A picture containing text, sign, blue

Description automatically generatedA picture containing text, sign, outdoor, blue

Description automatically generated**Shreenivasan**

**Cloud Data Engineer   
Employer-details:[Globexus963@gmail.com](mailto:Globexus963@gmail.com)Mobile :** 732-427-8156

**SUMMARY:**

* Over seven years of experience as a Cloud Data Engineer, specializing in high-volume data application development, addressing complex architectural and scalability challenges. Proficient in Spark, Databricks, the Azure suite, Snowflake, Kafka, and Airflow.
* Led the integration of advanced AI technologies, such as GENAI, into SharePoint on Azure, demonstrating proficiency in managing extensive data lakes using Azure Blob Storage.
* Skilled in streamlining data extraction and transformation processes using Python, regular expressions, and Spark data frames, significantly improving AI processing precision and speed.
* Experienced in development with a focus on Python and Cloud technologies, showcasing proficiency in tools like Spark, Databricks, the Azure suite, Snowflake, Kafka, and Airflow.
* In-depth experience in deploying Microsoft BI and Azure BI solutions, including Azure Data Factory, Azure Databricks, Azure Analysis Services, SQL Server Integration Services, SQL Server Reporting Services, and Tableau.
* Skilled in designing and implementing Azure Cloud Architecture for complex application workloads on MS Azure.
* Served as an Azure Data Engineer, working with Azure Cloud, Azure Data Factory, Azure Data Lake Storage, Azure Synapse Analytics, Azure Analytical services, Azure Cosmos, NO SQL DB, Azure HDInsight, Hadoop, Apache Spark, and Databricks.
* Engineered data pipelines leveraging Snowflake, Kafka, and Airflow for real-time data processing and ingestion.
* Implemented data workflow management using Apache Airflow, focusing on data quality, validation, and error management.
* Leveraged Snowflake to manage semi-structured data formats like JSON and Avro, ingested through Apache Kafka.
* Conducted complex data analysis and reporting using Snowflake's SQL capabilities.
* Utilized Snowflake for large-scale data warehousing and querying purposes.
* Optimized Hive queries by adjusting configuration settings for enhanced performance.
* Specialized in all aspects of Business Intelligence applications, including data extraction, visualization, report creation, infographics, and information visualization.
* Experienced in both Logical and Physical data modeling.
* Implemented Snowflake's security features like row-level security and data masking to uphold data governance and compliance.
* Proficient in developing Azure-based solutions, including Azure Blob Storage, Event Hubs, and Data Lake Analytics.
* Skilled in data transformation, cleansing, quality assurance, and error management.
* Proficient in SQL and Python for data manipulation and analysis, with experience in Spark and PySpark.
* Experience in using data visualization tools like PowerBI, capable of creating interactive dashboards and reports.
* Knowledgeable in data governance and security, with experience in data encryption and access control implementation.
* Capable of working both independently and collaboratively, with strong communication skills and a problem-solving mindset.
* Developed and maintained CI/CD pipelines for data applications using Gitlab and Jenkins.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| *Programming Languages:* | Python, JavaScript, SQL, Scala, C# |
| *Parallel & Syncing cloud outlineComputing:* | Spark, Databricks, Snowflake, Redshift, Kubernetes, Apache Hive, Apache Airflow, ML Flow, Teradata, AWS, Azure, GCP, Kafka, Jenkins |
| *SQL & NoSQL:* | MongoDB, Dynamo DB, PostgreSQL, MySQL, SQL Server, Oracle, FIAS |
| *Data Visualization:* | Tableau, Power BI, Cognos, Business Objects, Hyperion, Web Focus |
| *Tools:* | Lang Chain, AWS Glue, Informatica, Data Stage, APIGEE, Flask, Fast API, AWS API Gateway GITLAB, JIRA, AWS Code build, Splunk, Visio |

**Certifications**

* **Microsoft Certified: Power BI Data Analyst Associate**
* **Microsoft Certified: Azure Data Engineer**
* **Microsoft Certified: Azure AI Fundamentals**
* **University of Dayton: Certification in Autonomous Systems and Data Science**
* **IBM Data Engineering**
* **Databricks Generative AI Fundamentals**

**PROFESSIONAL EXPERIENCE**

Cloud Data Engineer

deutsche bank – NC, USA Dec 2022 to Present

Responsibilities:

* Spearheaded the integration of GENAI with SharePoint on Azure, effectively linking the AI bot to the company's extensive internal database. This database encompassed over 10,000 documents, which were efficiently organized into a scalable data lake in Azure Blob Storage.
* Utilized Python scripting, regular expressions, and Spark data frames to streamline the extraction and structuring of data from complex policy documents. This process transformed the data into a format conducive to AI processing, markedly enhancing GENAI's accuracy in responses.
* Managed the data ingestion process through Azure Data Factory, amalgamating data from diverse sources.
* Applied Azure Databricks for rigorous data processing, using Python in Databricks notebooks to execute Spark jobs, thereby converting raw data into organized formats ready for analysis.
* Executed ETL (Extract, Transform, Load) operations in Azure Databricks, using regular expressions and Spark data frames for parsing and organizing intricate policy documents.
* Created a comprehensive data warehouse utilizing Azure Synapse Analytics, which was effectively integrated with Databricks and Data Factory to facilitate swift data storage and retrieval.
* Enhanced GENAI's integration with the Azure platform, boosting its natural language processing (NLP) capabilities through Azure Cognitive Services, leading to improved accuracy and efficiency in responses.
* Upheld GDPR compliance and ensured data security within the Azure framework, employing encryption, access controls, and routine audits.
* Developed dynamic dashboards and automated reporting mechanisms in Power BI, linked with Azure Synapse Analytics, to deliver immediate insights and analytics.
* Tackled intricate data integration challenges, especially addressing inconsistencies in data formats, using Databricks for smooth integration and processing, which boosted GENAI's reliability and performance.
* Enhanced data retrieval processes using Spark SQL and Databricks' caching features, achieving a 20% increase in GENAI's response efficiency.
* Ensured end-to-end data encryption, both in transit and at rest, leveraging Azure's security tools, to maintain strict adherence to security standards in the GENAI project.
* Devised a dynamic data partitioning strategy in Databricks, improving data storage and access efficiency, leading to reduced query execution times and lower resource usage.
* Continuously updated and fine-tuned the data processing algorithms and AI models in response to changing data trends and business needs, ensuring GENAI's adaptability and relevance in the current market scenario.

**Environment:** Azure Data Factory, Azure Blob Storage, Azure Databricks, Python, Spark, Azure Synapse Analytics, Azure Cognitive Services, Power BI, Azure Stream Analytics, Azure Monitor, Azure DevOps, Machine Learning Algorithms, Third-Party APIs.

Cloud Data Engineer

Rocket mortgage - MI, USA Aug 2021 to Nov 2022

**Responsibilities:**

* Involved in Architecting and developing Azure data factory pipelines by creating datasets, source and destination linked services to move the data from oracle database to Azure Data Lake Store Raw Zone
* Employed Azure Databricks for executing data transformations and facilitating data loading into Azure Data Lake.
* Managed the migration of data from on-premises SQL servers to cloud databases, including Azure Synapse Analytics (DW) and Azure SQL DB.
* Gained substantial experience with Azure BLOB and Data Lake storage, proficient in loading data into Azure SQL Synapse Analytics (DW).
* Crafted and executed a variety of ETL pipelines using Azure Databricks, extracting, transforming, and loading data from diverse sources like flat files, databases, and APIs.
* Specialized in designing and maintaining reports in Power BI, leveraging Azure Synapse/Azure Data Warehouse, Azure Data Lake, and Azure SQL as data sources.
* Implemented Snowflake as a data warehousing solution for the storage and querying of large data volumes.
* Utilized Kafka for managing real-time data streaming, focusing on high volume, high throughput, and low latency data scenarios.
* Engaged with Snowflake's advanced security features, including row-level security and data masking, to uphold data governance and compliance.
* As a Power BI administrator, responsible for creating workspaces and designing report security, including row-level security.
* Involved in user requirement gathering and conducting user training sessions.
* Utilized Azure Cloud Technologies, including Azure Data Lake and Azure SQL Database, for data storage and management.
* Developed and maintained data pipelines to support business intelligence and reporting, integrating with Azure DevOps and using notebooks and clusters.
* Constructed pipelines in Azure Data Factory to run Databricks notebooks, focusing on robust pipeline design, handling unexpected scenarios, creating activity dependencies, and scheduling using triggers.
* Collaborated with diverse teams to understand data requirements, working closely with data analysts and business stakeholders to define project scope and requirements.
* Developed dashboards and reports in Power BI, showcasing interactive visualizations and reports for key stakeholders.
* Integrated Azure Storage into Databricks using service principles and securely managed secrets in Azure Key Vault.
* Worked closely with data analysts and business users to comprehend their data needs and provided technical guidance and support.
* Contributed to data modelling and architectural design discussions, offering insights on data storage and management solutions.
* Performed data quality checks and audits to ensure accuracy and completeness, and established procedures for continuous data monitoring and maintenance.
* Assisted in formulating and implementing data governance and security policies, including data encryption and access controls.
* Engaged in source/version control using tools like Perforce, validating code changes, managing check-ins/outs, versioning, and maintaining CI/CD pipelines for lower environments.

**Environment:** SQL Server, Azure SQL database, Snowflake, Kafka, airflow, Azure Delta, Azure Data warehouse, Azure Synapse Analytics, Azure Data Factory, Databricks, Azure Analysis Services, Azure cloud services, Azure Data Lake, Data Visualization, Agile, Jira, Python, GITLAB, JENKINS

Senior Data Engineer

InstaStores - NY, USA Jan 2020 to July 2021

**Responsibilities:**

* designing and executing comprehensive data solutions within Azure, incorporating expertise in Databricks for data processing, Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics, Azure SQL DW, HDInsight/Databricks, and NoSQL DB.
* Developed Python scripts for processing and transforming data from multiple sources into CSV format.
* Designed and deployed database solutions in Azure SQL Data Warehouse and Azure SQL, focusing on robust and scalable architectures.
* Utilized PySpark for JSON encoding/decoding and managing data frames in Apache Spark through Databricks notebooks.
* Configured Spark Streaming to process continuous data flows from Kafka, storing streamed data in HDFS.
* Conducted advanced statistical data analysis and created data visualizations using Python and R within Databricks notebooks.
* Managed on-premises API services using Azure API Manager, implementing various policies for enhanced service performance.
* Created data models in Splunk, utilizing pivot tables to analyze large datasets, extracting essential information for diverse business needs.
* Engineered and maintained ETL data workflows, integrating new data sources and refining existing workflows using Databricks.
* Constructed pipelines in Azure for transferring both hashed and un-hashed data from Azure Blob to Data Lake utilizing Databricks.
* Transferred and adapted existing application logic into Azure's Data Lake, Data Factory, SQL Database, and SQL Data Warehouse environments.
* Established Azure Repo and Pipelines for efficient CI/CD deployment of various objects.
* Formulated Hive queries within Databricks to assist market analysts in identifying trends, comparing new data against Teradata reference tables and historical metrics.
* Created normalized Logical and Physical database models for the design of an OLTP application.
* Collaborated with senior management to define and plan dashboard goals and objectives, leveraging Databricks for data analysis and visualization.
* Maintained regular communication with management and internal teams, providing updates on project statuses and tasks based on insights derived from data analysis in Databricks.

**Environment:** Azure Databricks, Python, Teradata, Azure SQL Data Warehouse, Azure SQL, Azure Data Platform, Azure Data Lake, Data Factory, Analytics, NoSQL Kafka, HDFS, Spark, Splunk

Data Engineer

Four soft - INDIA Jun 2016 to Aug 2019

**Responsibilities:**

* Built and maintained a wide range of ETL pipelines to extract, transform, and load data from various sources, including flat files, databases, and APIs.
* Developed ETL processes in line with mapping specifications for staging data from diverse sources like CSV, XML, XLSX, etc.
* Employed SQL and Python for effective data manipulation and analysis, leveraging widely used libraries including Spark and PySpark.
* Established data governance and security protocols to maintain data integrity and ensure regulatory compliance.
* Focused on constructing ETL pipelines for S3 parquet files within a data lake using AWS Glue.
* Authored Python-based lambda functions for tailored data transformation and ETL operations.
* Updated and refined data mapping documents to align with actual development practices.
* Utilized AWS Cloud Formation templates to enable consistent code deployment across various environments.
* Engaged with interdisciplinary teams to gauge data requirements and craft appropriate solutions.
* Created and maintained Power BI dashboards and reports for effective visualization and communication of data insights.
* Collaborated closely with data analysts and business users, understanding their data needs and providing necessary technical support and guidance.
* Actively involved in data modeling and architectural design discussions, offering valuable insights and recommendations.
* Crafted both simple and complex SQL scripts for validating Dataflow in multiple applications.
* Executed tasks involving Data Analysis, Data Migration, Data Cleansing, Transformation, Integration, as well as Data Import and Export.
* Developed PL/SQL Stored Procedures, Functions, Triggers, Views, and packages to enhance database functionality.
* Implemented strategies like Indexing, Aggregation, and Materialized Views for query performance optimization.
* Played a key role in developing and documenting the ETL strategy for populating the Data Warehouse from various source systems.
* Constructed logistic regression models using R and Python to forecast subscription response rates based on customer variables such as past transactions, promotional responses, demographics, interests, and hobbies.
* Developed and presented Tableau dashboards/reports for data visualization, Reporting, and Analysis to business stakeholders.
* Designed and developed end-to-end data pipelines for batch processing using Spark with Scala.
* Established Data Connections and published resources on Tableau Server for use with Operational or Monitoring Dashboards.

**Environment:** AWS, Spark, Python, PowerBi, Glue, S3, SQL, ETL, EC2, Cloud formation, Shell Scripting, GITLAB.

SQL, PL/SQL, R programming, Python, ETL (Extract, Transform, Load), Data Warehouse, Tableau, Spark

**Education and Training**

Master of Computer Science **May 2021**

University of Dayton

Bachelor of Technology**,** Electronics and Communication **June 2016**

SRM Institute of Science and Technology