Thakkar P

Senior Data Engineer, 9+ years experience
Ahmedabad, Gujarat, India
https://augmntx.com/profile/senior-data-engineer-thakkar-p/axo21948

AugmntX
+91 9820045154 | hr@augmntx.com

9+ Years of Experience, Highly skilled and results-oriented Databricks professional with a proven track record of designing and implementing efficient data pipelines, resulting in significant reductions in data processing time and improved accuracy. Adept at implementing advanced data quality and governance processes, ensuring compliance with industry regulations and minimizing data errors. Skilled in developing and maintaining machine learning models to drive customer retention and cross-selling opportunities, resulting in increased revenue and operational efficiency. Developed and maintained ETL processes to efficiently move data between systems, resulting in a 25% reduction in data processing time and improved data accuracy. Developed and maintained data governance and compliance processes, ensuring data integrity and compliance with industry regulations, resulting in a 15% improvement in data quality and a 10% decrease in compliance issues. Involved in monitoring the workflows and in optimizing the load times. Used Change Data Capture (CDC) to simplify ETL in data warehouse applications using Azure Databricks and Azure Data Factory. Worked on process improvements and helped streamlining CI/CD deployments helping implementing parameterized build pipelines. Knowledge of data visualization tools such as Tableau, Looker Studio and Apache Superset. Demonstrated technical, team, and solution leadership through strong communication skills to recommend actionable data-driven solutions. Build and maintained data warehousing solutions using Snowflake, allowing for faster data access and improved reporting capabilities. Developed and optimized complex SQL queries and stored procedures to extract insights from large datasets. Participated in the production on-call support and demonstrated ability to explain complex technical issues to both technical and non-technical audiences. Strong analytical skills with ability to work with business partners to create exceptional data solutions by understanding and refining requirements to generate expected outputs. Communicated with Data teams for implementing data mapping and transformation to generate quality of data-based business reports. Experienced SQL skills, including the ability to write, tune, and interpret SQL queries; tool specific experience in Microsoft SQL Server, Postgres SQL, Amazon Redshift, Azure SQL Server, DynamoDB and Bigquery. 5+ progressive years of strong data engineering, SQL scripting and tuning experience. Proficient in Elasticsearch, data modelling and querying using log aggregation, data extraction and reporting using Elasticsearch, Logstash, Kibana tools.

INDUSTRIES: Data Maintenance, Finance, Supply Chain Management, Healthcare, IOT Solutions, Ecommerce, Event Management

TECHNICAL SKILLS

• Python	9 years
• Pyspark	8 years
• JavaScript	7 years
• Type script	7 years
• AWS	5 years
• DevOps Practices	4 years
• AWS Glue	3 years
• DataBricks	2 years

SOFT SKILLS

Communication, Teamwork, Problem-solving, Time management, Critical thinking

PROJECT DETAILS

• Data Migration

Project Description: Data migration from on premise SQL server to Azure Data Lakes using Azure DataFactory and Datariccks

Responsibilities: Developed data pipelines following MEDALLION ARCHITECTURE, distributing data across bronze, silver and gold containers using (DLT) and generated materialized views. Created more than 20+ ADF data pipeline and performed data ingestion from SQL server to Azure ADLS. Created Azure App Function in python for Symmetric data decryption from sql server and exported data to ADLS in parquet format, implemented Azure Manage Identity and Principal roles for restricted access. Implemented ETL for data cleansing and developed algorithm to adhere the dynamic data nature, that is ingesting and updating the target UNITY CATALOG tables for new columns in data. Implemented DLT loads within Databricks UNITY CATALOG tables. Created a dynamic script in databricks notebook, that adheres the JSON schema and performs column mapping and export the tables to UNITY CATALOG tables. Implemented ETL, developed parameterized jobs that accepts JSON schema and string array, performs data ingestion. Developed a job in databricks implementing column level data encryption and avoiding data duplication with PII, using FERNET encryption libraries and developed customized UDFs Created Workflows and scheduled ETL jobs in Azure databricks, implemented autoloaders and applied parallelism over multiple scheduled jobs using workflows and increased data ingestion efficiency 80%. Generated more than 30+ SQL views / queries performing data modelling and transformed output for BA team. Converted, optimized and migrated more that 30+ SQL views to materialized views in Databricks. Debugged, monitored and fixed scheduled workflow jobs and Azure Data Factory pipelines. Reviewed and managed rebasing and resolving PR conflicts for the team.

Technologies: Azure Databricks, Azure Data Factory, Pyspark, Azure DataLake

Industry: Data Maintenance

• Extraction Concepts

Project Description: Data Ingestions from social platforms to Azure storage accounts using Azure Databricks

Responsibilities: Developed data pipelines ingesting data from varied sources to Azure data lake storage, and enriching data from Azure data lake to Neo4j Created process to pull data from social DB and upload into respective datalakes Converted manual process of uploading, processing and publishing data in azuredatafactory into python programs Scripts to clear and populate data in neo4j Used multithreading to reduce pipeline processing run time by 60% Designed and developed a decision tree application using Neo4J graph database to model the nodes and relationships for each decision. Implemented Neo4j to integrate graph databases with relational database and to efficiently store, handle and query highly connected elements in your data model especially for object oriented and relational developers. Exported SQL data from Elastic search to to Neo4j to load paged data from elasticsearch, and process one record at a time.

Technologies: Azure Databricks, Azure Data Lake, Neo4j, Python

Industry: Finance

• Data Modelling & Analytics

Project Description: Project is about delivery service company that specializes in delivering everyday essentials such as groceries, household items, and over-the-counter medications directly to consumers' doors.

Responsibilities: Written production ready scripts and test cases using DBT tool and maintained scalable data models and transformations using DBT in conjunction with Snowflake, ensure effective transformation and load data from diverse sources into data lake. Implemented and managed data models in DBT, guarantee accurate data transformation and alignment with business needs. Utilized DBT to convert raw unstructured data into structured datasets, enabling efficient analysis and reporting. Developed ETL Pipelines in and out of data warehouse, develop major regulatory and financial reports using advanced SQL queries in snowflake. Optimized SQL queries within DBT to enhance data transformation processes and improved all performance. Developed and Managed dashboard using Looker Studio, also created a calculated metrics, custom looker email Develop and maintain data warehouse reporting solutions, monitoring and alerting solutions

Technologies: DBT, Snowflake, Looker Studio (Google)

Industry: Supply Chain Management

• Hybrid Generative AI

Project Description: Analysis on Food Science Technology getting insights and intelligence on health and science

Responsibilities: Developed ETL Pipelines data ingestion from REST API and different sources (Social Media, Csv, XML, PDFs) Configured ETL server using apache airflow and developed and schedule pipelines and created workflows for data ingestion and transformation Developed CI /CD pipelines for FrontEnd, BackEnd and ETL pipelines using Kubernetes and podman Integrated backend in Django with third party rest API and used Django service as proxy Developed pipelines and transformed and preprocessed data from XML, PDF to JSON and performed data ingestion to Knowledge Graph Worked as scrum master and facilitate daily standups. Aid in estimation and sub task creation and ensured and maintained the scope and goal of sprints Identified the pain areas to ensure all team members and stakeholders understand the project;s goal and obstacles Translated and documented business requirements, technical and functional specifications to development

Technologies: DataDog, Apache Airflow, FastAPI, Django, Pandas, Python, Langchain, ChatGPT4, Ollama embeddings, Vector Database, Knowledge Graph DB

Industry: Healthcare

• Wavely

Project Description: Experience the Future of Connectivity with Our Comprehensive IoT Solutions Suite. From location tracking and robust security features to powerful analytics tools and resell opportunities, our platform empowers you to harness the full potential of IoT connectivity. Stay ahead with a versatile suite designed for seamless integration and business success.

Responsibilities: Currently leading a team of 3 data engineers and Led the migration of SQL DB, pulling data from sql tables in different modes. Created and managed data pipeline for full import of data from more than 100+ SQL tables, also created data pipeline for incremental data using CDC functions, that pulls changes on daily basis using Azure Data Factory and Azure Databricks and exported to Azure Data Lake. Created a 30+ data mapping templates using SQL scripts to create complex views for data insights, where data is distributed across different schemas and database. Designed more than 30+ parameterized data pipelines in ADF, enabling components reusability across pipelines

Technologies: Azure Databricks, Azure DataFactory, Pyspark, Python

Industry: IOT Solutions

• Retail Dashboard

Project Description: Ecommerce retail dashboard allowing to ingest product catalogs and allowing product manipulations for cataloging on web platform

Responsibilities: Managing and integrating of application requirements over frontend in ReactJS + Typescript + Redux Thunk and consuming rest API written in NodeJS. Authoring application development and deployment single handed from scratch. With code commit and maintaining standard practices. Writing unit test case using JEST and Enzyme. Developed code pipeline in AWS and Azure for application deployment based on cloud preference selected by client. Using DnD-Kit/core for allowing drag and drop of UI elements and customizing the products catalogs. Developed responsive UI for products catalog across different device and resolutions. Improved code quality following best practices and appropriate comment with clean and seamless readability, supporting reusability and scalability of components that adheres modern robustness requirements.

Technologies: React, Typescript, Redux Thunk, NodeJs, AWS, GCP

Industry: Ecommerce

• Zeus

Project Description: Project is a well known client offering hospitality for corporate serving entairtainment, venues, merchandise retail and stadium operations.

Responsibilities: Integrated secure connection bridge deployed over AWS ECS as micro service between salesforce and AWS data warehouse pulling all the benefits from kinesis streams SDK, by subscribing to salesforce channels and topics. Migrated Tableau Reports to open source data exploration and data visualization tool Apache Superset (written sql view and optimized SQL queries and developed metrics for complex viz. Elements). Defined AWS Glue Jobs that orchestrated data from various sources (Google Analytics, FTP Server and Salesforce) Implemented Custom Google Tags and Integrated Customized Google Analytics solution with existing system mapping every user's activity and leveraging the metrics from google analytics for insights. Integrated large set of SQL views for Apache Superset and implemented visualization. Designed and leveraged OData ecosystem for building REST API services. Developed glue jobs for data ingestion from google-analytics and ftp server to redshift data warehouse. Created database migrations scripts and contributed in designing schemas. Researched and documented POC tasks, application flow as well as application code in technical manuals within Jira confluence Implemented customized AWS Lambda function using python for centralized data management by creating a unique association ID shared across the system. Developed reusable web component in React using storybook. Managed REST API services using Flask, by engaging SQLAlchemy ORM design pattern making it easier to write secure, modular and fast applications in python language using MYSQL database.

Technologies: AWS ECS, AWS Glue, React, Pyspark, Postgres, NodeJs, Google Analytics, Salesforce

Industry: Event Management

EMPLOYMENT HISTORY

• Senior Data Engineer

Description: Designed and implemented Python APIs using fast api, ensuring continuous availability and reliability. Responsible for implementation of data viewers, Logging, error configurations for error handling the packages. Involved in the complete life cycle in creating SSIS packages, building, deploying, and executing the packages in both environments (Development and Production). Designed and developed scalable ETL pipelines using Azure Data Factory for ingesting large volume of data from onPrem SQL server. Developed and designed data models with complex SQL queries to extract insights from large datasets.Implemented and designed CI/CD release and deployment for different environments Developed data workflow using azure databricks for DLT loading increasing efficiency data load efficiency 60% Own the end to end delivery of the product. Defined system level requirements and architecture. Authored technical assessment for build / design and development discussions. Outlined design improvements and risk with mitigation strategies. Drived R&D process from feasibility through client delivery. Responsible for client communication and requirement analysis from client. Completed full redesigns of existing react application and refactored code and enhance visuals UI using MUI (Materia UI, React) Worked in a fast-paced Agile environment across different stages of application development to deployment. Responsible for development and code deployment periodically across different environments. Responsible for updating REST endpoints and frontend integration using React with smoke testing for development.

EDUCATION

• Bachelors degree in Computer Application from KSKVU 2008 - 2011

LANGUAGES

• English - native

END OF DOCUMENT
This PDF is automatically generated by AugmntX on 22 Nov 2026
Profile URL:https://augmntx.com/profile/senior-data-engineer-thakkar-p/axo21948