DIVYALATHA DEVANABOINA
Data Engineer

Atlanta, GA | 4708514108 | lathad2628@gmail.com | linkedin.com/in/divyalatha-devanaboina/

SUMMARY

- Experienced and results-driven **Data Engineer** with around **5 years** of hands-on experience in designing, developing, and implementing data solutions.
- Proficient in a wide range of programming languages including **R, Python, and SQL**, with expertise in **data manipulation**, **analysis**, and **visualization** using libraries such as **NumPy, Pandas**, **Matplotlib**, SciPy, Scikit-learn, Seaborn, TensorFlow, and PySpark.
- Extensive knowledge and practical experience in working with Big Data technologies, including **Apache Kafka**, **Hadoop, Hive**, **Apache Airflow**, **Snowflake**, **Teradata**, and **Spark SQL**.
- Skilled in building and optimizing data pipelines, **ETL** processes, and data workflows using tools like **Informatica**, Datastage, **Ab Initio**, **Talend**, and **SSIS**.
- Proven expertise in cloud technologies, with hands-on experience in **AWS** (RDS, EC2, Lambda, S3, DynamoDb, Redshift, SQS, SNS, Glue, Cloud Watch, CloudFormation, EMR, Athena, Kinesis, DAX) and **Azure** (Data Factory, Databricks, Data Lake Storage), enabling seamless integration and management of data across cloud platforms.
- Proficient in utilizing reporting tools such as **Tableau**, **Power BI**, and **SSRS** to create interactive dashboards and reports, providing actionable insights to stakeholders.
- Strong background in database management systems including **MySQL**, **MongoDB**, **SQL Server**, **PostgreSQL**, **Teradata**, **Oracle**, **Kusto**, **Cosmos DB**, and **Denodo**.
- Adhering to best practices and methodologies including **SDLC**, **Agile**, and **Waterfall**, with a focus on delivering high-quality solutions that meet business requirements and objectives.
- Proficient in version control systems like **Git**, and skilled in using **MS Office** and **Docker** for efficient collaboration and development processes.

PROFESSIONAL EXPERIENCE

Epsilon | GA   **Sep 2023 – Current**
Data Engineer

- Developed and optimized **ETL processes** in Snowflake, leveraging **SQL, Snowpark** and Snowflake's features to ensure seamless data extraction, transformation, and loading.
- Improved query performance by 30% through in-depth Performance optimization, including utilization of Snowflake's advanced SQL features.
- Designed, implemented, and managed a robust **data warehousing** solution using **Snowflake**, enabling the organization to efficiently store and analyze terabytes of data.
- Created new data validation processes using **Python** and **PySpark** to ensure data quality and integrity, conducting data profiling and cleansing activities.
- Designed, built, tested, and maintained data management systems using **Apache Spark** and **Apache Hadoop** ecosystem tools such as **Hive** and **HDFS**, enabling efficient data processing and analytics.
- Implemented data security and **governance** measures with modern-day security controls such as **encryption** and **access controls**, ensuring compliance with data privacy regulations.
- Implemented real-time **data ingestion** pipelines using **Snowpipe**, enabling near-real-time access to critical data for analytics and reporting and worked on Disaster Recovery and High availability.
- Enhanced data management efficiency and accuracy through proficient utilization of **Snowflake** features including data retention, Time Travel, clustering, and materialized views. Implemented dynamic **data masking** and fail-safe mechanisms to ensure data integrity and streamline workflows.
- Optimized decision-making processes by leveraging **Snowflake** functionalities such as **data sharing**, scheduling tasks, and dynamic **data masking**. Utilized features like clones and streams to facilitate seamless data operations and enhance overall system reliability.

Amazon | Seattle, WA,   **Jul 2022 – Aug 2023**
Data Engineer/ Software Development Engineer

- Implemented and optimized **Spark SQL** queries and data processing tasks within the **Spark ecosystem**, ensuring scalability and performance for processing and analyzing large-scale datasets.
- Identified and evaluated pertinent data sources, such as product catalogs, inventory databases, and user purchase history, optimizing **AWS** data storage and retrieval via services like **Amazon S3** and **RDS**.
- Played a pivotal role in designing and deploying **cloud-based** infrastructure solutions for **data pipelines**, leveraging various **AWS services** such as **CloudFormation**, **RDS, EC2, Lambda, Athena, S3, DynamoDB, SQS, Redshift, SNS, AWS Glue, CloudWatch, and IAM** , with proficiency in **Python** and **PySpark**.
- Spearheaded the development of an end-to-end data engineering pipeline, from data source identification to **schema development**, **API integration**, and data transformation, utilizing **AWS services** like **Amazon S3**, **RDS**, and **Glue**.
- Engineered an efficient data storage and processing system, optimizing algorithms for item buyability and ensuring data accuracy while seamlessly integrating with **AWS** services such as **EMR** and **Lambda**.
- Established robust data quality monitoring processes and maintained comprehensive documentation to enhance the user shopping experience, leveraging **AWS cloud services** for both batch and stream processing.
- Utilized **SSIS** to extract data from diverse sources such as databases, flat files, and APIs, performing data cleansing, validation, and transformation operations to ensure data quality and consistency.
- Designed and implemented backup and recovery strategies for, by ensuring data availability and integrity in the event of system failures or disasters.
- Implemented, maintained and managed large-scale **data warehousing** solutions, ensuring scalability, performance, and reliability.

Cognizant Technology Solutions (Novartis) | India   **Jul 2019 – Jul 2021**
Data Engineer/ Programmer Analyst

- Developed and optimized data processing workflows using Hadoop ecosystem components such as **HDFS**, **MapReduce**, and **Hive**, enabling batch processing and analytics on massive datasets.
- Led the development of end-to-end data pipelines in Azure, incorporating services like Azure Data Factory, Azure Databricks by using Pyspark and Python.

- Streamlined data processing by automating ETL workflows in **Informatica**, **Datastage**, and **Ab Initio**, resulting in a **30% reduction** in manual handling and improved data processing speed. This initiative enhanced data quality and integrity within **Snowflake**, with additional integration of **Informatica** with Airflow for enhanced workflow management.
- Developed backend applications to integrate real-time data and automate data loading processes using **Oracle SQL, PL/SQL, T-SQL**, and **Unix Shell** scripting. Engaged in **SQL** performance tuning, troubleshooting, and database administration to ensure optimal database performance.
- Designed, optimized, and secured **CI/CD** pipelines using **Apache Airflow**, **Docker**, **Kubernetes**, and **SQL**, collaborating cross-functionally for efficient data processing and analysis.
- Achieved operational cost reduction and improved data processing speed by 20% through pipeline optimizations, while mentoring junior data engineers to ensure team growth and development.
- Utilized **Snowflake's** architecture for external data sharing, supporting **data monetization** strategies, and collaborated on end-to-end **data pipelines**, optimizing ETL job performance and modernizing infrastructure, including ETL processes.
- Configured and managed **Apache Kafka** clusters to ensure high availability, fault tolerance, and scalability of data streaming infrastructure, handling terabytes of data per day.
- Implemented data visualization solutions with **Power BI** and **Tableau**, transforming complex datasets into actionable insights, and contributed to **ERD** development while focusing on data security and compliance.
- Maintained version control systems such as **Git** for code management and collaboration, ensuring code integrity and facilitating team collaboration.

Espire | India   **Aug 2018- Jun 2019**
Data Analyst
- Conducted exploratory data analysis using **Python** libraries such as **NumPy**, **Pandas**, and **Matplotlib**, resulting in a 20% increase in insights extracted from structured and unstructured datasets.
- Utilized **SQL** queries to retrieve, manipulate, and analyze data stored in **SQL** Server databases, ensuring a **15% improvement** in data integrity and accuracy.
- Developed interactive and visually appealing dashboards and reports using **Tableau**, improving stakeholder comprehension by 30%.
- Created interactive and insightful visualizations such as bar charts, line graphs, scatter plots, and geographical maps to present data **effectively by 15%**.
- Participated in **ETL process**, transforming and loading data from various sources into a database using SSIS and done transformations by using framework DBT and Dataforms.
- Collaborated with cross-functional teams to gather data requirements and ensure the accuracy and completeness of data.

TECHNICAL SKILLS

Programming Language:	R, Python, SQL
Packages:	NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow, PySpark
Big Data Ecosystem:	Apache Kafka, Hadoop, Hive, Apache Airflow, Snowflake, Teradata, Spark SQL
ETL Tools:	Informatica, Datastage, Ab Initio, Talend, SSIS (SQL Server Integration Services)
Cloud Technologies:	AWS (RDS, EC2, Lambda, S3, DynamoDB, Redshift, SQS, SNS, Glue, Cloud Watch, CloudFormation, Redshift, EMR, Athena, Kinesis, DAX), Azure (Data Factory, Databricks, Data Lake Storage)
Reporting Tools:	Tableau, Power BI, SSRS
Database:	MySQL, MongoDB, SQL Server, PostgreSQL, Teradata, Oracle, Kusto, Cosmos DB, Denodo
Methodology	SDLC, Agile, Waterfall
Other Tools:	Git, MS Office, Docker
IDE's:	PyCharm, Jupyter Notebook

EDUCATION

Master of Science   **Aug 2021 – Dec 2022**
University of Cincinnati | Cincinnati, OH
Bachelor of Technology   **May 2015 – Apr 2019**
Prasad V Potluri Siddhartha Institute of Technology | Vijayawada, India

PROJECTS

- **Implementation of an IOT Based Voice Controlled Smart Home:** Developed Voice-controlled home automation benefits adults and physically disabled individuals by enabling hands-free device control through Google Home, IFTTT, and RaspberryPi, enhancing accessibility and independence.
- **Rental Portal:** Designed a rental house search web application to efficiently connect owners and tenants in metropolitan cities.
- Developed **Fruit Basket project** for INMAR Technologies. Fruit Basket is an online application where people can buy and sell fruits. It's like a platform where buyers can buy fruits by just making online payment from the system. Similarly, sellers can login and set certain prices for the fruits which they want to sell. Fruits which are sold get deducted from the system and goes in the queue of buyer.
  Environments: HTML, CSS, JavaScript, MSSQL, Bootstrap, XML, JSON, React JS.

CERTIFICATIONS

- **Certified in "Microsoft Azure Data Fundamentals (DP-900)"**
- **Python bootcamp certification**
- **Certified in Informatica Power Center**