

RANJAN HRIDAY

Irving, TX-75038 || +1 (817)936-7096 || madasranjan09@gmail.com || [Linkedin: https://www.linkedin.com/in/ranjanmadas](https://www.linkedin.com/in/ranjanmadas)

PROFESSIONAL SUMMARY:

- Experienced Data Engineer with 5+ years of expertise in Python, SQL, AWS, Azure, GCP, Snowflake, ETL, Airflow, Power BI, Tableau.
- Proficient in designing data models, architectures for structured data in Oracle and PostgreSQL, schema-less designs for unstructured data in NoSQL.
- Adhered to Agile methodology, using Confluence to document user stories, sprint plans, and retrospectives, ensuring structured project management.
- Open to nationwide relocation within the USA.

EDUCATION:

Master's in computer and information Sciences | The University of Texas at Arlington.

Aug 2021- May 2023

Sreenidhi Institute of Science & Technology, Hyderabad, India | Bachelors.

Aug 2015- May 2019

TECHNICAL SKILLS:

Programming Languages	: Python, SQL, R, Java, Scala, Shell Scripting.
Big data technologies	: Apache Spark (Including Spark SQL), Apache Delta, Hadoop, Hive, Pig, MapReduce, Kafka, HDFS.
Cloud Technologies	: AWS (S3, Glue, Redshift), Azure (DataBricks, Synapse, ADF), Snowflake Data Warehouse (Snow pipe), GCP (BigQuery, GCS, Dataflow).
Data Visualization Tools	: Tableau, PowerBI (DAX, Power Query), Microsoft Excel (VLOOKUP, Pivot tables and formulas), Quick sight, Looker.
Data Engineering Skills	: Data Warehousing, Apache Airflow, Dimensional Data Modelling, ETL, Star/Snowflake Schema, ELT, Docker, Containerization.
Databases	: MySQL, PostgreSQL, SQLite3, Mongo DB, Microsoft SQL Server, Oracle, Teradata.
Data Analysis	: scikit-learn, hmmlearn, PyTorch, TensorFlow, NumPy, Pandas, XGBoost, NLP, Machine Learning.
Tools/OS	: Talend, Informatica, Great Expectations, Alteryx, Jenkins, Kubernetes, Linux, Windows.

PROFESSIONAL EXPERIENCE:

AMD | Senior Data Engineer | Irving, Texas:

Jan 2024 – July 2024

- Worked on an existing application by automating queries, integrating solutions for cloud data transfer, capturing outputs logs in AWS Cloud.
- Developed ETL pipelines using Python, SQL to automate data extraction, transformation, and loading resulting in 50% reduction in data handling.
- Executed ETL/ELT processes AWS Data Pipeline, integrating data from multiple sources (databases, APIs, and flat files) into a centralized data lake.
- Capitalized on Python scripting to extract data from web APIs and unstructured sources such as JSON and XML, increasing data availability by 45%.
- Exploited Amazon EMR, Apache Spark for ETL tasks in Amazon S3 Subsequently, the transformed data was sent to Snowflake Data Warehouse.
- Mastered in Creating, Debugging, Scheduling, and Monitoring jobs using Airflow for ETL batch processing to load into Snowflake data warehouse.
- Created efficient data pipelines between Snowflake and Tableau, automating report generation tasks decreasing report generation time by 40%.
- Configured automated build and deployment pipelines in Git, performed regular code reviews via pull requests, ensuring the code quality.
- Collaborated with stakeholders to gather requirements and translate them into epics and user stories in JIRA, ensuring alignment with project goals.

AMD | Data Engineer | Irving, Texas:

Jan 2023 – Dec 2023

- Spearheaded machine learning model for sentiment analysis that achieved a 90% accuracy rate in predicting emotions from human voice data.
- Conducted comprehensive data analysis using advanced SQL queries and performed performance tuning for stored procedures and SQL scripts.
- Extracted data from HDFS into Spark RDDs and performed predictive analysis using ML algorithms to forecast trends and identify patterns.
- Employed SSIS, SSAS, and SSRS tools to elevate operational efficiency by 20% and drive well-informed decision-making.
- Designed robust ETL workflows using Azure Data Factory, automating data ingestion from Azure cloud storage to Azure Data Lake Storage.
- Capitalized on Azure Data Lake to create staging layers for raw, transformed, and structured data, enhancing data organization by 30%.
- Developed PySpark scripts within Azure Databricks to efficiently transform and load structured data into Azure Synapse Analytics.
- Crafted powerful presentations, detailed maps and charts in Power BI, levitating decision-making by 60% with interactive dashboards.
- Deployed data applications using CI/CD pipelines and DevOps (Jenkins, Docker, Kubernetes) best practices, collaborating with Jira.

AMAZON | Data Engineer, India:

June 2019 – June 2021

- Made use of Python NumPy, pandas for Data Extraction, Data Cleaning, Data Transformation from sources leading to an overall improvement.
- Leveraged Hadoop to manage and process large-scale datasets, integrating with Apache Spark and Kafka to streamline data workflows.
- Proficiently used SQL, T-SQL, and crafted complex procedures for structured data storage in data warehouses, provoking a 40% improvement.
- Pioneered data modeling initiatives (Star/Snowflake Schema), designing, and implementing efficient data models that led to 30% improvement.
- Engineered automated data extraction from AWS RDS PostgreSQL to S3 buckets using AWS Lambda, optimizing ETL pipeline's initial stage.
- Formulated AWS Glue Catalog, crawlers for data extraction and transfer into AWS S3, and migrating data to AWS Redshift via Glue ETL.
- Crafted powerful presentations, detailed maps and charts in Power BI, levitating decision-making by 60% with interactive dashboards.

Studioline | Data Engineer, India:

Jan 2018 – Aug 2019

- Orchestrated data solutions with GCP Dataflow, Pub/Sub, Dataprep for ingestion and transformation, paired with BigQuery for data warehousing.
- Configured GCS buckets to store and manage unstructured data, enabling efficient data processing and analysis in BigQuery and Dataproc.
- Leveraged advanced PySpark techniques to design and implement complex data processing pipelines in BigQuery, enabling processing of large data.
- Integrated Dataproc with Pub/Sub and Dataprep to automate data workflows, streamlining data ingestion and transformation processes

ACADEMIC PROJECTS:

Big Data and Cloud Technologies Projects:

- Decreased processing time by 40% with synchronized matrix multiplication algorithms in Scala, Apache Spark, and Pig.
- Attained 50% faster processing of large-scale datasets by developing responsive approaches with Hadoop, MapReduce.

CERTIFICATIONS:

Certified AWS Cloud practitioner, Microsoft Certified: Azure Data Fundamentals | SQL(HackerRank) | Advance Excel | Python