Senapati Lokesh **Email: senapatilokesh61@gmail.com**

(Azure Data Engineer) **Mobile: +91-9032182330**

Career objective:

To work in a challenging and innovative IT environment with opportunities to utilize my Expertise in delivering quality software and contribute positively to the organization and enrich my knowledge, experience, and skills in emerging technologies.

Profile Summary:

A well versed Cloud data engineer with 4 Years and 5 months of Overall experience in the IT industry and having strong debugging and troubleshooting skills in Azure (Cloud), data Engineering with hands-on experience in designing and deploying pipelines for data ingestion using cloud native services like Azure Data Bricks, Azure Data Lake, Azure blob storage, Azure SQL Database, Azure Data Factory, MS-SQL Server, Oracle and python technologies.

Experience Summary:

- Extensive hands-on experience (4+ years) in data engineering and orchestration using Azure Databricks, Azure Data Lake Gen 2, Azure Blob Storage, and Snowflake for building scalable data pipelines and processing frameworks.
- Unified Data Framework: Designed and implemented a unified framework for ingestion, transformation, and loading (ETL) from multiple RDBMS sources (MySQL, PostgreSQL, Oracle, SQL Server, etc.) using Databricks, handling both full and incremental data loads.
- Spark & Delta Lake: Expertise in Apache Spark (PySpark & Scala) on Databricks for data transformation and analytics, leveraging Delta Lake to ensure ACID transactions, data reliability, and efficient time-travel queries.
- Source Integration: Extracted data from various sources, including RDBMS (SQL Server, Oracle, PostgreSQL), flat files, CSVs, JSON, and loaded into Azure Data Lake, and Snowflake via Databricks for further processing and analytics.
- Experience with Data Connectors: Worked with multiple data connectors and formats (Parquet, CSV, JSON, Delta) for high-performance data ingestion and transformation pipelines in Databricks.
- Cluster Management: Skilled in managing and optimizing Databricks clusters, ensuring high availability, performance tuning, and proper utilization of Spark resources.
- CI/CD Pipelines: Implemented CI/CD pipelines in Databricks using tools like Jenkins and GitLab for version control, testing, and deployment of production code. Automated data workflows using Azure Data Factory (ADF) triggers to ensure smooth job scheduling and orchestration.
- Experience with Incremental Data Loading: Developed robust frameworks for incremental data loading and handling Slowly Changing Dimensions (SCD1, SCD2) in Databricks to ensure efficient ETL and data integrity.
- Optimized Spark Jobs: Implemented performance optimization techniques and best practices in Spark jobs, including partitioning, caching, and broadcasting to reduce execution time and improve overall job performance.

- Data Orchestration in Databricks: Expertise in orchestrating data workflows using Azure Databricks for batch and streaming data pipelines, handling large datasets, and ensuring high performance with Spark SQL and DataFrames.

- Building JARs in IntelliJ: Skilled in Scala development with IntelliJ, building modular, reusable JAR files for Databricks jobs, ensuring efficient and clean deployment of production code.

- Scheduling & Automation: Automated Databricks pipelines using APPWORX(scheduled and event-based) and managed workflows for extracting and transforming data at regular intervals.

- Transformation Techniques: Worked on a variety of DataFrame transformations in Databricks including aggregates, joins, window functions, filters, and conditional transformations using Spark SQL and PySpark.

- Data Quality & Validation: Integrated data validation and error-handling mechanisms into ETL pipelines to ensure accurate and consistent data delivery to downstream systems.

- Snowflake Integration: Knowledge of Loading data from Azure Blob Storage to Snowflake and Azure Data Lake, managing efficient data flow between cloud platforms.

- Collaboration & Documentation: Strong team player, collaborating with cross-functional teams (data analysts, engineers, and business stakeholders) to deliver impactful data solutions. Skilled at documenting pipeline designs, processes, and best practices.

- Highly adaptable to different environments, demonstrating a continuous learning mindset to improve skills and successfully complete tasks in dynamic workspaces.

Education:

- B.COM (GENERAL) – 2018 – NEW SCHOLARS DEGREE COLLAGE-HYD affiliated to OSMANIA UNIVERSITY with an aggregate of 71.5%.
- Board of Intermediate Education (MPC) – 2015 – SRI CHAITANYA JUNIOUR KALASALA, HYDERABAD with an aggregate of 88.8%.
- Secondary School Certificate – 2013 - SHIVAJI VIDHYA PEETH SCHOOL(CBSE), HYDERABAD with 9.2 grade points.

Skill Summary:

- Operating Systems : Windows Family
- Databases : SQL Server 2016(SSMS)
- ETL Tools : Data Bricks, Azure Data Factory
- Storage : Azure Blob Storage and ADLS Gen2
- DWH : Snowflake
- Languages : Python, SQL

Experience:

- Working for Infosys, Hyderabad as Azure data engineer from Sept 2022 to till date.
- Previously worked in LARC Software Private Limited from July 2020 to Aug 2022

Cetifications:

- **Microsoft Certified: Azure Fundamentals (AZ-900)**
- **Microsoft Certified: Azure Data Fundamentals (DP-900)**
- **Microsoft Certified: Azure Data Engineer Associate (DP-203)**

Projects Details:

Project #1:

Project Title	:	German Data hub
Client	:	Daman Health Insurance Company
Role	:	Azure Data Engineer(Support and developer)
Team Size	:	8 members
Duration	:	June 2020 to August 2022
Environment	:	ADF, ADLS, AZURE Databricks Notebooks, MS-SQL SERVER.

Project Description:

Daman started its business in 2006, currently operated over 12 branches in UAE, Daman is UAE's first exclusive health insurance provider. Daman is a public joint-stock company which is 80% owned by the Abu Dhabi Government with the remaining 20% owned by Munich Re. Munich Re is one of the world's leading reinsurers in Germany. The primary objective of the project is to create a Data Hub for company German Region, data which will ensure that we have consistent, accurate and trustworthy data for fraud detection and risk mitigation, identifying new markets and their specific insurance needs etc.. to be utilized in applications, other warehouses and reports.

Roles &Responsibilities:

- Participated in Brainstorm discussions and designed effective solutions to move data from different data sources to cloud.
- Creating staging Database for temporary processing of data, consolidating data and loading into database using Integration Services.
- Creating Azure data lake store to maintain data and process data using Data lake analytics.
- Creating Activity's in Azure data factory like copy activity, stored procedure activity, Custom activity to process data.
- Created reusable pipelines in ADF by using different data transformations like derived columns, Lookup, Conditional Split, Join, Sort and execute Sql task to load data into database.
- Importing Source/Target tables from the respective Databases by using Execute SQL Task, using Control Tasks in SQL server 2016 Integration services.
- Analyze the requirement documents and get the concerns clarified with the business.
- Understanding Source data by discussing with the business users and built business logics as per their request in Azure data factory.

Project #2:

Project Title	:	Data Paltform Modernisation (DPM)
Track	:	BDS Track(Big Data Services)
Client	:	Cummins Ltd
Role	:	Developer
Team Size	:	12 members
Duration	:	October 2022 to till date
Environment	:	Databricks, ADLS, MS-SQL SERVER, Oracle, My-Sql, API, CFG, AppWorx, IntelliJ, GitLab.

Project Description:

Cummins Ltd, a leader in manufacturing **generators** and **large machinery**, required a **modern data architecture** to manage and process their growing data volumes for operational analysis, predictive maintenance, and real-time analytics. The project involved building a **unified data platform** using **Azure Databricks** to ingest and process data from **multiple RDBMS sources**, operational systems, and IoT sensors embedded in their machinery.

The primary goal was to provide the Cummins team with efficient access to real-time and historical data for better decision-making, **supply chain optimization**, and **predictive maintenance** of their heavy machinery.

Roles &Responsibilities:

Data Orchestration and Ingestion:

Our data engineering projects involve a diverse range of data sources and ingestion methods, including relational databases (RDBMS) such as MS SQL, MySQL, Oracle, and PostgreSQL, as well as Salesforce for CRM data. We also handle file-based data through SFTP (CFG), API-based data retrieval, and OData from SharePoint. Each data source and method is carefully orchestrated to ensure seamless integration into our data pipeline.

Intake and Analysis:

The project begins with receiving data ingestion requests through Jira cards. We perform a comprehensive analysis to ensure that all necessary details are provided and to assess the feasibility of the request. This stage is crucial for understanding the scope, defining requirements, and planning the development activities. It ensures that we have a clear understanding of the data needs and technical requirements before advancing.

Development Preparation

Following the analysis, we move to the development phase. For RDBMS sources, we use a unified framework to facilitate data ingestion, which involves creating JAR files with IntelliJ and uploading them to Azure Data Lake. For other sources like CFG (SFTP), APIs, and Salesforce, we develop customized solutions. This includes writing code in Scala or Python within Databricks notebooks or JARs and utilizing Spark SQL for data transformations.

Data Ingestion and Transformation

Data ingestion involves connecting to various sources and transferring data into a staging area in Azure Data Lake. We handle multiple file formats such as JSON, CSV, Excel, XML, and Parquet. Data transformation tasks include parsing and flattening, particularly for JSON files, to prepare the data for analysis.

Transformed data is then written to target systems, such as Azure Data Lake or Snowflake. We have transitioned from Parquet to Delta format to leverage the enhanced performance and management capabilities of Delta Lake, in line with advancements in Databricks and Unity Catalog.

Testing and Validation:

After development, we conduct extensive testing. This involves validating data pipelines and transformations in a staging environment and scheduling jobs using Appworx or Databricks Workflows. During User Acceptance Testing (UAT), we provide detailed design and implementation documentation and collaborate with users to ensure the solution meets their requirements. Feedback from UAT is incorporated to refine and finalize the solution.

Deployment and Migration:

Upon UAT approval, we proceed with deployment. This includes Knowledge Transfer (KT) to the production support team, explaining the development process and operational aspects. Code is merged into the master branch via GitLab, subject to administrative approvals. Jenkins is used to manage the deployment. Post-deployment, we validate the data and tables, ensuring everything is functioning correctly, and inform users of the project's completion.

Maintenance and Monitoring:
Ongoing maintenance and monitoring are critical for ensuring the reliability and performance of our data solutions:

Data Backups and Auto-Deletion:
We manage regular data backups and implement auto-deletion policies for old backup and log files. This helps optimize storage usage and ensures that only relevant data is retained.

Logging and Debugging:
Comprehensive logging is maintained for all data processes. This includes generating logs that aid in debugging and resolving any issues that arise. Effective logging is essential for tracking data flow, identifying problems, and ensuring smooth operations.

Alerting and Notifications:
We set up alerts to monitor for failures and other anomalies. This includes configuring automated notifications to quickly address any issues that impact data integrity or pipeline performance.

Unified Framework Transition:
We are transitioning API and CFG-based ingestion processes to a unified framework. This transition aims to enhance efficiency by standardizing processes and reducing complexity. The unified framework supports streamlined data handling and processing, improving overall operational effectiveness.

Place: Hyderabad **(Senapati Lokesh)**
Date: