|  |
| --- |
| **Akshita****akkireddy878@gmail.com****Phone: (512) 887-3667LinkedIn: linkedin.com/in/akshita-reddy-493a00299****Data Engineer** |

**Professional Summary:**

* 10+ years of experience in Azure, AWS, and GCP cloud ecosystems with expertise in building scalable data pipelines, hybrid cloud integration, and big data solutions.
* Experience working in reading Continuous Json data from different source system using Kafka into Databricks Delta and processing the files using Apache Structured streaming, Pyspark and creating the files in parquet format.
* Proficient in tools like Azure Data Factory, AWS Glue, GCP Dataflow, and Databricks for batch and streaming data workflows.
* Strong hands-on experience with storage services like Azure Data Lake, Amazon S3, and Google Cloud Storage.
* Well versed experienced in creating pipelines in Azure Cloud ADFv2 using different activities like Move &Transform, Copy, filter, for each, Data bricks etc.
* Providing Azure technical expertise including strategic design and architectural mentorship, assessments, POCs, etc., in support of the overall sales lifecycle or consulting engagement process.
* Hands on experience in Hadoop Ecosystem components such as Hadoop, Spark, HDFS, YARN, TEZ, Hive, Sqoop, MapReduce, Pig, OOZIE, Kafka, Storm, HBASE.
* Worked extensively with Big Data technologies including Hadoop, Spark, Kafka, and Databricks to process and analyze large-scale structured and unstructured datasets.
* Experience in designing and implementing data pipelines, data storage solutions, and data warehousing systems using AWS tools such as S3, RDS, DynamoDB, Redshift, and Athena.
* Experience in implementing data security and privacy policies to ensure the confidentiality and integrity of data using AWS tools such as IAM and VPC.
* Strong understanding of data architecture and design principles and the ability to develop and implement scalable data solutions using AWS services such as EC2, Glue, and Lambda.
* Ability to perform data analytics, predictive modeling, and data-driven decision-making using Azure tools such as HDInsight, Data Factory, and Synapse Analytics.
* Experience in working with data streams and real-time data processing systems using Azure tools such as Event Hubs, Stream Analytics, and Service Bus.
* Skilled in automating data migration processes using Azure Data Factory and scheduling pipelines for timely data updates.
* Strong experience in writing applications using Python using different libraries like **Pandas, NumPy**, SciPy, Matpotlib etc.
* Good understanding of NoSQL databases and hands on work experience in writing applications on NoSQL databases like CosmosDB.
* Much experience in performing Data Modelling by designing Conceptual, Logical data models and translating them to Physical data models for high volume datasets from various sources like Oracle, Teradata, Vertica, and SQL Server by using Erwin tool.
* Expert knowledge and experience in Business Intelligence Data Architecture, Data Management and Modeling to integrate multiple, complex data sources that are transactional and non-transactional, structured, and unstructured.
* Skilled in cloud-native analytics using Synapse, Redshift, and BigQuery along with automation using Terraform and Cloud Composer.Well versed with Relational and Dimensional Modeling techniques like Star, Snowflake Schema, OLTP, OLAP, Normalization, Fact and Dimensional Tables.
* Good knowledge in creating SQL queries, collecting statistics and Teradata SQL query performance tuning techniques and Optimizer/explain plan.

**Technical Skills:**

|  |  |
| --- | --- |
| Azure Cloud Platform | ADFv2, BLOB Storage, ADLS2, Azure SQL DB, SQL server, Azure Synapse, Azure Analytic Services, Data bricks, Mapping Dataflow (MDF), Azure Cosmos DB, Azure Stream Analytics, Azure Event Hub, Azure Machine Learning, App Services, Logic Apps, Event Grid, Service Bus, Azure DevOps, GIT Repository Management, ARM Templates |
| Reporting and BI Tools | Power BI, Tableau and Cognos |
| ETL Tools:  | ADFV2, Informatica Power Center 10.x/9.x, DataStage 11.x/9.x, SSIS, DBT, Apache Airflow, SSIS, AWS Glue, GCP Dataflow, Cloud Composer |
| Programming Languages | PySpark, Python, U-SQL, T-SQL, LINUX Shell Scripting, AZURE PowerShell, C#, Java |
| Big data Technologies | Hadoop, HDFS, Hive, Apache Spark, Apache Kafka, Pig, Zookeeper, Sqoop, Oozie, HBASE, YARN |
| Databases | Azure SQL Warehouse, Redshift, BigQuery, Cloud SQL, Azure SQL DB, Azure Cosmos No SQL DB, Oracle, Microsoft SQL Server, SQL MI |
| IDE and Tools | Code, Eclipse, SSMS, Maven, SBT, MS-Project, GitHub, Microsoft Visual Studio |
| Cloud Stack | AWS, GCP, Azure, Snowflake |
| Methodologies | Waterfall, Agile/Scrum, SDLC |

**Professional Experience:**

**Client: Mercer - Houston Tx Aug 2021 to Present**

**Role: Cloud Data Engineer**

**Responsibilities:**

* Collaborated with Business Analysts and Solution Architects to gather client requirements and translated them into Azure-based design architectures.
* Designed and maintained data pipelines using Azure Data Factory, achieving reduction in data processing time.
* Created High-Level Technical Design and Application Design documents, ensuring clear and comprehensive documentation for stakeholders.
* Engineered and optimized data pipelines leveraging the Medallion architecture, meticulously structuring data into Bronze, Silver, and Gold layers to ensure superior data management, enhanced accessibility, and streamlined analytics workflows.
* Implemented a 100% Reusable (plug & play) Python Pattern (Synapse Integration, Aggregations, ChangeDataCapture, Deduplication) and High Watermark Implementation. This process will accelerate the development time and standardization across teams in the Confidential project.
* Wrote Terraform code to automatically create Azure resources, enhancing infrastructure efficiency and reducing manual setup time.
* Successfully integrated Snowflake with Azure Databricks, enhancing big data processing capabilities and reducing processing time by 40%.
* Built automated data ingestion pipelines from AWS S3 to GCP Big Query using Glue and Dataflow for multi-cloud data warehousing.
* Successfully migrated data management operations from Hive Metastore to Unity Catalog within the Databricks environment. Leveraged Unity Catalog for centralized governance, comprehensive data lineage tracking, and managed identity.
* Built scalable ingestion workflows using AWS Glue and EMR to pull data from HDFS and API sources, storing transformed outputs in S3 and Redshift
* Wrote SQL stored procedures to enhance the performance of Azure Data Factory (ADF) pipeline runs, ensuring efficient data processing and reduced execution times.
* Worked with Azure Synapse Analytics to design and implement robust data warehousing solutions, supporting enterprise-level analytical needs.
* Engineered complex data transformations and manipulations using Azure Data Factory and PySpark with Databricks.
* Seamlessly integrated Azure Data Factory with Azure Logic Apps and Azure Functions to automate workflows and streamline data processes.
* Designed and implemented optimized ETL pipelines using Spark (Python) and Snowflake, improving data processing efficiency and reliability.
* Developed and managed Azure SQL Database and Blob Storage for robust data storage solutions.
* Managed end-to-end data workflows, including Data Transformation activities using Stored Procedures and Azure Functions.
* Successfully managed multiple projects concurrently, collaborating with on/offshore extended teams to meet deadlines and deliverables.
* Implemented real-time data streaming with Azure Stream Analytics and Event Hub, enabling timely analytics for critical operations.
* Wrote complex SQL queries, stored procedures, and performed performance tuning across Azure SQL, Snowflake, and SQL Server to support efficient data transformations, reporting layers, and real-time analytics.
* Developed and optimized Synapse SQL Pool for efficient querying and reporting, enhancing data retrieval speeds by 40%.
* Integrated GCP Big Query with Azure Data Factory for cross-cloud analytics; used GCP Dataflow for data transformation and Cloud Composer for orchestration.
* Leveraged U-SQL scripts to ingest and transform data into Azure Data Warehouse.
* Utilized DBT for transforming raw data into analytics-ready datasets within Azure Synapse and Snowflake environments, supporting standardized data modeling and transformation processes for enterprise data warehouses
* Managed data ingestion using Azure Data Lake Storage and created pipelines with Azure Data Factory v2, extracting data from diverse sources.
* Developed custom solutions and extensions using .NET with C# to meet specific client requirements.
* Collaborated with data science teams to support deployment of ML models by building scalable data pipelines using Azure Data Factory and Databricks, ensuring seamless flow of structured and unstructured data for model training and inference.
* Created numerous Databricks Spark jobs with PySpark for various data operations.
* Implemented a Power BI integration module for canned reports from Azure Data Lake Gen2.
* Proficiently used SQL Server Import and Export Data tool.
* Familiar with YAML for configuring Azure DevOps pipelines and infrastructure scripts; working knowledge of Data API patterns and deployment workflows.
* Designed and optimized Snowflake-based data pipelines for large-scale data processing, including data ingestion, transformation, and integration with Databricks and ADF
* Collaborated with stakeholders to gather data requirements and optimize SQL queries, resulting in improvement in query performance.
* Tuned Snowflake queries for performance and supported real-time analytics.

**Environment:** Azure Cloud, Azure Databricks, Azure Data Factory (ADF v2), Azure functions Apps, Azure Data Lake, Blob Storage, SQL server, Windows remote desktop, Unix, Azure PowerShell, Data bricks, Python, Pyspark Azure Cosmos DB, Azure Stream Analytics, Azure Event Hub, Power BI.

**Client: Extreme Networks - Morrisville, NC Nov 2019 to July 2021**

**Role: Sr. Data Engineer**

**Responsibilities:**

* Ensured Wells Fargo's customer eligibility system complied with Securities and Exchange Commission (SEC) regulations for data collection, storage, and usage within the Azure environment.
* Spearheaded data privacy and security measures, maintaining full compliance with SEC guidelines for handling securities-related data.
* Developed and implemented robust data governance policies and procedures within the Azure ecosystem, improving data quality, accuracy, and reliability.
* Established a culture of data stewardship, ensuring data usage remained aligned with SEC regulations.
* Created and managed Azure Synapse Pipelines to orchestrate end-to-end data workflows, ensuring seamless data integration and processing.
* Integrated diverse data sources, including on-premises databases, cloud storage, and third-party APIs, into ADF pipelines for centralized data processing.
* Led the development and implementation of a transformative customer eligibility project within the Azure platform.
* Proficient in SQL databases and queries, with knowledge of Change Data Capture (CDC) & Change Tracking (CT) inner workings, adeptly applied in Databricks environments for comprehensive data management and analytics.
* Enabled real-time data replication from AWS Kinesis and Azure Event Hub to GCP BigQuery using Kafka and Pub/Sub for unified analytics reporting.
* Designed and deployed serverless workflows using GCP Cloud Functions and AWS Lambda for event-driven data processing.
* Familiarity with SQL Server Reporting Services, Integration Services, and other relevant tools and technologies.
* Hands-on experience with NoSQL databases such as Cosmos DB, HBase, and DynamoDB for managing and processing semi-structured and unstructured data.
* Integrated streaming data from Kafka and Event Hub into data lakes and ML pipelines, enabling near real-time insights for fraud detection and customer behavior analysis.
* Loaded transformed data into target systems, including databases, data warehouses, and cloud storage, employing Python libraries and custom scripts.
* Designed and supported enterprise data warehouse solutions using Azure Synapse, Snowflake, and Databricks by implementing Medallion architecture and managing end-to-end data flows.
* Utilized cutting-edge data analytics techniques to analyze customers' financial and personal data, revolutionizing the process of determining eligibility for auto financing.
* Designed and implemented an efficient Extract, Transform, Load (ETL) architecture using Azure services for seamless data transfer from source servers to the Data Warehouse.
* Implemented automated data cleansing and integration processes, resulting in improved data quality.
* Implemented real-time event processing of data from multiple servers within the Azure environment, enabling rapid response to critical data events.
* Actively participated in designing and developing CI/CD pipelines for data engineering within the Azure ecosystem.
* Implemented automation from code commit to deployment using Azure DevOps.
* Managed a cloud data warehouse on Azure, facilitating batch processing and streaming.
* Enhanced data visualization for customer data using Azure Power BI.

**Environment:** Azure Data factory, Azure Databricks, Azure Event Hubs, Azure SQL Datawarehouse, Power BI.

**Client: Repco - Hyderabad, INDIA Aug 2017 to Feb 2019**

**Data Engineer**

**Responsibilities:**

* Demonstrated expertise in developing and deploying custom Hadoop applications within the AWS environment, ensuring seamless integration and performance optimization.
* Designed and implemented data architectures for managing large volumes of home loans data while adhering to GDPR and CCPA data privacy and security regulations.
* Developed and automated multiple ETL jobs using Amazon EMR, facilitating seamless data transfer from HDFS to S3.
* Created batch data pipelines for extracting data from S3 and loading it into RedShift using Glue jobs.
* experience in data ingestion, transformation, governance, and real-time processing across Azure, AWS, and on-prem platforms, ensuring data quality and compliance.
* Provided expert support in solving real business issues by leveraging knowledge of Hadoop distributed file systems and open-source frameworks, driving operational excellence and efficiency.
* Spearheaded design, development, and maintenance of dynamic data pipelines on Snowflake, ensuring seamless integration and analytics support.
* Configured Spark streaming to store and process real-time data from Kafka.
* Leveraged AWS EMR to store structured data in Hive and unstructured data in HBase.
* Cleaned and transformed data in HDFS using MapReduce (YARN) programs for ingestion into Hive schemas.
* Applied advanced SQL skills to fine-tune Snowflake performance and enhance query efficiency, bolstering data integrity and security across the lifecycle.
* Worked on ETL pipeline development using Informatica PowerCenter, building robust data integration workflows and managing data movement between on-prem and cloud platforms.
* Built Hadoop applications using HDFS, Hive, Pig, HBase, and MapReduce in AWS EMR; managed unstructured data with NoSQL databases like Cosmos DB and HBase.
* Created a data lake in Snowflake using Stitch, App Testing, and Production support.
* Managed S3 buckets, implemented policies, and utilized S3 and Glacier for storage and backup on AWS.
* Generated reports for the BI team by exporting analyzed data to relational databases for visualization using Sqoop.
* Created custom User Defined Functions (UDFs) to extend Hive and Pig core functionality.
* Enabled ODBC/JDBC data connectivity to Hive tables and worked with tools like Tableau and Flink

**Environment:** AWS S3, Glue, AWS EMR, Glacier, Redshift, Snowflake, Spark SQL, Sqoop, Flink, YARN, Kafka, MapReduce, Hadoop, HDFS, Hive, Tableau, Spotfire, HBase.

**Client: Colruyt Group– Hyderabad, INDIA June 2015 to July 2017**

**Data Engineer**

**Responsibilities:**

* Examined claims and supporting documentation to ensure policy compliance before processing.
* Developed a strong understanding of claim processing from both client and service provider perspectives, identifying key metrics for each.
* Managed policy servicing and maintenance operations, including coverage changes, beneficiary data updates, and premium payments.
* Processed claims data efficiently through the system.
* Possess a good understanding of Electronic Health Record (EHR) systems, including their functionalities, data models, data elements, and data privacy and security regulations.
* Designed high-performance batch ETL pipelines using AWS cloud services.
* Extracted data from relational databases and APIs with AWS Data Factory to store in AWS data lake storage.
* Developed and maintained ETL jobs using IBM DataStage for large volume batch processing and real-time data feeds, supporting enterprise reporting and data warehousing projects.
* Utilized PySpark scripts in AWS Databricks for data transformations and conversions.
* Designed data warehousing solutions using AWS Synapse Analytics for storing and analyzing transformed data.
* Implemented and designed Python microservices in the healthcare domain.
* Monitored productivity and resources using AWS Log Analytics.
* Implemented CI/CD pipelines with AWS DevOps for automated build, test, and deployment processes.
* Monitored and optimized cross-cloud data pipelines using AWS CloudWatch, Azure Monitor, and GCP Operations Suite (formerly Stackdriver).
* Utilized AWS Event Hub to capture real-time data streams and route them to the appropriate data stores.
* Monitored data pipeline performance using AWS Monitoring and Analytics tools to ensure seamless data flow and identify potential bottlenecks.
* Played a critical role in a data migration project involving EHR, ensuring accurate, efficient, and secure data migration.
* Ensured data pipeline security using AWS security features, including role-based access control and encryption, to safeguard data privacy and confidentiality.
* Managed encryption keys and passwords through AWS Key Vault.
* Utilized AWS Logic Apps to orchestrate complex business processes and workflows.
* Implemented serverless computing solutions using AWS Lambda Functions for cost-effective and scalable data processing.
* Designed visualization dashboards for data analytics using Power BI.
* Proficient in practicing Agile methodology to update workflows and manage project lifecycles and sprints.

**Environment:** AWS Data Factory, Data Lake storage, Synapse Analytics, Python, Event Hub, Logic Apps, Key Vault, Log Analytics, Scala, Power BI.

**Education:**

**Jawaharlal Nehru Technological University (JNTU), India 2011 to 2015**Bachelor of Technology in Computer Science