SWATHIREDDY BOBBALA

swathireddy112@gmail.com
Ph. No. : (+1 469-769-8727)

SUMMARY:

- Around 10 years of expertise in Analysis, Design, and Development of data platform on on-prem and Azure Cloud with high-quality Data Modelling, design, and development of Data Pipelines to ingest, store and transform data for data analytics and systems integration.
- Experience in migration of existing solutions on On-Premise systems/applications to Azure cloud.
- Worked on Azure Suite: Azure SQL Database, Azure DataBricks (ADB), Azure Data Lake (ADLS), Azure Data Factory (ADF) V2, Azure SQL Data Warehouse.
- Well versed experienced in creating pipelines in Azure Cloud ADFv2 using different activities like Data Bricks, Copy, Filter, for Each, Move &Transform etc.
- Develop prototypes using Azure Cloud services such as Azure Data factory (ADF), Logic Apps, Virtual machines, Key vaults, Azure SQL DB, Kubernetes services and Azure Data Lake storage (ADLS) to validate proposed solutions and seek feedback from stakeholders.
- Designing and Developing Azure Data Factory (ADF) pipelines to extract the data from Relational sources such as Teradata, Oracle, SQL Server, Sybase, Postgres, DB2 and non-relational sources like Flat files.
- Good Understanding of Azure Cosmos DB, Azure functions, Azure DevOps, and Blob storage.
- Involved in Data Migration from MS SQL server, Oracle, My SQL, and Postgre SQL to Teradata, Hadoop, and Cloud Lakes.
- Highly experienced in using Multi Cloud Services to achieve maximum efficiency and performance for data application, used Azure platform for building extensive data pipelines for doing terra scale of data analysis.
- Extract data from on-premise and cloud storages and Load data to Azure Data lake from On-Premise Databases, Azure SQL Databases, ADLS Gen 1, ADLS Gen 2, Azure BLOB Storage using Both ETL and ELT Fashion using Azure Data Factory (ADF) and Azure DataBricks (ADB) and PolyBase feature in SQL 2016 using External Tables.
- Moved SSIS packages from On-premise to cloud using Azure Data Factory (ADF) and Azure DataBricks (ADB).
- Well versed in creating 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.
- Created Dataframes in Databricks and applied various transformations like string functions, aggregations, window functions, Filtering, Splitting, Renaming, Removing duplicates etc
- Excellent communication skills with excellent work ethics and a proactive team player with a positive attitude. Domain Knowledge of Finance, Logistics and Health insurance.
- Strong and effective problem-solving, analytical and interpersonal skills, besides being a valuable team player.
- Excellent communication skills. Was the key member in all the projects in getting the requirements from the Client/Product Owner in Agile projects. Guided the offshore team

and monitored the offshore deliverables working from Onsite locations and also geologically different offshore locations.
- Involved in all the phases of Software Development Life Cycle (SDLC): Requirements gathering, analysis, design, development, testing, production and post-production support.

Education:
Bachelor of Computer Science Engineering - JNTU, INDIA (2004-2008).
Masters in Advanced Data Analytics – University of North Texas, USA(2023 – Present)

Technical Skills:
- Azure Cloud Services - Data Factory, Data Bricks, Synapse, Logic Apps, Azure Functions, Azure Sql, Hadoop/Big Data, HDFS, Map Reduce, Apache Spark, Apache kafka, Hive, Pig, Sqoop, Flume, Oozie, ZooKeeper, HBase, BMC ControlM, Machine Learning,
- Languages: Java, C/C++, Scala, Shell scripting, Python, Sql,
- Teradata, MS SQL Server, Oracle, PL/SQL, Informix, Sybase, Informatica, Datastage.
- JAVA, J2EE, Spring, Hibernate EJB, Webservices, Servlets, JSP, Jakarta Struts.
- JBoss, Tomcat.
- UML, OOAD.
- HTML, AJAX, CSS, XHTML, XML, XSL, XSLT, WSDL.
- Junit, MRUnit, Ant, Maven, Log4j, FrontPage.
- Eclipse, NetBeans.
- Linux, UNIX, Windows.

PROFESSIONAL EXPERIENCE:

Technology Lead
August 2020 to December 2022
Infosys Limited
Client: TMobile USA
Work Location: Vancouver, Canada and Hyderabad, India

Project Description: Worked for a project, migrating the legacy data warehouse which is On-Premise Teradata to Azure Cloud and bringing all the real time data of Prepaid, Postpaid, Financial, Insurance and Campaigning data from different sources and performing rigorous data cleansing, processing and re-engineering using Azure data factory, Azure Databricks and Azure Synapse dataware house and making the tables and views available for reporting.

Responsibilities:
- Working as a team lead, creating the new Data Warehousing solution on Azure cloud, for the migration of on-premise Dataware house to Cloud
- Worked directly with management to understand the requirement and proposed and developed best business solution that enabled effective decision-making, and drive business objectives
- Built conceptual and logical data models for stakeholders and management
- Created technical Design Specification documentation that clearly articulates the design and code being implemented.
- Designed and coded high quality database solutions within a fast paced sprint release cycle

- Created Azure data factory pipelines for migration of data from On-premise to ADLS gen2 layer
- Created Linked services and different datasets to connect to ADLS, databricks and Synapse using different authentication mechanisms.
- Implemented many new ADF Mapping Data Flow Components like New Branch, Join, Conditional Split, Exists, Union, Lookup, Derived Column, Surrogate Key, Window, Filter, Sort, Alter Row, For Each loop, Upsert which made ADF Code Free and well compatible with SSIS Transformations.
- The ADF Pipelines are made Dynamic with Parameters, Expressions, Functions and Column Patterns.
- The ADF Pipelines are Scheduled using Triggers and Monitored the Performance and Execution In Monitoring Logs and Dashboards.
- Created DataBrick Notebooks using Python, Scala and Spark SQL to read and write JSON, CSV and Parquet files
- Developed Pyspark scripts for transformations on data and re-engineered the Teradata MSS scripts on Azure Databricks.
- Created stored procedures with complex sql on Azure Synapse
- Created tables in Synapse with different constraints, different distribution techniques like hashing, roundrobin and replicated to improve the query performance
- Created several dimension, fact and snapshot tables and implemented SCD2 logic
- Involved in surrogate key generation for dimension tables based on the natural keys
- Worked with business teams to understand the business functionality and created views which are made available to reporting
- Analyzed and validated data accuracy of report results
- Deployed the databricks and synapse scripts/code to different environments using Azure devops CICD pipelines
- Used various complex data types like Structs, Arrays, Maps, Lists in Pyspark to handle complex data and flattened them using Split and explode
- Optimized Databricks notebooks tuning cluster configurations, Repartitioning, join types, Shuffling configurations, Memory settings and caching.
- Involved in the design process to improve on the prototype given by the architect design team.
- Develop External tables in Azure Synapse Analytics (SQL Data Warehouse) for data visualization and reporting purposes.
- Developed stored procedures to implement the complex business transformation in Azure Synapse Analytics.
- Create tasks in Azure DevOps (ADO) to track workload and deploy objects across environments.

Technology Analyst
February 2018 to July 2020
Infosys Limited
Client: British Telecom, UK
Work Location: Wales, UK and Hyderabad, India

Project Description:

Worked for Supply Chain project – Ingested the data from different vendors, both for current device deals and future device deals, from a network file share data source to Azure Synapse. Built a common unified data model for supply chain deals.
Business units across TMobile consume Assets/Devices data for a variety of use cases via the Enterprise Data Warehouse (EDW). The heaviest users of this data access it via the Finance Data Mart (FDM), which summarizes many disparate data sources into a more end user friendly format. The organization can now continue its purchase of financial assets from prospective vendors, which supports the enterprise goal of strengthening the balance sheet.

Responsibilities:

- Interacted with business stakeholders to understand business problems and requirements.
- Was responsible for converting business requirements to technical requirements.
- Was the only developer to design and develop end to end solutions for a common unified data model for BECU Loan deals.
- Created Infrastructure required and related key vaults, linked services, data sets, arm templates for multiple source systems and Azure services.
- Created Pipeline's to extract data from on premises source systems to Azure cloud data lake storage.
- Worked on copy activities and implemented the copy behavior's such as flatten hierarchy, preserve hierarchy and
- Merge hierarchy
- Exposure on Azure Data Factory activities such as Lookups, Stored procedures, if condition, for each, Set Variable, Append Variable, Get Metadata, Filter and wait.
- Configured the logic apps to handle email notification to the end users and key shareholders with the help of web services activities.
- Create dynamic pipelines to handle multiple source extracting to multiple target systems
- Extensively used Azure key vaults to configure the connections in linked services.
- Configured and implemented the Azure Data Factory Triggers and scheduled the Pipelines, monitored the scheduled Azure Data Factory pipelines and configured the alerts to get notification of failure pipelines.
- Implemented SCD-1, SCD-2 in Synapse for incremental data loads
- Implemented delta logic extractions for various sources with the help of control table
- Implemented the right distribution techniques for Synapse tables to reduce the latency time
- Implemented the Data Frameworks to handle the deadlocks, recovery, logging the data of pipelines.
- Deployed the codes to multiple environments with the help of CI/CD process and worked on code defect during the SIT and UAT testing and provide supports to data loads for testing; Implemented reusable components to reduce manual interventions
- Created Triggers, PowerShell scripts and the parameter JSON files for the deployments
- Reviewing work on ingesting data into azure data lake and provide feedbacks based on reference
- architecture, naming conventions, guidelines and best practices
- Implemented End-End logging frameworks for Data factory pipelines.

Environment: Azure Data Factory, Azure Data Lakes Gen1/Gen2, Visual Studio Code, Azure Synapse, Azure Dev-ops, Power BI, Azure Logic Apps and Azure Cloud Services, Azure functions Apps, Azure Monitoring, Key Vault.

Senior Software Developer
February 2016 to April 2017
Client: IT-Brianiac/Infosys/T-Mobile
Work Location: Bellevue, Washington, USA.

Project Description:
T-Mobile has a Hadoop cluster of 270-nodes that is populated with over 2 petabytes of customer transaction data, sales data. Used Hadoop to move all data to a common datalake in order to achieve a single point of data source about their customers, Products and their services. IDW (Integrated Data Warehouse) project is an effort to acquire data from multiple internal and external systems and ingest into Hadoop data lake – Hortonworks platform using Hadoop suite of tools.

Responsibilities:

- Worked on a 270 node Hadoop cluster using HortonWorks Distribution.
- Built real time pipelines for streaming data using Apache Kafka and Spark Structured Streaming.
- Created Input streams which allows Spark Structured Streaming to collect data from external source, Apache Kafka in near-real-time and performed necessary transformations and aggregation on the fly to build the common learner data model and persisted the data in Hadoop cluster.
- Created multiple Dataframes and Datasets by applying several transformations for input dataframes.
- Added new columns, renamed existing columns and dropped few columns for the datasets as per the project requirements using dataframe transformations.
- Created schemas for Datasets using Scala's case classes.
- Created Spark's managed and unmanaged tables and Spark UDF's.
- Applied different kinds of joins(Shuffle joins and Broadcast joins) available in Spark for datasets.
- Created Distributed shared variables wherever necessary
- Result datasets are written to external files in HDFS using different output modes.
- Performed performance tuning for Spark Streaming e.g. configuring degree of Parallelism, selection of correct Serialization format & by efficient memory management.
- Developed all the Spark use cases in Scala using functional and imperative programming paradigms. Created various named functions, anonymous function, traits in Scala for data transformations.
- Deployed Spark applications on YARN cluster manager.
- Involved in monitoring Spark jobs on Spark UI and debugging spark applications
- Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Spark Sql.
- Used Pig and Hive to do transformations, joins, filter traffic and some pre-aggregations before storing the data onto HDFS for legacy data.
- Created huge number of transactional and incremental hive tables to bring in history data and the current data from Oracle, Sql, Teradata databases.
- Involved in developing various shell and python scripts to invoke Oozie workflow and to perform various other transformations and cleanups on Hadoop edge node and on the cluster

- Involved in all phases of the project from Analysis, requirement gathering from external vendors/source teams until post-production support
- Involved in peer to peer reviews of the code
- Involved in deploying all the Spark application jobs on Hadoop Yarn cluster.
- Worked in a DevOps environment. Was responsible in pushing the source code into GitHub repository and used Jenkins to build rpm's and promoted the code to Production using Nexus.
- Worked on CntrlM scheduler tool to schedule jobs on a daily basis or as per requirement.
- Involved in monitoring the Hadoop cluster using Ambari.

Environment: Hadoop-2.4.2.0, Hive-1.2.1, Pig-0.15.1, Map Reduce, Sqoop-1.4.3, Zookeeper-3.4.5, Oozie-3.3.2, Spark 2.4, Scala 2.12, Sbt 1.1.1, Kafka 0.10, MySQL, DB2, Teradata, HBase, Linux, Eclipse Juno, JDK 1.8.

Project Engineer
December 2009 to December 2011
Client: Wipro Technologies/Microsoft
Location: Hyderabad, India

Project Description: The project in the process of upgrading the existing software system to cater to the Business requirement decisions

Responsibilities:

- Developed ETL best practices and standards.
- Design of source to target mapping (STM) document.
- Worked extensively on Source Analyzer, Mapping Designer, Target Designer, Workflow Manager and Workflow Monitor.
- Used various Transformations like Joiner, Aggregate, Java, Expression, Lookup, Filter, Union, Update Strategy, Stored Procedures, and Router etc. to implement the business logic.
- Created Complex mappings using Connected and Unconnected Lookups, Aggregate, Update Strategy, Stored Procedure and Router transformations for populating target table in efficient manner.
- Created Informatica mappings with PLSQL Procedures, Functions to build business rules to load data.
- Developed SCD Type1 and SCD Type2 Mappings to track the Change Data Capture (CDC).
- Prepared SDLC Document and conducted walk through while moving from Development to Test and Test to Integration Test environments and so forth.
- Co-ordinate with testing team in providing explanations and resolutions to the observations and defects raised by the testers.
- Involved in Performance Tuning of SQL Queries, Sources, Targets and sessions by identifying and rectifying performance bottlenecks.
- Creating job setup document for job scheduling tool.
- Worked on shell scripts used in scheduling Informatica pre/post session operations.
- Implemented different Tasks in workflows which included Session, Command, E-mail, Event-Wait etc.
- Migrated the Code from Informatica Power center 8.6.1 to 9.1.0.
- Involved in Folder Migrations from one environment to the other environments.

- Performed extensive Unit Testing on the developed Mappings and was also involved in the documentation of Test Plans and testing with the users (UAT).
- Extracted data from Flat files and Oracle and loaded them into Teradata.

Environment:
Informtica Power Center 9.x, SQL SERVER 2000, Oracle 10g, TOAD, AUTOSYS, Business Objects XI, Teradata.

Project Engineer
August 2008 to November 2009
Client: Wipro Technologies
Location: Bangalore, India

Project Description: This project was aimed at building a Data warehouse for the Sales & Marketing team for the Sales Analysis of their Products. OWB was used to transform a variety of Oracle relational data, transaction files from different product platforms into standardized data mart. The data mart has been designed using Erwin adopting Star Schema methodology. Models created based on the dimensions, levels and measures required for analysis.

Responsibilities:

- Extensively involved in extraction of data from Oracle, Flat files.
- Design of ETL process using Informatica.
- Developed various Mappings using Source Qualifier, Aggregator, Joiners, Lookups, Filters, Router and Update strategy.
- Extensively worked with joiner functions like normal join, full outer join, master outer join, and detail outer join in the joiner transformation.
- Used Update Strategy DD_INSERT, DD_DELETE, DD_UPDATE, AND DD_REJECT to insert, delete, update and reject the items based on the requirement.
- Extensively worked with aggregate functions like Avg, Min, Max, First, Last, and Count in the Aggregator Transformation.
- Extensively used SQL Override, Sorter, and Filter in the Source Qualifier Transformation.
- Extensively used Mapping Variables, Mapping Parameters, and Parameter Files for the capturing delta loads.
- Worked with various tasks like Session, E-Mail, Workflows, and Command.
- Optimized various Mappings, Mapplets, Sessions, Sources and Target Databases Unit testing performed on Mappings.
- Developed simple & complex mappings using Informatica to load Dimension & Fact tables as per STAR Schema techniques.
- Extensively used various transformations to load data into slowly changing dimensions (SCD).
- Code review for other developers and prepared Production Release sheet.
- Generated the standard reports – daily, weekly & monthly in excel format.

Environment:
Informatica 7, SQL, PL/SQL, Oracle 8i and Flat files.

Java Developer Internship
April 2007 to June 2008

Wipro Technologies
Location: Hyderabad, India

Project Description:
As a part of intern worked on real-time project. Developed a search engine portal for booking International and Domestic Airline tickets, Hotels, Cruise, Buses and Cars. The portal has been designed keeping the user in consideration at every step. Search engine displays the lowest fares from the major airline sites and offers industry leading travel planning tools.

Responsibilities:

- Used UML, C++,Java and NMDL, and Enterprise Architect and has written SQL queries
- Worked on DB2, implemented agile methodologies, developed Servlets and Java Server Pages(JSP).
- Provided dynamic content to the HTML pages developing JSPs and Servlets.
- Developed the web interfaces using JSP and business logic codes using Servlets at the back end of the system.
- Was also responsible for designing the front-end using HTML/Java Servlets.
- Developed PL/SQL queries to generate reports based on client requirements.
- Coded JDBC calls in the Servlets to access the Oracle database tables.
- Involved in developing the database tables to hold lender information and optimal storage of data.
- Enhanced the system according to the customer requirements.
- Designed the Entity beans and Session beans and deployed them on Web Logic Server.
- Developed interface to automatically forward quote requests to qualified customers using SMTP.
- Developed test cases to test the business logic and case scenarios.

Environment:
Core Java, Web Logic Server, SMTP, Servlets, Struts 1.2, PL/SQL, JSP, J2EE, XML, JMX, XSL, JDBC, HTML 5.1, UNIX,EJB, DB2, C++, NMDL, Oracle , UML, Python