SUSMITHA

Senior Data Engineer

+1(281)-750-6048| [susmithareddy311@gmail.com|](mailto:susmithareddy311@gmail.com|)

**PROFESSIONAL SUMMARY**:

* 5+ 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 modeling (Dimensional & Relational) concepts like Star - Schema Modeling, Snowflake Schema Modeling, Fact and Dimension tables.
* Strong understanding of the architecture and capabilities of BigQuery and Snowflake, enabling efficient storage, processing, and analysis of large-scale datasets.
* Hands-on experience working in various cloud (GCP, AWS and Azure) environments.
* 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.
* Constructing and manipulating large datasets of structured, semi-structured, and unstructured data and supporting system application architecture using tools like SAS, SQL, Python, R.
* 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 KNOWLEDGE:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | | **Tools/Technologies** | |
| **Visualization Tools** | Tableau, Looker, Power BI, Qlik Sense, Microsoft Excel, Python (Seaborn, Matplotlib), SSIS, SSRS | |
| **Tools** | Microsoft Office, DAX, PowerPoint, SharePoint, Hypothesis Testing | |
| **Database** | MYSQL, SQL Server, PostgreSQL, Azure ML Studio, Oracle, DB2 | |
| **Python Libraries** | Pandas, Numpy, TensorFlow, Keras, Scikit-learn, PyTorch. | |
| **Cloud Technologies** | AWS, GCP, Snowflake, Microsoft Azure | |
| **Environments** | GitHub, Google colab, PyCharm, VSCode, Jupyter, Rstudio | |
| **Data Modeling** | Logistic Regressions, Decision Trees, Multivariate Regression, Neural Networks, Random Forests | |
| **Methodologies** | Agile/Scrum, Waterfall. | |

**PROFESSIONAL EXPERIENCE:**

**Data Analyst** Jun 2018- Dec 2021

**First Republic New York, NY**

**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 in the retail industry.
* 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.

**Project Planning & Implementation Engineer | Tata Communication** Apr 2016- May 2018

**RESPONSIBILITIES:**

* Designed project plans that include project scope, timelines, budget, and resource requirements.
* Coordinated with vendors, contractors, and other stakeholders to ensure that projects are executed according to plan.
* Monitored project progress and adjustments as necessary to keep projects on track.
* Maintained clear and consistent communication with stakeholders throughout the project, including project sponsors, clients, and team members.

**EDUCATION**:

**MSc in Data Science** | **University of Houston** | Houston, Texas | GPA: 3.6

* Awards & Scholarships: Dean’s Honors List, Engineering Dean’s Master Scholarship, Master’s Competitive Scholarship
* **Relevant Courses**: Neural Networks, Machine Learning, Statistics, Predictive Analytics, Database Management, Big Data, Time series Forecasting, Machine Diagnostics, Hypothesis Testing

**CERTIFICATIONS:**

* Machine Learning (HarvardX)
* Google Data Analytics Professional Certificate
* Tableau Desktop Specialist (TableauSoftware, LLC)
* Snowflake-The Complete Masterclass (2023 edition)