Aswin Ravichandran

aswin.ravichandran06@gmail.com
LinkedIn
+1 (929) 595-3080
Harrison, NJ - USA

SUMMARY

- Extensive experience as a Data Engineer 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.
- Adept at collaborating with solution leads, project managers, data architects, and data scientists to create and integrate secure data solutions across diverse systems.
- Proven expertise in ETL processes, data transformation, and integration using tools such as Azure Data Factory, Redshift, and Informatica. Skilled in database design and modeling, with hands-on experience in SQL across multiple relational database platforms, including Microsoft SQL Server, Oracle, and PostgreSQL.
- Proficient in modern programming languages like Python, C#, and JavaScript, and experienced with cloud platforms such as AWS, Azure, and GCP. Familiar with containerization using Docker and orchestration with Kubernetes.
- Strong analytical abilities and a collaborative team player with a progressive mindset toward emerging technologies and operational improvements.

TECHNICAL SKILLS

Programming	Python (Pandas, NumPy, TensorFlow, Scikit-learn), SQL, Scala
ETL Tools	Apache Airflow, Apache Kafka, Azure Stream Analytics, AWS Step Functions
Big Data Technologies	Apache Spark (used on AWS EMR), PySpark, Amazon EMR (Elastic MapReduce), Hadoop, Databricks, Hive
Databases	Amazon (DynamoDB, S3, RedShift, Lambda), MySQL, NoSQL, Redis, PostgreSQL, MongoDB, Azure Cosmos DB, Oracle, HBase
Data Warehousing	Amazon Redshift, Snowflake, Azure Data Lake, Azure Synapse Analytics
Cloud Platforms	AWS (EC2, Amazon EBS, Lambda, RDS, Glue, Crawler, Quicksight) , Azure (VMs, Blobstorage, Data Lake Gen 2, Data Factory, Functions, Synapse Analytics, Databricks)
Containerization	Docker, Kubernetes
Data Modeling	ER diagrams, Star schema, Snowflake schema
Data Visualization	Power BI, Tableau
Others	Talend, Informatica, Jira, ServiceNow, Git, GitHub, Bitbucket

CERTIFICATIONS

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

WORK EXPERIENCE

MOUNT SINAI **MAR 2023 – PRESENT**

Role: Data Engineer / Analyst

Responsibilities:

- 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 Cosmos DB.

- Involved in developing scripts and indexing strategies to migrate on-premise SQL Server and MySQL databases to Azure Synapse Analytics data warehouse. Developed and automated data processes using programming languages like Python, C#, and JavaScript, enhancing operational efficiency.
- Implemented scripts and tools for automated data validation, monitoring, and reporting.
- Utilize Docker for containerization and Kubernetes for orchestration to enhance deployment processes and manage scalable applications.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.
- Automated data engineering tasks and workflows using programming languages like Python, C#, and JavaScript to enhance productivity and reduce manual effort.
- Developed **Power BI dashboards** fueled by Azure Analysis Services to provide real-time analytics and insights for healthcare providers and insurance firms.

VERSION 1 **JAN 2020 – JUL 2022**

Role: Data Engineer

Responsibilities:

- 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.
- Implemented **Lambda** functions in Python for data transformation tasks in AWS cloud.
- 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.
- Used **AWS CloudWatch** for monitoring pipeline performance, setting up alerts for job success, failures, and resource utilization.
- Designed and implemented end-to-end ETL pipelines in **AWS Glue** and Spark for data transformation.
- Utilized **AWS Data Pipeline** to automate the data processing workflows and ensure timely data availability.
- Good Understanding in tools like Apache Airflow, Apache Spark and PySpark.
- Monitored and optimized database migration tasks using **AWS DMS** Dashboard, CloudWatch metrics, and alerts, ensuring efficient and reliable data transfers.
- 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** to continuously ingest data into Snowflake.
- Utilized **Snow SQL** to write efficient SQL queries for data extraction and analysis, enabling rapid data-driven decision making across the organization.
- Developed CI/CD pipelines using **AWS code pipelines,** and automated testing to improve code quality and deployment frequency of data applications.
- Experienced in using version control tools like **GitHub** 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, external OLTP and RESTful APIs

***Environment*:** AWS, Python, spark, SQL, Migration, OLTP, OLAP, Code pipelines, Snowflake, GitHub.

EDUCATION

New Jersey Institute of Technology, Newark, NJ, USA | Master of Science in Computer Science

Coursework: Data Structures and Algorithms, Operating Systems, Data Mining, Advanced Database System Design, Web System Development, Java Programming

Rajalakshmi Engineering College, Chennai, TN, India | Bachelor of Engineering in Computer Science