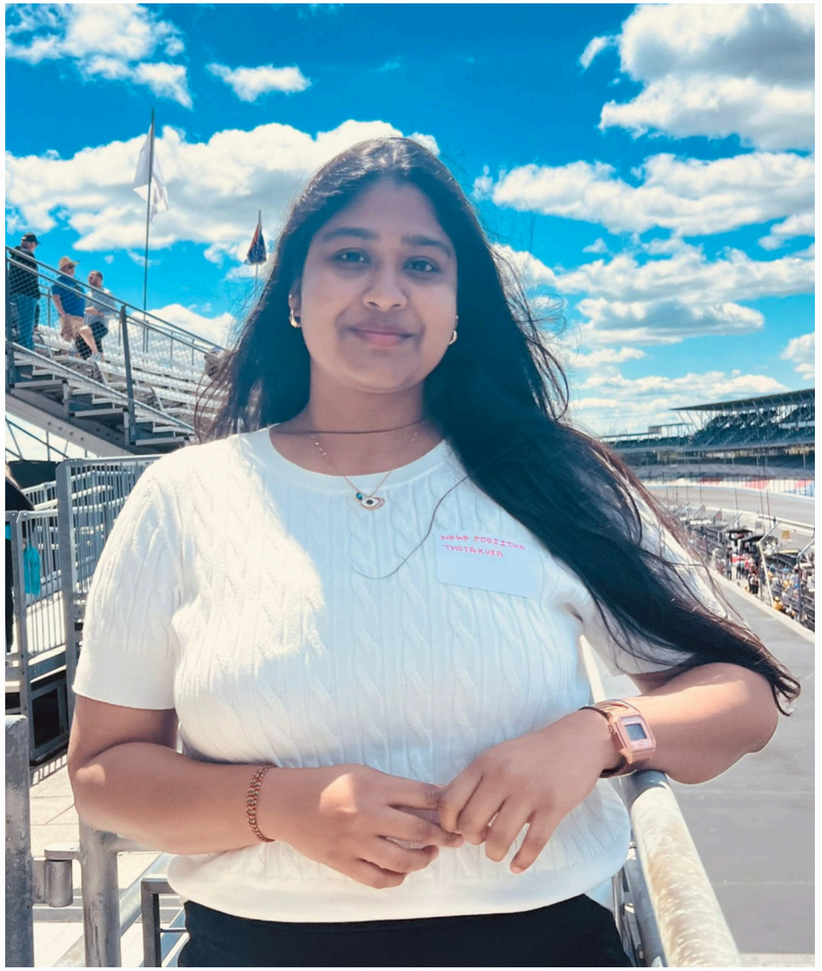
Naga Poojitha

Data engineer

Skilled and experienced Data Engineer with over 5+ years of expertise in designing and implementing robust data pipelines, scalable big data applications, and cloud-based solutions. Proficient in leveraging technologies like AWS (Glue, Redshift, S3, EMR), Azure (Data Factory, Synapse, Databricks), and Hadoop Ecosystem (PySpark, Hive, HDFS) for data processing, migration, and analytics.



Skill Set:

Programming Languages:
Python, PySpark,Scala, SQL,Linux

Databases:
Snowflake, MS SQL server, Oracle,Teradata, HBase, MySQL

Monitoring & Reporting
Tableau, Power BI, Quick Sight

IDE Tools:
Eclipse, Jupyter, Anaconda

AWS :
S3, EC2, EMR, Lambda, Glue, Kinesis, Athena, DynamoDB, C Auto Scaling, Redshift, Cloud Watch, RDS, SQS, SNS, Step functions

Azure:
Data Lake, Data Lake Analytics, Data Bricks, Data Factory, Synapse, Azure SQL, Cosmos DB, HD Insight, Azure DevOps (CI/ CD)

Big Data Tools:
Hive, Spark, Pyspark, Map Reduce,

Experience:

AWS Data Engineer

Pearson(Aug 2023 - Present)

- Used various AWS services including S3, EC2, AWS Glue, Athena, Redshift, EMR, SNS, SQS, DMS, Kinesis.
- Developed multiple Spark Streaming and batch processing jobs using Python on AWS EMR.
- Developed Data Models for RDBMS,NoSQL, and Cloud based data warehousing services like Redshift.
- Integrated Lambda functions with various AWS services, such as API Gateway, S3, DynamoDB, and SQS, to build scalable and resilient applications.
- Developed Airflow DAGs for Batch processing to coordinate Python Data pipelines for pre- ingestion preparation of CSV files, using configuration to parameterize many input files from various sources launching distinct Task instances.

Data Engineer

Ascension Health (Dec 2021 - July 2023)

- Worked on integration of Azure Data Factory and other Azure services enables to build modern data warehouse and real-time analytics solutions.
- Data was gathered, converted, and loaded from source systems to Azure Data Storage services using a combination of AzureData Factory, T-SQL, Spark SQL, and U-SQLAzure Data Lake Analytics.
- Build, deploy, and maintain ETL workflows using SQL, Unix shell script programming and the Informatica Big Data Management Platform.
- Created tables in Azure SQL Data Warehouse for the reporting and visualization of business- related data.
- Created Spark applications utilizing PySpark and Spark SQL in Databricks to modify and aggregate source data before importing it into Azure Synapse Analytics for reporting.

Hadoop Developer

EH Note (Dec 2019 - Jan 2021)

- Worked on migration projectf or converting SQL Server, SSIS applications into Hadoop framework using Python, Pyspark, Hive, Sqoop, Shell.
- Knowledge of monitoring, managing,and reviewing Hadoopclusters using Cloudera Manager.
- Developed database designin HDFS using HBase Architect design.
- Used varioustransformations like Filter,Expansion, Sequence Generator, Update Strategy, Joiner, stored procedure, and union to develop robust mappings in the Informatica designer.
- Collected differentreports with the deployment of different data into HDFS with the help of Tableau and Power BI.

Data Analyst

DITCOS(June 2018 - Oct 2019)

- Involved in the complete project life cycle,beginning with design discussions to project deployment.
- Analyzed Hadoop clusters using Big Data Analytic tools including Map Reduce, Pig and Hive.
- Worked with Hadoop Architect team in developing a Database design in HDFS using HBase Architecture design.
- Created Notebooks that extract raw data from a Data Bricks database,transform it, and then insert it into a cleansed Data Bricks database.
- Created and executed new or existing SQL queries to connect business intelligence (BI) tools like Jupyter notebooks (Python), Tableau Dashboard, and Excel reports.

Education

Master's in Computer Science
University of Dayton, Dayton, Ohio, USA

Bachelor's in Electronics & Commuication Engineering
VRSEC, Kanuru, AP, India

Get in Touch

npoojitha47@gmail.com

682-593-9292