

RUPAL SARKAR



Contact

Address:

Kolkata, West Bengal,
India -700156

Phone:

+91 8250757508

Email:

rupal.sarkar10@gmail.com

Additional Skills

Programming Language

C Python Groovy Java SQL

Framework

Flask Django Spring Boot

Tools & CI/CD

GitHub GitHub Actions

Certifications

- [Collibra Workflow Engineer](#)
- [Collibra Integration Engineer](#)
- [Collibra Solution Architect](#)
- [Crash Course on Python](#)

Technical Blogs

- [Collibra Workflow - "Modify Assets in Bulk"](#)
- [Collibra Data Analysis With Pyspark](#)

Summary

- Certified Collibra Developer with expertise in implementing Data Governance, Meta Data Management using Collibra DG Tool.
- Have strong knowledge of Enterprise Data Governance, Activiti BPMN 2.0 Workflows, and Reference & Master Data Management.
- Hands on experience with Java Core APIs and REST Core APIs.

Experience

Senior Consultant

Nov 2023 – Present

Ernst & Young (EY), India

Data Governance Project

(Client – [US Hospital](#))

- Engineered a **Java Spring Boot** integration coupled with **Collibra Integration Library 3.X** to seamlessly extract metadata from the **Radiant Logic** application. Additionally, engineered a tailored technical lineage through **Collibra's batch definition** option. Subsequently, deployed the integration as a packaged **JAR** on a **Red Hat Linux VM**. It has the capability to keep the metadata up to date.
- Crafted Collibra **BPMN Workflows** utilizing the modern workflow designer provided by Collibra, tailored to address diverse business use cases.
- Developed a bespoke **metamodel** specifically tailored to capture reference data pertinent to **OMOP** (Observational Medical Outcomes Partnership), facilitating the analysis of healthcare data to derive real-world evidence.
- Crafted a Python integration to meticulously generate **Databricks** lineage within Collibra DIC.
- Harvested metadata in Collibra Data Intelligence Cloud from **Databricks** Unity Catalog.
- Created Technical lineage for ETL Tool – **Informatica Intelligent Cloud Services**.
- Developed a python script using **Weasyprint, Jinja2** libraries to generate daily reports of Collibra Jobs and circulate among different stakeholders.

Consultant

Sep 2022 – Nov 2023

Virtusa, United Kingdom


Data Governance Project


(Client – [One of The Biggest Banks in UK](#))

- Harvested metadata and created Technical Lineage from different Data Sources (i.e., SQL Server, Teradata, Oracle, DB2 etc.).
- Developed a connector using **Java Spring Boot** framework & **Collibra Integration Library 3.X** to harvest metadata from XML, XLSX file and create technical lineage. Then compiled the source code to a JAR file and deployed this packaged version of integration with an external configuration properties file.
- Created 8 **BPMN Workflows** to address various use cases, with a notable challenge being the management of 22 million assets based on user input, skillfully resolved through the utilization of **OutputModuleApi** and **ImporterApi**.

- Developed **Power BI Dashboard** to visualize following key metrics:
 - Total Critical Data Elements (CDEs) by Application
 - Issues by Status (i.e., Minor, Moderate, Material, Severe)
 - Total asset ingested over time
- Developed a connector using **Flask API** & **BPMN** workflows and hosted that application on **Ubuntu VM** to serve following criteria:
 - While refreshing the metadata, it's able to identify the **obsolete** assets.
 - While refreshing, it's performing proper **UPSERT** operation (i.e., **Update** existing entries + **Insert** new entries)
 - Generating **Technical Lineage** to showcase column-to column transformations.
- Worked with different stakeholders to implement **GDPR** and another Privacy framework in Collibra to automate Data sharing Agreements.
- Created **Validation Rules** against individual assets to assure quality of assets.
- Implemented mappings between Technical Data Dictionary and Business Data Dictionary, so that consumers can conduct Impact Analysis.
- To manage issues developed different **Issue Management Workflows** using BPMN 2.0 framework.
- Developed Custom Connector using **Python** to harvest metadata from **COBOL Copybooks** for Mainframe applications.
- Customized technical lineage for System-to-System transformation.

Associate Consultant

 Oct 2020 – Aug 2022

 Capgemini, India


Data Governance Project

(Client – Leading German Pharmaceutical Company)

- Helped in implementing Collibra Operating Model.
- Defined & Implemented Glossary of Business Terms, Policy Domain, Rulebook, Validation Rule Domain.
- Configured Traceability diagrams to showcase assets & relations.
- Developed custom BPMN workflows – '**Asset Approval in Bulk**', modified version of '**Ask the Expert**', '**Onboard Business Asset**'
- Worked with other engineer to register data sources using JDBC Driver
- Developed Custom Integration. Using Collibra **REST Import API** (endpoint - '**/import/json-job**') harvested metadata formatted in JSON into Collibra Data Governance Center. Finally automated the process using **GitHub Actions**

Education

Bachelor of Engineering

 2016 – 2020

CGPA: 9.04

2016 - 2020

Jadavpur University, India