**Varshini Suryadevara**

**Phone No:** 7373287435 | Email: [varshinis1605@gmail.com](mailto:akkineni.saibhavana@gmail.com)

**Summary:**

* Over 9+ years of Professional work experience as Data Engineer in data analysis, developing machine learning and predictive models using Python, R, and data mining with large data sets of both structured and unstructured data.
* Experience in working with Big Data Technologies
* Experience in ETL data from various sources into Data Warehouse and processing data like collecting, aggregating, and moving data from various sources using Kafka, Flume.
* Developed Spark applications that can handle data from RDMS and Streaming sources.
* Able to use Sqoop to migrate data between RDBMS, NoSQL databases and HDFS.
* Experience in Machine Learning algorithms such as Linear Regression, Logistic Regressions, Decision Trees, Support Vector Machine and K-means clustering.
* Expertise in Python libraries such as NumPy, pandas, SciPy and Matplotlib.
* Proficient in databases like MySQL and MongoDB.
* Excellent working knowledge of Hadoop, Hive, Sqoop, HBase & Oozie in real time Environment.
* Worked on ETL processes to load data from various sources to HDFS using Sqoop, Pig & Oozie.
* Experience in Data pre-processing, Data transformation and Data cleansing procedures.
* Experience and knowledge of real time data streaming analytics using Spark Streaming, Kafka, Flume.
* Experienced in Spark Core, Spark RDD, Spark SQL and writing Spark Streaming and spark batch jobs using spark MLib for analytics.
* Hands on experience in developing Airflow for scheduling and orchestrating ETL process.
* Good knowledge in maintaining version controls system such as GIT, SVN, Bitbucket.
* Hands on experience working in GCP services like Big Query, Cloud Storage (GCS), cloud function, cloud dataflow, Pub/sub, Cloud Shell, GSUTIL, Big Query, Data Proc, Operations Suite (Stack driver).
* Hands on experience on Google Cloud Platform (GCP in all the big data products Big Query, Cloud Data Proc, Google Cloud Storage, Composer (Airflow as a service)
* Experienced in working with various Python Integrated Development Environments like IDLE, PyCharm and Atom.
* Experience on Migrating SQL database to Azure data Lake, Azure data lake Analytics, Azure SQL Database, Data Bricks and Azure SQL Data warehouse and controlling and granting database access and Migrating On premise databases to Azure Data Lake store using Azure Data factory.
* Strong development skills with Azure Data Lake, Azure Data Factory, Synapse, SQL Data Warehouse Azure Blob, and Data Bricks.
* Strong background in Agile/Scrum development methodologies.
* Outstanding communication skills, interpersonal abilities, problem-solving aptitude, and a team-oriented attitude, along with the capability to effectively communicate with various levels within an organization, including technical, management, and customers.

# Technical Skills:

Python Programming | SQL | R | Java | Data Visualization | Machine Learning | Microsoft Excel | MySQL | MongoDB | Snowflake | CSS | HTML | JavaScript | Django | Quantitative Methods | IDLE | Atom | Jupyter | GIT | Data Warehousing | Advanced Data Mining | Business Intelligence (BI) | Data Structures | Data Modeling | Data Visualization | Risk Management | Python | HDFS | MapReduce | Hive | Sqoop | Oozie | Flume | HBase | Kafka | Spark | Airflow | Amazon EC2 | S3 | BigQuery | Cloud Data Proc | GCS Bucket | G-Cloud Function | Apache Beam | Cloud Shell | GSUTIL | BQ Command Line | Cloud Data Flow | Microsoft Office Suite | QlikView | Tableau | Power BI | Microsoft Office Suite (Word, Excel, PowerPoint) | Windows

# Professional Experience:

## Bank of America - San Francisco, CA Aug 2021 to Present

**Data Engineer**

**Responsibilities:**

* Collaborated with the team to understand the requirements and data sources for processing healthcare claims data.
* Experience in building multiple Data pipelines, end to end ETL and ELT process for Data ingestion and transformation in GCP and coordinate task among the team.
* Develop and deploy the outcome using spark and Scala code in Hadoop cluster running on GCP.
* Design and implement various layers of Data Lake, Design star schema in Big Query.
* Using g-cloud function with Python to load data into big query for on arrival csv files in GCS bucket.
* Process and load bound and unbound data from Google pub/subtopic to big query using cloud Dataflow with Python.
* Got involved in migrating on prem Hadoop system to using GCP (Google Cloud Platform).
* Migrated previously written cron jobs to airflow/composer in GCP.
* Support existing GCP Data Management implementations.
* Created GCP Big Query authorized views for row level security or exposing the data to other teams.
* Have Extensive Experience in IT data analytics projects, Hands on experience in migrating on premise ETLs to Google Cloud Platform (GCP) using cloud native tools such as BIG query, Cloud Data Proc, Google Cloud Storage, and Composer.
* Designed Pipelines with Apache Beam, Kubeflow, and Dataflow and orchestrated jobs into GCP.
* Developed and Demonstrated the POC, to migrate on-prem workload to Google Cloud Platform using GCS, Big Query, Cloud SQL, and Cloud DataProc.
* Documented the inventory of modules, infrastructure, storage, components of existing On-Prem data warehouse for analysis and identifying the suitable technologies/strategies required for Google Cloud Migration.
* Design, development, and implementation of performing ETL pipelines using python API (pySpark) of Apache Spark.
* Build data pipelines in airflow in GCP for ETL related jobs using different airflow operators both old and newer operators.
* Created BigQuery authorized views for row level security or exposing the data to other teams.
* Developed ETL pipeline to extract healthcare claims data from multiple sources, including structured and unstructured data.

**Environment**: Spark, Airflow, GCP, Cloud SQL, Big Query, Cloud Data Proc, GCS, Cloud SQL, Cloud Composer, HDFS, GitHub, Hive, Python, Java, SQL Server.

## Macy’s, New York, NY Feb 2019 to July 2021

**Data Engineer**

## Responsibilities:

* As a Data Engineer, Responsible for designing and optimizing systems.
* Optimized system which processes huge amount of logs every day and loading it into the data warehouse.
* Used Spark, Airflow and Snowflake to design and implement data loading, aggregating and jobs that can handle hundreds of GBs amount of json files.
* Built tools using Tableau to allow internal and external teams to visualize and extract insights from big data platforms.
* Responsible for expanding and optimizing data and data pipelines, as well as optimizing data flow and collection for cross functional teams.
* Build ETLs with Apache Spark to load and transform raw data into easy-to-use dimensional data for self- service reporting.
* Improved the deployment and testing infrastructure within AWS, using tools like Jenkins, Puppet and Docker.
* Collaborated with teams to make sure data needs are considered during product development and to guide decisions related to data.

**Environment**: Scala, Spark, Spark SQL, Kafka, Airflow, AWS, Snowflake, Tableau.

## Accenture, Hyderabad, India Aug 2015 to May 2018

**Data Engineer**

**Responsibilities:**

* Collaborated with clients and businesses to understand expectations and requirements for Data and Analytics applications, including Data Services, Data Integration, Marketing, and Business Performance Metrics.
* Worked extensively with Azure Storage, Azure Data Lake, Azure File Share, and Azure Blob Storage for data file storage and retrieval.
* Employed Extract Transform and Load processes to transfer data from Source Systems to Azure Data Storage services using a mix of Azure Data Factory, T-SQL, Spark SQL, and U-SQL Azure Data Lake Analytics. Ingested data into various Azure Services (Azure Data Lake, Azure Storage, Azure SQL, and Azure DW) and processed it using Azure Databricks.
* Utilized Azure Key Vault to securely store connection string secrets and database passwords, which were then used to configure Linked services.
* Developed data pipelines in Azure Data Factory by leveraging linked services, datasets, and pipelines for ETL from Azure SQL, Blob storage, Azure SQL Data Warehouse, and write-back tools to handle forward and backward data flows.
* Successfully migrated a traditional on-premises warehouse to the scalable Azure Synapse Warehouse, ensuring smooth data integration and optimized storage solutions.
* Crafted complex SQL queries using stored procedures, common table expressions (CTEs), and temporary tables to support Power BI and SSRS reports.
* Skillfully used Microsoft Power BI for designing, developing, and maintaining dynamic reports and dashboards by integrating multiple data sources. Utilized Power BI Desktop for advanced data analysis, delivering insightful visualizations and actionable business intelligence.
* Experienced in migrating SQL databases to various Azure platforms such as Azure Data Lake, Azure Data Lake Analytics, Azure SQL Database, Data Bricks, and Azure SQL Data Warehouse, managing and granting database access. Successfully executed migration of on-premises databases to Azure Data Lake store using Azure Data Factory.
* Proficient in data analytics and engineering across multiple Azure platforms, including Azure SQL, Azure SQL Data Warehouse, Azure Data Factory, and Azure Storage Account. Expert in extracting, cleansing, consuming, and publishing data streams for diverse user bases.
* Designed and developed SQL objects and complex SQL queries using Microsoft Data Analysis, Data Modeling, Data profiling, T-SQL, and advanced functions.
* Replicated existing application logic and functionality in the Azure Data Lake, Data Factory, SQL Database, and SQL Data Warehouse environment.
* Architected and implemented medium to large-scale BI solutions on Azure using Azure Data Platform services (Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics, and Azure SQL DW).
* Created Databricks notebooks using PySpark for processing and formatting data from various files, leading to efficient data manipulation and rapid file operations.
* Analyzed source system data and designed the star schema for the data warehouse to be developed.
* Developed Dynamic Linked services and Datasets to reuse them in different pipelines by passing parameter values at runtime.

**Environment:** Azure SQL Server, Azure Synapse, T-SQL, SQL Server Integration services (SSIS), SQL Server Reporting Services (SSRS), Power BI, GIT, Azure Data Lake, Data Factory, Azure Blob, Azure Storage Explorer, PySpark, Power BI, PowerShell, JIRA, Devops.

## Mouritech, Hyderabad, India Jun 2013 - Jul 2015

**Python Developer**

## Responsibilities:

* Involved in interaction with client for requirement specifications and analyzed the system requirements and specifications.
* Maintaining and developing the products by using entire Backend.
* Worked on developing the front end of web application using HTML, CSS, JavaScript, jQuery, Bootstrap and created interactive webpages.
* Worked on creating RESTful http services to interact with user interface.
* Experience on using Django admin and created superusers, updated tables in database.
* Experience using many python modules and packages to build an application rapidly.
* Developed user interface using CSS, HTML, and JavaScript.
* Worked with Python modules to extract/load data from MySQL source database.
* Created entire application using python, Django, MySQL
* Worked with Purchase orders of the products by using Python Scripting as Backend and HTML, CSS, Bootstrap, JavaScript as Frontend.
* Experience in using the regular expression to match the pattern with the existing code.
* Maintaining and updating the products by the vendors and customer.
* Collaborated with teams to improve solutions and participated in status meetings, resolving ongoing problems and effective communication with Project Manager.

**Environment**: Python, Django, HTML, CSS, SQL, MySQL, JSON, REST.