***Senior Certified Azure Data Engineer***

Highly skilled Azure Data Engineer with extensive experience in designing, implementing, and managing robust data solutions. Proficient in developing and optimizing data pipelines, integrating diverse data sources, and ensuring data quality and security. Demonstrated expertise in leveraging Azure services to deliver scalable and high-performance data architectures.

**PROFILE SUMMARY**

* A Self-Motivated and Result-Oriented IT professional with **13+** Years of extensive SDLC experience in client-server and web applications on Big Data Systems.
* Extensive experience in designing and managing data pipelines using Azure Data Factory (ADF). Adept at orchestrating and automating data movement and transformation processes through the creation of data flows, parameterized pipelines, and custom activities for complex ETL processes.
* Expertise in data modelling and performance tuning, coupled with a strong ability to collaborate with cross-functional teams to meet business intelligence and analytical needs. Committed to driving data-driven decision-making through innovative and efficient data engineering practices.
* Proficient in utilizing Azure Databricks for big data analytics and transformation. Experienced in writing and optimizing Spark jobs, managing clusters, and performing complex data transformations and aggregations using Scala, Python, and SQL.
* Automated data quality monitoring and metadata tagging using Azure Data Factory Mapping Data Flows and Microsoft Purview policies.
* Skilled in creating and optimizing scalable ETL pipelines and data workflows to efficiently process and integrate data from various sources, including SQL databases, Azure Blob Storage, Azure Data Lake Storage, S3 buckets, and NoSQL databases like Cosmos DB and MongoDB.
* Adept at managing cross-cloud data connections and integrations between platforms such as Microsoft Azure, Amazon Web Services (AWS), and Google Cloud Platform (GCP). Capable of ensuring seamless data flows and interoperability across different cloud environments.
* Experienced in integrating data sources, including on-premises databases and cloud-based applications. Proficient in using Azure Data Sharing for secure data transfer and integration.
* Proficient in using Azure Stream Analytics and Event Hubs for real-time data ingestion and processing. Skilled in implementing notification mechanisms using Azure Logic Apps and Azure Functions to automate alerts and updates related to data pipeline statuses and system events.
* Experienced in implementing and managing data security measures, including access controls, encryption, and compliance with regulatory standards. Adept at applying Azure Active Directory (AAD) and Role-Based Access Control (RBAC) to secure data access and manage identities.
* Experienced in monitoring, tuning, and optimizing data pipelines and storage performance using Azure Monitor, Log Analytics, and Application Insights to ensure high efficiency and reliability.
* Proficient in implementing CI/CD pipelines for data engineering projects. Experienced with tools such as Azure DevOps and GitHub Actions to automate the build, test, and deployment processes of data solutions.
* Excellent communication and teamwork skills, with a proven track record of working closely with data scientists, analysts, and business stakeholders to deliver effective data-driven solutions.
* Proficient in utilizing GitLab for end-to-end DevOps lifecycle management in data engineering projects, including version control, issue tracking, CI/CD, and container registry.
* Proven track record of integrating Azure DevOps with other Azure services such as Azure Data Factory.
* Deploy, configure, maintain compute on azure cloud and automation for managed azure services like storage, accounts, virtual networks, network services, azure active directory, API management, azure websites. Responsible for creating multi-region, multi-zone Azure cloud infrastructure.

**EDUCATION**

* Masters in Information Systems & Technology from Wilmington University, USA – 2023
* Bachelors in Information Technology from Andhra University, India –2011

**TECHNICAL STACK**

|  |  |
| --- | --- |
| **Category** | **Skills** |
| Cloud Platforms | Microsoft Azure, Amazon Web Services (AWS) |
| Languages | Python, SQL, Scala, T-SQL,KQL |
| Version Control | Git, Azure DevOps, GitHub |
| Database | Azure SQL Database, Azure SQL Data Warehouse, Azure Cosmos DB, PostgreSQL, MySQL, Snowflake |
| Big Data | Azure Databricks, Apache Spark, Pyspark, Hadoop, Azure Stream Analytics, Event Hubs, Kafka, ETL, Structured Streaming |
| Reporting Tools | Power BI, Tableau |
| Documentation Tools | MS Office, MS SharePoint |
| Methodologies | Agile, Scrum |
| Operating Systems | Windows, Linux |
| Data Integration | Azure Data Factory (ADF), SSIS, Data Mapping |
| Data Storage | Azure Blob Storage, Azure Data Lake Storage, S3 |
| Security | Azure Active Directory (AAD), Role-Based Access Control (RBAC), Data Masking, Encryption |
| Real-Time Processing | Azure Stream Analytics, Event Hubs |
| CI/CD | Azure DevOps, GitHub Actions |
| Monitoring & Optimization | Azure Monitor, Log Analytics  |

**CERTIFICATIONS**

* Microsoft Azure Fundamentals (AZ-900)
* Azure Data Engineer Associate (DP-203)
* Microsoft Azure Developer (AZ-204)

**PROFESSIONAL EXPERIENCE**

**Client: National Life, TX Jun 2023 - Present
Role: Senior Azure Data Engineer**

Developed and maintained scalable data pipelines, optimized data storage solutions, and ensured data security and compliance with industry standards. Worked closely with stakeholders to integrate various data sources, enhancing real-time analytics and business intelligence capabilities. Leveraged Azure services to deliver efficient data solutions, supporting strategic decision-making processes and driving improvements in data-driven initiatives.

* Designed and built comprehensive end-to-end data pipelines using Azure Data Factory (ADF), orchestrating data movement and transformation across multiple sources to streamline data integration and processing.
* Integrated on-premises data sources with Azure cloud storage solutions, including Azure Blob Storage, Azure Data Lake Storage, and Azure SQL Database, MS SQL Database, to centralize data for analysis and reporting.
* Proficient in writing Python scripts/programs for automation, ETL, and data processing tasks. Data reconciliation scripts were built on spark to move data from Bronze to Silver layers as per Medallion architecture.
* Managed and optimized data storage solutions with Azure SQL Data Warehouse, Cosmos DB, and Azure Data Lake, implementing best practices for data partitioning, indexing, and compression to enhance performance and scalability.
* Designed and deployed a unified data architecture using Microsoft Fabric, integrating Synapse Data Engineering, Data Lakehouse, and Power BI for end-to-end analytics.
* Built scalable and secure data pipelines using Azure Data Factory within the Microsoft Data Fabric ecosystem, ingesting data from SQL Server, Oracle, and REST APIs.
* Applied advanced data partitioning, indexing, and compression techniques within Azure Synapse Analytics to improve query performance and storage efficiency, supporting high-speed analytics and reporting.
* Designed and implemented scalable data pipelines using Snowflake on Azure, ingesting and transforming data from Azure Data Lake Storage (ADLS) and Azure Synapse Analytics.
* Developed ETL/ELT processes using Snowflake's SQL capabilities, streamlining data integration across multiple Azure-based sources.
* Configured Snowflake warehouses, databases, and role-based access controls (RBAC) to optimize performance and security within Azure environments.
* Integrated Azure Data Factory (ADF) with Snowflake for automated data ingestion and transformation workflows.
* Implemented robust data security measures using Azure Active Directory (AAD) and Role-Based Access Control (RBAC) to manage identities and secure data access, ensuring compliance with organizational and regulatory standards.
* Created Pyspark frame to bring data from RDBMS to Amazon S3.
* Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL
* Ensured adherence to data governance policies by deploying data masking, encryption, and auditing solutions within Azure environments to protect sensitive information and maintain data integrity.
* Leveraged Azure Databricks and HDInsight for big data processing and real-time analytics, writing and optimizing Spark jobs, managing clusters, and creating interactive notebooks to facilitate data-driven insights and machine learning workflows.
* Implemented Pyspark scripts for efficient transformations on large datasets in Azure Blob storage via Azure Databricks. Built data layers for bronze, silver and gold layers as per medallion architecture.
* Proficient in writing Python scripts/programs for automation, ETL, and data processing tasks. Data reconciliation scripts were built on spark to move data from Bronze to Silver layers as per Medallion architecture.
* Programmed in Hive, Spark SQL, and Python to streamline the incoming data and build the data pipelines to get useful insights, and orchestrated pipelines using Azure Data Factory.
* Worked on ORC, Parquet file formats on HDInsight, Azure Blobs and Azure tables to store raw data.
* Developed Spark applications using Python (Pyspark) for easy Hadoop transitions and used Spark APIs over Hortonworks Hadoop YARN to perform analytics on data in Hive.
* Developed and maintained Lakehouse models in One Lake as part of Microsoft Fabric, supporting cross-team collaboration and governed self-service analytics.
* Leveraged Microsoft Purview for end-to-end data lineage, cataloguing, and governance, enhancing data discoverability and regulatory compliance.
* Created Azure SQL database, performed monitoring and restoring of Azure SQL database. Performed migration of Microsoft SQL server to Azure SQL database.
* Created reusable Dataflows Gen2 in Microsoft Fabric to support multiple analytical workloads with minimal redundancy, improving efficiency and consistency.
* Created data ingestion pipelines on Azure HDInsight Spark cluster using Spark SQL and integrated with Cosmos DB.
* Adept at collaborating with cross-functional teams to deliver scalable and efficient data structures using snowflake.
* Created many PySpark and SparkSQL scripts in synapse notebooks for doing data transformations as per the given business requirements.
* Creating Pipelines in Azure Data Factory (ADF) by configuring Linked Services/Integration Runtime to Extract, Transform and load data from different sources into Azure Data Lake Store (ADLS), Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool and backwards.
* Extracting, Parsing, Cleaning, and ingesting the incoming web feed data and server logs into the HDInsight and Azure Data Lake Store by handling structured and unstructured data.
* Setup Databricks Enterprise Platform environment created cross-account role in Azure for Databricks to provision Spark clusters to access Azure Data Lake Store (ADLS).
* Architect & implement medium to large scale BI solutions on Azure using Azure Data Platform services (Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics, Azure SQL DW, HDInsight/Databricks, NoSQL DB).
* Utilized Azure Stream Analytics and Azure Event Hubs for real-time data ingestion and processing, enabling timely and actionable insights from streaming data sources.
* Employed Azure DevOps for CI/CD automation, streamlining the deployment of data pipelines and reducing manual interventions to ensure consistent and reliable deployments across development, staging, and production environments.

**Client: BDSI, FL Jan 2023 – May 2023
Role: Data Engineer II**

Designed and developed a comprehensive data pipeline to integrate diverse data sources for retail sales analytics. Utilized Azure services to orchestrate data movement, transform data, and deliver actionable insights, supporting strategic decision-making and operational efficiency.

* Designed and implemented an end-to-end data pipeline using Azure Data Factory (ADF) for orchestrating and transforming data, seamlessly integrating various data sources for comprehensive retail sales analytics.
* Build python scripts to automate Databricks cluster creation process as per business requirement using Databricks API. Developed access policies for ML model execution as per corporate business units.
* Tuned Snowflake query performance by leveraging clustering keys, result caching, and warehouse sizing for cost-effective data processing.
* Implemented data partitioning and data sharing strategies using Snowflake's Zero-Copy Cloning and Secure Data Sharing features.
* Used Snow pipe to enable real-time data ingestion from Azure Blob Storage into Snowflake for near real-time analytics.
* Developed and maintained CI/CD pipelines for Snowflake object deployment using tools like Azure DevOps and dbt (Data Build Tool).
* Orchestrated the movement and integration of data from multiple sources, including SQL databases, Azure Blob Storage, Azure Data Lake Storage, and SFTP servers, through Azure Data Sharing to centralize and unify retail sales data for detailed analysis.
* Developed and managed data ingestion processes to collect and store raw sales data, ensuring data consistency and reliability across the entire pipeline, enabling accurate and timely analytics.
* Leveraged Azure Databricks to create and manage data transformation workflows, utilizing Spark for big data processing. Designed interactive notebooks and Spark jobs to clean, aggregate, and transform retail sales data into valuable insights.
* Utilized Azure Data Lake Storage and Azure SQL Database for efficient storage and management of transformed data. Implemented data partitioning, indexing, and compression techniques in Azure Synapse Analytics to enhance query performance and optimize storage efficiency.
* Ensured secure data access and compliance with data governance policies by applying Azure Active Directory (AAD) for authentication and Role-Based Access Control (RBAC) for authorization. Implemented data masking, encryption, and auditing practices within Azure environments to protect sensitive information.
* Created Python notebooks on Azure Databricks for processing the datasets and loading them into Azure SQL databases.
* Worked on advanced analytics using Python with Modules like Pandas, NumPy and Diplomate for data extraction and manipulation.
* Employed Azure Stream Analytics and Azure Event Hubs for real-time data ingestion and processing, facilitating timely updates and insights into sales performance to support dynamic business decisions.
* Created detailed reports and interactive dashboards using Power BI, integrating with Azure services to deliver actionable insights and enhance decision-making processes.
* Monitored and optimized data pipelines and storage performance using Azure Monitor, Log Analytics, and Application Insights to ensure high operational efficiency and address issues proactively.
* Implemented CI/CD pipelines with Azure DevOps to automate the deployment of data pipelines and Databricks workflows, reducing manual interventions and ensuring consistent, reliable deployments across development, staging, and production environments.
* Collaborated with cross-functional teams to gather requirements and deliver data solutions that align with business objectives, ensuring that the analytics solution effectively meets stakeholder needs and supports strategic goals.
* Used Scala for concurrency support, developing map-reduce jobs for JVM-based data processing.
* Executed SQL queries and published data for interactive Power BI dashboards and reporting.
* Documented technical designs, data flows, and standard operating procedures to facilitate knowledge sharing, maintain process consistency, and ensure effective project management.
* Applied cost-saving strategies by leveraging reserved capacity, scaling resources based on demand, and using Azure Cost Management tools to optimize expenditures and maximize return on investment.

**Client: ServiceNow. India Jun 2019 – Jun 2022
Role: Big Data Engineer**

Developed and maintained scalable data pipelines, optimized data storage solutions, and ensured data security and compliance with industry standards while working closely with stakeholders to integrate various data sources for enhanced real-time analytics and business intelligence. Leveraged Azure services to deliver efficient data solutions, supporting strategic decision-making processes for the bank.

* Utilized Agile methodologies to effectively manage the full life-cycle development of data engineering projects, ensuring iterative progress and timely delivery.
* Designed and developed SQL Server Integration Services (SSIS) packages to extract data from heterogeneous sources, including Excel, flat files, and DB2, and load it into destinations such as Azure SQL Database, Azure SQL Data Warehouse, and flat files.
* Created and published interactive Power BI dashboards from Azure Analysis Services in the Power BI workspace. Developed and maintained SQL tables, indexes, stored procedures, views, and triggers using T-SQL to support Power BI reporting and data visualization.
* Designed, maintained, and optimized databases, schema objects, SQL queries, stored procedures, indexes, functions, and views to facilitate data migration and ad-hoc reporting. Used SSIS to extract and transform data from various source systems, including Oracle, SQL Server, and flat files.
* Implemented Azure Key Vault to securely manage keys and passwords, providing encryption and protecting against vulnerabilities.
* Developed components using Python with Flask framework for data retrieval and manipulation.
* Worked on Python OpenStack APIs and used Python scripts for database updates and file manipulation.
* Developed and maintained ETL workflows and mappings to extract, transform, and load data into target environments, adhering to the current data loading architecture or on an ad-hoc basis as required.
* Migrated Databricks ETL jobs to Azure Synapse Spark pools, handling structured data from Oracle source systems and ingesting it through Synapse pipelines using the Oracle SQL connector.
* Implemented end-to-end ETL/ELT processes to extract, transform, and load data from diverse sources into target systems, achieving a 60% reduction in errors and a 30% improvement in data quality and consistency.
* Built data model in Azure Synapse Analytics (Data Warehouse) and developed PySpark ETL jobs to write data into Redshift post-transformation. Created SQL scripts for Change Data Capture (CDC) to de-duplicate data for Slowly Changing Dimension (SCD) Type-2 tables.
* Implemented an email notification feature using Azure Logic Apps to send pipeline status updates and alerts.
* Implemented Dimensional Data Modelling with multi-dimensional star and snowflake schemas.
* Conducted code change migration and executed SQL queries to query the Repository DB.
* Utilized Spark Data Frame API for data transformations and PySpark scripts for streaming data processing.
* Built data processing pipelines in PySpark, including data extraction, merging, enrichment, and loading into data warehouses.
* Worked on Kafka for live streaming data and conducted analysis on real-time events.
* Developed ETL jobs to process fixed-width files from a mainframe server and load them into Redshift on daily basis using batch processing.
* Applied Slowly Changing Dimensions techniques to incrementally load dimension tables into the target data warehouse, ensuring accurate historical tracking and data management.
* Implemented CI/CD pipelines for data engineering projects using Azure DevOps and GitHub Actions, automating the build, test, and deployment processes to ensure consistent and reliable delivery of data solutions.

**Client: Master Card, India Aug 2014 – Jun 2019
Role: Data Engineer/ETL Developer**

Demonstrated strong problem-solving abilities in ensuring data accuracy, consistency, and quality across all stages of ETL processes. Experienced in implementing CI/CD pipelines for automated deployment and continuous integration in Azure environments. Possesses excellent communication and collaboration skills, with a proven track record of working closely with data engineers, architects, and stakeholders to deliver scalable, efficient, and high-quality data solutions.

* Designed, developed, and optimized ETL processes using Microsoft technologies, including SQL Server Integration Services (SSIS), Azure Data Factory (ADF), and SQL Server. Efficiently extracted, transformed, and loaded data from diverse sources to meet business needs.
* Led the migration of data from legacy systems to modern Microsoft-based platforms, ensuring data integrity, accuracy, and minimal downtime throughout the migration process.
* Analysed and optimized existing ETL workflows to enhance performance, scalability, and reliability. Ensured that data processing workflows met business requirements and delivered high-quality results.
* Developed and implemented data mapping, transformation logic, and business rules to accurately migrate and integrate data into target systems. Adhered to data quality standards to ensure reliable and accurate data integration.
* Designed, developed, and optimized database schemas, tables, views, and stored procedures to support application and reporting requirements. Ensured efficient data storage and retrieval to meet performance needs.
* Optimized SQL queries and database performance by analysing query execution plans, indexing strategies, and server configurations. Reduced latency and improved processing times to enhance overall database performance.
* Collaborated with application developers to integrate database components into applications, providing guidance on best practices for database interaction and data access. Ensured seamless integration and efficient data handling.
* Regularly monitored database health and performance, performed routine maintenance tasks such as indexing and database cleanup, and addressed any issues that arose to maintain optimal database functionality.
* Developed and optimized complex SQL queries to retrieve and manipulate data efficiently. Applied techniques such as query indexing and execution plan analysis to enhance performance and reduce latency.
* Created complex queries based on the business requirement using PySpark/SparkSQL in Azure Synapse Spark pools.
* Designed and implemented stored procedures to encapsulate repetitive or complex business logic, improving code reusability and maintainability. Ensured procedures were optimized for performance and adhered to best practices.
* Utilized parameterized queries in stored procedures to enhance security (prevent SQL injection) and improve query performance by reusing execution plans.
* Implemented robust error handling within stored procedures using TRY...CATCH blocks to manage exceptions and maintain data integrity during data operations.
* Managed transactions within stored procedures to ensure data consistency and rollback operations in case of errors or failures.
* Employed dynamic SQL in stored procedures to handle scenarios requiring dynamic query construction while maintaining security and performance.
* Created and utilized scalar and table-valued user-defined functions to encapsulate reusable logic, perform complex calculations, and return results in a structured format.
* Leveraged functions to perform aggregations, data transformations, and calculations within queries, improving data analysis and reporting efficiency.

**Client: S&P Global Private Limited, India May 2011 – Jul 2014
Role: Software Developer**

* Provide technical support for RESTful APIs, diagnosing and resolving issues reported by users and clients.
* Utilize monitoring tools to track API performance, response times, and uptime, ensuring optimal operation.
* Manage incidents and service requests related to REST APIs, ensuring timely resolution and communication with stakeholders.
* Designed and developed requirements from insurance clients following Agile Software Development Life Cycle methodology including analysis, design, development, testing.
* Coordinated in a team to complete client requirements in Enterprise Product using Java, Groovy, and spring while handling the support tickets.
* Utilized my in-depth knowledge in redesigning the RESTful web APIs which in turn improved the response time by 25%.
* Recognized for excellent troubleshooting skills, and involved in the deployment of timely fixes. Implemented business rules using Drools which improved scalability of the application.
* Designed and Built Cognos Report and Query Studio reports. Actively participated in code reviews to achieve high code standard, increased code coverage to 85% using Test-Driven Development (TDD) approach in the application using JUnit.
* Assisted stake holders in taking crucial decisions by performing root cause analysis of business usage of applications.
* Involved in designing, developing and testing of J2EE components like Java Beans, Java, XML, Collection Framework, JSP, JMS, JDBC and deployments in WebLogic Server.
* Effectively developed Action classes, Action forms, JSP, JSF and other configuration files like struts-config.xml, web.xml.
* Worked on Backend Development, and fixed production issues.
* Implemented spend management system, designed invoice workflow functionalities thereby reducing human efforts.
* Create and maintain user manuals and API documentation to assist developers in understanding and utilizing API endpoints effectively.
* Assist in managing different versions of APIs, ensuring backward compatibility and smooth transitions for clients.
* Work closely with development teams to analyze and resolve API-related bugs and enhance functionality.