Swetha Adike   swethaad26@gmail.com
Ph: 848-216-5340,   https://www.linkedin.com/in/swetha-ad/

Professional Summary:

- Around 9 years of IT experience in design, development, and maintenance of Telecommunication, Banking, Pharmaceutical, and Retail Industry Software application needs.
- Strong experience in Data Analyst, Data mining with large data sets of Structured and Unstructured data, Data Acquisition, Data Validation, Predictive modeling, Statistical modeling, Data modeling, Data Visualization. Adept in statistical programming languages like R and Python, SAS, Apache Spark including Big Data technologies like Hadoop, Hive, Pig
- Experience in ETL tool development, design, implementation, and support with a major focus on Data Warehousing, and Database Applications.
- Implement test data to serve various IT environments (DEV, QA, UAT).
- Good experience on Bigdata projects, comprehensive experience as a Hadoop Developer in Hadoop Ecosystems like Hadoop Map Reduce, Hadoop distributed File System (HDFS), HIVE, IMPALA, PostgreSQL, Spark, Python.
- Strong Understanding of Data Warehousing concepts like OLAP, OLTP, Star Schema, Schema, and Dimension Data Modeling.
- Can work parallelly in both GCP and Azure Clouds coherently.
- Recreating existing application logic and functionality in the Azure Data Lake, Data Factory, SQL Database, and SQL Data warehouse environment.
- Experience in building and architecting multiple Data pipelines, end to end ETL and ELT process for Data ingestion and transformation in GCP.
- Well-versed in version control and CI-CD tools such as GIT, SourceTree, Bitbucket, etc, and Amazon Web Services (AWS) products S3, EC2, EMR, and RDS.
- Experience building distributed high-performance systems using Spark and Scala
- Propose architectures considering cost/spend in Azure and develop recommendations to right-size data infrastructure.
- Datacenter Migration and Azure Data Services have a strong virtualization experience.
- Use Kafka a publish-subscribe messaging system by creating topics using consumers and producers to ingest data into the
- Very keen in knowing newer techno stack that Google Cloud platform (GCP) adds.
- Experience managing Big Data platforms deployed in Azure Cloud.
- Implemented Copy activity, and Custom Azure Data Factory Pipeline Activities for On-cloud ET processing.
- Used output XML files, to remove empty delta files and to FTP the output XML files to a different server.
- Expertise in Importing and exporting data from different databases like Oracle, Teradata into HDFS, and Hive using Sqoop.
- Worked with an assortment of planning apparatuses including Autosys.
- Involved in creating Tables, Partitioning, bucketing of tables and creating UDFs in Hive.
- Good knowledge of Relational Data Modeling and expert-level understanding of Oracle databases.
- Data Conversation from Flat files to Oracle using Oracle SQL*Loader.
- Strong working experience in the use of SDLC methodologies, and standards and gathering the requirements for Data Analysis, Design, Development, Implementation, and Testing of Data Warehousing using Data Extraction, Data Transformation, and Data Loading (ETL).
- Strong knowledge of Oracle 10g/11g,12c.19c PL/SQL
- Install, Config, Test, Monitor, upgrade and tune new and existing PostgreSQL databases.
- Working experience with Agile Software development method.
- Experience in analyzing data using HiveQL, Spark SQL, PostgreSQL, Pig Latin, and custom Map Reduce program in Java.
- Involved in building the ETL design for job and Source to Target mapping to load data into Data Warehouse.

- Experience in Application Development and Maintenance using RDBMS-Oracle (PL-SQL), Business Intelligence, ETL, UNIX/LINUS.
- Skilled in UNIX/LINUX for writing shell scripts and Batch processes.
- Created mapping documents to outline data flow and metadata information from sources to targets.
- Extracted the data from multiple databases into the staging area and populated into the Data warehouse.
- Experience with Hadoop and related systems including distributed systems and data management.
- Experience designing, reviewing, implementing, and optimizing data transformation processes in the Hadoop/Informatica ecosystems. Able to consolidate validate and cleanse data from a vast range of sources – from applications and databases to files and web services.
- Capable of extracting data from an existing database. Web sources or APIs. Experience designing and implementing fast and efficient data acquisition using big data processing techniques and tools.

Technical Skills:

Hadoop Ecosystem	Map Reduce, Hive, Impala, Oozie, PIG, HBase, HDFS, Spark ,AWS
Databases	SQL Server 2008/2012/2014, MySQL, Oracle 10g, 11g,12c,19c
ETL Application Package	Azure SQL Database, Azure SQL Data warehouse , Talend Big Data Studio 6.0/6.3/6.5, Informatica Power center (6.2/7.1/8.1/8.6/9.1/9.5/9.6.1) IDQ, IDE, Data stage (7.5/8.1) MS Office Suite, Visio.
Operating Systems	Windows Server 2003, 2007, 2010, 2013, Windows 8/NT/XP/Professional, UNIX, Linux
Languages	SQL, PL/SQL, XML, Java, MS Access, HTML, Unix shell, Python Scripting
Tools	SQL Server Enterprise Manager, SQL Server Management Studio,
Methodologies	Agile, Waterfall

Client	**Barclays**
Role	Senior Data Engineer
Duration	**April 2022 to Present (1yr, 1 mo)**
Location	Whippany, New Jersey, United States

Responsibilities:
- Leading a team of 4 members
- Involved in requirements gathering, analysis, design, development, change management, deployment.
- Involved in the development of real time streaming applications using PySpark, Apache Flink, Kafka, Hive on distributed Hadoop Cluster.
- Utilized Apache Spark with Python to develop and execute Big Data Analytics and Machine learning applications, executed machine Learning use cases under Spark ML and Mllib.
- Extracted data from heterogeneous sources and performed complex business logic on network data to normalize raw data which can be utilized by BI teams to detect anomalies.
- Designed and developed Flink pipelines to consume streaming data from kafka and applied business logic to massage and transform and serialize raw data.
- Worked on ingesting the data from Production environments to staging area to create and maintain Test Data Management servers in Delphix Engines which serve the TDM team
- Worked on designing the automation preparation process involving Test data governance Run books to implement the test data management strategies, Inventory Lists on TDM Refresh Process, DB repeatable tasks.
- Developed common Flink module for serializing and deserializing AVRO data by applying schema.
- Developed Spark streaming pipeline to batch real time data, detect anomalies by applying business logic and write the anomalies to Hbase table.
- Created and maintained the procedure documentation for DevOps-Test Data Management activities in secured SharePoint web application storage system.
- Implemented layered architecture for Hadoop to modularize design. Developed framework scripts to enable quick development. Designed reusable shell scripts for Hive, Sqoop, Flink and PIG jobs. Standardize error handling, logging and metadata management processes.

- Indexed processed data and created dashboards and alerts in splunk to be utilized/ action by support teams.
- Responsible for operations and support of Big data Analytis platform, Splunk and Tableau visualization.
- Managed, developed, and designed a dashboard control panel for customers and Administrators using Tableau, PostgresSQL and RESTAPI calls.
- As assigned, develops code, analyzes, researches, and resolves system issues and Testing of applications developed using AGILE methodology.
- Experience in building and architecting multiple Data pipelines, end to end ETL and ELT process for Data ingestion and transformation.
- Attending a daily meeting with the Customer to find out the exact requirements and provide the technical solution to meet the customer's requirements.
- Build Complex distributed systems involving huge amount data handling, collecting metrics building data pipeline, and Analytics.
- Worked on installing new relic kafka plugin implementation for monitoring of kafka cluster.
- Review within developed code with the team members to make sure that each requirement is captured correctly. Make sure that our coding is as per the client standard platform and some hook and crook method for performance optimization to make sure that code during the run performs well or is complete within a given time frame.
- Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL.
- Experience developing Scala applications for loading/streaming data into NoSQL databases (MongoDB) and HDFS.
- Performed data collection, data cleaning, data profiling, data visualization and report creating.
- Outguessed the data from HDFS to Azure SQL data warehouse by building ETL pipelines using S worked on various methods including data fusion, machine learning and improved the accuracy of distinguished the right rules from potential rules.
- Build data pipelines in airflow in GCP for ETL related jobs using different airflow operators.
- Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL.
- Pulling data from oracle and dumping on HDFS as Avro format file and then again converting Avro format to Parquet to resolve the performance issue. We are loading data into hive/impala as parquet files.
- Designed and Developed applications using Apache Spark, Scala, Python, Nifi, S3, AWS EMR on AWS cloud to format, cleanse, validate, create schema and build data stores on S3.
- Used Spark and Scala for developing machine learning algorithms which analyses click stream data.
- Created and managed Source to Target mapping documents for all Facts and Dimension tables
- Experience in open-source Kafka, zookeepers, Kafka connects.
- Designed and Implemented ETL for data load from SQL Server and Oracle as target databases and for Fact and Slowly Changing Dimensions SCD-Type1 and SCD-Type2 to capture the changes to the snowflake cloud database.
- Performance tuning - Using the map cache properties, and multi-threading for better performance in case of huge source data. Tuning the SQL source queries to restrict unwanted data in the ETL process.
- Use a batch job to create a spark job to load data on HDFS.
- Adept at handling Missing Data by exploring the causes like MAR, MCAR, MNAR and analyzing Correlations and similarities, introducing dummy variables and various Imputation methods.
- Experience in GCP Dataproc, GCS, Cloud functions, BigQuery.
- Develop automation and data collection frameworks
- Experience in Migrating data from Oracle to Data Lake using Sqoop, Spark, and ETL Tool.
- Hands-on Experience on use batch Jobs to create spark jobs to load data on HDFS.
- Created Kafka producer API to send live-stream data into various Kafka topics.
- Install, Config, Test, Monitor, upgrade and tune new and existing PostgreSQL databases.
- Extracted data from databases like Oracle, SQL server, and DB2 using to load it into a single repository for data analysis.
- Partitioning Dynamic Partitions Buckets of Hive.

- Have designed and developed ETL mapping for data collection from various data feeds using REST API. The data sources include feeds from mobile, social Facebook Query Language, You tube, Twitter, web and other partner feeds.
- Closely work with the DBA team and carry out performance tuning for the DB objects whose performance is below SLA definitions.
- Designed and Developed Scala workflows for data pull from cloud based systems and applying transformations on it.
- Experience with Informatica Data Quality 9.6.1(IDQ) toolkit, Analysis, data cleansing, data matching, data conversion, exception handling, and reporting and monitoring capabilities of IDQ 9.6.1
- Experience in moving data between GCP and Azure using Azure Data Factory.
- Worked on Machine Learning Algorithms Development for analyzing click stream data using Spark and Scala.
- Primarily involved in Data Migration using SQL, SQL Azure, Azure storage, and Azure Data Factory, SSIS, PowerShell.
- Production Migration & Release Management: Coordinate deployment of applications in UAT/Production environments. Additionally, provide application-level support for UAT and post-production support.

Client	Northwell Health
Role	Sr Data Engineer
Duration	Feb 2021 to Apr 2022 (1yr 3mo)
Location	New York, US

Responsibilities:
- Study and understand the scope of the assignment.
- Created and managed Source to Target mapping documents for all Facts and Dimension tables.
- Used ETL methodologies and best practices to create ETL jobs. Followed and enhanced programming and naming standards
- Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL, and U-SQL Azure Data Lake Analytics.
- Implement NiFi to Spark streaming directly withtout using Kafka internally to provide various options to client in single Confidential.
- Implemented test data solutions by tiles creation for synthetic data generation using the GT Data Maker tool and CA TDM portal to improve the services for business users/QA/DEV Users.
- Primarily involved in Data Migration using SQL, SQL Azure, Azure Storage, Azure Data Factory, SSIS, PowerShell
- Pulling data from oracle and dumping on HDFS as Avro format file and then again converting Avro format to Parquet to resolve the performance issue. We are loading data into hive/impala as parquet files.
- Migrating data from Oracle to Data Lake using Sqoop, Spark, and ETL Tool.
- Used cloud shell SDK in GCP to configure the services Data Proc, Storage, BigQuery
- Mapping source files and generating Target files in multiple formats like XML, Excel, CSV, etc.
- Implement ad-hoc analysis solutions using Azure Data Lake Analytics/Store, HDInsight
- Transform the data and reports retrieved from various sources and generate derived fields.
- Reviewed the design and requirements documents with architects and business analysts to finalize the design.
- Created the jobs to load the data into Snowflake and SQL server tables. Creating new batches that utilized parallelism better (Multi Instances concept in DataStage).
- Migrate the code from development to the production environment using Nexus, Scheduling job to run in production through Control-M.
- Provided the test data manifests and folders as part of TDM Refresh process to the Release Management team periodically to execute the PL/SQL scripts, FDM batch files (to mask PHI/PII data) and Informatica jobs using TFS CARA tool.
- Experience in using features such as context variables, triggers, and connectors for Database and flat files.
- Extensively worked on migrating all the client data from the rom AWS S3 bucket to Snowflake DB.
- Modifying existing Talend mappings to load to Snowflake DB.
- Recreating existing AWS objects in Snowflake.

- Partitioning, Dynamic Partitions, Buckets of Hive.
- Implement Hive UDFs doe evaluation, filtering, loading, and storing of data.
- Load data from a different source (database and files) into Hive using the ETL tool (Standard, Map Reduce, and Spark Job), monitor system health and logs and respond to any warning or failure conditions.

Client	Northwell Health
Role	Data Engineer
Duration	May 2019 to Feb 2021(1 yr 10 mo)
Location	New York, US

Responsibilities:

- Performed Data Integration, Extraction, Transformation, and Load (ETL) Processes
- In preprocessing phase, used Pandas to remove or replace all the missing data and balanced the dataset with Over-sampling the minority label class and Under-sampling the majority label class to perform a well Data Cleaning Process.
- Developed models relying on Linear Regression, Multiple Regression, Decision Trees, Random Forest, Logistic Regression, Naive Bayes.
- Validated and selected models using k-fold cross validation, confusion matrices and worked on optimizing models for high recall rate and Implemented Ensemble models like Boosting and Bagging
- Provided data mining solutions on Association Rules, Classification and Clustering problems
- Used Python to identify trends and relationships between different pieces of data and drew appropriate conclusions.
- Developed Rest API to serve the data generated by prediction model to serve other customers/teams.
- Generated complex reports in various formats such as list reports, summary reports etc. using advanced data manipulation techniques from SAS Enterprise Guide.
- Analyzed data and generated analytical reports using SAS, MS SQL.
- Created Visual Charts, Graphs, Maps, Area Maps, Dashboards and Storytelling using Tableau.
- Develop, Maintain and support Continuous Integration framework (CI-CD) based on Jenkins.
- Implemented, tuned and tested the model on AWS EC2 with the best performing algorithm and parameters.

Client	ROKT
Title	Data Engineer
Duration	May 2017 - Feb 2019 · 1 yr 10 mos
Location	New York, United States

Responsibilities:
- Developed database objects like Tables, Views, Indexes, Synonyms, and Sequences.
- Involved in the creation of tables, Join conditions, correlated subqueries, nested queries, views, sequences, and synonyms for business application development.
- PL/SQL coding for Exception Handling and duplicate data maintenance.
- Used Triggers in Customizing Error conditions.
- Developed/modified PL/SQL code like stored procedures, and packages for interfaces.
- Developed UNIX shell scripts and PL/SQL procedures to extract and load data for month-end batch processing.
- Creation of Forms, Pop-up Menus, and maintaining Master-Detail relationships.
- Worked with Oracle Data stage for ETL processes.
- Testing all forms, PL/SQL code for logic correction.
- Transform the data and reports retrieved from various sources and generate derived fields.
- Reviewed the design and requirements documents with architects and business analysts to finalize the design.

- Analyzed business functionalities related to back-end database interfaces and utilized PL/SQL developed tool to develop all back-end database interfaces.
- Developed triggers for error handling and keeping track of changes to the fields of tables when changes are made.
- Developed shell scripts, and PL/SQL procedures for creating/dropping tables and indexes of performance.
- Attending technical review meetings.
- Designed the reusable Data Processor Transformation using Informatica Developer Tool and integrating that into Informatica Power Center TDM owned repository to parse various types of file formats (JSON, XML, EXCEL, and FLAT FILES etc.).
- Implemented Star Schema for De-normalizing data for faster data retrieval for Online Systems.
- Involved in unit testing and system testing and preparing Unit Test Plan (UTP) and System Test Plan (STP) documents.
- Responsible for monitoring all the jobs that are scheduled, running completed, and failed. Involved in debugging the jobs that failed using a debugger to validate the jobs and gain troubleshooting information about data and error conditions.
- Performed metadata validation, reconciliation, and appropriate error handling in ETL processes.

Client	Synchrony
Role	Big Data & Hadoop Developer
Duration	Sep 2013 - Jul 2016 · 2 yrs 11 mos
Location	Hyderabad, Telangana, India

Responsibilities:
- Involved in designing, reviewing, implementing, testing, and releasing databases, sign specifications, and application software.
- Responsible for installation, gradation of different environments, creating and maintaining oracle development and test databases.
- Wrote and modified Oracle PL/SQL, SQL statements procedures, and functions.
- Developed a data model for the payout stream using Erwin as a data modeling tool.
- Responsible for users, roles, and privilege management.
- Created PL/SQL sub setting scripts to mask the PHI columns across various IT environments as part of TDM Database Refresh Process occurring every quarter based on the business
- Migrated data from production (AIG) to test the deferred products with the development engine.
- Responsible for the definition and management of schemas and objects.
- Supporting and advising developers during databases related issues.
- Rewrite SQL queries to improve the performance of the application.
- Implementing version control on an application system and freezing previous versions.
- Worked on applications using SQL Server. Worked closely in building T-SQL and using SSRS to generate reports.

Education:

Marist College, Poughkeepsie, NY Aug 2016 - Dec 2017
Masters in Information Systems (Major: Business Analytics) (GPA:3.98)

Gokaraju Rangaraju Institute of Engineering and Technology, India Aug 2009- May 2013
Bachelor of Technology in civil engineering (GPA:3.81)