

Anurupa Velpula

Seattle, WA | (816) 217-3742 | velpulaanurupa4@gmail.com | <https://www.linkedin.com/in/anurupa-velpula-576a69206>

SUMMARY

With over 3 years of experience as a Data Engineer, I bring a wealth of expertise in data system design and management, ETL pipeline development, and data modeling. My proficiency in SQL, MySQL, Oracle, and SQL Server is complemented by my indepth knowledge of AWS technologies. I am adept at leveraging AWS tools like Redshift for data warehousing, Lambda for serverless computing, and S3 for efficient data storage. This expertise in AWS, combined with skills in Azure technologies, enhances my ability to optimize data storage, retrieval, and processing. My experience includes implementing machine learning algorithms to inform business decisions, and I excel in problem-solving, communication, and ensuring data integrity.

CERTIFICATIONS

- AWS Certified Developer - Associate
- AWS Certified Cloud Practitioner

TECHNICAL SKILLS

- **Programming Languages:** Python, Scala, Java, JavaScript, MySQL
- **Database | ETL:** Microsoft SQL Server, MySQL, Teradata SQL, DynamoDB, SSIS, Informatica, Cassandra, MongoDB.
- **Data Visualization:** Tableau, Power BI, Pandas, Advanced Excel, Amazon Quick Sight
- **Big Data:** Spark, HDFS, Map reduce, Yarn, Hive, Pig, Sqoop, Flume, Kafka, Oozie, Nifi, HBase, Pyspark
- **Cloud Platforms:** Amazon Web Services (EC2, EMR, S3, Redshift, Lambda, Glue, Athena, Kinesis), Microsoft Azure (Azure Data Factory, Azure Databricks, Azure Functions, Azure Synapse Analytics, HDInsight, Blob Storage, and Cosmos DB).
- **Packages:** NumPy, Pandas, Matplotlib, SciPy, Seaborn
- **Tools:** Alteryx, SSIS, SSRS, Talend, Informatica, Apache Airflow
- **Functional Skills:** Data Analysis, Data Pipelining, Data Visualization, Data Streaming, Data Integration, Data Transformation.

EXPERIENCE

Amazon, WA | Data Engineer

May 2022 - Present

- Involved in the entire Software Development Life Cycle – Analysis, Design, and Development – and coordinated with the team using **Agile Methodologies**.
- Implemented **PySpark** for data processing to handle data from various RDBMS and **streaming sources**; used Redshift for the Data Warehouse service.
- Built an end-to-end data pipeline and performed analytics using the AWS stack (**EMR, EC2, S3, RDS, Lambda, Glue, SQS, Redshift**).
- Implemented end-to-end data quality checks and monitoring alarms for **ETL** pipelines, minimizing data delay friction.
- Solely maintained Amazon **Redshift** cluster and AWS architecture, automating data pipelines and supporting data lakes.
- Developed data pipelines to extract data from various sources, transformed it as per the business requirements, and loaded it into **Redshift** for analysis.
- Employed **AWS Kinesis** to capture real-time data from various sources, handling terabytes of data.
- Worked on **CI/CD** solution, using **Git and Jenkins** to set up and configure the big data architecture on **AWS** cloud platform.
- Responsible for writing Unit Tests and deploying production level code with the assistance of **Git version** control.

HCL Global Systems, MI | Data Engineer

Jan 2022 to May 2022

- Served as a Data Engineer, responsible for **Data modelling, Data migration** and Designing ETL pipelines for both **cloud** and on-premises environments.
- Possess strong knowledge of **Apache Spark** components including **Spark CORE, Spark SQL, and Spark Streaming**
- Configured **Spark Streaming** to receive real time data from **Apache Kafka** and store streamed data to **HDFS** using **Scala** and **PySpark**.
- Involved in loading and transforming large datasets of **Structured, Semi-Structured** and **Unstructured** data and analyzed them using **Hive** queries.
- Performed Extract, Transform and Load operations from Source Systems to Azure Data Storage services leveraging **Azure Data Factory, Spark SQL** and **Azure Data Lake Analytics**.
- Executed Data Ingestion to Azure Services like **Azure Data Lake, Azure Storage, and Azure SQL DW**, processing data in Azure Databricks.
- Developed Pipelines in **ADF** using **Linked Services, Datasets, and Pipelines** to Extract, Transform, and load data from various sources like **Azure SQL, Blob storage, Azure SQL** Data warehouse.
- Imported data from multiple sources and created automated reports in **Power BI** after cleaning and preprocessing the data.

- Contributed to infrastructure design for optimal data ingestion, transformation, and loading using AWS big data technologies like **S3**, **Lambda**, and **Glue**.
- Reduced deployment time by 20% creating automated **CI/CD pipelines** in AWS CodePipeline using Powershell.
- Improved alert monitoring by 80% by designing a proactive alert management system using **AWS CloudFormation** to notify pipeline failures in real time.
- Improved data accuracy by 10% through exploratory data analysis on datasets.
- Developed **SQL** queries for complex data extraction and manipulation.
- Collaborated across ERP platforms and datasets to develop **ETL/ELT** processes reporting KPIs using AWS services and Power BI.

ACADEMIC PROJECT Real-time Social Media Sentiment Analysis Platform

- Built an end-to-end data pipeline for real-time ingestion from Twitter and Reddit using **Apache Kafka**.
- Utilized **Apache Spark Streaming** for efficient data processing, cleaning, and tokenization.
- Implemented sentiment analysis with **NLTK**, classifying emotions as Positive, Negative, or Neutral.
- Stored processed data in **Snowflake** for robust data management.
- Visualized sentiment trends in real-time using **Tableau** dashboards.
- Focused on storage optimization and compliance with data regulations.
- Deployed the application scalable using **Docker** and **Kubernetes**.

EDUCATION

Master’s in computer science
University of Missouri Kansas City

January 2021 – May 2022