ASWIN RAVICHANDRAN
DATA ENGINEER

E: aswin.ravichandran06@gmail.com|M: +1 (929) 595-3080 **|** LinkedIn **|** L: Harrison, NJ

PROFESSIONAL SUMMARY

- Extensive experience as a Data Engineering with 4 years of experience in Analyzing, Developing, Managing, and implementing various standards associated with the data set and client-server enterprise applications using Python Development, Cloud Platforms.
- Strong expertise in database management, **data integration, data modeling, and data pipeline orchestration**, coupled with a deep understanding of data quality and governance principles.
- Expertise in configuring and managing Directed Acyclic Graphs (DAGs) within Apache Airflow.
- Extracted data from SQL Server Database, copied it into different file systems, and used Hadoop tools like **Hive and Pig** to retrieve the data required for building models.
- Hands-on experience with different programming languages such as **Python, Scala, Java.** Enhancing adaptability to different project requirements.
- Skilled in data warehousing, working with platforms such as **Amazon Redshift, Snowflake**, and **Azure SQL Data Warehouse.**
- Proficient in working with a wide range of ETL tools, data warehousing solutions, and big data technologies, **including Apache Airflow, Amazon Redshift, Apache Spark.**
- Experienced in the development and design of various scalable systems using Hadoop technologies in various environments.
- Data integration platforms such as **Apache Nifi, Talend, Informatica and Microsoft SQL Server Integration Services (SSIS).**
- Worked with database technologies like PostgreSQL, MySQL to store and manage data in Python applications.
- Good knowledge of Hadoop Ecosystem components like HDFS, MapReduce, HBase, Hive, and Pig.
- Skilled in Dashboard, visualization tools such as **Tableau and Power BI.**
- Experienced in migrating the Cassandra, and Hadoop clusters on AWS.

TECHNICAL SKILLS

Programming: Python (Pandas, NumPy, Pytorch, TensorFlow, Scikit-learn), SQL, Scala
ETL Tools: Apache Airflow, Azure Data Factory, Apache Kafka, Azure Stream Analytics, AWS Step Functions, SSIS
Big Data Technologies: Apache Spark, PySpark, Amazon EMR, Hadoop, Databricks, DBT
Data Integration Platforms: Talend, Azure Data Factory
Databases: Amazon (DynamoDB, S3, RedShift, Lambda), MySQL, PostgreSQL, MongoDB, Azure Cosmos DB, HBase, Oracle
Data Warehousing: Snowflake(Snowpipe, Snowpark, SnowSQL), Azure Synapse Analytics, Azure SQL Data Warehouse, Redshift
Version Control: GitHub, Bitbucket
Cloud Platforms: AWS EC2, Elasticache, Amazon RDS, Azure VMs, Blob Storage, Databricks, API Management
Serverless Computing: AWS Lambda, Azure Functions
Containerization: Docker
Data Modeling: ER diagrams, Star schema, Snowflake schema
Bug Tracking: Jira, ServiceNow
CI/CD: Jenkins
Data Visualization: Power BI, Tableau
Workflow Automation: Apache Airflow
Monitoring: Splunk

CERTIFICATIONS

Azure Data Engineer Associate - **Microsoft**	Jan 18, 2024
Architecting with Google Compute Engine - **Coursera**	May 31, 2021

EDUCATION

New Jersey Institute of Technology, Newark, NJ, USA | Master of Science in Computer Science
Rajalakshmi Engineering College, Chennai, TN, India | Bachelor of Engineering in Computer Science
Coursework: *Data Structures and Algorithms, Operating Systems, Data Mining, Advanced Database System Design, Web System Development, Java Programming*

WORK EXPERIENCE

CIGNA- NJ, USA   **MAR 2023 – DEC 2023**
Data Engineer

- Collaborated with data analysts and other data engineers, I was part of a team to extract and work with marketing data.
- Utilized Azure data lake to store raw data gathered in JSON, XML, text, and other related file format.
- Creating Tables, Stored Procedures, and extracted data using **T-SQL** for business users to perform data analysis and generate technical solutions.
- Developed scalable ETL pipelines leveraging **Azure Databricks for distributed data processing.** Used **Python and PySpark APIs for**

efficient data transformations and loading into Azure Data Lake Store.

- Built machine learning models with Azure Databricks using Python and common data science libraries **like NumPy, Pandas, scikit-learn.** Productionized models into containers for operationalization.
- Designed and built a multi-terabyte **Azure Synapse Analytics** data warehouse from the ground up to handle large-scale ingestion and transformation of millions of records (ELT)**.**
- Working with Big Data and Azure cloud services including Virtual Machines, Blob Storage, Databricks and CosmosDB.
- Involved in developing scripts and indexing strategies to migrate on-premise SQL Server and MySQL databases to Azure Synapse Analytics data warehouse.
- Designed Azure CosmosDB databases with appropriate partitions and provisioned throughput for high-performance OLTP workloads.
- Benefited and helped in creating and presenting informational reports for management based on data.
- Resolved production incidents by investigating and troubleshooting issues related to data infrastructure, ETL processes, and data quality.
- Created workflows for a job and scheduled Databricks notebooks executing on PySpark job cluster
- Developed end-to-end data pipelines leveraging Azure services including Blob Storage, Data Factory, Databricks, SQL Data Warehouse, and Functions to ingest, process, and analyze datasets.
- Created reusable Python modules for common data cleansing tasks like handling missing values, outliers to rapidly prepare datasets for analysis in Azure Databricks.
- Implemented Delta Lake on Azure Databricks for **ACID transactions** on large datasets enabling scalable analytics.
- Developed Python jobs and notebooks scheduled with orchestration tools like **Apache Airflow** to automate and monitor Azure Databricks workflows.
- Built and maintained data models in healthcare data warehouses, **Snowflake**, Optimized for querying and analytics of clinical data.
- Developed **Power BI dashboards** fueled by Azure Analysis Services to provide real-time analytics and insights for healthcare providers and insurance firms.

SWAGG TECHNOLOGIES – Bangalore, India   **JAN 2020 – JUL 2022**
Data Engineer
- Designed and implemented scalable and fault-tolerant data processing pipelines using AWS Glue, AWS Lambda, AWS DynamoDB, AWS CloudWatch.
- Optimized analytical query performance by architecting ETL pipelines with **Amazon Redshift**, showcasing hands-on experience in designing scalable data warehousing solutions.
- Developed an end-to-end ETL pipeline for working with customer and campaign data.
- Orchestrated ETL processes with hands-on **integration of Amazon DynamoDB** for NoSQL data storage, accommodating diverse data types seamlessly.
- Applied hands-on expertise in Apache Spark, Hive, and other big data processing frameworks on **Amazon EMR,** to perform data transformation, aggregation, and analysis, enabling efficient and effective data-driven decision making.
- Experienced data cleaning and enrichment workflows, Worked on ETL Pipelines to consume the data from APIs using Python, transform such data and load it into AWS S3 to be consumed for Reporting and Analysis purposes by Data Scientists
- Designed and implemented end-to-end ETL pipelines in Azure Databricks using Spark for data transformation and analysis.
- Designed and implemented ETL data pipelines using Apache Airflow to automate data workflows and scheduling tasks.
- Good Understanding of data processing tools like Apache Kafka, Apache Spark and knowledge of PySpark.
- Experienced in DBT to build and manage large-scale data transformation pipelines for complex data models.
- Orchestrated the integration of AWS cloud services, strategically designing data processing workflows to ensure optimal utilization of EC2 resources and efficient data transfer between S3 and EMR.
- Ensured data consistency and durability by implementing appropriate redundancy and backup mechanisms in S3 and DynamoDB.
- Utilized Tableau to create interactive dashboards and reports for business stakeholders, enabling them to easily visualize and understand complex data insights.
- Implemented real-time data pipelines by leveraging **Snowpipe integration with Kafka** to continuously ingest streaming data into Snowflake.
- Utilized SnowSQL to write efficient SQL queries for data extraction and analysis, enabling rapid data-driven decision making across the organization.
- Developed CI/CD pipelines using Git, Jenkins**,** and automated testing to improve code quality and deployment frequency of data applications.
- Experienced in using version control tools like GitHub and Bitbucket enabling efficient and secure code.
- Created and maintained a Python based data ETL pipeline to extract semi-structured and structured data.
- Extracted data from multiple sources including AWS DynamoDB, AWS Elasticache Redis and RESTful APIs.