

MANASA NAMIREDDY

M: [+1-331 226-7959](tel:+13312267959) | E: namireddy.manasa@gmail.com | L: TX | Ln: <http://www.linkedin.com/in/manasanamireddy>

PROFESSIONAL SUMMARY:

- Experienced data engineer with over 4 years of expertise orchestrating end-to-end data pipelines.
- Proficient in **Python, Scala, Java, and SQL**, with expertise in building robust ETL processes and data pipelines. Proficiency in **Data Extraction, Ingestion, and Transformation**.
- Significant proficiency in **data engineering**, specializing in the creation of **ETL pipelines** for **batch** and **streaming data**, utilizing **PySpark**, and **SparkSQL**.
- Skilled in designing efficient **NoSQL storage solutions** using **Cassandra** and **MongoDB**. Experienced in building **real-time datastreaming pipelines** with **Apache Kafka** for advanced analytics.
- Data integration platforms such as **Apache Nifi**, **Talend**, **Informatica** and **Microsoft SQL Server Integration Services (SSIS)**.
- Proficient in leveraging **Hadoop technologies** for **distributed storage** and processing of large datasets, employing frameworks such as **HDFS**, **MapReduce**, **Hive**, and **Spark** for advanced analytics.
- Leveraged **Snowflake's** integration with BI tools like **Tableau** and **Power BI**.
- Engineered **ETL workflows**, extracting diverse data from multiple sources, refining it for analysis, and coordinating workflows **Apache Airflow** and **AWS Glue**.
- Experienced in utilizing **AWS cloud technologies**, including **EC2**, **EMR**, **S3**, **Lambda**, **SQS**, and **SNS for Data storage**. Leveraged AWS services like **S3** for data storage, **EC2** for computation, and **RDS** for hosting databases.
- Employed Azure-cloud platforms, with **ADFs**, **Log-Analytics**, and **Databricks APIs** to establish pipeline configurations.
- Strong expertise in **database management**, **data integration**, **data analysis**, **data ingestion**, **data modeling**, and **data pipeline orchestration**, coupled with a deep understanding of **data quality** and **governance** principles and Solid understanding of **data modeling** and **warehousing** concepts, crucial for designing and implementing effective **data solutions**.

TECHNICAL SKILLS:

- **Programming Languages:** Python, SQL, Java, C/C++, Scala
- **Data warehousing platforms:** Amazon RedShift, Google Big Query, Snowflake, Azure Synapse Analytics
- **Big Data Technologies:** Apache Hadoop, Hive, Apache Spark, PySpark, Apache Kafka, Metastore, Presto, Flume, ClickHouse, Flink
- **Databases:** PostgreSQL, SQL, MySQL, Oracle, DB Cassandra
- **ETL Tools:** Informatica, Talend, Apache NiFi, SQL Server Integration Services (SSIS)
- **Data visualization:** Power BI, Tableau
- **Cloud Platforms:**
 - **Amazon Web Services:** S3, RDS, DMS, EMR, EC2, AWS Glue
 - **Azure Cloud Platform:** Azure Data Lake, Azure Data Factory, Database, SQL Server, Azure Cloud Shell
- **Data Pipeline Orchestration:** Apache Airflow, AWS Step Functions
- **Scripting:** Bash Scripting, Shell scripting, PowerShell Scripting
- **Version Controls:** Git, Big Bucket
- **Other tools:** Linux, Windows, MS Office & Excel, Salesforce, Terraform, Agile Methodologies, Business Initiatives business intelligence.

EDUCATION DETAILS:

Master's in information technology, University of the Cumberlands, **USA**

May 2024

Bachelor's in computer science and engineering, Jawaharlal Nehru Technological University, **India**

April 2021

PROFESSIONAL EXPERIENCE:

LIBERTY MUTUAL, USA

APR 2023-APR2024

ROLE: DATA ENGINEER

- Extracted data from **SQL Server** and transferred it to **Hadoop**. Utilized **Hive**, **Pig Latin** to retrieve data, and **Apache Spark** for real-time/batch processing, reducing time by 25%.
- Automated **Python** and **Spark** script execution for scheduled data filtering, cleansing, mapping, and aggregation. Developed **Python** jobs and notebooks managed by **Apache Airflow** to automate and monitor **Azure Databricks** workflows.
- Employed **Azure**-cloud platforms, with **ADFs**, **Log-Analytics**, and **Databricks APIs** to establish pipeline configurations. Built and productized **machine learning** models with **Azure Databricks** using **Python** and common data science libraries like **NumPy**, **Pandas**, and **sci-kit-learn**.
- Designed and constructed a multi-terabyte **Azure Synapse Analytics** data warehouse for large-scale ingestion and transformation of millions of records (**ELT**). Developed robust, scalable data pipelines using **Azure Apache Beam**, **Azure Data Factory**, and **Azure Synapse Analytics** for real-time analytics.
- Acquired hands-on experience with **Snowflake** utilities, **Snow SQL**, **Snow Pipe**, and **Python-based Big Data** model techniques. Formulated **data acquisition strategies**, persuading stakeholders.
- Designed and conceptualized architecture, and performed analysis, **ETL operations**, and visualization of large datasets extracted via **RESTful API** web services. Worked hands-on with **SQL SSIS**, **Python**, and **SSMS**.
- Collaborated on metrics with **Analytics** and **BI** teams, reducing manual analysis. Created impactful **Power BI** dashboards. Designed and implemented reliable, scalable, robust, and extensible **big data systems**. Monitored systems and made trade-offs.
- Demonstrated track record in maintaining **legacy platforms** while formulating **data acquisition strategies**, and persuading stakeholders by elucidating **policies**.

ENVIRONMENT: Hadoop, Hive, SQL, Excel, Python, Azure Cloud Platform (Azure Data Lake, Azure Data Factory, Database, SQL Server, Azure HDInsight, Azure Cloud Shell), Snow SQL, SnowPipe, SQL SSIS, SSMS, Power BI.

C L INFOTECH PRIVATE LIMITED

APR 2023-APR2024

ROLE: DATA ENGINEER

- Prepared **Python scripts** for automated data ingestion from **API**, **AWS S3**, **PostgreSQL**, **Teradata**, and **Snowflake**.
- Developed and documented **ETL (Extract, Transform, and Load)** strategy for **Data Warehouse** population from various source systems.
- Developed data pipeline programs with **Spark Scala APIs**, data aggregations with **Hive**, and formatting data (**JSON**) for visualization, and generating.
- Engineered and maintained **Hadoop-based data lakes**, ensuring reliable and scalable storage. Utilized **MapReduce** for distributed processing, enabling efficient analysis of large-scale datasets.
- Implemented **Hive** for data warehousing, facilitating structured querying and reporting on **Hadoop** data. Architected end-to-end data processing pipelines using Hadoop components (**HDFS**, **MapReduce**, **Hive**) to efficiently handle large datasets.

- Deployed and maintained **AWS data solutions**, leveraging **S3**, **Redshift**, and **Lambda**. Utilized **Amazon S3** as a data lake to store raw data in various formats.
- Designed **ETL processes** in **AWS Glue** to migrate data from external sources (**S3**, **ORC/Parquet/Text Files**) into **AWS Redshift**.
- Built an **ETL framework** for migrating data from on-premises sources (**Hadoop**, **Oracle**) to **AWS** using **Apache Airflow**, **Sqoop**, and **Spark**.
- Scheduled **AWS Glue ETL jobs** to extract data from **S3** into **Redshift** data marts. Engineered **Amazon Redshift** as a data warehousing solution in the **AWS ETL** pipeline.
- Implemented **AWS CloudWatch** for monitoring, ensuring timely detection. Utilized **Apache Spark** and **Kafka** for processing large-scale datasets, enhancing processing speed.
- Created on-demand tables on **S3** with **Lambda** and **Glue** using **Python** and **PySpark**. Orchestrated workflows with **Apache Airflow**, optimizing task scheduling for efficiency.
- Designed and created insightful visualization dashboards in **Tableau**, translating complex data into actionable insights.

ENVIRONMENT: AWS (S3, RDS, Redshift, DMS, Lambda, EMR, EC2, Glue, CloudWatch), Airflow, HDFS, Hive, PySpark, SQL, Python, Git, Tableau, Advance Excel, Airflow, Oracle), Snowflake.

CERTIFICATIONS:

- **Microsoft Certified [Dp-203]:** [Associate Data Azure](#)
- **IBM Developer Skill Networks Certified:** [Python for Data Science](#)