**VENKATA SAINATH REDDY ARIKATLA**

**Email: Sainath.khs75@gmail.com PH: 469 846 8345**

**Data Engineer**

**Professional Summary**

* Around 8 years of work experience in Development and Implementations of Data Warehousing solutions.
* Experienced in Azure Data Factory and preparing CI/CD scripts for the deployment.
* Solid experience on building ETL ingestion flows using Azure Data Factory.
* Experience in building Azure Stream Analytics ingestion spec for data ingestion which helps users to get sub second results in Realtime.
* Experience in building ETL (Azure Data Bricks) data pipelines leveraging Spark, Spark SQL.
* Experience in building the Orchestration on Azure Data Factory for scheduling purposes.
* Experience working with Data warehouses like Teradata, Oracle, SAP.
* Experience on Implementation of Azure log analytics providing Platform as a service for SD-WAN firewall logs.
* Experience in building the data pipeline by leveraging the Azure Data Factory.
* Selecting appropriate low-cost driven AWS/Azure services to design and deploy an application based on given requirements.
* Expertise in working with databases like Azure SQL DB, Azure SQL DW.
* Solid programming experience working with Python.
* Strong expertise in Relational Data Base systems like Jenkins, MS SQL Server, MS Access, DB2 design and database development using SQL, PL/SQL, SQL PLUS, TOAD, SQL-LOADER. Highly proficient in writing, testing and implementation of triggers, stored procedures, functions, packages, Cursors using PL/SQL.
* Performed loads into Snowflake instance using snowflake connector in IICS for a separate project to support data analytics and insight use case for sales team.
* Hands-on experience with Snowflake utilities, Snows, Snow pipe, big data model techniques using Python.
* Experience with Snowflake multi-cluster warehouses and building snow pipes.
* Experience working in a cross-functional AGILE Scrum team.
* Happy to work with the team who are in middle of the road with some Big Data challenges for both on prem and cloud.
* Hands-on experience in Azure Analytics Services – Azure Data Lake Store (ADLS), Azure Data Lake Analytics (ADLA), Azure SQL DW, Azure Data Factory (ADF), Azure Data Bricks (ADB) etc.
* Orchestrated data integration pipelines in ADF using various Activities like Get Metadata, Lookup, For Each, Wait, Execute Pipeline, Set Variable, Filter, until, etc.
* Good knowledge of polybase external tables in SQL DW and involved in production support activities.
* Working knowledge on Azure services like Azure Data Factory, Azure Databricks and AWS services like EC2, S3, EMR, Redshift, Athena, Glue Meta store, Lambda etc.
* Expertise working on features like models and macros to create data pipelines using Data Build Tool (DBT) as transformation tool.
* Worked to extract data from various database sources like DB2, SQL, Main frame, CSV Files and Flat Files.
* Extensive experience in debugging, troubleshooting, monitoring and performance tuning using CA7, Control-M,
* Excellent understanding and knowledge of NOSQL databases like HBase and Mongo DB.
* Profound understanding of Partitions and Bucketing concepts in Hive and designed both Managed and External tables in Hive to optimize performance Jenkins, DBT Cloud.
* .
* Experience with ETL workflow Management tools like Apache Airflow and have significant experience in writing the python scripts to implement the workflow.
* Worked on AWS Data Pipeline to configure data loads from S3 into Redshift.
* Experience in big data technologies, such as Apache spark, Hadoop, or AWS EMR.
* Have good knowledge of Data frames, Data sets and RDD Concepts.
* Proficiency in SQL across several dialects (we commonly write MySQL, PostgreSQL, Redshift, SQL Server, and Oracle).
* Extensively worked with Teradata utilities Fast export, and Multi Load to export and load data to/from different source systems including flat files.
* Experienced in building Automation Regressing Scripts for validation of ETL process between multiple databases like Oracle, SQL Server, Hive, and Mongo DB using Python.
* Experience in analyzing, designing, developing, implementing, tuning, and supporting the Business Intelligence and Data Warehouse tools using Power BI.
* Good knowledge of detailed Data analysis and translates business requirements into logic.

data models and storytelling dashboards using Power BI**.**

* Sound knowledge on all Power BI scenarios like Calculations, Filters, Actions, Parameters, Groups, Sets, Graphs, Maps, DAX, Power Query, and blank query.
* Superior communication skills, strong decision making and organizational skills along with outstanding analytical and problem-solving skills to undertake challenging jobs.
* Experience in solving complex problems, strong analytical skills, and the ability to trace issues to the root source.
* Proficient in teaming up with the project sponsors, Analysts, QA, and Architecture teams to design and deliver outstanding business intelligence and data warehouse applications.

**Professional Experience**

**Capital One, Plano, Texas March 2022 to Present**

**Sr. Data Engineer**

**Responsibilities:**

* Create and maintain optimal data pipeline architecture for the Corporate Data Office.
* Assemble large, complex data sets that meet functional / non-functional business requirements.
* Identify, design, and implement internal process improvements: automating manual processes, optimizing data delivery, re-designing infrastructure for greater scalability, etc.
* designing and building new, cutting-edge Data Quality management product features
* Build the infrastructure required for optimal extraction, transformation, and loading of data from a wide variety of data sources.
* Build data analytics tools that utilize the data pipeline to provide actionable insights into data compliance, operational efficiency, and other key business performance metrics.
* Optimize the performance of corporate data platforms and services to ensure that they can handle large volumes of data and support real-time processing and analysis.
* Build out corporate data quality checks and develop processes for identifying and resolving data quality issues.
* Develop a corporate data library / metadata repository including best practice for development of data transformation specifications.
* Enable the engineering components to the rollout of the corporate data strategy framework, with a focus on improvement of data quality and the protection of sensitive data through related tools and architecture.
* Establishes measures to chart progress related to the completeness and quality of metadata for enterprise information; to support reduction of data redundancy and fragmentation and elimination of unnecessary movement of data; and to improve data quality.

**SBI General Insurance, Chennai, Tamil Nadu April 2020 to Dec 2021**

**Sr. Data Engineer**

**Responsibilities:**

* Created Linked Services for multiple source system (i.e.: Azure SQL Server, ADLS, BLOB, Rest API).
* Created Pipeline’s to extract data from on premises source systems to azure cloud data lake storage; Extensively worked on copy activities and implemented the copy behavior’s such as flatten hierarchy, preserve hierarchy and Merge hierarchy.
* Implemented Error Handling concept through copy activity.
* 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 activity; create dynamic pipeline to handle multiple source extracting to multiple targets; 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.
* Extensively worked on Azure Data Lake Analytics with the help of Azure Data bricks to implement SCD-1, SCD-2 approaches.
* Created Azure Stream Analytics Jobs to replication the real time data to load to Azure SQL Data warehouse.
* Implemented delta logic extractions for various sources with the help of control table; implemented the Data Frameworks to handle the deadlocks, recovery, logging the data of pipelines.
* Understand the latest features like (Azure DevOps, OMS, NSG Rules, etc..) introduced by Microsoft Azure and utilized for existing business applications.
* Worked on migration of data from On-prem SQL server to Cloud databases (Azure Synapse Analytics (DW) & Azure SQL DB).
* Deployed the codes to multiple envirsupportwith the help of CI/CD process and worked on code defects during the SIT and UAT testing and provide supports to data loads for testing; Implemented reusable components to reduce manual interventions.
* Developing Spark (Scala) notebooks to transform and partition the data and organize files in ADLS.
* Working on Azure Data bricks to run Spark-Python Notebooks through ADF pipelines.
* Using Data bricks utilities called widgets to pass parameters on run time from ADF to Data bricks.
* Created Triggers, PowerShell scripts and the parameter JSON files for the deployments.
* Worked with VSTS for the CI/CD Implementation.
* Reviewing individual work on ingesting data into azure data lake and providing feedback 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 Bricks, PolyBase, Azure DW, ADLS, Azure DevOps, BLOB, Azure SQL Server, Azure synapse.

**American International Group, Chennai, Tamil Nadu. March 2019 to April 2020**

**Data Engineer**

**Responsibilities:**

* Responsible for ingesting large volumes of user behavioral data and customer profile data to Analytics Data store.
* Developed custom multi-threaded Java based ingestion jobs as well as Sqoop jobs for ingesting from FTP servers and data warehouses.
* Developed Scala based Spark applications for performing data cleansing, event enrichment, data aggregation, de-normalization and data preparation needed for machine learning and reporting teams to consume.
* Worked on troubleshooting spark application to make them more error tolerant.
* Worked on fine-tuning spark applications to improve the overall processing time for the pipelines.
* Wrote Kafka producers to stream the data from external rest APIs to Kafka topics.
* Wrote Spark-Streaming applications to consume the data from Kafka topics and write the processed streams to HBase.
* Experienced in handling large datasets using Spark in Memory capabilities, using broadcasts variables in Spark, effective & efficient Joins, transformations, and other capabilities.
* Exercised the manipulation of raw data and ingesting it from AWS S3 into Snowflake.
* Responsible for ensuring data in Snowflake is updated timely and available for reporting and business analytics at the start of business daily.
* Experience working for EMR cluster in AWS cloud and working with S3, Redshift, Snowflake.
* Involved in creating Hive tables, loading, and analyzing data using hive scripts.
* Implemented Partitioning, Dynamic Partitions, Buckets in Hive.
* Good experience with continuous Integration of application using Bamboo.
* Used Reporting tools like Tableau to connect with Impala for generating daily reports of data.
* Collaborated with the infrastructure, network, database, application, and BA teams to ensure data quality and availability.
* Designed and documented operational problems by following standards and procedures using JIRA.

**Environment**: Snowflake, AWS EMR, Spark, Scala, Python, Hive, Sqoop, Oozie, Kafka, YARN, JIRA, S3, Redshift, Athena, Shell Scripting, GitHub, Maven.

**Oath-Map, Hyderabad, Telangana June 2018 to Dec 2019.**

**Java Developer**

**Responsibilities:**

* Involved in Business analysis and requirements gathering.
* Design and Implemented Food Court Web based Application.
* Played a key role in designing and implementing a Java based web application using Spring Boot, enabling real-time data visualization for customers.
* Utilized front end technologies like HTML, CSS, and JavaScript to create interactive and user-friendly interfaces.
* Collaborated with the QA team to develop comprehensive test cases and ensure high software quality.
* Implemented security measures and data encryption protocols to protect sensitive user information.
* Worked closely with product manager to gather requirements, estimate timeline, and prioritize tasks.
* Supported the team in documenting software functionalities and development processes.

**Technical Skills:**

* **Programming Languages:** Java, Python.
* **Web Technologies:** HTML, CSS, JavaScript**.**
* **Frame Works:** Spring, Spring Boot, Hibernate.
* **Databases:** MySQL, Oracle**.**