

NIHASH REDDY AKAVARAM | Data Engineer

+1 (682) 294-0711 | 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

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE | Jawaharlal Nehru Technological University, India

Professional Experience

University of North Texas, Denton, TX | Aug 2022 – May 2024

Teaching Assistant

- 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.
- Demonstrated strong communication and interpersonal skills in assisting students from diverse backgrounds in understanding complex data warehousing concepts.

Capgemini Technology Services India Limited | Nov 2021 – Aug 2022

Data Engineer | Associate Consultant

- Architected and executed end-to-end data solutions on Azure platform, leveraging Azure Databricks, Spark, and Azure Data Lake Storage (ADLS).
- Designed and implemented a job orchestration framework in Azure Databricks, optimizing Spark jobs for efficient data processing.

- Implemented Medallion Architecture for data ingestion, utilizing Azure Data Factory and Sqoop.
- Utilized optimization techniques in Spark, such as partitioning and caching, to enhance processing speed.
- Stored intermediate data in Parquet format with Snappy compression, minimizing storage costs and optimizing query performance.
- Ensured data integrity and reliability by leveraging Delta Lake as the final data format, enabling real-time analytics and achieving ACID compliance.
- Integrated Power BI with Azure Databricks for data visualization, reducing report generation time by 40% and enhancing decision-making agility.
- Collaborated with cross-functional teams to implement SQL Server-based data models using Azure Synapse Analytics for efficient data warehousing.
- Leveraged HDFS and ADLS GEN-2 for distributed storage and processing, implementing data replication strategies that improved fault tolerance and data durability.

Capgemini Technology Services India Limited | Jan 2020 – Oct 2021

Data Engineer | Senior Software Engineer

- Spearheaded the design and implementation of end-to-end data pipelines, ensuring seamless data ingestion, storage, and processing using AWS serverless technologies.
- Pioneered a serverless data lake architecture, enabling agile and self-service data analytics across the organization, significantly reducing operational overhead
- Orchestrated data ingestion from diverse sources, including batch and streaming data, into the data lake using AWS services like AWS DMS, Kinesis Data Firehose, and AWS AppFlow.
- Managed durable, scalable, and secure data storage solutions, utilizing Amazon S3 to handle petabytes of structured and unstructured data.
- Advanced the consumption readiness of datasets through validation, cleanup, normalization, and transformation, employing AWS Glue and Step Functions for efficient data processing workflows.
- Facilitated various analytics methods, including BI dashboarding, interactive SQL, and machine learning, by integrating with the consumption layer of the data lake.

Capgemini Technology Services India Limited | Aug 2019 – Dec 2019

Data Engineer | Intern

- 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

- Hands-On Essentials: Data Warehousing Workshop
- Hands-On Essentials: Data Lake Warehouse
- 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