NAZIA SYED

Email: syednazia@mymailsbox.net | LinkedIn | Ph No - (913)732-3511

SUMMARY: Microsoft-Certified Data professional with 7 years of Industry experience in Data Analytics, Data Engineering, and Business Intelligence. My focus on data integrity and cross-functional integration across all company departments has consistently boosted by >25% and optimized data pipelines.

WORK EXPERIENCE

U S Bank, Irving,TX, USA, *Data Engineer*   Dec 2022 – Present
- Implemented installation and configuration of multi-node cluster on Cloud using AWS on EC2.
- Wrote scripts and indexing strategy for a migration to Redshift from SQL Server and MySQL databases.
- Worked on AWS Data pipeline to configure data loads from S3 into RedShift.
- Configured Apache Airflow for AWS S3bucket and created DAGS to run Apache Airflow.
- Used AWS glue catalog with crawler to get the data from S3 and perform SQL query operations.
- Worked on PL/SQL Stored procedures, Functions, Triggers, Views and Packages and made use of Indexing, Aggregation and materialized views to optimize query performance.
- Worked on optimizing and tuning RedShift environment, enabled queries to perform faster for Tableau and SAS Visual Analytics.
- Redesigned ETL processes using AWS Glue, resulting in a 15% reduction in data inaccuracies.
- Worked with AWS services such as AWS Elastic File System (EFS) and Amazon S3 to implement scalable and durable storage solutions.
- Designed and implemented data partitioning and indexing strategies and optimized query performance.
- Implemented serverless data processing solutions using AWS Lambda and API Gateway
- Automated and orchestrated data pipelines with AWS Step Functions.
- Developed pyspark-based pipelines using pysark data frame operations to load data to EDL using EMR for jobs execution & AWS S3 a storage layer.
- Responsibe for designing,Development and testing of Developed Stored Procedures,Views and Triggers and implemented Copy activity, Custom Azure Data Factory Pipeline Activities.

Environment: Hive, AWS, Hadoop, HDFS, Python, Apache Airflow, Numpy, Pandas, Git, AWS, EMR, Glue, Jenkins, PL/SQL, Maven

HSBC, *Data Engineer*   Sep 2017 –Dec 2021
- Successfully implemented end-to-end data pipelines using Azure Data Factory, extracting, transforming, and loading data from diverse sources into Snowflake for efficient data processing.
- Integrated Azure Synapse Analytics, Azure Data Lake Storage, and Azure SQL Database to establish a unified data platform, ensuring seamless integration.
- Implemented Azure Data Factory comprehensively for diverse data ingestion, optimizing storage in Data Lake Storage Gen 2 for efficient handling of relational and unstructured data.
- Spearheaded the design and implementation of real-time streaming data pipelines using Apache Spark Structured Streaming, Kafka, or Azure Event Hubs, ensuring data consistency.
- Designed robust pipelines and executed complex data transformations and manipulations using ADF and PySpark with Databricks.
- Designed and implemented data processing workflows in Azure Databricks, harnessing the power of Spark to handle large-scale data transformations and analytics tasks.
- Leveraged Azure Data Lake Storage as a scalable data lake, implementing data partitioning and retention strategies for efficient storage and retrieval.
- Utilized Azure Blob Storage for cost-effective storage and retrieval of data files, implementing compression and encryption techniques for enhanced security.
- Integrated Azure Data Factory with Azure Logic Apps to orchestrate complex data workflows and trigger actions on specific events.
- Developed and deployed Azure Functions for data preprocessing, enrichment, and validation tasks within data pipelines.
- Optimized data pipelines and Spark jobs in Azure Databricks for improved performance, including tuning Spark configurations, leveraging caching techniques, and utilizing data partitioning.
- Collaborated closely with cross-functional teams, including data scientists, data analysts, and business stakeholders, to understand data requirements.

Environment: Tableau Desktop, Server, Online, AWS, DB2, Python, Power BI, Splunk, BigQuery. Azure Data Factory, PySpark, Databricks

Capgemini, Pune, India, *Data Analyst*   July 2016 –Aug 2017
- Developed PL/SQL programming that included writing Views, Stored Procedures, Packages, Functions, and Database Triggers.
- Performed data analysis and data profiling on various source systems including Oracle, SQL Server, and DB2.
- Wrote complex SQL scripts and PL/SQL packages, to extract data from various source tables of data warehouse.
- Designed Informatica Mappings and Workflows to be re-started after a failure.
- Built a POC on rounding transaction amounts to 6 decimal places in Informatica.
- Used XML Parsers and Generators to handle inbound and outbound XMLs in Informatica.
- Optimized the SQLs and Informatica mappings which handled millions of records.
- Developed reusable components in Informatica, Oracle, and UNIX.
- Involved in all phases of SDLC using Agile and participated in daily scrum meetings with cross-teams.

Environment: MS SQL, Informatica PowerCenter, SSIS, SSRS, Microsoft Excel, Power BI Desktop, Tableau Server, MySQL, T-SQL, Python, Jupyter Notebooks.

Master of Science – *CS, University of Central Missouri*   Jan 2022 – Dec 2022
Relevant Coursework: BIGData, Machine Learning, Foundation of Analytics, Statistics, Research Engineering & Project Management.

CERTIFICATIONS & ACTIVITIES
- *'Azure Data Engineer'* from Microsoft | *'Microsoft Azure'* from Coursera |'Big Data Analytics' from INSOFE

SKILLS
Programming: Python, R, DAX,SQL, PL-SQL & Libraries (NumPy, Pandas, Matplotlib, PySpark), Core Java , javascript , C++
Database: MySQL, MS SQL Server, PostgreSQL, Snowflake, Oracle
Tools: Tableau, Power BI , SSRS,PowerApps, Tableau Prep, Google Data Flow, Hadoop, HDFS, Spark, Hive, Pig, MapRaduce, EMR, Kafka, Sqoop
Data Science & Analytics: Machine Learning, Data Warehousing, Data Mining, Databricks, Data Analytics, Data Modeling, Data Engineering, Database Querying, Microsoft Office 365, Power Query, DAX Functions, ETL Processing, Data Pipelines, Data Transformation, Ad-Hoc Analytics, Git Repository, JIRA, Data Verse, Data Visualization Docker, Jenkins, REST API's.