Priya
Data Engineer
+1(347)607-5553
Priyalaki09@gmail.com
https://www.linkedin.com/in/priya0099/

Professional Summary:

- **4+ years of experience** as a **Data Engineer** specializing in **Business Intelligence (BI) solutions** with **Microsoft SQL Server 2016/2018**.
- Strong knowledge of **Azure SQL** and hands-on experience in building **end-to-end Azure data solutions**, including:
    - **Extracting data from SQL Server using Azure Data Factory (ADF)**
    - **Storing data in Azure SQL Server via Blob Storage**
    - **Developing interactive reports using Power BI Desktop**
- Expertise in **SQL Server (T-SQL), SSIS, SSAS, SSRS**, and **data visualization** with **Power BI**.
- Proficient in **ETL pipeline development** using **SSIS**, leveraging **For Each Loop, Conditional Split, Lookup, Derived Column, and Data Flow Transformations** for efficient data processing.
- Hands-on experience in **Azure Data Lake Storage (ADLS), Data Lake Analytics, and Databricks** for large-scale data management and integration.
- **Experience in migrating on-premises databases to Azure**, including **Azure SQL Database, Azure Data Warehouse, PowerShell, and SSIS in Azure**.
- Strong background in **Data Modeling** with **Star Schema, Slowly Changing Dimensions (SCDs), and OLTP/OLAP integration** for analytical reporting.
- Proficient in writing complex **T-SQL queries, Stored Procedures, Views, Common Table Expressions (CTEs), User-Defined Functions, Cursors, and Indexes**.
- Skilled in **SSRS reporting**—developing **Drill-Through, Parameterized, Linked, and Ad-Hoc Reports** based on business requirements.
- Experienced in **ETL job scheduling** using **SQL Server Agent**, with robust **logging and error handling** for monitoring.
- Expertise in **Power BI dashboard development**, creating rich, interactive, and visually appealing reports with **DAX, Power Pivot, and Tabular Models**.
- Experienced in **data extraction, transformation, and loading (ETL) using DTS/SSIS, Import Export Wizard, and Bulk Insert**.
- **Strong analytical and problem-solving skills**, with hands-on experience in **Data Warehousing and Performance Optimization**.
- **Knowledge of database migration strategies** across different SQL Server versions, ensuring **seamless data transformation and integrity**.

Technical Skills:

Big Data & Distributed Computing	Hadoop, Apache Spark, Hive, Pig, Sqoop, Kafka, Oozie, Delta Lake, Airflow,Apache Flink, IBM Streams, Spark Streaming

•

Databases & Data Warehousing	SQL Server, Azure SQL, MySQL, PostgreSQL, Netezza, Microsoft Access, MongoDB, DynamoDB, HBase, Snowflake, BigQuery, Redshift, Azure Synapse Analytics Azure Data Factory (ADF), AWS Glue, Informatica, SSIS, DBT
Programming & Scripting:	Python, Scala, SQL, PowerShell, VB
Cloud Technologies	Azure (Data Bricks, Data Factory, Data Lake), AWS (S3, RDS, Athena, Glue)
Data Visualization & Reporting	Power BI, Tableau, SSRS, Crystal Reports.

Education: Masters - Sacred Heart University

Professional Experience

Client: Inovalon - Bowie, MD
Role: Data Engineer | Dec 2023 – Present

Roles and Responsibilities:

- Designed and developed **ETL specifications**, gathered requirements, and prepared detailed documentation for data transformation processes.
- **Migrated on-premises SQL databases** to **Azure Synapse Analytics** using **Azure Data Factory (ADF)**, optimizing database architecture for performance and scalability.
- Developed **Azure Data Factory pipelines** for seamless **data movement between Azure BLOB Storage and SQL Server**.
- Implemented **ad-hoc data analysis solutions** using **Azure Data Lake Analytics, Azure Data Lake Storage (ADLS), HDInsight, and Databricks**.
- Worked extensively with **Microsoft on-prem data platforms**, including **SQL Server, SSIS, SSRS, and SSAS** for data integration and reporting.
- Developed **reusable ADF pipelines** to call **REST APIs** and consume **Kafka events** for real-time data processing.
- Utilized **Control-M for scheduling DataStage jobs** and **Logic Apps for automating ADF pipeline execution**.
- Configured **CI/CD pipelines in Azure DevOps**, ensuring secure and efficient build and release processes for **.NET and Java applications** using **Azure Git**.
- **Migrated legacy ETL workflows** from **SSIS and MS Access** to **Azure Data Factory pipelines** while preserving existing business logic.
- Built **high-performance data ingestion pipelines** using **Azure Data Factory and Azure Databricks**, handling large-scale data workloads efficiently.
- Developed **dynamic ADF pipelines** using **parameters and triggers** based on **event-driven logic (e.g., file availability on Blob Storage, schedule-based execution, and Logic Apps integration)**.
- Assisted in **SQL query optimization**, writing **DDL & DML scripts** to support system migration and **data mapping** between legacy and cloud-based data warehouses.
- Utilized **Spark Streaming APIs** for real-time data transformation, integrating **Azure Data Lake Store, PolyBase, and T-SQL** for efficient bulk data ingestion into **Azure SQL Data Warehouse**.

•

- Created **Azure Runbooks** to automate **scaling of Azure Analysis Services and Azure SQL Data Warehouse** for cost and performance efficiency.
- Led **Azure SQL Data Warehouse upgrades** from **Gen1 to Gen2**, optimizing **Fact and Dimension table design** with appropriate **distribution strategies (Hash, Replicated, Round-Robin)**.
- Developed **Power BI and SSRS reports** and designed **SSAS cubes** to facilitate **self-service BI and data visualization**.
- Built **Azure Data Factory pipelines** to extract, transform, and load data from **on-premises SQL Server to Azure Data Lake Storage**.
- Designed and implemented **massive data lake pipelines**, integrating **Kafka for near real-time data streaming** into **Cassandra and Amazon Redshift**.

Tech Stack: Azure Data Factory, Azure Data Lake, Azure Synapse Analytics, Databricks, SQL Server (SSIS, SSRS, SSAS), Power BI, Kafka, Redshift, Cassandra, Java, Selenium, ER Studio, Teradata 13.1, Oracle, Python, Tableau, Hadoop, Spark, Scala, Hive, Control-M, Logic Apps, PolyBase.

Wells Fargo, CA
Data Engineer | Aug 2022 – Nov 2023

Roles and Responsibilities:

- **Designed and maintained a scalable Azure data framework**, ensuring **data availability, reliability, and security** to support **business intelligence and analytics solutions**.
- **Developed and optimized ETL pipelines** using **Azure Data Factory (ADF)**, improving **data processing speed by** enhancing overall system performance.
- **Built and deployed resilient data pipelines** using **ADF and Cloud Functions**, integrating **Azure Storage, Azure Data Lake, and Azure SQL** for seamless data movement and storage.
- **Implemented Snowflake-based data solutions**, leveraging **SnowSQL and Snowpipe** for efficient **data ingestion, transformation, and querying**, improving data management workflows.
- **Configured monitoring and logging for Apache Flink jobs**, enabling **real-time performance tracking and issue diagnosis** in streaming data pipelines.
- **Developed adaptable data ingestion and transformation frameworks** to migrate **on-premises data to cloud-based platforms** in both **batch and real-time** processing modes.
- **Authored well-documented PowerShell scripts** for automated **data processing workflows**, ensuring maintainability and easy troubleshooting.
- **Managed fault tolerance mechanisms** in **IBM Streams**, ensuring high availability and reliability in real-time data processing.
- **Implemented dataflows and transformations in Azure Data Factory**, streamlining data integration and improving workflow efficiency.
- **Utilized various Azure services** (ADF, Azure SQL Server, Blob Storage, Azure Data Lake, Azure Synapse) to develop **data-driven solutions tailored to business needs**.
- **Designed and maintained CI/CD pipelines** using **Cloud Build and Cloud Run**, ensuring **secure and automated deployments** in compliance with industry best practices.
- **Collaborated in Agile Scrum environments**, actively participating in **daily stand-ups, sprint planning, and demo meetings**, ensuring efficient communication and project delivery.
- **Conducted end-to-end testing** of ADF data pipelines to **validate data accuracy, consistency, and completeness** from ingestion to reporting.
- **Optimized Snowflake Data Warehouse performance**, refining **Tables, Views, and Stored Procedures** to enhance query execution speed and database efficiency.
- **Developed and managed complex ADF pipelines** using **Linked Services, Datasets, and Dataflows** for efficient **ETL operations from diverse sources**, including **Azure SQL, Blob Storage, and Synapse Analytics**.

•

- **Performed load testing on Apache Flink applications**, ensuring system scalability and optimizing performance under high data loads.

Tech Stack: Azure Data Factory (ADF), Snowflake, SnowSQL, Snowpipe, Azure SQL Server, Blob Storage, Apache Flink, IBM Streams, PowerShell, CI/CD (Cloud Build, Cloud Run), DBT (Data Build Tool)

Michelin, INDIA
Data Engineer | June 2019 – July 2021

Roles and Responsibilities:

- **Collaborated with clients and stakeholders** as a senior BI developer, assisting architects in **designing technical solutions** that met business requirements.
- **Developed optimized Azure Data Factory (ADF) pipelines** with **parameterization**, enabling **parallel and sequential data loading** while performing **complex data transformations in Spark using Scala**.
- **Converted Hive/SQL queries into Spark transformations** using Scala, enhancing processing efficiency for large-scale datasets.
- **Implemented and scheduled ADF pipelines** using various triggers (**Schedule, Tumbling Window, and Event-based triggers**) to ensure automated and timely data processing.
- **Designed and implemented complete CI/CD pipelines** for ADF environments, utilizing **Azure DevOps and Azure Key Vault** for secure credential management.
- **Optimized ADF pipeline performance**, implementing **proactive monitoring solutions** that significantly reduced pipeline execution times.
- **Developed Azure Logic Apps** for automated monitoring, improving system efficiency by **45%** and reducing manual intervention.
- **Automated database monitoring alerts** to track **Azure SQL Database performance**, dynamically scaling resources **up/down** based on usage patterns.
- **Designed and implemented database solutions** in **Azure SQL Data Warehouse and Azure SQL**, creating **efficient stored procedures, triggers, user-defined functions, indexes, and views** for optimized data retrieval.
- **Performed ETL operations using Azure Data Factory and Databricks**, successfully loading data into **Azure SQL Data Warehouse** for business analytics.
- **Designed, reviewed, and created database objects** (views, indexes, partitions) based on logical and physical design models to **optimize query performance**.
- **Developed complex ETL packages using SSIS**, implementing **incremental data loading** and **seamless data extraction** from multiple sources.
- **Experienced in recursive and replicated joins in Hive**, ensuring efficient data transformations and analytics in big data environments.
- **Utilized advanced SQL queries** to validate data accuracy, ensure data consistency, and optimize database performance.
- **Worked with AWS cloud services** including **AWS Glue, Athena, S3, RDS, and DynamoDB**, integrating them into data processing workflows.
- **Utilized distributed computing technologies** such as **Spark, Hadoop, and Kafka** to build scalable and high-performance data solutions.
- **Developed interactive dashboards and reports** using **Power BI and Tableau**, enabling data-driven decision-making for stakeholders.

Tech Stack: Python, Power BI, AWS Glue, Athena, SSRS, SSIS, AWS S3, AWS RDS, DynamoDB, SQL, Tableau, Distributed Computing, Snowflake, Spark, Kafka, MongoDB, Hadoop