

CAREER SUMMARY – Lead Data Engineer with 4 years of experience in building data intensive applications, tackling challenging architectural and scalability problems in Insurance, Broadcasting and Supply Chain Domain.

Professional Skills

Data Governance & Security:
Apache Atlas, Apache Ranger, Alteryx Connect

Version Control System Tool:
GIT, GITHUB

Continuous Integration Tool:
GitLab

Continuous Monitoring:
CloudWatch

Project Delivery Strategy :
Scrum, Sprint

Cloud Computing Platform:
AWS

Big Data Technologies : **Apache Nifi , Hadoop (HDP), Hive, MapReduce**

Cluster Management Tool:
Apache Ambari

Advance Analytics Tools:
Alteryx

Scripting Languages:
SHELL, PYTHON, C++

Operating System: **LINUX, WINDOWS**

Database: **Oracle, MySQL, AWS DynamoDB**

Datawarehouse Technologies
AWS Redshift

Work Experience

HCL Technologies Ltd (Transitioned From Collinson Group) Lead Engineer (June'20 – Present)

Project Responsibilities

Domain: Insurance and Travel

- Developed Strategy and Roadmap to build Big data architecture and expertise in Analytics with a detailed execution plan.
- Automated the deployment of Ingestion pipelines via AWS Cloudformation to create auto-scaling enabled EMR stacks.
- Integrating external tools like Alteryx and Tableau with hive for ETL development , Reporting and Business Integration.
- Designed and integrated PySpark based custom modules to handle data for multiple structured and unstructured formats like XML, JSON, CSV, text files, MySQL , SQLServer and Oracle database sources.
- Planned the optimization and enhancement of the present features offered by internal Data Lake built.
- Continuous code enhancement and optimization as per the latest libraries and features.

Collinson Group

Associate Data Engineer (June'19 – May'20)

Project Responsibilities

Domain: Insurance and Travel

- Performing Source analysis and creating ingestion layout and pipelines.
- Aligning and understanding requirements with stakeholders and data scientists to develop ETL models.
- Developed secure data pipelines on the AWS platform to aggregate data from multiple sources like JDBC, API, FTP/SFTP, Azure Cloud and custom managed servers via IAC.
- Implemented ingestion framework to extract data from various source systems, data was stored in S3 with external Hive tables pointing to S3.
- The ingestion module addresses critical issues of schema evolution and compliance-oriented processes of filtering and de-identification using Spark and Nifi.
- Stabilizing, governing and enhancing the security and permissioning features on services and accesses using fine-grained IAM and Ranger policies.
- Research and implementation of custom-metric based Auto-Scaling on HDP cluster.
- Research and development of secure API to provide token based access to external users.

Issues, Task and Defect
Tracking: **GitLab** , **Jira**

Data Analytics Services: **AWS**
EMR, AWS Data Pipeline

Frameworks : **Spark** , **Django**,
Flask

DevOps Technologies:
Jenkins, Docker,
Cloudformation

LinkedIn Certified Technology:

1. **AWS Cloud**
2. **Hadoop**
3. **Python**
4. **Git**

Strength:

Team Player
Adaptability
Quick Learner
Confident
Leadership Qualities
Troubleshooting Skills
Debugging Skills

Data Peace AI Technologies

Software Developer (Dec'18 – May'19)

Project Responsibilities

Domain: Broadcasting(Datalake)

- Integrating Datalake procedure and handling the project flow from end to end on DEV and STAG Environments.
- Automating the data flow from ingestion, cleansing, transform and loading stages on EC2 and EMR Clusters using cron job scheduling on AWS Data Pipeline
- Working knowledge of build automation using Python scripting.
- Developed Spark code in Python using Spark SQL & Data Frames for aggregation.
- Worked on data from multiple sources with data volume of more than 40GB/Day.
- Actively participate in high level team activities such as suggesting architecture improvements, cost optimization, recommending process improvements and conducting tool evaluations.
- Quickly troubleshoot and resolve network, software and performance issues on AWS Data Centre.
- Setting up notification service using AWS SNS in case of job failures indicating the details of failure reason and timestamp details.

Capgemini India Pvt Ltd

Senior Software Engineer (Jun'17 – Nov'18)

Project Responsibilities

Domain: Insurance

- Release planning and coordination with Development and Testing teams. Maintaining Issues & Track Project tasks using Kanban.
- First line of support for all DEV/TEST/PROD environments.
- Coordinating the preparation of the target environments and deployment scripts for each release.
- Writing or validating release technical procedure for Production release of applications and support PROD team during PROD rollout.
- Responsible for the day to day operations of client servers on the cloud ensuring system availability, performance, capacity & continuity through proper response to incidents, events and problems.

Software Engineer (Oct'16 – May'17)

Project Responsibilities

Domain: Supply Chain

- Handling Master data Management on global scale which includes creation/Deletion/ Update/ Switch/ BOM of DFU, SKU.
- Work on various Change request to develop SQL procedure and writing PL/SQL Procedures, Functions for automation.
- Perform Dynamic data integration like Sales, Inventory, Shipments, Scheduled Receipts, Mass Events, Planned Arrivals and Planned Orders for various markets.

Educational Qualifications

Bachelor of Technology (ECE), 2016 from Guru Gobind Singh Indraprastha University, New Delhi (70%).

Class XII, 2012 from CBSE board, New Delhi (83.2%).

Class X, 2010 from CBSE board, New Delhi (83.6%).