Sources using Talend and Spark**.
- Worked on all phases of Integration development lifecycle like **gathering requirements, design, testing, implementation, and support of new and existing application.**
- Experience in migration of existing Hadoop application to Google Cloud Platform.
- Domain experience in **Healthcare** and **Pharma**

Technical Skills

- **Tool** – Talend Open Studio, Talend Real Time Big Data Platform (6.3.1) , Talend Data Fabric(7.3.1), Databricks, DBVisualizer, AzureDevOps, Databricks, Jira, Azure Boards, CA Agile.
- **Cloud Technology –** Google CloudPlatform, Azure
- **Cloud Services –** Big Query, DataFlow, DataProc, Pub/Sub, Composer, Cloud Function, Synapse, Blob, Synapse
- **Programming Language –** Java, Hadoop, Python, Spark
- **Scripting Language-** UNIX Shell Scripting
- **Query Language-** SQL, HSQL.
- **Operating System -** UNIX, WINDOWS.
- **Concepts-** Data Warehousing , ETL, OOPs, MD5
- **Scheduler -** Talend Administration Console (TAC), Control-M, Airflow
- **Methodology -**Agile

Certification:

- AHM250 Certified from PAHM
- AZ-900 certified from Microsoft
- Google Cloud Certified Associate Cloud Engineer
- Google Cloud Certified Data Engineer

Work Experience

Company
GlaxoSmithKline, Bangalore, India – 10/2021 to Present
Data Engineer

Project :- Code Orange

Roles and responsibilities:
- Create new and update existing java code to load data from various sources to ADLS and Synapse.
- Create and update code as per new business requirements.
- Add new functionalities to existing code Orange platform.
- Work with different teams for requirement gathering and analysis.
- Fixed Data Quality issue.

Tools & Technology: Talend Data Fabric 7.3.1, MySql , TMC, Git, Microsoft Azure, ADLS Gen2, Synapse, Blob, Java, Spark, Python, Databricks.

Company
Accenture, Bangalore, India – 11/2019 – 10/2021
Application Development Senior Analyst

Project :- BuyNow Analytics 1.0

Roles and responsibilities:
- Create **Talend** job to **load data from CSV file** to raw zone and then perform the required business transformation using HQLS and load final extracts in insight zone tables, which is thenreferred by DV developers for creation of Qlik viewDashboards.
- **Performance tuning** of the existing HQLs within application so as to ensure the minimum runtime of jobs and feed delivery to the downstream application at theearliest.
- Fix Data Quality issue.

Tools & Technology: Talend Open Studio, Hadoop, HDFS, Hive,HSQL, Windows 7,Linux, Unix shell Scripting, Control –M, Git, Azure DevOps, Google Cloud Platfrom, Airflow, Python, Spark

Project B :- BuyNow Analytics 2.0

Roles and responsibilities:
- Migration of existing BuyNow1.0 application from **Hadoop** to **Google Cloud Platfrom**.
- Transform existing hive queries to BigQuery.
- Work together with data architect team in finalizing GCP services that will be replacing existing Hadoop components
- Write Composer code to orchestrate the flow.

Tools & Technology: Google Cloud Platfrom, Composer, Bigquery, PubSub, Cloud Function, DataProc Stackdriver, Python, Airflow.

Company
United Health Group, NOIDA, India –07/2016 – 11/2019
Data Engineering Analyst
Project:- Individual Master Data Management

Roles and responsibilities:

♦ Create **Talend job** to extract data **from data warehouse**, perform the required business transformation use **MD5 concept** and create extracts, which were then loaded to tables and send to downstream users.
♦ Performance tuning of the existing jobs within application so as to ensure the minimum run time of jobs and feed delivery to the downstream application at the earliest.
♦ Worked on **Joblets** and Java **routines** in Talend.
♦ Fix Data Quality issue
♦ Involved in writing **unix shell scripts** using **Hive SQLs** as per requirements.

Tools & Technology: Talend Open Studio, Talend Real Time Big Data Platform (6.3.1), Hadoop, HDFS,Hive,HSQL, Windows 7,Linux, Unix shell Scripting, GitHub, TAC.

Achievements :

➢ Received multiple awards like Aquamarine, Sapphire, Opal recognitions for Quality, Innovation, Performance, Collaboration and Team work.
➢ Won Jupiter award for Q4 2018.
➢ Won Star award for Q3 2020

Education

B.Tech in Computer Science and Engineering, GNIOT, Greater Noida
Aug 2012- June 201.