Soni M

Data Engineer, 6+ years experience
Punjab, India
https://augmntx.com/profile/data-engineer-soni-m/axo25833

+91 9820045154 | hr@augmntx.com

Led and managed automation initiatives for data governance projects, utilizing tools for data quality checks, metadata management, and data lineage tracking. Gathered and cleaned data from CAFM/CMMS databases and other sources to ensure accuracy and consistency. Designed, built, and maintained data pipelines in the Azure environment, leveraging Azure Data Factory, Ascend, Databricks, and PySpark, GCP. Proficient in SQL, Python, and big data tools, with strong statistical methodology knowledge applied to pharmaceutical research. Developed and maintained automation dashboards and reports to monitor and enhance data governance processes. Contributed to developing BPMN workflow engines using Camunda, and created/modified BPMN workflows in Collibra. Played a pivotal role in data migration, conversion, and integration, ensuring data integrity and developing APIs using Spring Boot for customized Collibra data.

INDUSTRIES: Insurance, Finance

TECHNICAL SKILLS

• Data Engineer	5 years & 5 months
• GCP	5 years
• ETL	5 years & 5 months
• AWS	5 years
• Azure	5 years
• Ascend.io	3 years

PROJECT DETAILS

• Ciox January-2020 - February-2022

Project Description: This is an Insurance domain project where I Led end-to-end development of Electronic Medical Record (EMR) systems tailored to customer requests, ensuring comprehensive and efficient report generation and delivery.

Responsibilities: Led end-to-end development of Electronic Medical Record (EMR) systems tailored to customer requests, ensuring comprehensive and efficient report generation and delivery. Collaborated with cross-functional teams, including biostatisticians and clinical data managers, to integrate statistical models into the drug development process, optimizing decision-making. Employed diverse technologies such as Kubernetes clusters, Python Flask framework, Docker, and AWS services (SQS, Lambda, S3) to build scalable, resilient, and cloud-native EMR solutions. Developed and implemented ETL pipelines for insurance data using Azure Databricks and Snowflake, enabling efficient extraction, transformation, and loading for enhanced data management and analytics. Orchestrated the development of a customer-facing dashboard using Java Spring Boot and Reactjs, providing intuitive access to EMR reports, and enhancing data visualization for improved user experience. Utilized Ascend.io's Data Aware Scheduling to intelligently trigger pipeline runs based on upstream changes, minimizing unnecessary compute costs. Integrated Ascend.io with GCP services (BigQuery, GCS, Pub/Sub) to streamline ingestion, transformation, and loading processes across batch and streaming data sources.

Technologies: GCP, AWS, Azure, ETL

Industry: Insurance

• Citi Bank September-2022 - April-2025

Project Description: This is a Financial Sector Project for CITI Bank, where I worked as a Data Engineer.
Responsibilities: Spearheaded metadata management initiatives for a leading US bank, focusing on data cataloging, lineage, and quality processes. Collaborated closely with data owners and stewards to define project scope, identify critical data elements, and establish quality rules.Implemented rigorous testing protocols to identify and address integration challenges promptly, ensuring data reliability and functionality.
Technologies: GCP, AWS, Azure, ETL
Industry: Finance

EMPLOYMENT HISTORY

• Senior Data Engineer September-2022 - April-2025

Description: Led and managed automation initiatives for data governance projects, including data quality checks, metadata management, and data lineage tracking. • Gathered and cleaned data from CAFM/CMMS databases and other sources. • Designed, built, and maintained data pipelines in the Azure environment using tools like Azure Data Factory (ADF), Databricks, GCP, ascend and PySpark. • Proficient in SQL and Python, with strong knowledge of big data tools and statistical methodologies applied to pharmaceutical research. • Developed and maintained automation dashboards and reports to monitor data quality and governance processes. • Contributed to developing a comprehensive BPMN workflow engine using the Camunda platform and created/modified BPMN workflows in Collibra. • Played a pivotal role in data migration and conversion, ensuring data integrity and security, and developed APIs for data migration and customized Collibra data using Spring Boot.

EDUCATION

• Bachelors degree in B.tech (CSE) from Amity University 2016 - 2020

LANGUAGES

• English - advanced

END OF DOCUMENT
This PDF is automatically generated by AugmntX on 21 Nov 2026
Profile URL:https://augmntx.com/profile/data-engineer-soni-m/axo25833