Shravani Anupati

217-413-1331•shravanianupati18@gmail.com• www.linkedin.com/in/shravani-anupati18

SUMMARY:

- Over 3 years of professional experience in data engineering roles, specializing in designing and implementing robust data pipelines and ETL processes to streamline data processing workflows and ensure data reliability.
- Proficient in Python, SQL, and Spark for data integration, manipulation, and analysis, with hands-on experience in deploying Amazon EMR and working with AWS services such as Snowflake, S3, and DynamoDB for Big Data processing and management.
- Worked closely with data scientists and analysts to understand their data requirements and provided them with clean, reliable, and accurate data.
- Seeking a challenging role that allows me to leverage my technical expertise and analytical skills to drive innovation, solve complex problems, and contribute to the success of data-driven initiatives.

EXPERIENCE

Data Engineer, Zensar Technologies, Hyderabad, India   **Jun 2021 – Jul 2022**

- Designed and developed data pipelines to process, transform, and load large volumes of data from various sources into a centralized data warehouse.
- Implemented dbt (data build tool) for transforming and modeling data within the data warehouse, streamlining the ETL process, and enabling faster data insights and analysis for stakeholders.
- Worked on AWS Data pipeline to configure data load from S3 to Redshift.
- Implemented dynamic DAG configuration in Apache Airflow for real-time adaptability, ensuring agile and responsive workflow execution amidst changing data availability and processing needs.
- Hands-on experience with EC2 auto-scaling to dynamically adjust capacity based on application demand.
- Developed and implemented robust data models in Power BI to support effective visualization.
- Leveraged Kubernetes and Linux for containerized application deployment, optimizing resource utilization and system performance while ensuring high availability and fault tolerance.
- Managed data engineering and analytics projects on Databricks, utilizing Apache Spark for large-scale processing.
- worked with big data technologies such as Apache Hadoop, Apache Spark, and Apache Kafka to process and analyze large volumes of data.

Junior Data Engineer, Tech Mahindra, Hyderabad, India   **May 2020 – Jun 2021**

- Integrated Apache Cassandra as a distributed NoSQL database solution within the Hadoop ecosystem, leveraging its scalability and fault tolerance for high-throughput data storage.
- Built real-time streaming data pipelines on Azure using Scala and Azure Stream Analytics, processing high-volume data streams for timely insights and actionable intelligence.
- Established continuous integration and continuous deployment (CI/CD) pipelines using Azure DevOps, Git, and Azure Pipelines, accelerating software delivery and release cycles.
- Designed, developed, and tested dimensional data models using star and snowflake scheme methodologies.
- Participated in the design and implementation of data models, ensuring alignment with business requirements and scalability for future growth.
- Developed and maintained interactive dashboards using Tableau to provide insights into key business metrics.
- Experience with the full Microsoft BI stack, including SQL Server Integration Services (SSIS), SQL Server Analysis Services (SSAS), SQL Server Reporting Services (SSRS), and Power BI.
- Performed extreme coding to include Java map-reduce, Hive, Paquet & Shell scripting.
- Integrated and transformed data from diverse sources, including relational databases, flat files, APIs, and streaming data, to ensure unified and consistent data for analysis and reporting.

Data Engineer Intern, BHEL, Hyderabad, India   **Nov2018 - Apr 2019**

- Implemented Data Migration of Multistate level data from SQL server to Snowflake by using Python and Java methods.
- Strong background in designing, deploying, and maintaining PostgreSQL databases in production environments.
- Utilized AWS Lambda for serverless computing, enabling efficient and cost-effective execution of code in response to triggers from various AWS services or API calls, enhancing automation and scalability in data processing tasks.
- Organized the end-to-end ETL pipeline for transforming diverse customer sales data sourced from CSV files into a unified data warehouse using Python and pandas.

- Configured and monitored resource utilization and job scheduling with Apache YARN, optimizing cluster performance and resource allocation for maximum efficiency.

PROJECTS

Integration of Weather API

- I have completed a project that involved integrating a weather API, ingesting data into DynamoDB, and transmitting it seamlessly to Snowflake with Snow Pipe.
- This AWS project required me to build a robust real-time data pipeline using DynamoDB, Snowflake, and AWS Lambda. I leveraged AWS Lambda Functions to demonstrate its dynamic capabilities in orchestrating the real-time flow of data.
- I also integrated a weather API to ingest real-time data and used Snow Pipe for seamless transmission to Snowflake. Lastly, I optimized and scaled the data pipeline to ensure its efficiency.

EDUCATION

Masters in Computer Technology   **May 2024**
Eastern Illinois University, Charleston, IL

Bachelor's in computer science   **May   2020**
Vaageswari Engineering College, Karimnagar, India

Skills

- **Languages:** Python, Spark, SQL, Scala.
- **Tools & Technologies:** Snowflake, Airflow, MySQL, Hadoop, NoSQL, Tableau, Git, JSON Microsoft Office, Apache Hadoop, Kafka, Hive, Apache Spark.
- **Cloud Services:** Azure, AWS
- **Web programming:** jQuery, HTML, CSS
- **Analytical Skills:** Data Cleaning, Data Wrangling, Data Visualization, Data Extraction, Data Manipulation, Predictive Analysis, Data Analysis, Data Acquisition, Data Validation
- **Visualization:** PowerBI, Databricks, etc.
- **Soft Skills:** Critical Thinking, Communication, Risk Management, Problem-solving, Team workflows, Integration skills.