

Swetha Adike

Sr. Data Engineer

Tel : (848) 216-5340

E-mail: swethaad26@gmail.com

LinkedIn ID: <https://www.linkedin.com/in/swetha-ad/>

Web: <https://swetha-adike.vercel.app/>

PROFESSIONAL SUMMARY

- Senior Data Engineer with 8+ years of professional IT experience with Big Data experience in **Hadoop** ecosystem components in **ingestion, Data modeling, querying, processing, storage, analysis, Data Integration** and Implementing enterprise level systems spanning **Big Data**.
- Worked on extensive migration of **Hadoop** and **Spark** Clusters to **AWS** and **Azure**.
- Experience in **BI Development, Analysis Datacenter Migration, Azure Data Factory (ADF) V2**. Managing Database, Azure Data Platform services (**Azure Data Lake (ADLS), Data Factory (ADF), Data Lake Analytics, Stream Analytics, Azure SQL DW, Azure Synapse, HDInsight/Databricks, NoSQL DB**), **SQL Server, Oracle, Data Warehouse** etc.
- Experience in installing, configuring, and using Apache Hadoop ecosystem components like **Hadoop Distributed File System (HDFS), MapReduce, Yarn, Spark, Nifi, Pig, Hive, Flume, Hbase, Oozie, Zookeeper, Sqoop, Scala**.
- Experience with **AWS: EC2, VPC, EBS, S3, IAM, Auto Scaling Group, opsWorks, AMI** (Amazon Machine Image), Cloud formation template, CloudTrail, SNS, Route 53, cloud Formation, CloudTrail, CloudWatch
- Build multiple **Data Lakes**. Build how the data will be received, validated, transformed and then published.
- Experience in implementing hybrid connectivity between cloud and on - premises using virtual networks, VPN and Express Route.
- Extensive experience in working with different **ETL tool** environments like **SSIS, Informatica and reporting tool** environments like **SQL Server Reporting Services (SSRS)**.
- Expertise in **SQL Server Analysis Services (SSAS), SQL Server Reporting Services (SSRS)** tools and in development of **T-SQL, Oracle PL/SQL Scripts, Stored Procedures and Triggers** for business logic implementation.
- **Azure Analysis Services (SSAS) and SQL Server Reporting Services (SSRS)** and Familiarity with **Excel PowerPivot, Power View, and PowerBI**
- Experience in working with different data sources like **Flat files, XML files, log files and Database**.
- Architect & implement medium to large scale BI solutions on Azure using **Azure Data Platform services (Azure Data Lake, Data Factory, Azure Synapse, Data Lake Analytics, Stream Analytics, Azure SQL DW, HDInsight/Databricks, NoSQL DB, Cosmos DB)**
- Developed **ETL\Hadoop** related **java** codes, created **RESTful APIs** using **Spring Boot** Framework, developed web apps using **Spring MVC and JavaScript**, developed **coding framework**, etc.
- Hands-on experience working in migration and build for Snowflake environment. Built Non-functional requirements for **Snowflake (CI/CD, monitoring, orchestration, data quality, monitoring)**.
- Experience in developing POC's using **Scala, Spark SQL and MLib** libraries then deployed on the **Yarn** cluster.
- Proficient working on **NoSQL** technologies like **HBase, Cassandra and MongoDB**.
- Designing and implementing highly performant data ingestion pipelines from multiple sources using **Apache Spark, Azure Databricks, Data verse, Azure cosmos DB**.
- Experience in creating interactive **Dashboards** and Creative **Visualizations** using tools like **Tableau, Power BI using SSIS, Power BI Desktop, Power BI Services, M Language) Interactions, DAX**.
- Expert on maintaining and managing **Tableau** and **POWER BI-driven reports and dashboards**.
- Created various types of charts like **Heat Maps, Geocoding, Symbol Maps, Pie Charts, Bar Charts, Tree Maps, Gantt, Circle Views, Line Charts, Area Charts, Scatter Plots, Bullet Graphs and Histograms** in **Tableau Desktop, Power BI, and Excel** to provide better **data visualization**.
- Extensive skills on **LINUX** and **UNIX** Shell command.
- Experience in **Performance Tuning and Optimization (PTO)**, **Microsoft Hyper-V virtual infrastructure**, including **Performance Tuning of SQL's** by implementing the views, Clustered Column Store indexes, partitions, and aggregate tables. Develop **Dimensional Model** using **Kimball Model** and work on **Source/Target Data mapping**.
- Experience with scripting languages (**Scala, Pig, Python and Shell**) to manipulate data.
- Excellent working experience in **Scrum / Agile framework** and **Waterfall** project execution methodologies.

TECHNICAL SKILLS

- **Big Data Eco-system:** HDFS, MapReduce, Spark, Yarn, Hive, Pig, HBase, Sqoop, Flume, Kafka, Oozie, ZooKeeper, Airflow, Impala
- **Hadoop Technologies:** Apache Hadoop 1.x, Apache Hadoop 2.x, Cloudera CDH4/CDH5, Hortonworks
- **ETL Tools:** Azure Data Factory (ADF), Azure Database migration Service (DMS), ETL SQL Server Integration Services (SSIS), SQL Server Reporting Services (SSRS), ETL Extract Transformation and Load., Business Intelligence (BI).
- **Storage:** Azure Storage, Azure Blob Storage, Azure Backup, Azure Storage Disks- Premium, Azure Files, Azure Data Lake Storage Gen 1, AWS S3 Cloud Storage, AWS EBS, AWS Storage gateway, NETApp/EMC Storage Technologies, SAN
- **Cloud Computing:** Microsoft Azure Cloud technologies, Snowflake, Azure Backup AWS (Amazon web Services) Data Modeling logical and physical Data Modeling /Database design using Erwin and Visio etc.
- **Databases:** Azure SQL Database, AZURE data warehouse, Azure Data azure data factory, Azure Data Synch, Elastic Pools, SQL Server 2017, Oracle EBS R12, Oracle EBS 11i, SQL Server 2016, SQL Server 2014, SQL Server 2012, 2008 R2, SQL 2005, 2000, Azure SQL Database, RDBMS, Azure Synapse, Microsoft Azure VM, Business Intelligence (BI), Amazon Web Service (AWS), AWS Cloud Watch, Oracle 11g, MYSQL 5.x, MS Access, OLAP, OLTP, AWS RDS, AWS Cloud, Microsoft Azure Cloud, Cosmos DB, Mongo DB.
- **Scripting:** Windows Power Shell, Shell Script, AZURE CLI, Transact SQL, AWS CLI, UNIX, Python, Pyspark, SQL, Hive, Spark
- **Machine Learning:** Regression, Decision Tree, Clustering, Random Forest, Classification, SVM, NLP
- **Operating Systems:** Azure Virtual Machine (VM), Windows Server 2016, Windows Server 2012, Windows 2008 R2, /2000 Server(64bit), Windows 2000 Advanced Server, Windows XP, Unix, LINUX. Shell Script, AWS EC2
- **Container/Cluster Managers:** Docker, Kubernetes
- **BI Tool:** Tableau, Power BI
- **Web Development:** HTML, XML, CSS
- **IDE Tools:** Eclipse, Jupyter, Anaconda, PyCharm
- **Development Methodologies:** Agile, Waterfall

PROFESSIONAL EXPERIENCE

Client – Barclays, Whippany, New Jersey, United States
Sr. Data Engineer

Apr 2022 - Present

Responsibilities:

- Provided End to end **business intelligence** solutions by using Microsoft technologies using **Azure Data Lake, Azure Databricks, Azure Data Factory, Azure SQL data warehouse, and Azure Synapse.**
- **Extract Transform and Loaded 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.** Data Ingestion to one or more **Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW)** and processing the data in **In Azure Databricks.**
- Used **Azure Data Factory** or **Azure Databricks** to create **ETL workflows** that integrate with **PostgreSQL** instances.
- Optimized SQL queries that **retrieve, insert, update, or delete data in PostgreSQL** databases, using **Azure Data Studio** to write and debug **SQL** queries.
- The idea behind **Data Mesh** is to treat data as a product, with each domain or business unit responsible for its own data products, rather than relying on a central team to manage all data for the organization. This approach aims to improve data quality, ownership, and accountability, as well as to enable better collaboration between data producers and consumers.
- Created **CI/CD Pipelines** in **ADF** using **Linked Services/Datasets/Pipeline/** to **ETL** data from different sources like **Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool and backwards.**
- Integrated snowflake with a variety of **Azure** and **AWS** services, such as **Azure Data Factory** or **AWS Glue**, to enable **data ingestion** and **processing.**
- Used snowflake for data validation, data profiling, and data lineage to ensure the quality of the data.
- Developed application code that interacts with the **PostgreSQL database** and used a **PostgreSQL** driver to connect to the database and execute **SQL queries.**

- Developed **Azure Synapse Notebooks** using **Pyspark** and **Spark-SQL** for data extraction, transformation, and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
- Used **Oracle Database** in **Azure** to run mission-critical **Oracle workloads**, migrate **Oracle databases to Azure**, and integrate with other **Azure services** like **Azure Active Directory**, **Azure Monitor**, and **Azure Backup**.
- Created scripts as per **snowflake** supported datatypes and worked on **snowflake stored procedures framework**.
- Used **AWS** services as a back end for **Azure** application, such as using **AWS S3** for storage or **AWS Lambda** for serverless computing.
- Developed **ETL Processes** in **AWS Glue** to migrate data from external sources like **S3**, **ORC/Parquet/Text Files** into **AWS Redshift**.
- Worked on Ingesting data by going through cleansing and transformations and leveraging **AWS Lambda**, **AWS Glue** and **Step Functions**.
- Implemented Installation and configuration of multi-node cluster on **Cloud** using **Amazon Web Services (AWS) on EC2**.
- Responsible for developing data pipeline with **Amazon AWS** to extract the data from weblogs and store in **HDFS**.
- Used **Python** to develop **ETL pipelines** into **Azure data ecosystem** with libraries such as **Spark** or **Azure Databricks** to develop **ETL** workflows that handle data extraction, data transformation, and data loading into **Azure data** stores such as **Azure SQL Database**, **Azure Data Lake Storage**, or **Azure Synapse Analytics**.
- Ran **Oracle E-Business Suite**, **PeopleSoft**, **JD Edwards**, and **Siebel applications** on **Azure**, leveraging **Azure services** like **Azure Virtual Machines**, **Azure Blob Storage**, and **Azure Active Directory** to enable modernized **oracle** applications.
- Implemented **Python** framework to ingest the data into **snowflake**.
- Implemented **Data Sharing** and **Data cloning** for **Snowflake code**.
- Responsible for estimating the **cluster** size, and monitoring and troubleshooting the **Spark databricks cluster**.
- Experienced in performance tuning of **Spark Applications** for setting right Batch Interval time, correct level of Parallelism and memory tuning.
- To meet specific business requirements wrote **UDFs** in **Scala** and **Pyspark**.
- Used **Python libraries** and frameworks such as **Pandas**, **NumPy** to clean, filter, and aggregate data, and prepare it for further analysis or storage.
- Developed **JSON Scripts** for deploying the Pipeline in **Azure Data Factory (ADF)** that process the data using the **SQL Activity**.
- Hands-on experience in developing **SQL Scripts** for automation purpose.
- Stored data in **PCA** tables in **Azure SQL Database** or **Azure Synapse Analytics** and ensured that **PCA** tables containing sensitive data are encrypted at rest and in transit, using **Azure Disk Encryption**, **Azure Storage Service Encryption**, and **Azure Key Vault**.
- Created Build and Release for multiple projects (modules) in production environment using **Visual Studio Team Services (VSTS)**.
- Collaborated with users **on reports and dashboards utilizing workspace apps and content packs in Power BI**.
- **Created different visualization in the reports using custom visuals** like Bar Charts, Pie Charts, Line Charts, Cards, Slicers, etc. Also, using different transformation inside edit query into clean-up the data using **Power BI**.
- Created and managed data models using **Snowflake**.
- Performed data messaging operations using **DAX expressions** code in **Power BI**.
- Wrote complex **DAX expressions** to generate calculated columns and filters for role-based security.
- Utilized **Power BI** gateway to keep **dashboards and reports** up to-date with on premise data sources.
- Used **Python** to integrate with various **Azure services** such as **Azure Storage**, **Azure Event Hubs**, and **Azure Machine Learning**.
- Created hierarchies in **Power BI reports** using data visualizations such as **bar chart**, **line chart**, **pie charts** and **forecast charts**.

Environment: Azure Data Factory, Azure Synapse, Azure Databricks, Azure Lake Gen2, Azure SQL database, AWS, Principal component Analysis (PCA), Snowflake, PostgreSQL, Hadoop, Oracle, Map Reduce, Yarn, Hive, Python, Spark, Scala, SSMS, Horton Works, Data Lake, Databricks, Power BI Desktop, Power BI query, DAX.

Responsibilities:

- Involved in Requirement gathering, Business Analysis and translated business requirements into technical design in **Hadoop** and **Big Data**.
- Involved in **SQOOP** implementation which helps in loading data from various **RDBMS** sources to **Hadoop** systems.
- Developed **Python** scripts to extract the data from the web server output files to load into **HDFS**.
- Involved in **HBASE** setup and storing data into **HBASE**, which will be used for further analysis.
- Worked on Written a python script which automates to launch the **EMR** cluster and configures the **Hadoop** applications.
- Created various data pipelines using **PySpark**, **Scala** and **SparkSQL** for faster processing of data.
- Written **Spark-SQL** and embedded the **SQL** in **SCALA** files to generate jar files for submission onto the **Hadoop** cluster.
- Worked with **Avro** and **Parquet**, **ORC**, **XML**, **JSON** files and converted the data from either format Parsed Semi Structured **JSON** data or converted to **Parquet** using Data Frames in **PySpark**.
- Developed a **Python Script** to load the CSV files into the **S3 buckets** and created **AWS S3 buckets**, performed folder management in each bucket, managed logs, and objects within each bucket.
- Involved in Analyzing system failures, identifying root causes, and recommended course of actions, Documented the systems processes and procedures for future references.
- Experience in bulk loading and continuous ingestion of data using **COPY** and **Snow pipe** respectively.
- Involved in **Hadoop** installation, Commissioning, Decommissioning, Balancing, Troubleshooting, Monitoring and, debugging Configuration of multiple nodes using Hortonworks platform.
- Involved in working with **Spark** on top of **Yarn** for interactive and Batch Analysis
- Worked with a team to migrate from **Legacy/On prem** environment into **AWS**.
- Handled data modeling in **snowflake** and **Integrated Snowflake** with **AWS Lambda** and Step functions.
- Writing stored procedures in **snowflake** using **JavaScript/SQL**.
- Created **Dockized backend cloud applications** with exposed **Application Program Interface (API)** interfaces and deployed on **Kubernetes**.
- Experienced in analyzing and **Optimizing RDD's** by controlling partitions for the given data
- Experienced in writing live Real-time Processing using **Spark Streaming** with **Kafka**, **NiFi**.
- Used **HiveQL** to analyze the partitioned and bucketed data and compute various metrics for reporting.
- Worked with querying data using **SparkSQL** on top of **Spark** engine.
- Involved in managing and monitoring **Hadoop** cluster using **Cloudera** Manager.
- Developed data pipeline using **SQOOP**, **HQL**, **Spark** and **Kafka** to ingest Enterprise message delivery data into **HDFS**.
- Design and Develop **ETL** Processes in **AWS Glue** to migrate Campaign data from external sources like **S3**, **ORC/Parquet/Text Files** into **AWS Redshift**.
- Assisted in Cluster maintenance, cluster monitoring, adding, and removing cluster nodes and installed and configured **Hadoop**, **Map Reduce**, **HDFS**, developed multiple **Map Reduce** jobs in java for data cleaning and pre-processing.
- Automated **Snowpipe for Amazon S3**.
- Develop an understanding of **IT/Agile** processes.
- Build and transform large datasets for analytical use and synthesize data into products fit for analytics projects.
- Work directly with the **Data Governance** team to determine best practices for landing sensitive data
- Involved in file movements between **HDFS** and **AWS S3** and extensively worked with **S3 bucket in AWS**.
- Automated and monitored complete **AWS** infrastructure with terraform.
- Converted all **Hadoop** jobs to run in **EMR** by configuring the cluster according to the data size.
- Connected multiple data sources in **Tableau** and created reports using **Tableau** based on the client's needs.
- Created ad-hoc reports to users in **Tableau D-esktop** by connecting various **data sources**, **multiple views**, and **associated reports**.

Environment: **HDFS**, **Hive**, **Scala**, **Sqoop**, **Spark**, **Tableau**, **Yarn**, **Cloudera**, **SQL**, **Terraform**, **Splunk**, **RDBMS**, **Elastic search**, **Kerberos**, **Jira**, **Confluence**, **Shell/Perl Scripting**, **Zookeeper**, **AWS (EC2, S3, EMR, Redshift, ECS, Glue, S3, VPC, RDS etc.)**, **Snowflake**, **Snow pipe**, **Informatica**, **Git**, **Kafka**, **CI/CD(Jenkins)**, **Agile**, **Azure Databricks**, **Tableau**, **Java**.

Responsibilities:

- Writing Scope script to read and load the data into **Cosmos** portal.
- Extensive working experience in data warehousing and Business Intelligence Technologies with expertise in **SQL Server development, MSBI stack (TSQL, SSIS, SSAS, SSRS), Azure, Power BI for building, deploying, and managing applications** and services through a global network of Microsoft - managed data centers.
- Worked with business stakeholders and data analysts to define the **master data model (MDM)** for the project, involves identifying the key data entities and attributes, and defining relationships and hierarchies between them.
- Analyze, design, and build Modern data solutions using **Azure PaaS service** to support visualization of data. Understand current Production state of application and determine the impact of new implementation on existing business processes.
- 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. Data Ingestion** to one or more **Azure Services - (Azure Data Lake, Azure Storage, Azure Synapse, Azure SQL, Azure DW)** and processing the data in **Azure Databricks**.
- Created **Pipelines** in ADF using **Linked Services/Datasets/Pipeline/** to **Extract, Transform, and load data** from different sources like **Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool and backwards**.
- Developed **Spark** applications using **Pyspark** and **Spark-SQL** for **data extraction, transformation, and aggregation** from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
- Responsible for estimating the **cluster** size, monitoring, and troubleshooting of the **Spark data bricks cluster**.
- Experienced in **performance tuning** of **Spark Applications** for setting right Batch Interval time, correct level of Parallelism and memory tuning.
- To meet specific business requirements wrote **UDF's** in **Scala** and **Pyspark**.
- Developed **JSON** Scripts for deploying the **Pipeline** in **Azure Data Factory (ADF)** that process the data using the **SQL** Activity.
- Architected Cubes with **PowerPivot, MDX, SSAS**.
- Hands-on experience on developing **SQL Scripts** for automation purpose.
- Created **POWER BI Visualizations** and **Dashboards** as per the requirements
- **SQL Server Analysis Services (SSAS) and SQL Server Reporting Services (SSRS)** and Familiarity with **Excel PowerPivot, Power View, and PowerBI**.
- Created Build and Release for multiple projects (modules) in production environment using **Visual Studio Team Services (VSTS)**.

Environment: Azure Data Lake, Azure Storage, Azure Synapse, Azure SQL, Azure DW, Pyspark, SQL, MDM, ADF, T-SQL, Power BI Desktop, Power Bi Pro, SQL, SSIS, SSRS.

Responsibilities:

- Analyzed and reported customer using data transactional and analytical data to meet business objectives
- Worked on the entire **CRISP-DM** life cycle and actively involved in all the phases of project life cycle including data acquisition, data cleaning, data engineering.
- Developed weekly, monthly reports related to the **marketing and financial departments** using **Teradata SQL**.
- Designed high level **ETL** architecture for overall data transfer from the **OLTP** to **OLAP** with the help of **SSIS**.
- Extracted data from **SQL** Server using Talend to load it into a single data warehouse repository. Designed **tabular, matrix reports, drilldown, drill through, Parameterized and linked reports in SSRS**.
- Wrote **SQL** queries using **joins, grouping, nested sub-queries, and aggregation** depending on data needed from various relational customer databases.
- Created **V-Look Up** functions in **MS Excel** for searching data in large spreadsheets.
- Developed ad-hoc reports with **V-lookups, Pivot tables, and Macros** in **Excel** and recommended solutions to drive business decision making.

- Designed data models and data flow diagrams using **Erwin and MSVisio**.
- Involved in Trouble Shooting, Performance tuning of reports, resolving issues within **Tableau Server & Reports**.
- Wrote a complex **SQL, PL/SQL, Procedures, Functions, and Packages** to **validate data and testing process**.
- Done **SAS** programming such as **Proc Sql (Join/ Union), Proc Append, Proc Datasets, And Proc Transpose**.
- Created worksheets reports and converted into interactive dashboards by using **Tableau Desktop** and provided to **Business Users, Project Managers and End Users**.
- Designed, developed, and tested various **Tableau visualizations** for **dashboard** and **ad-hoc** reporting solutions by connecting from different data sources and databases.
- Responsible for submitting the change requests related to **Crystal Reports and Scripts** changes in Production
- Worked on requests to change the distribution list in **Schedule Reports**.
- Designed and documented the **test cases** to ensure data quality and performed thorough validation.
- Responsible to create test data for Reports during **UAT Testing**.
- Developed several scripts and **SQL** queries to create various reports for administration.
- Participated in **stakeholder** and project review meetings on weekly basis.

Environment: Crystal Reports, DB2 SQL, SQL Server 2008, SQL, Tableau, SSIS, SSRS.

Client – Synchrony Technologies, Hyderabad, India
Big Data & Hadoop Developer

Sep 2013 – Jul 2016

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
Masters in Information Systems (Major: Business Analytics) (GPA:3.98)

Aug 2016 - Dec 2017

Gokaraju Rangaraju Institute of Engineering and Technology, India
Bachelor of Technology in civil engineering (GPA:3.81)

Aug 2009- May 2013