NIHASH REDDY AKAVARAM | Data Engineer

akavaramnihashreddy@gmail.com | +1 (940) 843-1520 | www.linkedin.com/in/nihashreddy

Professional Summary

- Extensive experience in Data Engineering field including Ingestion, Data Lake, Datawarehouse, Reporting and Analytics.
- Solid Experience and understanding of Implementing large scale Data warehousing Programs and E2E Data Integration Solutions on Snowflake Cloud, AWS and Azure
- Proficient in designing and implementing data warehousing solutions on Snowflake, including data modeling, ETL/ELT processes, and performance optimization.
- Knowledge and experience on AWS services like Redshift, Redshift spectrum, S3, Glue, Athena, Lambda, cloudwatch and EMRs like HIVE, Hadoop and spark
- Hands on Experience on Python programming for data processing and to handle Data integration between On-prem and Cloud DB or Datawarehouse.
- Engineering Spark applications using PySpark and Spark-SQL for data extraction, transformation, and aggregation and PySpark for custom business logic, and implementing performance tuning strategies.
- Worked on Dimensional Data modelling in Star and Snowflake schemas and Slowly Changing Dimensions (SCD).
- Proficient in stream processing with tools like Apache Kafka, Spark Streaming, and Flink. Utilizes AWS Kinesis, Azure Stream Analytics, and GCP Cloud Pub/Sub for efficient cloud-based data streaming and analysis.
- Scheduling and orchestrating data workflows using tools like Apache Airflow, cron, and cloud-native services like AWS Step Functions and Azure Logic Apps for efficient automation.
- Proficient in managing SQL databases such as PostgreSQL, MySQL, and SQL Server, as well as NoSQL databases like MongoDB and Cassandra, ensuring efficient and scalable data storage solutions.
- Collaborating with cross-functional teams to analyze requirements and design data solutions, working across multiple accounts and projects for seamless integration and data sharing.
- Creating interactive dashboards and visualizations with tools like Tableau, Qlikview, Amazon QuickSight, and Power BI to support data-driven decision-making processes.
- Implementing encryption protocols, access controls, and data masking techniques to protect sensitive data and ensure compliance with regulations.

Education

MASTERS IN INFORMATION SYSTEMS AND TECHNOLOGIES | University of North Texas, Denton, TX | May 2024
BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE | Jawaharlal Nehru Technological University, India | May 2020

Professional Experience

University of North Texas, Denton, TX | Aug 2022 – May 2024
Teaching Assistant | Enterprise Data Warehousing

- Conducted regular office hours to provide one-on-one assistance, guidance, and clarification on content, assignments, and projects.
- Evaluated student performance through grading assignments, exams, and project submissions, providing constructive feedback to foster learning and improvement.
- Assisted students from diverse backgrounds in understanding complex data warehousing concepts, leading to higher student engagement and improved academic performance.

Capgemini Technology Services India Limited | Nov 2021 – Aug 2022
Data Engineer | Associate Consultant

- Architected and executed end-to-end data solutions on the Azure platform, leveraging Azure Databricks, Spark, and Azure Data Lake Storage (ADLS), which streamlined data workflows and improved data accessibility.
- Implemented Delta Lake and Delta Live Tables for reliable and consistent data pipelines, utilizing Databricks workflows to automate ETL processes, monitor task execution, and ensure data integrity and ACID compliance.

- Developed and exposed RESTful APIs to facilitate seamless data integration and retrieval for various applications and services, enhancing data accessibility and interoperability.
- Utilized optimization techniques in Spark, such as partitioning and caching, and stored intermediate data in Parquet format with Snappy compression, reducing latency, minimizing storage costs, and optimizing query performance.
- Integrated Power BI with Azure Databricks for data visualization, reducing report generation time by 40%, and collaborated with cross-functional teams to implement SQL Server-based data models using Azure Synapse Analytics for efficient data warehousing.
- Utilized Python and SQL for robust data processing solutions, including developing web applications using Python Django to enable efficient data interaction and management, and leveraged HDFS and ADLS GEN-2 for distributed storage, improving fault tolerance and data durability.
- Implemented data warehousing solutions using Azure Synapse Analytics, enhancing data consolidation, management, and reporting capabilities.
- Conducted ad-hoc analysis and offloaded repeatable processes by working closely with data analysts to validate data processes and results, ensuring accurate and timely insights for business stakeholders.

Capgemini Technology Services India Limited | Jan 2020 – Oct 2021
Data Engineer | Senior Software Engineer
- Architected and maintained a scalable data lake on AWS, utilizing Amazon S3 for storage, AWS Glue Data Catalog for metadata management, and Amazon Athena for interactive queries.
- Implemented Amazon Redshift for data warehousing and optimized query performance.
- Designed and developed ETL pipelines using AWS Glue and PySpark, transforming data across raw, conformed, and purpose-built zones.
- Implemented data masking for PII and optimized storage by converting CSV to Parquet format.
- Integrated Snowflake as a complementary data warehouse solution, leveraging Snowpipe for automatic and real-time data ingestion from AWS S3.
- Utilized Snowflake's time travel and zero-copy cloning features for efficient data management.
- Implemented cross-platform data sharing between Snowflake and Redshift using AWS Data Exchange, enabling secure and governed data collaboration across teams and external partners.
- Orchestrated workflows using AWS Step Functions and Lambda, with AWS AppFlow for SaaS data integration. Set up comprehensive monitoring and alerting using AWS CloudWatch and SNS.
- Employed AWS CDK Pipelines for CI/CD of data lake infrastructure and ETL jobs across multiple environments, ensuring consistent and repeatable deployments.
- Developed real-time data processing pipelines using Amazon Kinesis Data Streams and Kinesis Data Firehose, integrating with various data sources including APIs, MySQL databases, and Salesforce.
- Created interactive dashboards and reports using Tableau, connecting directly to Snowflake and Redshift for real-time data visualization and analysis. Optimized data models and implemented materialized views for improved query performance.
- Utilized Python and SQL for data manipulation and analysis.

Capgemini Technology Services India Limited | Aug 2019 – Dec 2019
Data Engineer
- Utilized Apache Sqoop to import customer transaction data from MySQL into Hadoop Distributed File System (HDFS) for scalable storage and efficient processing.
- Employed Apache Pig scripts to process and transform raw transaction data into structured formats, enabling easy querying and analysis.

- Designed and implemented Hive tables to store processed data, facilitating efficient querying and analytics through HiveQL.
- Leveraged Apache Spark to conduct in-memory data processing, significantly speeding up analysis and enabling real-time insights into customer behavior.
- Created interactive dashboards using Power BI to visualize key metrics and insights, such as customer purchase frequency and product preferences, providing actionable insights to business stakeholders.

Projects
Cloud Data Analytics: Snowflake, GitHub, DBT | GitHub URL
- Established Snowflake worksheet, loaded sample data, and connected dbt Cloud to Snowflake for seamless analytics.
- Transformed queries into dbt models, added and documented sources, ensuring data accuracy and clarity.
- Implemented generic and singular tests for models and sources, enhancing data quality.
- Employed GitHub for version control, fostering collaborative development.
- Automated jobs to run dbt models, optimizing data transformation processes for timely updates.

Certifications
- Snowflake Hands-On Essentials: Data Warehousing Workshop
- Snowflake Hands-On Essentials: Data Lake Warehouse
- Snowflake Hands-On Essentials: Data Engineering Warehouse
- Dbt Fundamentals

Technical Skills

Cloud Services	Amazon Web Services: S3 (Simple Storage Service), Glue, EMR, Redshift, AWS Data Pipeline, Kinesis (Streams, Firehose, Analytics), RDS, Glue DataBrew, Athena, Step Functions, DynamoDB, Lambda, Quicksight, Glue Elastic Views, CloudWatch Microsoft Azure: Data Factory, Databricks, Synapse Analytics, HDInsight, Stream Analytics, Data Lake Storage, SQL Database, Data Explorer, Data Catalog, Logic Apps. Google Cloud Platform: Cloud Storage, Cloud Dataflow, BigQuery, Cloud Dataprep, Cloud Dataproc, Cloud Composer, Cloud Pub/Sub, Cloud Bigtable, Cloud SQL, Cloud Spanner.
Data Integration and Transformation Tools	Talend, DBT (Data Build Tool)
Data Warehousing and Big Data	Snowflake, Databricks, Apache Spark, Kafka, Spark Streaming, Apache Airflow, Hadoop, Hive, Map Reduce, HUE, Sqoop
Programming Languages	Python, Pyspark, SQL, PL/SQL, shell Scripting, C, Java, Scala
Database Technologies	SQL Databases (PostgreSQL, MySQL, SQL Server), NoSQL Databases (MongoDB, Cassandra),
Data Visualization tools	PowerBI, Tableau, Qlikview, QuickSight
Other	Docker, Github, Jenkins
Operating System	Windows, Linux, Unix