**Susmitha **

**Susmithareddy311@gmail.com**

**Professional Summary:**

* 6+ Years of experience in designing data-intensive applications using **Big Data Analytical**, Cloud **Data engineering** (Aws,), Data Visualization, **Data Warehouse**, Reporting, and Data Quality solutions.
* Implemented large Lambda architectures using Azure Data platform capabilities like Azure Data Lake, **Azure Data Factory**, **HDInsight, Azure SQL Server, Azure ML** and Power BI.
* Practical understanding of the **Data modelling** (Dimensional & Relational) concepts like Star - Schema Modelling, **Snowflake Schema Modelling**, Fact and **Dimension tables**.
* Strong understanding of the architecture and capabilities of **Big Query** and **Snowflake**, enabling efficient storage, processing, and analysis of large-scale datasets.
* Hands-on experience working in various cloud (**GCP, AWS and Azure**) environments.
* Expertise in transforming business requirements into analytical models, designing algorithms, building models, developing **Data Mining**, Data Acquisition, Data Preparation, **Data Manipulation**, Feature Engineering, **Machine Learning Algorithms**, Validation and Visualization and reporting solutions that scales across massive volume of structured and unstructured Data.
* IT experience on Big Data technologies, **Spark**, database development. Good experience in Amazon Web Service (AWS) concepts like **EMR** and **EC2** webservices which provides fast and efficient processing **of Teradata Big Data Analytics**.
* Strong experience working with **HDFS, MapReduce, Flume, Kafka, Oozie, Pig and HBase**.
* Experience with creating **ETL** framework with **Python**.
* Experience with Requests, **NumPy, Matplotlib, SciPy, PySpark and Pandas** python libraries during development lifecycle and experience in developing APIs for the application using Python.
* Good Knowledge on Big Data Tools such as **Apache Kafka, Apache Spark, Airflow, Hive, Sqoop**, **Delta Lake** with different file formats like Avro, Csv and Parquet.
* Expert in designing Parallel jobs using various stages like Join, Merge, Lookup, remove duplicates, Filter, Dataset, look up file set, Complex flat file, Modify, Aggregator & XML.
* Extensive experience working with ETL tools **Airflow /SSIS** in **OLAP** and **OLTP** environments.
* Experience in building data ingestion, ETL and data processing pipelines using On Prem and Cloud services.
* Coordinating with Business Users, functional Design team and testing team during the different phases of project development and resolving the issues.
* Hands-on experience in handling database issues and connections with **SQL** and **No SQL** databases such as **MongoDB, SQL server**.
* Good knowledge of Teradata and Snowflake databases.

**Technical Skills:**

* **Big Data Tools**: Hadoop Ecosystem: Map Reduce, Spark 2.3, Airflow 1.10.8, Nifi 2, HBase 1.2, Hive 2.3, Pig 0.17 Sqoop 1.4, Kafka 1.0.1, Oozie 4.3, Hadoop 3.0
* **Visualization Tools:** Tableau, Looker, Power BI, Qlik Sense, Microsoft Excel, Python (Seaborn, Matplotlib), SSIS, SSRS
* **Database:** MYSQL, SQL Server, PostgreSQL, Azure ML Studio, Oracle
* **Python Libraries:** Pandas, NumPy, TensorFlow, Keras, Scikit-learn, PyTorch
* **Cloud Technologies:** AWS, GCP, Snowflake, Microsoft Azure
* **Data Modelling:** Logistic Regressions, Decision Trees, Multivariate Regression, Neural Networks, Random Forests
* **Methodologies:** Agile/Scrum, Waterfall
* **Version Tools:** GIT, SVN
* **Tools:** Microsoft Office, DAX, PowerPoint, SharePoint, Hypothesis Testing
* **Certification:**- Snowflake SnowPro Core Certified

**Professional Experience:**

**Kinder Morgan, Houston, Texas Aug 2023 – Till Date.**

**Senior Data Engineer**

**Responsibilities:**

* Designed, developed, and tested dimensional data models using Star and Snowflake schemas in accordance with the **Kimball methodology**.
* Managed AWS tools, including CloudWatch and CloudTrail, to monitor and track cloud resources and activities.
* Implemented user versioning in S3 buckets for highly sensitive data storage.
* Integrated **AWS DynamoDB** with **AWS Lambda** to manage data storage and backup through DynamoDB streams.
* Installed and configured **Apache Airflow** for AWS S3 buckets and created Directed Acyclic Graphs (DAGs) for workflow automation.
* Developed scripts for automating data ingestion processes using **PySpark** and **Scala**, connecting to various sources, including APIs, AWS S3, Teradata, and Redshift.
* Migrated on-prem Informatica **ETL** processes to **AWS cloud** and **Snowflake**.
* Employed Spark streaming for dividing streaming data into batches and processing them in the Spark engine.
* Wrote Spark applications for data validation, cleansing, transformation, custom aggregation, and analysis, making data available to data scientists.
* Proficient in building Snowpipe and experienced in Data Sharing within the **Snowflake Database**, Schema, and Table structures.
* Migrated on-prem Informatica ETL processes to **AWS cloud** and **Snowflake**.
* Expertise in migrating existing databases from on-premises to AWS Redshift using various AWS services.
* Created, modified, and executed DDL in **AWS Redshift** and **Snowflake** tables for data loading.
* Monitored resources and applications using AWS CloudWatch, including setting up alarms to track metrics for **EBS, EC2, ELB, RDS, S3**, and **SNS**, with customized notifications.
* Scheduled jobs using Oozie for efficient workflow management.
* Managed the import and export of data from **Snowflake**, **Oracle**, and **DB2** into **HDFS** and **HIVE** using Sqoop for analysis, visualization, and reporting.
* Used filters, quick filters, sets, parameters and calculated fields on **Tableau** and **Power BI** reports.

**Environment**: Python, Power BI, AWS Glue, Athena, SSRS, AWS S3, AWS Redshift, AWS EMR, AWS RDS, DynamoDB, SQL, Tableau, Distributed Computing, Snowflake, Spark, Kafka, MongoDB, Hadoop, Pyspark, Oozie, HBase, Hive, Pig, Docker.

**IBM, Hyderabad, India Jan 2018 – Dec 2021**

**Data Engineer**

**Responsibilities:**

* Created ETL mapping document and ETL design templates for the development team.
* Created external tables on top of S3 data which can be queried using AWS services.
* Developed ETL pipelines in and out of data warehouse using combination of **Python** and Snowflakes Snows Writing **SQL** queries against **Snowflake**.
* Utilized **Azure Machine Learning** to develop predictive models and apply advanced analytics techniques, empowering data-driven decision-making and optimizing business processes.
* Conducted ETL testing, including extracting data from databases, transforming data, and uploading it to data warehouse servers using **SQL** and **NoSQL**.
* Proficient in designing and developing multidimensional and tabular data models using SQL Server Analysis Services (**SSAS).**
* Created parameterized reports, drill-down reports, and sub reports in **SSRS** to cater to diverse reporting requirements.
* Implemented data visualization techniques for creating Dashboards and reports using **Python** and **Tableau** to present complex data insights to stakeholders, reduction in time spent on data analysis.
* Collaborated with cross-functional teams to develop and implement **machine learning algorithms** for maintenance and collect and report key metrics.
* Strong understanding of data mining algorithms and the ability to implement data mining models within **SSAS** for predictive analytics and pattern recognition.
* Hands-on experience with Snowflake utilities, SnowSQL, SnowPipe, Big Data model techniques using Python.

**Environment**: Amazon Web Services (AWS), Hadoop, Hive, Presto, PySpark, Databricks, Airflow, Apache Tez, MySQL, PostgreSQL, SQL Server, Python, Scala, Spark, Spark-SQL.

**Tata Communication Apr 2015 – Dec 2017**

**Project Planning & Implementation Engineer**

**Responsibilities:**

* Conducted risk assessments and developed **mitigation** strategies to address potential project challenges.
* Managed **project budgets**, ensuring efficient allocation of resources and **cost control**.
* Prepared and presented project status reports to senior management, outlining key milestones and performance metrics.
* Implemented project management tools and methodologies to streamline project workflows and enhance productivity.
* Collaborated with cross-functional teams to identify and resolve issues impacting project delivery.
* Led project kick-off meetings, ensuring alignment on project objectives and expectations.
* Conducted post-project evaluations to analyse outcomes, identify lessons learned, and provide recommendations for process improvements.
* Mentored and provided guidance to junior project team members, fostering professional development and knowledge sharing.
* Adapted to evolving project requirements and changes, demonstrating flexibility and **problem-solving skills.**
* Maintained up-to-date documentation of project plans, schedules, and deliverables to support project audits and compliance requirements.
* Stayed current with industry best practices and emerging trends in project planning and implementation to drive continuous improvement.

**Education:**

* Bachelors in Electronics and Communications from ANU, India
* Masters in Data Science from University of Houston, Houston, Texas